Energy-efficient tri-hybrid precoding with dynamic metasurface antennas

Miguel R. Castellanos, J. Carlson, and Robert W. Heath, Jr.
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27606
{mrcastel, jmcarls3, rwheathjr}@ncsu.edu

***Abstract*—Dynamic metasurface antennas (DMA) are smart leaky-wave antennas that enable low-power beamforming, albeit with particular constraints. In this paper, we develop algorithms to configure a tri-hybrid architecture that combines digital, analog, and DMA precoding. Our optimization prioritizes low power consumption to determine the most energy-efficient configurations of precoding weights. Further, we integrate the unique characteristics of the DMA architecture, such as the limited beamforming weight distribution and waveguide propagation, to generate jointly optimized precoding weights for a realistic DMA design. The key metrics of analysis are spectral and energy efficiency results to determine tradeoffs in performance between increased data rates and power consumption.**

I. Introduction

Large arrays are a key feature of modern communication systems but can be challenging to implement due to hardware limitations. Visions of future wireless systems for 5G and beyond leverage large antenna arrays in the sub-6 GHz band, the millimeter wave band, and even the upper mid-band [1]. One of the key limitations of large arrays, however, is the high power consumption at both the base stations and user devices [2]. The power consumption from components such as data converters, analog phase shifters, and even power amplifiers becomes overwhelming as arrays become larger. Power-efficient methods for scaling antenna arrays remains an important challenging in developing future wireless systems.

Dynamic metasurface antennas (DMAs) are low-power reconfigurable devices. The reconfigurable components are integrated into the individual antenna elements to dynamically tune the resonant frequency [3]. As the resonant frequency dictates the amplitude and phase of the DMA element radiation, it is possible to tune the DMA elements to establish a desired array gain pattern. Therefore, DMAs enable beamforming through the tuning of reconfigurable components, similar to analog phase shifters but with lower power consumption. Viewed in a different way, the elements in a DMA act like a reconfigurable antenna. The elements in a DMA may be configured in a variety of ways using, for example, varactor diodes to vary the capacitance of the element [3], [4] or PIN diodes to switch the elements off and on [5]. Overall, DMA arrays are flexible, compact, and consume less power than traditional analog phased arrays.

The full integration of DMAs into a MIMO system remains an open challenge. Prior work has analyzed the use of DMAs in place of analog precoding in the context of uplink massive MIMO [6] and energy efficiency maximization [7]. Additional studies with DMAs regarding channel estimation [8], near-field communication [9], MIMO-OFDM [10], and millimeter-wave communication [11] include a digital precoder to form a hybrid precoding architecture, where the digital and DMA analog precoders are jointly optimized. Going one step further, to our knowledge there is currently no work that studies a three-stage (tri-hybrid) precoding architecture involving digital precoding, analog precoding with phase shifters, and analog precoding with DMAs. Moreover, the prior work on DMA signal processing in [6]–[11] uses a simplified model for DMA tuning and lacks the realistic DMA constraints present in more rigorous DMA models [12], [13]. We aim to investigate the benefits of a tri-hybrid architecture with DMA precoding that incorporates a practical DMA design.

In this paper, we propose the *tri-hybrid* architecture that incorporates digital, analog, and DMA-based precoding. First, we formulate the signal model for a partially-connected tri-hybrid architecture and specify the different constraints for each signal processing component. We then propose a method for optimizing the tri-hybrid precoder using a codebook-based approach for DMA beamforming. A combinatorial search is used to find the best DMA configuration that maximizes the spectral efficiency, and the analog and digital precoders are then optimized based on the post-DMA-beamformed channel. We compare the power consumption and energy efficiency of three architectures: fully-digital, analog-hybrid, and tri-hybrid. Our numerical results demonstrate that the tri-hybrid outperforms the other architectures in terms of energy efficiency for practically relevant input power ranges due to the lower number of radio-frequency (RF) chains and phase shifters.

Notation: a bold lowercase letter $\mathbf{a}$ denotes a vector, a bold uppercase letter $\mathbf{A}$ denotes a matrix, and a script uppercase letter denotes a set $\mathcal{A}$. The matrix $\mathbf{A}^{\mathsf{T}}$ denotes the transpose of $\mathbf{A}$ and the matrix $\mathbf{A}^{*}$ denotes the conjugate transpose of $\mathbf{A}$. The matrix $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. The term j denotes the imaginary unit. Given M matrices $\{\mathbf{A}_i\}_{i=1}^{M}$, $\mathrm{blkdiag}\left(\mathbf{A}_1, \cdots, \mathbf{A}_M\right)$ denotes a block-diagonal matrix with $\{\mathbf{A}_i\}_{i=1}^{M}$ along the diagonal. The term $\|\mathbf{A}\|_{\mathsf{F}}^2$ denotes the Frobenius norm of $\mathbf{A}$.

