Distributed Space–Time Coding for Regenerative Relay Networks

Gesualdo Scutari, Student Member, IEEE, and Sergio Barbarossa, Member, IEEE

Abstract—Cooperation among mobile users (MUs) in a wireless network can be very useful to reduce the total radiated power necessary to insure the delivery of the information with the desired quality of service. A systematic framework for achieving such a gain consists in making the cooperating nodes act as the antennas of a virtual transmit array, operating according to a distributed space–time coding (DSTC) strategy. However, cooperation implies the allocation of dedicated resources, typically power and time slots, for the exchange of data between source and intermediate nodes (relays). It is then necessary to design the system properly to make possible a final net gain, taking into account all resources involved in the communication. In this paper, we consider regenerative relays and analyze the effect of intermediate decision errors at the relay nodes. We derive the optimal maximum-likelihood (ML) detector, at the final destination, in case of binary phase-shift keying (BPSK) transmission, and a suboptimal scalar detector, whose bit-error rate (BER) is expressed in (approximate) closed form. Since with DSTC the transmit antennas are not colocated, we show how to allocate the power among source and relay terminals in order to minimize the average BER at the final destination. Finally, we compare alternative cooperation and decoding strategies.

Index Terms—Cooperative communications, distributed space–time coding (DSTC), relay networks.

I. INTRODUCTION

COOPERATION among radio nodes in a wireless network can improve performance considerably with respect to the noncooperative case. Multihop packet radio networks based on relaying have been in fact proposed and analyzed for a long time (see, e.g., [27]). Fundamental theorems on the capacity of relay networks were established in [8] and [13]. Relaying can be very useful to combat shadowing effects, as shown by Eminian and Kaveh in [12]. Quite recently, Sendonaris et al. showed that cooperation among users can enlarge the capacity region of an uplink multiuser channel [22]–[24]. Laneman and Wornell showed, in a series of papers proposing alternative strategies, that cooperation can induce a spatial diversity gain, which can drastically reduce the power necessary to guarantee a desired bit-error rate (BER) at the final destination [17]–[19]. One of the ideas that proved to be more fruitful to design novel cooperation strategies is the idea of a virtual array (see, e.g., [2], [4], [9], [10], and [18]), where the antennas of cooperating nodes are seen as elements of a virtual array. This conceptual link is particularly important as it allows us to design cooperation protocols exploiting the powerful tools of space–time coding (STC). This strategy is known as distributed STC (DSTC), as the encoding operation is distributed among the cooperating nodes. The idea of cooperation through DSTC has been analyzed in several recent works, e.g., [2], [14], [18], [21], and [23]–[26]. Most of these works assumed that the relays had none of their own information to transmit. Hunter and Nosratinia considered a slightly different scenario, where two users cooperate with each other to transmit towards a common destination, and they both have data to transmit [14]–[16]. Within the DSTC perspective, conventional relaying can be seen as a kind of space–time repetition coding, where packets are transmitted from different points in space, at different times. But since repetition coding is known for being suboptimal, it is reasonable to expect an appreciable gain from the use of more appropriate DSTC strategies.

However, cooperation among radio nodes requires dedicated resources, typically power and time slots. The most critical aspect of DSTC is then the balance between the waste of resources necessary to establish a cooperative communication link and the advantages coming from cooperation. In the most typical implementation of a cooperative communication system, based on a time-division duplexing (TDD) strategy, there are three kinds of time slots dedicated to the following links: 1) from mobile users (MUs) to the access point (AP); 2) from the AP to the MUs; and 3) from MUs to other MUs. The first two slots are the conventional multiple access (uplink) and broadcast (downlink) channels, respectively, whereas the last one is the slot used for the exchange of data between sources and relays directly. To give a general formulation, valid for both uplink and downlink channels, we will denote by $S$ the source of information, whether user or AP, whereas $R$ is a relay, and $D$ is the final destination. Clearly, the insertion of the $S/R$ slot for the link between $S$ and $R$ (or vice versa) induces a rate loss with respect to conventional (nonrelay) networks. To reduce this loss, the $S/R$ slot must be necessarily shared among several pairs of sources and relays. In other words, the system must allow for spatial reuse of the $S/R$ slot. However, this implies that the $S/R$ link is subject to interference. The situation is pictorially sketched in Fig. 1, where there is an AP at the center of the cell and several sources and relays scattered around randomly. The sources transmitting to their relays at the same time interfere with each other. To limit the interference, the most crucial step is the rule that associates a relay to a source. A radio node should be selected from a given source as a relay...
only if the signal-to-noise-plus-interference ratio (SNIR) at the relay, relative to that given source, exceeds a given threshold. For example, without interference, the source \( S_2 \) in Fig. 1 would probably choose as relay \( R_3 \) as it is its nearest relay. However, because of interference, \( S_2 \) would more likely choose relay \( R_2 \) instead. Most available works on DSTC assume that a radio node acts as a relay only if it is error free. This simplifies the analysis, but it is clearly a too-strong assumption. This requirement would in fact prevent the use of a relay also in cases where relaying could still be beneficial, even though not necessarily error free. One important exception to the error-free assumption is [17], where the relay was allowed to make decision errors and the statistics of the relay decision errors were explicitly incorporated in the final detector. An elegant closed-form expression was derived in [17] for the maximum likelihood (ML) at the final destination. However, the coding strategy of [17] was essentially time repetition coding, as the information was sent over nonoverlapping time slots,\(^1\) and then suboptimal.

In this paper, we remove the assumption that the relay be error free and we determine the minimum SNIR that can be tolerated at the relay, while still guaranteeing a sufficient cooperation gain. Different from [17], we assume a DSTC scheme where the available power is optimally distributed among source and relay while they transmit to the destination. We compare the cooperative and noncooperative schemes, assuming the same overall radiated power (incorporating all time slots and links). We propose an optimal ML decoder, which exploits the knowledge of the error statistics at the relay, and a suboptimal decoder. The decoding errors at the relay induce an interfering term at the final destination, whose structure depends on the adopted space–time coding technique. We exploit such a structure to derive a closed-form expression for the final BER of the suboptimal decoder. Then, we show how to distribute the power between \( S \) and \( R \) in order to minimize the final average BER. Finally, we show the balance between the improvement due to the diversity gain offered by DSTC and the rate loss due to the insertion of the \( S/R \) slot.

II. DISTRIBUTED SPACE–TIME BLOCK CODING

We illustrate the proposed transmission protocol by referring to a TDD scheme, but the same considerations could apply to an frequency-division duplexing (FDD) mode. In a TDD system, each frame is subdivided in consecutive time slots: In the first slot, \( S \) transmits and \( R \) receives; in the second slot, \( S \) and \( R \) transmit simultaneously; \( D \) can be in a listening mode in both time slots or only in the second slot. What we propose in this work is that in the second time slot, similar to [3] and [19], \( S \) and \( R \) transmit using a block Alamouti scheme (see, e.g., [1] and [5]), as if they were the two antennas of a single node. The extension to other space–time coding techniques is almost straightforward (see, e.g., [6]). What is important to emphasize here are the differences between DSTC and conventional STC. Different from common STC, with DSTC: 1) regenerative relays might make decision errors, so that the symbols transmitted from \( R \) could be affected by errors; 2) the links between \( S \) and \( D \) and between \( R \) and \( D \) do not have the same statistical properties, in general; 3) even if \( S \) and \( R \) are synchronous, their packets might arrive at \( D \) at different times, as \( S \) and \( R \) are not collocated. In this paper, we will address all these problems specifically.

We describe our proposed solution within the following setup.

a1) All channels are finite-impulse response (FIR) of (maximum) order \( L_h \) and time invariant over at least a pair of consecutive blocks.
a2) The channel coefficients are independent identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance \( 1/d^2 \), where \( d \) is the link length.
a3) The information symbols are i.i.d. binary phase-shift keying (BPSK) symbols that may assume the values \( A \) or \(-A \) with equal probability.\(^2\)
a4) The received data are degraded by additive white Gaussian noise (AWGN).
a5) The channels are perfectly known at the receive sides and are unknown at the transmit side.
a6) The transmission scheme for all terminals is a block strategy, where each block is composed of \( M \) symbols, incorporating a cyclic prefix (CP) of length \( L \) equal to the sum of the relative delay with which packets from \( S \) and \( R \) arrive at \( D \) plus the maximum channel order \( L_h \).

We describe our distributed space–time protocol assuming that each terminal is equipped with one antenna, but the extension to multiple receive antennas at the destination is

\(^1\) Instead of time, one could use the frequency domain to separate the data coming from \( R \) and \( S \), but in any case, the strategy of [17] was based on the orthogonality between the source–relay and the relay–destination channels.

