An on-off receiver array for low-power scaling of mmWave massive MIMO

Maryam Eslami Rasekh*, Navid Hosseinzadeh†, Upamanyu Madhow‡, Mark Rodwell§
Department of Electrical and Computer Engineering
University of California Santa Barbara
Email: {*rasekh, †navid, ‡madhow, §rodwell}@ece.ucsb.edu

2021 IEEE Workshop on Signal Processing Systems (SiPS) | 978-1-6654-0144-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/SiPS52927.2021.00041

***Abstract*—Low-power solutions are crucial for realizing the throughput gains of mmWave massive MIMO in base station to mobile links, particularly for the battery-powered handset. LNA power consumption is a significant bottleneck in scaling to large arrays at high frequencies. We propose a simplified phased array architecture for a 140 GHz receiver that uses activation switches in place of noisy phase shifters, allowing the relaxation of LNA gain requirements and, consequently, power consumption. On-off beamforming is employed, which introduces a tradeoff between aperture utilization and power efficiency. Our results show the potential for 7X power savings with approximately 40% average utilization of die area.**

***Index Terms*—Low-power frontends, millimeter wave, Terahertz, massive MIMO, phased array antennas, analog beamforming, on-off beamforming.**

I. Introduction

The wireless data crunch is driving mass migration to higher frequency bands in the millimeter wave (mmWave) and Terahertz (THz) range, where large available bandwidth and immense potential for spatial multiplexing can provide the orders of magnitude increase in throughput necessary for meeting this ever growing demand [1]. At high frequencies, electronically large antenna arrays are required to overcome path loss and, in some cases, atmospheric absorption losses in the channel. With a fixed 2-dimensional die area (aperture size) and half wavelength spacing, the number of array elements scales quadratically with carrier frequency, and transceivers with hundreds or even thousands of elements are feasible in these bands. A fully digital or hybrid array architecture may be more suitable for the base station array to allow spatial multiplexing between different mobile users, whereas analog beamforming with a phased array is sufficient for a handset forming a beam towards the base station.

Power consumption is a major bottleneck in the deployment of THz frontends as the number of array elements grows large [2], [3]. This problem is especially pronounced in the battery-powered handset, which is our focus in this paper. Two possible architectures for the handset receiver are shown in Figure 1. In a conventional phased array architecture, the signal of each antenna goes through a phase shifter (passive or active) and low noise amplifier (LNA) to provide sufficient SNR in each channel before summation in the combiner. The amount of power burned in an LNA increases with its gain. If high gain is required in the LNA (e.g., to compensate for a noisy phase shifter), this consumption quickly adds up as the array grows large. In this paper, we propose an alternative architecture that allows significant reduction in the required LNA gain, providing an order of magnitude decrease in per-channel power consumption relative to the state of the art for 140 GHz phased array receivers [4], [5]. In the proposed architecture, phase shifters are replaced with a simple switch on each channel that can turn that channel ON or OFF, without the possibility of tuning its phase, as shown in Figure 1b. We argue that by taking advantage of the abundance of antennas, sufficient beamforming gains for closing the link can be achieved by activating only a *subset* of channels whose signals are sufficiently aligned in phase. Thus, we invoke a core principle of massive MIMO which allows relaxing per-element sophistication by leveraging scale.

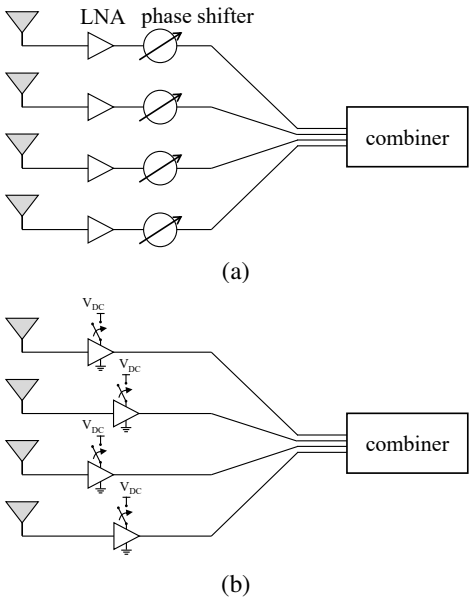

Fig. 1: (a) conventional phased array architecture, and (b) proposed switched architecture with on-off beamforming.