II. System Model

We consider a single-user narrowband MIMO link in which a tri-hybrid transmitter sends N_{s} data streams to a fully-digital

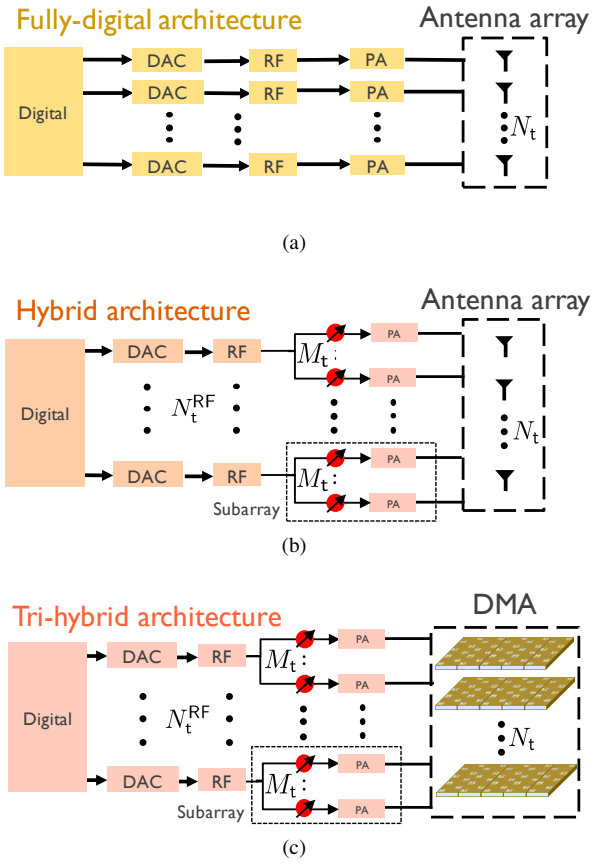

Fig. 1. Precoding architectures for (a) fully-digital precoding, (b) hybrid precoding, and (c) tri-hybrid precoding. The tri-hybrid architecture uses the least amount of RF components due to the additional DMA analog precoding from the reconfigurable components.

receiver. The transmitter is equipped with N_t antennas and the receiver is equipped with N_r antennas. The transmitter uses N_t^RF RF chains to convert signals from digital-baseband to analog-passband. We assume that $N_\mathsf{s} \le N_\mathsf{t}^\mathsf{RF} \le N_\mathsf{t}$, i.e, the number of transmit chains is at least the number of symbols and at most the number of antennas. We will compare the performance of a fully-digital architecture with $N_\mathsf{t} = N_\mathsf{t}^\mathsf{RF}$ with a tri-hybrid architecture with $N_\mathsf{t} > N_\mathsf{t}^\mathsf{RF}$.

Fig. 1 compares the fully-digital, hybrid, and tri-hybrid architectures, where we take a subarray approach for the hybrid and tri-hybrid cases due to the leaky-wave structure of DMAs. A key difference between the traditional hybrid architecture and the proposed tri-hybrid solution is the use of a reconfigurable antenna element at the transmitter. The reconfigurable antenna enables signal-processing by tuning the gain-pattern, polarization, and even operating frequency of the antenna. In this paper, we assume that the reconfigurable antenna is implemented using a block of DMAs, as shown in Fig. 1(c). The DMA is a type of tunable leaky-wave antenna that consists of a number of tunable unit cells that act as gain-pattern reconfigurable antennas. We assume each DMA block consists of N_t^uc unit cells, meaning the transmit array effectively contains $N_\mathsf{t}^\mathsf{ele} = N_\mathsf{t} N_\mathsf{t}^\mathsf{uc}$ total antenna elements.

The tri-hybrid transmitter consists of three separate signal processing blocks: a digital precoder $\mathbf{F}_\mathsf{dig} \in \mathbb{C}^{N_\mathsf{t}^\mathsf{RF} \times N_\mathsf{s}}$, an analog precoder $\mathbf{F}_\mathsf{ana} \in \mathbb{C}^{N_\mathsf{t} \times N_\mathsf{t}^\mathsf{RF}}$, and a DMA precoder $\mathbf{F}_\mathsf{dma} \in \mathbb{C}^{N_\mathsf{t}^\mathsf{ele} \times N_\mathsf{t}}$. The digital precoder operates at baseband and has no constraints in terms of the weights that can be applied. We assume a partially connected phase shifter analog precoder in which each RF chain is only connected to $M_\mathsf{t} = N_\mathsf{t}/N_\mathsf{t}^\mathsf{RF}$ antennas, where we assume that N_t is divisible by N_t^RF [14]. This imposes a block-diagonal structure on $\mathbf{F}_\mathsf{ana}$ in which the ith diagonal element is a vector $\mathbf{f}_{\mathsf{ana},i} \in \mathbb{C}^{M_\mathsf{t}}$, giving $\mathbf{F}_\mathsf{ana} = \mathrm{blkdiag}(\mathbf{f}_{\mathsf{ana},1}, \cdots, \mathbf{f}_{\mathsf{ana},N_\mathsf{t}^\mathsf{RF}})$. The phase shifter architecture further constrains each entry of $\mathbf{f}_{\mathsf{ana},i}$ to be unit-modulus. We denote the feasible analog precoder set as $\mathcal{F}_\mathsf{ana}$. We also place a power constraint on the signal entering the antenna as $\left\|\mathbf{F}_\mathsf{ana}\mathbf{F}_\mathsf{dig}\right\|_\mathsf{F}^2 = N_\mathsf{s}$. The input signal into the DMA will attenuate due to the waveguides and the passive unit cells. The power constraint here assumes that the power flowing into the antenna is not adjusted based on the DMA attenuation. We account for this via electromagnetic simulation results for the DMA beam patterns, as discussed in Section IV-B.

