High Sensitivity Time-Varying Systems in Photonics and Electronics

Thesis by
Parham Porsandeh Khial

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended September 23rd, 2021

Parham Porsandeh Khial
ORCID: 0000-0002-3242-8541

To my lovely parents,
Sima and Hassan

ACKNOWLEDGEMENTS

First, I have to express my deepest gratitude to Professor Hajimiri, who gave me this fantastic opportunity to learn and explore the world of science and engineering at Caltech. He is not only a great adviser and role model but also an amazing life coach. He is always open to new ideas and taught me that failing and trying again is the key to success. I will always be grateful for all the experience and knowledge I gained, which were beyond all I could ever imagine.

Next I would like to thank my candidacy and defense committee members: Professor Changhuei Yang, Professor Andrei Faraon, Professor Sander Weinreb, and Professor Azita Emami.

I have to thank all CHIC lab members who are keeping the ambiance of this amazing lab creative and inspiring. Dr. Behrooz Abiri, who is not only a great friend but also a top-notch scientist and engineer. Next, I want to express my huge appreciation to my friend Dr. Aroutin Khachaturian who always inspired and supported me, especially through tough times we all face in graduate schools. He never hesitated to share his boundless kindness and unlimited encouragement with me. My good friend Dr. Brian Hong who we share lots of memories together. I have treasured all the time we spent together exploring the outside world and having truly meaningful conversations. Without a doubt, I need to mention Dr. Alex Pai, who is an icon in our lab and always reminds me that we can have fun and pursue our dreams at the same time. I always appreciate our small group, and our shared memories will remain in my mind forever. My first few years at Caltech brought me an incredible experience of working with Alexander White. It was a privilege to know and collaborate with such a bright and intelligent friend. Austin Fikes, whose curiosity, discipline, and practical view I admire, and Samir Nooshabadi, who I had a valuable experience collaborating with. I must acknowledge Professor Constantine Sideris, who I learned a lot from when I joined the CHIC lab. Next, I have to thank Dr. Reza Fatemi for all his insightful thoughts and comments which led to a few joined projects.

I thank all my dear friends at Caltech and other schools for their support and for the fascinating times we had together: Dr. Ehsan Abbasi, Dr. Fariborz Salehi, Dr. Pooya Vahidi, Dr. Peyman Ayoubi, Dr. Pouria Dadras, Dr. Omid Samiee, and my old and special friend Aryan Hashemi.

ABSTRACT

Integrated electronics and photonics have been revolutionizing our daily lives for decades. However, the demand for high-speed communications, low-latency networks, and high-performance optical and electrical sensors continues to grow. In order to keep up with this demand as well as be able to address upcoming and unknown challenges, we need to explore unconventional solutions. Moving away from existing systems and traditional architectures allows us to take a deeper look at these challenges and potentially come up with nontrivial answers. In this thesis, unconventional approaches to implementing high-performance optical and electrical sensors and systems are investigated. Among these unorthodox solutions are time-varying architectures which led to completely new devices, sensors with dramatically improved sensitivity.

By developing a time-varying method that we call reciprocal sensitivity enhancement, we demonstrated a nanophotonic optical gyroscope (NOG) for the first time. The efficacy of this method is borne out of its ability to improve the performance of optical gyroscopes by two orders of magnitude. This sensitivity-enhancement method filters out reciprocal imperfections and noise, thereby increasing the overall signal-to-noise ratio. Next, the same approach is used to boost the performance of resonance-based magnetic biosensors. By merging two biosensors and taking advantage of the frequency response of magnetic beads, time-division switching cancels out most of the correlated noise. This solution pushes the sensitivity of this sensor below parts-per-million (PPM) levels for long periods of time—a property which is desirable in many biosensing applications.

Additionally, an electrical scalable router that mitigates line-of-sight issues in next-generation wireless systems is introduced. This novel design does not require any shared timing reference to form a coherent array and uses a time-varying baseband to create a proper true-time delay. Next, we discuss how radiating elements in silicon-photonics platforms can be engineered to create a passive lensless camera. By applying a robust reconstruction algorithm, the captured image can be faithfully recovered. The same concept can be used in multi-mode nanophotonic antennas to alleviate the field-of-view (FOV)-aperture trade-off.

Finally, a hybrid photonic transmitter/receiver architecture, an electrical full-duplex transceiver with one nonreciprocal element, and a nested-ring optical modulator are presented.

PUBLISHED CONTENT AND CONTRIBUTIONS

Khial, P. P., White, A. D., & Hajimiri, A. (2018). Nanophotonic optical gyroscope with reciprocal sensitivity enhancement. Nature Photonics, 12(11), 671-675. Available: https://www.nature.com/articles/s41566-018-0266-5.
10.1038/s41566-018-0266-5.
I designed and measured the device and wrote the paper.

Khial, P. P., White, A. D., & Hajimiri, A. (2019, April). A Chip-Scale Nanophotonic Optical Gyroscope. In 2019 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL) (pp. 1-3). IEEE. Available: https://ieeexplore.ieee.org/abstract/document/8739715.
10.1109/ISISS.2019.8739715.
I designed and measured the device and wrote the paper.

Sideris, C., Khial, P. P., & Hajimiri, A. (2018). Design and implementation of reference-free drift-cancelling CMOS magnetic sensors for biosensing applications. IEEE Journal of Solid-State Circuits, 53(11), 3065-3075. Available: https://ieeexplore.ieee.org/abstract/document/8490237.
10.1109/JSSC.2018.2865480.
I contributed to the design, measurement, and the writing the paper.

Sideris, C., Khial, P. P., Ling, B., & Hajimiri, A. (2018, February). A 0.3 ppm dual-resonance transformer-based drift-cancelling reference-free magnetic sensor for biosensing applications. In 2018 IEEE International Solid-State Circuits Conference-(ISSCC) (pp. 190-192). IEEE. Available: https://ieeexplore.ieee.org/abstract/document/8310248.
10.1109/ISSCC.2018.8310248.
I contributed to the design, measurement, and the writing the paper.

Fikes, A., Khial, P. P., Nooshabadi, S., & Hajimiri, A. (2020). Programmable active mirror: A scalable decentralized router. IEEE Transactions on Microwave Theory and Techniques, 69(3), 1860-1874. Available: https://ieeexplore.ieee.org/abstract/document/9303454.
10.1109/TMTT.2020.3042516.
I contributed to the design, measurement, and the writing the paper.

White, A., Khial, P., Salehi, F., Hassibi, B., & Hajimiri, A. (2020). A silicon photonics computational lensless active-flat-optics imaging system. Scientific reports, 10(1),

1-9. Available: https://www.nature.com/articles/s41598-020-58027-1. 10.1038/s41598-020-58027-1.

I contributed to the design, measurement, and the writing the paper.

Fatemi, R., Khial, P. P., Khachaturian, A., & Hajimiri, A. (2020). Breaking FOV-aperture trade-off with multi-mode nano-photonic antennas. IEEE Journal of Selected Topics in Quantum Electronics, 27(1), 1-14. Available: https://ieeexplore.ieee.org/abstract/document/9206075. 10.1109/JSTQE.2020.3026966.

I contributed to the design and writing the paper.

Hosseini, R., Khachaturian, A., Cătuneanu, M., Khial, P. P., Fatemi, R., Hajimiri, A., & Jamshidi, K. (2018, May). Compact, high extinction ratio silicon Mach-Zehnder modulator with corrugated waveguides. In CLEO: Science and Innovations (pp. SM3B-6). Optical Society of America. Available: https://www.osapublishing.org/abstract.cfm?uri=CLEO_SI-2018-SM3B.6. 10.1364/CLEO_SI.2018.SM3B.6.

I contributed to the measurement.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Silicon has been a game changer in the past few decades. Soon after the invention of the transistor at Bell labs in 1947 [1]–[4], the era of microelectronics began and these micro-scale transistors have been revolutionizing our world until now. Shortly after the first implementation of the bipolar junction transistors (BJTs) which were aimed to replace the famous vacuum tubes, in 1960 D. Kahng and M. M. Atalla introduced and implemented the first field-effect metal-oxide-semiconductor (MOS) transistor [5]. This was the first fully planer integrated electronic device and came with an amazing feature: scaling!

In 1965, Gordon Moore nicely predicted the technology scaling of the MOS transistors [6] which is known as the Moore's law. He stated that the number of components per chip would be double every year–He later corrected his prediction to every two years. Moore's law would anticipate a significant reduction in cost and power consumption for microelectronic devices. Therefore, the use of MOS transistors were becoming economically more favorable [7]. It was not until the 1980s that two different types of MOS transistors (n-channel and p-channel) came together on a single wafer and the dominance of complementary MOS (CMOS) technology started. Surprisingly, the earliest research on integrated optics had been already begun in 1970s [8], [9]. However, these early works were mostly associated with ferroelectric materials like Lithium Niobate ($LiNbO_3$) and III-V semiconductors. In 1980s, the powerful sound of Silicon's revolution in electronics enticed researchers and they started dreaming about the potential marriage of electronics and optics on a single "superchip". Then Soref's work in 1980s officially mark the birth of Silicon photonics [10], [11].

By 1950, scientists and researchers started looking back at the light for telecommunication purposes noting that utilizing optical wavelengths could increase the capacity of the communication by orders of magnitude [12]. However, neither a proper coherent light source nor a low-loss propagation medium were available at the time. T. H. Maiman solved the first issue by the invention of the laser, and its first successful demonstration in 1960 [13]. Nevertheless, the loss of the optical fibers still remained as the utmost bottleneck until 1966. In the same year, Charles Kao suggested that loss of the optical fibers can significantly be lowered by remov-

ing impurities from silica glass used in the core of the fiber [14]. Frankly, this idea brought him the 2009 Nobel prize regarding his influential contribution to the transmission of light in fiber for optical communication. Despite Kao's idea, it was not until 1979 that a Japanese group was able to reduce the loss of the fibers to an acceptable level. They reported an astonishing number for the loss: 0.2 dB/km at 1.55 μm wavelength [15].

Silicon's boom in electronics and the demonstration of a very low-loss optical fiber at 1.55 μm–Silicon's bandgap would allow Silicon itself to be a waveguide medium at 1.55 μm–fed the newborn Silicon photonics process as an ideal integrated optical platform. Since then, integrated photonics and electronics, specially Silicon photonics and CMOS process, have seen an immense expansion in technology and application. In addition, recently monolithic platforms are getting more attention and maturity. With further developments of these platforms, more devices and innovative solutions will come into the light. Ultra fast optical communication, high-performance optical imaging and ranging systems, like LiDARs, and high-precision optical sensors are a few example of such systems.

Both Silicon photonics and integrated electronics have seen an immense advancement after the 1980s. CMOS technology has arrived at sub-10 *nm* FinFET transistors and unlocks more integration and capabilities [16], [17]. These tiny devices improve the performance of electronic components ([18]) and bring new systems and architectures into the light according to today's needs [19], [20]. The transition from the 4G wireless network to the fifth generation of the wireless communication system is one of the research hot-spots in the field [21]. Digital back-end, fully integrated receiver/transmitter at different frequency bands, and demonstrating large, multi-beam phased arrays are a few examples [22]. These new potential uses bring new challenges with them. For example, moving to higher frequencies to achieve higher communication capacity comes with the drawback of a shorter range of such frequencies. Intelligent reflective surfaces (IRS) and scalable routers are some of the proposed solutions to mitigated these issues [23], [24].

Silicon photonics which started after integrated electronics, has seen improvements in few different areas. For instance, passive structures (filters, couplers, etc), modulators, and detectors on component levels, and optical sensors and high-speed electro-optical interconnects on system levels. Coupling light from a single-mode fiber (SMF) to a Silicon waveguide is not a trivial matter. However, grating couplers are shown to be a viable solution. Their efficiency spans from 65% at 1,550 *nm* to 90% [25], [26]. In addition, optical modulators and detectors are the key com-

ponents in photonics integrated circuits (PICs). Recently, researchers have been trying to boost the bandwidth of optical modulators and detectors while keeping their footprint and power consumption low. Thermal modulators, plasma dispersion (carrier depletion and accumulation), and non-linear modulators are some examples of these efforts [27]–[29]. On the detector side, researchers are trying to improve responsivity and bandwidth; however, decrease the dark current of the photodetector. Generally speaking, there are few types of detectors Silicon detectors, Germanium detectors, Schottky detectors, and III-V material-based detectors [30]–[34].

1.1 Contributions and Thesis Outline

In this thesis, I focused on unconventional and non-trivial architectures in electronics and photonics. By utilizing less investigated and implemented solutions such as time-varying techniques, the performance of electronic and photonic devices can be boosted. Also, new devices and system architectures according to the state-of-the-art technology are feasible to be fabricated. Using these innovative techniques manifests itself in high-performance devices, completely new conceptual systems, or demonstration of seemingly impossible apparatuses.

In chapter 3, I describe challenges regarding the fabrication of optical gyroscopes on Silicon photonics platform. Then, to alleviate these challenges, reciprocal sensitivity enhancement is introduced. This method is a time-varying scheme which takes advantage from the nonreciprocal nature of the Sagnac effect and filters out all other sources of noise. By using this method, we successfully demonstrated the first nanophotonic optical gyroscope on Silicon photonics platform.

Chapter 4 discusses the implementation of the resonance-based biosensor detection scheme and its limitation due to different noise sources. Despite all the existing solutions, thermal drift and correlated noise between the sensors' core remain as barriers. By designing a double-resonance sensor and switching between two different oscillation modes, we used the frequency response of the magnetic beads. We were able to cancel most of the long and short-term drifts and boost the sensitivity to the sub-ppm realm.

In the fifth chapter, we tackle the inevitable challenge of the line of sight and low penetration depth for high-frequency communication systems. To address this imminent challenge, we introduce and design a new scalable router system with time-varying base-band architecture. This specific base-band design would allow for the true-time delay correction and truly enables the scalability for these types of arrays.

Chapters 6 and 7 discuss the use of radiating elements with different radiation patterns in photonics processes to achieve high field-of-view lensless cameras and break the known field-of-view aperture trade-off. Finally, in the last chapter I briefly mention three other works: hybrid transmitter/receiver optical phased array with high field-of-view and small footprint, A full-duplex transceiver architecture which uses only one non-reciprocal element, and a nested-ring interferometer MZI that reduced the required voltage swing drastically.

Chapter 2

SILICON PHOTONICS

2.1 Silicon Photonics Platform

In recent years, research on integrated photonics has expanded dramatically, and integrated optical solutions have made their way to different applications and fields. However, among different approaches to this relatively new area, Silicon photonics remains as one of the most promising solutions [35], [36]. Silicon photonics applications broaden dramatically in the past few decades, high-performance optical computing, high-speed optical interconnects, on-chip optical sensors, and ultrafast optical signal processing, to name but a few [37], [38].

Today integrated device manufacturers (IDMs) are attempting to use the maturity of the complementary metal-oxide semiconductor (CMOS) process to boost the performance and reliability. In the same way, CMOS allowed for more complex integration and high-density fabrication; later Silicon photonics improvements can potentially enable high-yield and high-performance photonics devices on a chip-scale, opening doors to new applications and fields [39].

The building blocks of the Silicon photonics platforms include Silicon waveguides, modulators, detectors (Germanium detectors), and radiators. Silicon-based lasers are also being investigated to be added to this list. Moreover, due to the relatively low refractive index of SiO_2 ($n = 1.45$ at $1,550nm$) and the bandgap property of the Silicon, optical waveguides with SiO_2 cladding and Silicon core can be fabricated on Silicon photonics platforms. This enables a convenient high-speed and low-loss optical interconnect on this chip-scale platform.

Fig 2.1 demonstrates a Silicon waveguide in a typical Silicon photonics process. This waveguide with $500nm$ width and $220nm$ height only supports one mode at $1550nm$. As it is shown in Fig 2.1, this waveguide is surrounded by SiO_2 cladding and the higher refractive index of the Silicon confines the mode inside the waveguide. This process is a Silicon on insulator (SOI) process which can later be incorporated into other SOI processes with transistors and active components. Modulators are one of the main components in Silicon photonics platform since they play critical roles in optical communication as well as other applications. An ideal optical modulator is the one with the high modulation speed, low power consumption, and

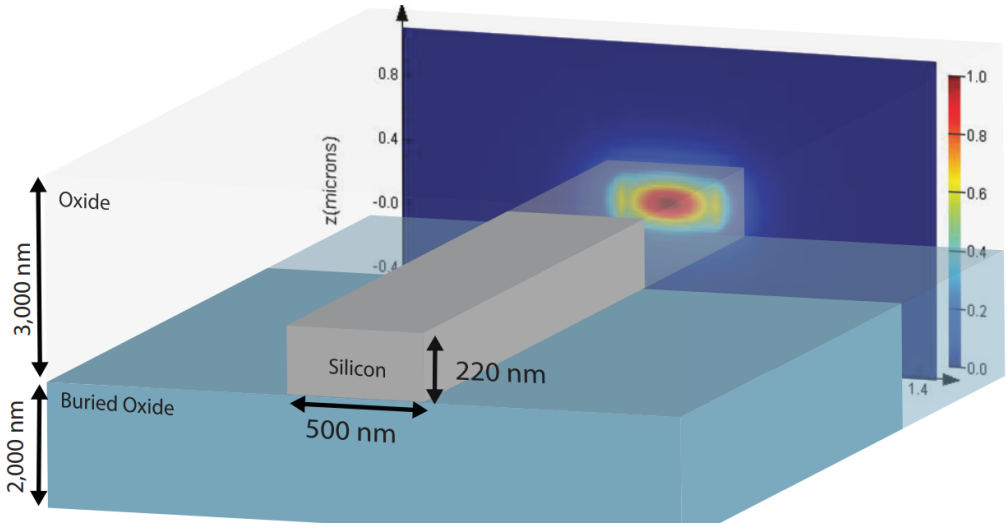

Figure 2.1: Cross section of a Silicon waveguide and its supported mode.

small footprint. Due to relatively high thermo-optic coefficient of the Silicon, thermal modulators are known to be one of the most easy-to-use modulators in Silicon photonics platform[40]. These modulators, however, suffer from low-bandwidth limitations and high-power consumption. Plasma dispersion modulators, which are the most common used modulators in Silicon photonics, operate based on changes in concentration of free carriers in Silicon [35].

Generally speaking, modulators can be used in two different configurations to create intensity modulation: 1- Mach-Zehnder interferometer (MZI) 2- Cavity based interferometer [41]. The latter one usually requires less voltage swing, but it demands tuning. Other modulation methods and architectures are proposed to push the bandwidth and integration density even further. Furthermore, modulators based on nonlinearity, for example, Silicon organic hybrid (SOH) modulators [42], [43], can provide the highest bandwidth as well as maintaining low power consumption and small footprint.

2.2 Future

The further development of Silicon photonics process will expand its application to more areas. Robust monolithic fabrication of electronics and photonics, potentially, can boost the performance of all the communication systems and also creates a suitable platform for better optical and electrical sensors. In addition, by optical quantum computing getting more attention, silicon photonics can host the future quantum computing realm on a chip-scale level.

Chapter 3

NANOPHOTONIC OPTICAL GYROSCOPE

3.1 Motivation

Gyroscopes are critical sensors for a variety of applications including high-precision navigation and consumer electronics [44], [45]. Recently, optical gyroscopes are getting more attention over micro-electromechanical (MEMS) counterparts due to their higher tolerance for overload and vibrations. Optical gyroscopes operate based on a relativistic phenomenon known as the Sagnac effect [46], [47]. When light travels around a closed path in a rotating frame, it accrues a phase proportional to its angular velocity multiplied by the area of the path it encloses, this will be discussed more in the next section. Optical gyroscopes can be categorized into two main groups: passive and active gyroscopes. Passive optical gyroscopes use an external laser source; however, active optical gyroscopes generate the beams inside the optical paths. The ladder either measures the beating frequency between clockwise (CW) and counter-clockwise (CCW) beams or resonance splitting in a passive optical cavity excited by two coherent beams. On the other hand, passive optical gyroscopes use phase interferometry to measure the phase shift between counter-propagating beams.

Technically speaking, in a typical interfrometry-based optical gyroscope (including fiber optic gyroscopes (FOGs)) architecture, by splitting light and sending the resultant beams in opposite directions around a fiber path, the recombined power at the output is proportional to the aforementioned Sagnac phase difference between the two beams of light. Due to the extremely high precision and stability of this type of measurement ([46]), it is highly desirable to miniaturize FOGs to the scales of their MEMS counterparts in order to extend the application space of FOGs and nanophotonics optical gyroscopes (NOGs) to other types of inertial systems including consumer electronics [48]–[50].

However, such miniaturization has several associated challenges. Most notably, since the enclosed area of the path dictates the magnitude of the output, the signal strength at the scale of integrated circuits can be so small that it is completely obscured by different sources of noise. This noise is due predominantly to thermal fluctuations. As silicon has a relatively high thermo-optic coefficient, thermal

changes can modulate the phase of the light and introduce a phase shift indistinguishable from that produced by the Sagnac effect. In addition to thermal effects, the difficulty of fabricating low-loss light paths in standard processes at this scale makes lossy waveguides an inevitability. Not only does loss reduce signal strength, but it also leads to back-scattering. If there is only one signal path, as is the case in most FOGs, this back-scattering and back-reflection–in case of using an external cavity– couples to the reverse direction of propagation. Thermal fluctuations can then change the phase of this back-reflection, further deteriorating the signal-to-noise ratio (SNR).

3.2 The Sagnac Effect

As mentioned above, optical gyroscopes determine angular velocity by measuring the relativistic Sagnac effect. When a beam of coherent light is split into two paths that traverse a closed circle (Fig 3.1), they meet with a certain phase relationship at the output. However, when the reference frame of the gyroscope is rotating, the effective path lengths experienced by the two beams change.

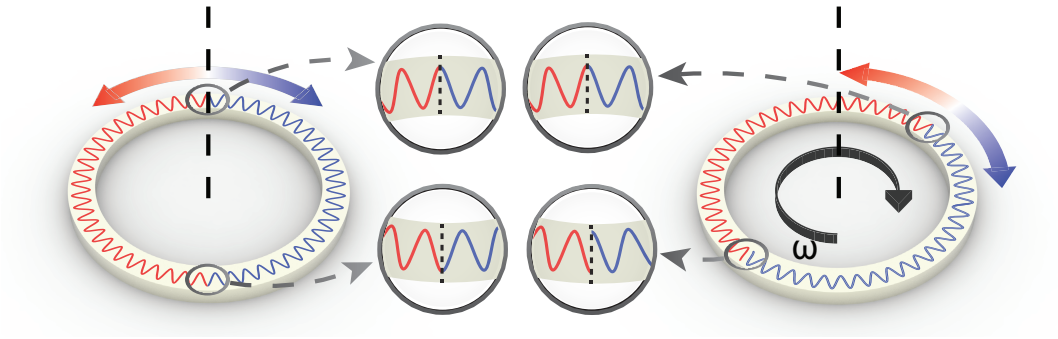

Figure 3.1: The Sagnac effect.

This causes them to experience an additional time delay with respect to one another that is proportional to the rate of rotation of the frame [51], [52]:

$$\Delta\varphi_{Sagnac} = \frac{8\pi}{\lambda c} m\mathbf{A}.\mathbf{\Omega} \tag{3.1}$$

where $m\mathbf{A}$ is the area vector with magnitude equal to the area of the closed path multiplied by m (the number of turns), $\mathbf{\Omega}$ is the angular velocity vector, λ is the wavelength, and c is the speed of light.

To measure the rates of rotation using the Sagnac effect, the input light beam in

a typical optical gyroscope is split into two paths: the signal path, which is often coiled to accumulate the Sagnac phase shift, and the reference path. These two paths combine inside a phase interferometer whose output is proportional to the phase difference between the input signals [45]:

$$Output = |Sagnac\ path + reference\ path|^2 \propto \Delta\varphi_{Sagnac}. \tag{3.2}$$

As the system size is reduced (for example, in a NOG), the performance of this topology suffers from the small magnitude of the phase shift, which implies a low SNR. This is because the Sagnac effect is proportional to the area enclosed by the optical path. This low SNR translates to a stronger random walk in the output signal of the gyro, resulting in a poor Allan variance, a standard measure for characterizing the stability of gyroscopes [53]. As we discussed earlier, among all imperfections and sources of noises, the inevitable thermal drift is more pronounced because of the relatively large thermo-optic coefficient of silicon at room temperature: $dn/dT = 1.86 \times 10^{-4} K^{-1}$, where n is the index of silicon and T is the temperature in Kelvin. Thus, even small temperature variations within different parts of the optical path can cause a considerable deviation in the measured rate of rotation.

However, these temperature fluctuations in silicon nanophotonics have a bandwidth of several kilohertz [54], are much less pronounced above a megahertz, and thus can be treated as constant over timescales of microseconds. This is one of the reasons that in Silicon Photonics platform, we can design a relatively simple, effective, and low-speed thermal modulators. In light of this observation, it is possible to significantly lower thermal fluctuations and other imperfections such as fabrication mismatch by intentionally alternating the polarity of the Sagnac phase shift at a fast rate, as shown conceptually in Fig 3.2, which will be discussed in the next section.

3.3 Reciprocal Sensitivity Enhancement

Since the Sagnac effect is inherently non-reciprocal, it is feasible to distinguish it from sources of noise that act reciprocally—namely, thermal fluctuations, system mismatch, and back-scattering. As our Sagnac path is a passive linear network in an isotropic media, the scattering parameters from the input to the output and from the output to the input are the same. Thus, if the direction of the Sagnac path is reversed, the effect of all mismatch due to thermal fluctuations and back reflection is the same. However, due to the nonreciprocity of the Sagnac effect, this feed reversal flips the sign of the signal generated by the Sagnac effect—i.e., the gyroscope is

virtually flipped.

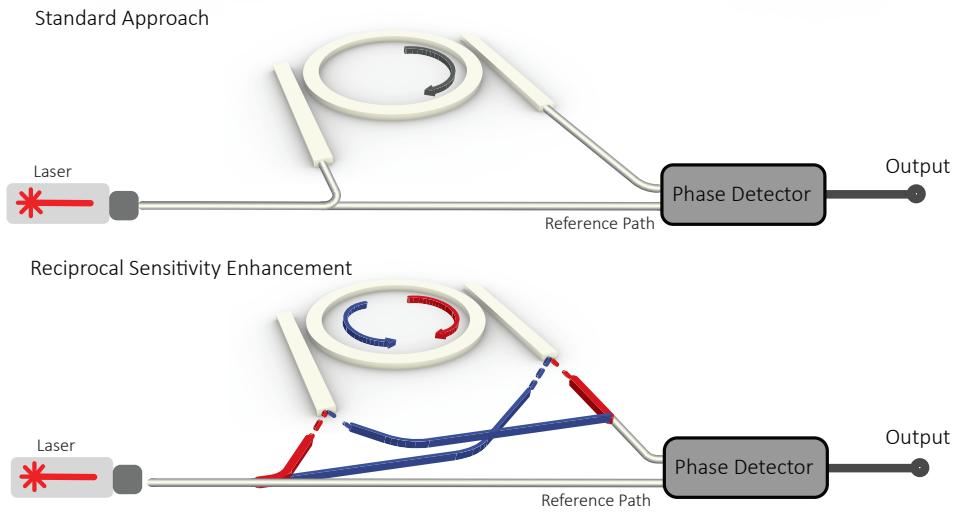

Figure 3.2: Standard approach vs reciprocal sensitivity enhancement.

We call this new approach: reciprocal sensitivity enhancement. Technically speaking, in a passive network made of isotropic elements (lossy or lossless), switching the input and output ports does not change the observed response from input to output. This is because the scattering matrix is symmetric, namely, $S_{ij} = S_{ji}$ (note that S_{ij} is a complex number representing the phase and amplitude of the outgoing wave). By alternating the paths (the blue and red paths, as shown in Fig. 3.2), two measurements are taken—one of S_{ij} and the other of S_{ji}—whereby the desired signal has its polarity flipped, but undesirable components such as thermally induced fluctuations and mismatch are common and can therefore be attenuated.

Reciprocal sensitivity enhancement also reduces the undesirable effects of reflections. Although we know that reciprocal networks satisfy $S_{ij} = S_{ji}$ for all i and j, in general $S_{ii} \neq S_{jj}$ for $i \neq j$. This causes mismatch between reflected light from different directions, which poses a challenge for using continuous laser sources in existing architectures of single-loop gyroscopes that utilize both clockwise and anticlockwise beams [44].

In a typical interferometry-based gyroscope architecture, light is sent in both clockwise and anticlockwise directions, and the phase shift induced by the Sagnac effect is measured at the output [55], [56]. The sensitivity of these types of gyroscope is limited by the Kerr effect, unequal thermal fluctuations between clockwise and anticlockwise propagating light, and Rayleigh scattering [57]. The latter two are usually

dominant. However, it has been demonstrated that the effect of back-scattering light can be reduced by using a non-coherent laser source. Another approach to alleviate the issue of scattering is to use a low-loss waveguide or a high quality factor external cavity, which usually demands a different waveguide core and clad material. In addition, some other methods have been proposed to enhance the Sagnac effect and to boost the overall SNR [58], [59]. Time-division switching can be used to reduce the back-scattering effect in fibre-optic gyroscopes [60]. Among all these methods, however, reciprocal sensitivity enhancement offers an architecture that greatly reduces all sources of noise. This is due to the temporal separation between the clockwise and anticlockwise propagating beams in each path as well as the cancellation of thermal fluctuations using high-frequency optical switching. Critically, this method is more tolerant to the propagation loss of the medium.
The signal can also be amplified without increasing the size of the interferometer by using ring resonators to replace the circular path [61]–[63].

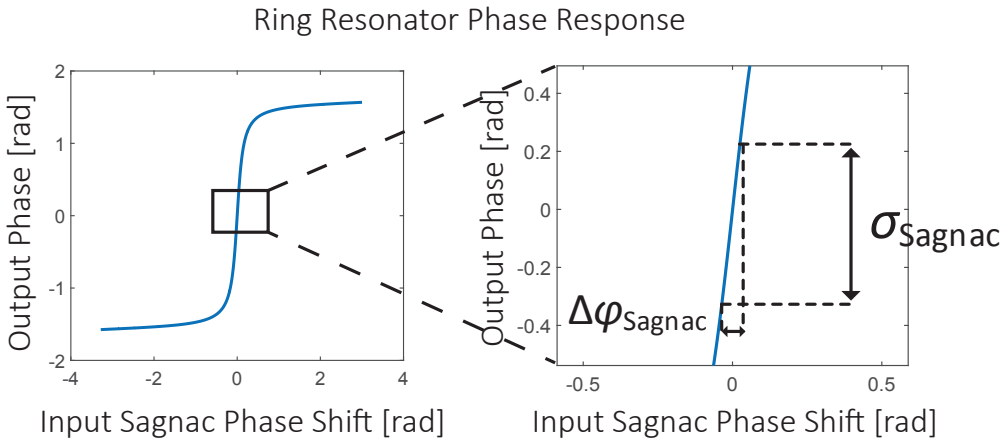

Figure 3.3: Ring resonator response.

The phase response of ring resonators amplifies the observed phase difference for a given Sagnac phase shift (Fig 3.3). Thus, to the first order, the phase at the output of the ring resonator, σ_{Sagnac}, is given by:

$$\sigma_{Sagnac} \approx S \times \Delta\varphi_{Sagnac} = \frac{Q\lambda}{\pi n L}\Delta\varphi_{Sagnac} \tag{3.3}$$

where Q is the quality factor of the ring resonator and L is the circumference of the ring. Note that S can be interpreted as the effective number of times that light travels around the ring before exiting. Another significant challenge with the

standard approach shown in Fig 3.2 is the large dn/dT in silicon that results in substantial temperature-dependent phase variation between the signal and reference paths. These two path lengths can be made equal by using two identical rings—in lieu of a reference path— which carry light in opposite directions (Fig 3.4). This can significantly alleviate such thermal fluctuations that are substantially different between the ring and the reference path (Fig 3.2). Thus, a system with two rings is only sensitive to spatial gradients in the temperature between the rings. In addition, because the two rings are in close proximity, the resonance peaks will move in a correlated fashion.

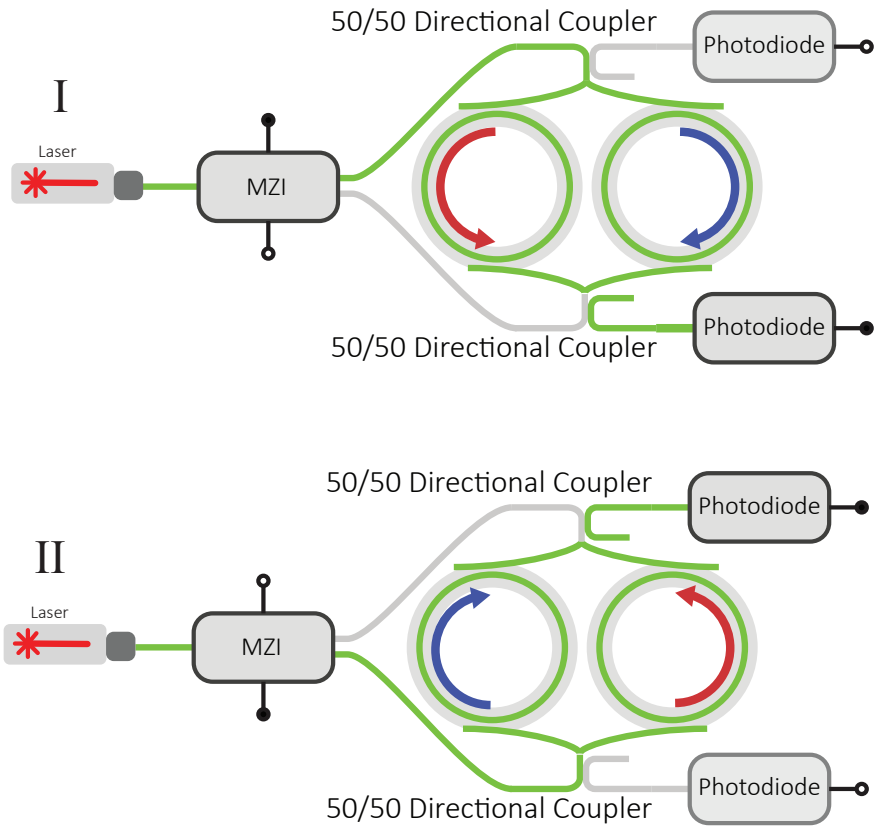

Figure 3.4: Reciprocal sensitivity enhancement with two ring resonators.

We computed the responses of the two rings. Note that, to have a measurable signal, we need to introduce a phase shift $\varphi_{applied}$ as an offset phase shift at the output of ring 1. Therefore, the output electric field phasors of ring 1 ($\mathbf{E_1}$) and ring 2 ($\mathbf{E_2}$)

can be written as:

$$\begin{aligned}\mathbf{E_1} &= E_1 e^{j(\varphi_1+\Delta\theta_{thermal}(t)+\varphi_{applied}+\sigma_{Sagnac})} \\ \mathbf{E_2} &= E_2 e^{j(\varphi_2-\sigma_{Sagnac})}\end{aligned} \tag{3.4}$$

where E_1 and E_2 are the magnitudes of the output fields, φ_1 and φ_2 are the phase responses of each ring resonator, and $\Delta\theta_{thermal}(t)$ is the phase shift difference between two rings due to temperature gradient fluctuations.
Next, the output signal, which we will denote $Y(t)$, is proportional to the square magnitude of the sum of these two electric fields:

$$\begin{aligned}Y(t) &= |\mathbf{E_1} + \mathbf{E_2}|^2 \\ &= E_1^2 + E_2^2 + 2E_1E_2cos[\varphi_1 - \varphi_2 + \varphi_{applied} \\ &+ \Delta\theta_{thermal}(t) + 2\sigma_{Sagnac}].\end{aligned} \tag{3.5}$$

Now by alternating the direction that rings are fed, we arrive at (Fig 3.5):

$$Y(t) = \begin{cases} +X(t)|\Delta\varphi_{Sagnac}| + U(t) & 0 < t < \frac{T}{2} \\ -X(t)|\Delta\varphi_{Sagnac}| + U(t) & \frac{T}{2} < t < T \end{cases} \tag{3.6}$$

where T is the alternating (that is, switching) period and we defined the signal amplitude, $X(t)$, and the undesired signal, $U(t)$, as

$$\begin{aligned}X(t) &= 4SE_1E_2sin[\beta(t)] \\ U(t) &= E_1^2 + E_2^2 + 2E_1E_2cos[\beta(t)] \\ \beta(t) &:= \varphi_1 - \varphi_2 + \varphi_{applied} + \Delta\theta_{thermal}(t).\end{aligned} \tag{3.7}$$

As is shown in equation above, $U(t)$ and $X(t)$ are, in general, stochastic processes because they depend on thermal fluctuations.
Moreover, if we alternate directions on a faster timescale than the timescale of the fluctuations (for example, 10MHz), the fluctuations remain mostly correlated. Therefore, these correlated fluctuations are mostly cancelled by adding the two outputs together and subtracting the common-mode signal (Fig 3.5).

3.4 Implementation

To demonstrate a NOG with Reciprocal Sensitivity Enhancement, a 1mm × 2mm NOG prototype was implemented on a silicon photonics platform. To guide the 1,550nm light around the device, 500nm by 220nm waveguides are fabricated on a

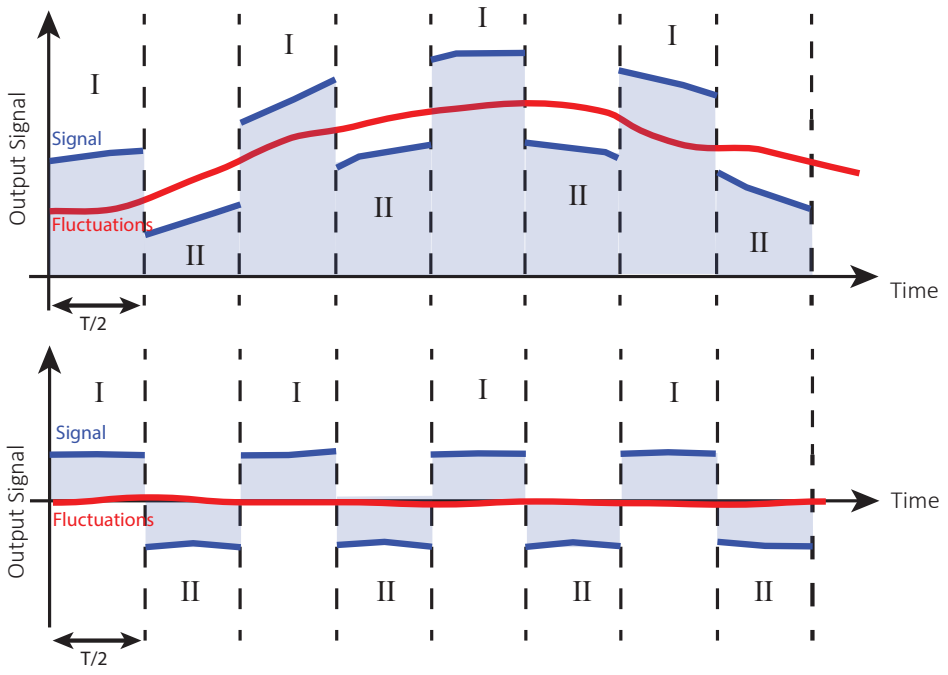

Figure 3.5: Signal and fluctuations in reciprocal sensitivity enhancement.

silicon-on insulator process with a $2\mu m$ bottom oxide and a 3µm passivation oxide. The Sagnac accumulation path consists of two identical ring resonators with radii of $500\mu m$. Using resonators allows for an increase in the effective area of the Sagnac path as the average photon circles more than once before coupling out. These ring resonators were tuned using thermal resistors positioned in $90nm$ thick silicon slabs encasing the rings. The two form the requisite two-port Sagnac network for Reciprocal Sensitivity Enhancemen. The rings are coupled to waveguides and attached via a y-plitter on either side (Fig 3.6).

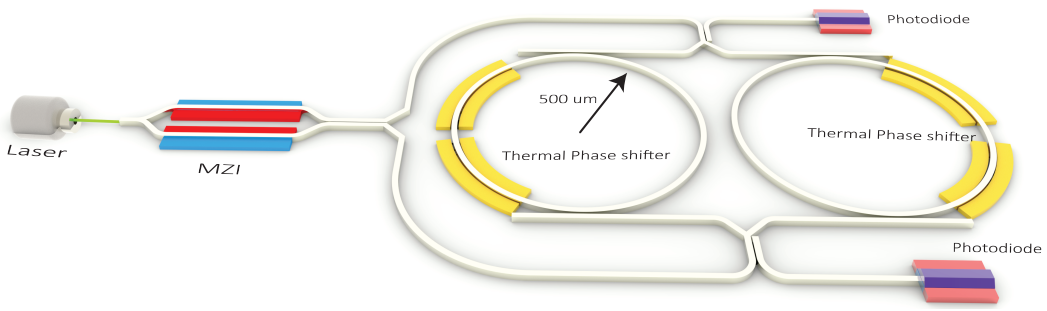

Figure 3.6: Designed NOG

To rapidly switch the feed direction, a Mach-Zehnder Interferometer with differential PIN diode phase shifters and a 50/50 directional coupler is used. The PIN diodes change the effective index of the silicon through carrier injection, thereby shifting the phase between the MZI paths and directing the power to one of the feeds (Fig 3.7). The signal is read from photodiodes coupled to the signal paths in the direction opposite the feed.

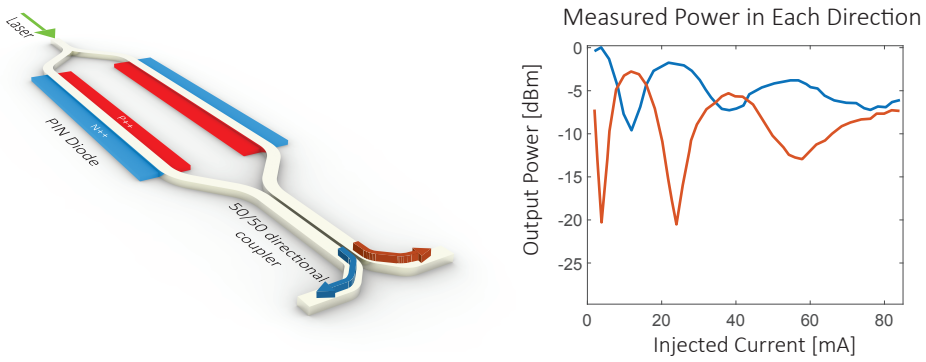

Figure 3.7: Optical switching using PIN diodes.

Fabricated device is shown in Fig 3.8 and a diagram of the measurement setup is shown in Fig 3.9. To account for any mismatch in power between the two feed

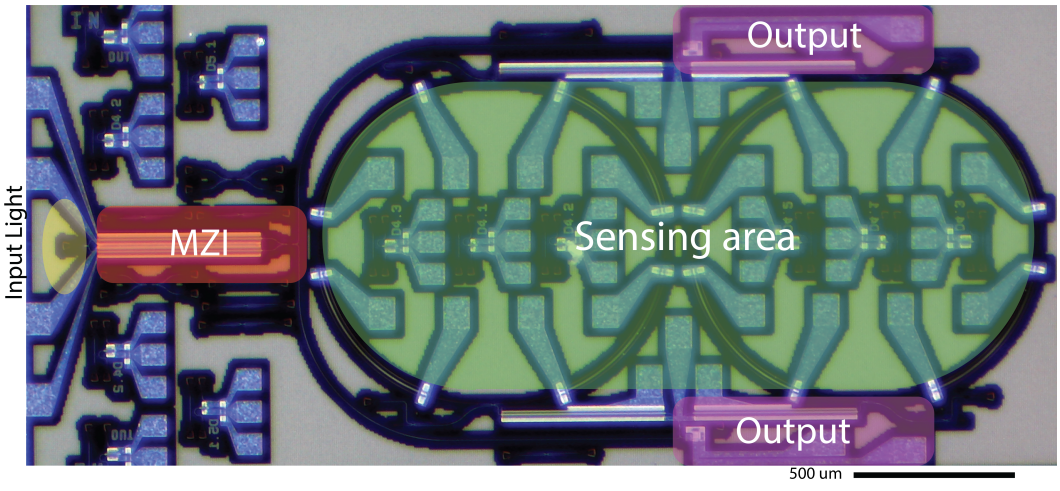

Figure 3.8: Fabricated device.

directions, the two gyroscope output signals are first amplified by transimpedance

amplifiers (TIAs), then they are fed into variable gain amplifiers (VGAs) tuned so that their sum contains only a DC component. Perturbation of this summation can then be heavily amplified to a readable voltage. As perturbations due to noise will be at lower frequency than the Sagnac signal (which will appear at the switching frequency), the fluctuations can be filtered out, leaving a direct measurement of the rate of rotation.

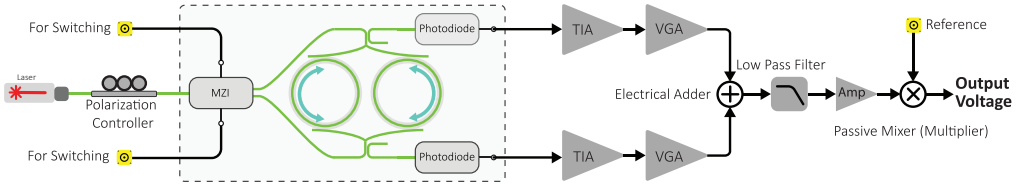

Figure 3.9: Measurement setup.

3.5 Measurement

We measured the sensitivity of the gyroscope by measuring the output voltage for different angular velocities (Fig 3.10). To observe the efficiency of our proposed

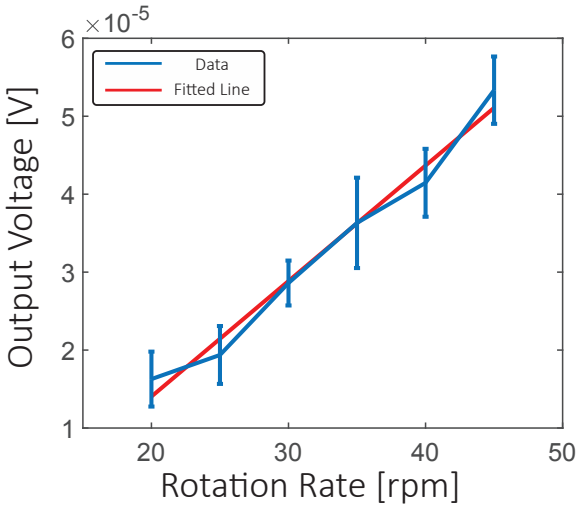

Figure 3.10: Output voltage vs rate of rotation.

approach, reciprocal sensitivity enhancement, we measured the Allan deviation to calculate the bias instability (BIS) and the angle random walk (ARW) for two separate scenarios (Fig 3.11): one where the switching frequency was set to 20 kHz, where thermal fluctuations have a more prevalent effect on the output, and one at

$1/T$=10 MHz, where thermal fluctuations are negligible. The results are shown in Fig 3.11. For the former setting ($1/T$=20 kHz), we measured a BIS of 105 r.p.m. and an ARW of $97,800°/\sqrt{hours}$, whereas the latter scenario ($1/T$=10MHz) resulted in BIS=1 r.p.m. and ARW = $650°/\sqrt{hours}$.

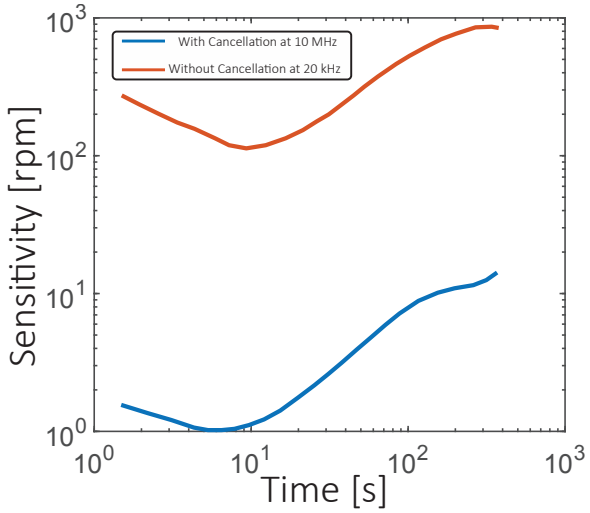

Figure 3.11: Allan deviation curve at 10 MHz and 20 kHz switching frequencies.

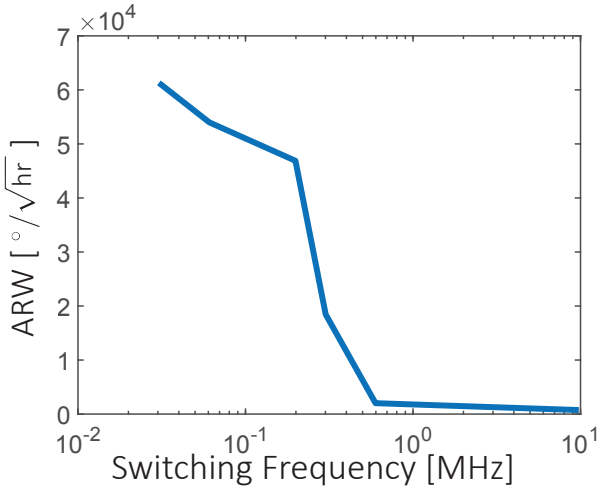

Figure 3.12: ARW vs switching frequency.

The sensitivity at 10 MHz improves by a factor of roughly 100 compared to that at 20 kHz. Moreover, measurements of ARW at different switching frequencies show

that the threshold frequency for reciprocal sensitivity enhancement is on the order of 100 kHz (Fig 3.12) which is around the bandwidth of thermal fluctuations on silicon nanophotonics platform.

3.6 Design Procedure

The fundamental limitations of the performance of a NOG include thermal fluctuations, backscattering, and process variation. However, another critical factor to SNR in all optical gyroscope systems is the shot noise, which is proportional to photodiode power and it is a fundamental quantum limit:

$$Total\ SNR = \frac{\Delta\varphi_{Sagnac}}{N_{shot} + N_{thermal} + N_{backscattering}}. \tag{3.8}$$

By utilizing the reciprocal sensitivity enhancement, we can mitigate thermal and backscattering effects, and we are left with:

$$Total\ SNR = \frac{\Delta\varphi_{Sagnac}}{N_{shot}} \propto \frac{Field\ Power}{\sqrt{FieldPower}} = \sqrt{FieldPower}. \tag{3.9}$$

Based on equation 3.9 in order to increase the overall SNR, having higher input power is desirable. On the other hand, the output power of the gyroscope can be written as:

$$Output\ Signal \propto P_0 e^{-\alpha L} \times L^2 \tag{3.10}$$

where L is the length of the optical path, α is the loss of the waveguide, and P_0 is the input power. The first term is due to the optical loss of the waveguide and the second term is related to the Sagnac effect, which is proportional to the area of the sensor. Fig 3.13 shows the output power for different waveguide losses. This analysis indicates that there is an optimum optical length that maximizes the output signal. Nevertheless, processes with lower waveguides loss and higher quality-factor resonators could potentially result in a more sensitive optical gyroscope, given the fact that there is an optimum size based on process specs.

3.7 Conclusion

In this work, we have introduced reciprocal sensitivity enhancement, which can be used in fibre-optic gyroscopes of all platforms and scales to increase their sensitivity significantly. Based on this new approach, we designed an all-integrated optical gyroscope which occupies only 2 mm^2 and detects the smallest recorded phase shift

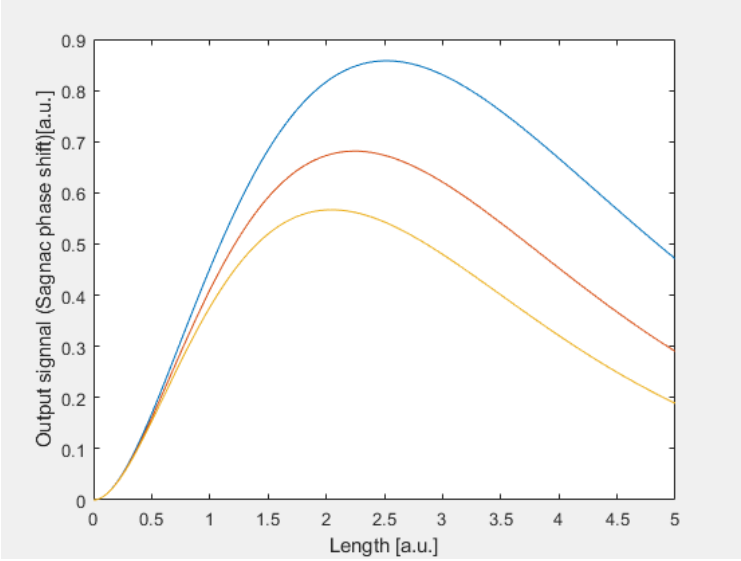

Figure 3.13: Output power of the optical gyroscope vs different waveguide loss.

(3 nrad) of all miniaturized approaches implemented in Silicon nanophotonics so far. This work shows the feasibility of an integrated Silicon nanophotonics platform for developing optical gyroscopes and paves the way towards having high-precision integrated optical gyroscope therefore, broadening their applications to other fields and areas including consumer electronics.

