Received 25 November 2020; accepted 19 December 2020. Date of publication 29 December 2020; date of current version 19 January 2021.

Digital Object Identifier 10.1109/OJCOMS.2020.3048063

Intelligent Surface-Aided Transmitter Architectures for Millimeter-Wave Ultra Massive MIMO Systems

VAHID JAMALI [1] (Member, IEEE), ANTONIA M. TULINO [2,3] (Fellow, IEEE), GEORG FISCHER [4] (Senior Member, IEEE), RALF R. MÜLLER [1] (Senior Member, IEEE), AND ROBERT SCHOBER [1] (Fellow, IEEE)

[1] Institute for Digital Communications, Friedrich-Alexander University Erlangen–Nürnberg, 91058 Erlangen, Germany
[2] Nokia Bell Labs, Holmdel, NJ 07733, USA
[3] Department of Electrical Engineering, Università degli Studi di Napoli Federico II, 80138 Naples, Italy
[4] Institute for Electronics Engineering, Friedrich-Alexander University Erlangen–Nürnberg, 91058 Erlangen, Germany
CORRESPONDING AUTHOR: V. JAMALI (e-mail: vahid.jamali@fau.de)
This article was presented in part at IEEE ICC 2019 [1].

ABSTRACT In this article, we study two novel massive multiple-input multiple-output (MIMO) transmitter architectures for millimeter wave (mmWave) communications which comprise few active antennas, each equipped with a dedicated radio frequency (RF) chain, that illuminate a nearby large intelligent reflecting/transmitting surface (IRS/ITS). The IRS (ITS) consists of a large number of low-cost and energy-efficient passive antenna elements which are able to reflect (transmit) a phase-shifted version of the incident electromagnetic field. Similar to lens array (LA) antennas, IRS/ITS-aided antenna architectures are energy efficient due to the almost lossless over-the-air connection between the active antennas and the intelligent surface. However, unlike for LA antennas, for which the number of active antennas has to linearly grow with the number of passive elements (i.e., the lens aperture) due to the non-reconfigurablility (i.e., non-intelligence) of the lens, for IRS/ITS-aided antennas, the reconfigurablility of the IRS/ITS facilitates scaling up the number of radiating passive elements without increasing the number of costly and bulky active antennas. We show that the constraints that the precoders for IRS/ITS-aided antennas have to meet differ from those of conventional MIMO architectures. Taking these constraints into account and exploiting the sparsity of mmWave channels, we design two efficient precoders; one based on maximizing the mutual information and one based on approximating the optimal unconstrained fully digital (FD) precoder via the orthogonal matching pursuit algorithm. Furthermore, we develop a power consumption model for IRS/ITS-aided antennas that takes into account the impacts of the IRS/ITS imperfections, namely the spillover loss, taper loss, aperture loss, and phase shifter loss. Moreover, we study the effect that the various system parameters have on the achievable rate and show that a proper positioning of the active antennas with respect to the IRS/ITS leads to a considerable performance improvement. Our simulation results reveal that unlike conventional MIMO architectures, IRS/ITS-aided antennas are both highly energy efficient and fully scalable in terms of the number of transmitting (passive) antennas. Therefore, IRS/ITS-aided antennas are promising candidates for realizing the potential of mmWave *ultra massive* MIMO communications in practice.

INDEX TERMS Intelligent reflecting/transmitting surfaces, reflect/transmit array, lens array, hybrid MIMO, mmWave communications, scalability, and energy efficiency.

I. INTRODUCTION

MILLIMETER wave (mmWave) communication systems are promising candidates for realizing the high data rates expected from the next generation of wireless communication networks [2], [3]. These systems will be equipped with a large array of antennas at the transmitter and/or the receiver to cope with the high path loss, limited scattering, and small antenna apertures at mmWave frequencies. However, conventional fully-digital

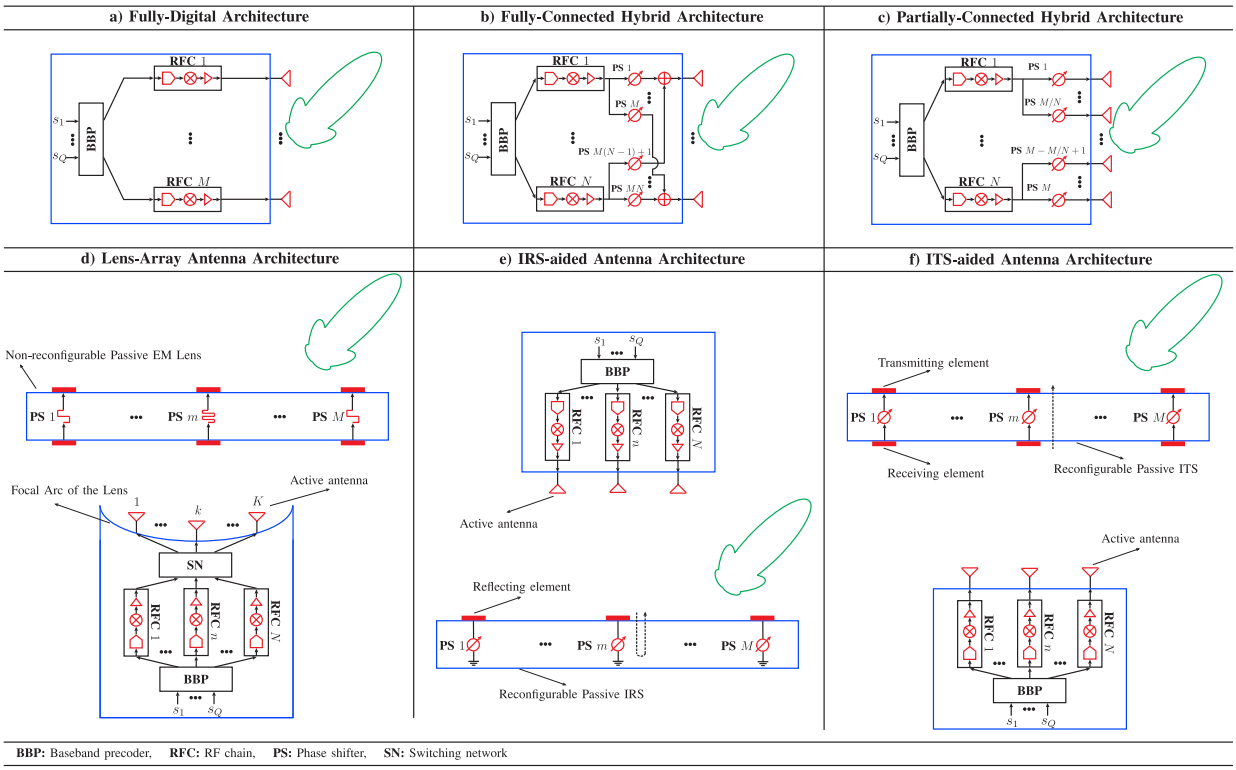


FIGURE 1. **Schematic illustration of the considered massive MIMO architectures. Here, *Q*, *N*, and *M* denote the numbers of data streams, RF chains, and transmit antennas (passive elements in LA and IRS/ITS-aided antennas), respectively, and *K* denotes the number of active antennas of the LA architecture. For the LA architecture, the EM lens is a non-reconfigurable surface and consists of fixed phase shifters, whereas the IRS/ITS-aided antenna architectures are equipped with intelligent surfaces employing reconfigurable phase shifters.**

(FD) multiple-input multiple-output (MIMO) systems, which connect each antenna to a dedicated radio frequency (RF) chain, see Fig. 1 a), are infeasible for mmWave systems due to the prohibitively high cost and energy consumption of the high resolution analog-to-digital/digital-to-analog converters required for each antenna element [4]. This has motivated researchers to consider hybrid analog-digital MIMO architectures, which tremendously reduce the number of RF chains by moving some of the signal processing operations into the analog domain [5]–[7].

Typically, in hybrid MIMO systems, it is assumed that the output of each RF chain is connected to all antennas via an analog network, see Fig. 1 b). This architecture is referred to as fully-connected (FC) hybrid MIMO and is able to realize the full beamforming gain of massive antenna arrays. Unfortunately, FC hybrid MIMO is not scalable due to the excessive power consumption of the analog network for large numbers of antennas [8]. In particular, the analog network is comprised of RF dividers, combiners, phase shifters, and line connections, which lead to high RF losses and hence reduce energy efficiency. To deal with this issue, partially-connected (PC) hybrid MIMO architectures were proposed in the literature where the output of each RF chain is connected to only a subset of the antennas [4], [9], see Fig. 1 c). Thereby, RF combiners are not needed, and the numbers of phase-shifters and RF lines are reduced. Nevertheless, as is shown in [8] and also in Section V of this article, the power consumption of PC hybrid MIMO still scales with the number of antennas in a similar manner as for FC hybrid MIMO.

To overcome the poor energy efficiency of analog networks, lens array (LA) antennas have been proposed in the literature [10]–[13]. LA antennas consist of two main components, namely an electromagnetic (EM) lens and several active antennas, which are connected by an almost lossless wireless link, see Fig. 1 d). EM lenses are phase-shifting devices which can be realized utilizing an array of passive antenna elements [13] or continuous aperture phase shifting [12]. The active antennas are placed on the focal arc of the EM lens and connected to a small number of RF chains via a switching network. The EM lens transmits the signals of different active antennas in different directions [11], [12]. Therefore, due to the sparsity of mmWave channels, only the few active antennas that lead to transmission in the directions of the scatterers in the channel have to be activated. Note that the number of passive antennas (i.e., the effective lens aperture) determines how narrow a beam can be made whereas the number of active antennas limits the number of beam directions (i.e., the resolution of the LA). Hence, as the number of passive elements increases, the number of available active antennas also has to increase in order to maintain a satisfactory performance. Although passive antennas can be

small and cheap (e.g., simple patch antennas), active antennas that can transmit with high power are typically bulky and expensive (e.g., horn antennas) [13]. Therefore, increasing the number of active antennas constitutes a bottleneck for the scalability of the LA architecture as its implementation becomes costly and bulky for massive MIMO systems.

In order to improve the scalability and energy efficiency of mmWave massive MIMO systems, in this article, we consider two novel massive MIMO transmitter architectures[1] which comprise few active antennas and a large intelligent reflecting surface (IRS), see Fig. 1 e), or a large intelligent transmitting surface (ITS), Fig. 1 f). Recently, intelligent surfaces or metasurfaces have been extensively investigated in the literature with the objective to improve the coverage, spectrum, and energy efficiency of wireless communication systems [15]–[32]. The deployment scenarios of intelligent surfaces can be roughly classified into two categories, namely intelligent surfaces as a part of the wireless channel and as a part of the transceiver architecture.[2] For the former case, the waves incident on the intelligent surface are emitted by a transmitter which is located far away, such that the channel between the transmitter and the intelligent surface is subject to fading, cannot be influenced, and has to be estimated for beamforming design [15]–[22], [24]. In contrast, for the latter case, the intelligent surface is embedded into the transmitter/receiver architecture [25]–[32]. Therefore, the waves incident on the intelligent surface are created by physically close active antennas, such that the channel between the active antennas and the intelligent surface is fixed and can be properly designed during manufacturing. The focus of this article is on the second category, namely IRS/ITS-aided transmitter antenna architectures, and their communication-theoretical modeling and system design.

In the considered IRS/ITS-aided MIMO architectures, each active antenna is equipped with a dedicated RF chain and illuminates the IRS/ITS. The IRS (ITS) consists of a large number of low-cost and energy-efficient passive antenna elements which are able to reflect (retransmit) a phase-shifted version of the incident electromagnetic field. In particular, each passive element receives a superposition of the signals transmitted (over the air) by the active antennas and adds a desired phase shift to the overall signal. In IRS-aided antennas, the phase-delayed signal is then reflected from the array whereas in ITS-aided antennas, the phase-delayed signal is transmitted in the forward direction.[3] Borrowing an analogy from optics, an IRS is analogous to a curved mirror whereas an ITS is analogous to a lens. The curvatures of this imagined mirror and lens are steerable via the reconfigurable phase shifters. Unlike in LA antennas, where the EM lens is non-reconfigurable, i.e., non-intelligent, and the direction of the beam is controlled by the location of the corresponding active antenna, in IRS/ITS-aided antennas, the direction of the beam is directly controlled by the reconfigurable intelligent surface. Therefore, the numbers of active antennas in IRS/ITS-aided antennas do not have to scale with the number of passive elements.[4] The performance gain of IRS/ITS-aided antennas compared to hybrid antennas is often attributed to the feed mechanism. In particular, the analog network which feeds the transmit antennas in hybrid architectures causes a severe power loss, which impedes its implementation in massive MIMO and high-frequency systems [25]. In contrast, the IRS/ITS-aided antennas feed the transmit antennas on the intelligent surface over the air, which is referred to as space feeding mechanism and inherently more energy efficient [25], [35].

We note that IRS/ITS-aided antennas have been widely investigated in the microwave and antenna community, where they are also known as reflect/transmit arrays [25], [26], quasi-optical arrays [27], and reconfigurable arrays [26], [28]. Moreover, various prototypes are available in the literature [26], [29], [30]. Thereby, the performance of these architectures is typically characterized in terms of the beamforming gain. In contrast, in this article, we are interested in multiplexing several data streams and the design of the corresponding precoder. In particular, this article makes the following contributions:

- We first model the precoder structure of IRS/ITS-aided antennas and show that the constraints it has to meet are different compared to those for conventional MIMO architectures. In addition, we introduce several illumination strategies (i.e., choices for the positions and orientations of the active antennas with respect to the intelligent surface) which affect the precoder structure.
- Taking the constraints on the precoder into account and exploiting the sparsity of mmWave channels, we design

1. Following the reciprocity theorem [14], the considered transmitter architectures can be also used as receivers. However, for concreteness, in this article, we focus on transmitter design and leave the receiver design for future work.

2. The term large intelligent surface (LIS) has been also used in the literature to refer to surfaces that are placed on e.g., walls and comprise a large array of transmitting/receiving antennas, see e.g., [33], [34]. Unlike the reflecting/transmitting intelligent surfaces considered in this article, the surfaces in [33], [34] neither reflect nor re-transmit their received signals, instead the input/output of the antenna elements is directly controlled/processed. The main purpose of LISs is to bring the massive MIMO transceivers close to the users and to ensure line-of-sight links.

3. We note that IRS/ITS-aided antennas have several advantages/disadvantages with respect to each other and which one is preferable depends on the particular implementation strategy. For instance, for IRS-aided antennas, the feed position introduces a blocking area whereas this issue does not exist for ITS-aided antennas. On the other hand, IRS facilitates the placement of the control system for the phase shifters on the back side of the surface [25]. Moreover, for the IRS-aided antennas, the magnitude of reflection coefficient is often large (close to one) due to the existence of a metal ground plane that reflects the entire incident wave, see Fig. 1 e). In contrast, for ITS-aided antennas, the intelligent surface has to be properly designed to ensure a large magnitude for the transmission coefficient which in general may lead to a higher implementation complexity [25].

4. LA antennas can be seen as special cases of ITS-aided antennas. In fact, ITS-aided antennas reduce to LA antennas when *i)* the surface is non-reconfigurable and is designed to focus a wavefront arriving perpendicular at the surface to its focal point and *ii)* the active antennas are placed on the focal arc of the surface (i.e., the lens).

two efficient precoders for IRS/ITS-aided antennas; one based on maximizing the mutual information (MI) and one based on approximating the optimal unconstrained FD precoder via orthogonal matching pursuit (OMP). The performance of the MI-based precoder serves as an upper bound for that of the OMP-based precoder; however, the computational complexity of the former is higher than that of the latter. Hence, using the MI-based precoder as a performance upper bound allows us to assess the efficiency of the OMP-based precoder.
- We develop a power consumption model for IRS/ITS-aided antennas that takes into account the impact of several IRS/ITS imperfections, namely the spillover loss, taper loss, aperture loss, and phase shifter loss, as well as the power consumption of the required digital signal processing and the power amplifiers. For a fair performance comparison, for conventional MIMO architectures, we adopt power consumption models from the literature [8], [36] that account for their unique characteristics, e.g., the losses in the RF feed networks of hybrid architectures and the losses in the switching network of the LA architecture. We show that the power consumption of the conventional FD, FC, and PC architectures significantly increases as a function of the number of transmit antennas whereas the power consumption of the LA and IRS/ITS-aided antennas is almost independent of the number of transmit antennas.[5]
- We study the impact of the system parameters on the achievable rate via simulations and show that a proper positioning of the active antennas with respect to the intelligent surface in IRS/ITS-aided antennas leads to a considerable performance improvement. In addition, our simulation results show that in contrast to the conventional FD, FC, PC, and LA MIMO architectures, the IRS/ITS-aided MIMO architectures are fully scalable in terms of the number of transmit antennas.[6] Therefore, IRS/ITS-aided antennas are promising candidates for realizing the potential of mmWave *ultra massive* MIMO in practice.

We note that the recent paper [37] also studied IRS-aided antennas (referred to as reflect arrays) and proposed a corresponding precoder design based on alternating optimization (AO). We employ this precoder as a benchmark and show that the proposed precoders outperform the AO-based precoder in [37] especially for environments with few scatterers. Moreover, the focus of this article is mainly on the scalability and energy efficiency of IRS/ITS-aided antennas which were not studied in [37]. Furthermore, compared to [37], in this article, more detailed models for the precoder structure and the power consumption of the IRS/ITS-aided antennas are provided. Such accurate models are needed for a fair performance comparison of different MIMO architectures and the design of IRS/ITS-aided MIMO structures (e.g., the adjustment of the relative positions and orientations of the active antennas with respect to the intelligent surface) which was not investigated in [37].

In our previous work [31], [32], we investigated symbol-level precoding of intelligent surfaces where the update rate was equal to the symbol rate. In contrast, in this article, we design the precoder for a given channel realization, which implies that the state of the intelligent surface has to be updated once per channel coherence time. Therefore, the complexity of the IRS/ITS-aided antennas considered in this article is lower than that of the symbol-level precoding schemes studied in [31], [32]. Finally, we note that this article extends its conference version [1] in the following directions: *i)* The precoder structure for the IRS/ITS-aided antenna architectures is formulated more carefully in this article (see assumptions A1-A4 in Section II-A) and the special case of hypothetical uniform illumination is considered, see Corollary 1, which is not included in [1]. *ii)* The power consumption of the IRS/ITS-aided antenna architectures is discussed in more detail via an example setup in Section III-D. Moreover, we establish lower and upper bounds on the power radiated by the active antennas in terms of the desired power to be radiated by the intelligent surface into the channel, see Lemma 1, which is not considered in [1]. *iii)* In this article, we propose two precoder designs, namely MI-based and OMP-based precoders, whereas in [1], only the OMP-based precoder is given. *iv)* We include the LA architecture as a new benchmark scheme in this article and provide corresponding models for the precoder structure and power consumption using the same unified framework as for the other benchmark architectures, see Table 2.

The rest of this article is organized as follows. In Section II, we provide the considered system, channel, and signal models. Mathematical models for the precoder structure and the power consumption of the considered IRS/ITS-aided antenna architectures are presented in Section III. In Section IV, two different precoder designs for IRS/ITS-aided antennas are developed. Simulation results are provided in Section V, and conclusions and directions for future research are presented in Section VI.