The DMA-based array places additional constraints on the transmit signal. The DMAs act as subarrays, where the nth DMA applies the weights $\mathbf{f}_{\mathsf{dma},n}$ to the transmit signal so that

$$\mathbf{F}_\mathsf{dma} = \mathrm{blkdiag}(\mathbf{f}_{\mathsf{dma},1}, \cdots, \mathbf{f}_{\mathsf{dma},N_\mathsf{t}}). \tag{1}$$

The weights applied by the DMAs depend on the tunable resonance of each unit cell and the effects of the waveguide on the signal. Let the weight applied by the kth unit cell of the nth be denoted as $\alpha_{n,k}$. We apply the Lorentzian resonator model for each unit cell as in [15], which constrains

$$\alpha_{n,k} \in \left\{ -\frac{\mathsf{j} + e^{\mathsf{j}\phi}}{2} \;:\; \phi \in [0, 2\pi) \right\}. \tag{2}$$

Tuning the DMA unit cell jointly changes its phase and amplitude due to the passive cell response. Prior to being radiated by the unit cell, however, the transmit signal must travel through the waveguide and experience distance-dependent phase offset and attenuation. We let $\eta_{n,k}$ denote the complex waveguide channel experienced by a transmit signal traveling to the kth unit cell in the nth DMA. We combine these effects and model

$$\mathbf{f}_{\mathsf{dma},n} = \left[\eta_{n,1}\alpha_{n,1}, \cdots, \eta_{n,N_\mathsf{t}^\mathsf{uc}}\alpha_{n,N_\mathsf{t}^\mathsf{uc}}\right]^\mathsf{T}. \tag{3}$$

We will let the set $\mathcal{F}_\mathsf{dma}$ denote the precoders that satisfy the block diagonal constraint in (1) and (3).

The transmit signal $\mathbf{x} \in \mathbb{C}^{N_\mathsf{t}^\mathsf{ele}}$ is formed by precoding the symbol vector $\mathbf{s} \in \mathbb{C}^{N_\mathsf{s}}$ as $\mathbf{x} = \mathbf{F}_\mathsf{dma}\mathbf{F}_\mathsf{ana}\mathbf{F}_\mathsf{dig}\mathbf{s}$. We assume that the symbol vector $\mathbf{s}$ is zero-mean and has variance $\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \mathbf{I}_{N_\mathsf{s}}$. Let $\mathbf{H} \in \mathbb{C}^{N_\mathsf{r} \times N_\mathsf{t}^\mathsf{ele}}$ be the channel matrix between each unit cell and the receiver, ρ be the average receive power, and $\mathbf{n}$ be the AWGN noise vector with entries with mean zero and variance σ^2. We define the average receiver power through the transmit power P_T and a channel gain G as $\rho = P_\mathsf{T} G$, which is further discussed in Section IV. The input-output model is

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_\mathsf{dma}\mathbf{F}_\mathsf{ana}\mathbf{F}_\mathsf{dig}\mathbf{s} + \mathbf{n}. \tag{4}$$

Assuming full channel knowledge and a fully-digital receiver, the transmitter finds the precoders that maximize spectral efficiency as

$$\begin{aligned}\max_{\mathbf{F}_{\mathsf{dig}},\mathbf{F}_{\mathsf{ana}},\mathbf{F}_{\mathsf{dma}}} \quad & \log_2\left|\mathbf{I}_{N_\mathsf{r}}+\frac{\rho}{\sigma^2}\mathbf{H}\mathbf{F}_{\mathsf{dma}}\mathbf{F}_{\mathsf{ana}}\mathbf{F}_{\mathsf{dig}}\mathbf{F}_{\mathsf{dig}}^*\mathbf{F}_{\mathsf{ana}}^*\mathbf{F}_{\mathsf{dma}}^*\mathbf{H}^*\right| \\ \text{s.t.} \quad & \mathbf{F}_{\mathsf{ana}}\in\mathcal{F}_{\mathsf{ana}} \\ & \mathbf{F}_{\mathsf{dma}}\in\mathcal{F}_{\mathsf{dma}} \\ & \left\|\mathbf{F}_{\mathsf{ana}}\mathbf{F}_{\mathsf{dig}}\right\|_{\mathsf{F}}^2=N_\mathsf{s}.\end{aligned} \tag{5}$$

We note that the power constraint in (5) limits the power entering the antennas, which is the input power after amplification and matching. Other options for the power constraint include the radiated power or the input power prior to amplification and matching. We leave the analysis of the effects of different power constraints for future work.