Chapter 4

MAGNETIC BIOSENSING

4.1 Introduction

Biology is one of the most rapidly growing fields which deals with the study of living things. Quantitative discoveries of the biology is mostly dominated by some of its sub-fields including bioengineering. Bioengineering tries to use provided tools by various fields of engineering to introduce more traditional and quantitative solutions to existing problems [1].

In the area of bioengineering, one area which is getting more attention due to its practicality is biosensing. Although biossensing is a general term, it usually refers to any methods and techniques for detecting biological molecules. These techniques will allow scientists to detect the presence of specific molecule or strand nucleotides like DNAs and RNAs in a solution full of many contaminants [64], [65]. Biosensing can be extended to detect more comlicated structures and molecules and even can be used to find new ones. Molecular biosensing has different applications including genetics and biomedical. Nowdays the medical workflow of medical diagnostics requires patient sample collection at specialized facilities, followed by analysis at centralized labs. The whole process is expensive and extremely slow. Sometimes it can take a few weeks for the results to become available. Fortunately, there has been a significant push toward point-of-care (PoC)-based medical diagnostics in the past decade, that brings the equipment into the doctor's office [66], [67]. While the decentralized PoC-based approach can be more time efficient, it is still expensive and needs some level of infrastructure, which makes it inadequate for deployment in underdeveloped countries [68]. This brings the attention to the need for point-of-use (PoU) diagnostics, which would introduce inexpensive, but highly specific and sensitive portable imagers into the homes of patients and to the field, facilitating affordable therapeutics for all.

This space is where biosensing meets electrical engineering, specifically integrated CMOS devices. CMOS chip fabrication processes have been around for a few decades. Barring a steep initial startup cost, thanks to the maturity of the fabrication process, modern CMOS chips generally cost on the order of a few cents per chip.

[1] This work was done in collaboration with Professor Constantine Sideris.

CMOS technology has become so mature that billions of transistors can be fit on a chip of few square millimeters of area, all of which function properly. This mind-blowing high yield allows for enormous, reliable computational power in a form factor smaller than a penny. Finding an effective way to design a CMOS chip which can be used as a biosensor would therefore solve both the cost and space issues.
Integrated biosensors using magnetic label as tags offer a viable approach to PoU diagnosis. This method grants high sensitivity despite being cheaper than their optical counterparts, namely fluorescent-based biomolecular detection devices. Part of this is because of utilizing CMOS process however, magnetic-based approach eliminates all expensive optics required by other optical methods. Moreover, recent studies indicate higher sensitivity as well as higher specificity of magnetic-based detection compared to fluorescent-based alternatives [69], [70].
Compared to the label-free or dielectric detection schemes, the magnetic-based approach offers higher sensitivity due to the lack of strong magnetic properties in biological samples. In conclusion, despite being cheaper and capable of being produced in large volumes, CMOS biosensor devices have lower detection limits. This includes methods with high-quality factor resonators such as MEMS-based sensors and optical ring-resonator sensors [71], [72].

4.2 Magnetic Biosensing

As mentioned above, magnetic biosensors use magnetic labels as tags. Furthermore, the transducer of magnetic-based sensors can be implemented completely on CMOS process with no post-processing beyond printing a surface chemistry of probe biomolecules. The introduced transducer here is an on-chip inductor. Also, the beads to be sensed are made of a paramagnetic material which becomes temporarily polarized in the presence of an external magnetic field. Passing any sort of current (whether it be static or alternating) through a spiral coil inductor generates a magnetic field due to Ampère's Law (Fig 4.1). This field magnetizes all beads sitting over the inductor surface, which causes an increase in the total magnetic energy stored in the system. Since effective inductance is proportional to the magnetic flux through the inductor surface, we can deduce that the total inductance of the coil will also increase in the presence of these beads. This change in magnetic energy can be calculated [65]:

$$\Delta E_m \approx \frac{\chi_{eff}}{2\mu_0} |\mathbf{B}|^2 V_p. \tag{4.1}$$

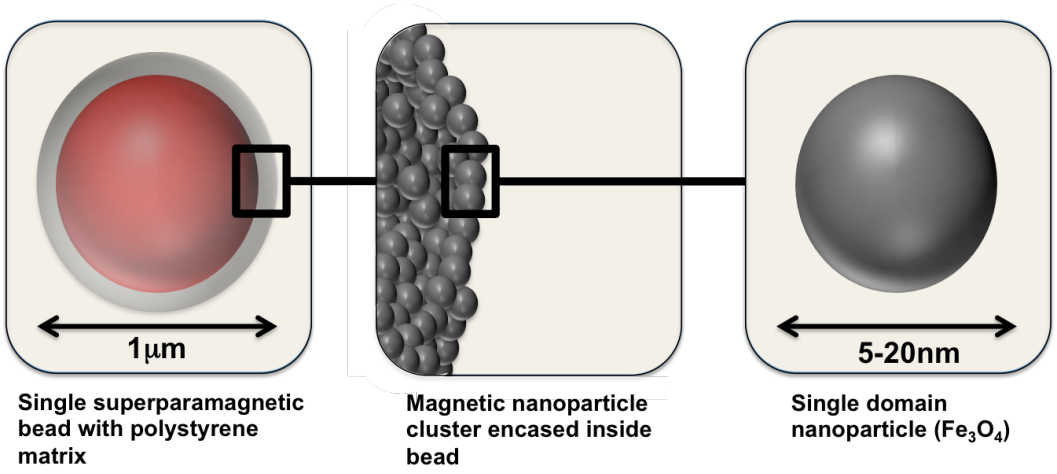

Figure 4.1: Construction of the magnetic bead from iron oxide nanoparticles encased in a non-magnetic polystyrene matrix. The nanoparticles are single-domain and paramagnetic.

Here, χ_{eff} is the susceptibility of the magnetic beads, **B** is the magnetic field vector, and V_p is the volume of the particle. On the other hand, we know:

$$\Delta E_m = \frac{1}{2} I^2 \Delta L \tag{4.2}$$

where I is the current going through the inductor and ΔL is the change in the value of the inductor. From the above equations we can conclude:

$$\Delta L \approx \frac{\chi_{eff}}{I^2 \mu_0} |\mathbf{B}|^2 V_p. \tag{4.3}$$

As we can see from Equation 4.3, the change in the value of the inductor is proportional to the magnetic susceptibility, which can be either positive or negative, as well as the magnitude of the magnetic field and the size of the particle–larger particles are expected to affect the value of the inductor more than smaller ones.
It is important to realize that due to the paramagnetic nature of these nanoparticles; the magnetic bead exhibits a rich frequency response in its magnetic susceptibility (χ) due to the finite magnetic dipole rotation time in the presence of an external, polarizing magnetic field (Fig 4.2). In fact, there exists a resonance frequency f_{res} where the real part of the magnetic susceptibility becomes 0 and the magnetic beads appear completely transparent to the sensing surface [73], [74]. This resonance occurs when the finite magnetic dipole rotation time of the magnetic dipoles leads them to rotate in time-quadrature with the external polarizing field. Furthermore, the magnetic beads again become transparent to the sensing inductor at very high

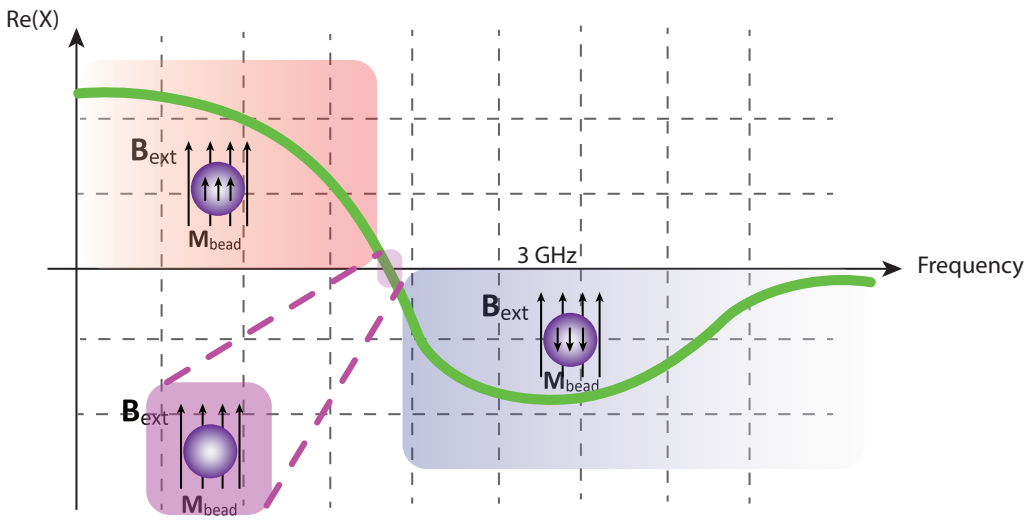

Figure 4.2: Magnetic susceptibility vs frequency.

frequencies due to the polarizing field going through a full cycle faster than the magnetic dipoles can react. A simplified magnetic biosensing approach is also presented in Fig 4.3.

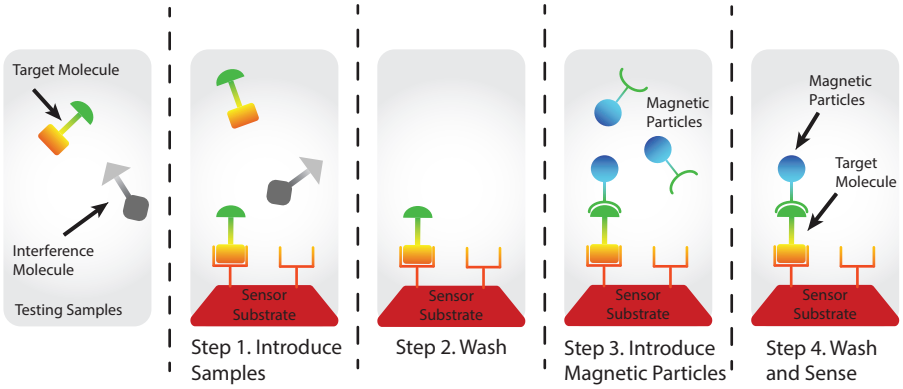

Figure 4.3: Magnetic biosensing scheme.

While a number of different magnetic detection approaches have been demonstrated in an integrated sense [75], [76], here we focus on resonance-shift-based sensors. Resonance-shift-based sensors are the simplest, yet perhaps most effective, approach for measuring magnetic particles in the context of a CMOS imager. Resonance-shift

sensors offer great sensitivity and can be implemented in standard, bulk CMOS processes without any post-process modifications, unlike alternative approaches such as Hall effect and GMR [77]. Furthermore, resonance-shift sensors can be readily extended to multiplex imagers, enabling the detection of multiple targets concurrently over a single-sensor surface [78]. Here, we will discuss the main working principle of the sensor (Fig 4.4). The core of the sensor consists of

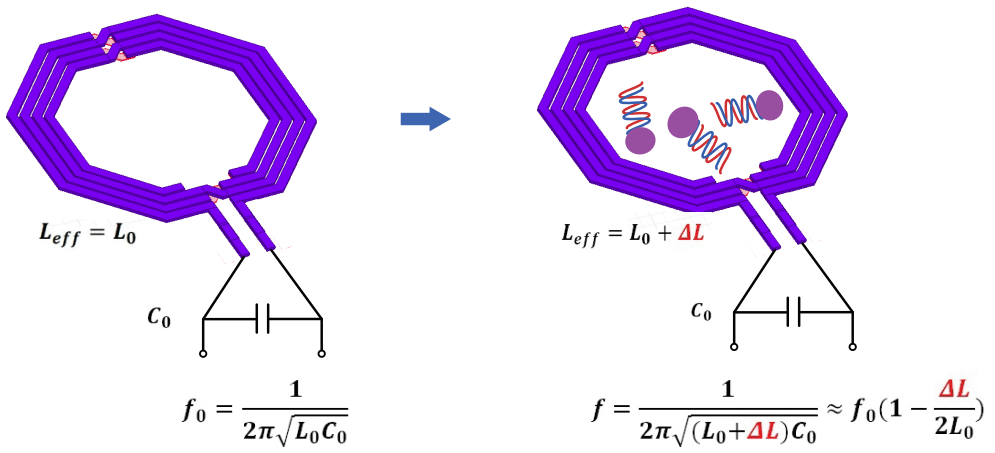

Figure 4.4: Resonance-shift-based CMOS sensor

an integrated LC resonant tank, with the center area of the inductor representing the functional sensing region. The current in the inductor generates a magnetic field which magnetizes the target paramagnetic particles within its vicinity. The magnetized particles, in turn, contribute their own magnetization field which adds a contribution to the total magnetic flux through the inductor, perturbing the total magnetic energy in the region.

The magnetic particles (contributing a non-unity μ_r) act as an effective magnetic core for the inductor, leading to an increase in inductance. Defining the nominal tank inductance and capacitance as L_0 and C_0, respectively, and the change in inductance due to magnetic content as ΔL, the new tank resonance frequency is

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(L_0 + \Delta L)C_0}} \approx f_0(1 - \frac{\Delta L}{2L_0}). \tag{4.4}$$

Thus, the change in resonance frequency over the center frequency is $(\Delta f / f_0) \approx -(\Delta L / 2L_0)$. While a direct tank-impedance measurement, e.g., using a Wheatstone

bridge could be used to quantify this resonance-shift [79]. Due to low on-chip inductor quality factor, this approach can be quite challenging and suffers from poor SNR. A simple alternative is to incorporate the LC tank as the resonator for a differential cross-coupled CMOS oscillator and measure the oscillation frequency in real-time, which corresponds to the dynamic resonance frequency of the tank. Moreover, other promising architectures including coupled inductive bridge, is proposed [80].
Despite the fact that the frequency-counting approach allows for very simple sensor circuitry and readout, it suffers from a significant drawback: CMOS-based oscillators typically have poor phase-noise performance and considerable random frequency fluctuations due to electrical noise and thermal fluctuations. Furthermore, CMOS devices contribute significant flicker noise which when combined with thermal variations contribute to slowly varying long-term drift in the oscillator frequency which cannot be removed with averaging. Fig 4.5(a) shows an idealized, noise-free sensor response due to magnetic beads and Fig 4.5(b) shows the same scenario, in a more realistic setting, with sensor transient noise superimposed. It is apparent that it can be quite difficult to detect small signals in the presence of significant noise.

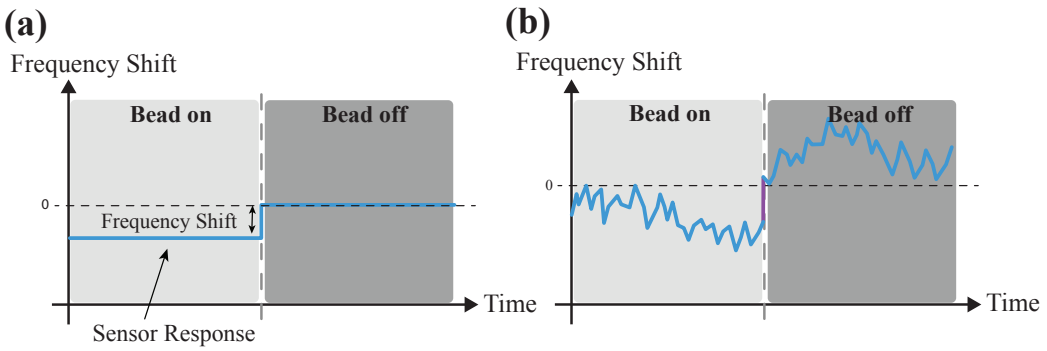

Figure 4.5: Sensor's response (a) Idealized, noise-free resonance-shift sensor time-domain response due to a magnetic bead and after removal. (b) Resonance-shift sensor response due to the same scenario in the presence of electrical and thermal noise fluctuations

The noise of the system can be separated into two categories: electrical noise sources from the circuit components and active devices and long-term instability due to component variations due to thermal variations. Assuming that one can measure the oscillator system for an arbitrary duration and perform time averaging, the dominant electrical noise sources become the flicker current noise from the active devices, as the white-noise sources which are dominant at higher frequencies

can be adequately averaged out [80]. Due to their slowly varying nature, the flicker noise sources, like the thermal variations, contribute to the long-term instability of the sensor system. Realistic biodetection experiments may require minute-to-hour-long timescales for completion, and thus stabilizing the oscillator over the duration of the whole experiment is crucial. Few methods are proposed in order to overcome this issues. One of the most effective one is called replica cell approach.

The most common approach for compensating long-term sensor drift is using a replica reference (Fig 4.6), which entails replicating a full sensor cell, along with all of its oscillator circuitry and buffer logic, and placing it in close proximity to the original sensor [81]. The close spatial proximity of the replica to the sensor allows tracking of some of the sensor drift due to process and thermal variations; however, the replica must be kept clean from magnetic content at all times and is, therefore, unusable for sensing, leading to an expensive doubling of chip area required per sensor. The active devices of the replica are completely uncorrelated with those of the sensor, leading to a 3 dB increase in noise power. Furthermore, process variation (usually resulting in differences in tank capacitance or threshold voltage of actives) and thermal gradients across the chip can be quite steep, rendering the replica incapable of fully cancelling the long-term variations of the system. This, combined with the additional uncorrelated slow-varying flicker noise from the reference, renders the replica approach unsuitable for achieving a low enough noise floor, which is stable over a useful time interval for realistic biodetection experiments, required for sub-parts-per-million (PPM) levels of signal detection. An approach which could measure the noise of the sensor circuit separately from

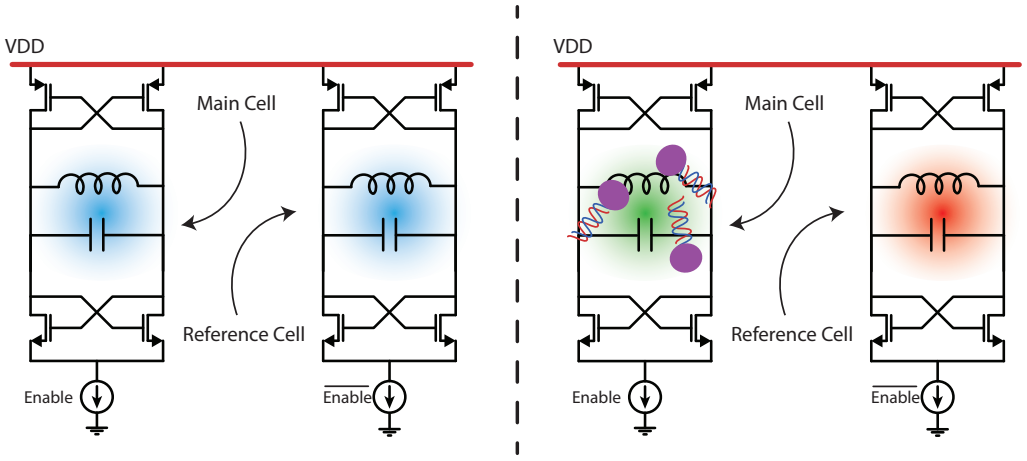

Figure 4.6: Replica reference cell technique.

the response due to the magnetic signal would obviate the need for a reference cell. This would simultaneously solve both the area doubling and noise increasing issues, and if the drift of the actual sensor itself were being monitored rather than that of a replica, the approach would further be immune to steep thermal gradients across the chip surface. We develop a technique capable of achieving splitting of signal and noise measurement in the context of a single-sensor circuit.

4.3 Drift Cancellation Method

We can develop a technique to measure signal and noise separately by taking advantage of the fact that the sensor will be utilized to measure specific magnetic beads suitable for use as tags for biodetection experiments. Indeed, as shown in Fig 4.1, a typical magnetic bead consists of a non-magnetic external polystyrene matrix with many superparamagnetic, single-domain iron oxide nanoparticles encased inside. As it was mentioned before, these magnetic beads have a rich frequency behaviour (Fig 4.2). Therefore, we can make use of the existence of this frequency behaviour and increase the overall SNR.

Indeed, suppose an oscillator is designed which can be switched to oscillate at either of two possible frequencies, which will hereafter be referred to as f_{LO} and f_{HI}. Further assume that the oscillator circuit is designed in a manner such that the slowly varying, colored noise is highly correlated between the two frequencies. Then, f_{LO} and f_{HI} can be designed such that f_{HI} happens to be at a magnetic sensing null (i.e., either at the resonance frequency of the magnetic beads or high enough in frequency where susceptibility becomes 0 again), and f_{LO} is low enough in frequency and responds significantly due to variations in magnetic content. In this scheme, f_{LO} would be affected by both magnetic signal and sensor electrical noise and thermal variations, while f_{HI} would only vary due to the noise and thermal effects since it is not sensitive to the magnetic beads. Thus, the sensing oscillator could be switched back and forth between f_{LO} and f_{HI}, counting each for a fixed time interval, measuring the drift response in f_{HI}, and using that information to compensate f_{LO}.

One potential approach for achieving this dual-frequency switching capability is by incorporating a switched capacitor which can be switched in for the system to oscillate at f_{LO} and left out for operation at f_{HI}. Assuming the sensor inductance is L, the fixed tank capacitance C, and the switched capacitance C_{SW}, this would

result (assuming a lossless LC tank) in

$$\begin{aligned} \omega_{LO} &= \frac{1}{\sqrt{(L+\Delta L)(C+\Delta C+C_{SW}+\Delta C_{SW})}} \\ \omega_{HI} &= \frac{1}{\sqrt{L(C+\Delta C)}} \end{aligned} \tag{4.5}$$

where ΔL represents variations in inductance due to magnetic beads and ΔC and ΔC_{sw} represent fluctuations in tank and switched-in capacitance due to noise and thermal drift. Notice here that we have made the assumption that $\omega_{HI}(f_{HI})$ is designed to be at a point where magnetic beads are transparent. Furthermore, we assume that the primary source of drift in the system is through the capacitance (i.e., the inductance only varies due to magnetic content), which may not be the case in practice. Despite these simplifying assumptions, since $C_{sw} \gg C$ (e.g., for $\omega_{HI} = 3\omega_{LO}$,$C_{sw} = 8C$), we expect that: $C_{sw} \gg C$. However, ω_{LO} does not see C_{sw} at all since the switch is OFF, meaning that the higher frequency does not contain any information about the drift in C_{sw}, which is expected to be the major contributor of drift at ω_{LO}. Furthermore, the switch transistors must be very large, having widths on the order of hundreds of micrometers, to achieve low ON-resistance for reducing degradation of tank Q, which leads to excess noise contribution in the ON-state (ω_{LO}) and significant parasitic capacitance while OFF (ω_{HI}) which will also drift. Thus, even though ω_{HI} indeed is "transparent" to magnetic beads (L), it is unsuitable for measuring the majority of the system noise and drift, and the switched capacitor approach is inadequate for realizing the aforementioned drift-cancellation technique.

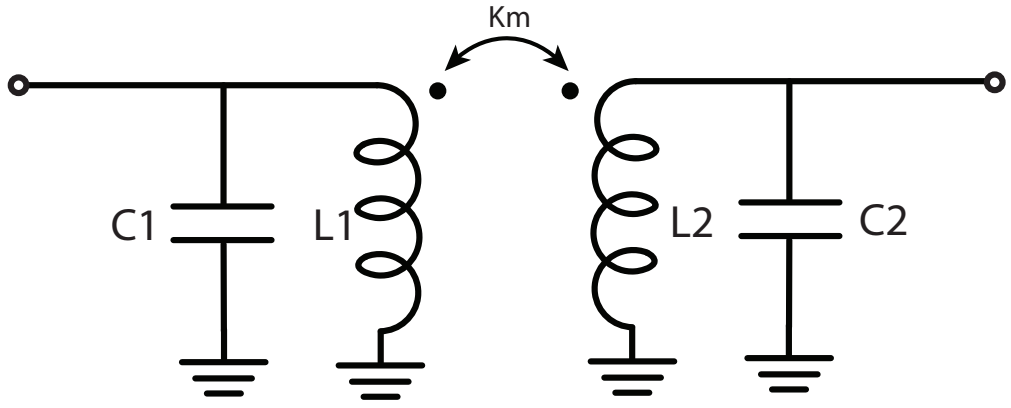

Figure 4.7: Fourth-order transformer-based resonant circuit.

An ideal topology for maximizing drift compensation not only requires that one frequency is transparent to the beads, but also that both frequencies vary in the same way due to all of the circuit components which may be sensitive to dynamic variations. In order to achieve this, we consider the fourth-order transformer-based resonant circuit shown in Fig 4.7. The circuit, in general, admits two resonant frequencies concurrently (assuming zero transformer series resistance) [82]:

$$f_{1,2}^2 = \frac{1 + (\frac{L_2C_2}{L_1C_1}) \pm \sqrt{1 + (\frac{L_2C_2}{L_1C_1})^2 + (\frac{L_2C_2}{L_1C_1})(4k_m^2 - 2)}}{8\pi^2 L_2C_2(1 - k_m^2)}. \tag{4.6}$$

Notice that both frequencies depend on all of the component parameters. Furthermore, the ratio f_1^2/f_2^2 is given by

$$\frac{f_1^2}{f_2^2} = \frac{1 + X + \sqrt{1 + X^2 + X(4k_m^2 - 2)}}{1 + X - \sqrt{1 + X^2 + X(4k_m^2 - 2)}} = \frac{g_1}{g_2} \tag{4.7}$$

where $X = (L_2C_2)/(L_1C_1)$ and $g_{1,2} = 1 + X \pm \sqrt{1 + X + X(4k_m^2 - 2)}$.
The ratio of the low and high frequencies only depends on the ratio of component parameters (X) and the transformer coupling factor km. It is well-known that with careful commoncentroid layouts, ratios between passives can be made considerably insensitive to process and thermal variations, unlike the absolute component values [83]. Furthermore, k_m exhibits almost no dependence due to temperature or process variations. Thus, assuming a proper layout, X is expected to vary much less than the individual passive inductor and capacitor values. Thus, for X constant, and L_1/L_2 being perturbed by L at the low frequency due to a magnetic signal

$$\frac{f_1^2}{f_2^2} = 1 + \frac{\Delta L}{L} \tag{4.8}$$

where L is either L_1 or L_2 and ΔL is the change in L. This shows that the ratio only depends on changes in the magnetic signal and not due to absolute variations of any of the reactive components due to noise, as long as X remains constant. This would be the case if the transformer is laid out so that the magnetic signal affects the inductances of the primary and secondary coils by the same proportion, leading to any changes in inductance to cancel out in X. In reality, X may indeed vary due to mismatch or thermal and process variation gradients, and thus, we can seek to further increase the robustness of the design by finding the nominal value of X for which $f_R^2 = f_1^2/f_2^2$ is least sensitive to deviations ΔX

$$\frac{d}{dX} f_R^2 = 0 \rightarrow g_1' g_2 = g_1 g_2'. \tag{4.9}$$

Therefore, in order for 4.9 to hold, the condition: $(X-1)(k_m^2-1)=0$ must be satisfied. Since $k_m < 1$, (df_R/dX) becomes exactly 0 only when $X = 1$.
This implies that the ratio of the two frequencies can be made insensitive to the first-order perturbations in the ratio of component values $((L_2C_2)/(L_1C_1))$, by choosing $L_1C_1 = L_2C_2$. Since we wish to have magnetic beads affect both sides of the transformer identically, we must have $L = L_1 = L_2$ which fixes $C = C_1 = C_2$ to result in $X = 1$, which satisfies the zero sensitivity to the first-order perturbations in X condition. In the remainder of this paper, we will, therefore, assume that both the inductances and capacitances on each side of the transformer are matched. When $X = 1$, the ratio f_R simplifies even further

$$f_R = \sqrt{\frac{1+k_m}{1-k_m}}. \tag{4.10}$$

Since the fourth-order circuit admits two parallel resonances concurrently, we must be careful when using it as the resonator for differential cross-coupled oscillator as either frequency can be a valid mode of oscillation.
These two possible oscillation frequencies can also be seen and analyzed by using T equivalent circuit of the tank in two different scenarios: 1- common-mode oscillation 2- differential-mode oscillation across the tank. Fig 4.8 demonstrates the equivalent inductances in both cases. Notice that the capacitor value is $2C$. Therefor, we will have

$$f_R = \frac{1/\sqrt{CL(1-k_m)}}{1/\sqrt{CL(1+k_m)}} = \sqrt{\frac{1+k_m}{1-k_m}} \tag{4.11}$$

which is equal to 4.10.
Assuming $L = L_1 = L_2$, $C = C_1 = C_2$, and if the circuit in Fig 4.7 is driven on both sides by two current sources (I_x) of the same amplitude and phase $(+I_x, +I_x)$, the higher frequency resonance disappears in the frequency response of the impedance, resulting in an impedance function

$$Z_0(j\omega) = \frac{\frac{1}{2}j\omega(1+k)L}{1+(j\omega)^2(1+k)LC} \tag{4.12}$$

which resonates at $\omega_0 = 1/\sqrt{(1+k_m)LC}$.
Conversely, if the transformer circuit is driven on both sides by current sources of the same amplitude but which are antiphase $(+I_x, -I_x)$, the lower frequency resonance disappears, resulting in the impedance function

$$Z_{180}(j\omega) = \frac{\frac{1}{2}j\omega(1-k)L}{1+(j\omega)^2(1-k)LC} \tag{4.13}$$

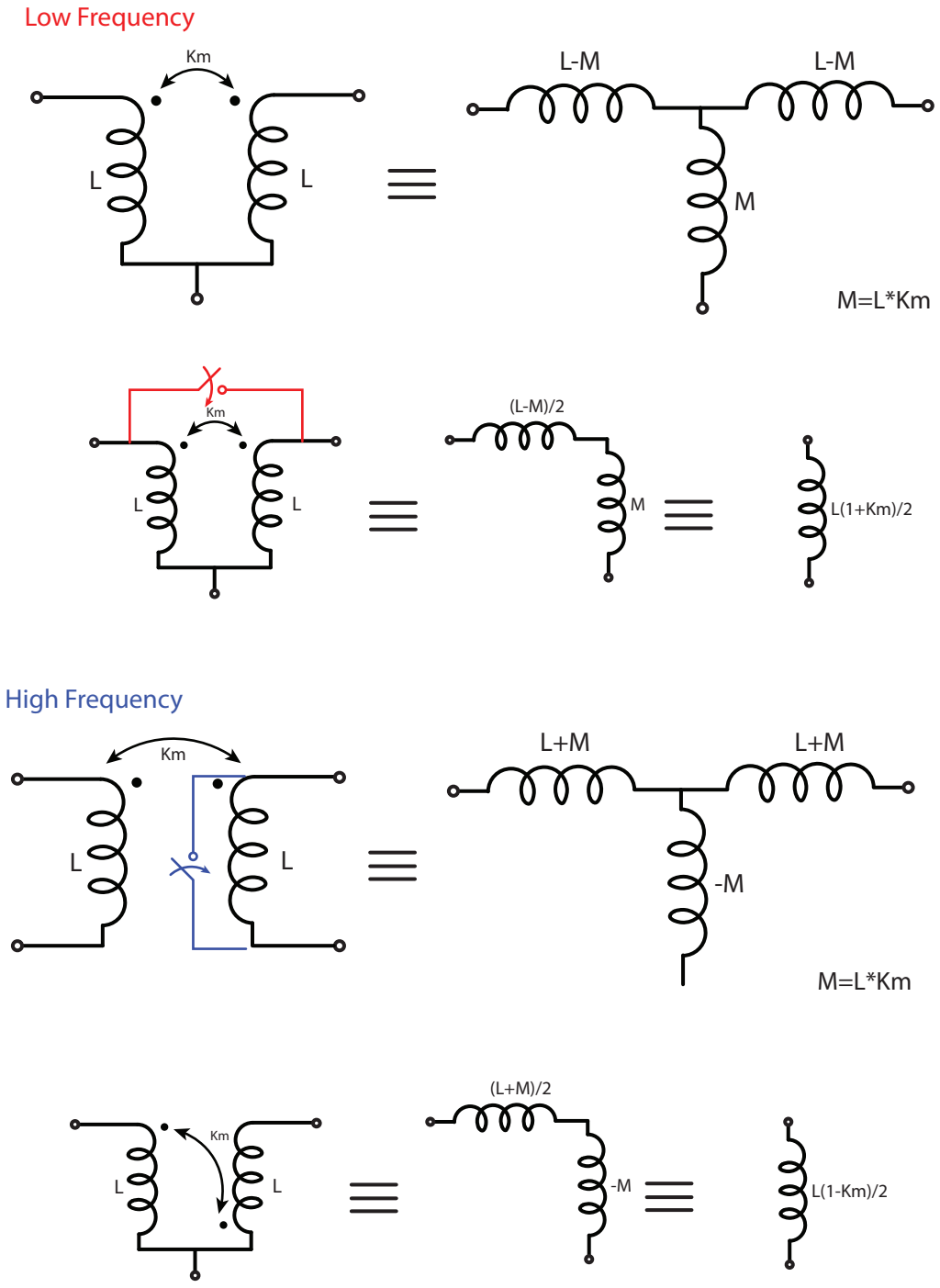

Figure 4.8: Common-mode and Differential-mode equivalent circuits

which resonates at $\omega_{180} = 1/\sqrt{(1-k_m)LC}$. These phase and anti-phase conditions imply that the circuit can only oscillate at ω_{LO} when the node voltages on the primary and secondary sides of the transformer are in-phase and are capable of

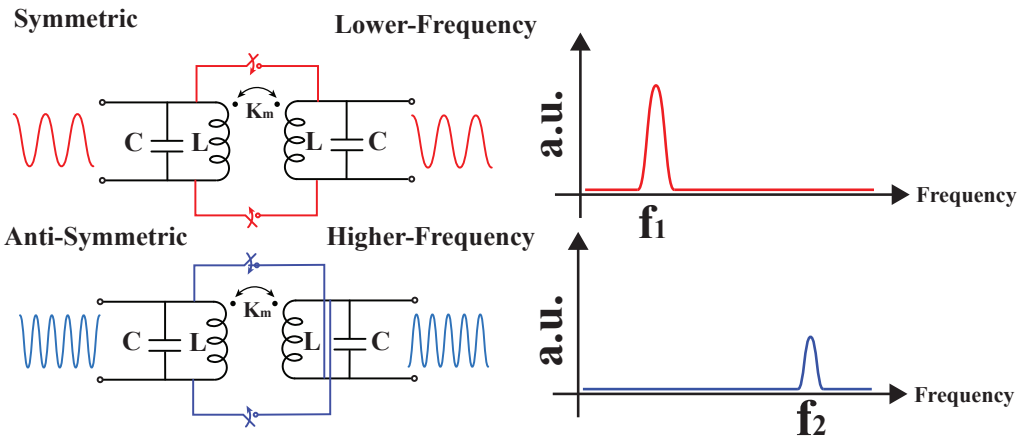

Figure 4.9: Switches enforce (a) in-phase or (b) anti-phase voltage boundary conditions between primary and secondary sides, allowing selection of desired oscillation frequency by "shorting out" undesired mode.

oscillating at ω_{HI} only when the node voltages on both sides are 180° out of phase with respect to each other.

Adding two switches, as shown in Fig 4.9(a), enforcing in-phase boundary conditions, thus forces the system to collapse to an effective second-order LC tank and oscillate at ω_{LO}, and likewise, adding two switches in a cross-coupled fashion (Fig 4.9(b)) which enforce anti-phase boundary conditions results in oscillation at ω_{HI}. It is important to note that the switches behave very differently in this circuit than those in the switched-capacitor oscillator discussed earlier. Indeed, the switches serve to "short out" the undesired frequency of operation and appear completely transparent to the desired, actively oscillating frequency. This is due to the fact that the switches which are ON have both sides attached to the "common mode" of the nodes between the primary and secondary sides of the transformer and thus the node voltages across both sides of each switch are always identical. This implies that the switches contribute no dynamic current and do not inject any extra noise into the system. This further implies that the switch does not impact the tank quality factor at all, unlike the switches in the switched-capacitor design which enable the active mode and can significantly $de-Q$ the LC tank due to their finite ON-resistance.

Up until this point, we have derived everything assuming ideal inductors without series resistance (infinite Q). In reality, on-chip inductors tend to have relatively low quality factors (usually 10–15 for symmetric spiral inductors in the 1–4-GHz frequency range), and therefore, this must be taken into consideration. Assuming that the loss of each side of the transformer is modeled by a non-zero series resis-

tance, $R = R_1 = R_2$, it can be shown that for switches enforcing in-phase operation the tank impedance frequency response becomes

$$Z_{R0}(j\omega) = \frac{\frac{1}{2}(R + j\omega(1+k)L)}{1 + (j\omega)^2(1+k)LC + j\omega RC} \tag{4.14}$$

which resonates at $\omega_{LO} = 1/\sqrt{(1+k_m)LC}\sqrt{1 - \frac{R^2C}{(1+k_m)L}}$. For the anti-phase scenario we have

$$Z_{R180}(j\omega) = \frac{\frac{1}{2}(R + j\omega(1-k)L)}{1 + (j\omega)^2(1-k)LC + j\omega RC} \tag{4.15}$$

which resonates at $\omega_{LO} = 1/\sqrt{(1-k_m)LC}\sqrt{1 - \frac{R^2C}{(1-k_m)L}}$. Now the ratio of the two frequencies becomes

$$f_R^2 = \frac{1+k_m}{1-k_m} \times \left(\frac{1 - \frac{\gamma^2 LC}{1-k}}{1 - \frac{\gamma^2 LC}{1+k}}\right) \tag{4.16}$$

where $\gamma = (R/L)$ is a proportionality constant of the series resistance to the inductance. Unfortunately, for non-zero series resistance of the inductors, the ratio of the two frequencies is not independent of the absolute values of the inductance and capacitance anymore. To minimize the sensitivity of f_R to drift in L and C, we must thus optimize the design for $\gamma^2 LC \ll 1 \rightarrow Q \gg 1$ since $\gamma = \omega/Q$ and $\omega \approx 1/\sqrt{LC}$. In order to increase the Q of the inductor, the number of turns, therefore its value, should be maximized ($L \propto N^2$ and $R \propto N$ where N is number of turns). In a fixed area, increasing N leads to an increase in inductance, which for a fixed design frequency would necessitate a corresponding decrease in tank capacitance. In summary, to mitigate the effects of absolute L and C variations on f_R given a fixed sensing area and operating frequency, L should be increased to maximize Q and then C selected to resonate at the correct frequency. However, as it is shown in [73], the phase-noise due to the flicker noise of the active devices increases with number of turns. Therefore, the overall system m must be carefully designed to find the optimal trade-off between decreasing sensitivity to component variations and reducing phase-noise contributions of the active devices.

Although the previous expressions derived describing the transformer-based sensor can lead to significant insight into the drift-cancellation properties of the circuit, any realistic implementation will have both finite inductor quality factor and finite switch ON-resistance. For this purpose, it is important to study the sensitivity of the sensor in response to capacitive perturbations under these non-ideal conditions.

The sensor realizes a Q of 10 and each of the NMOS switch devices for enforcing the in-phase or anti-phase boundary conditions between the primary and secondary has an ON-resistance of 7 Ω.

In Fig 4.10, we compare the sensitivity of the base transformer sensor (f_R/f_R)

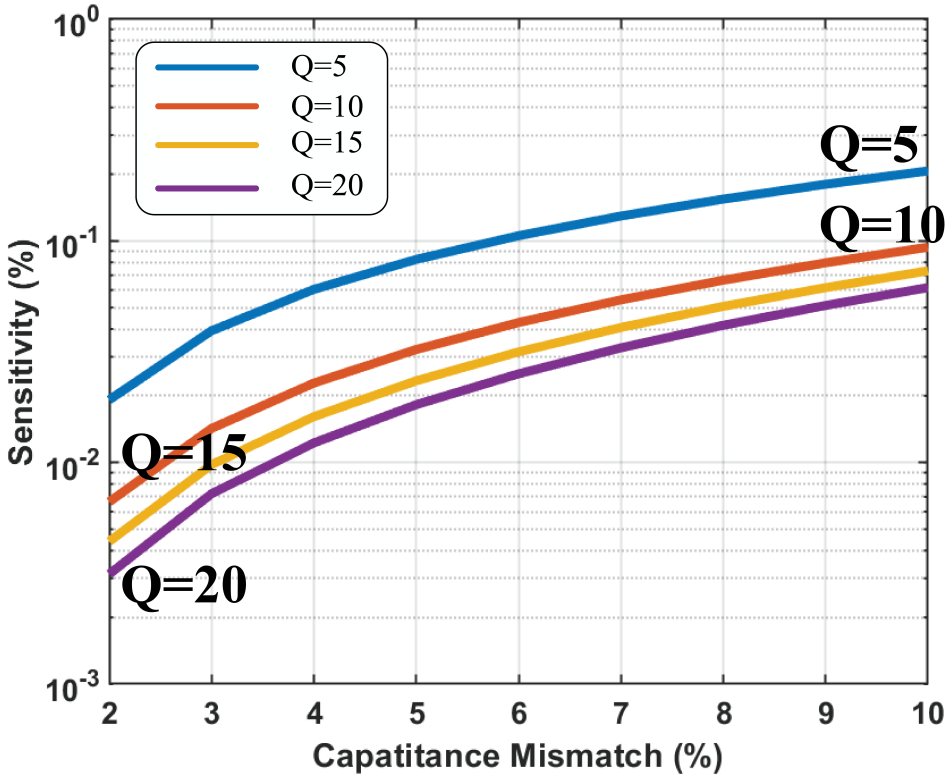

Figure 4.10: Sensitivity comparison of original reference-based approach, standard transformer-based self-referenced, and common-centroid layout transformerbased using nominal losses (inductor Q = 10 and switch R_{ON} = 7) with respect to the mismatch in capacitance on the sensor versus reference for original and mismatch in capacitance between primary and secondary for the transformer-based cases.

due to capacitive perturbation (up to 9% change in C on one side) to that of the original reference-based design $(f1 - f2)/f2$). We also show the sensitivity curve of the transformer sensor with a common-centroid layout of the tank capacitance, interleaving the primary and secondary capacitance (as shown in Fig 4.10). The common-centroid layout maximizes symmetry between the primary and secondary sides of the system, making it less sensitive to non-zero switch resistance and achieving a much lower sensitivity to perturbation than the original base design. The common-centroid-based transformer design achieves more than 2 orders of magnitude (220×) less sensitivity to component perturbations or drifts than the

original reference-based approach, highlighting the superiority of the design and its capability to attenuate drift due to process and thermal variations. In Fig 4.11(a),

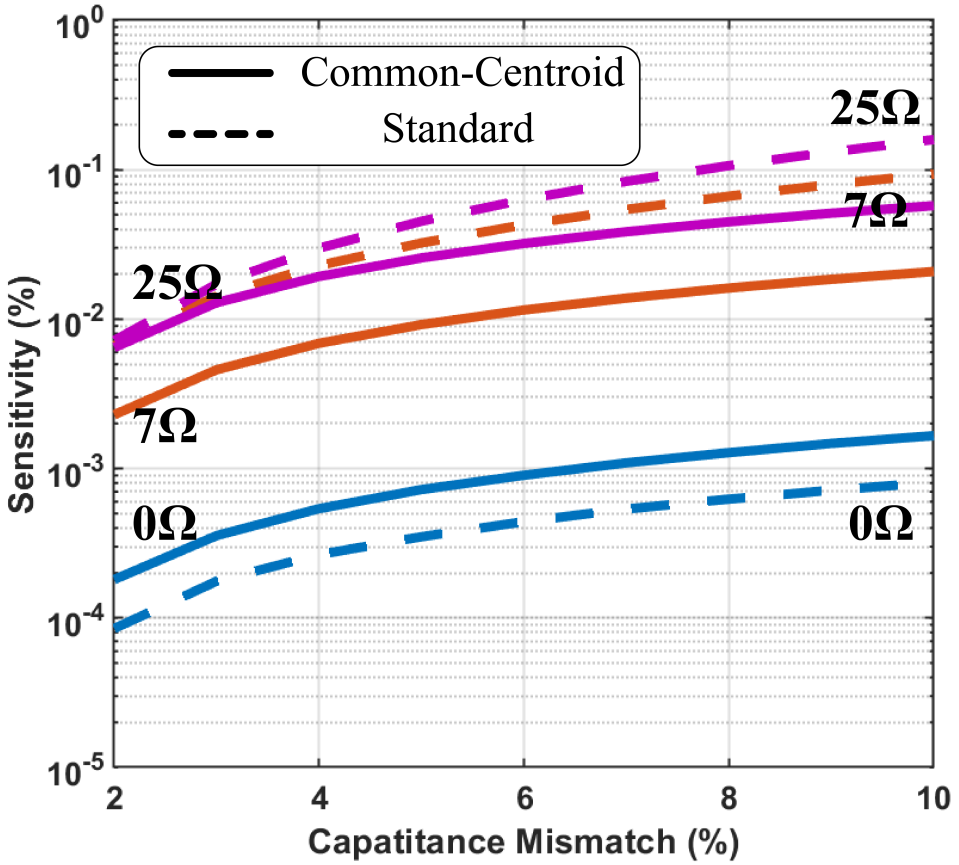

Figure 4.11: Sensitivity versus different switch ON-resistances (0, 7, and 20 Ω.) and nominal inductor Q of 10.

we compare the sensitivity of the circuit to capacitive perturbations for various different inductor quality factors and for a nominal switch ON-resistance of 7 Ω, demonstrating the importance of maximizing inductor Q. Interestingly, beyond a tank Q of 10 (which is readily realizable), the reduction in drift sensitivity is not significant. Finally, in Fig 4.11(b), we plot the sensitivity with a fixed nominal inductor Q of 10 for different switch ON-resistances (0, 7, and 20 Ω) for both the standard and common-centroid cases. Fig 4.11(b) clearly shows that the common-centroid version is more resilient to higher switch resistances compared to the base version and for non-ideal switches leads to significant performance improvements.

4.4 Device Implementation

To verify this sensing topology and quantify its efficacy for drift cancellation, we have designed and demonstrated a prototype 2×2 cell array of the transformer-based, self-referenced magnetic sensor in a standard, bulk 65-nm CMOS process. Similarly to other resonance-shift-based designs, the sensor prototype does not require any external biasing magnetic fields or exotic post fabrication processing. Unlike previous work, this design is completely self-referenced and achieves sub-PPM-level sensitivity over long measurement intervals, a property crucial for realistic biodetection experiments. As described earlier, maximum suppression of sensor noise

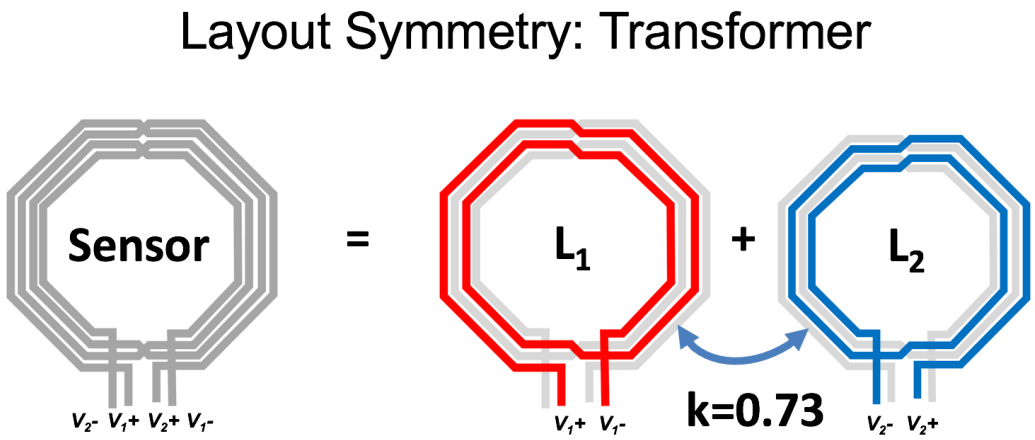

Figure 4.12: L_1 and L_2 are interleaved to maximize matching and achieve a high transformer coupling factor. $L = L_1 = L_2$ to make the ratio of the two tank resonance frequencies independent of inductance and so that magnetic content affects both primary and secondary sides identically.

and thermal drift occurs when the ratio X of the inductance and capacitance on the primary end of the transformer to the inductance and capacitance on the secondary is kept constant. To minimize variations in the ratio X due to absolute L and C variations, careful attention must be paid to the layout of every block of the sensing cell. To this extent, the sensing transformer is laid out as two interleaved symmetric spiral inductors (Fig 4.12) to maximize both matching of inductance on the primary and secondary sides in a radially symmetric fashion, but also to aid in achieving a large transformer coupling factor k_m. Furthermore, this interleaved topology results in magnetic beads placed in the empty center region affecting the inductance of both the primary and secondary sides identically, as is required for maintaining constant X. Each coil consists of two turns, achieving an inductance of 1.75 nH, quality factor of 10, and coupling coefficient of 0.73. The coupling coefficient fixes the

ratio of the two frequencies, f_R = 2.53, and the absolute frequencies are designed to be 1.44 and 3.65 GHz as a trade-off between maximizing response to magnetic beads at f_{LO} and minimizing response to the beads at f_{HI}. This fixes the total effective (including parasitics) tank capacitance on either side of the transformer at 4 pF.

Fig 4.13(a) shows the simulated magnetic field strength along a vertical line perpendicular to the sensing inductor, starting directly on the surface of the top metal (M9) and reaching up to 12 μm above the surface of the chip. The top surface of the chip, where magnetic material would be placed for detection, is 5 μm above the top inductor metal; however, the magnetic field strength at this height is 99% of the maximum field strength, indicating that any loss in sensitivity due to the passivation dielectric would be negligible. Fig 4.13(b) shows heat maps of the simulated magnetic field magnitude profiles on the chip surface, above the passivation, at both f_{HI} (3.65 GHz) and f_{LO} (1.44 GHz). Due to currents flowing in the opposing directions for the primary and secondary sides of the transformer, the magnetic field in the center sensing region is mostly cancelled out at f_{HI}. This is highly desirable as it further reduces sensitivity to magnetic content placed at the center at f_{HI}. Furthermore, the magnetic field magnitude at f_{LO} in the center region is quite uniform and varies less than 30% throughout the whole sensing region. This is important for minimizing gain variation due to magnetic bead position and maximizing sensor linearity and dynamic range for small bead concentrations. In a future implementation, the magnetic field uniformity could be enhanced even further, should better dynamic range be required, by using the inductor design approach of [84].

Complementary cross-coupled differential pairs are connected to both the primary and secondary sides to maximize capacitive matching (due to parasitic contributions of the active devices.) As shown in Fig 4.14, both the capacitors and active devices on each side are split up into two blocks and interleaved with each other in a common-centroid (ABBA) fashion to maximize matching and minimize variations due to on-chip process and thermal gradients. A binary switched resistor bank with five effective control bits is used to control the tail bias currents of the oscillator and minimize phase-noise contributions. The oscillator DC bias current can be adjusted from 3 to 30 mA depending on the switch settings. The sensor runs off a standard 1-V supply voltage and thus at the lowest current bias consumes only 3 mW of DC power. Each sensing cell has an open-drain analog buffer for time domain characterization of the oscillator circuitry, as well as a digital buffer chain with an on-chip digital divider which divides down the oscillation frequency by a factor of

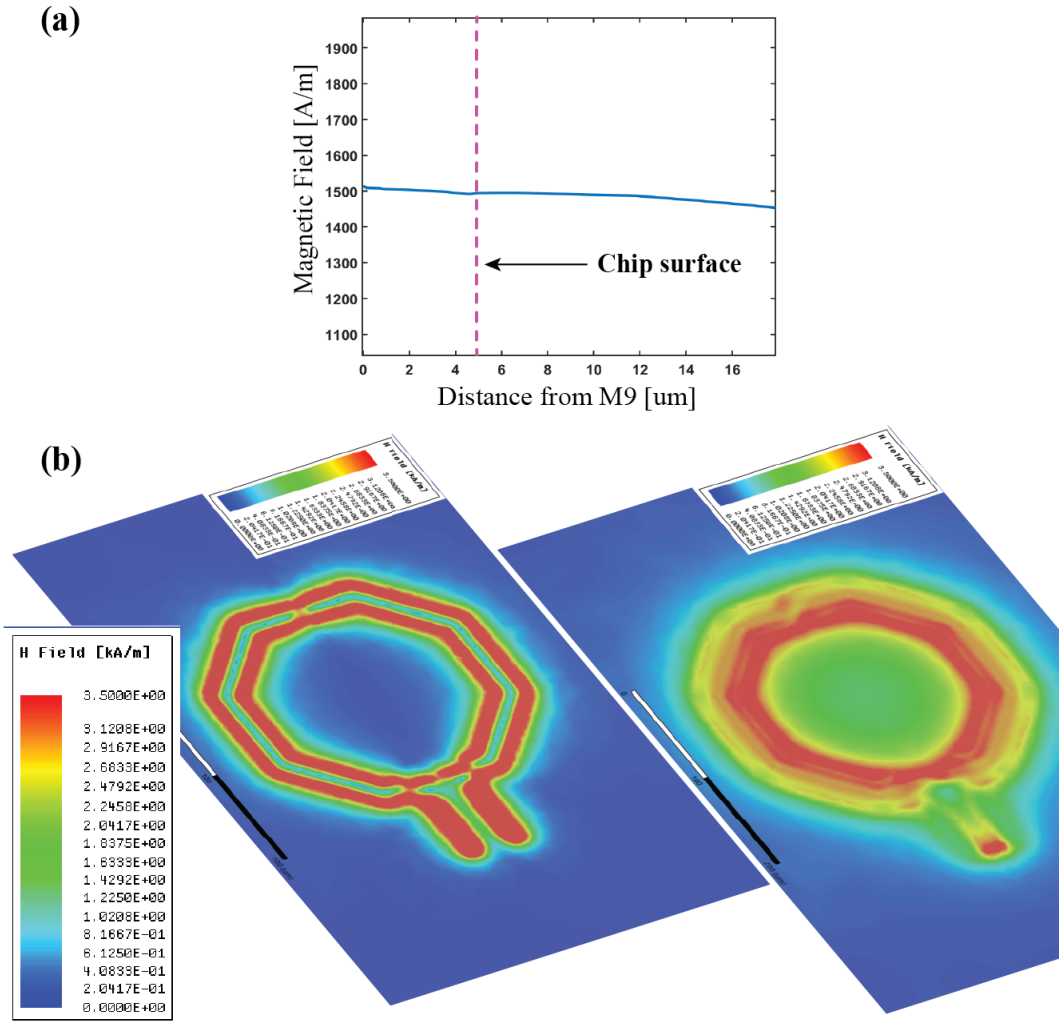

Figure 4.13: Magnetic field simulation (a) Magnetic field magnitude plotted on a vertical line over the center of the sensor, starting directly above metal trace (M9) and extending 12 μm above the surface of the chip. The top surface of the chip, due to passivation, is 5 μm above M9. The magnetic field strength at this distance is 99% of that directly over the surface of the metal, indicating that no appreciable sensitivity is lost due to the passivation layer. (b) Heat-map magnetic field magnitude profiles over the top surface of chip (high frequency on left, low on right), demonstrating the magnetic field uniformity in the center sensing region of the transformer. There is less than 30% variation in magnitude across the whole center region of the low frequency. Furthermore, the magnetic field in the center for the high frequency is very weak due to cancellation, favorably leading to even further reduction of the effects of magnetic material in the center sensing region on the high frequency. Excitation terminal voltage is normalized to $+/-1V$.

