Abstract—In this paper, we investigate the capacity of the Gaussian two-hop full-duplex (FD) relay channel with residual self-interference. This channel is comprised of a source, an FD relay, and a destination, where a direct source-destination link does not exist and the FD relay is impaired by residual self-interference. We adopt the worst-case linear self-interference model with respect to the channel capacity, and model the residual self-interference as a Gaussian random variable whose variance depends on the amplitude of the transmit symbol of the relay. For this channel, we derive the capacity and propose an explicit capacity-achieving coding scheme. Thereby, we show that the optimal input distribution at the source is Gaussian and its variance depends on the amplitude of the transmit symbol of the relay. On the other hand, the optimal input distribution at the relay is discrete or Gaussian, where the latter case occurs only when the relay-destination link is the bottleneck link. The derived capacity converges to the capacity of the two-hop ideal FD relay channel without self-interference and to the capacity of the two-hop half-duplex (HD) relay channel in the limiting cases when the residual self-interference is zero and infinite, respectively. Our numerical results show that significant performance gains are achieved with the proposed capacity-achieving coding scheme compared to the achievable rates of conventional HD relaying and/or conventional FD relaying.

Index Terms—Channel capacity, relay, full-duplex, self-interference.

I. INTRODUCTION

In wireless communications, relays are employed in order to increase the data rate between a source and a destination. The resulting three-node channel is known as the relay channel [2]. If the distance between the source and the destination is very large or there is heavy blockage, then the relay channel can be modeled without a source-destination link, which leads to the so called two-hop relay channel. For the relay channel, there are two different modes of operation for the relay, namely, the full-duplex (FD) mode and the half-duplex (HD) mode. In the FD mode, the relay transmits and receives at the same time and in the same frequency band. As a result, FD relays are impaired by self-interference, which is the interference caused by the relay’s transmit signal to the relay’s received signal. Latest advances in hardware design have shown that the self-interference of an FD node can be suppressed significantly, see [3]-[27], which has led to an enormous interest in FD communication. For example, [11] reported that self-interference suppression of 110 dB is possible in certain scenarios. On the other hand, in the HD mode, the relay transmits and receives in the same frequency band but in different time slots or in the same time slot but in different frequency bands. As a result, HD relays completely avoid self-interference. However, since an HD relay transmits and receives only in half of the time/frequency resources compared to an FD relay, the achievable rate of the two-hop HD relay channel may be significantly lower than that of the two-hop FD relay channel.

Information-theoretic analyses of the capacity of the two-hop HD relay channel were provided in [28], [29]. Thereby, it was shown that the capacity of the two-hop HD relay channel is achieved when the HD relay switches between reception and transmission in a symbol-by-symbol manner and not in a codeword-by-codeword manner, as is done in conventional HD relaying [30]. Moreover, in order to achieve the capacity, the HD relay has to encode information into the silent symbol created when the relay receives [29]. For the Gaussian two-hop HD relay channel without fading, it was shown in [29] that the optimal input distribution at the relay is discrete and includes the zero (i.e., silent) symbol. On the other hand, the source transmits using a Gaussian input distribution when the relay transmits the zero (i.e., silent) symbol and is silent otherwise.

The capacity of the Gaussian two-hop FD relay channel with ideal FD relaying without residual self-interference was derived in [2]. However, in practice, canceling the residual self-interference completely is not possible due to limitations in channel estimation precision and imperfections in the transceiver design [21]. As a result, the residual self-interference has to be taken into account when investigating the capacity of the two-hop FD relay channel. Despite the considerable body of work on FD relaying, see e.g. [5], [6], [9], [19], [26], the capacity of the two-hop FD relay channel with residual self-interference has not been explicitly characterized yet. As a result, for this channel, only achievable rates are known which are strictly smaller than the capacity. Therefore, in this paper, we study the capacity of the two-hop FD relay channel with residual self-interference for the case when the source-relay and relay-destination links are additive white Gaussian noise (AWGN) channels.

In general, the statistics of the residual self-interference depend on the employed hardware configuration and the...
adopted self-interference suppression schemes. As a result, different hardware configurations and different self-interference suppression schemes may lead to different statistical properties of the residual self-interference, and thereby, to different capacities for the considered relay channel. An upper bound on the capacity of the two-hop FD relay channel with residual self-interference is given in [2] and is obtained by assuming zero residual self-interference. Hence, the objective of this paper is to derive a lower bound on the capacity of this channel valid for any linear residual self-interference model. To this end, we consider the worst-case linear self-interference model with respect to the capacity, and thereby, we obtain the desired lower bound on the capacity for any other type of linear residual self-interference. For the worst-case, the linear residual self-interference is modeled as a conditionally Gaussian distributed random variable (RV) whose variance depends on the amplitude of the symbol transmitted by the relay.

For this relay channel, we derive the corresponding capacity and propose an explicit coding scheme which achieves the capacity. We show that the FD relay has to operate in the decode-and-forward (DF) mode to achieve the capacity, i.e., it has to decode each codeword received from the source and then transmit the decoded information to the destination in the next time slot, while simultaneously receiving. Moreover, we show that the optimal input distribution at the relay is discrete or Gaussian, where the latter case occurs only when the relay-destination link is the bottleneck link. On the other hand, the capacity-achieving input distribution at the source is Gaussian and its variance depends on the amplitude of the symbol transmitted by the relay, i.e., the average power of the source’s transmit symbol depends on the amplitude of the relay’s transmit symbol. In particular, the smaller the amplitude of the relay’s transmit symbol is, the higher the average power of the source’s transmit symbol should be, since in that case, the residual self-interference is small with high probability. On the other hand, if the amplitude of the relay’s transmit symbol is very large and exceeds some threshold, the chance for very strong residual self-interference is high and the source should remain silent and conserve its energy for other symbol intervals with weaker residual self-interference. We show that the derived capacity converges to the capacity of the two-hop ideal FD relay channel without self-interference [2] and to the capacity of the two-hop HD relay channel [29] in the limiting cases when the residual self-interference is zero and infinite, respectively. Our numerical results reveal that significant performance gains are achieved with the proposed capacity-achieving coding scheme compared to the achievable rates of conventional HD relaying and/or conventional FD relaying.

This paper is organized as follows. In Section II, we present the models for the channel and the residual self-interference. In Section III, we present the capacity of the considered channel and propose an explicit capacity-achieving coding scheme. Numerical examples are provided in Section IV, and Section V concludes the paper.

II. System Model

In the following, we introduce the models for the two-hop FD relay channel and the residual self-interference.

A. Channel Model

We assume a two-hop FD relay channel comprised of a source, an FD relay, and a destination, where a direct source-destination link does not exist. We assume that the source-relay and the relay-destination links are AWGN channels, and that the FD relay is impaired by residual self-interference. In symbol interval $i$, let $X_S[i]$ and $X_R[i]$ denote RVs which model the transmit symbols at the source and the relay, respectively, let $Y_R[i]$ and $Y_D[i]$ denote RVs which model the received symbols at the relay and the destination, respectively, and let $N_R[i]$ and $N_D[i]$ denote RVs which model the AWGNs at the relay and the destination, respectively. We assume that $N_R[i] \sim N(0, \sigma_R^2)$ and $N_D[i] \sim N(0, \sigma_D^2)$, where $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Moreover, let $h_{SR}$ and $h_{RD}$ denote the channel gains of the source-relay and relay-destination channels, respectively, which are assumed to be constant during all symbol intervals, i.e., fading is not considered. In addition, let $I[i]$ denote the RV which models the residual self-interference at the FD relay that remains in symbol interval $i$ after analog and digital self-interference cancelation.

Using the notations defined above, the input-output relations describing the considered relay channel in symbol interval $i$ are given as

$$
\hat{Y}_R[i] = h_{SR}X_S[i] + I[i] + \hat{N}_R[i] \tag{1}
$$
$$
\hat{Y}_D[i] = h_{RD}X_R[i] + \hat{N}_D[i]. \tag{2}
$$

Furthermore, an average “per-node” power constraint is assumed, i.e.,

$$
E\{X_S^2[i]\} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_S^2[k] \leq P_{S}, \quad \beta \in \{S, R\}, \tag{3}
$$

where $E\{\cdot\}$ denotes statistical expectation, and $P_S$ and $P_R$ are the average power constraints at the source and the relay, respectively.

B. Residual Self-Interference Model

Assuming narrow-band signals such that the channels can be modelled as frequency flat, a general model for the residual self-interference at the FD relay in symbol interval $i$, $I[i]$, is given by [11]

$$
I[i] = \sum_{m=1}^{M} h_{RR,m}[i](X_R[i])^m \tag{4},
$$

As customary for capacity analysis, see e.g. [31], as a first step we do not consider fading and assume real-valued channel inputs and outputs. The generalization to a complex-valued signal model is relatively straightforward [32]. On the other hand, the generalization to the case of fading is considerably more involved. For example, considering the achievability scheme for HD relays in [33], we expect that when fading is present, both HD and FD relays have to perform buffering in order to achieve the capacity. However, the corresponding detailed analysis is beyond the scope of this paper and presents an interesting topic for future research.
where $M \leq \infty$ is an integer and $h_{RR,m}[i]$ is the residual self-interference channel between the transmitter-end and the receiver-end of the FD relay through which symbol $(X_R[i])^m$ arrives at the receiver-end. Moreover, for $m = 1$, $(X_R[i])^m$ is the linear component of the residual self-interference, and for $m \geq 2$, $(X_R[i])^m$ is a nonlinear component of the residual self-interference. As shown in [11], only the terms for which $m$ is odd in (4) carry non-negligible energy while the remaining terms for which $m$ is even can be ignored. Moreover, as observed in [11], the higher order terms in (4) carry significantly less energy than the lower order terms, i.e., the term for $m = 5$ carries significantly less energy than the term for $m = 3$, and the term for $m = 3$ carries significantly less energy than the term for $m = 1$. As a result, in this paper, we adopt the first order approximation of the residual self-interference in (4), i.e., $\hat{I}[i]$ is modeled as

$$\hat{I}[i] \approx h_{RR}[i]X_R[i],$$

where $h_{RR}[i] = h_{RR,1}[i]$ is used for simplicity of notation. Obviously, the residual self-interference model in (5) takes into account only the linear component of the residual self-interference and assumes that the nonlinear components can be neglected. Such a linear model for the residual self-interference is particularly justified for relays with low average transmit powers [10].