The tri-hybrid architecture is a generalization of the fully-digital and analog-hybrid architectures. Letting $N_\mathsf{t}^{\mathsf{uc}}=1$, the DMA precoder becomes a diagonal matrix of dimensions $N_\mathsf{t}\times N_\mathsf{t}$. Assuming that the antennas in the array are not reconfigurable, $\mathbf{F}_{\mathsf{dma}}$ becomes a scaled identity matrix $\alpha\mathbf{I}_{N_\mathsf{t}}$ where the scalar can be absorbed into the channel. The signal model is then

$$\mathbf{y}=\sqrt{\rho}\mathbf{H}\mathbf{F}_{\mathsf{ana}}\mathbf{F}_{\mathsf{dig}}\mathbf{s}+\mathbf{n}, \tag{6}$$

which coincides with that of an analog-hybrid architecture. If we further assume that each antenna is connected to an RF chain, then $N_\mathsf{t}=N_\mathsf{t}^{\mathsf{RF}}$ and $\mathbf{F}_{\mathsf{ana}}=\mathbf{I}_{N_\mathsf{t}}$. This yields a fully-digital architecture with input-output model

$$\mathbf{y}\sqrt{\rho}\mathbf{H}\mathbf{F}_{\mathsf{dig}}\mathbf{s}+\mathbf{n}. \tag{7}$$

We will compare the spectral and energy efficiency of the tri-hybrid architecture to these two baselines to better find the regimes in which DMA-based processing is beneficial.

III. Tri-hybrid architecture analysis

A. DMA precoder optimization

The DMA precoder can be difficult to optimize due to the Lorentzian-constrained weights in (2) and the effects of the waveguide. These issues are compounded by the fact that the waveguide channel depends on the unit cell configuration because of mutual coupling. Rather than jointly optimizing all of the unit cells, we assume a codebook-based approach in which the weight vector from the nth is selected from a predefined codebook. Letting C denote the number of beams in $\mathcal{C}_{\mathsf{dma}}$, we can define a beamforming codebook $\mathcal{C}_{\mathsf{dma}}^{N_\mathsf{t}}$ with CN_t elements. Each codeword in $\mathcal{C}_{\mathsf{dma}}^{N_\mathsf{t}}$ defines a tuple $\left(\mathbf{f}_{\mathsf{dma},1},\cdots,\mathbf{f}_{\mathsf{dma},N_\mathsf{t}}\right)$ that generates the DMA precoder as $\mathbf{F}_{\mathsf{dma}}=\mathrm{blkdiag}(\mathbf{f}_{\mathsf{dma},1},\cdots,\mathbf{f}_{\mathsf{dma},N_\mathsf{t}})$. The codebook approach reduces the complexity of solving the DMA precoder optimization and has been shown to achieve beam patterns with similar gain and coverage as DFT codebooks [16].

The transmitter performs a combinatorial search to obtain the optimal DMA precoder. Let $\mathbf{F}_{\mathsf{dma}}^{(c)}$ denote the cth precoder obtained from $\mathcal{C}_{\mathsf{dma}}^{N_\mathsf{t}}$. Further, let $\mathbf{F}_{\mathsf{opt}}^{(c)}$ be the unconstrained precoding matrix obtained by waterfilling over the effective channel $\mathbf{H}\mathbf{F}_{\mathsf{dma}}^{(c)}$. The optimal DMA precoder is chosen as $\mathbf{F}_{\mathsf{dma,opt}}=\mathbf{F}_{\mathsf{dma}}^{(c_{\mathsf{opt}})}$ where

$$\begin{aligned}c_{\mathsf{opt}}=&\operatorname*{argmax}_{c} \\ &\log_2\left|\mathbf{I}_{N_\mathsf{r}}+\frac{\rho}{\sigma^2}\mathbf{H}\mathbf{F}_{\mathsf{dma}}^{(c)}\mathbf{F}_{\mathsf{opt}}^{(c)}\left(\mathbf{F}_{\mathsf{opt}}^{(c)}\right)^*\left(\mathbf{F}_{\mathsf{dma}}^{(c)}\right)^*\mathbf{H}^*\right|.\end{aligned} \tag{8}$$

We discuss the codebook design in detail in Section IV-B.

B. Hybrid precoder optimization

Given the optimal DMA precoder $\mathbf{F}_{\mathsf{dma,opt}}$ the transmitter computes the optimal digital and and analog precoders through the effective channel $\widetilde{\mathbf{H}}=\mathbf{H}\mathbf{F}_{\mathsf{dma,opt}}$. We apply the method described in [14] to solve the hybrid precoding optimization under the partially connected architecture. Letting $\widetilde{\mathbf{H}}_m$ denote the $N_\mathsf{r}\times M_\mathsf{t}$ effective channel matrix for the mth subarray, the effective channel can be written as

$$\widetilde{\mathbf{H}}=\left[\widetilde{\mathbf{H}}_1,\widetilde{\mathbf{H}}_2,\cdots,\widetilde{\mathbf{H}}_{N_\mathsf{t}^{\mathsf{RF}}}\right]. \tag{9}$$

The analog beamformer for each RF chain is found through the singular value decomposition (SVD) of the effective subarray channels. Let $\mathbf{v}_m$ denote the dominant right singular vector of $\widetilde{\mathbf{H}}_m$. Then the optimal analog precoder is computed as

$$\mathbf{F}_{\mathsf{ana,opt}}=\mathrm{blkdiag}\left(e^{\mathrm{j}\angle\mathbf{v}_1},\cdots,e^{\mathrm{j}\angle\mathbf{v}_{N_\mathsf{t}^{\mathsf{RF}}}}\right). \tag{10}$$

The digital precoder is then found by finding the optimal unitary precoder for the new effective channel after analog and DMA precoding. Let $\mathbf{V}^\star$ be the right singular matrix of $\mathbf{H}\mathbf{F}_{\mathsf{dma,opt}}\mathbf{F}_{\mathsf{ana,opt}}$. Then $\mathbf{F}_{\mathsf{dig,opt}}$ is obtained from the first N_s columns of $\mathbf{V}^\star$.