\(^2\) Assumption a3) is made only for simplifying our derivations, but there is no restriction to use higher order constellations.
IFFT matrix with lant Toeplitz matrices, with entries \( H \) denotes the pseudoinverse of \( h \)lowing notation. We denote with \( S \) and \( R \) the impulse responses between \( S \) and \( D \), \( S \) and \( R \), and \( R \) and \( D \), respectively, during the time slot index with \( \kappa_n \); \( i \) is the block index. Each block of symbols \( s(i) \) has size \( M \) and it is linearly encoded, so as to generate the \( N \)-size vector \( x_s(n) := Fs(n) \), where \( F \) is the \( N \times M \) precoding matrix. A CP of length \( L \geq L_h \) is inserted at the beginning of each block, to facilitate elimination of interblock interference (IBI), synchronization, and channel equalization at the receiver. \( A^\dagger \) denotes the pseudoinverse of \( A : \mathbb{R} \rightarrow \mathbb{R} \), \( A \) indicates the real part of \( A \); when applied to a vector, \( \mathbb{R} \rightarrow \mathbb{R} \) is the vector whose entries are the real part of the entries of \( x \).

With reference to Fig. 2, during the first time slot, \( S \) sends, consecutively, the two \( N \)-size information symbol blocks \( s(i) \) and \( s(i+1) \). The blocks are linearly encoded using the precoding matrix \( F \), so that the corresponding transmitted blocks are \( x_s(n) := Fs(n) \), with \( n = i, i+1 \). Within this time slot, the destination may be in a listening mode or not. Under a)6), after removing the guard interval at the receiver, the estimated vectors \( \hat{x}_s(n) \) and \( \hat{y}_s(n) \) are, respectively, with reference to Fig. 2, in the first half of the second time slot, \( S \) transmits \( x_s(i+2) = \alpha_1 Fs(i) \) and \( R \) transmits \( x_r(i+2) = \alpha_2 Fs(i+1) \). In the second half, \( S \) transmits \( x_s(i+3) = \alpha_1 Gs'(i+1) \) while \( R \) transmits \( x_r(i+3) = -\alpha_2 Gs'(i) \). To guarantee maximum spatial diversity, the two matrices \( G \) and \( F \) are related to each other by \( G = JF^\dagger \), as in [5], where \( J \) is a time-reversal (plus a one-chip cyclic shift) matrix. If \( N \) is even, \( J \) has all null entries except the elements of position \((1,1)\) and \((k, N-k+2)\), with \( k = 2, \ldots, N \), which are equal to \( 1 \). If \( N \) is odd, \( J \) is the antidiagonal matrix. The two real coefficients \( \alpha_1 \) and \( \alpha_2 \) are related to each other by \( \alpha_1^2 + \alpha_2^2 = 1 \). They are introduced in order to have a degree of freedom in the power distribution between \( S \) and \( R \), under a given total transmit power. In Section III, we will show how to choose \( \alpha_1 \) (and then \( \alpha_2 \)) in order to minimize the final average BER.

After discarding the CP, the blocks received by \( D \) in the two consecutive time slots \( i+2 \) and \( i+3 \) are

\[
y_d(i+2) = \alpha_1 H_{sd}^{n+2} Fs(i) + \alpha_2 H_{rd}^{n+2} Fs(i+1) + w_d(i+2)
\]

\[
y_d(i+3) = \alpha_1 H_{sd}^{n+3} Gs'(i+1) - \alpha_2 H_{rd}^{n+3} Gs'(i) + w_d(i+3)
\]

where \( H_{sd}^{n+2} \) and \( H_{rd}^{n+2} \) refer to the channels between \( S \) and \( D \), and between \( R \) and \( D \), respectively. Exploiting again the diagonalizations \( H_{sd}^{n+2} = WA_{sd}^{n+2} W^H \) and \( H_{rd}^{n+2} = WA_{rd}^{n+2} W^H \), if we premultiply in (3) \( y_d(i+2) \) by \( W^H \) and \( y_d(i+3) \) by \( W^T \), we get

\[
W^H y_d(i+2) = \alpha_1 A_{sd}^{n+2} \hat{F} s(i)
\]

\[
+ \alpha_2 A_{rd}^{n+2} \hat{F} s(i+1) + W^H w_d(i+2)
\]

\[
W^T y_d(i+3) = \alpha_1 A_{rd}^{n+3} \hat{G} s'(i+1)
\]

\[
- \alpha_2 A_{rd}^{n+3} \hat{G} s'(i) + W^T w_d(i+3)
\]

where \( \hat{F} := W^H F \) and \( \hat{G} := W^H G \). For the sake of simplicity, we assume that orthogonal frequency-division multiplexing (OFDM) is performed at both \( S \) and \( R \) nodes, so that \( N = M = F = W \), and thus, \( \hat{F} = I_N \) and \( \hat{G} = W \). We also introduce the orthogonal matrix

\[
A_{k_n} := \begin{pmatrix} \alpha_1 A_{sd}^{n} & \alpha_2 A_{rd}^{n} \\ -\alpha_2 A_{rd}^{n} & \alpha_1 A_{sd}^{n} \end{pmatrix}
\]

such that \( A_{k_n}^H A_{k_n} := I_2 \otimes \Lambda_{k_n}^2 \), where \( \Lambda_{k_n}^2 := \alpha_1^2 |A_{sd}^n|^2 + \alpha_2^2 |A_{rd}^n|^2 \), whereas \( \otimes \) denotes the Kronecker product. We introduce also the unitary matrix \( Q_{n}^{k_n} := A_{k_n} (I_2 \otimes \Lambda_{k_n}^{-1}) \), satisfying the relationships \( Q_{n}^{k_n} W_{n} = I_2 N \) and \( Q_{n}^{k_n} \Lambda_{k_n} = I_2 \otimes \Lambda_{k_n} \).

\footnote{The law of variation of \( \kappa_n \) versus \( n \) depends on the channel stationarity properties.}

\footnote{We suppose that the channels do not share common zeros on the grid \( x_d = e^{j2\pi q/N} \) with \( q \) integer, so that \( A_{k_n} \) is invertible.}
with covariance matrix $C$ where $I_2 \otimes \Lambda_{kn}$. Exploiting the above equalities and multiplying the vector $u := \left[ (W^H y_i, (i + 2))^T, (W^T y_i, (i + 3))^T \right]^T$ by the matrix $Q_{n+2}^H$, we get

$$
\begin{bmatrix}
\tilde{r}(i) \\
\tilde{r}(i + 1)
\end{bmatrix} := Q_{n+2}^H u

= \left[ \begin{array}{cc}
\tilde{\Lambda}_{sd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2} \\
\tilde{\Lambda}_{rd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2} + 2
\end{array} \right] s
$$

$$
+ \left[ \begin{array}{cc}
\tilde{\Lambda}_{sd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2} \\
\tilde{\Lambda}_{rd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2}
\end{array} \right] \tilde{s} + \tilde{w}
$$

where $s := [s(i)^T, s(i + 1)^T]^T$, $\tilde{s} := [\tilde{s}(i)^T, \tilde{s}(i + 1)^T]^T$, $\tilde{\Lambda}_{sd}^{k_2} := \alpha_1 \tilde{\Lambda}_{sd}^{k_2} + \frac{1}{\alpha_1}$, $\tilde{\Lambda}_{rd}^{k_2} := \alpha_2 \tilde{\Lambda}_{sd}^{k_2} + \frac{1}{\alpha_2}$, $\tilde{w} := [w(i)^T, w(i + 1)^T]^T = Q_{n+2}^H [w(i)^T, w(i + 3)^T]^T$. As expected, the previous equations reduce to the classical block Alamouti equations (see, e.g., [1] and [5]), if the two transmit antennas use the same power, i.e., $\alpha_1 = \alpha_2$, and there are no decision errors at the relay node, i.e., $\tilde{s}(n) \equiv s(n), n = i, i + 1$.

Since $Q_{n+2}^H$ is unitary, if $u$ is white, $\tilde{w}$ is also white, with covariance matrix $C_w = \sigma_w^2 I_{2N}$. Furthermore, since all matrices $\Lambda$ appearing in (5) are diagonal, the system (5) of 2N equations can be decoupled into N independent systems of two equations in two unknowns, each equation referring to a single subcarrier. More specifically, introducing the vectors $r_k := [r_k(i), r_k(i + 1)]^T$, $s_k := [s_k(i), s_k(i + 1)]^T$, $\tilde{s}_k := [\tilde{s}_k(i), \tilde{s}_k(i + 1)]^T$, and $\tilde{w}_k := [\tilde{w}_k(i), \tilde{w}_k(i + 1)]^T$, referring to the $k$th subcarrier, with $k = 0, \ldots, N - 1$ [for simplicity of notation, we drop the block index and we set $\tilde{\Lambda}_{sd} = \tilde{\Lambda}_{sd}^{k_2}(k, k)$ and $\tilde{\Lambda}_{rd} = \tilde{\Lambda}_{rd}^{k_2}(k, k)$], (5) is equivalent to the following systems of equations:

$$
r_k = \begin{bmatrix}
\tilde{\Lambda}_{sd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2} \\
\tilde{\Lambda}_{sd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2}
\end{bmatrix} s_k
$$

$$
+ \begin{bmatrix}
\tilde{\Lambda}_{sd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2} \\
\tilde{\Lambda}_{rd}^{k_2} & \tilde{\Lambda}_{rd}^{k_2}
\end{bmatrix} \tilde{s}_k + \tilde{w}_k.
$$

Since the noise vector $\tilde{w}_k$ is also white, with covariance matrix $C_w = \sigma_w^2 I_{2N}$, and there is no intersymbol interference (ISI) between vectors $s_k$ and $r_k$ corresponding to different subcarriers, $r_k$ represents a sufficient statistic for the decision on the transmitted symbol vector $s_k$.

Since $S$ and $R$ are not colocated, in general, the blocks transmitted from $S$ and $R$ arrive at $D$ at different times. This is a specific difference of distributed space–time block coding with respect to the classical DSTC. However, if the difference in arrival times $\tau_d$ is incorporated in the CP used from both $S$ and $R$, $D$ is still able to get $N$ samples from each received block, without IBI. In such a case, the different arrival time does not cause any trouble to the final receiver. In fact, let us take as a reference time the instant when the $i$th block coming from $R$ arrives at $D$. If the block coming from $S$ arrives with a delay of $L_d$ samples, the only difference, with respect to the case of perfect synchronization, is that the transfer function $\tilde{\Lambda}_{sd}(k)$ in (6) will be substituted by $\tilde{\Lambda}_{sd}(k)e^{-j2\pi L_d k/N}$. From (6), it is clear that such a substitution does not affect the useful term, as it only affects the interfering term. However, in the hypothesis of Rayleigh fading channel, $\tilde{\Lambda}_{sd}(k)$ is statistically indistinguishable from $\tilde{\Lambda}_{sd}(k)e^{-j2\pi L_d k/N}$. Hence, the combination of Alamouti (more generally, orthogonal STC) and OFDM is robust with respect to lack of synchronization between the time of arrivals of packets from $S$ and from $R$ [as long as $\alpha_6$ holds true]. The price paid for this robustness is the increase of the CP length $L$, which, in its turn, reflects into a rate loss. However, this loss can be made small by choosing a block length $N$ much greater than $L$ or by selecting only relays that are relatively close to the source, so as to make the relative delay small.

A. ML Detector

We derive now the structure of the ML detector at the final destination. Besides the previous assumptions, we assume also that $D$ has perfect knowledge of the vector of error probabilities $p_e(k)$ and $p_e(k)$, $k = 0, \ldots, N - 1$, occurring at the relay. This requires an exchange of information between $R$ and $D$. This information has to be updated with a rate depending on the channel coherence time. Later on, we will show an alternative (suboptimum) detection scheme that does not require such knowledge.

We denote with $S$ the set of all possible transmitted vectors $s_k$ and with $p_e(k)$ and $p_e(k)$ the conditional (to a given channel realization) error probabilities, at the relay node, on $s_k(1)$ and $s_k(2)$, respectively. After detection, at the node $R$, we have $\hat{s}_k(l) = s_k(l)$, with probability $1 - p_e(k)$, or $\hat{s}_k(l) = -s_k(l)$, with probability $p_e(k)$, $l = 1, 2$. Since the symbols are independent, the probability density function of the received vector $z$, conditioned to having transmitted $s_k$, is (see Appendix A)

$$
f_{z|s_k}(z|s_k) = \frac{1}{\pi^2 \sigma_n^2}
$$

$$
\times \left[ (1 - p_e(k_1)) (1 - p_e(k_2)) \exp \left\{ -\frac{|z - A_k(1, 1) s_k|^2}{\sigma_n^2} \right\} + p_e(k_1) p_e(k_2) \exp \left\{ -\frac{|z - A_k(-1, -1) s_k|^2}{\sigma_n^2} \right\} 
+ (1 - p_e(k_1)) p_e(k_2) \exp \left\{ -\frac{|z - A_k(1, -1) s_k|^2}{\sigma_n^2} \right\} + p_e(k_1) (1 - p_e(k_2)) \exp \left\{ -\frac{|z - A_k(-1, 1) s_k|^2}{\sigma_n^2} \right\} \right] \right)
$$

(7)
where \( A_k(\theta_1, \theta_2) \) is defined as follows:

\[
A_k(\theta_1, \theta_2) = \begin{bmatrix}
|\tilde{A}_{sd}|^2 + |\tilde{A}_{rd}|^2 \theta_1 & \tilde{A}_{sd} \tilde{A}_{rd} \theta_2 - \tilde{A}_{sd} \tilde{A}_{rd} \theta_1 \\
\tilde{A}_{sd} \tilde{A}_{rd} \theta_1 - \tilde{A}_{sd} \tilde{A}_{rd} \theta_2 & |\tilde{A}_{sd}|^2 + |\tilde{A}_{rd}|^2 \theta_2
\end{bmatrix}.
\]

Based on (7), the ML detector is

\[
\hat{s}_k = \arg \max_{s_k \in \mathbb{C}} \left\{ f_{r_k|s_k}(r_k|s_k) \right\}.
\] (8)

In deriving the optimal receiver (8), we assumed that \( D \) processes only the vectors received in the \((i + 2)\)th and \((i + 3)\)th time slots, ignoring the vectors (1) received in the slots \( i \) and \((i + 1)\). If \( D \) is in a listening mode also during the first time slot, since the vectors \( y_d(i) \) and \( y_d(i + 1) \) in (1) contain information about the transmitted symbols \( s(i) \) and \( s(i + 1) \), an appropriate combination of (1) and (3) improves the performance of the system, at the price of a slight increase of complexity. To this end, we introduce first the paraunitary matrix \( Q := \Lambda(I_2 \otimes \Lambda^{-1}) \), where

\[
\Lambda := \begin{pmatrix}
\Lambda_{sd} & 0 \\
\alpha_1 \Lambda_{sd} & \Lambda_{sd} \\
\Lambda_{rd} & \Lambda_{rd}
\end{pmatrix},
\]

\[
\Lambda^2 := |\Lambda_{sd}|^2 + |\Lambda_{sd}|^2 + |\Lambda_{sd}|^2 + |\Lambda_{sd}|^2,
\]

\[
\Lambda^H \Lambda = I_2 \otimes \Lambda^2,
\]

Then, proceeding as before, the optimal receiver, based on (1) and (3), is given by (8), where \( r_k := [r_k(i), r_k(i + 1)]^T \), with \([r(i), r(i + 1)]^T := Q^H(W^H y_d(i))^T, (W^H y_d(i + 1))^T, (W^H y_d(i + 2))^T, (W^T y_d(n + 3))^T \) given by

\[
\begin{bmatrix}
[r(i)] \\
r(i + 1)
\end{bmatrix} = \begin{bmatrix}
|\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 & -|\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 \\
|\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 & |\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 \\
-|\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 & |\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 \\
|\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2 & |\tilde{\Lambda}_{1d}|^2 & |\tilde{\Lambda}_{2d}|^2
\end{bmatrix}s + \tilde{w}
\] (9)

where \( s := [s(i)^T, s(i + 1)^T]^T \), \( \hat{s} := [\hat{s}(i)^T, \hat{s}(i + 1)^T]^T \) and (dropping the dependence on \( k \)) \( A_k(\theta_1, \theta_2) \) is given by the equation at the bottom of the page where \( \tilde{\Lambda}_{sd} := \Lambda_{sd}(k, k)\Lambda^{-1/2}(k, k) \), \( \tilde{\Lambda}_{sd} \) := \( \alpha_1 \Lambda_{sd}(k, k)\Lambda^{-1/2}(k, k) \), and \( \tilde{\Lambda}_{rd} := \alpha_2 \Lambda_{rd}(k, k) \times \Lambda^{-1/2}(k, k) \). Note that, thanks to the orthogonal space–time block coding strategy, the optimal detector preserves the receiver’s simplicity, because, under (a1)–(a6), the ML solution performs an exhaustive search only among four possible transmitted vectors \( s_k \).

**B. Suboptimum Detector**

The ML detector described above assume the knowledge, at the destination node, of the set of error probabilities \( p_{e1}(k) \) and \( p_{e2}(k) \), with \( k = 0, \ldots, N - 1 \). If this knowledge is not available, a suboptimum scalar detector can be implemented, instead of the ML detector. More specifically, the decision on the transmitted symbol \( s_k(n) \) can be simply obtained as

\[
\hat{s}(n) = \text{sign}\{\Re[r(n)]\}, \quad n = i, i + 1
\] (10)

where \( r(n) \) is given by (5) or (9). Note that, for high signal-to-noise ratio (SNR) at the relay (i.e., when \( R \) makes no decision errors), the symbol-by-symbol decision in \( D \) becomes optimal and, thus, the decoding rule (10) provides the same performance as the optimal receiver (8). When the decision errors at the relay side cannot be neglected, the suboptimal receiver introduces a floor in the BER curve, because the symbol-by-symbol decision (10) treats the wrong received symbols as interference. The choice between the decoding rules (8) and (10) should then result as a tradeoff between performance and computational complexity, taking into account the need for the ML detector to make available, at the destination node, the error probabilities of the relay node. We will show a comparison between ML and suboptimum strategies in Section IV.

**III. POWER ALLOCATION BETWEEN SOURCE AND RELAYS**

Whereas in conventional STC, the transmit antennas typically use the same power over all the transmit antennas, with DSTC, it is useful to distribute the available power between source and relay as a function of their relative position with respect to the final destination, since they are not colocated. In this section, we show how to distribute a given total power optimally between source and relay. We provide first a closed-form analysis in the ideal case where there are no decision errors at the relay, and then we will show some performance results concerning the real case where the errors are taken into account. We carry out our derivations considering the general case where the receiver combines all four vectors \( y_d(i) \div y_d(i + 3) \), as in (9).

**A. Error-Free S/R Link**

Under the assumption that there are no errors at the relay side, using the same derivations introduced in Section II,
the optimal detector is a symbol-by-symbol detector and the signal-to-noise ratio on the $k$th symbol in the $n$th block is

$$SNR_k(n) = \frac{A^2}{\sigma_n^2} \left( |\Lambda_{sd}^{k_i}(k, k)|^2 + \alpha |\Lambda_{sd}^{k_2}(k, k)|^2 + (1 - \alpha) |\Lambda_{rd}^{k_2}(k, k)|^2 \right)$$  \hspace{1cm} (11)$$

with $k = 0, 1, \ldots, N - 1$; $n = i, i + 1$, and $\alpha = \alpha_i^2$. The system performance depends on the channel stationarity properties. We distinguish between two cases: 1) fast-fading channels, where the channels are constant within each pair of blocks but they vary from any pair of blocks to the next one (this property could be implemented using a block interleaver with sufficient memory) and 2) slow-fading channels, where the channels are constant (at least within two consecutive time slots). In all cases, we assume that the variance of the channel-impulse-response coefficients is proportional to $1/d^2r$, with $r \geq 1$, where $d$ is the distance of the link.

1) Fast-Fading Channels: In such a case, the random variables $\Lambda_{sd}^{k_i}(k, k)$, $\Lambda_{sd}^{k_2}(k, k)$, and $\Lambda_{rd}^{k_2}(k, k)$ are independent. The error probability for binary antipodal constellation, conditioned to a given channel realization, is given by

$$P_{e/h}(k) = \frac{1}{2} \text{erfc}\left( \sqrt{0.5SNR_k} \right)$$  \hspace{1cm} (12)$$

where $SNR_k$ is given by (11). For each subcarrier $k$, $SNR_k$ is given by the sum of three statistically independent random variables, each one distributed according to a $\chi^2$ pdf with two degrees of freedom. Thus, using (11) and (12), the BER $P_b$ averaged over the channel realizations is [20]

$$P_b = \frac{1}{2} \sum_{k=1}^{Q} \pi_k \left[ 1 - \sqrt{\frac{\gamma_k}{1 + \gamma_k}} \right]$$  \hspace{1cm} (13)$$

where $Q = 3$, and

$$\pi_k := \prod_{i \neq k=1}^{Q} \frac{\gamma_i}{\gamma_i - \gamma_k}$$
$$\gamma_1 := \frac{A^2 \sigma_h^2}{\sigma_n^2 d_{sd}^2}$$
$$\gamma_2 := \frac{A^2 \alpha \sigma_h^2}{\sigma_n^2 d_{sd}^2}$$
$$\gamma_3 := \frac{A^2 (1 - \alpha) \sigma_h^2}{\sigma_n^2 d_{rd}^2}$$

$$\tilde{\sigma}_h^2 = \sigma_h^2 (L_h + 1).$$

From (13), we infer that, if the errors at the relay are negligible, our scheme achieves the maximum available diversity gain, equal to 3. The optimal value of $\alpha$ can be found by minimizing (13). In Appendix B, we prove that the average BER (13) is a convex function with respect to $\alpha$, so that the problem admits a unique solution.

2) Slow-Fading Channels: In case of slow fading [i.e., $\Lambda_{sd}^{k_i}(k, k) = \Lambda_{sd}^{k_2}(k, k)$ and $\Lambda_{rd}^{k_2}(k, k) = \Lambda_{rd}^{k_2}(k, k)$], the SNR on the $k$th block of the received block is [see (11)]

$$SNR_k = \frac{A^2}{\sigma_n^2} \left( (1 + \alpha) |\Lambda_{sd}^{k_2}(k, k)|^2 + (1 - \alpha) |\Lambda_{rd}^{k_2}(k, k)|^2 \right)$$  \hspace{1cm} (14)$$

and the BER averaged over the channel realizations is given by (13), with $Q = 2$ and

$$\gamma_1 := \frac{A^2 (1 + \alpha) \sigma_h^2}{\sigma_n^2 d_{sd}^2}, \quad \gamma_2 := \frac{A^2 (1 - \alpha) \sigma_h^2}{\sigma_n^2 d_{rd}^2}.$$

Hence, in this case, the maximum available diversity gain is 2. It is straightforward to show that, if $D$ is equipped with $R$ antennas, the achieved diversity gain is $3R$ for fast-fading channels, and $2R$ for slow-fading channels.

B. S/R Link With Errors

We derived a closed-form expression for the BER of the suboptimum scheme, in the presence of relay decision errors, conditioned to the channel (the details are in Appendix C). In such a case, the errors act as a bias on the final decision. The BER on the $i$th received bit $\delta_i(k)$ on the $k$th subcarrier, conditioned to the channel realizations, is (see Appendix C)

$$P_{e/h}(k) = \left[ \frac{1}{2} \text{erfc}\left( \frac{A \beta_k}{\sqrt{\sigma_n^2}} \right) \right] (p_{c1}(k)) (1 - p_{c2}(k))$$
$$+ \left[ \frac{1}{2} \text{erfc}\left( \frac{A \delta_k}{\sqrt{\sigma_n^2}} \right) \right] p_{c1}(k) (1 - p_{c2}(k))$$
$$+ \left[ \frac{1}{4} \text{erfc}\left( \frac{A (\beta_k + \gamma_k)}{\sqrt{\sigma_n^2}} \right) \right] p_{c2}(k) (1 - p_{c1}(k))$$
$$+ \left[ \frac{1}{4} \text{erfc}\left( \frac{A (\delta_k + \gamma_k)}{\sqrt{\sigma_n^2}} \right) \right] p_{c1}(k) p_{c2}(k)$$  \hspace{1cm} (15)$$

where [because of (a4)], for $n = i + 2, i + 3$

$$p_{c1}(k) = p_{c2}(k) = \frac{1}{2} \text{erfc}\left( \frac{A}{\sqrt{\sigma_n^2}} \right)$$

and

$$\beta_k = \sqrt{\alpha_i^2 |\Lambda_{sd}^{k_2}(k, k)|^2 + \alpha_i^2 |\Lambda_{rd}^{k_2}(k, k)|^2}$$
$$\gamma_k = \frac{2 \alpha_1 \alpha_2 \text{Re} \left\{ \Lambda_{sd}^{k_2}(k, k) \Lambda_{rd}^{k_2}(k, k) \right\}}{\sqrt{\alpha_i^2 |\Lambda_{sd}^{k_2}(k, k)|^2 + \alpha_i^2 |\Lambda_{rd}^{k_2}(k, k)|^2}}$$
$$\delta_k = \frac{\alpha_i^2 |\Lambda_{sd}^{k_2}(k, k)|^2 - \alpha_i^2 |\Lambda_{rd}^{k_2}(k, k)|^2}{\sqrt{\alpha_i^2 |\Lambda_{sd}^{k_2}(k, k)|^2 + \alpha_i^2 |\Lambda_{rd}^{k_2}(k, k)|^2}}$$  \hspace{1cm} (16)$$

with $\sigma_n^2 := \sigma_n^2 |\Lambda_{sd}^{k_2}(k, k)|^2, \quad \kappa_i + 2 = \kappa_{i+3}$. 
The average BER can then be obtained by averaging (15) over the channel realizations.

