Do Cooperative Radios Collide?

Andrea Rueetschi, *Student Member, IEEE,* and Anna Scaglione, *Fellow, IEEE*

***Abstract*—When cooperative terminals transmit concurrently, the receiver experiences additional distortion by the presence of radios carrier offsets and time delays. With the inevitable inaccuracy in estimating the channel at the receiving end, is it realistic to assume that relays transmissions do not collide?**

To address this question, we provide an analytical framework to model the effect of channel estimation methods that treat the channel as deterministic and unknown, and derive the average symbol error rate (SER) of a maximum likelihood sequence decoder (MLSD) and of a linear minimum mean squared error (LMMSE) equalizer using such estimates. We select the appropriate architecture by mapping two different channel estimation techniques onto the proposed analytical framework, and derive their diversity gain. First we analyze the well known linear least square channel estimation (LLSE), then, we show how to cast a compressed channel sensing (CCS) method onto our analytical model, by a novel approximation of the estimator noise and modeling mismatch. We observe that the latter is the key bottleneck in LLSE whereas the CCS appears sufficiently flexible to deliver the promised gains. This conclusion shows that collisions models are overly pessimistic, but it also underscores the need for an advanced receiver synchronization design.

***Index Terms*—Cooperative communication, channel estimation, diversity, distributed space-time coding.**

1. Introduction

COOPERATIVE transmission is an attractive method to form a multi-antenna arrays on the fly, using distributed wireless devices [1]. Several protocols have been suggested [2], among them, selective decode and forward (DF) has gained considerable attention [3] since it is relatively easy to integrate into existing wireless standards. In a first phase (P1) of DF, a single source broadcast a recruitment signal that includes the message to forward; a subset of relays, that have successfully decoded the message, participate in the second phase (P2) by forwarding the message (Fig. 1) toward the destination. When relays are uninformed about the channel, the source data are encoded, for example, using distributed coding [4], [5], distributed space-time block coding (DSTBC) [6], low density parity check codes (LDPC) [7] or convolutional encoding [8], to add spatio-temporal redundancy and achieve diversity at the cooperative receiver. Alternatively, if the relays have channel state information (CSI), distributed beam-forming is considered a viable option [9].

In most of the prior art, the average symbol error rate (SER) of the link is often derived assuming the receiver has perfect CSI [10]. This analysis is often considered sufficient when relays transmit in different times or frequency bands, like in [11]. This allows the end receiver to estimate each link individually, since each relay transmission is conveyed via a single-input-single-output (SISO) channel. Unfortunately, it is well known that this method reduces the spectral efficiency by a $1/V$ factor, where V is the number of relays. This is unlike MISO systems where the V transmit antennas are activated concurrently [12]. Is the usual assumption that concurrent transmissions *collide* really necessary? Indeed, physical layer cooperation with concurrent relay transmissions is looked with skepticism by some, who deem too difficult, if not outright impossible, to overcome the distortion of having different radios transmit at unison.

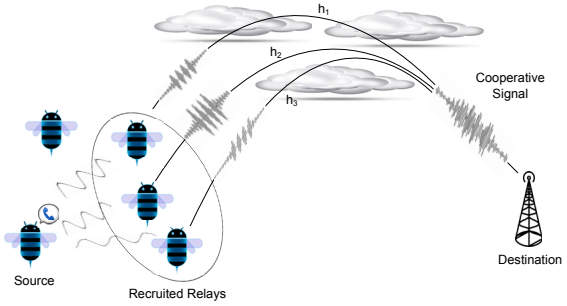

Fig. 1. Example of a cooperative transmission where $V = 3$ nodes, recruited during an initial phase (P1), concurrently transmit a space-time code over an HF channel (P2).

This background has motivated our work, since the current literature misses a practical case that proves the ability of a cooperative DF receiver to achieve the diversity gain, when the effect of the relay asynchronism is compounded with challenging time varying channels, like the so called Watterson model [13]. In fact, while forwarding the message to the recipient, each relay contributes adversely to the overall channel distortion due to their different location, mobility, adverse multi-path scattering, and other impairments that depend on the physical radio components, like the different processing delays and oscillator mismatches. All the effects combined in a single channel, contribute to the overall delay and frequency spread and may decrease significantly the odds of establishing a link [14].

Rather than embarking in exhaustive simulation campaigns, what we provide and propose in this paper is an analytical approach to determine how the cooperative link error rate and diversity of two decoders, namely Maximum Likelihood Sequence Estimation (MLSE) and linear minimum mean squared error (LMMSE) equalization followed by hard decisions, are affected by the use of channel estimates, in lieu of the actual CSI in the observation. When compared to similar studies, a

Paper approved by C. Tellambura, the Editor for Multicarrier Systems of the IEEE Communications Society. Manuscript received September 27, 2010; revised March 31, 2011 and October 1, 2011.

The authors are with the Department of Electrical and Computer Engineering, University of California, Davis, USA (e-mail : {arueetschi, ascaglione}@ucdavis.edu).

This work was supported by the National Science Foundation (NSF) by the grant 0905267.

Digital Object Identifier 10.1109/TCOMM.2012.022712.100556

fundamental difference in our approach is that we differentiate between the underlying actual linear time-varying channel that characterizes the received observation (whose parameters are in the vector h_d) and the channel model that is used by the receiver estimator to extract the estimate (the vector $\hat{h}_d$ used by the decoder). In fact, the analysis that we perform is based on modeling the resulting estimate $\hat{h}_d$ as having two components: one, denoted by $\psi_h \neq h_d$ that accounts for the basic model mismatch, or bias of the estimate, and the other one (w_h) that incorporates the effect of the observation noise during the training phase. To manage the complex interdependence between decision errors and channel estimation errors $\hat{h}_d$, we mainly rely on two approximations, (i) we neglect error terms that are of order higher than one [15]; (ii) we model the difference between the output of the presumed channel and actual channel as additional Gaussian independent noise (i.e. using maximum entropy equivalent), rather than using their exact statistics [16]. This leads to simplified expressions that closely approximate the performance, and that are easier to compare against the ideal performance than the exact ones. The main advantage of our approach is that we can test how robust certain hypotheses are when studying the error rate performance; in particular model mismatches can be easily compared to other factors, such as, for example, decoding strategy or simply noise.

We offer two concrete examples where the channel estimate can, in one case exactly and in the second case approximately, be mapped into the model we just mentioned, i.e. $\hat{h}_d = \psi_h + w_h$. The first case is the well known case of Linear Least Square Estimate (LLSE) when the channel is assumed Linear Time Invariant (see e.g. [17]), that we use as a benchmark. The second one, is the case of Compressed Channel Sensing (CCS) [18] using a time-frequency, or Gabor, dictionary [19], to better capture the time-varying nature of the channel. In the case of CCS, to provide the mapping of the estimate $\hat{h}_d = \psi_h + w_h$, we develop a novel *oracle estimator* to determine ψ_h. Numerical tests show that our mapping is effective in predicting the actual performance of the receiver, and that one can use analytical insights, rather than relying on heavy simulations of the CCS followed by the decoder of choice.

Our results highlight that it is nearly impossible to identify an SNR range for either decoding strategy, where the LLSE estimation mismatch does not propagate to the decoding stage so significantly, that diversity gains are still detectable. On the contrary, CCS combined with either decoding methods allows to reclaim the diversity that otherwise would have been lost. Hence, the answer to the question posed in the title is that cooperative relays *collide* if the receiver uses the linear time invariant (LTI) channel model, but that better signal processing at the receiver, such as for example CCS, can resolve the problem.