Notations: Bold capital and small letters are used to denote matrices and vectors, respectively. $\|\mathbf{A}\|_F$, $\mathrm{trace}(\mathbf{A})$, $\mathbf{A}^{\mathsf{T}}$, and $\mathbf{A}^{\mathsf{H}}$ denote the Frobenius norm, trace, transpose, and Hermitian of matrix $\mathbf{A}$, respectively. $\mathbb{E}\{\cdot\}$ represents expectation and $I(\mathbf{x};\mathbf{y})$ denotes the MI between random variables (RVs) $\mathbf{x}$ and $\mathbf{y}$. $|a|$ and $\angle a$ denote the absolute value and the angle of complex number a in polar coordinates, respectively. In addition, $|\mathbf{A}|$ denotes the determinant of square matrix $\mathbf{A}$. The big O notation, $g(x) = O(f(x))$, indicates $\lim_{x\to\infty} |g(x)/f(x)| \leq k$ for some fixed k, where $0 < k < \infty$. For a real number x, $[x]^+ = \max\{0, x\}$ and $y = \lceil x \rceil$ is the smallest integer number y for which $y \geq x$ holds. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

5. Throughout this article, we refer to the passive elements of the LA and IRS/ITS-aided architectures as transmit antennas, too.

6. We note that, at mmWave frequencies, hundreds and even thousands of passive antenna elements can be accommodated in a compact design. For example, a 10 cm-by-10 cm intelligent surface can contain approximately 350 and 1600 passive elements at frequencies of 28 GHz and 60 GHz, respectively, with element spacing of half a wavelength.

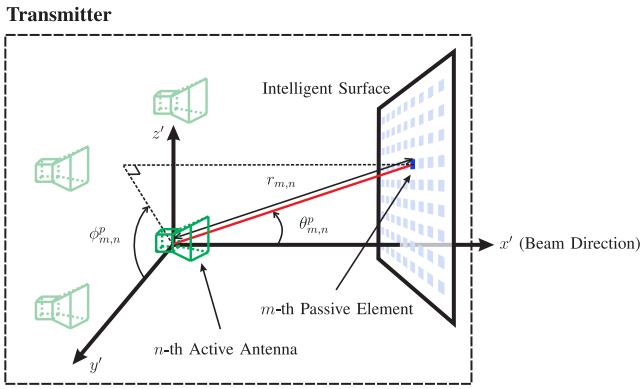


FIGURE 2. Schematic illustration of positioning of the active antennas and the intelligent surface.

denotes a complex Gaussian RV with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Furthermore, $\mathbf{0}_n$ and $\mathbf{0}_{n\times m}$ denote a vector of size n and a matrix of size $n \times m$, respectively, whose elements are all zeros. $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbb{C}$ represents the set of complex numbers. $[a(m,n)]_{m,n}$ represents a matrix with element $a(m,n)$ in its m-th row and n-th column. $\mathbf{A}_{m,n}$ and $\mathbf{a}_n$ denote the element in the m-th row and n-th column of matrix $\mathbf{A}$ and the n-th element of vector $\mathbf{a}$, respectively. Finally, vec($\mathbf{A}$) denotes a vector whose elements are the stacked columns of matrix $\mathbf{A}$, and $\otimes$ denotes the Kronecker product.

II. SIGNAL, SYSTEM, AND CHANNEL MODELS

In this section, we present the system, transmit signal, and channel models for the considered IRS/ITS-aided MIMO systems.

A. SYSTEM ARCHITECTURE

We assume that the considered IRS/ITS-aided antennas are equipped with N active antennas and that the intelligent surface comprises M passive antenna elements. Moreover, we assume that each active feed antenna is connected to a dedicated RF chain, i.e., there are N RF chains. To facilitate presentation, we characterize the positions of the passive antenna elements by $(r_{m,n}, \theta^p_{m,n}, \phi^p_{m,n})$, $r_{m,n} \geq 0, \theta^p_{m,n} \in [0, \frac{\pi}{2}], \phi^p_{m,n} \in [0, 2\pi]$, in N different spherical coordinate systems whose respective origins are the positions of the active antennas. Here, $\theta^p_{m,n}$ denotes the elevation angle of passive element m (with respect to the beam direction of active antenna n), $\phi^p_{m,n}$ represents the azimuth angle of passive element m (in the plane perpendicular to the beam direction of active antenna n), and $r_{m,n}$ is the distance between passive element m and active antenna n, see Fig. 2 for an illustration of $(r_{m,n}, \theta^p_{m,n}, \phi^p_{m,n})$. In a similar manner, we characterize the positions of the active antenna elements by $(r_{m,n}, \theta^a_{m,n}, \phi^a_{m,n})$, $r_{m,n} \geq 0, \theta^a_{m,n} \in [0, \frac{\pi}{2}], \phi^a_{m,n} \in [0, 2\pi]$, in M different spherical coordinate systems whose respective origins are the positions of the passive antennas. Here, $\theta^a_{m,n}$ denotes the elevation angle of active element n (with respect to the normal to the surface) and $\phi^a_{m,n}$ represents the azimuth angle of active element n (in the surface plane). Note that the values of $(r_{m,n}, \theta^p_{m,n}, \phi^p_{m,n})$ and $(r_{m,n}, \theta^a_{m,n}, \phi^a_{m,n})$ depend on the specific positioning of the feed antennas and the intelligent surface. Moreover, in order to rigorously present our results, we make the following assumptions.

- A1) We assume the same antenna pattern for all active antennas with antenna gain $G^a(\theta, \phi)$ for elevation angle θ and azimuth angle ϕ defined in the spherical coordinate system with the active antenna as the origin. Similarly, we assume the same antenna pattern for all passive antennas with antenna gain $G^p(\theta, \phi)$ for elevation angle θ and azimuth angle ϕ defined in the spherical coordinate system with the passive antenna as the origin.
- A2) We assume that while each element of the intelligent surface is in the far field of the active antennas; the entire surface is not in the far field. This implies that the electric-field power reaching each element decays with the distance square $\propto \frac{1}{r^2_{m,n}}$ (i.e., higher order terms $\propto \frac{1}{r^q_{m,n}}$, $q > 2$, are negligible); however, the wavefront phase curvature cannot be neglected. These are valid assumptions if $\lambda \ll r_{m,n} \leq 2d_e^2\lambda, \ \forall m, n$ (typically $r_{m,n} \geq 5\lambda, \ \forall m, n$), which holds in practice [14]. Here, λ and d_e denote the wavelength and the largest electric dimension (defined as the physical dimension normalized to λ) of the intelligent surface, respectively.
- A3) We neglect the mutual coupling between the active antennas (passive antenna elements) which is an accurate assumption when the antennas (passive elements) are sufficiently separated, i.e., typically by at least $\lambda/2$ [38].
- A4) We assume that the power radiated from the active antennas is either reflected/forwarded or absorbed by the intelligent surface such that no power from the active antennas directly arrives at the receiver.

B. TRANSMIT SIGNAL MODEL

Let $\mathbf{s} \in \mathbb{C}^{Q\times 1}$ denote the vector of Q independent data streams that we wish to transmit. Moreover, let $\mathbf{x} \in \mathbb{C}^{M\times 1}$ denote the transmit vector radiated from the intelligent surface. Assuming linear precoding, the relation between transmit vector $\mathbf{x}$ and data vector $\mathbf{s}$ is as follows

$$\mathbf{x} = \sqrt{P_{\text{tx}}}\mathbf{F}\mathbf{s}, \tag{1}$$

where $\mathbf{F} \in \mathbb{C}^{M\times Q}$ is the precoder which includes the baseband precoder, the impact of the channel between the active antennas and the intelligent surface, the imperfections of the IRS/ITS, and the phase change introduced by the intelligent surface, see Section III for the detailed modeling of precoder $\mathbf{F}$. Moreover, in (1), P_{tx} denotes the transmit power radiated by the intelligent surface and we assume that $\mathbb{E}\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\} = \mathbf{I}_Q$ and $\|\mathbf{F}\|_F = 1$ hold. In this article, we impose a constraint on the maximum power radiated from the intelligent surface into the channel which is typically enforced by regulations.

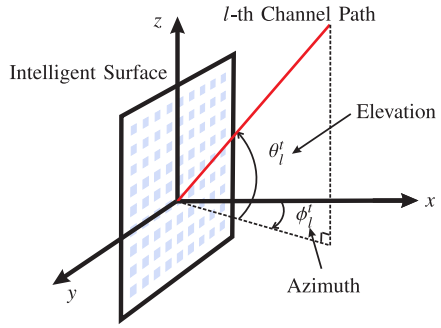


FIGURE 3. Schematic illustration of a given channel AoD with respect to the intelligent surface.

For example, for carrier frequencies 54-66 GHz, the United States Federal Communications Commission enforces a total maximum transmit power of 500 mW (27 dBm) for an emission bandwidth of more than 100 MHz [39]. Alternatively, one can impose a constraint on the power radiated by the active antennas.[7] Although our derivations in Section III and the proposed precoders in Section IV are applicable for both power constraints, we focus on the former power constraint for IRS/ITS-aided antennas since this enables a more straightforward comparison with conventional MIMO architectures.

C. CHANNEL MODEL

We consider a point-to-point MIMO system with the following input-output channel model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{2}$$

where $\mathbf{y} \in \mathbb{C}^{J\times 1}$ denotes receive vector and J is the number of receive antennas. Moreover, $\mathbf{z} \in \mathbb{C}^{J\times 1}$ denotes the additive white Gaussian noise vector at the receiver, i.e., $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_J, \sigma^2\mathbf{I}_J)$ where σ^2 is the noise variance at each receive antenna. Furthermore, $\mathbf{H} \in \mathbb{C}^{J\times M}$ is a low-rank channel matrix accounting for the limited number of scatterers in the channel. In particular, $\mathbf{H}$ is modeled as follows [2], [3], [40]

$$\mathbf{H} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} h_l \mathbf{h}_r\left(\theta_l^r, \phi_l^r\right)\mathbf{h}_t^{\mathsf{H}}\left(\theta_l^t, \phi_l^t\right). \tag{3}$$

where L is the number of effective channel paths and $h_l \in \mathbb{C}$ is the channel coefficient of the l-th path. Moreover, $\mathbf{h}_t(\theta_l^t, \phi_l^t)$ ($\mathbf{h}_r(\theta_l^r, \phi_l^r)$) denotes the transmitter (receiver) antenna array steering vector for angle-of-departure (AoD) (θ_l^t, ϕ_l^t) (angle-of-arrival (AoA) (θ_l^r, ϕ_l^r)) with elevation angle $\theta_l^t \in [-\pi/2, \pi/2]$ ($\theta_l^r \in [-\pi/2, \pi/2]$) and azimuth angle $\phi_l^t \in [-\pi/2, \pi/2]$ ($\phi_l^r \in [-\pi/2, \pi/2]$),[8] see Fig. 3. Assuming the passive elements are uniformly distributed on the intelligent surface which itself lies in the $y-z$ plane, $\mathbf{h}_t(\phi_l^t, \theta_l^t)$ is given as follows [2]

$$\begin{aligned}\mathbf{h}_t&\left(\theta_l^t, \phi_l^t\right)\\ &= \mathrm{vec}\left(\left[e^{j\frac{2\pi d}{\lambda}((m_y-1)\cos(\theta_l^t)\sin(\phi_l^t)+(m_z-1)\sin(\theta_l^t))}\right]_{m_y,m_z}\right)\\ &= \mathrm{vec}\left(\mathbf{h}_t\left(\theta_l^t\right) \otimes \mathbf{h}_t^{\mathsf{T}}\left(\phi_l^t|\theta_l^t\right)\right),\end{aligned} \tag{4}$$

where

$$\mathbf{h}_t\left(\theta_l^t\right) = \left[e^{j\frac{2\pi d}{\lambda}(m_z-1)\sin(\theta_l^t)}\right]_{m_z} \tag{5a}$$

$$\mathbf{h}_t\left(\phi_l^t|\theta_l^t\right) = \left[e^{j\frac{2\pi d}{\lambda}(m_y-1)\cos(\theta_l^t)\sin(\phi_l^t)}\right]_{m_y}. \tag{5b}$$

Here, d is the distance between the passive antenna elements. Moreover, assuming a square surface and that $\sqrt{M}$ is integer, the passive antenna indices along the y- and z-axes are denoted by $m_y, m_z = 1, \ldots, \sqrt{M}$. Furthermore, as shown in (4), the considered steering vector can be decomposed into the Kronecker product of two vectors, namely $\mathbf{h}_t(\theta_l^t)$ and $\mathbf{h}_t(\phi_l^t|\theta_l^t)$ which show the capability of the array to scan different elevation angles and different azimuth angles for a given elevation angle, respectively. We assume an FD receiver equipped with a uniform planar antenna array. Thus, $\mathbf{h}_r(\theta_l^r, \phi_l^r)$ is defined analogous to $\mathbf{h}_t(\theta_l^t, \phi_l^t)$ in (4).

III. MATHEMATICAL MODELING OF THE IRS/ITS-AIDED MIMO ARCHITECTURES

In this section, we first model the constraints that the IRS/ITS-aided antennas impose on the precoder. Then, we present different illumination strategies for the active antennas. Finally, we quantify the total power consumption of IRS/ITS-aided MIMO and further elaborate on the inherent losses for a special case.

A. CONSTRAINTS ON THE PRECODER

Let us define $\bar{\mathbf{x}} = [\bar{x}_1, \ldots, \bar{x}_N]^{\mathsf{T}} \in \mathbb{C}^{N\times 1}$ and $\bar{\mathbf{y}} = [\bar{y}_1, \ldots, \bar{y}_M]^{\mathsf{T}} \in \mathbb{C}^{M\times 1}$ where $\bar{x}_n$ and $\bar{y}_m$ denote the signal transmitted by the n-th active antenna and the signal received at the m-th passive antenna, respectively. As can be seen from Fig. 1, the data stream vector $\mathbf{s}$ is multiplied by the baseband precoder $\mathbf{B}$, fed to the RF chains, and then transmitted over the active antennas/illuminators, i.e., $\bar{\mathbf{x}} = \sqrt{P_{\mathrm{tx}}}\mathbf{B}\mathbf{s}$. We note that the maximum number of independent data streams Q that can be supported is limited by both the numbers of transmitter RF chains N and the rank of the channel matrix $\min\{M, J, L\}$. Moreover, the numbers of transmit and receive antennas are typically large

7. The maximum radiated power is also constrained by the antenna effective isotropic radiated power (EIRP) [39]. The transmit power and EIRP are related according to EIRP $= G_{\max}P_{\mathrm{tx}}$, where $G_{\max}$ is the overall maximum antenna gain of the passive elements on the intelligent surface.

8. We note that for notational convenience, in this article, the definitions of elevation and azimuth angles used for characterization of $(r_{m,n}, \theta_{m,n}^p, \phi_{m,n}^p)$ and $(r_{m,n}, \theta_{m,n}^a, \phi_{m,n}^a)$ are different from those used for characterization of the AoAs/AoDs of the channel, see Figs. 2 and 3. The latter follows the standard definition of spherical coordinate systems in the physics literature [14, Ch. 16] whereas the former is a popular convention used in the radar literature [11].

in mmWave communication systems in order to ensure a sufficient link budget, i.e., $M, J \gg N, L$, which implies that for the maximum number of independent data streams, we typically have $Q \leq \min\{N, L\}$. Based on assumptions A1-A3, the signal that is received at the m-th passive element, $\bar{y}_m$, is obtained as [28]

$$\bar{y}_m = \sum_{n=1}^{N} \sqrt{G^a\left(\theta^p_{m,n}, \phi^p_{m,n}\right) G^p\left(\theta^a_{m,n}, \phi^a_{m,n}\right)} \frac{\lambda e^{-j\frac{2\pi r_{m,n}}{\lambda}}}{4\pi r_{m,n}} \bar{x}_n. \quad (6)$$

The intelligent surface applies a phase shift of $2\pi\beta_m$ to the signal received at the m-th element before reflecting/transmitting it, i.e., $x_m = \bar{y}_m \sqrt{\rho_{\text{srf}}} \exp(j2\pi\beta_m)$, where x_m is the m-th element of $\mathbf{x}$. The signal attenuation caused by the aperture and phase shifter losses is captured by a power efficiency factor denoted by ρ_{srf}, see Section III-C for details. The following proposition formally characterizes the precoder structure for IRS/ITS-aided antennas in matrix form.

Proposition 1: Under Assumptions A1-A4, the precoder $\mathbf{F}$ for IRS/ITS-aided antennas has the form

$$\mathbf{F} = \mathbf{DTB}, \quad (7)$$

where $\mathbf{B} \in \mathbb{C}^{N \times Q}$ is the digital baseband precoder which controls the output signal of the active antennas. $\mathbf{D} \in \mathbb{C}^{M \times M}$ is a diagonal phase-shift matrix which controls the intelligent surface and is given by

$$\mathbf{D} = \text{diag}\left(e^{j2\pi\beta_1}, \ldots, e^{j2\pi\beta_M}\right) \quad (8)$$

with $\beta_m \in [0, 1]$. Moreover, $\mathbf{T} \in \mathbb{C}^{M \times N}$ is a fixed matrix, which depends on the power efficiency of the intelligent surface, denoted by ρ_{srf}, active and passive antenna gains, and the antenna positioning, i.e., $(r_{m,n}, \theta^p_{m,n}, \phi^p_{m,n})$ and $(r_{m,n}, \theta^a_{m,n}, \phi^a_{m,n})$, $\forall m, n$, and is given by

$$\mathbf{T} = \left[\frac{\lambda \sqrt{\rho_{\text{srf}} G^a\left(\theta^p_{m,n}, \phi^p_{m,n}\right) G^p\left(\theta^a_{m,n}, \phi^a_{m,n}\right)}}{4\pi r_{m,n}} e^{-j\frac{2\pi r_{m,n}}{\lambda}} \right]_{m,n}. \quad (9)$$

Proof: The proof follows from rewriting (6) in matrix form $\bar{\mathbf{y}} = \mathbf{T}\bar{\mathbf{x}}$, where the efficiency factor ρ_{srf} is included in $\mathbf{T}$, and combining it with $\mathbf{x} = \mathbf{D}\bar{\mathbf{y}}$, $\bar{\mathbf{x}} = \sqrt{P_{\text{tx}}}\mathbf{Bs}$, and the definition of the linear precoder $\mathbf{x} = \sqrt{P_{\text{tx}}}\mathbf{Fs}$ in (1). ■

We note that both IRS/ITS-aided antennas have identical precoder structures, as given by (7). The precoder for IRS/ITS-aided antennas consists of three parts, namely $\mathbf{B}$, $\mathbf{T}$, and $\mathbf{D}$. Among these three components, $\mathbf{T}$ is fixed and determined during manufacturing whereas $\mathbf{B}$ and $\mathbf{D}$ can be adjusted during online transmission based on the CSI. In Section III-B, we introduce different strategies for the design of matrix $\mathbf{T}$. Moreover, in Section IV, we propose two precoding schemes for optimization of matrices $\mathbf{B}$ and $\mathbf{D}$. We note that $\mathbf{T}$ and particularly the surface power efficiency factor ρ_{srf} may assume different values for IRS/ITS-aided antennas, respectively, see Section III-C for details.

B. ILLUMINATION STRATEGIES

Matrix $\mathbf{T}$ depends on how the active antennas illuminate the intelligent surface of the IRS/ITS-aided antennas. Thus, different illumination strategies, including the relative positioning and orientation of the active antennas and the intelligent surface, lead to different designs of matrix $\mathbf{T}$. In the following, we introduce several different illumination strategies. To do so, let us assume that the intelligent surface lies in the $y - z$ plane and its center is at the origin. Furthermore, let us assume that all active antennas have distance R_d from the intelligent surface and are uniformly distributed on a ring of radius R_r with center $\mathbf{c}$, see Fig. 4. We assume this specific geometry for the locations of the active antennas and the intelligent surface in order to be able to rigorously present the proposed illumination strategies. Nevertheless, we note that other geometries are also possible and only affect matrix $\mathbf{T}$ while the precoder designs proposed in Section IV are valid for any given matrix $\mathbf{T}$.