If $D$ performs the suboptimal symbol-by-symbol decision (10), using also the received blocks $y_d(i)$ and $y_d(i+1)$ [i.e., with $r(n)$ given by (9)], the BER conditioned to the channel realizations is given by (15) and (16), where $\alpha^2 d_{sd}^2 (k, k) + 2$ in (16) is replaced by $\alpha^2 |A_{sd}^n(k, k)|^2 + |A_{sd}^n(k, k)|^2$.

**Example 1—Optimal Power Allocation:** We show now the behavior of the final average BER as a function of the power allocation between $S$ and $R$, depending on the relative distances between $S$, $R$, and $D$. As an example, in Fig. 3, we report the average BER versus $\alpha$ (setting $\alpha = \alpha_1 = 1 - \alpha_2$), as defined in Section II, for different values of the distance $d_{rd}$ (and thus, of $\text{SNR}_R$) between $R$ and $D$ (all distances are normalized with respect to the distance $d_{rd}$ between $S$ and $D$). In Fig. 3(a), we consider the ideal case where there are no errors at the relay node. The $\text{SNR}_D$ at the final destination is fixed equal to 10 dB. We can observe that, when $d_{sd} = d_{rd} = 1$, the value of $\alpha$ that minimizes the average BER is $\alpha = 0.5$, i.e., the two transmitters use the same power. However, as $R$ gets closer to $D$, the optimal $\alpha$ tends to increase, i.e., the system allocates more power to $S$ with respect to $R$. The reverse happens when $d_{rd}$ is greater then 1. Thus, as expected, the system tends to somehow put $S$ and $R$ in the same conditions, with respect to $D$, in order to get the maximum diversity gain.

The real case, where there are decision errors at the relay node, is reported, as an example, in Fig. 3(b), where the average BER is again plotted as a function of $\alpha$, but for different values of the $\text{SNR}_R$ at the relay node. Interestingly, we can observe that, as $\text{SNR}_R$ decreases, the system tends to allocate less power to the relay node (the optimal value of $\alpha$ is greater than 0.5), as the relay node becomes less and less reliable.

**IV. Performance**

In this section, we compare alternative cooperative strategies. We assume a block length $N = 32$ and channel order $L = 6$. To make a fair comparison of the alternative transmission schemes, we enforce all systems to transmit with the same overall power. More specifically, if $P$ is the total power radiated by the noncooperative scheme, we denote by $P_1$ the power radiated by $S$ during the first time slot, and by $\alpha P_1$ and $(1-\alpha) P_1$ the power spent, respectively, by $S$ and $R$ in the second time slot. Since the overall radiated power is always $P$, it must be $P = P_1 + P_1$. The coefficient $\alpha$ is chosen in order to minimize the final average bit error probability (13) (see also Example 1). The power $P_1$ is chosen in order to achieve a required average average $\text{SNR}_R$ at the relay, defined as $\text{SNR}_R := P_1 / \sigma_n^2 d_{sd}^2$. All distances in the network are normalized with respect to the distance $d_{sd}$ between $S$ and $D$.

**Example 2—ML Versus Sub-Optimum Detector:** In Fig. 4, we compare the average BER obtained using alternative cooperative and noncooperative schemes in a slow-fading scenario. The BER is averaged over 2000 independent channel realizations. All curves are plotted versus the SNR in Fig. 4(a), and 25 dB in Fig. 4(b), for all values of $\text{SNR}_D$ reported in the abscissas. Since the noise power and $\text{SNR}_R$ are both fixed, increasing $\text{SNR}_D$ means that $P_1$ increases. In Fig. 4, we report, for the sake of comparison, the average BER obtained with the following schemes: 1) the single-hop method (dotted line); 2) the ideal ML detector for our DSTC scheme, with no errors at the relay (dashed and dotted line); 3) the real ML detector, incorporating the decision errors at the relay (dashed line); 4) the suboptimum scalar decoder for our DSTC scheme, showing both the theoretical average BER (solid line) and the corresponding simulation results (circles), obtained using a zero-forcing detector at the relay node; and 5) the

---

6We use our theoretical derivations, valid in the absence of errors at the relay, to simplify the strategy. One could improve upon this choice by using the BER resulting in the presence of errors at the relay.
SNR node. Also, the scheme of [17] exhibits a floor because a suboptimum receiver is due to the decision errors at the relay responding simulation results. The floor on the BER of the relay from $S$ that combines the signals coming from $R$ where $D$ increases, only $P_{II}$ increases, while $P_{I}$ is constant. As a consequence, the average number of errors at the relay does not decrease by increasing $SNR_D$. Thus, for a given BER at the relay, increasing $P_{II}$ does not yield a real benefit above a certain $SNR_D$. The BER floor can only decrease if the $SNR_R$ at the relay increases, as is evident by comparing Fig. 4(a) and (b). Conversely, in our DSTC scheme, even though the $SNR_R$ at the relay is fixed, the system has one more degree of freedom to be used against the effect of the relay decoding errors: the optimal power distribution between $S$ and $R$. It is this extra degree of freedom that avoids the final BER floor in the DSTC scheme. As an example, in the extreme case where the relay has a very low SNR, and thus a high error probability, the system would set $\alpha = 1$, i.e., it would shut off the relay by allocating all power to $S$ in the second slot. In such a case, the final destination would not be affected by the relay errors at all. The price paid for this advantage, with respect to [17], is that the optimal power distribution between $S$ and $R$ requires some extra signaling between $S$ and $R$. Comparing our suboptimum DSTC detector with the ML detector of [17], we note that [17] outperforms the suboptimum detector when both $SNR_R$ and $SNR_D$ are sufficiently high, whereas the situation is reversed at low/intermediate values of $SNR_R$. The comparison between our schemes and [17] is also strongly dependent on the relative distances between $S$, $R$, and $D$. As an example, in Fig. 5, we report the average BER obtained in the same scenario as Fig. 4, except that the distances are $d_{sr} = 0.3$ and $d_{rd} = 0.7$ in Fig. 5(a), and $d_{sr} = 0.7$ and $d_{rd} = 0.3$ in Fig. 5(b). We see that as $R$ moves away from $S$, our suboptimum detector loses with respect to [17], whereas our ML approach is always better than [17]. As before, the advantage, with respect to [17], is paid in terms of complexity, as it requires that $S$ and $R$ cooperate to select the most appropriate transmit power. Nevertheless, this is a low-rate signaling exchange that is not expected to affect the overall rate considerably.

It is also interesting to notice, from Figs. 4 and 5, that the suboptimum DSTC scheme exhibits performance very close to the optimal DSTC ML detector at low $SNR_D$, i.e., before the BER floor, when the relay is relatively close to the source. This indicates that the suboptimum detector is indeed a very good choice, under such a scenario, because it is certainly less complicated to implement than the ML detector. Most importantly, different from our ML and [17], the suboptimum scheme does not require any exchange of information between $R$ and $D$ about the BER in $R$. The price paid for this simplicity is that the $R$ node must have a sufficiently high SNR to guarantee that the BER of interest be above the floor. In Fig. 4, we have also reported the average BER (solid line with stars) obtained using a transmission strategy for our DSTC scheme, where instead of OFDM, in the $S/R$ slot, we used a linear precoding method that insures minimum BER at the relay, under the assumption of adopting a (suboptimal) minimum mean square error (MMSE) linear decoder (solid line with stars). In such a case, we observe that, with minimum additional complexity at the relay, the performance of the suboptimal DSTC scheme becomes closer to the ML decoder because of the lower BER at the relay.

scheme proposed by Laneman and Wornell in [17, Sec. C.1], where $R$ works as in 4) and $D$ implements the ML detector that combines the signals coming from $S$ in the first slot and from $R$ in the second slot (solid line with squares).