The residual self-interference channel gain in (5), $h_{RR}[i]$, is time-varying even when fading is not present, see e.g. [6], [10], [17], [22]. The variations of the residual self-interference channel gain, $h_{RR}[i]$, are due to the cumulative effects of various distortions originating from noise, carrier frequency offset, oscillator phase noise, analog-to-digital/digital-to-analog (AD/DA) conversion imperfections, I/Q imbalance, imperfect channel estimation, etc., see [6], [10], [17], [22]. These distortions have a significant impact on the residual self-interference channel gain due to the very small distance between the transmitter-end and the receiver-end of the self-interference channel. Moreover, the variations of the residual self-interference channel gain, $h_{RR}[i]$, are random and thereby cannot be accurately estimated at the FD node [6], [10], [17], [22]. The statistical properties of these variations are dependent on the employed hardware configuration and the adopted self-interference suppression schemes. In [10], $h_{RR}[i]$ is assumed to be constant during one codeword comprised of many symbols. Thereby, the residual self-interference model in [10] models only the long-term, i.e., codeword-by-codeword, statistical properties of the residual self-interference. However, the symbol-by-symbol variations of $h_{RR}[i]$ are not captured by the model in [10] since they are averaged out. Nevertheless, for a meaningful information-theoretical analysis, the statistics of the symbol-by-symbol variations of $h_{RR}[i]$ are needed. On the other hand, the statistics of the variations of $h_{RR}[i]$ affect the capacity of the considered relay channel. In this paper, we derive the capacity of the considered relay channel for the worst-case linear residual self-interference model, which yields a lower bound for the capacity for any other linear residual self-interference model.

To derive the worst-case linear residual self-interference model in terms of capacity, we insert (5) into (1), and obtain the received symbol at the relay in symbol interval $i$ as

$$\hat{Y}_R[i] = h_{SR}X_S[i] + h_{RR}X_R[i] + \tilde{N}_R[i].$$

(6)

Now, since in general $h_{RR}[i]$ can have a non-zero mean, without loss of generality, we can write $h_{RR}[i]$ as

$$h_{RR}[i] = \bar{h}_{RR} + \hat{h}_{RR}[i],$$

(7)

where $\bar{h}_{RR}$ is the mean of $h_{RR}[i]$, i.e., $\bar{h}_{RR} = E\{h_{RR}[i]\}$ and $\hat{h}_{RR}[i] = h_{RR}[i] - \bar{h}_{RR}$ is the remaining zero-mean random component of $h_{RR}[i]$. Inserting (7) into (6), we obtain the received symbol at the relay in symbol interval $i$ as

$$\hat{Y}_R[i] = h_{SR}X_S[i] + \bar{h}_{RR}X_R[i] + \hat{h}_{RR}[i]X_R[i] + \tilde{N}_R[i].$$

(8)

Given sufficient time, the relay can estimate any mean in its received symbols arbitrarily accurately, see [34]. Thereby, given sufficient time, the relay can estimate the deterministic component of the residual self-interference channel gain $\bar{h}_{RR}$. Moreover, since $X_R[i]$ models the desired transmit symbol at the relay, and since the relay knows which symbol it wants to transmit, the outcome of $X_R[i]$, denoted by $x_R[i]$, is known in each symbol interval $i$. As a result, the relay knows $\bar{h}_{RR}X_R[i]$ and thereby it can subtract $\bar{h}_{RR}X_R[i]$ from the received symbol $\hat{Y}_R[i]$ in (8). Consequently, we obtain a new received symbol at the relay in symbol interval $i$, denoted by $\hat{Y}_R[i]$, as

$$\hat{Y}_R[i] = h_{SR}X_S[i] + \hat{h}_{RR}[i]X_R[i] + \tilde{N}_R[i].$$

(9)

Now, assuming that the relay transmits symbol $x_R[i]$ in symbol interval $i$, i.e., $X_R[i] = x_R[i]$, from (9) we conclude that the relay “sees” the following additive impairment in symbol interval $i$

$$\hat{h}_{RR}[i]x_R[i] + \tilde{N}_R[i].$$

(10)

Consequently, from (10), we conclude that the worst-case scenario with respect to the capacity is if (10) is zero-mean independent and identically distributed (i.i.d.) Gaussian RV, which is possible only if $h_{RR}[i]$ is a zero-mean i.i.d. Gaussian RV. Hence, modeling the linear residual self-interference channel gain, $h_{RR}[i]$, as an i.i.d. Gaussian RV constitutes the worst-case linear residual self-interference model, and thereby, leads to a lower bound on the capacity for any other distribution of $h_{RR}[i]$.

Considering the developed worst-case linear residual self-interference model, in the rest of this paper, we assume $\hat{h}_{RR}[i] \sim \mathcal{N}(0, \hat{\alpha})$, where $\hat{\alpha}$ is the variance of $\hat{h}_{RR}[i]$. Since the average power of the linear residual self-interference at the relay is $\hat{\alpha}E\{X_R^2[i]\}$, $\hat{\alpha}$ can be interpreted as a self-interference amplification factor, i.e., $1/\hat{\alpha}$ is a self-interference suppression factor.

We note that similar distortions are also present in the source-relay and relay-destination channels. However, due to the large distance between transmitter and receiver, the impact of these distortions on the channel gains $h_{SR}$ and $h_{RD}$ is negligible.

3This is because a Gaussian RV has the highest uncertainty (i.e., entropy) among all possible RVs for a given second moment [31].
C. Simplified Input-Output Relations for the Considered Relay Channel

To simplify the input-output relation in (9), we divide the received symbol $Y_R[i]$ by $h_{SR}$ and thereby obtain a new received symbol at the relay in symbol interval $i$, denoted by $Y_R[i]$, and given by

$$Y_R[i] = X_S[i] + \frac{\hat{h}_{RR}[i]}{h_{SR}} X_R[i] + \hat{N}_R[i],$$

subject to $C1: E\{X_S^2\} \leq P_S$

$$X_R[i] = I[i] + N_R[i],$$

where

$$I[i] = \frac{\hat{h}_{RR}[i]}{h_{SR}} X_R[i]$$

is the normalized residual self-interference at the relay and $N_R[i] = N_R[i]/h_{SR}$ is the normalized noise at the relay distributed according to $N_R[i] \sim \mathcal{N}(0, \sigma_R^2)$, where $\sigma_R^2 = \hat{\alpha}^2/4$. The normalized residual self-interference, $I[i]$, is dependent on the transmit symbol at the relay, $X_R[i]$, and, conditioned on $X_R[i]$, it has the same type of distribution as the random component of the self-interference channel gain, $\hat{h}_{RR}[i]$, i.e., an i.i.d. Gaussian distribution. Let $\alpha$ be defined as $\alpha = \hat{\alpha}/h_{SR}^2$, which can be interpreted as the normalized self-interference amplification factor. Using $\alpha$ and assuming that the transmit symbol at the relay in symbol interval $i$ is $X_R[i] = x_R[i]$, the distribution of the normalized residual self-interference, $I[i]$, can be written as

$$I[i] \sim \mathcal{N}(0, \alpha x_R^2[i]), \text{ if } X_R[i] = x_R[i].$$

To obtain also a received symbol at the destination, we normalize $Y_D[i]$ in (2) by $h_{RD}$, which yields

$$Y_D[i] = X_R[i] + N_D[i].$$

In (14), $N_D[i]$ is the normalized noise power at the destination distributed as $N_D[i] \sim \mathcal{N}(0, \sigma_D^2)$, where $\sigma_D^2 = \hat{\alpha}^2/4$. Now, instead of deriving the capacity of the considered relay channel using the input-output relations in (1) and (2), we can derive the capacity using an equivalent relay channel defined by the input-output relations in (11) and (14), respectively, where, in symbol interval $i$, $X_S[i]$ and $X_R[i]$ are the inputs at source and relay, respectively, $Y_R[i]$ and $Y_D[i]$ are the outputs at relay and destination, respectively, $N_R[i]$ and $N_D[i]$ are AWGNs with variances $\sigma_R^2 = \hat{\alpha}^2/4$ and $\sigma_D^2 = \hat{\alpha}^2/4$, respectively, and $I[i]$ is the residual self-interference with conditional distribution given by (13), which is a function of the normalized self-interference amplification factor $\alpha$.

III. CAPACITY

In this section, we study the capacity of the considered Gaussian two-hop FD relay channel with residual self-interference.

A. Derivation of the Capacity

To derive the capacity of the considered relay channel, we first assume that RVs $X_S$ and $X_R$, which model the transmit symbols at source and relay for any symbol interval $i$, take values $x_S$ and $x_R$ from sets $X_S$ and $X_R$, respectively. Now, since the considered relay channel belongs to the class of memoryless degraded relay channels defined in [2], its capacity is given by [2, Theorem 1]

$$C = \max_{p(x_S,x_R) \in P} \min \{ I(X_S;Y_R|X_R), I(X_R;Y_D) \}$$

subject to $C1: E\{X_S^2\} \leq P_S$

$$C2: E\{X_R^2\} \leq P_R,$$

where $P$ is a set which contains all valid distributions. In order to obtain the capacity in (15), we need to find the optimal joint input distribution, $p(x_S,x_R)$, which maximizes the min function in (15) and satisfies constraints C1 and C2. To this end, note that $p(x_S,x_R)$ can be written equivalently as $p(x_S|x_R) = p(x_S|x_R)p(x_R)$. Using this relation, we can represent the maximization in (15) equivalently as two nested maximizations, one with respect to $p(x_S|x_R)$ for a fixed $p(x_R)$, and the other one with respect to $p(x_R)$. Thereby, we can write the capacity in (15) equivalently as

$$C = \max_{p(x_R) \in P} \max_{p(x_S|x_R) \in P} \min \{ I(X_S;Y_R|X_R), I(X_R;Y_D) \}$$

subject to $C1: E\{X_S^2\} \leq P_S$

$$C2: E\{X_R^2\} \leq P_R.$$  

(16)

Now, since in the min function in (16) only $I(X_S;Y_R|X_R)$ is dependent on the distribution $p(x_S|x_R)$, whereas $I(X_R;Y_D)$ does not depend on $p(x_S|x_R)$, we can write the capacity expression in (16) equivalently as

$$C = \max_{p(x_R) \in P} \min \{ I(X_S;Y_R|X_R), I(X_R;Y_D) \}$$

subject to $C1: E\{X_S^2\} \leq P_S$

$$C2: E\{X_R^2\} \leq P_R.$$  

(17)

Hence, to obtain the capacity of the considered relay channel, we first need to find the conditional input distribution at the source, $p(x_S|x_R)$, which maximizes $I(X_S;Y_R|X_R)$ such that constraint C1 holds. Next, we need to find the optimal input distribution at the relay, $p(x_R)$, which maximizes the min function in (17) such that constraints C1 and C2 hold.

B. Optimal Input Distribution at the Source $p^*(x_S|x_R)$

The optimal input distribution at the source which achieves the capacity in (17), denoted by $p^*(x_S|x_R)$, is given in the following theorem.

Theorem 1: The optimal input distribution at the source $p^*(x_S|x_R)$, which achieves the capacity of the considered relay channel in (17), is the zero-mean Gaussian distribution with variance $P_S(x_R)$ given by

$$P_S(x_R) = \alpha \max\{0, x_{th}^2 - x_R^2\},$$

(18)
where \( x_{th} \) is a positive threshold constant found as follows. If \( p(x) \) is a discrete distribution, \( x_{th} \) is found as the solution of the following identity

\[
\sum_{x_R \in X_R} \alpha \max \{0, x_{th}^2 - x_R^2\} p(x) = P_S,
\]

and the corresponding

\[
\max_{p(x) | x_R \in P} I(X_S; Y_R | X_R)
\]

is obtained as

\[
\max_{p(x) | x_R \in P} I(X_S; Y_R | X_R) = \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R).
\]

Otherwise, if \( p(x) \) is a continuous distribution, the sums in (19) and (20) have to be replaced by integrals.