Full Illumination (FI): Here, each active antenna fully illuminates the entire intelligent surface [37], [41]. To achieve this, we assume that all active antennas illuminate the center of the intelligent surface, see Fig. 4 b-i).

Partial Illumination (PI): Note that each passive element can only change the phase of the *superposition* of the signals that it receives from the active antennas. To compensate for this limitation, a natural option is to have the active antennas illuminate disjoint subsets of the intelligent surface. Assuming that the passive elements responsible for a given active antenna are physical neighbors, for PI, each active antenna illuminates the center of the area occupied by its respective passive elements, see Fig. 4 b-ii).

Separate Illumination (SI): SI is an extreme special case of PI where the signals of different active antennas are physically shielded such that each part of the intelligent surface only receives the signal of one active antenna, see Fig. 4 b-iii). Note that due to the wide beam patterns of the active antennas, even under PI, they illuminate not only their respective subset of passive elements, but all other passive elements as well. This causes interference between the signals from different RF chains which is avoided by SI.

Blockage-Free PI: In this article, in general, for the IRS-aided architecture, we neglect that the active antennas may partially block the reflected RF signal. Therefore, unless otherwise stated, we assume $\mathbf{c} = (R_d, 0, 0)$ which yields the minimum average distance between the active antennas and passive elements, i.e., $\frac{1}{MN}\sum_n \sum_m r_{m,n}$ and thus the over-the-air losses are minimized. In practice, IRS-aided antennas are designed to support transmission/reception for a limited range of AoDs/AoAs and the active antennas are placed outside this range to avoid blockage [41], [42]. To achieve this, for blockage-free illumination, we assume that the active antennas are placed on one side of the intelligent surface (e.g., $\mathbf{c} = (R_d, d\sqrt{M}/2, d\sqrt{M}/2)$) in order to avoid blocking the AoAs/AoDs pointing to the other side [41], [42], see

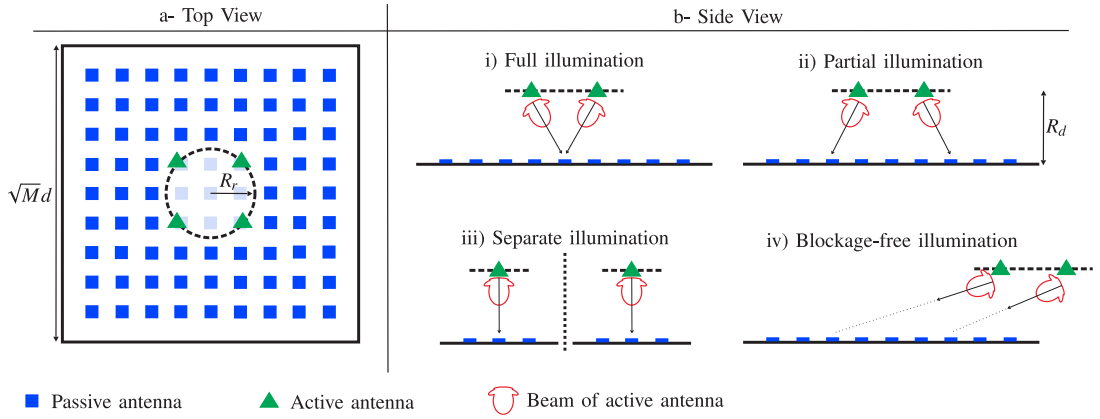


FIGURE 4. Schematic illustration of the arrangement of the active antennas and the intelligent surface for different illumination strategies.

Fig. 4 b-iv). For simplicity, in this article, we use blockage-free illumination only in combination with PI although in principle it can be also used in combination with FI and SI, e.g., see [41] for blockage-free FI.

Hypothetical Uniform SI: For practical illumination strategies, the distribution of the power received from each active antenna at the intelligent surface is non-uniform. For single-stream data transmission, non-uniform power distribution causes a decrease in the achievable antenna gain, which is known as taper loss [14, Ch. 15]. More generally, for multi-stream data transmission, a non-uniform power distribution reduces the achievable rate. In order to study the performance degradation due to non-uniform power distribution across the intelligent surface, we focus on SI and introduce a hypothetical illumination where the powers of the signals received from each active antenna at its respective passive elements are identical. To formally model uniform SI, let us rewrite matrix $\mathbf{T} = [c_{m,n}e^{-j2\pi r_{m,n}/\lambda}]_{m,n}$, where $c_{m,n} = \lambda\sqrt{\rho_{\text{srf}}G^a(\theta^p_{m,n},\phi^p_{m,n})G^p(\theta^a_{m,n},\phi^a_{m,n})}/(4\pi r_{m,n})$. For uniform SI, the phase $2\pi r_{m,n}/\lambda$ remains the same as for SI while $c_{m,n}$ is equal to a constant c, $\forall n, m \in \mathcal{M}_n$, and equal to zero otherwise. Here, $\mathcal{M}_n$ is the set of indices of the passive elements responsible for reflection/transmission of the signal emitted by active antenna n. To account for the over-the-air pathloss and surface efficiency, we set the value of c as $\lambda\sqrt{\rho_{\text{srf}}G^aG^p}/(4\pi r)$, where $r = \frac{1}{M}\sum_n\sum_{m\in\mathcal{M}_n} r_{m,n}$ and we assume $G^a(\theta^p_{m,n},\phi^p_{m,n}) \triangleq G^a$, $\forall\theta^p_{m,n} \in [0,\theta_0^{\text{SI}}]$, $\forall\phi^p_{m,n} \in [0,2\pi]$, and $G^p(\theta^a_{m,n},\phi^a_{m,n}) \triangleq G^p$, $\forall\theta^a_{m,n} \in [0,\frac{\pi}{2}]$, $\forall\phi^p_{m,n} \in [0,2\pi]$.[9] Here, G^a and G^p are constants and θ_0^{SI} is the elevation angular extent of the sub-surfaces in SI with respect to their feed antennas. The following corollary provides matrix $\mathbf{T}$ for uniform SI.

9. For the hypothetical uniform SI, for the active antennas, we assume a uniform antenna pattern with the elevation angle confined to $\theta \in [0,\theta_0^{\text{SI}}]$, whereas for the passive antennas, we assume a uniform antenna pattern with the elevation angle confined to $\theta \in [0,\frac{\pi}{2}]$. The reason for this choice is that the active antennas are usually horn antennas whose antenna gain is controllable, whereas the passive antennas are typically simple patch antennas whose antenna pattern is wide and not easy to control.

Corollary 1: For uniform SI, matrix $\mathbf{T}$ in (9) simplifies to

$$\mathbf{T} = \left[c\, e^{-j\frac{2\pi r_{m,n}}{\lambda}}\right]_{m,n} \tag{10}$$

with

$$c = \begin{cases} \frac{\lambda}{4\pi r}\sqrt{\frac{4\rho_{\text{srf}}}{1-\cos(\theta_0^{\text{SI}})}}, & \text{if } m \in \mathcal{M}_n \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Proof: The proof follows directly from noting that $G^a = \frac{2}{1-\cos(\theta_0^{\text{SI}})}$ and $G^p = 2$ has to hold such that $\int_\Omega \frac{1}{4\pi}G^q(\phi,\theta)\,\mathrm{d}\,\Omega = 1$, $q \in \{a,p\}$, holds where $\mathrm{d}\,\Omega = \sin(\theta)\,\mathrm{d}\,\theta\,\mathrm{d}\,\phi$ [14] and applying the simplifying assumptions of uniform SI in (9). ■

In Section V, we comprehensively study the aforementioned illumination strategies via simulations in order to obtain insights regarding their impact on performance for system design.

C. POWER CONSUMPTION AND LOSSES

The power consumption of IRS/ITS-aided antennas can be divided into two parts.

Baseline Circuitry: The circuit power consumption comprises the power consumed for digital baseband processing, denoted by P_{bb}, and by each RF chain (including the digital-to-analog converter, local oscillator, and mixer), denoted by P_{rfc}. Although, in principle, P_{bb} may depend on M and N, in the remainder of this article, we assume P_{bb} is constant since its impact is typically much smaller than that of P_{rfc} [43], [44].

Power Amplifier: The power consumed by the power amplifier (PA) is commonly modeled as $P_{\text{rd}}/\rho_{\text{pa}}$ where P_{rd} is the output power radiated by the active antennas and ρ_{pa} denotes the power amplifier efficiency [8], [44]–[46]. The power radiated by the active antennas is given by $P_{\text{rd}} = \mathbb{E}\{\bar{\mathbf{x}}^{\text{H}}\bar{\mathbf{x}}\} = P_{\text{tx}}\|\mathbf{B}\|_F^2$. Due to the losses incurred in the channel between the active antennas and the intelligent surface as well as the inefficiencies of the intelligent surface, $P_{\text{tx}} \leq P_{\text{rd}}$ holds. The main sources of the power loss for IRS/ITS-aided antennas are provided in the following:

- *Spillover Loss:* Since the effective area of the intelligent surface is finite, some of the power radiated by the active antennas will not be captured by the passive antennas, resulting in a spillover loss [42]. We define the efficiency factor ρ_S to take the spillover into account. We note that the value of ρ_S depends on the relative positions of the active antennas and the intelligent surface, i.e., $(r_{m,n}, \theta^p_{m,n}, \phi^p_{m,n})$, $\forall m, n$, and the radiation pattern of the active antennas, i.e., $G^a(\theta^p_{m,n}, \phi^p_{m,n})$. Hence, the impact of ρ_S is implicitly captured by matrix $\mathbf{T}$.
- *Taper Loss:* In general, the density of the received power differs across the intelligent surface as it depends on $G^a(\theta^p_{m,n}, \phi^p_{m,n})$ and $r_{m,n}$. As discussed earlier, for multi-stream transmission, taper loss leads to a reduction of the achievable rate. We define the efficiency factor ρ_T to account for this loss. Similar to ρ_S, the impact of ρ_T is implicitly captured by matrix $\mathbf{T}$.
- *Aperture Loss:* Ideally, for IRS/ITS-aided antennas, the total power captured by the aperture will be reflected/transmitted. In practice, however, the actual power transmitted into the wireless channel is smaller than the total power captured by the surface, which is due to inefficiencies in antenna reception/transmission, power absorption by the IRS/ITS, and unwanted reflection by the ITS. The aperture efficiency is taken into account by introducing efficiency factor ρ_A. For future reference, we decompose the aperture efficiency factor as $\rho_A = \rho_A^{\text{ant}} \rho_A^{\text{srf}}$, where ρ_A^{ant} is the antenna aperture efficiency of the individual passive antennas and ρ_A^{srf} denotes the surface efficiency factor which captures the remaining losses of the surface excluding the antenna aperture losses. Thereby, the impact of ρ_A^{ant} is implicitly captured via the passive antenna pattern $G^p(\theta^a_{m,n}, \phi^a_{m,n})$ by matrix $\mathbf{T}$.
- *Phase Shifters:* Each phase shifter introduces a certain loss once the RF signal passes through it. For the IRS-aided antennas, the signal passes twice through the phase-shifters whereas for the ITS-aided antennas, the signal passes only once through the phase-shifter, see the dashed arrows in Fig. 1e) and f). Hence, for the reflected wave in the IRS-aided antennas and the re-transmitted wave in the ITS-aided antennas to have the same overall phase shift, the phase shift induced by the phase shifters of the IRS-aided antennas should be half of that for the ITS-aided antennas. The impact of this difference between IRS- and ITS-aided antennas on the overall phase-shifter losses depends on the specific phase-shifter technology. Let $\rho_P(\beta)$ denote the phase-shifter power efficiency factor as a function of phase-shift value β. Then, the overall phase shifter efficiency factors of the IRS- and ITS-aided antennas are given by $\rho_P^2(\beta/2)$ and $\rho_P(\beta)$, respectively. In practice, for most phase-shifter realizations, the overall phase-shifter loss is almost independent of the phase-shift value and practically constant [47], [48]. For example, in switched-line phase shifters, the insertion loss is caused by switch losses (which are independent of the phase-shift values) and line losses (which increase with the phase-shift value) [49]. Nevertheless, the overall loss is dominated by the switch losses which implies that the phase-shifter loss is practically constant for different phase shifts. Therefore, in this article, we assume that the phase-shifter efficiency is constant for different phase-shift values, i.e., $\rho_P(\beta) = \rho_P$, $\forall \beta$.

Remark 1: In addition to *RF losses*, each phase-shifter *consumes* a certain amount of *power* in order to control its phase-shift states. The amount of power consumed depends on the adopted phase-shifter technology. For phase shifters realized by varactor diodes or micro-electro-mechanical system (MEMS) switches, the phase-shifter power consumption is almost negligible [47]. On the other hand, if the phase shifters are realized by positive-intrinsic-negative (PIN) diode switches, a constant direct current (DC) is needed to drive each switch which implies a constant power consumption per unit-cell element. In this case, the total power consumption of the surface for controlling the phase shifters increases linearly with the number of unit cells. Since we are interested in scalable MIMO transmitter architectures, we focus on the former realizations of the phase-shifters and neglect the corresponding power consumption.

Recall that the impact of the spillover, taper, and *antenna aperture* losses is included in matrix $\mathbf{T}$ in (9). The remaining losses are accounted for in the power efficiency of IRS and ITS defined as $\rho_{\text{srf}} = \rho_P^2 \rho_A^{\text{srf}}$ and $\rho_{\text{srf}} = \rho_P \rho_A^{\text{srf}}$, respectively, which accounts for the combined effects of the surface aperture and phase-shifter losses. In summary, the total power consumption of the IRS/ITS-aided MIMO architectures is obtained as

$$P_{\text{tot}} = P_{\text{bb}} + NP_{\text{rfc}} + \frac{P_{\text{rd}}}{\rho_{\text{pa}}}. \tag{12}$$

Due to the aforementioned power losses,[10] i.e., ρ_S, ρ_P, and ρ_A, the active antennas have to transmit with power P_{rd} to ensure the required power P_{tx} is radiated by the passive antennas where $P_{\text{rd}} \geq P_{\text{tx}}$ holds. The following lemma relates the power radiated by the active antennas, P_{rd}, to the power radiated by the intelligent surface P_{tx}, based on matrix $\mathbf{T}$ and power efficiency factors ρ_S, ρ_P, and ρ_A. For future reference, let $\varpi(\mathbf{T}) = \frac{\sigma_{\max}(\mathbf{T})}{\sigma_{\min}(\mathbf{T})} \geq 1$ denote the condition number of matrix $\mathbf{T}$, where $\sigma_{\max}(\mathbf{T})$ and $\sigma_{\min}(\mathbf{T})$ denote the maximum and minimum singular values of $\mathbf{T}$, respectively. Moreover, although the values of ρ_S and ρ_A may in general be different for each active antenna and depend on the position of the active antenna with respect to the intelligent surface, for simplicity of presentation, we assume identical ρ_S and ρ_A for all active antennas.

Lemma 1: The power radiated by the active antennas, P_{rd}, is bounded in terms of the transmit power radiated by the

10. Note that taper loss reduces the achievable rate but does not constitute a power loss.

intelligent surface, P_{tx}, as follows

$$\frac{P_{\text{tx}}}{\varpi^2(\mathbf{T})\rho_{\text{rts}}} \leq \frac{P_{\text{tx}}}{\sigma^2_{\max}(\mathbf{T})} \leq P_{\text{rd}} \leq \frac{P_{\text{tx}}}{\sigma^2_{\min}(\mathbf{T})} \leq \frac{\varpi^2(\mathbf{T})P_{\text{tx}}}{\rho_{\text{rts}}}, \quad (13)$$

where $\rho_{\text{rts}} = \rho_{\text{srf}}\rho_A^{\text{ant}}\rho_S$. Therefore, assuming an ideal well-conditioned matrix $\mathbf{T}$ with $\varpi(\mathbf{T}) = 1$, we obtain $P_{\text{rd}} = \frac{P_{\text{tx}}}{\rho_{\text{rts}}}$.

Proof: The proof is provided in Appendix A. ■

From (12) and Lemma 1, we can conclude that the total power consumption of IRS/ITS-aided antennas does not explicitly depend on the number of passive elements M (or equivalently the size of the IRS/ITS) which makes them energy efficient and scalable. Nevertheless, the condition number $\varpi(\mathbf{T})$ and the values of the spillover efficiency, ρ_S are determined by factors such as the size of the intelligent surface, the beam pattern of the active antennas, the distance between the active antennas and the intelligent surface, etc., which may in turn be *influenced* by M. Moreover, there is a trade-off between the spillover loss ρ_S and the taper loss ρ_T such that the former can be decreased at the expense of increasing the latter by employing a narrower beam for the active antennas [14, Ch. 15]. In the following subsection, we analytically show for a special case that ρ_S and ρ_T can be made independent of M by proper positioning of the antennas. Moreover, in Section V, we show via simulations that $\varpi(\mathbf{T})$ can be made small (i.e., close to one) when the antennas are properly positioned.

D. SPECIAL CASE

To illustrate the variation of the spillover and taper losses as a function of the feed pattern and the angular extent of the intelligent surface, we consider the following simple class of axisymmetric feed antenna patterns which have been widely-adopted by the antenna community[11] [14], [42]

$$G(\theta,\phi) = \begin{cases} 2(1+\kappa)\cos^{\kappa}(\theta), & \text{if } 0 \leq \theta \leq \frac{\pi}{2} \\ & \text{and } 0 \leq \phi \leq 2\pi \\ 0, & \text{otherwise,} \end{cases} \quad (14)$$

where $\kappa \geq 2$ is a real number and normalization factor $2(1+\kappa)$ ensures that $\int_\Omega \frac{1}{4\pi} G(\phi,\theta)\,\mathrm{d}\,\Omega = 1$ holds [14]. Therefore, the (maximum) antenna gain in dB is $10\log_{10}(2(1+\kappa))$. The value of κ (i.e., the gain of the active antenna) has to be jointly optimized with the relative position and orientation of the active antenna with respect to the intelligent surface such that the active antenna illuminates only the intended part of the intelligent surface, see illumination strategies in Section III-B. For the antenna pattern in (14) and assuming a *circular* planar surface where the feed antenna orthogonally illuminates the center of the surface, the spillover loss is obtained as [14, Ch. 15]

$$\rho_S = \frac{\int_\phi \int_{\theta=0}^{\theta_0} G(\phi,\theta)\,\mathrm{d}\,\Omega}{\int_\phi \int_{\theta=0}^{\pi} G(\phi,\theta)\,\mathrm{d}\,\Omega} = 1 - \cos^{\kappa+1}(\theta_0), \quad (15)$$

where θ_0 is the elevation angular extent of the intelligent surface with respect to the feed antenna. Similarly, the taper loss is obtained as [14, Ch. 15]

$$\rho_T = \frac{\left[\int_\phi \int_{\theta=0}^{\theta_0} \sqrt{G(\phi,\theta)}\,\mathrm{d}S\right]^2}{2\pi\left(\cos^{-1}(\theta_0)-1\right)\int_\phi \int_{\theta=0}^{\pi} G(\phi,\theta)\,\mathrm{d}S} = \frac{\kappa-1}{(\kappa/2-1)^2} \times \frac{\left[1-\cos^{\kappa/2-1}(\theta_0)\right]^2}{\left(1-\cos^{\kappa-1}(\theta_0)\right)\left(\cos^{-1}(\theta_0)-1\right)}, \quad (16)$$

where $\mathrm{d}S = \mathrm{d}S_\theta\,\mathrm{d}\,S_\phi$ is the normalized unit area covered by $[\theta, \theta+\mathrm{d}\,\theta]$ and $[\phi, \phi+\mathrm{d}\,\phi]$ on the intelligent surface, $\mathrm{d}S_\theta = \cos^{-2}(\theta)\,\mathrm{d}\,\theta$, and $\mathrm{d}S_\phi = \sin(\theta)\,\mathrm{d}\,\phi$. Moreover, the normalization factor $2\pi(\cos^{-1}(\theta_0)-1)$ ensures that $\rho_T = 1$ for uniform illumination. Note that choosing a larger κ decreases the spillover loss; however, it increases the taper loss.