We can observe a very good agreement between our theoretical derivations for the suboptimum detector and the corresponding simulation results. The floor on the BER of the suboptimum receiver is due to the decision errors at the relay node. Also, the scheme of [17] exhibits a floor because $SNR_R$ is fixed. In fact, as $SNR_D$ increases, only $P_{II}$ increases, while $P_I$ is constant. As a consequence, the average number of errors at the relay does not decrease by increasing $SNR_D$. Thus, for a given BER at the relay, increasing $P_{II}$ does not yield a real benefit above a certain $SNR_D$. The BER floor can only decrease if the $SNR_R$ at the relay increases, as is evident by comparing Fig. 4(a) and (b). Conversely, in our DSTC scheme, even though the $SNR_R$ at the relay is fixed, the system has one more degree of freedom to be used against the effect of the relay decoding errors: the optimal power distribution between $S$ and $R$. It is this extra degree of freedom that avoids the final BER floor in the DSTC scheme. As an example, in the extreme case where the relay has a very low SNR, and thus a high error probability, the system would set $\alpha = 1$, i.e., it would shut off the relay by allocating all power to $S$ in the second slot. In such a case, the final destination would not be affected by the relay errors at all. The price paid for this advantage, with respect to [17], is that the optimal power distribution between $S$ and $R$ requires some extra signaling between $S$ and $R$. Comparing our suboptimum DSTC detector with the ML detector of [17], we note that [17] outperforms the suboptimum detector when both $SNR_R$ and $SNR_D$ are sufficiently high, whereas the situation is reversed at low/intermediate values of $SNR_R$. The comparison between our schemes and [17] is also strongly dependent on the relative distances between $S$, $R$, and $D$. As an example, in Fig. 5, we report the average BER obtained in the same scenario as Fig. 4, except that the distances are $d_{sr} = 0.3$ and $d_{rd} = 0.7$ in Fig. 5(a), and $d_{sr} = 0.7$ and $d_{rd} = 0.3$ in Fig. 5(b). We see that as $R$ moves away from $S$, our suboptimum detector loses with respect to [17], whereas our ML approach is always better than [17]. As before, the advantage, with respect to [17], is paid in terms of complexity, as it requires that $S$ and $R$ cooperate to select the most appropriate transmit power. Nevertheless, this is a low-rate signaling exchange that is not expected to affect the overall rate considerably.

It is also interesting to notice, from Figs. 4 and 5, that the suboptimum DSTC scheme exhibits performance very close to the optimal DSTC ML detector at low $SNR_D$, i.e., before the BER floor, when the relay is relatively close to the source. This indicates that the suboptimum detector is indeed a very good choice, under such a scenario, because it is certainly less complicated to implement than the ML detector. Most importantly, different from our ML and [17], the suboptimum scheme does not require any exchange of information between $R$ and $D$ about the BER in $R$. The price paid for this simplicity is that the $R$ node must have a sufficiently high SNR to guarantee that the BER of interest be above the floor. In Fig. 4, we have also reported the average BER (solid line with stars) obtained using a transmission strategy for our DSTC scheme, where instead of OFDM, in the $S/R$ slot, we used a linear precoding method that insures minimum BER at the relay, under the assumption of adopting a (suboptimal) minimum mean square error (MMSE) linear decoder (solid line with stars). In such a case, we observe that, with minimum additional complexity at the relay, the performance of the suboptimal DSTC scheme becomes closer to the ML decoder because of the lower BER at the relay.

At low $SNR_D$ values, there are no corresponding BER because, for those values, the transmitted power is insufficient to guarantee the required $SNR_R$ at the relay and thus, no cooperative communication can take place with the given SNR at the relay.
Fig. 5. Comparison between average BER versus SNR (dB) achieved with different decision schemes: single S/D link (dotted line); ideal ML detector (dashed and dotted line); real ML detector (dashed line); suboptimum receiver—theoretical results (solid line) and simulation (circles); Laneman and Wornell’s scheme (squares); (a) $d_{sr} = 0.3, d_{rd} = 0.7$, (b) $d_{sr} = 0.7, d_{rd} = 0.3$; SNR$_R = 15$ dB.

Finally, looking at the slopes of the average BER curves of the ML DSTC detector shown in Fig. 4(a) and (b), it is worth noticing that, in the absence of errors at the relay, the cooperative scheme achieves full spatial diversity gain. However, in the presence of decoding errors, the cooperative schemes do not achieve the full diversity gain. Indeed, in the presence of decoding errors at the relay, all cooperative schemes exhibit an asymptotic average BER behavior proportional to $1/SNR_D$. Nevertheless, there is a considerable coding gain, which justifies the use of cooperation. Indeed, a more attentive look at the results shows that the average BER starts approaching the slope with maximum diversity, as far as the errors at the relay are negligible with respect to the errors at the destination. Then, when the errors at the relay become dominant, the final BER curve follows the $1/SNR$ behavior. The absence of full diversity gain (but high coding gain) was also observed in [6], where (reliable) relays were assumed available only with a certain probability, depending on the random distribution of the relays.

Example 3—Fast Versus Slowly Varying Channels: If the S/D channel varies fast enough to assume independent values over the two time slots where $S$ transmits alone or when $S$ and $R$ transmit together, we can achieve a diversity gain of 3, if the relay errors are negligible. An example is shown in Fig. 6, where we compare the following three cases: 1) no cooperation (circles); 2) ML DSTC detector with slow-fading channel; and 3) ML DSTC detector with fast-fading channel. The SNR$_R$ at the relay is high enough to make the relay errors negligible. In fact, we observe that the average BER, in the fast-fading case, exhibits a behavior proportional to $1/SNR^3$. Decreasing SNR$_R$, all the curves would smoothly change their slope until they follow the asymptotic behavior $1/SNR$.

V. CHOICE OF THE CONSTELLATION ORDER IN THE SOURCE–RELAY SLOT

The other major critical aspect of cooperative schemes is their rate loss due to the insertion of the S/R time slot. As an example, if all the links would use a BPSK constellation, the rate loss factor would be 1/2. To reduce this loss factor, we can use higher order constellations in the $S/R$ link, with respect to the constellations used in the other links, so that the duration of the $S/R$ slot can be made smaller than the duration of the other slots. In this section, we assume BPSK transmissions over all links, except the $S/R$ link, where the constellation order is allowed to increase. More specifically, using a constellation $\mathcal{A}$ of cardinality $M = 2^{n_b}$ in the $S/R$ link, the rate loss factor is $n_b/(n_b + 1)$. On the other hand, cooperation reduces the final BER and then it induces a capacity increase. To quantify the overall balance in terms of rate, we compared the maximum rate achievable by our DSTC system with the maximum rate achievable with a noncooperative scheme. We define as achievable
rate the maximum number of bits per symbol (bps) that can be decoded with an arbitrarily low error probability, provided that sufficient error-correction coding is incorporated in the system, conditioned to the assumptions a1)–a6). We have shown in Section II that the combination of orthogonal DSTC and OFDM makes the overall time-dispersive channel equivalent to a set of parallel nondispersive subchannels. The final $S/R$ link over each subchannel, in the presence as well as in the absence of the relay link, can always be made equivalent to a binary symmetric channel (BSC) with crossover probability depending on the specific cooperative (or noncooperative) scheme adopted. Let us indicate with $p_{e|h}(k)$ such a binary error probability, conditioned to the channel realization. The capacity of such an equivalent channel is [20] (dropping the dependence from $k$)

$$C_{\text{BSC}}(p_{e|h}) := 1 + p_{e|h} \log_2(p_{e|h}) + (1 - p_{e|h}) \log_2(1 - p_{e|h}).$$

Thus, the maximum rate $R(k|h)$ that can be reliably transmitted, over the $k$th subcarrier, for a given channel realization $h$, incorporating the rate loss due to the insertion of the $S/R$ slot, is

$$R(k|h) = \frac{1}{1 + \frac{1}{n_{h}}} C_{\text{BSC}}(p_{e|h}(k)) \text{ bps.} \quad (17)$$

If the suboptimal symbol-by-symbol detector (10) is performed at the receiver, the error probability $P_{e|h}(k)$ is given by (15), with $p_{c1}(k) = p_{c2}(k)$ well approximated by [11]

$$p_{c1}(k) \approx \frac{1}{N} \sum_{k=1}^{N} \left\{ \frac{\sqrt{M} - 1}{\sqrt{M} \log_2 \sqrt{M}} \text{erfc} \left( \frac{1}{\sqrt{M} \log_2 \sqrt{M}} \right) \right. \left. + \frac{\sqrt{M} - 2}{\sqrt{M} \log_2 \sqrt{M}} \text{erfc} \left( 3 \sqrt{\frac{\beta}{\sigma_w^2(k)}} \right) \right\} \quad (18)$$

where $\beta = 3\sigma_w^2/2(M - 1)$, $\sigma_w^2(k) = \sigma_n^2/|\Lambda_{\text{eq}}(k,k)|^2$, and $\sigma_n^2$ is the variance of the transmitted symbols. From (17), we infer that, because of the $S/R$ link, the cooperative transmission induces a systematic rate loss of $1/(1 + 1/n_{h})$ with respect to the noncooperative case. But, at the same time, cooperation yields a smaller error probability $P_{e|h}(k)$ and thus, a higher $C_{\text{BSC}}(p_{e|h}(k))$. Then, we may expect a tradeoff in the choice of $n_{h}$. This tradeoff can be better understood through the following example.