32. This results in a low frequency, CMOS-level digital output which can be easily

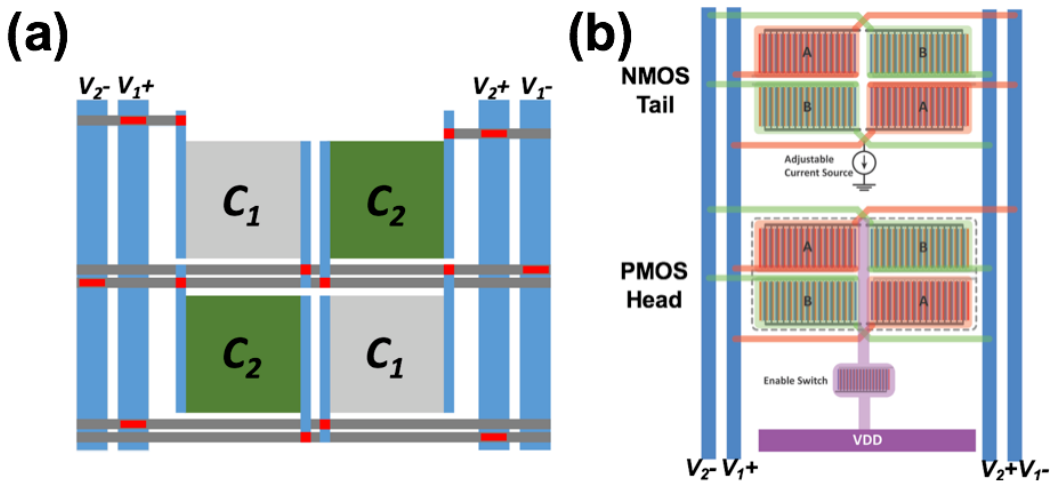

Figure 4.14: Layout considerations (a) C_1 and C_2 are each split in half and interleaved in a common-centroid fashion to maximize matching. (b) NMOS and PMOS active devices are also split into two blocks per side and laid out in an interleaved ABBA fashion to minimize effects of process and thermal gradients.

counted by an inexpensive off-chip digital circuit. Fig 4.15(a) shows the full sensor oscillator circuitry, and Fig 4.15(b) shows a high-level block diagram of the sensor array design.

4.5 Measurements

To characterize the drift-cancelling capability of the design, we measure a single sensing cell for an 11-h-long time period, alternating counting f_{LO} and f_{HI} for 0.5 s at a time (achieving 1 Hz counting resolution due to counting both rising and falling edges). Fig 4.16 shows the results of this experiment, plotting the raw f_{LO} measurement data and overlaying the reconstructed low frequency from the measured f_{HI} data using the known transformer coupling factor ratio relation (f_R).

This result demonstrates the excellent tracking ability of the sensor, despite the fact that absolutely no thermal compensation was employed and standard, noisy lab benchtop power supplies were used. The frequency noise floor of the drift-compensated signal was measured to be 500 Hz, corresponding to a 0.35-ppm detection capability. This means that the counting resolution only needs to be within 500-Hz accuracy, implying that each frequency can be counted for only 1 ms at a time (instead of 0.5 s). Faster counting can be quite useful for dynamic and cell-counting experiments. Next, the sensor response due to magnetic beads was characterized. For this experiment, 4.5-μm size Dynabeads were utilized due to

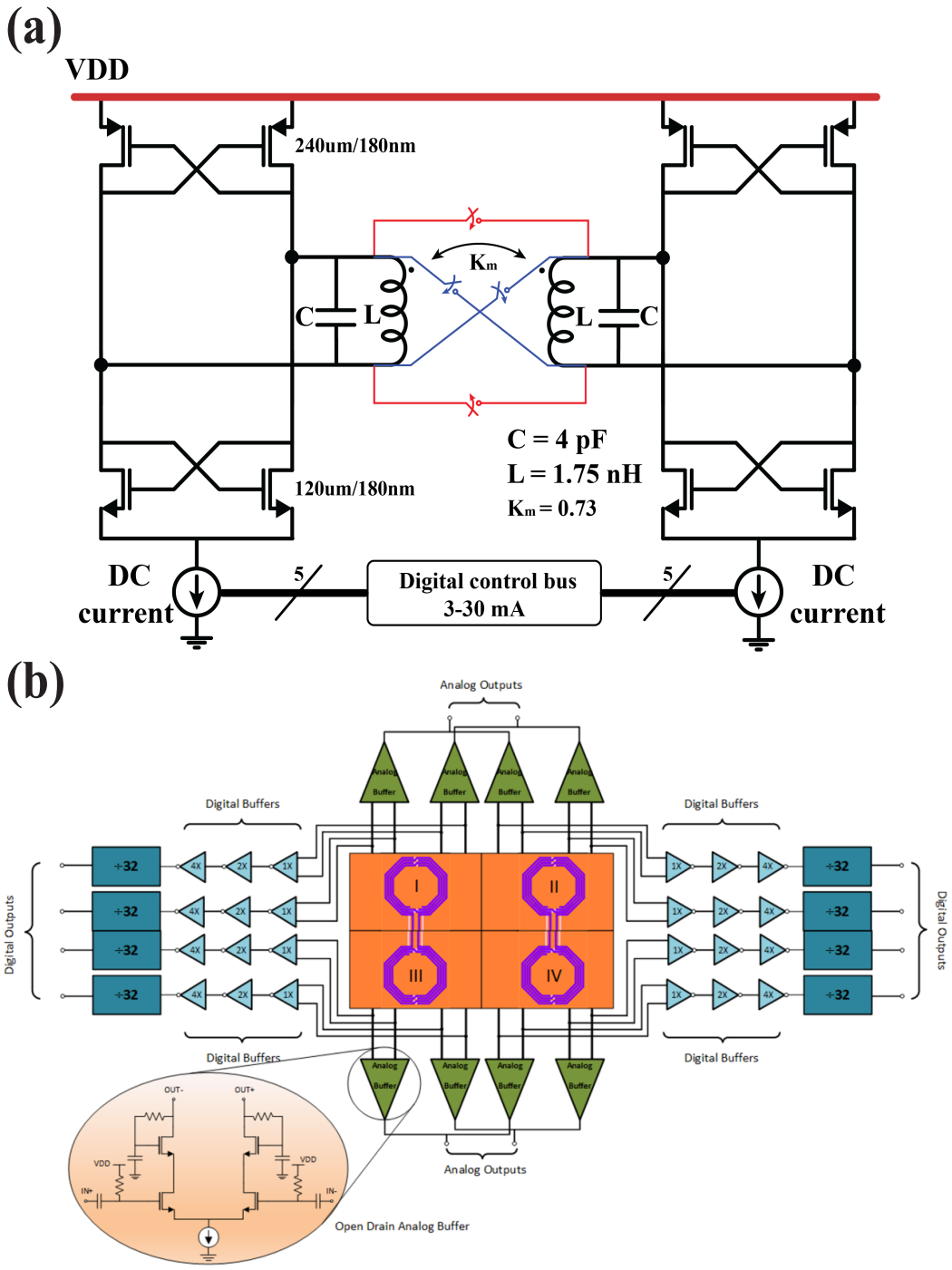

Figure 4.15: System implementation (a) Transformer-based magnetic sensor design with oscillator active devices. (b) High-level block diagram of 2×2 sensing system with corresponding analog and digital buffers.

their ease of manual handling under standard optical magnification. The beads were initially suspended in a buffer solution; however, the solution dried within seconds

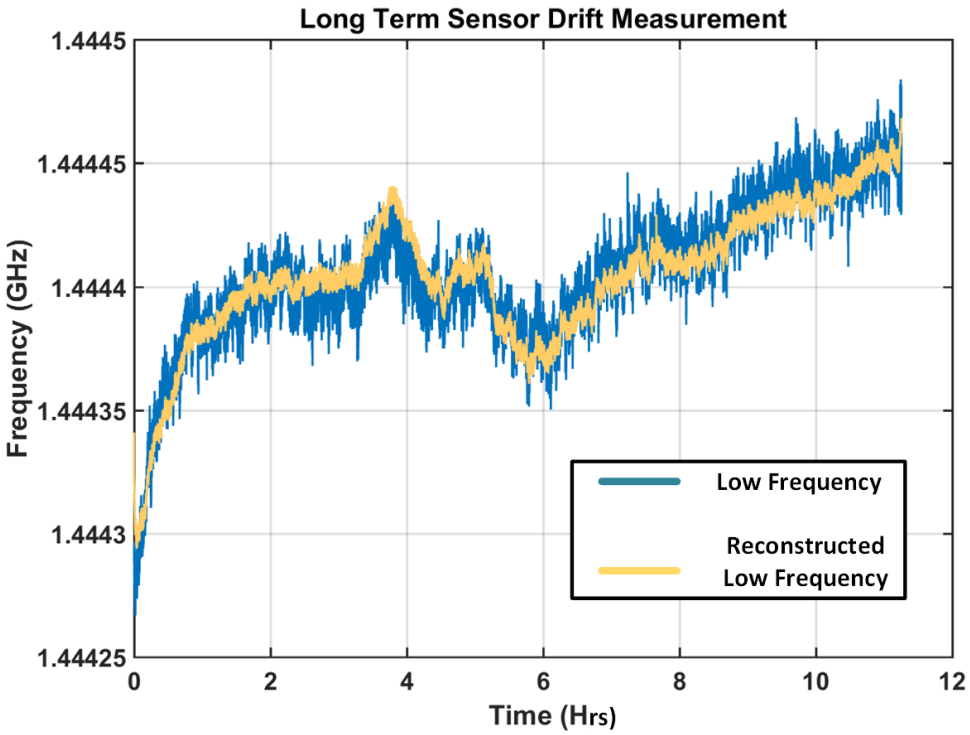

Figure 4.16: Sensor was measured over an 11-h-long time period, alternating counting f_{LO} and f_{HI} for 0.5 s each. Raw f_{LO} measured data is plotted with f_{LO} reconstructed from measured f_{HI} data overlayed, using the known resonant-frequency ratio relation f_R. The two curves track each other excellently, despite the fact that no thermal compensation was utilized.

after deposition onto the sensor surface and all bead measurements were taken dry. Fig 4.17(a) and (b) shows the raw measurement data from this experiment as well as the drift-compensated response. Important to note is that the amount of drift during the 10-min measurement time interval is of the same order of magnitude as the frequency shift due to the magnetic beads, highlighting the necessity of adequate drift-compensation for successful detection of small amounts of magnetic content. The sensor response due to differing amounts of beads (ranging over 2 orders of magnitude) was characterized and plotted in Fig 4.17(c). The sensor exhibits excellent linearity throughout the whole range, corresponding to a dynamic range of at least 62 dB. This measurement was limited only by practical considerations, due to the difficulty of accurate manual placement and counting under optical magnification of large numbers of magnetic beads. Thus, the actual sensor dynamic range is expected to be significantly higher.

The oscillator phase noise was measured to be -131.8 dBc/Hz at a 1-MHz offset

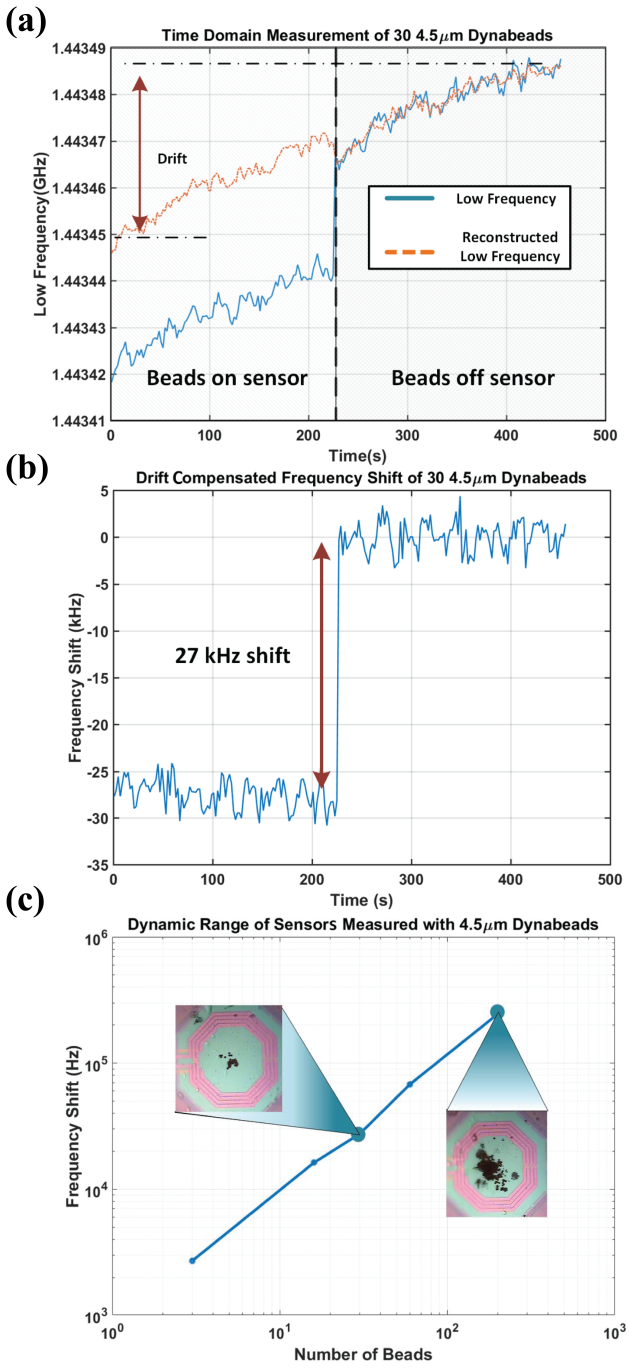

Figure 4.17: Sensor characterization.

for the 1.44-GHz frequency and -122.8 dBc/Hz at 1-MHz offset for the 3.65-GHz frequency. Fig 4.18 shows the phase-noise measurements of the sensor at both operating frequencies from 100-Hz to 10-MHz offset from the carrier. Fig 4.19 shows an optical micro-graph of the 2×2 prototype transformer-based sensing array.

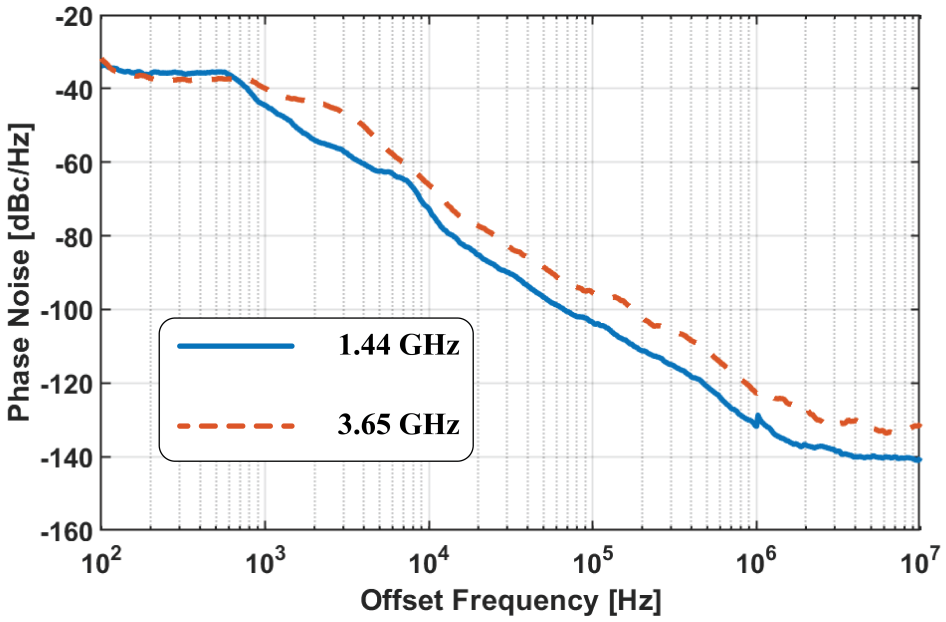

Figure 4.18: Phase noise measurement

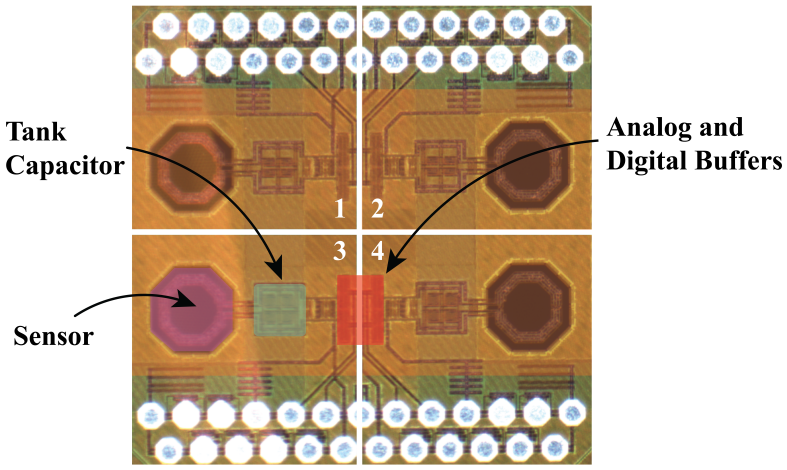

Figure 4.19: Chip micrograph.

Finally, a simple DNA biodetection experiment was performed to demonstrate the feasibility of the sensor to be used for *in vitro* PoU detection. Each sensing site is $250 \times 250\ \mu m^2$, offering adequate area for DNA, protein, and cell detection experiments. We designed a 90 nucleotide target DNA strand to be detected, a 42 nucleotide capture strand, complementary to the bottom half of the target and a 42 nucleotide probe strand, complementary to the top half of the target. The capture strand is attached to the chip surface via a simple epoxysilane surface chemistry

[85], and the target strand is attached to 1-μm Dynabeads using a Streptavidin–Biotin chemistry. A 1-μm concentration of target DNA was used in this experiment.

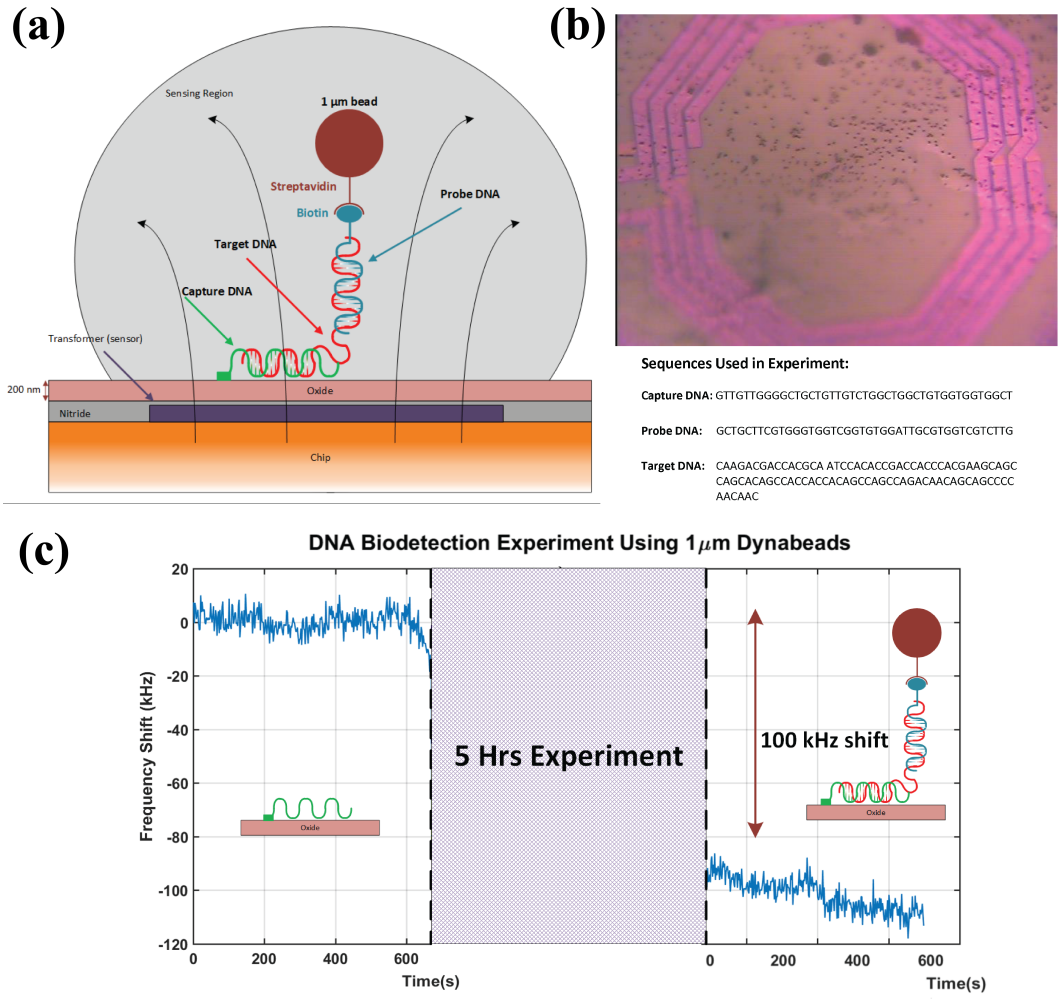

Figure 4.20: Bio experiment overview (a) DNA biodetection experimental setup. (b) Optical micrograph of sensor surface after hybridization of target DNA, attaching magnetic beads to the chip surface. (c) Drift-compensated time-domain sensor response, demonstrating sensor's capability to easily detect the target DNA, despite a 5-h time gap between the initial (pre-hybridization) and final (post-hybridization) measurements. No time-averaging was used. The 5-h window between the two measurements was used for the wet-lab experiment, and the sensor was not active during this time period.

Fig 4.20(a) shows a diagram of the experiment setup, along with the sequences of the DNA strands used. Fig 4.20(b) is an optical micrograph of the sensor surface after the experiment was performed, indicating that the hybridization worked, successfully binding the magnetic beads to the chip surface due to the present target DNA hybridizing to the capture and probe strands. Finally, Fig 4.20(c) shows the

drift-compensated sensor response, indicating that the sensor can easily detect the target DNA, despite the fact that there was a 5-h time period between the initial measurement (before the hybridization) and the final measurement (post-hybridization). Both measurements were done on a dry sensing surface after any buffering solutions had evaporated. The 5-h time window between the two measurements was utilized for the wet-lab experiment, and the sensor was inactive during this time. While the hybridization experiment could have been done in less time than 5 h, this extended time interval between the initial pre-attachment and final result measurement was chosen to show that this sensing approach can maintain excellent sensitivity and drift-compensation even for more complicated assays which may require many hours to reach completion and may need real time, dynamic monitoring of target concentrations throughout the whole experiment. This implies that this sensing approach is indeed suitable for performing such complex biodetection experiments, which crucially depend on the underlying long-term stability of the sensor.

4.6 Conclusion

In conclusion, we have developed a novel reference-free, transformer-based drift-cancellation technique for CMOS resonance-shift magnetic imagers. We have analyzed the proposed topology and compared it against previous reference-based approaches, demonstrating over 2 orders of magnitude of improvement in drift cancellation. Design considerations and trade-offs were discussed for the implementation of these transformer-based sensors. Finally, a prototype 2×2 cell sensing array was designed and demonstrated in a standard, 65-nm bulk CMOS process, and a simple bio experiment was performed to verify the feasibility of the approach to be used in realistic *in vitro* PoU settings. Furthermore, biodetection experiments for both DNA detection as well as immunoassays are currently in progress.

Chapter 5

SCALABLE DECENTRALIZED ROUTER

5.1 Motivation

As the communication systems and protocols are moving toward the next generations (5G, 6G, and beyond), engineers and system designers are trying to increase the bandwidth, overall speed, and the number of potential users. Few approaches are being considered such as having smaller communication cells, pico-cells, and increasing the operational frequency. However, this demand for bandwidth is regulated by Shannon's theorem that states that the maximum channel capacity is proportional to the available bandwidth and the logarithm of the signal-to-noise ratio (SNR) (plus 1)[1]. System and circuit designers have toiled to improve link budgets and achieve higher data rates and system capacities. Moreover, decreasing the overall latency is another side that needs to be worked on. On the circuit side, there has been a push to higher frequencies (where more bandwidth is available due to the fractional bandwidth) by improving critical circuit building block performance, e.g., power amplifier efficiency and linearity or the overall receiver sensitivity. On the system level, we have moved from single channel static transmitters and receivers to sophisticated programmable massive multiple-input multiple-output (MIMO) systems that can form larger apertures in transmitters and/or receivers to perform a slew of complex functions [86]–[88]. Regarding the attractiveness of the MIMO systems, phased arrays have several advantages. These arrays enhance the effective isotropic radiated power (EIRP) in transmitters and sensitivity of receivers to increase the available data rates through the second key parameter in Shannon's theorem, SNR. Also, the spatial directionality and beam confinement provided by arrays enables more effective spatial partitioning of the bandwidth, leading to higher frequency reuse ratios and smaller cell sizes which work toward having more overall users in a specific cell [2].

Despite these clear advantages, existing phased arrays face aperture scaling limitations inherent to their centralized architecture. The challenge of aggregating every element's signals within the array grows perniciously at higher element numbers and aperture sizes. Furthermore, for high-speed data communication, the difference in

[1] $R_B = B \cdot \log_2(SNR + 1)$.

[2] This work was done in collaboration with Austin Fikes and Samir Nooshabadi.

data arrival (and departure) times between elements generates dispersion manifested as ISI [89] that must be dealt with through array level delays or complex equalization schemes in a centralized fashion. Even clever designs rapidly reach practical limitations of signal routing density, interface bandwidth, and data synchronization. Additionally, a centralized approach precludes spatially and/or electrically separated apertures working together.

5.2 Scalable Decentralized Router

In here, we propose the scalable router: a decentralized relay array architecture that can selectively receive multiple signals from several desired incident angles and re-transmit them in other arbitrary directions with minimal data distortion. Such scalable routers unite smaller spatially and electrically separated apertures to produce an effective large aperture at high data rates in a decentralized and dynamic fashion.

Intuitively, a scalable router is analogous to a programmable mirror that can be dynamically redirected in different directions for different incident signals that will bounce each one of those incoming beams towards different targets in various locations. Unlike a standard passive mirror, this programmable active one amplifies and conditions the signals it reflects. This needs special treatment for data and carrier signals. Furthermore this mirror can be constructed out of multiple disconnected elements that could also move in real time. The ability to operate multiple elements to form a decentralized, non-uniform, and/or dynamically changing array can open a plethora of new opportunities. For instance, locally-powered arrays of elements with no need for timing reference synchronization can be deployed, gradually built-up, and constantly changed across unused walls, ceilings, and buildings surfaces at multiple locations (Fig 5.1). Furthermore, future infrastructure can allow such systems to be incorporated into various platforms, such as mobile devices, vehicles, building infrastructures, airborne systems, and satellites. It is obvious that such arrays would greatly benefit from architectures that allow decentralized operation with dynamically moving elements.

The scalable router architecture can be seen as evolution and marriage of two existing microwave concepts: the bent pipe relay and the standard centralized phased array. The bent pipe relay is essentially a receive antenna, an amplifier, and transmit antenna connected in series (Fig. 5.2). This simple system re-amplifies incident signals and redirects them with a static, unchanging radiation pattern.

A bent pipe relay can be modified to use two standard M-element phased arrays,

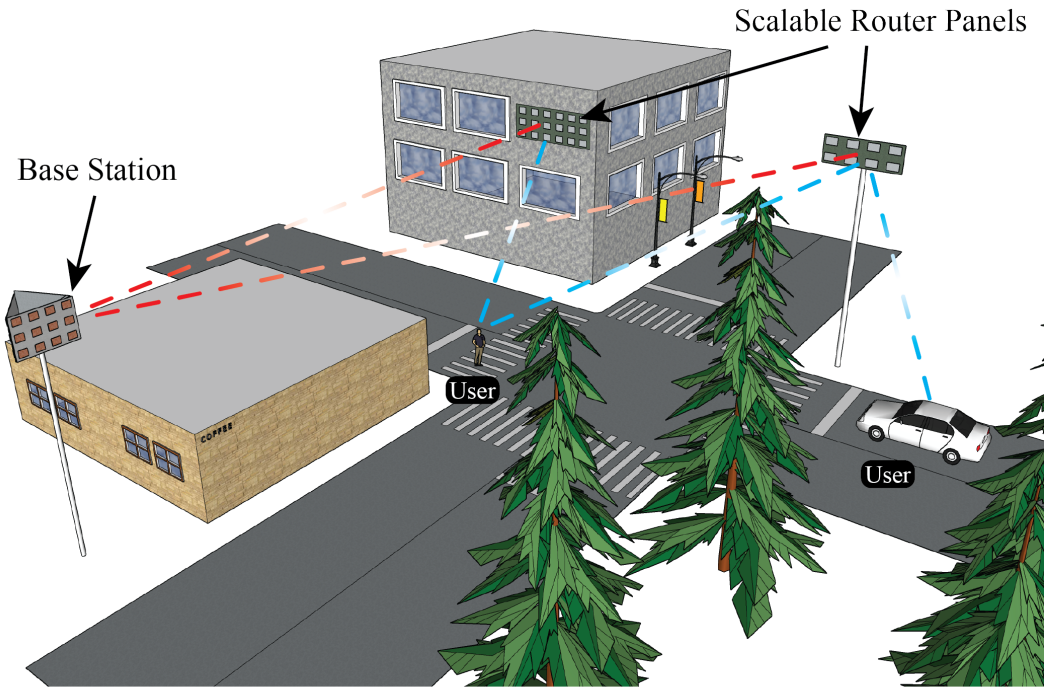

Figure 5.1: The scalable router can extend the effective reach of basestations to greater distances or areas blocked by obstructions.

one used as a receiver, the other used as a transmitter (Fig. 5.2(b)). Each phased array creates an electronically steered beam of microwave power by controlling the phase of each element within an antenna array. Now the system can electronically steer the transmit and receive beams.

To create the scalable router, we split the centralized aggregation node that is shared by all elements in the array and add tunable broadband time delays to each branch, as shown in the next step in the progression shown in Fig. 5.2(c). For an idealized, conceptual model in which amplification, summation, and delay are linear operations, this change is simply an application of the distributive property to a delayed-array. Rather than sum, then split the incident signals within the system, signal summation occurs only in the re-radiated beam. The tunable delay elements allow the receive and transmit beams to be steered without centralized signal aggregation occurring within the router[3].

[3]There are some similarities (and differences) between the scalable router architecture and early space fed phased arrays [90]. The scalable router architecture and space fed arrays both have tunable

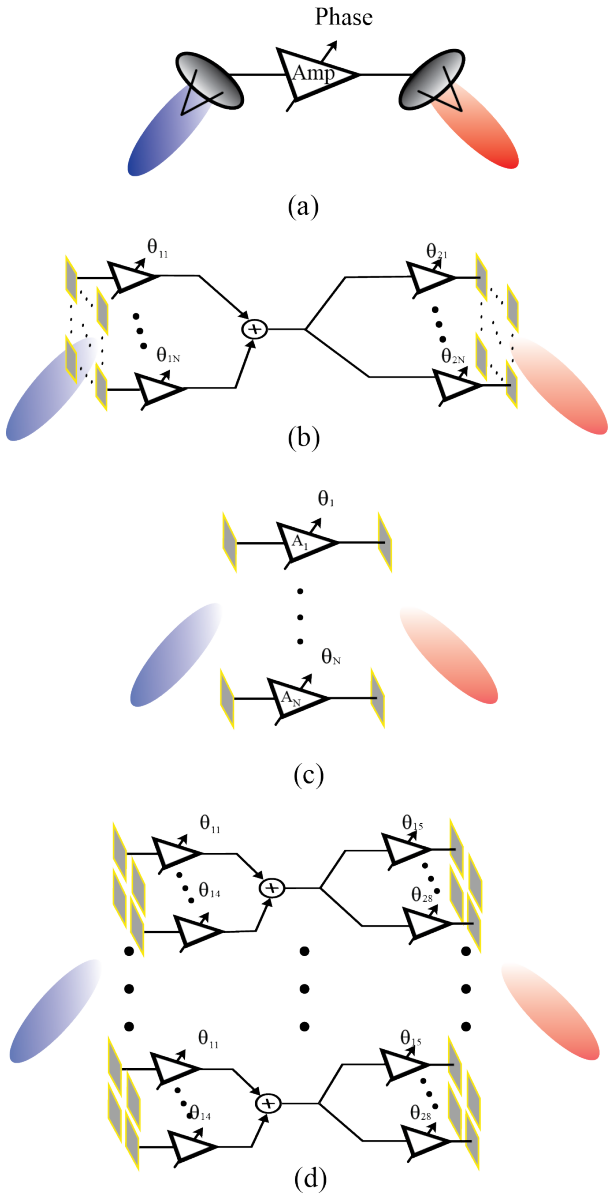

Figure 5.2: Relay architectures (a) Bent pipe relay, (b) Centralized phased array implementation of bent pipe relay, (c) Scalable router system architecture, (d) Hybrid scalable router.

While at a highly abstracted level the scalable router can perform all of the functions of a phased array relay, in a real system there are profound implications for noise, linearity, isolation, and other system parameters which we will explore in this work. Fig. 5.2(d) described as a hybrid scalable router represents a syncretic state of local centralization and system level decentralization in which the router

elements with radiative inputs and outputs, but differ as the scalable router is decentralized, uses true time delay, and operates in the radiative far field.

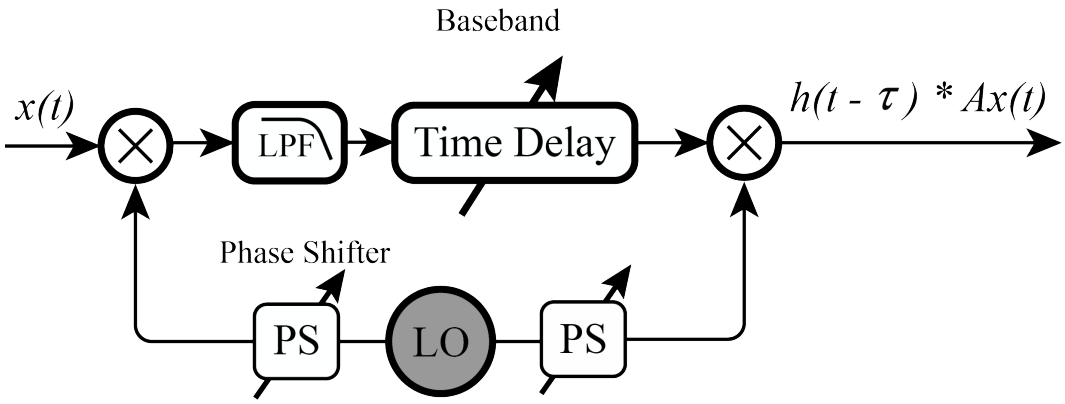

Figure 5.3: Possible integrated circuit branch implementation using baseband time delay. A scalable router branch provides amplification, time delay, and filtering to the signal it receives and transmits.

branches are composed of smaller phased arrays. The benefits offered by this array of arrays are described later. A critical feature of the scalable router is that each branch (receive element connected to a transmit element) does not interface with other branches within the array. The transmit and receive beams are steered entirely by setting the delay within each branch (possibly as low frequency digital signals). Not only does this architecture bypass the challenge of centralizing data, but also means that the branches can be implemented without a shared timing reference. The router is fully decentralized: a router may be formed by apertures which are physically separated. The decentralized operation of the scalable router emerges from each branch performing a decentralized function, not reliant on information from any other branch. Fig. 5.3 models idealized operation of a single branch where amplification, delay, and filtering are performed. While these functions can be accomplished by systems at any frequency built with a variety of technologies, Fig. 5.3 shows an down-conversion/up-conversion implementation well suited for integrated circuits at microwave frequencies. Interestingly, the decentralized architecture does not require the local oscillator (LO) signals within a branch to be phase or frequency locked to any other branch. Branch independence enables routers formed by a combination of static arrays, satellites, autonomous aircraft, ground vehicles, or any other surface which can support an aperture. Table I summarizes the differences between the scalable router architecture and conventional phased arrays. The scalable router architecture is well suited for integration in emerging mm-wave communication infrastructure. Universal adoption of mm-wave systems is hindered by their line-of-sight nature and the high absorption of walls and other obstructions

Table 5.1: Scalable Router vs. Conventional Phased Array

	Scalable Router	Conventional Phased Array
Beam forming/steering	Carrier and data	Carrier only
Avoids internal, centralized signal combining and distribution	Yes	No
Aperture scalability	Yes	Limited by ISI
Operate without elements sharing timebase	Yes	No
Fully decentralized	Yes	No

at these frequencies. Attempting to overcome these issues in multi-room indoor settings, dense urban environments, remote areas, and flying systems using traditional phased arrays can lead to unwieldy apertures and power requirements (Fig. 5.1). The scalable router excels in this niche. It can dynamically bounce signals around obstacles or extend and fortify low quality communication links. The scalable decentralized router architecture can enable uninterrupted high-speed connectivity in the presence of large path loss as well as static and/or dynamic obstructions.

Several possible scalable router use-cases are examples of cooperative diversity, which is used as an umbrella term for multi-antenna, relay-reliant, multi-user, or multi-hop schemes intended to increase channel capacity in communication networks [91]–[94]. A substantial body of theoretical cooperative diversity research exists, often focusing on optimizing the capacity of a hypothetical network consisting of a base-station and several cellphone users given power constraints and incomplete channel state information [95]–[99]. While these analytical works have not explored the challenges, potential, and emergent capabilities of large scale relay array hardware such as the scalable router, their analysis might be fruitfully adapted for specific scalable router use scenarios.

Electronically Steerable Microwave Mirror

The scalable router can act as an electronically steerable mirror at microwave frequencies. As the delay within each branch is electronically changed (mirror is rotated), the incident signal is conditioned and re-routed (reflected) to a new direction. Fig. 5.4 models the electronically steerable mirror analogy for a 16 element scalable router. Despite our desire for tidy comparisons, the mirror analogy elides

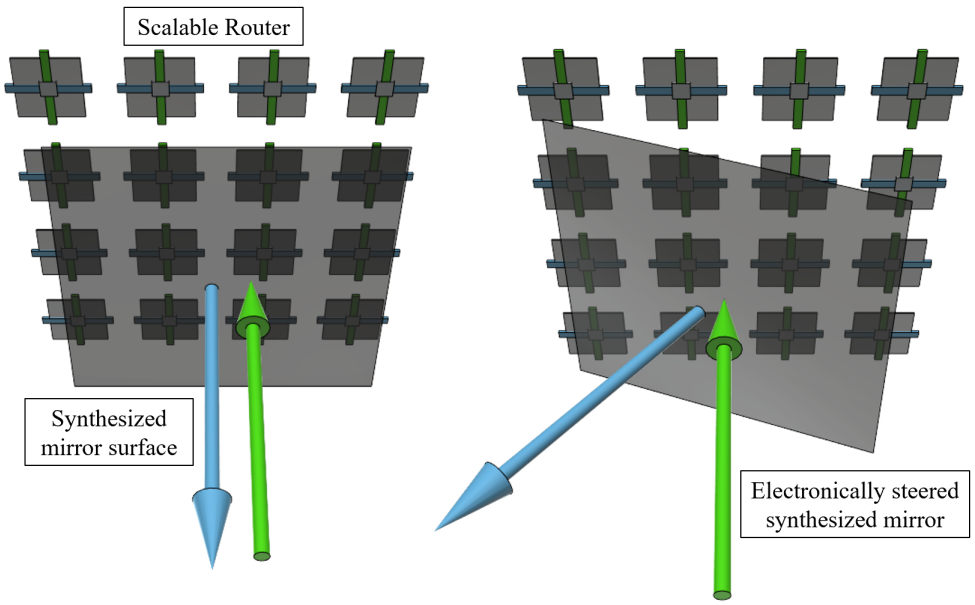

Figure 5.4: A 16 element scalable router synthesizes a microwave mirror (shown as transparent) which can be electronically steered.

subtle but critical aspects of scalable router beamforming.

The beam patterns of dynamic, spatially decentralized routers can be determined with a geometric derivation. Unlike a conventional centralized phased array, no aggregation occurs within the scalable router, intimately linking the receive and transmit gain beam patterns and deviating from the behavior predicted by our earlier geometric optics analogy. Fig. 5.5(a) shows a general, decentralized array structure. The relationship between the intended direction of the received beam pattern, $\hat{R}_{rx}$, the intended direction of the transmit pattern, $\hat{R}_{tx}$, and the unwrapped phase (a surrogate for the delay), δ_m, of each branch at location $\vec{r}_m$, can be derived as follows. Considering the origin of our coordinate system as a phase reference, we note that the difference in propagation length to a point $\vec{R}_{tx}$, between a wave radiated by an emitter at $\vec{r}_m$ and the origin is

$$|\vec{R}_{tx} - \vec{r}_m| - |\vec{R}_{tx}| = |\vec{R}_{tx}|\sqrt{1 - 2\hat{R}_{tx} \cdot \hat{r}_m \frac{|\vec{r}_m|}{|\vec{R}_{tx}|} + \frac{|\vec{r}_m|^2}{|\vec{R}_{tx}|^2}} - |\vec{R}_{tx}|. \tag{5.1}$$

Under the special case $\left|\vec{r}_m\right| / \left|\vec{R}_{tx}\right| \ll 1$ (which implies that the array aperture is much smaller than the distance to the intended beamforming point), (5.1) can be

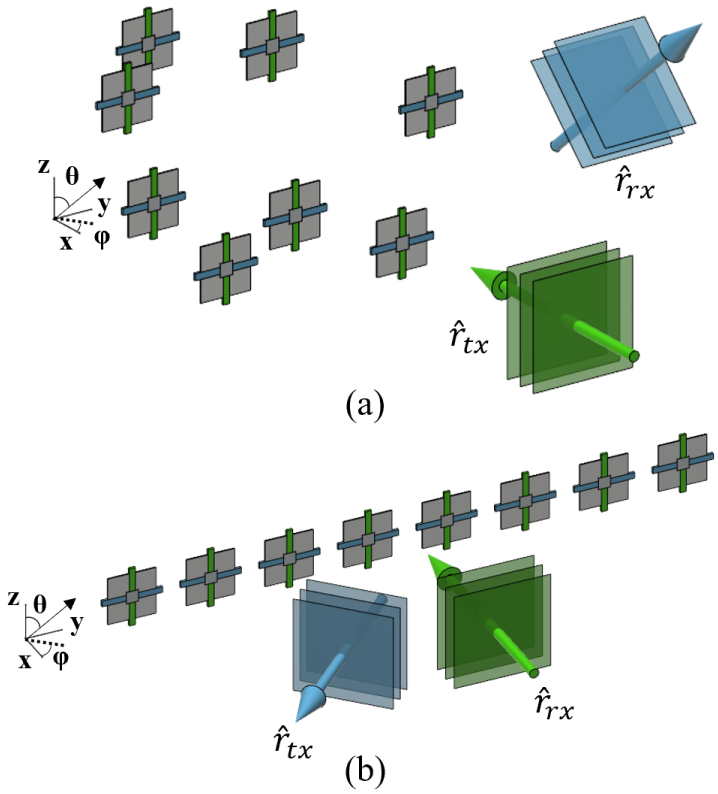

Figure 5.5: Router examples (a) An 8 branch router arranged in 3D space (b) A 2D 8 branch router with equidistant branch spacing. That spacing is chosen to be $d = \lambda/2$ for the simulated patterns shown in Fig. 5.6.

Taylor-expanded to yield:

$$\left|\vec{R}_{tx} - \vec{r}_m\right| - \left|\vec{R}_{tx}\right| = -\hat{R}_{tx} \cdot \hat{r}_m \left|\vec{r}_m\right| + O\left(\frac{\left|\vec{r}_m\right|^2}{\left|\vec{R}_{tx}\right|}\right). \tag{5.2}$$

The above propagation length variation manifests itself in the phase propagation term of electromagnetic waves, which under substitution of (5.2) becomes

$$\exp\left[jk(-\hat{R}_{tx} \cdot \hat{r}_m \left|\vec{r}_m\right| + O\left(\frac{\left|\vec{r}_m\right|^2}{\left|\vec{R}_{tx}\right|}\right))\right] \approx \exp\left[-jk\hat{R}_{tx} \cdot \hat{r}_m \left|\vec{r}_m\right|\right] \tag{5.3}$$

where the above approximation can be made under the far-field condition $\left|\vec{r}_m\right|^2/\lambda \ll \left|\vec{R}_{tx}\right|$. We note that (5.3) is the phase difference incurred during transmission of each emitter with respect to the origin. The dual set of phase differentials can be found for the case of an incident wave from a receive direction, $\hat{R}_{rx}$, in an analogous

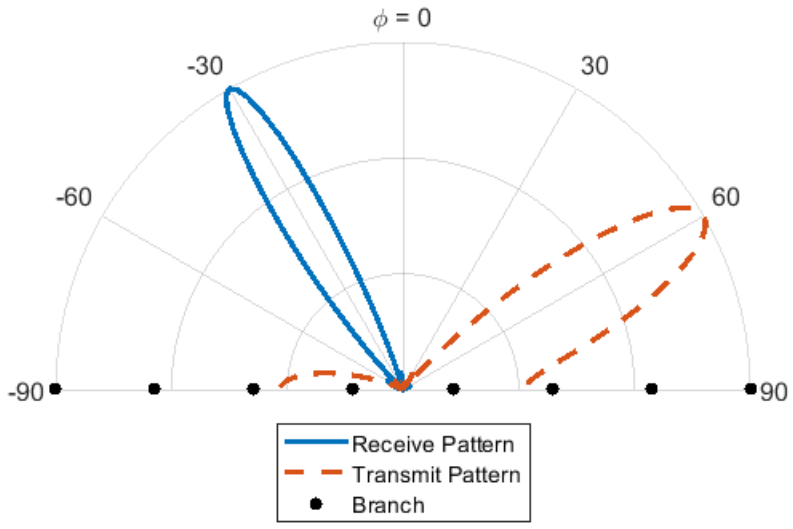

Figure 5.6: Normalized linear magnitude pattern plot for 1D 8-branch linear array with $\lambda/2$ branch spacing, and branch delays programmed for an intended $\phi_{rx} = -30°$ and $\phi_{tx} = 60°$.

manner. To keep the notation simpler, we define our new $\hat{R}_{rx}$ to be pointing outward. Combining the two phase terms results in the following phase propagation value:

$$\exp\left[-jk\hat{R}_{tx}\cdot\hat{r}_m\left|\vec{r}_m\right|\right]\exp\left[-jk\hat{R}_{rx}\cdot\hat{r}_m\left|\vec{r}_m\right|\right]. \tag{5.4}$$

For coherent beamforming to occur, (5.4) needs to equal 0 for the desired beamforming direction. If a variable unwrapped phase, $e^{-j\delta_m}$, is added to the mth emitter, then coherent beamforming will occur for

$$\delta_m = -k\left|\vec{r}_m\right|(\hat{R}_{tx}\cdot\hat{r}_m + \hat{R}_{rx}\cdot\hat{r}_m). \tag{5.5}$$

We note that in (5.5), the required delay for beamforming is given in terms of unwrapped phase δ_m. Noting that $k = \omega/c$, the required phase delay is a frequency dependent term. As noted earlier, for wideband signals a frequency-independent phase delay will result in data decoherence/ISI. Thus, the delay in (5.5) is best implemented via a time-delay, which is given by the recast form of (5.5)

$$t_m = -\frac{\left|\vec{r}_m\right|}{c}(\hat{R}_{tx}\cdot\hat{r}_m + \hat{R}_{rx}\cdot\hat{r}_m). \tag{5.6}$$

As a simple and familiar example, the case of a 1D M element array in Fig. 5.5(b), with branch pitch d, is studied. For simplicity, we assume that the array coordinates

are given by $\vec{r}_m = [0, md, 0]$. This results in (5.6) being evaluated as

$$t_m = -\frac{md}{c}(\sin\theta_{tx}\sin\phi_{tx} + \sin\theta_{rx}\sin\phi_{rx}) \tag{5.7}$$

and since Fig. 5.5(b) describes the $x-y$ plane, we set $\theta_{rx} = \theta_{tx} = \pi/2$ which results in (5.7) reducing to

$$t_m = -\frac{md}{c}(\sin\phi_{tx} + \sin\phi_{rx}) \tag{5.8}$$

where t_m is the delay of the mth branch with respect to the $m = 0$ branch at the origin. To ensure that all delays are positive, a common delay to all branches may be added. Thus (5.8) can be used to set the internal delay of each emitter to achieve desired reception and transmission angles. Using this expression allows the scalable router to operate as a programmable microwave mirror—the user can set the direction in which reflections should be sent. While the above example is for a 1D array, this derivation can be used for 2D and 3D routers by using the general form found in (5.6).

To produce a conventional beam pattern, we must choose a specific direction for either the receive or the transmit. Consider a 1D 8-branch array with $d = \lambda/2$ branch spacing, such as that shown in Fig. 5.5(b), with intended receive direction, $\phi_{rx} = -30°$, and intended transmit direction, $\phi_{tx} = 60°$. The needed branch delays are calculated using (5.8). Fig. 5.6 shows the transmit and receive beam patterns for the programmed array. The transmit beam pattern shows the relative strength of the radiated beam from the router in any given direction when a signal is incident on the router at $-30°$. The receive beam pattern shows how energy incident on the router from any given direction contributes to the transmitted beam at $60°$.

Peripheral Vision

While the patterns in Fig. 5.6 describe the intended behavior of the router, attentive readers may note that for a given set of branch delays, signals may be received from, and transmitted to, directions other than the intended. We describe this as peripheral vision, since signals incident from outside the directions from which the array is "looking" may be redirected as well. The mathematical justification is apparent from (5.8) as there are many pairs of ϕ_{rx} and ϕ_{tx} that satisfy the equation for a given t_m. While the peripheral vision does not interfere with the primary function of the system, it may be undesirable in certain situations. Fortunately, element position can be used to suppress the router's peripheral vision.

Router peripheral vision can be quantified for the more general case of a router with steering capability encompassing the entire range of azimuths and elevations. Since

the unwanted coherent combination of power is of concern, peripheral vision occurs wherever the carrier signal coherently combines, even though the data signal may be incoherent. Focusing on the carrier signal and assuming far-field conditions, the field at a point in space due to a uniformly excited router is proportional to the summation of the propagation phases of each branch

$$\left|\vec{E}(\hat{R}_{tx}, \hat{R}_{rx})\right| \propto \left|\sum_{n} \exp[-j(k\vec{r}_m \cdot \hat{R}_{tx} + \delta_n + k\vec{r}_m \cdot \hat{R}_{rx})]\right| \tag{5.9}$$

where $\vec{r}_m$ denotes the location of the mth branch, and $\hat{R}_{tx}$ and $\hat{R}_{rx}$ denote the instantaneous transmit and receive beam directions, respectively. The δ_m term quantifies the added unwrapped phase by each branch that is used to steer the transmitted beam to a desired $\hat{R}_{\widetilde{tx}}$ for a given $\hat{R}_{\widetilde{rx}}$, and was defined in (5.5). In this framework, the problem of minimizing peripheral vision reduces to minimizing (5.9) for a given set of $\hat{R}_{\widetilde{tx}}$, $\hat{R}_{\widetilde{rx}}$, $\hat{R}_{tx}$, $\hat{R}_{rx}$ by varying $\vec{r}_m$. [4]

As an example of the effect of branch position, $\vec{r}_m$, on peripheral vision, the maximum transmitted power over all $\hat{R}_{tx}$ as a function of $\hat{R}_{rx}$ is shown in Fig. 5.7 for both a circular and a square router of 9 branches. The branches in the routers were programmed to transmit at $\hat{R}_{tx} = [\phi_{tx} = 45°, \theta_{tx} = 60°]$, and were intending to receive at $\hat{R}_{rx} = [\phi_{rx} = -45°, \theta_{rx} = 30°]$. Contours shown in Fig. 5.7 correspond to the maximum transmitted power in any $\hat{R}_{tx}$ for the given received direction, $\hat{R}_{rx}$, which is described by a point in the $\phi_{rx} - \theta_{rx}$ plane. Note that the router is programmed to receive a beam in only a desired $\hat{R}_{\widetilde{rx}}$, which corresponds to a single point in the $\phi_{rx} - \theta_{rx}$ plane in Fig. 5.7. Thus, any contours in Fig. 5.7 that lie on points in the $\phi_{rx} - \theta_{rx}$ other than the intended $\hat{R}_{\widetilde{rx}}$ represent power that is being received from directions other than $\hat{R}_{\widetilde{rx}}$ and subsequently routed to some unintended transmit direction. The higher the amount of this power (contour level in Fig. 5.7), the more peripheral vision is present in the router system. The goal of peripheral vision reduction is to minimize the contours in the Fig. 5.7 so that power is only transmitted when the received beam direction is the intended received beam direction $\hat{R}_{\widetilde{rx}}$. To normalize the comparison between the square and circle routers in Fig.

[4] In scenarios where there is a maximum undesired power level that can be transmitted due to the peripheral vision, the minimization can be explicitly stated over all space, for a set $\hat{R}_{\widetilde{tx}}$, $\hat{R}_{\widetilde{rx}}$, as

$$\min_{\vec{r}_m} \oiint_S \left|\vec{E}(\hat{R}_{tx}, \hat{R}_{rx})\right|^2 \mathcal{H}\left(\left|\vec{E}(\hat{R}_{tx}, \hat{R}_{rx})\right|^2 - P_{max}\right) dS_{rx}\, dS_{tx} \tag{5.10}$$

where integration over S, the unit sphere, captures different $\hat{R}_{tx}$, $\hat{R}_{rx}$ directions, $\mathcal{H}$ is the Heaviside operator and P_{max} is the maximum undesired power level.

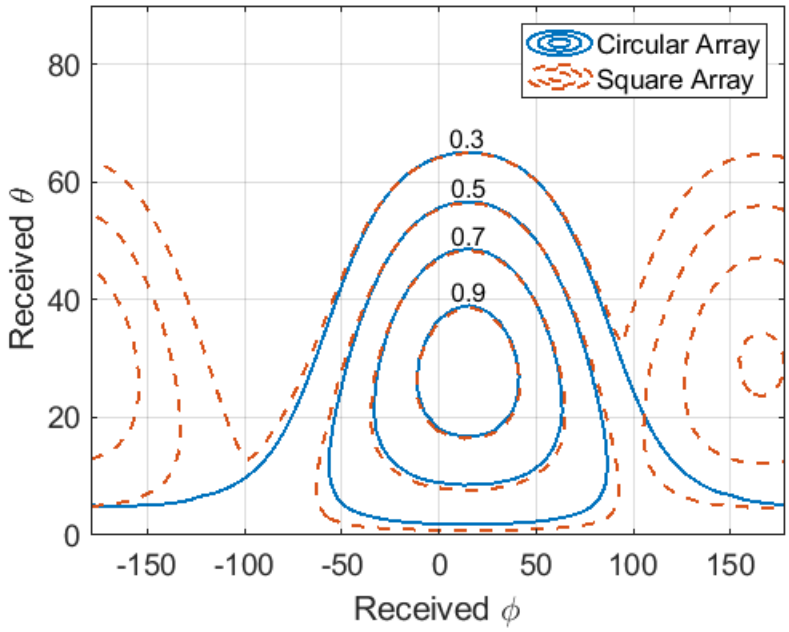

Figure 5.7: Contours showing the normalized maximum transmitted power in any $\hat{R}_{tx}$ as a function of $\hat{R}_{rx}$ for both a circular and square router of 9 branches with a fixed aperture size of λ^2.

5.7, the aperture size of the two arrays are held constant—for a square router with 9 branches and $\lambda/2$ branch pitch, the circular router has $\lambda/2.25$ branch pitch. Router radiative elements are simulated with a $\cos\theta$ element pattern. As can be seen, the circular router has a better peripheral vision rejection, and highlights the importance of branch placement on minimizing peripheral vision. Note that the actual peak of transmitted power does not occur for the intended receive direction, this is due to the effect of the $\cos\theta$ element pattern. More insight into peripheral vision suppression could be obtained by further analysis of (5.10). Fig. 5.8 shows a subset of the above analysis, where instead of finding the maximum transmitted power over all $\hat{R}_{tx}$ as a function of $\hat{R}_{rx}$, the transmitted power in the $\hat{R}_{\widetilde{tx}}$ as a function of $\hat{R}_{rx}$ is shown. This is effectively the amount of undesirable power, save for that from the intended receive signal $\hat{R}_{\widetilde{rx}}$, that is transmitted in the desired transmit direction, $\hat{R}_{\widetilde{tx}}$. Once again, to minimize the peripheral vision we want to minimize the contours in Fig. 5.8 so that power is only transmitted when the received beam direction is the intended received beam direction $\hat{R}_{\widetilde{rx}}$.

Scalable router peripheral vision is relevant when blocker signals may be present. While out of band blockers are filtered by the frequency selectivity of the branch

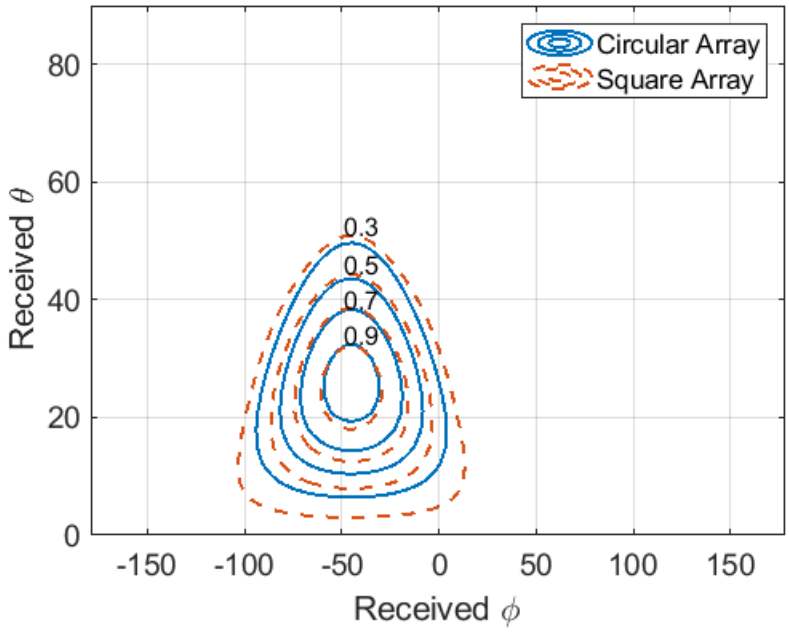

Figure 5.8: Contours showing the normalized transmitted power in $\hat{R}_{\widetilde{tx}}$ as a function of $\hat{R}_{rx}$ for both a circular and square router of 9 branches with a fixed aperture size of λ^2.

antennas and circuits, large in-band blockers could degrade router performance. The lack of centralization in scalable routers has linearity and blocker tolerance advantages compared to a router constructed from conventional arrays. The greatest amplitude for a blocker signal can occur at the centralized summation node in a conventional array which is avoided in scalable routers. If a blocker signal is high enough power to cause non-linear effects within individual branch circuitry the scalable router will exhibit the same signal intermodulation and gain reduction, which occurs in conventional arrays. In addition to frequency and power, blocker angle of arrival is critical to determining its effect. For the standard, fixed pitch, linear array, a blocker arriving from a direction outside of the intended receive direction will be redirected away from the intended target and will be unlikely to cause any negative effects. For a scalable router with suppressed peripheral vision, a blocker arriving from a direction outside the intended receive direction may not be coherently redirected in any direction. A thorough, probabilistic analysis of blocker suppression and redirection could be performed if router geometry and the likely positions of other relevant transmitters and receivers are known. Unsurprisingly, a high power, in-band blocker arriving at the same orientation as the intended receive direction constitutes a worst case scenario, where the router would rely on branch

circuit linearity and the linearity/selectivity of the system it is routing a signal to for successful operation.

Finally, it should be noted that peripheral vision is only a concern for the router architecture of Fig. 5.2(c). For example, in a conventional phased array, such as that shown in Fig. 5.2(b), signal aggregation is done before transmit, and peripheral vision is nonexistent. The hybrid architecture of Fig. 5.2(d) thus results in a lower peripheral vision than Fig. 5.2(c) for the same number of elements. Additionally, as mentioned before, the peripheral vision described above relates to the unwanted coherent combination of the carrier. Peripheral vision where data coherence is maintained is only a subset of the points in the carrier peripheral vision space and is less of an issue for large, spatially distributed arrays operating in wideband networks.

Data Coherence

The scalable router architecture enables the creation of large aperture arrays (which may be contiguous or physically separated). Data coherence degradation is a natural concern for such systems as ISI and beamsquint occur if within each branch phase delay is used instead of true time delay [89]. These effects are more pronounced when the wavelength of the highest frequency components of an incident signal's modulation is comparable to array aperture size. This makes large aperture arrays steering high bandwidth beams most susceptible.

For a given beam direction, pure phase control maintains perfect coherence only at a single frequency. In order to preserve beam coherence in a band of frequencies and prevent ISI, an additional degree of freedom must be added. This can be achieved by controlling the slope of each branch's phase response with respect to frequency (i.e., adjusting group delay). Programmable time delay within each branch unlocks system scalability—the primary motivation for the scalable router. While true time delay enables high bandwidth arrays, the additional degree of freedom it affords can alternatively be used to simultaneously and independently control two separate, full power beams. Dual beam capability is further explained and demonstrated in the first implementation section of this work.

Noise Implications

To live up to the scalability potential of the distributed router architecture, the highly complex branch circuits must be manufacturable at a low cost and high

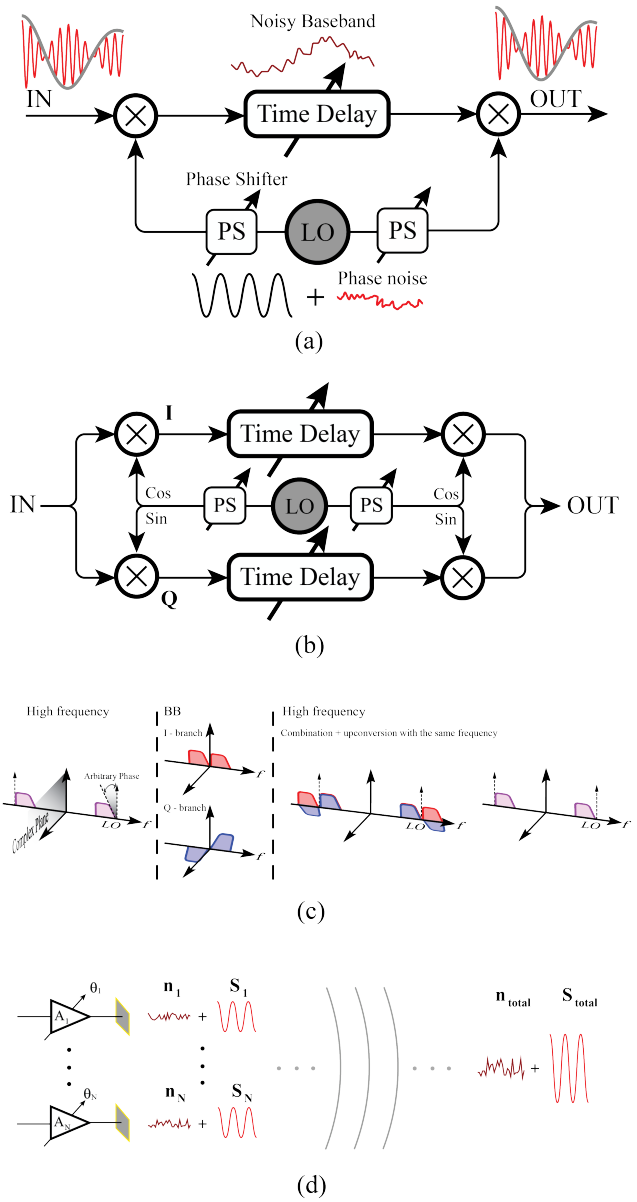

Figure 5.9: System architecture (a) A typical branch implementation includes down-conversion and up-conversion of the signal by an LO tone and the application of true time delay. (b) To avoid the image issues associated with single sideband mixing, separate I/Q paths may be used. (c) Visual representation of image rejection by the I/Q architecture. (d) Because the scalable router branches are independent and decentralized, the noise added within each branch is uncorrelated.

volumes. Integrated circuits processes, especially general purpose complementary metal oxide semiconductor (CMOS), can deliver this complexity and volume at an attractive cost. An integrated circuit based scalable router also has the potential to reduce implementation cost and printed circuit board complexity when compared to a conventional two-phased array relay. Without a centralization node, the receive and transmit circuitry can be combined within a single integrated circuit die. An integrated circuit-based implementation also reduces the marginal cost of additional circuits (such as programmable time delay) needed for a scalable router.

While an integrated circuit implementation has a lot of advantages, it presents a challenge to achieving programmable true time delay with wide range and high resolution at microwave frequencies. Hence, it is preferable to down-convert the received microwave signal and apply true time delay at lower frequencies. This architecture is shown in Fig. 5.9(a). To suppress the signal image, the architecture can incorporate an in-phase/quadrature (I/Q) scheme as depicted in Fig. 5.9(b) and (c).

It is noteworthy that the local oscillator (LO) within each branch of the scalable router does not have to be phase or frequency locked to the data carrier frequency or other branches. Also, the down-conversion and up-conversion branch architecture serendipitously suppresses the effect of phase noise in the branch LO. To understand this effect, consider an input to the I/Q branch as:

$$X(t) = I(t)cos(2\pi f_0 t) + Q(t)sin(2\pi f_0 t), \tag{5.11}$$

where f_0 is the incoming wave frequency. After down-conversion, we would have:

$$\begin{aligned} X_I &= \frac{I(t)}{2}cos[2\pi(f_{LO} - f_0)t + \phi_{LO}(t) + \phi_{PR}] \\ &\quad - \frac{Q(t)}{2}sin[2\pi(f_{LO} - f_0)t + \phi_{LO}(t) + \phi_{PR}] \\ X_Q &= \frac{I(t)}{2}sin[2\pi(f_{LO} - f_0)t + \phi_{LO}(t) + \phi_{PR}] \\ &\quad + \frac{Q(t)}{2}cos[2\pi(f_{LO} - f_0)t + \phi_{LO}(t) + \phi_{PR}], \end{aligned} \tag{5.12}$$

here f_{LO} is the local oscillator frequency, $\phi_{LO}(t)$ is the associated phase noise, and ϕ_{PR} is the applied phase shift[5]. The output signal after applying true time delay and

[5]It is implicitly assumed in (5.12) that the phase noise of the I and Q are correlated. This is generally a valid assumption if they are generated within the same oscillator core.

up-converting is:

$$\begin{aligned} X_{out} = \frac{I(t-\tau)}{2} cos[2\pi f_0(t-\tau) - \phi_{PR} \\ + \phi_{LO}(t-\tau) - \phi_{LO}(t)] \\ + \frac{Q(t-\tau)}{2} sin[2\pi f_0(t-\tau) - \phi_{PR} \\ + \phi_{LO}(t-\tau) - \phi_{LO}(t)]. \end{aligned} \tag{5.13}$$

As above equation shows the transmitted frequency is exactly at f_0. This result is independent of each branch local oscillator frequency (f_{LO}). Since the delay, (τ), is on the order of pico-seconds to few nano-seconds, the resultant additional phase noise, due to the term $\phi_{LO}(t-\tau)-\phi_{LO}(t)$, is negligible up to offset frequencies in the giga-hertz range. This near-complete phase noise cancellation bolsters the scalable routers potential for distributed operation as inexpensive reference oscillators with relaxed stability (such as cheap crystal oscillators or on-chip free-running voltage controlled oscillator (VCO)) may be used. A measurement of the implemented branch circuit shown in Fig. 5.10 demonstrates this phenomenon clearly. A tone at 25.01 GHz is sent through a branch of the IC (whose details will be discussed in the next section) with a free running VCO around 24.98 GHz. The image at 24.95 GHz exhibits twice the phase noise of the VCO while the re-transmitted tone has a clean spectrum without any of the VCO phase noise. In addition to phase noise cancellation, the scalable router architecture also provides mitigation of added amplitude noise within each branch. Due to the transceiver branches being fully separate and the absence of any kind of physical summation node within the system, the added noises (antenna noise temperature, noise added by amplifiers, etc...) are uncorrelated. This lack of correlation due to decentralization, shown in Fig. 5.9(d), results in higher SNR at the target compared to a relay constructed from traditional, centralized arrays. This noise reduction can be leveraged to trade component level noise performance for other system benefits, for example, reduction of capacitance in a switched capacitor filter to increase bandwidth and reduce on-chip area at the cost of added uncorrelated noise. Decentralization can also help reduce the effect of delay and phase shift quantization noise or setting errors. Uncorrelated stochastic variations or deterministic errors in individual branches are incoherently combined in the transmitted beam of the router, blunting their effect.

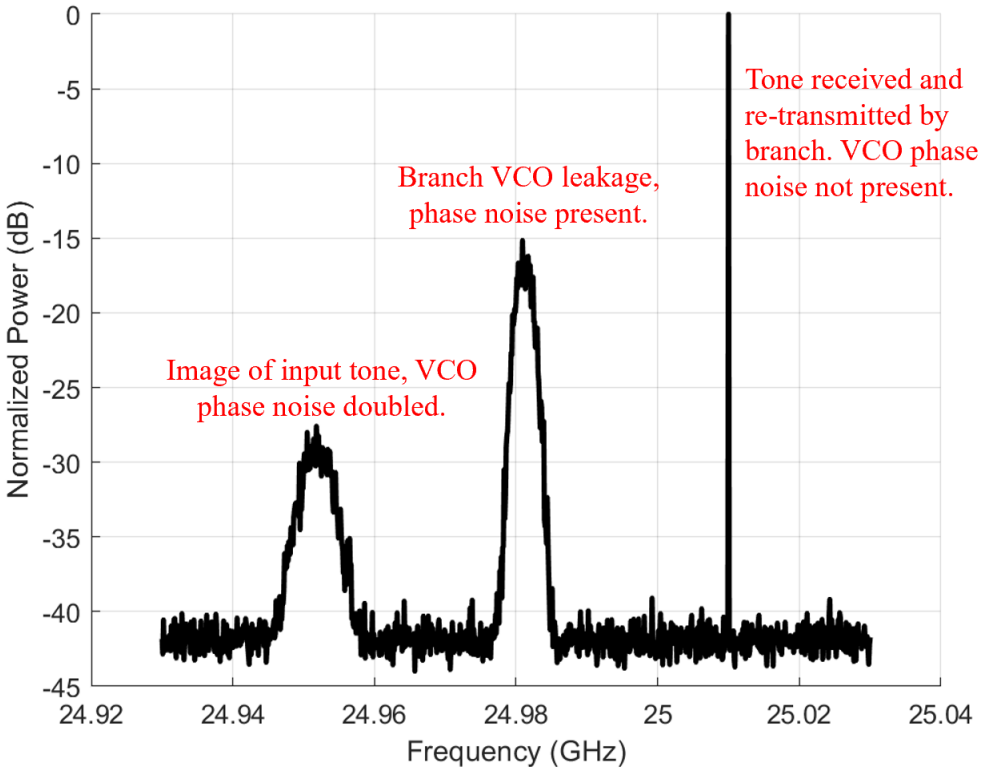

Figure 5.10: Radiative measurement of output spectrum of a branch with free running VCO and excited by an external source. The branch is digitally configured to maximize VCO leakage through up-conversion mixer for better observation of phase noise cancellation.

Branch Isolation and Self-Interference

Scalable routers are not immune to the self-interference issues that plague many simultaneous transmision/reception (full-duplex) systems. In particular, parasitic feedback from the transmitter output back to the input of the receive chain interferes with system function even if it is far below the levels necessary to cause oscillation. Consider the simplified, frequency independent branch with forward path gain α and parasitic feedback β shown in Fig. 5.11. The delay element within the branch is an ideal constant amplitude phase rotator. The branch's closed loop transfer function phase and normalized amplitude are plotted vs. phase rotator setting for several open loop gains ($\alpha * \beta$) in Fig. 5.11. The parasitic feedback introduces non-idealities to the previously ideal phase rotator. An open loop gain of -20 dB produces nearly ideal behavior, but non-ideality quickly emerges as this gain rises. The dependence of these non-idealities on open loop gain can be observed in Fig. 5.12. Fig. 5.12 shows the amplitude and phase error (deviation from the

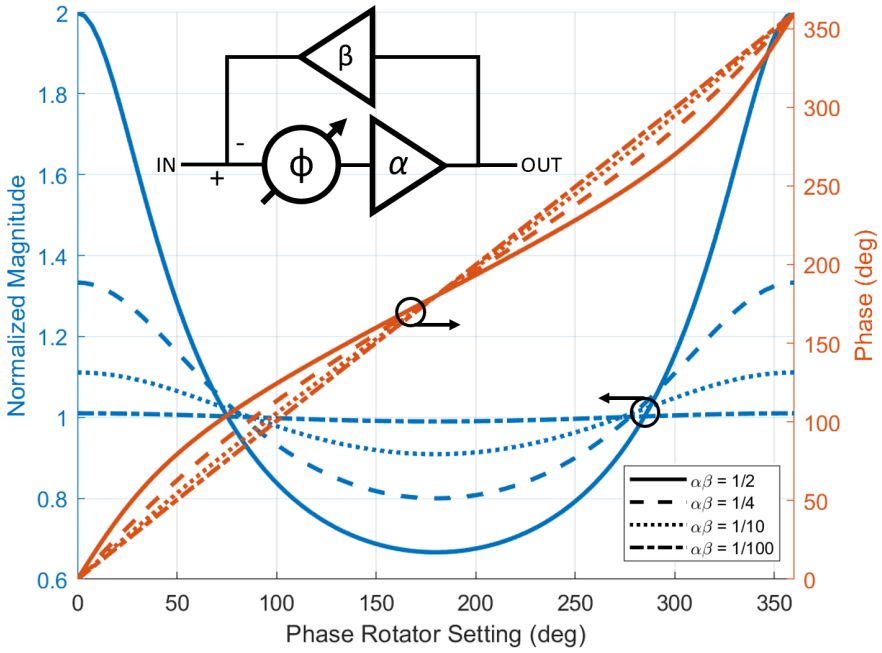

Figure 5.11: Simplified branch transfer function phase and amplitude as ideal, unity gain phase rotation occurs under presence of parasitic feedback

ideal) as open loop gain is changed. Even at an open loop gain of 0.1, the peak phase error exceeds 5°. These non-idealities limit the achievable branch forward path gain, as a gain of 30 dB would require isolation of close to 50 dB for the peak amplitude and phase variations to be rendered unnoticeable. Polarization isolation, isolating radiators on opposite sides of a ground plane, or active feedback cancellation techniques can reduce parasitic feedback to acceptable levels. Provided the branches are implemented by integrated circuits, the additional complexity of active cancellation circuits comes at low marginal cost. While coupling between adjacent branches may also be a concern, the isolation within a branch is likely to be worse than the isolation between branches of even a dense (0.5λ pitch) array.

5.3 First Implementation

The first implementation is done on the CMOS process with phase shifters and a true-time-delay unite. CMOS ICs are suitable for scalable routers as they not only compactly combine the many digital and analog functions needed for array operation but also offer low cost at high volume. Favorable cost scaling is crucial for the scalable router architecture as cost is as formidable an obstacle as any circuit level performance specification for widespread adoption of arrays of hundreds or thousands of elements. The primary function of this implementation is to provide

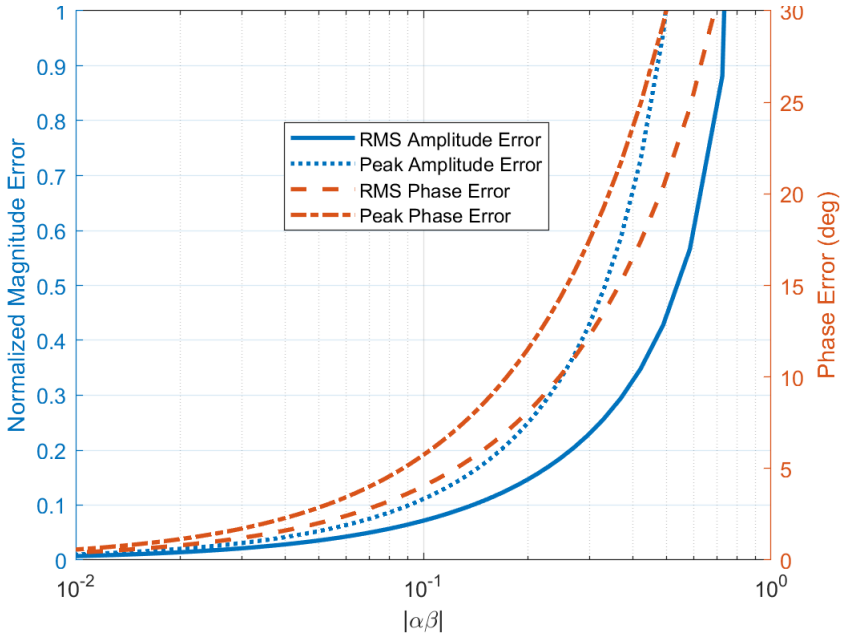

Figure 5.12: Peak and RMS transfer function phase and amplitude error vs the open loop gain of the simplified branch model.

amplification and programmable true time delay to its received signal as well as a demonstration to show its practicality and functionality. Here, we presented a hybrid analog/digital true-time-delay unit to generate enough true-time-delay for meter-scale arrays (HTDU). The branch architecture is shown in Fig. 5.13(a). A branch begins with a low noise amplifier (LNA) with simulated noise figure of about 4 dB followed by an I/Q down-conversion mixer. The I/Q mixer's LO is generated on chip by having a free running voltage-controlled-oscillator (VCO) (which is not locked to a specific synchronization clock signal). The LO signals pass through vector sum phase rotators that provide 360° phase control. Variable gain amplifiers condition the baseband I/Q signals before they are sent to the HTDU. After the delay unit, the I/Q signals are up-converted using similar LO signals that undergo independently controlled phase rotation. The up-converted I/Q signals are recombined at RF in a vector summer and transmitted by the driver and the PA. Fig. 5.13(b)-(c) shows LNA input matching, receiver chain IIP3, and system gain, output power, and compression curves. The die micrograph of the branch IC is shown in Fig 5.14. Time delay with fine resolution control and broad variation range is critical to the decentralized router's scalability. Large scale timed arrays are challenging to implement because many existing integrated delays are power hungry, area inefficient, or lacking sufficient range and resolution. Non-IC solutions, such

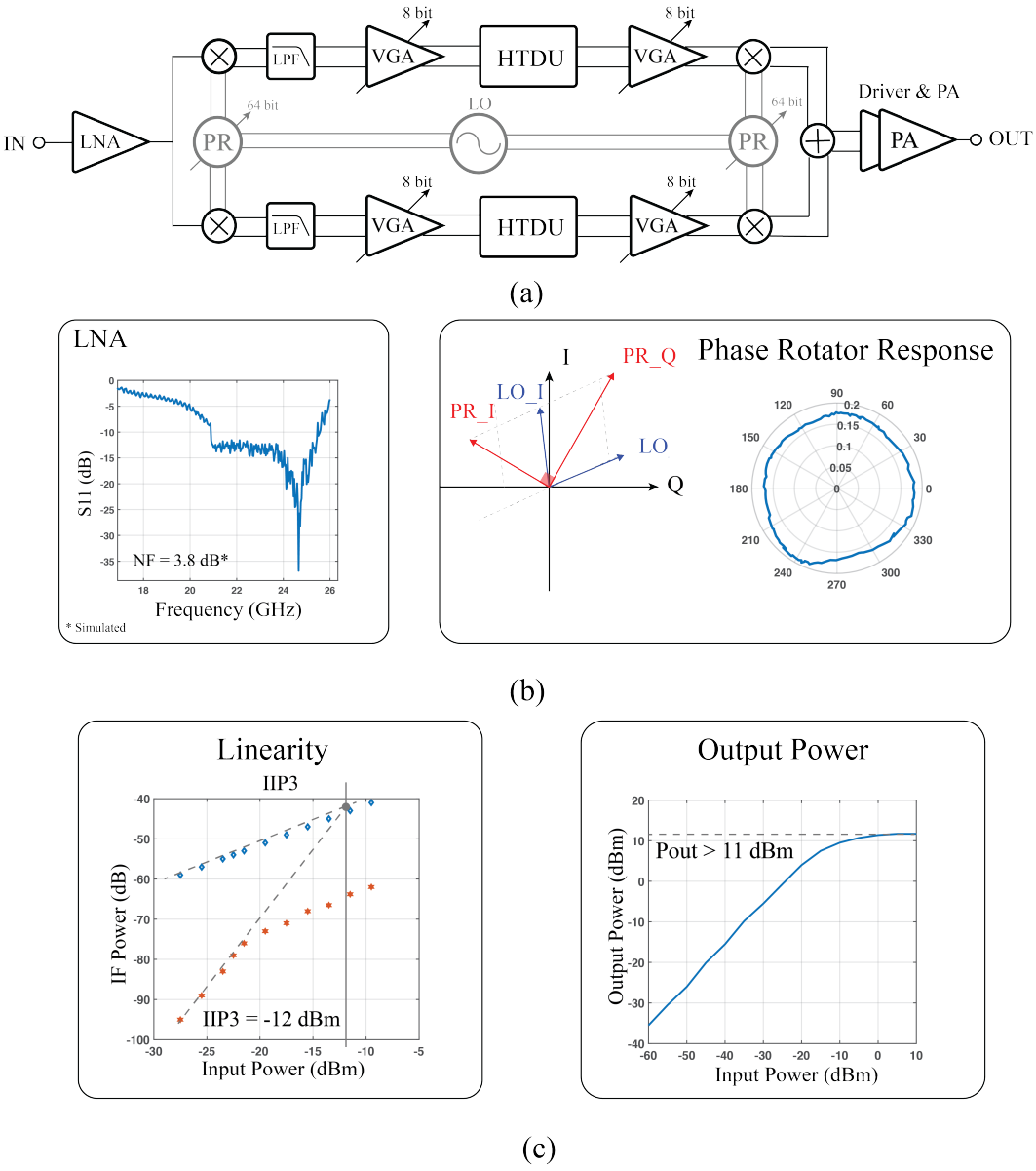

Figure 5.13: System architecture and block measurements (a) Branch integrated circuit architecture, (b) LNA matching and Vector sum phase rotator performance, (c) System linearity and PA output power.

as switchable transmission lines or optical delays, are physically and economically impractical for arrays with hundreds or thousands of elements.

The scalable router HTDU performance requirements are determined by the physical array aperture size and bandwidth requirements of the intended application. For instance, in a 1 m aperture span array with 500 MHz baseband bandwidth, two

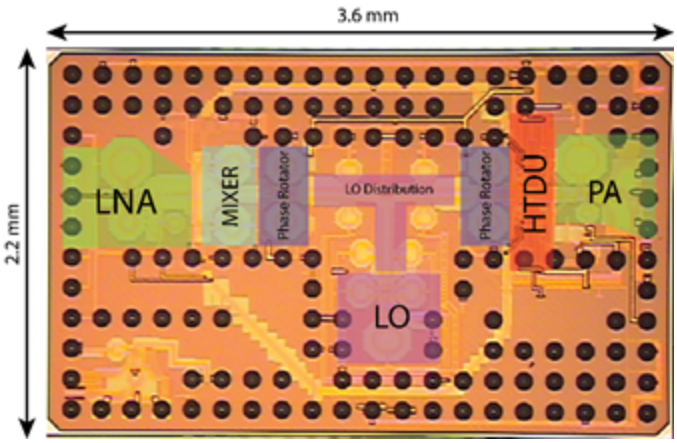

Figure 5.14: Die photo of branch integrated circuit implemented in standard 65nm CMOS process.

elements on opposite sides of the array may experience time offset of up to 3 ns while requiring delay adjustment resolution of <20 ps for temporal alignment within 1% of the maximum data frequency content. The 3 ns delay range requirement is larger than existing analog delay solutions [100]–[102] while the <20 ps resolution is challenging in purely digital solutions due to unrealistic digital clock requirements [103]. This work uses a hybrid analog/digital switched capacitor delay unit, which fulfils both requirements. A hybrid analog/digital delay unit concept intended for large array applications was recently presented in [104]. Any switched capacitor circuit can be considered as a time delay. The signal is sampled onto the capacitor by a clock edge delivered to the input switch and accessed later by a subsequent clock edge. By controlling the delay between the sample and access clock edges, the delay of the circuit can be changed. To provide sufficient sample rate for the input signal bandwidth, multiple switched capacitors are placed in parallel in an n-path configuration. The input and output switches are controlled by two separate non-overlapping clock generators (NOCs). The time delay is controlled by the relative phase of the two NOCs. The HTDU (schematic shown in Fig. 5.15) uses 8 switched capacitors in parallel and has fine, medium, and coarse control of the relative phase of the NOCs. The fine and medium control change the phase of the clock driving the output NOC. The fine control uses a DAC to change the bias of a chain of current starved inverters carrying the clock signal. The medium control adds or removes inverters to/from the output clock signal path. The coarse control changes the location of the pulse in the output NOC. The delay element measurements in Fig. 5.16 show 5 ns of range with a minimum step size of 5 ps for the fine step.

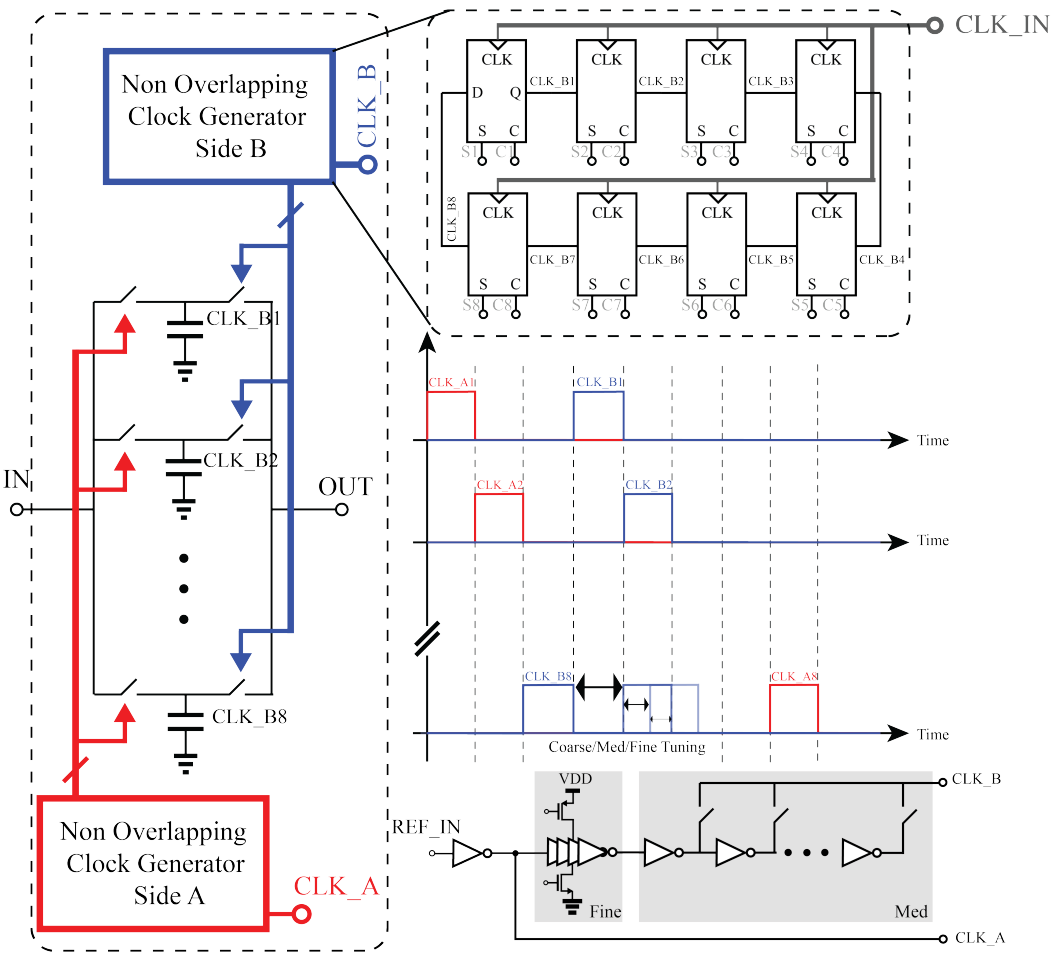

Figure 5.15: Hybrid analog/digital time delay unit (HTDU). Coarse delay is controlled by the initial set/reset state of the NOCs. Medium delay adds or removes inverters and fine delay changes a voltage controlled delay line.

Minimum step is defined as the smallest step that the non-linear fine delay setting control could be linearized to while still utilizing its full range. The measured delays are determined by fitting a line to the measured phase response and taking its slope (group delay). Digital code "0" for the coarse delay represents a state when the clock pulse for reading from the delay capacitors overlaps the writing clock pulse as such digital code "1" is used for normalization instead. Simulation/analysis curves for coarse, medium, and fine delay are based on extracted delay cell simulations from which the delay range is calculated rather than full system top level simulations. The coarse delay step and range can be increased or decreased by adjusting the

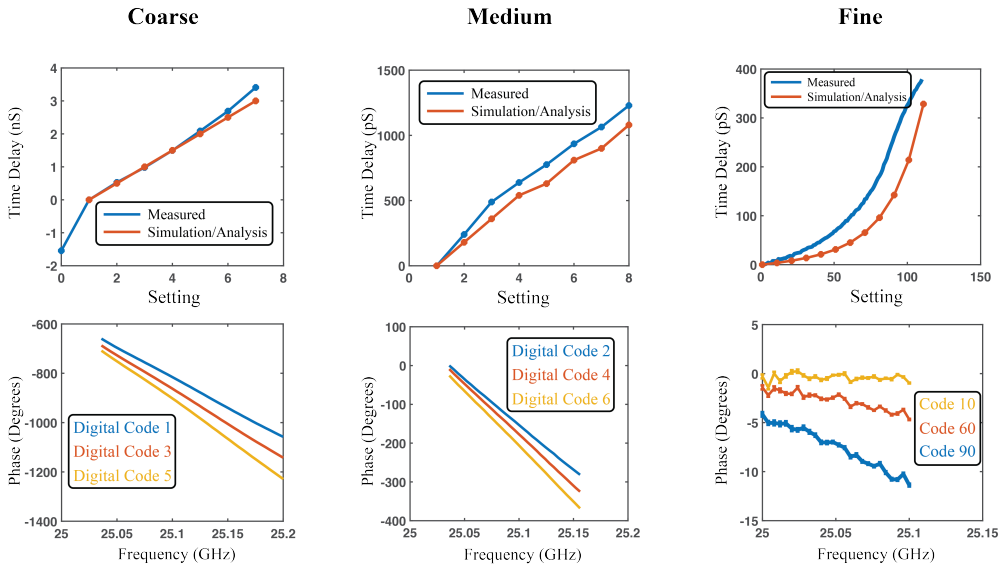

Figure 5.16: Coarse, medium, and fine true time delays and measured phase responses for span of coarse, medium, and fine delay settings. All measurements are referenced to the lowest delay setting.

NOC clock frequency, with the maximum achievable 10 ns of delay demonstrated in the radiative measurements. In our implementation, the input clock can operate in any frequency from 650 MHz to 4 GHz. The clock frequency of 650 MHz (corresponding to 10 ns of delay) is the minimum value for which all the delays in the range can be generated without a gap. A radiative, 4-branch, receive and transmit capable, scalable router prototype was built. Each branch IC is mounted on a printed circuit board (PCB) with orthogonally polarized patch antennas. The scalable router is formed using a number of these branch PCBs arranged in the desired spatial configuration. The branch circuit board, the simulated patch antenna S-parameters (input matching and isolation), and radiation pattern are presented in Fig. 5.17. The simulated isolation between antennas is ~50 dB—high enough to not induce significant feedback effects.

The scalable router architecture's potential in large scale array applications where there is no shared timing reference between branches is demonstrated in the test set-up depicted in Fig. 5.18. The branches are placed in two pairs separated by 1.5 m. A transmit/receive horn antenna pair is placed 1 m from the leading branch pair. This transmit and receive horn antenna pair is used to excite the router and measure its re-radiated beam. The total round-trip path length difference between the front pair and back pair is approximately 3 m, which corresponds to 10 ns of delay. The

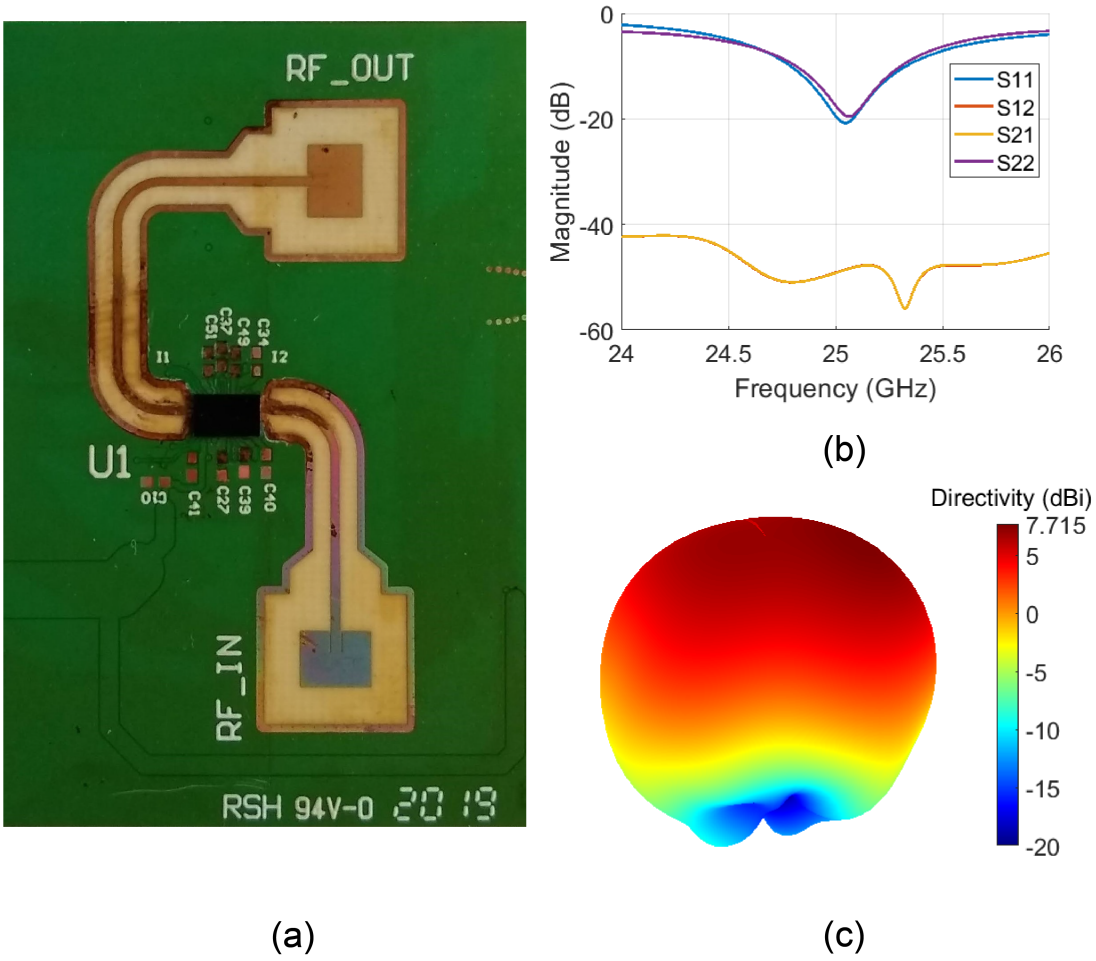

Figure 5.17: Measurement setup (a) Fabricated branch transceiver PCB for radiative measurements of the designed branch integrated circuit. The TX and RX antennas are orthogonally polarized. (b) The simulated S-Parameters of the PCB. (c) The simulated radiation pattern of the designed patch antenna.

branch circuits share *no* timing information and use internal free running VCOs to provide the LO signal for the circuit.

To illustrate the routers functionality and the importance of true time delay for the scalable router, two digital configurations of the router were measured. The first configuration steers the routed beam using only the phase rotators. This matches the elements' phases at only a single frequency point. The 10 ns delay mismatch between the branch pairs causes a difference in group delay (slope of the phase response) and prevents coherent combination of the branches' signals outside of a narrow bandwidth. The second configuration uses true time delay in addition to the phase rotators to match the branch phase responses over a frequency band. By

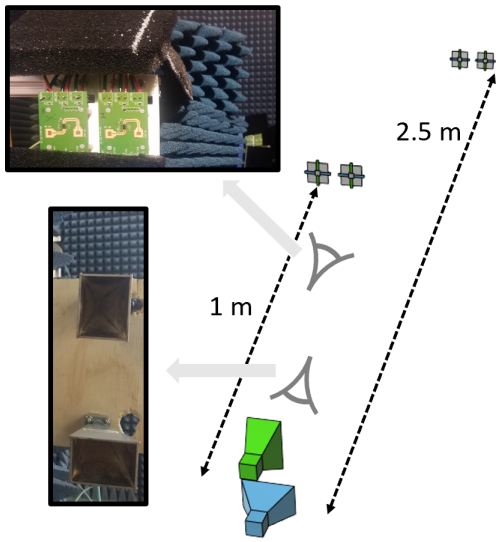

Figure 5.18: Radiative scalable router test set-up. Two pairs of branch circuits are excited by a horn antenna and their re-radiation is measured by the other horn antenna. The branches are not colocated and do not share a timing/phase reference.

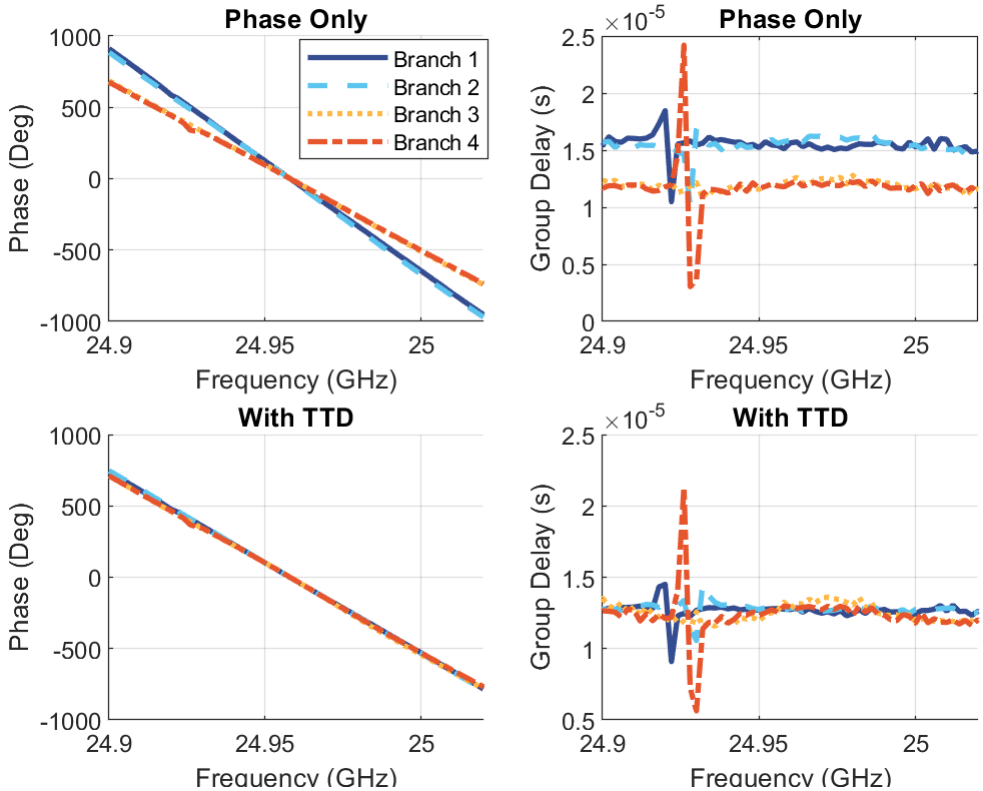

Figure 5.19: Radiatively measured branch phase response and group delays with and without true time delay correction.

connecting a vector network analyzer (VNA) to the transmit and receive antennas of the set-up, the response of each branch can be measured individually. Fig. 5.19 shows the measured phase response and group delay of each branch with and without true time delay correction. The measurements with TTD clearly illustrate matched phase and group delay for all four branches, demonstrating the true time delay adjustment capability of the branch circuit. The peaks in group delay are caused by LO leakage of each branch circuit.

The coherence restored by the branch circuit true time delay is critical for the transmission of data by large arrays. Without this correction, ISI degrades the re-routed data. The same test set-up and branch configurations discussed above were excited by a 24.96 GHz signal modulated with BPSK, 16-QAM and 64-QAM data streams at 45 Msps. The re-routed signals were measured and demodulated. No equalization was used in the measurement setup. The BPSK eye diagram and 16-QAM and 64-QAM constellations and results of the demodulation are shown in Fig. 5.20. The images on the left hand side correspond to phase-only steering while the images on the right show the results with combined phase and time delay steering. The addition of true time delay noticeably improves the BPSK eye diagram and improves its EVM from 11.4% to 5.2%, while the 16-QAM EVM is improved from 8.6% to 4.4%. The 64-QAM constellation is changed from nearly unrecognizable with phase only steering to an EVM of 4% with phase and time delay steering[6].
While 45 Msps is sufficient to observe the importance of true time delay correction, it is even more critical at higher modulation rates. The bandwidth of the presented system is limited by unintentional down-tuning of the branch LOs in the branch integrated circuit.

Dual Beam Demonstration

The true time delay capability within each branch can also be used to independently steer beams at two different frequencies. A phased array forms a beam when the phases of the signals radiated by each element in the array match in the desired direction, creating constructive interference. A dual beam array requires this constructive interference to occur in two desired and potentially arbitrary directions at two different frequencies. Programmable true time delay allows for the phase response slope (group delay) of an element to be changed, while a programmable

[6]The 4% EVM may be slightly optimistic as several points at the edges of the constellation may be misidentified as the incorrect symbol. Despite this, the improvement provided by true time delay is undeniable.

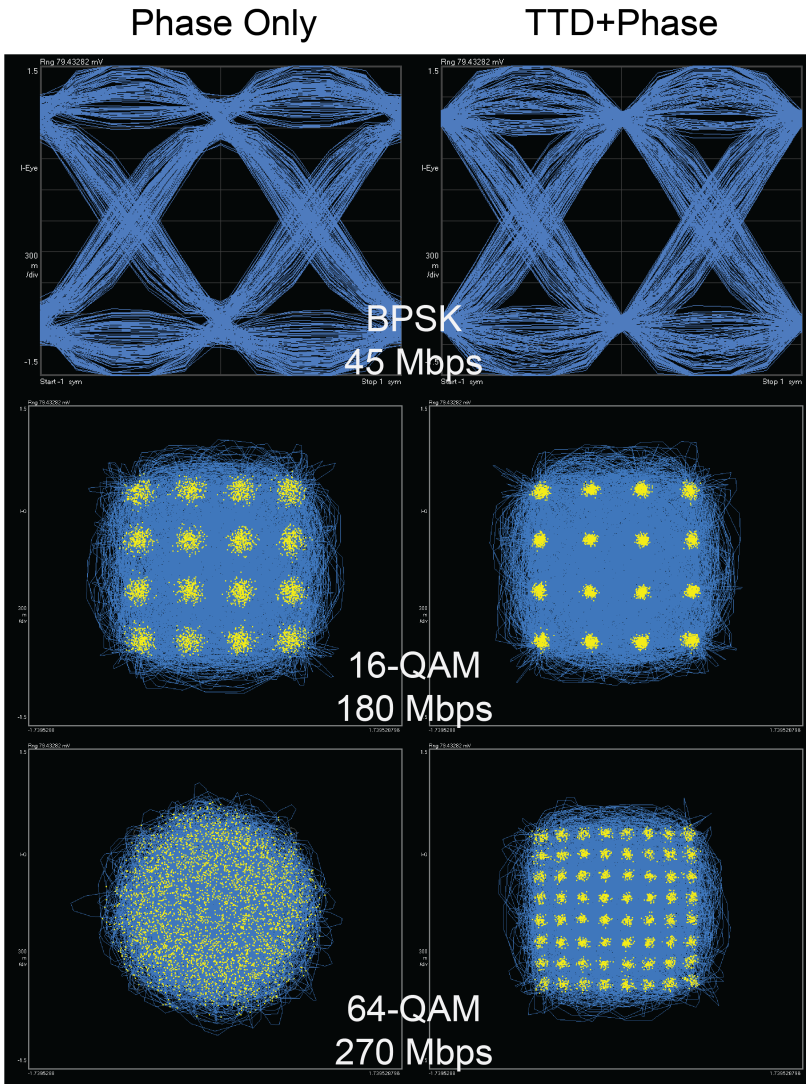

Figure 5.20: Received eye diagrams/constellations for BPSK, 16-QAM and 64-QAM at 45 Msps with beamforming achieved via phase-only steering (left side) or true time delay (TTD) and phase steering (right side).

phase rotator changes the phase response offset or intercept. In the previous measurement, we used these two degrees of freedom to match the offset and slope of multiple branches over a band of frequencies to prevent ISI, but they can also be used to match the phase response of the branches at one frequency in one direction and another frequency in another direction. This, in effect, creates two independently controlled full power beams from the array.