Contributions: We propose a low-power on-off array architecture for reducing power consumption of phased array receivers at the cost of hardware redundancy and under-utilization of antenna die area. We provide a coarse comparison of power consumption in our architecture with that of a conventional

978-1-6654-0144-9/21/$31.00 ©2021 IEEE
DOI 10.1109/SiPS52927.2021.00041

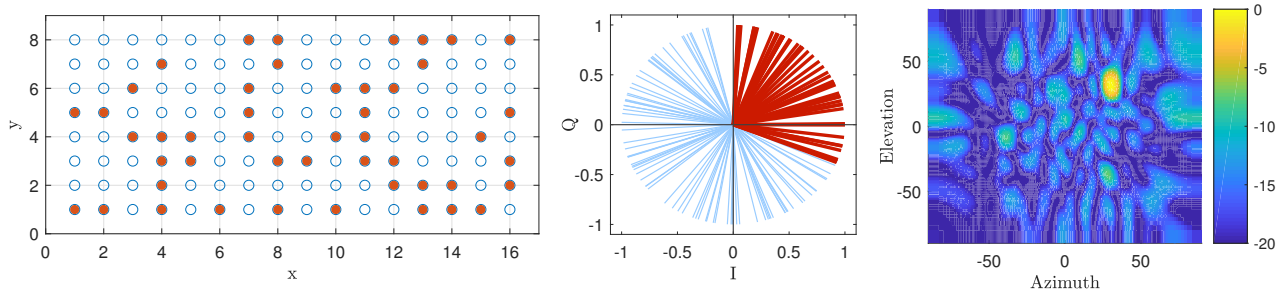

Fig. 2: Example of on-off beamforming toward 30° azimuth and 30° elevation in a 16×8 element rectangular array. Active elements are marked in red (left), and corresponding phasors of inactive and active channels are depicted in the I-Q plane, in blue and red, respectively (center). The resulting radiation pattern is depicted in the azimuth-elevation plane (right).

phased array, as well as analytical and numerical assessment of the performance of this scheme, laying out the inherent power-redundancy tradeoff that it entails. Our calculations predict that 7X power savings can be attained with 40% utilization of die area or, equivalently, 2.5X hardware redundancy.

II. System model

A conventional phased array architecture is depicted in Figure 1a. The signal of each antenna goes through a low noise amplifier and phase shifter. The phase-aligned signals are added in the combiner. The phase shifters may be *active* or *passive*. Passive phase shifters consume no power themselves but impose attenuation on the signal, placing more stringent gain and noise figure requirements on the LNA. Active phase shifters, on the other hand, offload part of the amplification requirements from the LNA and, consequently, reduce LNA wattage. However, as active components, they consume additional power. This problem is accentuated at higher mmWave and THz frequencies as building suitable RF phase shifters becomes more challenging in these bands. To date, no CMOS prototypes of low-noise active phase shifters, or low-attenuation and hence low-noise *passive* phase shifters, have been demonstrated at 140 GHz. Most existing designs downconvert the RF signal to an intermediate frequency (IF) band where phase correction and combining is performed [4]–[6]. This approach requires incorporating an additional mixer and LO driver *on each channel* which further adds to the per-channel power consumption of the phased array. Thus, removing the phase shifter altogether has the potential to reduce power consumption significantly – by an order of magnitude or more – both directly by eliminating active components, and indirectly by relaxing the gain and noise figure requirements of the LNA.

We note that in our analysis, we exclude the power consumption of components downstream of the combiner, i.e., "shared" devices, and only account for the power consumption of components that are duplicated for each channel. Since we do not assume a specific array size, this approach simplifies our analysis and facilitates an informative comparison between the different architectures. Thus, by "per-channel power consumption" we refer to the consumption of LNAs, phase shifters, RF to IF converters, and any block that the signal passes *before* arriving at the combiner.

