

Joint Beamforming and Matching for Ultra-Dense Massive Antenna Arrays

Carolina Nolasco-Ferencikova^{*,1}, Georg Schwan^{*,1}, Raphael Rolny², Alexander Stutz-Tirri¹, and Christoph Studer¹

¹ETH Zurich, Switzerland; ²armasuisse Science and Technology, Switzerland

email: cnolasco@ethz.ch, gschwan@ethz.ch, raphael.rolny@armasuisse.ch, alstutz@ethz.ch, studer@ethz.ch

Abstract—Massive multiple-input multiple-output (MIMO) offers substantial spectral-efficiency gains, but scaling to very large antenna arrays with conventional all-digital and hybrid beamforming architectures quickly results in excessively high costs and power consumption. Low-cost, switch-based architectures have recently emerged as a potential alternative. However, prior studies rely on simplified models that ignore (among others) antenna coupling, radiation patterns, and matching losses, resulting in inaccurate performance predictions. In this paper, we use a physically consistent electromagnetic modeling framework to analyze an ultra-dense patch-antenna array architecture that performs joint beamforming and matching using networks of inexpensive RF switches. Our results demonstrate that simple, switch-based beamforming architectures can approach the antenna-gain of all-digital solutions at significantly lower cost and complexity.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) with hundreds of antenna elements can dramatically increase spectral efficiency and reliability in wireless systems [2]. However, deploying large antenna arrays remains challenging due to the cost and power consumption of existing beamforming architectures. All-digital solutions that allocate one RF chain per antenna element quickly become expensive and power-hungry as the array size grows; hybrid analog-digital architectures mitigate these issues by reducing the number of RF chains, but still rely on networks of high-resolution phase shifters that remain costly and power hungry [3], [4].

In order to address these limitations, recent research has explored hybrid architectures that employ inexpensive RF components, such as RF switches. For example, Eslami Rasekh *et al.* [5] propose replacing conventional phase shifters with low-resolution, switch-based binary phase shifters, offering a viable approach towards low-cost and energy-efficient beamforming. Despite their promise, existing research on such low-cost architectures relies on simplified and physically inconsistent models

that neglect real-world effects, including mutual coupling, non-ideal element patterns, ohmic losses, and matching losses. These effects become particularly important with inexpensive and highly constrained RF hardware, and ignoring them leads to inaccurate performance predictions. As a result, the real-world efficacy of such low-cost architectures is currently unclear.

A. Contributions

In this paper, we demonstrate that physically consistent modeling enables the systematic design, optimization, and analysis of novel beamforming architectures capable of scaling MIMO arrays to extreme dimensions. Specifically, we utilize the framework introduced in [1, Sec. II] to model an ultra-dense massive patch-antenna array, capturing mutual coupling, antenna element patterns, losses, etc. We then propose a new hybrid beamforming architecture that performs joint beamforming and matching using low-cost RF components (i.e., networks of inexpensive RF switches and transmission lines), thereby eliminating the need for costly and power-hungry phase shifters or other analog circuitry. Our results confirm that simple beamforming architectures can approach the antenna gain of all-digital architectures at significantly lower costs.

B. Notation

We use lowercase boldface for general vectors (e.g., $\mathbf{a}$) and uppercase boldface for general matrices (e.g., $\mathbf{A}$). We use pink sans-serif (e.g., $\mathbf{a}$) and pink sans-serif boldface (e.g., $\mathbf{a}$) for phasors (cf. [1, Def. 1]) and vectors containing phasors, respectively. The superscripts $\top$ and H indicate transpose (e.g., $\mathbf{A}^\top$) and conjugate transpose (e.g., $\mathbf{A}^\mathrm{H}$), respectively. Given a vector $\mathbf{a}$, we use $\text{diag}(\mathbf{a})$ to indicate the diagonal matrix with the elements of $\mathbf{a}$ in the main diagonal. We denote the Euclidean norm by $\|\cdot\|_2$. We use blackboard bold font for operators (e.g., $\mathbb{S}$). For a complex number $z \in \mathbb{C}$, the real part is $\Re\{z\}$. Given $N \in \mathbb{N}$, we define the set $[N] \triangleq \{1, \dots, N\}$.

II. MODELING METHOD

A. REMS Model

In [1, Sec. II], we proposed an *efficient and physically consistent* far-field modeling framework for *general reconfigurable electromagnetic structures (REMSs)*. Throughout this paper, we will utilize parts of this model for our analysis.

^{*}equal contribution

For self-consistency, parts of Sec. II intentionally overlap with [1, Sec. II]. The authors thank Prof. Hua Wang for providing the scattering parameters of the RF switches used in this work. The work of AST was funded in part by armasuisse. The work of AST, GS, and CS was funded in part by the Swiss State Secretariat for Education, Research, and Innovation (SERI) under the SwissChips initiative, the Swiss National Science Foundation (SNSF) grant 200021_207314, by a CHIST-ERA grant for the project CHASER (CHIST-ERA-22-WAI-01) through the SNSF grant 20CH21_218704, and by the European Commission within the context of the project 6G-REFERENCE (6G Hardware Enablers for Cell Free Coherent Communications and Sensing), funded under EU Horizon Europe Grant Agreement 101139155.

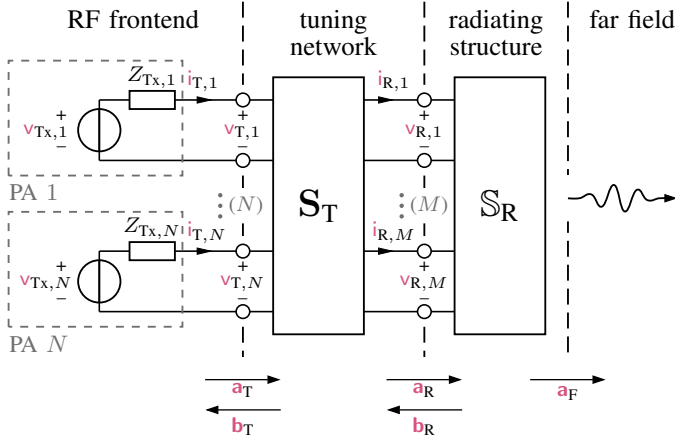


Figure 1. Structure of the model used in this paper. The *RF frontend* represents N power amplifiers; the *tuning network* represents reconfigurable analog beamforming and matching RF circuitry; and the *radiating structure* represents M antenna elements. The original structure shown in [1, Fig. 2] additionally models low-noise amplifiers (for receivers), noise from various sources, and incoming far-field waves—all of which are not used here.