**Proof:** Please refer to Appendix A.

From Theorem 1, we can see that the source should perform power allocation in a symbol-by-symbol manner. In particular, the average power of the source’s transmit symbols, \( P_S(x_R) \), given by (18), depends on the amplitude of the transmit symbol at the relay, \( |x_R| \). The lower the amplitude of the transmit symbol of the relay is, the higher the average power of the source’s transmit symbols should be, in that case, there is a high probability for weak residual self-interference. Conversely, the higher the amplitude of the transmit symbol of the relay is, the lower the average power of the source’s transmit symbols should be, in that case, there is a high probability for strong residual self-interference. Moreover, solving (24) reveals that constraint C1 has to hold with equality and that \( p^*(x_R) \) has the following discrete form

\[
p^*(x_R) = p_{R,0} \delta(x_R) + \sum_{j=1}^J \frac{1}{2} p_{R,j} \left( \delta(x_R - x_{R,j}) + \delta(x_R + x_{R,j}) \right),
\]

where \( p_{R,j} \in [0,1] \) is the probability that \( X_R = x_{R,j} \) will occur, where \( x_{R,j} > 0 \) and \( \sum_{j=0}^J p_{R,j} = 1 \) hold. With \( p^*(x_R) \) as in (25), the capacity has the following general form

\[
C = \frac{P_R}{2} \log_2 \left( 1 + \frac{\alpha x_{th}^2}{\sigma_R^2 + \alpha x_{th}^2} \right) + \sum_{j=1}^J \frac{1}{2} p_{R,j} \log_2 \left( 1 + \frac{\alpha x_{th}^2}{\sigma_R^2 + \alpha x_{th,j}} \right).
\]

**Proof:** Please refer to Appendix B.

From Theorem 2, we can draw the following conclusions. If condition (21) holds, then the relay-destination channel is the bottleneck link. In particular, even if the relay transmits with a zero-mean Gaussian distribution, which achieves the capacity of the relay-destination channel, the capacity of the relay-destination channel is still smaller than the mutual information (i.e., data rate) of the source-relay channel. Otherwise, if condition (21) does not hold, then the optimal input distribution at the relay, \( p^*(x_R) \), is always discrete and symmetric with respect to \( x_R = 0 \). Moreover, in this case, the mutual informations of the source-relay and relay-destination channels have to be equal, i.e., \( I(X_S; Y_R | X_R) |_{p(x_R) = p^*(x_R)} = I(X_R; Y_D) |_{p(x_R) = p^*(x_R)} \) has to hold. In addition, we note that constraint C2 in (24) does not always have to hold with equality, i.e., in certain cases it is optimal for the relay to reduce its average transmit power. In particular, if the relay-destination channel is very strong compared to the source-relay channel, then, by reducing the average transmit power.
of the relay, we reduce the average power of the residual self-interference in the source-relay channel, and thereby improve the quality of the source-relay channel. We note that this phenomenon was first observed in [3] and [5], where it was shown that, in certain cases, it is beneficial for FD relays to not transmit with the maximum available average power. However, even if the average transmit power of the relay is reduced, \( I(X_S; Y_D|X_R) = I(X_R; Y_D) \) still has to hold, for the data rates of the source-relay and the relay-destination channels to be equal.

**Remark 1:** From (22), it can be observed that threshold \( x_{th} \) is inversely proportional to the normalized self-interference amplification factor \( \alpha \). In other words, the smaller \( \alpha \) is, the larger \( x_{th} \) becomes. In the limit, when \( \alpha \to 0 \), we have \( x_{th} \to \infty \). This is expected since for smaller \( \alpha \), the average power of the residual self-interference also becomes smaller, which allows the source to transmit more frequently. If \( \alpha \to 0 \) the residual self-interference tends to zero. Consequently, the source should never be silent, i.e., \( x_{th} \to \infty \), which is in line with the optimal behavior of the source for the case of ideal FD relaying without residual self-interference described in [2]. On the other hand, inserting the solution for \( x_{th} \) from (22) into (21), and then evaluating (21), it can be observed that the right hand-side of (21) is a strictly decreasing function of \( \alpha \). This is expected since larger \( \alpha \) result in a residual self-interference with larger average power and thereby in a smaller achievable rate on the source-relay channel.

**D. Achievability of the Capacity**

The source wants to transmit message \( W \) to the destination, which is drawn uniformly from a message set \( \{1, 2, ..., 2^{nR}\} \) and carries \( nR \) bits of information, where \( n \to \infty \). To this end, the transmission time is split into \( B + 1 \) time slots and each time slot is comprised of \( k \) symbol intervals, where \( B \to \infty \) and \( k \to \infty \). Moreover, message \( W \) is split into \( B \) messages, denoted by \( w(1), ..., w(B) \), where each \( w(b) \), for \( b = 1, ..., B \), carries \( kR \) bits of information. Each of these messages is to be transmitted in a different time slot. In particular, in time slot one, the source sends message \( w(1) \) during \( k \) symbol intervals to the relay and the relay is silent. In time slot \( b \), for \( b = 2, ..., B \), source and relay send messages \( w(b) \) and \( w(b-1) \) to relay and destination during \( k \) symbol intervals, respectively. In time slot \( B + 1 \), the relay sends message \( w(B) \) to the destination during \( k \) symbol intervals and the source is silent. Hence, in the first time slot, the relay is silent since it does not have information to transmit, and in time slot \( B + 1 \), the source is silent since it has no more information to transmit. In time slots 2 to \( B \), both source and relay transmit. During the \( B + 1 \) time slots, the channel is used \( k(B + 1) \) times to send \( nR = BKkR \) bits of information, leading to an overall information rate of

\[
\lim_{B \to \infty} \lim_{k \to \infty} \frac{BKkR}{k(B + 1)} = R \quad \text{bits/symbol.} \quad (27)
\]

A detailed description of the proposed coding scheme for each time slot is given in the following, where we explain the rates, codebooks, encoding, and decoding used for transmission. We note that the proposed achievability scheme requires all three nodes to have full channel state information (CSI) of the source-relay and relay-destination channels as well as knowledge of the self-interference suppression factor \( 1/\alpha \).

**Rates:** The transmission rate of both source and relay is denoted by \( R \) and given by

\[
R = C - \epsilon, \quad (28)
\]

where \( C \) is given in Theorem 2 and \( \epsilon > 0 \) is an arbitrarily small number.

**Codebooks:** We have two codebooks, namely, the source’s transmission codebook and the relay’s transmission codebook. The source’s transmission codebook is generated by mapping each possible binary sequence comprised of \( kR \) bits, where \( R \) is given by (28), to a codeword \( x_S \) comprised of \( kp_T \) symbols, where \( p_T \) is the following probability

\[
p_T = \Pr \{ |x_R| < x_{th} \}. \quad (29)
\]

Hence, \( p_T \) is the probability that the relay will transmit a symbol with an amplitude which is smaller than the threshold \( x_{th} \). In other words, \( p_T \) is the fraction of symbols in the relay’s codeword which have an amplitude which is smaller than the threshold \( x_{th} \). The symbols in each codeword \( x_S \) are generated independently according to the zero-mean unit variance Gaussian distribution. Since in total there are \( 2^{kR} \) possible binary sequences comprised of \( kR \) bits, with this mapping, we generate \( 2^{kR} \) codewords \( x_S \) each comprised of \( kp_T \) symbols. These \( 2^{kR} \) codewords form the source’s transmission codebook, which we denote by \( C_S \).

On the other hand, the relay’s transmission codebook is generated by mapping each possible binary sequence comprised of \( kR \) bits, where \( R \) is given by (28), to a transmission codeword \( x_R \) comprised of \( k \) symbols. The symbols in each codeword \( x_R \) are generated independently according to the optimal distribution \( p^*(x_R) \) given in Theorem 2. The \( 2^{kR} \) codewords \( x_R \) form the relay’s transmission codebook denoted by \( C_R \).

The two codebooks are known at all three nodes. Moreover, the power allocation policy at the source, \( P_S(x_R) \), given in (18), is assumed to be known at source and relay.

**Remark 2:** We note that the source’s codewords, \( x_S \), are shorter than the relay’s codewords, \( x_R \), since the source is silent in \( 1 - p_T \) fraction of the symbol intervals because of the expected strong interference in those symbol intervals. Since the relay transmits during the symbol intervals for which the source is silent, its codewords are longer than the codewords of the source. Note that if the silent symbols of the source are taken into account and counted as part of the source’s codeword, then both codewords will have the same length.

**Encoding, Transmission, and Decoding:** In the first time slot, the source maps \( w(1) \) to the appropriate codeword \( x_S(1) \) from its codebook \( C_S \). Then, codeword \( x_S(1) \) is transmitted to the relay, where each symbol of \( x_S(1) \) is amplified by \( \sqrt{P_S(x_R = 0)} \), where \( P_S(x_R) \) is given in (18). On the other hand, the relay is scheduled to always receive and be silent (i.e., to set its transmit symbol to zero) during the first time slot. However, knowing that the codeword transmitted by the source in the first time slot, \( x_S(1) \), is comprised of \( kp_T \)
symbols, the relay constructs the received codeword, denoted by \( y_R(1) \), only from the first \( kp_{\text{FP}} \) received symbols.