A. A Low-Power Architecture

Our proposed architecture, depicted in Figure 1b, replaces the phase shifter in each channel with a simple switch that, in the ON position, activates the channel and allows its signal to propagate to the combiner, and deactivates it when turned OFF. Deactivated channels draw near zero current while active channels consume a small amount of power, relative to a conventional phased array channel, enough to provide the gain required to overcome the loss of the combiner.

In the designs published to date at 140 GHz, the LNA typically provides 17 to 20+ dB of gain and burns around 30 to 60+ mW of DC power. The mixer and IF beamformer consume around 14 mW and 30 mW, respectively, bringing the overall per-channel consumption to approximately 76-100+ mW [4], [5].[1] By avoiding the phase shifter, only a few dB of gain is required, reducing LNA power consumption to below 10 mW. The power consumption per active channel is therefore an order of magnitude lower than existing designs. Since this architecture frees up the budget for increasing the number of channels, the overall consumption per active channel would be even lower.

Thus, we predict that with the on-off configuration, the per-channel power consumption is reduced by a factor of 10X or more, but due to the lack of phase control, a portion of the power of each active channel is effectively "wasted" due to imperfect phase alignment. In the following section,

[1]These numbers do not include the power consumption of the LO that drives the IF mixer, as the values reported in these references were lumped up with that of the LO multiplier, which is a shared component. With the optimistic assumption of only a few mW consumption by each channel's LO driver, the state of the art design can be assumed to burn 80+ mW per channel (excluding consumption of shared components).

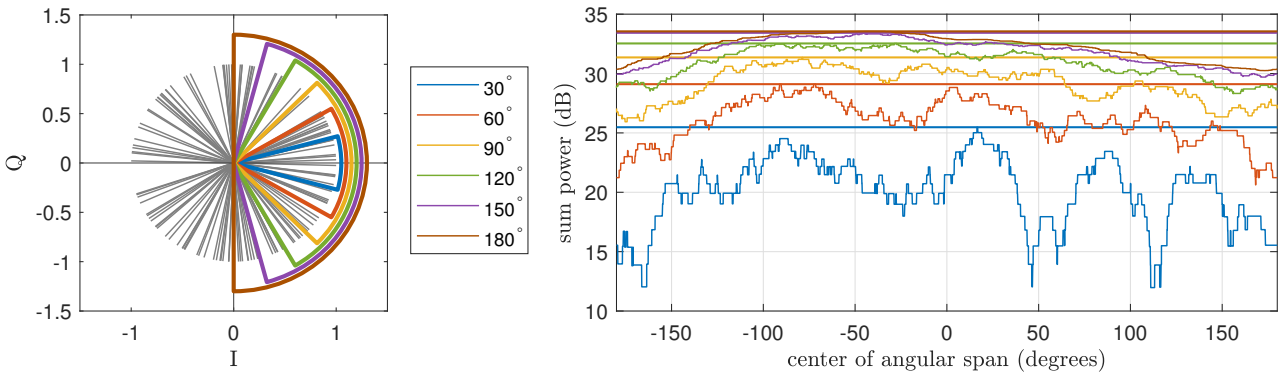

Fig. 3: Typical example of phasor distribution and output power with different choices for angular span of phasors activated. The maximum possible output for each span is marked by a solid horizontal line. Extending the active span beyond 120° provides negligible gain in output power.

we discuss the beamforming procedure and gain-efficiency tradeoffs in the on-off architecture.

B. On-Off Beamforming in the Switched Array

Without phase shifters, it is not possible to align all of the channels to maximize array gain. In this setup, "beamforming" constitutes choosing the switch position of each antenna such that the maximum number of channels add up constructively. Since some elements are switched off, and the active channel phasors are not perfectly aligned, the overall beamforming gain provided by an N element on-off receiver is a fraction of N. Thus, this architecture forces a tradeoff between die area utilization (and hardware redundancy) and power efficiency: We can get away with switches instead of phase shifters at the cost of leaving a portion of our aperture area unsampled.