With *efficient*, we mean that (i) the model parameters can be obtained with acceptable computational effort at design time, and (ii) the model enables real-time prediction of a REMS's behavior with very low complexity. Concretely, our approach allows each REMS in a wireless communication or sensing system to be characterized separately using a *single* full-wave EM simulation or measurement campaign. These characterizations can then be used to quickly compute the system-level behavior for arbitrary REMS positions, RF circuitry configurations, and RF frontend output signals.

With *physically consistent*, we mean that our model's predictions adhere to the physical laws dictated by Maxwell's equations. This includes effects such as mutual (inter-antenna) coupling, nonreciprocal materials, polarization, ohmic and matching losses, the influence of metallic housing, noise contributed by low-noise amplifiers, and noise generated within or received by antennas. Furthermore, when predicting the properties of wireless channels, our approach is significantly more numerically stable and, consequently, more accurate, while running orders of magnitude faster than full-wave EM simulations, as we demonstrate in [1, Sec. III].¹

With *general REMSs*, we refer to a wide class of wireless communication or sensing systems, including all-digital² or hybrid multi-antenna systems, reconfigurable reflectarrays, and passive or active reconfigurable intelligent surfaces. Practically, almost any system with a linear, time-invariant *radiating structure* can be described by our model.³

¹A detailed discussion on numeric stability and required computational resources can be found in [1, Sec. III-C].

²We speak of an *all-digital* architecture when the number of RF chains N is equal to the number of antenna elements M .

³To be specific, any system for which Assumptions 1–3 in [1] hold can be described by our model.

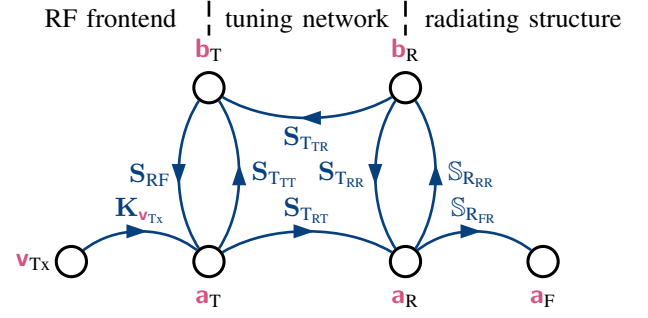


Figure 2. Signal-flow graph of the model used in this paper. The nodes represent (i) complex vectors corresponding to voltages and circuit-theoretic power waves, and (ii) elements in an L^2 -space that describe the far-field radiation pattern of the outgoing electromagnetic waves. The edges denote bounded linear operators acting between the respective spaces. The original graph shown in [1, Fig. 5] additionally contains received voltages, influence from various noise sources, and incoming far-field waves.

B. Aspects of the REMS Model Used in This Work

We now review the components of our REMS model that we use in Sec. III.⁴

Our model characterizes the REMS's behavior at an arbitrary frequency $f \in \mathbb{R}_{>0}$.⁵ For a given frequency f , we denote the free space wavenumber, impedance, and wavelength by $k \triangleq 2\pi f \sqrt{\mu_0 \epsilon_0}$, $Z_0 \triangleq \sqrt{\mu_0 / \epsilon_0}$, and $\lambda \triangleq (\sqrt{\epsilon_0 \mu_0} f)^{-1}$ respectively, where μ_0 is the permeability and ϵ_0 is the permittivity of free space. We use the physicist's convention for spherical coordinate systems [6], with r as the *radial distance*, θ as the *polar angle*, and φ as the *azimuthal angle*. Furthermore, for each coordinate (r, θ, φ) , we denote the local orthogonal unit vectors in the directions of increasing r , θ , and φ as $\hat{r}$, $\hat{\theta}$, and $\hat{\varphi}$, respectively. Finally, we denote the set of all angle pairs (θ, φ) on the unit sphere by $\Omega \triangleq [0, \pi] \times [0, 2\pi]$.

To describe the internal behavior of a REMS, we rely on a circuit-theoretic approach based on scattering parameters and *circuit-theoretic power waves* as defined in [1, Def. 2].

To describe the interaction between the radiating structure and electromagnetic waves in its far-field region⁶, we rely on the formalism introduced in [1, Sec. II-B4]. To this end, we note that, far away from a radiating structure, the electric field can be approximated using

$$\lim_{r \rightarrow \infty} r \mathbf{E}(r, \theta, \varphi) = \mathbf{E}^{\leftarrow}(\theta, \varphi) e^{+jkr} + \mathbf{E}^{\rightarrow}(\theta, \varphi) e^{-jkr}, \quad (1)$$

where $\mathbf{E}^{\leftarrow}(\theta, \varphi)$ and $\mathbf{E}^{\rightarrow}(\theta, \varphi)$ are complex-valued three-dimensional vectors [7, Eq. 5.12]. Consequently, we define the *outgoing far-field power wave pattern* as

$$\mathbf{a}_F : \Omega \rightarrow \mathbb{C}^2, (\theta, \varphi) \mapsto \frac{1}{\sqrt{Z_0}} \begin{bmatrix} \hat{\theta}^\top \mathbf{E}^{\rightarrow}(\theta, \varphi) \\ \hat{\varphi}^\top \mathbf{E}^{\rightarrow}(\theta, \varphi) \end{bmatrix}. \quad (2)$$

⁴To maintain consistency, in this paper, we repeat selected material from [1].

⁵This does not limit our model to narrowband analysis; it simply means that if multiple frequencies are of interest, the modeling procedure must be applied separately to each one.

⁶Specifically, our formalism captures the interaction with (i) waves incident from sufficiently far away and (ii) waves radiated into the far-field region.

In [1], we treat a REMS as a general transceiver and scatterer. Because the case study presented in this paper considers a purely transmitting system, our model reduces to the system depicted in Fig. 1 and described in the following.⁷

Our model comprises three subsystems: the *RF frontend*, the *radiating structure*, and the *tuning network*.