**Lemma 1**: The codeword \( x_S(1) \) sent in the first time slot can be decoded successfully from the codeword received at the relay, \( y_R(1) \), using a typical decoder [31] since \( R \) satisfies

\[
R < \max_{p(x)} \left( I(X_S;Y_R) \right)_{p(x)=p^*(x_R)} = \frac{1}{2} \log_2 \left( 1 + \frac{\alpha x_{1h}^2}{\sigma_R^2} \right) p^*(x_R). \tag{30}
\]

**Proof**: Please refer to Appendix D.

In time slots \( b = 2, ..., B \), the encoding, transmission, and decoding are performed as follows. In time slots \( b = 2, ..., B \), the source and the relay map \( w(b) \) and \( w(b-1) \) to the appropriate codewords \( x_S(b) \) and \( x_R(b) \) from codebooks \( C_S \) and \( C_R \), respectively. Note that the source also knows \( x_R(b) \) since \( x_R(b) \) was generated from \( w(b) \) which the source transmitted in the previous (i.e., the \( (b-1) \)-th) time slot. As a result, both source and relay know the symbols in \( x_R(b) \) and can determine whether their amplitudes are smaller or larger than the threshold \( x_{1h} \). Hence, if the amplitude of the first symbol in codeword \( x_R(b) \) is smaller than the threshold \( x_{1h} \), then in the first symbol interval of time slot \( b \), the source transmits the first symbol from codeword \( x_S(b) \) amplified by \( \sqrt{p_S(x_R)} \), where \( x_{R,1} \) is the first symbol in relay’s codeword \( x_R(b) \) and \( p_S(x_R) \) is given by (18). Otherwise, if the amplitude of the first symbol in codeword \( x_R(b) \) is larger than \( x_{1h} \), then the source is silent. The same procedure is performed for the \( j \)-th symbol interval in time slot \( b \), for \( j = 1, ..., k \). In particular, if the amplitude of the \( j \)-th symbol in codeword \( x_R(b) \) is smaller than \( x_{1h} \), then in the \( j \)-th symbol interval of time slot \( b \), the source transmits its next untransmitted symbol from codeword \( x_S(b) \) amplified by \( \sqrt{p_S(x_{R,j})} \), where \( x_{R,j} \) is the \( j \)-th symbol in relay’s codeword \( x_R(b) \). Otherwise, if the amplitude of the \( j \)-th symbol in codeword \( x_R(b) \) is larger than \( x_{1h} \), then for the \( j \)-th symbol interval of time slot \( b \), the source is silent. On the other hand, the relay transmits all symbols from \( x_R(b) \) while simultaneously receiving. Let \( \hat{y}_R(b) \) denote the received codeword at the relay in time slot \( b \). Then, the relay discards those symbols from the received codeword, \( \hat{y}_R(b) \), for which the corresponding symbols in \( x_R(b) \) have amplitudes which exceed threshold \( x_{1h} \), and only collects the symbols in \( \hat{y}_R(b) \) for which the corresponding symbols in \( x_R(b) \) have amplitudes which are smaller than \( x_{1h} \). All symbols collected from \( \hat{y}_R(b) \) constitute the relay’s information-carrying received codeword, denoted by \( y_R(b) \), which is used for decoding.

**Lemma 2**: The codewords \( x_S(b) \) sent in time slots \( b = 2, ..., B \) can be decoded successfully at the relay from the corresponding received codewords \( y_R(b) \), respectively, using a jointly typical decoder since \( R \) satisfies

\[
R < \sum_{x_R \in X_R} \max_{p(x_R)} \left( I(X_S;Y_R) \right)_{p(x_R)=p^*(x_R)} = \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max\{0, x_{1h}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p^*(x_R). \tag{31}
\]

**Proof**: Please refer to Appendix E.

On the other hand, the destination listens during the entire time slot \( b \) and receives a codeword denoted by \( y_D(b) \). By following the “standard” method for analyzing the probability of error for rates smaller than the capacity, given in [31, Sec. 7.7], it can be shown in a straightforward manner that the destination can successfully decode \( x_R(b) \) from the received codeword \( y_D(b) \), and thereby obtain \( w(b-1) \), since rate \( R \) satisfies

\[
R < \max_{p(x_R)} I(X_R;Y_D)_{p(x_R)=p^*(x_R)} = \frac{1}{2} \log_2 \left( 1 + \frac{\alpha x_{1h}^2}{\sigma_R^2} \right) p^*(x_R). \tag{32}
\]

where \( I(X_R;Y_D) \) is given in Theorem 2.

In the last (i.e., the \( (B+1) \)-th) time slot, the source is silent and the relay transmits \( w(B) \) by mapping it to the corresponding codeword \( x_R(B+1) \) from codebook \( C_R \). The relay transmits all symbols in codeword \( x_R(B+1) \) to the destination during time slot \( B+1 \). The destination can decode the received codeword in time slot \( B+1 \) successfully, since (32) holds.

Finally, since both relay and destination can decode their respective codewords in each time slot, the entire message \( W \) can be decoded successfully at the destination at the end of the \( (B+1) \)-th time slot.

A block diagram of the proposed coding scheme is presented in Fig. 1. In particular, in Fig. 1, we show schematically the encoding, transmission, and decoding at source, relay, and destination. The flow of encoding/decoding in Fig. 1 is as follows. Messages \( w(b-1) \) and \( w(b) \) are encoded into \( x_R(b) \) and \( x_S(b) \) at the source using the encoders \( C_R \) and \( C_S \), respectively. Then, an inserter In is used to create a vector \( x_S(b) \) by inserting the symbols of \( x_S(b) \) into the positions of \( x_S(b) \) for which the corresponding elements of \( x_R(b) \) have amplitudes smaller than \( x_{1h} \) and setting all other symbols in \( x_S(b) \) to zero. Hence, vector \( x_S(b) \) is identical to codeword \( x_S(b) \) except for the added silent (i.e., zero) symbols generated at the source. The source then transmits \( x_S(b) \) and the relay receives the corresponding codeword \( y_R(b) \). Simultaneously, the relay encodes \( w(b-1) \) into \( x_R(b) \) using encoder \( C_R \) and transmits it to the destination, which receives codeword \( y_D(b) \). Next, using \( x_R(b) \), the relay constructs \( y_R(b) \) from \( y_R(b) \) by selecting only those symbols from \( y_R(b) \) for which the corresponding symbols in \( x_R(b) \) have amplitudes smaller than \( x_{1h} \). Using decoder \( D_R \), the relay then decodes \( y_R(b) \) into \( w(b) \) and stores the decoded bits in its buffer \( Q \). On the other hand, the destination decodes \( y_D(b) \) into \( w(b-1) \) using decoder \( D_D \).

E. Analytical Expression for Tight Lower Bound on the Capacity

For the non-trivial case when condition (21) does not hold, i.e., the relay-destination link is not the bottleneck, the capacity of the Gaussian two-hop FD relay channel with residual self-interference is given in the form of an optimization problem, cf. (24), which is not suitable for analysis. As a result, in this subsection, we propose a suboptimal input distribution at the relay, which yields an analytical expression for a lower bound on the capacity, derived in Theorem 2. Our numerical results...
show that this lower bound is tight, at least for the considered numerical examples, cf. Fig. 3. In particular, we propose that

The relay uses the following input distribution

\[ p(x_R) = \frac{1}{2\pi p_R/q} \exp \left( -\frac{x_R^2}{2p_R/q} \right) + (1-q)\delta(x_R), \]  

where the value of \( p_R \) is optimized in the range \( p_R \leq P_R \) in order for the rate to be maximized. Hence, with probability \( q \), the relay transmits a symbol from a zero-mean Gaussian distribution with variance \( p_R/q \), and is silent with probability \( 1-q \). Since the relay transmits only in \( q \) fraction of the time, the average transmit power when the relay transmits is set to \( p_R/q \) in order for the average transmit power during the entire transmission time to be \( P_R \). Now, with the input distribution \( p_B(x_R) \) in (33), we obtain the mutual information of the source-relay channel as

\[
\max_{p(x_S|x_B) \in P} I(X_S;Y_R|X_R)
\begin{aligned}
&= q \int_{-x_{th}}^{x_{th}} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha(x_{th}^2 - x_R^2)\sigma_R^2}{\sigma_R^2 + \alpha x_{th}^2} \right) \\
&\quad \times \frac{1}{\sqrt{2\pi p_R/q}} \exp \left( -\frac{x_R^2}{2p_R/q} \right) dx_R \\
&\quad + (1-q) \frac{1}{2} \log_2 \left( 1 + \frac{\alpha x_{th}^2}{\sigma_R^2} \right),
\end{aligned}
\]  

(34)

and the mutual information of the relay-destination channel as

\[
I(X_R;Y_D)|_{p(x_R)=p_B(x_R)}
\begin{aligned}
&= -\int_{-\infty}^{\infty} \left[ q \frac{1}{\sqrt{2\pi(p_R/q + \sigma_D^2)}} \exp \left( -\frac{y_D^2}{2(p_R/q + \sigma_D^2)} \right) \\
&\quad + (1-q) \frac{1}{\sqrt{2\pi\sigma_D^2}} \exp \left( -\frac{y_D^2}{2\sigma_D^2} \right) \right] \\
&\quad \times \log_2 \left( q \frac{1}{\sqrt{2\pi(p_R/q + \sigma_D^2)}} \exp \left( -\frac{y_D^2}{2(p_R/q + \sigma_D^2)} \right) \\
&\quad + (1-q) \frac{1}{\sqrt{2\pi\sigma_D^2}} \exp \left( -\frac{y_D^2}{2\sigma_D^2} \right) \right) dy_D - \frac{1}{2} \log_2 (2\pi e\sigma_D^2),
\end{aligned}
\]

(35)

The threshold \( x_{th} \) in (34) and the probability \( q \) in (34) and (35) are found from the following system of two equations

\[
\begin{align*}
q \left( \sqrt{\frac{2p_R/q}{\pi}} \alpha x_{th} \exp \left( -\frac{x_{th}^2}{2p_R/q} \right) \right) \\
+\alpha (x_{th}^2 - p_R/q) \operatorname{erf} \left( \frac{x_{th}}{\sqrt{2p_R/q}} \right) + (1-q)\alpha x_{th}^2 = P_S
\end{align*}
\]

\[
\max_{p(x_S|x_B) \in P} I(X_S;Y_R|X_R)|_{p(x_R)=p_B(x_R)}
\begin{aligned}
&= I(X_R;Y_D)|_{p(x_R)=p_B(x_R)}.
\end{aligned}
\]

(36)

Thereby, \( x_{th} \) and \( q \) are obtained as a function of \( p_R \). Now, the achievable rate with the suboptimal input distribution \( p_B(x_R) \) is found by inserting \( x_{th} \) and \( q \) found from (36) into (34) or (35), and then maximizing (34) or (35) with respect to \( p_R \) such that \( p_R \leq P_R \) holds.

### IV. Numerical Evaluation

In this section, we numerically evaluate the capacity of the considered two-hop FD relay channel with self-interference and compare it to several benchmark schemes. To this end, we first provide the system parameters, introduce benchmark schemes, and then present the numerical results.

#### A. System Parameters

We compute the channel gains of the source-relay (SR) and relay-destination (RD) links using the standard path loss model

\[ h_L^2 = \left( \frac{c}{f_c \lambda} \right)^2 d_L^\gamma, \quad \text{for } L \in \{SR, RD\}, \]

(37)

where \( c \) is the speed of light, \( f_c \) is the carrier frequency, \( d_L \) is the distance between the transmitter and the receiver of link \( L \), and \( \gamma \) is the path loss exponent. For the numerical examples in this section, we assume \( \gamma = 3 \), \( d_{SR} = 500 \)m, and \( d_{RD} = 300 \)m. Moreover, we assume a carrier frequency of \( f_c = 2.4 \) GHz. Furthermore, we assume that the noise power per Hz is \(-170\) dBm, which for 200 kHz leads to a total noise power of \(2 \times 10^{-15}\) Watt. Finally, the normalized self-interference amplification factor, \( \alpha \), is computed as \( \alpha = \hat{\alpha} / h_{SR}^2 \), where \( \hat{\alpha} \) is the self-interference
amplification factor. For our numerical results, we will assume that the self-interference amplification factor \( \hat{\alpha} \) ranges from \(-110 \, \text{dB} \) to \(-140 \, \text{dB} \), hence, the self-interference suppression factor, \( 1/\hat{\alpha} \), ranges from \(110 \, \text{dB} \) to \(140 \, \text{dB} \). We note that self-interference suppression schemes that suppress the self-interference by up to \(110 \, \text{dB} \) in certain scenarios are already available today [11]. Given the current research efforts and the steady advancement of technology, suppression factors of up to \(140 \, \text{dB} \) in certain scenarios might be possible in the near future.