A. Related Work and Paper Contribution

The end-receiver design for cooperative links has yet to receive the deserved attention in the literature. Many articles address the subject of cooperation from an information theoretic standpoint, mainly by analyzing different protocols or source/channel encoding schemes that enhance capacity, outage probability and diversity [12], [20]–[24]. In many instances, the analysis assumes perfect CSI, slow fading and perfectly synchronized receivers. In [25] cooperative gains are achieved through delay diversity, but the scheme suffers from low spectral efficiency and relies on the receiver prior knowledge of the transmission delays. A similar approach also appears in [26]–[28] which also conclude that through an asynchronous transmission, the cooperative receiver will perform comparably as if the system were to transmit synchronously over orthogonal slots. Delays in this model belong to a finite discrete set, known by the receiver. Recently, [29] studied the effects of channel estimation errors and synchronization misalignments of an asynchronous concurrent transmission. By modeling the channel estimation errors as Gaussian random variables, that are independent from the channel fading, the author concludes that an increasing number of relays reduces the effects of errors, but at the cost of a poor time of arrival estimation.

In addition to many other subtle differences in capturing the effect of channel estimation onto the receiver performance, at present none of the previous literature has addressed the following key question: should cooperative relay transmissions be viewed as *collisions*, from an error rate performance standpoint, or not? Our paper tries to provide analytical insights and a compelling case study to answer the question, indicating what receiver designs are up to the task of decoding the signal as well as harvesting the cooperative gains.

Since traditional estimation methods for dispersive channels, based on a block LTI Finite Impulse Response (FIR) channel model, are not up to the task of synchronizing the receiver, we provide insights on how state of the art channel methods can overcome the obstacle. We focus on the emerging technique of compressed channel sensing (CCS) [18], [30], proposed for several different contexts, such as underwater acoustic communications [31]–[33], ultrawideband (UWB) [34] and mobile multiple-input-multiple-output (MIMO) multi-carrier communication [35], [36]. Our work, to the best of our knowledge, is the first providing an analytical framework to establish what is the effect of CCS on symbol decoding.

The organization of the article is as follows. In Sec. 2 we introduce the time-varying receiver models, Sec. 3 presents a general additive error model for the estimated channel and derives bounds on the average error rate for both the MLSD and LMMSE; their diversity is analyzed in Sec. 4 for a specific channel distribution. Sec. 5 examines the linear and sparse channel estimators and, for each, specifies the error model. In Sec. 6 we simulate the specific case of an HF channel applied to the each receiver configuration to draw our conclusions on the ability of a cooperative receiver to capture the diversity under time-varying channel conditions.

2. The Cooperative Model

We consider a set of V relays, successfully recruited during the broadcast phase (P1), and available to concurrently transmit a shared message during the cooperative phase (P2). Since, during P1, nodes receive from a single source, the

source to relay link performance is rather straightforward. Hence, our paper focuses on the performance of P2 only, rather than studying the end to end link, since we are interested in grasping the challenge of synchronizing the end receiver when multiple cooperating nodes transmit concurrently, and to determine its implications on the validity of the *collision model* for the cooperative relays. Other aspects that play a role in the end-to-end performance, are irrelevant and obnubilate our ability to address this compelling question. Hence, assume in the following that all relay decoded correctly the data after P1.

A block transmission system is implemented, where [37] known pilot symbols, $t \in \mathbb{C}^{N_t}$, and unknown data, $d \in \mathbb{C}^{N_d}$, alternate over time. To ease the notation we omit the index of the block. The length of each block, $N_t + N_d$, and the amount of training, N_t, are chosen parameters and, at the receiver side, there are K_t and K_d receiver samples that correspond to the training and communication signals. The channel encoding for in t and d may differ; for example, a system could map d on a QAM constellations and select a chirp waveform samples for t [38]. In this paper, we specify a common average symbol energy, $\mathcal{E}_s \in \mathbb{R}^+$ for both training samples and data, and assume that the data vector that is drawn uniformly at random from a constellation $\mathcal{A}$, such that $\mathbb{E}\{d\} = 0$ and $\mathbb{E}\{dd^H\} = \mathcal{E}_s I_{N_d}$. In modeling the channel, we assume that there is a time varying frequency selective channel between each transmitter and the end receiver, with $h_{t,v} \in \mathbb{C}^{K_t L}$ and $h_{d,v} \in \mathbb{C}^{K_d L}$ vectors stacking K_t and K_d, L taps long, impulse responses of the vth user, excited during the K_t samples of the training, and the K_d samples of the data, respectively. The entire set of impulse response for all samples and all antennas are denoted by h_t and h_d. Due to the linearity of the channel, the vectors of samples, $y_t \in \mathbb{C}^{K_t}$ and $y_d \in \mathbb{C}^{K_d}$ capture training and data phases, respectively, in two equivalent forms. In the first form the noiseless channel output is written as a time-varying convolution/coding matrix, built with the training or the data symbols, multiplying h_t and h_d respectively; in the second form, the roles are reversed, and a time-varying channel matrix multiplies training or the data symbols. More specifically, for the training phase, $\Psi'_v(t) \in \mathbb{C}^{K_t \times K_t L}$, represents the time-varying convolution matrix that is applied to the vth channel $h_{t,v}$,

$$y_t = \sum_{v=1}^{V} \Psi'_v(t) h_{t,v} + w_t = \Psi'(t) h_t + w_t, \qquad (2.1)$$

where $w_t \sim \mathcal{CN}(0, N_0 I_{K_t})$ is the receiver noise of variance N_0, and $\Psi'(t) \triangleq (\Psi'_1(t), \ldots, \Psi'_v(t), \ldots, \Psi'_V(t)) \in \mathbb{C}^{K_t \times K_t LV}$ is the aggregate training matrix. During data transmission the relays implement a distributed linear cooperative code of V dimensions, like the multi-antenna space-time (ST) scheme [39]. Then, the model y_d is equivalently described by a superposition of V signals as follows,

$$y_d = \sum_{v=1}^{V} \Phi'_v(d) h_{d,v} + w_d = \Phi'(d) h_d + w_d \qquad (2.2)$$

where $w_d \sim \mathcal{CN}(0, N_0 I_{K_d})$ is the noise during the data interval, uncorrelated with w_t, $\Phi'(d) \in \mathbb{C}^{K_d \times K_d LV}$ is the aggregate equivalent code matrix, where the operators $\Phi'_v(\cdot)$ and $\Phi'(\cdot)$ implement the time-varying convolution and rearrange the data elements according to the selected ST code. Alternatively to (2.1) and (2.2) we equivalently specify y_t and y_d by

$$y_t = \Psi(h_t) t + w_t \qquad y_d = \Phi(h_d) d + w_d \qquad (2.3)$$

where $\Psi(h_t) \in \mathbb{C}^{K_t \times N_t}$ and $\Phi(h_d) \in \mathbb{C}^{K_d \times N_d}$ are constructed by the operators $\Psi(\cdot)$ and $\Phi(\cdot)$ such that both equalities, $\Psi(h_t) t \triangleq \Psi'(t) h_t$ and $\Phi(h_d) d \triangleq \Phi'(d) h_d$ hold.