The *RF frontend* consists of $N \in \mathbb{N}$ power amplifiers (PAs), which we model using their Thévenin-equivalent circuits. The respective equivalent voltages and impedances are denoted by

$$\mathbf{v}_{\text{Tx}} \triangleq [\mathbf{v}_{\text{Tx},1} \cdots \mathbf{v}_{\text{Tx},N}]^T \quad (3)$$

$$\mathbf{Z}_{\text{Tx}} \triangleq \text{diag}([Z_{\text{Tx},1} \cdots Z_{\text{Tx},N}]). \quad (4)$$

The *radiating structure* consists of $M \in \mathbb{N}$ antennas which we model with (i) the *inter-element coupling operator* $\mathbb{S}_{\text{RRR}}: \mathbb{C}^M \rightarrow \mathbb{C}^M$ and (ii) the *transmitting operator* $\mathbb{S}_{\text{RFR}}: \mathbb{C}^M \rightarrow L^2$.⁸ Given the circuit-theoretic power waves $\mathbf{a}_{\text{R}} \in \mathbb{C}^M$ incident on the M antenna ports, then the circuit-theoretic power waves $\mathbf{b}_{\text{R}} \in \mathbb{C}^M$ leaving the ports, and the far-field power-wave pattern $\mathbf{a}_{\text{F}} \in L^2$ radiated into the far-field region are given as

$$\begin{bmatrix} \mathbf{b}_{\text{R}} \\ \mathbf{a}_{\text{F}} \end{bmatrix} = \begin{bmatrix} \mathbb{S}_{\text{RRR}} \\ \mathbb{S}_{\text{RFR}} \end{bmatrix} \mathbf{a}_{\text{R}}. \quad (5)$$

The *tuning network* contains the (reconfigurable) RF circuitry used for analog beamforming and matching. We represent the tuning network as a multiport, with the input-output relation

$$\begin{bmatrix} \mathbf{b}_{\text{T}} \\ \mathbf{a}_{\text{R}} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbb{S}_{\text{T TT}} & \mathbb{S}_{\text{T TR}} \\ \mathbb{S}_{\text{TR T}} & \mathbb{S}_{\text{RRR}} \end{bmatrix}}_{\triangleq \mathbb{S}_{\text{T}}} \begin{bmatrix} \mathbf{a}_{\text{T}} \\ \mathbf{b}_{\text{R}} \end{bmatrix}. \quad (6)$$

Here, $\mathbf{a}_{\text{T}} \in \mathbb{C}^N$ and $\mathbf{b}_{\text{T}} \in \mathbb{C}^N$ represent the incoming and outgoing waves (from the tuning network's perspective) present at the ports connecting the tuning network to the RF frontend.

C. Input-Output Relationship

The behavior of the REMSs considered here can be described by a system of linear equations, including (5) and (6). This system of equations can be visualized as the signal-flow graph depicted in Fig. 2 where we use the auxiliary matrices

$$\mathbf{K}_{\mathbf{v}_{\text{Tx}}} \triangleq (\mathbf{Z}_{\text{Tx}} + R_0 \mathbf{I}_N)^{-1} \sqrt{R_0} \quad (7)$$

$$\mathbf{S}_{\text{RF}} \triangleq (\mathbf{Z}_{\text{Tx}} + R_0 \mathbf{I}_N)^{-1} (\mathbf{Z}_{\text{Tx}} - R_0 \mathbf{I}_N). \quad (8)$$

Utilizing this signal-flow graph offers the advantage that the input-output relationship(s) can be derived quickly.⁹

Since this paper focuses on transmitting REMSs, the REMS input consists only of the PA's equivalent voltages $\mathbf{v}_{\text{Tx}}$, and

⁷For simplicity, some definitions used here are not fully consistent with [1].

⁸To be specific, the L^2 space we use here is the Bochner space induced by the measure space $(\Omega, \mathcal{A}, \mu)$ and the Hilbert space $\mathbb{C}^2$, where $\mathcal{A}$ is the respective σ -algebra of Lebesgue measurable sets, and μ is the measure defined by $\mu: \mathcal{A} \rightarrow [0, \infty]$, $X \mapsto \iint_{(\theta, \varphi) \in X} \sin(\theta) d\theta d\varphi$.

⁹The return loop method introduced in [8], for example, can be used to derive input-output relationships. The return loop method generalizes Mason's gain formula [9] to nodes in *vector spaces* of arbitrary dimension.

the only output is the outgoing far-field radiation pattern $\mathbf{a}_{\text{F}}$. The corresponding input-output relationship is given by

$$\mathbf{a}_{\text{F}} = \mathbb{G}_{\mathbf{v}_{\text{Tx}}}^{\mathbf{a}_{\text{F}}} \mathbf{v}_{\text{Tx}}. \quad (9)$$

Here, we use the gain operator

$$\mathbb{G}_{\mathbf{v}_{\text{Tx}}}^{\mathbf{a}_{\text{F}}} \triangleq \mathbb{S}_{\text{RFR}} (\mathbf{I}_M - \mathbf{L}_2)^{-1} \mathbf{S}_{\text{TRT}} (\mathbf{I}_N - \mathbf{L}_1 - \mathbf{L}_3)^{-1} \mathbf{K}_{\mathbf{v}_{\text{Tx}}}, \quad (10)$$

where we use the following auxiliary matrices:

$$\mathbf{L}_1 \triangleq \mathbf{S}_{\text{RF}} \mathbf{S}_{\text{T TT}} \quad (11)$$

$$\mathbf{L}_2 \triangleq \mathbf{S}_{\text{TRR}} \mathbb{S}_{\text{RRR}} \quad (12)$$

$$\mathbf{L}_3 \triangleq \mathbf{S}_{\text{RF}} \mathbf{S}_{\text{T TR}} \mathbb{S}_{\text{RRR}} (\mathbf{I}_M - \mathbf{L}_2)^{-1} \mathbf{S}_{\text{TRT}}. \quad (13)$$

D. Power Metrics

In [1, Def. 5], we define the following *interface-related power metrics* to analyze the power flow between a REMS's subsystems: The *tuning network accepted power* is¹⁰

$$P_{\text{T}} \triangleq \Re\{\mathbf{v}_{\text{T}}^H \mathbf{i}_{\text{T}}\} = \|\mathbf{a}_{\text{T}}\|_2^2 - \|\mathbf{b}_{\text{T}}\|_2^2, \quad (14)$$

the *radiating structure accepted power* is

$$P_{\text{R}} \triangleq \Re\{\mathbf{v}_{\text{R}}^H \mathbf{i}_{\text{R}}\} = \|\mathbf{a}_{\text{R}}\|_2^2 - \|\mathbf{b}_{\text{R}}\|_2^2, \quad (15)$$

and the *total far field radiated power* of a pure transmitter is

$$P_{\text{F}} = P^{\nearrow} = \|\mathbf{a}_{\text{F}}\|_{L^2}^2. \quad (16)$$

Here, with $P^{\nearrow}$ we denote the power radiated by the radiating structure in all directions.

Moreover, in [1, Def. 6], we define the *power amplifiers available power* to analyze the potentially available power flow out of the PAs: For a fixed PA output voltage vector $\mathbf{v}_{\text{Tx}} \in \mathbb{C}^N$, the *power amplifiers available power* is given as

$$P_{\text{A}} = \frac{1}{4} \mathbf{v}_{\text{Tx}}^H \Re\{\mathbf{Z}_{\text{Tx}}\}^{-1} \mathbf{v}_{\text{Tx}}. \quad (17)$$

E. Radiating Intensity and Gain of a REMS

We use the *radiation intensity* $I(\theta, \varphi) \triangleq \|\mathbf{a}_{\text{F}}(\theta, \varphi)\|_2^2$ to determine how much power is transmitted per unit solid angle into the direction (θ, φ) .