B. Benchmark Schemes

**Benchmark Scheme 1 (Ideal FD Transmission without Residual Self-Interference):** The idealized case is when the relay can cancel all of its residual self-interference. For this case, the capacity of the Gaussian two-hop FD relay channel without self-interference is given in [2] as

\[
C_{\text{FD,Ideal}} = \min \left\{ \frac{1}{2} \log_2 \left( 1 + \frac{P_S}{\sigma_R^2} \right), \frac{1}{2} \log_2 \left( 1 + \frac{P_R}{\sigma_D^2} \right) \right\}. \tag{38}
\]

The optimal input distributions at source and relay are zero-mean Gaussian with variances \( P_S \) and \( P_R \), respectively.

**Benchmark Scheme 2 (Conventional FD Transmission with Self-Interference):** The conventional FD relaying scheme for the case when the relay suffers from residual self-interference uses the same input distributions at source and relay as in the ideal case when the relay does not suffer from residual self-interference, i.e., the input distributions at source and relay are zero-mean Gaussian with variances \( P_S \) and \( P_R \), respectively, where the relay’s transmit power, \( P_R \), is optimized in the range \( P_R \leq P_H \) such that the achieved rate is maximized. Thereby, the achieved rate is given by

\[
R_{\text{FD,Conv}} = \max_{P_H \geq P_R} \min \left\{ \int_{-\infty}^{\infty} \frac{1}{2} \log_2 \left( 1 + \frac{P_S}{\sigma_R^2 + \alpha x^2_R} \right) \frac{e^{-x_R^2/(2P_R)}}{\sqrt{2\pi P_R}} \, dx_R : \frac{1}{2} \log_2 \left( 1 + \frac{P_R}{\sigma_D^2} \right) \right\}. \tag{39}
\]

**Benchmark Scheme 3 (Optimal HD Transmission):** The capacity of the Gaussian two-hop HD relay channel was derived in [29], but can also be directly obtained from Theorem 2 by letting \( \alpha \rightarrow \infty \). This capacity can be obtained numerically and will be denoted by \( C_{\text{HD}} \). In this case, the optimal input distribution at the relay is discrete. On the other hand, the source transmits using a Gaussian input distribution with constant variance. Moreover, the source transmits only when the relay is silent, i.e., only when the relay transmits the symbol zero, otherwise, the source is silent. Since both source and relay are silent in fractions of the time, the average powers at source and relay for HD relaying are adjusted such that they are equal to the average powers at source and relay for FD relaying, respectively.

**Benchmark Scheme 4 (Conventional HD Transmission):** The conventional HD relaying scheme uses zero-mean Gaussian distributions with variances \( P_S \) and \( P_R \) at source and relay, respectively. However, compared to the optimal HD transmission in [29], in conventional HD transmission, the relay alternates between receiving and transmitting in a codeword-by-codeword manner. As a result, the achieved rate is given by [30]

\[
R_{\text{HD,Conv}} = \max_{t} \min \left\{ \frac{1}{2} \log_2 \left( 1 + \frac{P_S/(1-t)}{\sigma_R^2} \right) ; \frac{t}{2} \log_2 \left( 1 + \frac{P_R/t}{\sigma_D^2} \right) \right\}. \tag{40}
\]

In (40), since source and relay transmit only in \((1-t)\) and \(t\) fraction of the time, the average powers at source and relay are adjusted such that they are equal to the average powers at source and relay for FD relaying, respectively.

**Remark 3:** We note that Benchmark Schemes 1-4 employ DF relaying. We do not consider the rate achieved with amplified-and-forward (AF) relaying because it was shown in [2] that the optimal mode of operation for relays in terms of rate for the class of degraded relay channels, which the investigated two-hop relay channel belongs to, is the DF mode. This means that for the considered two-hop relay channel, the rate achieved with AF relaying will be equal to or smaller than that achieved with DF relaying.

C. Numerical Results

In this subsection, we denote the capacity of the considered FD relay channel, obtained from Theorem 2, by \( C_{\text{FD}} \).

In Fig. 2, we plot the optimal input distribution at the relay, \( p^*(x_R) \), for \( d_{SR} = d_{RD} = 500 \, \text{m} \), \( P_S = P_R = 25 \, \text{dBm} \), and self-interference suppression factor of \( 1/\hat{\alpha} = 130 \, \text{dB} \). As can be seen from Fig. 2, the relay is silent in 40% of the time, and the source transmits only when \( |x_R| < x_{\text{th}} = 0.9312 \). Hence, similar to optimal HD relaying in [29], shutting down the transmitter at the relay in a symbol-by-symbol manner is important for achieving the capacity. This means that in a fraction of the transmission time, the FD relay is silent and effectively works as an HD relay. However, in contrast to optimal HD relaying where the source transmits only when the relay is silent, i.e., only when \( x_R = 0 \) occurs, in FD relaying, the source has more opportunities to transmit since it can transmit also when the relay transmits a symbol whose amplitude is smaller than \( x_{\text{th}} \), i.e., when \(-x_{\text{th}} \leq x_R \leq x_{\text{th}} \) holds. For the example in Fig. 2, the source transmits 96% of the time.

In Fig. 3, we compare the capacity of the considered FD relay channel, \( C_{\text{FD}} \), with the achievable rate for the suboptimal input distribution, \( p_{\text{th}}(x_R) \), given in Section III-E, denoted by \( R_{\text{FD,B}} \), the capacity achieved with ideal FD relaying without residual self-interference, \( C_{\text{FD,Ideal}} \), cf. Benchmark Scheme 1, the rate achieved with conventional FD relaying, \( R_{\text{FD,Conv}} \), cf. Benchmark Scheme 2, the capacity of the two-hop HD relay channel, \( C_{\text{HD}} \), cf. Benchmark Scheme 3, and the rate achieved with conventional HD relaying, \( R_{\text{HD,Conv}} \), cf. Benchmark Scheme 4, for \( d_{SR} = d_{RD} = 500 \, \text{m} \) and a self-interference suppression factor, \( 1/\hat{\alpha} \), of 130 dB as a function of the average source and relay transmit powers \( P_S = P_R \). The figure shows that indeed the achievable rate with the suboptimal input distribution given in Section III-E, \( R_{\text{FD,B}} \), is a tight
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2017.2653112, IEEE Transactions on Communications

Fig. 2. Optimal input distribution at the relay, $p^*(x_r)$, for $d_{SR} = d_{PD} = 500m$, $P_S = P_R = 25 \text{ dBm}$, and self-interference suppression factor, $1/\hat{\alpha} = 130 \text{ dB}$.  

Fig. 3. Comparison of the derived capacity with the rates of the benchmark schemes as a function of the source and relay transmit powers $P_S = P_R$ in dBm for a self-interference suppression factor, $1/\hat{\alpha} = 130 \text{ dB}$.  

lower bound on the capacity $C_{FD}$. Hence, this rate can be used for analytical analysis instead of the actual capacity rate, which is hard to analyze. In addition, the figure shows that for $P_S = P_R > 20 \text{ dBm}$, the derived capacity $C_{FD}$ achieves around 1.5 dB power gain with respect to the rate achieved with conventional FD relaying, $R_{FD,Conv}$, using Benchmark Scheme 2. Also, for $P_S = P_R > 20 \text{ dBm}$, the derived capacity $C_{FD}$ achieves around 5 dB power gain compared to capacity of the two-hop HD relay channel, $C_{HD}$, and around 10 dB power gain compared to rate achieved with conventional HD relaying, $R_{HD,Conv}$.

In Fig. 4, the same parameters as for Fig. 3 are adopted, except that a self-interference suppression factor, $1/\hat{\alpha}$, of 120 dB is assumed. Thereby, Fig. 4 shows that for $P_S = P_R > 20 \text{ dBm}$, the derived capacity $C_{FD}$ achieves around 2 dB gain with respect to capacity of the two-hop HD relay channel, $C_{HD}$, and around 6 dB gain with respect to the rates achieved with conventional FD relaying, $R_{FD,Conv}$, and conventional HD relaying, $R_{HD,Conv}$. In this example, we can see that, due to the strong residual self-interference, the rate achieved with conventional FD relaying is considerably lower than the derived capacity of FD relaying and even the capacity of HD relaying.

**Remark 4:** From Figs. 3 and 4, we observe that the multiplexing gain of the derived capacity is 1/2, i.e., the same as the value for the HD case. In fact, when $P_S = P_R$, the derived capacity for FD relaying achieves only several dB power gain compared to the capacity for HD relaying. This means that for the adopted worst-case linear residual self-interference model, a self-interference suppression factor of 130 dB is too small to yield a multiplexing gain of 1 in the considered range of $P_S = P_R$. Intuitively, this happens since for $P_S = P_R > 15 \text{ dBm}$, the average power of the residual self-interference, $\hat{\alpha}E[X_{FD}^2|\hat{\alpha}]$, exceeds the average power of the Gaussian noise. In fact, in general, for a fixed self-interference suppression factor $1/\hat{\alpha}$ and $P_S = P_R \to \infty$, the power of the residual self-interference at the relay also becomes infinite. As a result, the corresponding multiplexing gain is limited to 1/2.

In Fig. 5, we show the capacity gain of the two-hop FD
relay channel compared to the two-hop HD relay channel as a function of the self-interference suppression factor, $1/\hat{\alpha}$, for different average transmit powers at source and relay $P_S = P_R$ and $d_{SR} = d_{RD} = 500$m$^4$. As can be seen from Fig. 5, for a self-interference suppression factor of 120 dB, we obtain only a 5 percent capacity increase for FD relaying compared to HD relaying. In contrast, for a self-interference suppression factor of 130 dB, we obtain around 10 to 15 percent increase in capacity depending on the average transmit power. A 50 percent increase in capacity is possible if $P_S$ and $P_R$ are larger than 25 dBm and the self-interference suppression factor is larger than 150 dB. However, such large self-interference suppression factors might be difficult to realize in practice.

In Fig. 6, we compare the capacity of the considered FD relay channel, $C_{FD}$, with the capacities of the ideal FD relay channel without self-interference, $C_{FD,\text{Ideal}}$, and the HD relay channel, $C_{HD}$, for $d_{SR} = 500$m, $d_{RD} = 300$m, and $P_R = 25$ dBm as a function of the average transmit power at the source, $P_S$. This models a practical scenario where the transmission from a source, e.g., a base station, is supported by a dedicated low-power FD relay. Different self-interference suppression factors are considered for this scenario. For this example, since the relay transmit power is fixed, the capacity of the relay--destination channel is also fixed to around 1.84 Mbps. As a result, the capacity of the considered relay channel cannot surpass 1.84 Mbps. In addition, it can be observed from Fig. 6 that the derived capacity of the considered FD relay channel, $C_{FD}$, is significantly larger than the capacity of the HD relay channel, $C_{HD}$ when the transmit power at the source is larger than 30 dBm. For example, for 1.5 Mbps, the power gains are approximately 30 dB, 25 dB, 20 dB, and 15 dB compared to HD relaying for self-interference suppression factors of 140 dB, 130 dB, 120 dB, and 110 dB, respectively. This numerical example shows the benefits of using a dedicated low-power FD relay to support a high-power base station.