The spillover and taper efficiencies in (15) and (16), respectively, were derived for a circular planar surface. For square planar surfaces, we can obtain approximate expressions for ρ_S and ρ_T from (15) and (16), respectively, by approximating the square surface with a circular surface having the same area. In particular, for a square surface with area $(\sqrt{M}d)^2$, the elevation angular extent θ_0 of the approximately equivalent circular surface is obtained as

$$\theta_0 \approx \tan^{-1}\left(\frac{d}{R_d}\sqrt{\frac{M}{\pi}}\right), \quad (17)$$

where R_d is the distance between the active antenna and the intelligent surface. Therefore, for square surfaces, if R_d is chosen to be proportional to $\sqrt{M}$, the value of θ_0 is independent of M. Hence, in this case, the spillover and taper losses do not scale with M, see (15) and (16).

IV. PRECODER DESIGN FOR IRS/ITS-AIDED MIMO ARCHITECTURES

In this section, we propose two linear precoders for IRS/ITS-aided antennas exploiting the sparsity of the mmWave channel. We assume that CSI $\mathbf{H}$ is available at the transmitter and is used for precoder design. Therefore, similar to the precoder designs in [5]–[7], [50] for conventional MIMO architectures, the frequency with which the proposed precoders (including the phase shifters at the intelligent surface) are updated should be chosen in accordance with the channel coherence time. This is in contrast to load modulated arrays [51], media-based modulation [52], and symbol-level precoding for IRS/ITS-aided MIMO [31], [32], where antenna loads or phase shifters change at the symbol rate. Exploiting the CSI, ideally, we would like to design the optimal precoder which maximizes the achievable rate or equivalently the MI between $\mathbf{s}$ and $\mathbf{y}$, i.e.,

11. We note that our modeling and derivations are valid for general feed antenna patterns and only for the analysis in Section III-D and the simulation results in Section V, we adopt the example antenna pattern in (14).

$R(\mathbf{F}) \triangleq I(\mathbf{s};\mathbf{y}) = \log_2|\mathbf{I}_J + \gamma\mathbf{H}\mathbf{F}\mathbf{F}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}|$, for $\mathbf{y}$ given in (2) and Gaussian $\mathbf{s}$, as follows

$$\begin{aligned} &\underset{\mathbf{F}\in\mathcal{F}}{\text{maximize}} \ \log_2\left|\mathbf{I}_J + \gamma\mathbf{H}\mathbf{F}\mathbf{F}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right| \\ &\text{subject to :} \ \|\mathbf{F}\|_F^2 \le 1, \end{aligned} \tag{18}$$

where $\gamma = \frac{P_{\text{tx}}}{\sigma^2}$, constraint $\|\mathbf{F}\|_F^2 \le 1$ enforces the transmit power constraint, $\mathcal{F} = \{\mathbf{F} = \mathbf{D}\mathbf{T}\mathbf{B} | \mathbf{B} \in \mathbb{C}^{N\times Q}$ and $\mathbf{D} = \text{diag}(d_1,\ldots,d_M),\ d_m \in \mathbb{A}\}$ is the set of feasible precoders, and $\mathbb{A} = \{x | x \in \mathbb{C}$ and $|x| = 1\}$ is the set of unit-modulus numbers. We note that the problem in (18) is different from those considered in [5], [9], [45], [53]–[55] for conventional MIMO architectures due to the different constraints imposed on the precoder via $\mathcal{F}$. Hence, the solutions proposed in the literature for conventional MIMO architectures are not directly applicable to (18).

Unfortunately, problem (18) is not tractable since set $\mathcal{F}$ is not convex due to the unit-modulus constraint on the elements of phase-shift matrix $\mathbf{D}$. A similar challenge exists for conventional hybrid precoders where the elements of the corresponding analog precoder have to be unit-modulus. Since the global optimal solution of such non-convex problems cannot be found in a computationally efficient manner, suboptimal solutions have been pursued in the literature. These solutions can be classified into two categories: *i)* solutions that find a local optimal/stationary point of the original problem [9], [53], [54], and *ii)* greedy solutions which exploit certain useful properties of the original problem (e.g., the sparsity of the mmWave channel) [5], [45], [55]. In this section, we propose new precoder designs for IRS/ITS-aided MIMO which in part belong to both aforementioned categories. In particular, in each iteration, for a given phase-shift matrix, the digital baseband precoder is found as the global optimum of a corresponding sub-problem; however, for the phase-shift matrix, we only allow transmission in the direction of the AoDs of the channel which is an intuitive but in general heuristic choice. We note that IRS-aided antennas have been considered previously in [37] and a local optimal solution based on approximating the optimal FD precoder, instead of maximizing the MI as in (18), was derived. We show in Section V via simulations that the proposed precoders outperform the benchmark scheme from [37].

A. RATIONALE BEHIND THE PROPOSED PRECODERS

For the spatially sparse channel model introduced in (3), $\mathcal{H}_t = \{\mathbf{h}_t(\theta_l^t,\phi_l^t), \forall l = 1,\ldots,L\}$ forms a vector space for the rows of $\mathbf{H}$. Moreover, in practice, (θ_l^t,ϕ_l^t) is an RV that takes its values from a continuous distribution. Therefore, since $L \ll M$, the elements of $\mathcal{H}_t$ are with probability one linearly independent [5]. Let $\mathcal{H}_t^{\perp}$ denote the null space of $\mathcal{H}_t$. Thereby, any precoder $\mathbf{F} = \mathbf{F}_{\mathcal{H}_t} + \mathbf{F}_{\mathcal{H}_t^{\perp}}$ can be decomposed into matrix $\mathbf{F}_{\mathcal{H}_t}$ belonging to space $\mathcal{H}_t$ and matrix $\mathbf{F}_{\mathcal{H}_t^{\perp}}$ belonging to space $\mathcal{H}_t^{\perp}$. The following lemma formally characterizes the impact of $\mathbf{F}_{\mathcal{H}_t}$ and $\mathbf{F}_{\mathcal{H}_t^{\perp}}$ on the cost function and the constraint in (18).

Lemma 2: For any given precoder $\mathbf{F}$, the relations $R(\mathbf{F}) = R(\mathbf{F}_{\mathcal{H}_t})$ and $\|\mathbf{F}\|_F \ge \|\mathbf{F}_{\mathcal{H}_t}\|_F$ hold.

Proof: The proof is given in Appendix B. ■

Motivated by the above results, we limit our attention to precoders of the form $\mathbf{F} = \mathbf{F}_{\mathcal{H}_t}$ which includes the optimal FD precoder [5], [7]. More explicitly, the precoder is rewritten as $\mathbf{F} = \mathbf{H}_t\mathbf{C}$, where $\mathbf{H}_t = [\mathbf{h}_t(\theta_1^t,\phi_1^t),\ldots,\mathbf{h}_t(\theta_L^t,\phi_L^t)] \in \mathbb{C}^{M\times L}$ and $\mathbf{C} \in \mathbb{C}^{L\times Q}$ contains the corresponding coefficients. For the FC hybrid MIMO architecture, the similarity of the structure of the optimal precoder $\mathbf{F} = \mathbf{H}_t\mathbf{C}$ and the corresponding hybrid precoder $\mathbf{F} = \mathbf{R}\mathbf{B}$, where $\mathbf{B}$ and $\mathbf{R}$ denote the digital and analog precoders, receptively, has motivated researchers to use the channel response vectors $\mathbf{h}_t(\theta_l^t,\phi_l^t)$ as the columns of $\mathbf{R}$ [3]–[5]. Since $\mathbf{R}$ has N columns (i.e., there are N RF chains), the hybrid precoder problem simplifies to choosing the best N columns of $\mathbf{H}_t$ and the corresponding coefficients. Unfortunately, this concept is not directly applicable to the IRS/ITS precoder in (7) because of its different structure. Hence, we rewrite $\mathbf{F} = \mathbf{H}_t\mathbf{C}$ in a more useful form. Let us divide the index set of the passive antennas $\{1,\ldots,M\}$ into N mutually exclusive sets $\mathcal{M}_n,\ n = 1,\ldots,N$. We note that depending on the adopted illumination scenario, the passive antenna elements corresponding to set $\mathcal{M}_n$ may also receive signals from active antennas $n' \ne n$. Now, $\mathbf{F} = \mathbf{H}_t\mathbf{C}$ can be rewritten as

$$\mathbf{F} = \sum_{n=1}^{N} \mathbf{H}_t^{\mathcal{M}_n}\mathbf{C}, \tag{19}$$

where $\mathbf{H}_t^{\mathcal{M}_n} = \mathbf{I}_{\mathcal{M}_n}\mathbf{H}_t \in \mathbb{C}^{M\times L}$ and $\mathbf{I}_{\mathcal{M}_n} \in \{0,1\}^{M\times M}$ is a diagonal matrix whose m-th diagonal entry is one if $m \in \mathcal{M}_n$ and zero otherwise. In other words, we decompose $\mathcal{H}_t$ into N subspaces, denoted by $\mathcal{H}_t^{\mathcal{M}_n},\ n = 1,\ldots,N$, which have mutually exclusive non-zero supports and are fully characterized by $\mathbf{H}_t^{\mathcal{M}_n},\ n = 1,\ldots,N$, respectively. In a similar manner, let us rewrite the precoder in (7) as

$$\mathbf{F} = \sum_{n=1}^{N} \mathbf{D}^{\mathcal{M}_n}\mathbf{T}^{\mathcal{M}_n}\mathbf{B}, \tag{20}$$

where $\mathbf{D}^{\mathcal{M}_n} = \mathbf{I}_{\mathcal{M}_n}\mathbf{D} \in \mathbb{A}^{M\times N}$ and $\mathbf{T}^{\mathcal{M}_n} = \mathbf{I}_{\mathcal{M}_n}\mathbf{T} \in \mathbb{C}^{M\times N}$. Comparing (19) and (20) motivates us to choose $\mathbf{D}^{\mathcal{M}_n}$ such that $\mathbf{D}^{\mathcal{M}_n}\mathbf{T}^{\mathcal{M}_n}$ becomes similar to $\mathbf{H}_t^{\mathcal{M}_n}$. To do this, we have to address the following two challenges. First, since $\mathbf{D}^{\mathcal{M}_n}$ has only M/N non-zero elements and $\mathbf{H}_t^{\mathcal{M}_n}$ has ML/N non-zero elements, $\mathbf{H}_t^{\mathcal{M}_n}$ cannot be fully reconstructed via $\mathbf{D}^{\mathcal{M}_n}\mathbf{T}^{\mathcal{M}_n}$ as matrix $\mathbf{T}^{\mathcal{M}_n}$ is fixed. Hereby, we choose to reconstruct only one column of $\mathbf{H}_t^{\mathcal{M}_n}$ via $\mathbf{D}^{\mathcal{M}_n}\mathbf{T}^{\mathcal{M}_n}$. The unmatched columns of $\mathbf{D}^{\mathcal{M}_n}\mathbf{T}^{\mathcal{M}_n}$ are treated as interference. Fortunately, for $M \gg N$, the interference approaches zero due to channel hardening. Second, we have to choose which column of $\mathbf{H}_t^{\mathcal{M}_n}$ to reconstruct. In the following, we introduce two approaches, namely MI-based and OMP-based strategies, to choose the best N columns of $\mathbf{H}_t^{\mathcal{M}_n}$. Note that the above precoder design effectively reduces the search

space for the diagonal phase-shift matrix $\mathbf{D}$ from multi-dimensional continuous set $\mathbb{A}^{M\times 1}$ to the finite elements of sets $\mathcal{H}_t^{\mathcal{M}_n}$, $n = 1, \ldots, N$, i.e., NL elements in total. Therefore, as typically $N, L \ll M$, one can adopt an exhaustive search over this reduced space to obtain the optimal (e.g., rate-maximizing) phase-shift matrix $\mathbf{D}$ for a given baseband precoder $\mathbf{B}$.

B. MI-BASED PRECODER

For the proposed MI-based precoder, we design the phase-shift matrices $\mathbf{D}^{\mathcal{M}_n}$, $n = 1, \ldots, N$, and the corresponding baseband precoder in an iterative manner, such that the MI expression in (18) is maximized [50]. In particular, the proposed precoder design consists of an inner loop and an outer loop. The outer loop involves N iterations where in the n-th iteration, we choose the best channel path for the design of phase-shift matrix $\mathbf{D}^{\mathcal{M}_n}$, and in the inner loop, we determine the corresponding baseband precoder, denoted by $\mathbf{B}_n$. In particular, the inner loop involves L iterations where in the l-th iteration, we maximize the achievable rate by optimizing the baseband precoder $\mathbf{B}$ assuming $\mathbf{D}^{\mathcal{M}_n}$ is one of the elements of $\mathcal{H}_t^{\mathcal{M}_n}$. Therefore, we have to consider the following two problems.

Optimizing the Baseband Precoder: Here, we assume the phase-shift matrix $\mathbf{D} = \sum_{n=1}^{N} \mathbf{D}^{\mathcal{M}_n}$ is given. Since the logarithm is a monotonically increasing function, the optimization problem for finding the digital baseband precoder $\mathbf{B}$, which maximizes the MI $I(\mathbf{s}; \mathbf{y})$, simplifies to

$$\begin{aligned} &\underset{\mathbf{B}\in\mathbb{C}^{N\times Q}}{\text{maximize}} \quad \left|\mathbf{I}_J + \gamma \mathbf{H}\mathbf{C}_1\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{C}_1^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right| \\ &\text{subject to :} \quad \text{trace}\left(\mathbf{C}_1\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{C}_1^{\mathsf{H}}\right) \le 1, \end{aligned} \tag{21}$$

where $\mathbf{C}_1 = \mathbf{D}\mathbf{T} \in \mathbb{C}^{M\times N}$. Let us define matrix $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{C}_1(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1)^{-\frac{1}{2}} \in \mathbb{C}^{J\times N}$ and its corresponding singular value decomposition (SVD) $\widetilde{\mathbf{H}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}}$, where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_J] \in \mathbb{C}^{J\times J}$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N] \in \mathbb{C}^{N\times N}$ are unitary matrices containing the left and right singular vectors, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values $\sigma_1, \ldots, \sigma_N$ in descending order. The solution of (21) is given in the following lemma.

Lemma 3: For a given phase-shift matrix $\mathbf{D}$, the optimal baseband precoder $\mathbf{B}$ as a solution of (21) is given by

$$\mathbf{B} = \left(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1\right)^{-\frac{1}{2}}[\mathbf{v}_1, \ldots, \mathbf{v}_Q]\mathbf{Z}, \tag{22}$$

where $\mathbf{Z} = \text{diag}(\sqrt{z_1}, \ldots, \sqrt{z_Q}) \in \mathbb{C}^{Q\times Q}$, $z_q = [\mu - \frac{1}{\gamma\sigma_q^2}]^+$, and threshold μ is chosen such that constraint $\sum_{q=1}^{Q} z_q = 1$ is met.

Proof: The proof is given in Appendix B. ∎

Note that the baseband precoder $\mathbf{B}$ effectively eliminates the interference between the data streams.

Optimizing the Phase-Shift Matrix: As discussed earlier, we decompose $\mathbf{D}$ into N components $\mathbf{D}^{\mathcal{M}_n}$, $n = 1, \ldots, N$, which are initialized to the identity matrix $\mathbf{I}_M$ and their values are updated in each iteration. In particular, in the n-th iteration, the following problem is solved

$$\underset{\mathbf{D}^{\mathcal{M}_n}\in\mathcal{D}_n}{\text{maximize}} \quad \left|\mathbf{I}_J + \gamma \mathbf{H}\mathbf{D}\mathbf{C}_2\mathbf{D}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right|, \tag{23}$$

where $\mathbf{C}_2 = \mathbf{T}\mathbf{B}(\mathbf{D})\mathbf{B}(\mathbf{D})^{\mathsf{H}}\mathbf{T}^{\mathsf{H}} \in \mathbb{C}^{M\times M}$. Here, $\mathbf{B}(\mathbf{D})$ denotes the optimal baseband precoder as a function of a given phase-shift matrix $\mathbf{D}$ which is obtained from (22). Moreover, set $\mathcal{D}_n$ is given by

$$\begin{aligned} \mathcal{D}_n &= \left\{\mathbf{D}^{\mathcal{M}_n}[l] \in \mathbb{A}^{M\times M}, \ \forall l = 1, \ldots, L\right| \\ \mathbf{D}^{\mathcal{M}_n}_{m,m'}[l] &= \begin{cases} \exp\left(j\left[\angle\left(\mathbf{H}_t^{\mathcal{M}_n}\right)_{m,l} - \angle\mathbf{T}_{m,n}\right]\right), \\ \qquad\qquad \forall m = m' \in \mathcal{M}_n \\ 0, \qquad\qquad \text{otherwise} \end{cases}\Bigg\}. \end{aligned} \tag{24}$$

As can be seen, the cardinality of $\mathcal{D}_n$ is L which allows us to solve (23) via an exhaustive search. Algorithm 1 summarizes the main steps of the proposed MI-based precoder design.

Algorithm 1 MI-Based Precoder Design

initialize: $\mathbf{D}^{\mathcal{M}_n} = \mathbf{I}_M$, $\forall n$.
for $n = 1, \ldots, N$ **do**
  for $l = 1, \ldots, L$ **do**
    Set $\mathbf{D}^{\mathcal{M}_n} = \mathbf{D}^{\mathcal{M}_n}[l]$ from $\mathcal{D}_n$ in (24).
    Find $\mathbf{B}[l]$ using (22) for $\mathbf{D}[l] = \sum_{i=1}^{n} \mathbf{D}^{\mathcal{M}_i}$.
    Set $R[l] = \log_2\left(\left|\mathbf{I}_J + \gamma\mathbf{H}\mathbf{F}\mathbf{F}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right|\right)$ for $\mathbf{F} = \mathbf{D}[l]\mathbf{T}\mathbf{B}[l]$.
  end for
  Update $\mathbf{D}^{\mathcal{M}_n} = \mathbf{D}^{\mathcal{M}_n}[l^*]$ and $\mathbf{B}_n = \mathbf{B}[l^*]$ for $l^* = \underset{l}{\text{argmax}}\ R[l]$.
end for
Return $\mathbf{D} = \sum_{n=1}^{N} \mathbf{D}^{\mathcal{M}_n}$ and $\mathbf{B} = \mathbf{B}_N$.