Furthermore, we use the *REMS gain*, which, for a given PA output voltage vector $\mathbf{v}_{\text{Tx}} \in \mathbb{C}^N$, and a given direction $(\theta, \varphi) \in \Omega$ is given as

$$G_{\text{REMS}}(\mathbf{v}_{\text{Tx}}; \theta, \varphi) = \frac{1}{P_{\text{A}}} 4\pi I(\theta, \varphi) \Big|_{\mathbf{v}_{\text{Tx}} \text{ fixed}} \quad (18)$$

$$= 16\pi \frac{\|(\mathbb{G}_{\mathbf{v}_{\text{Tx}}}^{\mathbf{a}_{\text{F}}} \mathbf{v}_{\text{Tx}})(\theta, \varphi)\|_2^2}{\mathbf{v}_{\text{Tx}}^H \Re\{\mathbf{Z}_{\text{Tx}}\}^{-1} \mathbf{v}_{\text{Tx}}}. \quad (19)$$

As shown in Fig. 3, the *REMS gain* characterizes the combined effect of a general REMS's (i) *matching efficiency* $\eta_{\text{matching}} \triangleq P_{\text{T}}/P_{\text{A}}$, (ii) *tuning network efficiency* $\eta_{\text{tuning}} \triangleq P_{\text{R}}/P_{\text{T}}$, (iii) *radiation efficiency* $\eta_{\text{radiating}} \triangleq P_{\text{F}}/P_{\text{R}}$, and (iv) *directivity* $D(\theta, \varphi) \triangleq (4\pi I(\theta, \varphi))/P_{\text{F}}$.

¹⁰For definitions of the voltage and current vectors $\mathbf{v}_{\text{T}}$, $\mathbf{i}_{\text{T}}$, $\mathbf{v}_{\text{R}}$, and $\mathbf{i}_{\text{R}}$, we refer the reader to Fig. 1 or [1].

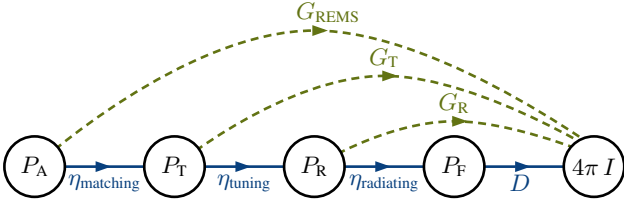


Figure 3. The relationships between the power metrics, including the radiating intensity, are illustrated with solid blue lines. The REMS gain, tuning gain, and radiating gain, which relate the radiation intensity to the corresponding power quantities, are shown with a dashed green line. This figure extends [1, Fig. 6] by additionally including the tuning and radiating gain.

Remark 1 (cf. [1, Rem. 6]). *The REMS gain G_{REMS} can be interpreted as the radiation intensity gain of the REMS relative to a reference system containing a single lossless, isotropic radiator¹¹, whose feed is perfectly matched to the reference system P_A 's output stage provided that P_A of the REMS is equal to P_A of the reference system.*

To complement the REMS gain, we now introduce two additional gain metrics: the *tuning gain* and the *radiating gain*.

Definition 1 (Tuning Gain). *Given a tuning network and a radiating structure. Let $\mathbf{a}_T \in \mathbb{C}^N$ be the power waves flowing into the tuning network and $(\theta, \varphi) \in \Omega$ specify a direction. We define the tuning gain corresponding to $\mathbf{a}_T$ and (θ, φ) as*

$$G_T(\mathbf{a}_T; \theta, \varphi) \triangleq \frac{1}{P_T} 4\pi I(\theta, \varphi) \Big|_{\mathbf{a}_T \text{ fixed}} \quad (20)$$

$$= 4\pi \frac{\|(\mathbb{G}_{\mathbf{a}_T}^{\mathbf{a}_F} \mathbf{a}_T)(\theta, \varphi)\|_2^2}{\|\mathbf{a}_T\|_2^2 - \|\mathbb{G}_{\mathbf{a}_T}^{\mathbf{b}_T} \mathbf{a}_T\|_2^2}, \quad (21)$$

with

$$\mathbb{G}_{\mathbf{a}_T}^{\mathbf{a}_F} \triangleq \mathbf{S}_{\text{RFR}} (\mathbf{I}_M - \mathbf{L}_2)^{-1} \mathbf{S}_{\text{TRT}} \quad (22)$$

$$\mathbb{G}_{\mathbf{a}_T}^{\mathbf{b}_T} \triangleq \mathbf{S}_{\text{TTR}} + \mathbf{S}_{\text{TRR}} \mathbf{S}_{\text{RRR}} (\mathbf{I}_M - \mathbf{L}_2)^{-1} \mathbf{S}_{\text{TRT}}. \quad (23)$$

Definition 2 (Radiating Gain). *Given a radiating structure. Let $\mathbf{a}_R \in \mathbb{C}^M$ be the power waves flowing into the radiation structure and $(\theta, \varphi) \in \Omega$ specify a direction. We define the radiating gain corresponding to $\mathbf{a}_R$ and (θ, φ) as*

$$G_R(\mathbf{a}_R; \theta, \varphi) \triangleq \frac{1}{P_R} 4\pi I(\theta, \varphi) \Big|_{\mathbf{a}_R \text{ fixed}} \quad (24)$$

$$= 4\pi \frac{\|(\mathbb{G}_{\mathbf{a}_R}^{\mathbf{a}_F} \mathbf{a}_R)(\theta, \varphi)\|_2^2}{\|\mathbf{a}_R\|_2^2 - \|\mathbb{G}_{\mathbf{a}_R}^{\mathbf{b}_R} \mathbf{a}_R\|_2^2}, \quad (25)$$

with

$$\mathbb{G}_{\mathbf{a}_R}^{\mathbf{a}_F} \triangleq \mathbf{S}_{\text{RFR}} \quad \text{and} \quad \mathbb{G}_{\mathbf{a}_R}^{\mathbf{b}_R} \triangleq \mathbf{S}_{\text{RRR}}. \quad (26)$$

III. A LOW-COMPLEXITY BEAMFORMING ARCHITECTURE

We now demonstrate that with *physically consistent* modeling, one can jointly perform hybrid beamforming and matching in an ultra-dense massive antenna array using inexpensive RF

¹¹A coherent isotropic radiator cannot exist [10].