V. CONCLUSION

We studied the capacity of the Gaussian two-hop FD relay channel with linear residual self-interference. For this channel, we considered the worst-case linear residual self-interference model, and thereby, obtained a capacity which constitutes a lower bound on the capacity for any other linear residual self-interference model. We showed that the capacity is achieved by a zero-mean Gaussian input distribution at the source whose variance depends on the amplitude of the transmit symbols at the relay. On the other hand, the optimal input distribution at the relay is Gaussian only when the relay--destination link is the bottleneck link. Otherwise, the optimal input distribution at the relay is discrete. Our numerical results show that significant performance gains are achieved with the proposed capacity-achieving coding scheme compared to the achievable rates of conventional HD and/or FD relaying. In addition, we proposed a suboptimal input distribution at the relay, which, for the presented numerical examples, achieves rates that are close to the capacity achieved with the optimal input distribution at the relay.

APPENDIX

A. Proof of Theorem 1

We first assume that $p(x_R)$ is discrete. In addition, we assume that $p(x_S|x_R)$ is a continuous distribution, which will turn out to be a valid assumption. Now, from (17), the corresponding maximization problem with respect to $p(x_S|x_R)$ is given by

$$\max_{p(x_S|x_R) \in \mathcal{P}} \sum_{x_R \in \mathcal{X}_R} I(X_S; Y_R|X_R = x_R)p(x_R)$$

Subject to C1:
$$\sum_{x_R \in \mathcal{X}_R} \left[ \int_{x_S} x_S^2 p(x_S|x_R) dx_S \right] p(x_R) \leq P_S. \tag{41}$$

Since $I(X_S; Y_R|X_R = x_R)$ is the mutual information of a Gaussian AWGN channel with noise power $\sigma_R^2 + \alpha x_R^2$, cf. (11), the optimal distribution $p(x_S|x_R)$ that maximizes $I(X_S; Y_R|X_R = x_R)$ is the zero-mean Gaussian distribution with variance $P_S(x_R)$. The variance $P_S(x_R)$ has to satisfy constraint C1 in (41). Hence, to find the variance $P_S(x_R)$, we first substitute $p(x_S|x_R)$ in (41) with the zero-mean Gaussian distribution with variance $P_S(x_R)$. Thereby, we obtain the following optimization problem

$$\max_{P_S(x_R)} \sum_{x_R \in \mathcal{X}_R} \frac{1}{2} \log_2 \left( 1 + \frac{P_S(x_R)}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R)$$

Subject to C1:
$$\sum_{x_R \in \mathcal{X}_R} P_S(x_R)p(x_R) \leq P_S \tag{42}$$

$$C2: P_S(x_R) \geq 0, \forall x_R.$$ 

Since (42) is a concave optimization problem, it can be solved in a straightforward manner using the Lagrangian method, which results in (18). In (18), $x_{th}$ is a Lagrange multiplier which has to be set such that constraint C1 in (42) holds with equality. Inserting (18) into constraint C1 in (42), we obtain (19). Whereas, inserting $P_S(x_R)$ in (18) into the objective function in (42), we obtain (20).

Following a similar procedure as above for the case when $p(x_R)$ is assumed to be continuous, we arrive at the same solution for $P_S(x_R)$ and $\max_{p(x_S|x_R) \in \mathcal{P}} I(X_S; Y_R|X_R)$ as in (18) and (20), respectively, but with the sums replaced by integrals. This concludes the proof.
B. Proof of Theorem 2

Assuming \( p(x_R) \) is discrete, the corresponding capacity expression for the \( p(x_S|x_R) \) given in Theorem 1 is given by

\[
C = \max_{p(x_R) \in \mathcal{P}} \min_{p(x_R) \in \mathcal{P}} \left\{ \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R) \right. \\
\left. : I(X_R;Y_D) \right\}
\]

Subject to C1: \( \sum_{x_R \in X_R} \alpha \max \{0, x_{th}^2 - x_R^2\} = P_S \)

C2: \( \sum_{x_R \in X_R} x_R^2 p(x_R) \leq P_R \) \hspace{1cm} (43)

Using its epigraph form, the optimization problem in (43) can be equivalently represented as

Maximize : \( u \)

Subject to

C1 : \( u - \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R) \leq 0 \)

C2 : \( u - I(X_R;Y_D) \leq 0 \)

C3 : \( \sum_{x_R \in X_R} \alpha \max \{0, x_{th}^2 - x_R^2\} p(x_R) = P_S \)

C4 : \( \sum_{x_R \in X_R} x_R^2 p(x_R) \leq P_R \)

C5 : \( \sum_{x_R \in X_R} p(x_R) - 1 = 0 \). \hspace{1cm} (44)

In the optimization problem (44), constraint C2 is convex with respect to \( p(x_R) \), and constraints C1, C3, C4, and C5 are affine with respect to \( p(x_R) \). Hence, the optimization problem in (44) is a concave optimization problem and can be solved using the Lagrangian method. The Lagrangian function of the optimization problem in (44) is given by

\[
L = u - \xi_1 \left( u - \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R) \right) - \xi_2 (u - I(X_R;Y_D)) - \lambda_1 \left( \sum_{x_R \in X_R} \alpha \max \{0, x_{th}^2 - x_R^2\} p(x_R) - P_S \right) - \lambda_2 \left( \sum_{x_R \in X_R} x_R^2 p(x_R) - P_R \right) - \nu \left( \sum_{x_R \in X_R} p(x_R) - 1 \right),
\]

where \( \xi_1, \xi_2, \lambda_1, \lambda_2, \) and \( \nu \) are Lagrange multipliers corresponding to constraints C1, C2, C3, C4, and C5, respectively.

Due to the KKT conditions, the following has to hold

\[
\xi_1 \left( u - \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R) \right) = 0, \hspace{1cm} (46a)
\]

\[
\xi_2 (u - I(X_R;Y_D)) = 0, \hspace{1cm} (46b)
\]

\[
\lambda_1 \left( \sum_{x_R \in X_R} \alpha \max \{0, x_{th}^2 - x_R^2\} p(x_R) - P_S \right) = 0, \hspace{1cm} (46c)
\]

\[
\lambda_2 \left( \sum_{x_R \in X_R} x_R^2 p(x_R) - P_R \right) = 0, \hspace{1cm} (46d)
\]

\[
\nu \left( \sum_{x_R \in X_R} p(x_R) - 1 \right) = 0. \hspace{1cm} (46e)
\]

Differentiating L with respect to \( u \), we obtain that \( \xi_1 = 1 - \xi_2 = \xi \) has to hold, where \( 0 \leq \xi \leq 1 \). Inserting this into (45), then differentiating with respect to \( p(x_R) \), and equating the result to zero we obtain the following

\[
\xi \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) + (1 - \xi) I'(X_R;Y_D) - \lambda_1 \alpha \max \{0, x_{th}^2 - x_R^2\} - \lambda_2 x_R^2 - \nu = 0, \hspace{1cm} (47)
\]

where \( I'(X_R;Y_D) = \partial I(X_R;Y_D)/\partial p(x_R) \). We note that there are three possible solutions for (47) depending on whether \( \xi = 1, \xi = 0, \) or \( 0 < \xi < 1 \), respectively. In the following, we analyze these three cases.

Case 1: Let us assume that \( \xi = 1 \) holds. Then, from (46), we obtain that

\[
u < I(X_R;Y_D) \quad \text{and} \quad u = \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R), \hspace{1cm} (48)
\]

which means that for the optimal \( p(x_R) \) the following holds

\[
I(X_R;Y_D) \bigg|_{p(x_R) = p^*(x_R)} > \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p^*(x_R). \hspace{1cm} (49)
\]

The optimal \( p^*(x_R) \) in this case has to maximize the right hand side of (49), i.e.,

\[
\sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R). \hspace{1cm} (50)
\]

It turns out that the optimal \( p(x_R) \) which maximizes (50) is \( p^*(x_R) = \delta(x_R) \), i.e., the relay is always silent and never transmits. However, if we insert \( p^*(x_R) = \delta(x_R) \) in \( I(X_R;Y_D) \) in (49), we obtain the following contradiction

\[
I(X_R;Y_D) \bigg|_{p(x_R) = \delta(x_R)} = 0 > \frac{1}{2} \log_2 (1 + P_S/\sigma_R^2) > 0. \hspace{1cm} (51)
\]

Hence, \( \xi = 1 \) is not possible. The only remaining possibilities are \( \xi = 0 \) and \( 0 < \xi < 1 \).
Case 2: Let us assume that $\xi = 0$ holds. Then, from (46), we obtain that

$$u = I(X_R; Y_D)$$

and

$$u < \sum_{x_R \in \mathcal{X}_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R),$$

(52)

has to hold, which means that for the optimal $p(x_R)$ the following holds

$$I(X_R; Y_D) \bigg|_{p(x_R) = p^*(x_R)} < \sum_{x_R \in \mathcal{X}_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p^*(x_R).$$

(53)

The optimal $p(x_R)$ in this case is the one which maximizes the left hand side of (53), i.e., maximizes $I(X_R; Y_D)$. Since the relay-destination link is an AWGN channel, $I(X_R; Y_D)$ is maximized for $p^*(x_R)$ being the zero-mean Gaussian distribution with variance $P_R$. As a result, the capacity is given by

$$I(X_R; Y_D) \bigg|_{p(x_R) = p^*(x_R)} = \frac{1}{2} \log_2 \left( 1 + \frac{P_R}{\sigma_D^2} \right).$$

(54)

Hence, (54) is the capacity if and only if (iff) after substituting $p^*(x_R)$ with the zero-mean Gaussian distribution with variance $P_R$, (53) holds, i.e., (21) holds.