Due to the limitations of the DMA, a standard hybrid precoder will generally achieve higher spectral efficiency than the tri-hybrid precoder. The main benefit of the DMA is the passive tunability, which drastically reduces the power consumption. In the next section we overview the power consumption of each of the different architectures.

C. Power consumption calculation

In this section, we derive expressions for the power consumption of the general tri-hybrid architecture, a hybrid precoding architecture with no DMA, and a full-digital precoding architecture. We denote the power consumption of a number of transmit components as follows: P_{PA} for a single power amplifier, P_{DAC} for a digital-to-analog converter (DAC) P_{LO} for a local oscillator, P_{RF} for the combination of the mixer, low-pass filter and hybrid with buffer, P_{PS} for a phase shifter. Each architecture consumes a different amount of power due to the difference in the quantity of each components. We use the power consumption and component loss values described in [17] for each architecture and the simulation results.

Fully-digital architecture: The fully-digital architecture attaches each antenna to its own RF chain. All processing is performance at digital baseband and no phase shifters or DMAs are used. Since there are N_t antennas, the power consumption is

$$P_{\mathsf{FD}}=P_{\mathsf{LO}}+N_\mathsf{t}\left(2P_{\mathsf{DAC}}+P_{\mathsf{RF}}+P_{\mathsf{PA}}\right). \tag{11}$$

This architecture consumes the most power due to the large number of RF chains and DACs required.

Hybrid-analog architecture: The traditional hybrid architecture reduces the number of RF chains by translating some of the signal processing to the RF domain. Since the transmitter uses a partially-connected configuration, N_{t} phase shifters are used for analog precoding. The power consumption is then

$$P_{\mathsf{HA}} = P_{\mathsf{LO}} + N_{\mathsf{t}}^{\mathsf{RF}} \left(2P_{\mathsf{DAC}} + P_{\mathsf{RF}}\right) + N_{\mathsf{t}} \left(P_{\mathsf{PS}} + P_{\mathsf{PA}}\right). \quad (12)$$

The phase shifters increase the system power consumption, but the hybrid-analog architecture contains less RF chains than the fully-digital architecture, consuming less overall power.

Tri-hybrid architecture: The tri-hybrid architecture combines analog precoding with antenna-based processing to further reduce power consumption. The only difference in the power consumption between this and the hybrid-analog architecture is the tunable components required to reconfigure the DMA. Here, we assume the tunable component is a varactor diode with power consumption P_{VAR}. Each DMA block contains $N_{\mathsf{t}}^{\mathsf{uc}}$ unit cells, so the tri-hybrid power consumption is

$$\begin{aligned} P_{\mathsf{TH}} =& P_{\mathsf{LO}} + N_{\mathsf{t}}^{\mathsf{RF}} \left(2P_{\mathsf{DAC}} + P_{\mathsf{RF}}\right) + N_{\mathsf{t}} \left(P_{\mathsf{PS}} + P_{\mathsf{PA}}\right) \quad (13) \\ &+ N_{\mathsf{t}} N_{\mathsf{t}}^{\mathsf{uc}} P_{\mathsf{VAR}}. \quad (14) \end{aligned}$$

The power consumed by each individual varactor diode is negligible compared to the other transceiver components. This means that a large number of unit cells can be used to offset the reduced digital and analog processing.

Lastly, we integrate the losses from RF components into the model and define a metric for energy efficiency. We include loss from power dividers and phase shifters to model the total transmit power into the antenna arrays. For an input power P_{IN} and total component loss L, we define the transmit power for the different antenna arrays as

$$P_{\mathsf{T}} = \frac{P_{\mathsf{IN}}}{L}. \quad (15)$$

Since the fully-digital, hybrid-analog, and tri-hybrid architectures contain different amounts of power dividers and phase shifters, the component loss and transmit power for each architecture will vary. We define the component loss from phase shifters as L_{PS}, and the component loss due to an N_{a}-way power divider as $L_{\mathsf{D}}(N_{\mathsf{a}})$. We can now establish the component loss models for the three architectures. For fully-digital precoding with no phase shifters, the total component loss for the N_{t}-way power divider is

$$L_{\mathsf{FD}} = L_{\mathsf{D}}(N_{\mathsf{t}}). \quad (16)$$

For hybrid-analog precoding, the component loss for an M_{t}-way power divider and phase shifters is

$$L_{\mathsf{HA}} = L_{\mathsf{D}}(M_{\mathsf{t}}) L_{\mathsf{PS}}. \quad (17)$$

Lastly, for tri-hybrid precoding, the total component loss incorporates both the M_{t}-way power divider and the additional power divider for N_{y} waveguides as

$$L_{\mathsf{TH}} = L_{\mathsf{D}}(M_{\mathsf{t}}) L_{\mathsf{D}}(N_{\mathsf{y}}) L_{\mathsf{PS}}. \quad (18)$$

The total transmit power is then calculated for each architecture using the component loss, and the transmit power is integrated into the average receive power for spectral efficiency results. For a spectral efficiency R for each architecture, we analyze the energy efficiency of the system via the ratio of spectral efficiency to total power consumption P_{cons} as

$$\eta = \frac{R}{P_{\mathsf{cons}}}. \quad (19)$$

We present simulation results for the spectral and energy efficiency of the fully-digital, hybrid-analog, and tri-hybrid architectures in the following section.