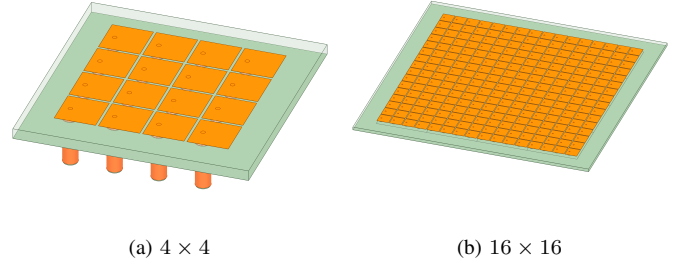


Figure 4. Screenshots from Ansys HFSS of the used radiating structures: A 4×4 and a 16×16 ultra dense arrays of square patch antennas with an inter-element spacing of $\frac{\lambda}{4}$. The arrays lie on the $\theta = 90^\circ$ plane.

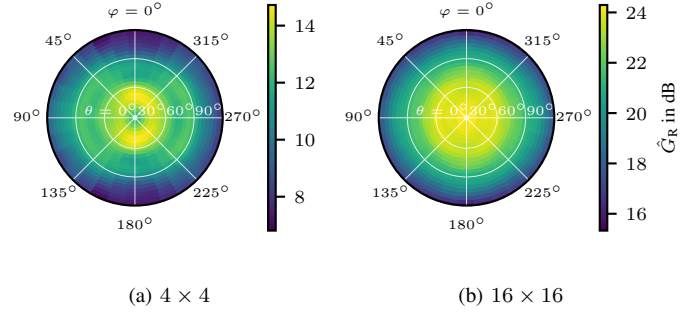


Figure 5. Radiating structure gain $\hat{G}_R(\theta, \varphi)$ of the used radiating structures for $(\theta, \varphi) \in [0, \pi/2] \times [0, 2\pi]$.

circuitry. Concretely, given the radiating structure detailed next, we propose a *cost-efficient*¹² beamsteering architecture.

A. Radiation Structure

We consider the 4×4 and 16×16 antenna arrays depicted in Fig. 4 operating at 12 GHz. We use square patch antennas with side lengths of approximately $\frac{\lambda}{4}$. The antennas are excited by a coaxial port, and the input impedance of a single element, when embedded in the array, is roughly 50Ω .¹³ In both arrays, the inter-element spacing between element centers is $\frac{\lambda}{4}$; hence, they are strongly coupled, ultra-dense arrays.

We characterized both arrays, specifically their *inter-element coupling operator* $\mathbf{S}_{\text{RRR}}$ and *transmitting operator* $\mathbf{S}_{\text{RFR}}$, using full-wave EM simulations in Ansys HFSS.

B. Gain Metrics

The *REMS gain* $G_{\text{REMS}}(\mathbf{v}_{\text{Tx}}; \theta, \varphi)$ defined in [1, Def. 7] provides a comprehensive metric for evaluating hybrid beamformers, as it accounts for directivity, ohmic losses, impedance mismatch, and the effects of both analog and digital beamforming. Two further gain metrics useful for assessing individual subsystems of a REMS are the *tuning gain* $G_T(\mathbf{a}_T; \theta, \varphi)$ (see Def. 1) and the *radiating gain* $G_R(\mathbf{a}_R; \theta, \varphi)$ (see Def. 2).

All three gain metrics depend not only on the direction (θ, φ) but also on the current beamforming scheme, via $\mathbf{v}_{\text{Tx}}$, $\mathbf{a}_T$, and $\mathbf{a}_R$. In the following, when evaluating a gain metric in

¹²We use *cost-efficient* to emphasize that our hybrid beamformer needs only a fraction of the RF chains of an *all-digital* architecture and relies on low-cost RF switches rather than expensive phase shifters for analog beamforming.

¹³The exact input impedances vary due to mutual coupling.

a given direction (θ, φ) , we choose $\mathbf{v}_{\text{Tx}} \in \mathbb{C}^N$, $\mathbf{a}_{\text{T}} \in \mathbb{C}^N$, and $\mathbf{a}_{\text{R}} \in \mathbb{C}^M$ so as to maximize $G_{\text{REMS}}(\mathbf{v}_{\text{Tx}}; \theta, \varphi)$, $G_{\text{T}}(\mathbf{a}_{\text{T}}; \theta, \varphi)$, and $G_{\text{R}}(\mathbf{a}_{\text{R}}; \theta, \varphi)$, respectively. We now show how to obtain these voltage and power wave vectors.

It follows directly from (19), (21), and (25) that, in all three cases, the respective maximization problem can be expressed as

$$\arg \max_{\mathbf{x}} G(\mathbf{x}; \theta, \varphi) = \frac{\mathbf{x}^H \mathbf{A}(\theta, \varphi) \mathbf{x}}{\mathbf{x}^H \mathbf{B} \mathbf{x}}, \quad (27)$$

where $\mathbf{x}$ denotes the voltage or power-wave vector being optimized, and $\mathbf{A}(\theta, \varphi)$ and $\mathbf{B}$ are Hermitian complex matrices chosen such that $\mathbf{x}^H \mathbf{A}(\theta, \varphi) \mathbf{x} = 4\pi I(\theta, \varphi)$ and $\mathbf{x}^H \mathbf{B} \mathbf{x}$ equals a power metric.¹⁴ We note that the radiation intensity $I(\theta, \varphi)$ is nonnegative by definition, and we assume the power metrics P_{A} , P_{T} , and P_{R} are strictly positive. From $I(\theta, \varphi) \geq 0$, it follows that $\mathbf{A}(\theta, \varphi)$ is positive semidefinite and from $P_{\text{A}}, P_{\text{T}}, P_{\text{R}} > 0$ it follows that $\mathbf{B}$ is positive definite. Consequently, (27) is a Rayleigh-quotient maximization whose solution is well known [11]:

$$\hat{\mathbf{x}}(\theta, \varphi) \triangleq \mathbf{B}^{-\frac{1}{2}} \mathbf{c}_1(\theta, \varphi) \in \arg \max_{\mathbf{x}} G(\mathbf{x}; \theta, \varphi). \quad (28)$$

Here, the matrix $\mathbf{B}^{-\frac{1}{2}}$ is a positive definite square root of $\mathbf{B}^{-1}$ and $\mathbf{c}_1(\theta, \varphi)$ is the dominant eigenvector of $\mathbf{C}(\theta, \varphi) \triangleq \mathbf{B}^{-\frac{1}{2}} \mathbf{A}(\theta, \varphi) \mathbf{B}^{-\frac{1}{2}}$.

Let $\hat{\mathbf{v}}_{\text{Tx}}(\theta, \varphi)$, $\hat{\mathbf{a}}_{\text{T}}(\theta, \varphi)$, and $\hat{\mathbf{a}}_{\text{R}}(\theta, \varphi)$ be optimal voltage and power wave vectors according to (28). For the rest of this paper, we use the following shorthand notation:

$$\hat{G}_{\text{REMS}}(\theta, \varphi) \triangleq G_{\text{REMS}}(\hat{\mathbf{v}}_{\text{Tx}}(\theta, \varphi); \theta, \varphi) \quad (29)$$

$$\hat{G}_{\text{T}}(\theta, \varphi) \triangleq G_{\text{T}}(\hat{\mathbf{a}}_{\text{T}}(\theta, \varphi); \theta, \varphi) \quad (30)$$

$$\hat{G}_{\text{R}}(\theta, \varphi) \triangleq G_{\text{R}}(\hat{\mathbf{a}}_{\text{R}}(\theta, \varphi); \theta, \varphi). \quad (31)$$

Remark 2. By definition, for every REMS, it holds that

$$\hat{G}_{\text{REMS}}(\theta, \varphi) \leq \hat{G}_{\text{T}}(\theta, \varphi) \leq \hat{G}_{\text{R}}(\theta, \varphi) \quad \forall (\theta, \varphi) \in \Omega. \quad (32)$$

Remark 3. For a given radiating structure, the radiating gain $\hat{G}_{\text{R}}(\theta, \varphi)$ quantifies the fundamentally maximal REMS gain $\hat{G}_{\text{REMS}}(\theta, \varphi)$ that can be achieved with that radiating structure. Consequently, the deviation between $\hat{G}_{\text{R}}(\theta, \varphi)$ and $\hat{G}_{\text{REMS}}(\theta, \varphi)$ quantifies losses due to non-ideal matching and due to a non-ideal tuning network.¹⁵

Remark 4. For a given tuning network and radiating structure, the tuning gain $\hat{G}_{\text{T}}(\theta, \varphi)$ quantifies the fundamentally maximal REMS gain $\hat{G}_{\text{REMS}}(\theta, \varphi)$ that can be achieved with that tuning network and radiating structure. The deviation between $\hat{G}_{\text{T}}(\theta, \varphi)$ and $\hat{G}_{\text{REMS}}(\theta, \varphi)$ quantifies the losses due to non-ideal matching.

¹⁴Explicitly, for a fixed direction $(\theta, \varphi) \in \Omega$, these matrices are given as follows: When maximizing G_{REMS} , we use $\mathbf{A} = 4\pi(\mathbb{G}_{\text{vTx}}^{\text{aF}})^H \mathbb{G}_{\text{vTx}}^{\text{aF}}$ and $\mathbf{B} = 1/49\{\mathbf{Z}_{\text{Tx}}\}^{-1}$; when maximizing G_{T} , we use $\mathbf{A} = 4\pi(\mathbb{G}_{\text{aT}}^{\text{aF}})^H \mathbb{G}_{\text{aT}}^{\text{aF}}$ and $\mathbf{B} = \mathbf{I}_N - (\mathbb{G}_{\text{aT}}^{\text{bT}})^H \mathbb{G}_{\text{aT}}^{\text{bT}}$; and when maximizing G_{R} , we use $\mathbf{A} = 4\pi(\mathbb{G}_{\text{aR}}^{\text{aF}})^H \mathbb{G}_{\text{aR}}^{\text{aF}}$ and $\mathbf{B} = \mathbf{I}_M - (\mathbb{G}_{\text{aR}}^{\text{bR}})^H \mathbb{G}_{\text{aR}}^{\text{bR}}$. Note that, to simplify notation, we refrain from explicitly indicating the dependence on the direction (θ, φ) in this footnote.

¹⁵Note that an ideal tuning network ensures optimal beamforming, regardless of the number of RF chains used.

Fig. 5 shows the radiating gain $\hat{G}_{\text{R}}(\theta, \varphi)$ for the two used radiating structures.

C. Objective

Given the radiating structures described in Sec. III-A, our objective is to achieve a high REMS gain $\hat{G}_{\text{REMS}}(\theta, \varphi)$ using a *cost-efficient* architecture for all directions in front of the arrays (i.e., we consider $(\theta, \varphi) \in [0, \pi/2] \times [0, 2\pi)$, with the arrays lying in the $\theta = 90^\circ$ plane).

D. Proposed Hybrid Beamformer Architecture

To achieve the objective in Sec. III-C, we propose the following hybrid beamformer architecture.

We divide the antenna array into 4×4 antenna tiles. Accordingly, the array shown in Fig. 4a contains one tile, whereas the array in Fig. 4b contains sixteen tiles. Each tile is fed by one PA, whose output impedance is set to $50/16 \Omega$.

The tuning network connecting each PA to sixteen antenna elements is shown in Fig. 6a. It comprises three main components: (i) a matching network located directly after the PA, (ii) a 16-way power splitter that distributes the signal to the individual antenna elements, and (iii) a switch unit placed ahead of each antenna, replacing traditional, costly phase shifters.

The proposed matching-network architecture is shown in Fig. 6b. The network is a triple-stub tuner, in which RF switches determine whether each stub is active and whether it is terminated in an open or short circuit. Using three stubs provides an acceptable trade-off between high matching efficiency and the additional losses introduced by (realistic and thus, slightly lossy) switches.

The proposed switch-unit architecture is depicted in Fig. 6c. Each switch unit comprises two RF switches, enabling four states: (i) the signal passes through largely unchanged when only the top switch is closed; (ii) the signal passes through with a relative phase shift of approximately 180° when only the bottom switch is closed; (iii) the input signal is reflected at an open-circuit when both switches are open; and (iv) the signal is reflected at a short-circuit when both switches are closed.

We use realistic RF-switch models, which introduce ohmic losses, imperfect isolation, and unwanted reflections.¹⁶ While these effects are significant, they can be modeled exactly with our circuit-theoretic approach.

E. Benchmark Architectures

In our experiments, we compare the proposed architecture in Sec. III-D to the following benchmark architectures.

1) *All-Digital, Ideal Matching*: An all-digital beamformer in which the N PAs are ideally matched to the $M = N$ antenna elements (see Fig. 7a) represents the fundamentally best system that can theoretically be realized with a given radiating structure. Consequently, following Rem. 3, its REMS gain $\hat{G}_{\text{REMS}}(\theta, \varphi)$ is equal to the radiating gain $\hat{G}_{\text{R}}(\theta, \varphi)$ of the used radiating structure. Following Rem. 2, we use the

¹⁶The switch parameters were obtained from a Cadence Virtuoso simulation of a realistic RF switch.

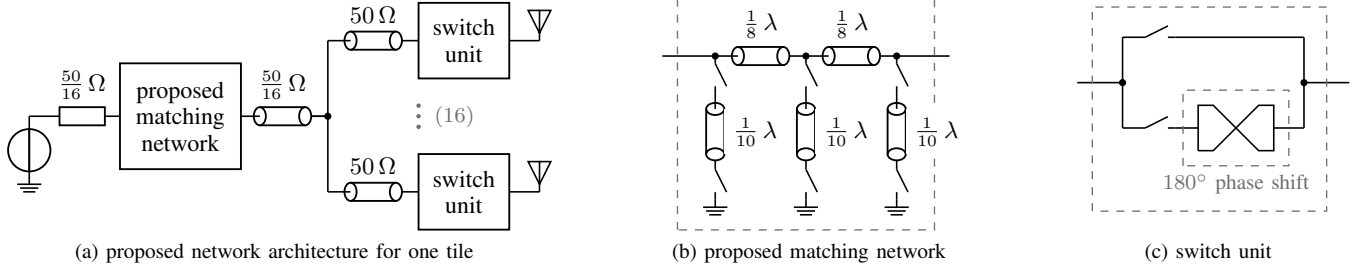


Figure 6. Proposed *cost-efficient* hybrid beamformer architecture. Sixteen antenna elements form one tile. Each tile is fed by one RF chain, which we represent by the Thévenin equivalent of the RF chain's PA. The PAs are connected to the antenna elements via (i) a reconfigurable matching network, (ii) a 16-way power splitter, and (iii) reconfigurable switch units, with which one can perform analog beamforming. When implementing multiple tiles in a system, one can additionally perform hybrid digital-analog beamforming.

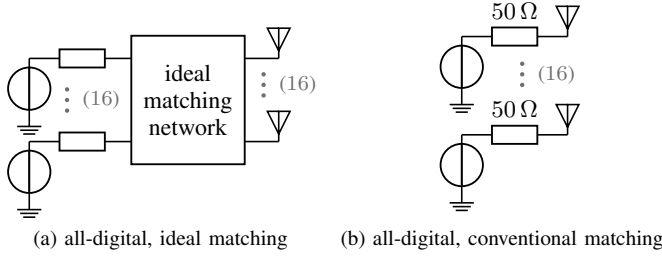


Figure 7. Architecture of the all-digital with ideal matching and the all-digital with conventional matching benchmark architectures.

performance of this all-digital, ideally matched benchmark to quantify how close a given system comes to the fundamentally best achievable performance for the given radiating structure.

2) *All-Digital, Conventional Matching*: The ideal benchmark system described above in Sec. III-E1, with current technologies, is either not practically realizable or only realizable at high costs. A more realistic benchmark is an all-digital beamformer with conventional matching, i.e., in which the RF chains are directly connected to the antenna array and the PA output impedance is approximately matched to the antenna input impedance, which in our case is close to $50\ \Omega$ (see Fig. 7b).

3) *Proposed Architecture with Ideal Matching*: This benchmark system corresponds to the proposed architecture in Fig. 6a with the proposed matching network replaced by an ideal matching network. Following Rem. 4, its REMS gain $\hat{G}_{\text{REMS}}(\theta, \varphi)$ is equal to the tuning gain $\hat{G}_T(\theta, \varphi)$ of the proposed architecture. Following Rem. 2 and Rem. 4, we use the performance of this benchmark to quantify how close our proposed architecture comes to the fundamentally best achievable performance for the given radiating structure with ideal PA-matching assumed.

F. Optimization of RF Switch Positions

Thanks to the efficiency of our REMS model, for the 4×4 array, the optimal switch positions¹⁷ can be found in a few days via exhaustive search. For the 16×16 array, however, the search space is too large for exhaustive search to remain feasible. Thus, we use a heuristic algorithm based on coordinate ascent to determine favorable switch states.

¹⁷There exist in total 2^{38} different combinations for the 4×4 tile.

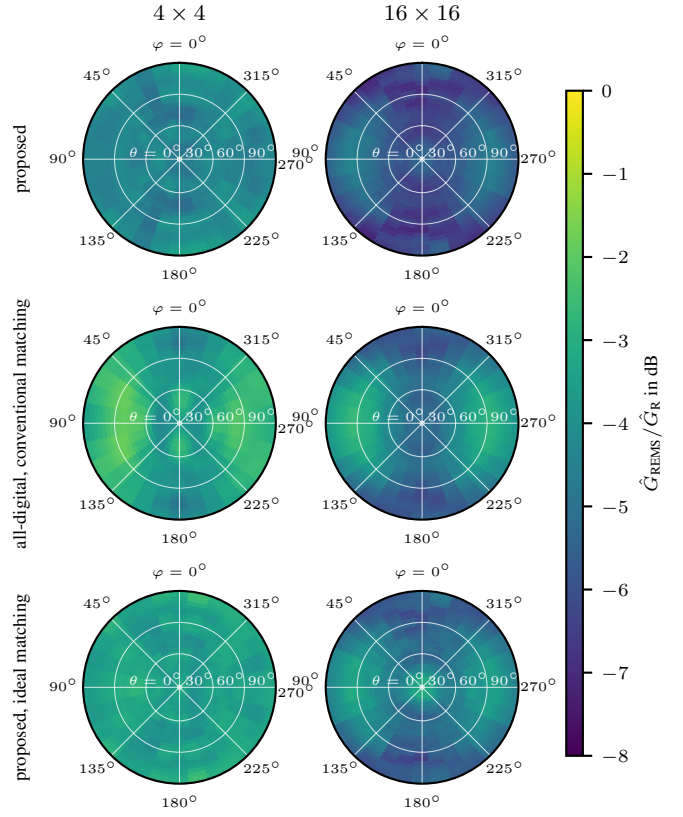


Figure 8. Relative REMS gain, $\hat{G}_{\text{REMS}}(\theta, \varphi) / \hat{G}_R(\theta, \varphi)$ of the various analyzed architectures. This ratio between REMS gain and radiating gain quantifies how closely each architecture approaches the maximal performance achievable with the given arrays. The results are shown for $(\theta, \varphi) \in [0, \pi/2] \times [0, 2\pi]$.