We assume here that each channel includes a constant but random phase shift in its path to the combiner. This will most likely occur naturally due to variability in the path lengths and layout of the RF circuits, but it can also be imposed by design. As a result, the phasors arriving at the combiner from the N channels – which include the circuit path phase shifts plus the phases of the channel response on the array – are randomly distributed. This is shown for a typical example in Figure 2. The actual phasor directions vary with the angle of arrival, but their distribution remains uniformly random on the unit circle. In order to beamform toward a given direction, we look at the positions of these phasors for that angle of arrival and activate the largest possible set of channels whose signals add up with sufficient coherence. One possible choice is shown in Figure 2 for a sample pointing direction along with the resulting beamforming pattern.

In the next section, we provide analytical insight into the beamforming capacity of this architecture and the tradeoff between gain and power efficiency. We also discuss some of the open questions and practical challenges that we hope to address in the future.

III. Performance Analysis and Numerical Results

On-off beamforming involves selecting the switch position of each antenna for maximizing the output signal power of the combiner. Assuming the wireless channel response and circuit path offsets are known, the phasor of the RF signal of each antenna as it arrives at the combiner can be calculated. In order to direct the beam toward the desired channel, we select a portion of these phasors that are "sufficiently" aligned and activate the corresponding antennas. This selection may be tweaked to steer a null toward an undesired interferer or jammer (we leave a detailed analysis of nullforming to future work). Increasing the number of activated channels increases the power taken in by the array, but it also diminishes the coherence between the active phasors. This brings about a tradeoff between gain and efficiency which we discuss below.

A. The gain-efficiency tradeoff

The angular range of phasors activated for beamforming decides a tradeoff between the total beamforming gain and power efficiency, i.e., beamforming gain per unit of power burned. Clearly, all activated phasors must be within one half-plane in I-Q space, but phasor alignment and power efficiency can be improved by reducing the maximum angular divergence of activated phasors. In Figure 3, the diminishing returns of expanding the angular range beyond 120° are clearly shown.

In order to quantify our activation efficiency, we define the contribution of each channel, or the "per-channel gain" (PCG) as the normalized inner product between its phasor and the sum of all other active phasors. This value captures the effective benefit we get from burning power to activate that channel, conditioned on the set of already active channels. Consequently, the *minimum* per-channel gain (MPCG) over all active channels is a suitable criterion for the efficiency

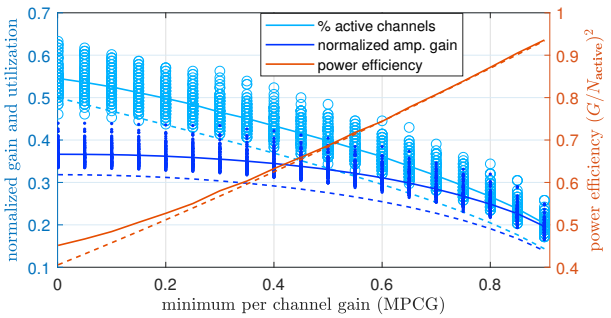

Fig. 4: Left axis: amplitude of on-off beamformed signal (as a fraction of N) and fraction of active antennas as a function of efficiency criterion, MPCG. Right axis: power efficiency, i.e., ratio of beamformed power to that of a standard phased array with the same number of active antennas. Solid lines are numerical results averaged over 100 realizations, dashed lines are theoretical lower bounds.

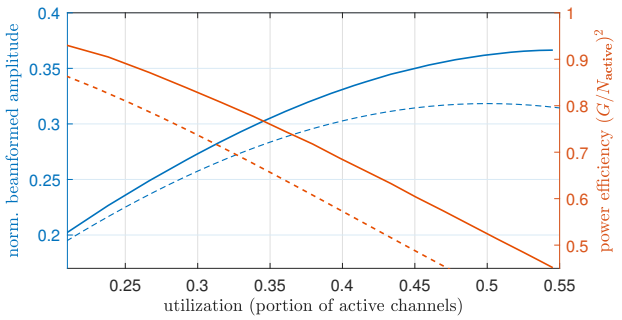

Fig. 5: The efficiency-utilization tradeoff. Left axis: beamforming gain, and right axis: power efficiency, as a function of aperture utility (portion of active channels). Numerical means and theoretical lower bounds depicted by solid and dashed lines, respectively.

