Intelligent Reflecting Surface vs. Decode-and-Forward: How Large Surfaces Are Needed to Beat Relaying?

Emil Björnson, Senior Member, IEEE, Özgecan Özdogan, Student Member, IEEE, Erik G. Larsson, Fellow, IEEE

Abstract—The rate and energy efficiency of wireless channels can be improved by deploying software-controlled metasurfaces to reflect signals from the source to the destination, especially when the direct path is weak. While previous works mainly optimized the reflections, this letter compares the new technology with classic decode-and-forward (DF) relaying. The main observation is that very high rates and/or large metasurfaces are needed to outperform DF relaying, both in terms of minimizing the total transmit power and maximizing the energy efficiency, which also includes the dissipation in the transceiver hardware.

Index Terms—Intelligent reflecting surface, DF relaying.

I. INTRODUCTION

A reflectarray is a surface that "reflects" an impinging plane wave in the shape of a beam [1]. Different from parabolic reflectors, whose physical curvature and direction determine the beamforming, a reflectarray is flat and consists of an array of discrete elements that each scatter and phase-shift the impinging waves differently [2]. The phase-shift pattern among the elements determine in which direction the reflected beam is formed. While the surface can be large, the individual elements are typically sub-wavelength sized [3], [4]. Reflectarrays with real-time reconfigurable properties have recently gained interest in mobile communications, under names such as intelligent reflecting surface (IRS) [5], [6] and software-controlled metasurfaces [7], [8]. The main idea is to support the transmission from a source to a destination by adapting the propagation environment; that is, to configure the IRS to beamform its received signal towards the destination.

This is the same use case as for half-duplex relays [10], with the key difference that a relay actively processes the received signal before retransmitting an amplified signal, while an IRS passively reflects the signal without amplification but with beamforming. The relay achieves a higher signal-to-noise ratio (SNR) at the cost of a pre-log penalty due to the two-hop transmission. A comparison with an ideal amplify-and-forward (AF) relay was made in [11], showing large energy efficiency gains by using IRSs. However, decode-and-forward (DF) relaying is known to outperform conventional relaying. To this end, we optimize both technologies by computing the optimal transmit powers and the optimal number of elements in an IRS.

II. SYSTEM MODEL

We consider communication from a single-antenna source to a single-antenna destination. The deterministic flat-fading channel is denoted by \( h_{sd} \in \mathbb{C} \). The received signal at the destination is

\[
y = h_{sd}\sqrt{p}s + n,
\]

where \( p \) is the transmit power, \( s \) is the unit-power information signal, and \( n \sim \mathcal{N}(0, \sigma^2) \) is the receiver noise. For notational convenience, the antenna gains are included in the channels. The capacity of this single-input single-output (SISO) channel is

\[
R_{SISO} = \log_2 \left( 1 + \frac{p|h_{sd}|^2}{\sigma^2} \right).
\]

The capacity can potentially be increased by involving additional equipment in the communication. In this paper, we consider two such setups: An IRS that is configured to "reflect" the signal towards the destination or a relay that operates in DF mode. The respective achievable rates (also known as spectral efficiencies) are derived below and then optimized analytically to enable a fair comparison. However, we stress that the choice of system model is biased in favor of the IRS; in particular, the assumption of deterministic flat-fading channels is ideal for an IRS since it is less capable of handling channel estimation and frequency-selective fading than relays.