IV. Numerical results

We now present some simulation results to evaluate the tri-hybrid architecture performance. We compare both spectral and energy efficiency for the fully-digital, hybrid, and tri-hybrid architectures to determine the tradeoff in performance and power consumption savings.

A. Antenna design

We first describe the antenna designs and characteristics used in the simulation results. We assume that the tri-hybrid architecture uses a DMA and that the fully-digital and hybrid architectures use patch antennas. We alter a previously-designed DMA in [4] to obtain gain patterns for a realistic DMA. We design each DMA block to consist of $N_{\mathsf{y}} = 5$ waveguides, each having 8 unit cells, which gives $N_{\mathsf{t}}^{\mathsf{uc}} = 40$ elements. Placing 8 elements per waveguide ensures that nearly all of the input power from the waveguide excitation is radiated out through the elements. The operating frequency of the DMA design is $f_0 = 15$ GHz. We assume quarter-wavelength spacing between DMA elements and half-wavelength spacing between waveguides. Additional information regarding the DMA design can be found in [16].

The next step is to design a comparable antenna array to use with the fully-digital and hybrid architectures. As DMA elements have a directional beam pattern, the elements in the antenna array should also be directional in nature. Therefore, we choose to design a patch antenna element at $f_0 = 15$ GHz. The designed antenna element is a square copper patch with length 3.5 mm atop a square substrate with length 10 mm. The resulting patch antenna yields the desired resonance at $f_0 = 15$ GHz and has a high overall efficiency.

B. DMA codebook design

We now define the codebook used for DMA precoding as a part of the tri-hybrid precoding architecture. For simplicity, we assume the elevation angle to be zero and focus solely on precoding in the azimuth direction with a DFT-based codebook. For $N_{\mathsf{x}} = 8$ unit cells per waveguide, the DFT codebook consists of 8 beam patterns that span the azimuth angular region from $\phi \in [-90°, 90°]$. We define the array

steering vector for azimuth steering angle Ω, element spacing d, and wavenumber k_0 as [18]

$$\mathbf{a}(\Omega) = \frac{1}{\sqrt{N_{\mathsf{x}}}}\left[e^{\mathrm{j}k_0 d0\sin\Omega}, \ldots, e^{\mathrm{j}k_0 d(N_{\mathsf{x}}-1)\sin\Omega}\right]^T. \quad (20)$$

We define the beamforming vector for codeword c as [18]

$$\mathbf{f}_{\mathsf{dft}}(c) = \mathbf{a}\left(-1 + \frac{2c-1}{N_{\mathsf{x}}}\right), \quad (21)$$

where $c \in \{1, \cdots, 8\}$. We use the ideal DFT codebook vectors to create the DMA codebook.

Next, we convert the weights described in (21) to fit the Lorentzian-constrained distribution in (2). We use a method known as Euclidean modulation to map the DFT weights onto the DMA-constrained weights by minimizing the Euclidean distance [4]. We also apply an optimization strategy involving a phase rotation ζ of the desired, DFT codeword weights prior to mapping to improve beamforming gain. For a DFT codeword element defined by the cth codeword and kth unit cell as $f_{\mathsf{dft},k}(c)$, we define the DMA weight as a function of ζ for the nth DMA precoder $\eta_{n,k}\alpha_{n,k}$ as [4]

$$\alpha_{n,k}^{\mathsf{EM}}(\zeta) = \underset{\alpha_{n,k}}{\operatorname{argmin}} |\eta_{n,k}\alpha_{n,k} - f_{\mathsf{dft},k}(c)e^{\mathrm{j}\zeta}|^2. \quad (22)$$

We simulate beam patterns with the designed DMA in target directions for the DFT codebook. For a resulting beam pattern gain $v(\zeta, \phi_0)$ in a target direction ϕ_0, we define the phase rotation that maximizes beamforming gain within a set of possible phase rotations $\mathcal{Z}$ as

$$\hat{\zeta}(\phi_0) = \underset{\zeta\in\mathcal{Z}}{\operatorname{argmax}}\, v(\zeta, \phi_0). \quad (23)$$

We then use $\alpha_{n,k}^{\mathsf{EM}}(\hat{\zeta}(\phi_0))$ as the weights for the target direction ϕ_0 to form the DMA codebook. Additional information regarding this phase rotation optimization strategy and its implementation can be found in [16]. We use the beam patterns optimized with the phase rotation through electromagnetic simulation results as the final, DFT codebook. Next, we will discuss the integration of the beam patterns into a channel model for spectral and energy efficiency simulation results.