IV. RESULTS

We begin by analyzing the ratio of the REMS gain $\hat{G}_{\text{REMS}}(\theta, \varphi)$ to the radiating gain $\hat{G}_R(\theta, \varphi)$. As discussed in Rem. 3, this ratio quantifies how closely a system approaches the fundamental performance limit for a given radiating structure and direction (θ, φ) . We compare three specific systems: (i) the proposed architecture (see Sec. III-D), (ii) the all-digital architecture with conventional matching

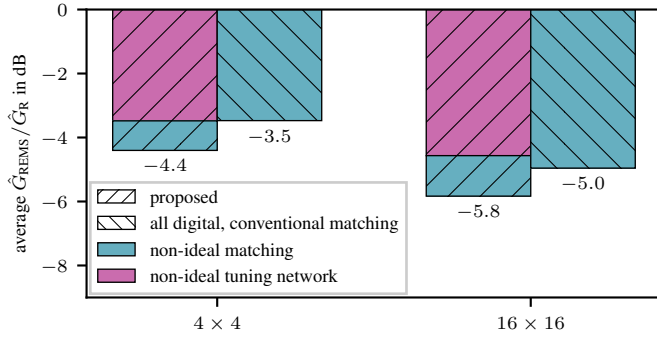


Figure 9. Averaged relative REMS gain, $\hat{G}_{\text{REMS}}(\theta, \varphi) / \hat{G}_{\text{R}}(\theta, \varphi)$, for the proposed architecture and for an all-digital, conventionally matched architecture. The average is defined as the median over $(\theta, \varphi) \in [0, \pi/2] \times [0, 2\pi]$. For the all-digital architecture with a conventional matching network, gain is lost purely due to non-ideal matching. However, for the proposed architecture, we can distinguish between gain losses due to non-ideal matching and those due to the non-ideal tuning network, which includes, among other effects, losses from non-ideal beamforming. The proposed architecture achieves only approximately 1 dB lower gain, while using $16 \times$ fewer RF chains.

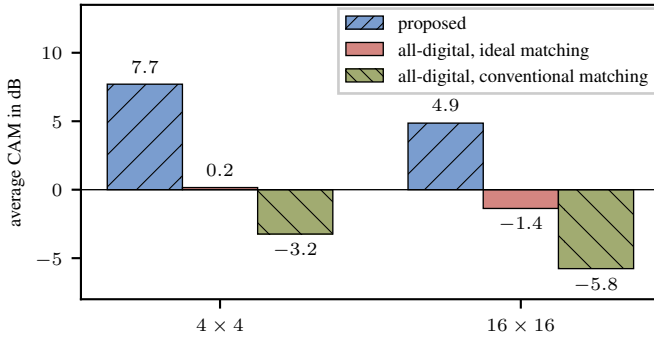


Figure 10. Averaged cost-aware metric CAM $\triangleq \hat{G}_{\text{REMS}}(\theta, \varphi) / \# \text{RF chains}$ for the proposed architecture, the all-digital, ideally matched architecture, and the all-digital, conventionally matched architecture. The average is defined as the median over $(\theta, \varphi) \in [0, \pi/2] \times [0, 2\pi]$. For both arrays, the proposed architecture offers significantly higher gain per RF chain compared to both of the considered all-digital architectures.

(see Sec. III-E2), and (iii) the proposed architecture with ideal matching (see Sec. III-E3). Furthermore, using the fact that the REMS gain of the proposed architecture with ideal matching corresponds to the tuning gain of the proposed architecture and following Rem. 4, one can quantify the proposed architecture's losses due to non-ideal matching.

Fig. 9 illustrates the average ratio $\hat{G}_{\text{REMS}}(\theta, \varphi) / \hat{G}_{\text{R}}(\theta, \varphi)$ for the proposed architecture and the all-digital architecture with conventional matching. Remarkably, despite using $16 \times$ fewer RF chains, the proposed architecture performs within 4.4 dB (for the 4×4 array) and 5.8 dB (for the 16×16 array) of the fundamentally best system (all-digital with ideal matching). Moreover, when compared to the practical benchmark of an all-digital beamformer with conventional matching, the proposed architecture achieves only approximately 1 dB lower gain, while again using $16 \times$ fewer RF chains.

To further demonstrate the cost efficiency of our architecture,

we introduce a cost-aware metric $\text{CAM}(\theta, \varphi) \triangleq \frac{\hat{G}_{\text{REMS}}(\theta, \varphi)}{\# \text{RF chains}}$, which quantifies the achievable gain per RF chain. In Fig. 10, we visualize the average value of this cost-aware metric for (i) the proposed architecture, (ii) the all-digital architecture with ideal matching, and (iii) the all-digital architecture with conventional matching. For both arrays, the proposed architecture offers approximately 7 dB higher gain per RF chain compared to the all-digital architecture with ideal matching and approximately 11 dB higher gain per RF chain compared to the all-digital architecture with conventional matching.

V. CONCLUSIONS

We have applied the *efficient* and *physically consistent* model proposed in [1] to design and analyze a novel beamforming architecture. Specifically, we have proposed a joint hybrid beamforming and matching architecture that relies on cost- and power-efficient RF switches and transmission lines. Numerical results based on full-wave EM simulations have demonstrated that the proposed architecture attains a beamforming gain comparable to all-digital architectures with conventional matching, while requiring only a fraction of the expected cost. Furthermore, we have shown that our architecture scales favorably with the number of antennas, making it a promising candidate for very large antenna arrays.

While this case study analyzed an inter-antenna spacing of $\frac{\lambda}{4}$, the trade-off between the added complexity of this antenna spacing relative to a standard $\frac{\lambda}{2}$ array warrants further investigation. Overall, our results highlight how physically consistent modeling enables the design and analysis of new beamforming architectures, opening up novel opportunities for energy- and cost-efficient large-scale antenna systems.

REFERENCES

- [1] A. Stutz-Tirri, G. Schwan, and C. Studer, "Efficient and physically consistent modeling of reconfigurable electromagnetic structures," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 1610–1633, Feb. 2025.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] 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 Sys. Mag.*, vol. 19, no. 2, pp. 33–58, May 2019.
- [4] S. Dutta, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter, and S. Rangan, "A case for digital beamforming at mmWave," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 756–770, Feb. 2020.
- [5] M. Eslami Rasekh, N. Hosseinzadeh, U. Madhow, and M. Rodwell, "Low-resolution architectures for power-efficient scaling of mmWave phased array receivers," *J. Signal Process. Syst.*, vol. 94, pp. 1005–1014, Jun. 2022.
- [6] *Quantities and units Part 2: Mathematics*, ISO 80000-2, 2019.
- [7] M. Nietro-Vesperinas, *Scattering and Diffraction in Physical Optics*, 2nd ed. Singapore: World Scientific, 2006.
- [8] D. Riegler and P. Lin, "Matrix signal flow graphs and an optimum topological method for evaluating their gains," *IEEE Trans. Circuit Theory*, vol. 19, no. 5, pp. 427–435, Sep. 1972.
- [9] S. J. Mason, "Feedback theory—further properties of signal flow graphs," *Proc. IRE*, vol. 44, no. 7, pp. 920–926, Jul. 1956.
- [10] H. F. Mathis, "A short proof that an isotropic antenna is impossible," *Proc. IRE*, vol. 39, no. 8, p. 970, Aug. 1951.
- [11] R.-C. Li, "Rayleigh quotient based optimization methods for eigenvalue problems," in *Matrix Functions and Matrix Equations*, Z. Bai, W. Gao, and Y. Su, Eds. Singapore: World Scientific, 2015, pp. 76–108.