Remark 1: The expressions (2.1) and (2.2) represent the ground truth, as they account for all the possible variations. Clearly, channel estimation is not a tractable problem without further restrictions on this model that reduce the number of parameters. In Sec. 5, we review methods that reduce the unknowns in the true model to make the estimation problem solvable, whose performance we compare in our analysis. There, and in the following section that defines two possible cooperative detectors, we only consider designs that treat the channel deterministic but unknown. Approaches using stochastic channel models, though interesting, result in receiver architectures that are not amenable to the type of analysis we carry out next.

3. Error Rate of the Cooperative Receiver with Imperfect CSI

We measure the cooperative gains of different receiver configurations through the average symbol error rate (SER), $\bar{P}_e(\rho) \in \mathbb{R}^+$, expressed as a function of the signal-to-noise (SNR) ratio $\rho \in \mathbb{R}^+$. To derive the SER we assume a channel estimator that treats the channel as deterministic, but unknown. We assume that there is a true vector of channel parameters h_d and call its estimate $\hat{h}_d \in \mathbb{C}^{K_d LV}$. Our analysis is based on a model for the estimate that includes: (i) a zero mean *noise error* with a Gaussian distribution, and (ii), an unknown *modeling error* term, independent of the noise, that arises after the simplifications of the true model (2.1) made by the channel estimator. Both contributions are included in the additive model,

$$\hat{h}_d = \psi_h + w_h \qquad (3.1)$$

where, (i) $\psi_h(h_t) \in \mathbb{C}^{K_d LV}$, is the bias in the approximation of the channel h_d, ψ_h is a function of h_t but is independent of the noise v_t embedded in the training interval observations; and (ii) $w_h(h_t, v_t) \in \mathbb{C}^{K_d LV}$, is the noise component of the channel estimate, which depends on both h_t and v_t. More specifically, we assume that the noise contribution is Gaussian, $w_h \sim \mathcal{CN}(0, N_0(A_h A_h^H))$, with a normalized covariance determined by the matrix $A_h \in \mathbb{C}^{K_d LV \times K_t}$, whose structure depends on the channel estimator of choice and on the properties of the training design. When the noise in (3.1) vanishes, $N_0 \to 0$, the difference, $h_d - \psi_h$ represents the residual *modeling error*, i.e. a mismatch between the true channel h_d and its best noiseless approximation ψ_h. The persistence of such error is an indicator of the quality of the selected channel estimator. We derive the conditional SER, $\bar{P}_e(h_d, \psi_h; \rho) \in \mathbb{R}$, based on a single realization of h_t and h_d, and average it so that:

$$\bar{P}_e(\rho) = \mathbb{E}_{h_d, \psi_h}\big\{\bar{P}_e(h_d, \psi_h; \rho)\big\}. \qquad (3.2)$$

In the next section, we analyze the effects of $h_d - \psi_h$ on $\bar{P}_e(h_d, \psi_h; \rho)$ for two popular decoding options, (i) the MLSD; and (ii) the LMMSE, which is computationally attractive but a sub-optimal decoding method. Using specific encoding structures embedded in the mapping $\Phi(\cdot)$, one can harvest spatial but also delay diversity even with the LMMSE; however, the MLSD is the premier choice to gather all channel diversity at the receiver.

Remark 2: Although it is out of the scope of this paper, our expressions of the error rate on the cooperative link (P2) can be extended to the general case of a DF end-to-end error analysis by, (i) conditioning $\bar{P}_e(h_d, \psi_h; \rho)$ on the event that the relay has erroneously decoded the broadcast message (P1), (ii) by modeling each broadcast link as a SISO channel and averaging on that error event. Accounting for large-scale fading and shadowing effects, the end-to-end analysis can provide trade-offs between the network topology and the physical layer performance.

A. The MLSD Cooperative Link Error Rate

The ML detector with imperfect CSI, searches across a finite lattice $\mathcal{A}$ for the optimal vector

$$\hat{d} = \arg\min_{d' \in \mathcal{A}} \left\| y_d - \Phi(\hat{h}_d) d' \right\|^2 \tag{3.3}$$

where $\hat{d} \in \mathbb{C}^{N_d}$, $d' \in \mathbb{C}^{N_d}$ and $\Phi(\hat{h}_d) \in \mathbb{C}^{K_d \times N_d}$ is the coded channel matrix built upon $\hat{h}_d$. Then, the average error rate for a single channel realization is $\bar{P}_e(h_d, \psi_h; \rho) = \mathbb{E}_d\{\mathbb{P}(\hat{d} \neq d | h_d, h_t; \rho)\}$ and is typically approximated by the union bound of the pairwise error event [40]. After substituting the true model y_d (2.2) in (3.3), the error event, $\{\hat{d} \neq d\}$, will have statistics that depend on a weighted sum of non-central chi-square random variables $\|v_d\|^2$ and on a mixture of interference terms (see Appendix A). The distribution of the former exists but as an infinite sum [41] that, unless it is implemented numerically, does not provide insights about the impact of the modeling error $h_d - \psi_h$ on $\bar{P}_e(h_d, \psi_h; \rho)$. Instead, we approximate $\bar{P}_e(h_d, \psi_h; \rho)$ at an high SNR regime so that the contribution from $\|v_d\|^2$ becomes negligible and model the interference term as Gaussian, based on a maximum entropy model. Lastly, we bound the event $\{\hat{d} \neq d\}$ by typical events consisting of a single letter error in the decoded vector compared to the original one. Though in general this may not be the largest term in the union bound, it can be used to gauge the trend of the new approximated SER, with the advantage of leading to an expression that solely depends on the channel realization. This approach leads to the following proposition:

Proposition 1: Let $\delta_{min} e_m \in \mathbb{C}^{N_d}$ be a single error pairwise distance, where e_m is the mth coordinate vector, weighted by the minimum distance $\delta_{min} \in \mathbb{C}$ between two symbols of the constellation $\mathcal{A}$. The approximated SER is bounded by an expression that depends on the realization of the true channel h_d and of its noiseless estimate ψ_h

$$\bar{P}_e(h_d, \psi_h; \rho) \lessapprox \frac{\bar{N}_e}{N_d} \sum_{m=1}^{N_d} \mathrm{Q}\left(\frac{\gamma_{min} \|\phi_m(\psi_h)\|^2}{\sqrt{\rho^{-1} \|\phi_m(\psi_h)\|^2 + \|\Phi^H(h_d - \psi_h)\phi_m(\psi_h)\|^2}} \right) \tag{3.4}$$

where $\Phi(h_d - \psi_h) \in \mathbb{C}^{K_d \times N_d}$, $\phi_m(\psi_h) \triangleq \{\Phi(\psi_h)\}_m \in \mathbb{C}^{K_d}$, the SNR $\rho = \frac{\mathcal{E}_s}{N_0} \in \mathbb{R}^+$, a constant $\gamma_{min} = \frac{|\delta_{min}|}{\sqrt{2\mathcal{E}_s}} \in \mathbb{R}^+$ from $\mathcal{A}$, and $\bar{N}_e$ is the average number of nearest neighbors per dimension.