C. Simulation results

To provide a realistic analysis of the three precoding architectures in a wireless system, we use the channel modeling software QuaDRiGa [19]. QuaDRiGa incorporates a realistic wireless channel model between a transmit and receive array, allowing us to integrate the simulated beam patterns for the DMA codebook and patch antenna element and obtain a simulated channel gain G. We assume a single-antenna receiver at a distance $r = 500$ meters away from the transmit array, and consider a line-of-sight channel model where the elevation angle is zero and the receiver is within an azimuth angular distribution $\phi \in [-70°, 70°]$. Using the simulated the beam patterns for an entire 5×8 DMA in HFSS, the channel we obtain from QuaDRiGa represents a post-beamforming channel for the DMA. We use this channel to apply the subsequent analog and digital precoding for the tri-hybrid architecture. We choose the codeword in the DMA codebook that maximizes the spectral efficiency.

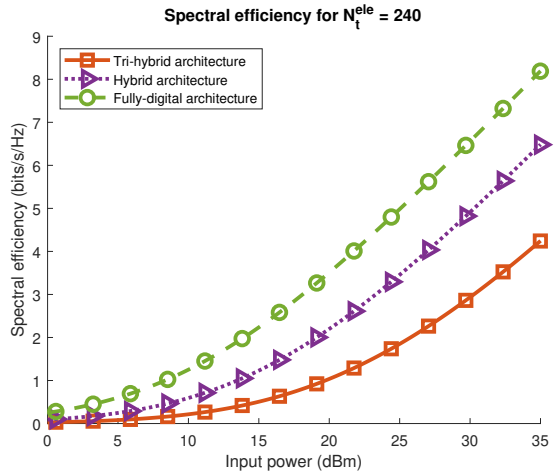

Fig. 2. Spectral efficiency results for a scenario with $N_{\mathsf{t}}^{\mathsf{ele}} = 240$ elements. We find that both the fully-digital and hybrid architectures outperform the tri-hybrid architecture. This is expected, as these precoding methods consume more power than the tri-hybrid approach.

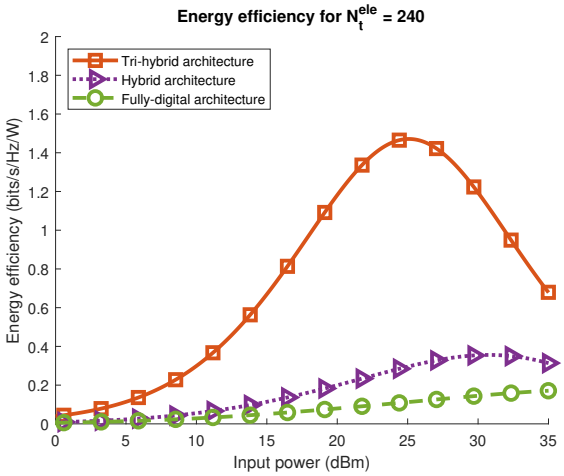

Fig. 3. Energy efficiency results for a scenario with $N_{\mathsf{t}}^{\mathsf{ele}} = 240$ elements. The tri-hybrid provides the best energy efficiency compared to the fully-digital and hybrid architectures due to the decreased power consumption of the DMA.

We calculate the spectral and energy efficiency results for the three precoding architectures from (8) and (19) across input power and the number of antenna elements. We define the number of antenna elements as $N_{\mathsf{t}}^{\mathsf{ele}} = N_{\mathsf{t}}N_{\mathsf{t}}^{\mathsf{uc}}$, where $N_{\mathsf{t}}^{\mathsf{uc}} = 1$ for the fully-digital and hybrid-analog cases, and $N_{\mathsf{t}}^{\mathsf{uc}} = 40$ for the tri-hybrid case. We show the results for $N_{\mathsf{t}}^{\mathsf{ele}} = 240$ elements in Figs. 2 and 3. As anticipated, fully-digital precoding provides the best spectral efficiency results, followed by hybrid precoding and tri-hybrid precoding. This directly follows the power consumed by each precoding architecture, where fully-digital consumes the most power due to the large number of required RF chains, and tri-hybrid precoding consumes the least power from the incorporation of

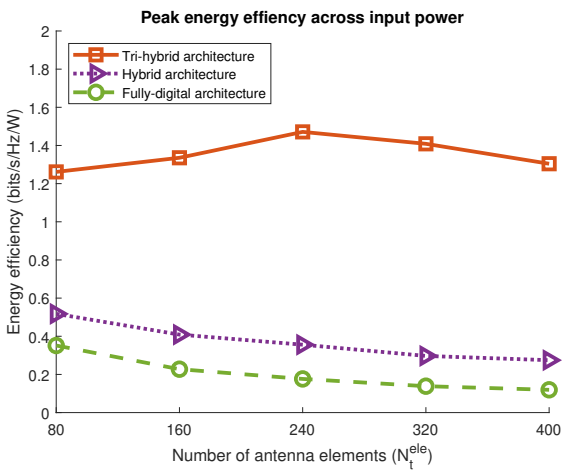

Fig. 4. Peak energy efficiency for the three precoding architectures as a function of the number of antenna elements. As the number of antennas increases, the peak energy efficiency of the fully-digital and hybrid architectures decreases, while the tri-hybrid architecture remains approximately the same.

DMA precoding. Therefore, in terms of energy efficiency in Fig. 3, we find that the tri-hybrid architecture provides the best performance due to the lower power consumption compared to the fully-digital and hybrid precoding architectures.