A. IRS-supported Transmission

In this setup, we have an IRS with \( N \) discrete elements, as illustrated in Fig. 1(a). The deterministic channel from the source to the IRS is denoted by \( h_{sr} \in \mathbb{C}^N \), where \( [h_{sr}]_n \) denotes the \( n \)-th component. The channel between the IRS and the destination is denoted by \( h_{rd} \in \mathbb{C}^N \). Each element has a smaller size than the wavelength, thus it scatters the incoming signal with approximately constant gain in all directions of interest [13]. The IRS's properties are therefore fully represented by the diagonal matrix

\[
\Theta = \alpha \text{diag}(e^{j\theta_1}, \ldots, e^{j\theta_N}),
\]

where \( \alpha \in (0, 1] \) is the fixed amplitude reflection coefficient and \( \theta_1, \ldots, \theta_N \) are the phase-shift variables that can be optimized by the IRS. Following the system model derived in [13] (which was earlier used in e.g. [5], [11], [14]), the received signal at the destination is

\[
y = (h_{sd} + h_{sr}^{\ast} \Theta h_{rd})\sqrt{p}s + n,
\]

©2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Please refer to 10.1109/LWC.2024.3453787 for additional clarification.

The authors are with the Department of Electrical Engineering (ISY), Linköping University, SE-58183 Linköping, Sweden [emil.bjornson,ozgecan.ozdogan,erik.g.larsson]@liu.se.
Lemma 1. The channel capacity of the IRS-supported network is
\[
R_{\text{IRS}}(N) = \max_{\theta_1, \ldots, \theta_N} \log_2 \left( 1 + \frac{p|h_{sd}|^2 \Theta_h h_{rd}|^2}{\sigma^2} \right)
\]
\[
= \log_2 \left( 1 + \frac{p}{\sigma^2} |h_{sd}|^2 + \alpha \sum_{n=1}^{N} |[h_{sr} n | h_{rd} n]|^2 \right).
\]

Proof: For any given \( \Theta \), the rate expression in (5) is achieved from the capacity of an additive white Gaussian noise channel. Notice that \( h_{sr} \Theta_h h_{rd} = \alpha \sum_{n=1}^{N} e^{j \theta_n} |h_{sr} n | h_{rd} n| \). The maximum rate, which is the capacity, is achieved when the phase-shifts are selected as \( \theta_n = \arg(h_{sd}) - \arg([h_{sr} n | h_{rd} n]|) \) to give every term in the sum the same phase as \( h_{sd} \).

B. Relay-supported Transmission

In this alternative setup, we make use of a half-duplex relay that is deployed at the same location as the IRS. This setup is illustrated in Fig. 1(b). We consider the classic repetition-coded DF relaying protocol where the transmission is divided into two equal-sized phases. In the first phase, the source transmits and the received signal at the destination is
\[
y_{1d} = h_{sd} \sqrt{p_1} s + n_{1d}, \tag{7}
\]
where \( p_1 \) is the transmit power, \( s \) is the unit-power information signal, and \( n_{1d} \sim \mathcal{N}(0, \sigma^2) \) is the receiver noise. In the same phase, the received signal at the relay is
\[
y_{1r} = h_{sr} \sqrt{p_1} s + n_{1r}, \tag{8}
\]
where \( h_{sr} \in \mathbb{C} \) denotes the channel between the source and relay, while \( n_{1r} \sim \mathcal{N}(0, \sigma^2) \) is the receiver noise. The DF relay uses \( y_{1r} \) to decode the information and then encodes it again for transmission in the second phase. Note that the relay can be compact; an antenna, transceiver chains, and a baseband unit fit into the dimensions of a small mobile phone.

In the second phase, the relay transmits \( \sqrt{p_2} s \) and the received signal at the destination is
\[
y_{2d} = h_{rd} \sqrt{p_2} s + n_{2d}, \tag{9}
\]
where \( p_2 \) is the transmit power, \( h_{rd} \in \mathbb{C} \) denotes the channel between the relay and destination, while \( n_{2d} \sim \mathcal{N}(0, \sigma^2) \) is the receiver noise. By utilizing (7) and (9) for maximum ratio combining, the following rate is achievable at the destination.

**Lemma 2.** The achievable rate with repetition-coded DF relaying is
\[
R_{\text{DF}} = \frac{1}{2} \log_2 \left( 1 + \min \left( \frac{p_1 |h_{sr}|^2}{\sigma^2}, \frac{p_1 |h_{rd}|^2}{\sigma^2} + \frac{p_2 |h_{rd}|^2}{\sigma^2} \right) \right). \tag{10}
\]

Proof: This is a classical result found in \([10]\) Eq. (15).

