Power Scaling Laws and Near-Field Behaviors of Massive MIMO and Intelligent Reflecting Surfaces

EMIL BJÖRNSON [1] (Senior Member, IEEE), AND LUCA SANGUINETTI [2] (Senior Member, IEEE)

[1]Department of Electrical Engineering (ISY), Linköping University, 58183 Linköping, Sweden

[2]Dipartimento di Ingegneria dell'Informazione, University of Pisa, 56122 Pisa, Italy

CORRESPONDING AUTHOR: E. BJÖRNSON (e-mail: emil.bjornson@liu.se)

ABSTRACT The use of large arrays might be the solution to the capacity problems in wireless communications. The signal-to-noise ratio (SNR) grows linearly with the number of array elements N when using Massive MIMO receivers and half-duplex relays. Moreover, intelligent reflecting surfaces (IRSs) have recently attracted attention since these can relay signals to achieve an SNR that grows as N^2, which seems like a major benefit. In this article, we use a deterministic propagation model for a planar array of arbitrary size, to demonstrate that the mentioned SNR behaviors, and associated power scaling laws, only apply in the far-field. They cannot be used to study the regime where $N \to \infty$. We derive an exact channel gain expression that captures three essential near-field behaviors and use it to revisit the power scaling laws. We derive new finite asymptotic SNR limits but also conclude that these are unlikely to be approached in practice. We further prove that an IRS-aided setup cannot achieve a higher SNR than an equal-sized Massive MIMO setup, despite its faster SNR growth. We quantify analytically how much larger the IRS must be to achieve the same SNR. Finally, we show that an optimized IRS does not behave as an "anomalous" mirror but can vastly outperform that benchmark.

INDEX TERMS Intelligent reflecting surface, reconfigurable intelligent surface, software-controlled metasurface, massive MIMO, regenerative MIMO relays, asymptotic limits, power scaling law, near-field, far-field.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (mMIMO) is the key physical layer technology in 5G [2]. In a nutshell, mMIMO uses a base station with many antennas (e.g., ≥ 64) to deliver large array gains and perform spatial multiplexing of many users on the same time-frequency resource [3]–[5]. In this way, the spectral efficiency (SE) can be increased by, at least, an order of magnitude compared to 4G and mmWave communications can be enabled in mobile networks. Due to the success of mMIMO, it is expected that beyond 5G systems will make use of even larger arrays and wider spectrum ranges [6], [7]. The arrays can either consist of active or passive elements, and both cases are considered in this article.

The active arrays are essentially mMIMO transceivers but with many more antenna elements than what is considered in 5G (i.e., much more than 64). To make this clear, the research community has recently used new names to describe this category; for example, *large intelligent surfaces* [8], *extremely large aperture arrays* [9], and *holographic MIMO* [10]. However, in this article, we will simply refer to it as mMIMO since asymptotically large arrays have been analyzed since the inception of mMIMO [3], [11], [12].

The passive arrays are large metasurfaces [13], [14] that are deployed somewhere in the propagation environment (in between the transmitter and receiver) to support the transmission from a source to a destination by creating and shaping additional propagation paths. A metasurface

consists of many sub-wavelength-sized elements that each acts as a diffuse scatterer [13] but with the special feature of being able to adjust the phase (i.e., time delay) and polarization. By controlling the phase-shifts of the individual elements, the metasurface can "reflect" an incident wave as a beam in the desired direction [14]; the physics is the same as for beamforming using a phased array, except that the array then generates the signal locally. The concept of real-time controllable metasurfaces has recently received much attention in the communications society and is called *intelligent reflecting surface (IRS)* [15], [16], *software-controlled metasurface* [17], [18], and *reconfigurable intelligent surface* [19]–[21]. We will call it IRS in this article and consider the case when the metasurface is used as a relay that reflects the signal from the source towards the destination in an effort to maximize the signal-to-noise ratio (SNR). Using conventional relaying terminology [22], [23], an IRS is a transparent relay since it processes the received signal in the analog domain and operates in full duplex since the signals are received and reflected simultaneously. Different from conventional relays, the signal is not amplified in the IRS but it instead improves the SNR by capitalizing on the array power gain achieved when having a large surface. We will compare the IRS with the use of a conventional half-duplex mMIMO relay, which is also deployed in between a source and destination to improve the propagation conditions; see [24]–[26] and references therein for prior work on mMIMO relays. We refer to [7], [16], [17], [21] for an overview of other prospective use cases of the IRS technology beyond relaying, including how it can be used in conjunction with other technologies.

A fundamental benefit of using large arrays is that the SNR grows with the number of elements N. In mMIMO setups, the SNR is proportional to N if optimal beamforming is applied [11], [12], [24]–[26]. This implies that the transmit power needed to achieve a target SNR value during data transmission reduces as $1/N$, which is a so-called asymptotic power scaling law.[1] In contrast, the SNR grows as N^2 when using an IRS that is optimally configured [15]. Hence, a more aggressive power scaling law can be formulated where the transmit power is reduced as $1/N^2$ [15], [20], [27].

A main limitation of the aforementioned SNR analyses and power scaling laws is that they are derived under an implicit assumption of far-field operation. When considering arrays, the far-field refers to the propagation range at which the direction and channel gain are approximately the same from all elements in the array to the transmitting/receiving antenna. This is different from the Fraunhofer distance, which measures the radiative far-field from a single antenna element and will not be considered in this article. Since the array size grows with N, we will inevitably operate in the array's near-field as $N \to \infty$. The near-field behavior

of the SNR is uncertain since different papers have put forward different hypotheses. For example, several papers have studied the IRS behavior in the far- and near-field assuming that it operates as a specular reflector (also called an "anomalous" mirror) [20], [28], [29] and made parallels to geometrical physics to support this assumption.

The conference version [1] of this article was the first attempt to mathematically derive the near-field behavior in both the mMIMO and IRS setups, but the results are approximate since they relied on the propagation model from [8] that neglects polarization effects. The mismatch between the polarization of an antenna and of the incident wave is approximately the same for all antennas in the far-field, thus one can compute the SNR without polarization and then multiply with a coefficient accounting for the mismatch loss [4, Sec. 7.4]. The situation is more complicated in the near-field (i.e., for large arrays) where the incident wave is arriving from distinctly different angular directions to different elements, thus one must model the polarization on an element-by-element basis to obtain accurate results. Recently, [30]–[32] provided numerical studies and discussions regarding the near-field behavior, but the results are approximate since polarization is neglected in these works. In [30], [32], the effective areas of the elements are also assumed constant in the array, which is not the case in the near-field since the elements are observed from distinctly different angles. This is an additional source of approximation errors. Nevertheless, the experimental results presented in [30] show that there are finite-sized setups where the approximate formulas are matching quite well with measurements. However, we will show later that none of these prior analytical results can be used to characterize the asymptotic limit where $N \to \infty$.

A. CONTRIBUTIONS
We derive a novel closed-form expression for the channel gain when communicating between a single-antenna device and a planar array of arbitrary size, by taking the varying distances to the elements, polarization mismatches, and effective areas into account. We demonstrate that it is necessary to utilize this model to rigorously study the signal propagation in the array's near-field and the asymptotic limits where the array dimensions grow large, because the approximate models in previous work give different results. We use the derived expression to mathematically derive the near-field and far-field behaviors in three key setups: conventional mMIMO, half-duplex mMIMO relaying, and IRS-aided communications. In particular, we explain under which conditions the SNR grows with N in the ways described above, and when we instead need to consider the alternative near-field behaviors that we establish. The analysis shows that the far-field approximation is accurate when the distance to the array is larger than its height/width, which holds in many practical scenarios but will not hold when studying the asymptotic limit when $N \to \infty$. We derive new power scaling laws that are asymptotically accurate. Furthermore, we prove that an

---

1. If one also reduces the transmit power in the channel acquisition phase, the power scaling law changes; we refer to [4], [11], [12] for details.

IRS cannot achieve a higher SNR than any of the mMIMO setups when the array sizes are equal, despite the fact that the SNR in the IRS setup grows as N^2 in the far-field. We derive closed-form expressions for how large an IRS must be to beat conventional mMIMO or half-duplex mMIMO relaying. We provide a geometric interpretation of an SNR-maximizing IRS, which is different from the specular reflector scenario that has been assumed in some prior work. While the main theory is developed for a free-space line-of-sight setup, we also extend the results to consider arbitrary deterministic channel models.

B. OUTLINE

Preliminaries on signal propagation and array gains are provided in Section II. The channel between a single-antenna transmitter and a planar antenna array of arbitrary size is derived in Section III, assuming a free-space line-of-sight scenario. The asymptotic limits are derived and the asymptotic deficiencies of previously used models are exemplified. We define the system models and achievable spectral efficiencies of three different setups in Section IV: conventional mMIMO, half-duplex mMIMO relaying, and IRS-aided communications. The power scaling laws and near/far-field behaviors of these setups are uncovered in Section V, by utilizing the results from Section III. The case with different array sizes is studied in Section VI, to quantify how much larger an IRS must be to match the spectral efficiencies achieved by the other setups. Next, in Section VII, we provide a new geometric interpretation of an IRS that is configured to maximize the SNR. Section VIII discusses the extension to more general channel models. Finally, the main results and conclusions are summarized in Section IX.

C. REPRODUCIBLE RESEARCH

The simulation results can be reproduced using code available at: https://github.com/emilbjornson/near-field-behavior.

D. NOTATION

Boldface lowercase letters, $\mathbf{x}$, denote column vectors and boldface uppercase letters, $\mathbf{X}$, denote matrices. The superscripts ^{T}, *, and ^{H} denote transpose, conjugate, and conjugate transpose, respectively. The $n \times n$ identity matrix is $\mathbf{I}_n$, $\mathrm{mod}(\cdot, \cdot)$ indicates the modulo operation, and $\lfloor \cdot \rfloor$ rounds the argument to the closest smaller integer. The multi-variate circularly symmetric complex Gaussian distribution with covariance matrix $\mathbf{R}$ is denoted $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R})$. We define $\|\mathbf{x}\|$ as the Frobenius norm of vector $\mathbf{x}$.

II. PRELIMINARIES

This article analyzes the wireless propagation when using arrays of different sizes and transmitters/receivers at different distances. We begin by considering the free-space propagation scenario shown in Fig. 1(a), where an ideal isotropic transmit antenna sends a signal to a receive antenna located at distance d. Assume that the receive antenna is lossless,

has an (effective) area A perpendicular to the direction of propagation, and has a polarization perfectly matching that of the transmitted signal. Then, from Friis' formula [33], the received power is

$$P_{\mathrm{rx}} = \frac{A}{4\pi d^2} P_{\mathrm{tx}} \tag{1}$$

where P_{tx} denotes the transmit power and the factor

$$\beta_d = \frac{A}{4\pi d^2} \tag{2}$$

is the free-space channel gain, also known as pathloss. Note that this factor is given by the area A of the receive antenna divided by the total surface area of a sphere with radius d. We use the subscript d in β_d to express that the channel gain is a function of d. Since the received power P_{rx} can never be higher than the transmit power P_{tx} (due to the law of conservation of energy), it is evident that $\beta_d \in [0, 1]$. In most cases, β_d is much smaller than one, as we will now exemplify.

Example 1: If the receive antenna is isotropic, its area is $A = \lambda^2/(4\pi)$ where $\lambda = c/f$ is the wavelength and c is the speed of light. If the transmission has a carrier frequency of $f = 3\,\mathrm{GHz}$, then $\lambda = 0.1\,\mathrm{m}$. For propagation distances $d \in [2.5, 25]\,\mathrm{m}$, the channel gain β_d ranges from $-40\,\mathrm{dB}$ to $-60\,\mathrm{dB}$. If the carrier frequency is increased to $f = 30\,\mathrm{GHz}$, the antenna area becomes 100 times smaller and thus β_d will instead range from $-60\,\mathrm{dB}$ to $-80\,\mathrm{dB}$ for $d \in [2.5, 25]\,\mathrm{m}$.

A way to increase the channel gain in (2) is to make the receive antenna area larger. In particular, we can deploy N antennas of the same kind in an array. If they are deployed on the sphere in Fig. 1(a) and are non-overlapping, and if each antenna has an orientation and polarization that match the locally received signal, the total received power is N times the value in (1):

$$P_{\mathrm{rx}}^{\mathrm{spheric}\text{-}N} = N P_{\mathrm{rx}} = N \beta_d P_{\mathrm{tx}}. \tag{3}$$

This setup is illustrated in Fig. 1(b) and the channel gain is $N\beta_d = \frac{NA}{4\pi d^2}$. Note that the channel gain is proportional to the total antenna area NA, thus one can achieve the same result with many physically small receive antennas or a few large antennas. Clearly, no more than $NA = 4\pi d^2$ non-overlapping receive antennas can be deployed on the sphere in the way shown in Fig. 1(b). In that case, $N\beta_d = 1$ and $P_{\mathrm{rx}}^{\mathrm{spheric}\text{-}N} = P_{\mathrm{tx}}$, so that all the transmitted power is received. Observe that very many antennas are needed to make this happen. Under the assumptions in Example 1, we need 10^4 antennas to cover the entire sphere for $d = 2.5\,\mathrm{m}$ when communicating at $f = 3\,\mathrm{GHz}$, and 10^6 antennas for $d = 25\,\mathrm{m}$. Both values increase by 100 times when communicating at $f = 30\,\mathrm{GHz}$ since the area of an isotropic receive antenna becomes 100 times smaller.
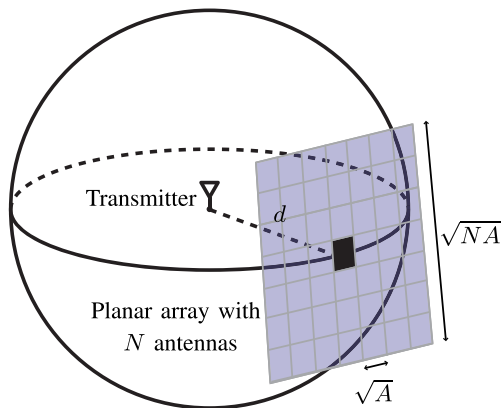
The linear growth with N in (3) is called an *array gain* and is the key motivating factor behind mMIMO communications using antenna arrays with a large (possibly infinite) number of antennas. A common assumption in such systems
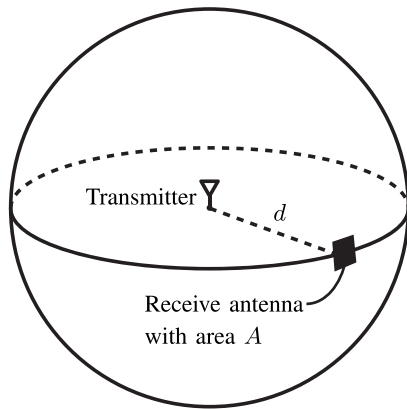
(a) One receive antenna with area A.
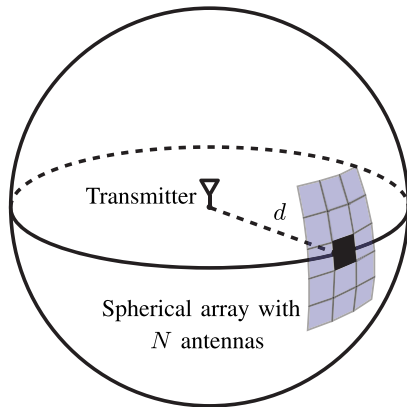
(b) Spherical array with N equal-sized receive antennas.