C. OMP-BASED PRECODER

In this subsection, we propose a second precoder, namely the OMP-based precoder, which is computationally less complex than the MI-based precoder (see Section IV-D) but achieves a similar performance in poor scattering environments (see Section V). In particular, the OMP-based precoder attempts to approximate the optimal unconstrained precoder for the FD MIMO architecture, denoted by $\mathbf{F}^{\text{opt}}$, using the OMP algorithm. Minimization of $\|\mathbf{F}^{\text{opt}} - \mathbf{F}\|_F$ has been commonly adopted as design criterion for constrained hybrid precoders for conventional MIMO architectures, see e.g., [5], [37], [44]. Motivated by this, we consider the following optimization problem for the precoder of IRS/ITS-aided MIMO systems

$$\begin{aligned} &\underset{\mathbf{B}\in\mathbb{C}^{N\times Q}, \mathbf{D}^{\mathcal{M}_n}\in\mathcal{D}_n}{\text{minimize}} \quad \left\|\mathbf{F}^{\text{opt}} - \mathbf{D}\mathbf{T}\mathbf{B}\right\|_F^2 \\ &\text{subject to :} \quad \|\mathbf{D}\mathbf{T}\mathbf{B}\|_F^2 \le 1. \end{aligned} \tag{25}$$

Again, let us fix sets $\mathcal{M}_n$, $n = 1, \ldots, N$, a priori. The proposed precoder employs N iterations where in each iteration, the following two problems are solved.

Algorithm 2 OMP-Based Precoder Design

initialize: $\mathbf{F}_0^{\text{res}} = \mathbf{F}^{\text{opt}}$ and $\mathbf{D}^{\mathcal{M}_n} = \mathbf{0}_{M\times M}$, $\forall n$.
for $n = 1, \ldots, N$ **do**
  $l_n^* = \text{argmax}_{l=1,\ldots,L} \, (\mathbf{\Psi}\mathbf{\Psi}^{\mathsf{H}})_{l,l}$ for $\mathbf{\Psi} = \mathbf{H}_t^{\mathsf{H}}\mathbf{F}_{n-1}^{\text{res}}$.
  Update $\mathbf{D}^{\mathcal{M}_n}$ as the element of set $\mathcal{D}_n$ in (24) corresponding to l_n^*.
  Update $\mathbf{B}_n$ using (28) for $\mathbf{C}_3 = \sum_{i=1}^{n} \mathbf{D}^{\mathcal{M}_i}\mathbf{T}^{\mathcal{M}_i}$.
  Update $\mathbf{F}_n^{\text{res}} = \mathbf{F}^{\text{opt}} - \sum_{i=1}^{n} \mathbf{D}^{\mathcal{M}_i}\mathbf{T}^{\mathcal{M}_i}\mathbf{B}_n$.
end for
Return $\mathbf{D} = \sum_{n=1}^{N} \mathbf{D}^{\mathcal{M}_n}$ and $\mathbf{B} = \mathbf{B}_N$.

Optimizing the Phase-Shift Matrix: Let $\mathbf{F}_n^{\text{res}} = \mathbf{F}^{\text{opt}} - \sum_{i=1}^{n} \mathbf{D}^{\mathcal{M}_i}\mathbf{T}^{\mathcal{M}_i}\mathbf{B}_n$ denote the residual precoder in iteration n where $\mathbf{B}_n$ is the baseband precoder designed in iteration n. In each iteration, we project the residual matrix from the previous iteration onto the space defined by $\mathbf{H}_t$ and find the direction l^* that has the maximum projected value. This can be mathematically formulated as follows

$$l_n^* = \underset{l=1,\ldots,L}{\text{argmax}} \left(\mathbf{\Psi}\mathbf{\Psi}^{\mathsf{H}}\right)_{l,l}, \tag{26}$$

where $\mathbf{\Psi} = \mathbf{H}_t^{\mathsf{H}}\mathbf{F}_{n-1}^{\text{res}} \in \mathbb{C}^{L\times Q}$. Therefore, $\mathbf{D}^{\mathcal{M}_n}$ is selected as the element of $\mathcal{D}_n$ corresponding to the l^*-th channel path, see (3).

Optimizing the Baseband Precoder: By defining $\mathbf{C}_3 = \sum_{i=1}^{n} \mathbf{D}^{\mathcal{M}_i}\mathbf{T}^{\mathcal{M}_i} \in \mathbb{C}^{M\times N}$, we can formulate the optimization problem for $\mathbf{B}_n$ as follows

$$\begin{aligned} &\underset{\mathbf{B}\in\mathbb{C}^{N\times Q}}{\text{minimize}} \quad \left\|\mathbf{F}^{\text{opt}} - \mathbf{C}_3\mathbf{B}\right\|_F^2, \\ &\text{subject to :} \quad \|\mathbf{C}_3\mathbf{B}\|_F^2 \le 1, \end{aligned} \tag{27}$$

which has the following well-known normalized least square solution [5]

$$\mathbf{B}_n = \frac{\left(\mathbf{C}_3^{\mathsf{H}}\mathbf{C}_3\right)^{-1}\mathbf{C}_3^{\mathsf{H}}\mathbf{F}^{\text{opt}}}{\left\|\mathbf{C}_3\left(\mathbf{C}_3^{\mathsf{H}}\mathbf{C}_3\right)^{-1}\mathbf{C}_3^{\mathsf{H}}\mathbf{F}^{\text{opt}}\right\|_F}. \tag{28}$$

Algorithm 2 summarizes the main steps for the proposed OMP-based precoder design.

D. COMPLEXITY ANALYSIS

Let us assume that $M \gg J \ge N \ge Q$ and $M \gg L \ge N \ge Q$ hold. Moreover, we use the following results: The SVD of matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ of rank p has complexity order $O(mnp)$, the inversion of matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$ has complexity order $O(m^3)$, and the multiplication $\mathbf{AB}$ of matrices $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{B} \in \mathbb{C}^{n\times p}$ has complexity order $O(mnp)$ [56], [57]. The MI-based precoder involves NL iterations (i.e., inner and outer loops) where each iteration comprises the SVD of matrix $\widetilde{\mathbf{H}}$ (i.e., $O(JN^2)$), the inversion of matrix $\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1$ (i.e., $O(N^3)$), and matrix multiplications (i.e., $O(MNJ)$). Hence, recalling $M \gg J, N, Q$, the overall complexity order of the MI-based precoder is $O(MN^2JL)$. On the other hand, the OMP-based precoder requires the SVD of matrix $\mathbf{H}$ (i.e., $O(MJL)$) and involves N iterations where each iteration comprises the inversion of matrix $\mathbf{C}_3^{\mathsf{H}}\mathbf{C}_3$ (i.e., $O(N^3)$) and matrix multiplications (i.e., $O(MLQ + MN^2)$). Assuming $N = Q$, the overall complexity order of the OMP-based precoder simplifies to $O(M(N^2 + J)L)$. In summary, the complexity of both proposed algorithms is linear in M, which is a crucial advantage for ultra massive MIMO transmitters employing several hundreds (or even thousands) of transmit antennas. Moreover, computing the OMP-based precoder entails a lower complexity than computing the MI-based precoder as $N^2 + J < N^2J$ holds for typical values of N and J, i.e., $N, J > 1$.

V. SIMULATION RESULTS

In this section, we first describe the considered simulation setup and benchmark schemes. Subsequently, we study the performance of the proposed precoders and the impact of the system parameters. Then, we compare the performance of IRS/ITS-aided MIMO with that of the conventional MIMO architectures. Finally, we study the impact of imperfect CSI on the performance of the proposed precoder design.

A. SIMULATION SETUP

We generate the channel matrices according to (3). Thereby, we assume that the AoAs/AoDs θ_l^t, θ_l^r, ϕ_l^t, and ϕ_l^r are uniformly distributed RVs in the intervals given in Table 1. Moreover, we use the square uniform planar array in (4), i.e., a $\sqrt{M}d \times \sqrt{M}d$ planar surface. The channel coefficient for each effective path is modeled as $h_l = \sqrt{\bar{h}_l}\tilde{h}_l$, where $\bar{h}_l$ and $\tilde{h}_l$ are the path loss and the random fading components, respectively, and are given by

$$\bar{h}_l = \left(\frac{\lambda}{4\pi\ell}\right)^{\eta} \quad \text{and} \quad \tilde{h}_l = \mathcal{CN}(0,1), \tag{29}$$

respectively. In (29), ℓ denotes the distance between the transmitter and the receiver and η represents the path-loss exponent. The noise power at the receiver is given by $\sigma^2 = WN_0N_F$, where W is the bandwidth, N_0 represents the noise power spectral density, and N_F denotes the noise figure. All results shown in this section have been averaged over 10^3 random realizations of channel matrix $\mathbf{H}$ from (3) which includes the random realizations of the AoAs (θ_l^r, ϕ_l^r), the AoDs (θ_l^t, ϕ_l^t), and the corresponding path coefficients h_l.

We arrange the active antennas with respect to the intelligent surface as shown in Fig. 4 and described in Section III-B. We neglect the impact of the blockage of the active antennas in IRS-aided antennas; nevertheless, we study the performance of IRS-aided antennas under blockage-free illumination which is designed to avoid the blockage of desired AoAs/AoDs [41], [42]. Moreover, we adopt the feed antenna pattern in (14), which is widely used in the antenna community [14], [42]. For the passive antenna elements, we assume a uniform antenna pattern in the *half-space* where the intelligent surface faces the active antennas, which implies a constant antenna gain of 3 dB. Unless otherwise stated, we adopt the default values of the system parameters provided

TABLE 1. Default values of system parameters [8], [28], [36], [40], [44], [45], [58].

Parameter	ℓ	η	L	θ_l^t, θ_l^r	ϕ_l^t, ϕ_l^r	N_0	N_F	W	λ	d	R_d	R_r	κ
Value	100 m	2	8	$[-2\pi/3, 2\pi/3]$	$[-\pi/2, \pi/2]$	-174 dBm/Hz	6 dB	100 MHz	10 mm (28GHz)	$\lambda/2$	S1:$\frac{4d\sqrt{M}}{\sqrt{\pi}}$, S2:$\frac{4d\sqrt{M}}{\sqrt{N\pi}}$	S1:$2d$, S2:$\frac{d\sqrt{2M}}{4}$	49 (20 dB gain)

Parameter	P_{bb}	P_{rfc}	P_{sw}	P_{tx}	P_{amp}	G_{amp}	L_D	L_C	L_P ($1/\rho_P$ in dB)	$1/\rho_A^{\text{srf}}$ in dB	ρ_{amp}	Q	N	M	J
Value	200 mW	100 mW	5 mW	20 dBm	40 mW	10 dB	3.6 dB	3.6 dB	2 dB	IRS: 0.5 dB, ITS: 1.5 dB	0.3	4	4	256	16

in Table 1, which include the values of the parameters of the conventional MIMO architectures too, see Section V-B. Recall that the analysis in Section III-D revealed that R_d should scale with the square of the area of the passive surface that is intended to be illuminated. Therefore, for FI and blockage-free PI, the value of R_d has to scale with $\sqrt{M}$ such that the elevation angular extent θ_0 of the *entire array* with respect to any of the active antennas remains approximately constant, see (17). Furthermore, we assume a fixed value for R_r which is larger than $\lambda/2$ to ensure negligible mutual coupling among the active antennas, see Assumption A3. In contrast, for PI and SI, the value of R_d has to scale with $\sqrt{M/N}$ such that the elevation angular extent of each *sub-array* with respect to its dedicated active antenna remains approximately constant. Moreover, we scale R_r with M such that each active antenna remains close to the center of the area where the passive elements it serves are located. Therefore, in Table 1, we provide the values of R_r and R_d for two scenarios, namely scenario S1 for FI and blockage-free PI, and scenario S2 for PI and SI.

B. BENCHMARK SCHEMES

We consider the FD, FC hybrid, PC hybrid, and LA antennas as benchmark architectures. The precoder structures and power consumption models for these architectures are summarized in Table 2. In the following, we briefly explain the assumptions made to arrive at these expressions for conventional MIMO systems.

Fully-Digital MIMO: Here, we have $N = M$ RF chains which enable FD precoding, i.e., $\mathbf{F} = \mathbf{B}$, where $\mathbf{B}$ is the digital precoder. For FD MIMO, we adopt the optimal unconstrained precoder obtained from the SVD of the channel and water filling power allocation [7].

Fully-Connected Hybrid MIMO: In the FC hybrid architecture, we have N RF chains whose outputs are connected to M antennas via passive analog dividers, phase shifters, and combiners [5]. For this MIMO architecture, the precoder is given by $\mathbf{F} = \mathbf{RB}$, where $\mathbf{B} \in \mathbb{C}^{N\times Q}$ denotes the digital precoder and $\mathbf{R} \in \mathbb{A}^{M\times N}$ represents the analog RF precoder. We adopt the spatially-sparse precoder introduced in [5]. For large RF networks, the insertion loss may easily exceed 20-30 dB which makes a one-shot power compensation infeasible due to amplifier nonlinearities at high gains [8]. In practice, to compensate for this insertion loss, multiple gain-compensation amplifiers (GCAs) are cascaded to ensure that a minimum power is delivered to drive the PAs before transmission via the antennas [8], [46]. Assuming that the signal is amplified by GCAs before being fed to the PAs to compensate for the RF losses,[12] the power consumption of the RF network is given by $\lceil \frac{L_{\text{rf}}}{G_{\text{amp}}} \rceil P_{\text{amp}}$, where L_{rf} is the total loss in dB occurring in the RF network, G_{amp} denotes the maximum amplification gain of the GCAs in dB, and P_{amp} is their respective power consumption. Assuming that the power dividers (combiners) are implemented by a cascade of two-port power dividers (combiners), we need at least $\lceil \log_2(M) \rceil$ ($\lceil \log_2(N) \rceil$) stages of division (combining) [46], [59]. Therefore, the total power loss for the signal flowing towards each antenna is obtained as $L_{\text{rf}} = \lceil \log_2(M) \rceil L_D + \lceil \log_2(N) \rceil L_C + L_P$, where L_D (L_C) is the power loss of each three-port divider (combiner) in dB and $L_P = 10\log_{10}(1/\rho_P)$ [46], [59].

Partially-Connected MIMO: The signal model for the PC architecture is identical to that of the FC architecture, i.e., $\mathbf{F} = \mathbf{RB}$, with the difference that $\mathbf{R}$ is now a block-diagonal matrix $\mathbf{R} = \text{diag}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$, $\mathbf{r}_n \in \mathbb{A}^{r_n\times 1}$, where $\mathbf{r}_n \in \mathbb{A}^{r_n\times 1}$ is the RF precoder vector which connects the output of the n-th RF chain to r_n antennas [9], [54]. Note that $\sum_{n=1}^{N} r_n = M$ has to hold. We assume that all RF chains are connected to the same number of antennas, i.e., $r_n = M/N$, $\forall n$, where we assume that N is a divisor of M. Equivalently, the precoder for the PC architecture can be rewritten as $\mathbf{F} = \mathbf{D}\tilde{\mathbf{T}}\mathbf{B}$ where $\tilde{\mathbf{T}}$ is a fixed matrix whose element in the m-th row and n-th column is one if the m-th antenna is connected to the n-th RF chain and zero otherwise and diagonal matrix $\mathbf{D}$ is the corresponding analog precoder. Considering the structure of $\mathbf{F}$, the proposed Algorithms 1 and 2 can be exploited for precoder design for the PC hybrid MIMO architecture. In Fig. 8, where we compare PC MIMO and IRS/ITS-aided MIMO, we employ Algorithm 2 for both architectures to ensure a fair comparison. Noting that the PC architecture does not include a power combiner and assuming that the power dividers are implemented by a cascade of two-port power dividers, we obtain the total power consumption given in Table 2.

Lens Array MIMO: EM lenses are phase-shifting devices which can be designed employing either an array of passive antenna elements (similar to the considered ITS) [13] or continuous aperture phase shifting [4], [12]. We consider the

12. In practice, multiple stages of power amplification are needed within the RF network to ensure that the signal power does not get too weak, see [8] for examples of multiple-stage power amplification. Our motivation for considering single-stage power amplification in this article is two-fold. First, the exact design of multiple-stage amplification crucially depends on the specific system parameters, e.g., M, N, G_{amp}, and L_{rf}, and cannot be easily generalized. Second, since the number of required GCAs is larger for multiple-stage amplification, single-stage amplification constitutes a favorable choice for hybrid MIMO architectures, which we consider as performance benchmarks.

TABLE 2. Comparison of different MIMO architectures, namely FD, FC, PC, LA, IRS-aided, and ITS-aided antennas.

Architecture	Precoder $\mathbf{F}$	Constraints	Total Power Consumption P_{tot}
FD	B	$\mathbf{B} \in \mathbb{C}^{M\times Q}$	$P_{\text{bb}} + MP_{\text{rfc}} + \frac{P_{\text{tx}}}{\rho_{\text{pa}}}$
FC	RB	$\mathbf{B} \in \mathbb{C}^{N\times Q}$, $\mathbf{R} \in \mathbb{A}^{M\times N}$	$P_{\text{bb}} + NP_{\text{rfc}} + \left\lceil \frac{\lceil \log_2(M)\rceil L_D + \lceil \log_2(N)\rceil L_C + L_P}{G_{\text{amp}}} \right\rceil MP_{\text{amp}} + \frac{P_{\text{tx}}}{\rho_{\text{pa}}}$
PC	RB	$\mathbf{B} \in \mathbb{C}^{N\times Q}$, $\mathbf{R} = \text{diag}(\mathbf{r}_1, \ldots, \mathbf{r}_N), \mathbf{r}_n \in \mathbb{A}^{r_n\times 1}, \sum_{n=1}^{N} r_n = M$	$P_{\text{bb}} + NP_{\text{rfc}} + \left\lceil \frac{\lceil \log_2(M/N)\rceil L_D + L_P}{G_{\text{amp}}} \right\rceil MP_{\text{amp}} + \frac{P_{\text{tx}}}{\rho_{\text{pa}}}$
LA	DTSB	$\mathbf{B} \in \mathbb{C}^{N\times Q}$, $\mathbf{S} \in \{0,1\}^{K\times N}$, fixed matrix $\mathbf{T} \in \mathbb{C}^{M\times N}$, fixed $\mathbf{D} = \text{diag}(d_1, \ldots, d_M), d_m \in \mathbb{A}$	$P_{\text{bb}} + NP_{\text{rfc}} + NP_{\text{sw}} + \frac{P_{\text{tx}}}{\rho_{\text{rts}}\rho_{\text{pa}}}$
IRS & ITS	DTB	$\mathbf{B} \in \mathbb{C}^{N\times Q}$, fixed matrix $\mathbf{T} \in \mathbb{C}^{M\times N}$ (cf. (9)), $\mathbf{D} = \text{diag}(d_1, \ldots, d_M), d_m \in \mathbb{A}$	$P_{\text{bb}} + NP_{\text{rfc}} + \frac{P_{\text{tx}}}{\rho_{\text{rts}}\rho_{\text{pa}}}$

former option since it allows us to employ similar surfaces for LA and ITS-aided antennas. Therefore, the precoder for LA antennas is given by $\mathbf{F} = \mathbf{DTSB}$, where $\mathbf{B} \in \mathbb{C}^{N\times Q}$ is the digital precoder matrix, $\mathbf{S} \in \{0,1\}^{K\times N}$ is a binary switching matrix which specifies which RF chain is connected to which active antenna, $\mathbf{T} \in \mathbb{C}^{M\times K}$ is a fixed matrix which models the channel between the active and the passive surface (similar as for IRS/ITS-aided antennas in (9)), and $\mathbf{D} \in \mathbb{A}^{M\times M}$ is a fixed diagonal matrix with unit-modulus elements designed with the objective to focus the wavefront perpendicular to the lens plane at the focal point of the lens (i.e., a non-reconfigurable/non-intelligent surface). For LA antennas, we use a modification of the proposed OMP-based precoder obtained with Algorithm 2. Since for LA antennas $\mathbf{D}$ is fixed and different active antennas are selected via matrix $\mathbf{S}$ for transmission in different directions, the main change required when adapting Algorithm 2 to LA antennas is that line 4 is replaced with the selection of the active antenna which is used to transmit the signal of RF chain n along the AoD $(\theta^t_{l^*_n}, \phi^t_{l^*_n})$ chosen in line 3. Here, we adopt the following antenna selection strategy:

$$[\mathbf{S}]_{k,n} = \begin{cases} 1, & \text{if } k = \underset{\tilde{k}=1,\ldots,K}{\text{argmax}} \left| (\mathbf{h}_t^{\mathsf{H}}(\theta^t_{l^*_n}, \phi^t_{l^*_n})\mathbf{DT})_{\tilde{k}} \right| \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

Moreover, we assume that in the LA architecture, the active antennas are placed on the focal arc of the lens as given in [11, Eqs. (1) and (2)] with focal distance $4\lambda\sqrt{M/\pi}$, and illuminate the center of the passive EM lens. For the design in [11], the number of active antennas is a linear function of the effective lens aperture or equivalently M, which we refer to as LA antennas with full K. In addition, we consider the case where K is fixed independent of M and refer to it as LA antennas with fixed K. The power consumption model for LA antennas is similar to that for ITS-aided antennas except for the additional power consumed by the N active switches, i.e., NP_{sw}, where P_{sw} denotes the power consumption of each switch [36].