Remark 1. For brevity, the analysis in this letter assumes deterministic channels, but the extension to fading channels with perfect channel knowledge is straightforward; we only need to take expectations of the rate expressions in (6) and (10). Hence, all the conclusions apply to this case as well.

III. Analytical Performance Comparison

In this section, we compare the three achievable rates that were presented in Section I. Interestingly, the expressions only depend on the amplitudes of the channel elements, but not on their phases. For brevity, we introduce the notation \( |h_{sd}| = \sqrt{\beta_{sd}}, |h_{sr}| = \sqrt{\beta_{sr}}, |h_{rd}| = \sqrt{\beta_{rd}}, \) and \( \frac{1}{2} \sum_{n=1}^{N} |[h_{sr} n | h_{rd} n]| = \sqrt{\beta_{\text{IRS}}} \). We can now rewrite (2), (6), and (10) in the more compact forms
\[
R_{\text{SISO}} = \log_2 \left( 1 + \frac{p \beta_{sd}}{\sigma^2} \right), \tag{11}
\]
\[
R_{\text{IRS}}(N) = \log_2 \left( 1 + \frac{p \sqrt{\beta_{sd}} + N \alpha \sqrt{\beta_{\text{IRS}}}}{\sigma^2} \right), \tag{12}
\]
\[
R_{\text{DF}} = \frac{1}{2} \log_2 \left( 1 + \min \left( \frac{p_1 \beta_{sr}}{\sigma^2}, \frac{p_1 \beta_{rd}}{\sigma^2} + \frac{p_2 \beta_{rd}}{\sigma^2} \right) \right). \tag{13}
\]

It is obvious that \( R_{\text{IRS}}(N) \geq R_{\text{SISO}} \) since equality is achieved for \( N = 0 \) and \( R_{\text{IRS}}(N) \) is an increasing function of \( N \). In fact, the rate grows as \( O(\log_2(N^2)) \) when \( N \) is large, as previously noted in \([5]\) and further explained in \([10]\). The comparison between the IRS and DF relay cases is non-trivial. To make it fair, we first select \( p_1 \) and \( p_2 \) optimally, while having the same average power \( p \) as when using the IRS.
Proposition 1. Assume that \( p_1, p_2 \geq 0 \) are selected under the constraint \( p = \frac{p_1 + p_2}{2} \). If \( \beta_{sd} > \beta_{sr} \), it holds that \( R_{SISO} > R_{DF} \) for any selection of \( p_1, p_2 \), thus DF relaying is suboptimal.

If \( \beta_{sd} \leq \beta_{sr} \), the rate with DF relaying is maximized by \( p_1 = \frac{2p\beta_{sr} - \beta_{sd}}{\beta_{sr} + \beta_{rd} - \beta_{sd}} \) and \( p_2 = \frac{2p\beta_{sd} - \beta_{sr}}{\beta_{sr} + \beta_{rd} - \beta_{sd}} \), leading to

\[
R_{DF} = \frac{1}{2} \log_2 \left( 1 + \frac{2p\beta_{sr}\beta_{rd}}{(\beta_{sr} + \beta_{rd} - \beta_{sd})\sigma^2} \right). \tag{14}
\]

Proof: If \( \beta_{sd} > \beta_{sr} \), \( \min \left( \frac{p_1\beta_{sr}}{\sigma^2} + \frac{p_2\beta_{sd}}{\sigma^2}, \frac{p_1\beta_{sr}}{\sigma^2} + \frac{p_2\beta_{sd}}{\sigma^2} \right) = \frac{p_1\beta_{sr}}{\sigma^2} \), which is maximized by \( p_1 = 2p \) and \( p_2 = 0 \). Hence, the relay is not used and obviously \( R_{SISO} > R_{DF} \). If \( \beta_{sd} \leq \beta_{sr} \), \( R_{DF} \) is maximized by selecting \( p_1, p_2 \) to achieve \( \frac{p_1\beta_{sr}}{\sigma^2} = \frac{p_2\beta_{sd}}{\sigma^2} \) under the constraint \( p = \frac{p_1 + p_2}{2} \). This gives a linear system of equations with the solution that is stated in the proposition.