of switched beamforming. MPCG of ϵ can be translated to angular span ψ of activated phasors using the formula,

$$\epsilon = \cos\frac{\psi}{2}, \qquad \psi = 2\cos^{-1}\epsilon, \tag{1}$$

since any phasor within $[-\frac{\psi}{2}, \frac{\psi}{2}]$ of the sum phasor provides a contribution larger than or equal to $\epsilon = \cos\frac{\psi}{2}$ to the sum phasor. With the pessimistic assumption that the angular span is chosen uniformly at random, i.e., without trying to focus on the area with highest phasor density, the statistical average of the number of active elements, N_{active}, and sum of added phasors, G, is given by

$$N_{\text{active}} = \frac{\psi}{2\pi}N, \tag{2}$$

$$G = \frac{N}{2\pi}\int_{-\frac{\psi}{2}}^{\frac{\psi}{2}} \cos\varphi d\varphi = \frac{\sin\frac{\psi}{2}}{\pi}N, \tag{3}$$

bringing the average power efficiency of the scheme to

$$\eta = \left(\frac{G}{N_{\text{active}}}\right)^2 = \text{sinc}^2\frac{\psi}{2\pi}. \tag{4}$$

Figure 4 depicts numerical results for the variation of beamforming gain and power efficiency as functions of MPCG. This figure shows that, as MPCG increases from 0 up to around 0.5, there is very little loss in the overall beamforming gain but significant improvement in power efficiency. Thus we choose this threshold for beamforming and performance analysis in our system. With MPCG $= 0.5$, we have $\psi = 2\pi/2$ by (1). That is, all active phasors are guaranteed to reside within a 120° span in I-Q space. Although phasors are distributed uniformly, the selection process favors areas with higher density than average[2], and we may conclude from (2) that *at least* $1/3$ of antennas are active. A *lower bound* for the combined phasor amplitude is calculated by the statistical average described in (3) as follows:

$$G = \frac{N}{2\pi}\int_{-\frac{\pi}{3}}^{\frac{\pi}{3}} \cos\varphi d\varphi = \frac{\sqrt{3}}{2\pi}N \approx 0.28N. \tag{5}$$

Thus, by setting MPCG $= 0.5$, we activate at least $1/3$ of the antennas, the effective utilization of the antenna aperture (or die area) is greater than 0.28, and the effective power efficiency is at least $(0.28\times 3)^2 \approx 70\%$. From the results shown in Figure 4 for a 128 element array (arranged as 16×8), we see that, in fact, 40% of channels are active on average, and the mean normalized gain is 0.34, providing the same power efficiency as the analytical prediction, 70%. Note that without setting a threshold for MPCG, this efficiency would be around 40%.

Thus, with a choice of MPCG = 0.5, the switched architecture provides a factor of $10 \times 0.7 = 7$X reduction in power consumption relative to a phased array (for providing the same beamforming gain) at the cost of 2.5X hardware redundancy and increase in die area. The full tradeoff curve between power efficiency and hardware utilization is depicted in Figure 5. The switched architecture is, therefore, a promising approach for low-power scaling of the handset array. In order to effectively deploy such a system, however, several issues in signal processing and hardware must be addressed.

B. On-off nullforming for interference suppression

Under-utilization of the aperture in the manner shown in Figure 2 does not manifest as a wider beam or lower angular resolution, since active elements are scattered across the entire aperture; rather, it results in *higher side lobe levels*. In fact, as shown in Figure 6, a 128 element on-off array (with 47 elements active, in this instant) has the same beamwidth as a fully sampled 128 element array, but this results in higher energy outside the main beam, even relative to a fully sampled 47 element array. Thus, if the array is required to suppress interference from undesired sources in the environment, active *nullforming* is likely to be necessary.

[2]In this paper, our focus is on the performance outcome of optimal beamforming. The algorithmic details of finding the best possible phasor selection are not included here. We note, however, that this task can be done with $\mathcal{O}(N)$ computational complexity, and only needs to be repeated on the time scale of channel coherence.

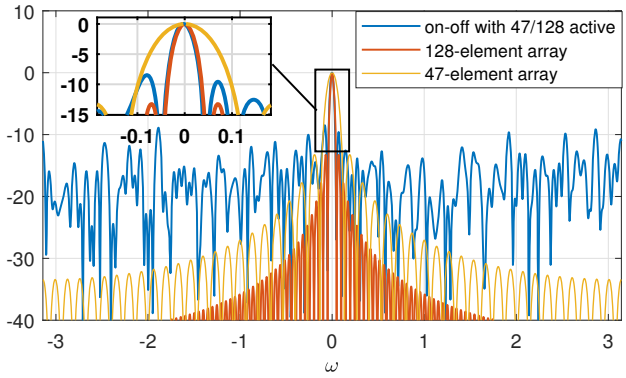

Fig. 6: Normalized radiation pattern of a linear 128-element on-off array with 47 active channels, alongside that of standard 128- and 47-element phased arrays. Patterns are plotted as a function of normalized spatial frequency in radians. We observe that the main beam tracks a 128 element array closely because of equal aperture size, but side lobe levels are, on average, higher than the 47-element array: with equal radiated power (number of active elements), a narrower main beam results in more power spread onto the sidelobes.

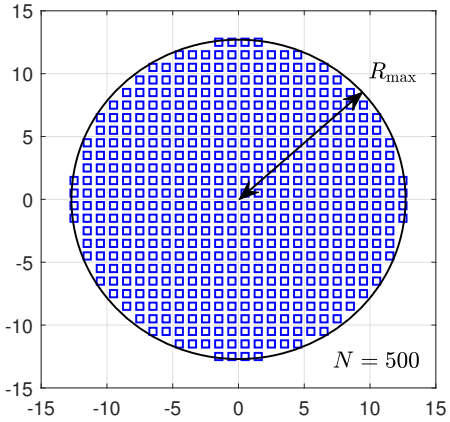

Fig. 7: Number of elements in 2D array limited by maximum length of transmission line to chip (center). At 0.1 dB/mm, maximum allowable radius for $\leq$ 1.5 dB loss is around 13 mm, corresponding to around 500 elements at 140 GHz assuming (near) half-wavelength spacing between elements in each dimension.

Nullforming is a straightforward task in conventional phased arrays where beamforming weights can be tuned to project onto the nullspace of the interferer's channel, with little cost to beamforming gain (as long as the desired and undesired sources are not too close in angular domain). When limited to switches instead of phase shifters, however, conventional techniques are not applicable. The problem of null steering in on-off arrays is therefore an interesting open problem that we plan to address in future work.

C. Calibration and channel estimation

A prerequisite for any beam or null formation, of course is knowledge of the phasor response from each antenna. In order to determine these phasors, we not only need to know the channel from the transmitter to the array (channel estimation), but we also require the relative phase offsets induced by the signal chain of each channel (calibration). Assuming the calibration offsets are known, compressive channel estimation techniques developed in prior work can easily be modified for the on-off architecture. The conventional Newtonized orthogonal matching pursuit (NOMP) algorithm [7] can be applied if the link maintains coherence between switch updates, otherwise a noncoherent scheme may be adapted for this purpose [8], [9].

Calibrating the array in a *controlled* environment (i.e., inside an anechoic chamber where the channel vector is known) and *coherent* measurements is fairly straightforward: After taking $K \geq N$ measurements of the channel with different (random) switch positions, we solve (or find the MMSE fit to) the system of linear equations to obtain the overall array response and, thereby, calibration coefficients. It may, however, be desirable to calibrate arrays *in the field* with unknown channels and, more importantly, without relying on phase coherence across measurements. This is also an open issue that will be the focus of future work.

D. The limits of scaling

Assuming the above mentioned challenges in signal processing are met, the switched beamforming architecture is a promising approach to low-power scaling of arrays in size and frequency; but how far can this architecture be pushed in practice? We consider a 2D circular array to understand the practical limits of scaling. As the array expands, the length of the paths that route outer elements to the combiner becomes large, and the propagation loss of these paths becomes a bottleneck. Our electromagnetic simulations of a high performance laminate circuit board show that at 140 GHz, the transmission line exhibits propagation loss of 0.1 dB/mm. In order to ensure routing loss below 1.5 dB, a maximum radius of 15 mm is imposed on the array which, assuming half-wavelength spacing between elements, limits the number of antennas that can efficiently be connected to one RF combiner (around 500 elements for our case study). Interestingly, the loss per unit distance tends to scale linearly with carrier frequency, f. Therefore, as we scale up in frequency, the allowable array diameter reduces linearly (and its area quadratically) with f. Since the half-wavelength element spacing *reduces* proportionally to $1/f$, the maximum allowable number of elements in the phased array remains relatively constant as we move higher in the spectrum! This imposes a natural, fixed "tile size" on phased array antennas, unless a more complex hierarchical combining strategy is employed. Increasing the number of elements beyond this limit can also be done via a modular approach of cascading, or "tiling", multiple on-

off arrays that are synchronized to emulate a larger array. A tiled architecture with a separate RF chain on each tile would allow *hybrid* beamforming with the possibility of spatial multiplexing and greater capability for beam and null steering.

IV. Conclusions and Future work

In this paper, we demonstrate a scalable, low-power architecture for mmWave massive MIMO frontends that utilizes on-off switches in place of noisy phase shifters for each array element. We show that this architecture reduces per-channel power consumption by an order of magnitude, and is a promising approach for scaling to large arrays and higher frequencies. These power savings come at the cost of under-utilization of the array aperture and, consequently, die area. We observe that approximately 40% of antennas are active for efficient beamforming, meaning 2.5X hardware redundancy is needed to provide the desired beamforming gain relative to a conventional phased array. On each channel, the switched architecture provides 10X reduction in power consumption, which is reduced to 7X savings in *beamformed* power consumption due to the reduced power efficiency of on-off beamforming. This efficiency can be increased by placing more stringent requirements on the phase alignment of active channels, which would result in a smaller percentage of channels being activated at any instant, i.e., lower die utilization and greater hardware redundancy.

In future work, we plan to develop efficient algorithms for calibration and channel estimation in the on-off array, and to extend the beamforming strategy to incorporate nullforming for interferer rejection. An interesting hardware design question is whether binary phase shifting, i.e., switching between beamforming weights of ± 1, can be implemented at comparable power consumption; this would double the effective number of antennas used for beamforming and increase aperture utilization by a factor of 2.

V. Acknowledgments

This work was supported in part by ComSenTer, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

References

[1] M. J. Rodwell, "100-340GHz spatially multiplexed communications: IC, transceiver, and link design," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[2] P. Skrimponis, S. Dutta, M. Mezzavilla, S. Rangan, S. H. Mirfarshbafan, C. Studer, J. Buckwalter, and M. Rodwell, "Power consumption analysis for mobile mmwave and sub-THz receivers," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. IEEE, 2020, pp. 1–5.

[3] P. Skrimponis, N. Hosseinzadeh, A. Khalili, E. Erkip, M. J. Rodwell, J. F. Buckwalter, and S. Rangan, "Towards energy efficient mobile wireless receivers above 100 GHz," *IEEE Access*, vol. 9, pp. 20 704–20 716, 2020.

[4] S. Li and G. M. Rebeiz, "A 134-149 GHz IF beamforming phased-array receiver channel with 6.4-7.5 dB NF using CMOS 45nm RFSOI," in *2020 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. IEEE, 2020, pp. 103–106.

[5] S. Nicolson, A. Tomkins, K. Tang, A. Cathelin, D. Belot, and S. Voinigescu, "A 1.2 V, 140GHz receiver with on-die antenna in 65nm CMOS," in *2008 IEEE Radio Frequency Integrated Circuits Symposium*. IEEE, 2008, pp. 229–232.

[6] T. Heller, E. Cohen, and E. Socher, "A 102–129-GHz 39-dB gain 8.4-dB noise figure I/Q receiver frontend in 28-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 5, pp. 1535–1543, 2016.

[7] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Frequency estimation for a mixture of sinusoids: A near-optimal sequential approach," in *3rd IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015.

[8] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*, 2017, pp. 13–18.

[9] M. E. Rasekh and U. Madhow, "Noncoherent compressive channel estimation for mm-wave massive MIMO," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 889–894.