**Example 4—Rate and Diversity Gain:** We report in Fig. 7(a) and (b) the achievable rate versus SNR, for an SNR$_R$ in $R$ equal to 10 and 3 dB, respectively, for different choices of the constellation used in the $S/R$ link, achieved with or without cooperation. In both cases, we assume that the $S/D$ and $R/D$ links assume the same (BPSK) constellation, in both cases of cooperation or no cooperation. We can see that, at high SNR$_D$, the noncooperative case approaches the maximum value, equal to 1 bps, whereas the cooperative cases tend to an asymptote less than 1, depending on the constellation used in the $S/R$ slot. We observe from Fig. 7(a) that, for SNR$_R = 10$ dB, increasing the constellation order in the $S/R$ slot from BPSK to 16-QAM (quadrature-amplitude modulation) improves the achievable rate; however, passing from 16-QAM to 64-QAM does not induce any further gain because of the higher BER at the relay. For lower values of SNR$_R$, i.e., SNR$_R = 3$ dB, there is no appreciable rate gain in increasing the constellation order.

---

8. It is important to remark that the rate defined above is smaller than the capacity of the system, because the proposed scheme is designed to maximize the spatial diversity gain and not to maximize information rate.

9. We could indeed use higher order constellations in the cooperative case, with respect to the noncooperative case. But this aspect is beyond the subject of this paper.
because of the excessive BER at the relay. Nevertheless, even though the cooperative scheme proposed in this paper was not aimed at improving the rate, it is interesting to notice that, at low/medium SNR (within a range depending on SNR), the cooperative case can outperform the noncooperative case also in terms of achievable rate, because the BER decrease can more than compensate the rate insertion loss due to the $S/R$ slot.

VI. CONCLUDING REMARKS

In this work, we have shown that a DSTC scheme using different nodes to build a virtual transmit array can be very effective to reduce the overall power radiated in a network, which is necessary to guarantee a desired final BER. However, the gain is achievable if the relays are chosen appropriately. In particular, source and relay should be close to each other for a series of convergent reasons: 1) less power is wasted to send data from the source to the relays; 2) less interference is generated towards other source/relay links sharing the same send data from the source to the relays; 3) synchronization problem is reduced. We have shown, by theory and simulation, that a suboptimum DSTC receiver, at the final destination, has a close behavior to the optimal ML detector at low SNR, but it exhibits a BER floor, whose value depends on the SNR at the relay. The floor can be eliminated using an ML decoder, but this comes at the price of increased complexity and the need of sending the estimated relay error probabilities at the destination. Overall, we may say that, if the SNR at the relay is sufficiently high, the suboptimum scheme is a good choice. We have then shown how to distribute the transmit power between source and relay in order to minimize the final average BER. In particular, we have shown that the overall system is relatively robust with respect to the choice of the power balance as the average BER curve is relatively flat for a large interval of $\alpha$. Finally, we have shown how to find a good balance, in terms of rate, between the loss due to the insertion of cooperation time slots and the performance improvement resulting from cooperation. Some preliminary results on alternative space–time coding strategies have been reported in [6].

APPENDIX A

We derive the optimal minimum BER detector in the case where $D$ takes the decision on the transmitted symbols by collecting all the received blocks. The optimal detector based only on the received blocks in the time slots $(i+2)$th and $(i+3)$th can be derived as a particular case. Given $[r^{T}(i), r^{T}(i + 1)]^{T} = q^{[l]}[W^{H}y(i)]^{T}, (W^{H}y(i + 1)]^{T}, (W^{H}y(i + 2)]^{T}, (W^{H}y(i + 3)]^{T}$, since the additive noise in (5) is Gaussian and white and no ISI occurs in the received blocks $r(i)$ and $r(i + 1)$, $r_k$ constitutes a sufficient statistic for deciding upon the transmitted symbols $s_k(i)$ and $s_k(i + 1)$. The vector $r_k$, assumes, in general, the form shown at the bottom of the page (with $A_k(\theta_1, \theta_2)$ defined right below), $s_k := [s_k(i), s_k(i + 1)]^{T}$, $\tilde{\Lambda}_sd := \Lambda^{\varsigma}_sd(k, i + 2) - x^{-1}(k, i)$, $\tilde{\Lambda}_sd := \alpha_1 \Lambda^{\varsigma}_sd(k, i + 2) - x^{-1}(k, i)$, and $\tilde{\Lambda}_rd := \alpha_2 \Lambda^{\varsigma}_rd(k, i + 2) - x^{-1}(k, i)$; $\tilde{\Lambda}_sd := \Lambda^{\varsigma}_sd(k, i + 1)$, and $\theta_1$ and $\theta_2$ are two independent random variables, with pdf given by $f_{\Theta_1}(\theta_1) = p_{\epsilon_1}(\theta_1)$ and $f_{\Theta_2}(\theta_2) = p_{\epsilon_2}(\theta_2)$, respectively; $p_{\epsilon_1}(\theta_1)$ and $p_{\epsilon_2}(\theta_2)$ are, for a given channel realization, the average error probability at the relay terminal on the $k$th bit in the blocks $s(i)$ and $s(i + 1)$, respectively. The optimal decision rule minimizing the average BER $P_{b|h}$, conditioned to a given channel realization, is given by the ML decision on the block $r_k = r_k(x(i), r_x(i + 1)^T)$. Denoting by $S$ the set of all possible vectors $s_k$, and by $s_k^* \in S$ the $i$th vector in the set $S$, the vector $r_{k,x}$, given $s_k^*$, $\Theta_1 = \theta_1$, and $\Theta_2 = \theta_2$, is a Gaussian random vector with mean $A_k(\theta_1, \theta_2)s_k^*$ and covariance matrix $C_w = \sigma_w^2 I_2$. Thus, the probability density function $f_{r_{k,x}|s_k^*}(z|s^*)$ of receiving the vector $z$, conditioned to having transmitted $s_k^*$, is (we drop the dependence from $k$) $f_{r_{k,x}|s_k^*}(z|s^*) = \frac{1}{\pi C_w} [1 - p_{\epsilon_1}(1 - p_{\epsilon_2}) g(1, 1, z, s^*)] \sum_{i} \left(1 - p_{\epsilon_1}(1 - p_{\epsilon_2}) g(0, 1, 1, z, s^*) + p_{\epsilon_1}(1 - p_{\epsilon_2}) g(1, 1, z, s^*) + (1 - p_{\epsilon_1}) p_{\epsilon_2} g(0, 1, 1, z, s^*) + (1 - p_{\epsilon_1}) p_{\epsilon_2} g(1, 1, z, s^*) \right)$

where $g(p, q, z, s^*) = \exp \{-|z - A(p, q) s^*|H C_w^{-1} [z - A(p, q) s^*] \} = \exp \{-|z - A(p, q) s^*|^2 / \sigma_n^2 \}$.

$$r_k = \left[ \begin{array}{c} \left(\tilde{\Lambda}_sd^2 + \tilde{\Lambda}_rd^2\right)^2 s(i) + \tilde{\Lambda}_sd^2 \tilde{\Lambda}_rd^2 \tilde{s}(i) + \tilde{\Lambda}_sd^2 \tilde{\Lambda}_rd^2 \tilde{s}(i + 1) - \tilde{\Lambda}_sd^2 \tilde{\Lambda}_rd^2 \tilde{s}(i + 1) \end{array} \right] + \tilde{\omega}_k$$

$$A_k(\theta_1, \theta_2) = \left[ \begin{array}{c} \left(\tilde{\Lambda}_sd^2 + \tilde{\Lambda}_rd^2\right)^2 \theta_1 + \tilde{\Lambda}_sd^2 \tilde{\Lambda}_rd^2 \theta_2 - \tilde{\Lambda}_sd^2 \tilde{\Lambda}_rd^2 \theta_1 \end{array} \right]$$
APPENDIX B

We prove that the average BER (13) is a convex function in $\alpha$. In fact, the function $g(x)$ defined as

$$g(x) = \int_{-\infty}^{\infty} w(y) f(x, y) dy$$

(19)

is convex in $x$ if, $\forall y$, $f(x, y)$ is convex in $x$ and $w(y) \geq 0$ [7]. From (19) and

$$P_b = E_h \{ P_{b|h} \} = \int_{\mathbb{R}} P_{b|h} \, f_h(h) \, dh$$

(20)

it follows that $P_b$ is convex in $\alpha$, if $P_{b|h} = 0.5 \operatorname{erfc} \left( \sqrt{0.5 \text{SNR}_k(\alpha)} \right)$ is convex in $\alpha$, where SNR$_k(\alpha)$ is given by (11) or (14). Introducing the function $f = h \circ g : \mathbb{R} \rightarrow \mathbb{R}$ defined as $f(x) = h(g(x))$, with $\text{dom} f = \{ x \in \text{dom} g | g(x) \in \text{dom} h \}$, where $\text{dom} f$ denotes the domain of the function $f$, the second derivative $f''(x)$ of the composition function $f$ is given by

$$f''(x) = h''(g(x)) (g'(x))^2 + h'(g(x)) g''(x).$$

(21)

From (21), we infer that $f(x)$ is convex if $h(x)$ is a convex ($h''(x) \geq 0$) nonincreasing ($h'(x) \leq 0$) function and $g(x)$ is concave ($g''(x) \leq 0$). Defining $h(x) := \operatorname{erfc} \left( \frac{\sqrt{n}}{\sqrt{2}} x \right)$, we have

$$h'(x) = -\frac{1}{2} \sqrt{\frac{a_1}{\pi x}} e^{-a_1 x^2} < 0 \quad \forall x > 0$$

$$h''(x) = \frac{1}{2} \sqrt{\frac{a_1}{2\pi x}} e^{-a_1 x^2} (a_1 + \frac{1}{x}) > 0 \quad \forall x > 0.$$

Thus, $h(x)$ is a convex nondecreasing function in $x$. As SNR$_k(\alpha)$, given by (11) or (14), is concave in $\alpha$ (because it is linear in $\alpha$) for any given channel realization, from (19) and (21), it follows that $f(\alpha) := P_b(\alpha)$ is convex in $\alpha$.