Case 3: Let us assume that $0 < \xi < 1$. Then, from (46), we obtain that

$$u = I(X_R; Y_D)$$

and

$$u < \sum_{x_R \in \mathcal{X}_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R),$$

(55)

which means that for the optimal $p(x_R)$, the following holds

$$I(X_R; Y_D) \bigg|_{p(x_R) = p^*(x_R)} = \sum_{x_R \in \mathcal{X}_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p(x_R).$$

(56)

For $0 < \xi < 1$, we can find the optimal distribution $p^*(x_R)$ as the solution of (47). To this end, we need to compute $I'(X_R; Y_D)$. Since for the AWGN channel, $I(X_R; Y_D) = H(Y_D) - H(Y_D|X_R)$, where $H(Y_D|X_R) = \frac{1}{2} \log_2 (2\pi e \sigma_D^2)$ holds, we obtain that $I'(X_R; Y_D) = H'(Y_D)$. On the other hand, $H'(Y_D)$ for the AWGN channel is found as

$$H'(Y_D) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \sigma_D^2}} \exp \left( -\frac{(y_D - x_R)^2}{2\sigma_D^2} \right) \log_2(p(y_D)) dy_D - \frac{1}{\ln(2)}.$$  

(57)

Inserting (57) into (47), we obtain

$$\frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max \{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) - (1 - \xi)$$

$$\times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \sigma_D^2}} \exp \left( -\frac{(y_D - x_R)^2}{2\sigma_D^2} \right) \log_2(p(y_D)) dy_D$$

$$- (1 - \xi) - \frac{1}{\ln(2)} - \lambda_1 \alpha \max \{0, x_{th}^2 - x_R^2\} - \lambda_2 x_R^2 - \nu = 0.$$  

(58)

Hence, the optimal $p(x_R)$ has to produce a $p(y_D)$ for which (58) holds. In Appendix C, we prove that (58) cannot hold if $p(x_R)$ is a continuous distribution and that (58) can hold if $p(x_R)$ is a discrete distribution since then it has to hold only for the discrete values $x_R \in \mathcal{X}_R$.

Remark 5: Although we derived (47) assuming that $p(x_R)$ was discrete, we would have arrived at the same result if we had assumed that $p(x_R)$ was a continuous distribution. To do so, we first would have to replace the sums in the optimization problem in (44) with integrals with respect to $x_R$. Next, in order to obtain the stationary points of the corresponding Lagrangian function, instead of the ordinary derivative, we would have to take the functional derivative and equate it to zero. This again would have led to the identity in (47). Hence, the conclusions drawn from the Lagrangian and (47) are also valid when $p(x_R)$ is a continuous distribution.

C. Proof That $p(x_R)$ is Discrete When $0 < \xi < 1$

This proof is based on the proof for the discreteness of a distribution given in [35]. Furthermore, similar to [35], to simplify the derivation of the proof, we set $\sigma_D^2 = 1$.

First, we decompose the integral in (58) using Hermitian polynomials. To this end, we define

$$\log_2(p(y_D)) = \sum_{m=0}^{\infty} c_m H_m(y_D),$$

(59)

where the $c_m$, $\forall m$, are constants and $H_m(y_D)$, $\forall m$, are Hermitian polynomials, see [35]. Note that $\ln(p(y_D))$ is square integrable with respect to $e^{-\frac{y_D^2}{2\sigma_D^2}}$ and hence can be decomposed using a Fourier-Hermite series decomposition, see [35]. Then, the integral in (57) with $\sigma_D^2 = 1$, can be written as

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y_D - x_R^2)}{2\sigma_D^2}} \log_2(p(y_D)) dy_D$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y_D^2}{2\sigma_D^2}} \log_2(p(y_D)) dy_D$$

$$= \sum_{m=0}^{\infty} c_m x_R^m,$$

(60)

where (a) is obtained by inserting (59) and using the generating function of Hermitian polynomials, given by

$$e^{(-\frac{1}{2} + tx)} = \sum_{m=0}^{\infty} H_m(x) \frac{t^m}{m!}.$$  

(61)
Furthermore, \((b)\) in \((60)\) follows since Hermitian polynomials are orthogonal with respect to the weight function \(\omega(x) = e^{-x^2}, \) i.e.,
\[
\int_{-\infty}^{\infty} H_m(x)H_n(x)\omega(x)dx = \begin{cases} m!\sqrt{2\pi} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{62}
\]
holds. By inserting \((60)\) into \((58)\), we obtain
\[
(1 - \xi) \sum_{m=0}^{\infty} c_m x_R^m = \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max\{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) - \lambda_1 \alpha \max\{0, x_{th}^2 - x_R^2\} - \lambda_2 x_R^2 - \frac{1}{\ln(2)} - \nu.
\tag{63}
\]

Now, we have two cases for \(|x_R|\), one when \(|x_R| \geq x_{th}\) and the other one when \(|x_R| < x_{th}\). Also, we have two cases for \(\lambda_2\), one when \(\lambda_2 > 0\) (constraint \(C2\) in \((43)\) holds with equality) and the other are when \(\lambda_2 = 0\) (constraint \(C2\) in \((43)\) does not hold with equality). The resulting four cases are analyzed in the following.

**Case 1:** If \(|x_R| \geq x_{th}\) and \(\lambda_2 > 0\) hold, then \((63)\) simplifies to
\[
\sum_{m=0}^{\infty} c_m x_R^m = - \frac{\lambda_2}{1 - \xi} x_R^2 - \frac{1}{\ln(2)} - \nu.
\tag{64}
\]
Comparing the exponents in \((64)\), we obtain
\[
c_0 = - \frac{1}{\ln(2)} - \nu; \quad c_1 = 0;
\]
\[
c_0 = \frac{\lambda_2}{1 - \xi}; \quad c_n = 0, \quad \forall n > 2.
\tag{65}
\]
Inserting \((65)\) into \((59)\), we obtain \(p(y_d)\) as
\[
p(y_d) = e^{\ln(2)(c_0 H_0(y_d) + c_2 H_2(y_d))} \overset{(a)}{=} e^{\ln(2)(c_0 - c_2)} e^{\ln(2)c_2 y_d^2},
\tag{66}
\]
where \((a)\) follows from \(H_0(y_d) = 1\) and \(H_2(y_d) = y_d^2 - 1\). This solution for \(p(y_d)\) can be a valid probability density function only for \(c_2 < 0\), which yields a Gaussian distribution for \(p(y_d)\). Now, for \(Y_D\) to be Gaussian distributed and \(Y_D = X_R + N_D\) to hold, where \(N_D\) is also Gaussian distributed, \(p(x_R)\) also has to be Gaussian distributed. However, since the Gaussian distribution is unbounded in \(x_R\), the Gaussian distribution \(p(x_R)\) cannot hold only in the domain \(|x_R| \geq x_{th}\) but has to hold in the entire domain \(|x_R| \leq \infty\). Hence, we have to see whether a Gaussian \(p(x_R)\) is also optimal for \(|x_R| < x_{th}\). If we obtain that \(p(x_R)\) is not Gaussian for \(|x_R| < x_{th}\), then \(p(x_R)\) can only be discrete in the domain \(|x_R| \geq x_{th}\) for any \(x_{th} > 0\).

**Case 2:** If \(|x_R| < x_{th}\) and \(\lambda_2 > 0\) hold, \((63)\) simplifies to
\[
\sum_{m=0}^{\infty} c_m x_R^m = \frac{\xi}{1 - \xi} \log_2 \left( 1 + \frac{\alpha x_{th}^2 - x_R^2}{\sigma_R^2 + \alpha x_R^2} \right) - \frac{\lambda_1 \alpha + \lambda_2}{1 - \xi} x_R^2 - \frac{1}{1 - \xi} (1/\ln(2)) + \nu + \lambda_1 \alpha x_{th}^2.
\tag{67}
\]
We now represent the \(\log_2(\cdot)\) function in \((67)\) using a Taylor series expansion as
\[
\log_2 \left( 1 + \frac{\alpha x_{th}^2 - x_R^2}{\sigma_R^2 + \alpha x_R^2} \right) = \sum_{n=0}^{\infty} (-1)^n a_n x_R^{2n},
\tag{68}
\]
where \(a_n > 0, \forall n\), and the exact (positive) values of these coefficients are not important for this proof. Inserting \((68)\) into \((67)\), we obtain
\[
\sum_{m=0}^{\infty} c_m x_R^m = \frac{\xi}{1 - \xi} \log_2 \left( 1 + \frac{\lambda_1 \alpha + \lambda_2}{1 - \xi} x_R^2 - \frac{1}{1 - \xi} (1/\ln(2)) + \nu + \lambda_1 \alpha x_{th}^2. \right)
\tag{69}
\]
Comparing the exponents on the left hand side and the right hand side of \((69)\), we can find \(c_m\) as
\[
c_m = \begin{cases} 
\frac{\xi}{1 - \xi} \log_2 \left( \frac{1}{1/\ln(2) + \nu + \lambda_1 \alpha x_{th}^2} \right) & \text{if } m = 0 \\
\frac{\xi}{1 - \xi} \log_2 \left( \frac{1}{1/\ln(2) + \nu + \lambda_1 \alpha x_{th}^2} \right) & \text{if } m = 1 \\
\frac{\xi}{1 - \xi} \log_2 \left( \frac{1}{1/\ln(2)} \right) & \text{if } m > 1 
\end{cases}
\tag{70}
\]
Inserting \((70)\) into \((59)\), we obtain \(p(y_d)\) as
\[
p(y_d) = e^{\ln(2)\sum_{n=0}^{\infty} c_n H_n(y_d)}
\]
\[
= e^{\ln(2)\sum_{n=0}^{\infty} q_n y_d^2} = \prod_{n=0}^{\infty} e^{\ln(2)q_n y_d^2},
\tag{71}
\]
where \(q_n\) are known non-zero constants. Now, since \(q_n > 0\) for some \(n \to \infty\), \(p(y_d)\) in \((71)\) cannot be a valid distribution since \(p(y_d)\) becomes unbounded. As a result, \(p(x_R)\) has to be discrete in the domain \(|x_R| < x_{th}\). Consequently, \(p(x_R)\) also has to be discrete in the domain \(|x_R| \geq x_{th}\). This concludes the proof for the case when \(\lambda_2 > 0\). Following a similar procedure for \(\lambda_2 = 0\) as for the case when \(\lambda_2 > 0\), we obtain that again \(p(x_R)\) has to be discrete in the entire domain of \(x_R\).

On the other hand, \(p^*(x_R)\) has to be symmetrical with respect to \(x_R = 0\). To prove this, assume that we have an unsymmetrical \(p(x_R)\), denoted by \(p_u(x_R)\), with only one unsymmetrical mass point \(x_{Ru}\) which has probability \(p_{Ru}\). Now, let us construct a new, symmetrical \(p(x_R)\), denoted by \(p_s(x_R)\), by making \(p_{Ru}(x_R)\) symmetrical. In particular, in \(p_{Ru}(x_R)\), we first reduce the probability of the mass point \(x_{Ru}\) to \(p_{Ru}/2\). Next, we add the mass point \(-x_{Ru}\) to \(p_{Ru}(x_R)\) and set its probability to \(p_{Ru}/2\). Now, it is clear that the average power of the relay is identical for both \(p_u(x_R)\) and \(p_s(x_R)\). On the other hand, by making \(p(x_R)\) symmetrical, we have increased the entropy of \(X_R\), i.e., \(H(X_R)p(x_R) = p_{Ru}(x_R) \leq H(X_R)p(x_R) = p_s(x_R)\) holds. Consequently, we have increased the differential entropy of \(Y_D\), i.e., \(h(Y_D)p(x_R) = p_{Ru}(x_R) \leq h(Y_D)p(x_R) = p_s(x_R)\) holds. Now, since for the AWGN channel \(h(Y_D|X_R)\) is independent of \(p(x_R)\), it follows that \(I(X_R;Y_D)p(x_R) = p_{Ru}(x_R) \leq I(X_R;Y_D)p(x_R) = p_s(x_R)\) holds. This concludes the proof of the symmetry of \(p^*(x_R)\).
D. Proof of Lemma 1

Here, we only prove the non-trivial case when (21) does not hold. The trivial case is identical to the case without self-interference and its achievability is shown [2].