C. PERFORMANCE OF PROPOSED PRECODERS AND IMPACT OF THE SYSTEM PARAMETERS

In Fig. 5, we show the spectral efficiency R (bits/s/Hz) from (18) versus the number of transmit antennas M for SI, $Q = N = 4$, and $J \in \{16, M/4\}$. Note that both IRS/ITS-aided antennas yield the same spectral efficiency since we neglect the impact of the blockage of the active antennas in IRS-aided antennas and the difference in the array efficiency factor ρ_A for these architectures only influences their power consumption but does not impact their spectral efficiency. We observe that, as the number of antennas M increases, the spectral efficiency increases. However, the slope of the increase is larger for $J = M/4$ than for $J = 16$. As can be seen from Fig. 5, the MI-based procoder outperforms the OMP-based precoder in terms of spectral efficiency. This is expected since the MI-based procoder is optimized for maximization of the achievable rate whereas the OMP-based precoder is obtained by approximating the optimal unconstrained FD precoder. Nevertheless, the additional gain of the MI-based procoder is small and decreases as M increases. This can be attributed to the fact that both the MI- and OMP-based precoders search over the same sets for their respective phase-shift matrices whose cardinality is rather small, i.e., NL. In fact, we show later in Fig. 10 that for rich scattering environments (i.e., large L), the proposed MI-based precoder achieves larger performance gains over the OMP-based precoder. For the remainder of this section (except for Fig. 10), we consider poor scattering mmWave channels (specifically $L = 8$), and hence focus on the OMP-based precoder.

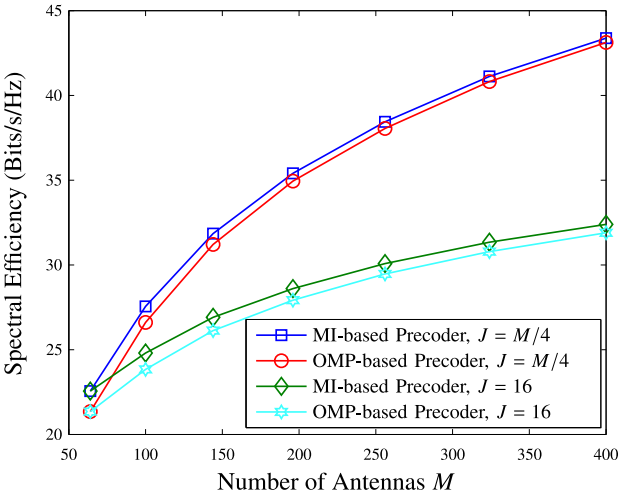


FIGURE 5. Spectral efficiency (bits/s/Hz) versus number of transmit antennas M for SI, $Q = N = 4$, and $J \in \{16, M/4\}$.

In Fig. 6, we show the spectral efficiency (bits/s/Hz) versus the number of transmit antennas M for $Q = N = 4$ and $J = 16$, and study the impact of the different illumination strategies introduced in Section III-B, namely FI, PI, SI, blockage-free PI, and uniform SI, see Fig. 4. As can be observed from this figure, PI achieves a better performance than FI. This can be explained as follows. For the proposed precoder, the passive elements are partitioned into N mutually exclusive subsets, i.e., $\mathcal{M}_n$, $n = 1, \ldots, N$, where each

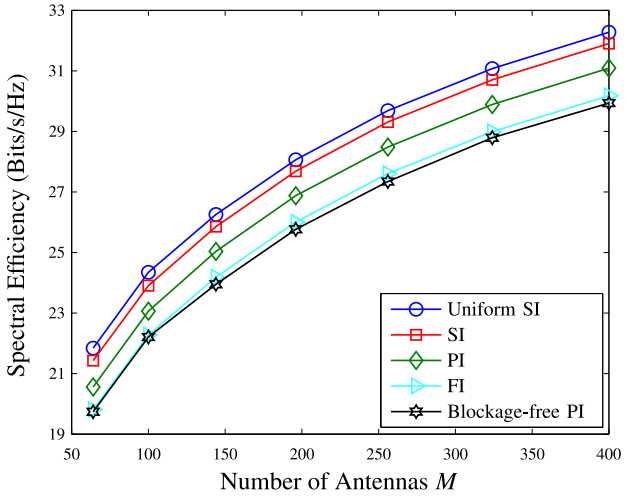


FIGURE 6. Spectral efficiency (bits/s/Hz) versus number of transmit antennas M for $Q = N = 4$, $J = 16$, and different illumination strategies.

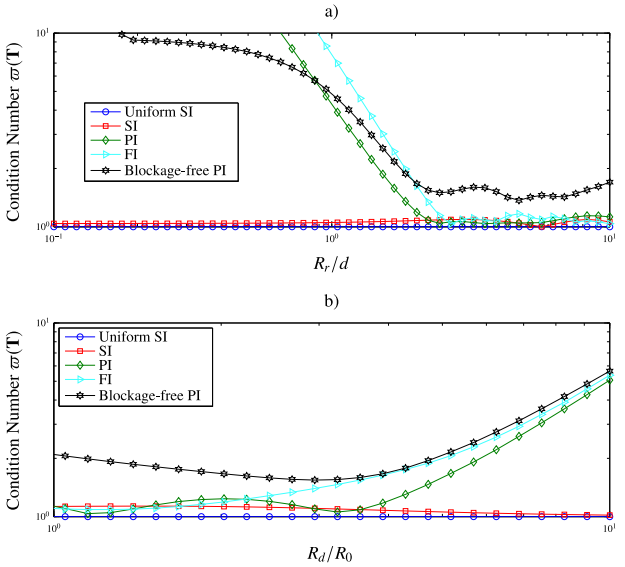


FIGURE 7. Condition number $\varpi(\mathbf{T})$ versus a) R_r for $R_d = 4R_0$ and b) R_d for $R_r = 2d$ assuming $M = 256$, $N = 4$, and different illuminations.

subset is responsible for reflection/transmission of the signal of one of the active antennas, see (20). For PI, the positioning of the active antennas minimizes the interference between the different subsets of passive antennas $\mathcal{M}_n$, $n = 1, \ldots, N$, which is beneficial for performance, whereas for FI there is significant interference between different subsets of passive antennas, which cannot be mitigated by the precoder. To study the impact of the power distribution across the passive antennas, we also show the achievable rate for uniform SI, see Proposition 1 and Corollary 1. As expected, uniform SI outperforms SI; nevertheless, the additional gain is small. Furthermore, we observe from Fig. 6 that blockage-free PI, which is designed to avoid the blockage of desired AoAs/AoDs by the active antennas for the IRS-aided architecture, achieves the lowest spectral efficiency. This is due to the higher over-the-air power loss and the more non-uniform power distribution across the intelligent surface compared to the other illumination strategies.

Next, we study the impact of the positioning of the active antennas and the intelligent surface via parameters R_r and R_d. Unfavorable positioning of the active antennas and the intelligent surface causes matrix $\mathbf{T}$ to be ill-conditioned which in turn decreases the achievable rate of any precoder design due to the reduced degrees of freedom in $\mathbf{F} = \mathbf{DTB}$. Moreover, an ill-conditioned matrix $\mathbf{T}$ leads to an increase of the power that has to be radiated by the active antennas to achieve a certain transmit power for the intelligent surface, see Lemma 1. Therefore, we study the condition number of matrix $\mathbf{T}$, i.e., $\varpi(\mathbf{T})$, for different illumination scenarios. Recall that a well-conditioned matrix $\mathbf{T}$ has a condition number close to one. In Fig. 7, we show the condition number $\varpi(\mathbf{T})$ versus a) R_r for $R_d = 4R_0$ where $R_0 \triangleq d\sqrt{\frac{M}{\pi N}}$ (see (17)) and b) R_d for $R_r = 2d$ assuming $M = 256$, $N = 4$, and different illumination strategies, namely FI, PI, SI, blockage-free PI, and uniform SI. As can be seen from Fig. 7 a), for FI, PI, and blockage-free PI, the condition number of $\mathbf{T}$ improves (i.e., decreases) as R_r increases. Whereas for SI (uniform SI), the condition number of $\mathbf{T}$ is close to one (exactly one) for the entire considered range of R_r. On the other hand, Fig. 7 b) shows that, for FI, PI, and blockage-free PI, the condition number of $\mathbf{T}$ generally increases as R_d increases which is expected since the columns of matrix $\mathbf{T}$ become more similar. Interestingly for SI (uniform SI), the condition number of $\mathbf{T}$ remains again close to one (exactly one) for the entire considered range of R_d. From Figs. 6 and 7, we conclude that SI yields a better performance than FI and PI. This makes SI a suitable illumination option especially for the ITS-aided architecture which does not face the issue of the blockage of AoAs/AoDs by the active antennas. More importantly, as far as hardware implementation is concerned, SI is simpler than FI and PI since each active antenna and its respective passive elements can be manufactured independent of the other active and passive antennas.

D. COMPARISON OF DIFFERENT MIMO ARCHITECTURES

For IRS/ITS-aided antennas, we adopt the proposed OMP-based precoder (except for Fig. 10) and the SI illumination strategy. For comparison, in addition to the benchmark schemes discussed in Section V-B, we consider the AO-based precoder in [37] for IRS-aided antennas. In this case, we adopt the FI strategy, not SI, since this precoder was not designed for SI and, as a result, has a poor performance in this case.

In Fig. 8, we show a) the spectral efficiency R (bits/s/Hz) given in (18), b) the corresponding total consumed power P_{tot} (dBm), and c) the corresponding energy efficiency WR/P_{tot} (bits/joule) versus the number of transmit antennas M for $Q = N = 4$ and $J = 16$. Our observations from Fig. 8 can be summarized as follows:

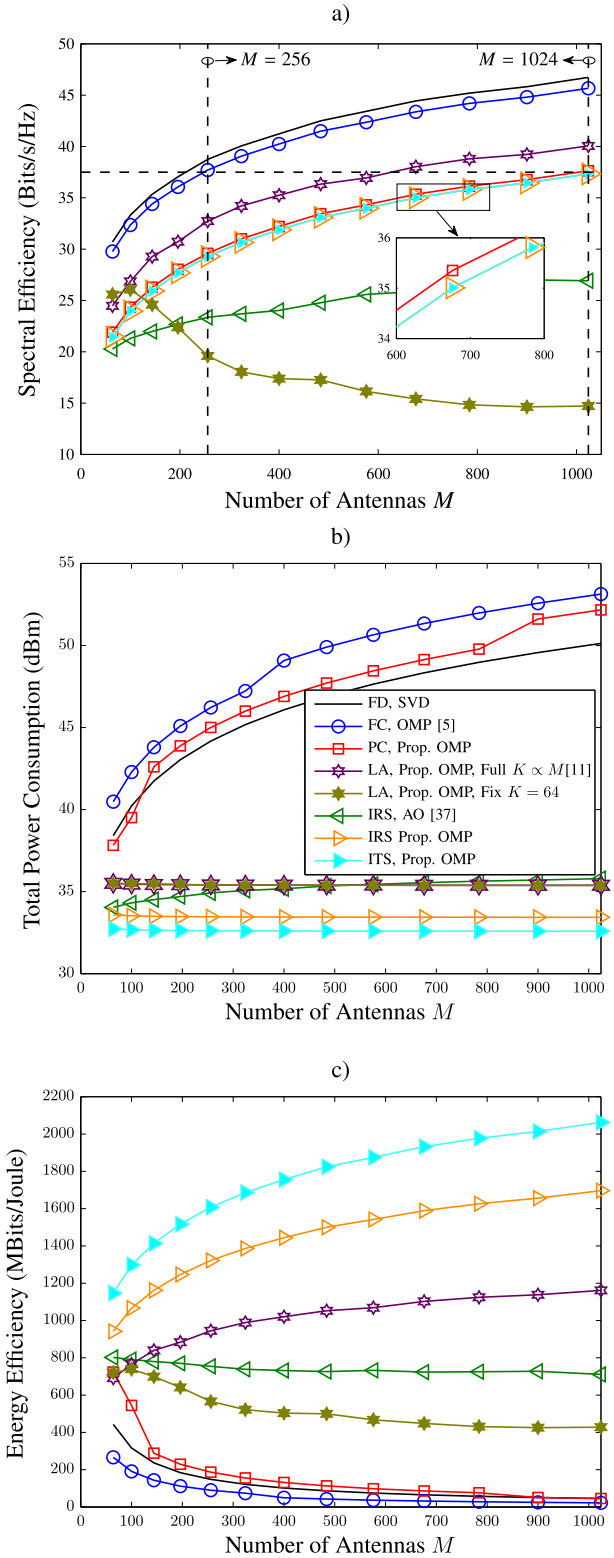


FIGURE 8. a) Spectral efficiency (bits/s/Hz), b) total consumed power P_{tot} (dBm), and c) energy efficiency (MBits/Joule) versus number of transmit antennas M for $Q = N = 4$ and $J = 16$.

- As can be seen from Fig. 8 a), the FC hybrid architecture can closely approach the spectral efficiency of the FD architecture. As expected, PC hybrid MIMO has a lower spectral efficiency compared to FC hybrid MIMO due to the fewer degrees of freedom available for beamforming in PC MIMO, i.e., only M phase shifters are used in PC MIMO whereas MN phase shifters are employed in FC MIMO. Although the IRS/ITS-aided architectures also have M phase shifters, they achieve a slightly lower spectral efficiency compared to the PC architecture due to non-uniform power distribution across the intelligent surface. The LA antennas with full K outperform the PC and IRS/ITS-aided antennas since the entire surface/lens can be used for transmission of the signal from each active antenna due to their placement on the focal arc of the lens. However, in practice, linearly increasing K with M is infeasible as active antennas are costly and bulky. For a fixed K of 64, the achievable rate of the LA antenna even decreases with M which is due the fixed number of supported AoDs and the narrow beam generated by the lens. We investigate the impact of K on the performance of LA antennas in more detail in Fig. 9.
- We also observe from Fig. 8 a) that the proposed OMP-based precoder outperforms the AO-based precoder in [37] by a large margin. This might be attributed to the fact that the iterative AO-based algorithm in [37] is more prone to getting trapped in a local optimum which is avoided by the proposed OMP-based precoder as it efficiently exploits the sparsity of mmWave channels.
- The main advantage of the IRS/ITS-aided architectures is their scalability in terms of the number of antennas M which is evident from Figs. 8 b) and c). In fact, IRS/ITS-aided MIMO (using the proposed precoder) achieve similar spectral efficiency as FD and FC MIMO if they are equipped with N times more antennas, e.g., in Fig. 8 a), FD and FC MIMO with $M = 256$ antennas and IRS/ITS-aided MIMO with $M = 1024$ antennas achieve the same spectral efficiency of 37.5 bits/s/Hz. However, from Fig. 8 b), we observe that the total transmit power of the conventional FD, FC, and PC architectures significantly increases as M increases which makes their implementation quite costly or even infeasible.[13] We note that the high power consumption of FD antennas can be attributed to the large number of DACs whereas that of the PC and FC hybrid antennas is mainly caused by the GCAs needed to compensate for the power loss in the analog network [8]. On the other hand, the total power consumption of the LA and IRS/ITS-aided architectures remains almost constant as M increases, which is mainly due to the efficient over-the-air connection between the active and passive antennas [25], [35].
- We observe from Fig. 8 b) that the power consumption of the LA architecture does not depend on the number of active antennas K. However, the overall

13. We note that the jumps in the power consumptions of the FC and PC MIMO architectures in Fig. 8 b) are due to an increase of the number in required GCAs per antenna, i.e., $\lceil \frac{L_{\text{rf}}}{G_{\text{amp}}} \rceil$, as M increases.

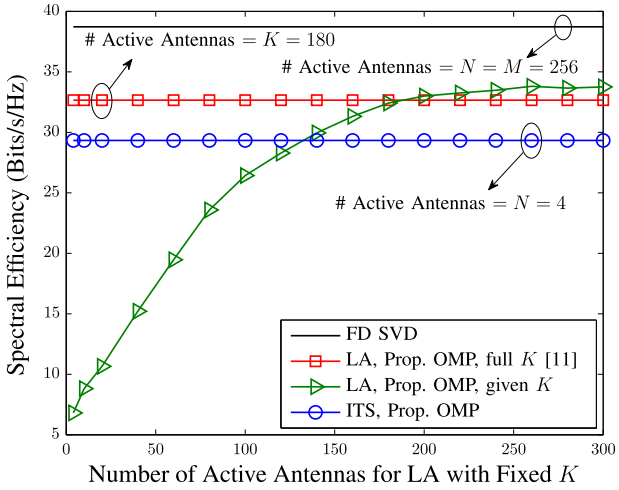


FIGURE 9. Spectral efficiency (bits/s/Hz) versus number of active antennas K in the LA MIMO architecture for $Q = N = 4$, $M = 256$, and $J = 16$.

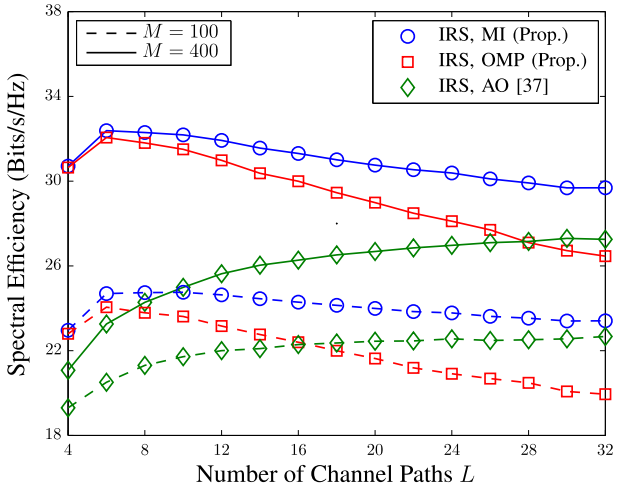


FIGURE 10. Spectral efficiency (bits/s/Hz) versus number of channel paths L for $Q = N = 4$, $M \in \{100, 400\}$, and $J \in \{16, 36\}$.

power consumption of LA antennas is higher than that of IRS/ITS-aided antennas since the active antennas have to transmit with higher power P_{rd} to achieve the same transmit power of the intelligent surface P_{tx}. In particular, for the LA architecture, the active antennas have to be placed on the focal arc of the lens, which leads to a severely non-uniform power distribution across the intelligent surface for the received signal of some of the active antennas. This issue does not exist for IRS/ITS-aided antennas which is the reason for their lower power consumption compared to LA antennas. In addition, from Fig. 8 b), we observe that the proposed ITS-aided antennas have a lower power consumption compared to the proposed IRS-aided antennas due to their higher array efficiency factor, i.e., $[\rho_{\text{srf}}]_{\text{dB}} = 2[\rho_P]_{\text{dB}} + [\rho_A^{\text{srf}}]_{\text{dB}} = -4.5$ dB and $[\rho_{\text{srf}}]_{\text{dB}} = [\rho_P]_{\text{dB}} + [\rho_A^{\text{srf}}]_{\text{dB}} = -3.5$ dB, see Table 1. Furthermore, the precoder in [37] leads to a higher power consumption than the proposed precoder since the latter employs SI whereas the former employs FI.