Proof: See the Appendix A ∎

The bound (3.4) shows that the modeling error, $h_d - \psi_h$, introduces an error floor when $\rho \to \infty$

$$\bar{P}_e(h_d, \psi_h; \rho) \xrightarrow{\rho \to \infty} \frac{\bar{N}_e}{N_d} \sum_{m=1}^{N_d} \mathrm{Q}\left(\frac{\gamma_{min} \|\phi_m(\psi_h)\|^2}{\|\Phi^H(h_d - \psi_h)\phi_m(\psi_h)\|} \right) \tag{3.5}$$

The magnitude of this error does not depend on the SNR but on the set of angles between the column space of $\Phi(\psi_h)$ and $\Phi(h_d)$: the smaller the angles, the smaller is the error. The worse case occurs when the two column spaces are orthogonal to each other.

B. The LMMSE Receiver Cooperative Link Error Rate

If $\hat{d}$ in (3.3) is chosen over the entire complex field, the solution is the output of the LMMSE equalizer [17]

$$z_d = \Phi^\dagger(\hat{h}_d) y_d = \Phi^\dagger(\hat{h}_d)\Phi(h_d) d + \Phi^\dagger(\hat{h}_d) w_d \tag{3.6}$$

where $z_d \in \mathbb{C}^{N_d}$ are the equalized symbols whose metric matches the $\mathcal{A}$ constellation. A decision is made by searching for the vector $d' \in \mathbb{C}^{N_d}$ with the smallest Euclidean distance from z_d

$$\hat{d} = \arg\min_{d' \in \mathcal{A}} \|z_d - d'\|^2 \tag{3.7}$$

where $\hat{d} \in \mathbb{C}^{N_d}$ are the symbol decisions.

After expanding the pseudo-inverse operator with (3.1), discarding the second order noise terms and applying a maximum entropy model to the interference, we study the high SNR behavior of the detector and identify the contribution of both modeling error and noise. The outcome of these approximations is in the following statement:

Proposition 2: Let $\delta_{min} e_m \in \mathbb{C}^{N_d}$ be a single error pairwise distance, where e_m is the mth coordinate vector, weighted by the minimum distance $\delta_{min} \in \mathbb{C}$ between two symbols of the constellation $\mathcal{A}$. The approximated SER of the LMMSE is bounded by an expression that depends on the realization of the of the true channel h_d and of its noiseless estimate ψ_h

$$\bar{P}_e(h_d, \psi_h; \rho) \lessapprox \frac{\bar{N}_e}{N_d} \sum_{m=1}^{N_d} \mathrm{Q}\left(\sqrt{\frac{\gamma_{min}^2 \rho}{\{\rho C_\zeta + C_u\}_{m,m}}} \right) \tag{3.8}$$

with the covariances, $C_\zeta(h_d, \psi_h) \in \mathbb{C}^{N_d \times N_d}$ and $C_u(h_d, \psi_h) \in \mathbb{C}^{N_d \times N_d}$, given by

$$\begin{aligned} C_\zeta &= \Phi^\dagger(\psi_h)\Phi(h_d - \psi_h)\Phi^H(h_d - \psi_h)\big(\Phi^\dagger(\psi_h)\big)^H \quad \text{and} \\ C_u &= \Phi^\dagger(\psi_h)\Big(I_{N_d} + \mathcal{E}_s \sum_{i=1}^{N_d} \big(\Phi(\mathrm{a}_i)\Gamma\Gamma^H\Phi^H(\mathrm{a}_i) \\ &\quad + \big(\Phi(\mathrm{a}_i)\Phi^\dagger(\psi_h)\big)^H \Omega\Omega^H \Phi(\mathrm{a}_i)\Phi^\dagger(\psi_h)\big)\Big)\big(\Phi^\dagger(\psi_h)\big)^H, \end{aligned} \tag{3.9}$$

with, $\Gamma(h_d, \psi_h) \in \mathbb{C}^{N_d \times N_d}$ and $\Omega(h_d, \psi_h) \in \mathbb{C}^{K_d \times N_d}$, given by, $\Gamma = I_{N_d} + \Phi^\dagger(\psi_h)\Phi(h_d - \psi_h)$, and

$$\Omega = \big(I_{K_d} - \Phi(\psi_h)\Phi^\dagger(\psi_h)\big)\Phi(h_d - \psi_h), \tag{3.10}$$

and $\Phi(\mathrm{a}_i) \in \mathbb{C}^{K_d \times N_d}$ argument is the ith column of $\mathrm{a}_i \triangleq \{A_h\}_i \in \mathbb{C}^{K_d LV}$, where A_h is defined after (3.1), as part of the statistics of the channel estimate noise term $w_h \sim \mathcal{CN}(0, N_0(A_h A_h^H))$.

Proof: See the Appendix B ∎

The matrix Ω measures the subspace error between $\Phi(h_d)$ and its reconstruction $\Phi(\psi_h)$, which vanishes when $\Phi(\psi_h)$ is square or under moderate channel variations. Naturally, all the noise terms, adding to the equalization error the term $N_0 C_u$, vanish when $\rho \to \infty$; however, the receiver error asymptotically saturates, driven by the covariance C_ζ of the residual inter-symbol interferer (ISI),

$$\bar{P}_e(h_d, \psi_h; \rho) \xrightarrow{\rho \to \infty} \frac{\bar{N}_e}{N_d} \sum_{m=1}^{N_d} \mathrm{Q}\left(\sqrt{\frac{\gamma_{min}^2}{\{C_\zeta\}_{m,m}}}\right) \quad (3.11)$$

4. Diversity Analysis

It is well known that cooperative communication provides spatial diversity. This enhancement is measured by averaging the SER $\bar{P}_e(h_d, \psi_h; \rho)$, over the channel realizations, h_d, and its noiseless estimate $\psi_h(h_t)$, and by tracing the asymptotic decay on a logarithmic scale [42]. Both (3.5) and (3.8) indicate the presence of an error floor, that makes the conventional asymptotic (high SNR) diversity analysis superfluous. For this reason a more informative measure is the *local diversity* $\eta(\rho) \in \mathbb{R}$, defined as:

$$\eta(\rho) = -\frac{1}{\log \rho} \log \mathbb{E}_{h_d, \psi_h}\{\bar{P}_e(h_d, \psi_h; \rho)\}, \quad (4.1)$$

since it identifies the SNR range where diversity gains are available. However, to analyze the decay of $\eta(\rho)$ and identify the best operating point for ρ, we have to introduce a stochastic model for both h_t and h_d and for the resulting channel estimate $\hat{h}_d$, to average the conditional SER, $\bar{P}_e(h_d, \psi_h; \rho)$ over the channel realizations. Thus, in this section we take the deterministic channels responses used in the previous analysis and replace them with random outcomes of an underlying wide-sense stationary (WSS) correlated complex Gaussian random model. The end receiver is agnostic of this model, and considers the channel as a deterministic block fading model. For instance, the example used in the numerical section is a time-varying multi-path narrowband channel often encountered in cooperative HF transmission. All the impulse responses of a single block can be expressed as a linear combination of an i.i.d. Gaussian vector $a \in \mathbb{C}^{QV}$, with a being stationary on a coherence interval that spans over several blocks and with Q as the number of channels per user, therefore

$$h_t = \Pi_t\, a \qquad h_d = \Pi_d\, a \qquad a \sim \mathcal{CN}\big(0, (1/\bar{\eta})\, I_{\bar{\eta}}\big) \quad (4.2)$$