We further elaborate on the energy efficiency results in Fig. 4 by showing the peak energy efficiency across input power as a function of the number of antennas. We find here that the tri-hybrid architecture has an approximately flat peak energy efficiency as the number of antennas increases. This indicates that the spectral efficiency increases proportionally to the increase in power consumption from a larger number of RF chains. The hybrid and fully-digital architectures, however, have a peak energy efficiency that decreases with the number of antennas, meaning that the increase in power consumption for more antennas is larger than the increase in spectral efficiency. Overall, the power consumption savings for the tri-hybrid architecture leads to a significant increase in energy efficiency compared to the fully-digital and hybrid precoding architectures for a realistic wireless system.

V. Conclusion

Tri-hybrid precoding offers a flexible and energy-efficient alternative to conventional hybrid architectures. The key idea is to leverage novel reconfigurable antennas to reduce the number of RF chains and phase shifters at the trasnceiver. A critical challenge to realizing tri-hybrid MIMO lies in jointly optimizing the digital, analog, and digital precoders. In this paper, we showed that even with a suboptimal precoding approach, the tri-hybrid architecture offers much higher energy efficiency. In the future, we plan to develop more sophisticated methods for tri-hybrid precoding for a variety of MIMO applications, including wideband and multi-user communication.

VI. Acknowledgements

This material is based upon work supported by the National Science Foundation under grant nos. NSF-ECCS-2153698, NSF- CCF-2225555, NSF-CNS-2147955 and is supported in part by funds from federal agencies and industry partners as specified in the Resilient & Intelligent NextG Systems (RINGS) program. This work is also supported in part by the Qualcomm Innovation Fellowship.

References

[1] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Fiveq promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, Nov. 2019.

[2] O. Kanhere, H. Poddar, Y. Xing, D. Shakya, S. Ju, and T. S. Rappaport, "A power efficiency metric for comparing energy consumption in future wireless networks in the millimeter-wave and terahertz bands," *IEEE Wireless Commun.*, vol. 29, no. 6, pp. 56–63, Dec. 2022.

[3] D. R. Smith, O. Yurduseven, L. P. Mancera, P. Bowen, and N. B. Kundtz, "Analysis of a waveguide-fed metasurface antenna," *Physical Review Applied*, vol. 8, no. 5, p. 054048, Nov.

[4] M. Boyarsky, T. Sleasman, M. F. Imani, J. N. Gollub, and D. R. Smith, "Electronically steered metasurface antenna," *Scientific reports*, vol. 11, no. 1, pp. 1–10, Feb. 2021.

[5] H. Li, S. Li, B. Hou, X. Zhang, W. Wen, and C. Hu, "A digital SIW-slot antenna array with FPGA implementation of beamforming," *Scientific Reports*, vol. 12, no. 1, p. 8927, May 2022.

[6] N. Shlezinger, O. Dicker, Y. C. Eldar, I. Yoo, M. F. Imani, and D. R. Smith, "Dynamic metasurface antennas for uplink massive MIMO systems," *IEEE Trans. commun.*, vol. 67, no. 10, pp. 6829–6843, Oct. 2019.

[7] L. You, J. Xu, G. C. Alexandropoulos, J. Wang, W. Wang, and X. Gao, "Energy efficiency maximization of massive MIMO communications with dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, Aug. 2022.

[8] M. Rezvani and R. Adve, "Channel estimation for dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, pp. 1–1, Nov. 2023.

[9] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Beam focusing for near-field multiuser MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7476–7490, Mar. 2022.

[10] H. Wang, N. Shlezinger, Y. C. Eldar, S. Jin, M. F. Imani, I. Yoo, and D. R. Smith, "Dynamic metasurface antennas for MIMO-OFDM receivers with bit-limited ADCs," *IEEE Trans. commun.*, vol. 69, no. 4, pp. 2643–2659, Apr. 2020.

[11] W. Huang, H. Zhang, N. Shlezinger, and Y. C. Eldar, "Joint microstrip selection and beamforming design for MmWave systems with dynamic metasurface antennas," *arXiv preprint arXiv:2210.12390*, Oct. 2022.

[12] R. J. Williams, P. Ramírez-Espinosa, J. Yuan, and E. de Carvalho, "Electromagnetic based communication model for dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8616–8630, Apr. 2022.

[13] I. Yoo, D. R. Smith, and M. Boyarsky, "Experimental characterization of a waveguide-fed varactor-tuned metamaterial element using the coupled dipole framework," *IEEE Antennas Wireless Propag. Lett.*, vol. 22, no. 2, pp. 387–391, Oct. 2023.

[14] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[15] D. R. Smith, O. Yurduseven, L. P. Mancera, P. Bowen, and N. B. Kundtz, "Analysis of a waveguide-fed metasurface antenna," *Physical Review Applied*, vol. 8, no. 5, p. 054048, Nov. 2017.

[16] J. Carlson, M. R. Castellanos, and R. W. Heath Jr, "Hierarchical codebook design with dynamic metasurface antennas for energy-efficient arrays," *arXiv preprint arXiv:2305.09016*, 2023.

[17] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. of Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 298–312, Apr. 2018.

[18] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[19] F. Burkhardt, S. Jaeckel, E. Eberlein, and R. Prieto-Cerdeira, "QuaDRiGa: A MIMO channel model for land mobile satellite," in *Proc. of EuCAP 2014*, Apr. 2014, pp. 1274–1278.