- We observe in Fig. 8 c) that the energy efficiency of the conventional FD, FC, and PC architectures decreases as M increases whereas the energy efficiency of the proposed IRS/ITS-aided architectures increases. This is mainly due to the high power consumption of the conventional architectures for large M. For LA antennas with fixed $K = 64$ (full K), the energy efficiency decreases (increases) due to the decreasing (increasing) spectral efficiency as M increases.

In Fig. 9, we show the spectral efficiency (bits/s/Hz) of the LA architecture versus the number of active antennas K for $Q = N = 4$, $M = 256$, and $J = 16$. In addition, we also show the performance of the LA architecture with the full K design in [11] ($K = 180$ for $M = 256$ and the considered range of AoDs), the proposed ITS-aided architecture (with 4 active antennas), and the FD architecture (with 256 active antennas). As expected the achievable rate of LA antennas improves with K due to the larger number of supported AoDs. However, the curve saturates for large K (approximately $K > M$) since in this regime, the bottleneck is the passive lens (i.e., how narrow the beam can be made). Fig. 9 shows that at $M = 256$, the proposed ITS-aided architecture with only 4 active antennas outperforms the LA architecture with $K = 132$ antennas.

In Fig. 10, we compare the performance of the AO-based precoder in [37] and the proposed MI-based and OMP-based precoders in more detail for different scattering environments and different numbers of transmit antennas. In particular, in Fig. 10, we show the spectral efficiency (bits/s/Hz) versus the number of channel paths L for $Q = N = 4$, $M \in \{100, 400\}$, and $J = 16$. From this figure, we observe that as the number of channel paths L increases, the spectral efficiency of the proposed MI-based and OMP-based precoders first increases and then decreases. This behavior is due to the fact that the proposed precoders select the N best paths. Therefore, for larger L, we have more paths to select from, which yields a diversity gain, but selecting only N out of L paths becomes a limiting factor. In contrast, the spectral efficiency of the AO-based precoder in [37] increases as L increases. This is due to the fact that the AO-based precoder in [37] does not explicitly choose its phase-shift matrix based on the transmit array response of the available paths. In fact, increasing L leads to a better conditioned channel matrix which improves the convergence behavior of the precoder in [37]. In addition, we observe from Fig. 10 that the performance gain of the proposed precoder over the AO-based precoder increases with the number of transmit and receive antennas. This behavior is in line with the results reported in the literature which state that for large M, even the optimal unconstrained precoder transmits the data over at most the N strongest channel paths [5], [7]. Finally, we observe that for all ranges of the parameters considered in Fig. 10, the proposed MI-based precoder outperforms both the proposed OMP-based precoder and the AO-based precoder in [37].

E. CSI IMPERFECTION

Next, we investigate the impact of channel estimation errors on the performance of the proposed precoder design. We assume that the channel is reconstructed by estimating the AoAs (θ_l^r, ϕ_l^r), the AoDs (θ_l^t, ϕ_l^t), and the corresponding channel coefficients h_l. The estimated parameters are modeled as $\hat{x} = x + \epsilon_x,\ x \in \{\theta_l^r, \phi_l^r, \theta_l^t, \phi_l^t, h_l\}$, where ϵ_x denotes the estimation error and is a (real for AoAs/AoDs and complex for h_l) zero-mean Gaussian RV. For simplicity, we assume that the estimation errors of different parameters are independent, the estimation error variances for the AoAs and AoDs are identical, and the number of paths L is correctly estimated. As a measure for the estimation quality, we use the widely-adopted normalized mean square error (NMSE), defined as $\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2$, where $\hat{\mathbf{H}}$ denotes the estimated channel matrix [60], [61]. We use the estimated channel matrix $\hat{\mathbf{H}}$ for the precoder design, but compute the spectral efficiency for the actual channel matrix $\mathbf{H}$. Fig. 11 shows the spectral efficiency of the FD and IRS-aided architectures versus the NMSE for $M = 400$, $Q = N = 4$, and $J = 16$. To separately investigate the impact of estimation errors of the AoAs/AoDs and the path coefficients, we consider three cases, namely *i)* the AoAs and AoDs are estimated and the path coefficients h_l are known, *ii)* the path coefficients h_l are estimated and the AoAs and AoDs are known, and *iii)* the path coefficients h_l and the AoAs/AoDs are estimated and the estimation error ϵ_x has the same variance, $\forall x \in \{\theta_l^r, \phi_l^r, \theta_l^t, \phi_l^t, h_l\}$. In addition to FD antennas and IRS-aided antennas using MI-based and OMP-based precoders, we also show results for IRS-aided antennas with a random precoder as a baseline. The random precoder $\mathbf{F} = \mathbf{DTB}$ is computed by initially generating the entries of $\mathbf{D}$ and $\mathbf{B}$ as zero-mean unit-variance complex Gaussian RVs, then normalizing each entry of $\mathbf{D}$ to ensure the unit-modulus constraint, and finally normalizing matrix $\mathbf{B}$ to ensure the transmit power constraint $\|\mathbf{F}\|_F = 1$, such that the random precoder satisfies $\mathbf{F} \in \mathcal{F}$. From Fig. 11, we observe that for small NMSE (approx. below 0.1), the performances of the considered precoders designed based on estimated channel $\hat{\mathbf{H}}$ and perfect channel $\mathbf{H}$ are similar. For the estimated AoAs/AoDs, the achievable spectral efficiency decreases as the NMSE increases. Note that, in this case, the NMSE is upper bounded by 2. Here, the upper bound corresponds to the extreme case where the estimated and the actual AoAs/AoDs are independent RVs, which yields $\mathbb{E}\{\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2\} = 2\mathbb{E}\{\|\mathbf{H}\|_F^2\}$. In contrast, for estimated h_l, although the achievable spectral efficiency decreases as NMSE increases, it saturates to a larger value. The reason for this behavior is that the values of h_l determine the N best directions among the known AoDs/AoAs for transmission and the corresponding power allocation to the beams. Since the number of paths is small, i.e., $L = 16$, and the actual AoAs/AoDs are known, we can achieve a large rate even by random path selection and random power allocation where the randomness is due to severely imperfect $\hat{h}_l$. On the other hand, if the AoDs are incorrectly estimated,

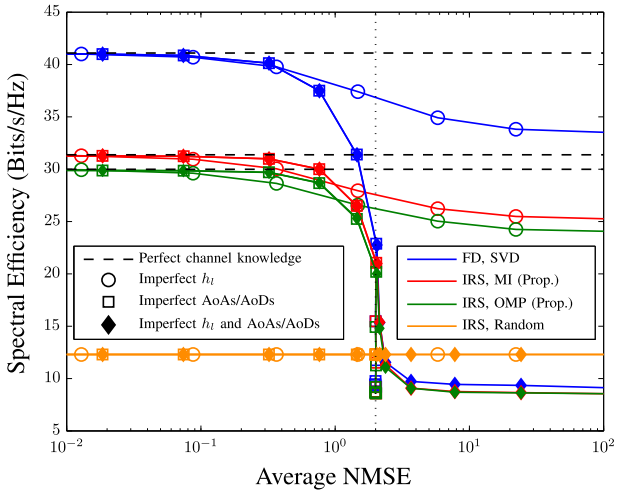


FIGURE 11. Spectral efficiency (bits/s/Hz) versus average NMSE for $M = 400$, $Q = N = 4$, $J = 16$, and $L = 16$.

the precoder transmits the signal in wrong directions which severely reduces the amount of signal power that reaches the receiver. In fact, for large NMSEs, precoder designs based on imperfect CSI are even outperformed by the random precoder which scatters the signal in all directions. Moreover, we observe from Fig. 11 that the imperfect AoAs/AoDs are the performance bottleneck when both the AoAs/AoDs and the path coefficients are imperfect. Furthermore, Fig. 11 suggests that the impact of channel estimation errors on the achievable spectral efficiency of the FD and IRS-aided architectures is quite similar.

VI. CONCLUSION AND FUTURE WORK

In the following, we first briefly present the major conclusions drawn from this article, and then we provide some directions for future research.

A. CONCLUSION

In this article, we proposed to employ IRS and ITS to realize the full potential of mmWave *ultra massive* MIMO in practice. In particular, we derived models for the corresponding precoder structure and the consumed power that accounted for the imperfections of IRS/ITS. Furthermore, we proposed different illumination strategies for the active antennas to realize the full potential of IRS/ITS-aided MIMO. Based on the derived precoder structure and exploiting the sparsity of mmWave channels, we designed two efficient precoders for IRS/ITS-aided antennas; namely MI- and OMP-based precoders. Our comprehensive simulation studies provided the following interesting insights for system design: 1) A proper positioning of the active antennas with respect to the intelligent surface leads to a considerable improvement in spectral efficiency of IRS/ITS-aided MIMO architectures. 2) In poor scattering environments, the proposed OMP-based precoder approaches the performance of the proposed MI-based precoder, whereas in rich scattering environments, there is a performance gap between these two precoders. Therefore, in channels with poor scattering, the proposed

OMP-based precoder is preferable due to its lower computational complexity, whereas in channels with rich scattering, the proposed MI-based precoder may be preferable due to its superior performance. 3) For large numbers of transmit antennas, the conventional MIMO architectures are either energy inefficient (due to the high power consumption of the large numbers of RF chains of the FD MIMO architecture and the large power loss in the RF network of the hybrid MIMO architectures) or expensive and bulky (due to the large numbers of active antennas required for the LA MIMO architecture). 4) The proposed IRS/ITS-aided MIMO architectures are highly energy efficient (because of the almost lossless over-the-air connection between the active antennas and the passive intelligent surface) and fully scalable in terms of the number of transmit antennas (since unlike for LA antennas, only few active antennas are required).

B. FUTURE WORK

The results of this article can be extended in several directions.

Reception Design: In modern systems, the same node usually serves as both transmitter and receiver, e.g., a base station (BS) in cellular networks. Since the resources used for transmission and reception are typically orthogonal in time and/or frequency, it is customary to treat the transmission and reception problems separately. In this article, we studied the use of IRS/ITS-aided antennas at the transmitter and focused on the modeling and design of the precoder. An important direction for future research is to investigate the application of IRS/ITS-aided antennas at the receiver and to develop corresponding combining schemes.

Multi-user Communications: In this article, we considered a point-to-point MIMO system, which is a suitable model for, e.g., data backhauling from a macro BS to a small-cell BS. Other interesting network architectures include the uplink and downlink between a BS and multiple mobile users. The design of corresponding precoders and combiners when the BS employs IRS/ITS-aided antennas constitutes an interesting research problem for future work, see [44]. Hereby, one may consider different precoder design criteria including maximization of the users' sum rate, fairness among the users, and the coverage area.

Channel Estimation: In this article, we assumed that knowledge of channel matrix $\mathbf{H}$ is available at both transmitter and receiver, which is a widely-adopted assumption for precoder design in the literature [5], [9], [37], [44], [45], [50], [53]–[55], [62], [63]. In practice, channel matrix $\mathbf{H}$ has to be estimated from pilot symbols. A simple approach is based on beam steering whereby the transmitter and the receiver create pencil beams along specific AoDs (θ_l^t, ϕ_l^t) and AoAs (θ_l^r, ϕ_l^r), respectively, and estimate the corresponding channel coefficient h_l. Channel matrix $\mathbf{H}$ can be estimated by repeating this procedure for all possible AoDs and AoAs with a predefined resolution. Since this simple approach may cause a huge training overhead for high-quality channel estimation, more efficient channel estimation techniques have been developed in [3], [61], [64] using compressed sensing, which exploits the sparsity of the mmWave channel. Although the beam steering technique can be directly applied to IRS/ITS-aided antennas, the existing compressed sensing-based schemes rely on the precoder structure of the conventional hybrid beamforming architectures, and hence, have to be suitably modified for application to IRS/ITS-aided antennas. Moreover, since the number of pilot symbols is limited due to the finite coherence time of the channel, estimation errors are unavoidable. This motivates the design of robust precoders for IRS/ITS-aided architectures which account for estimation errors, see [65] for robust precoder designs for conventional hybrid and FD architectures.

APPENDIX A
PROOF OF LEMMA 1

The power radiated by the active antennas is obtained as $P_{\text{rd}} = \text{trace}\{\mathbb{E}\{\bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathsf{H}}\}\} \overset{(a)}{=} P_{\text{tx}}\,\text{trace}\{\mathbf{B}\mathbf{B}^{\mathsf{H}}\} = P_{\text{tx}}\|\mathbf{B}\|_F^2$, where equality (a) follows from $\bar{\mathbf{x}} = \sqrt{P_{\text{tx}}}\mathbf{B}\mathbf{s}$ and $\mathbb{E}\{\mathbf{s}\mathbf{s}^{\mathsf{H}}\} = \mathbf{I}_Q$. Similarly, the transmit power radiated by the intelligent surface is given by $P_{\text{tx}} = \text{trace}\{\mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathsf{H}}\}\} \overset{(a)}{=} P_{\text{tx}}\,\text{trace}\{\mathbf{F}\mathbf{F}^{\mathsf{H}}\} \overset{(b)}{=} P_{\text{tx}}\,\text{trace}\{\mathbf{D}\mathbf{T}\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\} \overset{(c)}{=} P_{\text{tx}}\,\text{trace}\{\mathbf{T}\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{D}\} = P_{\text{tx}}\|\mathbf{T}\mathbf{B}\|_F^2$, where equality (a) follows from $\bar{\mathbf{x}} = \sqrt{P_{\text{tx}}}\mathbf{F}\mathbf{s}$ in (1), equality (b) follows from $\mathbf{F} = \mathbf{D}\mathbf{T}\mathbf{D}$ from Proposition 1, and equality (c) follows from the matrix equality $\text{trace}\{\mathbf{A}\mathbf{B}\} = \text{trace}\{\mathbf{B}\mathbf{A}\}$. Therefore, $\|\mathbf{T}\mathbf{B}\|_F^2 = 1$ has to hold. Moreover, the Frobenius norm of the product of two matrices are bounded as [66]

$$\sigma_{\min}^2(\mathbf{T})\|\mathbf{B}\|_F^2 \le \|\mathbf{T}\mathbf{B}\|_F^2 \le \sigma_{\max}^2(\mathbf{T})\|\mathbf{B}\|_F^2. \tag{31}$$

Using the above bounds and recalling that $\|\mathbf{T}\mathbf{B}\|_F^2 = 1$ holds, we obtain bounds $\sigma_{\max}^{-2}(\mathbf{T}) \le \|\mathbf{B}\|_F^2 \le \sigma_{\min}^{-2}(\mathbf{T})$ whose substitution in $P_{\text{rd}} = P_{\text{tx}}\|\mathbf{B}\|_F^2$ leads to the tighter lower and upper bounds in (13). Moreover, $\sigma_{\min}^2(\mathbf{T})$ and $\sigma_{\max}^2(\mathbf{T})$ are bounded respectively as

$$\begin{aligned}\|\mathbf{T}\|_F^2 &= \sum_n \sigma_n^2(\mathbf{T}) \\ &= \sigma_{\min}^2(\mathbf{T})\sum_n \frac{\sigma_n^2(\mathbf{T})}{\sigma_{\min}^2(\mathbf{T})} \le N\sigma_{\min}^2(\mathbf{T})\varpi^2(\mathbf{T}) \qquad (32a)\\ \|\mathbf{T}\|_F^2 &= \sum_n \sigma_n^2(\mathbf{T}) \\ &= \sigma_{\max}^2(\mathbf{T})\sum_n \frac{\sigma_n^2(\mathbf{T})}{\sigma_{\max}^2(\mathbf{T})} \ge \frac{N\sigma_{\max}^2(\mathbf{T})}{\varpi^2(\mathbf{T})}. \qquad (32b)\end{aligned}$$

On the other hand, $\|\mathbf{T}\|_F^2$ can be computed as

$$\begin{aligned}\|\mathbf{T}\|_F^2 &= \sum_n \rho_{\text{srf}} \\ &\quad \times \underbrace{\sum_m \left(\frac{\lambda}{4\pi r_{m,n}}\right)^2 G^a\left(\theta_{m,n}^p, \phi_{m,n}^p\right) G^p\left(\theta_{m,n}^a, \phi_{m,n}^a\right)}_{A_n} \\ &= N\rho_{\text{srf}}\rho_A^{\text{ant}}\rho_S, \qquad (33)\end{aligned}$$

where A_n is the fraction of power radiated from active antenna n and collected by the surface. A_n is related to spillover loss ρ_S and the passive antenna aperture loss ρ_A^{ant}. Assuming that ρ_A^{ant} and ρ_S are identical for all active antennas, we obtain $A_n = \rho_A^{\text{ant}}\rho_S$, $\forall n$. Substituting (33) in (32a) and then in (31) leads to the looser lower and upper bounds in (13) and concludes the proof.

APPENDIX B
PROOF OF LEMMA 2

The proof follows from similar arguments as those provided in [5]–[7], [50] for the precoder design of conventional MIMO systems. In particular, for the achievable rate $R(\mathbf{F})$, the precoder appears in the term $\mathbf{HF} = \mathbf{HF}_{\mathcal{H}_t} + \mathbf{HF}_{\mathcal{H}_t^\perp}$. The elements of matrix $\mathbf{HF}_{\mathcal{H}_t^\perp}$ are obtained based on the multiplication of the rows of $\mathbf{H}$ and the columns of $\mathbf{F}_{\mathcal{H}_t^\perp}$. Moreover, the rows of $\mathbf{H}$ belong to space $\mathcal{H}_t$ whereas the columns of $\mathbf{F}_{\mathcal{H}_t^\perp}$ belong to space $\mathcal{H}_t^\perp$. Since $\mathcal{H}_t$ and $\mathcal{H}_t^\perp$ are orthogonal, we obtain $\mathbf{HF}_{\mathcal{H}_t^\perp} = \mathbf{0}_{J,N}$. Therefore, $\mathbf{F}_{\mathcal{H}_t^\perp}$ does not impact the achievable rate in (18). Moreover, we have $\|\mathbf{F}\|_F^2 = \text{trace}(\mathbf{FF}^{\mathsf{H}}) = \text{trace}((\mathbf{F}_{\mathcal{H}_t} + \mathbf{F}_{\mathcal{H}_t^\perp})(\mathbf{F}_{\mathcal{H}_t} + \mathbf{F}_{\mathcal{H}_t^\perp})^{\mathsf{H}}) = \text{trace}(\mathbf{F}_{\mathcal{H}_t}\mathbf{F}_{\mathcal{H}_t}^{\mathsf{H}} + \mathbf{F}_{\mathcal{H}_t^\perp}\mathbf{F}_{\mathcal{H}_t^\perp}^{\mathsf{H}}) = \|\mathbf{F}_{\mathcal{H}_t}\|_F^2 + \|\mathbf{F}_{\mathcal{H}_t^\perp}\|_F^2 \geq \|\mathbf{F}_{\mathcal{H}_t}\|_F^2$ where we used $\text{trace}(\mathbf{F}_{\mathcal{H}_t}\mathbf{F}_{\mathcal{H}_t^\perp}^{\mathsf{H}}) = \text{trace}(\mathbf{F}_{\mathcal{H}_t^\perp}\mathbf{F}_{\mathcal{H}_t}^{\mathsf{H}}) = 0$. This concludes the proof.