with $\bar{\eta} = QV$ as the available spatial degrees of freedom; and where the matrices, $\Pi_t \in \mathbb{C}^{K_t LV \times \bar{\eta}}$ and $\Pi_d \in \mathbb{C}^{K_d LV \times \bar{\eta}}$, capture the time-frequency variations over a block period. When these variations are well localized in time and frequency we should expect $\Pi_d^H \Pi_d \propto I_{\bar{\eta}}$ and $\Pi_t^H \Pi_t \propto I_{\bar{\eta}}$ to be diagonally dominated. Together with the distribution, the analysis also requires a partial knowledge of the structure of $\Phi(\cdot)$. To simplify our analysis we consider linear space-time block codes (STBC), like [43], and introduce two binary matrices, $J_\Phi \in \mathbb{N}^{K_d LV \times K_d LV}$ and $J_{\phi_m} \in \mathbb{N}^{K_d LV \times K_d LV}$, that select all the active elements of, $\Phi(h_d)$ and $\phi_m(h_d)$, such that, $\|\Phi(h_d)\|_F^2 \triangleq \|J_\Phi h_d\|^2$ and $\|\phi_m(h_d)\|^2 \triangleq \|J_{\phi_m} h_d\|^2$, hold. In the next section the diversity of the MLSD is analyzed; for the LMMSE, instead, the evaluation is carried out numerically.

A. MLSD Receiver

Proposition 3: A lower bound on the local diversity, $\eta^{(MLSD)}(\rho) \in \mathbb{R}$, for the MLSD receiver is

$$\eta^{(MLSD)}(\rho) \geq \frac{-1}{\log \rho} \log \frac{\bar{N}_e}{N_d} \sum_{m=1}^{N_d} \int_0^1 \left(1 - \sum_{i=1}^{\bar{\eta}} \frac{(\sigma_i)^{\bar{\eta}}}{|\sigma_i| \prod_{i' \neq i}(\sigma_{i'} - \sigma_i)} \exp\left(\frac{-1}{\rho\, \sigma_i}\right) u(\sigma_i)\right) dy \quad (4.3)$$

with the integral calculated over the set of eigenvalues, $\sigma_i(y; m) \in \mathbb{R}$, whose dependency on the mth column index and on the variable, y, is specified by,

$$\sigma_i(y; m) = \lambda_i\Big(\big(\gamma_{min}/\,\mathrm{Q}^{-1}(y)\big)^2\, \Pi_h^H J_{\phi_m} \Pi_h - (\Pi_h - \Pi_d)^H J_\Phi\, (\Pi_h - \Pi_d)\Big) \quad (4.4)$$

where $\Pi_h = A_h \Pi_t \in \mathbb{C}^{K_d LV \times \bar{\eta}}$, $\lambda_i(\cdot)$ as the operator which returns the ith eigenvalue $\forall i \in \{1, \ldots, \bar{\eta}\}$, $u(\cdot)$ as the unit step function, and with $\mathrm{Q}^{-1}(\cdot)$ as the inverse of the Q-function.

Proof: See the Appendix C ∎

A deeper look at this expression shows that the MLSD receiver is bounded by a sum of terms $\mathcal{O}(\rho^{-k})$ with $k \leq \bar{\eta}$. This is apparent using a lower bound for $\eta^{(MLSD)}(\rho)$, obtained replacing the Chernoff bound [40] in lieu of the probability expression in (A.16),

$$\mathbb{P}\Big(\sum_{i=1}^{\bar{\eta}} \sigma_i(y; m) |\{\tilde{a}\}_i|^2 \leq \rho^{-1}\Big) \leq \min_{s>0} e^{-s} \prod_{i=1}^{\bar{\eta}} \mathbb{E}\big\{e^{-s\rho\sigma_i(y;m)\,|\tilde{a}_i|^2}\big\} = \min_{s>0} e^{-s} \prod_{i=1}^{\bar{\eta}} \big(1 + s\rho\, \sigma_i(y; m)\big)^{-1}$$

with parameter $s > 0$ and, $\tilde{a} \in \mathcal{CN}(0, I_{\bar{\eta}}/\bar{\eta}) \sim a$, as the equivalent channel fades after the eigenvalue decomposition (4.4). The random event $(\sum_{i=1}^{\bar{\eta}} \sigma_i(y; m) |\{\tilde{a}\}_i|^2 \leq \rho^{-1})$ can be interpreted as the event that estimated channel is in *deep fade* and, hence, its probability decays as an $\mathcal{O}(\rho^{-\bar{\eta}})$, if and only if all eigenvalues are positive, $\sigma_i(y; m) > 0$. This is consistent with (4.3), since the unit step functions $u(\sigma_i)$ set to zero the contributions from terms having $\sigma_i(y; m) < 0$. Instead, to have $\sigma_i(y; m) > 0$, requires a positive definite matrix argument in (4.4),

$$\big(\gamma_{min}/\,\mathrm{Q}^{-1}(y)\big)^2\, \Pi_h^H J_{\phi_m} \Pi_h - (\Pi_h - \Pi_d)^H J_\Phi\, (\Pi_h - \Pi_d) \succ 0. \quad (4.5)$$

Solving for y, returns an interval, $y \in (\mathrm{Q}(\pi^*), \mathrm{Q}(-\pi^*)) \subseteq (0, 1)$, where the contribution from every $\sigma_i(y; m)$ reaches full diversity, with the boundaries, $\mathrm{Q}(\pi^*)$ and $\mathrm{Q}(-\pi^*)$, specified by

$$\pi^* = \gamma_{min} \min_{i \in \{1, \ldots, \bar{\eta}\}} \lambda_i^{1/2}\big((\Pi_d - \Pi_h)^H\, J_\Phi (\Pi_d - \Pi_h)\big)^{-1} \times (\Pi_h^H\, J_{\phi_m} \Pi_h)\big) \xrightarrow{(\Pi_d - \Pi_h) \to 0} \infty \quad (4.6)$$

Under perfect CSI conditions, $\Pi_h = \Pi_d$, and, as expected, every $\sigma_i(y;m)$ contributes with full diversity. Otherwise, there will be intervals of y where their contribution have an $\mathcal{O}(\rho^{-k})$ with $k < \bar{\eta}$, lowering the bound for $\eta^{(MLSD)}(\rho)$. The overall loss in diversity will depend on how narrow is the range $(\mathrm{Q}(\pi^*), \mathrm{Q}(-\pi^*))$ relative to $(0,1)$. The width of this interval depends on the ability of the channel estimator to decrease the mismatch between Π_d and Π_h, as can be seen in (4.5).

B. LMMSE Receiver

A lower bound of the local diversity $\eta^{(LMMSE)}(\rho)$ for the LMMSE receiver with imperfect CSI,

$$\eta^{(LMMSE)}(\rho) \geq \frac{-1}{\log \rho} \log \sum_{m=1}^{N_d} \mathbb{E}_{h_d,\psi_h} \left\{ \exp\left(-\frac{\gamma_{min}^2 \rho}{2\{C_u + \rho C_\zeta\}_{m,m}} \right) \right\} - \frac{1}{\log \rho} \log\left(\frac{\bar{N}_e}{N_d}\right), \tag{4.7}$$

will be identified numerically, as the expression does not lend itself to a simple analytical evaluation, given the several terms in C_u and C_ζ. A perfect CSI condition, $\psi_h = h_d$, implies that $C_u = (\Phi^H(h_d)\Phi(h_d))^{-1}$ and $C_\zeta = 0$; therefore, to have $\eta^{(LMMSE\text{-}CSI)}(\rho) \to \eta^{(MLSD\text{-}CSI)}(\rho)$, requires C_u to be diagonal and full rank. Under moderate channel variations, codes like TR-STBC [44], meant for LTI channels, capture only part of the available cooperative diversity [45].

5. Channel Estimation Error Models

So far we have given for granted that the estimator output $\hat{h}_d$ can be modeled as postulated in (3.1). The resulting analysis of the error rate and of the local diversity suggests to minimize the strength of the modeling error $\|h_d - \psi_h\|^2$. However, this aim needs to be balanced with the need of containing the number of parameters used in fitting the channel, since the larger is the number of parameters, the more sensitive the estimate is to noise.

To give concrete examples on how the model (3.1) can be used in practice, in the next sections we consider two possible estimators that result from fitting the actual channel with two approximations, specifically: (i) the well known LLSE, that assumes an underlying block LTI FIR channel and (ii) CCS, which assumes that the channel is governed by a finite number of unknown parameters and exploits a time-frequency basis [19] to derive a sparse model, comparable to the work in [35]. In both cases we map the output of the estimator into the model for $\hat{h}_d$ proposed in (3.1). That includes specifying estimate bias, ψ_h, and the normalized noise covariance $A_h A_h^H$ of the noise term w_h, for a given deterministic pair h_t, h_d; this model is used to relate the statistics of h_t used to compute the diversity to the statistics of $\hat{h}_d$, as motivated by Remark 1. The specific expressions for ψ_h and A_h that correspond to these two channel estimators are used in the MLSD and LMMSE expressions of $\bar{P}_e(h_d, \psi_h; \rho)$ (3.2) and averaged over h_t and h_d, to study the average performance and diversity.

A. Error Model of the Linear Channel Estimator

A classic method in receiver design is to assume a block fading model, meaning that the state of the cooperative channel remains approximately invariant over at least the duration of a single block. This simplification greatly reduces the number of unknowns in (2.1) from $K_t LV$ down to LV, and the estimation problem becomes solvable by minimizing the least-squares (LS) error,

$$\hat{h}^{(LTI)} = \underset{h \in \mathbb{C}^{LV}}{\arg\min} \| y_t - \Psi^{(LTI)}(t) h \|^2 \tag{5.1}$$

with $\Psi^{(LTI)}(t) \in \mathbb{C}^{K_t \times LV}$ designed such that $\Psi^{(LTI)}(t) h^{(LTI)} \triangleq \Psi'(t)(1_{K_t} \otimes h^{(LTI)})$ holds, with $1_{K_t} \in \mathbb{N}^{K_t}$, as the all-ones vector.

It is known that in this case the model in (3.1) is exact, and the specific parameters of the channel estimate model are specified by $\hat{h}_d = A_h\, y_t$ with the decomposed terms given by,

$$\psi_h = A_h \Psi'(t) h_t \quad A_h = 1_{K_d} \otimes (\Psi^{(LTI)}(t))^\dagger \tag{5.2}$$

The selection of the training t is such that $\left(\Psi^{(LTI)}(t)^H \Psi^{(LTI)}(t)\right) \propto I_{LV}$ to meet the rank requirement and to minimize the noise contribution to the estimate [46]. Alternatively, by a simple linear interpolation [47], the impulse responses between two consecutive training transmissions, of $t^{(n)}$ and $t^{(n+1)}$, separated by $K_t + K_d$ symbols, are reconstructed by linearly weighting the corresponding LTI estimates, $\hat{h}^{(LTI,n)}$ and $\hat{h}^{(LTI,n+1)}$, at any instant $k \in \{0, 1, \ldots, K_d - 1\}$, by the coefficient $\beta_k = (K_t/2 + k)/(K_t + K_d)$ and $(1 - \beta_k)$, respectively.[1] The error model becomes,

$$\begin{aligned} \hat{h}_{d,k} &= (1-\beta_k)\, \hat{h}^{(LTI,n)} + \beta_k\, \hat{h}^{(LTI,n+1)} \\ \psi_{h,k} &= (1-\beta_k)\, \psi_h^{(LTI,n)} + \beta_k\, \psi_h^{(LTI,n+1)} \\ A_{h,k} &= \left((1-\beta_k)(\Psi^{(LTI)}(t^{(n)}))^\dagger \;,\; \beta_k\, (\Psi^{(LTI)}(t^{(n+1)}))^\dagger \right) \end{aligned} \tag{5.3}$$

B. Error Model of the Compressed Channel Sensing

Compressed Channel Sensing (CCS) can be effectively applied to estimate cooperative channels by assuming that the contribution from each user to the cooperative channel consists on a limited number of paths, and that each path contributes independently to the uncertainty in h_t by a single delay, Doppler and fading coefficient, with a common offset added to each Doppler, that models the local oscillator offset. Specifically, the cooperative receiver further assumes that, (i) all the delay, Doppler's and fading's are stationary over several block periods, (ii) each delay and Doppler pair belongs to a finite and quantized set of P elements that is common to all the V users, and form a time-frequency dictionary $G_k \in \mathbb{C}^{L \times P}$ [48] where each atom models a single impulse, of L taps, associated to a single quantized delay-Doppler pair. Then, each single instant impulse response, $h_{t,v,k} \in \mathbb{C}^L$ of $h_{t,v}$, is approximated by a linear combinations of few atoms, weighted by the unknown coefficients in $\alpha_v \in \mathbb{C}^P$. [2] CCS is most effective if an accurate sparse approximation of the impulse

[1] For this case we need to deviate from Remark 1 and assign block indices to both the LTI estimate and the training

[2] Our description of G_k is general but an example of a Gabor frame [19] is included in our numerical section.

response exists in the dictionary G_k. This is typically true if the dictionary is highly redundant, meaning that $L << P$, so that every impulse response is compressible with respect to G_k [49]. Within this framework, the estimation of $\alpha \in \mathbb{C}^{PV}$ can be cast into a compressed sensing (CS) problem [50],

$$\min_{\alpha} \|\alpha\|_0 \quad \text{subject to} \quad \| y_t - \Psi'(t) G_t \alpha \|^2 \leq \epsilon_0 \qquad (5.4)$$

where $G_t = \text{diag}\left(G_{t,1}, \ldots, G_{t,v}, \ldots, G_{t,V}\right) \in \mathbb{C}^{K_t LV \times PV}$ is a block diagonal matrix that extends the dictionary G_k to the training interval and to the V users, with $G_{t,v} = \left(G_1^T, \ldots, G_k^T, \ldots, G_{K_t}^T\right)^T \in \mathbb{C}^{K_t L \times P}$, $\Psi'(t) \in \mathbb{C}^{K_t \times K_t LT}$ is the training matrix now acting as a sensing matrix [49], and together with G_t forms an overcomplete CS basis $\Psi'(t)G_t \in \mathbb{C}^{K_t \times PV}$ with $\alpha = \text{vec}(\alpha_1, \ldots, \alpha_v, \ldots, \alpha_V) \in \mathbb{C}^{PV}$ as the unknown sparse support, and with $\epsilon_0 \in \mathbb{R}^+$ as a noise tolerance. CCS solutions $\hat{\alpha} \in \mathbb{C}^{PV}$ are found by relying on greedy algorithms [51] or ℓ_1 minimizers [52]. Their ability to identify the sparsest solution relies on the training design and, often but not always [53], on the restricted isometry property (RIP) [49]. Another advantage of this method is that it naturally leads to an interpolation strategy, by simple continuation of the basis function evolution during the data period. Hence, by the stationary assumption, the sparse vector estimated during training, is used to estimate the channel response with the appropriate *continuation basis* $G_d \in \mathbb{C}^{K_d LV \times PV}$,

$$\hat{h}_d = G_d \, \hat{\alpha} \qquad (5.5)$$

In the case of CCS the application of the model in (3.1) is anything but obvious, given that the recovery methodology is highly non-linear. In the following proposition, we approximate the model by resorting to an approach similar to those adopted to obtain *oracle bounds* on the performance, where, in essence, the channel support is known a priori through some *oracle* [54]–[56]. Numerically we observe that the method proposed in the following proposition closely approximates the CCS estimator performance based on the greedy solution of (5.4).

Proposition 4: For a given channel realization, h_d and h_t, and noise variance N_0, the CCS estimate $\hat{h}_d$ that solves (5.4) can be approximated by

$$\hat{h}_d = G_{d,\mathcal{S}} \left(\Psi'(t) G_{t,\mathcal{S}}\right)^{\dagger} y_t = \psi_h + w_h \qquad (5.6)$$

with the following terms

$$\psi_h = A_h \Psi'(t) h_t \qquad A_h = G_{d,\mathcal{S}} \left(\Psi'(t) G_{t,\mathcal{S}}\right)^{\dagger} \qquad (5.7)$$

where $\mathcal{S} \in \mathcal{P} = \{1, 2, \ldots, PV\}$ is a subset of columns, $G_{t,\mathcal{S}} \in \mathbb{C}^{K_t LV \times |\mathcal{S}|}$ and $G_{d,\mathcal{S}} \in \mathbb{C}^{K_d LV \times |\mathcal{S}|}$ are the selected subspaces of G_t and G_d, respectively. The support $\mathcal{S}$, provided by the *oracle* and used in the model in (5.6), is selected after solving the minimization

$$\mathcal{S} = \arg\min_{\mathcal{S}' \in \mathcal{P}} \|h_d - \psi_h(\mathcal{S}')\|^2 + N_0 \, \text{Trace}\left(A_h(\mathcal{S}') A_h(\mathcal{S}')^H\right) \qquad (5.8)$$

Proof: See the Appendix D ∎

The selection of $\mathcal{S}$ depends on the modeling error and shows that, in the absence of noise, the scheme selects the largest number of columns to better approximate h_d. The second term, instead, acts as a regularization parameter which, instead, grows along with $|\mathcal{S}|$ to prevent an overfitting. At low SNR, the second term becomes dominant, and few columns will be selected. In practice solving for $\mathcal{S}$ could require an exhaustive search, however, when (5.8) is aided by an efficient recovery algorithm, like the orthogonal matching pursuit (OMP) [57], the search over the domain of $\mathcal{S}$ can be replaced by a search over $|\mathcal{S}|$, with great computational savings,

$$\begin{aligned} \mathcal{S} = &\arg\min \left\|h_d - \psi_h(h_t, \mathcal{S}')\right\|^2 + N_0 \, \text{Trace}\left(A_h(\mathcal{S}') A_h^H(\mathcal{S}')\right) \\ &\text{subject to } \mathcal{S}' = \text{OMP}\left(\Psi(h_t)t, \Psi'(t)G_t, \left|\mathcal{S}'\right|\right) \\ &\text{and } \left|\mathcal{S}'\right| \in \{1, 2, \ldots, K_t\} \end{aligned} \qquad (5.9)$$

where OMP$(\cdot)$ is the OMP algorithm that, for every noiseless observation $\Psi(h_t)t$, basis $\Psi'(t)G_t$, and sparsity $|\mathcal{S}'|$ returns a single set of indices $\mathcal{S}'$. The stopping criteria is based on the maximum number of iterations, set to the desired sparsity $|\mathcal{S}'|$. [3] An example of the cost function (5.9), traced with respect to $|\mathcal{S}'|$ and N_0, appears in Fig. 2.

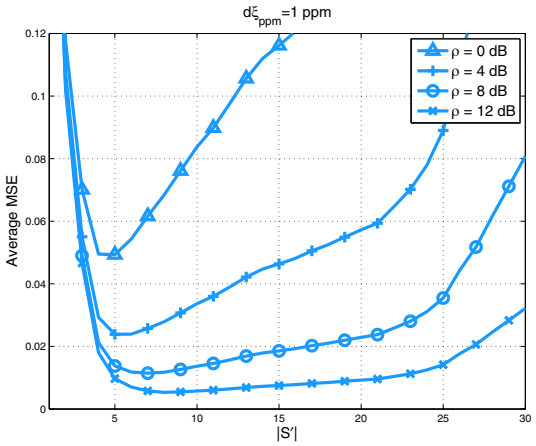

Fig. 2. Cost function (5.9) measured on a single realization of the cooperative channel, under a moderate Doppler ($d\xi_{ppm} = 1$ ppm) effect, traced as a function of the cardinality, $|\mathcal{S}'| = \{1, 2, \ldots, 30\}$, and SNR, $\rho = \{0, 4, \ldots, 12\}$ dB.

Although the number of active elements $|\mathcal{S}|$ does not necessarily match the total number of delay and Doppler values that are unknowns in the channel model, due to a quantization error, this approach remains valid since we expect the quantized parameters to be clustered around the true unknown ones [49]; thereby recovering $\mathcal{S}$ leads to an estimate of h_d.

6. Numerical Results

To benchmark our receiver we simulate a cooperative channel designed for a two antenna system, $V = 2$. The transmitted cooperative block is formed by a data vector d of $N_d = 16 \times 2 = 32$ symbols modulated with QPSK and TR-STBC encoded [44], thus $\bar{N}_e = 2$ and $V = 2$; and by a $N_t = 32 \times 2$ symbol length training, $t = \text{vec}(t_1, t_2)$, constructed from two cubic sequences [58], $\{t_1\}_k = \exp(j2\pi k^3/37) \in \mathbb{C}$ and $\{t_2\}_k = \exp(j2\pi k^3/47) \in \mathbb{C}$ with $0 \leq k < N_t/2$. The

[3] Alternatively, with a small performance loss, the minimization (5.9) can be replaced by the set of columns returned by OMP for the average number of selected columns, $\mathbb{E}_{h_t, h_d}|\mathcal{S}|$, for a given SNR, $\mathcal{S} = \text{OMP}(\Psi(h_t)t, \Psi'(t)G_t, \mathbb{E}_{h_t,h_d}|\mathcal{S}|)$.

block has unitary average power ($\mathcal{E}_s = 1$) and, consequently, $\rho = 1/N_0$ and $\gamma_{min} = 1$.

To decode a single data vector, the receiver stores two consecutive blocks and concatenates the training observations, that are adjacent to the data interval, into a single vector y_t, for a total of $K_t = 2(N_t/2 + L - 1)$ symbols, where $L = 6$ is the length of the time-varying impulse response. For the data, instead, there are $K_d = 2(N_d/2 + L - 1)$ symbols. Therefore, the time index, k, of the stored symbols belongs either to the set, $\mathcal{K}_t = \{0, 1, \ldots, K_t/2-1\} \cup \{K_t/2+K_d, \ldots, K_t+K_d-1\}$ for the training or , $\mathcal{K}_d = \{0, \ldots, K_d - 1\}$ for the data, respectively.