One important implication of Proposition 1 is that the relay-supported network needs to switch between two modes: SISO transmission and DF relaying. It is only when the channel from the source to the relay is stronger than the direct path from the source to destination (i.e., \( \beta_{sr} \geq \beta_{sd} \)) that DF relaying might provide \( R_{DF} > R_{SISO} \).

Proposition 2. The IRS-supported transmission provides the highest rate for any \( N \geq 1 \) if \( \beta_{sd} > \beta_{sr} \). In the case \( \beta_{sd} \leq \beta_{sr} \), it provides the highest rate if and only if

\[
N > \frac{\sqrt{\left(1 + \frac{2p\beta_{sr}\beta_{rd}}{(\beta_{sr} + \beta_{rd} - \beta_{sd})\sigma^2} - 1\right)} \sigma^2}{p - \sqrt{\beta_{sd}}} \alpha \sqrt{\beta_{IRS}}. \tag{15}
\]

Proof: Since \( R_{IRS}(N) > R_{SISO} \) for \( N \geq 1 \), the IRS-supported case gives the highest rate if and only if \( R_{IRS}(N) > R_{DF} \). This always occurs for \( \beta_{sd} > \beta_{sr} \) since \( R_{SISO} > R_{DF} \) due to Proposition 1. If \( \beta_{sd} \leq \beta_{sr} \), the inequality \( R_{IRS}(N) > R_{DF} \) can be simplified to (15) by utilizing (12) and (14).

To interpret the result in Proposition 2, we now consider that there is line-of-sight to and from the IRS. We assume that each IRS element has the same size as the relay’s antenna, thus it follows that all elements in \( h_{sr} \) have the same magnitude as \( h_{rd} \) and all elements in \( h_{sd} \) have the same magnitude as \( h_{rd} \). Consequently, \( \beta_{rs} = \beta_{sr} \beta_{rd} \). We first notice that although IRS-supported transmission provides the highest rate for \( \beta_{sd} > \beta_{sr} \), the difference between \( R_{IRS}(N) \) and \( R_{SISO} \) is small in this case since \( \sqrt{\beta_{sd}} \gg N\alpha \sqrt{\beta_{IRS}} = N\alpha \sqrt{\beta_{sr} \beta_{rd}} \) for most practical values of \( N \) because \( \beta_{rd} \) is a very small number in practice; note that a “large” channel gain in wireless communications is \(-60 \text{dB}\). Hence, it is in the case \( \beta_{sd} \leq \beta_{sr} \) that an IRS can provide an appreciable performance gain.

The right-hand side of (15) depends on the transmit SNR \( p/\sigma^2 \), the amplitude reflection coefficient \( \alpha \), and the channel gains \( \beta_{sd}, \beta_{sr}, \) and \( \beta_{rd} \) (recall that \( \beta_{IRS} = \beta_{sr} \beta_{rd} \)). Note that the right-hand side approaches \( \frac{\sqrt{\beta_{sd}}}{\alpha \sqrt{\beta_{sr} \beta_{rd}}} \) as \( p \to \infty \), which implies that the IRS-supported transmission achieves the largest rate at high SNR for any \( N \). In contrast, the inequality in (15) becomes

\[
N > \frac{1}{\sqrt{(\beta_{sr} + \beta_{rd} - \beta_{sd})}} - \sqrt{\beta_{sr} \beta_{rd}}. \tag{16}
\]

as \( p \to 0 \), which can be a very large number if \( \beta_{sd} \ll \beta_{sr} \). For example, (16) becomes \( N > 963 \) for \( \alpha = 1, \beta_{sd} = -110 \text{dB}, \beta_{sr} = -80 \text{dB}, \) and \( \beta_{rd} = -60 \text{dB} \).