APPENDIX C
PROOF OF LEMMA 3

Let us define $\widetilde{\mathbf{B}} = (\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1)^{\frac{1}{2}}\mathbf{B} \in \mathbb{C}^{Q\times N}$. Assuming $\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1$ is a non-singular matrix, the constraint in (21) is rewritten as

$$\begin{aligned}
&\text{trace}\Big(\mathbf{C}_1\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{C}_1^{\mathsf{H}}\Big)\\
&= \text{trace}\Bigg(\mathbf{C}_1\Big(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1\Big)^{-\frac{1}{2}}\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^{\mathsf{H}}\Big(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1\Big)^{-\frac{\mathsf{H}}{2}}\mathbf{C}_1^{\mathsf{H}}\Bigg)\\
&\overset{(a)}{=} \text{trace}\Bigg(\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^{\mathsf{H}}\Big(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1\Big)^{-\frac{\mathsf{H}}{2}}\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1\Big(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1\Big)^{-\frac{1}{2}}\Bigg)\\
&= \text{trace}\Big(\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^{\mathsf{H}}\Big), \qquad (34)
\end{aligned}$$

where for equality (a), we used the relation $\text{trace}(\mathbf{XY}) = \text{trace}(\mathbf{YX})$ for $\mathbf{X} \in \mathbb{C}^{n\times m}$ and $\mathbf{Y} \in \mathbb{C}^{m\times n}$. Based on this result and using the definition $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{C}_1(\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1)^{-\frac{1}{2}} \in \mathbb{C}^{J\times N}$, the problem in (21) is rewritten as

$$\begin{aligned}
&\underset{\widetilde{\mathbf{B}}\in\mathbb{C}^{N\times Q}}{\text{maximize}} \quad \Big|\mathbf{I}_J + \gamma\widetilde{\mathbf{H}}\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^{\mathsf{H}}\widetilde{\mathbf{H}}^{\mathsf{H}}\Big|\\
&\text{subject to :} \quad \text{trace}\Big(\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^{\mathsf{H}}\Big) \leq 1. \qquad (35)
\end{aligned}$$

The above problem has the form of MI maximization for an FD MIMO system with equivalent channel matrix $\widetilde{\mathbf{H}}$. Thus, the solution is found via the waterfilling algorithm as $\widetilde{\mathbf{B}} = [\mathbf{v}_1, \ldots, \mathbf{v}_Q]\mathbf{Z}$ [7]. Then, the optimal baseband precoder is given by $\mathbf{B} = (\mathbf{C}_1^{\mathsf{H}}\mathbf{C}_1)^{-\frac{1}{2}}\widetilde{\mathbf{B}}$ which concludes the proof.

REFERENCES

[1] V. Jamali, A. M. Tulino, G. Fischer, R. Müller, and R. Schober, "Scalable and energy-efficient millimeter massive MIMO architectures: Reflect-array and transmit-array antennas," in *Proc. IEEE ICC*, May 2019, pp. 1–7.

[2] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 836–869, 2nd Quart., 2017.

[3] A. V. Delgado, M. Sanchez-Fernandez, J. Llorca, and A. Tulino, "Feasible transmission strategies for downlink MIMO in sparse millimeter-wave channels," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 49–55, Jul. 2018.

[4] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[6] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[7] R. Ghanaatian, V. Jamali, A. Burg, and R. Schober, "Feedback-aware precoding for millimeter wave massive MIMO systems," in *Proc. IEEE PIMRC*, Sep. 2019, pp. 1–7.

[8] 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 Syst. Mag.*, vol. 19, no. 2, pp. 33–58, 2nd Quat., 2019.

[9] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[10] Y. Zeng and R. Zhang, "Cost-effective millimeter-wave communications with lens antenna array," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 81–87, Aug. 2017.

[11] Y. Zeng, L. Yang, and R. Zhang, "Multi-user millimeter wave MIMO with full-dimensional lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2800–2814, Apr. 2018.

[12] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[13] D. Popovic and Z. Popovic, "Multibeam antennas with polarization and angle diversity," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 651–657, May 2002.

[14] C. A. Balanis, "Antenna theory, analysis and design," 3rd Edition. NJ, USA: Wiley, 2016.

[15] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE GLOBECOM*, Dec. 2018, pp. 1–6.

[16] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A novel communication paradigm for high capacity and security via programmable indoor wireless environments in next generation wireless systems," *Ad Hoc Netw.*, vol. 87, pp. 1–16, May 2019.

[17] M. D. Renzo *et al.*, "Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, p. 129, May 2019.

[18] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[19] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 5394–5409, Nov. 2019.

[20] M. Najafi and R. Schober, "Intelligent reflecting surfaces for free space optical communications," in *Proc. GLOBECOM*, Dec. 2019, pp. 1–7.

[21] R. Karasik, O. Simeone, M. Di Renzo, and S. Shamai, "Beyond max-SNR: Joint encoding for reconfigurable intelligent surfaces," in *Proc. ISIT*, May 2020, pp. 2965–2970.

[22] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physical modeling and scalable optimization of large intelligent reflecting

surfaces," *IEEE Trans. Commun.*, early access. [Online]. Available: https://arxiv.org/abs/2004.12957

[23] V. Jamali, M. Najafi, R. Schober, and H. V. Poor, "Power efficiency, overhead, and complexity tradeoff in IRS-assisted communications–quadratic phase-shift design," *IEEE Commun. Lett.*, early access. [Online]. Available: https://arxiv.org/pdf/2009.05956

[24] J. Zuo, Y. Liu, E. Basar, and O. A. Dobre, "Intelligent reflecting surface enhanced millimeter-wave NOMA systems," 2020. [Online]. Available: arXiv:2005.01562.

[25] A. H. Abdelrahman, F. Yang, A. Z. Elsherbeni, and P. Nayeri, "Analysis and design of transmitarray antennas," *Synth. Lectures Antennas*, vol. 6, no. 1, pp. 1–175, 2017.

[26] H. Luyen, Z. Zhang, J. H. Booske, and N. Behdad, "Wideband, beam-steerable reflectarrays based on minimum-switch topology, polarization-rotating unit cells," *IEEE Access*, vol. 7, pp. 36568–36578, 2019.

[27] Z. Popovic and A. Mortazawi, "Quasi-optical transmit/receive front ends," *IEEE Trans. Microw. Theory Techn.*, vol. 46, no. 11, pp. 1964–1975, Nov. 1998.

[28] L. Di Palma, "Reconfigurable transmitarray antennas at millimeter-wave frequencies," Ph.D. dissertation, Dept. Comput. Sci., Univ. Rennes, Rennes, France, 2015.

[29] K. T. Pham, A. Clemente, E. Fourn, F. Diaby, L. Dussopt, and R. Sauleau, "Low-cost metal-only transmitarray antennas at ka-band," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 6, pp. 1243–1247, Jun. 2019.

[30] H. Hasani, J. S. Silva, S. Capdevila, M. Garcma-Vigueras, and J. R. Mosig, "Dual-band circularly polarized transmitarray antenna for satellite communications at 20/30 GHz," *IEEE Trans. Antennas Propag.*, vol. 67, no. 8, pp. 5325–5333, Aug. 2019.

[31] A. Bereyhi, V. Jamali, R. R. Müller, G. Fischer, R. Schober, and A. M. Tulino, "PAPR-limited precoding in massive MIMO systems with reflect-and transmit-array antennas," in *Proc. Asilomar Conf. Signal Syst. Comput.*, Nov. 2019, pp. 1690–1694.

[32] A. Bereyhi, V. Jamali, R. R. Müller, A. M. Tulino, G. Fischer, and R. Schober, "A single-RF architecture for multiuser massive MIMO via reflecting surfaces," in *Proc. ICASSP*, May 2020, pp. 8688–8692.

[33] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2746–2758, May 2018.

[34] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Performance analysis of large intelligent surfaces (LISs): Asymptotic data rate and channel hardening effects," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2052–2065, Mar. 2020.

[35] P. Nayeri, F. Yang, and A. Z. Elsherbeni, *Reflectarray Antennas: Theory, Designs and Applications*. Hoboken, NJ, USA: Wiley, 2018.

[36] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.

[37] Z. Zhou, N. Ge, Z. Wang, and S. Chen, "Hardware-efficient hybrid precoding for millimeter wave systems with multi-feed reflectarrays," *IEEE Access*, vol. 6, pp. 6795–6806, 2018.

[38] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, Jul. 2013.

[39] F. C. Commission, "Use of spectrum bands above 24 GHz for mobile radio services," *Fed. Regist.*, vol. 81, no. 164, pp. 58270–58308, 2016.

[40] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[41] M. Arrebola, J. A. Encinar, and M. Barba, "Multifed printed reflectarray with three simultaneous shaped beams For LMDS central station antenna," *IEEE Trans. Antennas Propag.*, vol. 56, no. 6, pp. 1518–1527, Jun. 2008.

[42] D. M. Pozar, S. D. Targonski, and H. Syrigos, "Design of millimeter wave microstrip reflectarrays," *IEEE Trans. Antennas Propag.*, vol. 45, no. 2, pp. 287–296, Feb. 1997.

[43] G. Fischer, "Next-generation base station radio frequency architecture," *Bell Labs Tech. J.*, vol. 12, no. 2, pp. 3–18, 2007.

[44] C. Lin and G. Y. Li, "Energy-efficient design of indoor mmWave and sub-THz systems with antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4660–4672, Jul. 2016.

[45] 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. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 298–312, May 2018.

[46] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog-digital precoding revisited under realistic RF modeling," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 528–531, Oct. 2016.

[47] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 183–198, Jan. 2013.

[48] L. G. Maloratsky, "Electrically tunable switched-line diode phase shifters," *High Frequency Electronics*. Bedford, NH, USA: Summit Technical Media, 2010, pp. 16–21.

[49] Switched Line Phase Shifters. (2020). *Microwave Encyclopedia*. [Online]. Available: https://www.microwaves101.com/encyclopedias/switched-line-phase-shifters

[50] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[51] M. A. Sedaghat, V. I. Barousis, R. R. Müller, and C. B. Papadias, "Load modulated arrays: A low-complexity antenna," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 46–52, Mar. 2016.

[52] A. K. Khandani, "Media-based modulation: A new approach to wireless transmission," in *Proc. ISIT*, Jul. 2013, pp. 3050–3054.

[53] G. Zang, Y. Cui, H. V. Cheng, F. Yang, L. Ding, and H. Liu, "Optimal hybrid beamforming for multiuser massive MIMO systems with individual SINR constraints," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 532–535, Apr. 2019.

[54] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[55] D. Castanheira, P. Lopes, A. Silva, and A. Gameiro, "Hybrid beamforming designs for massive MIMO millimeter-wave heterogeneous systems," *IEEE Access*, vol. 5, pp. 21806–21817, 2017.

[56] M. Najafi, V. Jamali, D. W. K. Ng, and R. Schober, "C-RAN with hybrid RF/FSO fronthaul links: Joint optimization of fronthaul compression and RF time allocation," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8678–8695, Dec. 2019.

[57] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU, 2012.

[58] J. Y. Lau, "Reconfigurable transmitarray antennas," Ph.D. dissertation, Dept. Elect. Eng., Univ. Toronto, Toronto, ON, CANADA, 2012.

[59] D. M. Pozar, *Microwave Engineering*. New York, NY, USA: Wiley, 2009.

[60] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[61] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Select. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jun. 2017.

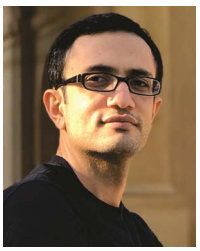
[62] J. Mirza, G. Zheng, K.-K. Wong, and S. Saleem, "Joint beamforming and power optimization for D2D underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8324–8335, Sep. 2018.

[63] M. Zeng, W. Hao, O. A. Dobre, and H. V. Poor, "Energy-efficient power allocation in uplink mmWave massive MIMO with NOMA," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3000–3004, Mar. 2019.

[64] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[65] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 484–498, Jun. 2018.

[66] E. Bodewig, *Matrix Calculus*. Amsterdam, The Netherlands: Elsevier, 2014.



VAHID JAMALI (Member, IEEE) received the B.Sc. and M.Sc. degrees (Hons.) in electrical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2010 and 2012, respectively, and the Ph.D. degree (with Distinction) from the Friedrich-Alexander University (FAU) Erlangen–Nürnberg, Erlangen, Germany, in 2019.

In 2017, he was a Visiting Research Scholar with Stanford University, Stanford, CA, USA. He is currently a Postdoctoral Researcher with the Institute for Digital Communication, FAU. His research interests include wireless and molecular communications, Bayesian inference and learning, and multiuser information theory. He received several awards for his publications and research work, including the Best Paper Award from the IEEE International Conference on Communications in 2016, the Doctoral Research Grant from the German Academic Exchange Service (DAAD) in 2017, the Goldener Igel Publication Award from the Telecommunications Laboratory (LNT), FAU, in 2018, the Best Ph.D. Thesis Presentation Award from the IEEE Wireless Communications and Networking Conference in 2018, the Best Paper Award (co-recipient) from the ACM International Conference on Nanoscale Computing and Communication in 2019, the Best Journal Paper Award (Literaturpreis) from the German Information Technology Society (ITG) in 2020, the Best Paper Award (third place, co-recipient) from the Asilomar Conference on Signals, Systems, and Computers in 2020, and the Postdoctoral Research Fellowship by the German Research Foundation (DFG) in 2020. He has served as a member of the Technical Program Committee for several IEEE conferences and he is currently an Associate Editor of the IEEE COMMUNICATIONS LETTERS and IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.

ANTONIA M. TULINO (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the Seconda Universita degli Studi di Napoli, Italy, in 1999.

She held research positions with the Center for Wireless Communications, Princeton University, Oulu, Finland, and also with the Universita degli Studi del Sannio, Benevento, Italy. From 2002 to 2016, she was an Associate Professor with the Università degli Studi di Napoli Federico II, where she has been a Full Professor since 2017. Since 2009, she has been collaborating with Nokia Bell Labs. Since October 2019, she has been a Research Professor with the Department of Electrical and Computer Engineering, NYU Tandon School of Engineering, New York, NY, USA. Her research interests lie in the area of communication systems approached with the complementary tools provided by signal processing, information theory, and random matrix theory. She has received several paper awards and among the others the 2009 Stephen O. Rice Prize in the Field of Communications Theory for the best paper published in the IEEE TRANSACTIONS ON COMMUNICATIONS in 2008. She was a recipient of the UC3M-Santander Chair of Excellence from 2018 to 2019. She has been a Principal Investigator of several research projects sponsored by the European Union and the Italian National Council, and was selected by the National Academy of Engineering for the Frontiers of Engineering Program in 2013. From 2011 to 2013, she was a member of the Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY. Since 2019, she has been the Chair of the Information Theory Society Fellows Committee.

GEORG FISCHER (Senior Member, IEEE) was born in Wachtendonk, Germany, in 1965. He received the Diploma degree in electrical engineering with a focus on communications and microwave from RWTH Aachen University, Aachen, Germany, in 1992, and the Dr.Ing. degree in electrical engineering from the University of Paderborn, Paderborn, Germany, in 1997.

From 1993 to 1996, he was a Research Assistant with the University of Paderborn, where he was involved in adaptive antenna array systems for mobile satellite communications. From 1996 to 2008, he performed research with Bell Laboratories, Lucent (later Alcatel-Lucent), where he focused on the RF and digital architecture of mobile communication base stations for global system for mobile communications (GSM), universal mobile telecommunications system, and features for network coverage and capacity enhancements. In 2000, he became a Bell Laboratories Distinguished Member of Technical Staff, and in 2001, he became a Bell Laboratories Consulting Member of Technical Staff. He also acted as the Chairman of the European Telecommunications Standards Institute during the physical-layer standardization of the GSM-EDGE system. From 2001 to 2007, he was a part-time Lecturer with the University of Erlangen–Nuremberg, Erlangen, Germany, during that time, he lectured on base station technology. Since 2008, he has been a Professor of Electronics Engineering with the University of Erlangen–Nuremberg. He holds over 50 patents concerning microwave and communications technology. His research interests are transceiver design, analog/digital partitioning, converters, enhanced amplifier architectures, duplex filters, metamaterial structures, GaN transistor technology and circuit design, and RF microelectromechanical systems with a specific emphasis on frequency agile, tunable, and reconfigurable RF systems for software-defined radio applications. His current research interests concentrate on medical electronics, such as using microwaves for detection of vital parameters. He is also a Senior Member of the IEEE Microwave Theory and Techniques, Antennas and Propagation, Computer Society, and Vehicular Technology Society. He is also a member of VDE-ITG and the European Microwave Association.

RALF R. MÜLLER (Senior Member, IEEE) was born in Schwabach, Germany, in 1970. He received the Dipl.-Ing. and Dr.-Ing. degrees (with Distinction) from Friedrich-Alexander-Universität (FAU) Erlangen–Nuremberg, Erlangen, Germany, in 1996 and 1999, respectively.

From 2000 to 2004, he directed a research group with the Telecommunications Research Center Vienna, Austria, and taught as an Adjunct Professor with TU Wien. In 2005, he was appointed as a Full Professor with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway. In 2013, he joined the Institute for Digital Communications, FAU Erlangen–Nuremberg. He held visiting appointments with Princeton University, USA; Institute Eurécom, France; the University of Melbourne, Australia; the University of Oulu, Finland; the National University of Singapore; Babeş-Bolyai University; Cluj-Napoca, Romania; Kyoto University, Japan; FAU Erlangen–Nuremberg; and TU München, Germany. He received the Leonard G. Abraham Prize (jointly with Sergio Verdú) for the paper "Design and analysis of low-complexity interference mitigation on vector channels" from the IEEE Communications Society. He was presented awards for his dissertation "Power and bandwidth efficiency of multiuser systems with random spreading" by the Vodafone Foundation for Mobile Communications and the German Information Technology Society (ITG). He received the ITG Award for the paper "A random matrix model for communication via antenna arrays" as well as the Philipp-Reis Award (jointly with Robert Fischer). He served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2003 to 2006 and on the executive editorial board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2014 to 2016.

ROBERT SCHOBER (Fellow, IEEE) received the Diplom (Univ.) and Ph.D. degrees in electrical engineering from the Friedrich-Alexander University (FAU) Erlangen–Nürnberg, Erlangen, Germany, in 1997 and 2000, respectively.

From 2002 to 2011, he was a Professor and a Canada Research Chair with the University of British Columbia (UBC), Vancouver, Canada. Since January 2012, he has been an Alexander von Humboldt Professor and the Chair for Digital Communication with FAU. His research interests fall into the broad areas of communication theory, wireless communications, and statistical signal processing. He received several awards for his work, including the 2002 Heinz Maier-Leibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, the 2011 Alexander von Humboldt Professorship, the 2012 NSERC E.W.R. Stacie Fellowship, and the 2017 Wireless Communications Recognition Award by the IEEE Wireless Communications Technical Committee. He is listed as a 2017 Highly Cited Researcher by the Web of Science and a Distinguished Lecturer of the IEEE Communications Society (ComSoc). From 2012 to 2015, he served as the Editor-in-Chief for the IEEE TRANSACTIONS ON COMMUNICATIONS. He is currently the Chair of the Steering Committee of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS, a member of the Editorial Board of the PROCEEDINGS OF THE IEEE, a Member-at-Large of the Board of Governors of ComSoc, and the ComSoc Director of Journals. He is a Fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada.