Scaling Phased Array Receivers to Massive MIMO and Wide Bandwidth with Analog Baseband Beamforming

Emily Naviasky

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2023-44
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-44.html

May 1, 2023

Scaling Phased Array Receivers to Massive MIMO and Wide Bandwidth with Analog Baseband Beamforming

by

Emily Lauren Naviasky

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering — Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali Niknejad, Co-chair
Professor Elad Alon, Co-chair
Professor Martin White

Spring 2022

Scaling Phased Array Receivers to Massive MIMO and Wide Bandwidth with Analog Baseband Beamforming

Abstract

Scaling Phased Array Receivers to Massive MIMO and Wide Bandwidth with Analog Baseband Beamforming

by

Emily Lauren Naviasky

Doctor of Philosophy in Engineering — Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali Niknejad, Co-chair

Professor Elad Alon, Co-chair

Massive multi-user MIMO is a promising technique to increase capacity with spatial filtering for spectrum reuse, and increasing SNR through array gain. There are many advantages to increasing antenna and user count in these arrays, but there is also much debate on how best to do so efficiently. An especially important decision is where to place the beamformer in the signal chain. This work proposes a hardware informed model for comparison between analog and digital beamforming, breaking down the decision into intuitive designer chosen specifications and comparing in power. The model finds that analog baseband beamforming is critical for scaling to wide-bandwidth and large arrays, especially in the presence of interferers. The hardware assumptions in the model are tested in a 16x16 beamforming receiver array in 28nm silicon. Design techniques such as noise or bandwidth limited design, two stage beamforming, and offset correction are used in the circuit implementation. The ASIC was packaged and used in a real time demonstration which performs multi-user and multi-panel operation.

For my family.

Contents

List of Figures

List of Tables

Acknowledgments

A PhD is a fascinatingly contradictory experience. The research that comes out of our lab is simultaneously bold and prestigious, yet often limited by mundane needs like time and funding. I have grown confident in my expertise over this narrow sliver of my field, yet over the years I have watched with a distant dismay as more general knowledge has become slippery and difficult to exercise. So it is fitting that a PhD should also be a contradiction in that it is both uniquely isolating and deeply connecting. I have felt hopeless in ways only my dear friends in the department could understand and get me through. I have felt insufficient, then grown from the help and patience of those more experienced. I have felt lost more times than I thought endurable, but was sheltered by the love of friends and family. My name is alone on a diploma, and yet the accomplishment would be impossible without the support, time, help, friendship, and love from so many. The least I can try to do is thank you all here.

To the BWRC admin, who keep the whole ship afloat in ways sometimes obvious and sometimes invisible but always important. Candy and James, you are some of the kindest people I know and it has been a blessing to get to know and work with you.

To my advisors. Elad, I disagree with you often, but you told me to switch my application from a Master's to a PhD. Now that I'm out, it was probably the right call. Ali, thank you for taking on so many orphaned graduate students and for being open to feedback when I worked up the courage. I admire your obvious brilliance, and how humble and kind you are in your explanations.

To my friends in BWRC. Thank you to Nathan, Bonjern, Fil, and Greg Wright for fun lunches and conversations over beers. To Ben for recommending I be an EECS Peer even when I felt grossly underqualified, and for always remembering me when it's time to make pirogi. To Nima, for always being prompt and kind with your expert technical advice. To Sidney and Panos for conversations and commiseration. To Bob, Rebekah, and Alisha who carry on the good work of EECS Peers and women's lunch. To Ozzy, for introducing me to some of my favorite Bay Area experiences, and Antonio, for many silly debates. You both set an impossible to follow example, which ultimately made me better for both the attempt to measure up and the realization that I didn't need to. All paths through grad school are unique and impossible to follow, and I'm proud of mine, just as I'll always be impressed by yours. To Matt, for being my absolute favorite person to talk for way too long about weird technical ideas, good questions, and hard papers. I thought I would never love circuits again until you came to lab and made me feel sane again. I'll owe you forever for that.

To those who worked on EE16B with me. First Nathan Mailoa, I will always remember our time yelling at cars in the Cory basement fondly. It was worth getting tricked into a PhD for. To Forrest and Simon who helped me finally realize my dream of running the class, and being the absolute best team when it suddenly turned virtual.

To Regina, Hani, Laura, Nate, Josh, Sandra, and Vasuki for your tireless contributions to Bias Busters and for letting me join you. I learned so much, and it was a much needed outlet; you're all so amazing.

To my friends outside of work, thank you for fun and getting me away from lab for a bit. To Regina, Sarah, Mindy, Laurel, Emma thank you for sharing your stories, baked goods, and friendship. To Andy, Chris, Jon, Sarah, and Zoe for being the best DnD group, and for the chance every week to laugh until I can't breathe. To my majaoes: Kara, Destiny, Clare, Katie, Jessi, and Noël for rants, memes, halp, and always letting me know I'm a goddess for everything I do. To my house mates Grace, Regina, and Sarah. I'm so glad I got to talk to you in the evenings, I couldn't have asked for better partners to survive all the oddities of renting life. To Hayes, for answering midnight calls, curating the best of anime and professional wrestling, and starting your PhD before me so I at least got a warning of what was coming. To Judy, for being just as bad at remembering to call as me, and never letting it matter. To Noël, I can't believe you decided to get a PhD, too, but I'm selfishly so glad that I can commiserate with you about all of the ridiculousness that is grad school.

Thank you to the the friends that I would not have been able to finish this PhD without. Lorenzo, I could not have asked for a better collaborator. You're endlessly smart, competent, and kind. You've listened in ways no one else does. You made it easy to laugh off mistakes and fun to push through bugs. I am so fortunate to have gotten to work closely with you, and I appreciate you so much. Keertana, you are no longer my desk-mate but you're the desk-mate of my heart and forever, forever my dear friend. Thank you for always being in my corner and listening when I needed to rant, and letting me be that for you too. Thank you for always being up to do something big and sometimes down-grading to scones. I learned so much from you, and am always so impressed at what you've accomplished.

Finally thank you to my family. Mom, Dad, Andrew, I'm always trying to make you proud. Thanks for making me feel like I've succeeded in that no matter what I do. Love, Dr. Naviasky.

Chapter 1

Introduction

Wireless capacity capabilities have increased with each mobile wireless generation, and yet new applications and modalities continue to demand more. Fully autonomous vehicles are approaching on the technical horizon, and consensus points to some portion of the processing will need to be done remotely on the cloud [1]. A global pandemic has spurred the advent of remote work that will continue even afterwards, and may even grow into a more connected virtual reality [2]. All of these emerging technologies and applications will require major advances in increasing capacity, reducing latency, and increasing reliability.

However, increasing capacity to meet increasing demand requires re-imagining the typical wireless network. For example, state-of-the-art coding techniques have pushed spectral efficiency close to the Shannon limit [3]. Capacity improvement beyond that is currently sought through two methods: wider channel bandwidth (BW) and spatial multiplexing. Wider bandwidths are achieved by moving to higher frequency bands of the wireless communications spectrum, particularly those in the mm-wave range [4]. However, as available bandwidth increases with carrier frequency, so too does attenuation; mm-wave links suffer from significant propagation loss from atmospheric absorption, which limits range and capacity. This can be overcome using phased-array transceivers. An array of M antennas is employed to create a high-directivity beam and electronically steer it towards a user. This introduces a boost of the signal-to-noise ratio (SNR) to improve link range [5]. A system which is already using phased-arrays is also able to leverage spatial multiplexing, introducing another multiplexing domain to further increase capacity with spectrum reuse [6].

The combination of wide bandwidth and spatial multiplexing has the potential to supply incredible capacity, and continue the expansion of wireless communication [7, 8]. However, both wide-bandwidth and phased arrays come with many design challenges. To get the full advantage of increased capacity, the array must be designed as a full system, identifying the best architecture for the application.

The research presented in this dissertation will demonstrate analog baseband beamforming in a receiver for many-user beamforming, and show that it is more power-efficient than digital in the case of large interferers and wide bandwidth. By creating the models for the research presented, we can develop insights on the different application classes and wireless

environments as well as insight on how to continue scaling beyond the current state of the art.

This chapter will introduce the concepts and beamforming architectures relevant to this discussion. Chapter 2 identifies the ADC as a important block for comparison and how the ADC specifications change between analog and digital beamforming. Chapter 3 presents a deeply hardware informed model of the baseband amplifiers, wiring, and analog or digital beamformer. Chapter 4 combines the models and examines the results and application spaces that are best suited to baseband analog or digital beamforming. The design methodology for analog beamformer design is used in practice in Chapter 5 to implement a fully-connected, 16 user beamformer in 28nm CMOS as a part of a two panel, 32 antenna receiver. Chapter 6 reports the measurement results of the ASIC as well as the demonstration of multi-user wireless communication at 73GHz. Finally, Chapter 7 gives a summary, future directions, and final thoughts on on the work.

1.1 Phased Arrays and MIMO

History

Arrays of antennas are nearly as old as wireless communication itself. In 1902, Guglielmo Marconi's pioneering transatlantic wireless experiment featured an antenna array. Many wires slung between tall poles were to transmit a signal from Cornwall, UK to a second array in Cape Cod, Massachusetts. The Cornwall array was downsized, due to destruction by high winds, and the smaller surviving array was expected to have a more limited range, making it unable to reach the Cape Cod station [9]. Thus, the final experiment featured a receiver attached to a kite in St. John's Newfoundland.

A few years later in 1905, Marconi's Co-Nobel Laureate, Karl Ferdinand Braun, demonstrated the first manually steerable phased array [10]. Two matching signals and a phase shifted signal could be delivered to each of three antennas spaced on an equilateral triangle. The signals could be manually moved between antennas to direct the signal one of three directions.

Since then, passive and active phased arrays have been used mainly in military radar, and radio broadcast. As we move into the 21st century, however, the size and cost of phased arrays has been shrinking with wavelength, and phased arrays have begun to break into everyday commercial spaces and wireless standards [11].

Phased Arrays

To begin understanding antenna receive arrays, imagine two receiving antennas placed equidistant from a single transmitting antenna, measuring distance in the wavelength of the carrier frequency. The equidistant receivers each expect to get the same signal from transmitter at exactly the same time. If the signal at each receiver is then summed together,

the two sinusoidal signals combine constructively in voltage. The combined signal power is then 4 times larger than at one receiver alone. The noise received at the antennas, however, is uncorrelated, and so the noise power only doubles. The result is an increase in not just received power, but signal to noise ratio (SNR) compared to a single antenna. This can be expanded not just for 2 antennas, but for any number of M antennas.

$$SNR_{M \text{ antennas}} = \frac{(Signal_{1 \text{ antenna}} \times M)^2}{Noise^2_{1 \text{ antenna}} \times M} = M \times SNR_{1 \text{ antenna}} \tag{1.1}$$

However, say the transmitter moves so now receiver 1 is a half wavelength closer to the transmitter than receiver 2. This results in a half wavelength delay between the signal at each receiver. Now, when the receivers are combined, the two sinusoids are perfect inverses of each-other and sum destructively. To fix this, a delay element could be introduced to re-align the two receiver signals before they are combined. This is the concept behind directional phase arrays.

The phase delay with which a signal arrives at a series of antennas can be calculated for a given angle of arrival. Fig. 1.1 shows the geometry used to calculate the phase delay between each antenna element in a linear array.

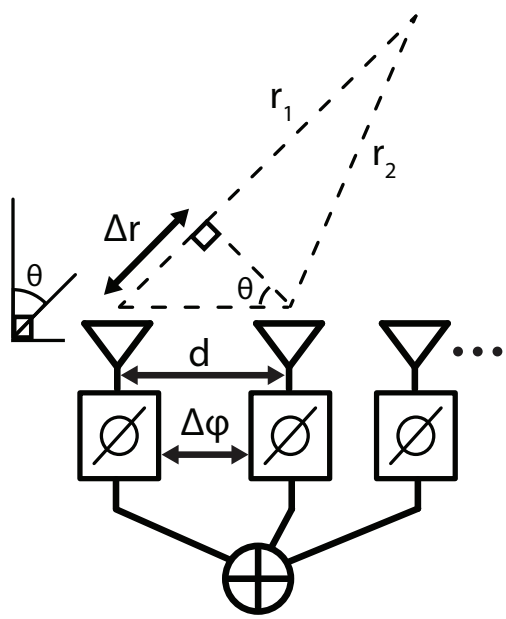

Figure 1.1: Phase delay between antenna elements calculated from signal angle of arrival for an array with d spacing.

The difference in distance from the transmitter to two adjacent elements of the receiving array is Δr. The spacing between antennas is d, and the signal's angle of arrival is θ, measured from the perpendicular to the array plane. We would like to know the difference

in phase between the signals due to delay, which we find using convenient right angles and knowing that the angle across from the distance of interest is the signal's angle of arrival: θ.

$$\Delta r = d \sin\theta \tag{1.2}$$

If Δr is measured in wavelengths, then the conversion to difference in phase is $\Delta\phi = 2\pi\,\Delta r/\lambda$. Antenna spacing is measured in wavelengths and frequently chosen to be $\lambda/2$. This spacing prevents spatial equivalent of under-sampling and aliasing in the beam, known as grating lobes. With these values we can solve for the phase difference between each antenna.

$$\Delta\phi = \frac{\lambda}{s}\sin\theta \tag{1.3}$$

When the transmitter is positioned directly in front of the array, $\theta = 0$, from Eq. 1.3, we see that, likewise, the phase delay between each antennas is $\Delta\phi = 0$. By physical intuition, for small arrays the signal must travel the same distance to every antenna. Mathematical intuition says that $\Delta\phi = 0$ over the entire array by the small angle approximation. However, for very large antenna arrays, where the width of the array is on the order of r, the small angle approximation no longer applies and we see that Eq. 1.3 is an approximation that applies only for far field.

Furthermore, $\Delta\phi$ is equivalent to time delay only for the frequency corresponding to wavelength λ. However, for very wide-band signals, the difference between phase delay $\Delta\phi$ and the true time delay becomes significant. This is called beam-squint, where the effective angle-of-arrival begins to spread over the frequency range. To fix this, a true time delay can be used instead of phase for either very wide-band signals. This is also necessary for very large arrays, where the time delay from one end of the array to the other is a significant percent of a symbol time. The true time delay is calculated using the speed of light, c, as: $\Delta t = \Delta r/c$. The specifics of when true time delay is necessary is covered more in Section 5.3.

Channel and Beamforming Arrays

Many In, Single Out

We first examine the case of a single transmitter antenna and a receiver array of M antenna elements, known as Many In, Single Out (MISO). The phase delays corresponding to each antennas can be conceived of as a vector of complex weights. For a transmitted signal x and the vector of received signals $\hat{y}$, the channel vector represents the phase delay and amplitude attenuation due to angle and distance of the signal through the wireless channel, $\hat{H}$.

$$\hat{y} = \hat{H}\,x \tag{1.4}$$

The simplest model of $\hat{H}$ comes from the angle of arrival phase delay discussed in Fig. 1.1, where the phase difference between each pair of adjacent antennas is $\Delta\phi$ and the attenuation due to far field travel is ρ. Thus, $\hat{H} = \rho\,[0, e^{j\Delta\phi}, e^{j2\,\Delta\phi}, \ldots, e^{j(M-1)\,\Delta\phi}]^T$.

The phase delay applied to each received signal before summation can also be represented as a vector, which will be referred to as $\hat{\omega}$. Choosing the weights for $\hat{\omega}$ can be done with a variety of algorithms tailored to different goals. Applying a weight matrix which is the conjugate transpose of the channel matrix, $\hat{\omega} = \hat{H}^*$, will correct the delay resulting from a known angle of arrival as described with Eq.1.3. This gives the final estimate of the transmitted value x'.

$$x' = \hat{H}^* \, \hat{y} \tag{1.5}$$

This choice of weights for $\hat{\omega}$ is known as conjugate beamforming. If the signal arrives with different amplitude weights at the antennas, as occurs in arrays much larger than the carrier wavelength, then there is an amplitude component. This is the most basic version of a solution where signals are combined in proportion to their rms signal power known as Maximal Ratio Combining (MRC), called maximal as it maximizes for SNR. For most arrays the transmitter is much further, in wavelength, than the width of the array, and amplitude difference is negligible. Thus, MRC beamforming and conjugate beamforming are frequently equivalent.

The array gain from MRC beamforming is largest in the direction of angle of arrival, θ, and decreases outside of that, creating a spatial filter. Illustrated in Fig. 1.2 is the beam shape for conjugate beamforming antenna weights. The antennas act as discrete spatial samples, creating a beam that resembles an FIR filter. With more samples, a filter has more resolution, and so more antennas in the array can create a narrower beam. If spatial samples from the antenna elements are the equivalent to time samples, then the multiplication with the conjugate weights transforms from spatial domain to beam domain.

Many In, Many Out

The beamforming weights need not be limited to a vector, but can be a matrix as well. In hardware, this equates to sending an antenna signal to multiple phase or time delay blocks, which are then combined into separate beams. Intuitively, this looks like a channel with rich channel diversity, such as from many scattering elements. This creates many angles of arrival, which can be spatially distinguished from each other[12, 13], but the array cannot distinguish more spatial channels than there are antennas. The spatial diversity of an environment can be improved with a transmitter capable of steering to angles of departure, and utilizing reflective surfaces in the environment to create multiple paths to the receiver, thus increasing the number of angles of arrival. To do so, the transmitter uses the same concept as the receiver array to create a transmitter array. This is known as Multiple In, Multiple Out (MIMO). The channel can be represented as transmit and receiver vectors, $\hat{x}$ and $\hat{y}$, as well as the channel matrix, $\mathbf{H}$.

$$\hat{y} = \mathbf{H}\hat{x} \tag{1.6}$$

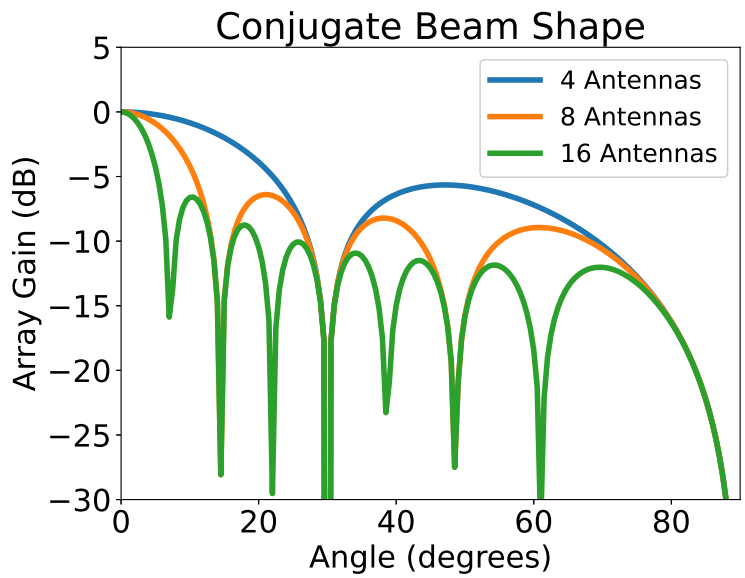

Figure 1.2: Spatial filtering response for MRC beamforming, with varying array size.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{M_r} \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,M_t} \\ h_{2,1} & h_{2,2} & \dots & h_{2,M_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_r,1} & h_{M_r,2} & \dots & h_{M_r,M_t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{M_t} \end{bmatrix} \tag{1.7}$$

Then the beamforming matrices are $\boldsymbol{\omega}_t$ for the transmitter and $\boldsymbol{\omega}_r$ for the receiver. With the desired signal, s encoded by the transmitter beam weights, transformed by the channel, and decoded by the receiver beam weights.

$$s' = \boldsymbol{\omega}_r \, \mathbf{H} \, \boldsymbol{\omega}_t \, s \tag{1.8}$$

The representation of the array in linear algebra allows us to apply fundamentals from linear algebra. First, an orthogonal spatial channel requires an independent column in the matrix. Examining Eq. 1.7, we see quickly that the channel cannot support more orthogonal channels than the rank of $\mathbf{H}$, which is fundamentally limited by M_r and M_t. A full rank $\mathbf{H}$ could represent the entire spatial domain in beam domain without any loss of spatial information. Alternatively, in a far-field, line-of-sight channel with no reflecting objects in the environment, all signals would arrive at the same angle and the channel matrix would have rank 1. The requirement of independent columns in $\mathbf{H}$, gives us intuition on the importance of a channel matrix with a rich scattering coefficient, as from many reflective surfaces.

From linear algebra, we also see that if one has perfect knowledge of the channel matrix, it would be possible to invert the channel using some combination of $\boldsymbol{\omega}_r$ and $\boldsymbol{\omega}_t$ to obtain the

original signal s. This channel inverse would require that the channel be invertible, which requires $M_r = M_t$. More leniently, a Moore-Penrose pseudo inverse, represented as $A^\dagger$, can achieve a similar result, with a number of independent channels up to the minimum of M_r, M_t, or rank of $\mathbf{H}$. For the receiver beamforming matrix this is found as shown.

$$\boldsymbol{\omega}_r = \mathbf{H}^\dagger = \mathbf{H}^*(\mathbf{H}^*\mathbf{H})^{-1} \tag{1.9}$$

A beam weight matrix which is a pseudo inverse of the channel is known as zero forcing (ZF) beamforming, so called because it forces inter-beam-interference to zero, maximizing SINR.

Single User, Multi-User

Much like separate frequency channels, multiple orthogonal spatial channels can be used to increase capacity in a few ways. A TX array can send the same symbols along multiple spatial channels, thus increasing the SNR through "transmit diversity". Alternatively, each channel can be used to transmit separate and parallel streams of data, known as "spatial multiplexing". Transmit diversity is generally useful for applications where the desired SNR is unreliable. However, coding techniques generally handle maximizing capacity with unreliable SNR, thus spatial multiplexing is more effective at increasing capacity [14]. Thus, the spatial channels are best utilized for spatial multiplexing, and enabling capacity increase from spectrum reuse.

As discussed, depending on the multi-path richness in a given wireless environment, the transmitting array can steer beams to support a maximum of M_r orthogonal links which can be used to spatially encode data. This concept is at the base of Single-User MIMO (SU-MIMO), which has been successfully employed in sub-6GHz wireless communications. Fig. 1.3.(a) illustrates a MIMO uplink taking advantage of multi-paths in the wireless channel between a single user and a base station to send spatially multiplexed information. However, this solution becomes less applicable as carrier frequencies increase. At carriers approaching mm-wave and THz, the signal experiences significant attenuation, even for line-of-sight. This reduces the channel's spatial diversity, thus severely limiting the capacity of SU-MIMO at high frequencies.

Multi-user MIMO (MU-MIMO), shown in Fig. 1.3.(b), re-introduces spatial diversity to a mm-wave frequency wireless channel by serving K spatially distant users using beamforming[15]. The base station supports a large array of M antennas simultaneously steering $K \leq M$ independent beams. Array gain at the RX still boosts SNR to compensate for propagation losses, thus increasing link range. In addition, spatial multiplexing allows users to occupy the same frequency spectrum at the same time, thus increasing capacity by $K\times$ for well spatially separated users.

Beamforming and user tracking are performed at the base station, so user transceivers can be power-efficient and simple, with single or low numbers of antennas.

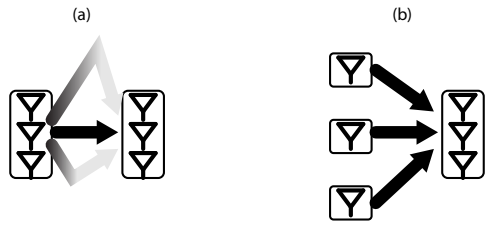

Figure 1.3: Spatial diversity for (a) SU-MIMO and (b) MU-MIMO.

Searching for users and characterizing the wireless channel can be done in many ways, including pilot signals [16], searching using sweeps in frequency [17], with beam codebooks [18], or aided by AI [19].

Massive MIMO

MIMO systems further benefit from the Massive MIMO scenario, i.e. $M \gg K$ [20]. Under this condition, computationally inexpensive, linear algorithms such as MRC and ZF, can achieve near-optimal user tracking and spatial interference rejection [21]. Intuitively, Massive MIMO is the case when the spatial resolution, or beam width, is much finer than the number of users, decreasing the chance that any two users are interfering. To aid in discussing the "massive-ness" of an array, we'll use $\alpha = M/K$. This is also the inverse of the loading factor. So a massive MIMO case might have an α of 10 or 100. Whereas an α of 1 indicates the same number of users as antennas, and would require a very spatially diverse channel.

Although promising for increasing range and capacity, Massive MU-MIMO systems pose substantial challenges in hardware development [22]. These challenges are generally related to the large array size and the complexity of handling M signal streams or K user streams, especially at mm-wave frequencies and wide signal bandwidth.

1.2 Architecture

Serving many users with many elements comes with many system design decisions which will have significant implications on scalability, flexibility, and maximum capacity of the array. Understanding the advantages and shortcomings of different architectures is critical to choosing the best array for the application. We will begin first with how a massive array can be divided into sub-arrays or not, and then examine, at a high level, different options for beamformer location in the signal chain.

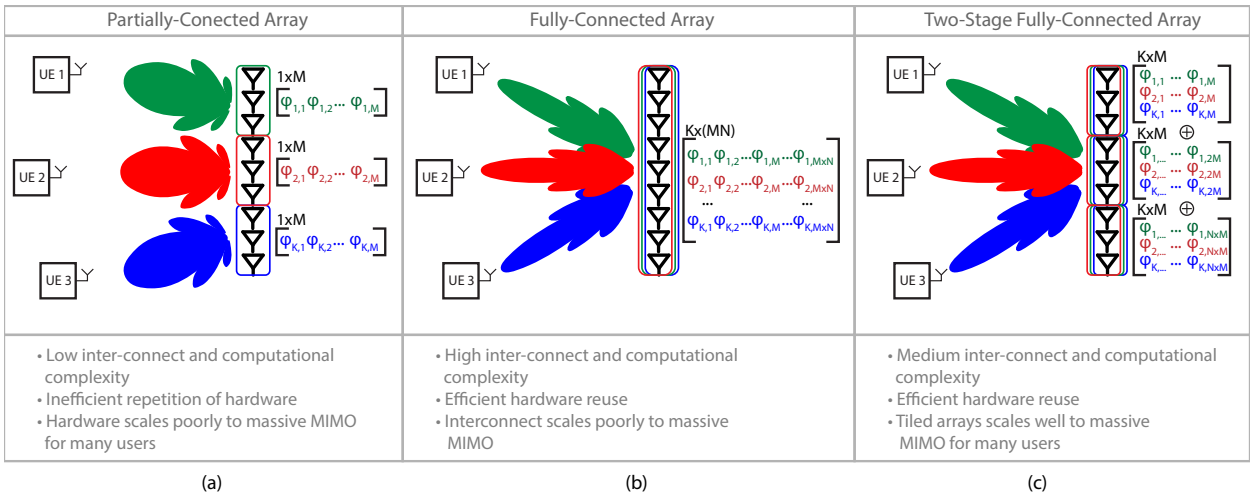

Figure 1.4: System architectures for MU-MIMO.

Array Division

Different architectures can be used to achieve Massive MU-MIMO. To compare across the architectures considered in Fig. 1.4, the number of antennas in the full array will be constant and discussed in terms of $M \times N$ where N is the number of sub-arrays when applicable.

The partially-connected array, Fig. 1.4.(a), is a conventional architecture for systems that prefer to perform beamforming in analog. The partially connected architecture subdivides the full array into N sub-arrays, frequently serving a single user [23]. This creates an easily modularized array, where the number of users can be increased by adding additional sub-arrays, but the full array aperture is never available to any user. This results in reduced array gain and wide beams. While this architecture has fewer logistical problems due to interconnect complexity, the large amount of hardware and associated power costs required for this architecture do not scale well to many users or else cannot support the large antenna/user ratios of massive MIMO.

Alternatively, a fully-connected array, Fig. 1.4.(b), provides all $M \times N$ antenna elements of an equally sized array to each user, enabling better hardware efficiency [24]. However, a massive fully-connected array results in significant interconnect complexity where every antenna signal must be carried to a central location for processing at great routing and power cost. The LO distribution and corresponding phase noise likewise scales in difficultly as the array size grows and must be carefully managed [25]. In addition, the size of the matrix operation can be difficult to scale to wide bandwidths, for either analog or digital processing. This complexity scales poorly to massive arrays with many users, and ultimately limits the array size that can be achieved.

The two-stage fully-connected architecture, shown in Fig. 1.4.(c), is an architecture that

enables hardware efficiency by providing the entire array aperture to all users, and which has manageable interconnect complexity that scales with number of users as opposed to number of antennas [26, 24, 27, 28]. Two-stage beamforming is divided into N modularizable sub-arrays, and beamforming computation is divided between the local sub-array and the back-end. The sub-array performs lower resolution conjugate beamforming on an $M \times K$ sized matrix. The parallelization of beamforming makes the smaller beamforming operations feasible for wide bandwidths and at the local sub-array level without sacrificing spatial resolution. The second stage of beamforming can then perform higher resolution beamforming on a smaller $K \times K$ matrix. Two-stage beamforming further has the advantage of being able to estimate the channel as frequency flat at the sub-array level, and performing frequency dependent beamforming in the second stage with little loss in channel capacity[27]. The LO distribution is often challenging to scale, so it is broken into cross-array and local distribution. Phase noise is managed at the transition between local beamforming and back-end beamforming [22]. The architecture then favors sub-arrays with a single LO, which combine many antennas in the first local stage of beamforming [25]. These advantages make two-stage fully connected architecture the most feasible to scale efficiently to massive MU-MIMO.

Fig. 1.5 illustrates a more detailed implementation of the distributed, modularized beamforming for multiple users. At the sub-array level, conjugate beamforming points beams towards each user, using all M local antennas and outputting K unique user streams. The first stage of analog baseband beamforming boosts SNR before A/D conversion, while decreasing the cross-array interconnect. The second stage of the Two-Stage architecture is performed in the digital back-end where a digital matrix multiplication completes the ZF operation, $H^*(HH^*)^{-1}$ [22]. This further reduces inter-user interference, can be performed as frequency dependent beamforming for channels with frequency fading, and fully realizes the near optimal beamforming of a massive array for K users.

A key element of the Two-Stage Architecture is the sub-array radio module, which implements a multi-output BF matrix capable of supporting K user beams, where K is the maximum number of users the system plans to support. To achieve high capacity with spectral efficiency, K should be large. Mm-wave sub-array modules based on custom CMOS/BiCMOS ASICs, which tile together to create a massive array, have recently been proposed [29, 23, 30, 31, 32, 33, 34]. However, each sub-array only supports one output stream (or two streams for dual polarization [23, 33]), thus limiting operation to single-user or partially-connected multi-user. A few multiple-user-output RX beamforming ASICs for fully-connected arrays have been shown at 28GHz, although with limited user count. In [24], two outputs are extracted from a 8-element array. In [35], up to 4 concurrent users are supported with a 4-element array; however, the baseband bandwidth demands of the frequency-multiplexed user streams prevent the architecture from scaling to larger K. Chapter 5 discusses a 28nm CMOS ASIC for use as a sub-array in a fully connected mm-wave Massive MU-MIMO basestation E-band receiver[36]. The ASIC has multiple-output baseband analog beamformer capable of supporting up to 16 concurrent user streams with 16 RX elements, allowing for future investigation of the Two-Stage Beamforming architecture.

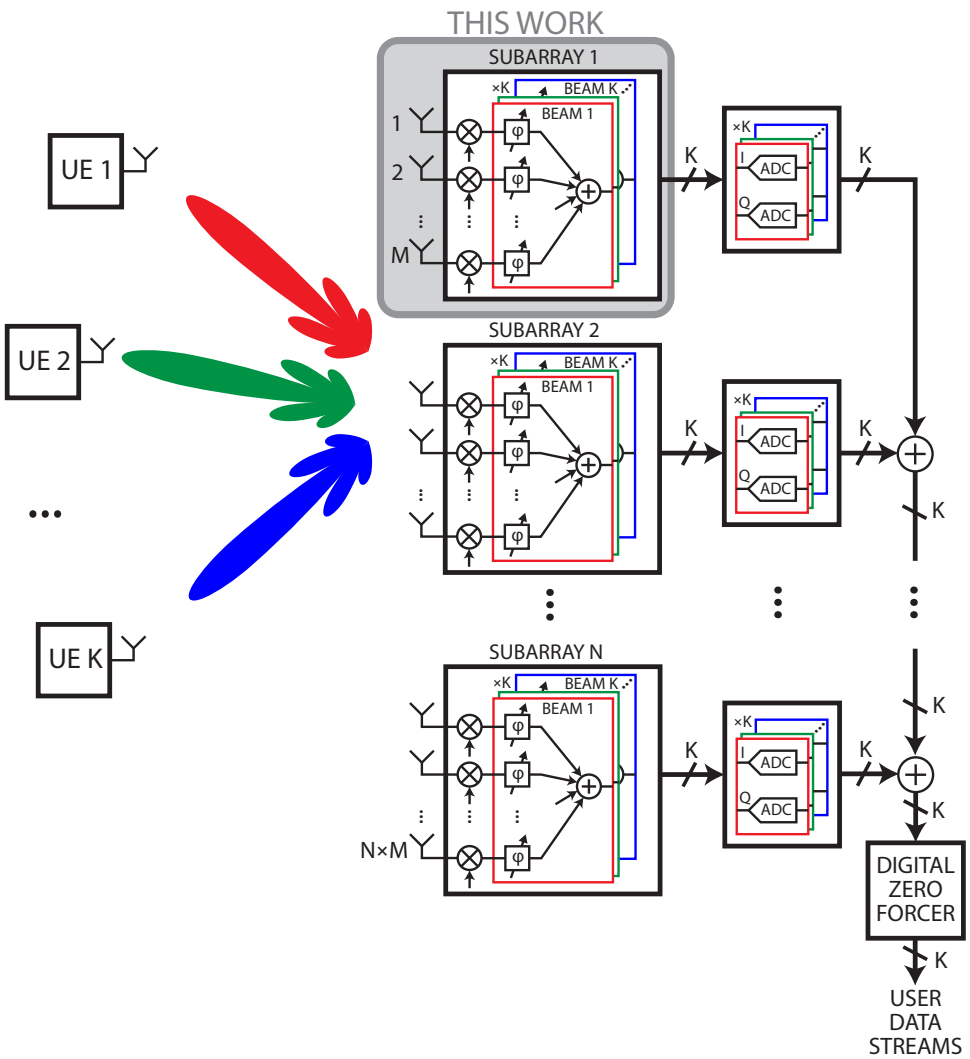

Figure 1.5: Massive MU-MIMO architecture with two-step fully-connected beamforming.

Beamformer Placement

Beamforming can be performed at nearly any stage in the signal path, with many of the options shown in Fig. 1.6-1.8. Beamforming transforms from an array of antennas to one or more beams, so the advantage of early beamforming is it decreases the number of signal chains after, as $K \leq M$. However, beamforming performed later has the advantage of the beamformer itself being smaller, lower power, and able to perform more complex weighting and processing operations.

Digital beamforming is able to perform the highest resolution with multiple stages of processing, alignment, or learning. For some beamforming or channel learning algorithms,

this level of intense processing may be a requirement. Thus the system must manage as many signal chains as there are antennas, and the amount of hardware and cross array transport of many signals may limit the power efficiency, size, or number of antenna elements that a digital beamforming array can support.

Analog baseband beamforming is the next step up from digital beamforming and reduces the number of ADCs in the array. Baseband analog is more efficient for low/medium resolution beamforming operations and signal bandwidth greater than a few hundred of MHz [32][29]. It can still handle a large matrix of beamforming weights, but is best suited to frequency flat channels and a two rather than three dimensional beamforming arrays. In addition, the control loop for estimating the channel matrix and setting the weights in the BF matrix is a larger loop to close and may make rapid beam switching more difficult.

Beamforming can be performed with the mixer as the phase shifting element and the summation in baseband. This does take the phase shifting stage out of the signal path, but has trouble with amplitude control and thus is limited to only conjugate beamforming. Importantly, however, this complicates the LO distribution and phase noise management.

RF beamforming has the significant advantage of reducing the number of mixers and thus the complexity of the LO distribution, which is a high power and challenging design block. RF beamforming can achieve impressively wide bandwidth beamforming [34][33]. However, the loss of passive elements at mm-wave and above, as well as the size of RF beamformers disqualifies them from fully connected, many user ($K > 2$) applications.

Antenna level beamforming is the furthest end of moving the beamformer up the signal path[37]. This reduces the number of every block in the signal chain, however, the NF hit from beamforming before the LNA, means there is a trade-off in the LNA between reduced number but tighter noise requirement.