To demonstrate the dual beam capability, a test setup with 4 elements was built. A 4 branch scalable router is radiatively excited by an antenna connected to one port of a VNA, the re-routed signal is measured by an antenna which is connected to the

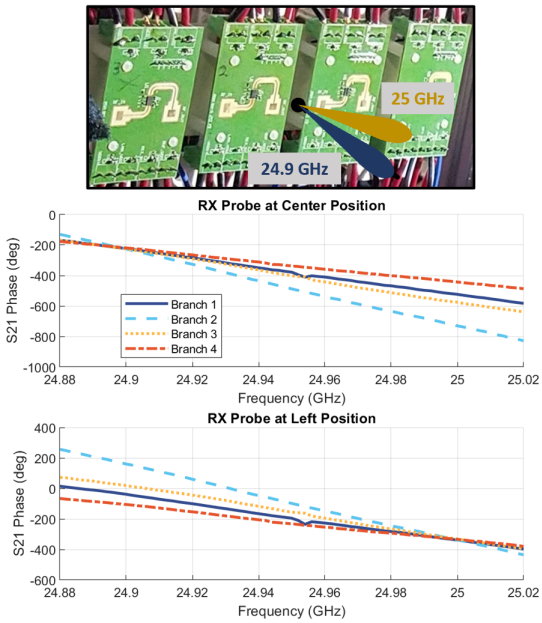

Figure 5.21: Branch phase responses measured at the center and left RX probe positions for the router digital configuration which steers 24.9 GHz to the center and 25 GHz left.

other port of the VNA and mounted on a linear scanning platform. The dual beam capability of the scalable router is demonstrated by keeping the first beam, which is at 24.9 GHz, at the broadside while we steer the second beam, at 25 GHz, to three different directions. The steering positions are separated by 5 cm (close to 5° off the broadside direction) and are chosen in order to stay within the grating-lobes caused by the transmit antenna pitch of 5.5 cm.

Fig. 5.21 shows the measured S_{21} phase for all branches for the configuration where the 24.9 GHz beam is steered broadside and the 25 GHz beam is steered left. The phase is measured at two locations: broadside and the position corresponding to the left steered beam. The constructive interference responsible for beamforming is evident by the matched phase for all elements at 24.9 GHz for the center probe position and at 25 GHz for the left probe position. Fig. 5.22 shows successful steering of the beam at 25 GHz while the broadside beam at 24.9 GHz stays constant. The left and right steered traces have been trimmed to prevent grating lobes (at ~15° away) from appearing on the opposite side. Static reflections caused by other objects near the set-up were measured separately and subtracted from the presented results. In order to arbitrarily steer beams at two frequencies the branch circuits must be

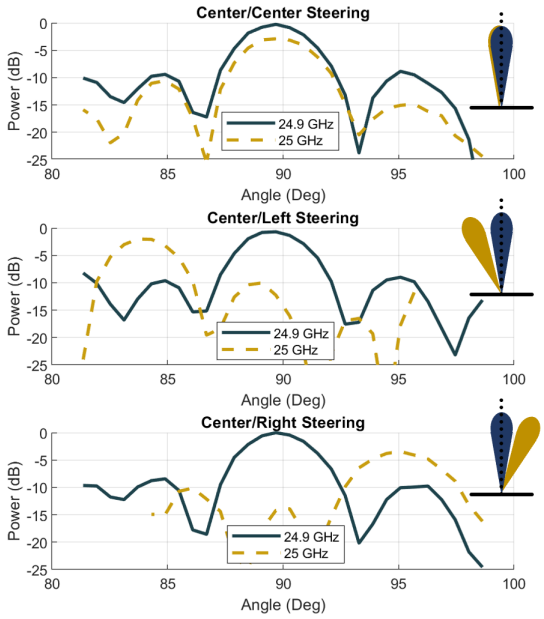

Figure 5.22: Measured patterns at 24.9 and 25 GHz for three different steering configurations. Demonstrates simultaneous, independent control of two receive/transmit beam pairs. The powers are normalized to the same global maximum occuring on the 24.9 GHz beam during center/center steering.

able to change their relative phase to any value from 0 to 360°. Since our system accomplishes this relative change using time delay, the period of the minimum frequency separation of two frequencies that can be fully independently steered is the maximum achievable time delay. Thus 100 MHz separation is the smallest achievable for 10 ns of delay control.

Given the obvious advantages of a single array serving multiple users, multi-beam microwave communication systems have been an active area of research for several decades. An overview of state of the art multi-beam approaches as of 2017 can be found in [105]. When considering the wide variety of approaches and subsequent trade-offs, direct comparison between systems is not always apt. Dividing a larger array into independent sub-arrays is a common technique usable with no additional hardware but it divides power and aperture between the beams. Also, smaller aperture for each beam translates to broader beam for each user which might not be suitable for some applications. The following array hardware paradigms (as well as the scalable router dual beam capability) achieve multiple beams without sacrificing power in the beams.

A well established family of multi-beam systems is multi-port passive (or semi-active) networks used to create a predetermined set of beam patterns. These arrays

can be transmit or receive, have been fully integrated [106], and can create a multitude of beams at the cost of design complexity. However, these arrays are not electronically steerable and require separate input drivers to achieve their multi-beam capability. Moreover, these designs are vulnerable to scalability and simulation time. Another common family of multi-beam arrays are the digital arrays, which process the same received signals in several parallel channels [103], [107]. While these systems can create as many steerable beams as processing power and time are available, the topology has only been shown for receive arrays, not transmit arrays. Furthermore, there can be dynamic range limitations due to the analog-to-digital conversion process. While digital processing is potent, it introduces its challenges. This can manifest itself in speed limitations, scalability, and complexity.

The scalable router dual-beam capability (enabled by programmable time delay) differs from the previously described paradigms as it derives its two beams by "frequency multiplexing" the array. In most previous demonstrations, all beams were at one frequency, and there was not any frequency multiplexing. By tuning the phase response of the element's at two frequencies, two independent beams are created. It should be noted that transmit dual beam capability is not unique to the scalable router architecture. While programmable true time delay means the router is naturally suited to the task, any transmit array with independent phase and group delay control within each element could achieve this task. Because this control is established through analog circuits at baseband it can be used for transmit or receive arrays. Programmable time delay only grants a second beam to control, but additional degrees of freedom for controlling the element phase response could added. This includes any extra group delay controls, dispersion engineering, and utilizing a non-linear phase response.

While each approach has distinct disadvantages, a future multi-beam paradigm could provide many full-power, electronically steerable, receive and transmit capable beams. Introducing additional degrees of freedom to the element phase response (building on the technique presented in this work) is a promising pathway to fully capable multi-beam arrays.

5.4 Second Implementation

For this new router architecture to be fully adapted into the currently active areas, including 5G and unmanned aerial routers, as well as the future standards (6G and beyond), some challenges have to be addressed. To name but a few: 1-Increasing the number of users in a scalable fashion, 2-Having less complicated and power efficient

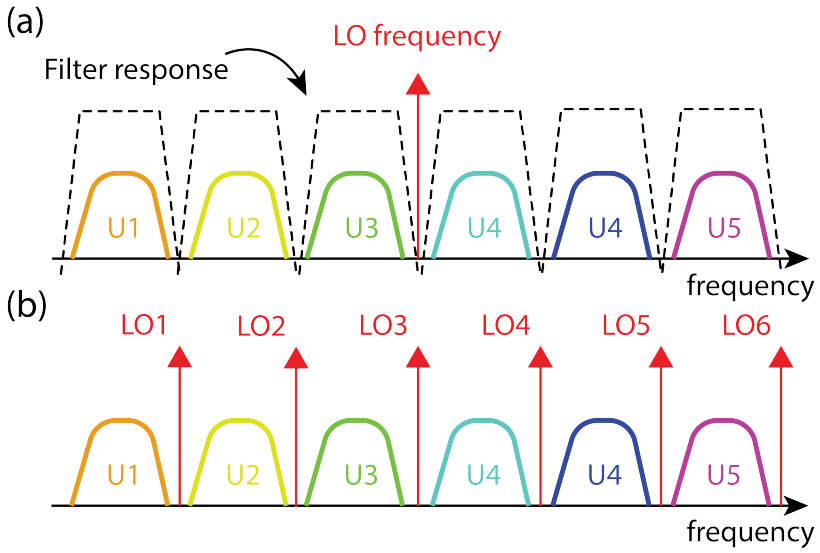

Figure 5.23: Multi-user architectures (a) Multi-channel approach with corresponding filtering needed in order to separate the channels. (b) Multi-oscillator approach with the same base-band architecture.

base-band architecture, 3-Addressing self-interference and feedback phenomena in each branch.

Here I am going to introduce some new possible architectures to mitigate some of these challenges. We also designed and fabricated a new chip with two branches with 3-channel architecture operating at 28 GHz, which will be discussed later.

Increasing the number of users

As mentioned before, providing a multi-beam and multi-user decentralized router without sacrificing the total power and the beam size is beneficial in many ways. This is in line with later use of these devices in wireless communication systems. To expand the number of connections in a specific area (increasing the connection density), reusing the frequency band in micro-cells and pico-cells, and providing directive beams for each user, different solutions and architectures are offered [108], [109]. Moreover, there have been a tremendous effort toward designing and simulating antennas for 5G and MIMO applications [110], [111]. However, phased arrays are always considered to be a viable solution for all these applications [112]. There are two main paths to achieve this goal: 1. Multi-channel architecture, 2. Multi-oscillator architecture (Fig 5.23). The former solution works mostly based on conventional multi-channel transmitters/receivers. However, the frequency bands, the amount of digital processing needed for each channel, isolation/rejection be-

tween channels, and linearity/noise requirements are dictated by different standards and applications. Any of these requirements can change the architecture dramatically. For instance, required bandwidth and frequency separation between different channels (users) impose various practical limitations, specifically on the circuit level. On this matter, filtering of channels can happen in three different frequency bands: 1. Radio frequency, (RF) 2. Intermediate frequency (IF), 3. Base band (BB). Any of these options have their pros and cons and should be chosen based on a specific application or requirements.

Furthermore, the multi-channel architecture can allow for a more straightforward base-band by combining multiple tasks into a single unit. As we will see later in this section, we can merge the filtering and phase shifting into one N-path filter unit. Therefore, this architecture can lead to a more power-efficient system. On the other hand, this idea will suffer from the fact that it should be designed for an exact number of users and the number of channels cannot be expanded easily.

Multi-oscillator approach however, provides a simpler solution to increase the number of users. This architecture can benefit from having the same base-band design for all the users and maximizes the isolation between two adjacent channels. Indeed, the mixing and shifting the frequency can happen in all mentioned bands in order to optimize the power consumption, linearity, and area on chip (since most of the components at higher frequency require inductors and transmission lines). Moreover, multi-oscillator architecture opens up a space for micro-wave and RF combs in scalable router realm. There have been multiple studies toward creating efficient and robust electrical combs [113], [114]. Most of the works in this area have tried to adapt novel electrical systems to the existing high-accuracy optical combs or using a high quality factor micro-resonator [115].

Power efficiency, system complexity, and channel cross-talk are some of the challenges this approach faces with. However, given a specific application, some of these specifications can be relaxed.

Less complexity in base-band

As, discussed before, multi-channel approach can increase the total number of users as well as having a simpler base-band architecture. Less complex base-band can be achieved by merging different blocks into one block with multiple tasks which also can be a more power efficient solution. In order to observe this opportunity, we can take a quick look at the hybrid analog/digital true-time-delay unit. This unit is a specific type of N-path filter. A typical N-path filter consists of two mixers and a

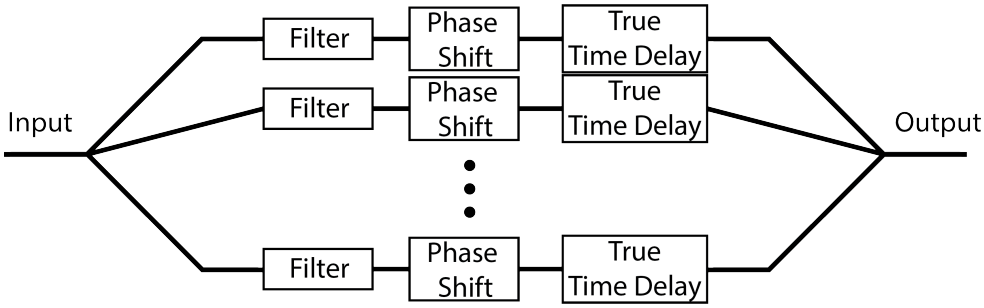

Figure 5.24: Multi-channel structure in scalable routers.

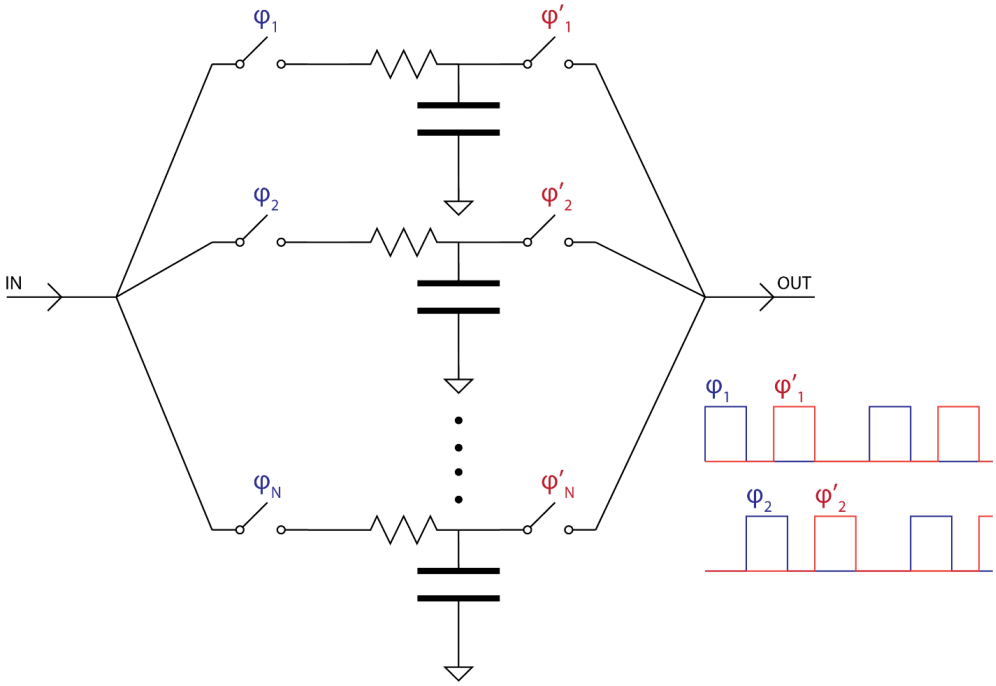

Figure 5.25: True-time delay N-path filter.

filter between them. In a typical N-path filter, the profile of the filter will move up in frequency and different filter responses can be implemented. Furthermore, the relative phase between two LO signals manifests itself at the output phase in two different ways: 1. phase shift 2. true time delay.

As it is shown in Fig 5.25 and Fig 5.26, if there is a continuous electrical connection between the input and the output of the N-path filter, the relative phases between the LOs manifest itself as a true-time delay—also known as group delays. On the other hand, if a noncontinuous N-path filter—with extra phases to achieve different filter

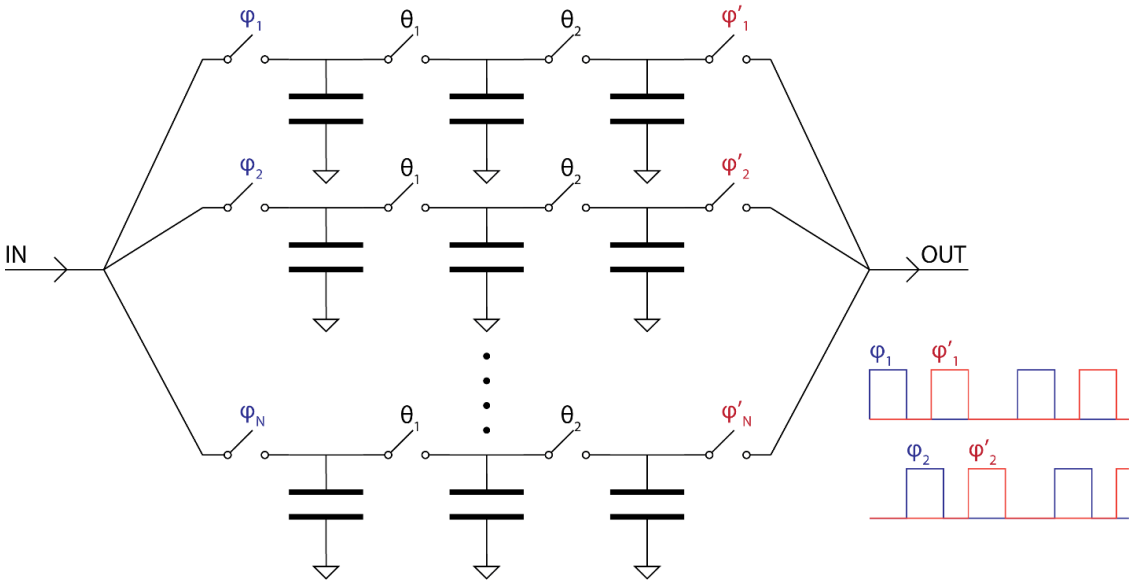

Figure 5.26: Phase shifting N-path filter.

responses—is used, in that case, the simulation shows that the mentioned phase shift will also behave like a regular phase shift at the output. This behavior allows us to merge the channel filtering and phase shifting into one unit. Therefore, we can design a base-band with two N-path filters which also perform as phase shifters and true-time delay units.

For the second implementation, we designed a 3-channel scalable router operating at 28 GHz. Each channel contains a full transceiver. On the receiver side, this implementation includes a LNA with 7 dB extracted gain from the PCB trace to the input of the down-conversion mixers (5.27). Simulation shows a noise figure of about 4 dB and the S_{11} of -18 dB. The down-conversion mixers—for I and Q paths—achieve 5 dB single side-band conversion gain and they are followed by three VGAs which input three base-band filters.

Each channel encompasses two N-path filters to select the desired frequency band and provide the required phase shift and true-time delay. Fig 5.28 illustrates the simulated response of the channels at the maximum programmable gain. Channels' frequencies are chosen based on the about 20 dB isolation between two adjacent users in order to minimize the ISI. Each user has around 80 MHz bandwidth and the reference frequency of the filters is at 3.5 GHz. Fig 5.29 shows the overall architecture of the implemented chip. Each chip contains two branches which

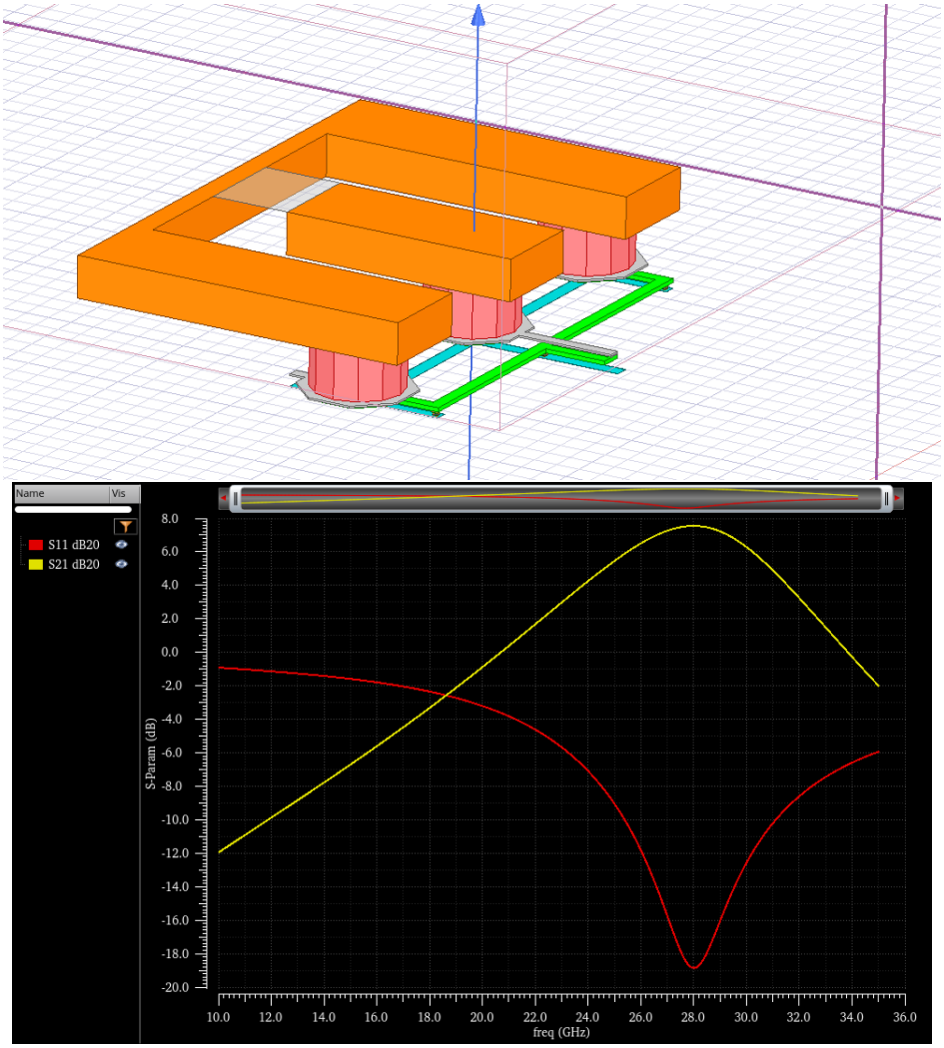

Figure 5.27: LNA gain and input matching characteristic

share the same LO at 28 GHz. Two RF phase shifters are placed before the down-conversion mixers to provide phase shifting capability for the first channel (low pass channel).

5.5 Conclusion

This chapter describes the scalable router architecture and its implications. The scalable router is a decentralized relay array architecture, whose elements (branches) do not need shared timing or phase references to perform beamforming. True time delay control within each branch is critical to ensure data coherence. Here we implemented two versions of the scalable router with multi-channel and multi-beam capability.

While the scalable router offers intriguing advantages, it bolsters rather than replaces

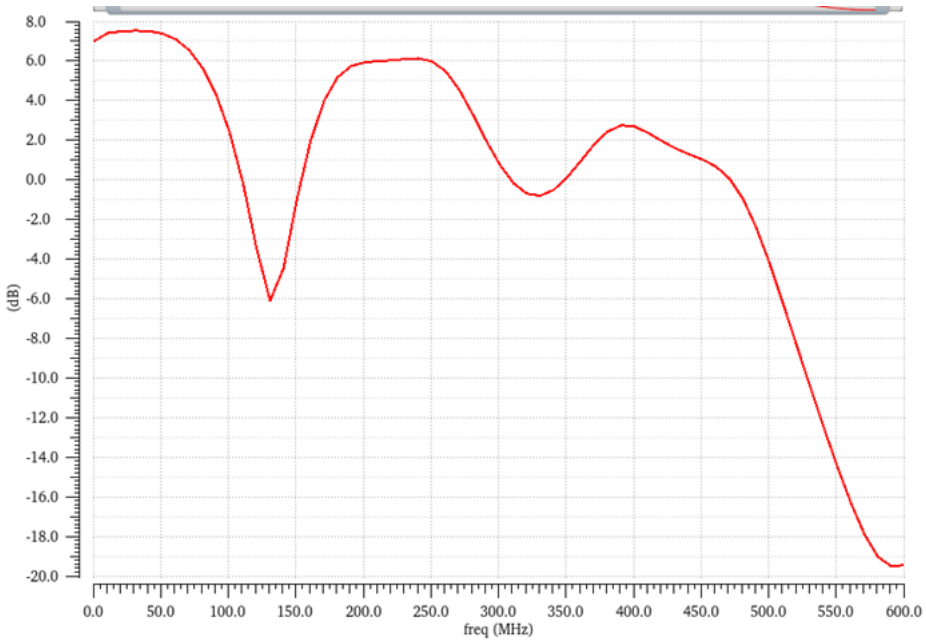

Figure 5.28: Extracted simulation of the frequency response of the channels.

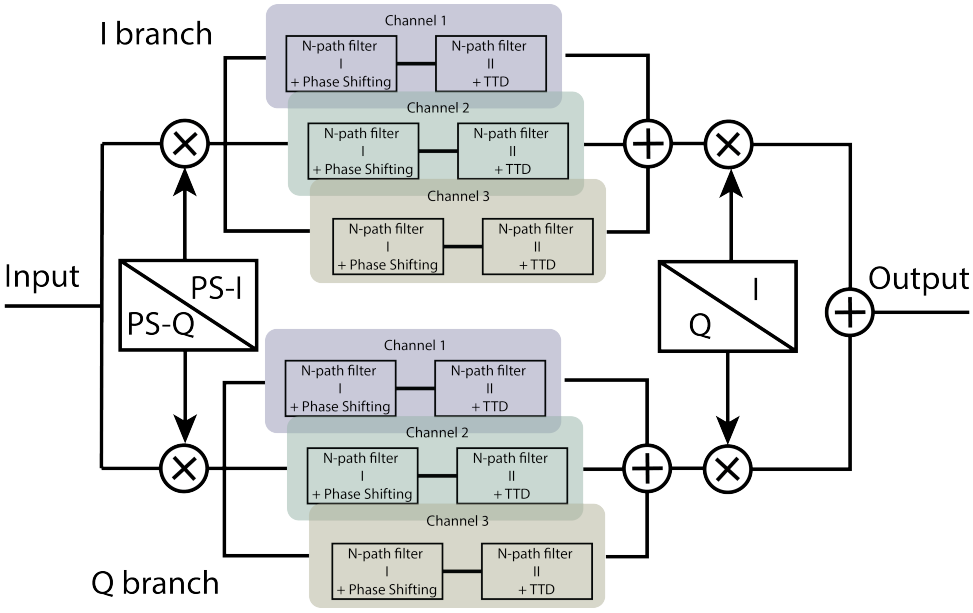

Figure 5.29: High-level architecture of the second implementation.

existing systems. The scalable router naturally fits into emerging communication networks with smaller cells, higher frequencies, and line of sight issues as well as follows broader trends towards distributed systems. The scalable router's fully distributed and decentralized nature, adaptability, and capacity to operate as a single aperture spread across a variety of physically separated surfaces offer a valuable

piece in the larger puzzle of future communication networks.

Chapter 6

SILICON PHOTONICS COMPUTATIONAL ACTIVE FLAT-OPTIC IMAGING SYSTEM

6.1 Introduction

Lightweight miniature imaging systems are becoming essential to the development of wearable electronics, IoT devices, and drones [116]–[118]. In order to boost the performance of these devices, computation can be used to estimate information that is not captured by the imager, as in phase-retrieval imaging [119], three dimensional imaging [120], and encoded/coded aperture imaging [121]. Since imaging systems contain expensive and bulky components, cleverly designed computational processes—specifically engineered surfaces—can be used in imaging systems to replace standard optical components like lenses [122]. Traditionally, many of these types of systems use conventional CMOS or CCD sensors for data collection [119], [121], [123]. While the use of these prepackaged sensors allows for rapid development of large-scale systems, it fails to realize the full potential of a co-designed system-sensor in terms of versatility, performance, size, and mass [120], [124][1].

Here, we present an imaging system based on a custom-designed array of unique diffraction gratings (acting as optical receivers) co-designed with an adaptive reconstruction algorithm and implemented on an integrated photonics platform. This approach enables imaging within a very small volume, without the need for the depth typically required for most imaging systems. Integrated photonics allows for the capturing, manipulation, and sensing of light within an active layer only a few microns thick on a chip using dielectric waveguides manufactured with standard CMOS-type processing [125]. Furthermore, the built-in directionality of the custom integrated photonics micrograting can serve as a means of enhancing the computational image-recovery performance. These diffraction gratings can be engineered to have sensitivity in nearly arbitrary patterns and to admit a large span of wavelengths [126], [127], and as we show, can respond to incoherent light as well as coherent light. Using micrograting couplers co-designed with computational processing, we demonstrate that silicon photonics is a viable platform for computational imaging using a prototype lensless imaging device.

[1] This work was done in collaboration with Alexander White.

6.2 Sensor Design

Images are representations of light field patterns where each point represents the intensity (and possibly other properties, such as color) of light impinging on the imager from a different bearing. When an image is displayed, these points replicate the original light, and the eye can process the image similarly to the original scene, using a lens to direct light from each angle to a different photoreceptor cell on the retina. Conventional cameras work in the same way, using a lens to redirect light impinging at different angles onto different pixels. However, this is not the only way to capture an image. As long as enough independent information is captured from a light field pattern, an image can be formed. For instance, this can be achieved, as in the retinal cells in the compound eye of an insect, with sensors/pixels individually sensitive to light impinging from different directions [128], [129], or even in a different domain, such as Fourier space [130]. While the data itself is not in the form of an image, various computation algorithms can be used to generate a recognizable image from the captured information [122].

Here, we use an array of individually designed silicon-on-insulator (SOI) diffraction gratings to capture spacial intensity information from incoherent light with a wavelength centered at $1550nm$. By using an array of custom micrograting couplers, each with a different sensitivity profile vs. angle of incidence, we can capture enough independent information to linearly map our data into spatial pixels.

Micrograting couplers are versatile devices used to couple light from free space to guided modes in integrated photonics waveguides. Basic grating couplers consist of a periodic structure that diffracts incoming light such that it interferes coherently into the plane of the grating and a taper or other mode matching structure that guides the light into a waveguide mode, Figure 1a. The periodicity and effective indexes of its internal components serve as design parameters that can be used to achieve the desired angular sensitivity profile of a coupler. For instance, in the case of uniformly spaced grating elements, the angle of maximum sensitivity is given by:

$$\theta_{max} = sin^{-1}(n_{eff} - \frac{k\lambda_0}{d}) \tag{6.1}$$

where n_{eff} is the effective index of the grating structure, λ_0 is the free space wavelength, d is the distance between gratings, and k is any integer. The one-dimensional radiation pattern of a uniform coupler in the direction parallel to its main axis can be approximated by assuming an exponential optical loss along the

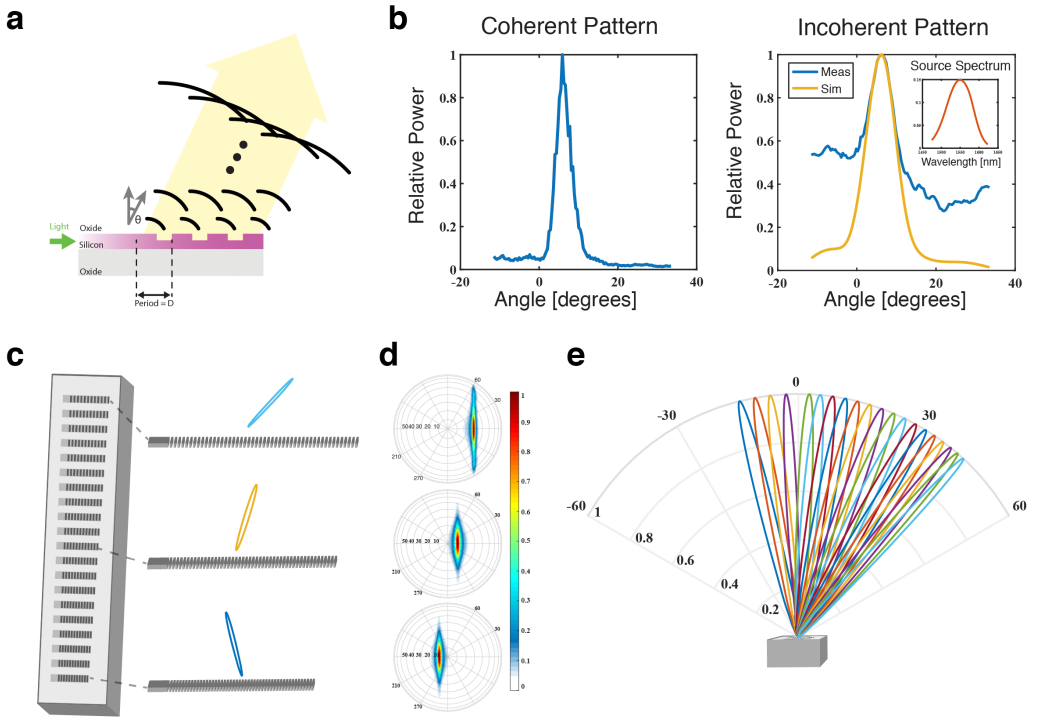

Figure 6.1: Lenssless camera concept (a) Operation of a silicon photonics grating coupler. Light guided in a silicon waveguide clad by silicon dioxide is scattered by regularly spaced etched gratings. These scattered waves interfere coherently in a direction θ dependent on the period of gratings D. (b) The measured coherent and incoherent patterns of a grating coupler. The incoherent pattern is measured with the source spectrum shown. The simulated data is the theoretical pattern given the coherent pattern, the source pattern, and equation (5). (c) Visual representation of the fabricated device containing 20 uniquely spaced grating couplers. Three example couplers are shown with their simulated angular reception patterns. (d) Hemispherical projection of angular reception patterns of the corresponding grating couplers in c. (e) Simulated angular reception patterns of the full 20 grating couplers, in the direction of the gratings.

coupler in conjunction with the propagation phase shift:

$$h(\theta) \propto \sum_k (1-\alpha)^{k-1} e^{i(\frac{d}{\lambda_{eff}} - \frac{d}{\lambda_0} sin(\theta))} \tag{6.2}$$

where α is the loss per grating tooth and λ_{eff} is the effective wavelength inside the grating structure. Given such a pattern $h(\theta)$, the power received at time, t, due to an arbitrary farfield illumination $I(\theta, t)$ can be given as

$$P(t) = |\int_\theta h(\theta) I(\theta, t) d\theta|^2. \tag{6.3}$$

Under spatially incoherent illumination, $I(\theta)$ and $I^*(\theta')$ are uncorrelated except when $\theta = \theta'$. If the illumination is also temporally incoherent, we can represent the total power as

$$P = \int_\theta \int_\omega |h(\theta, \omega)|^2 |I(\theta, \omega)|^2 d\omega d\theta. \tag{6.4}$$

With a single illumination source, we can separate the angular and frequency dependence of I, and define a modified sensitivity pattern $h'(\theta)$,

$$|h'(\theta)|^2 \equiv \int_\omega |h(\theta, \omega)|^2 S(\omega) d\omega \tag{6.5}$$

where $S(\omega)$ is the frequency spectrum of the scene illumination. Under narrowband illumination, the pattern of a uniform coupler can be approximated as the convolution of the coherent pattern and the optical power spectrum, Fig 6.1b. For non-uniform couplers or extremely broadband illumination, the sensitivity pattern can be calculated by explicit simulation and weighted averaging of the patterns and the wavelength component of the optical power spectrum. As incoherent illumination leads to power addition and not electric field addition (as it would under coherent illumination), the power received by an arbitrary far-field distribution is given by the inner product of the sensitivity pattern and the far-field pattern. We can discretize this inner product to have the following approximation,

$$P \approx \sum_\theta |h'(\theta)|^2 |I(\theta)|^2 \Delta\theta \ , \tag{6.6}$$

where the number of discretization points determines the accuracy of our estimation. Consequently, we can rewrite (6.6) as the following,

$$P = \mathbf{h}^T \mathbf{x}, \tag{6.7}$$

where $\mathbf{h}$ is an n dimensional positive pattern vector ($\mathbf{h} \in \mathbb{R}^n_+$), and $\mathbf{x}$ is the intensity vector, where n is given by the number of the distinct values of θ we have used in the approximation (6.6). If we combine the results from multiple sensors, we can get the following linear system of equations,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \ , \tag{6.8}$$

where the m-dimensional positive vector $\mathbf{y}$ is the measured power, m is the number of sensors, the m-dimensional vector $\mathbf{z}$ is the unknown noise vector, and $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_m^T \end{bmatrix}$

is the matrix whose rows correspond to the pattern of each sensor. As we construct **H** through the design of our grating couplers, we can reconstruct the image **x** by inverting the linear system.

To form an imager, we can modulate the grating pitch across a number of couplers, effectively generating a set which is sensitive to different illumination angles. In this way, the sensors mimic the effect of a lens on traditional pixels. As each grating coupler is designed to be most sensitive to a different angle, the approximated transfer matrix **H** (between discrete angles and sensor power) can be inverted effectively. Heuristically, overlapping sensor patterns allow for the data capture and localization of the full imaging area, and single angle reception allows for more independent measurements, improving the resolution or minimum resolvable angle. Thus it is favorable to strike a balance between overlapping and narrow sensor patterns.

As a proof of concept, we designed and fabricated a device with 20 of these

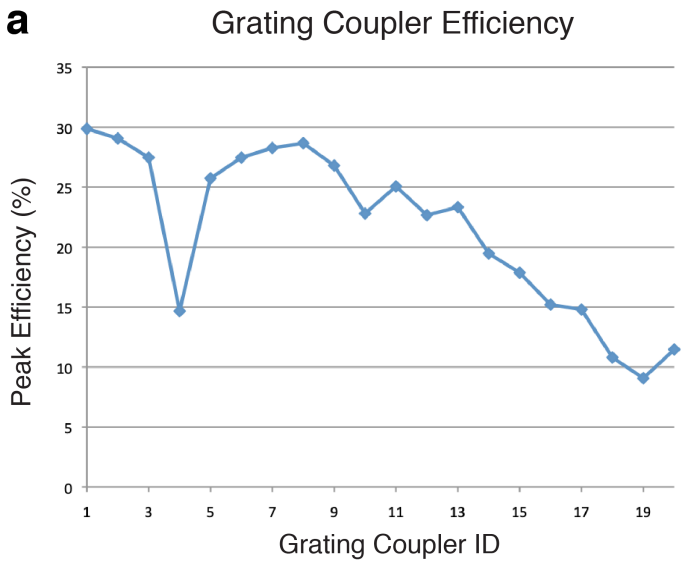

Figure 6.2: Peak efficiency of grating couplers. Efficiency drops for the coupler with near perpendicular optimal incidence angle and drops for steeper angles.

sensors on a silicon photonics SOI platform, Fig 6.1c. Each sensor consists of a uniform grating coupler and a germanium photodiode, sensitive to light with a wavelength around 1,550nm. The grating couplers guide impinging light into 500nm by 220nm transverse-electric single mode silicon waveguides, which are then fed to the germanium photodiodes. The photodiodes convert the optical power received to an electric current, then this current can be amplified by using a transimpedance amplifier (TIA). These sensors have a far-field coherent full-width half-maximum

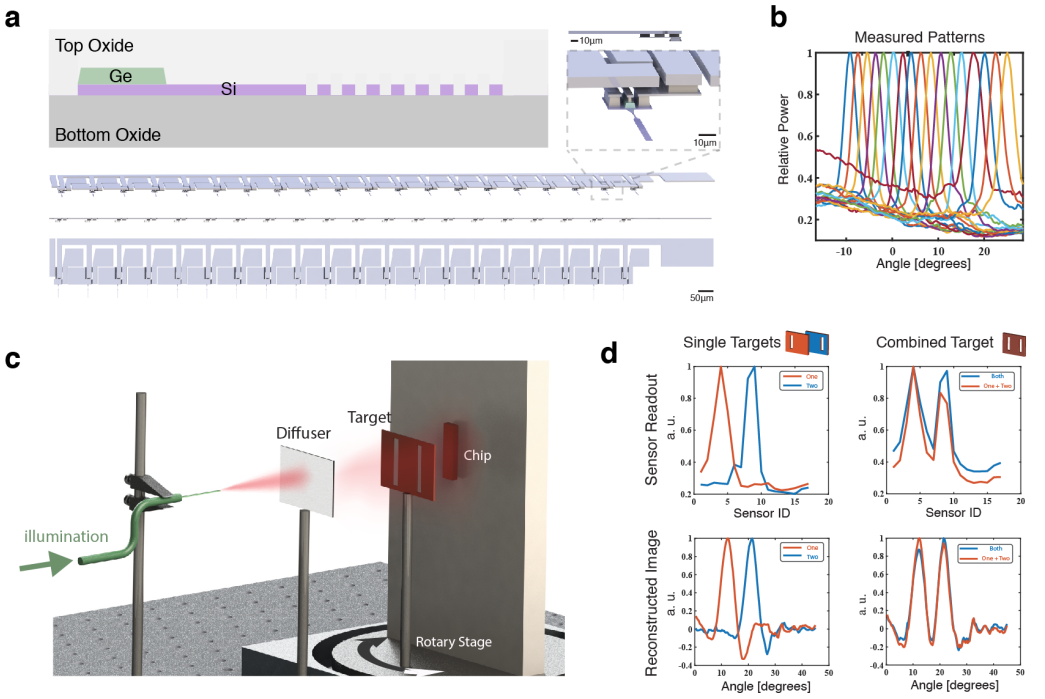

Figure 6.3: Measurement setup and characterization (a) Diagram of silicon on insulator platform with germanium photodiodes (top left). Scale isometric rendering of fabricated device from angle, sides, and top. Silicon shown in purple, germanium in green, and metal connections, pads, and vias in grey. (b) Measured angular reception patterns of the 17 functional sensors. Three sensors were not functioning due to photodiode connections. (c) Representation of test setup. The chip is illuminated through a cleaved optical fiber, diffuser, and target. The copper target blocks light from some the angle it is present, forming a negative of the barcode image seen by the chip. The target and chip are mounted to a rotary stage so that the light spread from the diffuser can be averaged to effectively provide an even illumination across all visible angles. (d) Data collected (top row) and reconstruction (bottom row) of three targets: two with slits in different positions (left column) and one with both slits (right column).

sensitivity of approximately 5 degrees in the grating direction and 30 degrees in the transverse direction, Fig 6.1d, and an average peak efficiency of 21.5%. Due to reflections from the silicon substrate, the peak efficiency varies over the couplers, Fig 6.2. To minimize the effect of stray light and reflections, the photodiodes and area surrounding the grating couplers are covered by metals and an absorptive germanium layer. By sweeping the grating pitch, the maximum reception angle is swept from -10 to 30 degrees (to normal) in 2 degree increments, Fig 6.1e.

The fabricated device is shown in Fig 6.3a. The reception patterns of 17 functional

gratings were measured with a super luminescent diode (SLD) with a $50nm$ bandwidth, and are shown in Fig 6.3b. To verify the functionality and linearity of the sensor, we imaged light impinging on a double slit and each of its constituents, Fig 6.3c-d. By imaging two slits and their combination, we can test the approximation in equation 6.6, as the image formed by the combination should be the sum of the constituents, which exhibits the linearity of the imager and also follows the mathematical model. To do this, light from the SLD is first projected from a cleaved optical fiber onto a diffuser to perturb the spacial coherent of the beam. The diffuse light then travels through a copper complementary image (negative) of a target image and impinges on the chip. The target and chip are positioned on a rotary stage so that the illumination from the diffuser can be integrated over the entire angular sensitivity range (45 degrees). The final images can then be reconstructed by inverting the measurement in Fig 6.3b, our new $\mathbf{H}$, and multiplying by the data collected.

6.3 Reconstruction Algorithm

To capture images with this sensor, we must reconstruct the unknown intensity vector $\mathbf{x}$, from the (noisy) linear system (6.8). This reconstruction problem faces two main challenges. First, because we need the discrete approximation in (6.6) to be accurate, the number of measurements $m = 17$ is much smaller than the underlying dimension $n = 120$. Second, our algorithm must be tuned properly to adjust for the impact of the noise in the linear system.

To address both of these challenges, we employ regularized least squares [131], [132], which provides us with the following estimate,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2 + \lambda \, ||\mathbf{\Gamma}\mathbf{x}||^2 \tag{6.9}$$

where $\mathbf{\Gamma}$ is an $n \times n$ matrix that is chosen based on the structure of the underlying dataset and λ is the regularization parameter that must be tuned properly. Note that the objective function in (6.9) consists of two terms. The first term in the objective function, the least-squares term, essentially captures the power of the noise vector $\mathbf{z}$. The second term in the optimization, the regularization term, enforces a structure dictated by $\mathbf{\Gamma}$ on the unknown vector $\mathbf{x}$.

For any set of parameters, we can represent the solution to the optimization (6.9) in the following closed-form,

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{H} + \lambda \, \mathbf{\Gamma}^T\mathbf{\Gamma}\right)^{-1} \mathbf{H}^T\mathbf{y} \tag{6.10}$$

which leads to very computationally efficient solutions. To generate an optimal

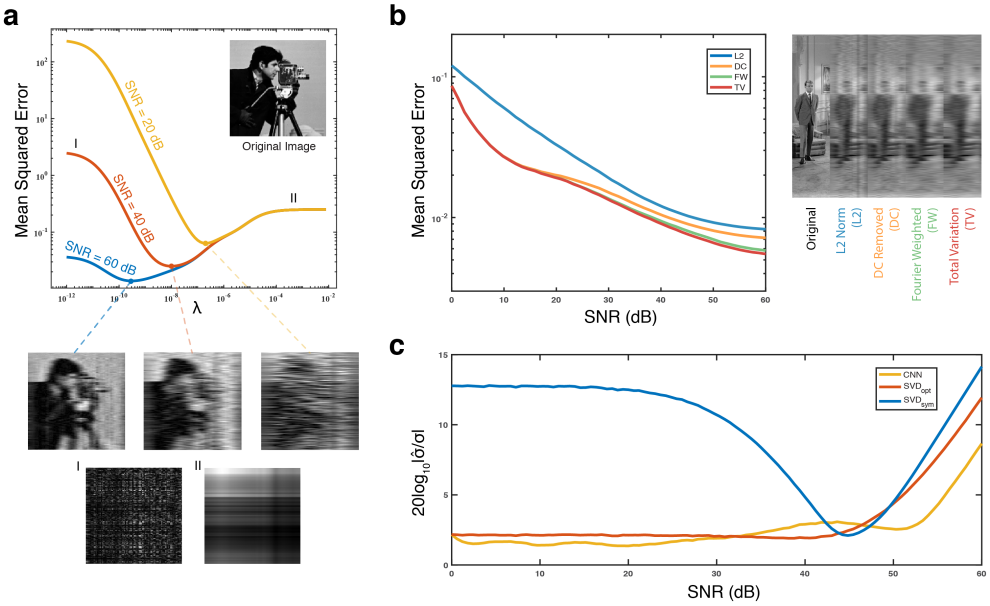

Figure 6.4: Performance (a) Comparison of simulated reconstruction error across λ for data with different SNRs. Images shown for λ_{opt} of each SNR as well as in the low lambda and high lambda regimes. As this is a 1D imager, an image consists of a single slice of 1x120 pixels. The 2D images are representations of the concatenated 1D slices. (b) Comparison of reconstruction error with different $\mathbf{\Gamma}$ matrices at λ_{opt}. (c) Comparison of SNR estimators: CNN, SVD with a heuristic filter (SVD_{sym}), and SVD with an optimized filter (SVD_{opt}).

reconstruction, we must choose an optimal regularization strength λ and regularization matrix $\mathbf{\Gamma}$. The choice of these parameters is critical, as can be seen in Fig 6.4a.

The regularization matrix, $\mathbf{\Gamma}$, should be chosen based on the structure of the underlying data [133]. Since there is no specific structure imposed on the data, there is no strictly optimal $\mathbf{\Gamma}$. However, we can use general properties of images to chose better matrices. Here, we utilize four matrices for $\mathbf{\Gamma}$ which have desirable properties for image reconstruction:

- The identity or L2 norm (L2)
- DC removed (DC)
- Fourier Weighted (FW)
- Total Variation (TV)

Fig 6.4b shows a comparison of the reconstruction algorithms for the above choices of the matrix $\mathbf{\Gamma}$.

After choosing a favorable regularization matrix, $\mathbf{\Gamma}$, we must tune λ to its optimal value, λ_{opt} (Fig 6.4a). In the optimization program (6.9), it can be observed that the first term, $||\mathbf{y} - \mathbf{H}\mathbf{x}||^2$, measures the power of the noise vector, and the second term, i.e., $||\mathbf{\Gamma}\mathbf{x}||^2$, measures the power of the signal. Hence, we expect the optimal value of the λ to be inversely proportional to the signal-to-noise ratio (SNR), i.e., $\lambda_{\text{opt}} \propto \frac{1}{SNR}$.

As it was mentioned before, to reconstruct an image based on its predicted SNR, we must map this SNR to a regularization strength λ. While the optimal mappings are close to linear, a linear approximation is not exact, Fig 6.5 and Fig 6.6. Thus, for optimal reconstruction, we can use more simulated mappings between SNR and λ. These can be generated by numeric optimization of λ across a set of noisy data.

As regularizers perform differently across data types, we generate two sets of mappings. One set maps SNR to λ in natural images, optimized over the 18432 images used in simulations (Fig 6.5 top). The other set maps SNR to λ in sharp images, like the barcodes physically measured. These mappings were optimized over 15000 randomly generated barcode images (Fig 6.5, bottom).

As each $\mathbf{\Gamma}$ matrix will perform differently on a given dataset, a separate mapping is required for each one. While we do need multiple maps, using SNR as a mediator between image analysis and regularization strength allows for a single SNR prediction algorithm for all $\mathbf{\Gamma}$ matrices.