Let us assume that condition (21) does not hold. Then, according to Theorem 1, $p(x_R)$ is discrete and the capacity $C$ is given in (24). Moreover, for the considered coding scheme, $R$ satisfies the following

$$R < C = \max_{p(x_R)} I(X_S; Y_R | X_R)$$

$$= \sum_{x_R \in X_R} \max_{p(x_R)} I(X_S; Y_R | X_R = x_R) p^*(x_R)$$

$$\leq \sum_{x_R \in X_R} \max_{p(x_R)} I(X_S; Y_R | X_R = x_R = 0) p^*(x_R)$$

$$= \max_{p(x_R)} I(X_S; Y_R | X_R = 0) p_T$$

$$= \frac{1}{2} \log_2 \left( 1 + \frac{\alpha x_{th}^2}{\sigma_R^2} \right) p_T,$$

where (a) follows since for the considered coding scheme the source is silent when $|X_R| \geq x_{th}$ and as a result $I(X_S; Y_R | X_R = x_R) = 0$ for $|X_R| \geq x_{th}$, (b) follows since, for the considered relay channel, $I(X_S; Y_R | X_R = x_R)$ is maximized for $x_R = 0$, because in that case there is no residual self-interference at the relay, and (c) follows from (20).\[593/\]

Now, note that for the considered coding scheme in time slot 1, the source-relay channel can be seen as an AWGN channel with a fixed channel gain $\sqrt{P_S(x_R = 0)} = \alpha x_{th}$ and AWGN with variance $\sigma_R^2$ which is used $k_{PT}$ times. Hence, any codeword selected uniformly from a codebook comprised of $2^{kR}$ Gaussian distributed codewords, where each codeword is comprised of $k_{PT}$ symbols, with $k \to \infty$ and $R$ satisfying

$$kR/(k_{PT}) < \frac{1}{2} \log_2 \left( 1 + \frac{\alpha x_{th}^2}{\sigma_R^2} \right),$$

can be successfully decoded at the relay using a jointly-typical decoder, see [31]. Noting that the proposed coding scheme satisfies the properties outlined above, we can conclude that the codeword transmitted in time slot 1 can be decoded successfully at the relay.

E. Proof of Lemma 2

Again, we only prove the non-trivial case when (21) does not hold.

In time slot $b$, for $2 \leq b \leq N$, the source-relay channel can be seen equivalently as an AWGN channel with states $X_R$, where a different state $X_R = x_R$ produces a different channel gain and a different noise variance. In particular, for channel state $X_R = x_R$, the channel gain of the equivalent AWGN channel is $\sqrt{P_S(x_R)}$ and the variance of the AWGN is $\sigma_R^2 + \alpha x_R^2$. Moreover, for this equivalent AWGN channel with states, the source has to transmit unit-variance symbols in order for the average power constraint of the original source-relay channel, given by $E[X_R^2] \leq P_S$, to be satisfied. Furthermore, for the equivalent AWGN channel with states, note that both source (i.e., transmitter) and relay (i.e., receiver) have CSI in each channel use and thereby know that the channel gain and the noise variance in channel use $j$ will be $\sqrt{P_S(x_{R,j})}$ and $\sigma_R^2 + \alpha x_{R,j}^2$, respectively. Now, instead of deriving a capacity-achieving coding scheme for the original source-relay channel, we can find equivalently a capacity-achieving coding scheme for the equivalent AWGN channel\[593/\] with states. To this end, we will first find the capacity of an “auxiliary AWGN channel”, using results which are already available in the literature. Then, we will modify the capacity-achieving coding scheme of the “auxiliary AWGN channel” in order to obtain a capacity-achieving coding scheme for the equivalent AWGN channel with states.\[593/\]

The “auxiliary AWGN channel” is identical to the equivalent AWGN channel but without CSI at the source (i.e., transmitter). The channel coding scheme that achieves the capacity of the “auxiliary AWGN channel” in $k \to \infty$ channel uses is the following, see [36], [37] for proof. The codebook is comprised of $2^{kR}$ codewords, where each codeword is comprised of $k$ symbols and each symbol is generated independently according to the zero-mean unit-variance Gaussian distribution. Moreover, the parameter $R$ of the channel code has to satisfy

$$R < \max_{p(x_R)} I(X_S'; Y_D | X_R) \bigg|_{E[X_R^2] = 1}$$

$$= \sum_{x_R \in X_R} \max_{p(x_R)} I(X_S'; Y_D | X_R = x_R) p^*(x_R)$$

$$\leq \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{P_S(x_R)}{\sigma_R^2 + \alpha x_R^2} \right) p^*(x_R),$$

$$\leq \sum_{x_R \in X_R} \frac{1}{2} \log_2 \left( 1 + \frac{\alpha \max\{0, x_{th}^2 - x_R^2\}}{\sigma_R^2 + \alpha x_R^2} \right) p^*(x_R),$$

where $X_R$ is the input at the source of the “auxiliary AWGN channel”, (a) follows due to the unit-variance constraint $E[X_R^2] = 1$ and since for each state $X_R = x_R$ the channel is AWGN with channel gain $\sqrt{P_S(x_{R,j})}$ and noise variance $\sigma_R^2 + \alpha x_{R,j}^2$, and (b) follows from (18). Any codeword selected uniformly from this codebook and transmitted in $k$ channel uses can be successfully decoded at the relay (i.e., receiver) using a jointly typical decoder, see [36], [37], [31]. Now, for the “auxiliary AWGN channel” note that the source transmits a symbol during channel uses for which the channel gain is zero, i.e., $\sqrt{P_S(x_R)} = 0$ holds. Obviously, the symbols transmitted when $\sqrt{P_S(x_R)} = 0$ do not reach the relay due
to the zero channel gain, i.e., the relay receives only noise during these channel uses.

Now, for the equivalent AWGN channel, we can use the same coding scheme as for the “auxiliary AWGN channel”, but, since in this case the source has CSI, the source can choose not to transmit during a channel use for which $\sqrt{P_S(x_{bT})} = 0$ holds. Moreover, since the source has knowledge that $\sqrt{P_S(x_{bT})} > 0$ holds in a $p_T$ fraction out of the $k$ channel uses, the source can reduce the length of the codewords from $k$ to $kp_T$. Thereby, the channel code for the equivalent AWGN channel has a codebook comprised of $2^{kp_T}$ Gaussian distributed codewords, where each codeword is comprised of $kp_T$ symbols. Moreover, for the equivalent AWGN channel, the source is silent for states for which $\sqrt{P_S(x_{bT})} = 0$ holds, i.e., $|x_{bT}| \geq x_{th}$ holds, and transmits a symbol from the selected codeword only when $\sqrt{P_S(x_{bT})} > 0$, i.e., $|x_{bT}| < x_{th}$ holds, which is exactly the proposed scheme. Hence, for the proposed scheme, we can conclude that the codeword transmitted in time slot $b$, for $2 \leq b \leq N$, can be decoded successfully at the relay.

REFERENCES


Nikola Zlatanov (S’06, M’15) was born in Macedonia. He received the Dipl.Ing. and Master degree in electrical engineering from Ss. Cyril and Methodius University, Skopje, Macedonia in 2007 and 2010, respectively, and his PhD degree from the University of British Columbia (UBC) in Vancouver, Canada in 2015. He is currently a Lecturer (Assistant Professor) in the Department of Electrical and Computer Systems Engineering at Monash University in Melbourne, Australia. His current research interests include wireless communications and information theory. Dr. Zlatanov received several scholarships/awards for his work including UBC’s Four Year Doctoral Fellowship in 2010, UBC’s Killam Doctoral Scholarship and Macedonia’s Young Scientist of the Year in 2011, the Vanier Canada Graduate Scholarship in 2012, best journal paper award from the German Information Technology Society (ITG) in 2014, and best conference paper award at ICNC in 2016. Dr. Zlatanov serves as an Editor of IEEE Communications Letters. He has been a TPC member of various conferences, including Globecom, ICC, VTC, and ISWCS.

Robert Schober (S’98, M’01, SM’08, F’10) was born in Neuenstadt, Germany, in 1971. He received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from the University of Erlangen-Nuernberg in 1997 and 2000, respectively. From May 2001 to April 2002 he was a Postdoctoral Fellow at the University of Toronto, Canada, sponsored by the German Academic Exchange Service (DAAD). From 2002 to 2011, he was a Professor and Canada Research Chair at the University of British Columbia (UBC), Vancouver, Canada. Since January 2012 he is an Alexander von Humboldt Professor and the Chair for Digital Communication at the Friedrich Alexander University (FAU), Erlangen, Germany. His research interests fall into the broad areas of Communication Theory, Wireless Communications, and Statistical Signal Processing.

Dr. Schober received several awards for his work including the 2002 Heinz Maier-Leibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, a 2011 Alexander von Humboldt Professorship, and a 2012 NSERC E.W.R. Steacie Fellowship. In addition, he has received several best paper awards for his research. Dr. Schober is a Fellow of the Canadian Academy of Engineering and a Fellow of the Engineering Institute of Canada. From 2012 to 2015, he served as Editor-in-Chief of the IEEE Transactions on Communications and since 2014, he is the Chair of the Steering Committee of the IEEE Transactions on Molecular, Biological and Multiscale Communication. Furthermore, he is a Member at Large of the Board of Governors of the IEEE Communications Society.

Erik Sippel was born in Frth, Germany on February 21th 1991. He received the M.Sc. degree in electronic engineering from Friedrich Alexander University (FAU), Erlangen, Germany in 2015 and is currently working towards the Ph.D. degree at FAU, Erlangen. He has joined the Institute of Microwaves and Photonics at FAU in 2016. His research interests are radar for near field localization, data transmission, and time interleaved analog to digital conversion.

Vahid Jamali (S’12) was born in Fasa, Iran, in 1988. He received his B.S. and M.S. degrees both with first class honors in electrical engineering from K. N. Toosi University of Technology (KNTU), in 2010 and 2012, respectively. Currently, he is working toward his Ph.D. degree at the Friedrich-Alexander University (FAU), Erlangen, Germany. His research interests include wireless communications, molecular communications, multiuser information theory, free space optical communications, cognitive radio network, LDPC codes, and optimization theory. He served as member of Technical Program Committees for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, special issue on “Recent Advances in Heterogeneous Cellular Networks” in 2015 and International Conference on Computing, Networking, and Communication (ICNC) in 2015 and 2016. He received best paper award from IEEE International Conference on Communications (ICC) 2016, was awarded as “Exemplary Reviewer” of the IEEE COMMUNICATIONS LETTERS in 2014, and received student travel grants for “SP Coding and Information School,” Sao Paulo, Brazil and “Training School on Optical Wireless Communications,” Istanbul, Turkey in 2015.