At this point, it is also worth mentioning hybrid beamforming which is a combination of any of the analog options, and digital beamforming. Analog beamforming is used to reduce the spatial resolution to something $< M$ and $\geq K$, getting some of the advantage of reduced number of signal chains, while maintaining enough spatial resolution to perform

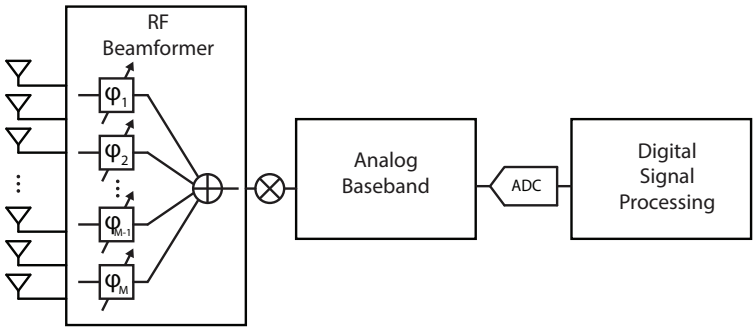

Figure 1.6: Array architecture with RF beamforming.

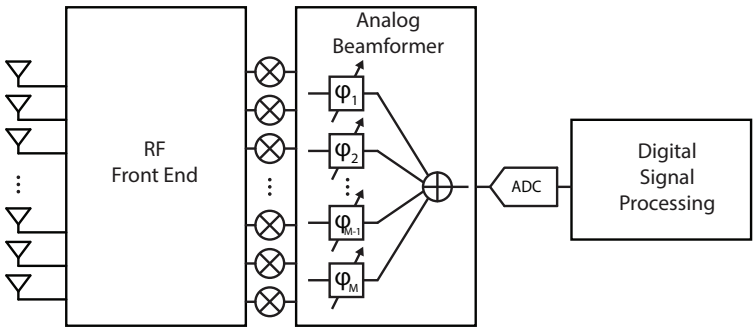

Figure 1.7: Array architecture with analog baseband beamforming.

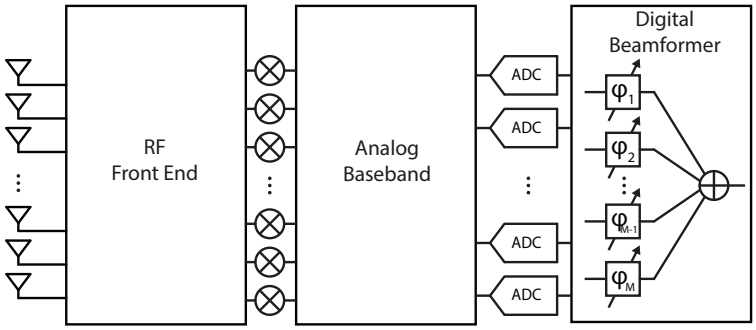

Figure 1.8: Array architecture with digital beamforming

a second more complex step of beamforming in digital. This does balance the trade-offs between number of signal chains and spatial resolution in beamforming, which trades power and hardware complexity for spectral efficiency. The Two-Stage Beamforming is a type of hybrid beamforming, as the ZF operation is broken into a local analog step and a back-end digital step, and it does so without loss of spatial resolution when the analog BF step is fully connected. This is a linear beamforming operation, but it only supports ZF or MMSE, and requires a fully connected beamformer in analog. There are hybrid beamforming options that opt for reduced spatial resolution, or a weight matrix that is less easily divided. However, calculating the analog and digital weight matrices can be very computationally expensive [38].

From this examination, two stage beamforming is an excellent architecture for scaling to massive MIMO with many users, and will be the assumed architecture moving forward. Two stage beamforming could support other forms of hybrid beamforming, but this is a discussion left to other work, this dissertation will assume ZF beamforming, with conjugate BF performed at the subarray level. We can also exclude RF and mixer BF placement options, but the optimal decision between analog baseband and digital BF for the local

conjugate BF operation is not immediately obvious and is investigated further in Chapters 2, 3, and 4.

Chapter 2

ADC Modeling

As the bandwidth available at high carrier frequencies increases, the ADC becomes a significant power consumer in Massive MU MIMO. As such, the choice between analog and digital BF has major implications on the ADC. As seen in Fig. 2.1, the first and most obvious effect is the number of ADCs needed. In digital beamforming, the signal from each antenna is digitized, so digital beamforming needs M ADCs. Whereas, both RF and baseband analog beamforming transform the channel information from antenna domain to beam domain when they perform the beamforming matrix operation. This means that there are K ADCs needed for analog beamforming architectures. This comparison initially looks very challenging for digital beamforming, as in Massive MIMO $M \gg K$, and in mm-wave or THz applications, the wide bandwidth needed makes even a single ADC a very high power block.

Changing the data dimensionality, however, is not the only effect of beamforming before the ADC. The array gain also increases the SNR of the signal. Thus, in the analog beamformer case the signal going into the ADC is higher SNR than in the digital beamformer case. This implies that the analog beamfomer needs fewer in number but higher resolution ADCs, while the digital beamformer needs more lower resolution ADCs. This has led to an interest in very low resolution digital beamforming schemes in the interest of making digital beamforming arrays power feasible [39][40]. Low ADC resolution does, however, come with a reduction in SNR, which is further discussed in Section 2.4 and Section 2.7.

Thus, we see that there is a trade-off between number of ADCs and their resolution. There have been many published comparisons of throughput for RF, hybrid, and digital beamforming[41, 42, 43, 44, 45, 46, 47]; as well as analyses of capacity in low-resolution digitization and beamforming[48, 49, 50, 51]. However, innate to these analyses is an assumption that analog beamforming is not capable of handling a fully connected beamformer, resulting in the conclusion that digital beamforming is more spectrally efficient. More recent publications have, however, shown that analog beamforming can handle large beamforming matrices at baseband, with sufficient amplitude and phase resolution. Unfortunately, decoupling the results of throughput from assumptions about beamforming placement and spatial resolution loss is very difficult. To more intuitively compare the analog and digital beamforming architectures, this chapter will define required ADC resolution in terms of common

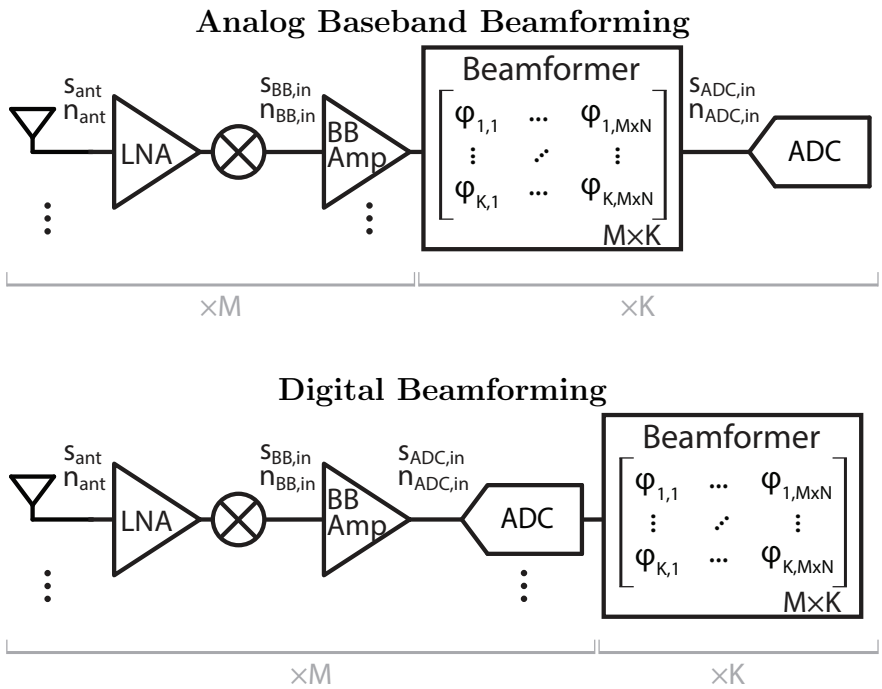

Figure 2.1: High level model of analog baseband beamformer(top) and digital beamformer(bottom) signal chain.

design variables, such as bandwidth, signal SNR, M, K, and SNR degradation. These design variables are held constant between architecture types, and by doing so the capacity is held constant. (For another example of holding output SNR constant and comparing only power, see [52].)

2.1 ADC Figure of Merit

The ADC model can be defined by three specifications: sampling frequency (f_s), number of bits (B), and power consumption (P_{ADC}). The Walden FOM (FOM_W) for ADCs is a good estimate of ADC power per resolution and sampling frequency for $B \leq 12$ and $f_s < 440MHz$ [53]. The definition of the Walden FOM follows.

$$\text{FOM}_W = \frac{P_{ADC}}{fs2^B} \tag{2.1}$$

The Schreier FOM ($\text{FOM}_S = \text{SNDR}_{ADC} + 10\log_{10}(f_s/(2P_{ADC}))$) is best used for dynamic range $> 74dB$ [53], thus the ADC model switches from Walden to Schreier FOM for ADCs beyond $B = 12$. In cases when this is necessary, the text will explicitly note the switch. How-

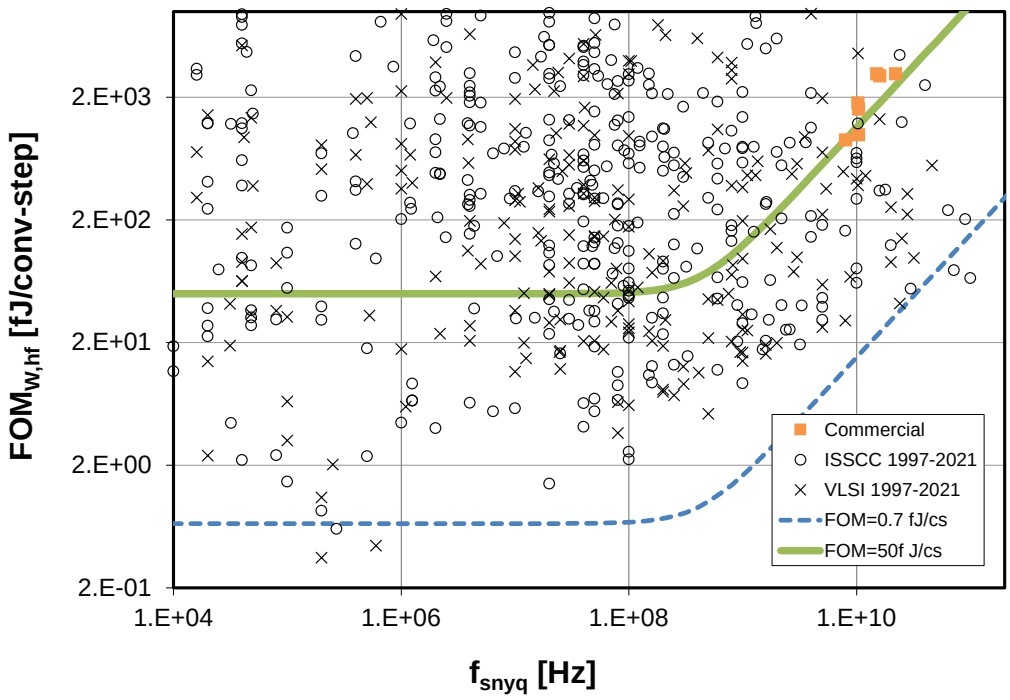

Figure 2.2: Murmann survey with commercially available high sampling frequency ADCs added.

ever, this dynamic range is relevant only very rarely in mm-wave, where signals experience high attenuation, and is considered an edge case.

Notably, both FOMs, have limited accuracy at high frequency. As seen in Fig. 2.2, ADCs published between 1997-2021 with a competitive FOM_W become less available after $f_s = 440MHz$.

This indicates a departure of real world designs from the FOM estimate, often due to various timing and technology limitations. To correct for this, the model presented here will follow the fitting curve used in the Murmann ADC survey, to model an appropriately scaled ADC FOM ($\text{FOM}_{W,adj}$) [53].

$$\text{FOM}_{W,adj} = \frac{P_{ADC}}{f_s 2^B}\sqrt{1+\left(\frac{f_s}{440MHz}\right)^2} = \text{FOM}_W \eta_{adj} \tag{2.2}$$

The high frequency adjustment term will be simplified as η_{adj} for brevity. The adjustment term η_{adj} encapsulates the efficiency limitations of high frequency ADCs in an easily modeled equation. The $\text{FOM}_{W,adj}$ allows us to extend the ADC model to wide-bandwidth applications, while comparing in power. This enables fair comparison of low bandwidth arrays to high bandwidth arrays, where an equally power efficient ADC may not exist.

Thus, we can write the single ADC power model as:

$$P_{ADC} = 2^B \, \mathrm{FOM}_{W,adj} \, f_s \, \eta_{adj} \tag{2.3}$$

From Eq. (2.3), we see that the power of a single ADC relies on bandwidth ($f_s = 2\, f_{BW}$), ADC figure of merit ($\mathrm{FOM}_{W,adj}$), and the number of bits (B). Of those, only B will change with array architecture.

2.2 ADC SINAD and Signal Degradation

The number of bits required of an ADC is calculated from the maximum input signal and maximum acceptable quantization noise. The quantization noise is defined by the RMS quantization error in terms of least-significant-bit (LSB) is $n_\Delta = (\mathrm{LSB}/\sqrt{12})^2$ (note that it is noise power, not spectral density). We assume a sinusoidal input signal occupying the full swing of the ADC, $s_{FS} = (2^B \mathrm{LSB}/(2\sqrt{2}))^2$. Because SNR is typically given in dB, we let function $\mathrm{dB}(x) = 10\log_{10}(x)$. The inverse is $\mathrm{dB}^{-1}(x) = 10^{x/10}$. Then the ADC SINAD is:

$$\mathrm{SINAD}_{ADC} = \mathrm{dB}\left(\frac{(2^B \,\mathrm{LSB}\, \sqrt{12})^2}{(\mathrm{LSB}\; 2\sqrt{2})^2}\right) \tag{2.4}$$

$$= B\,\mathrm{dB}\,(4) + \mathrm{dB}\left(\frac{3}{2}\right) \tag{2.5}$$

Which is typically written as:

$$\mathrm{SINAD}_{ADC} = 6.02B + 1.76 \tag{2.6}$$

We would like to use SNR degradation as a design variable, and thus need to be able to calculate the number of bits required to achieve a given SNR degradation. This requires a few steps, as follows.

We start by calculating the SNR degradation (SD) of the input signal due to additive white quantization noise. Let the SNR before the ADC be $\mathrm{SNR}_{ADC,in}$ and the SNR after be $\mathrm{SNR}_{ADC,out}$.

$$\mathrm{SD}_{ADC} = \mathrm{SNR}_{ADC,in} - \mathrm{SNR}_{ADC,out} \tag{2.7}$$

The SD is obviously very similar to the definition of NF, but NF is defined for a standard noise temperature input. Since we specifically need to examine effect of beamforming before or after the ADC, we must define a non-standard noise input. Thus, we prefer to use SD so as to not to overload the term NF.

Equation (2.7) can be rewritten in terms of input noise $n_{ADC,in}$ and quantization noise n_Δ, both in power.

$$SD_{ADC} = \mathrm{dB}\left(\frac{n_{ADC,in} + n_\Delta}{n_{ADC,in}}\right) = \mathrm{dB}\left(1 + \frac{n_\Delta}{n_{ADC,in}}\right) \tag{2.8}$$

We see then that the SNR degradation due to the ADC can be fully defined by the ratio of the input noise to the quantization noise. This will be a useful value which we define as:

$$q = \frac{n_\Delta}{n_{ADC,in}} \tag{2.9}$$

We can solve for q in terms of design variable SD_{ADC} only, imparting design variable status to q.

$$q = \mathrm{dB}^{-1}(\mathrm{SD}_{ADC}) - 1 \tag{2.10}$$

Note that we will generally use a $q < 1$ which is equivalent to $\mathrm{SD}_{ADC} < 3dB$. Intuitively, if $q = 1$ then the quantization noise is the same power as the noise at the input, and the output noise power is then doubled as expected at the 3dB point. Smaller values of SD_{ADC} require smaller q, which means $n_\Delta < n_{ADC,in}$.

At this point, we still need to incorporate SINAD_{ADC} so we can solve for the number of bits. Towards that goal, we rearrange q. We assume that the maximum input signal occupies the entire ADC dynamic range ($s_{ADC,in} = s_{FS}$), enforced by programmable gain in the signal chain before-hand. Then, we can write a relation between SINAD_{ADC} and q.

$$\mathrm{dB}\,(q) = \mathrm{dB}\left(\frac{n_\Delta}{n_{ADC,in}}\right) = \mathrm{dB}\left(\frac{\frac{s_{ADC,in}}{n_{ADC,in}}}{\frac{s_{FS}}{n_\Delta}}\right) = \mathrm{SNR}_{ADC,in} - \mathrm{SINAD}_{ADC} \tag{2.11}$$

We can then define SINAD_{ADC} in terms of design variables $\mathrm{SNR}_{ADC,in}$ and q.

$$\mathrm{SINAD}_{ADC} = \mathrm{SNR}_{ADC,in} - \mathrm{dB}\,(q) \tag{2.12}$$

Spending a little time getting intuition with Eq. 2.12, we see that as expected SINAD_{ADC} increases with the SNR of the signal that must be represented, $\mathrm{SNR}_{ADC,in}$. In addition, as the desired SD_{ADC} gets smaller, q becomes much less than one. This leads to $dB(q)$ becoming a negative number with large absolute value. It follows intuitively that to get low SNR degradation from the ADC, we need an ADC with higher SINAD and thus higher resolution.

All of this, combined with Eq.2.6, allows us to finally solve for the number of bits needed in terms of $\mathrm{SNR}_{ADC,in}$ and q.

$$B = \frac{\mathrm{SNR}_{ADC,in} - \mathrm{dB}\,(q) - 1.76}{6.02} \tag{2.13}$$

The use of $\mathrm{SNR}_{ADC,in}$ and SD_{ADC} as design constants allows us to define ADC resolution specifications separately from the signal chain design. The noise or gain of the preceding analog signal chain can be ignored as long as it: (1) Guarantees enough gain to fill the dynamic range of the ADC, and (2) Agrees on a $\mathrm{SNR}_{ADC,in}$ at the input of the ADC. In actual use, the signal power received at the antenna array is expected to change, then the SD_{ADC} is the maximum acceptable SNR degradation for the largest expected $\mathrm{SNR}_{ADC,in}$, as B will then be sufficient to deal with lower power signals.

2.3 ADC Total Power Budget

The total ADC power budget can now be written for n ADCs. Digital BF requires n = M ADCs, one for each antenna; while analog BF requires n = K ADCs, one for each user.

$$P_{ADC} = \mathrm{n}\, 2^{\frac{\mathrm{SNR}_{ADC,in} - dB(q) - 1.76}{6.02}}\, \mathrm{FOM}_W\, f_s\, \eta_{adj} \tag{2.14}$$

This can be further simplified by using Eq. (2.5), and exercising the logarithm change of base rule ($\log_{10}(x) = \log_2(x)/\log_2(10)$) to get rid of the exponents and see a more intuitive equation for ADC power.

$$P_{ADC} = \mathrm{n} \sqrt{\frac{s_{ADC,in}}{n_{ADC,in}} \frac{1}{q} \frac{2}{3}}\, FOM_W\, f_s\, \eta_{adj} \tag{2.15}$$

The values of FOM_W, f_s, and η_{adj} are design constants, common to analog and digital beamforming architectures for fair comparison. The design variable q will also be held constant, but benefits from additional explanation as to why. The SNR of the signal that arrives at the antenna array is constant because we are comparing the same number of users with the same transmit power. The output SNR is held constant to hold capacity constant. This means that the SNR degradation of the signal chain must be held constant, by definition. The analog and digital beamforming architectures could distribute SNR degradation differently between baseband blocks: amplifiers, beamformer, ADC. As will be shown, however, the ADC is a significant power block, especially at wide-bandwidth. (Comparison to RF blocks is performed later in the chapter.) Thus, to a first order, is it preferable to choose the largest acceptable SD_{ADC} regardless of the beamformer placement, and so q is held constant in the comparison. With these variables held constant, P_{ADC} changes only with n and $\mathrm{SNR}_{ADC,in}$ between the analog and digital beamforming cases. The following sections will address the effects of beamforming on these values.

2.4 ADC SNR Input

The $\mathrm{SNR}_{ADC,in}$ is set by many design variables in the signal chain, such as expected signal power at the antenna input, NF of the RF blocks, and SD of the baseband blocks. In the case of analog beamforming, however, array gain must be included as well. In addition, this section will examine the effect of interferers on the necessary ADC resolution.

Signal Dominated Environment

Let the average signal power of a single user at the antenna array be $s_{ant,1}$, and let all user signals be independent. For K users, the total signal at the input of the array is a combination of all user signals such that $s_{ant,tot} = K s_{ant,1}$. The SNR at the antenna array is then $\mathrm{SNR}_{ant} = K s_{ant,1}/n_{ant}$.

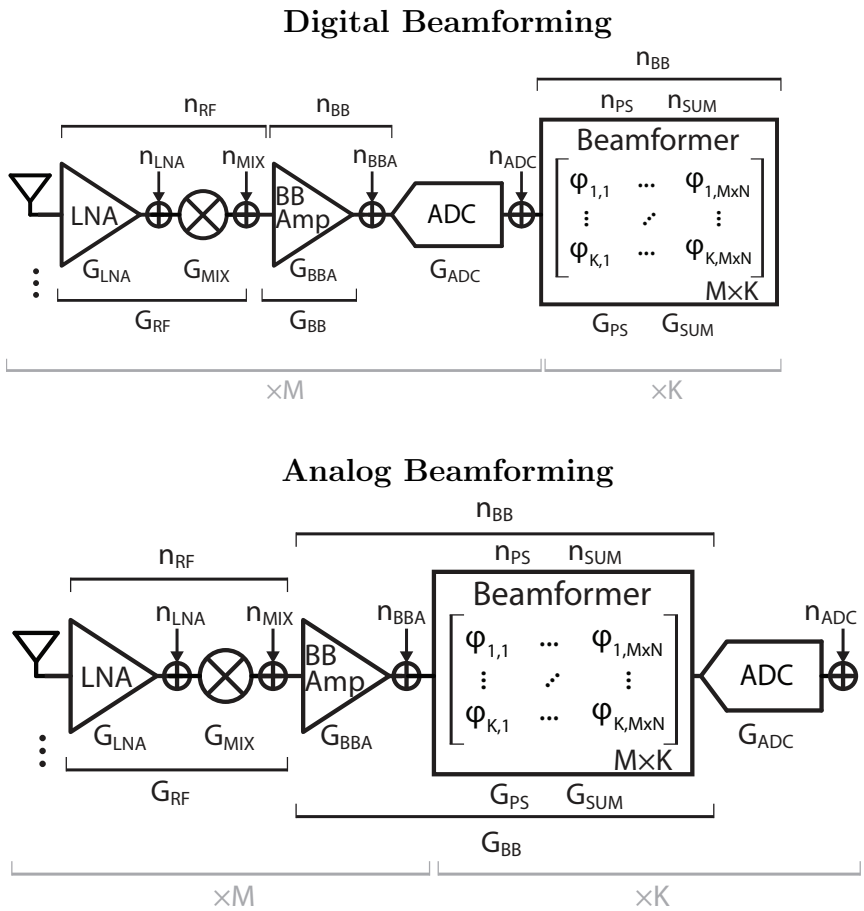

Figure 2.3: Noise contribution before ADC in baseband analog signal chain, comparison between analog and digital beamforming.

In the digital beamforming case, a signal path from antenna to ADC applies the same SNR degradation to each of the K signals. Then the input ADC SNR can be written as:

$$\mathrm{SNR}_{ADC,in,D} = \mathrm{SNR}_{ant} - \mathrm{NF}_{RF} - \mathrm{SD}_{BB} \tag{2.16}$$

$$= \mathrm{dB}\left(\frac{K\,G_{RF}\,G_{BB}s_{ant,in,1}}{G_{RF}\,G_{BB}n_{ant} + G_{BB}n_{RF} + n_{BB,1}}\right) \tag{2.17}$$

$$= \mathrm{dB}\left(\frac{K s_{ADC,in,1}}{n_{ADC,in}}\right) \tag{2.18}$$

The NF_{RF} and SD_{BB} are design variables that will be held constant between architectures for fair comparison. Note, that n_{BB} in the digital case is separated into two parts in digital beamforming. This is not a problem as $n_{BB,2}$ is assumed to be zero. Further explanation of balancing the baseband noise budget can be found in Section 3.1.

Recall that we define $\text{SNR}_{ADC,in,D}$ at the threshold between the baseband chain and the ADC, so the terms $s_{ADC,in,1}$ and $n_{ADC,in}$ are design variables. The specific implementation of the signal chain does not matter as long as the ADC design and the analog signal chain design agree on the same $\text{SNR}_{ADC,in,D}$ at the interface. Then, the key observation in Eq. (2.18) is that the $\text{SNR}_{ADC,in,D}$ scales with K.

For analog BF, each ADC serves one of K beams pointed at each user. As shown in Eq. (1.1), the in-beam user signal is summed coherently and boosted by M^2, while the noise is summed non-coherently and experience a gain of M. This provides an SNR boost to the in-beam signal.

The rejection of out of beam signals depends on their angle of arrival, as the beamformer provides some spatial filtering of the other users. This work assumes the application is the first stage of local beamforming in a two-stage beamforming architecture. Thus, the beamforming is MRC, which maximizes SNR, but not SINR. The MRC beam has a regular beam shape, shown again in Fig. 2.4, which depends only on the angle of arrival of the desired user signal. The out-of-beam rejection of inter-user interference due to beamforming is then highly dependent on the angle of arrival of the interferer. However, for a uniformly distributed, random angle of arrival given many users, we estimate an out of beam rejection of dB(M). As seen in Fig. 2.4, the side lobe peaks asymptotically approach dB(M). This also aligns with the intuition that out of beam signals should combine non-coherently like noise.

Thus, we have one user which experiences M^2 array gain, and $(K-1)$ users which experience M array gain.

$$SNR_{ADC,in,A} = \text{SNR}_{ant} - \text{NF}_{RF} - \text{SD}_{BB} + \text{Array Gain} \tag{2.19}$$

$$= \text{dB}\left(\frac{G_{RF}\,G_{BB}M^2 s_{ant,in,1} + G_{RF}\,G_{BB}M(K-1)s_{ant,in,1}}{G_{RF}\,G_{BB}Mn_{ant,in} + G_{BB}Mn_{RF} + Mn_{BB}}\right) \tag{2.20}$$

$$= \text{dB}\left(\frac{(M+K-1)s_{ADC,in,1}}{n_{ADC,in}}\right) \tag{2.21}$$

It is important to note that the signal and noise power, $s_{ADC,in,1} = G_{RF}\,G_{BB}s_{ant,in,1}$ and $n_{ADC,in} = G_{RF}\,G_{BB}n_{ant} + G_{BB}n_{RF} + n_{BB}$, are the same in Eq. 2.18, and Eq. 2.21. This allows us to see that the only difference between $\text{SNR}_{ADC,in,A}$ and $\text{SNR}_{ADC,in,D}$ comes from array gain and filtering of inter-user interference in terms of M and K. If a beamforming other than MRC is used, such as ZF or an algorithm with more interference cancellation, then the $(K-1)$ term would disappear from the numerator.

From this we see that $\text{SNR}_{ADC,in,A}$ is larger than $\text{SNR}_{ADC,in,D}$ and depends on values of K and M. Thus, the analog beamforming ADCs will require more resolution than the ADC for digital beamforming.

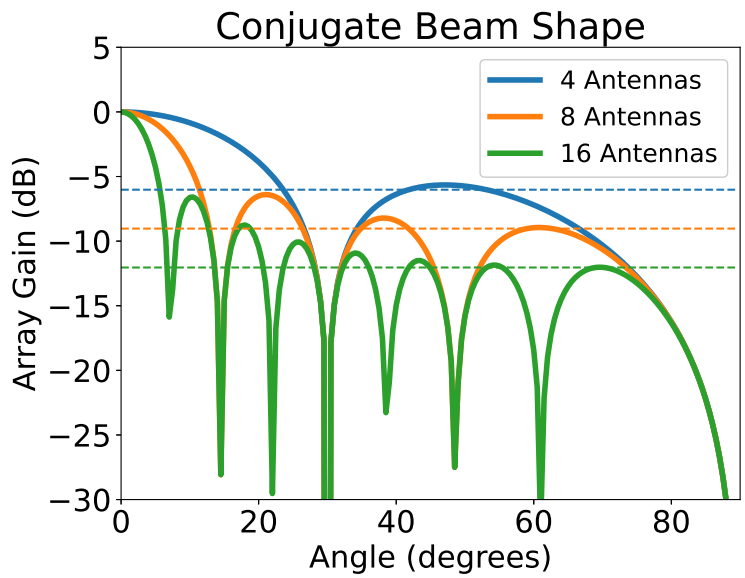

Figure 2.4: Beam shape shows that side-lobe peaks asymptotically $10\log_{10}(M)$. Blue:$10\log_{10}(4) = 6.02dB$, Orange:$10\log_{10}(8) = 9.03dB$, Green:$10\log_{10}(16) = 6.02dB$

Interference Dominated Environment

A wireless channel with large interference present will effect the ADC resolution required. The interferer is assumed to be an out-of-network interferer which cannot be handled with power control within the network, such as an out of network base station in the same frequency band.

It may not be not immediately clear why a large interferer must change the ADC resolution and not just the gain of the baseband amplifiers. Recall that the design variable q is held constant for analog and digital beamforming, and sets the required ratio between the input noise $n_{ADC,in}$ and the quantization noise n_Δ. Decreasing the amplifier gain would allow the ADC to accommodate a larger interferer without saturating, but not without also decreasing $n_{ADC,in}$. Thus the ADC resolution must also increase to decrease n_Δ and maintain the same q.

When calculating B for a signal dominated environment, any interference power (s_I) not filtered before the ADC must be added to the desired signal $s_{ADC,in}$. The ADC SNR in the digital beamforming case is then

$$\mathrm{SNR}_{ADC,in,D} = dB\left(\frac{Ks_{ADC,in,1} + s_I}{n_{ADC,in}}\right) \tag{2.22}$$

The analog beamforming provides some spatial filtering of out-of-beam interferers, as seen with the inter-user interference in Equation (2.21). The out of network interferer combines

non-coherently, and thus, similar to noise, does not experience an SNR boost from the beamforming.

$$\mathrm{SNR}_{ADC,in,A} = dB\left(\frac{(M+K-1)s_{ADC,in,1} + s_I}{n_{ADC,in}}\right) \tag{2.23}$$

The analog beamformer still requires more resolution from the ADC than the digital case, but as s_I becomes $\gg (M+K-1)s_{ADC,in,1}$ the difference becomes smaller relative to the interferer. This is a problem for digital beamforming, as the lower resolution required in ADCs is expected to make up for needing more ADCs. This means that in an application space with large interferers expected, digital beamforming ADCs may not expect the same lower resolution advantage. This helps motivate our interest in modeling the ADC power in detail, so that we may characterize at what interferer size this begins to matter

It should also be noted that this is only the case for MRC beamforming, and that a beamforming algorithm that is able to place nulls at large interferers will look more like the signal dominated environment.

2.5 ADC Power Comparison

Having shown how the ADC power is modeled using the frequency adjusted Walden figure of merit and how the SNR at the input of the ADC varies with M and K, we can now examine how the total ADC budget changes with array size, number of users, and analog or digital beamforming.

Signal Dominated Environment

We can combine ADC power, Eq. (2.3), and signal at the input to the ADC in the case of digital beamforming, Eq. (2.6). For digital beamforming, there is one ADC for each antenna (n = M) which must represent input signal SNR that scales with K.

$$P_{ADC,D} = M\sqrt{\frac{K s_{ADC,in,1}}{n_{ADC,in}}\frac{1}{q}\frac{2}{3}}\,FOM_W\, f_s\, \eta_{adj} \tag{2.24}$$

For analog beamforming, there is an ADC for each beam (n = K) which must represent an SNR that scales with $M+K-1$.

$$P_{ADC,A} = K\sqrt{\frac{(M+K-1)s_{ADC,in,1}}{n_{ADC,in}}\frac{1}{q}\frac{2}{3}}\,FOM_W\, f_s\, \eta_{adj} \tag{2.25}$$

The FOM_W, f_s, and η_{adj} are variables chosen by the system designer and common to both architectures. This leaves only the signal power scalar and number of ADCs that changes with beamformer placement. We can more clearly compare how P_{ADC} scales with array size in both cases by removing the common terms.

$$P_{ADC,A} \propto K\sqrt{M+K-1} \qquad P_{ADC,D} \propto M\sqrt{K} \tag{2.26}$$

This matches our expectation, that for large arrays serving a single user we prefer analog beamforming. We can also see that we prefer digital beamforming only when $\alpha \approx 1$ (i.e. $M \approx K$), which is conceivable in applications with a very spatially diverse channel or well distributed users. However, this comparison also rapidly shows that for massive MIMO, where α is large (i.e. $M \gg K$) analog beamforming is going to be frequently preferable with respect to the ADC budget.

Interference Dominated Environment

Recall from Section 2.4 an interference dominant environment is one in which the ADC dynamic range is set mostly to keep from saturating on a large interferer. Eq. (2.22) and Eq. (2.23) then are written with the assumption that $s_I \gg s_{ADC,in}$. We can then re-write the estimated ADC power accounting for interferer power.

$$P_{ADC,D} = M \sqrt{\frac{K s_{ADC,in,1} + s_I}{n_{ADC,in}} \frac{1}{q} \frac{2}{3}} FOM_W \, f_s \, \eta_{adj} \tag{2.27}$$

$$P_{ADC,A} = K \sqrt{\frac{(M+K-1) s_{ADC,in,1} + s_I}{n_{ADC,in}} \frac{1}{q} \frac{2}{3}} FOM_W \, f_s \, \eta_{adj} \tag{2.28}$$

However, since $s_I \gg s_{ADC,in}$, we see that the SNR which effectively sets the required ADC resolution is about common to both power terms. Then the approximate proportionality from before can be rewritten.

$$P_{ADC,A} \propto K \qquad\qquad P_{ADC,D} \propto M \tag{2.29}$$

In which case we see that in the presence of interferers, we expect the preference for analog beamforming to increase.

2.6 Achieving High Capacity

It is interesting to take a quick tangent to apply this ADC power model to a broader question. In the case that the ADC is a dominant power block, how might this effect even higher level system decisions. For example, as ADCs become more expensive at wide bandwidth, is the pursuit of wider bandwidth which has fueled interest in these high carrier frequencies still an efficient means to increase capacity.

The Shannon capacity shows that increases in capacity can come from an increase in bandwidth (f_{BW}) or an increase in SNR. By utilizing spatial multiplexing, a term for the number of independent user streams can increase the achievable capacity. We specify the effective number of users K_{eff}, where $K_{eff} = K$ only when there is negligible inter-user interference due to $M >> K$ or from other user isolation scheme.

$$C = f_{BW} \log_2(1 + \frac{s}{n}) K_{eff} \tag{2.30}$$

Common practice says when increasing capacity, bandwidth is more power efficient than SNR, as capacity increases linearly rather than logarithmically. Of course, with the caveat that the SNR is not too excessively low, at which point the capacity is power limited. Thus, high capacity communication seeks to utilize all bandwidth available in the spectrum, and then to move to higher carrier frequencies where more bandwidth is available. However, in Section 2.1, the FOM is adjusted to model the reduction in ADC efficiency after 440MHz. If the ADC is becoming a power dominant block, it is worth asking how this reduction in ADC power efficiency effects the trade-off between bandwidth and SNR needed to achieve a certain capacity.

Equation 2.14 gives the ADC power in terms of $\mathrm{SNR}_{ADC,in}$ and bandwidth. We can re-write the long form of η_{adj} to get all SNR and f_s terms.

$$P_{ADC} = \mathrm{n}\, 2^{\frac{\mathrm{SNR}_{ADC,in} - dB(q) - 1.76}{6.02}}\, \mathrm{FOM}_W\, f_s \sqrt{1 + \left(\frac{f_s}{440MHz}\right)^2} \tag{2.31}$$

The ADC power depends on both SNR and bandwidth, which can achieve a given capacity with infinite combinations. To search for the minimum ADC power efficiently, we would like to put the ADC power in terms of capacity. This allows us to sweep a single variable (bandwidth or SNR) to find the optimum bandwidth and SNR pair to achieve the desired capacity.

The Shannon capacity is in terms of final SNR, $\mathrm{SNR}_f = dB(\frac{s}{n})$, but we need to relate it to $\mathrm{SNR}_{ADC,in}$. However, the 1+ term inside of the logarithm in the Shannon capacity means that the the same logarithm rules as 2.14 do not apply, and we must assume $s/n >> 1$ to make the simplification.

$$C = f_{BW} \log_2(1 + \frac{s}{n}) \tag{2.32}$$

$$\approx f_{BW}\, 0.33\, \mathrm{SNR}_f \tag{2.33}$$

This simplification is within 80% of the true value for $\mathrm{SNR}_f > 5$dB, and 90% for $\mathrm{SNR}_f > 7$dB. The final SNR, SNR_f, comes after array gain from beamforming, so this is frequently a valid assumption. However, we must know the array size (M) and placement of the BF to relate SNR_f to $\mathrm{SNR}_{ADC,in}$.