While the noise level could be calibrated, chip-to-chip variation, temperature fluctuation, varying levels of electrical interference, and numerous additional environmental factors can affect the noise present in a measurement. As optimal reconstruction is highly sensitive to SNR, to create a more robust system, we devise two algorithms to estimate the noise variance in a given measurement. Note that the noise in the system of equations (6.8) is caused by two different sources: 1. The measurement noise stemming from the device. Note that this noise can have impact of both $\mathbf{H}$ and the measurement vector $\mathbf{y}$. 2. The approximation error caused by discretizing the model in (6.6). Note that in order to have a better approximation in the discrete model, one needs to increase the signal dimension n; however, this would have a negative impact on the reconstruction as $\frac{m}{n}$ decreases. For our analysis purposes, we assume that the additive noise, $\mathbf{z}$, has independently and identically distributed Gaussian entries with mean zero and variance equal to σ^2. In order to find an estimate for SNR we employ two different algorithms. Our first proposed algorithm relies on the

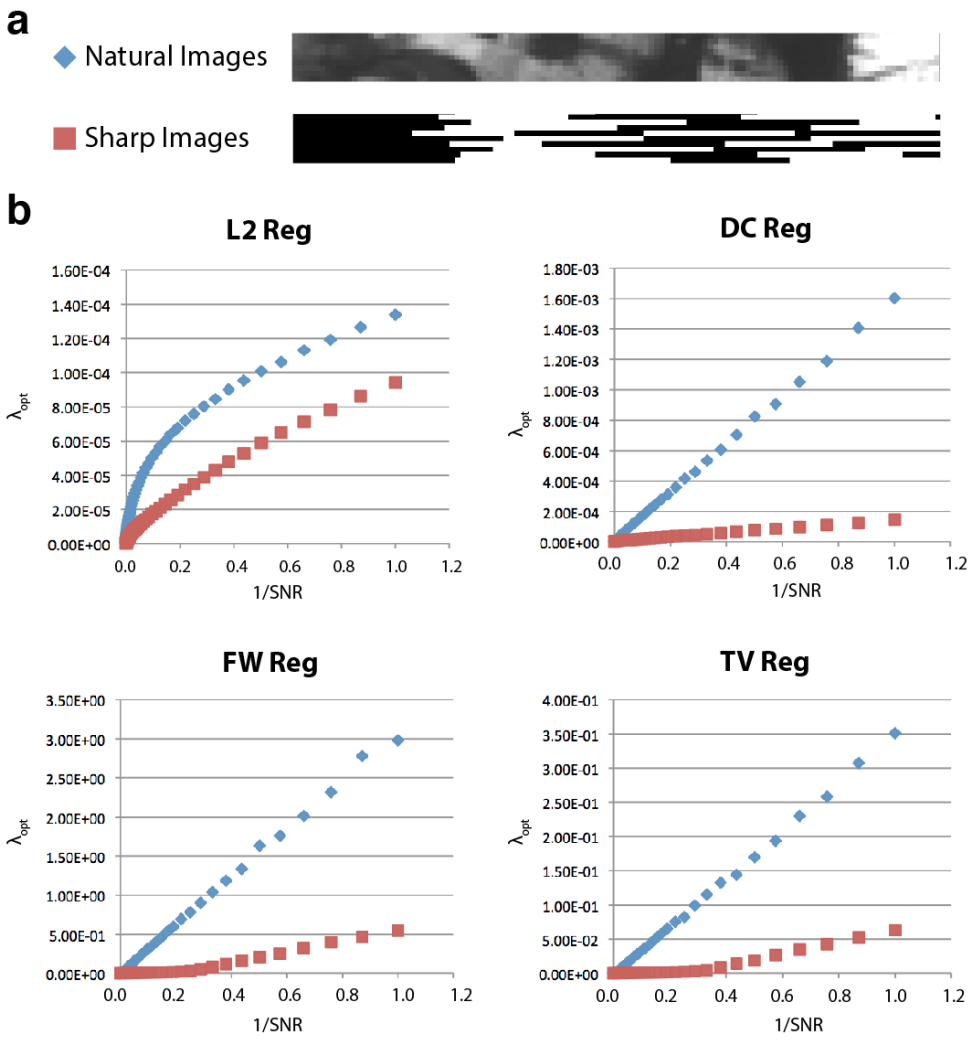

Figure 6.5: Natural and sharp images (a) Examples of natural images and randomly generated sharp images used in 1/SNR to λ_{opt} mapping. (b) Plots of the optimal lambda across 1/SNR from 10^{-6} to 1. This mapping is shown for each of the four $\mathbf{\Gamma}$ matrices and for natural and sharp images.

assumption that the underlying signal $\mathbf{x}$ is smooth. This assumption is commonly used in imaging applications [133]. Consider the singular value decomposition of the measurement matrix $\mathbf{H}$ as follows,

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T , \tag{6.11}$$

where $\mathbf{U}$ is an $m \times m$ orthogonal matrix whose i^{th} column is represented by $\mathbf{u}_i$, $\mathbf{\Sigma} = \text{diag}(s_1, s_2, \ldots, s_m)$, and $\mathbf{V}$ is an $m \times n$ matrix with orthonormal columns, where $\mathbf{v}_i$ denotes its i^{th} column. We then consider the impact of the reconstruction

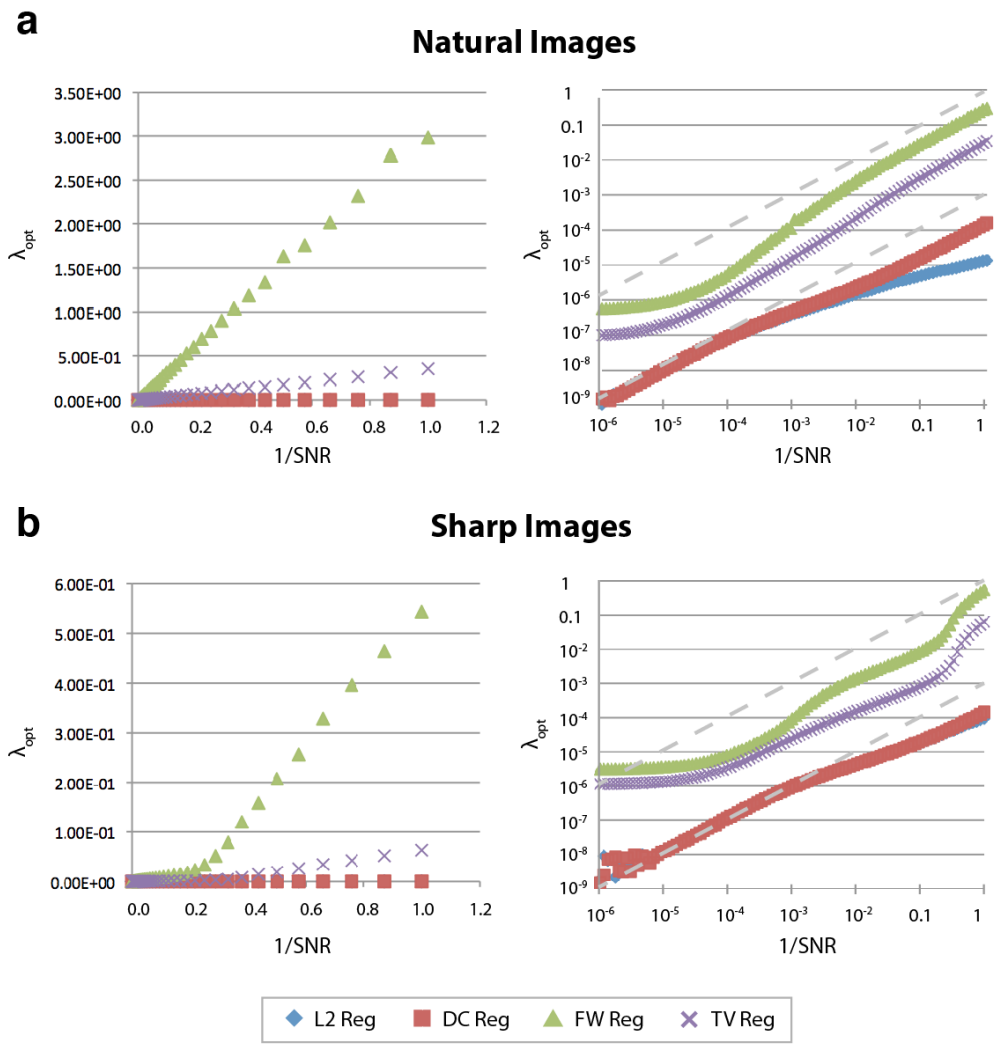

Figure 6.6: 1/SNR to λ_{opt} simulation (a) Linear and log-log plots of 1/SNR to λ_{opt} for natural images using all four $\mathbf{\Gamma}$ matrices. Dashed lines in log-log plot indicate linear slope. (b) Linear and log-log plots of 1/SNR to λ_{opt} for sharp images using all four $\mathbf{\Gamma}$ matrices. Dashed lines in log-log plot indicate linear slope.

algorithm on the noise vector. We have,

$$\mathbf{z}^{\dagger} = \mathbf{H}^{\dagger}\mathbf{z} = \sum_{i=1}^{m} s_i^{-1}(\mathbf{u}_i^T\mathbf{z})\mathbf{v}_i \ , \tag{6.12}$$

where $\mathbf{H}^{\dagger}$ is the pseudo-inverse of $\mathbf{H}$. If we filter out the low frequency components of this reconstruction (assuming that the underlying signal is smooth), we are left with an approximation of the reconstruction contribution of noise, and σ^2 can be

approximated as,

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(s_i\ \mathbf{v}_i^T\mathbf{z}^\dagger)^2\ . \tag{6.13}$$

The second proposed algorithm is to train a convolutional neural network [134], [135]. This approach is particularly useful when we have enough samples from the dataset. Unlike the first approach, here we do not need any assumption on the structure of the underlying signal. In our implementation, we trained a 4-layer network with two convolution layers and two fully-connected layers. The details of the training is mentioned in [136].

Fig 6.4c compares the result of the above-mentioned algorithms in estimating the SNR. As expected, the CNN approach often provides a better estimate. To form an image, the data collected from the sensor is fed through the SNR estimator, the SNR is mapped to λ_{opt}, and the data is finally processed with the reconstruction (6.9), Fig 6.7a.

To test the full system, we measure 6 barcode patterns, Fig 6.7c (Target), that spell the letters C, I, T when concatenated. The collected sensor data is then multiplied by a reconstruction matrix $\mathbf{H}$ and fed through the trained CNN or SVD algorithm to estimate the noise variance. This noise variance is then used to map to an optimal lambda estimate through SNR. Finally, a new $\mathbf{H}_{opt}$ is calculated using this optimal lambda and the image is reconstructed, Fig 6.7a. Fig 6.7c-e show the resulting images reconstructed using each of the four $\mathbf{\Gamma}$ matrices, SVD and CNN algorithms, and two SNR to λ mappings. We achieved a minimum mean squared error of 11.7%, equivalent to a peak SNR of 9.3 dB.

6.4 Conclusion

Here we develop an imaging technique that utilizes an engineered set of grating couplers in integrated photonics to capture images without the need for a lens or active beam-steering. We demonstrate this technique with a 1-dimensional imager that occupies only $2000 \times 200 \times 20\ \mu m^3$ and can capture 20 data points a 45-degree field of view. This type of miniature image sensor could enable imaging in systems that require extremely light-weight and compact sensors, such as wearable electronics, drones, and health monitors. In addition, the demonstrated algorithms can be used in other computational imaging systems, and in many cases can easily coordinate with existing infrastructure using linear regularizers to improve robustness to noise. Because it is implemented in an integrated photonics process, this type of sensor can be augmented by photonic processing techniques such as filtering and passive

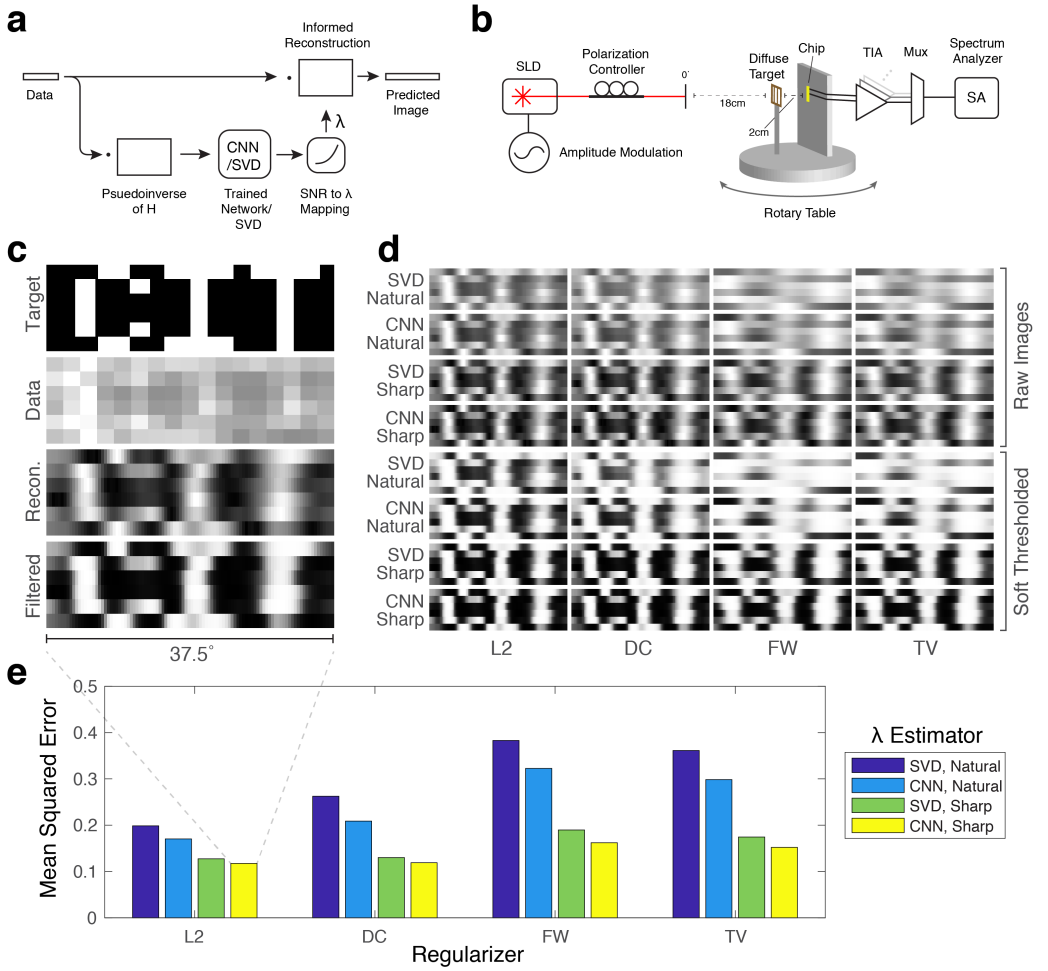

Figure 6.7: (a) Block diagram of reconstruction algorithm. (b) Measurement setup. (c) Reconstruction example with measured CIT barcode patterns. 'Target' shows the barcodes used, displayed horizontally and concatenated vertically. 'Data' shows the data collected from each of the 17 sensors. 'Recon.' shows the image reconstructed using L2 regularization, CNN noise estimation, and an SNR-λ mapping for sharp images. 'Filtered' shows the reconstruction after passing through a sigmoidal amplitude filter. (d) Reconstructed images using different $\mathbf{\Gamma}$ matrices, SNR estimators, and SNR to λ_{opt} mappings. 'Natural' is using the SNR to λ_{opt} mapping optimized for natural images, and 'Sharp' is using the SNR to λ_{opt} mapping optimized for random binary images. (e) Quantified reconstruction error for plots in d. Each bar represents the average of the 6 barcode patterns.

beam-forming [137]–[139], and can be generalized to 2D and 3D imaging. By optimizing gratings for transverse as well as longitudinal angles, the same type of imaging can applied to a 2D field of view, and by using parallax or optimization of near field reception patterns, this type of imager could also capture 3D data.

For high resolution in the extra dimensions, the imager system requires many more gratings, and thus larger sensors for a similar effective aperture. To get around this problem, it may be necessary to use grating couplers optimized for multiple gratings to increase the camera's effective aperture. To reconstruct these 2D and 3D images, we can take the same algorithmic approach as an the 1D case by adding the extra dimensions in the transfer matrix H (replacing it with a tensor or flattened tensor). While the reconstruction will be more computationally intensive, the longer computational time can be shortened using pre-calculated matrix inverses. In addition, photonics information processing techniques could be used in as a prepossessing step or in conjunction with the reconstruction algorithm [140], [141].

Chapter 7

BREAKING FOV-APERTURE TRADE-OFF

7.1 Introduction

An integrated nano-photonic antenna is a planar structure fabricated on a chip [142]–[150] which is defined as a transducer that couples the guided mode into the free-space mode and vice versa[1]. In other words, an antenna is an on-chip component that in the transmitter setting is fed with a guided mode and radiates into the free-space, and in the receiver setting, captures the free-space incident wave and couples it into a waveguide (Fig. 7.1). Nano-photonic antennas have application in many recently demonstrated integrated photonic systems such as optical phased array transmitters [142], [144] and receivers [145], nano-photonic coherent imager [150], wireless optical communication [149], lens-less cameras [136], [139], etc. For implementing most of today's low cost, high-yield, and commercial systems, silicon photonic integration technology with a single photonic layer of silicon is used [145]. A silicon photonic platform is often realized on a silicon-on-insulator (SOI) process (Fig. 7.1) in which the top silicon layer is used to realize photonic components, including antennas, through etching and patterning. Moreover, various etching levels, doping layers, and metal layers are provided to realize photonic components and interfacing, as well as germanium layer for photo-detection in some near-infrared wavelengths. Moreover, to achieve low loss and visible range photonic components, silicon nitride has been used as the photonic layer material in some platforms [151][2].

In the early implementations of integrated photonic systems, conventional grating couplers were used as transmitter antennas. However, the large footprint of a grating coupler is a limiting factor in many systems such as dense antenna arrays. Moreover, the need to selectively/collectively improve specifications of the antennas such as field-of-view (the angular range that the antenna is an effective radiator/receiver), radiation efficiency (the fraction of power coupled to free-space), and form factor has recently motivated the researchers to devise novel antenna topologies with improved performance. Deviating from grating structures allows achieving higher radiation efficiency [145]. Moreover, an ultra-compact broadband antenna is realized via optimization of the silicon slab etching pattern [152]. In addition to devising novel

[1] This is the same definition used for conventional microwave antennas.

[2] This work was done in collaboration with Dr. Reza Fatemi and Dr. Aroutin Khachaturian.

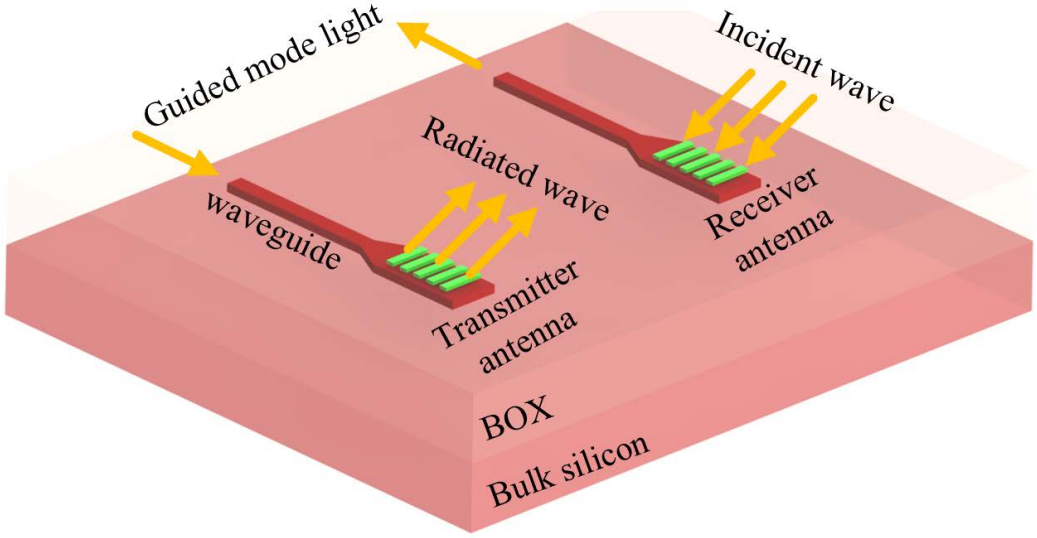

Figure 7.1: Schematic of a standard SOI silicon photonic platform consisting a bulk silicon substrate and a buried oxide layer (BOX). Optical components including transmitter and receiver nano-photonic antennas are fabricated by patterning the silicon layer on the BOX.

antenna structures on a single photonic layer process, by adding a high contrast grating on top of the antenna structure [146] and bottom reflectors [153] the radiation efficiency of the antenna is improved significantly. However, most of these works are limited to antennas with large footprints which often have an FOV that is narrow in one of the dimensions.

In addition to the transmitter antenna, a photonic receiver system [139], [145], [150] also incorporates antennas that collect the incident power and output to waveguides. The waveguides are then used to route the optical signals for further optical or digital processing. Since the received signal is not very strong in many applications, for a given physical size of the antenna, the largest effective aperture is desired to maximize the signal-to-noise ratio, sensitivity of the system, and robustness to different noise sources. While techniques such as heterodyne detection [139] can be used to increase the sensitivity of the system, a larger collection area is always beneficial to a receiver system and allows for operation at lower received signal levels and noisier environments. Moreover, the FOV of the antenna needs to be in harmony with the FOV of the system. A limited antenna FOV directly affects the performance of the system. For instance, in an optical phased array, small antenna FOV translates, directly, to a limited steering range. Conventional photonic antenna structures impose a compromise between FOV and effective aperture, i.e.,

a standard antenna with a large effective aperture is limited in the angular range that it can operate and vice versa.

7.2 FOV-Aperture Trade-Off

In a single-port receiving antenna, incoming light couples into a guided mode within an on-chip planar waveguide. Its effective aperture is defined as

$$A(\theta, \phi) = P_o / S_i, \tag{7.1}$$

where P_o is the total power coupled into the waveguide attached to the antenna port and S_i is the power density (defined by the Poynting vector) of a plane wave arriving from the direction defined by (θ, ϕ), assuming they are polarization matched. If the polarization state of the incident light, on the Poincare sphere representation, deviates from the polarization that results in the maximum power collection by an angle ψ, a polarization matching factor of $\cos^2(\psi/2)$ must be added to the above equation [154], which can reduce the effective aperture (down to zero) for a mismatched polarization. Generally speaking, having a receiver/transmitter antenna that can couple in both polarization is desirable for many applications. Whenever there is a need for high signal strength and a large signal-to-noise ratio (almost always), a high effective receiving aperture is desirable. Effective aperture as a function of θ and ϕ defines the receiving pattern of the antenna, which also known as the antenna pattern. For any given application, a minimum FOV must be achieved, while maximizing the collection area. Increasing the effective aperture faces limitations in different applications and scenarios. For instance, in an optical phased array, if the individual element antenna size is increased to collect more power, the element spacing needs to be increased accordingly to fit the antennas. Increasing the element spacing reduces the grating lobe spacing in the array pattern, which can limit the effective grating-lobe-free beam steering range (FOV) of the optical phased array [145]. In addition, forced by conservation of energy, the effective aperture of each antenna in a very large array cannot exceed the unit grid area of the array, defined by the array pitch. If an array is made of individual elements whose standalone effective aperture exceeds this limit, the total effective aperture will be reduced, and it manifests itself through coupling between elements. On the other hand, in the transmitter scenario, it is desired to maximize the radiation efficiency of the antenna which is defined as the ratio of the total power radiated by an antenna to the net power fed into the antenna port[3]. The radiated power is

[3]In some literature, the port mismatch is not included in the radiation efficiency and the net power delivered to the antenna is considered.

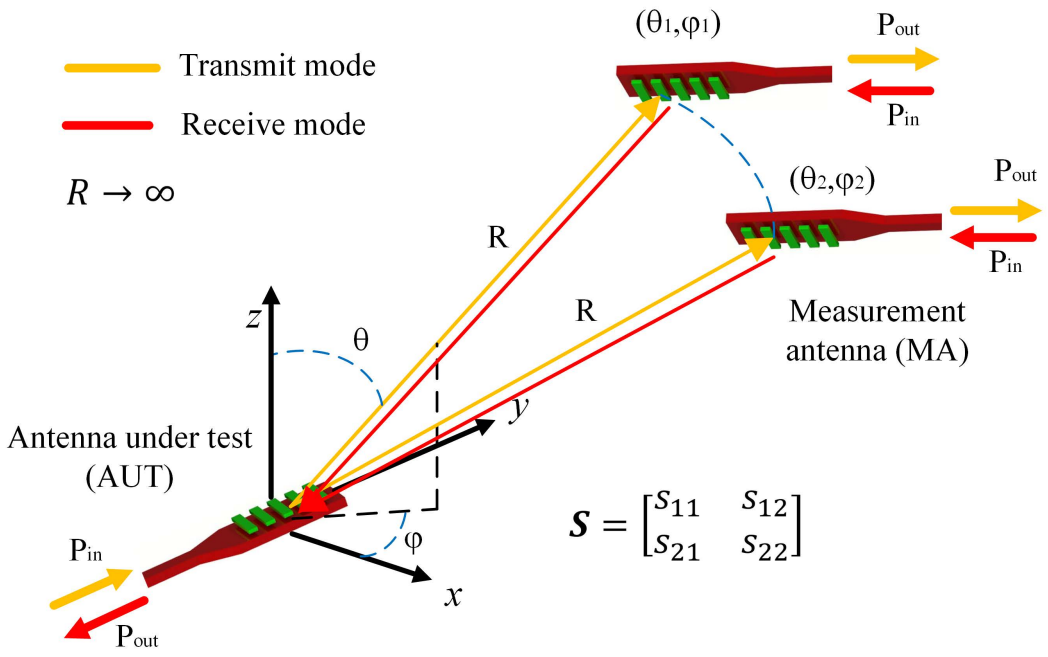

Figure 7.2: Antenna setting for deriving radiation pattern in the transmit mode and effective aperture pattern in the receive mode. Radiation pattern, $D(\theta, \phi)$, and receiving pattern, $A(\theta, \phi)$, of an antenna in a reciprocal medium are linearly related.

desired to be distributed on a certain angular range or FOV which is a property of the antenna pattern. Antenna pattern or directivity pattern, $D(\theta, \phi)$, is a function of elevation angle, θ, and azimuth angle, ϕ. Directivity characterizes the relative power density in each direction to the average power density radiated by the antenna.

There is a well-known antenna reciprocity relationship between the receive and transmit patterns of a single-port antenna that imposes a trade-off between the mentioned FOV and effective aperture.

Consider the setting represented in Fig. 7.2, with the antenna under test (AUT) and a measurement antenna (MA) at a very large distance R at (θ, ϕ) (shown at two different locations in the figure). MA is oriented such that it is polarization matched to the AUT. Both antennas have a single-mode port. Thus, a two-port network model with a scattering matrix of **S** is sufficient to analytically represent the system. The parameters of the scattering matrix depend on the location of the MA and are functions of (θ, ϕ). When AUT is in the transmit mode, the antenna pattern (directivity) of AUT can be measured by sweeping MA over (θ, ϕ). The power transfer from AUT to MA is given by $|s_{21}(\theta, \phi)|^2$ at each (θ, ϕ). Therefore,

the directivity of the antenna is proportional to

$$D(\theta,\phi) \propto |\frac{s_{21}(\theta,\phi)}{s_{21}(\theta_1,\phi_1)}|^2, \tag{7.2}$$

in which (θ_1, ϕ_1) is the reference angle (usually the angle with maximum power). Conversely, with AUT in the receive mode, the MA radiates power and the AUT receives it. Now, the power transfer from MA to AUT is given by $|s_{12}(\theta,\phi)|^2$. The effective aperture pattern (receive pattern) $A_{eff}(\theta,\phi)$ is proportional to

$$A_{eff}(\theta,\phi) \propto |\frac{s_{12}(\theta,\phi)}{s_{12}(\theta_1,\phi_1)}|^2. \tag{7.3}$$

In a reciprocal medium, the scattering parameter matrix is symmetric, i.e. $s_{12} = s_{21}$ [155]. Therefore, the directivity pattern in the transmit mode and effective aperture pattern in the receive mode are linearly related as

$$A_{eff}(\theta,\phi) \propto D(\theta,\phi). \tag{7.4}$$

Hence, the well-known relation between directivity and effective aperture for a single-port antenna that can be obtained by calculating the effective aperture and directivity pattern of an arbitrary antenna which yields [156]

$$A_{eff}(\theta,\phi) = \frac{\lambda^2}{4\pi} D(\theta,\phi), \tag{7.5}$$

in which λ is the wavelength of the electromagnetic wave. In other words, the normalized radiation pattern of the antenna in the transmit mode and its effective aperture pattern in the receive mode are the same. Therefore, a single-port antenna that has a good performance in sending the wave towards a certain direction is also a good antenna for collecting the light impinging from the same direction.

Multi-Mode antennas

To push the performance of the antennas beyond the conventional limits, the derived trade-off (equation (7.5)) has to be overcome. This would only be possible by violating the assumptions leading to (7.5). The single-mode antenna port assumption for the configuration of Fig. 7.2 resulted in the two-port network model, which forms the foundation of our derivation. Extending the antenna structure to support multi-mode guided waves could potentially open a path for designing antennas with performance metrics beyond the conventional limits. To capture the effect of the extra modes of the antennas in the scattering matrix model, a new row and column should be added for every extra mode on each antenna. It can be shown that for

a reciprocal medium, the relationship of equation (7.5) still holds for every pair of transmitter-receiver modes. However, combining the contribution of different modes *collectively* can result in a larger effective aperture and FOV, simultaneously. In other words, adding a second mode to the antenna could in theory up to double the FOV (in solid angle) with the same effective aperture. Alternatively, it could up to double the effective aperture with the same FOV or increase both FOV and effective aperture partially. Increasing the number of modes supported by the antenna relaxes the trade-off with the same trend by each mode bringing its share of FOV and effective aperture. It is worth noting that dual-polarization grating couplers investigated in the literature [157]–[159] are a small subset of multi-mode antenna structures. In these couplers, the two orthogonal incident polarizations are captured by the two antenna modes and coupled into the super-modes formed by two single-mode waveguides. While the peak value of the effective aperture is not increased, it is extended to cover the whole polarization state-space uniformly rather than having a peak at a single polarization and dropping to zero for the corresponding orthogonal polarization. As a side note, another assumption leading to (7.5) is the reciprocity of the medium, which can potentially be broken by incorporating non-reciprocal materials in the antenna structure.

The multi-mode antennas are studied in the microwave region and used for various purposes such as designing loaded scatterers [160], [161], pattern synthesis [162], realizing antennas for multiple-input multiple-output (MIMO) systems [163], [164], etc. For metallic antennas operating in the microwave region, design methods such as characteristic modes can be used to analyze the antenna and realize single- and multi-mode antennas[165]–[167]. A class of electromagnetic structures that share similarities with multi-mode antennas are multi-beam antennas. In microwave antenna design, multi-beam antenna is referred to a *collection* of independent antennas that can send independent waves in different directions [168]. It should be noted that in a reciprocal medium, the power collected and coupled to multiple modes can not be fully combined into a single mode and carried by a single-mode waveguide, as that entire system would be tantamount to a single-mode single-port antenna. While an ideal electromagnetic power combiner does not exist [155], it is possible to combine the power of multiple modes into a single one, if the phases and amplitudes of the input modes are defined or controlled [169] and novel techniques such as inverse design and computational design method can be used to improve the performance [170]–[172]. However, in the case of the multi-mode antenna, the relative phase and amplitude of the excited modes changes versus the incident angle

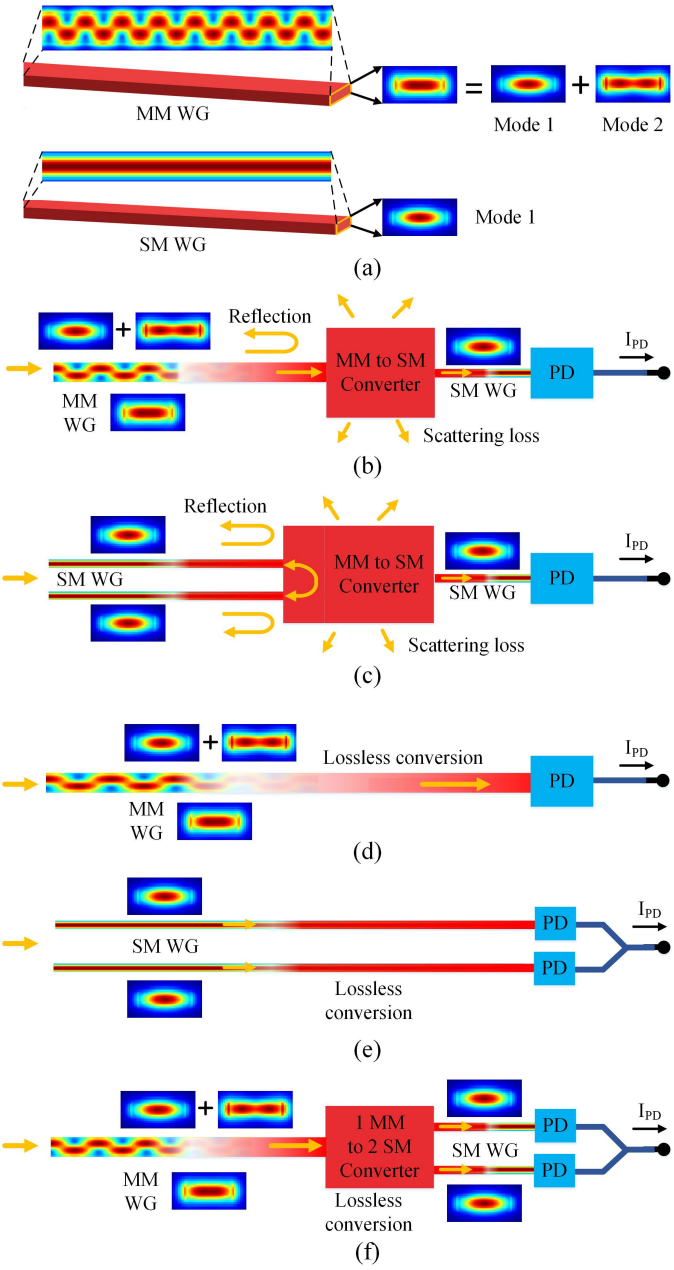

Figure 7.3: Mode simulations(a) Multi-mode and single-mode waveguides (b) The two modes carried by a multi-mode waveguide combined into a single mode waveguide. (c) The two super-modes supported by the two single-mode waveguide combined into a single mode waveguide. (d) The power carried by a multi-mode waveguide input to a power detector that outputs the combined power. (e) The two super-modes of the structure are input to two power detectors and power combining happens after down-conversion to DC. (f) A multi-mode to multi-mode (super-mode of two single modes) is used before feeding the power into the photodetector.

and rules out this possibility. In other words, such a power combiner cannot be lossless or it will not have matched ports that causes power reflection at the ports, Fig. 7.3(b) & (c). This fact can be easily understood by changing the boundaries of the network and including the combiner inside the system, which yields an antenna with a single-mode port and leads to the limitations discussed. Therefore, the collected power should either be transferred using a multi-mode waveguide or should be down-converted by photodetectors first and then combined, Fig. 7.3(d) & (e). The non-linearity of the photo-detection process breaks the assumptions in our derivations and invalidates the discussed constraints. While lossless conversion of multiple modes into a single-mode is not possible, converting n modes carried by a structure into n or more modes supported by a different structure is possible. Examples of such mode converter structures are presented in [173]–[176] which convert the two modes of a multi-mode waveguide into two super-modes supported by two single-mode waveguides. Therefore, to improve the photodetection process, multi-mode to multi-mode conversion can be used first, Fig. 7.3(f).

There are several ways to add more modes to the structure of the antenna. A simple illustrative structure is the dual-port antenna of Fig. 7.4(a), where a super-mode is formed using two single-mode waveguides such that the field distribution of each waveguide is not disturbed. This is done by placing the two waveguides on opposite sides of the antenna. The two states of the super-mode formed by the two waveguides are shown in Fig. 7.4(b) which can be approximated with the individual modes of the two waveguides with high accuracy. To keep the design simple and the number of degrees of freedom small, a periodic grating of 6 silicon slabs is used to shape the field distribution on the antenna aperture. Since only two modes are assumed, the design space provided by the grating structure is flexible enough to yield high radiation efficiency and effective aperture for both modes. To maximize the FOV, the FOV range of the two modes are designed to be separated, Fig. 7.4(c). Therefore, the FOV of the antenna is doubled compared to a single-mode antenna of the same kind in a low-cost silicon photonics process. The grating period is $665nm$ with 25% duty cycle. Moreover, an etched silicon slab facilitates the forward propagation of the field as well as providing a good matching at the input ports. This antenna achieves a peak effective area of $2.9\,\mu m^2$ with total FOV of 30° by 30°.

Separation of the FOV of the two modes maximizes the functional angular range of the antenna but results in a deep notch at 0°. Using the same structure, it is also possible to achieve a monotonic angular range for FOV and adjust the ripple of effective aperture/directivity inside the FOV range. Figure 7.5 shows a design

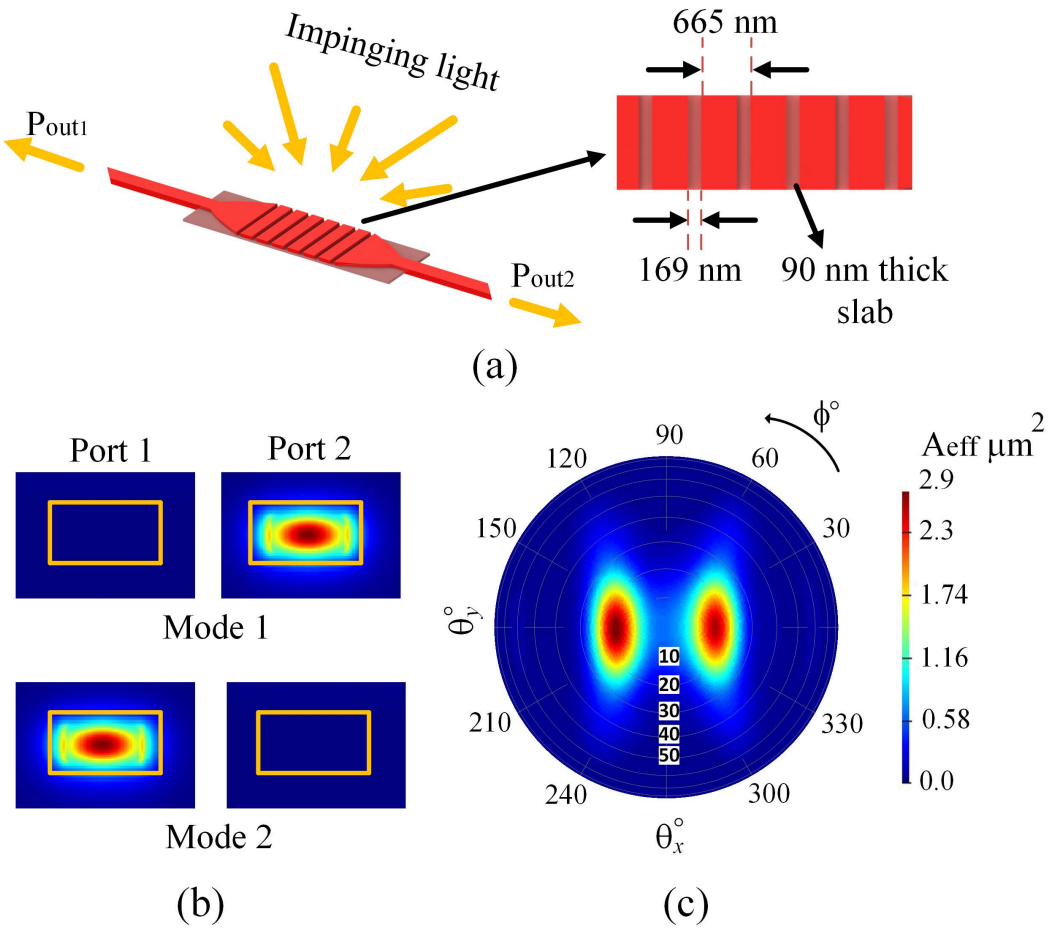

Figure 7.4: Multi-mode radiators (a) Dual-mode antenna which uses two single mode waveguides to form two super-modes carrying the optical power output by the antenna (b) Cross section of the supper-modes of the antenna propagating in the waveguides (c) Total FOV and effective aperture of the antenna versus azimuth and elevation angles.

with a single large angular range of FOV of 40° by 15° achieving effective area of 2.5 μm^2. Since the FOV of the two modes overlap, the peak effective aperture regarding the total collected power by the two modes can be more than the design of Fig. 7.4. In other words, since the trade-off is relaxed by a factor of two in the 2-mode structures, if FOV is reduced the effective area inside the FOV will increase.

Extrapolating the methodology used for the dual-port antenna of Fig. 7.4 and Fig. 7.5, an example of a 3-mode antenna is shown in Fig. 7.6(a). It constructs super-modes using three single-mode waveguides with 0.5 μm gaps between them to increase the product of FOV and effective aperture. The three waveguides collectively support the three modes of the antenna port, as shown in Fig. 7.6. For

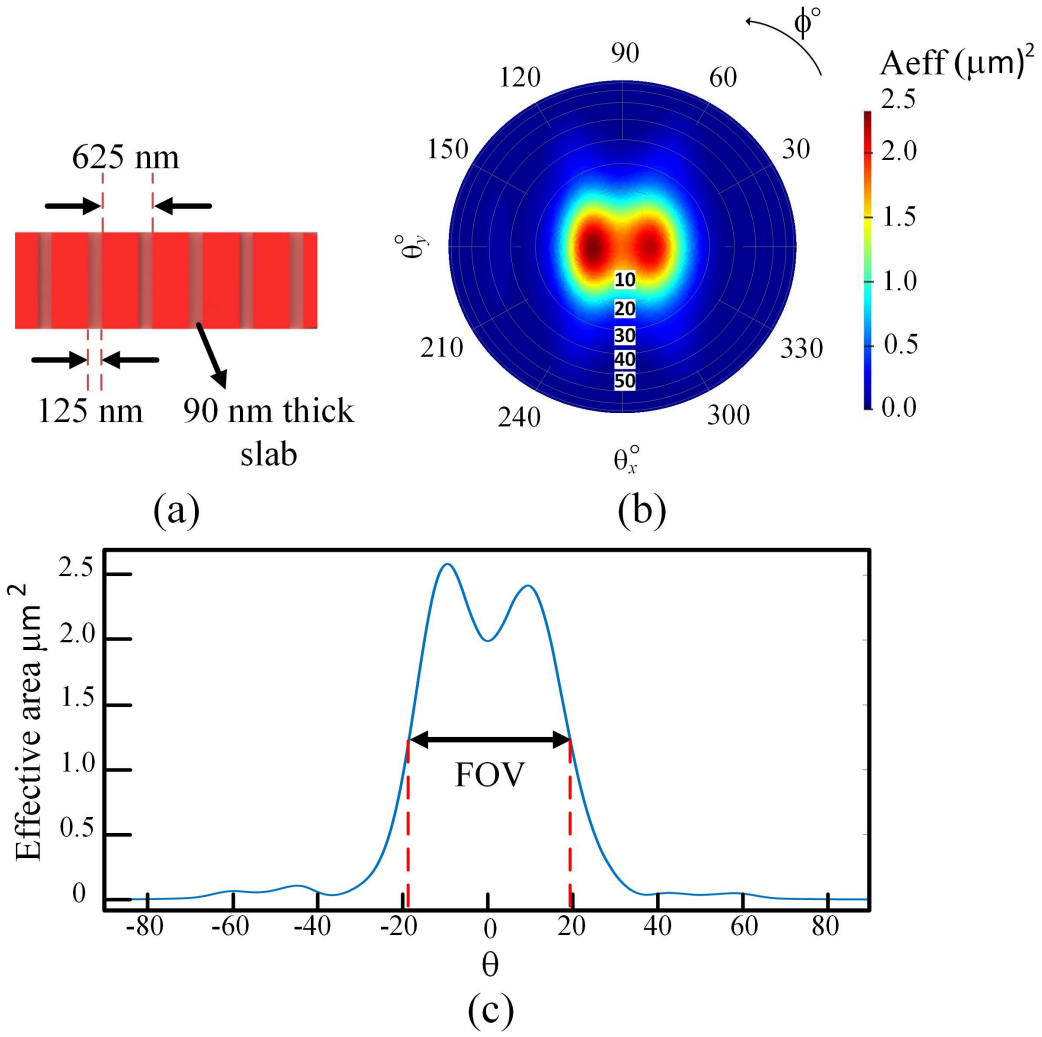

Figure 7.5: Multi-mode grating coupler simulation (a) Design parameters of the dual-mode antenna with a monotonic FOV (b) Total FOV and effective aperture of the antenna versus azimuth, ϕ, and elevation, θ, angles (c) Effective area of the antenna versus θ for $\phi = 0$ which shows a 40° of FOV.

adjusting the electric field distribution on the antenna aperture, a standard grating structure has a limited design space in this case and cannot be used to achieve good performance over a broad range of angles. Therefore, we adopted a patterned dielectric slab for the antenna area and exploited an optimization method to generate an etching pattern on the slab. The etching pattern consists of 150 nm blocks that are either etched or left unetched. Therefore, the etching pattern can be represented by a binary pattern, which makes genetic algorithm a suitable optimization method for the problem [145]. The physical antenna aperture is chosen as a relatively small

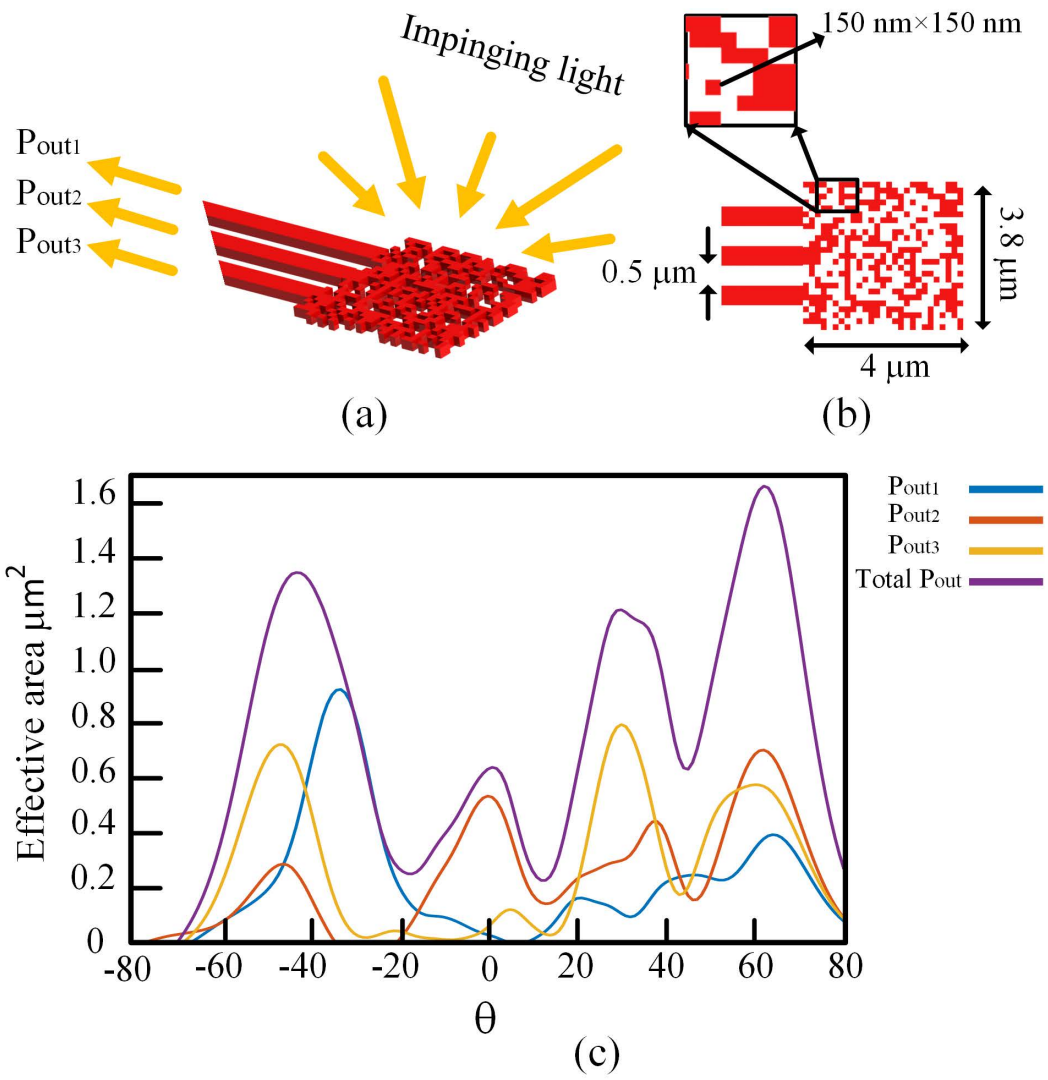

Figure 7.6: 3-mode radiator (a) Schematic of the designed 3-mode antenna which uses three single mode waveguides to form the super-modes carrying the optical power output by the antenna(b) Design details of the antenna aperture formed by etching a binary pattern of 150 nm squares (c) Total effective aperture and individual effective aperture provided by each supported mode versus elevation angle.

square size of 3.8 μm by 4 μm. Therefore, FOV of each mode is larger than the 2-mode antenna presented above. The effective aperture of the antenna is shown in Fig. 7.6(c) exhibiting a broad FOV range with a peak effective aperture of 1.6 μm^2. Figure 7.6(c) also shows the effective aperture versus angle for the three modes individually which illustrates the contribution of each mode in the overall power collection capability of the antenna.

Constructing a super-mode using multiple uncoupled single modes yields a larger

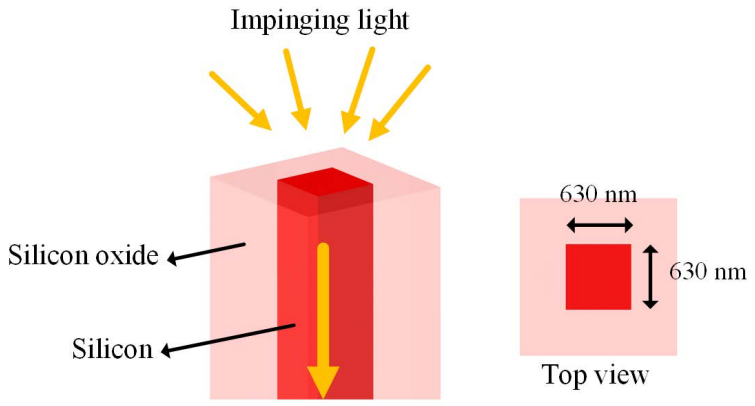

Figure 7.7: A single pillar as a non-planar antenna that supports 4 mode pairs that each pair collect power from two orthogonal polarization.

antenna footprint as the number of modes increases. An alternative structure is an antenna with a single physical port that supports multiple modes. To collect the output signal of the antenna, a single waveguide that supports those modes and is matched to the input port of the antenna at each mode is needed. In a standard silicon photonics process with a silicon layer thickness of 220 nm, the width of a single-mode waveguide is 0.5 µm and larger waveguide width introduces more modes that are supported by the waveguide. Since the group indexes of different modes of a multi-mode waveguide differ a lot, a grating structure is not very effective to construct the antenna aperture and a patterned slab similar to Fig. 7.6 is needed to achieve high performance.

While an integrated planar multi-mode antenna significantly increases the collection area compared to the conventional nano-photonic antennas, non-planar antenna designs can potentially lead to solutions with higher fill factor and broader FOV due to the addition of the third dimension that can be used to support additional modes. To demonstrate this possibility through an example, we propose a multi-mode silicon pillar antenna with a square cross section of 630 nm by 630 nm surrounded by silicon oxide. Although air as the surrounding material results in better confinement of the optical field in the silicon pillar, silicon oxide is used here for better mechanical stability and fabrication yield. The dimensions and the higher dielectric constant of silicon compared to the surrounding silicon oxide defines the supported modes of the antenna. The antenna supports four pairs of modes and each pair covers two orthogonal polarizations of the same kind, as in Fig. 7.8(a)-(h). The difference between the electric field distribution of the mode profiles, the amplitude distributions

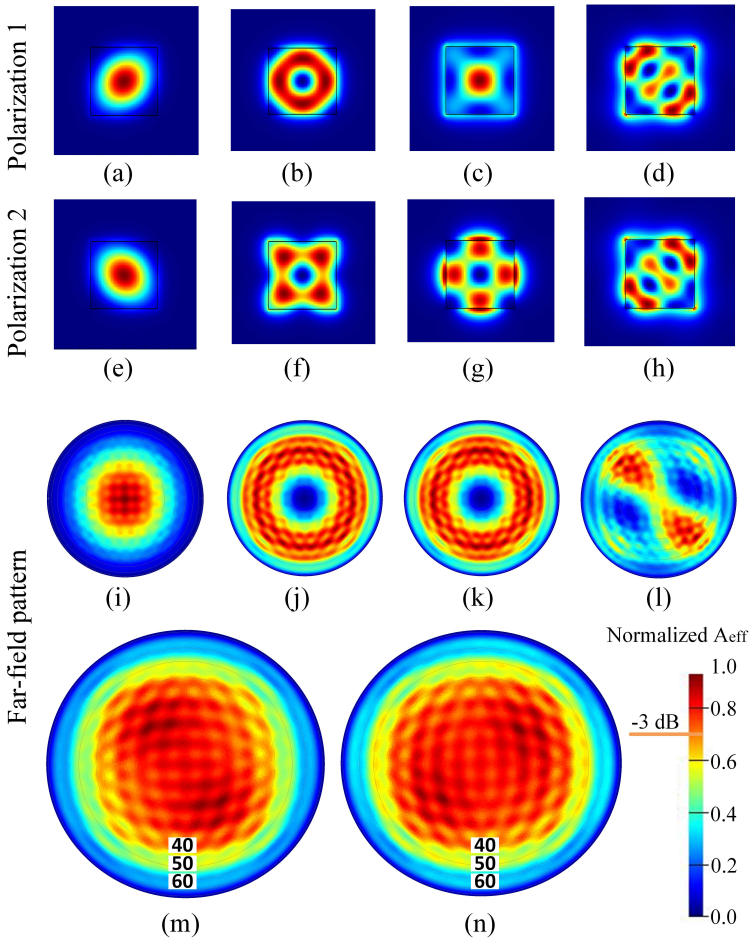

Figure 7.8: The graphs are normalized and the color bar shows the relative intensity (a)-(d) Modes of the polarization set 1 (e)-(h) modes of the polarization set 2 (i)-(l) Effective area of the polarization set 1 modes (effective area for the set 2 is counterpart of these graphs) (m) Collective effective area of set 1 (n) collective area of set 2.

as well as their phase distributions, leads to a different far-field radiation pattern for each mode, where the effective aperture of each mode covers different angular ranges. The reciprocity theorem can be used for each mode separately to obtain the antenna effective aperture corresponding to that mode, which are shown in Fig. 7.8(i)-(l). The collective effective aperture for the two polarization sets are shown in Fig. 7.8(m) and (n) achieving a broader overall antenna FOV[4]. As mentioned before, a single-mode antenna has a maximum collection area for a particular polarization and zero collection capability for the orthogonal one for any given (θ, ϕ). However,

[4]The radiation patterns associated with different modes have finite overlap, which can never be fully eliminated [177]. However, such overlap does not present a fundamental challenge to increasing the FOV.

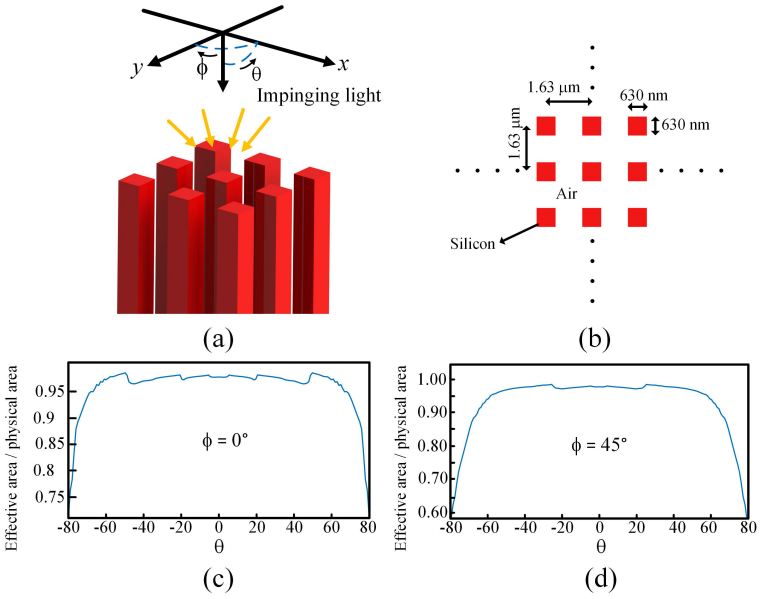

Figure 7.9: Array simulation (a) Dense array of pillar antennas (b) top view of the array with dimensions of each unit cell (c)&(d) power collection efficiency of the nit cell versus azimuth angles 0° and 45°

in this structure, not only a large FOV is achieved, but also the pillar antennas have large effective aperture and FOV for both polarizations of the incident wave.

The compact form factor of this pillar antenna makes it a suitable candidate for a dense receiver array aperture with broad FOV and large fill factor. Due to the large index contrast of silicon with the surrounding oxide medium, the modes of the single antennas will not get affected significantly by the neighboring array elements [178], and the major modes of the array will essentially resemble the modes of a single antenna. Figure 7.9(a) shows the antenna array incorporating pillar antennas as array elements. The element spacing of the array and the dimensions of the antenna are optimized to maximize the power collected by each unit cell. The ratio of the collected power by each unit cell to the total incident power (Fig. 7.9(b) & (c)) is more than 95% for a broad range of FOV exceeding 140° by 140°. It should be noted that the effective area of each antenna is close to its upper bound in an array, as discussed previously. Noticeably, both FOV and effective apertures are pushed to their limits independently in this architecture.

In a photonic receiver array, the area of the photo-detector is generally a fraction of the unit cell due to the space occupied by various other photonic and electronic components as well as metal interconnects and interfaces. To achieve a large fill

factor, an antenna should receive and confine the incident light into a smaller volume so that it can be coupled into the photodetector efficiently. Hence, the dimensions of the pillar antenna and the array element spacing are optimized such that it confines the optical power into a 0.6 µm by 0.6 µm window in the vicinity of the pillar while maintaining a large effective aperture. While the exact sizing of the focusing window is determined by the photodetector area in the unit cell, the 0.6 µm by 0.6 µm window size is a good estimate for the detector size (considering the wavelength-scale unit cell size to have a dense enough array for sampling the wavefront) to demonstrate the focusing capability of the antennas. Figure 7.10 shows the power focusing capability of the array into a small region of the antenna. Since the propagation speeds of different modes are not the same, a periodic intensity distribution is formed that peaks where all the modes are in phase. Since the effective apertures of the modes vary versus the reception angle, the relative magnitude of the modes changes accordingly. Therefore, the focusing capability of the antenna is affected for larger angles. A metric for an effective aperture of the focusing, which takes into account only the fraction of the power in the 0.6 µm by 0.6 µm window is a more precise measure of the performance of the antennas in an array. Figure 7.10(e) & (f) show the focusing effective aperture of the antenna versus the reception angle. While the uniformity of the effective aperture is not the same as the full unit cell, the 3 dB FOV of the aperture exceeds 160°. The full width at half maximum (FWHM) window size of the focused power is 0.58 µm by 0.58 µm which contains 72% of the total power received by the unit cell. This corresponds to the effective area of 1.92 µm^2. The performance of such a radiator must be evaluated within the array, as the interactions with the adjacent element change the individual antenna behavior compared to a standalone pillar. Therefore, each unit cell is an antenna unit with periodic boundary conditions. For a photodetector size of $0.58 \times 0.58\,\mu m^2 = 0.34\,\mu m^2$, the dielectric antenna array increases the effective multi-mode collection area by

$$\eta_{ap} = \frac{A_{eff}}{A_{phy}} = 5.7, \tag{7.6}$$

in which A_{eff} is the effective collection area and A_{Phy} is the physical area of the photodetector. Large aperture efficiency and FOV suggest the application of this structure for a dense optical phased array coherent receiver. To further clarify the merits of this structure, it can be compared to the recently published optical phased array receivers [139] which uses conventional nano-photonic antennas in an array for collecting the incident power. The antenna in [139] has 0.92 µm^2 of peak effective aperture at −6° with unit cell size of 11.2 by 11.2 µm^2, where it achieves an aperture

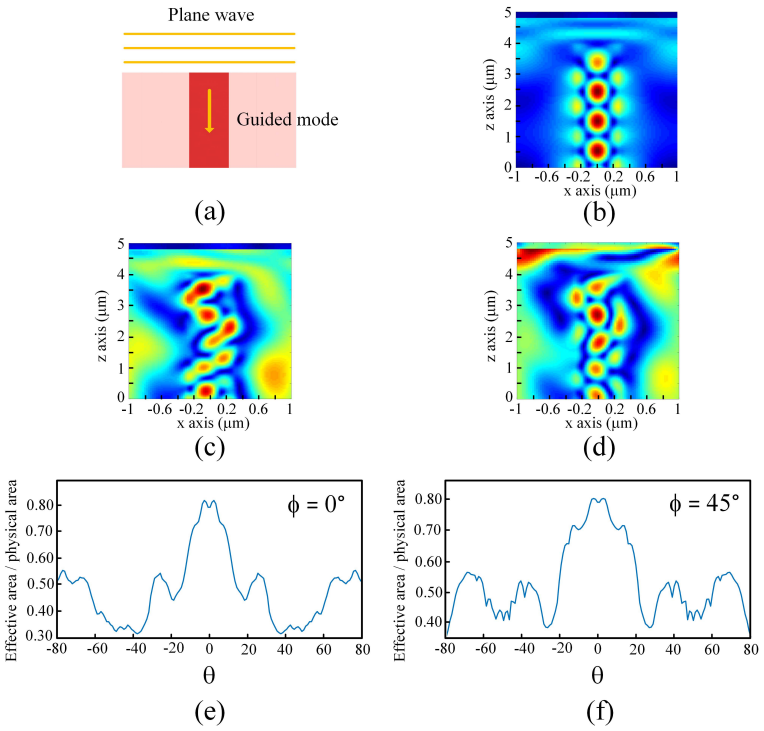

Figure 7.10: Power collection efficiency of the array in a 0.6 μm by 0.6 μm window around the pillar antenna (a) azimuth angle = 0° (b) azimuth angle = 45°.

efficiency of 0.0074 while the array element spacing limits its FOV to 8°.

7.3 Coherent Detection with Large FOV and Large A_{eff}

The proposed multi-mode pillar antennas forming a wide-FOV dense receiving array are not only compatible with a direct detection architecture (implicitly discussed above), but also can be used with a heterodyne detection scheme to realize a coherent imager by recording both phase and amplitude information of the wavefront. Nano-photonic receiver architectures featuring heterodyne detection have been shown to achieve high sensitivity, and robustness to interference and noise which makes them a suitable candidate for various applications such as imaging [139] and ranging. Figure 7.11 shows the system level diagram of the proposed multi-mode coherent detection array in which each port represents a single mode and each mode has its own received pattern. If an optoelectronic mixer [139] is used before the power summation node, then the phase of the wave components in each mode is preserved after mixing. The optical mixer includes a directional coupler to combine a reference optical signal acting as a local oscillator (LO) (with an optical frequency of ω_L) and the input wave, as well as a photodiode pair (*e.g.*, made of germanium in a

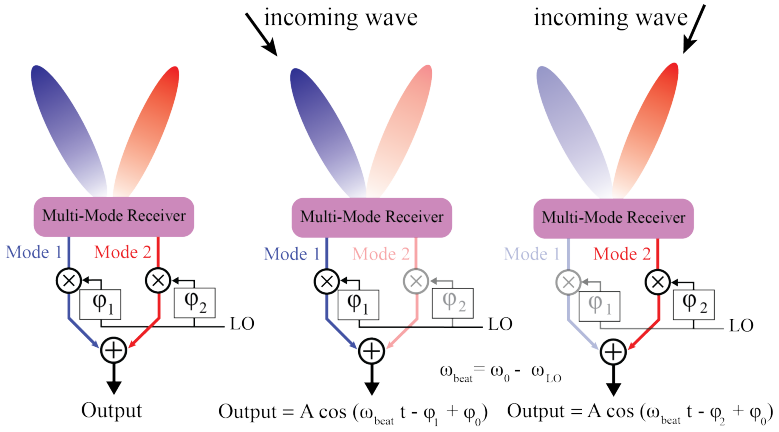

Figure 7.11: Heterodyne detection system model in multi-mode antennas.

silicon photonics process) to convert the optical wave into an electrical current. Moreover, the phase shifters in the LO path, symbolically shown as ϕ_1 and ϕ_2, allow for controlling the output phase of the mixer. As shown in [139], [179], the output current of an optoelectronic mixer is proportional to the signal strength at the beat frequency, $\omega_b = |\omega_0 - \omega_L|$, namely,

$$I_{out} \propto cos((\omega_0 - \omega_L)t + \phi_0 - \phi_L), \tag{7.7}$$

where R is the responsivity of the photo-diode and ω_0 and ω_L are illumination and LO wave frequencies, respectively.

While the modes are assumed independent in the system-level representation of Fig. 7.11, they share the same spatial region and have considerable overlap in the multi-mode pillar antenna discussed in the previous section. To increase the separation between the modes, a pair of coupled pillars can be used as an antenna, Fig. 7.12. This system consists of several pillars that are arranged in close vicinity of each other, where light can couple from one to another. There are several supported modes in this structure and each mode has a different reception pattern. Figure 7.12 shows two 0.6 µm by 0.6 µm coupled pillars with 0.2 µm gap between them are used as a unit cell of an infinite array with 3 µm cell pitch. This structure supports two main modes, the symmetric and anti-symmetric modes (Fig. 7.12). The symmetric mode refers to a situation where the excited waves in both pillars have the same phase and amplitude, also known as the common mode. On the other hand, the anti-symmetric mode (a.k.a. the differential mode), corresponds to the excitation in which the phases of the wave in the pillars are 180° out of phase. The other supported modes by the coupled pillar antenna are shown in Fig. 7.12(a). These modes have effective indices

near silicon which means they are mostly confined inside the silicon pillars. Also, polarization of these modes are close to the direction of y axis. For an incident plane wave with an electric field polarization in y axis direction, E_y, only symmetric and anti-symmetric modes have the exact same polarization matching. The far-field radiation patterns of these two modes are shown in Fig. 7.12(b) showing that they collectively cover a larger FOV than individually. Therefore, the light coming from different angles couples into different modes of the antenna structure.

To understand the behavior of this structure, we decomposed the electric field excitation inside the silicon pillars into the supported modes for two exemplary cases: a plane wave with E_y polarization impinging at 0° and a plane wave with E_y polarization arriving at 20°. As shown in Fig. 7.13(a), for the first scenario, most of the power couples to the symmetric mode (mode 1). As the illumination angle increases, some of the power shifts to the anti-symmetric mode (mode 4). Since the polarization only matches with these two modes, the coupled power to the other supported modes is negligible. Also, part of the power is coupled to the extra supported modes associated with the array structure which are mainly concentrated in the inter-pillar space with effective indices around 1. The properties of these modes can be manipulated by covering the pillars with silicon oxide (Fig. 7.14). Figure 7.13(b) illustrate the total electric field underneath the silicon pillars for the two illuminations at 0° and 20°.

Since for illumination at 0° the optical power mostly couples into the symmetric mode, light travels through both pillars. Figure 7.13(b) shows the electric field magnitude for this case in the X-Z cut. As the angle of illumination increases, both symmetric and anti-symmetric modes are excited. Therefore, the total electric field is the superposition of these two modes. Due to the phase distribution of the two mode profiles, superposition results in having optical power mostly in one pillar rather than both of them, as seen in Fig. 7.13(b), while the effective aperture of the antenna is maintained. According to the simulation results, the 3 dB FOV of this structure is more than 50°.

The discussed heterodyne detection scheme can be used with this pillar array configuration by routing a waveguide carrying the LO signal underneath the pillar pairs as shown in Fig. 7.14 (a). Here, a slab of germanium operates as the mixer and the combiner, shown with building blocks in Fig. 7.11. The mixing is achieved through the beating between the LO and the coupled incident light followed by the absorption within the germanium slab. The absorption of light generates temporally varying

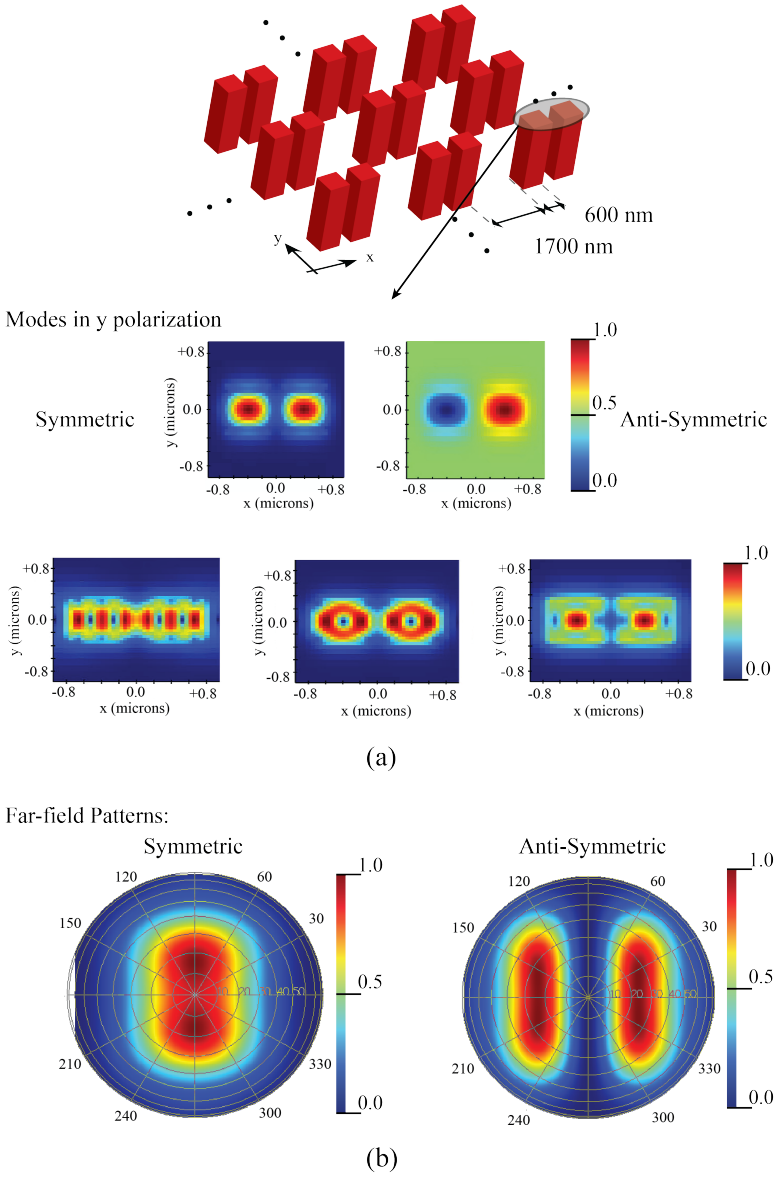

Figure 7.12: Coupled pillars (a) coupled pillar structure with supported modes, (b) far-field pattern of symmetric and anti-symmetric modes.

electron-hole pairs with phases determined by the LO and input wave phases at each point. The electrons and holes are then swept to the two ports of the photodetector and the total output current is the coherent sum of all these regional currents. As mentioned, for different illumination angles, the superposition of symmetric and anti-symmetric modes controls the power balance in the two pillars, Fig. 7.13(a) (b). This means that there are two main hot spots underneath each pillar within the germanium slab where most of the mixing and combining occurs. These two spots are not in-phase and there is a phase difference for the electromagnetic field present

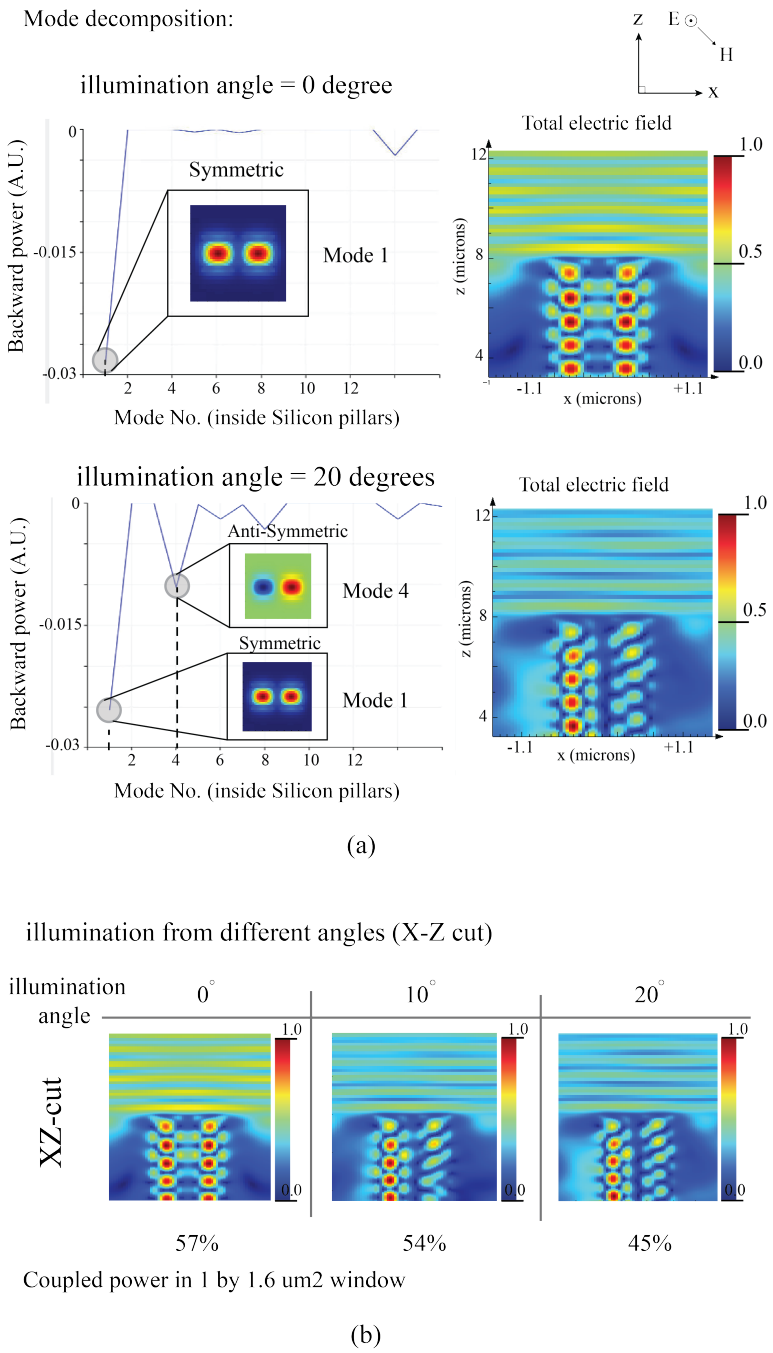

Figure 7.13: Coupled pillars mode expansion (a) Mode expansion simulation at different illumination angles. (b) Electric field magnitude in X-Z cut vs different illumination angles.

at these two places. Therefore, to achieve a coherent sum and combine the powers of the two modes, as the LO signal propagates in the structure, it should experience a phase shift equal to the phase difference of the two hot spots which yields

in-phase beat components. This is done by adjusting the physical dimensions of the structure and controlling the propagation constant of the waves. Consequently, the LO waveguide, its wavelength, and the gap between two pillars as well as their sizes must be designed in a way that two beat-frequency signals have the same phases (Fig. 7.14) and add constructively. To combine the LO wave with the received wave through pillars, the polarization of the electric fields inside the pillars and the LO waveguide have to be aligned (direction of y axis in this case). Here, the electric fields are combined inside the waveguide and a directional coupler used in opto-electronic mixer [145] can be eliminated. The germanium slab is added underneath the LO waveguide to convert the combined electric fields into electrical current. Connections to the germanium can be made by having doped p- and n-type regions next to the LO waveguide forming a PIN diode.

7.4 Conclusion

In this chapter, we studied the fundamental limits of antenna design for photonic transmitter and receiver systems and proposed architectures for designing high-performance antennas. For the receiver antenna, maximizing the collection efficiency while maintaining the FOV of the receiver is the desired design target due to the typically weak received signal levels. However, the strong relationship between the antenna characteristics in the transmit and receive modes introduce trade-offs in the design space of conventional antennas. In addition, a photonic coherent receiver achieving more than 95% collection efficiency and 170° of FOV utilizing a multi-mode dense pillar antenna array is demonstrated.

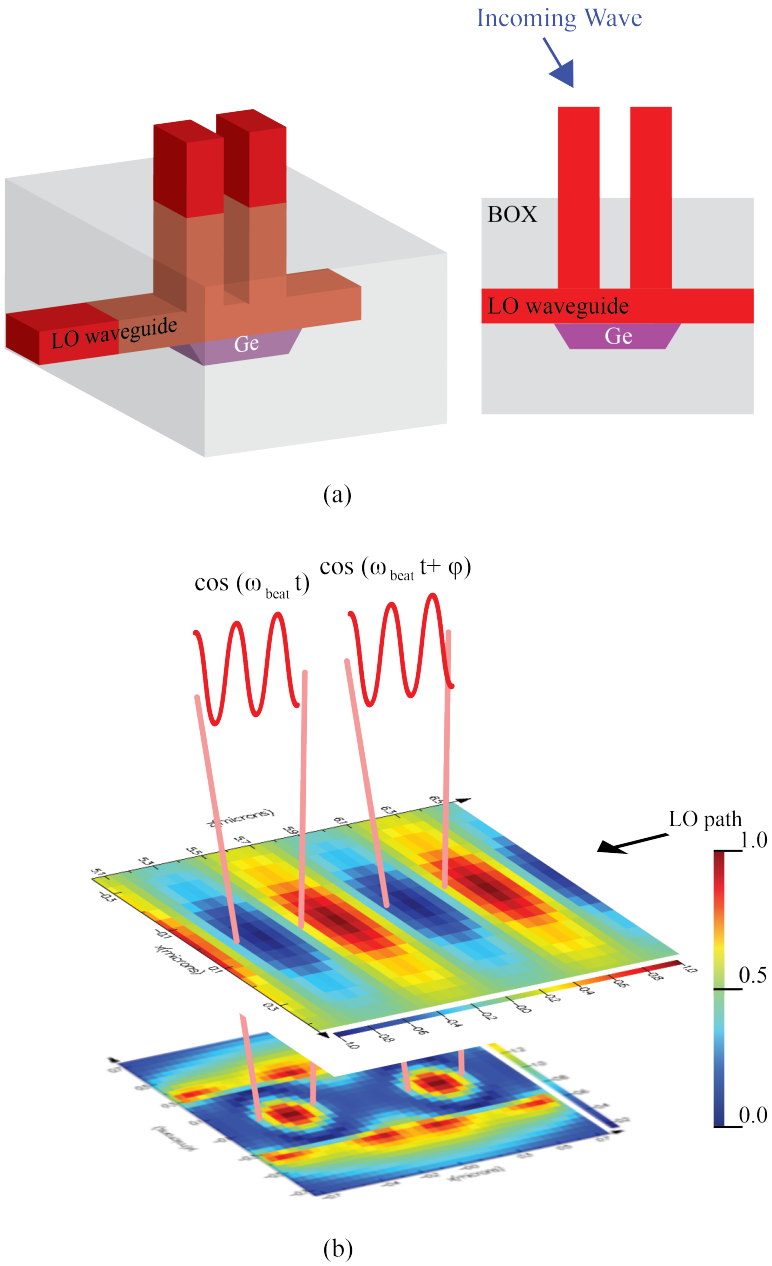

Figure 7.14: Proposed heterodyne detection scheme (a) Heterodyne detection combined with two-pillar structure. (b) simulated electric field strength at zero-angle illumination. Fields maximize under the pillars and the relative phase between LO wave and the illumination has to be the same at the hot spots.

Chapter 8

OTHER WORKS

8.1 Hybrid Transmitter and Receiver Optical Imaging System

Introduction

Optical phased arrays have made their way to various applications and areas [148], [150]. Despite many developments in photonics processes, fabrication limitations impose restrictions on photonics devices, including optical phased arrays. In a typical phased array configuration, $\lambda/2$ spacing between the elements results in the maximum greating-lob-free FOV ([180]). However, fabrication imposed limitations would not allow nanophotonic antennas to have $\lambda/2$ spacing ([145]). Routing nanophotonic waveguides and a single layer fabrication process severely decrease the FOV of optical phased arrays.

To mitigate this problem and increase the FOV of the optical phased arrays, a few methods have been proposed [145], [181]. Here, we present a novel architecture based on transmitter/receiver pattern overlap in order to increase the FOV with minimum complexity.

X Transceiver Architecture

Friis equation ([155]) describes the received power at the receiver in a transmitter/receiver system as:

$$P_r(\theta, \varphi, r) = \frac{P_t G_t(\theta, \varphi) G_r(\theta, \varphi) \lambda^2}{(4\pi r)^2} \tag{8.1}$$

where θ, φ, and r are the spherical coordinates. P_t represents the transmitted power, G_t and G_r are the transmit and receive antenna gain, respectively. In addition, λ is the operating wavelength of the system.

As equation 8.1 indicates, received power at the transceiver depends on θ and φ–two independent variables–to cover the whole space. It is noticeable that the dependency comes from $G_t(\theta, \varphi)$ and $G_r(\theta, \varphi)$. However, as it was mentioned before, the optical phased arrays are very limited in providing wide two-dimensional antenna patterns. Therefore, the performance of the transceiver will be limited to the transmitter/receiver overall FOV. This limitation is mainly dominated by the routing of the nanophotonic antennas in the phased array aperture [181]. Nevertheless, one-dimensional phased arrays can enhance the FOV because the elements are being fed

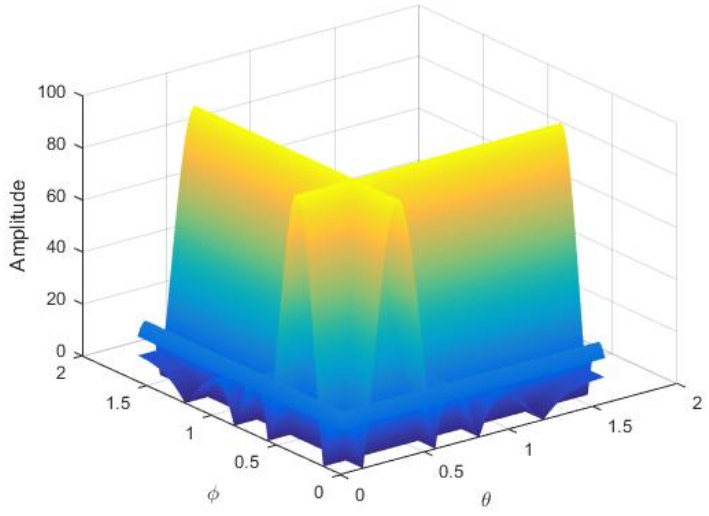

Figure 8.1: Transmitter and receiver patterns overlapping on top of each other.

from one end, and the radiation is from the top or the other end [182], [183]. Moreover, one-dimensional phased arrays necessitate variable frequency (wavelength) lasers to provide enough degree of freedom to sweep the two-dimensional space; however, using a single wavelength source is always desirable.

Here, we can combine the large FOV of the one-dimensional phased arrays and the Friis equation (8.1) by taking advantage of the $G_t(\theta,\varphi)G_r(\theta,\varphi)$ factor. Since the received power in a transceiver system depends on $G_t(\theta,\varphi)G_r(\theta,\varphi)$, we can distribute the θ and φ dependency between receiver and transmitter patterns. Therefore we will have:

$$P_r(\theta,\varphi,r) = \frac{P_t G_t(\theta) G_r(\varphi) \lambda^2}{(4\pi r)^2}. \tag{8.2}$$

Here $G_t(\theta)$ and $G_r(\varphi)$ are one-dimensional separated phased arrays. The overall transceiver system has a $2N$ degree of freedom, and it is equivalent to the N^2 single transmitter/receiver phased array. Fig 8.1 and Fig 8.2 demonstrate an example of two one-dimensional phased array patterns. The overall transceiver pattern is equivalent to a two-dimensional phased array with N^2 elements, and a beam can be directed into different directions.

Implementation and Simulation

Based on the proposed idea, we designed and implemented a transceiver optical phased array system with 16 elements on the receiver and 16 elements on the transmitter side. The total FOV is limited by the grating lobes of the phased array as

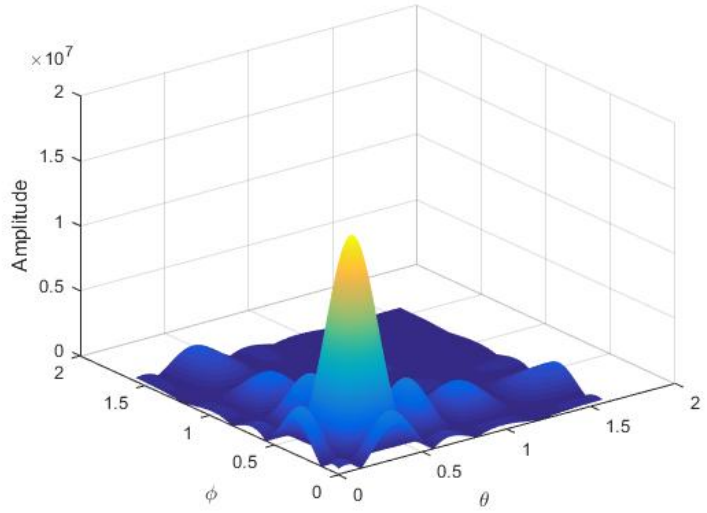

Figure 8.2: Overall transceiver pattern.

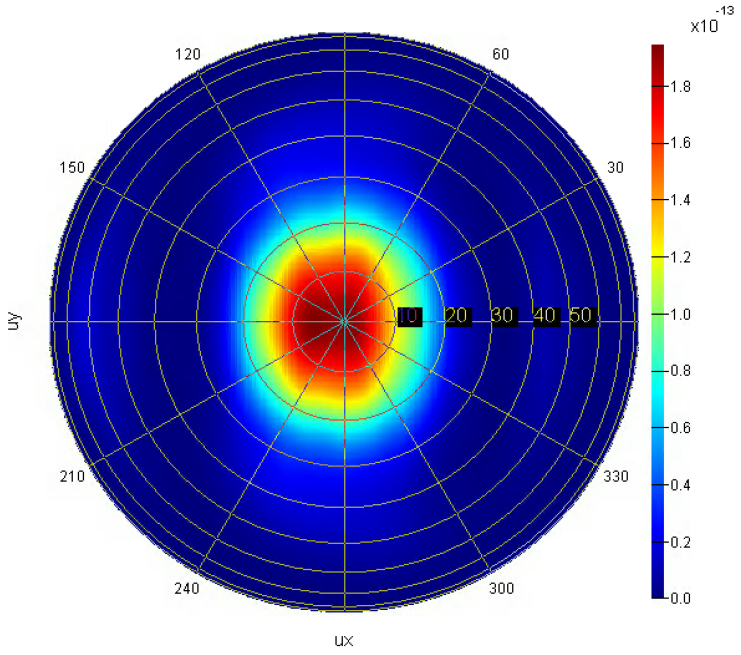

Figure 8.3: Grating coupler used in the transceiver.

well as the pattern of each transmitter element, mostly grating couplers. Therefore, it is desirable to use a grating coupler with a wide pattern. Fig 8.3 shows the radiation pattern of the used grating coupler.

Fig 8.5 shows the layout of the transceiver. The transmitter phased array is a typical

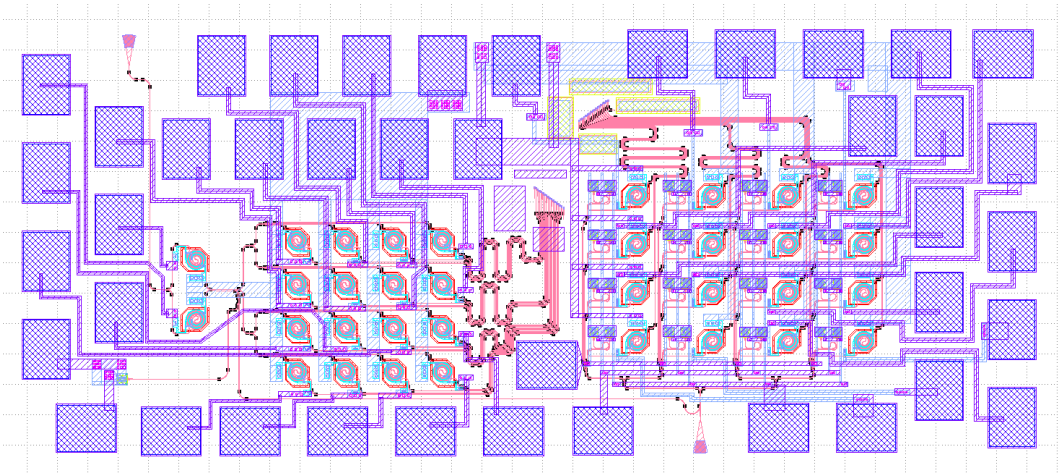

Figure 8.4: Implemented transceiver.

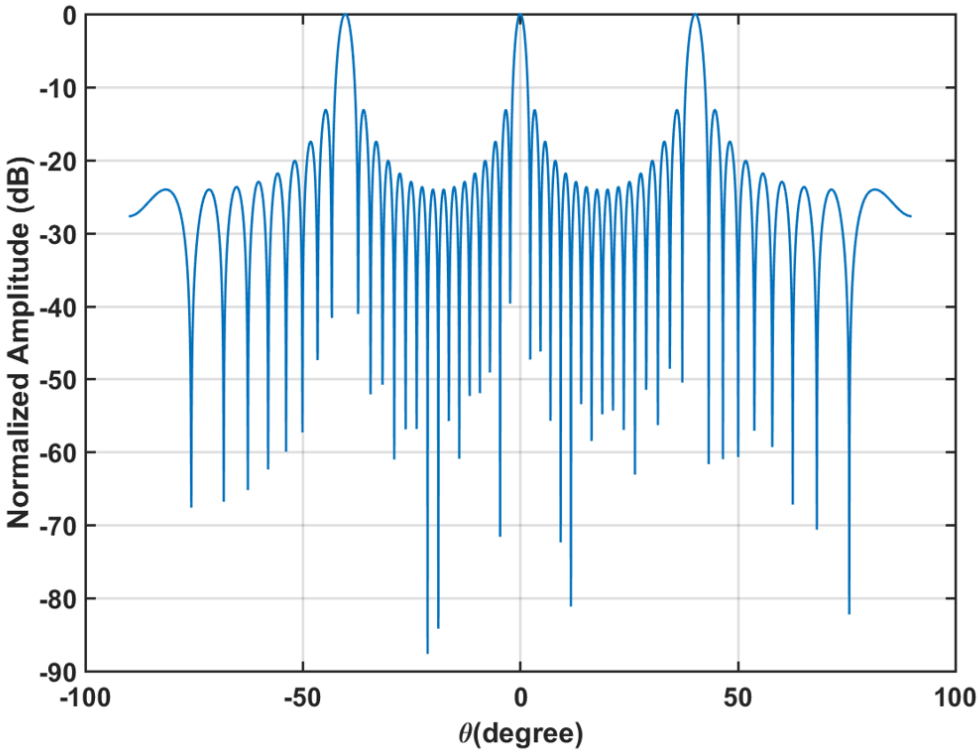

Figure 8.5: The pattern of the implemented receiver/transmitter phase array with 30° FOV.

optical phased array with 16 elements and an on-chip amplitude modulator–to separate the coupled light vs uncoupled light [150]. On the other hand, the receiver phased array, is an optical receiver with heterodyne scheme [144]. This architecture will reduce the receiver noise to the shot noise of the Germanium photodetector. the

beam-width of this structure around 2° and this transceiver potentially can achieve the resolution of 200 (resolvable) pixels.

Conclusion

In this work, we introduced and implemented a hybrid transmitter/receiver imaging system. This system distributes the degrees of freedom between transmitter and receiver in order to provide enough controllability. Friis transmission equation would ensure that the received signal is equivalent to a N^2 typical phased array system.

8.2 Nonreciprocal Transmitter/Receiver Architecture

Introduction

Phased arrays get lots of attentions because of their widespread use in communication, arbitrary field manipulation, ranging, and sensing applications like radar. All-electronic beam-forming capability, without having any mechanical parts, increases the flexibility and robustness simultaneously. Also, an electronically controlled beam can provide a higher resolution scan, faster scanning rate, and boosts the overall performance of the phased array.

Systems like radar and telecommunication need the ability to both rapidly transmit and receive data. Such systems can employ two antenna systems or use one whose time is split between receiving and transmitting, but this either occupies twice the amount of space or has half the bandwidth of a single antenna. Thus, it is highly desirable to implement a full duplex system: one that is able to transmit and receive data continuously with the same antenna. Such a system is realizable if both the transmission chain and reception chain can be connected to the antenna with a high degree of isolation between the transmitter and receiver.

In order to achieve full transmission from the transmitter to antenna and antenna to receiver, reciprocity of the system must be broken. Because of the "Lorentz reciprocity" theorem for time-invariant linear media, this cannot be done without using magnetic elements, active circuits, or time-varying circuits [184]–[186]. These non-reciprocal elements are usually some of the largest and most expensive parts in the system. Here, we present a scalable system that allows for isolated transmission and reception with independent N-element phased array patterns using a single non-reciprocal element. Such a system would allow for full duplex beam steering, critical to many radar systems and for the next generation of cellular communication

Nonreciprocal Transmitter and Receivers

A typical ring circulator with a non-reciprocal element is shown and being investigated in various fields and areas [187]–[189]. Here, we present a double-ring architecture which is shown in Fig 8.6. In this new configuration, there are two transmitter ports, two antenna ports, and one receiver port. However, in this design, both receivers share one non-reciprocal element.

Intuitively speaking, the working principle of this design can be seen in Fig 8.7 and Fig 8.8. For example, all the signal paths going from the antenna 1 to transmitter 1 is out of phase from all the signal paths from antenna 2 to transmitter 1. Therefore, the total signals from the antennas to the transmitter 1 cancel out. Moreover, it can

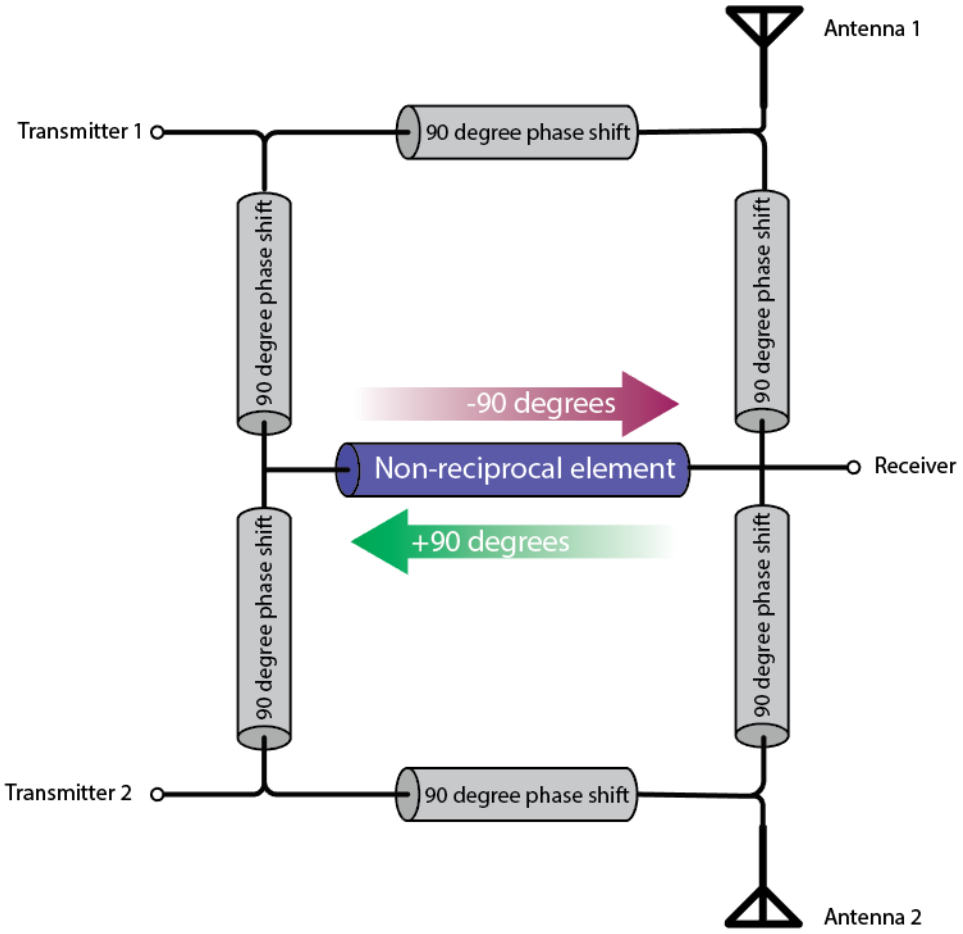

Figure 8.6: The new proposed double-ring circulator.

be seen that the signals from antenna 1 and 2 will add up in phase at the receiver. In total, all the in-phase signals from the antennas end up at the receiver and create isolation between the antenna transmitters.

Simulations

As an example, we designed such a system with ideal elements at $1GHz$ operational frequency. Fig 8.9 shows the input matching of the design to 50Ω except the receiver port with is matched to 50||50 = 25Ω. Fig 8.10 demonstrate the isolation from the transmitter ports to the antenna ports and Fig 8.11 shows the antenna to antenna isolation. According to these results, this architecture can behave as two separate

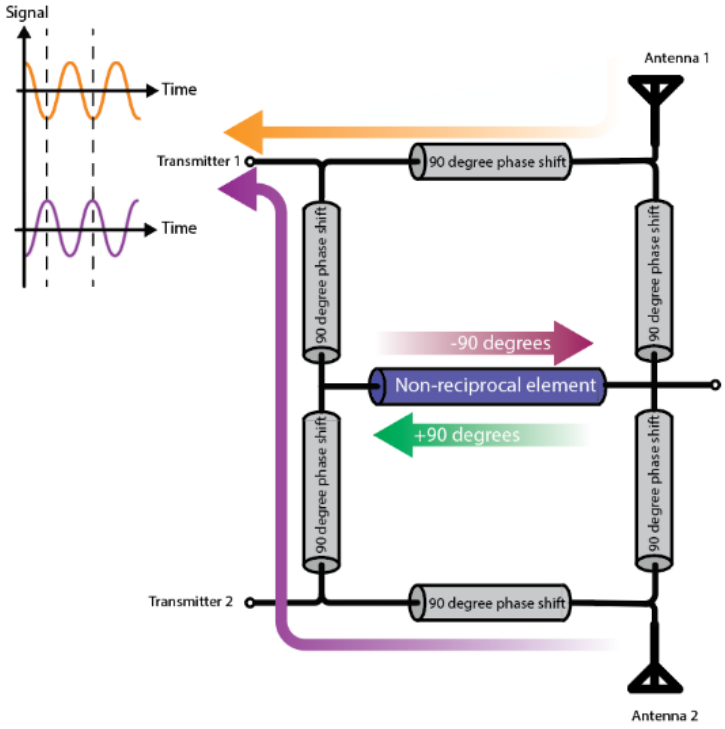

Figure 8.7: Cancellation from the antenna to transmitter ports.

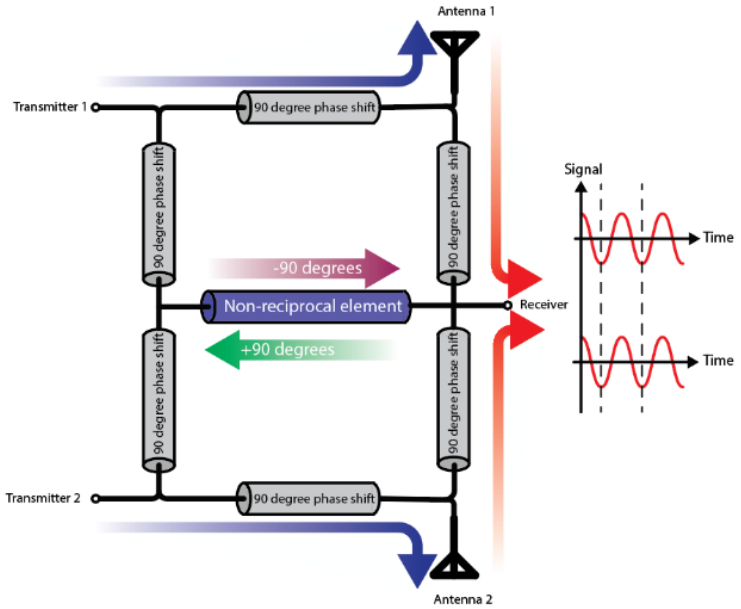

Figure 8.8: Addition at the receiver port.

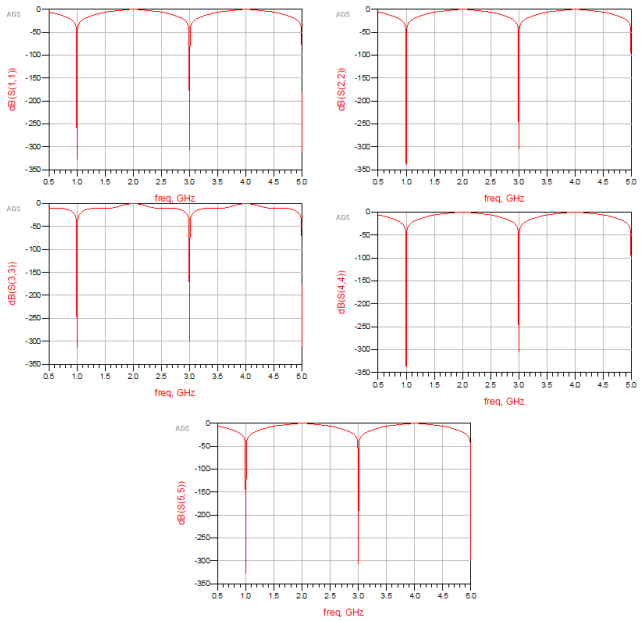

Figure 8.9: Input matching of all ports.

circulators with only one non-reciprocal element.

The concept of sharing one non-reciprocal element can be extended to N transceivers. Also, flipping the direction of the non-reciprocal element would change the receiver port to a transmitter port and all the transmitter ports to receiver ports. This scalable architecture is shown in Fig 8.12 and Fig 8.13. It is noticeable that in this design the shared receiver/transmitter port has to be matched to $50/N$ Ω.

Conclusion

Integrated circulators are one of the key components toward having full-duplex communication in future systems. Here, we proposed, designed, and simulated a multi-ring circulator which only uses one non-reciprocal element. Future wireless communication systems (5G and beyond) can potentially host this new design in their infrastructures.

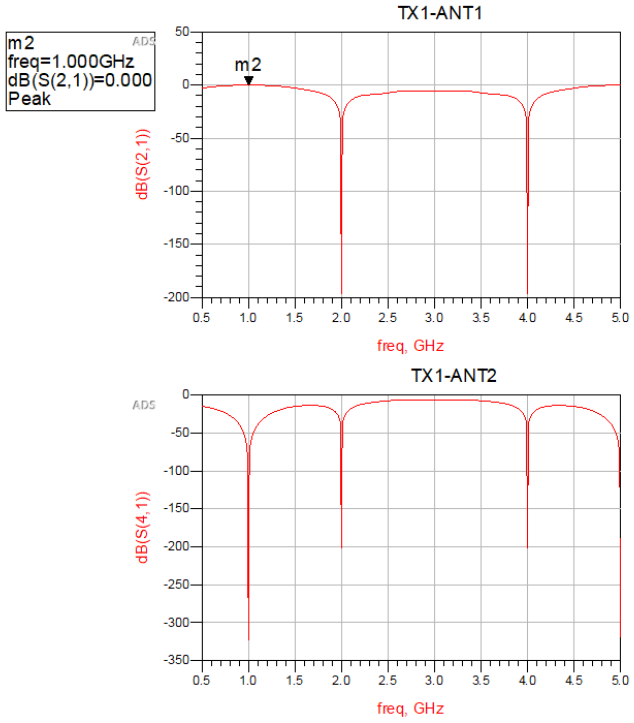

Figure 8.10: Transmitter to antenna simulation. All the power injected to transmitter 1 goes into antenna 1.

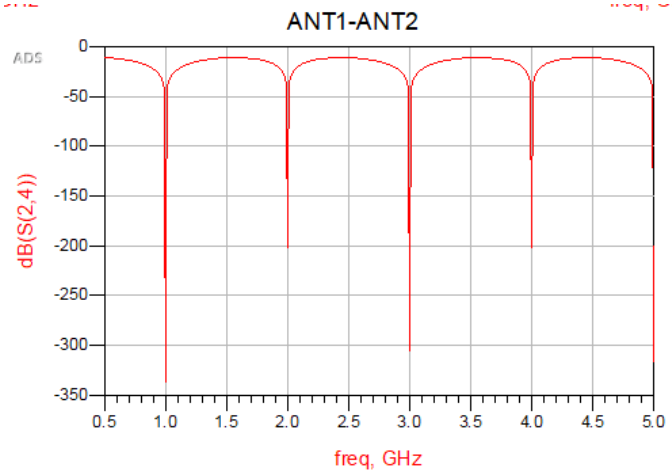

Figure 8.11: Antenna to antenna isolation.

8.3 High-Speed Nested-Ring-Assisted MZI Modulator

Introduction

High-performance electro-optic modulators are one of the fundamental building blocks that will enable the realization of high-density and complex integrated electro-

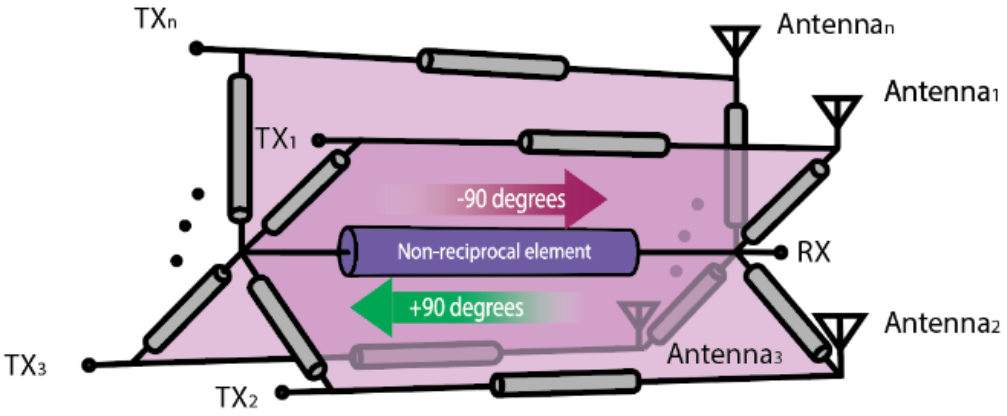

Figure 8.12: N-element transmitter architecture.

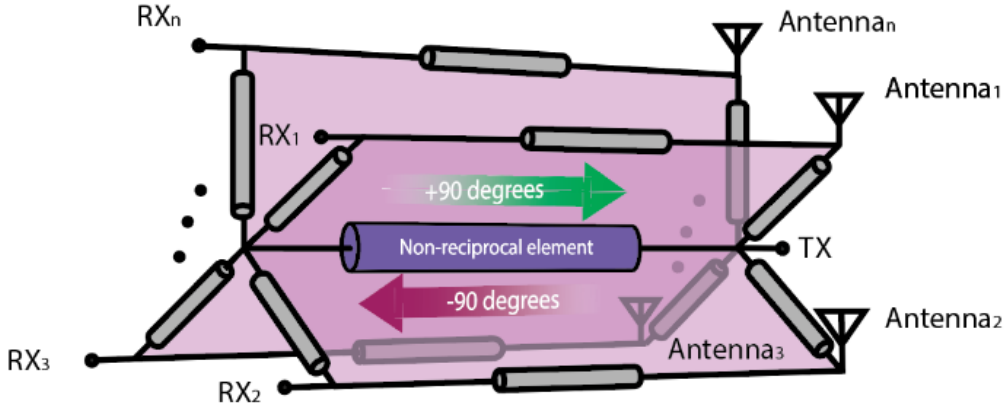

Figure 8.13: N-element receiver architecture.

optical systems [190], [191]. For example, data-communication applications require high-bandwidth and low drive-voltage modulators to improve the energy efficiency per bit of data [192]. There are typically two approaches to address these needs. First, it is the device choice for the electro-optic modulator, which comes with a broad range of trade-offs in efficiency, footprint, loss, and bandwidth. For instance, plasma dispersion modulators such as PN or MOS devices [190], [193] offer high bandwidth but have a large footprint and suffer from undesired phase to amplitude conversion due to the Kramers-Kronig relation. Alternatively, non-silicon materials with strong $\chi^{(2)}$ nonlinearity [192] have high bandwidth and do not suffer from phase and amplitude conversion, but come at an increased device cost.

The second approach for addressing the performance of integrated modulators is the architectural choice. In this perspective, most modulators can be grouped into

the two basic categories of resonance and interferometric structures. Resonance base modulators are very compact but require active resonance tuning to correct for undesired temperature variations. On the other hand, interferometer-based devices typically require large drive voltages and have a large footprint, but they are much less sensitive than resonance structures. In these modulators, the large size of the devices necessitates more complex electrical drive methodologies such as traveling-wave [194] structures.

Nested-Ring-Assisted MZI Modulator

For this design, each arm of the interferometer is a multi-ring serially-coupler filter (Fig. 8.14). The coupling coefficients of these modulators are optimized to achieve a flat pass-band in the intensity response and a linear behavior in the phase response of the structure's drop port. We designed the ring filter based on maximally-flat filter ring coupling coefficients in [195] and analyzed the ring frequency response using the wave propagation method described in [196]. The plot of the intensity and phase response at the drop port of the modulator is shown in Fig. 8.15(a),(b) for the different number of rings in the chain. In this simulation, the propagation loss is assumed to be 1 dB cm^{-1} in a 10 µm radius ring resonator. Fig. 8.15(b) shows that as the number of rings in the chain increases, the slope of the phase response also increases. This result is shown in 8.15(c). This linear response suggests that for a given desired phase response, the drive voltage requirement decreases linearly as the number of rings increases. Therefore, for a given drive voltage (V) and modulator capacitance (C), the energy required for modulation ($E_0 = 1/2CV^2$) drops linearly as the number of rings (N) in the chain increases ($E = 1/2(NC)(V/N)^2 = 1/NE_0$).

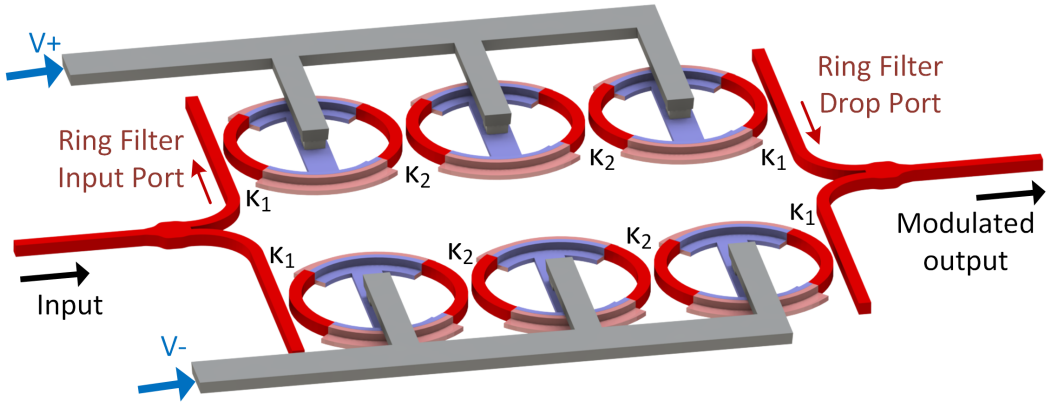

Figure 8.14: Serially-coupled ring resonator modulators as the arms of an MZI modulator.

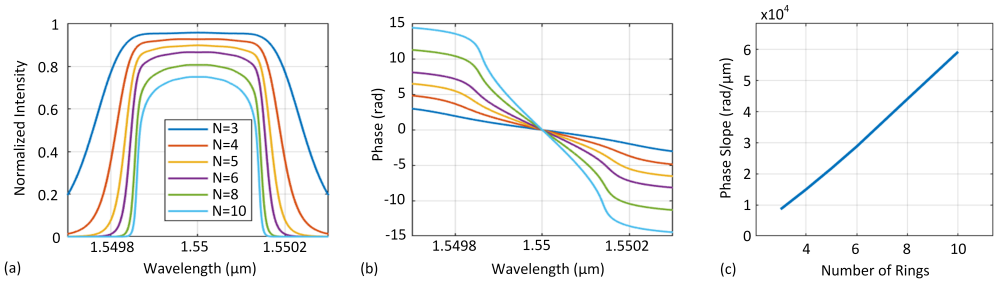

Figure 8.15: Serially-coupled ring modulator performance. (a) Drop-port intensity response. (b) Drop-port phase response. (c) Phase slope of the serially-coupled ring structure for the different number of rings.

We then proceeded to incorporate a three-ring resonator with maximally-flat response in the arms of an MZI interferometer as phase modulators as shown in Fig. 8.14 with $\kappa_1 = 0.4668$ and $\kappa_2 = 0.0990$ [195]. For this simulation, we assumed $2\,\mathrm{dB\,cm^{-1}}$ and $10\,\mathrm{dB\,cm^{-1}}$ for the waveguide loss and PN modulator propagation loss, respectively. Furthermore, we assumed $20\,\mathrm{V}\cdot\mathrm{mm}$ modulation efficiency. In addition, the ring radius was set to $20\,\mu\mathrm{m}$ with an $80\,\mu\mathrm{m}$-long modulator. From these values, we computed that a 2.5 V swing on these PN modulators will result in 0.01π phase shift in the ring. We then plotted the change in ring response for different applied voltages (Fig. 8.17(a) and (b)).

Based on these specifications, we designed a three-ring nested MZI modulator shown in Fig 8.16. In this design each coupling region is designed to satisfy the mentioned coupling coefficients. Also, due to fabrication imperfections, they can be further tuned to achieve the optimum performance. For tuning purposes, thermo-optical modulators are used in every coupling region. In addition, sniffer photodetectors are added to ensure that the maximum power couples into each ring resonator.

It is clear from 8.17(b) that a differential driver with 2.5 V swing is sufficient for inducing differential π phase shift in the arms of the MZI. The insertion loss of this architecture is 2.6 dB. If the same MZI were implemented with standard PN modulators, the modulator length would be 4 mm long, and the structure will have an insertion loss of 4 dB.

Conclusion

We proposed and demonstrated a compact, low drive-voltage, and low insertion-loss optical modulator based on serially-coupled maximally-flat ring resonator filters that have a much smaller footprint and 1.4 dB lower insertion-loss compared to the standard MZI modulators.

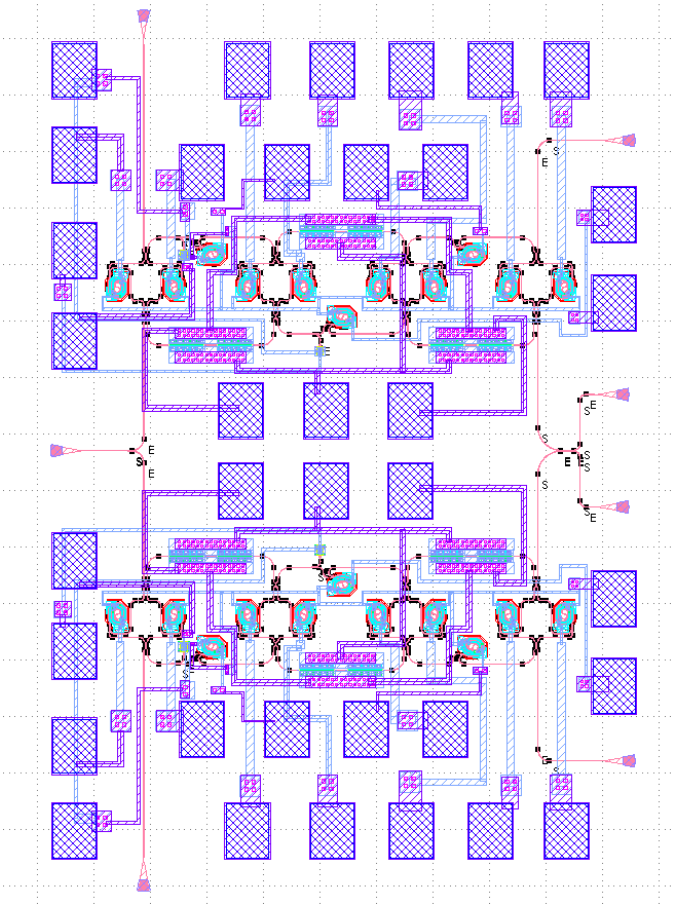

Figure 8.16: Designed MZI modulator with three nested ring architectures.

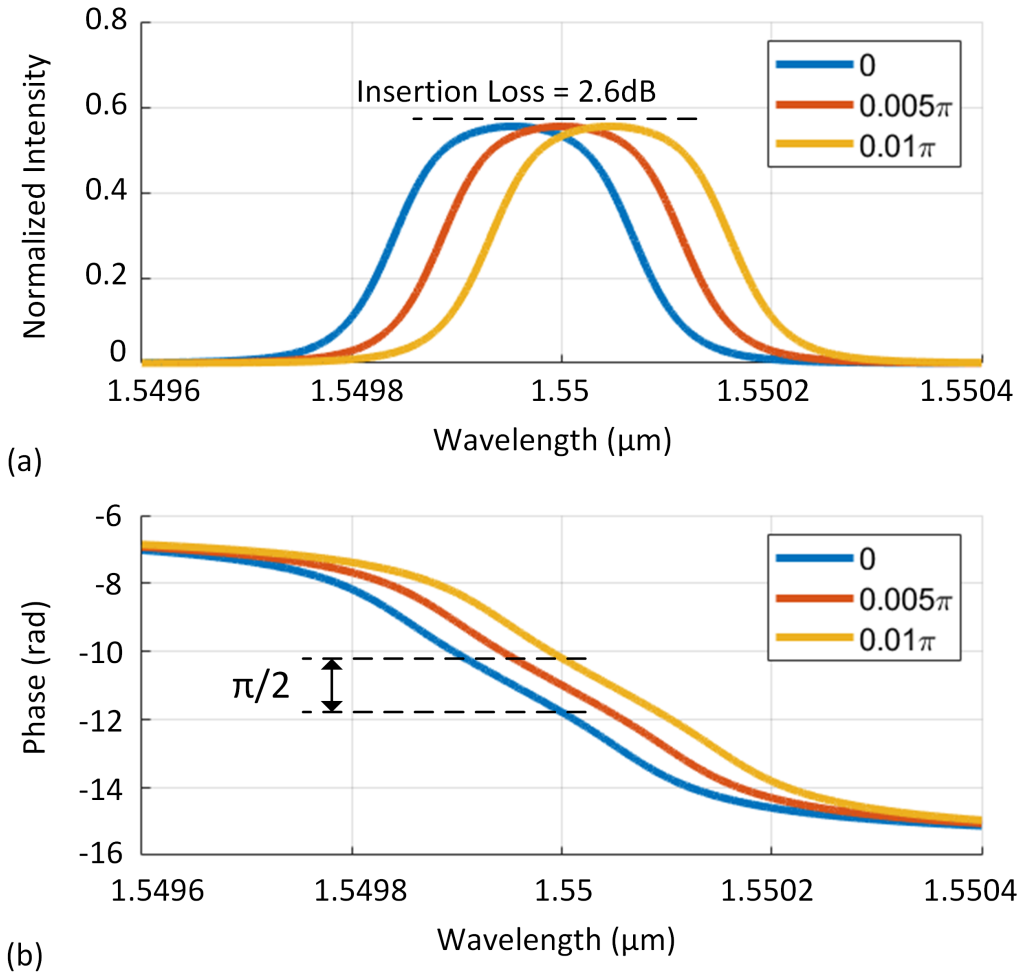

Figure 8.17: Frequency response of the proposed structure (a) Intensity response of the proposed serially coupled three-ring structure for different values of phase shift $(0, 0.005\pi, 0.01\pi)$ in the rings corresponding to 0 V,1.25 V, and 2.5 V bias voltages. (b) Phase response of the proposed structure for different bias voltages.

BIBLIOGRAPHY

[1] J. Bardeen and W. H. Brattain, "The transistor, a semi-conductor triode," *Physical Review*, vol. 74, no. 2, p. 230, 1948.

[2] J. Bardeen and W. H. Brattain, "Physical principles involved in transistor action," *Physical Review*, vol. 75, no. 8, p. 1208, 1949.

[3] W. H. Brattain and W. Shockley, "Density of surface states on silicon deduced from contact potential measurements," *Physical Review*, vol. 72, no. 4, p. 345, 1947.

[4] W. Shockley and G. Pearson, "Modulation of conductance of thin films of semi-conductors by surface charges," *Physical Review*, vol. 74, no. 2, p. 232, 1948.

[5] D. Kahng, "Silicon-silicon dioxide field induced surface devices," in *the Solid State Device Research Conf., Pittsburgh, PA. June 1960*, 1960.

[6] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998.

[7] S. E. Thompson and S. Parthasarathy, "Moore's law: The future of si microelectronics," *Materials today*, vol. 9, no. 6, pp. 20–25, 2006.

[8] S. E. Miller, "Integrated optics: An introduction," *The Bell System Technical Journal*, vol. 48, no. 7, pp. 2059–2069, 1969.

[9] G. T. Reed, W. R. Headley, and C. J. Png, "Silicon photonics: The early years," in *Optoelectronic Integration on Silicon II*, International Society for Optics and Photonics, vol. 5730, 2005, pp. 1–18.

[10] R. Soref and J. Lorenzo, "Single-crystal silicon: A new material for 1.3 and 1.6 um integrated-optical components," *Electronics Letters*, vol. 21, no. 21, pp. 953–954, 1985.

[11] R. Soref and J. Larenzo, "All-silicon active and passive guided-wave components for lambda= 1.3 and 1.6 um," *IEEE Journal of Quantum Electronics*, vol. 22, no. 6, pp. 873–879, 1986.

[12] G. P. Agrawal, "Optical communication: Its history and recent progress," in *Optics in Our Time*, Springer, Cham, 2016, pp. 177–199.

[13] T. H. Maiman *et al.*, "Stimulated optical radiation in ruby," 1960.

[14] K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," in *Proceedings of the Institution of Electrical Engineers*, IET, vol. 113, 1966, pp. 1151–1158.

[15] T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, "Ultimate low-loss single-mode fibre at 1.55 um," *Electronics Letters*, vol. 15, no. 4, pp. 106–108, 1979.

[16] A. Asenov, Y. Wang, B. Cheng, X. Wang, P. Asenov, T. Al-Ameri, and V. P. Georgiev, "Nanowire transistor solutions for 5nm and beyond," in *2016 17th International Symposium on Quality Electronic Design (ISQED)*, IEEE, 2016, pp. 269–274.

[17] F.-L. Yang, D.-H. Lee, H.-Y. Chen, C.-Y. Chang, S.-D. Liu, C.-C. Huang, T.-X. Chung, H.-W. Chen, C.-C. Huang, Y.-H. Liu, *et al.*, "5nm-gate nanowire finfet," in *Digest of Technical Papers. 2004 Symposium on VLSI Technology, 2004.*, IEEE, 2004, pp. 196–197.

[18] E. Swindlehurst, H. Jensen, A. Petrie, Y. Song, Y.-C. Kuan, Y. Qu, M.-C. F. Chang, J.-T. Wu, and S.-H. W. Chiang, "An 8-bit 10-ghz 21-mw time-interleaved sar adc with grouped dac capacitors and dual-path bootstrapped switch," *IEEE Journal of Solid-State Circuits*, 2021.

[19] Y. Wang, R. Wu, J. Pang, D. You, A. A. Fadila, R. Saengchan, X. Fu, D. Matsumoto, T. Nakamura, R. Kubozoe, *et al.*, "A 39-ghz 64-element phased-array transceiver with built-in phase and amplitude calibrations for large-array 5g nr in 65-nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 5, pp. 1249–1269, 2020.

[20] A. Hajimiri, B. Abiri, F. Bohn, M. Gal-Katziri, and M. H. Manohara, "Dynamic focusing of large arrays for wireless power transfer and beyond," *IEEE Journal of Solid-State Circuits*, 2020.

[21] J. Pang, R. Wu, Y. Wang, M. Dome, H. Kato, H. Huang, A. T. Narayanan, H. Liu, B. Liu, T. Nakamura, *et al.*, "A 28-ghz cmos phased-array transceiver based on lo phase-shifting architecture with gain invariant phase tuning for 5g new radio," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 5, pp. 1228–1242, 2019.

[22] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE journal on selected areas in communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[23] D. Kudathanthirige, D. Gunasinghe, and G. Amarasuriya, "Performance analysis of intelligent reflective surfaces for wireless communication," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, IEEE, 2020, pp. 1–6.

[24] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, "Large intelligent surface/antennas (lisa): Making reflective radios smart," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 40–50, 2019.

[25] D. Taillaert, F. Van Laere, M. Ayre, W. Bogaerts, D. Van Thourhout, P. Bienstman, and R. Baets, "Grating couplers for coupling between optical fibers and nanophotonic waveguides," *Japanese Journal of Applied Physics*, vol. 45, no. 8R, p. 6071, 2006.

[26] X. Chen, C. Li, and H. K. Tsang, "Two dimensional silicon waveguide chirped grating couplers for vertical optical fibers," *Optics Communications*, vol. 283, no. 10, pp. 2146–2149, 2010.

[27] F. Gardes, G. Reed, N. Emerson, and C. Png, "A sub-micron depletion-type photonic modulator in silicon on insulator," *Optics Express*, vol. 13, no. 22, pp. 8845–8854, 2005.

[28] J.-B. You, M. Park, J.-W. Park, and G. Kim, "12.5 gbps optical modulation of silicon racetrack resonator based on carrier-depletion in asymmetric pn diode," *Optics express*, vol. 16, no. 22, pp. 18 340–18 344, 2008.

[29] J. W. Park, J.-B. You, I. G. Kim, and G. Kim, "High-modulation efficiency silicon mach-zehnder optical modulator based on carrier depletion in a pn diode," *Optics express*, vol. 17, no. 18, pp. 15 520–15 524, 2009.

[30] M. Safavi-Naeini, D. Franklin, M. L. Lerch, M. Petasecca, G. U. Pignatel, M. Reinhard, G.-F. Dalla Betta, N. Zorzi, and A. B. Rosenfeld, "Evaluation of silicon detectors with integrated jfet for biomedical applications," *IEEE Transactions on Nuclear Science*, vol. 56, no. 3, pp. 1051–1055, 2009.

[31] P. Burger, M. Keters, O. Evrard, and L. Van Buul, "Industrial silicon detectors, advancements in planar technology," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 591, no. 1, pp. 1–5, 2008.

[32] L. Vivien, J. Osmond, J.-M. Fédéli, D. Marris-Morini, P. Crozat, J.-F. Damlencourt, E. Cassan, Y. Lecunff, and S. Laval, "42 ghz pin germanium photodetector integrated in a silicon-on-insulator waveguide," *Optics express*, vol. 17, no. 8, pp. 6252–6257, 2009.

[33] Z. Sheng, L. Liu, J. Brouckaert, S. He, and D. Van Thourhout, "Ingaas pin photodetectors integrated on silicon-on-insulator waveguides," *Optics express*, vol. 18, no. 2, pp. 1756–1761, 2010.

[34] J. Brouckaert, G. Roelkens, D. Van Thourhout, and R. Baets, "Thin-film iii–v photodetectors integrated on silicon-on-insulator photonic ics," *Journal of Lightwave Technology*, vol. 25, no. 4, pp. 1053–1060, 2007.

[35] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE journal of quantum electronics*, vol. 23, no. 1, pp. 123–129, 1987.

[36] R. Soref, "Silicon photonics: A review of recent literature," *Silicon*, vol. 2, no. 1, pp. 1–6, 2010.

[37] A. E.-J. Lim, J. Song, Q. Fang, C. Li, X. Tu, N. Duan, K. K. Chen, R. P.-C. Tern, and T.-Y. Liow, "Review of silicon photonics foundry efforts," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 4, pp. 405–416, 2013.

[38] W. N. Ye and Y. Xiong, "Review of silicon photonics: History and recent advances," *Journal of Modern Optics*, vol. 60, no. 16, pp. 1299–1320, 2013.

[39] S. Y. Siew, B. Li, F. Gao, H. Y. Zheng, W. Zhang, P. Guo, S. W. Xie, A. Song, B. Dong, L. W. Luo, *et al.*, "Review of silicon photonics technology and platform development," *Journal of Lightwave Technology*, 2021.

[40] Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometre-scale silicon electro-optic modulator," *nature*, vol. 435, no. 7040, pp. 325–327, 2005.

[41] Z. Fang and C. Z. Zhao, "Recent progress in silicon photonics: A review," *International Scholarly Research Notices*, vol. 2012, 2012.

[42] J. Clark and G. Lanzani, "Organic photonics for communications," *Nature photonics*, vol. 4, no. 7, pp. 438–446, 2010.

[43] J. Leuthold, C. Koos, and W. Freude, "Nonlinear silicon photonics," *Nature photonics*, vol. 4, no. 8, pp. 535–544, 2010.

[44] C. Ciminelli, F. Dell'Olio, C. E. Campanella, and M. N. Armenise, "Photonic technologies for angular velocity sensing," *Advances in Optics and Photonics*, vol. 2, no. 3, pp. 370–404, 2010.

[45] H. C. Lefevre, *The fiber-optic gyroscope*. Artech house, 2014.

[46] B. Culshaw and A. Kersey, "Fiber-optic sensing: A historical perspective," *Journal of lightwave technology*, vol. 26, no. 9, pp. 1064–1078, 2008.

[47] B. Barrett, R. Geiger, I. Dutta, M. Meunier, B. Canuel, A. Gauguet, P. Bouyer, and A. Landragin, "The sagnac effect: 20 years of development in matter-wave interferometry," *Comptes Rendus Physique*, vol. 15, no. 10, pp. 875–883, 2014.

[48] J. Li, M.-G. Suh, and K. Vahala, "Microresonator brillouin gyroscope," *Optica*, vol. 4, no. 3, pp. 346–348, 2017.

[49] Y.-H. Lai, Y.-K. Lu, M.-G. Suh, and K. Vahala, "Enhanced sensitivity operation of an optical gyroscope near an exceptional point," *arXiv preprint arXiv:1901.08217*, 2019.

[50] B. Wu, Y. Yu, J. Xiong, and X. Zhang, "Silicon integrated interferometric optical gyroscope," *Scientific Reports*, vol. 8, no. 1, pp. 1–7, 2018.

[51] G. Sagnac, "Effet tourbillonnaire optique. la circulation de l'éther lumineux dans un interférographe tournant," *J. Phys. Theor. Appl.*, vol. 4, no. 1, pp. 177–195, 1914.

[52] E. J. Post, "Sagnac effect," *Reviews of Modern Physics*, vol. 39, no. 2, p. 475, 1967.

[53] R. J. Vaccaro and A. S. Zaki, "Statistical modeling of rate gyros," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 3, pp. 673–684, 2011.

[54] N. C. Harris, Y. Ma, J. Mower, T. Baehr-Jones, D. Englund, M. Hochberg, and C. Galland, "Efficient, compact and low loss thermo-optic phase shifter in silicon," *Optics express*, vol. 22, no. 9, pp. 10 487–10 493, 2014.

[55] T. Buret, D. Ramecourt, J. Honthaas, E. Willemenot, Y. Paturel, and T. Gaiffe, "Fiber optic gyroscopes for space application," in *Optical Fiber Sensors*, Optical Society of America, 2006, p. MC4.

[56] J. Davis and S. Ezekiel, "Closed-loop, low-noise fiber-optic rotation sensor," *Optics Letters*, vol. 6, no. 10, pp. 505–507, 1981.

[57] C. Cutler, S. Newton, and H. J. Shaw, "Limitation of rotation sensing by scattering," *Optics Letters*, vol. 5, no. 11, pp. 488–490, 1980.

[58] C. Ciminelli, F. Peluso, and M. N. Armenise, "A new integrated optical angular velocity sensor," in *Integrated Optics: Devices, Materials, and Technologies IX*, International Society for Optics and Photonics, vol. 5728, 2005, pp. 93–100.

[59] C. Vannahme, H. Suche, S. Reza, R. Ricken, V. Quiring, and W. Sohler, "Integrated optical ti: Linbo3 ring resonator for rotation rate sensing," *Proc. 13th ECIO*, vol. 2007, 2007.

[60] K. Suzuki, K. Takiguchi, and K. Hotate, "Monolithically integrated resonator microoptic gyro on silica planar lightwave circuit," *Journal of Lightwave Technology*, vol. 18, no. 1, pp. 66–72, 2000.

[61] J. Scheuer and A. Yariv, "Sagnac effect in coupled-resonator slow-light waveguide structures," *Physical review letters*, vol. 96, no. 5, p. 053 901, 2006.

[62] S. Ezekiel and S. Balsamo, "Passive ring resonator laser gyroscope," *Applied Physics Letters*, vol. 30, no. 9, pp. 478–480, 1977.

[63] H. Zhang, J. Chen, J. Jin, J. Lin, L. Zhao, Z. Bi, A. Huang, and Z. Xiao, "On-chip modulation for rotating sensing of gyroscope based on ring resonator coupled with mach-zehnder interferometer," *Scientific reports*, vol. 6, no. 1, pp. 1–9, 2016.

[64] J. Wang, "Survey and summary: From dna biosensors to gene chips," *Nucleic acids research*, vol. 28, no. 16, pp. 3011–3016, 2000.

[65] C. Sideris, "Electromagnetic field manipulation: Biosensing to antennas," Ph.D. dissertation, California Institute of Technology, 2017.

[66] C. D. Chin, V. Linder, and S. K. Sia, "Commercialization of microfluidic point-of-care diagnostic devices," *Lab on a Chip*, vol. 12, no. 12, pp. 2118–2134, 2012.

[67] S. K. Vashist, P. B. Luppa, L. Y. Yeo, A. Ozcan, and J. H. Luong, "Emerging technologies for next-generation point-of-care testing," *Trends in biotechnology*, vol. 33, no. 11, pp. 692–705, 2015.

[68] C. P. Price, "Point of care testing," *Bmj*, vol. 322, no. 7297, pp. 1285–1288, 2001.

[69] A. Pai, A. Khachaturian, S. Chapman, A. Hu, H. Wang, and A. Hajimiri, "A handheld magnetic sensing platform for antigen and nucleic acid detection," *Analyst*, vol. 139, no. 6, pp. 1403–1411, 2014.

[70] J. Schotter, P.-B. Kamp, A. Becker, A. Pühler, G. Reiss, and H. Brückl, "Comparison of a prototype magnetoresistive biosensor to standard fluorescent dna detection," *Biosensors and Bioelectronics*, vol. 19, no. 10, pp. 1149–1156, 2004.

[71] E. Timurdogan, B. E. Alaca, I. H. Kavakli, and H. Urey, "Mems biosensor for detection of hepatitis a and c viruses in serum," *Biosensors and Bioelectronics*, vol. 28, no. 1, pp. 189–194, 2011.

[72] A. L. Washburn, L. C. Gunn, and R. C. Bailey, "Label-free quantitation of a cancer biomarker in complex media using silicon photonic microring resonators," *Analytical chemistry*, vol. 81, no. 22, pp. 9499–9506, 2009.

[73] C. Sideris and A. Hajimiri, "Design and implementation of an integrated magnetic spectrometer for multiplexed biosensing," *IEEE transactions on biomedical circuits and systems*, vol. 7, no. 6, pp. 773–784, 2013.

[74] P. Fannin, L. Cohen-Tannoudji, E. Bertrand, A. Giannitsis, C. Mac Oireachtaigh, and J. Bibette, "Investigation of the complex susceptibility of magnetic beads containing maghemite nanoparticles," *Journal of Magnetism and Magnetic Materials*, vol. 303, no. 1, pp. 147–152, 2006.

[75] G. Li, V. Joshi, R. L. White, S. X. Wang, J. T. Kemp, C. Webb, R. W. Davis, and S. Sun, "Detection of single micron-sized magnetic bead and magnetic nanoparticles using spin valve sensors for biological applications," *Journal of applied physics*, vol. 93, no. 10, pp. 7557–7559, 2003.

[76] J.-C. Chien and A. M. Niknejad, "Oscillator-based reactance sensors with injection locking for high-throughput flow cytometry using microwave dielectric spectroscopy," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 2, pp. 457–472, 2015.

[77] S.-J. Han, L. Xu, H. Yu, R. J. Wilson, R. L. White, N. Pourmand, and S. X. Wang, "Cmos integrated dna microarray based on gmr sensors," in *2006 International Electron Devices Meeting*, IEEE, 2006, pp. 1–4.

[78] C. Sideris and A. Hajimiri, "An integrated magnetic spectrometer for multiplexed biosensing," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, IEEE, 2013, pp. 300–301.

[79] K. Hoffmann, *Applying the Wheatstone bridge circuit*. HBM Germany, 1974.

[80] M. Gal-Katziri and A. Hajimiri, "Analysis and design of coupled inductive bridges for magnetic sensing applications," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 7, pp. 1883–1894, 2019.

[81] H. Wang, S. Kosai, C. Sideris, and A. Hajimiri, "An ultrasensitive cmos magnetic biosensor array with correlated double counting noise suppression," in *2010 IEEE MTT-S International Microwave Symposium*, IEEE, 2010, pp. 616–619.

[82] M. Babaie and R. B. Staszewski, "A class-f cmos oscillator," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3120–3133, 2013.

[83] M. P.-H. Lin, Y.-T. He, V. W.-H. Hsiao, R.-G. Chang, and S.-Y. Lee, "Common-centroid capacitor layout generation considering device matching and parasitic minimization," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 7, pp. 991–1002, 2013.

[84] H. Wang, C. Sideris, and A. Hajimiri, "A frequency-shift based cmos magnetic biosensor with spatially uniform sensor transducer gain," in *IEEE Custom Integrated Circuits Conference 2010*, IEEE, 2010, pp. 1–4.

[85] G. T. Hermanson, *Bioconjugate techniques*. Academic press, 2013.

[86] M. Huang and H. Wang, "A mm-wave wideband MIMO RX with instinctual array-based blocker/signal management for ultralow-latency communication," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 12, pp. 3553–3564, 2019.

[87] S. Shahramian, M. J. Holyoak, A. Singh, and Y. Baeyens, "A fully integrated 384-element, 16-tile, W-band phased array with self-alignment and self-test," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 9, pp. 2419–2434, 2019.

[88] J. Pang, Z. Li, R. Kubozoe, X. Luo, R. Wu, Y. Wang, D. You, A. A. Fadila, R. Saengchan, T. Nakamura, J. Alvin, D. Matsumoto, B. Liu, A. T. Narayanan, J. Qiu, H. Liu, Z. Sun, H. Huang, K. K. Tokgoz, K. Motoi, N. Oshima, S. Hori, K. Kunihiro, T. Kaneko, A. Shirane, and K. Okada, "A 28-GHz CMOS phased-array beamformer utilizing neutralized bi-directional technique supporting dual-polarized MIMO for 5G NR," *IEEE Journal of Solid-State Circuits*, pp. 1–1, 2020.

[89] A. Hajimiri, A. Komijani, A. Natarajan, R. Chunara, X. Guan, and H. Hashemi, "Phased array systems in silicon," *IEEE Communications Magazine*, vol. 42, no. 8, pp. 122–130, 2004.

[90] M. I. Skolnik, *Radar Handbook, 2nd Ed.* New York, NY: McGraw-Hill, 1990, ISBN: 0-07-057913-X.

[91] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Proceedings. 1998 IEEE International Symposium on Information Theory (Cat. No.98CH36252)*, 1998, p. 156.

[92] ——, "User cooperation diversity. Part I. System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[93] ——, "User cooperation diversity. Part II. Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.

[94] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 65–71, 2012.

[95] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No.01CH37252)*, 2001, p. 294.

[96] A. Scaglione and Yao-Win Hong, "Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2082–2092, 2003.

[97] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4306–4316, 2008.

[98] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2499–2517, 2009.

[99] A. Dimas, D. S. Kalogerias, and A. P. Petropulu, "Cooperative beamforming with predictive relay selection for urban mmWave communications," *IEEE Access*, vol. 7, pp. 157 057–157 071, 2019.

[100] M. Cho, I. Song, and J. D. Cressler, "A true time delay-based SiGe bidirectional T/R chipset for large-scale wideband timed array antennas," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2018, pp. 272–275.

[101] I. Mondal and N. Krishnapura, "A 2-GHz bandwidth, 0.25–1.7 ns true-time-delay element using a variable-order all-pass filter architecture in 0.13 μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 8, pp. 2180–2193, 2017.

[102] S. K. Garakoui, E. A. M. Klumperink, B. Nauta, and F. E. van Vliet, "Compact cascadable gm-C all-pass true time delay cell with reduced delay variation over frequency," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 3, pp. 693–703, 2015.

[103] S. Jang, R. Lu, J. Jeong, and M. P. Flynn, "A 1-GHz 16-element four-beam true-time-delay digital beamformer," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 5, pp. 1304–1314, 2019.

[104] E. Ghaderi, A. Sivadhasan Ramani, A. A. Rahimi, D. Heo, S. Shekhar, and S. Gupta, "An integrated discrete-time delay-compensating technique for large-array beamformers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 9, pp. 3296–3306, 2019.

[105] W. Hong, Z. H. Jiang, C. Yu, J. Zhou, P. Chen, Z. Yu, H. Zhang, B. Yang, X. Pang, M. Jiang, Y. Cheng, M. K. T. Al-Nuaimi, Y. Zhang, J. Chen, and S. He, "Multibeam antenna technologies for 5G wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6231–6249, 2017.

[106] T. Chu and H. Hashemi, "True-time-delay-based multi-beam arrays," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 8, pp. 3072–3082, 2013.

[107] T. Nishio, Hsiao-Ping Tsai, Yuanxun Wang, and T. Itoh, "A high-speed adaptive antenna array with simultaneous multibeam-forming capability," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2483–2494, 2003.

[108] Y. Huang, Y. Li, H. Ren, J. Lu, and W. Zhang, "Multi-panel mimo in 5g," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 56–61, 2018.

[109] F. W. Vook, A. Ghosh, and T. A. Thomas, "Mimo and beamforming solutions for 5g technology," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, IEEE, 2014, pp. 1–4.

[110] N. Shoaib, S. Shoaib, R. Y. Khattak, I. Shoaib, X. Chen, and A. Perwaiz, "Mimo antennas for smart 5g devices," *IEEE Access*, vol. 6, pp. 77 014–77 021, 2018.

[111] H. Huang, X. Li, and Y. Liu, "5g mimo antenna based on vector synthetic mechanism," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 6, pp. 1052–1055, 2018.

[112] A. H. Naqvi and S. Lim, "Review of recent phased arrays for millimeter-wave wireless communication," *Sensors*, vol. 18, no. 10, p. 3194, 2018.

[113] X. Xu, J. Wu, T. G. Nguyen, M. Shoeiby, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, and D. J. Moss, "Advanced rf and microwave functions based on an integrated optical frequency comb source," *Optics Express*, vol. 26, no. 3, pp. 2569–2583, 2018.

[114] X. Xu, J. Wu, M. Shoeiby, T. G. Nguyen, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, and D. J. Moss, "Reconfigurable broadband microwave photonic intensity differentiator based on an integrated optical frequency comb source," *Apl Photonics*, vol. 2, no. 9, p. 096 104, 2017.

[115] X. Yi, Q.-F. Yang, K. Y. Yang, M.-G. Suh, and K. Vahala, "Soliton frequency comb at microwave rates in a high-q silica microresonator," *Optica*, vol. 2, no. 12, pp. 1078–1085, 2015.

[116] N. T. Jafferis, E. F. Helbling, M. Karpelson, and R. J. Wood, "Untethered flight of an insect-sized flapping-wing microscale aerial vehicle," *Nature*, vol. 570, no. 7762, p. 491, 2019.

[117] S. K. H. Win, L. S. T. Win, D. Sufiyan, G. S. Soh, and S. Foong, "Dynamics and control of a collaborative and separating descent of samara autorotating wings," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 3067–3074, 2019.

[118] D. Tseng, O. Mudanyali, C. Oztoprak, S. O. Isikman, I. Sencan, O. Yaglidere, and A. Ozcan, "Lensfree microscopy on a cellphone," *Lab on a Chip*, vol. 10, no. 14, pp. 1787–1792, 2010.

[119] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE signal processing magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[120] B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. Padgett, "3d computational imaging with single-pixel detectors," *Science*, vol. 340, no. 6134, pp. 844–847, 2013.

[121] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2016.

[122] J. N. Mait, G. W. Euliss, and R. A. Athale, "Computational imaging," *Advances in Optics and Photonics*, vol. 10, no. 2, pp. 409–483, 2018.

[123] A. El Gamal, "Trends in cmos image sensor technology and design," in *Digest. International Electron Devices Meeting,*, IEEE, 2002, pp. 805–808.

[124] M. D. Robinson and D. G. Stork, "Joint digital-optical design of superresolution multiframe imaging systems," *Applied optics*, vol. 47, no. 10, B11–B20, 2008.

[125] B. Jalali and S. Fathpour, "Silicon photonics," *Journal of lightwave technology*, vol. 24, no. 12, pp. 4600–4615, 2006.

[126] J. Jiang and J. A. Fan, "Global optimization of dielectric metasurfaces using a physics-driven neural network," *Nano Lett.*, vol. 19, 53665372, 2019.

[127] N. V. Sapra, D. Vercruysse, L. Su, K. Y. Yang, J. Skarda, A. Y. Piggott, and J. Vučković, "Inverse design and demonstration of broadband grating couplers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 3, pp. 1–7, 2019.

[128] A. Brückner, J. Duparré, R. Leitel, P. Dannberg, A. Bräuer, and A. Tünnermann, "Thin wafer-level camera lenses inspired by insect compound eyes," *Optics Express*, vol. 18, no. 24, pp. 24 379–24 394, 2010.

[129] Y. M. Song, Y. Xie, V. Malyarchuk, J. Xiao, I. Jung, K.-J. Choi, Z. Liu, H. Park, C. Lu, R.-H. Kim, *et al.*, "Digital cameras with designs inspired by the arthropod eye," *Nature*, vol. 497, no. 7447, p. 95, 2013.

[130] P. R. Gill, C. Lee, D.-G. Lee, A. Wang, and A. Molnar, "A microscale camera using direct fourier-domain scene capture," *Optics letters*, vol. 36, no. 15, pp. 2949–2951, 2011.

[131] A. N. Tikhonov, "Solution of incorrectly formaulated problems and the regularization method," *Dokl. Akad. Nauk.*, vol. 151, 1963.

[132] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM review*, vol. 40, no. 3, pp. 636–666, 1998.

[133] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, pp. 1–111, 2018.

[134] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[135] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[136] A. White, P. Khial, F. Salehi, B. Hassibi, and A. Hajimiri, "A silicon photonics computational lensless active-flat-optics imaging system," *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.

[137] R. Fatemi, B. Abiri, and A. Hajimiri, "A one-dimensional heterodyne lensfree opa camera," in *CLEO: Science and Innovations*, Optical Society of America, 2016, STu3G–3.

[138] F. Horst, W. M. Green, S. Assefa, S. M. Shank, Y. A. Vlasov, and B. J. Offrein, "Cascaded mach-zehnder wavelength filters in silicon photonics for low loss and flat pass-band wdm (de-) multiplexing," *Optics express*, vol. 21, no. 10, pp. 11 652–11 658, 2013.

[139] R. Fatemi, B. Abiri, and A. Hajimiri, "An 8 * 8 heterodyne lens-less opa camera," in *CLEO: QELS$_F$undamentalScience*, Optical Society of America, 2017, JW2A–9.

[140] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature communications*, vol. 5, p. 3541, 2014.

[141] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, *et al.*, "Deep learning with coherent nanophotonic circuits," *Nature Photonics*, vol. 11, no. 7, p. 441, 2017.

[142] J. K. Doylend, M. Heck, J. T. Bovington, J. D. Peters, L. Coldren, and J. Bowers, "Two-dimensional free-space beam steering with an optical phased array on silicon-on-insulator," *Optics express*, vol. 19, no. 22, pp. 21 595–21 604, 2011.

[143] K. Van Acoleyen, W. Bogaerts, J. Jágerská, N. Le Thomas, R. Houdré, and R. Baets, "Off-chip beam steering with a one-dimensional optical phased array on silicon-on-insulator," *Optics letters*, vol. 34, no. 9, pp. 1477–1479, 2009.

[144] R. Fatemi, A. Khachaturian, and A. Hajimiri, "A nonuniform sparse 2-d large-fov optical phased array with a low-power pwm drive," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 5, pp. 1200–1215, 2019.

[145] R. Fatemi, B. Abiri, A. Khachaturian, and A. Hajimiri, "High sensitivity active flat optics optical phased array receiver with a two-dimensional aperture," *Optics express*, vol. 26, no. 23, pp. 29 983–29 999, 2018.

[146] P. Wang, G. Luo, H. Yu, Y. Li, M. Wang, X. Zhou, W. Chen, Y. Zhang, and J. Pan, "Improving the performance of optical antenna for optical phased arrays through high-contrast grating structure on soi substrate," *Optics express*, vol. 27, no. 3, pp. 2703–2712, 2019.

[147] R. Fatemi, A. Khachaturian, and A. Hajimiri, "A low power pwm optical phased array transmitter with 16° field-of-view and 0.8° beamwidth," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, IEEE, 2018, pp. 28–31.

[148] B. Abiri, R. Fatemi, and A. Hajimiri, "A 1-d heterodyne lens-free optical phased array camera with reference phase shifting," *IEEE Photonics Journal*, vol. 10, no. 5, pp. 1–12, 2018.

[149] C. V. Poulton, M. J. Byrd, P. Russo, E. Timurdogan, M. Khandaker, D. Vermeulen, and M. R. Watts, "Long-range lidar and free-space data communication with high-performance optical phased arrays," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–8, 2019.

[150] F. Aflatouni, B. Abiri, A. Rekhi, and A. Hajimiri, "Nanophotonic coherent imager," *Optics Express*, vol. 23, no. 4, pp. 5117–5125, 2015.

[151] C. V. Poulton, M. J. Byrd, M. Raval, Z. Su, N. Li, E. Timurdogan, D. Coolbaugh, D. Vermeulen, and M. R. Watts, "Large-scale silicon nitride nanophotonic phased arrays at infrared and visible wavelengths," *Optics letters*, vol. 42, no. 1, pp. 21–24, 2017.

[152] J. L. Pita, I. Aldaya, P. Dainese, H. E. Hernandez-Figueroa, and L. H. Gabrielli, "Design of a compact cmos-compatible photonic antenna by topological optimization," *Optics express*, vol. 26, no. 3, pp. 2435–2442, 2018.

[153] Y. Wang, J. P. van Engelen, S. F. Reniers, M. B. van Rijn, X. Zhang, Z. Cao, V. Dolores-Calzadilla, K. A. Williams, M. K. Smit, and Y. Jiao, "High resolution grating antennas for beam steering on the imos platform," in *Asia Communications and Photonics Conference*, Optical Society of America, 2019, M4B–5.

[154] D. K. John and J. M. Ronald, "Antennas: For all applications," *Mc Graw Hill*, 2002.

[155] D. Pozar, *Microwave Engineering, 4th Edition*. Wiley, 2011, ISBN: 9781118213636. [Online]. Available: https://books.google.com/books?id=JegbAAAAQBAJ.

[156] D. Cheng, *Field and Wave Electromagnetics*, ser. Addison-Wesley series in electrical engineering. Addison-Wesley Publishing Company, 1989, ISBN: 9780201128192. [Online]. Available: https://books.google.com/books?id=KL%5C_vAAAAMAAJ.

[157] Y. Luo, Z. Nong, S. Gao, H. Huang, Y. Zhu, L. Liu, L. Zhou, J. Xu, L. Liu, S. Yu, and X. Cai, "Low-loss two-dimensional silicon photonic grating coupler with a backside metal mirror," *Opt. Lett.*, vol. 43, no. 3, pp. 474–477, Feb. 2018. DOI: 10.1364/OL.43.000474. [Online]. Available: http://ol.osa.org/abstract.cfm?URI=ol-43-3-474.

[158] W. Bogaerts, D. Taillaert, P. Dumon, D. V. Thourhout, R. Baets, and E. Pluk, "A polarization-diversity wavelength duplexer circuit in silicon-on-insulator photonic wires," *Opt. Express*, vol. 15, no. 4, pp. 1567–1578, Feb. 2007. DOI: 10.1364/OE.15.001567. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-15-4-1567.

[159] T. Watanabe, Y. Fedoryshyn, and J. Leuthold, "2-D grating couplers for vertical fiber coupling in two polarizations," *IEEE Photonics Journal*, vol. 11, no. 4, pp. 1–9, 2019.

[160] J. Mautz and R. Harrington, "Modal analysis of loaded n-port scatterers," *IEEE Transactions on Antennas and Propagation*, vol. 21, no. 2, pp. 188–199, 1973.

[161] R. Harrington and J. Mautz, "Pattern synthesis for loaded n-port scatterers," *IEEE Transactions on Antennas and Propagation*, vol. 22, no. 2, pp. 184–190, 1974.

[162] R. G. Vaughan, "Two-port higher mode circular microstrip antennas," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 3, pp. 309–321, 1988.

[163] C. Waldschmidt and W. Wiesbeck, "Compact wide-band multimode antennas for MIMO and diversity," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 8, pp. 1963–1969, 2004.

[164] T. Svantesson, "Correlation and channel capacity of MIMO systems employing multimode antennas," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 6, pp. 1304–1312, 2002.

[165] M. Cabedo-Fabres, E. Antonino-Daviu, A. Valero-Nogueira, and M. F. Bataller, "The theory of characteristic modes revisited: A contribution to the design of antennas for modern applications," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 5, pp. 52–68, 2007.

[166] R. Harrington and J. Mautz, "Theory of characteristic modes for conducting bodies," *IEEE Transactions on Antennas and Propagation*, vol. 19, no. 5, pp. 622–628, 1971.

[167] R. J. Garbacz, "Modal expansions for resonance scattering phenomena," *Proceedings of the IEEE*, vol. 53, no. 8, pp. 856–864, 1965.

[168] W. Hong, Z. H. Jiang, C. Yu, J. Zhou, P. Chen, Z. Yu, H. Zhang, B. Yang, X. Pang, M. Jiang, Y. Cheng, M. K. T. Al-Nuaimi, Y. Zhang, J. Chen, and S. He, "Multibeam antenna technologies for 5G wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6231–6249, 2017.

[169] P. Velha, V. Sorianello, M. V. Preite, G. D. Angelis, T. Cassese, A. Bianchi, F. Testa, and M. Romagnoli, "Wide-band polarization controller for Si photonic integrated circuits," *Opt. Lett.*, vol. 41, no. 24, pp. 5656–5659, Dec. 2016. DOI: 10.1364/OL.41.005656. [Online]. Available: http://ol.osa.org/abstract.cfm?URI=ol-41-24-5656.

[170] Y. Xu, J. Wang, J. Xiao, and X. Sun, "Design of a compact silicon-based slot-waveguide crossing," *Appl. Opt.*, vol. 52, no. 16, pp. 3737–3744, Jul. 2013. DOI: 10.1364/AO.52.003737. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-52-16-3737.

[171] L. F. Frellsen, Y. Ding, O. Sigmund, and L. H. Frandsen, "Topology optimized mode multiplexing in silicon-on-insulator photonic wire waveguides," *Opt. Express*, vol. 24, no. 15, pp. 16 866–16 873, Jul. 2016. DOI: 10.1364/OE.24.016866. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-15-16866.

[172] J. Lu and J. Vučković, "Nanophotonic computational design," *Opt. Express*, vol. 21, no. 11, pp. 13 351–13 367, Jul. 2013. DOI: 10.1364/OE.21.013351. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-21-11-13351.

[173] D. Dai and H. Wu, "Realization of a compact polarization splitter-rotator on silicon," *Opt. Lett.*, vol. 41, no. 10, pp. 2346–2349, May 2016. DOI: 10.1364/OL.41.002346. [Online]. Available: http://ol.osa.org/abstract.cfm?URI=ol-41-10-2346.

[174] W. D. Sacher, T. Barwicz, B. J. F. Taylor, and J. K. S. Poon, "Polarization rotator-splitters in standard active silicon photonics platforms," *Opt. Express*, vol. 22, no. 4, pp. 3777–3786, Feb. 2014. DOI: 10.1364/OE.22.003777. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-22-4-3777.

[175] Z. Lu, M. Ma, H. Yun, Y. Wang, N. A. F. Jaeger, and L. Chrostowski, "Silicon photonic polarization beamsplitter and rotator for on-chip polarization control," in *2016 IEEE 13th International Conference on Group IV Photonics (GFP)*, 2016, pp. 70–71.

[176] M. Ma, A. H. K. Park, Y. Wang, H. Shoman, F. Zhang, N. A. F. Jaeger, and L. Chrostowski, "Sub-wavelength grating-assisted polarization splitter-rotators for silicon-on-insulator platforms," *Opt. Express*, vol. 27, no. 13, pp. 17 581–17 591, Jun. 2019. DOI: 10.1364/OE.27.017581. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-13-17581.

[177] W. White, "Pattern limitations in multiple-beam antennas," *IRE Transactions on Antennas and Propagation*, vol. 10, no. 4, pp. 430–436, 1962.

[178] C. J. Chang-Hasnain and W. Yang, "High-contrast gratings for integrated optoelectronics," *Adv. Opt. Photon.*, vol. 4, no. 3, pp. 379–440, Sep. 2012. DOI: 10.1364/AOP.4.000379. [Online]. Available: http://aop.osa.org/abstract.cfm?URI=aop-4-3-379.

[179] B. Abiri, R. Fatemi, and A. Hajimiri, "A 1-D heterodyne lens-free optical phased array camera with reference phase shifting," *IEEE Photonics Journal*, vol. 10, no. 5, pp. 1–12, 2018.

[180] W. L. Stutzman and G. A. Thiele, *Antenna theory and design*. John Wiley & Sons, 2012.

[181] A. Khachaturian, R. Fatemi, and A. Hajimiri, "Achieving full grating-lobe-free field-of-view with low-complexity co-prime photonic beamforming transceivers," *arXiv preprint arXiv:2108.10223*, 2021.

[182] S. A. Miller, Y.-C. Chang, C. T. Phare, M. C. Shin, M. Zadka, S. P. Roberts, B. Stern, X. Ji, A. Mohanty, O. A. J. Gordillo, *et al.*, "Large-scale optical phased array using a low-power multi-pass silicon photonic platform," *Optica*, vol. 7, no. 1, pp. 3–6, 2020.

[183] C.-P. Hsu, B. Li, B. S. Rivas, A. Gohil, P. H. Chan, A. D. Moore, and V. Donzella, "A review and perspective on optical phased array for automotive lidar," *IEEE Journal of Selected Topics in Quantum Electronics*, 2020.

[184] N. Reiskarimian and H. Krishnaswamy, "Magnetic-free non-reciprocity based on staggered commutation," *Nature communications*, vol. 7, no. 1, pp. 1–10, 2016.

[185] N. Reiskarimian, J. Zhou, and H. Krishnaswamy, "A cmos passive lptv nonmagnetic circulator and its application in a full-duplex receiver," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 5, pp. 1358–1372, 2017.

[186] N. Reiskarimian, M. B. Dastjerdi, J. Zhou, and H. Krishnaswamy, "18.2 highly-linear integrated magnetic-free circulator-receiver for full-duplex wireless," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, 2017, pp. 316–317.

[187] J. A. Weiss, G. F. Dionne, and D. H. Temme, "The ring-network circulator for integrated circuits: Theory and experiments," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 12, pp. 2743–2748, 1995.

[188] J. A. Weiss, "Circulator synthesis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 13, no. 1, pp. 38–44, 1965.

[189] T. Dinc, M. Tymchenko, A. Nagulu, D. Sounas, A. Alu, and H. Krishnaswamy, "Synchronized conductivity modulation to realize broadband lossless magnetic-free non-reciprocity," *Nature communications*, vol. 8, no. 1, pp. 1–9, 2017.

[190] L. Vivien and L. Pavesi, *Handbook of silicon photonics*. Taylor & Francis, 2016, ISBN: 9781439836101.

[191] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. Thomson, "Silicon optical modulators," *Nature photonics*, vol. 4, no. 8, pp. 518–526, 2010.

[192] C. Kieninger, C. Füllner, H. Zwickel, Y. Kutuvantavida, J. N. Kemal, C. Eschenbaum, D. L. Elder, L. R. Dalton, W. Freude, S. Randel, and C. Koos, "Soh mach-zehnder modulators for 100 gbd pam4 signaling with sub-1 db phase-shifter loss," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3, ISBN: 978-1-9435-8071-2.

[193] J. Fujikata, S. Takahashi, M. Noguchi, and T. Nakamura, "High-efficiency of narrow-width mos capacitor type si optical modulator with tm mode excitation," in *2019 IEEE 16th International Conference on Group IV Photonics (GFP)*, 2019, pp. 1–2. DOI: 10.1109/GROUP4.2019.8926106.

[194] A. Khachaturian, B. Abiri, and A. Hajimiri, "A compact spiral mzi modulator on soi process," in *2015 IEEE 12th International Conference on Group IV Photonics (GFP)*, 2015, pp. 151–152. DOI: 10.1109/Group4.2015.7305996.

[195] Y. Chen and S. Blair, "Nonlinearity enhancement in finite coupled-resonator slow-light waveguides," *Opt. Express*, vol. 12, no. 15, pp. 3353–3366, Jul. 2004. DOI: 10.1364/OPEX.12.003353. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-12-15-3353.

[196] J. K. S. Poon, J. Scheuer, S. Mookherjea, G. T. Paloczi, Y. Huang, and A. Yariv, "Matrix analysis of microring coupled-resonator optical waveguides," *Opt. Express*, vol. 12, no. 1, pp. 90–103, Jan. 2004. DOI: 10.1364/OPEX.12.000090. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-12-1-90.