(c) Planar array with $\sqrt{N} \times \sqrt{N}$ equal-sized receive antennas.

FIGURE 1. Examples of basic antenna scenarios.

is that the linear scaling holds true even in the asymptotic regime where $N \to \infty$ [3], [11], [12], [34]–[37], which has been utilized to define properties such as channel hardening and favorable propagation, as well as studying the fundamental impact of pilot contamination. This assumption is physically incorrect since the total channel gain would be higher than one, thereby invalidating the law of conservation of energy. On the other hand, the analysis above has

shown that a very large number of antennas is needed to receive all the transmitted power. Hence, the linear scaling might hold in practical mMIMO communications, even if thousands of antennas are used and the propagation distance is short. The aim of the next section is to revisit the asymptotic regime with practical planar antenna arrays and prove under what conditions the linear scaling is inaccurate versus approximately correct. These results will be fundamental in Section V when studying the power scaling laws and near-field behaviors of different mMIMO systems and IRS-aided setups.

III. PLANAR ANTENNA ARRAYS

We now turn the attention to the planar array illustrated in Fig. 1(c), particularly because such arrays are commonly used in practical mMIMO deployments [6]. The transmit antenna is at distance d from the center of the array. For notational convenience, we make the following assumption that will be considered in the remainder of this article.

Assumption 1: The planar array consists of N antennas that each has area $A \le (\lambda/4)^2$. The antennas have size $\sqrt{A} \times \sqrt{A}$ and are equally spaced on a $\sqrt{N} \times \sqrt{N}$ grid. The antennas are deployed edge-to-edge, thus the total area of the array is NA.

These assumptions[2] are important when quantifying the channel gain, because the *effective* area of each receive antenna will depend on its physical location and rotation, with respect to the direction of the transmitter. The physical area of each antenna is A but the effective area seen from the transmitter varies. If the receive antenna is fully perpendicular to the direction of propagation, then the effective area also equals A. In any other case, the effective area is smaller than A. The antenna gain in a particular direction is determined by the effective antenna area seen from transmitter and also by the polarization loss caused by having a rotated aperture [38].[3] When the transmitter is in the near-field of the array, three fundamental properties must be taken into account:

1) The distances to the elements vary over the array;
2) The effective antenna areas vary since the elements are seen from different angles;
3) The losses from polarization mismatch vary since the signals are received from different angles.

A. EXACT EXPRESSION FOR THE CHANNEL GAIN
The following lemma extends the prior work in [38] to the case when the transmitter and the receiver are arbitrarily

---

2. Assumption 1 restricts N to be the square of an integer, but the analytical results of this article only require a quadratic planar array with dimension $\sqrt{NA} \times \sqrt{NA}$. For a given array area NA, we can always adapt A to make N be the square of an integer.

3. Only the components of the field vectors that are perpendicular to the boresight of the antenna can be received, irrespective of whether linearly or circularly polarized signals are considered. When deriving the analytical results, we consider linear polarization along the Y direction but since we assume a square array, the results are rotationally invariant. Therefore, another choice of polarization will lead to the same end results, even if the individual elements will contribute differently.

located, thereby providing a general way of computing the channel gains to each of the N elements of a planar array.[4]

Lemma 1: Consider a lossless isotropic antenna located at $\mathbf{p}_t = (x_t, y_t, d)$ that transmits a signal that has polarization in the Y direction when traveling in the Z direction. The receive antenna is located in the XY-plane, is centered at $\mathbf{p}_n = (x_n, y_n, 0)$, and has area $a \times a$. The free-space channel gain is upper bounded by

$$\zeta_{\mathbf{p}_t, \mathbf{p}_n, a} = \frac{1}{4\pi} \sum_{x \in \mathcal{X}_{t,n}} \sum_{y \in \mathcal{Y}_{t,n}} \left(\frac{\frac{xy}{d^2}}{3\left(\frac{y^2}{d^2} + 1\right)\sqrt{\frac{x^2}{d^2} + \frac{y^2}{d^2} + 1}} + \frac{2}{3} \tan^{-1}\left(\frac{\frac{xy}{d^2}}{\sqrt{\frac{x^2}{d^2} + \frac{y^2}{d^2} + 1}} \right) \right)$$

$$(4)$$

where $\mathcal{X}_{t,n} = \{a/2 + x_n - x_t, a/2 - x_n + x_t\}$ and $\mathcal{Y}_{t,n} = \{a/2 + y_n - y_t, a/2 - y_n + y_t\}$.

The upper bound is tight when the antenna area is sufficiently small compared to the wavelength: $a \le \lambda/4$.

Proof: The proof is given in Appendix A in two steps. The impact of the three fundamental properties mentioned above is clearly pointed out in (69). ∎

This lemma provides an upper bound on the channel gain by assuming the received signal's phase-variations are negligible over the antenna area, which is commonly assumed in the literature but is only a tight bound for sub-wavelength-sized antennas, as will be assumed in this article. We will use the general formula from Lemma 1 later in the paper, particularly when analyzing an IRS-aided setup. However, we first notice that a compact expression for the channel gain and, thus, the total received power can be obtained when the transmitter is centered in front of the planar array.

Corollary 1: Under Assumption 1, when the transmitter is centered in front of the planar array, the received power is

$$P_{rx}^{\text{planar-}N} = \alpha_{d,N} P_{tx} \tag{5}$$

where the total channel gain is

$$\alpha_{d,N} = \frac{N\beta_d}{3(N\beta_d\pi + 1)\sqrt{2N\beta_d\pi + 1}} + \frac{2}{3\pi} \tan^{-1}\left(\frac{N\beta_d\pi}{\sqrt{2N\beta_d\pi + 1}} \right) \tag{6}$$

with β_d given in (2).

Proof: This formula follows from Lemma 1 by setting $x_t = y_t = 0$, $x_n = y_n = 0$, and $a = \sqrt{NA}$, in which case $\mathcal{X}_{t,n} = \mathcal{Y}_{t,n} = \{\sqrt{NA}/2, \sqrt{NA}/2\}$. Even if we might have $a > \lambda/4$, the expression of $\alpha_{d,N}$ is a tight upper bound on the true channel gain since the individual antennas satisfy $\sqrt{A} \le \lambda/4$ and we are summing up their received powers. By replacing d with $\sqrt{A/(4\pi\beta_d)}$ and rearranging the terms, we obtain (6) from (4). ∎

4. We disregard the mutual coupling effect in this article, to focus on the asymptotic behaviors with ideal hardware.

The channel gain in (6) is valid for arbitrarily large planar arrays, which is different from the models considered in [30]–[32] that assume equal effective areas of all elements. The new expression supports the case when the transmitter is in the near-field of the array.[5] We will explore the far-field approximation and large-array limit appearing in the near-field.

Remark 1: The exact expression in (6) depends on $N\beta_d$, thus it is the total array area NA that matters and not the individual values of N and A. Hence, the results in this article hold for any frequency band and choice of individual antenna areas, as long as the total area is the same and $A \le (\lambda/4)^2$ so Lemma 1 provides a tight bound for the individual antennas. As the wavelength reduces, the area A shrinks and then more elements are needed to fill the same total array area. In all the simulation figures, we are considering $\lambda = 0.1\,\text{m}$ (3 GHz) when reporting the number of elements N, but the same behaviors appear in any frequency band under the condition that the total area NA of the array is the same.

B. FAR-FIELD APPROXIMATION AND LARGE-ARRAY LIMIT

Suppose the planar array considered in Corollary 1 is in the far-field of the transmitter in the sense that $d \gg \sqrt{NA}$. In this case, $N\beta_d\pi + 1 \approx 1$ and $\sqrt{2N\beta_d\pi + 1} \approx 1$. By using the first-order Taylor approximation $\tan^{-1}(x) \approx x$, which is tight when the argument is close to zero (as is the case when $N\beta_d\pi$ is small), it follows from (5) that

$$P_{rx}^{\text{planar-}N} \approx \left(\frac{N\beta_d}{3} + \frac{2}{3\pi}N\beta_d\pi \right) P_{tx} = N\beta_d P_{tx} \tag{7}$$

which is equal to $P_{rx}^{\text{spheric-}N}$ in (3). Hence, for relatively small planar arrays, the received power is proportional to N. Both terms in (6) contribute to the result, but not equally much.

If N grows large, the far-field approximation is no longer valid and we instead notice that as $N \to \infty$ it holds that

$$\frac{N\beta_d}{3(N\beta_d\pi + 1)\sqrt{2N\beta_d\pi + 1}} \to 0, \tag{8}$$

$$\tan^{-1}\left(\frac{N\beta_d\pi}{\sqrt{2N\beta_d\pi + 1}} \right) \to \frac{\pi}{2}. \tag{9}$$

Hence, the received power in (5) saturates and has the asymptotic limit

$$P_{rx}^{\text{planar-}N} \to \frac{2}{3\pi}\frac{\pi}{2}P_{tx} = \frac{P_{tx}}{3} \quad \text{as } N \to \infty. \tag{10}$$

This value satisfies the law of conservation of energy since only one third of the transmitted power is received. An intuitive explanation for why the limit is finite, although the array is infinitely large, is that each new receive antenna is deployed further away from the transmitter; the effective area (perpendicularly to the direction of propagation) becomes gradually smaller and the polarization loss also increases.

5. Note that we assume throughout this article that $d \gg \lambda$, so the system does not operate in the reactive near-field of the transmit antenna (even if it is in the near-field of the array). In fact, this assumption was made in the proof of the expression in Lemma 1.
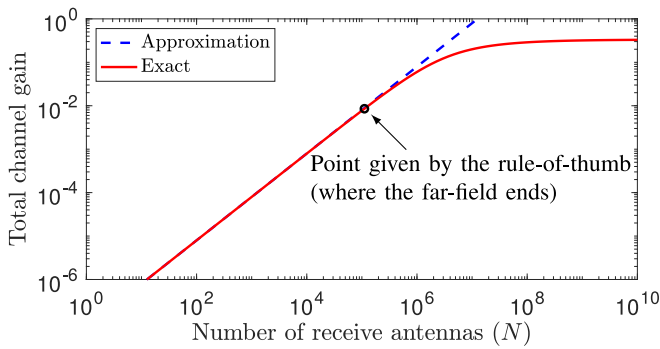
FIGURE 2. The total channel gain $P_{rx}^{planar\text{-}N}/P_{tx}$ with a planar array with $\sqrt{N} \times \sqrt{N}$ equally spaced antennas. Each antenna has area $A = (\lambda/4)^2$, the wavelength is $\lambda = 0.1$ m, and $d = 25$ m.

FIGURE 3. The total channel gain $P_{rx}^{planar\text{-}N}/P_{tx}$ with a planar array with a varying number of antennas. The exact curve ("All three properties") is compared with approximations that are obtained by neglecting some of the near-field propagation properties.

From the above discussion, a natural question arises: *Will the received power grow linearly with N for practical array sizes, so that we can utilize the approximation in* (7), *or do we need to use the exact expression?*

To answer this question, Fig. 2 shows the total channel gain $P_{rx}^{planar\text{-}N}/P_{tx} \in [0, 1]$ as a function of N, using either the exact expression in (5) or the far-field approximation in (7). We consider a setup with $d = 25$ m, $A = (\lambda/4)^2$, and $\lambda = 0.1$ m (corresponding to $f = 3$ GHz). The results of Fig. 2 show that 10^5 antennas are needed before the far-field approximation error is noticeable (above 5%), and 10^8 antennas are needed to approach the upper limit of 1/3.

As a rule-of-thumb, the far-field approximation in (7) is accurate for all N satisfying $NA/9 \leq d^2$ or, equivalently, satisfying

$$\sqrt{NA} \leq 3d. \tag{11}$$

The value $N = 9d^2/A$ that gives equality in this rule-of-thumb is indicated by a circle in Fig. 2. The interpretation is that the far-field approximation is accurate as long as the width/height $\sqrt{NA}$ of the array is smaller than three times the distance d to the transmitter. Hence, if $d = 25$ m, then the approximation can be applied for arrays up to 75×75 m. As the distance d increases or the carrier frequency increases, the maximum number of antennas that satisfies the rule-of-thumb grows quadratically, but the area remains constant. In conclusion, the far-field approximation is usually accurate and might be used to predict scaling behaviors, but the exact expression in Corollary 1 is needed to study the asymptotic limit.

Several recent works have also analyzed the propagation effects in the array's near-field [1], [8], [30]–[32] but using less detailed models. Recall that three near-field properties were listed earlier in this section. The models in [30]–[32] only capture the first property: that the distances to the antennas are different in large arrays. The models in [1], [8] also capture the second property: that the effective antenna areas vary over the array. A main novelty of this article is that we also include the third property: variations in the polarization mismatch over the array. The importance of considering all
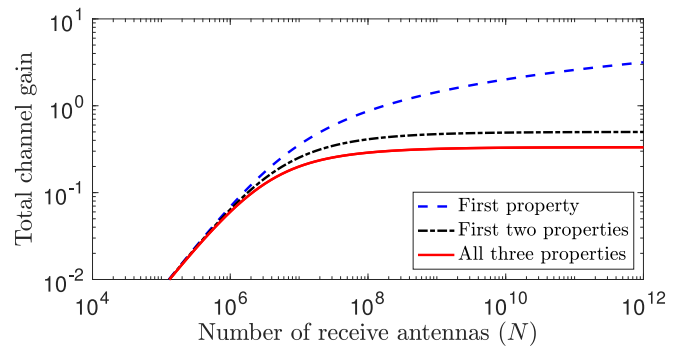
three properties when studying the near-field and asymptotic limits is emphasized in Fig. 3. The figure considers the same setup as in Fig. 2 but focuses on the upper tail where the near-field behavior occurs. If the polarization effects are neglected ("First two properties"), then the channel gain converges to 1/2 as $N \to \infty$, as previously shown in [1], [8]. The general near-field behaviors are correct but not the channel gain values. If also the variations in the effective areas are neglected ("First property"), then the channel gain diverges as $N \to \infty$ and thereby breaks the law of conservation of energy. Hence, the models from [30]–[32] are not recommended to use when studying the asymptotic limits (or the near-field in general). However, all the models are accurate in the far-field.

Remark 2: In this article, we define the far-field as being the propagation range between the array and the single-antenna transmitter where the far-field approximation gives accurate results. When the transmit antenna is isotropic, the border between the near-field and far-field is frequency-independent and proportional to the width/height $\sqrt{NA}$ of the receiving array, as seen from (11). Note that the far-field definition for an array is conceptually different from the Fraunhofer distance, which is a frequency-dependent limit for the radiative far-field from a single antenna element. It describes the range at which one can neglect the reactive electromagnetic phenomena that appear within few wavelengths from an antenna. When considering large arrays in this article, we will be in the far-field of the individual elements but in the near-field of the array.

Remark 3: The propagation models presented in this section are physically accurate, under the given assumptions, and will be used in the remainder of this article. However, this does not mean that the assumptions are applicable in any conceivable practical setup—no model is generally applicable. For example, there can be other antenna gains, other polarization directions, and channels consisting of multiple paths. We are not focusing on these generalizations since we aim to provide an intuitive exposition of the fundamental behaviors. However, a generalization of the results are discussed in Section VIII.

IV. THREE DIFFERENT MIMO SETUPS

Next, we introduce the three different setups that are analyzed and compared in this article, which are all illustrated in Fig. 4. In the conventional mMIMO setup of Fig. 4(a), a single-antenna source transmits a signal that is received by a planar array with N antennas, in the same form as in Fig. 1(c). In the half-duplex mMIMO relay setup shown in Fig. 4(b), the same planar array receives the signal from the source and retransmits it to a single-antenna destination. In the IRS-aided setup in Fig. 4(c), the planar array is replaced by an IRS with N passive elements that operate as a full-duplex relay that "reflects" the incoming signal in a controllable manner. The signal comes from the source and is supposed to reach the destination. The IRS is intelligent in the sense that each of the N reflecting elements can control the individual phase of its diffusely reflected signal.

Line-of-sight (LoS) propagation is considered in all setups. Since the channels are deterministic and thus can be estimated arbitrarily well from pilot signals, perfect channel state information is assumed. Despite simple, the three setups in Fig. 4 are sufficient to develop the fundamental scaling laws and near-field behaviors (see Remark 3) and to compare the setups.

A. CONVENTIONAL UPLINK mMIMO

In the LoS scenario, the deterministic flat-fading channel is represented by the vector $\mathbf{h} = [h_1, \ldots, h_N]^T \in \mathbb{C}^N$, where $h_n = |h_n| e^{-j\phi_n}$ is the channel from the source to the nth receive antenna with $|h_n|^2 \in [0, 1]$ being the channel gain and $\phi_n \in [0, 2\pi)$ an arbitrary phase shift. In the uplink, the received signal $\mathbf{r}_{\text{mMIMO}} \in \mathbb{C}^N$ is

$$\mathbf{r}_{\text{mMIMO}} = \mathbf{h}\sqrt{P_{\text{tx}}}s + \mathbf{n} \tag{12}$$

where P_{tx} is the transmit power, s is the unit-norm information signal, and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ is the independent receiver noise. Under the assumption of perfect channel knowledge, linear receiver processing is optimal [4], [39] and we let $\mathbf{v} \in \mathbb{C}^N$ denote the receive combining vector. It is well-known that the maximum SNR is achieved with maximum ratio (MR) combining, defined as $\mathbf{v} = \mathbf{h}^*/\|\mathbf{h}\|$ [4]. The SE is

$$\log_2(1 + \text{SNR}_{\text{mMIMO}}) \tag{13}$$

with

$$\text{SNR}_{\text{mMIMO}} = \frac{|\mathbf{v}^T\mathbf{h}|^2}{\|\mathbf{v}\|^2} \frac{P_{\text{tx}}}{\sigma^2} = \|\mathbf{h}\|^2 \frac{P_{\text{tx}}}{\sigma^2} = \left(\sum_{n=1}^{N} |h_n|^2\right) \frac{P_{\text{tx}}}{\sigma^2}. \tag{14}$$

B. HALF-DUPLEX mMIMO RELAY

The half-duplex relay transmission takes place over two phases: 1) transmission from the source to the relay; 2) transmission from the relay to the destination. No direct link is present. Among the different relaying protocols (e.g., [40]–[42] among others), we consider the
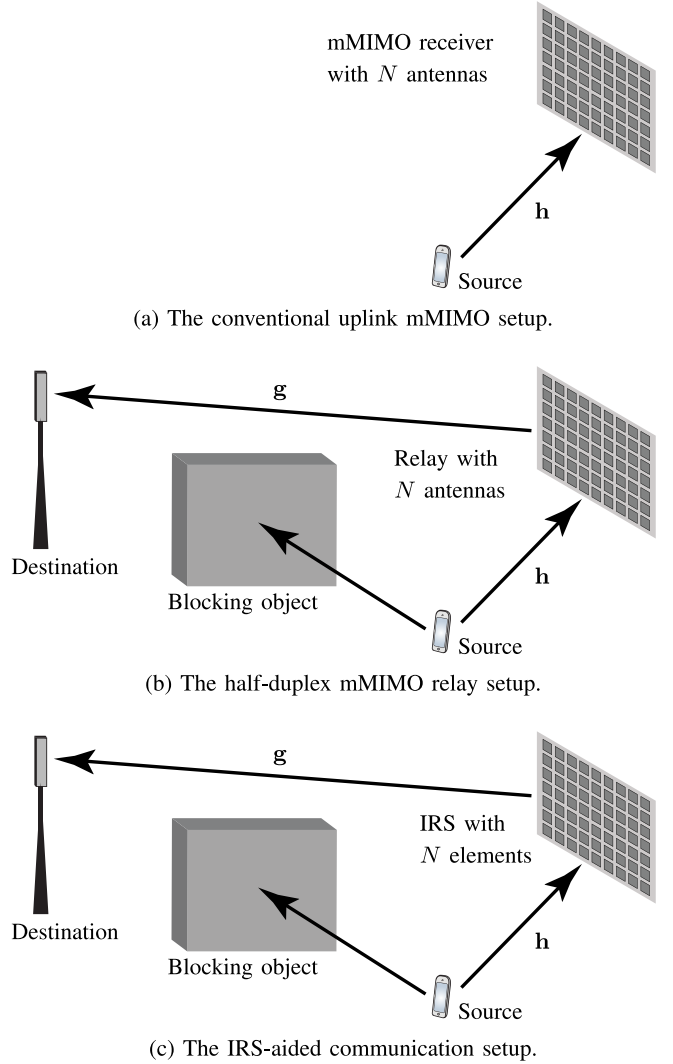

(a) The conventional uplink mMIMO setup.

(b) The half-duplex mMIMO relay setup.

(c) The IRS-aided communication setup.

FIGURE 4. Illustration of the three different MIMO setups compared in this article.

basic repetition-coded decode-and-forward protocol where equal time is allocated to the two phases. The first phase achieves the same SNR as in the mMIMO setup considered above. Therefore, the SE is $\frac{1}{2}\log_2(1 + \text{SNR}_{\text{mMIMO}})$ where $\text{SNR}_{\text{mMIMO}}$ is given in (14) and the pre-log factor represents the fact that each phase is allocated half of the time resources. In the second phase, the relay retransmits the signal s with power P_{relay} using a unit-norm precoding vector $\mathbf{w}$. The LoS channel from the array to the destination is represented by the deterministic vector $\mathbf{g} = [g_1, \ldots, g_N]^T \in \mathbb{C}^N$, where $g_n = |g_n|e^{-j\psi_n}$ represents the channel from the nth antenna to the receiver. The received signal $r_{\text{relay}} \in \mathbb{C}$ at the single-antenna destination is

$$r_{\text{relay}} = \mathbf{g}^T\mathbf{w}\sqrt{P_{\text{relay}}}s + n \tag{15}$$

where $n \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ is the independent receiver noise. It is well-known that the SNR is maximized by MR precoding

with $\mathbf{w} = \mathbf{g}^*/\|\mathbf{g}\|$ [4], which leads to

$$\text{SNR}_{\text{relay}} = \|\mathbf{g}\|^2 \frac{P_{\text{relay}}}{\sigma^2} = \left(\sum_{n=1}^{N} |g_n|^2\right) \frac{P_{\text{relay}}}{\sigma^2}. \quad (16)$$

The SE of the end-to-end mMIMO relay channel is then given by the minimum of the two phases:

$$\text{SE}_{\text{relay}} = \frac{1}{2}\log_2\big(1 + \min(\text{SNR}_{\text{mMIMO}}, \text{SNR}_{\text{relay}})\big). \quad (17)$$

C. IRS-AIDED COMMUNICATION

The IRS-aided communication is also a relaying setup, thus the system model resembles the half-duplex mMIMO relay case with the key differences that each element in the IRS scatterers the incoming signal with a controllable phase-shift but without increasing its power or requiring a separate retransmission phase. The received signal $r_{\text{IRS}} \in \mathbb{C}$ can be modeled as [31], [43]

$$r_{\text{IRS}} = \mathbf{g}^{\mathrm{T}}\boldsymbol{\Theta}\mathbf{h}\sqrt{P_{\text{tx}}}s + n \quad (18)$$

where P_{tx} and s are the same as in the previous setups and $n \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ is the noise at the receiver. The reflection properties are determined by the diagonal matrix

$$\boldsymbol{\Theta} = \text{diag}\left(\mu_1 e^{j\theta_1}, \ldots, \mu_N e^{j\theta_N}\right) \quad (19)$$

where $\mu_1, \ldots, \mu_N \in [0, 1]$ are the amplitude scattering variables (describing the fraction of the incident signal power that is scattered) and $\theta_1, \ldots, \theta_N \in [0, 2\pi)$ are the phase-shift variables (describing the delays of the scattered signals). These parameters can be optimized based on $\mathbf{g}$ and $\mathbf{h}$. With perfect channel knowledge [15], [44], an achievable SE is[6]

$$\log_2(1 + \text{SNR}_{\text{IRS}}) \quad (20)$$

where

$$\begin{aligned}
\text{SNR}_{\text{IRS}} &= \left|\mathbf{g}^{\mathrm{T}}\boldsymbol{\Theta}\mathbf{h}\right|^2 \frac{P_{\text{tx}}}{\sigma^2} \\
&= \left|\sum_{n=1}^{N} \mu_n |h_n||g_n| e^{j(\theta_n - \phi_n - \psi_n)}\right|^2 \frac{P_{\text{tx}}}{\sigma^2} \quad (21)
\end{aligned}$$

is the SNR at the receiver. We will optimize the amplitude and phase-shift variables in the next section.

V. POWER SCALING LAWS AND NEAR-FIELD BEHAVIORS

We will now investigate the asymptotic behaviors of the three setups defined in Section IV. Particularly, the power scaling laws, near-field behaviors, and asymptotic SE limits will be analyzed as N increases. New insights into the fundamental properties will be obtained by utilizing the deterministic propagation model derived in Section III. The following assumption is made for all setups.

6. Note that we are considering relaying operation where the IRS is unaware of the information. A higher SE can be achieved by also encoding information into the matrix $\boldsymbol{\Theta}$ [45].
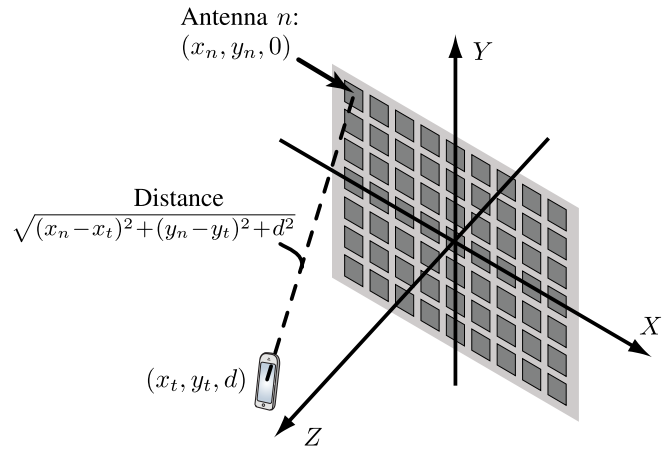
(a) A source at an arbitrary location (x_t, y_t, d) transmits to an planar array located in the XY-plane.
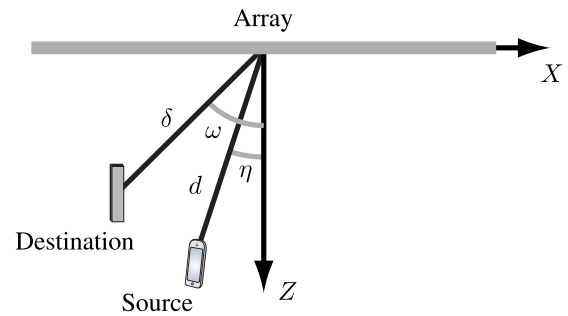
(b) In the analytical parts, the source is at distance d in angle η, while the destination is at distance δ in angle ω.

FIGURE 5. Geometric illustration of the setup defined by Assumptions 1 and 2 that is used for analyzing the SNR behavior in the near- and far-fields.

Assumption 2: The planar array is centered around the origin in the XY-plane, as illustrated in Fig. 5(a). The source is located in the XZ-plane at distance d from the center of the array with angle $\eta \in [-\pi/2, \pi/2]$, as illustrated in Fig. 5(b). It sends a signal that has polarization in the Y direction when traveling in the Z direction.

Under Assumption 2, the source is located at $\mathbf{p}_t = (d\sin(\eta), 0, d\cos(\eta))$ and the nth antenna is centered at $\mathbf{p}_n = (x_n, y_n, 0)$. If we number the antennas from left to right, row by row, according to Fig. 5(a), the coordinates x_n and y_n of the nth receive antenna for $n = 1, \ldots, N$ are

$$x_n = -\frac{(\sqrt{N} - 1)\sqrt{A}}{2} + \sqrt{A}\,\text{mod}(n - 1, \sqrt{N}) \quad (22)$$

$$y_n = \frac{(\sqrt{N} - 1)\sqrt{A}}{2} - \sqrt{A}\left\lfloor\frac{n - 1}{\sqrt{N}}\right\rfloor. \quad (23)$$

A. CONVENTIONAL UPLINK mMIMO

We will now study the mMIMO setup in detail. Following the geometry stated in Assumption 2, we have that $\mathbf{p}_t = (d\sin(\eta), 0, d\cos(\eta))$ and $\mathbf{p}_n = (x_n, y_n, 0)$ where the coordinates x_n and y_n are defined in (22) and (23). By using Lemma 1, the channel $h_n = |h_n|e^{-j\phi_n}$ to the nth receive antenna is obtained as

$$|h_n|^2 = \zeta_{(d\sin(\eta), 0, d\cos(\eta)),(x_n, y_n, 0),\sqrt{A}} \quad (24)$$

with the phase computed based on the propagation delay as

$$
\begin{aligned}
\phi_n &= 2\pi \cdot \mathrm{mod}\left(\frac{\|\mathbf{p}_t - \mathbf{p}_n\|}{\lambda}, 1 \right) \\
&= 2\pi \cdot \mathrm{mod}\left(\frac{\sqrt{(x_n - d\sin(\eta))^2 + y_n^2 + d^2\cos^2(\eta)}}{\lambda}, 1 \right) \\
&= 2\pi \cdot \mathrm{mod}\left(\frac{\sqrt{x_n^2 + y_n^2 + d^2 - 2dx_n\sin(\eta)}}{\lambda}, 1 \right). \quad (25)
\end{aligned}
$$

The following result is then obtained.

Proposition 1: Under Assumptions 1 and 2, in the mMIMO setup, the SNR with MR combining becomes

$$
\mathrm{SNR}_{\mathrm{mMIMO}} = \xi_{d,\eta,N} \frac{P_{\mathrm{tx}}}{\sigma^2} \quad (26)
$$

where the total channel gain $\xi_{d,\eta,N}$ is given by

$$
\begin{aligned}
&\xi_{d,\eta,N} \\
&= \sum_{i=1}^{2} \Bigg(\frac{B + (-1)^i\sqrt{B}\tan(\eta)}{6\pi(B+1)\sqrt{2B + \tan^2(\eta) + 1 + 2(-1)^i\sqrt{B}\tan(\eta)}} \\
&\quad + \frac{1}{3\pi}\tan^{-1}\left(\frac{B + (-1)^i\sqrt{B}\tan(\eta)}{\sqrt{2B + \tan^2(\eta) + 1 + 2(-1)^i\sqrt{B}\tan(\eta)}} \right) \Bigg)
\end{aligned}
$$
$$(27)$$

with $B = N\pi\beta_{d\cos(\eta)} = \frac{NA}{4d^2\cos^2(\eta)}$.

Proof: This result follows from Lemma 1 with $\mathbf{p}_t = (d\sin(\eta), 0, d\cos(\eta))$, $\mathbf{p}_n = (0, 0, 0)$, and $a = \sqrt{NA}$. ∎

We stress that the channel gain in (27) depends only on the total array area NA (see Remark 1), thus the choice of frequency band only affects how many antennas are needed to achieve that area. By using Corollary 1, a more compact expression can be obtained when the transmitter is centered in front of the array (i.e., $\eta = 0$).

Corollary 2: When the transmitter is located in direction $\eta = 0$, the SNR in (26) simplifies to

$$
\mathrm{SNR}_{\mathrm{mMIMO}} = \alpha_{d,N} \frac{P_{\mathrm{tx}}}{\sigma^2} \quad (28)
$$

where the total channel gain $\alpha_{d,N}$ is given in (6).

We will now use the general expression in Proposition 1 for an arbitrary η to study the far-field behavior in the next corollary. Note that $d\cos(\eta)$ is the distance from the transmitter to the plane where the array is deployed and $\sqrt{NA}$ is the width/height of the array.

Corollary 3 (Far-Field Approximation): If the transmitter is in the far-field of the mMIMO receiver, in the sense that $d\cos(\eta) \gg \sqrt{NA}$, then (28) is well approximated as

$$
\mathrm{SNR}_{\mathrm{mMIMO}} \approx \mathrm{SNR}_{\mathrm{mMIMO}}^{\mathrm{ff}} = N\varsigma_{d,\eta} \frac{P_{\mathrm{tx}}}{\sigma^2} \quad (29)
$$

where

$$
\varsigma_{d,\eta} = \beta_{d\cos(\eta)}\cos^3(\eta) \quad (30)
$$

and $\beta_{d\cos(\eta)}$ is given in (2).

Proof: The derivation can be found in Appendix B. ∎

Plugging (2) into (30), we have that (29) reduces to

$$
\begin{aligned}
\mathrm{SNR}_{\mathrm{mMIMO}}^{\mathrm{ff}} &= N\frac{A}{4\pi(d\cos(\eta))^2}\cos^3(\eta)\frac{P_{\mathrm{tx}}}{\sigma^2} \\
&= N\frac{P_{\mathrm{tx}}}{\sigma^2}\overbrace{A\cos(\eta)}^{\text{Effective area}}\underbrace{\frac{1}{4\pi d^2}}_{\text{Free-space propagation}} \quad (31)
\end{aligned}
$$

which is equivalent to [46, eq. (17)]. This shows that, in the far-field, the channel gain per antenna is computed according to Friis' formula with the effective antenna area being $A\cos(\eta)$ [33]. This is a consequence of the fact that, although each antenna has a physical size of $\sqrt{A} \times \sqrt{A}$, its effective size shrinks to $\sqrt{A}\cos(\eta) \times \sqrt{A}$ when observed from the direction of the transmitter.

From Corollary 3, we notice that the far-field SNR in (29) is proportional to N, which is consistent with previous work in the mMIMO literature [4], [11], [12]. Hence, when N increases, the system can either benefit from a linearly increasing SNR or reduce P_{tx} as $1/N$ to keep the SNR constant. The latter is the conventional power scaling law for mMIMO, which first appeared in [11], [12]. However, when computing the asymptotic behavior as $N \to \infty$, these prior works implicitly assumed the transmitter remains in far-field of the array and thus that the SNR goes to infinity as $N \to \infty$ (or the power can be brought down to zero following the scaling law, while the SNR remains strictly non-zero). This is not physically possible. As N increases, the far-field approximation eventually breaks down and the total channel gain saturates in the near-field, as illustrated in Fig. 2. We provide the following novel asymptotic limit and power scaling characterization for the mMIMO receiver.

Corollary 4 (Asymptotic Analysis): As $N \to \infty$ with a constant transmit power P_{tx}, the SNR with MR combining satisfies

$$
\mathrm{SNR}_{\mathrm{mMIMO}} \to \frac{1}{3}\frac{P_{\mathrm{tx}}}{\sigma^2}. \quad (32)
$$

If the transmit power is reduced with N as $P_{\mathrm{tx}} = P/N^\rho$ for some constant $P > 0$ and exponent $\rho > 0$, then as $N \to \infty$

$$
\mathrm{SNR}_{\mathrm{mMIMO}} = \xi_{d,\eta,N}\frac{P}{\sigma^2 N^\rho} \to 0. \quad (33)
$$

Proof: The limit in (32) is computed in the same way as the finite limit in (10). Since $P\xi_{d,\eta,N}$ has a finite limit and $1/N^\rho \to 0$ as $N \to \infty$, the result in (33) follows directly. ∎

This corollary shows that any power scaling for which $P_{\mathrm{tx}} \to 0$ as $N \to \infty$ will asymptotically lead to zero SNR. Hence, the asymptotic motivation behind the power scaling laws in the mMIMO literature [11], [12] is not correct. The scaling laws are, nevertheless, useful in many practical situations. To demonstrate this, Fig. 6 shows $\mathrm{SNR}_{\mathrm{mMIMO}}$ in (28) when we scale down the transmit power as $P_{\mathrm{tx}} = P/N^\rho$ for $\rho \in \{0, 1/2, 1\}$, where $\rho = 0$ corresponds to constant power. We consider a setup where the transmitter location is given by $d = 25\,\mathrm{m}$ and $\eta = 0$, the antenna area is $A = (\lambda/4)^2$, the
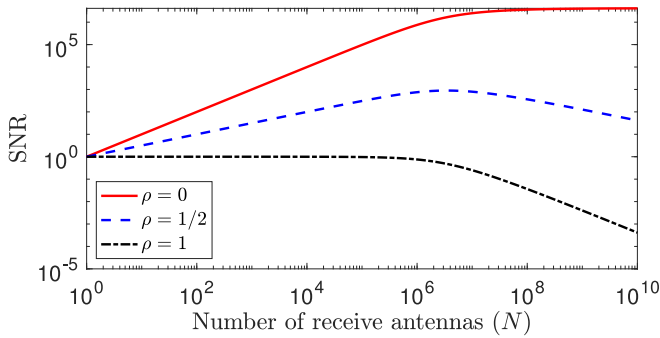
FIGURE 6. The SNR value $\text{SNR}_{\text{mMIMO}}$ in (28) when scaling down the transmit power as $P_{\text{tx}} = P/N^\rho$ for $\rho \in \{0, 1/2, 1\}$. The setup is given by $d = 25\,\text{m}$, $\eta = 0$, $A = (\lambda/4)^2$, $\lambda = 0.1\,\text{m}$, and P/σ^2 is selected to give $\text{SNR}_{\text{mMIMO}} = 0\,\text{dB}$ for $N = 1$.

wavelength is $\lambda = 0.1\,\text{m}$, and the transmit power is selected so that $P\xi_{d,\eta,N}/\sigma^2 = 0\,\text{dB}$ for $N = 1$. We observe that for $\rho = 0$, the far-field behavior, namely, an SNR that grows linearly with N, approximately holds true for any $N \leq 10^6$. This probably includes all cases of practical interest since the array can be up to $25 \times 25\,\text{m}$. If one selects $\rho = 1/2$, the SNR will instead grow as $\sqrt{N}$ for $N \leq 10^6$. Moreover, for $\rho = 1$, the SNR is approximately constant for $N \leq 10^6$. For larger values of N, the SNR goes to zero whenever $\rho > 0$, as proved in Corollary 4.

Since this example considers $\eta = 0$, we know from Corollary 2 that $\xi_{d,0,N} = \alpha_{d,N}$. It is the relation between N and d in $\alpha_{d,N}$ that determines when the far-field behavior breaks down. Since these variables enter into $\alpha_{d,N}$ as $N\beta_d = \frac{NA}{4\pi d^2}$, the far-field behavior appears as long as $N/d^2 \leq 10^6/25^2 = 1600$. Hence, even if we would reduce the propagation distance to $d = 2.5\,\text{m}$, the approximate scaling laws will be accurate for $N \leq 10^4$ or arrays being up to $2.5 \times 2.5\,\text{m}$. In conclusion, the conventional power scaling laws *can be safely applied in many practical scenarios*, but if we truly want to let $N \to \infty$ or study the case where the transmitter is very close to the array, the asymptotically accurate behavior derived in Corollary 4 must be considered.

B. HALF-DUPLEX mMIMO RELAY

We now turn the attention to the half-duplex mMIMO relay setup. We assume the destination is equipped with a lossless isotropic antenna located in the XZ-plane at distance δ from the center of the array with angle $\omega \in [-\pi/2, \pi/2]$, as shown in Fig. 5(b). This means that it is located at $(\delta\sin(\omega), 0, \delta\cos(\omega))$. According to the geometry stated in Assumption 2, the nth transmit antenna at $(x_n, y_n, 0)$, where x_n and y_n are defined in (22) and (23), respectively. We assume each antenna is sufficiently small to radiate the signal isotropically into the half-plane in front of it.[7] From Lemma 1, it follows that the channel $g_n = |g_n|e^{-j\psi_n}$ from

---

7. This property holds for antennas with $a \leq \lambda/4$ for essentially the same reason as the approximation in Lemma 1 being tight in this interval. An illustration of the radiation pattern for $a = \lambda/5$ is found in [43, Fig. 2].

the nth antenna to the destination is given by

$$|g_n|^2 = \zeta_{(\delta\sin(\omega),0,\delta\cos(\omega)),(x_n,y_n,0),\sqrt{A}} \quad (34)$$

where the propagation delay implies that

$$\psi_n = 2\pi \cdot \text{mod}\left(\frac{\sqrt{x_n^2 + y_n^2 + \delta^2 - 2\delta x_n \sin(\omega)}}{\lambda}, 1\right). \quad (35)$$

The following result is then obtained.

Proposition 2: Suppose the destination is located at distance δ in angle $\omega \in [-\pi/2, \pi/2]$ from the center of the mMIMO relay. Under Assumptions 1 and 2, the SNR with MR precoding becomes

$$\text{SNR}_{\text{relay}} = \xi_{\delta,\omega,N}\frac{P_{\text{relay}}}{\sigma^2} \quad (36)$$

where $\xi_{\delta,\omega,N}$ is given in (27) with $B = N\pi\beta_{\delta\cos(\omega)} = \frac{NA}{4\delta^2\cos^2(\omega)}$.

Proof: This result follows from Lemma 1 with $\mathbf{p}_t = (\delta\sin(\omega), 0, \delta\cos(\omega))$, $\mathbf{p}_n = (0, 0, 0)$, and $a = \sqrt{NA}$. ∎

By utilizing the results in Proposition 1 and Proposition 2, the end-to-end SE in (17) can be rewritten as

$$\text{SE}_{\text{relay}} = \frac{1}{2}\log_2\left(1 + \frac{\min(\xi_{d,\eta,N}P_{\text{tx}}, \xi_{\delta,\omega,N}P_{\text{relay}})}{\sigma^2}\right). \quad (37)$$

Note that, when the receiver is centered in front of the array (i.e., $\omega = 0$), we have $\xi_{\delta,0,N} = \alpha_{\delta,N}$.

Just as in the uplink mMIMO setup, the SNR expression takes a simpler approximate form in the far-field, which is now jointly represented by $d\cos(\eta) \gg \sqrt{NA}$ and $\delta\cos(\omega) \gg \sqrt{NA}$. Following a similar approach as in the proof of Corollary 3, we have that

$$\text{SNR}_{\text{relay}} \approx N\varsigma_{\delta,\omega}\frac{P_{\text{relay}}}{\sigma^2} \quad (38)$$

in the far-field, where $\varsigma_{\delta,\omega}$ is defined as in (30). In conjunction with the far-field result in Corollary 3, we obtain the following result in the mMIMO relay setup.

Corollary 5 (Far-Field Approximation): If the source and destination are both in the far-field of the mMIMO relay, in the sense that $d\cos(\eta) \gg \sqrt{NA}$ and $\delta\cos(\omega) \gg \sqrt{NA}$, then (37) is well approximated as

$$\text{SE}_{\text{relay}} \approx \frac{1}{2}\log_2\left(1 + N\frac{\min(P_{\text{tx}}\varsigma_{d,\eta}, P_{\text{relay}}\varsigma_{\delta,\omega})}{\sigma^2}\right). \quad (39)$$

This corollary shows that the end-to-end SNR grows proportionally to N whenever the far-field approximation is applicable. Hence, one can either keep the transmit powers fixed and achieve an SNR that grows proportionally to N, or reduce the transmit powers P_{tx} and P_{relay} as $1/N$ and achieve the same SNR as with $N = 1$. Since the half-duplex relay channel is a time-multiplexed composition of one uplink and one downlink mMIMO channel, the insights from the last subsection still apply: the far-field approximation and the power scaling law hold in most cases of practical interest.

However, in the asymptotic limit as $N \to \infty$, we have that $\xi_{\delta,\omega,N} \to \frac{1}{3}$ which is the same asymptotic limit as in the first phase where it holds that $\xi_{d,\eta,N} \to \frac{1}{3}$. The following corollary shows that the power scaling law breaks down asymptotically.

Corollary 6 (Asymptotic Analysis): As $N \to \infty$ with constant transmit powers P_{tx} and P_{relay}, the SE with the mMIMO relay satisfies

$$\text{SE}_{relay} \to \frac{1}{2} \log_2\left(1 + \frac{1}{3}\frac{\min(P_{tx}, P_{relay})}{\sigma^2}\right). \quad (40)$$

If the transmit powers are reduced with N as $P_{tx} = P_1/N^{\rho_1}$ and $P_{relay} = P_2/N^{\rho_2}$ for some constants $P_1, P_2 > 0$ and exponents $\rho_1, \rho_2 > 0$, then as $N \to \infty$ it follows that

$$\text{SE}_{relay} \to 0. \quad (41)$$

Proof: The proof follows that of Corollary 4 and is therefore omitted. ∎

This scaling behavior is essentially the same as the one illustrated in Fig. 6, thus we postpone the numerical comparison with uplink mMIMO to Section VI. There are plenty of previous works that study mMIMO relays and the related power scaling laws [24]–[26], often in more general setups (e.g., full-duplex or two-way relaying) than those considered in this article. Although the power scaling laws developed in those papers are practically relevant, the non-zero asymptotic limits are incorrect since the channel models that are used are asymptotically inaccurate. Since the total channel gain is upper bounded by one, any power scaling law that leads to zero transmit power as $N \to \infty$ must also have a zero-valued asymptotic SE. Corollary 6 demonstrates this in a simple decode-and-forward relay setup but the result naturally extends to more complicated setups.

C. IRS-AIDED COMMUNICATION

We begin by observing that the SNR is maximized in (21) when all the terms in the summation has the same phase [15], [44]. This is achieved, for example, by selecting $\theta_n = \phi_n + \psi_n$ for $n = 1, \ldots, N$. In this case, all the N terms have a positive contribution to the SNR, which is maximized by setting $\mu_1 = \ldots = \mu_N = 1$ so that all terms take their maximum achievable value.[8] In doing this, (21) becomes

$$\text{SNR}_{IRS} = \frac{P_{tx}}{\sigma^2}\left(\sum_{n=1}^{N} |h_n||g_n|\right)^2. \quad (42)$$

We can compute this expression exactly using (24) and (34), which provides the values of $|h_n|$ and $|g_n|$. We can also obtain

8. Depending on the IRS implementation, the phases and amplitudes might not be independently controllable but coupled [47]. Moreover, there are some implementations that allow for controlling the state of an element over a continuous range, while other implementations have a discrete number of possible states [48], [49]. To characterize the ultimate performance, we consider the ideal case when we can optimize the phases and amplitudes independently with arbitrary precision. A practical IRS might require more elements to deliver the same SNR as described in this article.

the following simple upper bound that does not involve any summations and will be shown to be tight.

Proposition 3: The SNR in (42) with optimal phase-shifts can be upper bounded as

$$\text{SNR}_{IRS} \leq \text{SNR}_{IRS}^{upper} = \left(\sum_{n=1}^{N} |h_n|^2\right)\left(\sum_{n=1}^{N} |g_n|^2\right)\frac{P_{tx}}{\sigma^2}$$
$$= \xi_{d,\eta,N}\xi_{\delta,\omega,N}\frac{P_{tx}}{\sigma^2}$$
$$= \xi_{\delta,\omega,N}\text{SNR}_{mMIMO} \quad (43)$$

with SNR_{mMIMO} given in (26). The equality holds if and only if the vectors $[|h_1|, \ldots, |h_N|]^T$ and $[|g_1|, \ldots, |g_N|]^T$ are parallel.

Proof: The inequality is a direct application of Hölder's inequality, followed by computing the total channel gains of the two links using Proposition 1. ∎

Interestingly, the upper bound in (43) is the product of the SNR in the mMIMO setup and the term $\xi_{\delta,\omega,N}$, which is the total channel gain from the IRS to the destination. Section III described that the value of $\xi_{\delta,\omega,N}$ must be below one (or rather $1/3$) due to the law of conservation of energy. Therefore, Proposition 3 implicitly states that the IRS-aided setup cannot achieve a higher SNR than the corresponding mMIMO setup, if the array sizes are equal. One way to interpret this result is that the IRS acts as an uplink mMIMO receiver that uses the receive combining $\mathbf{v} = \Theta^T\mathbf{g}$, which has a different directivity than the channel $\mathbf{h}$, except when $[|h_1|, \ldots, |h_N|]^T$ and $[|g_1|, \ldots, |g_N|]^T$ are parallel vectors. Moreover, it also incurs an additional SNR loss given by

$$\|\mathbf{v}\|^2 = \|\Theta^T\mathbf{g}\|^2 = \|\mathbf{g}\|^2 = \xi_{\delta,\omega,N} < 1. \quad (44)$$

Similar conclusions hold when the IRS is compared to the half-duplex mMIMO relay. To see this, assume for simplicity that the transmit power is the same in the two phases (i.e., $P_{relay} = P_{tx}$). In this case, we may equivalently rewrite (43) as

$$\text{SNR}_{IRS}^{upper} = \xi_{d,\eta,N}\text{SNR}_{relay} \quad (45)$$

which can never be higher than SNR_{relay}, based on the same arguments as above. Since the end-to-end SNR of the mMIMO relay channel is the minimum of the SNRs in the two phases, i.e., $\min(\text{SNR}_{relay}, \text{SNR}_{mMIMO})$, and both are higher than SNR_{IRS}^{upper}, we can conclude that the IRS can never achieve a higher SNR than the corresponding mMIMO relay setup with a matching array size and transmit power. However, the half-duplex relay suffers from the $1/2$ pre-log factor in (17), which can potentially make the IRS more spectrally efficient, even if the SNR is lower. To investigate this further, assume that $\text{SNR}_{relay} > \text{SNR}_{mMIMO}$ so that (17) becomes

$$\text{SE}_{relay} = \frac{1}{2}\log_2(1 + \text{SNR}_{mMIMO}). \quad (46)$$

From (43), the SE with the IRS is upper bounded by $\log_2(1 + \xi_{\delta,\omega,N}\text{SNR}_{\text{mMIMO}})$, which is higher than (46) when

$$\xi_{\delta,\omega,N} > \frac{\sqrt{1 + \text{SNR}_{\text{mMIMO}}} - 1}{\text{SNR}_{\text{mMIMO}}}. \tag{47}$$

This condition will be satisfied if $\text{SNR}_{\text{mMIMO}}$ is sufficiently large. Hence, there are high-SNR cases when the IRS-aided setup outperforms the mMIMO relay. This observation is in line with previous results in [19], [44].

We will now study the power scaling law. Recall from (42) that the SNR is proportional to the square of a sum with N terms. Intuitively, the SNR may then grow quadratically with N. That behavior can in fact be observed in the far-field.

Corollary 7 (Far-Field Approximation): If both the source and destination are in the far-field of the IRS, in the sense that $d\cos(\eta) \gg \sqrt{NA}$ and $\delta\cos(\omega) \gg \sqrt{NA}$, the SNR in (42) can be approximated as

$$\text{SNR}_{\text{IRS}} \approx \text{SNR}_{\text{IRS}}^{\text{ff}} = N^2 \varsigma_{d,\eta}\varsigma_{\delta,\omega}\frac{P_{\text{tx}}}{\sigma^2}. \tag{48}$$

Proof: This result is proved in the same way as Corollary 3 and Corollary 5. ∎

The quadratic scaling with N in (39) has been recognized in several recent works [15], [20], [27], but without explaining that it only holds when the far-field approximation applies. Moreover, those papers noticed that SNR growth is faster than the linear scaling with N observed for mMIMO receiver in (29) and for the mMIMO relay in (39). Although that implies that an IRS benefits more from increasing the array size, it does not mean that it will achieve a higher SNR when N is large. Indeed, we already know from Proposition 3 and the subsequent discussion that this cannot happen in neither the far-field nor the near-field.

An instructive way of interpreting the N^2 scaling can be found by factorizing the far-field SNR in (48) into two factors:

$$\text{SNR}_{\text{IRS}}^{\text{ff}} = \overbrace{N\varsigma_{\delta,\omega}}^{\leq 1,\ \text{Fraction of reflected power reaching destination}} \times \underbrace{N\varsigma_{d,\eta}\frac{P_{\text{tx}}}{\sigma^2}}_{=\text{SNR}_{\text{mMIMO}}^{\text{ff}}}. \tag{49}$$

The first factor contains one N-term and describes the fraction of power received at the IRS that also reaches the destination. Since this term is fundamentally upper bounded by one, this N-term describes a drawback rather than a benefit of using the IRS-type of relay (it is the fraction of power that is *not* lost). The second factor in (49) equals the far-field mMIMO SNR in (29) and its N-term represents the array power gain that is achieved when having a large array.

To demonstrate these properties, Fig. 7 shows the total channel gains obtained by the mMIMO receiver and the IRS-aided setup for a varying number of antennas/elements N. We consider the setup in Fig. 5(b) with $d = 25\,\text{m}$, $\eta = \pi/6$, $\delta = 2.5\,\text{m}$, and $\omega = -\pi/6$. Each element in the array has area $A = (\lambda/4)^2$ with $\lambda = 0.1\,\text{m}$. We stress that in the IRS-aided setup, the destination is in the vicinity of the IRS. The
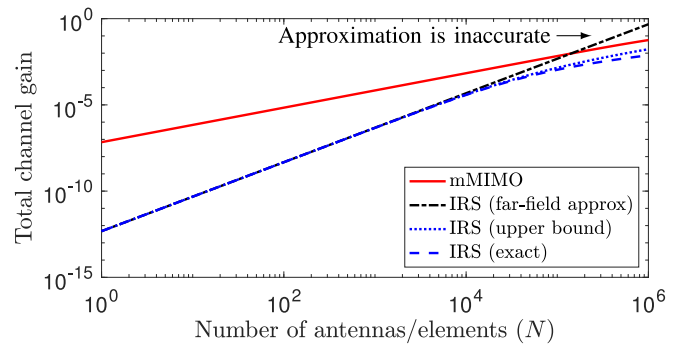

FIGURE 7. The total channel gain obtained with the mMIMO receiver and with the IRS-aided setup, for different number of antennas/elements N. The setup in Fig. 5(b) is considered with $d = 25\,\text{m}$, $\eta = \pi/6$, $\delta = 2.5\,\text{m}$, $\omega = -\pi/6$, $A = (\lambda/4)^2$, and $\lambda = 0.1\,\text{m}$.

figure shows that the total channel gain grows as N^2 with the IRS and as N with the mMIMO receiver, which is consistent with the respective far-field approximations derived in Corollaries 3 and 7. Nevertheless, mMIMO provides a much larger channel gain for most values of N, which is consistent with Proposition 3. The advantage remains asymptotically. The reason is that each element of the IRS acts as an isotropic scatterer, thus the IRS is a full-duplex relay that forwards the signal to the destination without amplifying it [44]. Even if the destination is close to the IRS, the far-field approximation in (48) is accurate until the IRS has roughly 10^4 elements. The upper bound in (43) follows the exact curve closely, even for $N > 10^4$, which is why we called it a tight bound.

Remark 4: The upper bound in Proposition 3 contains the product of the total channel gain $\xi_{d,\eta,N}$ between the source and IRS and the total channel gain $\xi_{\delta,\omega,N}$ between the IRS and the destination. This is the same structure as for the far-field SNR in (48), which has been analyzed in a series of previous works (e.g., [30]–[32], [43]). However, the near-field behavior has not be analytically studied with the same rigor. The IRS was approximated as a specular reflector (i.e., an ideal mirror) in [20], [28]–[30] and the channel gain in the near-field can then be made proportional to $1/(d + \delta)^2$ [29]. This expression is different from the upper bound in Proposition 3, where there are no terms that depend on both d and δ. The following conclusion can be made: If one can operate an IRS to get a channel gain of the kind in [20], [28]–[30], the SNR is likely not maximized by doing so. We return to this matter in Section VII.

We conclude this section by using the upper bound in Proposition 3 to study the asymptotic behavior of an IRS, particularly in the near-field.

Corollary 8 (Asymptotic Analysis): As $N \to \infty$ with constant transmit power P_{tx}, the SNR in the IRS setup is asymptotically upper bounded since

$$\text{SNR}_{\text{IRS}}^{\text{upper}} \to \frac{1}{9}\frac{P_{\text{tx}}}{\sigma^2}. \tag{50}$$

If the transmit power is reduced with N as $P_{\text{tx}} = P/N^\rho$ for some constant $P > 0$ and exponent $\rho > 0$, then as $N \to \infty$

it follows that

$$\text{SNR}_{\text{IRS}} = \left(\sum_{n=1}^{N} |h_n||g_n|\right)^2 \frac{P}{N^\rho \sigma^2} \to 0. \qquad (51)$$

Proof: The upper bound follows from the fact that $\xi_{d,\eta,N}, \xi_{\delta,\omega,N} \to 1/3$ as $N \to \infty$, which was also utilized in Corollary 4. Since the channel gain is upper bounded, the SNR goes to zero if P_{tx} goes asymptotically to zero. ∎

This corollary shows, once again, that the asymptotic SE limit of any conventional power scaling law is zero. Nevertheless, we can expect the SNR in the IRS-aided setup to grow as N^2 for most practical array sizes. In those cases, it is also possible to reduce the transmit power as $1/N^2$ and keep the SNR constant. In agreement with Proposition 3, Corollary 8 also shows that an IRS-aided setup can never reach the same SNR as the mMIMO receiver for any common value of N. The difference remains as $N \to \infty$ since the limits are different.

VI. HOW LARGE IRS IS NEEDED TO ACHIEVE THE SAME SNR?

When looking for suitable use cases for the IRS technology, one needs to ask the question: How *large* must the IRS be to achieve the same performance as with an active mMIMO receiver or a regenerative half-duplex mMIMO relay? To answer this question, we now let N_{mMIMO}, N_{relay}, and N_{IRS}, denote the number of elements of the mMIMO receiver, the mMIMO relay, and the IRS, respectively. We can then determine how many elements are needed in the IRS to achieve the same or higher SE than with the competing technologies.

Corollary 9: When operating in the far-field, the IRS case provides higher SE than the mMIMO receiver if

$$N_{\text{IRS}} \geq \sqrt{\frac{N_{\text{mMIMO}}}{\varsigma_{\delta,\omega}}}. \qquad (52)$$

Similarly, the IRS case provides higher SE than the half-duplex mMIMO relay if

$$N_{\text{IRS}} \geq \frac{\sigma^2}{P_{\text{tx}}\varsigma_{d,\eta}\varsigma_{\delta,\omega}} \sqrt{1 + N_{\text{relay}} \frac{\min\left(P_{\text{tx}}\varsigma_{d,\eta}, P_{\text{relay}}\varsigma_{\delta,\omega}\right)}{\sigma^2}}. \qquad (53)$$

Proof: This follows from comparing the expressions in Corollaries 3, 5, and 7. ∎

By inserting values into the expressions in Corollary 9, Fig. 8 shows how many antennas are needed to achieve a particular SE in each of the three setups. The same simulation parameters as in Fig. 7 are considered with $P_{\text{tx}} = P_{\text{relay}}$. The first observation is that the IRS needs more than 100 elements before it provides an SE that is clearly above zero. After that, the number of elements grows more gracefully with the SE than for the half-duplex relay and mMIMO setups, since the SNR grows as N^2 for the IRS. However, it is only for SEs greater than 4.4 bit/s/Hz that
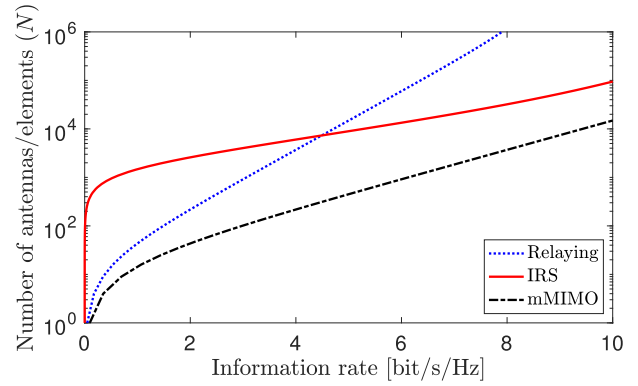

FIGURE 8. The number of elements/antennas needed to achieve a given information rate in the different setups. The same simulation parameters as in Fig. 7 are considered with $P_{\text{tx}} = P_{\text{relay}}$.

$N_{\text{IRS}} < N_{\text{relay}}$ in this example. The IRS must always be larger than the mMIMO array to deliver the same SE. For example, $N_{\text{mMIMO}} = 100$ delivers 3 bit/s/Hz, while $N_{\text{IRS}} \approx 4000$ is needed to achieve the same SE. Since this example considers a 3 GHz carrier frequency, this corresponds to a mMIMO receiver that is 0.25×0.25 m and an IRS that is 1.6×1.6 m, thus it is practically possible deploy such an IRS even in an indoor environment. The difference in physical size reduces asymptotically but will not vanish, as proved in the previous section. If one would consider a different carrier frequency, the physical array dimensions remain the same but the number of elements required to build the array change.

Remark 5: Even if the IRS must be physically larger than the mMIMO counterparts to achieve a given SNR, it might still be practically preferable since the surface can be thin, integrated into existing construction elements, and, hopefully, cheap and energy-efficient. While the first generation of mMIMO technology is commercially available, the IRS technology is in its infancy which makes it impossible to quantify its cost and energy consumption [23]. It has been possible to build metasurfaces for many years but the IRS operation also requires real-time channel estimation and reconfigurability. This is an active research area [27], [50], [51] that has not converged to a mature solution yet and it is potentially the implementation of these functionalities that will dominate the hardware cost and energy consumption [23].

VII. GEOMETRIC INTERPRETATION OF OPTIMIZED IRS

Some recent works model an IRS as a specular reflector or an "anomalous mirror" (i.e., a mirror with an unusual reflection angle) [20], [28]–[30]. This basically means that the IRS reflects the incoming signal towards the destination as a flat and perfectly rotated plane mirror would do. Under these conditions, the total channel gain of the IRS setup would converge to

$$\varsigma_{d,\delta}^{\text{mirror}} = \left(\frac{\lambda}{4\pi(d+\delta)}\right)^2 \qquad (54)$$

as the array size grows large and the near-field is considered. This asymptotic formula can be motivated by geometrical
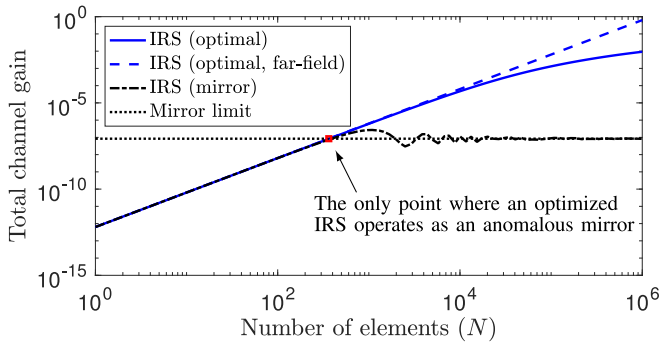
FIGURE 9. The total channel gain obtained with the IRS-aided communication using either optimal phases or mirror-mimicking phases, for different number of elements N. The setup in Fig. 5(b) is considered with $d = 25\,\text{m}$, $\eta = 0$, $\delta = 2.5\,\text{m}$, $\omega = 0$, $A = (\lambda/4)^2$, and $\lambda = 0.1\,\text{m}$.

physics if one considers an equivalent setup where the destination is behind the mirror and the total propagation distance is $d + \delta$. Interestingly, the limit in (54) differs from the asymptotic upper bound derived in Corollary 8. More precisely, the two distances d and δ appear in (54) in a joint factor $(d + \delta)^2$ and not within separate multiplicative factors as in Proposition 3. This reveals that an IRS that is optimized to maximize the SNR does not operate as a plane mirror. If a plane wave is impinging on a plane mirror, its specular reflection is also a plane wave. In contrast, if a plane wave is impinging on an IRS, each element will scatter a piece of the wave with a particular phase-shift. By optimizing the phase-shifts so that the N scattered waves add constructively at the destination, the IRS effectively operates as a concave mirror that focuses the incoming wave at the point of the destination. The phase-shift optimization finds the SNR-maximizing curvature of the concave mirror and the IRS synthesizes such a mirror without actually changing its physical shape.

Fig. 9 demonstrates this in a setup where both the source and destination are centered in front of the array: $d = 25\,\text{m}$, $\eta = 0$, $\delta = 2.5\,\text{m}$, $\omega = 0$ following the notation in Fig. 5(b). In this case, a mirror-mimicking IRS has $\theta_n = 0$ and $\mu_n = 1$ for all n and the corresponding total channel gain can be computed using (21) as $|\sum_{n=1}^{N} |h_n||g_n|e^{-j(\phi_n + \psi_n)}|^2$. Fig. 9 also reports the total channel gain $(\sum_{n=1}^{N} |h_n||g_n|)^2$ of an optimized IRS with $\theta_n = \phi_n + \psi_n$. Each element has area $A = (\lambda/4)^2$ with $\lambda = 0.1\,\text{m}$. There is no noticeable difference in Fig. 9 for small IRSs because when the source and destination are in the far-field, focusing the incoming plane wave on a far-away point is approximately the same as mimicking a plane mirror that reflects the signal towards a point infinitely far away in the same angular direction. However, at around $N = 360$, the channel gain of the mirror-mimicking IRS starts to converge to (54), while the channel gain of the optimized IRS continues to increase. At $N = 10^4$, the optimized IRS has a 500 times better channel gain than the mirror limit in (54). We conclude that the SNR achieved by an *optimized IRS* can generally not be described using the mirror limit; particularly not in the near-field since the

dashed far-field approximation is accurate far beyond the point where optimized SNR surpasses the mirror limit. This conclusion is consistent with the results in [31], which were derived by neglecting polarization effects. However, one can certainly use the mirror analogy to identify the approximately optimal phase-shifts when operating in the far-field [43].

By setting the far-field approximation in (48), for an optimized IRS, equal to the mirror limit in (54), we obtain that

$$NA = \left(\frac{1}{d} + \frac{1}{\delta}\right)^{-1} \lambda \qquad (55)$$

is the largest array area that a mirror-mimicking IRS can make use of in this example.[9] This point is indicated by a square in Fig. 9. If the IRS is larger, the remaining area is essentially wasted on scattering signals in other directions. The same phenomenon appears when a person is looking into a large plane mirror and only sees his/her reflection in a small part of it. Hence, if one uses (54) as a proxy for the channel gain of an optimized IRS (e.g., as done in [20], [28]), then the results only hold when the IRS has exactly the area in (55) and the source/destination are centered in front of it. If we change δ or λ, the curves in Fig. 9 will be shifted in different directions, but the quantitative conclusions remain the same. Note that a larger area NA is obtained in (55) as the wavelength is increased, thus a much larger area is needed to mimic a mirror in radio spectrum than in visible light. In summary, the correct geometric interpretation of an optimized IRS is that it synthesizes the scattering off an optimally shaped concave mirror that can focus the incoming wave onto the point of destination.

A. RECONFIGURABILITY UNDER MOBILITY
The SNR-maximizing configuration focuses the reflected signal at the location of the destination, while the mirror-mimicking configuration forms a beam in the angular direction of the destination. One reason to consider the latter configuration is that only the angle must be known, thus the IRS must not be reconfigured if the destination moves along a trajectory where the angle is constant [21]. However, even if the optimal focusing requires a continuous reconfiguration to be withheld under user mobility, one can also focus the signal at a point in the vicinity of the destination and keep this IRS configuration fixed as the destination moves.

An example of this is provided in Fig. 10 for a setup where the source is located at $d = 25\,\text{m}$ from the IRS and the distance δ to the destination is varied. The figure shows how the total channel gain varies for $\delta \in [1, 100]\,\text{m}$ when using an IRS with $N = 10^4$ elements. The setup in Fig. 5(b) is considered with $d = 25\,\text{m}$, $\eta = 0$, $\omega = 0$, $A = (\lambda/4)^2$, and $\lambda = 0.1\,\text{m}$. The upper curve represents the optimal SNR-maximizing configuration which requires reconfiguration of the IRS as the destination moves (i.e., a different Θ for

---

9. The expression is generalized in [31, eq. (35)] to cases where the source and destination are located in different directions.
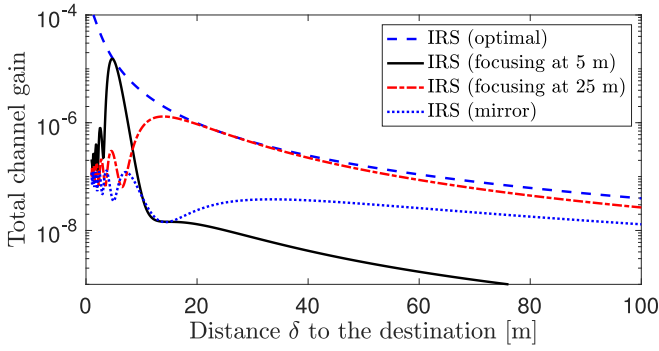
FIGURE 10. The total channel gain when the destination is at different distances δ from the IRS. The SNR-maximizing configuration (utilizing the value of δ) is compared with the mirror-mimicking case (only utilizing the angle ω) and two cases where the signal is focused at a fixed point.

every value of δ). There is a large gap to the dotted curve that represents the mirror-mimicking case that approximates specular reflection and beamforms the signal towards a point infinitely far away in the right angular direction. There are also two curves where the reflected signal is focused at a point that is either given by $\delta = 5\,\mathrm{m}$ or $\delta = 25\,\mathrm{m}$. These curves intersect with the optimal curve in the respective points but are otherwise below that curve, but the practical benefit is that the IRS is not reconfigured as the destination moves. We notice that focusing at a nearby point (5 m) leads to an array gain that is only obtained when the destination is close to the IRS, while the mirror-mimicking configuration outperforms it at larger distances. When the focus point is at an intermediate distance (25 m), the achievable channel gain is larger than or approximately the same as in the mirror-mimicking case at all the considered distances and, particularly, preferable when the destination is at distances above 10 m. The reason is that a specular reflector focuses the reflected signal at a point infinitely far away, thus it is only at very large distances (compared to the size of the IRS) that the mirror-approximation coincides with the SNR-maximizing configuration. Whenever there is side-information regarding the distance interval of where the destination might be, it is better to focus the signal at some point in that interval and then keep the configuration fixed even under mobility.

VIII. EXTENSION TO GENERAL PROPAGATION SETUPS
The analytical results have been derived in the free-space line-of-sight scenario, where the near-field behavior and asymptotic limits can be rigorously derived. Since the near-field behavior occurs when the propagation distance is comparable to the width/height of the array, it will mostly appear over short distances where the line-of-sight path exists. However, the channel can also contain additional scattered paths and the line-of-sight path can be partially blocked. Each such path can be modeled in the way described in Section III, making the channel vector a summation of deterministic multi-path components arriving from different angles with different amplitude and phase. Any such setup

can be described by a pair of deterministic channel vectors **h** and **g** that satisfy the law of conservation of energy: $\|\mathbf{h}\|^2 \leq 1$ and $\|\mathbf{g}\|^2 \leq 1$. For any such setup, it follows from (13) that the SE of the uplink mMIMO is

$$\mathrm{SE}_{\mathrm{mMIMO}} = \log_2\left(1 + \|\mathbf{h}\|^2 \frac{P_{\mathrm{tx}}}{\sigma^2}\right). \tag{56}$$

From (17), the SE of the half-duplex mMIMO relay is

$$\mathrm{SE}_{\mathrm{relay}} = \frac{1}{2}\log_2\left(1 + \min\left(\|\mathbf{h}\|^2\frac{P_{\mathrm{tx}}}{\sigma^2}, \|\mathbf{g}\|^2\frac{P_{\mathrm{relay}}}{\sigma^2}\right)\right). \tag{57}$$

Furthermore, the SE of the IRS-aided setup is

$$\mathrm{SE}_{\mathrm{IRS}} = \log_2\left(1 + \frac{P_{\mathrm{tx}}}{\sigma^2}\left(\sum_{n=1}^{N}|h_n||g_n|\right)^2\right)$$

$$\leq \log_2\left(1 + \|\mathbf{h}\|^2\|\mathbf{g}\|^2\frac{P_{\mathrm{tx}}}{\sigma^2}\right) \tag{58}$$

where the upper bound follows from Proposition 3. The SE will be strictly lower than that in (58) if the IRS has a limited ability to control the phase-shift variables, but it seems that only a few discrete levels are sufficient to keep the SNR loss below 1 dB [49]. In any case, (58) remains a rigorous upper bound and the asymptotic limits are theoretically achievable.

If the arrays are equal-sized in all three setups, then the uplink mMIMO setup achieves the highest SE since $\|\mathbf{h}\|^2\|\mathbf{g}\|^2 \leq \|\mathbf{h}\|^2$ for any practical channel setup. Whether the half-duplex mMIMO relay or the IRS-based relay achieves the highest SE depends on the channel model and transmit powers. Based on the previous results, we can expect the half-duplex relay to perform better when the arrays are small, while the IRS is preferable when the arrays are sufficiently large. The way to understand this is that the smaller pre-log factor of the half-duplex relay is particularly detrimental when the SNR is high, because then its higher SNR cannot compensate for it. If the IRS is physically larger than the mMIMO array, it can achieve the same or higher SNRs. No general relationship can be obtained. When it comes to the asymptotic SE limits, the mMIMO and IRS-aided setups will give convergence to the same limit if $\|\mathbf{g}\|^2 \to 1$ as $N \to \infty$. However, this would require that the receiver captures all of the power that is scattered by the IRS, which might not occur in practice.

IX. CONCLUSION
The limit of a large number of antennas has been studied in the multi-antenna literature for decades and is a core motivation behind the mMIMO technology. In this article, we have noticed that previous asymptotic analyses have used channel models that are only accurate in the far-field, while the asymptotic limit can only be approached when operating in the near-field. Hence, the asymptotic SE behaviors and asymptotic power scaling laws in the existing literature can potentially be misleading. To determine when the asymptotic behaviors break down, we have derived a physically accurate

channel gain expression for planar arrays, taking both polarization and near-field conditions (such as varying effective antenna areas) into account. We have used this model to revisit the power scaling laws and asymptotic limits in three MIMO setups: conventional mMIMO, half-duplex mMIMO relays, and IRS-aided communications.

The main observations are as follows. The total channel gains in the two mMIMO setups grow as N in the far-field, where N is the number of antennas/elements, while it grows as N^2 in the IRS setup. Numerical results showed that these behaviors are accurate even when the arrays have many thousands of elements/antennas, thus the classical scaling results are accurate in most practical deployments. However, the growth rate eventually tapers off when entering the near-field, and the channel gain converges to 1/3 as $N \to \infty$ in the mMIMO setups, and is upper bounded by 1/9 in the IRS setup. The near-field behavior begins when the width/height of the array is comparable (or larger) than the distance to the transmitter/receiver. A consequence is that any power scaling law that lets the transmit power go asymptotically to zero will also lead to zero asymptotic SE.

The IRS will provably always achieve a lower SNR than the two mMIMO setups for any common value of N, despite the faster growth rate observed in the far-field. The reason is that one of the N-terms in the SNR accounts for the fraction of power that is lost in the IRS's reflection, thus it represents a drawback rather than a benefit. However, if the IRS has a larger array size than in the mMIMO setups, it can achieve a higher SNR. This qualitative conclusion is previously known, but we have substantiated it by deriving exact expressions for when the breaking point occurs and shown that it appears in practically relevant cases. Needless to say, the cost per element is lower with an IRS than in mMIMO, thus future work needs to consider if the total cost of the IRS technology will also be smaller.

By using the analytical expressions, we have proved that the SNR of an *optimized* IRS contains the product of the channel gains from the source to the IRS and from the IRS to the destination, in both the near-field and the far-field. Previous works have interpreted the IRS as being an anomalous plane mirror (specular reflector) that can control the angular direction of the "reflected" signal, but we stress that an optimized IRS synthesizes the scattering of a concave mirror that can also focus the signal on a point in the near-field. The optimal concave mirror is approximately plane when it is physically small and/or when the point is far away. In these cases, the optimal SNR does not match the "sum-of-distances" expression in (54) for an infinitely large plane mirror. In the near-field, the IRS can achieve SNRs that are orders-of-magnitude above the plane mirror limit.

The asymptotic analysis of this article relies on a deterministic channel model, which is valid only for the considered propagation scenarios with fixed locations of the transmitter, receiver and arrays. Unlike deterministic models, stochastic approaches are independent of a particular propagation environment and allow to model random reflections and

scattering, and the channel fading they give rise to. Although there is no apparent reason to question the asymptotic findings of this article under stochastic propagation conditions, it must be clear that the classical stochastic channel models cannot be used for the analysis of near-field behaviors since they do not capture the essential near-field propagation properties, such as the three key properties listed in Section III. The research on this subject is generally open and nontrivial. In fact, the classical stochastic channel models do not well-reflect the physical properties of large arrays with a massive number of antennas in a compact space, not even in the far-field. A recent attempt to address this deficiency is [52], where a spatially-stationary model for the small-scale fading in the far-field of non-isotropic random scattering environments is developed on the basis of a Fourier plane-wave spectral representation.

APPENDIX A
PROOF OF LEMMA 1

The proof proceeds in two steps. In the first step, we compute the channel gain of the nth antenna element using electromagnetic arguments (extending the work by [38]) and show numerically how it can be tightly upper bounded when its area $a \times a$ is small (compared to the wavelength λ). In the second step, the upper bound on the channel gain is computed in closed form.

A. CHANNEL GAIN COMPUTATION AND UPPER BOUND
The electric field $\mathbf{E}(\mathbf{p}_t, \mathbf{p}_n) \in \mathbb{C}^3$ generated in a point $\mathbf{r} = [r_x, r_y, 0]$ from a *point source* (isotropic antenna) located in $\mathbf{p}_t = [x_t, y_t, d]$ is [38, eq. (4)]

$$\mathbf{E}(\mathbf{p}_t, \mathbf{r}) = \mathbf{G}(\mathbf{r} - \mathbf{p}_t)\mathbf{J}(\mathbf{p}_t) \tag{59}$$

where $\mathbf{J}(\mathbf{p}_t) = J_x(\mathbf{p}_t)\hat{\mathbf{u}}_x + J_y(\mathbf{p}_t)\hat{\mathbf{u}}_y + J_z(\mathbf{p}_t)\hat{\mathbf{u}}_z$ (with $\hat{\mathbf{u}}_x, \hat{\mathbf{u}}_y, \hat{\mathbf{u}}_z$ representing the unit vectors in the x, y, z directions) and $\mathbf{G}(\mathbf{r} - \mathbf{p}_t) \in \mathbb{C}^{3\times3}$ is the Green function which, under the condition that $\|\mathbf{r} - \mathbf{p}_t\| \geq \lambda$, is well-approximated as [38, eq. (3)]

$$\mathbf{G}(\mathbf{r}) = -j\eta \frac{e^{-j\frac{2\pi}{\lambda}\|\mathbf{r}\|}}{2\lambda\|\mathbf{r}\|} \left(\mathbf{I}_3 - \hat{\mathbf{r}}\hat{\mathbf{r}}^{\mathrm{H}}\right) \tag{60}$$

with $\hat{\mathbf{r}} = \frac{\mathbf{r}}{\|\mathbf{r}\|}$. This approximation is tight when the transmitter is beyond the reactive near-field of the receive antenna.

If only the Y direction of $\mathbf{J}(\mathbf{p}_t)$ is excited at the point source, then we have that $\mathbf{J}(\mathbf{p}_t) = J_y(\mathbf{p}_t)\hat{\mathbf{u}}_y$. The electric field reduces to

$$\mathbf{E}(\mathbf{p}_t, \mathbf{r}) = \mathbf{G}_y(\mathbf{r} - \mathbf{p}_t)J_y(\mathbf{p}_t) \tag{61}$$

where $\mathbf{G}_y(\mathbf{r} - \mathbf{p}_t) = \mathbf{G}(\mathbf{r} - \mathbf{p}_t)\hat{\mathbf{u}}_y$ is the second column of the Green function in (60). The complex-valued channel from a point source located in $\mathbf{p}_t$ to the receive point $\mathbf{r}$ located in the XY-plane is

$$h(\mathbf{p}_t, \mathbf{r}) = |h(\mathbf{r} - \mathbf{p}_t)|e^{-j\frac{2\pi}{\lambda}\|\mathbf{r} - \mathbf{p}_t\|} \tag{62}$$

where

$$|h(\mathbf{r} - \mathbf{p}_t)|^2 = \overbrace{\frac{4}{\eta^2} \|\mathbf{G}_y(\mathbf{r} - \mathbf{p}_t)\|^2}^{\text{Power gain}} \overbrace{\frac{(\mathbf{r} - \mathbf{p}_t)^{\mathsf{T}} \hat{\mathbf{u}}_z}{\|\mathbf{r} - \mathbf{p}_t\|}}^{\text{Projection on the Z direction}}$$

$$= \frac{1}{4\pi} \frac{d\big((r_x - x_t)^2 + d^2\big)}{\big((r_x - x_t)^2 + (r_y - y_t)^2 + d^2\big)^{5/2}} \quad (63)$$

denotes the channel gain along the Z direction (perpendicular to the array) where $\frac{\mathbf{r} - \mathbf{p}_t}{\|\mathbf{r} - \mathbf{p}_t\|}$ denotes the pointing direction of the electric field. If the nth antenna has the area $a \times a$, the effective channel is

$$h_n(\mathbf{p}_t) = \frac{1}{a} \int_{x_n - a/2}^{x_n + a/2} \int_{y_n - a/2}^{y_n + a/2} h(\mathbf{p}_t, \mathbf{r}) \partial r_x \partial r_y. \quad (64)$$

The total channel gain is thus

$$|h_n(\mathbf{p}_t)|^2 = \left| \int_{x_n - a/2}^{x_n + a/2} \int_{y_n - a/2}^{y_n + a/2} h(\mathbf{p}_t, \mathbf{r}) \partial r_x \partial r_y \right|^2$$

$$\le \int_{x_n - a/2}^{x_n + a/2} \int_{y_n - a/2}^{y_n + a/2} |h(\mathbf{p}_t, \mathbf{r})|^2 \partial r_x \partial r_y = \zeta_{\mathbf{p}_t, \mathbf{p}_n} \quad (65)$$

where the inequality follows from the Cauchy-Schwarz inequality and $|h(\mathbf{p}_t, \mathbf{r})|^2 = |h(\mathbf{r} - \mathbf{p}_t)|^2$ in (63). Fig. 11 numerically evaluates the normalized channel gain $|h_n(\mathbf{p}_t)|^2 / \zeta_{\mathbf{p}_t, \mathbf{p}_n}$ as a function of a/λ when $\mathbf{p}_n = (x_n, 0, 0)$ with $x_n = 0, 5$ and 10 m and $\mathbf{p}_t = (0, 0, d)$ with $d = 10$ m. The carrier frequency is $f = 3$ GHz. The results of Fig. 11 show that $\zeta_{\mathbf{p}_t, \mathbf{p}_n}$ is a tight upper bound of $|h_n(\mathbf{p}_t)|^2$ for $a \le \lambda/4$ since the relative error in the channel gain is below 1 dB. An antenna element $a \le \lambda/10$ is needed to approach 0 dB. Motivated by this, we assume $a \le \lambda/4$ and replace $|h_n(\mathbf{p}_t)|^2$ with $\zeta_{\mathbf{p}_t, \mathbf{p}_n}$, which can be computed in closed form as shown next.

B. CLOSED-FORM EXPRESSION OF CHANNEL GAIN BOUND

We will make use of the following primitive functions:

$$\int \frac{\partial x}{(x^2 + a)^{3/2}} = \frac{x}{a\sqrt{x^2 + a}} + C \quad (66)$$

$$\int \frac{\partial x}{(x^2 + a)^{5/2}} = \frac{x}{3a(x^2 + a)^{3/2}} + \frac{2x}{3a^2 \sqrt{x^2 + a}} + C \quad (67)$$

$$\int \frac{\partial x}{(x^2 + a)\sqrt{x^2 + a + b}}$$
$$= \frac{1}{\sqrt{ab}} \tan^{-1} \left(\frac{\sqrt{b}x}{\sqrt{a}\sqrt{x^2 + a + b}} \right) + C \quad (68)$$

where a, b are arbitrary scalars and C is an arbitrary constant.

From (63), it follows that $\zeta_{\mathbf{p}_t, \mathbf{p}_n}$ in (65) requires to solve the following integral:

$$\zeta_{\mathbf{p}_t, \mathbf{p}_n}$$
$$= \frac{1}{4\pi} \int_{x_n - a/2}^{x_n + a/2} \int_{y_n - a/2}^{y_n + a/2} \frac{d\big((r_x - x_t)^2 + d^2\big) \partial r_x \partial r_y}{\big((r_x - x_t)^2 + (r_y - y_t)^2 + d^2\big)^{5/2}}$$
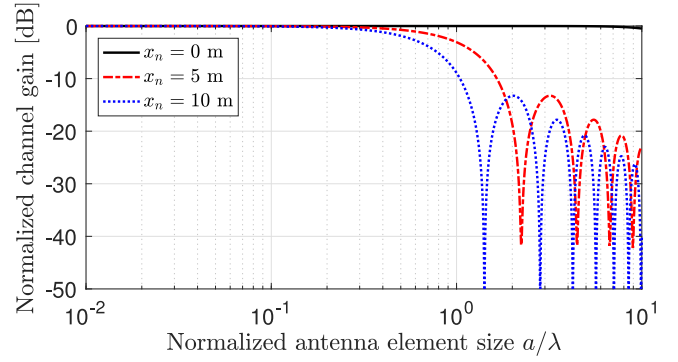

FIGURE 11. Normalized channel gain $|h_n(\mathbf{p}_t)|^2 / \zeta_{\mathbf{p}_t, \mathbf{p}_n}$ in dB as a function of a/λ when $\mathbf{p}_n = (x_n, 0, 0)$ with $x_n = 0, 5$ and 10 m and $\mathbf{p}_t = (0, 0, d)$ with $d = 10$ m.

$$= \int_{x_n - a/2}^{x_n + a/2} \int_{y_n - a/2}^{y_n + a/2} \underbrace{\frac{d}{\sqrt{(r_x - x_t)^2 + (r_y - y_t)^2 + d^2}}}_{\text{Reduction in effective area from directivity}}$$

$$\times \underbrace{\frac{(r_x - x_t)^2 + d^2}{(r_x - x_t)^2 + (r_y - y_t)^2 + d^2}}_{\text{Polarization loss factor}}$$

$$\times \underbrace{\frac{\partial r_x \partial r_y}{4\pi \big((r_x - x_t)^2 + (r_y - y_t)^2 + d^2\big)}}_{\text{Free-space pathloss}} \quad (69)$$

where r_x, r_y are integration variables representing the location of the receive antenna. The contributions of the three fundamental properties when operating in the near-field of the array (i.e., the distance to the elements, the effective antenna areas, the loss from polarization) are clearly explicated. Next, we make the change of variables $\chi = r_x - x_t$ and $\upsilon = r_y - y_t$, so that (69) becomes

$$\frac{1}{4\pi} \int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \int_{y_n - a/2 - y_t}^{y_n + a/2 - y_t} \frac{d(\chi^2 + d^2) \partial \upsilon \partial \chi}{(\chi^2 + \upsilon^2 + d^2)^{5/2}}$$

$$= \frac{1}{4\pi} \int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \left[\frac{\upsilon d}{3(\chi^2 + \upsilon^2 + d^2)^{3/2}} \right]_{y_n - a/2 - y_t}^{y_n + a/2 - y_t} \partial \chi$$

$$+ \frac{1}{4\pi} \int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \left[\frac{2\upsilon d}{3(\chi^2 + d^2)\sqrt{\chi^2 + \upsilon^2 + d^2}} \right]_{y_n - a/2 - y_t}^{y_n + a/2 - y_t} \partial \chi$$

$$\quad (70)$$

$$= \frac{1}{4\pi} \left(\sum_{y \in \mathcal{Y}_{t,n}} \int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \frac{yd}{3(\chi^2 + y^2 + d^2)^{3/2}} \partial \chi \right.$$

$$\left. + \sum_{y \in \mathcal{Y}_{t,n}} \int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \frac{2yd}{3(\chi^2 + d^2)\sqrt{\chi^2 + y^2 + d^2}} \partial \chi \right)$$

$$\quad (71)$$

by utilizing (67). The first integral in (71) can now be computed using (66) as

$$\int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \frac{yd}{3(\chi^2 + y^2 + d^2)^{3/2}} \partial \chi$$

$$= \sum_{x \in \mathcal{X}_{t,n}} \frac{xyd}{(y^2 + d^2)\sqrt{x^2 + y^2 + d^2}}. \qquad (72)$$

Moreover, the second integral in (71) can be computed using (68) as

$$\int_{x_n - a/2 - x_t}^{x_n + a/2 - x_t} \frac{2yd}{3(\chi^2 + d^2)\sqrt{\chi^2 + y^2 + d^2}} \partial \chi$$

$$= \sum_{x \in \mathcal{X}_{t,n}} \frac{2}{3} \tan^{-1}\left(\frac{xy}{d\sqrt{x^2 + y^2 + d^2}} \right). \qquad (73)$$

Substituting (72) into (70) and (73) into (71) yield the final result in (4), after dividing the numerators and denominators by d.

APPENDIX B
PROOF OF COROLLARY 3

When $d\cos(\eta) \gg \sqrt{NA}$, it follows that $B + 1 \approx 1$ and $2B + 1 \approx 1$. We can then utilize that $\tan^{-1}(x) \approx x$ for $x \approx 0$ to approximate (27) as

$$\xi_{d,\eta,N} \approx \sum_{i=1}^{2} \frac{B + (-1)^i \sqrt{B} \tan(\eta)}{2\pi \sqrt{\tan^2(\eta) + 1 + 2(-1)^i \sqrt{B} \tan(\eta)}}. \qquad (74)$$

Furthermore, we can utilize that $\sqrt{1+x} \approx 1 + x/2$ for $x \approx 0$ to approximate the denominator of (74) and obtain

$$\xi_{d,\eta,N} \approx \sum_{i=1}^{2} \frac{B + (-1)^i \sqrt{B}\tan(\eta)}{2\pi\sqrt{1 + \tan^2(\eta)}\left(1 + \frac{(-1)^i \sqrt{B}\tan(\eta)}{1 + \tan^2(\eta)}\right)}$$

$$= \frac{2B - \frac{2B\tan^2(\eta)}{1 + \tan^2(\eta)}}{2\pi\sqrt{1 + \tan^2(\eta)}\left(1 + \frac{\sqrt{B}\tan(\eta)}{1 + \tan^2(\eta)}\right)\left(1 - \frac{\sqrt{B}\tan(\eta)}{1 + \tan^2(\eta)}\right)}$$

$$\approx \frac{B}{\pi(1 + \tan^2(\eta))^{3/2}} = N \underbrace{\beta_{d\cos(\eta)} \cos^3(\eta)}_{=\zeta_{d,\eta}} \qquad (75)$$

where we first simplified the expression by writing the two fractions as a single fraction and then utilized that $1 - \frac{(-1)^i \sqrt{B}\tan(\eta)}{1 + \tan^2(\eta)} \approx 1$ and finally that $1 + \tan^2(\eta) = 1/\cos^2(\eta)$.

REFERENCES

[1] E. Björnson and L. Sanguinetti, "Demystifying the power scaling law of intelligent reflecting surfaces and metasurfaces," in *Proc. IEEE Int. Workshop Comput. Adv. Multi Sens. Adapt. Process.(CAMSAP)*, Le Gosier, Guadeloupe, 2019, pp. 549–553.

[2] S. Parkvall, E. Dahlman, A. Furuskär, and M. Frenne, "NR: The new 5G radio access technology," *IEEE Commun. Stand. Mag.*, vol. 1, no. 4, pp. 24–30, Dec. 2017.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.

[5] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

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

[7] N. Rajatheva *et al.*, "White paper on broadband connectivity in 6G," 6G Res. Visions, Oulu, Finland, White Paper, 2020.

[8] 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.

[9] A. Amiri, M. Angjelichinoski, E. de Carvalho, and R. W. Heath, "Extremely large aperture massive MIMO: Low complexity receiver architectures," in *Proc. IEEE Global Commun. Conf. Workshops (GLOBECOM Workshops)*, Abu Dhabi, UAE, 2018, pp. 1–6.

[10] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, "Spatially-stationary model for holographic MIMO small-scale fading," 2019. [Online]. Available: arXiv:1911.04853.

[11] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[12] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[13] L. Liang *et al.*, "Anomalous terahertz reflection and scattering by flexible and conformal coding metamaterials," *Adv. Opt. Mater.*, vol. 3, no. 10, pp. 1374–1380, 2015.

[14] A. M. H. Wong and G. V. Eleftheriades, "Perfect anomalous reflection with a bipartite huygens' metasurface," *Phys. Rev. X*, vol. 8, Feb. 2018, Art. no. 011036.

[15] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[16] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," 2020. [Online]. Available: arXiv:2007.02759.

[17] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[18] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 129, May 2019.

[19] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

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

[21] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead," 2020. [Online]. Available: https://arxiv.org/abs/2004.09352.

[22] M. Dohler and Y. Li, *Cooperative Communications: Hardware, Channel and PHY.* New York, NY, USA: Wiley, 2010.

[23] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three myths and two critical questions," 2020. [Online]. Available: https://arxiv.org/abs/2006.03377.

[24] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.

[25] X. Yang, X. Liang, X. Zhang, S. Jin, and T. Song, "Power scaling laws for massive MIMO relay systems with linear transceivers," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Orlando, FL, USA, Dec. 2015, pp. 38–42.

[26] C. Kong, C. Zhong, M. Matthaiou, E. Björnson, and Z. Zhang, "Multipair two-way half-duplex DF relaying with massive arrays and imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3269–3283, May 2018.

[27] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[28] I. Yildirim, A. Uyrus, and E. Basar, "Modeling and analysis of reconfigurable intelligent surfaces for indoor and outdoor applications in future wireless networks," 2019. [Online]. Available: https://arxiv.org/abs/1912.07350.

[29] M. D. Renzo, F. H. Danufane, X. Xi, J. de Rosny, and S. Tretyakov, "Analytical modeling of the path-loss for reconfigurable intelligent surfaces—Anomalous mirror or scatterer?" 2020. [Online]. Available: https://arxiv.org/abs/2001.10862.

[30] W. Tang *et al.*, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," 2019. [Online]. Available: https://arxiv.org/abs/1911.05326.

[31] S. W. Ellingson, "Path loss in reconfigurable intelligent surface-enabled channels," 2019. [Online]. Available: https://arxiv.org/abs/1912.06759.

[32] J. Garcia, A. Sibille, and M. Kamoun, "Reconfigurable intelligent surfaces: Bridging the gap between scattering and reflection," 2019. [Online]. Available: https://arxiv.org/abs/1912.05344.

[33] H. T. Friis, "A note on a simple transmission formula," *Proc. IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[34] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[35] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[36] R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[37] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[38] D. Dardari, "Communicating with large intelligent surfaces: Fundamental limits and models," 2019. [Online]. Available: https://arxiv.org/abs/1912.01719.

[39] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[40] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[41] G. Farhadi and N. C. Beaulieu, "On the ergodic capacity of multi-hop wireless relaying systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2286–2291, May 2009.

[42] M. N. Khormuji and E. G. Larsson, "Cooperative transmission based on decode-and-forward relaying with partial repetition coding," *IEEE Trans. Commun.*, vol. 8, no. 4, pp. 1716–1725, Apr. 2009.

[43] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 581–585, May 2020.

[44] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

[45] R. Karasik, O. Simeone, M. D. Renzo, and S. S. Shitz, "Beyond Max-SNR: Joint encoding for reconfigurable intelligent surfaces," 2019. [Online]. Available: https://arxiv.org/abs/1911.09443.

[46] E. Björnson and L. Sanguinetti, "Utility-based precoding optimization framework for large intelligent surfaces," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 863–867.

[47] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," 2020. [Online]. Available: https://arxiv.org/abs/1907.06002.

[48] O. Tsilipakos *et al.*, "Toward intelligent metasurfaces: The progress from globally tunable metasurfaces to software-defined metasurfaces with an embedded network of controllers," *Adv. Opt. Mater.*, vol. 8, Sep. 2020, Art. no. 2000783.

[49] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[50] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," 2019. [Online]. Available: arXiv:1904.10136.

[51] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[52] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, "Spatially-stationary model for holographic MIMO small-scale fading," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1964–1979, Sep. 2020.

EMIL BJÖRNSON (Senior Member, IEEE) received the M.S. degree in engineering mathematics from Lund University, Sweden, in 2007, and the Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Sweden, in 2011.

From 2012 to 2014, he held a joint post-doctoral position with the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France, and the KTH Royal Institute of Technology. He joined Linköping University, Sweden, in 2014, where he is currently an Associate Professor and a Docent with the Division of Communication Systems. In September 2020, he became a Visiting Full Professor with the KTH Royal Institute of Technology. He has authored the textbooks *Optimal Resource Allocation in Coordinated Multi-Cell Systems* (2013) and *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017). He is dedicated to reproducible research and has made a large amount of simulation code publicly available. He has performed MIMO research for over ten years and has filed more than twenty MIMO related patent applications. He performs research on MIMO communications, radio resource allocation, machine learning for communications, and energy efficiency.

Prof. Björnson has received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA, the 2015 Ingvar Carlsson Award, the 2016 Best Ph.D. Award from EURASIP, the 2018 IEEE Marconi Prize Paper Award in *Wireless Communications*, the 2019 EURASIP Early Career Award, the 2019 IEEE Communications Society Fred W. Ellersick Prize, and the 2019 IEEE Signal Processing Magazine Best Column Award. His coauthored papers that received Best Paper Awards at the conferences, including WCSP 2009, the IEEE CAMSAP 2011, the IEEE WCNC 2014, the IEEE ICC 2015, WCSP 2017, and the IEEE SAM 2014. Since 2017, he has been on the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2016.

LUCA SANGUINETTI (Senior Member, IEEE) received the Laurea Telecommunications Engineer degree (*cum laude*) and the Ph.D. degree in information engineering from the University of Pisa, Italy, in 2002 and 2005, respectively.

In 2004, he was a visiting Ph.D. student with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. From June 2007 to June 2008, he was a Postdoctoral Associate with the Department of Electrical Engineering, Princeton. From July 2013 to October 2017, he was with the Large Systems and Networks Group (LANEAS), CentraleSupélec, France. He is currently an Associate Professor with the Dipartimento di Ingegneria dell'Informazione, University of Pisa. He has coauthored the textbook *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017). His expertise and general interests span the areas of communications and signal processing.

Dr. Sanguinetti received the Marconi Prize Paper Award in *Wireless Communications* in 2018, and coauthored a paper that received the Young Best Paper Award from the ComSoc/VTS Italy Section. He was the co-recipient of two best conference paper awards: IEEE WCNC 2013 and IEEE WCNC 2014. He was the recipient of the FP7 Marie Curie IEF 2013 Dense deployments for green cellular networks. He served as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE SIGNAL PROCESSING LETTERS, and as Lead Guest Editor of IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS Special Issue on "Game Theory for Networks" and as an Associate Editor for IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS (series on Green Communications and Networking). He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and is a member of the Executive Editorial Committee of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.