Let us begin with examining the case of digital beamforming for a single user: $\mathrm{n} = M$, $K = 1$. Then $\mathrm{SNR}_f = \mathrm{SNR}_{ADC,in} + dB(M)$ for a single user after the ADC. We will disregard other SNR modifications such as $dB(q)$ or oversampling, as they are constants and can be added separately.

$$P_{ADC} = M\, 2^{\frac{\mathrm{SNR}_f - dB(M) - 1.76}{6.02}}\, \mathrm{FOM}_W\, f_s \sqrt{1 + \left(\frac{f_s}{440MHz}\right)^2} \tag{2.34}$$

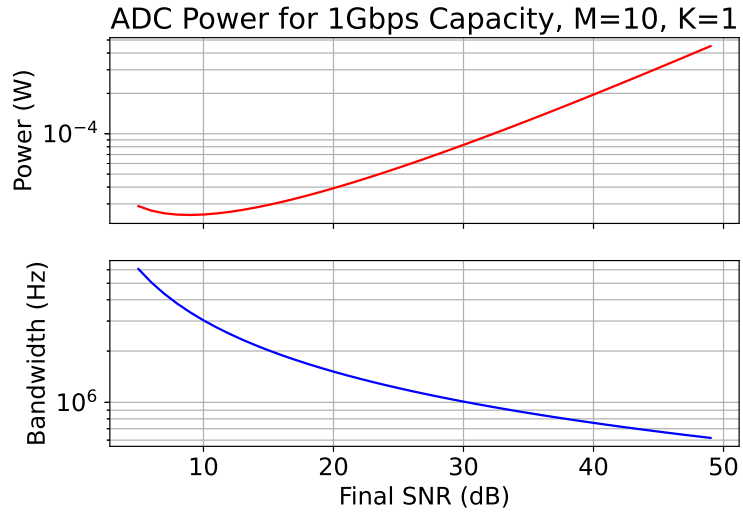

Figure 2.5: ADC Power (red) for a given desired capacity and swept SNR, required bandwidth (blue) to meet capacity. For 1Gbps capacity, minimum power: 25uW at SNR: 9dB and BW: 3.4MHz.

We will use 2.34 for the optimization that follows, but it is interesting to simplify this result a little further for intuition.

$$P_{ADC} \approx \sqrt{M}\, 2^{\frac{\mathrm{SNR}_f - 1.76}{6.02}}\, \mathrm{FOM}_W\, \frac{f_s^2}{440MHz} \tag{2.35}$$

We can see quickly that an increase in capacity with SNR requires an exponential increase in ADC power, which is expensive. However, as bandwidth exceeds 440MHz, the ADC requires a square increase in power, which is significantly more than the linear increase in power at lower bandwidths.

The optimal bandwidth and SNR combination for ADC power is still not obvious, so we wish to plot the range of combinations as desired capacity increases. To do so, we put sampling frequency ($f_s = 2\, f_{BW}$) in terms of capacity and SNR_f using Eq. 2.32. We then define a desired capacity of $C = 1$Gbps and sweep SNR_f. The results for 10 antennas, with digital BF, serving a single user with 1Gbps, 10Gbps, and 100Gbps capacity is shown in Figs. 2.5–2.7. Power is calculated from Eq. 2.34 with $FOM_W = 50fJ$/conv-step and n = 10.

As expected, large values of SNR are always sub-optimal in terms of ADC power. However, as the desired capacity increases, the minimum ADC power occurs at larger SNR. The optimal SNR occurs at 9dB for 1Gbps, 11dB for 10Gbps, and 17dB for 100Gbps. The 1Tbps case is not shown, but the optimum remains at 17dB, suggesting that this effect saturates. Looking at the blue bandwidth in all figures, we can see that the optimum SNR moves up as the required ADC sampling frequency approaches 100s of MHz. Interestingly, the opti-

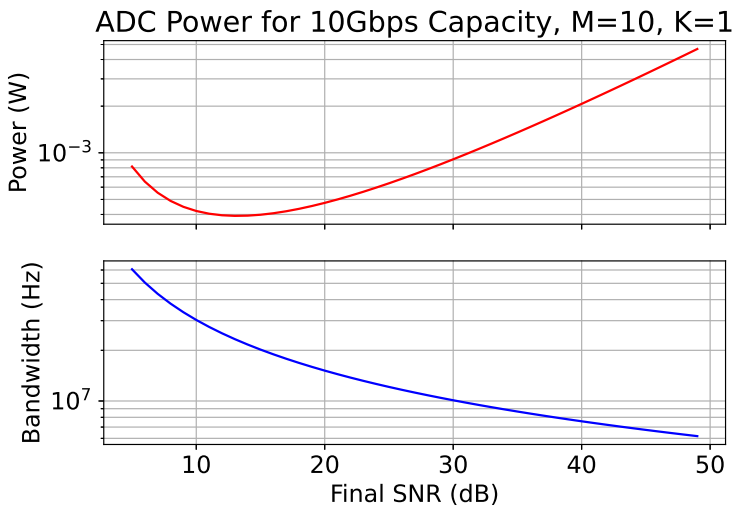

Figure 2.6: For 10Gbps capacity, minimum power: 390uW at SNR: 13dB and BW: 23MHz.

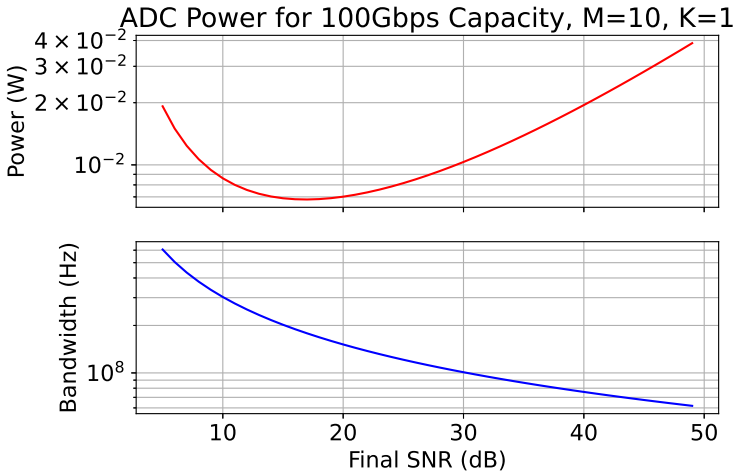

Figure 2.7: For 100Gbps capacity, minimum power: 27mW at SNR: 17dB and BW: 180MHz.

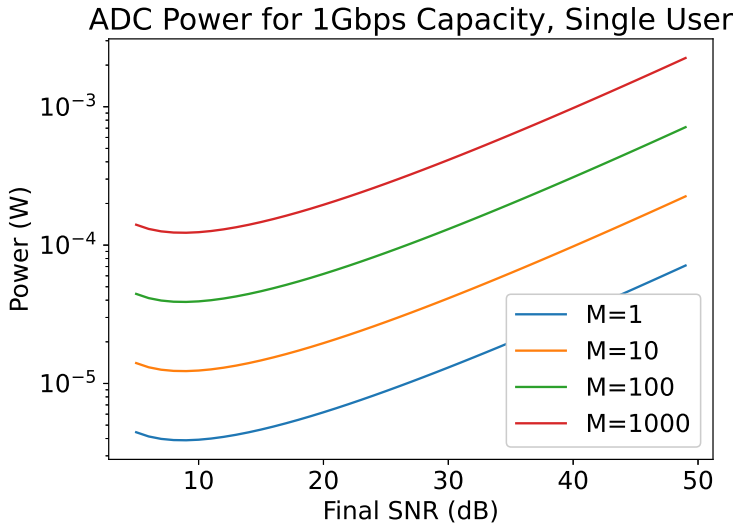

Figure 2.8: ADC Power for multiple array sizes serving a single user with 1Gbps desired capacity

mum is also shallower as the required capacity increases and the ADC η_{adj} term begins to dominate. This suggests, that for very large capacity demands, very close to the minimum ADC power can be achieved over a relatively wide range of SNR/bandwidth combinations. The system can then be optimized for other blocks or specifications in the signal chain.

We also plot the power curve as array size increases for a single user in Fig. 2.8. As expected from Eq. 2.35, increasing array size increases the ADC power by $\sqrt{M}$, and for a single user, this simply shifts the power up.

However, more fairly, as the array size is increased, the number of users can be increased. To keep $K_{eff} \approx K$, we specify a massive MIMO array with 10 antennas per user, $\alpha = 10$. This increases the capacity of larger arrays, and in Fig. 2.9 we see that a large array is more power efficient as it achieves the same capacity with lower bandwidth.

It is worth noting that these trade-offs for optimizing for minimum ADC power are less interesting for low carrier frequencies where the channel bandwidth is small and ADC power is negligible. This analysis is most useful for high carrier frequency bands where more bandwidth is available and ADC power begins to dominate.

Imagining a slightly strained design space where ADC power is the primary concern and a bandwidth $\gg 220MHz$ is always available, a final SNR of around 17dB is optimum for power. The shallow optimum could justify pushing to closer to 20dB to support a 16-QAM constellation with acceptable BER, but higher order constellations will still be less power efficient in the ADC. More realistically, however, bandwidth is frequently highly constrained by spectrum requirements and other blocks in the signal chain, and ADC efficiency is far

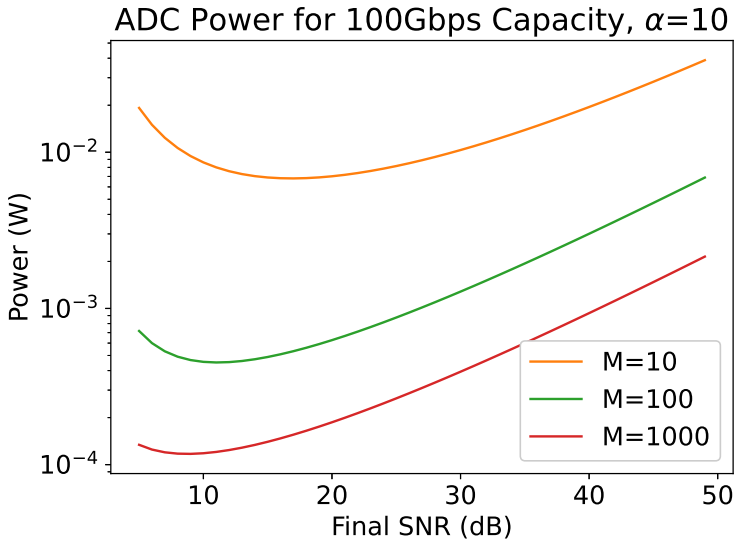

Figure 2.9: ADC Power for multiple array sizes multiple users at $\alpha = 10$ with 100Gbps desired total capacity

from the only design concern. This leads to the softer conclusion that applications seeking to push above 10s of Gbps of capacity may want to consider increasing SNR for a bit longer at mm-wave before extending to THz, for example. Or may prefer to explore more system solutions to create frequency channels, rather than supporting all of the available bandwidth in a single signal chain.

2.7 Budgeting SNR Degredation

Multi-user massive MIMO creates a somewhat unique trade-off in the RX signal chain when it comes to noise. Large arrays have the dual benefit of increasing SNR with array gain and creating narrower beams with more spatial filtering. There are then two requirements which may set array size: first where the number of antennas is chosen to close the link budget with SNR boost from array gain; second where the the number of antennas is chosen to meet the number of desired users and inter-user interference reduction needs. In the first case, it is desirable to minimize SD through the entire signal chain, as it reduces the number of antennas needed to meet a given SNR. In the second case, the number of antennas is chosen to meet a required SINR, and decreasing the signal chain SD has rapidly diminishing gains after the noise floor is pushed below inter-user interference. This implies that in the many-user massive MIMO case, there may be spare SNR in the signal chain that can be used to trade for other system specifications such as power, bandwidth, or linearity.

Since over-designing SNR degradation in the signal chain does not give any benefit when

inter-user interference will be the main noise source, the next question is where is most efficient to budget the additional SNR degradation? Common design intuition says that all the SNR slack should be spent on the NF of the LNA or front-end. However, while there does not exist the same agreement on LNA FOM as there is for ADCs, a survey of hundreds of CMOS LNAs and examination over multiple FOMs, shows that the LNA does not experience the same FOM corner in bandwidth[54]. The LNA FOMs that include bandwidth, show only a linear relationship between DC power consumption and bandwidth. Whereas, this chapter has spent a significant amount of time showing that the ADC is a significant power consumer at wide bandwidth. It may be worth taking a moment to check common design intuition as bandwidth increases and ADC power begins to become non-negligible compared to the LNA. For this discussion we will assume CMOS LNAs, as more exotic process can more easily reach high carrier frequencies, but come with off-chip routing loss and packaging concerns that are significant variables with less well established norms and best practices.

There is a simple and common FOM for LNA, using LNA power($P_{LNA,DC}$), gain ($G_{LNA,dB}$), and noise figure (NF_{LNA}):

$$P_{LNA,DC}\ (\text{mW}) = \frac{G_{LNA,dB}}{FOM_{LNA,1}(NF_{LNA} - 1)} \tag{2.36}$$

A second, recently published FOM in [54] is specifically for CMOS LNAs and includes signal bandwidth (f_{BW}), technology (L), carrier frequency (f_0), and noise measure ($F_M = (dB^{-1}(NF) - 1)/(1 - G_{LNA,dB}^{-1})$) :

$$P_{LNA,DC}\ (\text{mW}) = \left(\frac{L^{4/3} f_0^{2/3} f_{BW}^{1/3}}{F_M\, FOM_{LNA,2}} \right)^3 \tag{2.37}$$

It is worth noting that this second LNA FOM better matches the performance seen in published works, and thus provides a better power estimate. Thus we will be using the second FOM for this discussion.

The comparison is performed for multiple f_{BW}, as well as $L = 28$nm, and $G_{LNA,dB} =$ 20dB, ADC $FOM_S = 50$fJ/conv-step, LNA FOM= -15dB. The carrier frequency is adjusted for an appropriate carrier for the specified bandwidth; for example, $f_0 = 70$GHz is used for $f_{BW} = 1 - 5$GHz, and $f_0 = 160$GHz for $f_{BW} = 5 - 10$GHz. Obviously, exact results will change for different values, but these represent a somewhat common set of ADCs and LNAs, and reveal an interesting trend. The ADC curve is defined with $f_s = 2f_{BW}$, digital beamforming, $\text{SNR}_{ADC,in} = 11dB$, and $\text{FOM}_W = 50fJ$/conv-step.

Fig. 2.10 shows the estimated power spent in the LNA and the ADC to meet a range of SNR degradation specifications. The results reinforce the common intuition that the LNA power efficiency benefits the most from a large NF specification. For example let us focus on the 1GHz case. At the given gain and technology, a common NF for carrier frequencies around $f_0 = 70$GHz is $\approx 5dB$ in published works[55]. We'll assume an ADC with SD_{ADC}3dB to start, so $q = 1$. Say there are 3dB of spare SNR in the system, then Fig. 2.10 shows

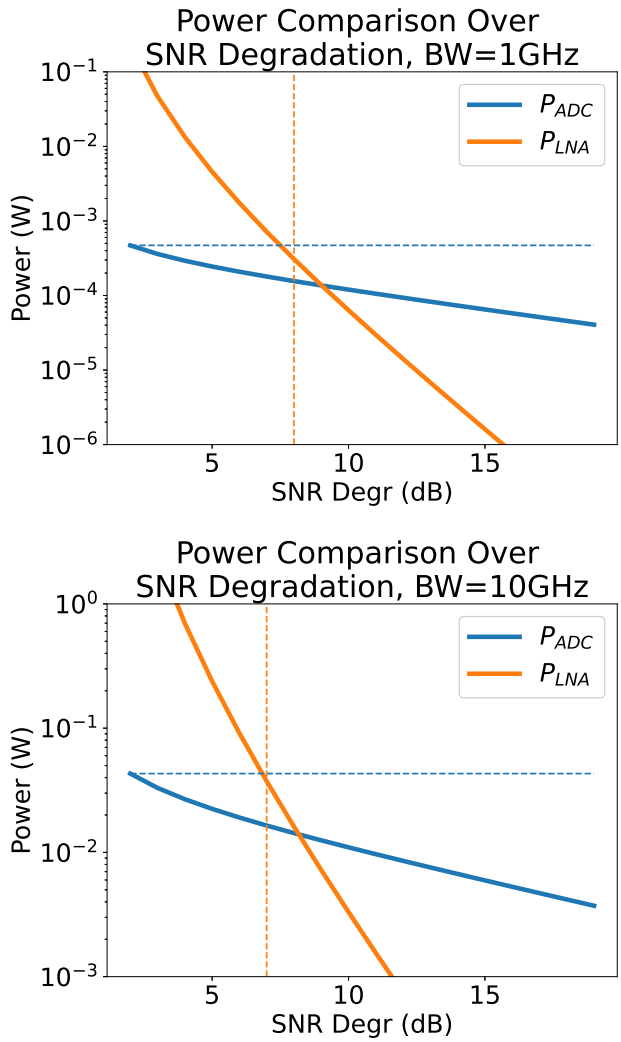

Figure 2.10: Comparison of power spent to achieve NF or SD in LNA or ADC

significant power savings if the LNA NF is allowed to increase from 5dB to ≈ 8dB, more than allowing SD_{ADC} to move from 3dB to 6dB. Indicating that it is more power efficient to invest in a lower power LNA than a lower resolution ADC. However, say there are an additional 6dB of spare SNR in the system. SD_{ADC} begins to become comparable in power around $8dB$ of SNR degradation. This indicates that an LNA with NF beyond $8dB$, should start to be balanced against the acceptable SD_{ADC} and it may depend on other design decisions which is the best place to budget the SNR degradation.

This is especially relevant in the 10GHz case, at carrier frequencies around 160GHz. The common NF values published for CMOS LNAs start to creep up 8dB, while the curves in Fig. 2.10 remain similar. Indicating that the ADC has a more immediate claim on that

same spare 3dB or 6dB of SNR degradation at higher carrier frequencies.

This result shows that low resolution ADCs are a power efficient solution for arrays with wide bandwidth above 1GHz, but only once the LNA NF has been optimized. This is more likely to be relevant at high carrier frequencies, where LNA NF is already lower by design necessity. This is good news for low resolution ADC algorithms, as the high carrier, wide bandwidth applications are the ones most in need of low resolution ADCs. It is important to recall, however, that this is only applicable in channels where the inter-user interference and spatial resolution requirements are setting the number of antennas. Otherwise, the lower resolution ADCs cut directly into the link budget.

Chapter 3

Beamformer Modeling

The ADC power budget becomes large as bandwidth, user count, and interferers increase, but to get a better idea of how large this is compared to the rest of the system, it is valuable to form a model of the other blocks in the baseband. It's an important comparison if analog beamforming saves power in the ADC, but costs more power to implement a fully connected beamforming matrix with equivalent resolution to a digital implementation. The beamformer is the only block that scales with antennas, M, times users, K, which for many-user massive MIMO, can be significant. In addition, the baseband amplifiers must provide enough gain and linearity to fill the ADC swing, which is non-trivial at very wide-bandwidths. Creating a hardware informed model of how the beamformer power scales with array size allows the designer to proceed with complicated architecture decisions in an informed manner.

We will make a few assumptions in the comparison. First, that the RF blocks in the signal chain are unchanged by the decision between analog and digital beamforming, and thus can be excluded from the comparison. Next, to help manage the cross connect between many antennas, and the inter-user interference between many, fully-connected users we will tailor the discussion towards the two-stage beamforming architecture. As discussed in Section 1.2, performing local beamforming over many antennas allows the system to manage many antennas and phase noise achieve high capacity[25]. However, many of the results still apply broadly. Finally, we will assume interferers are present when specified, but that the massiveness factor, $\alpha = M/K$, has been chosen sufficiently large such that they are out-of-beam for the intended use.

3.1 Baseband Signal Chain Challenges

While holding system specifications constant, the model performs the comparison of analog and digital beamforming in power. Similar to the ADC comparison, this provides a way to form intuition about the system by providing results in a unit which is rich with design intuition, while assuring identical signals on the output. Initial intuition may suggest that the digital beamformer is at an advantage. At absolute minimum, digital beamforming

needs only a single amplifier before the ADC, where as the analog beamformer by definition must meet those same specifications with baseband amplifier, phase shifters, summation stages, and significant interconnect parasitics before digitization. However, because the digital beamformer does not benefit from array gain, the baseband signal chain must supply more active gain than the analog beamformer case. For low bandwidths and low resolution, this gain can be very significant and may be impossible to accomplish in a single stage for a given technology. Thus, there exist several trade-offs in the baseband design. While much of the signal chain power optimization is performed using code to search the design space, there a several interesting insights into the challenges of balancing design specifications that fall out of the model.

Bandwidth

The primary challenge meeting a given signal chain bandwidth lies in balancing the many poles present in the signal chain in tandem with other design concerns such as gain and signal transport. For a desired signal bandwidth, f_{BW}, a simple bandwidth specification is to set the 3dB point of the signal chain about equal to f_{BW}. This is not a perfectly flat frequency channel for the desired signal, but results in acceptable level of noise amplification from any equalization in the following digital signal processing. For a signal chain with multiple poles and no feedback, there are multiple ways to accomplish this specification. For example, all but one or two dominant poles can be placed very far out in frequency, or all the poles can be placed moderately far out and none are particularly dominant. Both can result in the same 3dB point, but have very different implications on power and group delay. Fig. 3.1 shows the comparison of three extreme cases that attain 1GHz 3dB point. It shows that equally placed poles result in a sharp drop off, beneficial for filtering, and desirable group delay properties for eye width. However, the power spent to obtain such a frequency response is only optimal if the load, gain, and noise of each stage is equal. This is relevant in the analog beamformer case especially. For example, the input and output nodes of the beamformer are very high parasitic nodes, but there is some balancing that can be done in the design, by moving high gain and low SNR degradation requirements off of stages with high parasitic nodes and onto low parasitic stages where ever possible. Thus, there is a trade-off, managing low power operation along with pole placement.

Despite the difficulty of balancing many design requirements, the many stages of the analog beamformer can actually be advantageous. Seven stages of amplification and beamforming, provides a 7th order filter before the ADC, removing the need for or reducing the requirements of an active filter stage. There are not as many stages in the digital beamforming signal chain, and thus the frequency filter is lower order. However, the model will assume this is a secondary effect. Given the high attenuation of mm-Wave frequencies, the signal chain naturally provides enough rejection of out-of-band noise and interference. This may produce slightly optimistic results, especially for the digital beamforming case where a better frequency filter may need to be added in actual operation.

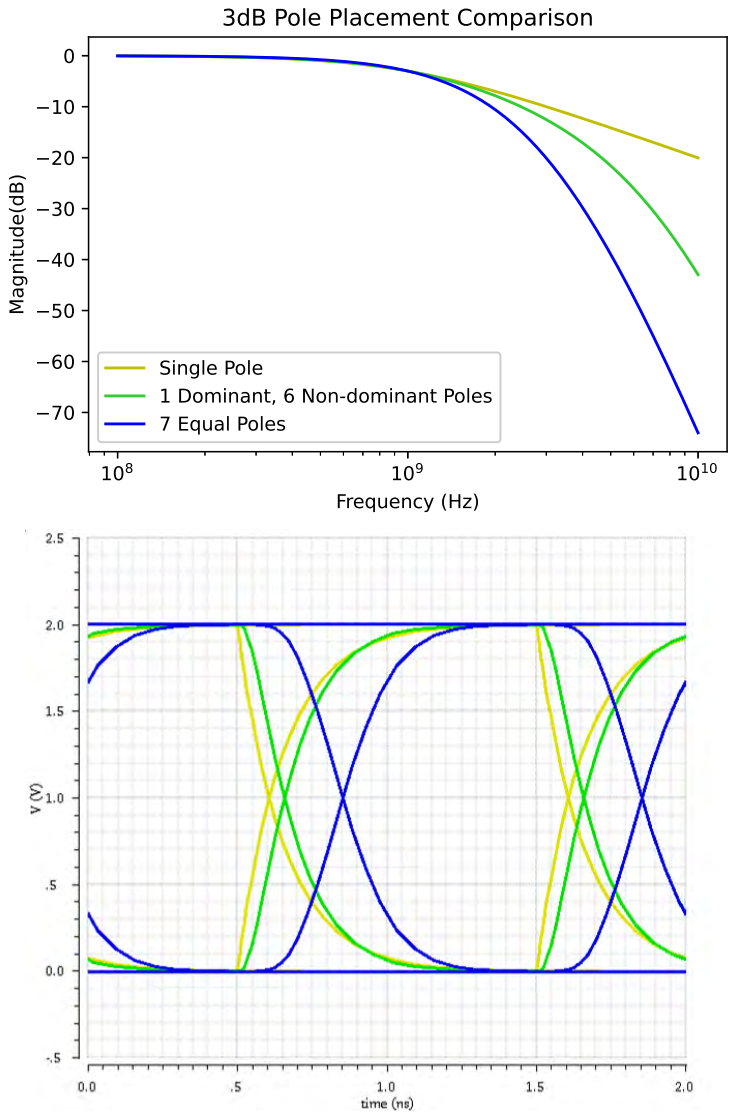

Figure 3.1: Comparison of signal chains with 3dB point at 1GHz. (Top) Frequency Domain Comparison, (Bottom) Time Domain Comparison. Yellow is a single pole at 1GHz for comparison. Green has one dominant pole at 1.1GHz and 6 non-dominant poles at 8.2GHz. Blue has 7 poles at 3.1GHz.

Noise

The model will budget the same SNR degradation between analog and digital beamforming. All SD is budgeted to the analog baseband in the digital beamformer case, assuming that the digital beamformer increases the number of bits as needed to maintain all available resolution in the beamforming operation. This is to the advantage of the baseband amplifiers in the digital BF case, as they have a looser noise requirement to make up for their higher gain requirement, as shown in Section 3.2. On the other hand the analog beamforming case budgets the given SD between all baseband stages, including the beamformer stages. This creates a more difficult noise specification for the amplifier stages, but the gain is lower, and divided between many stages. This requires careful budgeting and design in the analog beamformer in order to achieve low power in the analog beamformer case.

Linearity

Linearity of the baseband stages is of most concern at the end of the signal chain when all of the gain has been applied before the ADC. The maximum signal is limited by the ADC, but the signal chain before should target very linear operation for signals inside of the ADC range. Linearity inside of the beamformer is of great concern. Larger gain before the beamformer reduces power in the beamformer in the case the beamformer is noise limited, however, after a certain point, the power of the beamformer increases again as the linearity requirements begins to dominate. The design variable $V_x^* = 2I_D/g_m$, which is proportional to the inverse of transconductance efficiency and equal to overdrive voltage in saturation, is proportional to the input linear range and headroom of the transistor. This allows a designer to also examine the trade-off between linearity, noise, and power, as explored further in Section 3.3. Thus, the linearity is tightly tied to the noise and bandwidth specs and will be tracked through the signal chain.

Another important linearity consideration is the offset due to mismatch in the beamformer. To save power, stages in the beamformer should not target significantly more linearity than necessary. To help in reducing the linearity requirement, several stages of offset correction are necessary. Otherwise, mismatch that is gained up by subsequent stages will drive devices out of their linear range. This is not considered in depth in the model, but is very relevant in the silicon implementation.

Balancing Noise and Bandwidth Limited Design

An array of size M will increase signal power by a factor of M more than noise power, by combining M stages distributed across the array. Both the noise and the wiring parasitics on the summation node are dependent on the number of signals summed. This leads to the question of whether summation stage power is limited by noise or by bandwidth, which is highly dependent on summation stage active gain and wiring parasitic, and valuable to estimate early when designing for low power.

Section 3.4 discusses how to put the SNR degradation of each stage in terms of design variables. If a stage designed to meet the bandwidth has larger SNR degradation than the given design specification SD_{BB}, then the system is noise limited and the noise specification will define the minimum power. Otherwise the system is BW limited and the wiring parasitics and f_{BW} will set the minimum power. The beamformer is quite large, and the input and output stages of the beamformer must drive very large cross wires; this makes them a good example of a bandwidth limited stage. Generally, the system will only be noise limited for stages with low wiring parasitics or very small M and K. The phase shifter is a good example of such a stage: the phase shifter scales with $M \times K$, so there are more phase shifters than any other stage and should be placed as close to the subsequent summation stage as possible. Thus, the phase shifter is noise limited and is budgeted as much of the noise figure as is available to drive the power even lower.

By identifying the power limiting specification, the designer can then make budget specifications appropriately. For example, since the stage which drives the beamformer input is strongly bandwidth limited, it should be made a dominant pole in the bandwidth budget. On the other hand, for the noise limited phase shifter, the noise requirement can be relaxed by budgeting as much gain before the beamformer as linearity requirements will allow, which amplifies the noise floor and reduces the impact of the phase shifter's added noise. The following sections will go into further detail on how this design process is performed in detail.

3.2 Baseband Voltage Gain

In Section 2.2, defining $\mathrm{SNR}_{ADC,in}$ and SD_{ADC} as design variables allows the designer to separate the ADC specifications from the analog signal chain design. To do so, the analog signal chain must guarantee $s_{ADC,in}$ fills the ADC dynamic range. It is then necessary to solve for the appropriate voltage gain from the signal chain. When considering quantization noise ratio q, this is equivalent to the baseband voltage gain being chosen to amplify the integrated noise to the desired ratio with an LSB.

As mentioned, the model assumes that the RF blocks are identical to analog and digital beamforming signal chains. The LNA and mixer power gain G_{RF} and noise figure NF_{RF} are often designed as well as can be, but are limited by the technology. They are not considered a design variable in this work, because the optimum is generally as large as the technology will allow and the RF designer can actualize. It is useful, then, to break the design into three parts, as illustrated in Fig. 2.1. The minimum SNR is then defined at the antenna input (SNR_{ant}), the BB input ($\mathrm{SNR}_{BB,in}$), and the ADC input ($\mathrm{SNR}_{ADC,in}$). With this, we can define an expected noise and signal power at the input and output of the baseband stage.

To begin, the antenna input spectral noise in dBm is defined for room temperature (T_0) as:

$$n_{ant,dBm} = dB(k_B T_0 \Delta f / 1 \times 10^{-3}) \tag{3.1}$$

And the total signal power of K users at the baseband input for the expected SNR_{ant} as:

$$s_{BB,in,dBm} = n_{ant,dBm} + \mathrm{SNR}_{ant} + G_{RF,dB} \tag{3.2}$$
$$= K\, s_{BB,in,1,dBm} \tag{3.3}$$

The signal is then amplified by the RF stages and converted to V_{RMS}^2 on a 50Ω load. Which will allow, for comparison with the ADC voltage swing ($V_{ADC,swing}$),

$$v_{s,BB,in}^2 \Delta f = 10^{s_{BB,in,dBm}/10} \times 1 \times 10^{-3} \times 50 \tag{3.4}$$
$$= K\, v_{s,BB,in,1}^2 \Delta f. \tag{3.5}$$

We see then how to solve for a necessary baseband voltage gain to utilize the full swing of the ADC. We define total baseband voltage gain for digital beamforming case, $A_{V,BB,D}^2$; and analog beamforming case, $A_{V,BB,A}^2$.

This means that for a given signal bandwidth f_{BW} and given $V_{ADC,swing}$, the digital $A_{V,BB,D}$ is calculated straight-forwardly. With a $2\sqrt{2}$ to convert from RMS power to peak-to-peak voltage.

$$A_{V,BB,D} = \frac{V_{ADC,swing}}{2\sqrt{2}\, v_{s,BB,in,1}\sqrt{K f_{BW}}} \tag{3.6}$$

While the active gain provided by the baseband signal chain in the anlog beamforming case is:

$$A_{V,BB,A} = \frac{V_{ADC,swing}}{2\sqrt{2}\, v_{s,BB,in,1}\sqrt{(M+K-1) f_{BW}}} \approx \frac{A_{V,BB,D}}{\sqrt{\alpha}} \tag{3.7}$$

The $A_{V,BB,A}$ is then shown to be smaller than the $A_{V,BB,D}$. One intuitive interpretation is that array gain coherently combines one of the user signals and non-coherently combines the others. Thus, the voltage gain for the analog beamforming case must provide less active gain than the digital beamforming case. For an alternative explanation this can also be viewed from the noise requirement: analog beamforming requires more bits, which for the same $V_{ADC,swing}$ means a smaller LSB. The input noise to quantization noise ratio q is the same for both cases, so the noise does not need to be amplified as large relative to an LSB as the digital case does. This makes the digital baseband amplifier gain requirement harder to meet, especially in very low resolution cases, or low bandwidth cases when the noise is integrated over a smaller bandwidth. This conclusion, applies only when the SNR degradation from each stage (RF, BB, ADC) is held constant for analog and digital beamforming cases, but is a good assumption in order to make a fair comparison when the only differences in architectures are the location of the beamformer.

3.3 Baseband Amplifiers

The baseband signal chain will be modeled with two baseband amplifiers and approximations of the loads they must drive. Between the two of them, they should conveniently divide the

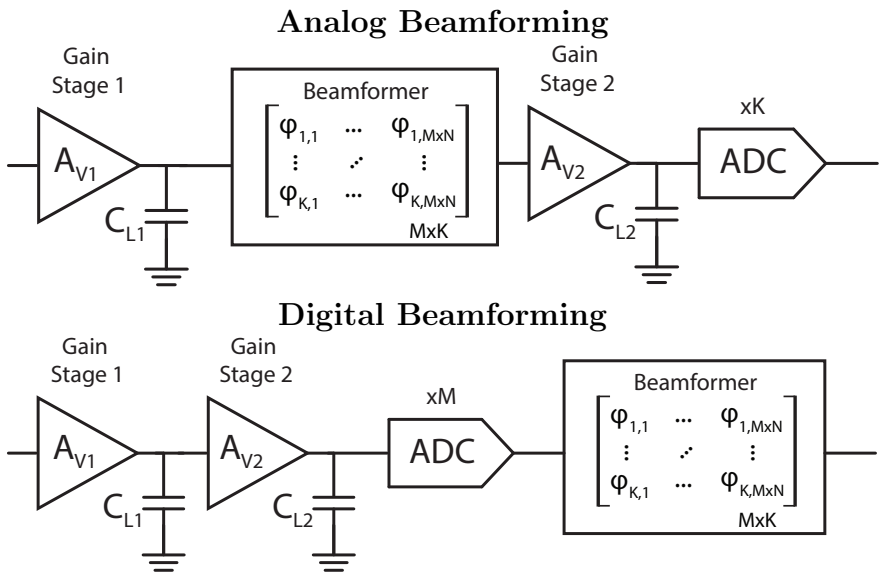

Figure 3.2: Baseband signal chain model for digital beamforming architecture

gain and cross chip wiring. The digital beamforming signal chain, shown in Fig. 3.2, is a simple implementation of this. The analog beamforming signal chain, shown in Fig. 3.2, must balance the two amplifiers with the noise, gain, and bandwidth of the beamformer as well.

The amplifiers can be modeled as a differential pair, to get a rough idea of how the gain, bandwidth, and noise will scale with power. The details of which are covered below.

Differential Amplifier with Resistive Load

As mentioned, there are two design spaces of interest: noise limited and bandwidth limited, which will define the minimum power necessary in a stage. We begin with familiar equations for gain, bandwidth, and output referred noise density in a restively loaded differential pair, as well as the definition of design variable V^* mentioned in the linearity discussion. These are obviously first order models of an already simplified amplifier, but they should give a good idea of how power scales with design requirements within reasonable design specifications for each stage.

$$A_V = g_m R_L \tag{3.8}$$

$$f_p = \frac{1}{2\pi R_L C_L} \tag{3.9}$$

$$v_n^2 = 8kT(R_L + \gamma g_m R_L^2) \tag{3.10}$$

$$V^* = \frac{2I_D}{g_m} \tag{3.11}$$

Noise Limited

The above Equation (3.8) and Equation (3.10) can be manipulated to remove R_L, which will be used to meet the bandwidth requirements.

$$v_n^2 = 8kTR_L(1 + \gamma A_V) \tag{3.12}$$

$$g_m = \frac{8kTA_V(1 + \gamma A_V)}{v_n^2} \tag{3.13}$$

Then, from Equation (3.11) the stage power can be written in terms of current, which for a constant V_{DD} is proportional to power.

$$I_D = \frac{4kTV^* A_V(1 + \gamma A_V)}{v_n^2} \tag{3.14}$$

Finally, R_L is found from the gain in Equation (3.8), and checked against the bandwidth requirement Equation (3.9). If the bandwidth is insufficient, then the stage is bandwidth limited.

Bandwidth Limited

In the case the stage is bandwidth limited, we can derive the current required with Equation (3.9) and Equation (3.11).

$$f_p = \frac{g_m}{2\pi A_V C_L} \tag{3.15}$$

$$I_D = f_p \pi V^* A_V C_L \tag{3.16}$$

To a first order, and ignoring house keeping circuits, power is seen to be proportional to gain and capacitive load, and inversely proportional to v_n^2.

This model will run into some trouble in cases of very large gain. For a given technology, the gain-bandwidth product of a single stage is limited, especially assuming a resistively loaded differential stage. We can quickly show the maximum gain of a resistively loaded differential pair from 3.8, 3.11, and Ohm's Law on R_L, $(V_{DD} - V_{CM}) = R_L I_D$, where V_{CM} is the DC common mode output voltage.

$$A_V = \frac{2}{V^*}(V_{DD} - V_{CM}) \tag{3.17}$$

In that case, the stage must be split into two. If bandwidth is the limiting design factor for that amplifier, then there is a great power advantage as the power now scales with approximately $\sqrt{A_V}$. However, in the case of a noise limited stage, the power is at a disadvantage, as optimal noise favors large gain in early stages of the signal chain. This is a difficult case to model, and can be very technology dependent. It also only applies in the case of very large gain, which only occurs for small arrays, few users, and low bandwidth as per Equation (3.6) and Equation (3.7). Thus, the model will be more pessimistic in these cases than what may be actually achievable. However, the model is already optimistic as it does not model housekeeping and reference circuits which also increase as additional stages are added.

Differential Amplifier with Active Load

There is another point to call attention to, which is that passive loads are not commonly used in actual design. As shown in Equation(3.17) their common mode output voltage is tightly tied to gain. Active loads solve this problem and the power, gain, noise, and bandwidth relations can be found similarly to above.

$$A_V = g_{m,n} r_o \tag{3.18}$$

$$v_n^2 = 8kT(\gamma g_{m,n} + \gamma g_{m,p})\frac{r_o}{2} \tag{3.19}$$

$$V^* = \frac{2I_D}{g_{m,n}} \tag{3.20}$$

$$r_o = \frac{V_A + V_{DS}}{I_D} \tag{3.21}$$

Noise to Power Derivation

$$I_D = 4kT\gamma(A_V + \frac{V_A}{V_p^*})\frac{V_A}{v_n^2} \tag{3.22}$$

Maximum Gain Derivation

$$A_V = \frac{2}{V_n^*}(V_A - (V_{DD} - V_{CM})) \tag{3.23}$$

These equations show that the maximum gain in the case of small V_{DD} is larger as generally $V_A > 2(V_{DD} - V_{CM})$, where V_A is the Early voltage. Note that this does not assume a common-mode feedback loop, but the inequality still applies when it is present. However, the active load amplifier's large output impedance can be troublesome when driving large capacitive loads. To obtain lower output impedance, a voltage follower can be added as a second stage, but this adds complexity to the model which must add another power consuming stage (where the current is chosen to bias a low enough $1/g_m$ to meet the required output impedance) and adds another pole to the frequency response of the entire chain. For simplicity, we can add a shunt resistor to the differential output. In which case, r_o is replaced with R_L and the active load diff pair modeling equations for I_D reduce back to the passive load equations. Thus, to a first order, it is acceptable to estimate baseband amplifiers as resistively loaded differential pairs, and to still assume single stage gain beyond Equation (3.17).

ADC Driver and Sampling Noise

The ADC driving stage is considered differently from the first amplifier stage. As the bandwidth and noise of the stage must be considered in relation to the ADC sampling capacitor. The output resistance of the stage is set by the acceptable settling error of the sampled

voltage onto the capacitor, which must be $\ll$ LSB. The transient voltage on a capacitor is described as follows.

$$V_{out} = V_{in}(1 - e^{-t/\tau}) \tag{3.24}$$

Then the goal is to have $V_{in}e^{-t/\tau} \ll$ LSB at sampling time $t = T_s/2$, where $f_s = 1/T_s$, for a worst case $V_{in} = V_{ADC,swing}$. The necessary τ is then:

$$\tau \ll \frac{T_s}{2}\frac{1}{\ln(V_{in}/\text{LSB})} = \frac{T_s}{2\ln(2^B)} \tag{3.25}$$

Then $\tau = R_L C_s$ can be rewritten in terms of design variables of sampling frequency $f_s = 2f_{BW}$ and sampling capacitor C_s, and solve for the driver load resistance (inclusive of switch resistance, but which we assume is comparatively small). We also define a scalar $\rho \ll 1$, so that R_L is sufficiently less.

$$R_L = \frac{\rho}{f_s C_s\, 2\ln(2^B)} \tag{3.26}$$

The C_s is then chosen for noise requirements. We begin by assuming that the switch resistance is negligible in series amplifier output resistance, thus f_s is set by R_L. Importantly, this is not the signal bandwidth as this stage has no filtering before the ADC and thus all noise will alias into the digitized signal. The added noise of the second amplifier used as the ADC driver stage is: $n_t = n_s + n_d$ in power. The first term is the sampling noise from the series resistance on the capacitor $n_s = 2kT/C_s$, with the two coming from the differential amplifier. Note that this is in units of power. Additionally, the device noise is $n_d = 2kT\gamma A_V/C_s$, which has also been integrated onto the the capacitor over the full spectrum. These will be used to solve for the necessary value of C_s given a desired noise spectrum density n_t and signal bandwidth f_{BW}, where f_{BW} is used instead of a brick-wall noise equivalent for simplicity.

$$n_t f_{BW} = \frac{2kT}{C_s} + \frac{2kT\gamma A_V}{C_s} \tag{3.27}$$

$$C = \frac{2kT(1+\gamma A_V)}{n_t f_{BW}} \tag{3.28}$$

The ADC driver stage is finally sized to provide the residual gain in the signal chain, and n_t is chosen to take up the residual noise factor left over in the signal chain, assuming the previous amplifier stage was bandwidth limited (a reasonable assumption in a large array). This will be shown in more detail in the next section.

3.4 Baseband Beamformer

At baseband, the signal is decomposed into I and Q parts, so an efficient implementation of a phase shifter is a vector modulator. Fig. 3.3 shows the vector modulator equation, which is implemented mathematically in digital, and as a combination of two amplitude modulated signals in analog.

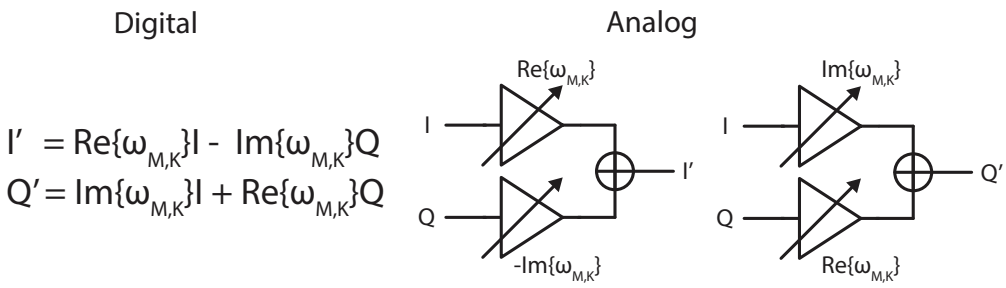

Figure 3.3: Vector Modulator implementations in digital and analog.

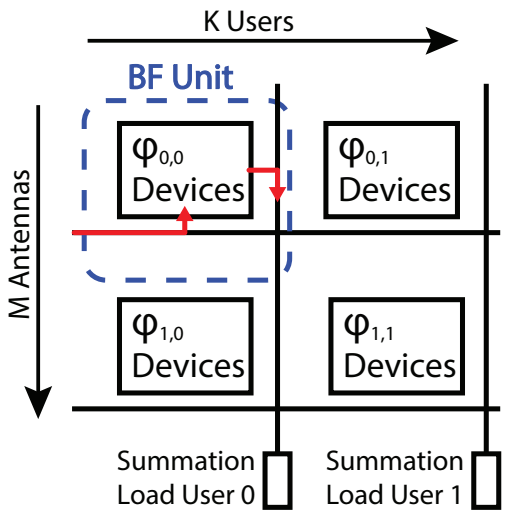

Figure 3.4: Analog BF unit blocks, show wiring and capacitance scales with M and K.

Analog

The analog vector modulator is modeled in the system above as a unit block with a unit input and output capacitance, and footprint.

Fig. 3.4 shows a first order sketch of an analog beamformer layout. This is over simplified, but enough to accomplish the modeling goal of predicting wiring capacitance from estimated footprint size. A BF unit cell must contain the phase shifter for a given phase resolution, the local transamplifier cell of the summation stage, the footprint for the wiring to get the baseband signal in and the user summation node out, and the local digital control and wiring. This is a non-trivial number of circuits to estimate, and the wiring is a significant amount of the area. The analog BF unit cell uses the layout for the ASIC from Chapter 5, to more accurately estimate these values. The ASIC used a phase shifter with 8.5 deg of resolution, so the model will likewise assume a phase shifters in analog and digital with that resolution.

To change this variable, the layout BF unit cell should be scaled to reflect the additional devices to increase as resolution.

The BF unit cell is also defined by an input and output capacitance. This is important as the wiring and device parasitics will define the input load that the stage before must drive which scales with K, as well as the output load that the summation stage must drive which scales with M.

Digital

The digital beamformer is modeled as the logic gates needed to perform the mathematical operations to implement a vector modulator and a summation of M elements. Operational blocks are defined by the number of NAND gates needed to build a typical implementation. A half adder is $C_{HA} = 5C_{NAND}$, a full adder is $C_{FA} = 9C_{NAND}$, a 4-bit carry look ahead is $C_{CLA} = 20C_{NAND}$, and a simple combinational multiplier is $C_{MP} = N \times BC_{NAND} + BC_{HA} + N(B-1)C_{FA} \approx N \times B10C_{NAND}$ for an N and B-bit input. Then the capacitive load of the digital operation is defined with the capacitive load of a NAND gate for the desired technology.

The vector modulator operation requires 2 multiplies and 1 addition for each weight element of the full matrix operation. The number of bits in the signal, B, is defined with the number of bits out the ADC, and increased to preserve resolution as the summation stage increases the SNR of the signal. The real (I) and imaginary (Q) part of the signal is multiplied by beamformer weights ($Re\{\phi_{M,K}\}$ or $Im\{\phi_{M,K}\}$). Beamformer weights are $N = 4$ bits in order to meet the 8.5 deg of resolution for the vector modulator. The vector modulator summation is performed by $\lfloor (B+N)/4 \rfloor$ instances of a 4-bit carry look ahead adders, rounded down with the $\lfloor x \rfloor$ operator, and a full adder handles the remainder, calculated with the modulo operation (%).

$$C_{AD,VM} = \left\lfloor \frac{B+N}{4} \right\rfloor C_{CLA} + (B+N)\%4\, C_{FA} \tag{3.29}$$

$$C_{MP,VM} = N \times B\ 10C_{NAND} \tag{3.30}$$

The summation is performed for K users, and requires $M-1$ addition operations. The number of bits is increased to keep all the SNR boost from the array gain, rounded up with the $\lceil x \rceil$ operator and represented by the function: $\text{bits}(M,K) = \left\lceil \frac{dB(M+K-1)-1.76}{6.02} \right\rceil$.

$$C_{AD,SUM} = \left\lfloor \frac{\text{bits}(M,K)+B+N}{4} \right\rfloor C_{CLA} + (\text{bits}(M,K)+B+N)\,\%4\, C_{FA} \tag{3.31}$$

The digital beamformer increases the number of bits, and does not attempt to decrease the SNR of the signal to save power. This is something that could be implemented, but it is left for future work and not implemented in the system model. Since the BF does not reduce resolution for power savings, the analog BF must implement an equivalent trade to maintain

the constant SNR at the output of both architectures. Thus the baseband signal chain SNR degredation, SD_{BB} for both analog and digital cases is held the same, even though the analog BF architecture has more stages contributing noise.

Then for a total collected capacitance:

$$C_D = K(M-1)C_{AD,SUM} + KM(2C_{MP,VM} + C_{AD,VM}) \tag{3.32}$$

The estimated power to perform the full beamformer matrix operation is:

$$P_{BF,D} = 0.25\, C_D\, V_{DVDD}^2 f_s + \frac{C_D}{C_{NAND}} i_{leak} \tag{3.33}$$

Where i_{leak} is the expected leakage for a NAND gate in the given technology, f_s is the sampling frequency of the ADC, and V_{DVDD} is the digital power supply, expected to be lower than the V_{DD} used for analog blocks.

This model is a dramatically optimistic estimate, as this does not include clock distribution, or the power of any other overhead circuits. This will generally favor the digital beamforming for power estimation, as the clock distribution is generally a significantly larger portion of a digital beamformer power budget, than say the equivalent current distribution or common-mode feed-back of an analog beamformer would be. However, including the clock tree would be a constant scalar along with other technology specific variables, and we are more interested in characterizing how the beamformer scales with M and K.

3.5 Noise in a Single Summation Stage

The noise factor is usually calculated for a single signal chain in non-MIMO applications. However, in the context of RF beamforming, the noise factor is calculated for the array. The array noise factor (F_{arr}) is obtained by multiplying the single element noise factor (F_{se}) by the number of array elements $F_{arr} = F_{se}M$. This separates noise figure from array gain[56], which is accounted for instead in the antenna gain [57], and avoids negative noise figure values. We expand this analysis into baseband beamforming, using a similar technique but for $dB^{-1}(\mathrm{SD}_{arr})$ instead of F_{arr}.

Let us start with a toy example, shown in Figure 3.5, a summation block has M coherently summed inputs of $\mathrm{SNR}_{in} = s_{in}/n_{in}$, active gain or passive loss modifier G_{sum}, and output referred added noise n_{sum}, such that $\mathrm{SNR}_{out} = M^2 G_{sum} s_{in}/(M G_{sum} n_{in} + n_{sum})$. It then follows:

$$dB^{-1}(\mathrm{SD}_{arr}) = \frac{s_{in}/n_{in}}{M^2 G_{sum} s_{in}/(M G_{sum} n_{in} + n_{sum})} M = 1 + \frac{n_{sum}}{M G_{sum} n_{in}} \tag{3.34}$$

This applies, as long as the signal is correlated and noise is uncorrelated between elements. At first, the array gain then appears to reduce the effect of noise further down in the chain. However, it is important to note that the single element gain G_{sum} of the summation stage is often a function of M as well.

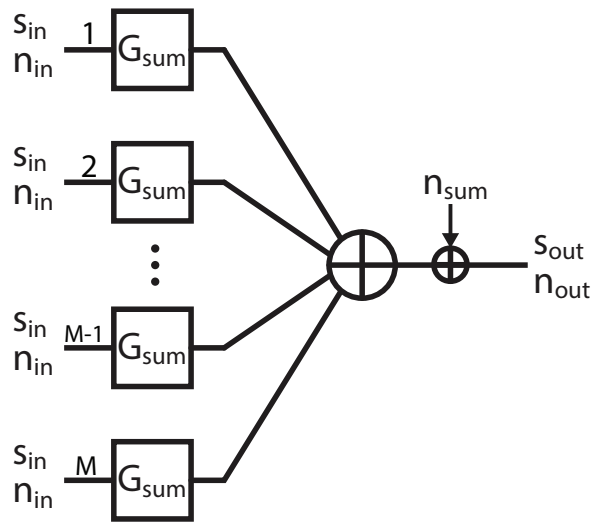

Figure 3.5: High level model of single stage of summation.

This conclusion relies on the assumption that n_{in} across all elements is uncorrelated, which is generally a good assumption in analog and RF blocks as long as good layout practices are implemented to keep cross coupling between elements low. This is a slightly more relevant consideration though at baseband where, for very large q, the ADC noise may be correlated for digital beamforming [50]. However, this is only relevant in very low resolution applications.

Passive Summation

Passive summation is most commonly used in RF beamforming, but is useful to examine for intuition. We will begin with an ideal passive summation stage with no loss, where $G_{sum} = 1/M$. It is then apparent that the array gain M provides no benefit or detriment to the noise of later stages. SNR degradation works simply as it would in a normal signal chain. More realistically, the passive summation stage has some loss, $G_{sum} = L_{sum}/M$, where $L_{sum} < 1$. The loss in effect makes the summation stage and following stages' noise contribute more to SNR degradation.

Single Stage, Active Summation

In the case of an active current summation stage, it is possible for the gain to be $1/M$ or larger. We will us a restively loaded current summation stage as an example, but this analysis can be performed for any other type of active summation.

We begin by defining active summation specific variables. The active voltage gain of the summation stage is defined as $A_V = g_m R_L$, with a single cell having gain $A_{V,sc} = g_m R_L/M$.

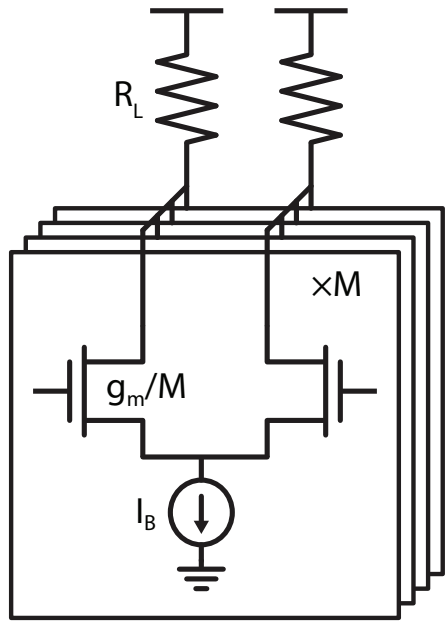

Figure 3.6: Generalized active current summation stage, summing M signals with active gain $A_V = g_m R_L$

This means that the signal and noise input to each single cell experiences power gain: $G_{sum} = A_{V,sc}^2 = A_V^2/M^2$. Looking at Equation 3.34, we see then that active summation achieves a $G_{sum} = 1/M$, the equivalent for ideal passive summation, when $A_V^2 = M$. The active gain can also have $A_V^2 > M$, where the n_{sum} term and noise from later stages benefits from the large gain. Then, the summation noise and following stages' noise contributes less to the signal chain SNR degradation. However, $A_V^2 \geq M$ can be a difficult gain requirement to hit for large M. Especially when the wiring parastics of the summation stage also scale with M. For a desired bandwidth and given technology, the gain-bandwidth product needed may not be achievable. Therefore, it is not unreasonable, for arrays with large M to consider a design where $A_V^2 < M$ or chosen independently of M. However, the added noise should then be carefully considered as the active stage will act like a passive summation stage with loss. Equation (3.34) can be re-written to emphasize the effect of "lossy" summation on later stages.

$$dB^{-1}(\mathrm{SD}_{arr}) = 1 + \frac{n_{sum}}{M\left(\frac{A_V}{M}\right)^2 n_{in}} \tag{3.35}$$

$$= 1 + \frac{M\, n_{sum}}{A_V^2 n_{in}} \tag{3.36}$$

The key here, is how stage gain is defined, where each input to a single cell experiences

a gain of $A_{V,sc} = g_m R_L/M = A_V/M$. The reason to choose to define gains this way is not arbitrary, but because this allows us to treat the entire summation stage as a single amplifier with gain $g_m R_L$. In addition, the noise of the stage, n_{sum} is defined simply with g_m and does not contain a factor of M. Then, as usual for a resistively loaded differential pair:

$$n_{sum} = 8kT(R_L + \gamma M(g_m/M)R_L^2) \tag{3.37}$$
$$= 8kT(R_L + \gamma g_m R_L^2) \tag{3.38}$$
$$= 8kTR_L(1 + \gamma A_V) \tag{3.39}$$

3.6 Noise in a Two Stage Summation

To ameliorate the noise or gain requirement in active summation, the summation stage can be broken into multiple stages.

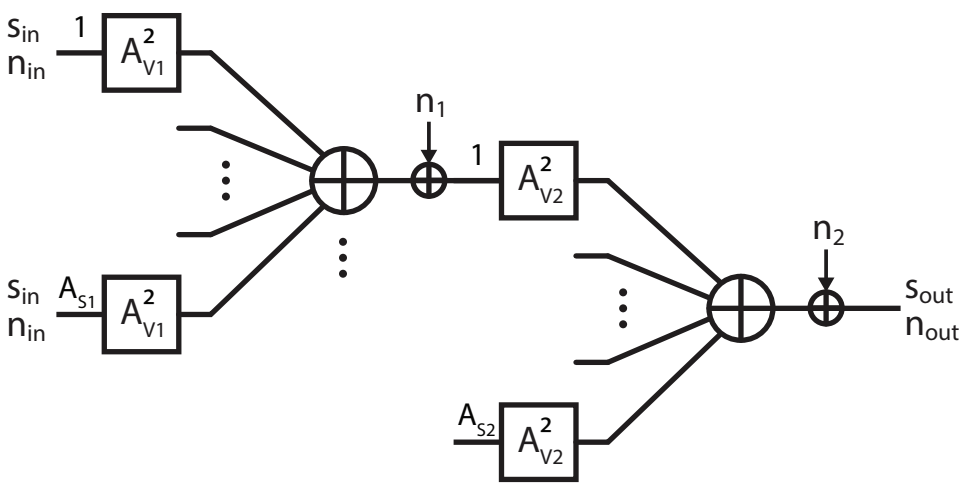

Figure 3.7: Circuit model of two stage active summation in baseband.

However, splitting into two stages is not without consequence. The SD of a two stage summation can be calculated similarly to a one stage. Let the active gain of each stage be A_{V1} and A_{V2} with the total voltage gain of the summation stage as $A_{VT} = A_{V1} \times A_{V2}$. The number of signals summed in each stage are A_{S1} and A_{S2}, with $M = A_{S1} \times A_{S2}$. The values of g_{m1}, g_{m2} and R_{L1}, R_{L2} are defined directly from the circuit in Fig. 3.6, with corresponding added noise n_1, n_2.

$$dB^{-1}(\mathrm{SD}_{arr,2stg}) = 1 + \frac{A_{S1}n_1}{A_{V1}^2 n_{in}} + \frac{A_{S1}A_{S2}n_2}{A_{V1}^2 A_{V2}^2 n_{in}} \tag{3.40}$$

We can examine Equations (3.40) and (3.35) side by side:

$$dB^{-1}(\mathrm{SD}_{arr,1stg}) = 1 + \frac{M\, n_{sum}}{A_V^2 n_{in}} \tag{3.41}$$

$$dB^{-1}(\mathrm{SD}_{arr,2stg}) = 1 + \frac{A_{S1} n_1}{A_{V1}^2 n_{in}} + \frac{M n_2}{A_{VT}^2 n_{in}} \tag{3.42}$$

It is apparent then that if n_1 and n_2 are chosen so that both stages consume the same power as a single stage with n_{sum}, then $dB^{-1}(\mathrm{SD}_{arr,2stg}) > dB^{-1}(\mathrm{SD}_{arr,1stg})$. Thus, in the case that the summation stage is noise limited, two stage summation costs more power. In a noise-limited design space, two-stage summation is then only preferred in cases where large M puts $A_V^2 \geq M$ out of the capability of the technology. However, this is not the same in a bandwidth limited case. Determining if the summation stage is noise limited or bandwidth limited is performed similarly to the amplifier, but with the additional consideration of array gain, as shown in the next section.

Active Summation, Noise Limited Design

We can estimate the theoretical power of a single summation stage that meets both gain, A_{VT}, and noise, $\mathrm{SD}_{arr,1stg}$, design specifications. We begin by re-writing Equation (3.39) using $A_{V,x}$ and then V_x^*.

$$n_x = 8kT\frac{A_{Vx}}{g_{mx}}(1 + \gamma A_{Vx}) \tag{3.43}$$

$$g_{mx} = \frac{8kTA_{Vx}(1 + \gamma A_{Vx})}{n_x} \tag{3.44}$$

$$I_{Dx} = \frac{4kTV_x^* A_{Vx}(1 + \gamma A_{Vx})}{n_x} \tag{3.45}$$

Substituting Equation (3.45) into Equation (3.35) and Equation (3.42), can be used to compare the power of one or two stage summation to meet a given SD. However, we will show only the single stage summation below.

$$dB^{-1}(\mathrm{SD}_{arr,1stg}) - 1 = \frac{M\, n_{sum}}{A_V^2 n_{in}} \tag{3.46}$$

$$n_{in} = \frac{M\, n_{sum}}{A_V^2(dB^{-1}(\mathrm{SD}_{arr,1stg}) - 1)} \tag{3.47}$$

$$\frac{4kTV_x^* A_V(1 + \gamma A_V)}{I_D} = \frac{M\, n_{sum}}{A_V^2(dB^{-1}(\mathrm{SD}_{arr,1stg}) - 1)} \tag{3.48}$$

$$I_D = \frac{4kTV^* A_V(1 + \gamma A_V)(dB^{-1}(\mathrm{SD}_{arr,1stg}) - 1)}{M\, n_{sum}} \tag{3.49}$$

The result, Equation (3.49) is proportional to the power of a resitively loaded differential stage in terms of design variables gain, noise, V^* and targeted SD. Similar to the amplifier of Section 3.3, the R_L required can then be solved for from I_D and A_V. If R_L is too large to meet the bandwidth requirements of the stage, then the stage is bandwidth limited instead.

Active Summation, Bandwidth Limited Design

The bandwidth limited power estimate is identical to the amplifier of Section 3.3, and is simply re-written here to save the reader the effort of scrolling.

$$I_D = f_p \pi V^* A_V C_L \tag{3.50}$$

If Equation (3.50) is greater than Equation (3.49) for the desired specifications, then the stage is bandwidth limited. Alternatively, if Equation (3.50) is less than Equation (3.49), then it is noise limited.

3.7 Signal Chain Optimization

The signal chain given in Fig. 3.2 must provide $A_{V,BB}$ total gain, for a signal with bandwidth f_{BW} and SNR degradation SD_{BB}. The model then optimizes the division of those specifications between the available stages, for power.

To meet the desired signal bandwidth, the system takes into account the number of poles in choosing the specified bandwidth for each stage. For example, a digital beamforming signal chain targeting f_{BW} has two poles, not including those in the RF chain. Thus, each stage should target a bandwidth of $1.5f_s$, so that the 3dB point is at f_{BW}. The analog beamforming signal chain, however, has 4 poles assuming the beamformer itself has a separate phase shifting stage and summation stage. Thus, each stage should target a bandwidth of $2.3f_s$, so that the 3dB point is held constant. This model will ignore the difference in filter order, between the two architectures.

For both cases, the $A_{V,BB}$ is the gain needed to amplify the signal at the input of the baseband stage to the full swing of the ADC, as found in Equation (3.6) and (3.7). However, the ratio of gain $A_{V,A1}$ in amplifier stage 1 and $A_{V,A2}$ in amplifier stage 2 is a variable that can be optimized over for a desired noise and parasitic load on each stage's output. The analog beamforming case also includes the gain of the beamformer itself, $A_{V,BF}$, which includes the active phase shifter and active current summation. As discussed in Section 3.5, the summation stage should have $A_{V,sum} > \sqrt{M}$ to improve noise contribution of later stages. However, the beamformer is the largest block in the chain, and scales with $M \times K$, thus there is significant power savings to be had by shifting difficult specifications out of the beamformer and onto the surrounding amplifiers, thus decreasing the overall power of the signal chain. The optimum depends on whether the beamformer block is more power limited by noise or bandwidth. For example, a bandwidth limited beamformer is spending power to drive the large wiring load and so has low noise contribution, which leaves more

for later stages, which then reduces the need for high gain in the beamformer. To explore these different design spaces, the gain of the beamformer is swept over $\sqrt{M}$, 1, and $1/\sqrt{M}$. These steps in design space may create small discontinuities in the results, but provide a good enough range for the various array size and bandwidth options.

The next goal of the model's power optimization is to meet the SNR degradation specification. We begin by writing out the equation for the baseband noise in both the analog and digital cases.

Analog Beamforming:

$$dB^{-1}(\mathrm{SD}_{BB}) = \frac{A_{V,A1,A}^2 M \left(\frac{A_{V,BF,A}}{M}\right)^2 A_{V,A2,A}^2 n_{in} + M \left(\frac{A_{V,BF,A}}{M}\right)^2 A_{V,A2,A}^2 n_{A1,A} + A_{V,2,A}^2 n_{BF,A} + n_{A2,A}}{A_{V,1,A}^2 M^2 \left(\frac{A_{V,BF,A}}{M}\right)^2 A_{V,A2,A}^2 n_{in}} M \tag{3.51}$$

Digital Beamforming:

$$dB^{-1}(\mathrm{SD}_{BB}) = \frac{A_{V,A1,D}^2 A_{V,A2,D}^2 n_{in} + A_{V,A2,D}^2 n_{A1,D} + n_{A2,D}}{A_{V,A1,D}^2 A_{V,A2,D}^2 n_{in}} \tag{3.52}$$

We will begin by assuming that all of the stages are bandwidth limited. With this assumption, the noise of each stage can be calculated from the defined values $C_{L,x}$ and bandwidth, and the yet unspecified gain of each stage. Then, the remainder of the noise is budgeted for the ADC driver stage and sampling capacitor as discussed in Section 3.3. The noise of the stage is written as below in terms of the gains and noise of previous stages.

Analog Beamforming:

$$\begin{aligned} n_{A2,A} = dB^{-1}(\mathrm{SD}_{BB} - 1) A_{V,1,A}^2 \left(\frac{A_{V,BF,A}^2}{M}\right) A_{V,2,A}^2 n_{in} - \\ \left(\frac{A_{V,BF,A}^2}{M}\right) A_{V,2,A}^2 n_{A1} - A_{V,2,A}^2 n_{BF,A} - n_{A2,A} \end{aligned} \tag{3.53}$$

Digital Beamforming:

$$n_{A2,D} = A_{V,1,D}^2 A_{V,2,D}^2 n_{in}^2 (\mathrm{SD}_{BB} - 1) - A_{V,2}^2 v_{n,1}^2 - v_{n,2}^2 \tag{3.54}$$

The gain of each stage is still unspecified, but the total gain $A_{V,BB}$ is given. To constrain the system of equations, we define the ratio of $A_{V,1}^2 : A_{V,2}^2$, and sweep this ratio, and find the minimum power. If $n_{A2,D}$ is negative and there is no noise budget remaining, then the signal chain is noise limited, and more gain needs to be pushed to the beginning of the signal chain

to reduce the noise contribution of later stages. This may occur if the total gain is too low, the SD_{BB} specification is too low, or the bandwidth and array size are very small.

The model presented in this chapter is a deeply hardware informed estimation of a baseband signal chain performing analog or digital beamforming. It allows us to predict how the signal chain will respond to changes in array size, signal bandwidth, and changes in technology specific parameters. In exploring the model, we are able to identify useful design mindsets for creating a large array, such as determining whether a block is noise or bandwidth limited, and how gain inside of the beamformer effects requirements of stages before and after. The model does not reflect all parts of the circuit design. For example, the model does not consider two stage summation, even though there is a power advantage to splitting stages that must supply large gain. This is left off due to the complexity this creates in the design space, and does mean that the beamforming power model is pessimistic in the case of low bandwidth and large K, and large M for analog baseband beamforming. However, the model is optimistic in that it leaves off house keeping circuits like clock trees and current distribution. This allows us to focus more on how power scales to a first order with M and K, and a few important design variables such as input SNR, bandwidth, and noise. The main pieces of intuition and how the optimization work were shown here, the results will be discussed next chapter, so that they are viewed in the context of beamformer power in relation to ADC power.

Chapter 4

Analog and Digital Beamforming Comparison

Having formed a model of the blocks and specifications that are expected to change between analog and digital beamforming architectures, we would now like to examine the results for a particular technology and design.

The code which implements this model and reports these results can be found in: `https://github.com/enaviasky/bf_comparison`.

Variables

We begin with an examination of the relevant design variables, to be clear about what is a technology or design dependent variable which could be changed, and what is a swept variable which gives us an idea of how array power scales. The table will also indicate the values of constants and the range of variables used to generate the results, with values for an approximately 28nm process.

The variables which are swept are array size, which is specified in the number of users K, and the number of antennas per users α, from which the number of antennas M is calculated. Also swept is bandwidth and interferer power. Power is then calculated for the ADC, and optimized in the signal chain and beamformers to meet the given specifications.

4.1 Results of Beamforming Comparison

Recall that all of the specifications in Table 4.1 and Table 4.2 are held constant between the two architectures to ensure that the throughput of both architectures is equal and thus a comparison in only the dimension of power is valid.

SNR_{ant}	10 dB
NF_{LNA}	9 dB
G_{LNA}	20 dB
Phase Shifter Resolution	8.5 deg
NF_{BB}	0.4 dB
γ factor	1.5
V^* Before Beamforming	0.2 V
V^* After Beamforming	0.25 V
Unit C_L to BF Input	28fF V
Unit C_L to BF Output	56fF V
NF_{ADC}	0.4 dB
ADC V_{swing}	0.8 V
ADC Oversampling Factor	2
ADC Walden FOM_W	50 fJ/conv-step
ADC Walden FOM_W Corner	440 MHz
ADC Schreier FOM_S	170 dB
ADC Schreier FOM_S Corner	30MHz
DVDD	0.8 V
C_{NAND}	2 fF
i_{leak}	20 nA

Table 4.1: Design Constants In System Model

Users (K)	2-40
α (M/K)	1-50
Signal Bandwidth	[.01, .1, 1, 10] GHz
Signal to Interference Ratio	[0, -10, -30, -40] dB

Table 4.2: Design Variables In System Model

10MHz Signal Bandwidth and Signal Dominated Environment

To begin we examine large arrays, with many users, no interferers, and signal bandwidth < 10MHz. The first result of interest is the power comparison of the baseband amplifiers and beamformer. As discussed, in Chapter 3, the amplifier is included as, especially in the analog beamforming case, the baseband amplifier chain is deeply intertwined with the design of the beamformer.

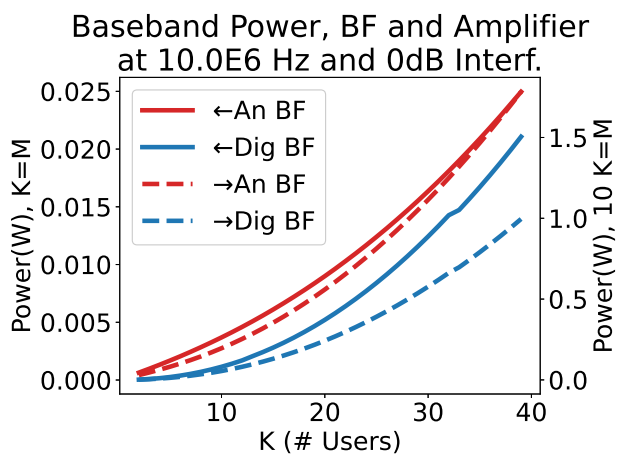

Figure 4.1: Baseband amplifier and beamformer power at 10MHz in a signal dominated environment. Solid line: $\alpha = 1$ with power on left axis. Dotted line: $\alpha = 10$, with power on right axis.

Figure 4.1 plots the power of baseband power consumption excluding the ADC for analog beamforming in red, and digital beamforming in blue. The solid line is for $\alpha = 1$ with power on the left axis. The dotted line shows the model's power estimate for $\alpha = 10$, with power plotted on the right axis. The result shows that for low signal bandwidth of 10MHz and an environment with no significant interferers, the digital beamformer is roughly lower power. This of course depends on many technology dependent settings for the constants, and does not include housekeeping circuits for analog and digital circuits. This is true at 10MHz for the range of K examined. We note that while the digital beamformer option displays the K^2 curve that we expect for a fixed α and a beamformer that scales with $M \times K$. However, the analog beamforming power, especially for $\alpha = 1$, is slightly shallower than K^2, which comes from being able to offload some of the beamformer gain and noise requirements onto the amplifiers. The small discontinuity in the digital beamforming result comes from the minimum step size in the power optimization.

A wider range of α can be shown for the baseband amplifier and beamformer power for the analog beamforming case utilizing a contour graph to see the difference in power. Figure 4.2 plots the difference in power for the baseband excluding the ADC over the entire swept

K and α space. The contour is specifically $P_{BB,A} - P_{BB,D}$, so positive (red) values indicate that $P_{BB,D}$ is lower, and negative (blue) values indicate that $P_{BB,A}$ is lower. Figure 4.2 shows that the modeled digital beamformer baseband power for the given design variables is lower for all ranges of array size at 10MHz in a signal dominant environment.

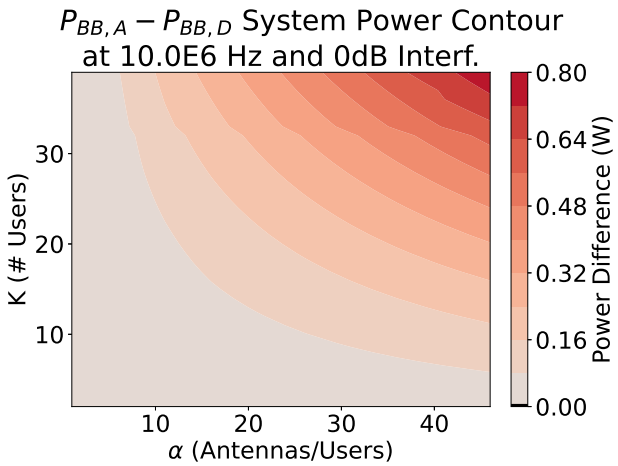

Figure 4.2: Baseband amplifier and beamformer power for analog BF and digital BF difference at 10MHz in a signal dominated environment

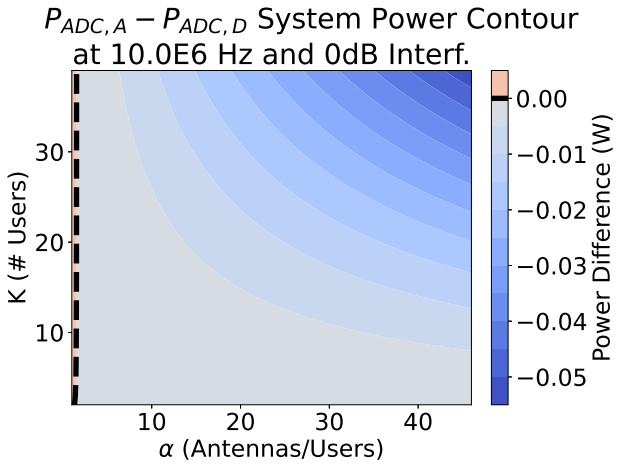

Figure 4.3: ADC power difference at 10MHz in a signal dominated environment

To now compare the baseband and beamformer power to the ADC power, Figure 4.3 plots the difference in ADC power over the same K and α space. The contour is specifically

$P_{ADC,A} - P_{ADC,D}$, so positive (red) values indicate that $P_{ADC,D}$ is lower, and negative (blue) values indicate that $P_{ADC,A}$ is lower. The black dashed line is where $P_{ADC,A} = P_{ADC,D}$. Then, Figure 4.3 shows that ADC power is lower at all but $\alpha = 1$, as expected from 2.26.

It is rapidly visible from these graphs that the difference in beamfomer power is always larger than the difference in ADC power between digital and analog beamforming architectures. This is seen at a glance in total baseband power contour, combining amplifier, beamformer, and ADC in Figure 4.4.

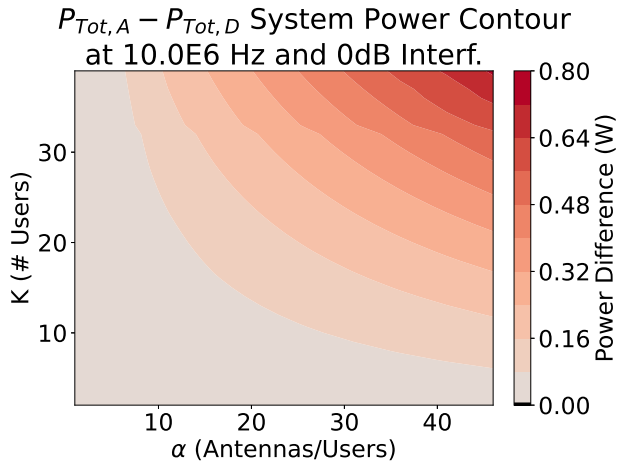

Figure 4.4: Total power difference at 10MHz in a signal dominated environment

Again, the red is area where the digital beamformer architecture is preferred as it delivers the same spatially filtered and SNR output signal for less power. Area in blue, which is not present in this contour, is area that analog beamforming is preferred, and the black dotted line is where the power of the two architectures is equal. Indeed, for these specifications and technology variables, the analog beamforming option is always more power expensive than the digital beamformer. However, the expected power savings are less than half a watt, even at very large array sizes. From this result, we see that in this case of low bandwidth and no interference, the selection of beamformer location is dependent on designer preference and concerns other than the baseband circuit design.

10MHz Signal Bandwidth and Interference Dominated Environment

Maintaining the same bandwidth as before, we now introduce dominant out of network interferers.

The plots in Figure 4.5 show that as the interferer grows larger, the analog beamformer baseband signal chain power increases more slowly than the digital beamforming case, as it

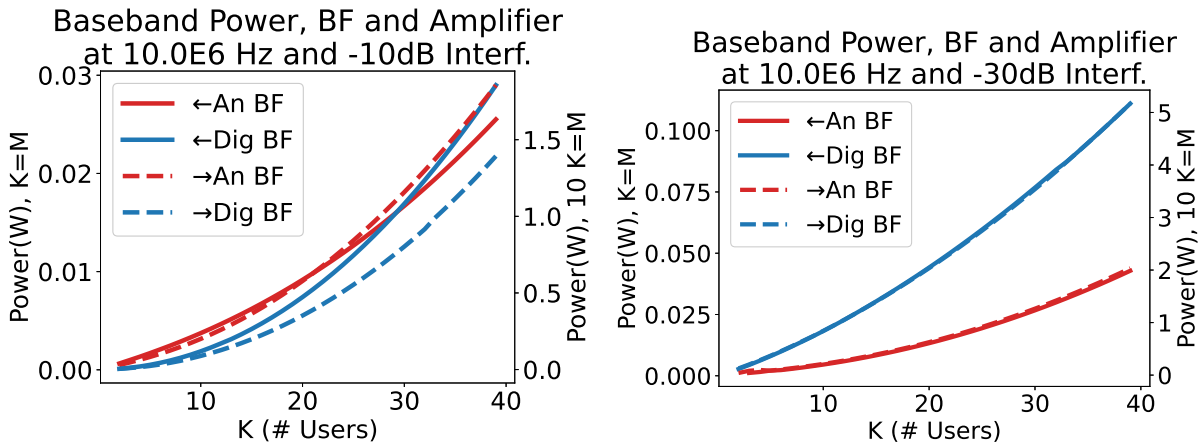

Figure 4.5: Baseband amplifier and beamformer power at 10MHz in a interferer dominated environment

is more power intensive to increase the number of bits in the digital beamformer, than in the analog beamformer, which can offset some of the power of noise reduction by increasing the gain of the first amplifier stage.

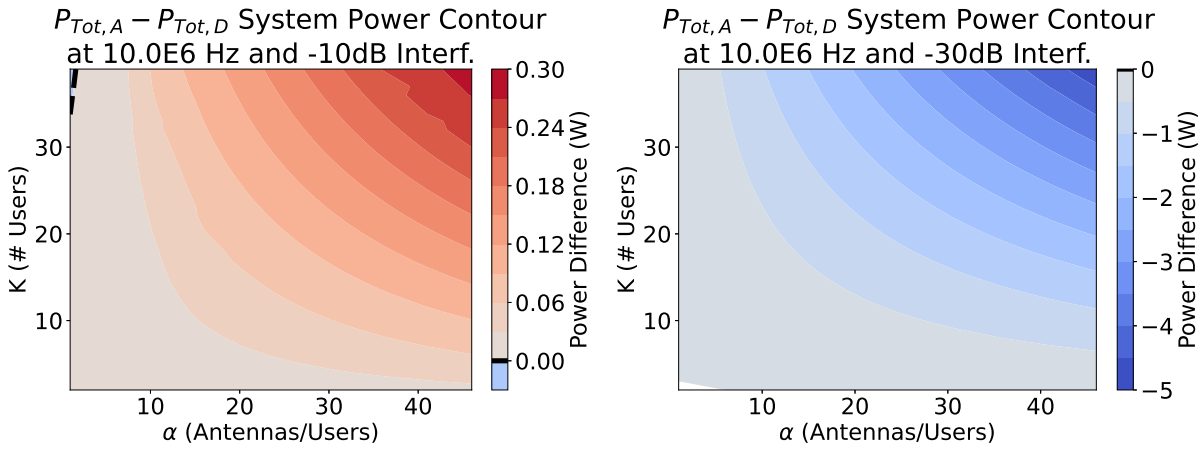

Figure 4.6: Total power difference at 10MHz in an interference dominated environment

As the digital beamformer power begins to grow larger than the analog beamformer due to the interferer, the analog beamforming architecture becomes more power efficient. However, the difference in power is still negligible and likely more dependent on other factors.

One more point of interest, is the results of pushing the SIR to -40dB. In Figure 4.7, the interferer is large enough that the ADC model has switched from using the Walden FOM to

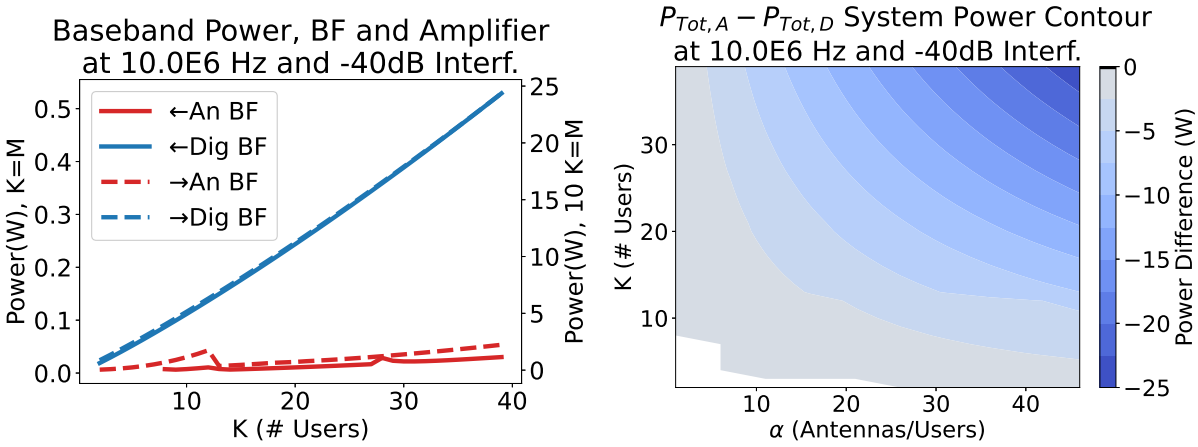

Figure 4.7: Power difference at 10MHz in a deeply interference dominated environment

the Schreier FOM, as the Schreier FOM is a better model at high resolution. This results in the discontinuities in the amplifier and beamfomer power. This is a result of pushing the model to the edge of functionality, since the crossover between Walden and Schreier FOM is not an exact point as it is modeled. Fortunately, large arrays generally seek to keep ADC resolution low, preferring to handle interferers with frequency or spatial filtering, or accepting lower noise performance and increasing SD_{ADC}.

Wide Bandwidth and Signal Dominated Environment

Moving to mmWave and beyond is done to take advantage of the much wider bandwidth available at these carrier frequencies. Thus, the model next explores the results of increasing the signal bandwidth to 100MHz and 1GHz. This is done first without any interferers.

Figure 4.8 shows that as the bandwidth increases, the baseband power of analog beamforming is closer to the digital beamforming baseband power, and by 1GHz the preferred architecture for baseband power has switched. The total power contours of figure 4.9 show digital beamforming is still power optimal at 100MHz, but at 1GHz analog beamforming is more efficient for all but $\alpha = 1$. However, looking only at the total baseband power masks a very important aspect of the architecture's power budget.

The ADC power budget is growing very rapidly between the two examples due to the corner in ADC FOM at bandwidths above 440MHz. Figure 4.10 shows for a single point on the contour of $\alpha = 2$ and $K = 10$ that the ADC begins to dominate the power budget as the bandwidth increases.

We can examine the ADC power in relation to the baseband amplifiers over a wider range of K. Figure 4.11 shows the baseband amplifier and beamformer power normalized to an ADC sized for the associated architecture. Normalizing the power allows us to get an idea

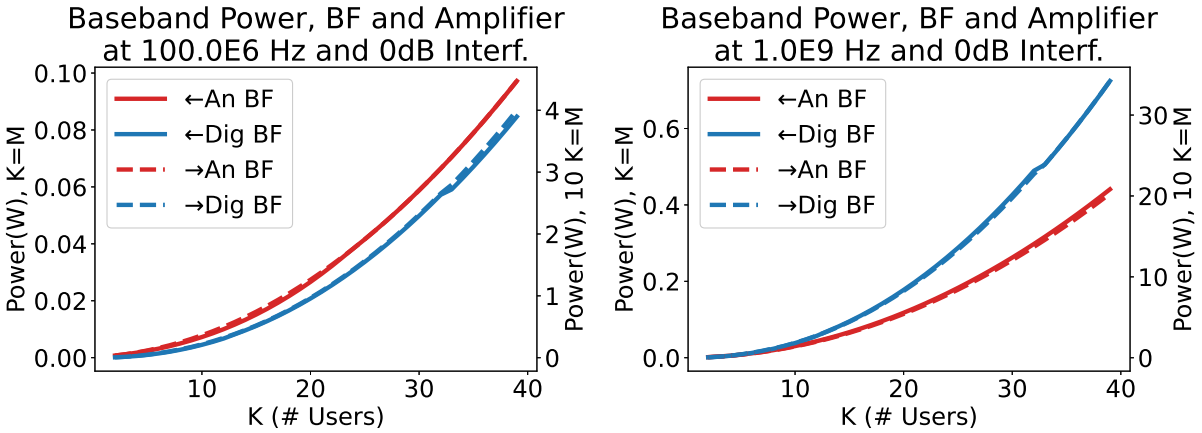

Figure 4.8: Baseband amplifier and beamformer power at 100MHz and 1GHz in a signal dominated environment

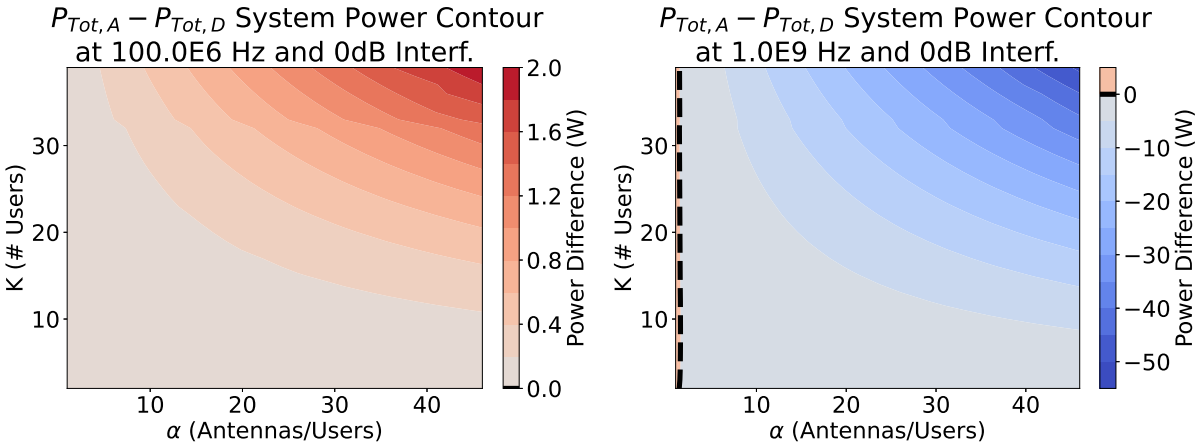

Figure 4.9: Power difference at 1GHz in a signal dominated environment

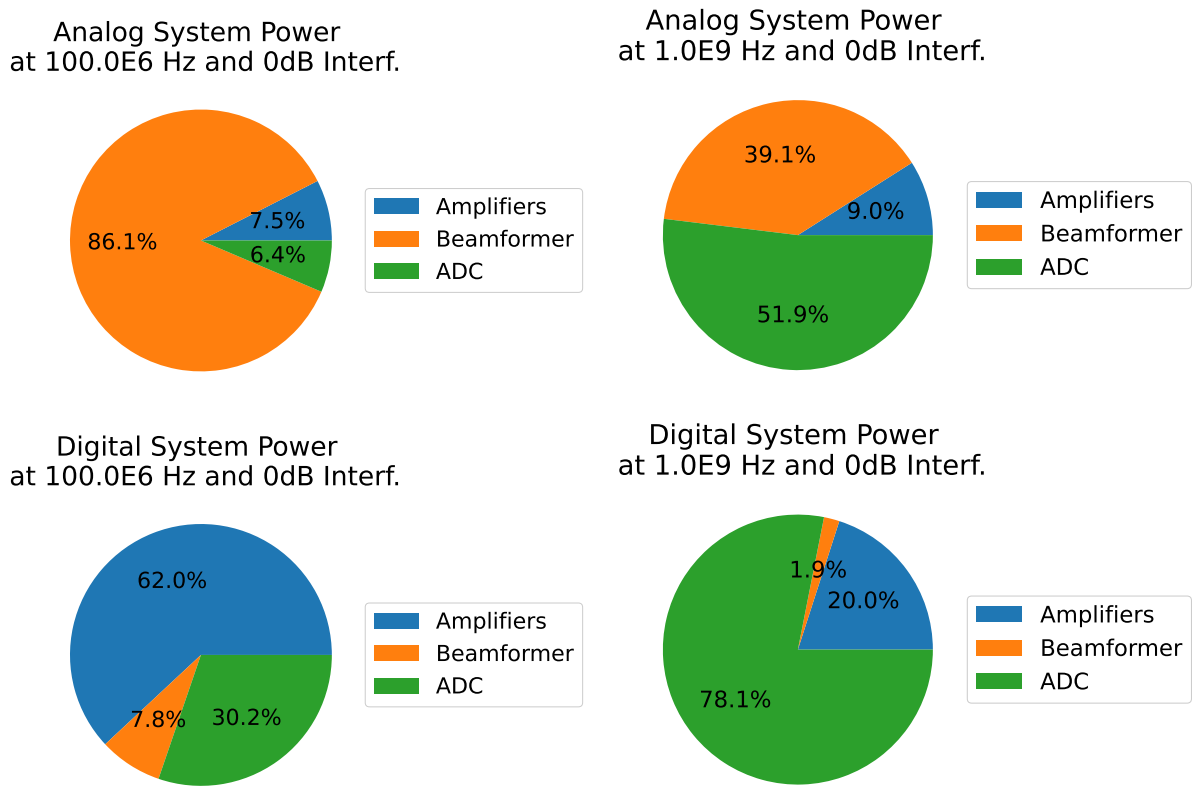

Figure 4.10: Power breakdown in baseband model for $\alpha = 2$, $K = 10$.

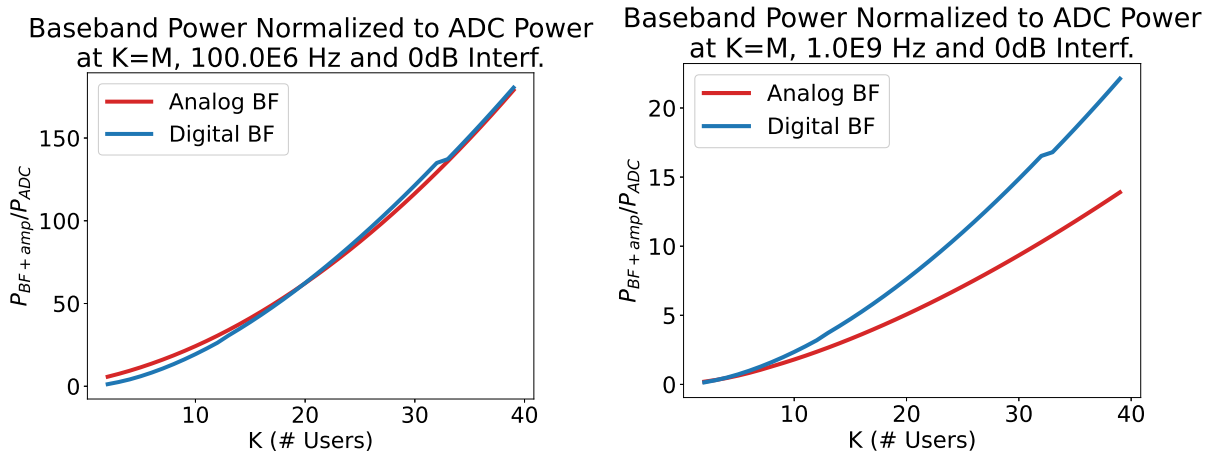

Figure 4.11: Power difference at 1GHz in a signal dominated environment

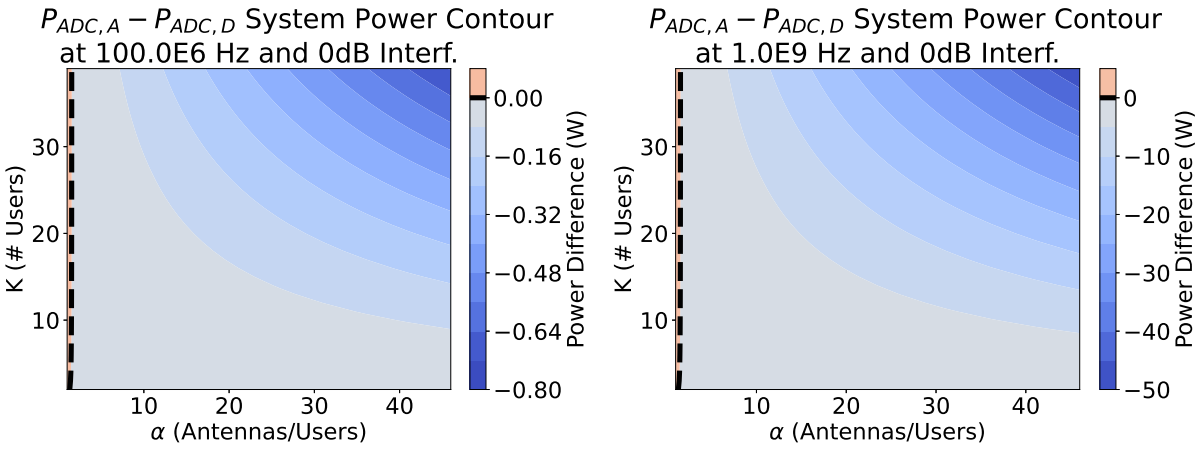

Figure 4.12: Difference in ADC power budget for a 100MHz and 1GHz bandwidth

of whether the beamformer or ADC dominates the power budget, and the importance of alpha in the case the ADC dominates. For example, examining the 100MHz case: for large numbers of users, the difference between digital and analog beamforming is 100s of ADCs. This means that α can be quite large before the increase in the number of ADCs incurred from a switch from analog to digital beamforming has a significant impact on the power. However, for 1GHz case, where the ADC is becoming a significant portion of the power budget, the switch from analog to digital not only increases the power of the beamformer and amplifier stages, but also the difference between K or M ADCs becomes a larger concern. For example, at 1GHz and 20 users, Figure 4.11 shows that the difference in baseband power can be measured in about 4 ADCs. This translates to an $\alpha > 4$, indicating for more Massive MIMO applications at these wide bandwidths, digital beamforming can expect to spend significantly more power or else must deliver a lower SNR signal. Intuitively, this can also be seen in how similar the 1GHz total power contour looks to the ADC only power contour.

Wide Bandwidth and Interference Dominated Environment

The presence of large interferes increases the number of bits required in the ADC, making the ADC more power expensive. As one might expect then, since a larger ADC power budget increases the preference for analog beamforming, a system which expects interferers prefers analog beamforming for even 100MHz.

Very Wide Bandwidth and Signal Dominated Environment

It is of some interest to push the model even further to make sure that the trend remains. Thus, we repeat for 10GHz of signal bandwidth.

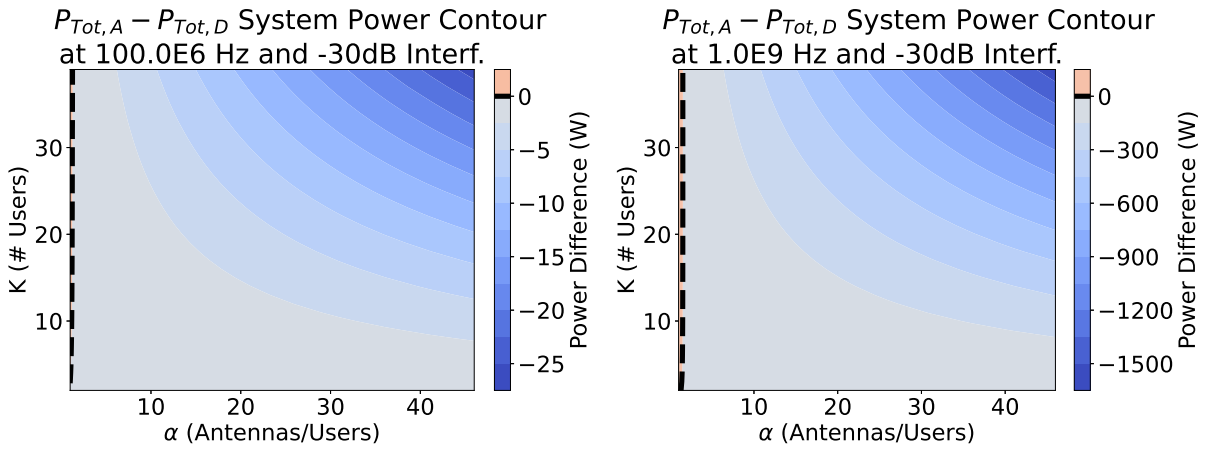

Figure 4.13: Difference in total power budget for a 100MHz and 1GHz bandwidth in an interference dominated environment.

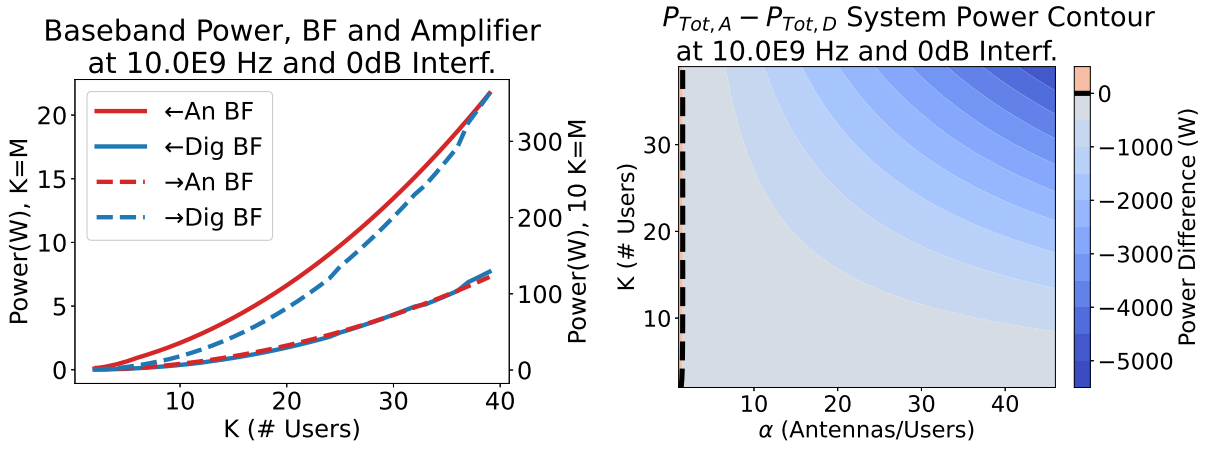

Figure 4.14: Power difference at 10GHz in a signal dominated environment

This shows that despite the extremely large approximations of baseband amplifier and beamforming power, the ADC power is growing more rapidly. As exemplified by the total power comparison contour, which looks the same as an ADC power comparison contour.

4.2 Power Savings from Interference Filtering

As the previous results illustrated, the ADC power consumption is a major concern as signal bandwidth increases. This is also true as the expected SIR increases. Figure 4.15 shows the power of a single ADC as the maximum allowable SIR is increased, assuming a 1dB SNR desired signal and 50fJ/conv-step Walden FOM. The ADC is considered outside of the array context for a single antenna. (Although, section 2.5 details the effects of array gain on ADC resolution in the context of interferers, and shows resolution increases directly with SIR before and after beamforming.)

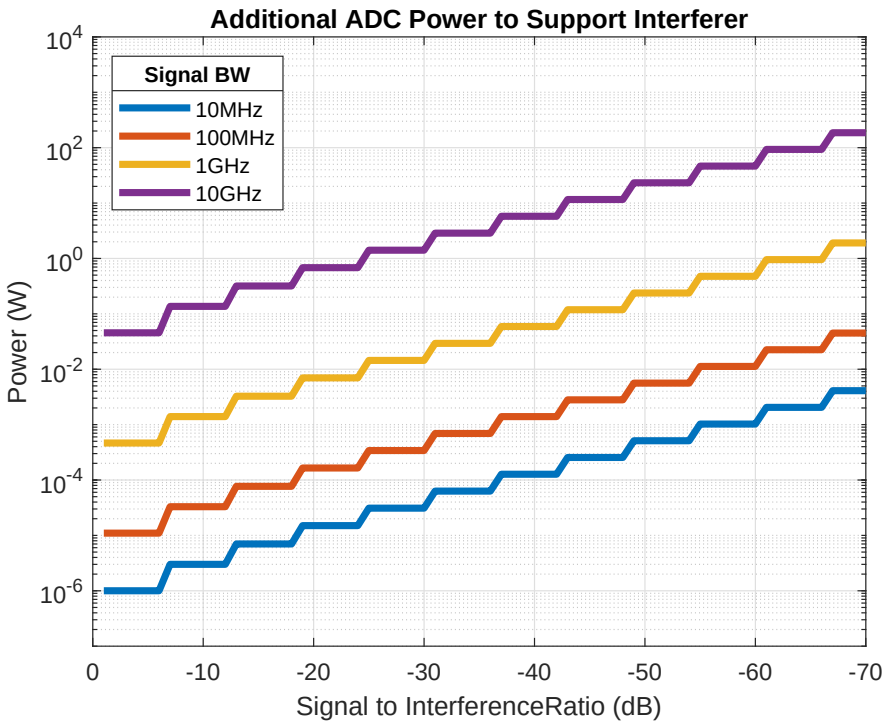

Figure 4.15: Power of a single ADC needed to support various SIR and signal bandwidth

Comparing the 0dB SIR case to the expected SIR, shows the power savings to be had from filtering of interferers. For 10MHz, the power lost to accommodating even a -50dB SIR interferer is negligible, less than a mW. However, as the bandwidth increases, it becomes clear that the power lost to over-design of ADC resolution or to accommodating interferers

becomes a cause for significant concern to the system designer, potentially more than a Watt for a single ADC at 10GHz. To be beneficial, then, a filter need only be lower power than the power that would be spent on the ADC. At wide bandwidths, this means there is a large power budget for filters, as filtering interferers before the ADC is a very important task to keep array power low. This suggests that a digital beamformer would benefit from not only frequency filtering, but from a spatial filter such as [58] or [59] which achieve significant suppression of out of beam interferers without reducing spatial resolution in a later stage of beamforming.

4.3 Results Summary

The model results shown here map the cases in which it is power optimal for a system to choose analog or digital beamforming. The specifications of the signal chain are held constant from the start of the baseband blocks to the digital output of K beamformed outputs from M antennas, so that the power comparison is fair.

The model finds that for low signal bandwidth, cases below the ADC sampling frequency corner (440MHz), the ADC does not take a significant share of the system power. Therefore, at low bandwidths the placement of a baseband beamformer in the signal chain depends on designer preference and other system factors.

However, factors that increase the power of the ADC, such as wide signal bandwidth, or large interferers quickly push preference towards baseband analog beamforming, as the number of ADCs grows faster than the resolution of the ADCs. Thus, as beamforming arrays reach towards higher frequency spectrum bands, if arrays wish to utilize the full GHz of signal bandwidth available, the need increases for baseband analog beamforming matrices. The power savings from an analog baseband array can be the difference between a build-able or completely infeasible array.

There are many challenges, even with these insights. For example, the actual implementation of a many user, many antenna beamforming array in baseband analog. The observations from this model are simply arm-chair theorizing without showing that a fully connected beamforming array serving these large array sizes can be feasibly built for the wide bandwidths targeted. Therefore, the next chapters will detail a silicon implemented array serving 16 users with 16 antennas with analog baseband beamforming, and demonstrate its use in a multi-user wireless setup.

Chapter 5

Many-User, Massive MIMO Receiver ASIC

By modeling the baseband blocks of a receiver array in the previous chapter, we have identified the power savings to be had from performing the beamforming operation in analog baseband, especially in the case of wide bandwidth signals or interferer dominated environments. However, published beamforming arrays at mmWave carrier frequencies, which are capable of supporting wide bandwidths, demonstrate limited user count. In [24], two outputs are extracted from an 8-element array. In [35], up to 4 concurrent users are supported with a 4-element array; although, the baseband bandwidth demands of the frequency-multiplexed user streams prevent the architecture from scaling to larger K. A 4-element, RF beamforming array achieved 2 beams, but using dual polarization in the antennas, which does not scale to more beams, even though it achieved very wide bandwidth[33]. The dearth of many-user mmWave sub-arrays suggests the need to actually implement a large, many-user array with analog BF. Putting a many-user analog beamformer on silicon, confirms that the theory can make it out of the arm-chair and into practice, where non-idealities may stress the model.

This chapter will test the design techniques developed in the baseband amplifier and analog beamformer model to scale a beamformer to large K. The baseband analog beamformer performs a fully connected matrix operation, serving 16 unique user beams, and is designed to achieve low power operation with tight footprint. The beamformer is a part of a full chip which receives an array of 16 antennas and outputs 16 unique user beams, as shown in 5.1. Off-chip, each of the K users are digitized and aggregated across each of N sub-arrays, as per the two-stage beamforming architecture. In this way, interconnect complexity is decreased while the fully-connected nature of the array is preserved. The ASIC is intended to be a sub-array of a much larger array, totalling 128 antennas. The two-stage architecture is chosen to take advantage of the scalability and additional inter-user interference mitigation in the back-end.

The ASIC discussed in this thesis will perform conjugate beamforming, as the first stage of local beamforming. The second stage of beamforming, zero forcing, needs high resolution, and so is best performed in digital, after the power efficiency gained from decreased ADC

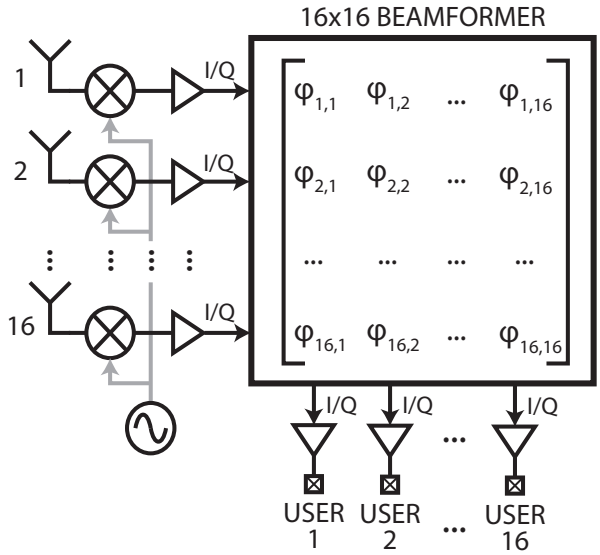

Figure 5.1: Overview of the proposed 16-output RX sub-array ASIC.

count has occurred. The ASIC targets a maximum of 10dB SNR at the antenna and up to 16 users with 1GHz single-sided bandwidth each. While SNR is boosted by combining additional subarrays, the array seeks to attain high capacity through wide-band and many user operation rather than pursuing higher order constellations. The ASIC while showing the ability to perform a large matrix multiplication, unfortunately, is unable to demonstrate ADC reduction from $\alpha > 1$. A 32 antennas subarray would be optimal for phase noise[25] in addition to power efficiency, however, packaging and available area have limited the chip to 16 antennas. The ASIC still serves as a proof of concept of many massive MIMO design concepts, such as trading off NF in the front-end for wide-band operation and power savings, scaling a analog baseband beamformer to many users, and demonstrating multi-user at mm-Wave.

5.1 Beamformer Top Level

The baseband beamformer performs a 16×16 matrix multiplication operation, as shown in Fig.5.2, transforming 16 antenna signals into 16 beams using complex weights. The beamformer is composed of three stages: 1) the distribution stage, 2) the phase shifters (PS), and 3) the summation stage. To implement a fully connected baseband analog beamformer for this application, three design challenges are of particular concern. First, keeping system power low, as the beamformer is fully connected and the number of operations scales with $M \times K$. This is of less concern in a smaller array or a single sub-array, but in a massive tiled array, the power of the matrix operation can rapidly become untenable if each stage is not

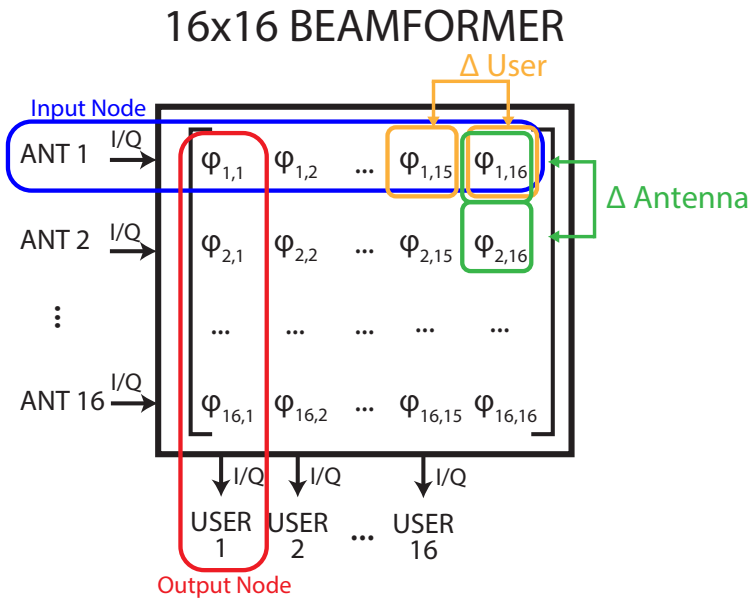

Figure 5.2: Matrix illustration of analog BF operation. Blue square is the input node, and red is the output node. The delay Δ User between the weights in yellow are the user mismatch, and Δ Antenna between weights in green are the antenna mismatch.

carefully optimized. Second, the design should prioritize reducing the beamformer's cross array routing parasitics. The beamformer input node, blue in Fig.5.2, and output node, red, must route signals across the entire array. This can result in significant loads, especially when one considers that the beamformer occupies a 0.7×2 mm^2 area, which accommodates the 512 phase shifters, the cross beamformer wiring, and the summation stages. The third and last design concern is the signal delay between matrix elements. As illustrated in Fig. 5.2, there are two types of possible delay mismatch: delay between different antennas delivered to the same user, in green, and the delay between different users receiving the same antenna signal, in yellow. Antenna mismatch results in malformed beams and can only be corrected with the intensive calibration of each of 512 phase shifters within the beamformer. On the other hand, user mismatch results in delay and bandwidth mismatch between users, but this can be corrected with back-end timing recovery and equalization, performed on only 16 users. Thus, we choose to sacrifice user delay matching in favor of antenna delay matching wherever applicable.

5.2 Cross Beamformer Signal Distribution

The signal distribution stage, Fig.5.3, provides high variable gain (13dB/20dB) with two stages of VGAs, and drives the BF input load with a buffer stage. The high gain in the

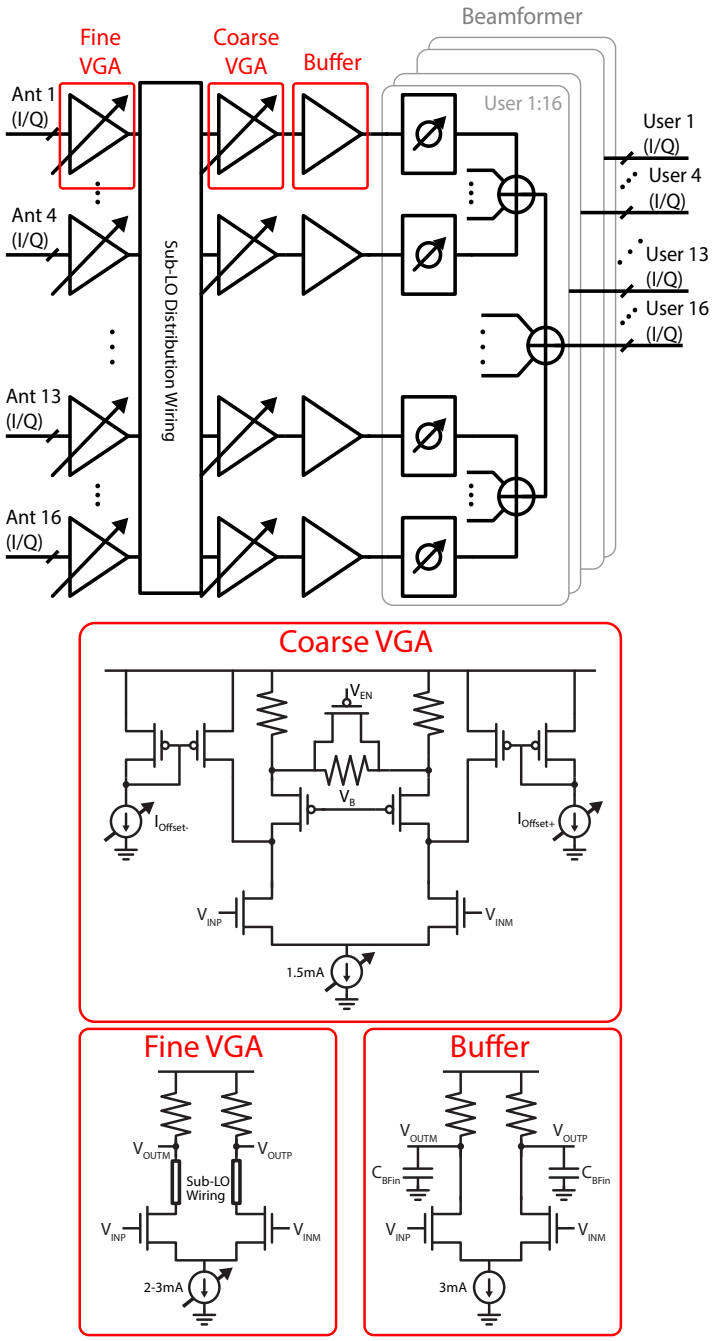

Figure 5.3: Schematic of the distribution chain driving a high-level schematic of the BF.

distribution stage serves the usual noise purposes in a signal chain i.e, much like an LNA, gain early in the RX chain reduces total system power by relaxing noise requirements on

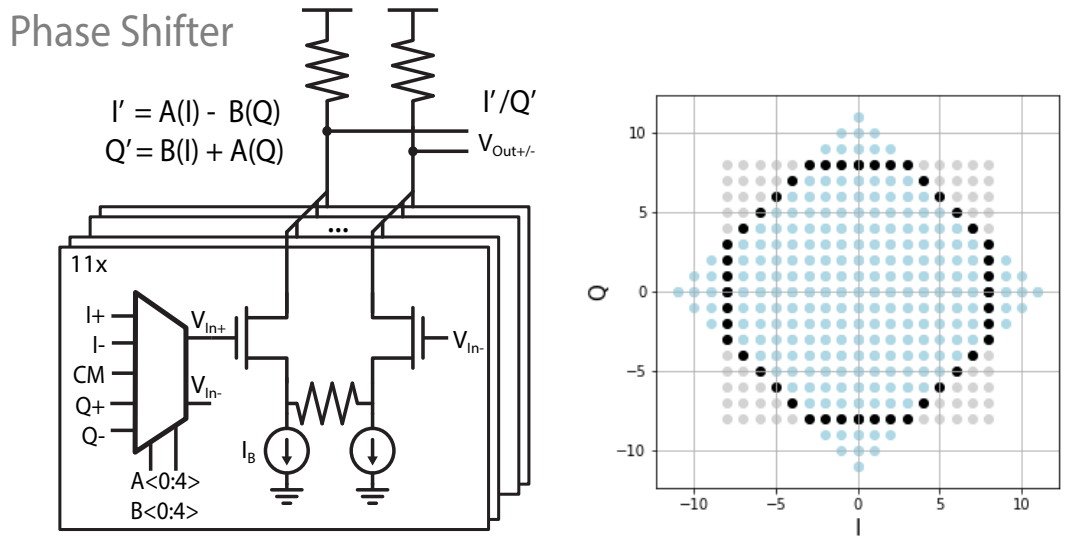

Figure 5.4: Circuit detail of the vector modulator used as the phase shifter. The 11 shared g_m cells map to the blue diamond of reachable points.

subsequent stages. The fully connected nature of the BF, however, lends this technique additional efficiency. The key concept to note here is that the power of a distribution stage scales with M, so 32 I/Q signals for 16 antennas; whereas the BF scales with $M \times K$, so the power of a single phase shifter is multiplied by 512. There is an additional factor of K in the power efficiency of reducing the noise requirements of the phase shifters. Thus, the phase shifter should be designed first to achieve very low power, and then the gain of the distribution stage is set to place the input noise floor well above the PS's noise contribution.

The first stage of the distribution stage, the fine VGA of Fig. 5.3, is a resistively loaded differential pair which amplifies the signal by a simulated 5.5-8.5dB, provides a simulated 3dB of fine amplitude control through tunable tail current for I/Q amplitude adjustment, and transports antenna signals beneath the LO distribution. The second VGA, provides a coarse tradeoff between amplitude (10dB / 14dB) and BW, to provide some noise filtering depending on the expected signal BW (250Msym/s or 1Gsym/s). In addition, the coarse VGA uses current mirrors in parallel with the cascode devices to correct voltage offset that has propagated from the two high gain VGA stages, protecting the linearity of the beamformer. The buffer stage is a low output impedance, resistively-loaded differential pair, which drives the large BF input load composed of cross BF wiring parasitics and 16 I and Q phase shifters inputs. In this way the distribution stage accomplishes the first two design goals of low power design and driving input parasitics.

Design for ADC: Noise, Gain, and Linearity

The signal is expected to arrive as a maximum of -58dBm signal at the input (-55dBm at the antennas with 3dB of routing loss). The baseband signal chain comes after 20dB of RF front-end gain and the output buffer has a peak to peak signal swing of $0.6V$. This means the baseband signal chain needs a minimum gain of $46dB$, though more is necessary for smaller signals. As discussed in Chapter 3, signal chain gain and linearity are designed in the baseband stages to meet the needs of the ADC. However, high frequency ADCs are difficult to design or power inefficient, so the optimum ADC and actual ADC specs differ. The available ADC is a 14-bit ADC, with $V_{swing} = 1.7Vpp$. The ASIC operates on a 1V supply, so an off-chip VGA before the makes up the difference ADC to make sure that the full dynamic range of the ADC is utilized, but not saturated. Next, a 14-bit ADC provides $SINAD_{ADC} = 86dB$; with a system designed for 16 antennas with an array gain of $12dB$, and an $SNR_{ADC,in} = 12.5dB$, then from Eq. (2.13), the $SD_{ADC} = 2 \times 10^{-7}dB$, which is obviously unnecessarily small. Furthermore, the ADC has a sampling frequency of 1 Gsps, and is over-sampled by two for clock recovery. This limits the multi-user test setup to 250 GHz bandwidth. Ideally the system would utilize an ADC closer to 4 or 5 bits, trading off resolution for higher sampling frequency and power savings. However, this ADC still allows for multi-user testing, and single user testing is performed with lab equipment (See Chapter 6).

Table 5.1 reports the simulated gain, bandwidth and linearity budgeted for each of the stages discussed. The simulations were done with extracted devices, π-models of wires $> 100\mu m$ for worst case layout, and at nominal corner. The cumulative row includes gain and BW from the receiver beforehand of 20dB and 1.5GHz respectively. BF design included balancing a few low BW and higher BW stages to achieve a simulated cumulative 3dB corner around 850MHz.

Table 5.1: Baseband Simulated Gain, BW, and Linearity.

Stage Name	**Fine VGA**	**Coarse VGA**	**Buffer**	**Phase Shifter**	**Sum 1**	**Sum 2**	**Pad Driver**
Cumulative Gain (dB)	25-28	31-42	30.5-41.5	30.5-41.5	33.5-44.5	42-53	42-58
Cumulative BW (GHz)	1.2	1	0.95	0.9	0.9	0.9	0.85
Stage Gain (dB)	5-8	6/14	-0.5	0	3	8.5	0-5
Stage BW (GHz)	3.5	4.5/2.5	2	4	3.5	1.5	3
Input Referred P1dB (mVpp)	70	130	150	140	110	200	220

The Table 5.1 illustrates that summation stage 2 has the least slack in linearity due to significant gain before and long cross beamformer wires. As mentioned the signal chain is expected to support an input signal power up to -55dBm on the antenna input. A full uplink system must then be able to limit cumulative user TX power so that received power is not saturated on each antenna. If additional SNR is required, it the full RX array must include enough sub-array panels to provide enough array gain from summation after digitization. The following sections will go into circuit detail on each stage of the baseband beamformer and explain design decisions in context of low power, managing input and output parasitics, or delay matching.

5.3 Phase Shifter

The phase shifters, Fig. 5.4, are digitally controlled 5-bit I/ 5-bit Q vector modulators capable of phase and amplitude modulation. These implement the complex beamforming weight multiplication for each element of the array, with 8.5° of phase resolution which is sufficient for the conjugate beamforming with 16 antennas in the intended two-stage beamforming architecture. There are 512 phase shifters on the chip, all in simultaneous use during full operation. Thus, design of the phase shifter block focuses on keeping its own input parasitics, area, and power as low as possible.

This is done, first, by reducing phase shifter area. To implement 8.5° of phase resolution, illustrated in Fig. 5.4 as a desired perimeter of black dots, a typical Cartesian vector modulator needs 16 g_m cells, illustrated as an 8×8 square of reachable grey dots per quadrant. Instead, the phase shifters used here are implemented with a shared g_m cell scheme[60]. I and Q inputs are shared between 11 cells to create a diamond of reachable points with the same phase resolution, illustrated in blue. This enables both phase and amplitude manipulation, which is convenient for beamforming algorithms other than conjugate beamforming. However, lower amplitude perimeters will have lower phase resolution and such algorithms may have limited spectral efficiency as a result. Cells are shared using a small transmission gate multiplexer to provide any of I+/I-/Q+/Q- to reach all four quadrants, or common mode voltage to disable a cell. The reduced number of cells necessary reduces the total area of the BF and thus the wiring parasitics necessary to cross it.

Second, the phase shifters are designed to be as low power as possible, with the assumption that gain in the distribution stage will handle the noise requirement. The tail current can be lowered by decreasing device size, which is limited by matching and decreasing overdrive voltage, which are limited by linearity. To further reduce the tail current, resistive degeneration was used to maintain linearity with a lower bias current without increasing the gate capacitance as a longer device would. This has the added benefit of improving matching. Resistive degeneration enabled a g_m cell with $16\mu A$ tail current, and shared cell design means there are only 11 g_m cells, so each phase shifter burns only $176\mu A$ while maintaining small area and input parasitics.

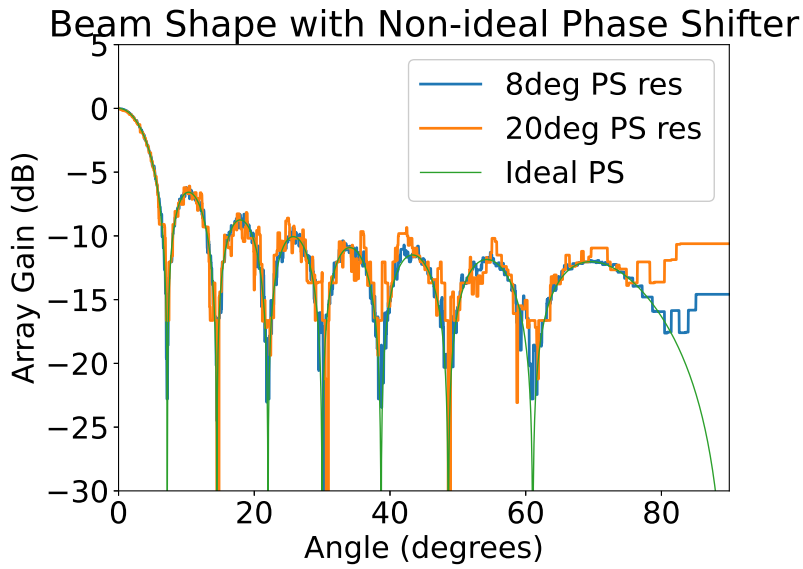

Figure 5.5: Simulated phase error of beam in Python, found by rounding beam weight to nearest available phase step with given resolution.

Phase Shifter Degrees of Resolution

Phase shifter resolution is another important concern, as it effects beamforming accuracy. Conjugate beamforming with many antennas, unlike spatial notching or other beamforming algorithms, is very robust to low phase shifter resolution within the beam. The main effect, as shown in the beam figure below, is at the edges of the beam pattern. There is also deviation from ideal at the beam nulls, which may worsen inter-user interference, but which can be improved with a second stage of zero-forcing beamforming as in a two stage architecture.

The desired resolution in degrees is chosen by the desired inter-user interference, and the number of cells needed to support this minimum resolution is found numerically by calculating the maximum phase step on the grid of reachable points for a given number of cells. System design determined 8.5° of phase resolution should be sufficient for up to 16 users with ZF. Thus 11 cells were chosen, which provides 8.2°.

The phase shifters suffer, intentionally, from particularly small gate size and need relatively good matching to achieve the desired 8.5° of phase shifting resolution. Thus, Monte Carlo simulations were run over FF, SS, and TT process corners; and 1.2V, 1V, and 0.8V voltage corners. We have left out SF and FS corners because the phase shifter has a resistive load instead of an active load, so N/P mismatch is less relevant when all the devices in the amplifier are NMOS. We have also left out temperature since we have no temperature correction on-chip and the chip is tested in a controlled environment. In a product, such an ASIC should include temperature compensation, but we leave that for future work.

VDD (V)	Corner	RMS error (%)
0.8	FF	0.9
1	FF	1.8
1.2	FF	2.0
0.8	SS	8.4
1	SS	1.2
1.2	SS	1.6
1	TT	1.5

Simulation results show a nominal 1.5% and worst case 8% RMS error vector. Worst case occurs for SS and VDD=0.8V. Fortunately, the BF has its own power supply and this is an easy corner to avoid by increasing BF VDD. The measured RMS error from testing is 3% for a TT chip. This is sufficient to achieve our desired phase resolution.

Phase Shift vs. True Time Delay

The vector modulator is an I/Q phase shifter, not true time delay, so for extremely large arrays or extremely wide bandwidth signals, there is some beam squinting as a phase shift is equivalent to a time shift for only a single frequency, and the error becomes larger further from the center frequency. To get an idea of how much phase shifting vs time delay matters, we can calculate the frequency offset that results in a maximum acceptable error. Beam squinting acts like an incorrect angle of arrival on a beam. For an angle of arrival θ, the phase difference between two antennas is $\Delta\phi$ for a carrier frequency f_c from d wavelengths away.

$$\Delta\phi = \frac{2\pi f_c}{c} d \sin(\theta) \tag{5.1}$$

An error in f_c can be compared to an equivalent shift in $\sin(\theta)$ to estimate the expected array gain degradation, as seen in Fig. 5.6.

As the figure shows, the error from beam-squint becomes more pronounced at large arrays. This necessitates time delay correction between sub-arrays and puts a limit on the maximum sub-array size. An error of $-1dB$ for 4 antennas occurs at 19% of f_c. As the number of antennas increases, the $-1dB$ error point occurs sooner. As an example: for 8 antennas, the $-1dB$ error is at 9.2% of f_c, and for 16 antennas at 4.6%. For a carrier frequency of 71GHz with 16 antennas, as is used here, the error in array gain from beam squint is negligible for bandwidths below $0.046 \times 71GHz = 3.3GHz$. If bandwidths beyond this are targeted, a smaller subarray or frequency dependent beamforming should be used.

5.4 Summation

After phase shifter weights are applied, there are two stages of active current summation across the BF, with circuit detail shown in Fig.5.7. The minimum summation stage power

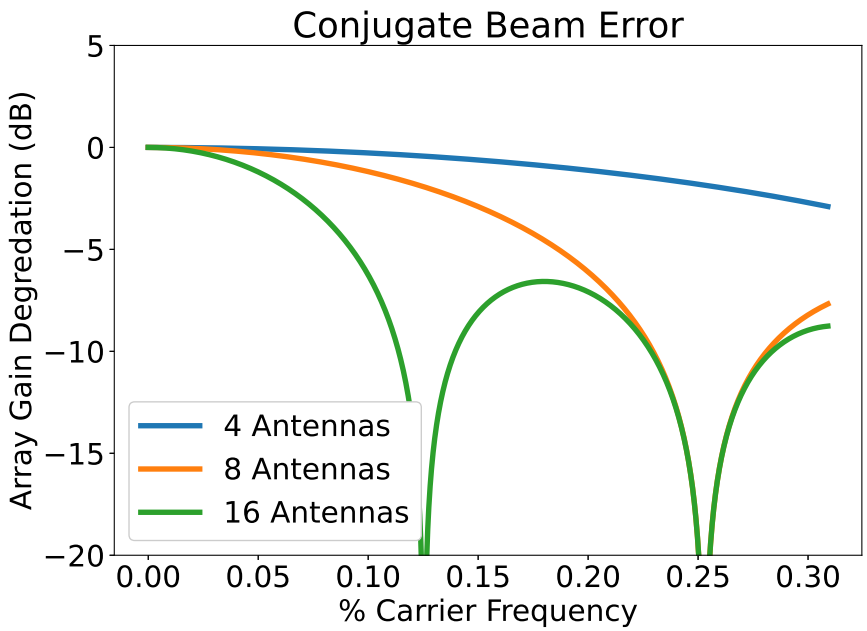

Figure 5.6: Array Gain Error, plotted over distance from carrier frequency in $\%f_c$

is limited by bandwidth, due to the large cross-array wiring parasitics. The analysis used to determine whether noise or bandwidth is limiting power is performed in Section 3.5. In the bandwidth limited design case, two-stage summation provides improved gain-bandwidth and ultimately lowers the summation power. The first stage provides 12dB of gain, the second stage 0dB, and both stages sum four signals. The first stage of summation performs local quadrant summation, and cross beamformer wiring is done in the second stage, routed on top metals to reduce parasitics. Both summation stages feature a cascode to further decouple gain and bandwidth design requirements and reduce summation power.

The further advantage of this two stage summation is the layout symmetry enabled by doing summation in sets of four. Summation layout is illustrated in Fig. 5.8, where antenna delay matching on the summation nodes ends up dictating low level BF layout. The first stage of summation, in the grey box, sums four local weighted antenna signals, while the second stage, on the far right end of the layout, sums the four quadrants across the full beamformer. In red, we show four phase shifters of User 1 for Antennas 1-4, their output are fed to the transconductor of the first summation stage and currents are summed symmetrically on to a load, at the center. This gives us the first stage summation output for user 1. Routing length, and thus delay is equal for all four antenna inputs. The summation for the same user's Q output, in blue, is done as similarly as possible, just below the I signals. User 2, in green, is also symmetric around the load, but with longer routing than User 1. This results in slight mismatch between users, but antenna mismatch is tightly controlled within each

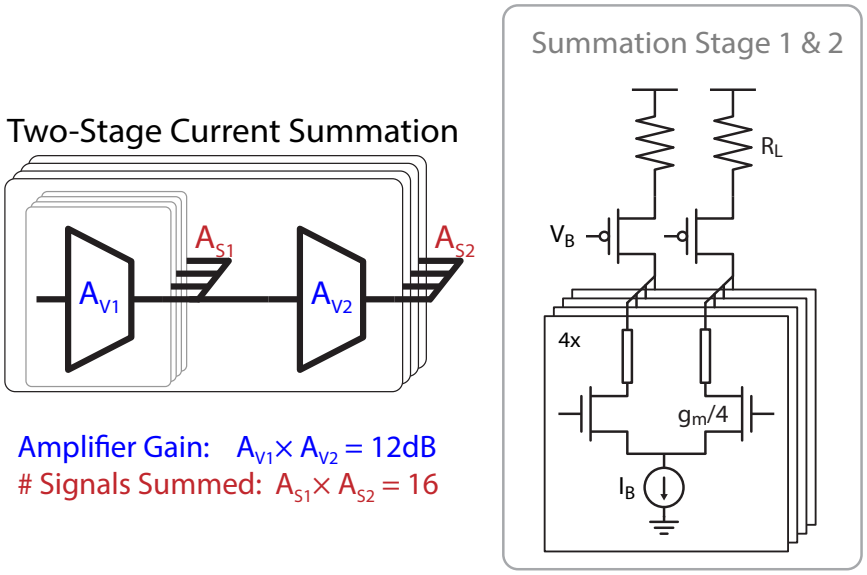

Figure 5.7: Two stage summation circuit with high-level illustration, where A_S is the number of signals summed in a stage, and active voltage gain is A_V. Active current summation circuit detailed for both stages.

user. In this way, the two stage beamformer allows for lower power design, while matching antenna delay across a single user.

5.5 Output Buffers

The output buffer chain, shown in Fig. 5.9, includes a Cherry-Hooper pre-driver and a terminated pad driver. 32 of such chains are included on-chip, to drive the 16 I and 16 Q differential baseband output pads. The Cherry-Hooper stage drives the 2-3mm routing lines between the beamformer outputs, at the center of the chip, and the output cells on the IC periphery. Transconductors are differential stages with active load and common-mode feedback loop, to optimize output bias for linearity. Feedback resistors are digitally-controlled to achieve 0-5dB programmable gain. Pad drivers are resistively-terminated differential pairs, designed to provide 0dB voltage gain when driving an external 100Ω load. Each output chain burns 15mW nominal power, and achieves simulated 3GHz BW and 300mV differential zero-peak output-referred 1dB compression point (OCP1dB). Bias current can be digitally increased to improve bandwidth and linearity, e.g. raising power to 20mW results in ∼30% improvement of both BW and OCP1dB.

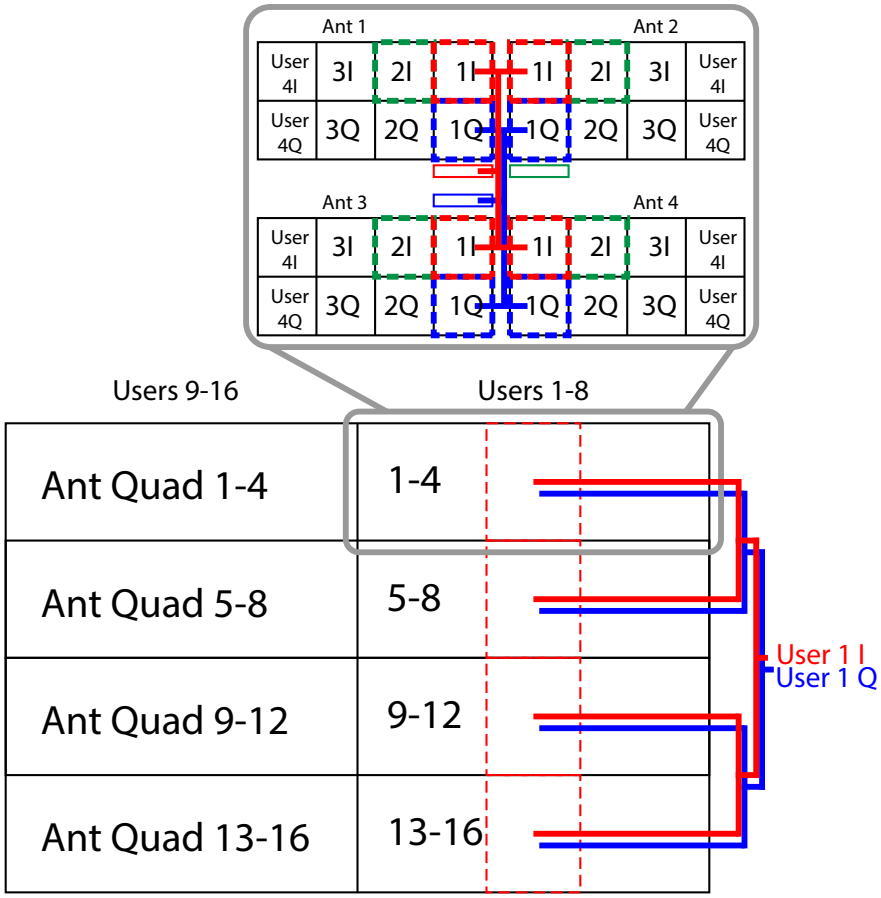

Figure 5.8: Beamformer layout routing example, illustrating antenna delay match and user delay mismatch for an antenna quadrant. User 1 I signal is in red, and user 1 Q signal is in blue, while user 2 I signal is in green.

5.6 Offset Correction and Digital Interface

Since the whole signal path is DC-coupled, on-chip offset correction is required to prevent railing and linearity degradation. Offset compensation is performed by three independently-controlled differential current DACs for each channel: one at the RX front-end output, one at the VGA (see Fig. 5.3), and one at the output pre-driver. StrongARM comparators are included at the output pads, as well as two internal nodes, to perform digital offset calibration.

The comparator outputs are read through a digital serial interface, also used to write

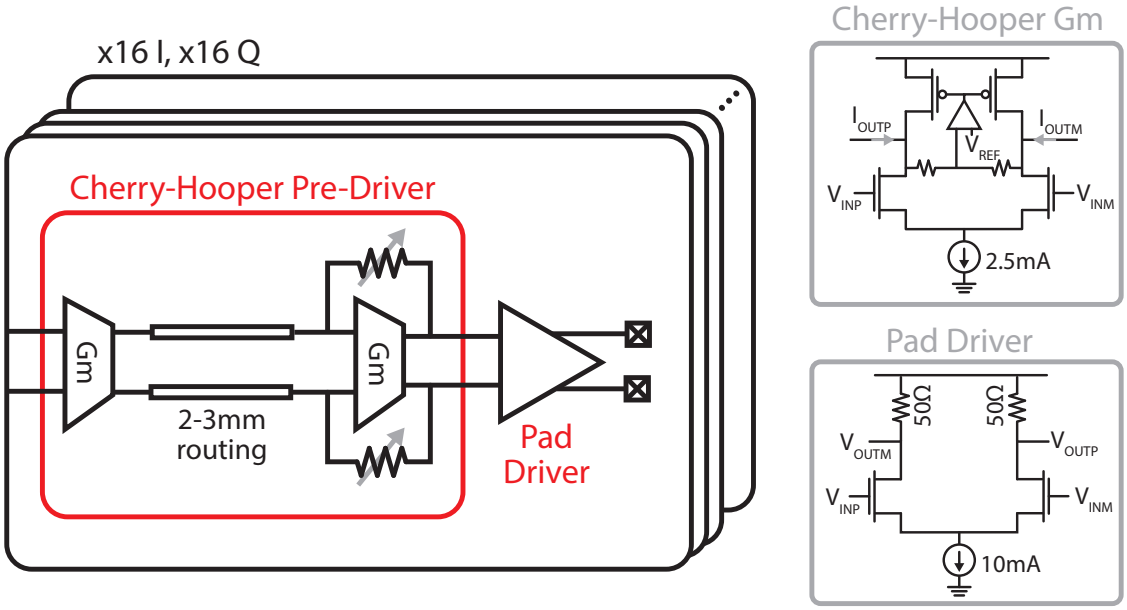

Figure 5.9: Schematic of the output buffer chain.

digital registers that control current DACs, beamformer settings, VCO frequency, etc. All digital calibration loops such as offset correction and PLL lock detection are run off-chip for simplicity.

5.7 RX Front-End

The RX front-end, is particularly suited to for massive MIMO systems, where minimizing area and power is key, while noise figure requirements can be relaxed due to array gains [22]. It is based on the direct-conversion mixer-first topology described in [61]. The RX targets moderate noise figure (8-10dB) with small footprint and low power consumption (12mW). It employs frequency-translational feedback and an input impedance transformation network to scale down the mixer size and reduce LO buffer power consumption, while preserving wideband input matching.

The reason to use an RX topology that trades NF for power may be subtle. The mixer-first front-end results in ∼3dB higher NF than other integrated arrays such as [32, 33], which requires an array to double the number of antennas to meet a given link budget. As discussed in Section 1.1, massive MIMO targets linear BF weight derivation algorithms, and thus must meet the $M \gg K$ condition (typically $\alpha \approx 10$ or larger in practical cases) which is set by inter-user interference rejection requirements [20] [21]. For example, a system with half the number of antennas but 3dB higher single-channel NF would have the same SNR, but worse SINR. In that case, better NF would only improve system performance

in a noise-limited scenario, not in an inter-user-interference-limited one. Given that this work specifically targets many user operation where inter-user interference is the dominant source of noise, NF is traded off for power, and area. The mixer-first RX enables tighter packing of RX channels into a transceiver ASIC, not just because of the power saving but also because of the very compact footprint. For cases with smaller K, however, placing an LNA before the mixer will result in better system performance. The exact number of K for which an LNA is preferred depends on link-budget requirements such as maximum distance, bandwidth and modulation order. The LNA power can be estimated at ∼10mW, as is the case for [62]. Bandwidth extension of [62] to cover the whole 71-86GHz BW with margin could be achieved by using staggered transformer-coupled interstage networks, at expense of extra power and area, and higher NF (see for example [63]).

However, for massive MU-MIMO systems targeting a large user pool ($K \gg 1$), the number of antennas is set by the $M \gg K$ condition required to achieve sufficient inter-user interference rejection [21], rather than by noise-related link budget requirements. This results in significant array gain, which in turn allows the system to relax the front-end NF requirement. In this case, a mixer-first RX which trades NF for power and area is advantageous.

5.8 LO Generation

In the massive MIMO architecture shown in Fig. 1.5, the LO reference must be delivered to N sub-arrays. Cross array LO distribution can be very expensive if performed at high frequencies where the cross array interconnect results in high loss and the gain available to clock buffers is limited[22]. Lower frequency distribution, however, results in large multiplication factor of reference phase noise outside of the clock recovery bandwidth, and requires more power to achieve acceptable phase noise[25]. For an array of this expected size and frequency, power is minimized for a cross-array sub-6GHz reference frequency. Therefore, LO generation circuits, shown in Fig. 5.10(a), are designed to derive a 72-85GHz tone from a 3-4GHz external reference. The LO generation block consists of two subcircuits: a 24-28GHz integer-N PLL, and an injection-locked frequency tripler (ILFT). Running the PLL at ∼25GHz benefits from higher quality factor in the oscillator tank, allowing one to meet tuning range and phase noise requirements with low power consumption [64]. In Two-Stage Beamforming, the PLL BW and the carrier recovery BW impact the amount of correlated and uncorrelated phase noise in the final, aggregated beams, which can be optimized to a particular signal bandwidth [25]. Therefore, the PLL loop BW is designed to be programmable as well as frequency tunable over the entire Eband. Frequency tuning of the VCO is performed using a digitally-controlled switched-capacitor bank, and a small varactor for analog tuning is used to adjust the VCO frequency gain (K_{VCO}) as oscillation frequency changes. This achieves constant K_{VCO} across the tuning range.

The VCO is followed by a frequency-divide-by-8 chain, including a CML-to-CMOS converter buffer, a high-speed TSPC divider [65] and two static CMOS dividers. The buffered

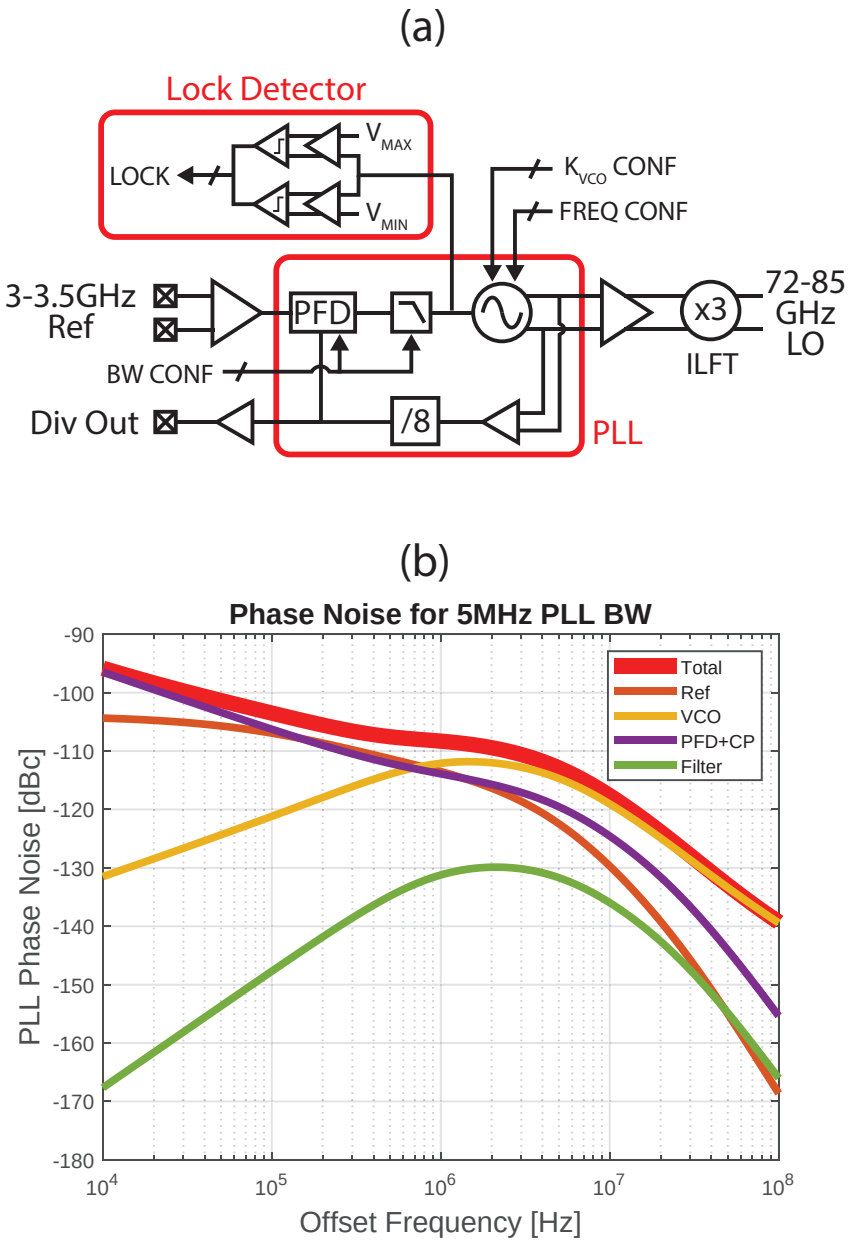

Figure 5.10: (a) LO Generation schematic. (b) Simulated PLL phase noise.

divider output is available on a pad for measurement purposes. The phase-frequency detector (PFD) is a typical D flip-flop. The charge pump and low pass filter, while also of typical make, are programmable to provide a PLL BW between 1-10MHz by configuring the charge pump current and the filter's pole and zero. On chip, the PLL is the dominant source of phase noise, broken down by contribution in Fig. 5.10(b). This simulated PLL phase

noise plot was obtained by simulating noise floor, flicker corner, and bandwidth values from extraction, and then generating the combined phase noise from theory in Matlab. Phase noise from the tripler is negligible [66]. LO distribution contributes uncorrelated phase noise from parallel noise sources, but has a negligible contribution compared to correlated phase noise.

A lock detector is connected to the PLL control signal V_{CTRL}. The lock detect consists of two StrongARM comparators [67], monitoring if $V_{CTRL} < V_{MIN}$ or $V_{CTRL} > V_{MAX}$, which indicates that the charge pump has railed. When a lock fail is detected, the VCO digital frequency control is adjusted to a higher or lower code accordingly. Single-stage open-loop buffers between V_{CTRL} and the comparator isolate the PLL from potential kickback coupling of the comparator clock.

The PLL VCO drives the ILFT to generate the mm-wave LO reference. The tripler is based on the dual-injection topology presented in [66], which achieves wide operation range without the use of tuning elements and calibration loops. All circuit details are identical to [66], except for the center frequency, which was readjusted to match the required LO range.

5.9 LO Distribution

The LO distribution architecture is shown in Fig. 5.11. It consists of a 1:2 power splitter, a regeneration buffer, and a 1:4 power splitter, whose outputs drive two RX elements each. Quadrature generation is performed at each RX element, to simplify distribution. The LO distribution circuits employ impedance-scaling techniques to reduce power consumption. Unlike off-chip microwave circuits that are usually designed for 100Ω, on chip the designer has the freedom to scale impedances to optimize performance. The 1:2 and 1:4 splitters, are based on a non-isolated $\lambda/4$ power splitter structure (see Fig. 5.11), which unlike a conventional Wilkinson does not require the outputs to be physically close. In this way, $\lambda/4$ lines can be used for both matching and routing, thus minimizing power loss in cross-chip distribution [68].

Quadrature generation is based on a transformer-coupled implementation of a quadrature-hybrid coupler, which features compact area and low insertion loss [69]. At the hybrid output, LO buffers identical to the design in [61] drive the RX mixer gates.

5.10 Package and Antenna

The 4mm×4mm ASIC, realized in CMOS 28nm technology, is shown in Fig. 5.12. The 16 mm-wave inputs are located on the left side, while baseband outputs occupy the other three sides. The ASIC is flip-chip attached with C4 solder bumps to a 3.3cm×3.3cm organic interposer, which includes a linear array of 16 patch antennas, as shown in Fig. 5.13.(a). The 4-layer organic substrate interposer, shown in Fig. 5.13.(b), has a thick core dielectric and thin top/bottom build-up films, featuring low loss at mm-waves and high-density traces

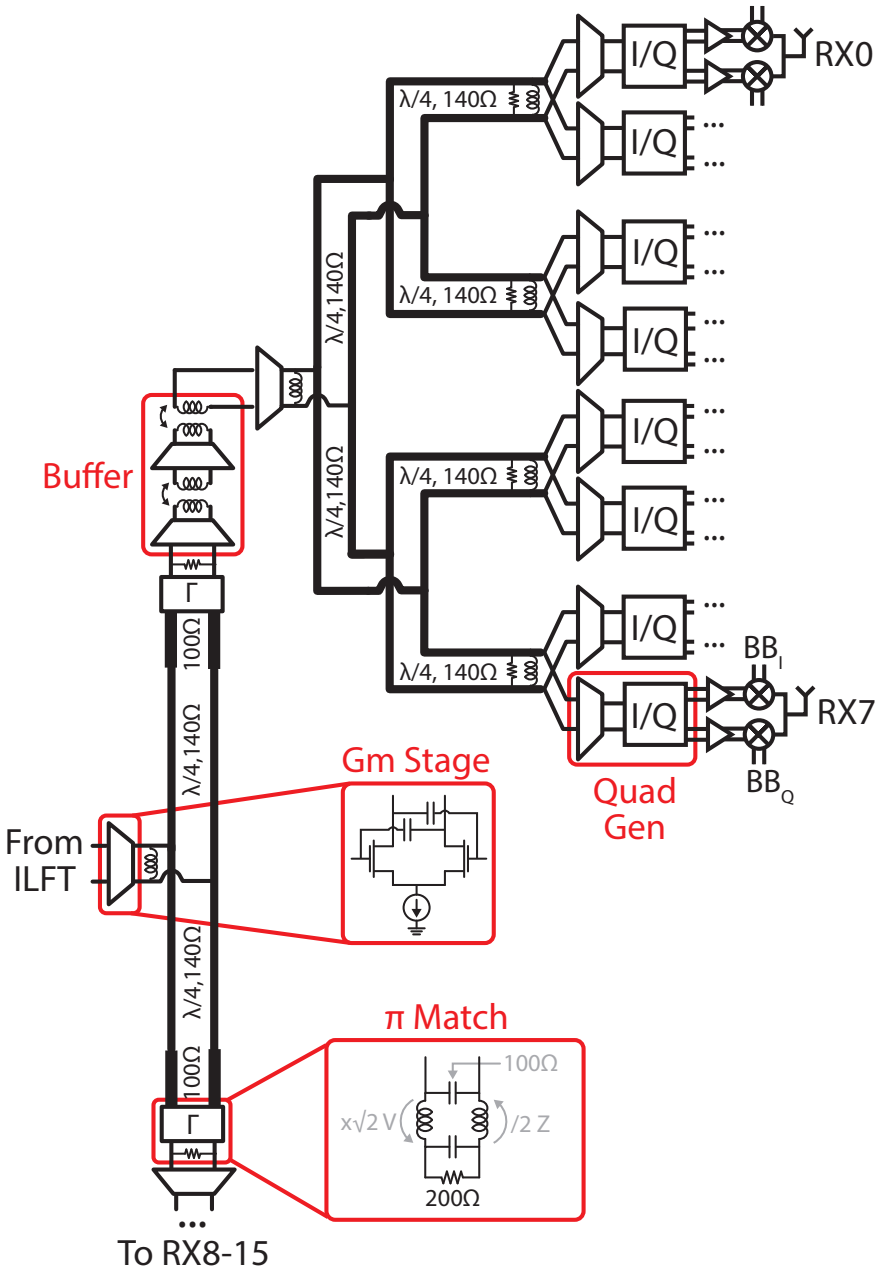

Figure 5.11: LO distribution architecture. Schematic geometry matches layout floorplan.

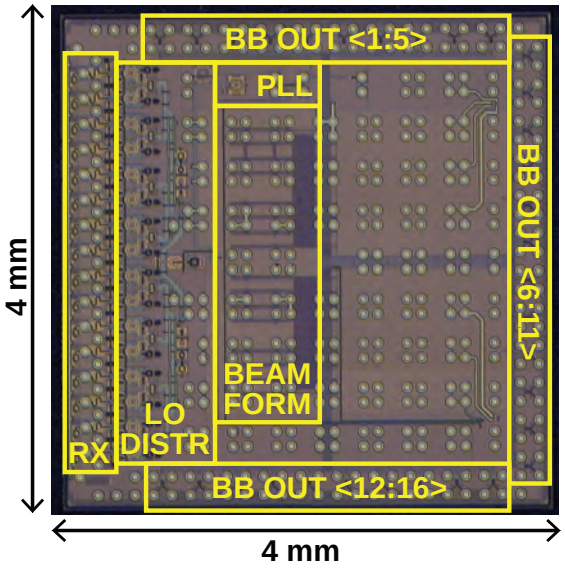

Figure 5.12: Chip microphotograph.

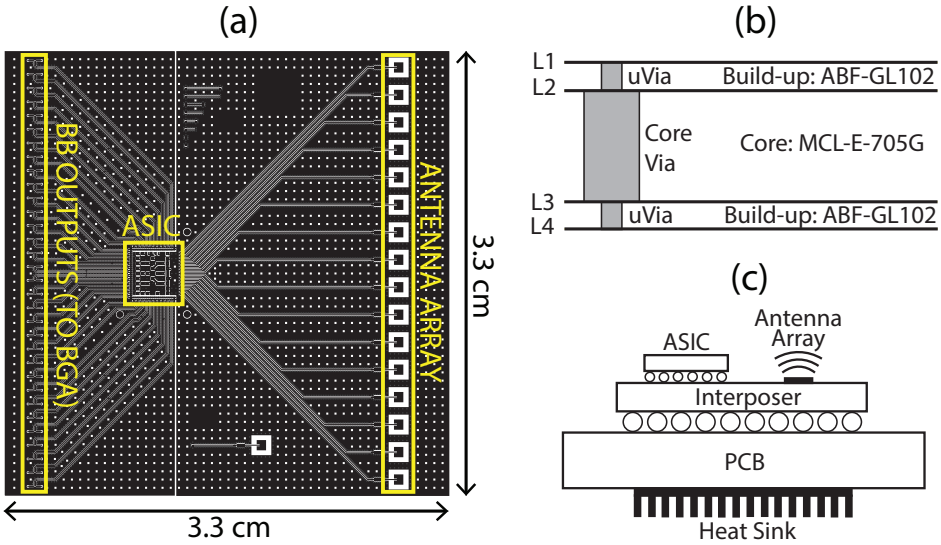

Figure 5.13: ASIC package overview: (a) interposer layout, (b) interposer stackup and (c) assembly diagram.

and microvias. As shown in Fig. 5.13.(c), the interposer is mounted on the PCB as a BGA module. Thermal vias through the interposer and PCB create a low-thermal-resistance path between the ASIC and a heat sink on the PCB bottom layer. Multiple modules can be tiled together on a PCB to create a massive MIMO array.

Antennas are routed to the ASIC using 50Ω shielded microstrip TLines, with 1.5dB/cm simulated loss. To reduce routing loss and layout complexity, TLine lengths are not equalized across the array. The worst-case length mismatch (5mm) corresponds to a ∼30ps time delay, which is a small fraction of the baseband data Unit Interval (UI). Thus, true-time-delay effects within the signal BW are negligible [70, 71], and the phase mismatch is corrected by a one-time beamformer calibration (see Chapter 6).

Chapter 6

Measurements and MU Demonstration

Measured results from the Multi-user, Massive MIMO Receiver ASIC are reported in this chapter.

6.1 Testing PCB

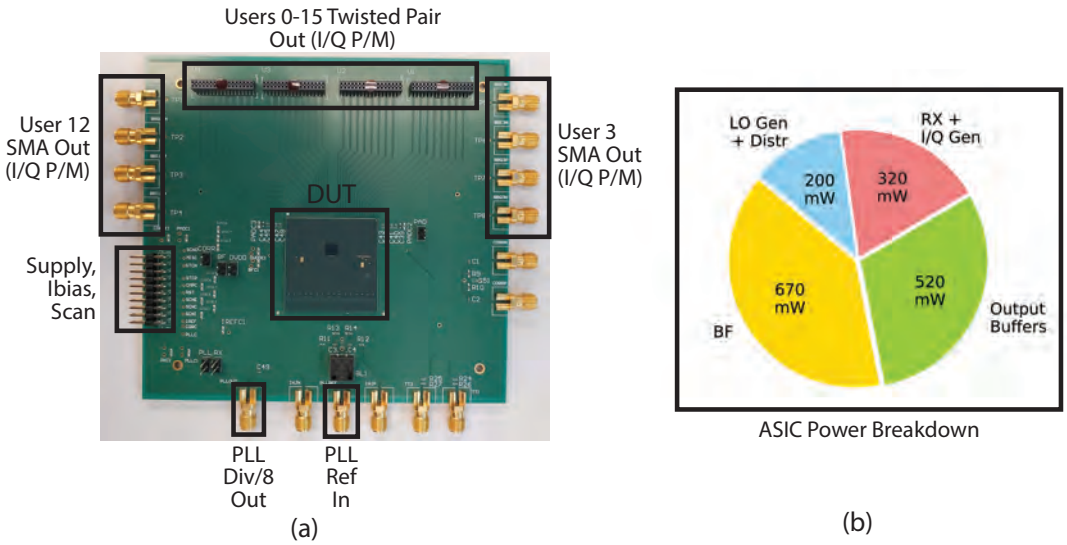

Figure 6.1: (a) Test PCB. (b) Measured power breakdown.

The testing PCB is shown in Fig. 6.1.(a). A single ASIC on a PCB is referred to as a panel, and can be daisy chained together with other panels, to increase antenna count

in a two-stage beamforming architecture. The 16 baseband beam outputs are routed to TigerEye twisted pair sockets, and two outputs (user 4 and 13) are also accessible via SMA connectors. A separate board provides 5 separate power supplies through dedicated LDOs, 10 bit programmable bias current, and level shifting for digital control signals from a Spartan6 FPGA. Total power consumption was measured at 1.7 W, which includes all 16 input and 16 user outputs active. Power breakdown is shown in Fig. 6.1.(b), and measured using current monitors on each power domain's LDO. For operation above 82GHz, LO distribution current is increased by 100mA to achieve sufficient LO swing at the RX, resulting in 1.8 W total power.

6.2 Single-User Continuous-Wave Measurements

The setup in Fig. 6.2 was also used for continuous-wave wireless measurements. A mm-wave tone was transmitted to the ASIC using a signal generator, mixer, ×6 multiplier, and horn antenna. SMA baseband outputs were analyzed using a real-time scope or a spectrum analyzer.

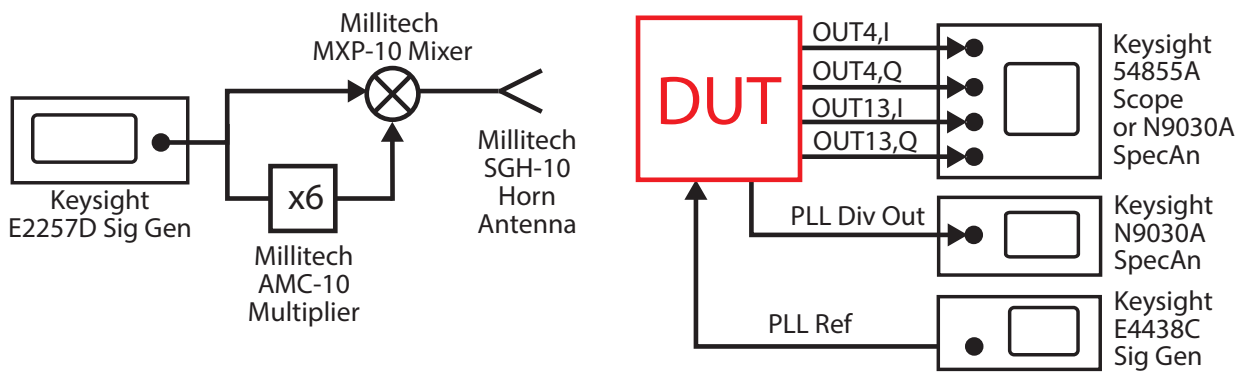

Figure 6.2: Setup for PLL and continuous-wave wireless measurements.

Phase Shifter Characterization

Characterization of individual phase shifters, shown in Fig. 6.3., was obtained by disabling (setting to CM input) all phase shifters in the BF, save a reference phase shifter and the one under test. The TX was placed at broadside of the RX array and set to transmit a single-side-band tone at 20MHz IF. The beamformer matrix was configured to output antenna RX1 on both baseband output 13, used as a reference, and baseband output 4, used to characterize the phase shifter. All available phase shifter settings for output 4 were swept and the magnitude and phase information of the downconverted tone was measured against

the reference signal. Noise was averaged over multiple waveforms collected on the scope. Phase shifter characterization showed an RMS Vector Error of 3%, which is close to the RMS Vector Error of 1.5% from Monte Carlo simulations over process variation.

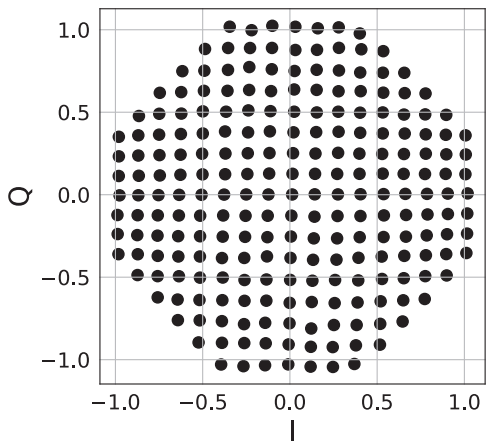

Figure 6.3: Single phase shifter vector modulation characterization for Antenna 1 and User 4.

Antenna Calibration

A similar process as above was used to perform a one-time antenna delay calibration, before beamforming, to compensate for phase mismatches between ASIC-to-antenna routing traces on the package (see section IV). The beamformer was configured to deliver RX1 as a reference to baseband output 4. RX1-16 were iteratively delivered to baseband output 13, and the phase shifter setting of output 13 was swept until the phase of both waveforms aligned. Phase shifter offsets for several carrier frequencies were stored in a lookup table. Note, that because phase offsets are dominated by systematic TLine delay mismatch on the package, they are stable across on-chip PVT variations, so a one-time calibration is sufficient for reliable operation. Moreover, because antenna delays within a single user are matched inside the beamformer (see section III.D), calibration values derived for beam output 13 are also valid for the other outputs.

Beam Pattern

Beam patterns, reported in Fig. 6.5, were derived by transmitting a mm-wave tone at 100MHz IF and measuring the downconverted RX power on the spectrum analyzer. Measurements show single-user BF operation with 10dB worst-case spatial rejection at first

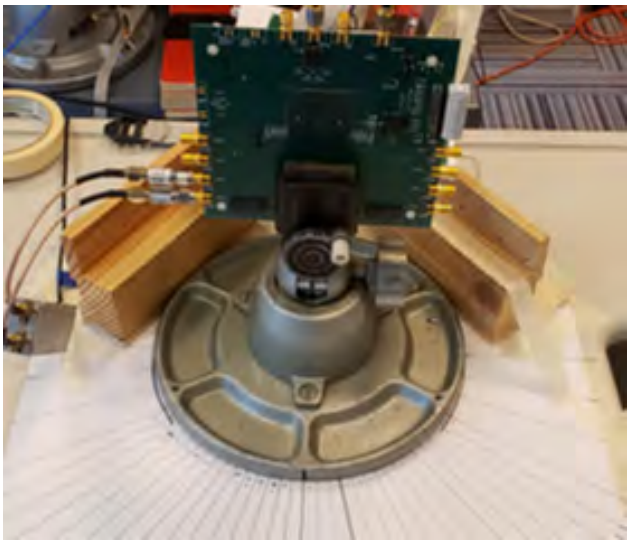

Figure 6.4: Beam patterns measurement setup pictured

sidelobe. Out-of-beam sidelobe suppression is shown to be around -20dB, which is expected for 16 antenna conjugate beamforming. On-chip cross-user coupling is then negligible in comparison to inter-user interference innate to the wireless channel, and is dealt with in the back-end by the second stage of zero-forcing. Consistent performance is shown across different carrier frequencies and for different output channels where user 4 and user 13 have worst and best case layout respectively.

6.3 PLL Measurements

The PLL was tested with the setup in Fig. 6.2, using a signal generator as a reference and analyzing the divider output signal (see Fig. 5.10). The measured PLL range is 21.8-29GHz, corresponding to a 65.5-87GHz LO range after the tripler. Phase noise of this divided PLL output is shown in 6.6.(a), with different loop bandwidth values. The spur at 70kHz is from the power source, and is filtered by the digital carrier recovery loop during wireless link operation [25, 71]. All following measurements that require the PLL, use a PLL BW of 5MHz. The LO generation and distribution are on the separate power domains and consume 200mW of total power while active, which is measured to be 20mW for the PLL, 17mW for the tripler, and 163mW for the distribution.

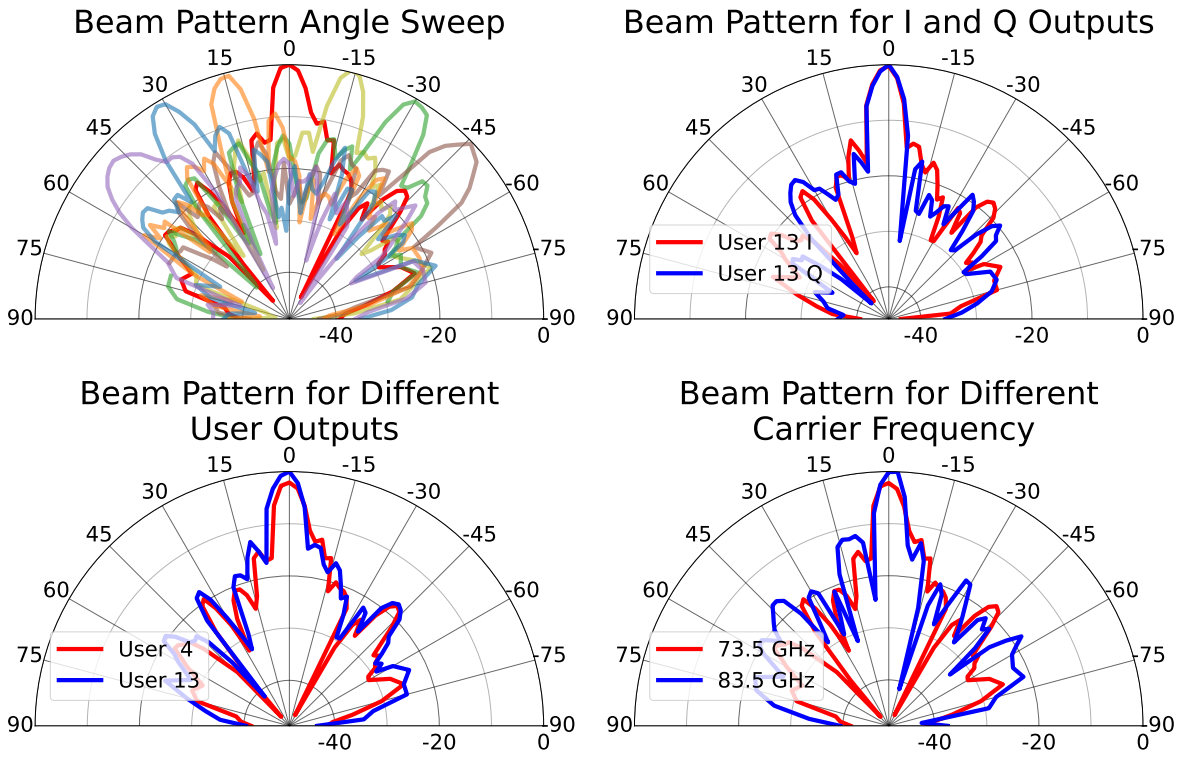

Figure 6.5: Beam patterns taken at 73.5GHz carrier on User 4 (unless otherwise specified)

6.4 Single-User Wireless Link Measurements

The setup for modulated-data wireless measurements is shown in Fig. 6.7. I/Q PRBS10 streams were provided by a pattern generator, and upconverted to E-Band using a sliding-IF TX. The RX ASIC output waveform was acquired on the real-time scope and saved. Afterwards, a Python DSP chain performed standard radio back-end operations, such as I/Q correction, filtering and timing recovery. A decision-directed phase tracking loop was also included to cancel low-frequency phase noise [71]. For a fair comparison with the state of the art, no DSP equalization was performed. The link distance was limited to 60cm due to setup constraints, and the TX output power was limited to -12dBm. This results in an estimated -58dBm at the ASIC input.

Received constellations for 73.5GHz and 83.5GHz carriers are shown in Fig. 6.8; similar performance was measured for other carrier frequencies across the 71-76GHz and 81-86GHz

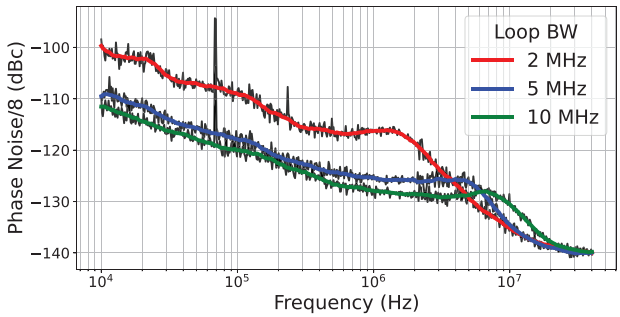

Figure 6.6: Phase Noise of PLL Output/8 measured for different loop BW settings with 3.125GHz Reference.

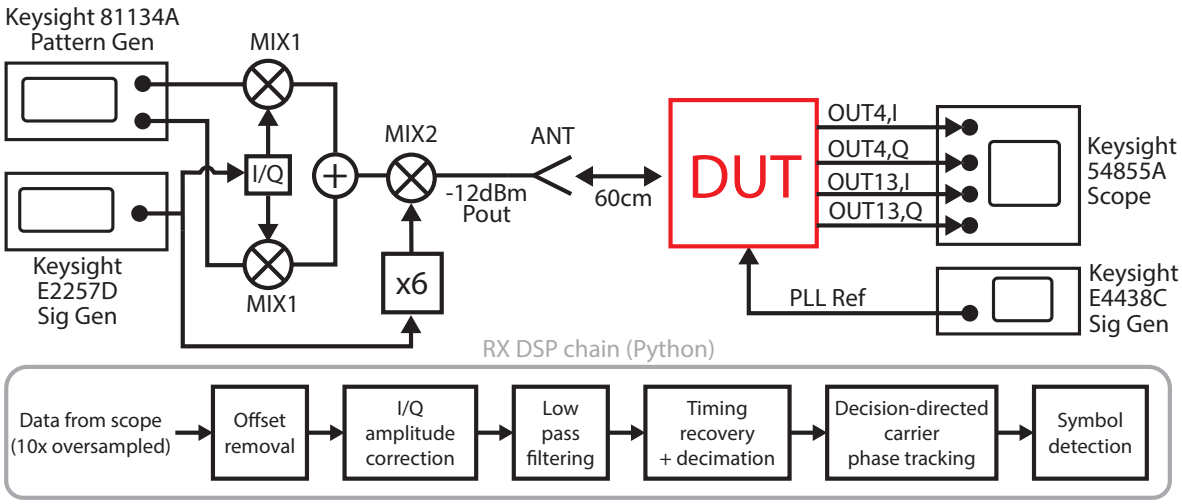

Figure 6.7: Measurement setup for QPSK single-user wireless link. In the 16QAM setup, two identical pattern generators were used, combining the outputs to obtain I/Q PAM4 data streams. Components: MIX1 - Minicircuits ZX05-24MH mixer; MIX2 - Millitech MXP-10 mixer; x6 - Millitech AMC-10 multiplier; I/Q - Meca 705S-11.750 90° coupler; ANT - Millitech SGH-10 horn antenna.

bands. Beamformed single-user wireless links up to 2Gbps for QPSK and 16QAM are supported on the 71-76GHz band with $<10^{-3}$ BER. For the lower band, the signal throughput is limited by the on-chip baseband signal chain BW, which results in self-interference and reduced EVM. Signal BW could be extended with equalization, but results are shown here without equalization for fair comparison with state of the art. 16-QAM measurements show early signs of gain compression, limited by the ASIC, as RX power is near design limit.

However, in a full multi-panel system, RX power would be decreased as many panel array gain will provide more SNR boost. In the 81-86GHz band, the QPSK data rate is limited to 1Gbps due to a decrease in the RX mixer bandwidth, likely due to frequency misalignments in the on-chip LO chain. 16QAM measurements at upper E-Band could not be obtained due to linearity limitations in the TX setup, caused by off-the-shelf mm-wave waveguide components operating outside of the intended range. Although not allocated for wireless communications, the ASIC also covers the 76-81GHz band, where single-user QPSK constellations were measured with performance comparable to the 71-76GHz range.

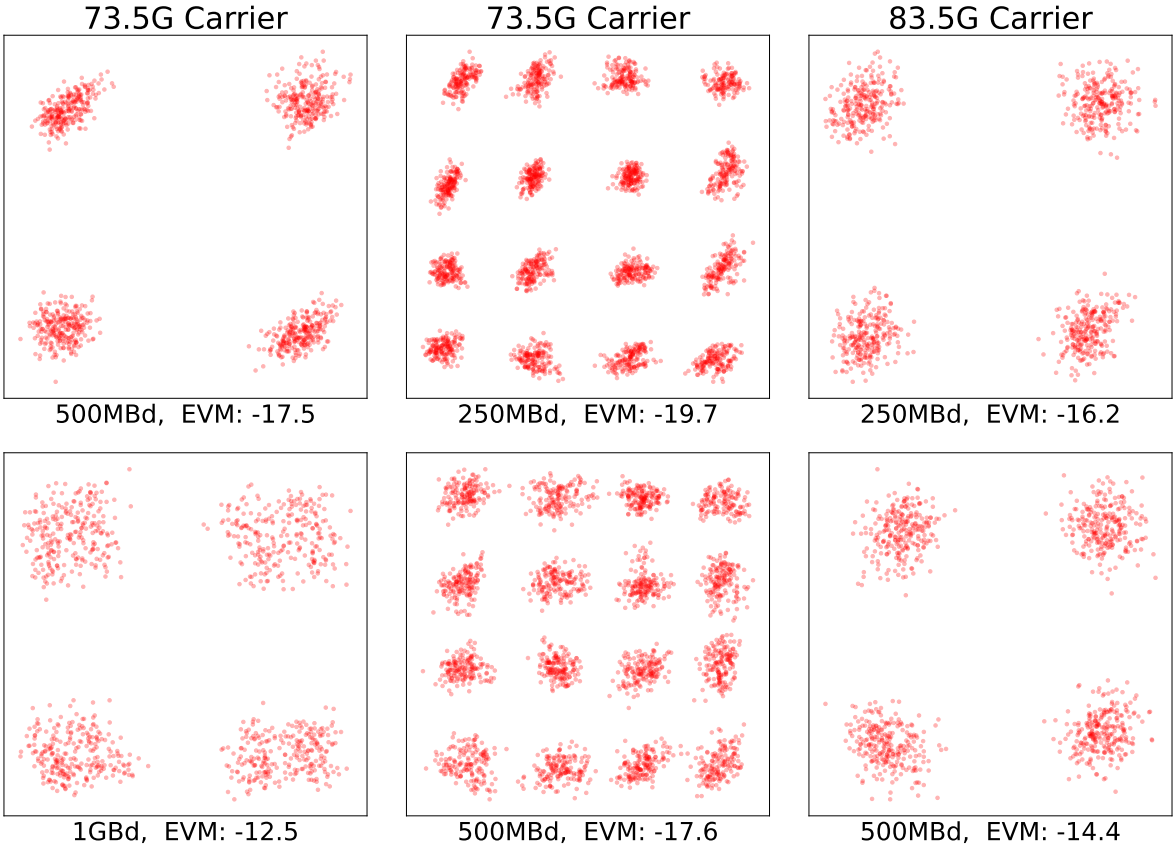

Figure 6.8: Singe-user constellations

6.5 Multi-User Wireless Link Measurements

The multi-user wireless setup for one and two panels is based on the MIMO testbed presented in [27]. Four COTS TX user-equipment (UE) devices were assembled using FPGAs and custom PCBs that include DACs, E-Band quadrature upconverter, PA, and patch antenna.

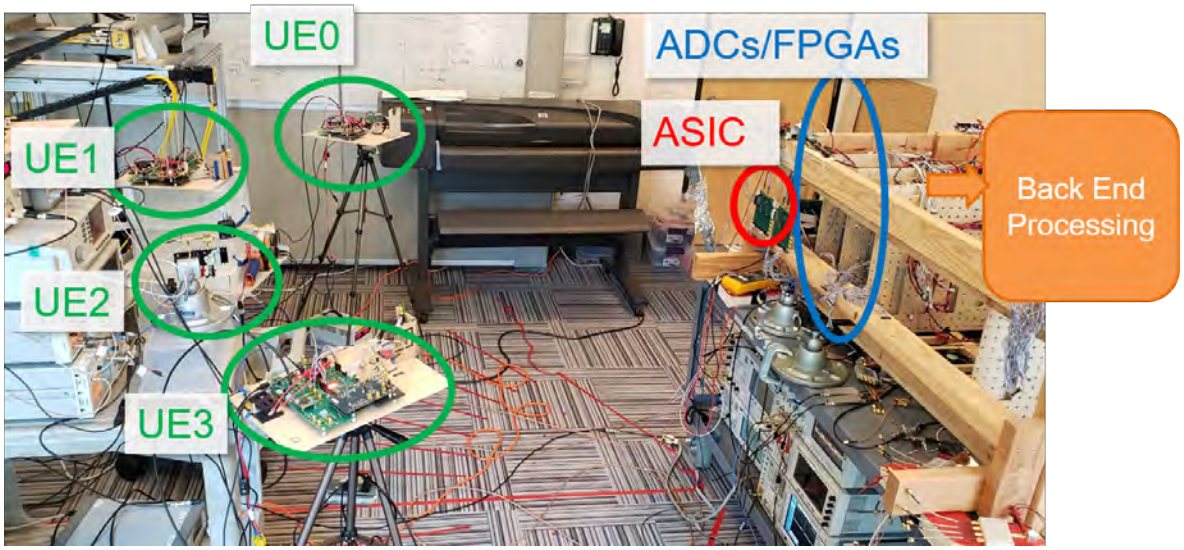

Figure 6.9: Photo of 4 user demonstration. Users positioned at approximately 15 degree increments (-36 deg, -15 deg, 13 deg, 40 deg) and approximately 1.5m distance from receiver

The UEs, which transmit ∼8dBm power each, were placed at increments of about 30 degrees, 1.5m away from the RX array. The ASIC's analog BF was open-loop configured to point beams in the direction of the corresponding UEs, and the 16 baseband outputs from the ASIC were digitized using a custom ADC board and fed to an FPGA for timing correction. Back-end DSP is performed in post in Python, with mean and power calculated over the span of a frame. More details on the UE, ADC custom hardware, and back-end DSP are presented in [27]. The back-end DSP is similar to the single-user measurements, with the addition of a pilot-based zero forcing stage, which completes the two-stage beamforming algorithm described in Section II to reduce inter-user interference. As shown in Fig. 6.10, UEs begin by sending time-interleaved pilots in the frame header, which are used to compute the zero forcer coefficients. After the pilots, UEs transmit their spatially multiplexed payloads at the same time and frequency.

Single Panel Results

Illustrated in Fig. 6.10, is the multi-user setup for a single panel communicating with 4 simultaneous users.

Multi-user, single panel results in Fig. 6.11 show single-user and multi-user, 73GHz carrier, QPSK constellations. The single user constellation provides a comparison with single user test setup, without inter-user interference. EVM is lower than single-user test setup at -15.3dB, and is limited by the high-pass corner of the AC-coupling from the DAC eval board used in the UEs. This lowers TX SNR and precludes testing with higher order constellations.

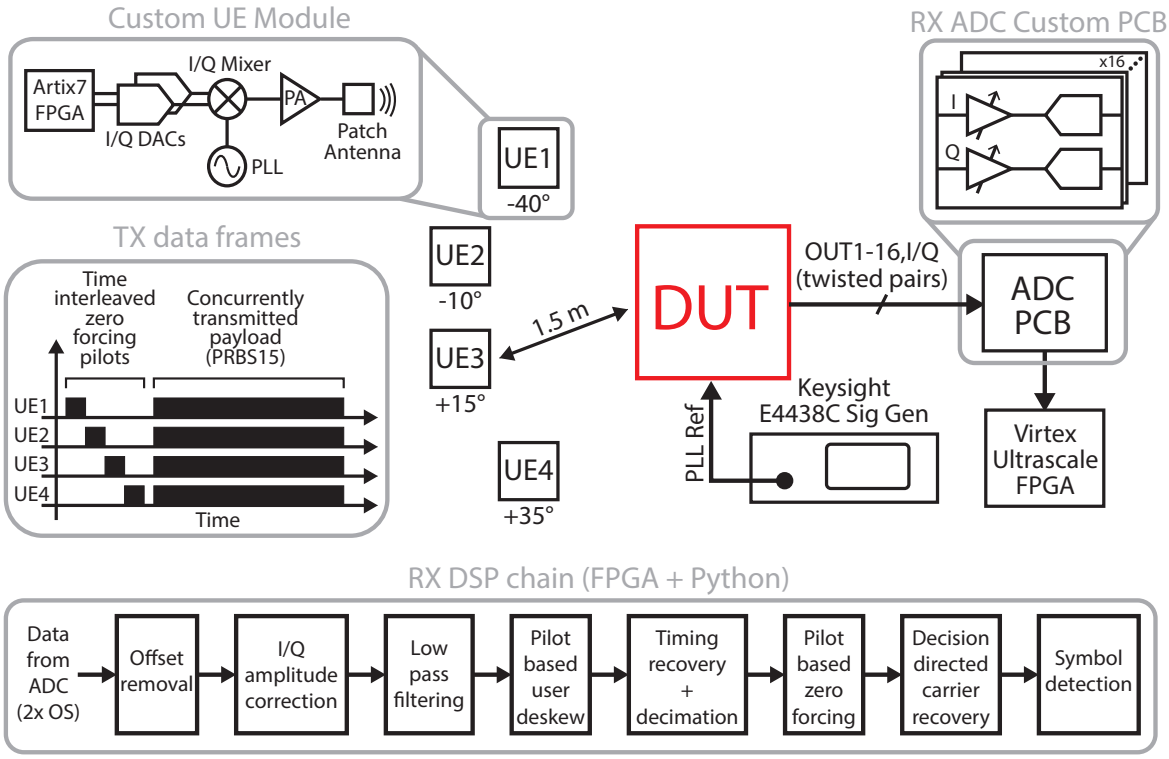

Figure 6.10: Measurement setup for a single panel, multi-user wireless link.

The four simultaneous users constellations show correct beamforming and spatial filtering of user streams, with an average EVM of -11.8dB. EVM is lower than the single user case, as expected, due to inter-user interference in the wireless channel. The four users demonstrate 0.5 Gb/s throughput each, for a total of 2 Gb/s from four ASIC outputs. The data rate is limited by the available ADC/DAC sampling rate. Multi-user wireless measurements were limited to 4 simultaneous users due to the availability of TX hardware.

Inter-user interference is handled at the full system level, where multiple panels will reduce beam-width and improve ZF. Shown in Figure 6.12 are two users transmitting to a single panel, before and after ZF. As shown, the ZF improves EVM significantly. The two user signals' EVM without ZF is significantly worse than the four user case. Note that the pseudo-inverse performed by ZF is prone to over-amplifying noise if the SINR of users is too low. Thus, ZF is more successful as α increases, which is why the 2 user case is better than the 4 user case, even though ZF has been performed on both.

Massive MIMO for extremely large α allows near optimal beam isolation with conjugate beamforming alone [20]. Use of ZF allows for smaller α, but since ZF still requires a minimum SINR, $\alpha \gg 1$ is still required. Thus, full 16 user operation is only feasible in a multi-panel setup. Instead, all the 16 ASIC outputs were tested by pointing each of the 16 beams

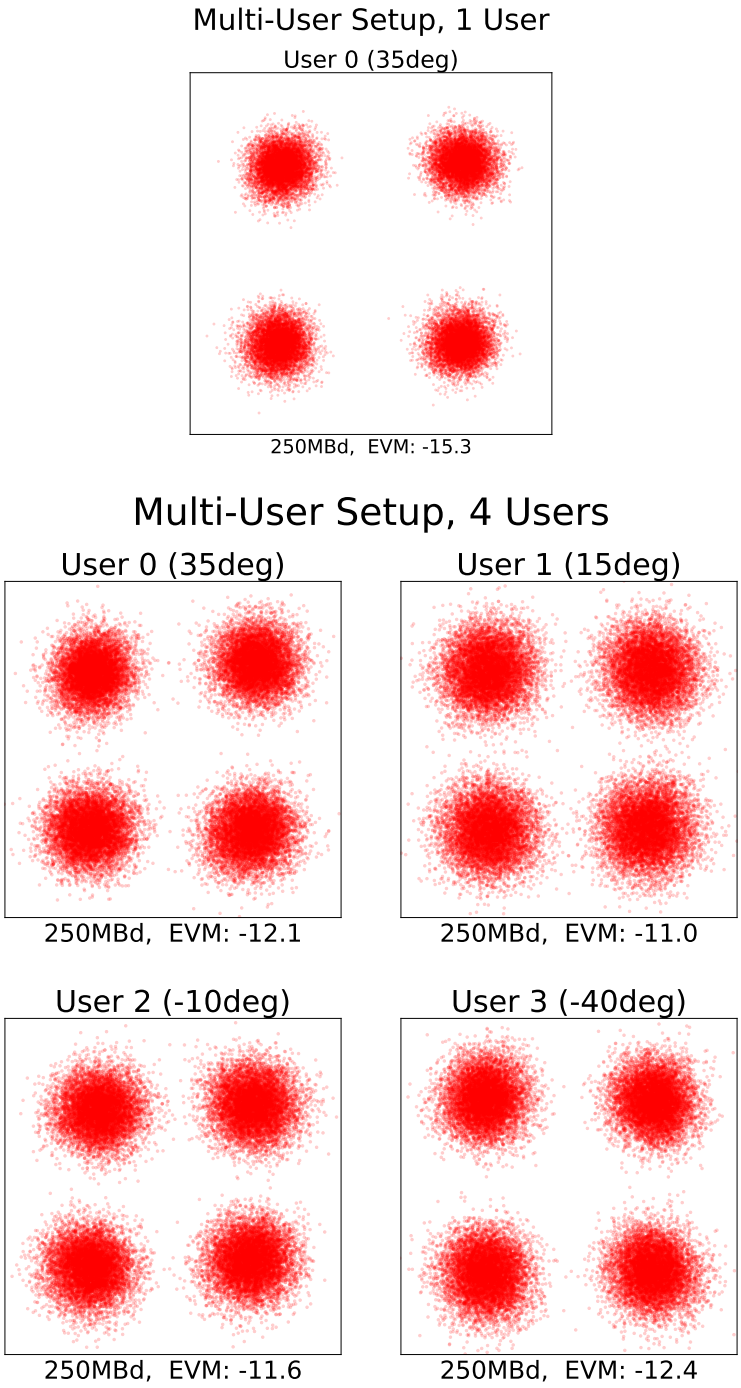

Figure 6.11: 73GHz carrier multi-user constellations. (Top) Multi-user setup with a single user on. (Bottom) Four users in the multi-user setup with zero-forcing on.

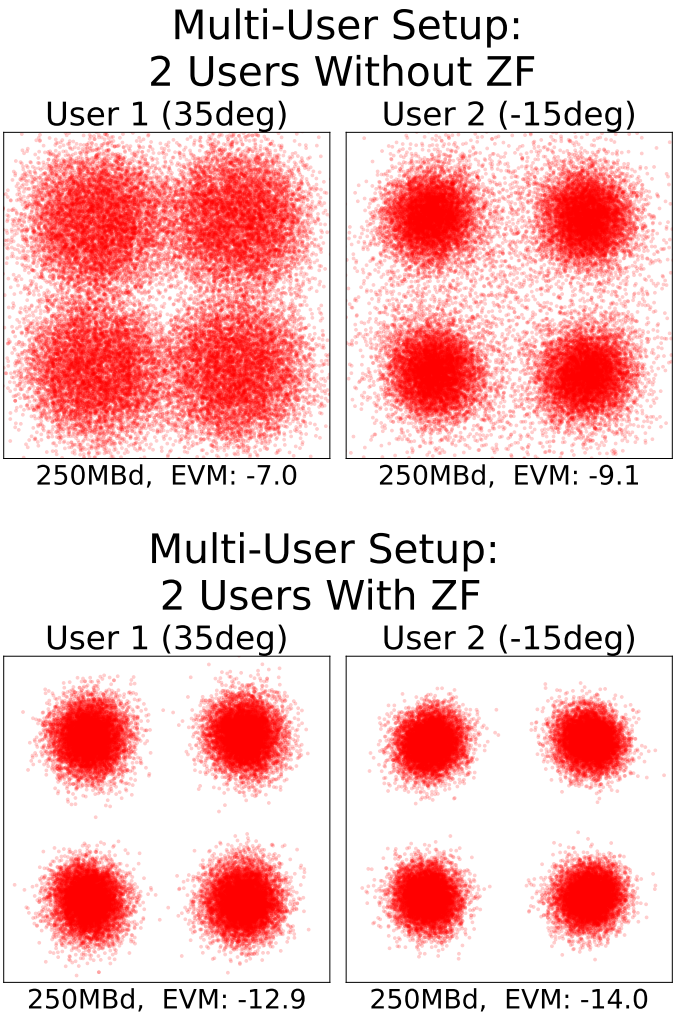

Figure 6.12: 73GHz carrier 2-user constellations, with and without zero forcing, showing the importance of inter-user interference mitigation.

at the same user, and only ±0.5dB EVM variation was measured across outputs. The remaining EVM variation shown in the constellation results is a result of TX power and distance variation. This together with single-user testing results shows the capability of the sub-array ASIC to serve more users and higher data rates in a full array system.

Multi-Panel Results

The multi-panel, multi-user test setup is shown in in Figure 6.13. Panels are daisy chained together, with each user time aligned and summed across panels. This provides fully-connected conjugate beamforming for each user, and increases SNR of each user from array gain.

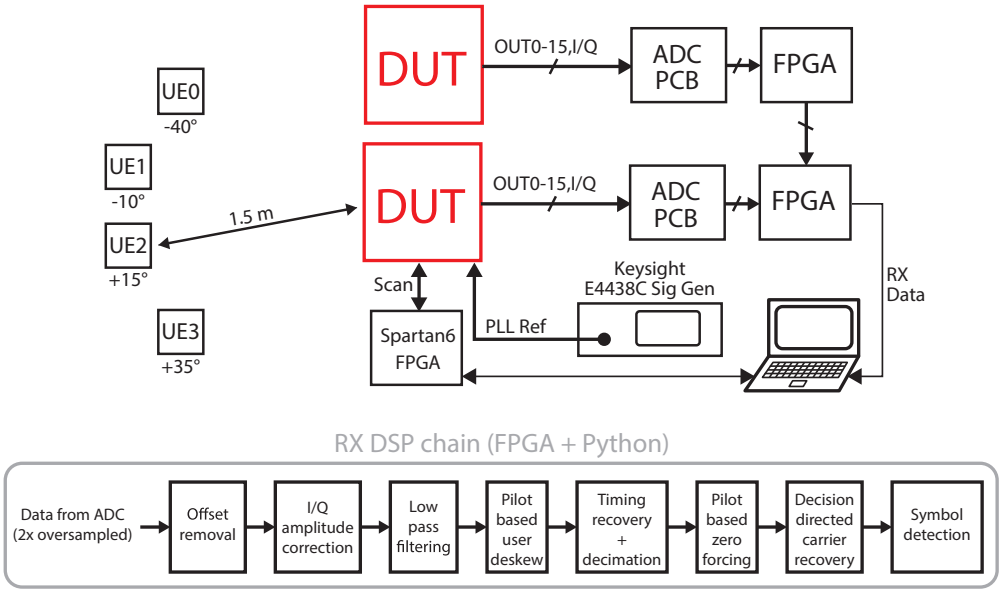

Figure 6.13: Measurement setup for a two panel, multi-user wireless link.

As discussed, multi-panel operation improves ZF operation. Shown in Figure 6.14, we see that two users with two panels improves the measured EVM. As the number of users is increased, the inter-user interference does reduce the EVM, but the two panel measurements outperform the single panel, as expected.

6.6 Comparison with State of the Art

This chapter presents the measured results of a fully packaged ASIC sub-array which enables efficient Two-Stage Architecture for Massive MU-MIMO at E-band. Results are collected in Table 6.1 and compared with state-of-the-art integrated 60-100GHz RX arrays. The sub-array is able to serve 16 beams on a single IC, which is to our knowledge the highest number reported at mm-wave. The ASIC demonstrated a maximum 2Gb/s per-user data rate, and while other papers show higher single user data rate they can only serve one or two users. This work demonstrated 4 simultaneous wireless user links, with the potential to serve up to 16 users enabling a many-user tiled array with $M \gg K$. The ASIC achieves wideband operation covering the entire 71-86GHz range. Finally, the power/antenna and power/antenna/user metrics, which are of key interest in massive MU-MIMO systems, are competitive with state-of-the-art and enable feasible power scaling for massive arrays.

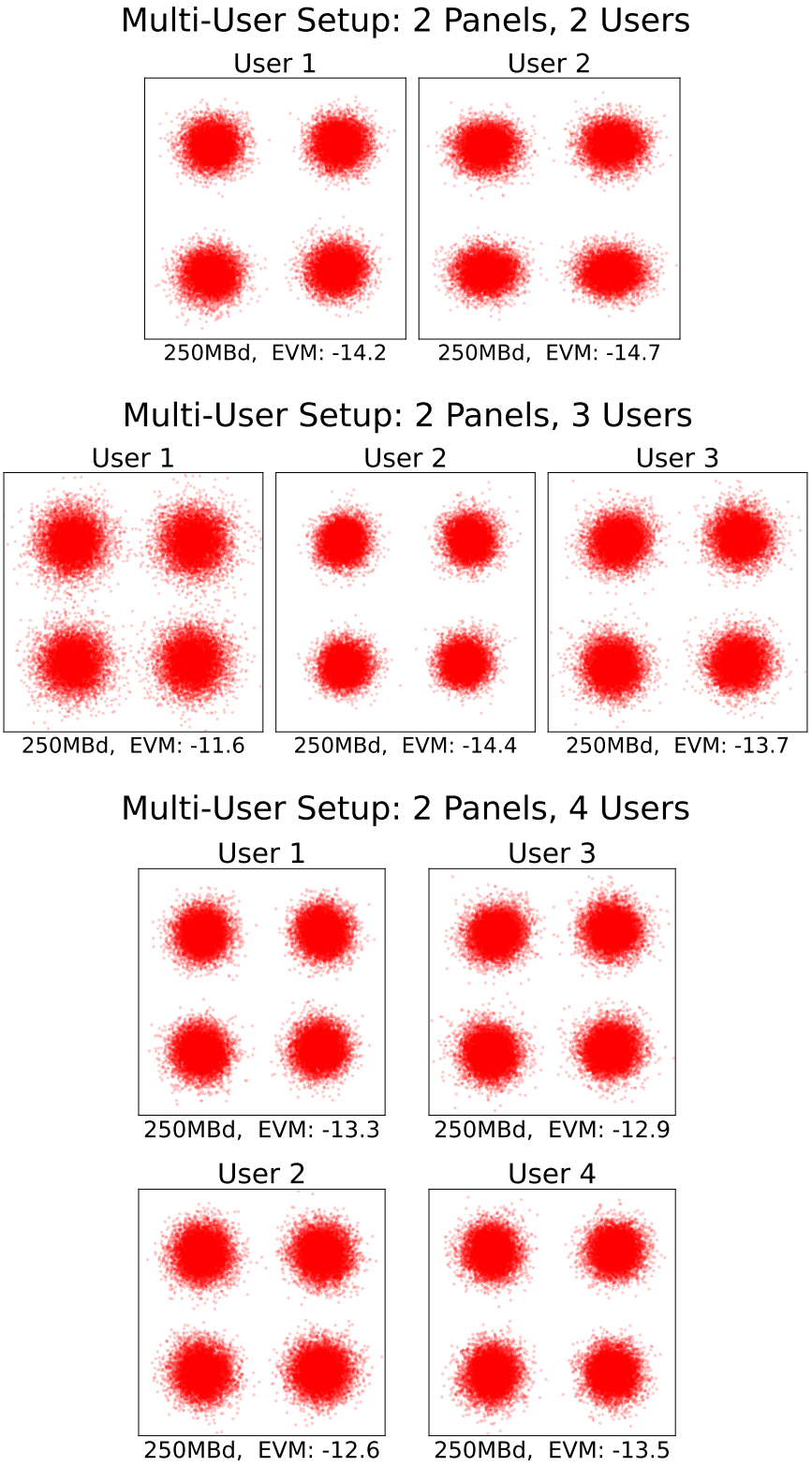

Figure 6.14: 73GHz carrier, two panel (32 antenna), multi-user constellations.

Table 6.1: Comparison with State of the Art.

	This Work	[31]	[29]	[32]	[33]	[34]
RX/TX	RX	RX/TX	RX/TX	RX/TX	RX/TX	RX/TX
Freq. (GHz)	71-86	80-100	57-66	71-76	57-71	57-66
Phase Shifter Type	BB	RF	BB	BB	RF	RF
No. RX Elements/IC	**16**	16	4	4	4	12
No. of Beams/IC	**16**	1	1	1	2	1
Total RX Power /IC* (mW)	1710	4300	508	672	355	550
RX Power /Ant. (mW)	107	269	127	168	88	46
RX Power /Ant./Beam (mW)	**7**	269	127	168	44	46
NF (dB)	9-11†	6.5-8	4.8-6.2	6	5.8	6.5
Single User Data Rate (Gbps)	2	10	7	7.2 (16 ICs)	32	4.6
RX EVM (dB)	-18	-22	-20	-24	-19	-22
Die Area (mm^2)	16	35.2	7.9	5.04	8.75	20.25
Process	28nm CMOS	0.18 μm BiCMOS	45 nm CMOS	22nm FinFET	28nm CMOS	28nm CMOS

* Includes LO circuits and BB output buffers

† Based on single-channel measurements of the RX front-end, plus 0.5dB simulated contribution of baseband circuits

Chapter 7

Conclusion

This dissertation has discussed the importance of wide-bandwidth, massive, multi-user MIMO arrays to meet the continually increasing capacity demands of wireless communication. However, the multiplicity of antennas and cross array signal transport lead to many practical problems such as packaging, noise, and distribution problems that can get quickly out of hand. Thus, implementing these large arrays is a systems problem. Designing them efficiently requires not just well designed individual blocks, but a system which conscientiously integrates and budgets specifications between these blocks.

7.1 Summary and Contributions

This dissertation began by summarizing phased array receivers, then outlining a range of beamforming architectures and their benefits and short comings. The Two-Stage Beamforming architecture was identified in particular for its flexible and power efficient scalability.

For multi-user and wide-bandwidth applications, the ADC was then identified as a key block when performing a comparison of architectures. The design requirements and specifications for an ADC were discussed and a model for ADC power using FOM and the number of bits in terms of M and K was identified and discussed.

A deeply hardware informed model of baseband amplifiers and analog and digital beamformers was designed to examine how cross array wiring, gain and bandwidth requirements of amplifiers, and the number of complex multiplications required scales with M and K. Results showed that baseband amplifier and beamfomer power is comparable for the analog and digital beamforming case, except for extremely large arrays where analog BF may scale better due to integration with the amplification stage. Both analog and digital beamforming models were optimistic in power estimation, but show interesting design insights as bandwidth and array size change.

The combination of the ADC and BF models showed that analog beamforming reduces ADC power, but this power savings is only significant for environments with very large

interferers, or very wide-bandwidth applications, such as those targeted in mmWave and beyond.

An ASIC was designed to show in practice the design lessons from the analog beamformer and baseband amplifier models on silicon. The chip shows the feasibility and power savings of many-user beamforming in analog. The design targeted an E-band RX with wide-bandwidth, 16 fully connected antenna beamforming for 16 users performed in baseband analog, and low power operation.

The 16 antenna, 16 user ASIC enabled a wireless demonstration of 2 receiver panels, serving 4 simultaneous users in a two-stage beamforming array.

7.2 Future Directions and Final Thoughts

Wide-bandwidth is an effective way to increase capacity, and as designers push transceivers to higher frequencies, the spectrum allows bandwidths up to several GHz. Much of the goal of this thesis is to emphasize that designing a transceiver capable of reaching higher carrier frequencies is not the only limitation in the system.

ADCs begin to become dominant in the signal chain power budget around 1GHz of bandwidth. Suggestions to reduce ADC resolution make high sampling frequencies more feasible, but the SNR degradation that comes with this low resolution is most power efficient if budgeted for in the LNA until over 100GHz.

Spatial and frequency interference mitigation saves power at wide-bandwidth by reducing the number of bits required in the ADC to support interferers. In addition, the most problematic source of interferes is a out-of-network base station transmitting a wide-band interferer. However, truly wide-bandwidth spatial notch filters with high rejection require very careful design, and only solve the problem for out-of-beam interferers. Published mm-wave spatial notch filters which do not reduce spatial resolution either do not report or demonstrate reduced interference rejection for wide-band interferers [72][59][73][58]. Thus, a wide-band, high-rejection spatial notch filter is likely an important direction for continued research.

Finally, analog beamformers reduce the ADC cost for large arrays with $\alpha > 1$ and wide bandwith or large interference requirements. However, wide bandwidth increases design effort as gain, bandwidth, and routing in each stage must be carefully managed.

All of this combined, suggests there is value in rethinking the traditional RF signal chain as the field pushes towards and beyond 1GHz of bandwidth. There is likely power savings to be had from breaking the signal chain into separate channels, either with separate base-stations and user-equipment, or with multiple signal chains. This is an interesting space for continued thought and research as we move towards wide-bandwidth wireless communication.

Bibliography

[1] L. Eliot, "The Autonomous Vehicular Cloud Is Steering Into View," *Forbes*, Mar 2021.

[2] S. Sowmyanarayan, "Return to business as unusual : Workplace of the future," tech. rep., Verizon, 2020.

[3] A. Viterbi, "Approaching the shannon limit: Theorist's dream and practitioner's challenge," in *Mobile and Personal Satellite Communications 2* (F. Vatalaro and F. Ananasso, eds.), (London), pp. 1–11, Springer London, 1996.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] A. Hajimiri, H. Hashemi, A. Natarajan, X. Guan, and A. Komijani, "Integrated phased array systems in silicon," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1637–1655, 2005.

[6] J. H. Winters, "On the Capacity of Radio Communication Systems with Diversity in a Rayleigh Fading Environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, 1987.

[7] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-blast: an architecture for realizing very high data rates over the rich-scattering wireless channel," *1998 URSI International Symposium on Signals, Systems, and Electronics. Conference Proceedings*, pp. 295–300, 1998.

[8] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[9] P. K. Bondyopadhyay, "The First Application of Array Antenna," *IEEE International Conference on Phased Array Systems and Technology*, pp. 29–32, 2000.

[10] R. L. Haupt and Y. Rahmat-samii, "Antenna Array Developments : A Perspective on the Past , Present and Future," *IEEE Antennas and Propagation*, vol. 57, no. 1, pp. 86–96, 2015.

[11] "Ieee standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements-part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 3: Enhancements for very high throughput in the 60 ghz band," *IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012)*, pp. 1–628, 2012.

[12] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[13] G. Foschini and M. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, no. 6, pp. 311–335, 1998.

[14] A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexing in modern MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 186–197, 2010.

[15] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 60–67, 2004.

[16] T. L. Marzetta, "How Much Training Is Required For Multiuser MIMO?," in *Asilomar Conference on Signals, Systems and Computers*, pp. 359–363, 2006.

[17] C.-c. Lin, C. Puglisi, and E. Ghaderi, "A 4-Element 800MHz-BW 29mW True-Time-Delay Spatial Signal Processor Enabling Fast Beam-Training with Data Communications," *2021 IEEE 47th European Solid State Circuits Conference*, pp. 287–290, 2021.

[18] J. Wang, Z. Lan, and K. Shuzo, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, 2009.

[19] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine Learning Paradigms for Next-Generation Wireless Networks," *IEEE Wireless Communications*, no. April, pp. 98–105, 2017.

[20] E. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.

[21] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, Nov. 2010.

[22] A. Puglielli, A. Townley, G. LaCaille, V. Milovanović, P. Lu, K. Trotskovsky, A. Whitcombe, N. Narevsky, G. Wright, T. Courtade, E. Alon, B. Nikolić, and A. M. Niknejad, "Design of energy- and cost-efficient massive mimo arrays," *Proceedings of the IEEE*, vol. 104, pp. 586–606, March 2016.

[23] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. K. Reynolds, Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J. E. Thillberg, L. Rexberg, M. Yeck, X. Gu, M. Ferriss, D. Liu, D. Friedman, and A. Valdes-Garcia, "A 28-GHz 32-Element TRX Phased-Array IC With Concurrent Dual-Polarized Operation and Orthogonal Phase and Gain Control for 5G Communications," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3373–3391, 2017.

[24] S. Mondal, R. Singh, A. I. Hussein, and J. Paramesh, "A 25–30 GHz Fully-Connected Hybrid Beamforming Receiver for MIMO Communication," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 5, pp. 1275–1287, 2018.

[25] G. LaCaille, A. Puglielli, E. Alon, B. Nikolic, and A. Niknejad, "Optimizing the LO Distribution Architecture of mm-Wave Massive MIMO Receivers," 2019. arXiv:1911.01339.

[26] L. Liang, W. Xu, and X. Dong, "Low-Complexity Hybrid Precoding in Massive Multiuser MIMO Systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, 2014.

[27] G. LaCaille, J. Dunn, A. Puglielli, L. Iotti, S. Ramakrishnan, L. Calderin, Z. Lin, E. Naviasky, B. Nikolic, A. Niknejad, and E. Alon, "Design and Demonstration of a Scalable Massive MIMO Uplink at E-Band," *2020 IEEE International Conference on Communications Workshops*, pp. 1–6, 2020.

[28] S. Mondal, L. R. Carley, and J. Paramesh, "4.4 a 28/37ghz scalable, reconfigurable multi-layer hybrid/digital mimo transceiver for tdd/fdd and full-duplex communication," *2020 IEEE International Solid- State Circuits Conference*, pp. 82–84, 2020.

[29] G. Mangraviti, K. Khalaf, Q. Shi, K. Vaesen, D. Guermandi, V. Giannini, S. Brebels, F. Frazzica, A. Bourdoux, C. Soens, W. Van Thillo, and P. Wambacq, "A 4-antenna-path beamforming transceiver for 60GHz multi-Gb/s communication in 28nm CMOS," *2016 IEEE International Solid-State Circuits Conference*, pp. 246–247, 2016.

[30] K. Kibaroglu, M. Sayginer, and G. M. Rebeiz, "A Low-Cost Scalable 32-Element 28-GHz Phased Array Transceiver for 5G Communication Links Based on a Beamformer Flip-Chip Unit Cell," *IEEE Journal of Solid-State Circuits*, vol. 53, pp. 1260–1274, May 2018.

[31] 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.

[32] S. Pellerano, S. Callender, W. Shin, Y. Wang, S. Kundu, A. Agrawal, P. Sagazio, B. Carlton, F. Sheikh, A. Amadjikpe, W. Lambert, D. S. Vemparala, M. Chakravorti, S. Suzuki, R. Flory, and C. Hull, "A Scalable 71-to-76GHz 64-Element Phased-Array Transceiver Module with 2×2 Direct-Conversion IC in 22nm FinFET CMOS Technology," *2019 IEEE International Solid-State Circuits Conference*, pp. 174–176, Feb. 2019.

[33] A. Chakrabarti, C. Thakkar, S. Yamada, D. Choudhury, J. Jaussi, and B. Casper, "A 64Gb/s 1.4pJ/b/element 60GHz 2×2-Element Phased-Array Receiver with 8b/symbol Polarization MIMO and Spatial Interference Tolerance," *2020 IEEE International Solid-State Circuits Conference*, pp. 84–86, 2020.

[34] T. Sowlati, S. Sarkar, B. G. Perumana, W. L. Chan, A. Papio Toda, B. Afshar, M. Boers, D. Shin, T. R. Mercer, W. H. Chen, A. Grau Besoli, S. Yoon, S. Kyriazidou, P. Yang, V. Aggarwal, N. Vakilian, D. Rozenblit, M. Kahrizi, J. Zhang, A. Wang, P. Sen, D. Murphy, A. Sajjadi, A. Mehrabani, E. Kornaros, K. Low, K. Kimura, V. Roussel, H. Xie, and V. Kodavati, "A 60-GHz 144-element phased-array transceiver for backhaul application," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3640–3659, 2018.

[35] R. Garg, G. Sharma, A. Binaie, S. Jain, S. Ahasan, A. Dascurcu, H. Krishnaswamy, and A. S. Natarajan, "A 28-GHz Beam-Space MIMO RX With Spatial Filtering and Frequency-Division Multiplexing-Based Single-Wire IF Interface," *IEEE Journal of Solid-State Circuits*, 2020.

[36] E. Naviasky, L. Iotti, G. LaCaille, B. Nikolić, E. Alon, and A. Niknejad, "A 71-to-86GHz Packaged 16-Element by 16-Beam Multi-User Beamforming Integrated Receiver in 28nm CMOS," *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, vol. 64, pp. 218–220, 2021.

[37] M. G. Anderson, S. Member, A. Thielens, S. Wielandt, A. Niknejad, J. Rabaey, and A. Theory, "Ultralow-Power Radio Frequency Beamformer Using Transmission-Line Transformers and Tunable Passives," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 158–160, 2019.

[38] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Communications Surveys and Tutorials*, vol. 20, pp. 3060–3097, Oct 2018.

[39] J. Mo and R. W. Heath, "Capacity Analysis of One-Bit Quantized MIMO Systems With Transmitter Channel State Information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, 2015.

[40] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput Analysis of Massive MIMO Uplink With Low-Resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 4038–4051, 2017.

[41] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive mimo systems: Digital versus hybrid analog-digital," *2014 IEEE Global Communications Conference*, pp. 4066–4071, 2014.

[42] M. N. Kulkarni, S. Member, A. Ghosh, and J. G. Andrews, "A Comparison of MIMO Techniques in Downlink Millimeter Wave Cellular Networks With Hybrid Beamforming," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1952–1967, 2016.

[43] C. Lin and G. Y. Li, "Energy-Efficient Design of Indoor mmWave and Sub-THz Systems with Antenna Arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4660–4672, 2016.

[44] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver comparison under energy vs spectral efficiency trade-off," *European Wireless 2017; 23th European Wireless Conference*, pp. 1–7, 2017.

[45] H. Yan, S. Ramesh, T. Gallagher, C. Ling, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," *IEEE Circuits and Systems Magazine*, vol. 19, pp. 33–58, Secondquarter 2019.

[46] W. Zhang, X. Xia, Y. Fu, and X. Bao, "Hybrid and Full-Digital Beamforming in mmWave Massive MIMO Systems : A Comparison Considering," *China Institute of Communications*, pp. 91–102, 2019.

[47] S. Dutta, S. Member, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter, S. Member, and S. Rangan, "A Case for Digital Beamforming at mmWave," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 756–770, 2020.

[48] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution adcs," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 8131–8146, Dec 2017.

[49] M. Sarajlić, S. Member, and L. Liu, "When Are Low Resolution ADCs Energy Efficient in Massive MIMO ?," *IEEE Access*, vol. 5, pp. 14837–14853, 2017.

[50] A. Molev-shteiman, X.-f. Qi, and B. M. Hochwald, "New Equivalent Model of a Quantizer With Noisy Input and Its Applications for MIMO System Analysis and Design," *IEEE Access*, vol. 8, 2020.

[51] S. Dutta, C. N. Barati, A. Dhananjay, and S. Rangan, "5G Millimeter Wave Cellular System Capacity with Fully Digital Beamforming," 2018. arXiv:1711.02586v2.

[52] P. Skrimponis, N. Hosseinzadeh, A. Khalili, E. Erkip, M. J. W. Rodwell, J. F. Buckwalter, and S. Rangan, "Towards energy efficient mobile wireless receivers above 100 ghz," *IEEE Access*, vol. 9, pp. 20704–20716, 2021.

[53] B. Murmann, "Adc performance survey 1997-2021," [Online]. Available: http://web.stanford.edu/ murmann/adcsurvey.html.

[54] L. Belostotski, S. Member, and E. A. M. Klumperink, "Figures of Merit for CMOS Low-Noise Amplifiers and Estimates for Their Theoretical Limits," *IEEE Transactions on Circuits and Systems II: Express Briefs*, no. Early Access, 2022.

[55] L. Belostotski, "Low-noise-amplifier (lna) performance survey.," [Online]. Available: https://schulich.ucalgary.ca/contacts/leo-belostotski.

[56] A. Natarajan, S. K. Reynolds, M.-D. Tsai, S. T. Nicolson, J.-H. C. Zhan, D. G. Kam, D. Liu, Y.-L. O. Huang, A. Valdes-Garcia, and B. A. Floyd, "A Fully-Integrated 16-Element Phased-Array Receiver in SiGe BiCMOS for 60-GHz Communications," *IEEE Journal of Solid-State Circuits*, vol. 46, pp. 1059–1075, May 2011.

[57] J. Lee, "G/T and Noise Figure of Active Array Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 2, pp. 241–244, 1993.

[58] M. Y. 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.

[59] L. Zhang, S. Member, and H. Krishnaswamy, "Arbitrary Analog / RF Spatial Filtering for Digital MIMO Receiver Arrays," *JSSC*, vol. 52, no. 12, pp. 3392–3404, 2017.

[60] C. Marcu, D. Chowdhury, C. Thakkar, J. Park, L. Kong, M. Tabesh, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. Zamani, E. Alon, and A. M. Niknejad, "A 90 nm cmos low-power 60 ghz transceiver with integrated baseband circuitry," *IEEE Journal of Solid-State Circuits*, vol. 44, pp. 3434–3447, Dec 2009.

[61] L. Iotti, S. Krishnamurthy, G. LaCaille, and A. M. Niknejad, "A Low-Power 70–100-GHz Mixer-First RX Leveraging Frequency-Translational Feedback," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 8, pp. 2043–2054, 2020.

[62] W. Shin, S. Callender, S. Pellerano, and C. Hull, "A compact 75 ghz lna with 20 db gain and 4 db noise figure in 22nm finfet cmos technology," *2018 IEEE Radio Frequency Integrated Circuits Symposium*, pp. 284–287, 2018.

[63] M. Vigilante and P. Reynaert, "On the Design of Wideband Transformer-Based Fourth Order Matching Networks for E-Band Receivers in 28-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, pp. 2071–2082, Aug. 2017.

[64] A. Musa, R. Murakami, T. Sato, W. Chaivipas, K. Okada, and A. Matsuzawa, "A Low Phase Noise Quadrature Injection Locked Frequency Synthesizer for MM-Wave Applications," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 11, pp. 2635–2649, 2011.

[65] B. Razavi, "TSPC Logic [A Circuit for All Seasons]," *IEEE Solid-State Circuits Magazine*, vol. 8, no. 4, pp. 10–13, 2016.

[66] L. Iotti, G. LaCaille, and A. M. Niknejad, "A 57–74-GHz Tail-Switching Injection-Locked Frequency Tripler in 28-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 115–118, 2019.

[67] B. Razavi, "The StrongARM Latch [A Circuit for All Seasons]," *IEEE Solid-State Circuits Magazine*, vol. 7, no. 2, pp. 12–17, 2015.

[68] A. Townley, P. Swirhun, D. Titz, A. Bisognin, F. Gianesello, R. Pilard, C. Luxey, and A. M. Niknejad, "A 94-GHz 4TX–4RX Phased-Array FMCW Radar Transceiver With Antenna-in-Package," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 5, pp. 1245–1259, 2017.

[69] R. Frye, S. Kapur, and R. Melville, "A 2-GHz quadrature hybrid implemented in CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 550–555, Mar 2003.

[70] H. Hashemi, X. Guan, A. Komijani, and A. Hajimiri, "A 24-GHz SiGe phased-array receiver-LO phase-shifting approach," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 2, pp. 614–626, 2005.

[71] U. Mengali and A. D'Andrea, *Synchronization techniques for digital receivers*. Springer Science & Business Media, 2013.

[72] L. Zhang, A. Natarajan, and H. Krishnaswamy, "Scalable Spatial Notch Suppression in Spatio-Spectral-Filtering MIMO Receiver Arrays for Digital Beamforming," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 12, pp. 3152–3166, 2016.

[73] R. W. Irazoqui and C. J. Fulton, "Spatial Interference Nulling before RF Frontend for Fully Digital Phased Arrays," *IEEE Access*, vol. 7, pp. 151261–151272, 2019.