APPENDIX C

We derive the closed-form expression for the BER in the case of suboptimal symbol-by-symbol detection (10), with $\beta(n)$ given by (5). The decision on the transmitted symbol $s_k(i)$ is performed by taking a hard decision on

$$r_k(i) = [\hat{\Lambda}^{\kappa_u+2}_{sd}]^2 s_k(i) + [\hat{\Lambda}^{\kappa_u+2}_{rd}]^2 \hat{s}_k(i)$$

$$+ [\hat{\Lambda}^{\kappa_u+2}_{sd}] \hat{\Lambda}^{\kappa_u+2}_{rd} \hat{s}_k(i+1) - [\hat{\Lambda}^{\kappa_u+2}_{sd}] \hat{\Lambda}^{\kappa_u+2}_{rd} s_k(i+1) + \tilde{w}_k(i)$$

where $\hat{\Lambda}^{\kappa_u+2}_{sd} := \hat{\Lambda}^{\kappa_u+2}_{sd}(k, k)$, $\hat{\Lambda}^{\kappa_u+2}_{rd} := \hat{\Lambda}^{\kappa_u+2}_{rd}(k, k)$. For the ensuing derivations, it is useful to introduce the following events: $\mathcal{H}_1 \equiv \{ s_k(i) = s_k(i), \hat{s}_k(i+1) = s_k(i+1) \}$, $\mathcal{H}_2 \equiv \{ \hat{s}_k(i) = -s_k(i), \hat{s}_k(i+1) = s_k(i+1) \}$, $\mathcal{H}_3 \equiv \{ \hat{s}_k(i) = s_k(i), \hat{s}_k(i+1) = -s_k(i+1) \}$, $\mathcal{H}_4 \equiv \{ \hat{s}_k(i) = -s_k(i), \hat{s}_k(i+1) = -s_k(i+1) \}$, and $\mathcal{E}_{11} \equiv \{ s_k(i) = A, s_k(i+1) = A \}$, $\mathcal{E}_{1-1} \equiv \{ s_k(i) = A, s_k(i+1) = -A \}$, $\mathcal{E}_{-1,1} \equiv \{ s_k(i) = -A, s_k(i+1) = A \}$, $\mathcal{E}_{-1,-1} \equiv \{ s_k(i) = -A, s_k(i+1) = -A \}$.

Given the error probabilities $p_{e1}(k)$ and $p_{e2}(k)$, introduced in Section II-A, we have $P(\mathcal{H}_1) = (1 - p_{e1}(k))(1 - p_{e2}(k))$, $P(\mathcal{H}_2) = p_{e1}(k)(1 - p_{e2}(k))$, $P(\mathcal{H}_3) = (1 - p_{e1}(k))p_{e2}(k)$, and $P(\mathcal{H}_4) = p_{e1}(k)p_{e2}(k)$. Denoting, for simplicity, $u_i := \mathbb{R}\{ r_k(i) \}$, and using a3), the error probability $P_{e,h}(k)$ on $s_k(i)$, conditioned to a given channel realizations, is then

$$P_{e,h}(k) := \frac{1}{4} \sum_{j=1}^{4} P(\mathcal{H}_j)$$

$$\times \{ P\{ u_i > 0 | \mathcal{E}_{1-1}, \mathcal{H}_j \} + P\{ u_i > 0 | \mathcal{E}_{-1,1}, \mathcal{H}_j \}$$

$$+ P\{ u_i < 0 | \mathcal{E}_{1,1}, \mathcal{H}_j \} + P\{ u_i < 0 | \mathcal{E}_{-1,-1}, \mathcal{H}_j \} \}.$$

(22)

We derive now, in detail, all the terms in the sum of (22). Setting $\tilde{w}_r := \mathbb{R}\{ \tilde{w}_k(i) \}$, we have

$$P\{ u_i > 0 | \mathcal{E}_{1-1}, \mathcal{H}_1 \}$$

$$= P \left\{ \left( \frac{\hat{\Lambda}^{\kappa_u+2}_{rd}^2}{\hat{\Lambda}^{\kappa_u+2}_{sd}^2} \right)^2 - A + \tilde{w}_r > 0 \right\}$$

$$:= \frac{1}{2} \operatorname{erfc} \left( \frac{A\beta_k}{\sigma_n} \right)$$

where $\beta_k$ is given by (16). Using this approach, we find

$$P\{ u_i > 0 | \mathcal{E}_{1-1}, \mathcal{H}_1 \} = \frac{1}{2} \operatorname{erfc} \left( \frac{A\beta_k}{\sigma_n} \right)$$

$$P\{ u_i > 0 | \mathcal{E}_{-1,1}, \mathcal{H}_1 \} = \frac{1}{2} \operatorname{erfc} \left( \frac{A\beta_k}{\sigma_n} \right)$$

$$P\{ u_i > 0 | \mathcal{E}_{1,-1}, \mathcal{H}_2 \} = \frac{1}{2} \operatorname{erfc} \left( \frac{A\delta_k}{\sigma_n} \right)$$

$$P\{ u_i > 0 | \mathcal{E}_{-1,-1}, \mathcal{H}_2 \} = \frac{1}{2} \operatorname{erfc} \left( \frac{A(\beta_k - \gamma_k)}{\sigma_n} \right)$$

$$P\{ u_i > 0 | \mathcal{E}_{1,1}, \mathcal{H}_3 \} = \frac{1}{2} \operatorname{erfc} \left( \frac{A(\beta_k + \gamma_k)}{\sigma_n} \right)$$

$$P\{ u_i > 0 | \mathcal{E}_{-1,-1}, \mathcal{H}_4 \} = \frac{1}{2} \operatorname{erfc} \left( \frac{A(\delta_k - \gamma_k)}{\sigma_n} \right)$$

(22)

where $\delta_k$ and $\gamma_k$ are defined as in (16). Repeating this same approach, we are able to find all the terms in (22), which then give rise to (15).
REFERENCES


Gesualdo Scutari (S’01) received the electrical engineering and Ph.D. degrees from University of Rome “La Sapienza,” Rome, Italy, in 2001 and 2004, respectively.

During 2003, he held a visiting research appointment at the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley. He is currently a member of the Cooperative Communications Laboratory at INFOCOM Department, University of Rome “La Sapienza.” He has participated in two European projects on multi-antenna systems and multihop systems (IST SATURN and IST-ROMANTIK). His primary research interests include communication aspects of wireless multiple-input multiple-output (MIMO) channels, with special emphasis on convex optimization theory and game theory applied to communications systems, cooperative communications, distributed space–time coding (DSTC), and sensor networks.

Sergio Barbarossa (S’82–M’88) received the electrical engineering degree in 1984 and the Ph.D. degree in electrical engineering in 1988, both from the University of Rome “La Sapienza,” Rome, Italy. He joined the Radar System Division of Selenia in 1985, as a Radar System Designer. In 1987, he was a Research Engineer at the Environmental Research Institute of Michigan (ERIM), Ann Arbor, MI. From 1988 until 1991, he was an Assistant Professor at the University of Perugia. He has held positions as Visiting Scientist and Visiting Professor at the University of Virginia (1995, 1997), the University of Minnesota (1999), and the Polytechnic University of Catalonia (2001). In November 1991, he joined the University of Rome “La Sapienza,” where he is currently a Full Professor and Director of graduate studies. He has been involved in two European projects on multi-antenna systems and multihop systems. His current research interests lie in the areas of cooperative communications, distributed space–time coding, and sensor networks.

From 1998 to 2004, he has been a member of the IEEE Signal Processing for Communications Technical Committee. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1998 to 2001. He received the 2000 IEEE Best Paper Award from the IEEE Signal Processing Society, for a coauthored paper in the area of signal processing for communications. He is the author of a research monograph on Multiantenna Wireless Communication Systems. Since 2004, he has been serving as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, focusing on sensor networks.