In summary, the choice between an IRS and a relay depends on the SNR and number of elements. In Section IV, we assess if practical setups operate in the low or high SNR regime; that is, if any of the asymptotic results above can applied.

A. Transmit Power Minimization Under Rate Constraints

If the destination requires a particular data rate \( \bar{R} \), the rate expressions in (11)–(14) can be used to identify the required transmit power for each of the three communication setups.

Corollary 1. To achieve a data rate \( \bar{R} \), the SISO case requires the power

\[
p_{SISO} = \left(2\bar{R} - 1\right) \frac{\sigma^2}{\beta_{sd}}, \tag{17}
\]

the IRS-supported transmission requires the power

\[
p_{IRS}(N) = \left(2\bar{R} - 1\right) \frac{\sigma^2}{\sqrt{\beta_{sd} + N\alpha \sqrt{\beta_{IRS}}}}. \tag{18}
\]

and the relay-supported transmission requires the power

\[
p_{DF} = \begin{cases} \left(2\bar{R} - 1\right) \frac{\sigma^2}{\beta_{sd}} & \text{if } \beta_{sd} > \beta_{sr} \ , \\ \left(2\bar{R} - 1\right) \frac{\sigma^2}{\beta_{sd} + \beta_{sd} - \beta_{sd}} & \text{if } \beta_{sd} \leq \beta_{sr} . \end{cases} \tag{19}
\]

If the relay-supported system switches between SISO and DF relaying mode to minimize the transmit power, its required transmit power is

\[
p_{\text{DF mode}} = \min(p_{SISO}, p_{DF}).
\]

B. Total Power Minimization Under Rate Constraints

The total power consumption, \( P_{\text{total}} \), of the system consists of both transmit power and dissipation in hardware components. In the SISO case, it is \( P_{\text{total}}^{SISO} = P_s + P_d + P_r \), where \( P_r \in [0, 1] \) is the efficiency of the power amplifier while \( P_s \) and \( P_d \) are the hardware-dissipated power at the source and destination, respectively. In the IRS case, it becomes \( P_{\text{total}}^{IRS} = \frac{p_{IRS}(N)}{\nu} + P_s + P_d + N \rho e \).

\[
P_{\text{total}}^{IRS}(N) = \frac{p_{IRS}(N)}{\nu} + P_s + P_d + N \rho e, \tag{20}
\]

where \( P_e \) is the power dissipation per element, which is caused by the circuitry required for adaptive phase-shifting. In the relaying case, the source is only active half of the time, thus

\[
p_{\text{total}}^{DF} = \frac{p_{DF}}{\nu} + \frac{1}{2} P_s + P_d + P_t, \tag{21}
\]

where \( P_t \) is the hardware-dissipated power at the relay.

Proposition 3. Suppose \( \beta_{IRS} \) is a constant independent of \( N \). For a given data rate \( \bar{R} \), the total power \( P_{\text{total}}^{IRS}(N) \) when using an IRS is a convex function and minimized by

\[
N_{\text{opt}} = \sqrt{\frac{2\bar{R}}{\nu \sigma^2 \beta_{IRS} P_e}} - \frac{1}{\alpha} \sqrt{\frac{\beta_{sd}}{\beta_{IRS}}}. \tag{22}
\]

Proof: The function is convex since \( \frac{\sigma^2}{\nu} p_{IRS}(N) > 0 \). The solution (22) is then obtained from \( \frac{\sigma^2}{\nu} p_{IRS}(N) = 0 \).
We consider the simulation setup in Fig. 3, where the source and IRS/relay are deployed at fixed locations, while the location of the destination is determined by the variable \(d_1\). We use (23) to compute the channel gains based on the distances, assuming equal-sized 5 dBi antennas at the source and IRS/relay, while the destination is a handset with an omnidirectional antenna with 0 dBi. The IRS and relay are deployed to have LOS channels to the source, and the destination has a LOS channel to the IRS/relay. We can therefore assume \(\beta_{\text{IRS}} = \beta_{\text{sr}} \beta_{\text{rd}}\). Moreover, there is an NLOS channel between the source and destination, which leads to a weaker channel gain and motivates the use of an IRS or relay to support the transmission.

Fig. 4 shows the transmit power that is needed to achieve the rate \(\bar{R}\) in the scenario shown in Fig. 3 as a function of the distance \(d_1\).

**A. Energy Efficiency**

It was shown in [11] that an IRS can improve the energy efficiency (EE), which is defined as \(B \cdot \bar{R}/P_{\text{total}}\). We will...
make a similar analysis using the models in Sec. III-B with $\nu = 0.5$, $P_s = P_d = P_t = 100\text{ mW}$, $P_c = 5\text{ mW}$ [11], and $d_1 = 70\text{ m}$. Fig. 5 shows the EE as a function of $\tilde{R}$. The number of elements in the IRS, $N$, is optimized for maximal EE using Proposition 3. The SISO case provides the highest EE for $\tilde{R} \in (0, 3.47]$ bit/s/Hz, while the DF relaying case provides the highest EE for $\tilde{R} \in (3.47, 8.41]$ bit/s/Hz. It is only for $\tilde{R} > 4$ bit/s/Hz that the IRS has $N_{\text{opt}} > 0$ and it is only for $\tilde{R} > 8.41$ bit/s/Hz that it provides higher EE than DF relaying. Hence, a system that switches between the SISO and DF relaying modes is preferable both in terms of minimizing the transmit power and maximizing the energy efficiency, except when very high rates are required.

V. CONCLUSION AND DISCUSSION

We have compared classic repetition-coded DF relaying with the new concept of IRSs. The key observation is that an IRS needs hundreds of reconfigurable elements (each of the size of an antenna) to be competitive—even if we considered ideal phase-shifting and frequency-flat channels, which are two assumptions that clearly benefit the IRS. The reason is that the source’s transmit power must travel over two channels to reach the destination, leading to a very small channel gain $\beta_{sr}$ per element in the IRS—the SNR becomes almost the same as for amplify-and-forward relaying without amplification. Hence, the IRS needs many elements to compensate for the low channel gain. In contrast, with DF relaying, we first transmit over a channel with gain $\beta_{sr}$ and then transmit again over a channel with gain $\beta_{rd}$. While the large number of elements is a weakness for IRSs, the advantage is that an IRS requires no power amplifiers in its ideal form; however, in practice, active components are needed for adaptive phase-shifting. Even if the power dissipation per element is low, the total power is non-negligible. An IRS only achieves higher EE than DF relaying if very high rates are needed. Note that we only considered repetition-coded DF relaying, but other DF protocols achieve higher rates by optimizing the coding of the two hops and, thus, are even more competitive against an IRS-supported transmission [13].

The fact that the source and destination are physically separated from the IRS is the key feature—it allows for controlling the propagation environment—but also the reason for the large pathlosses. Classical reflectarrays are using nearby sources equipped with high-gain horn antennas to manage the pathlosses. Although the IRS will be larger than a relay, it is important to notice that an IRS with hundreds of elements, which was necessary to beat DF relaying in the simulations of this paper, can be still physically rather small since each element is assumed to have a sub-wavelength size [3], [4]. In general, it is the total size of the IRS (and not the number of elements or their individual size) that determines the pathloss, as explained in detail in [13]. For the sizes considered in this paper, the IRS will not behave as a specular reflector but reflect the incident wave as a beam; however, there are other scenarios where that might occur, particularly when operating in the THz range.

Reproducible research: The simulation code can be downloaded from [https://github.com/emilbjornson/IRS-relaying](https://github.com/emilbjornson/IRS-relaying)