To generate the cooperative channels, as in (4.2), it requires Π_t (and Π_d) to be organized as $\Pi_t = \text{diag}(\Pi_{t,1}, \ldots, \Pi_{t,v}, \ldots, \Pi_{t,V}) \in \mathbb{C}^{K_t LV \times QV}$ with $\Pi_{t,v} = \left(\Pi_{t,v,0}^T, \ldots, \Pi_{t,v,k}^T \ldots, \Pi_{t,v,K_t+K_d-1}^T\right)^T \in \mathbb{C}^{K_t L \times Q}$ ($\forall\, k \in \mathcal{K}_t$). The same applies to Π_d after swapping the set $\mathcal{K}_d$ with $\mathcal{K}_t$. Then, the vth channel, $h_{t,v} = \text{vec}(h_{t,v,0}, \ldots, h_{t,v,k}, \ldots, h_{t,v,K_t+K_d-1})$ with $h_{t,v,k} \in \mathbb{C}^L$ is specified by $h_{t,v,k} = \Pi_{t,v,k}\, a_v$ with

$$\left\{\Pi_{t,v,k}\right\}_{l,q} = \exp(j2\pi\xi_v k T_s)\, g(lT_s - \tau_{v,q}), \tag{6.1}$$

$\forall\, k \in \mathcal{K}_t$ and where $l \in \{0, \ldots, L-1\}$ is the tap index, $q \in \{1, \ldots, Q\}$ the path index of $Q = 2$ independently fading paths per user, specified by the vth fading coefficients, $a_v \in \mathbb{C}^Q$, of $a = \text{vec}(a_1, \ldots, a_v, \ldots, a_V)$. The same approach is valid for $h_{d,v}$ after substituting $\mathcal{K}_t$ with $\mathcal{K}_d$.

We select the symbol rate $1/T_s = 2.4$ kHz, a typical choice in HF communication [59]. The set of propagation delays, $\tau_{v,q} \sim \mathcal{U}(0, \tau_{max}) \in \mathbb{R}^+$, is uniformly distributed and has a maximum spread of $\tau_{max} = (L-1)T_s \approx 2.1$ ms. The carrier offset for the v-th oscillator, $\xi_v \sim \mathcal{U}(-\frac{1}{2}\xi_{max}, \frac{1}{2}\xi_{max}) \in \mathbb{R}$, is uniformly distributed over a range of $\xi_{max} = (\xi_c d\xi_{ppm})10^{-6}$ Hz, defined by the maximum deviation of $d\xi_{ppm} \in \{1, 2, 3\}$ ppm of a $\xi_c = 10$ MHz HF carrier.

Lastly, $g(\cdot) \in \mathbb{R}$ is a truncated raised-cosine function such that $g(lT_s) = 0$ when $0 > l > L$. In a narrowband transmission $\tau_{v,q}$ is usually approximated as time invariant, $\tau_{v,q} \approx \tau_{v,q}(k)$; moreover, the delays have small impact on the performance when TR-STBC is used, due to its ability to extract delay diversity [45]. Therefore simulating a single τ_{max} is sufficient.

The CCS receiver is implemented with a time-frequency dictionary [19] $G_k \in \mathbb{C}^{L \times P}$

$$\{G_k\}_{l,p} = \exp\left(j2\pi\Delta\nu(-F/2 + f')kT_s\right) g(lT_s - \Delta\varepsilon\, d') \tag{6.2}$$

of dimensions $(D, F) = (\tau_{max}/\Delta\varepsilon + 1, \xi_{max}/\Delta\nu + 1) \in \mathbb{N}^2$, $P = FD$, resolution $(\Delta\varepsilon, \Delta\nu) = (T_s/2, d\xi_{ppm}/2) \in \mathbb{R}^2$, and delay-frequency indices $d' \in \{0, 1, \ldots, D-1\}$ and $f' \in \{0, 1, \ldots, F-1\}$; therefore the p-th column is selected by $p = Df' + d'$. The basis G_t and G_d are constructed by concatenating (as described in Sec. 5) the set of $\{G_k\}$ matrices generated at every instant $k \in \mathcal{K}_t$, and at $k \in \mathcal{K}_d$ during the data interval, respectively.

The oracle aided estimator (5.9), denoted by SP-ORCL, is compared against the sparse estimator, denoted SP-OMP, that approximates the solution of (5.4) also with OMP, implemented as function $\hat{\alpha} = \text{OMP}_2(y_t, \Psi'(t)G_t, N_0)$, that returns the estimate, $\hat{\alpha}$, from the noisy y_t, basis $\Psi'(t)G_t$, and noise level N_0. For this practical implementation the algorithm stops when the residual error falls below a threshold equal to N_0. The LLSE, instead, is implemented with either the LTI model (5.2), denoted by LS-LTI, or by the linear interpolation (5.3), denoted by LS-INTP.

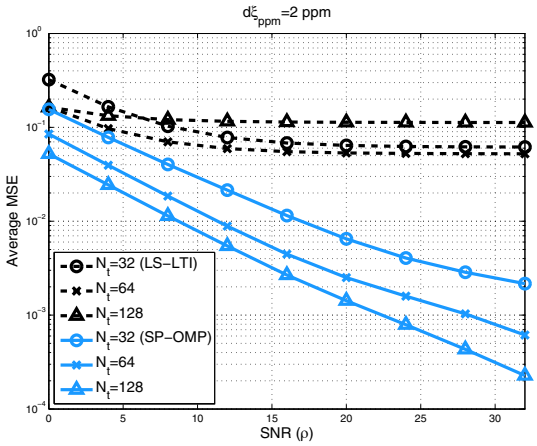

Fig. 3. Channel MSE, $\frac{1}{K_d V}\mathbb{E}\|h_d - \hat{h}_d\|^2$, comparing the LLSE (LS-LTI) (dashed black) against the sparse estimator (SP-OMP) (solid blue), with different training lengths $N_t = \{32, 64, 128\}$ at $d\xi_{ppm} = 2$ ppm.

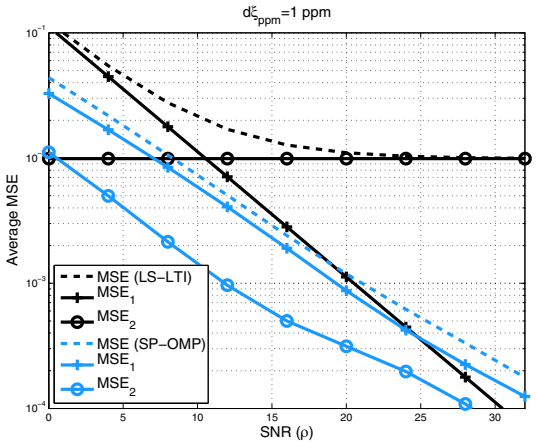

Fig. 4. Channel MSE of LS-LTI (dashed black) and SP-OMP (dashed blue), decomposed as MSE = MSE_1 + MSE_2 with contribution from the modeling error, $\text{MSE}_1 = \frac{1}{K_d V}\mathbb{E}\|h_d - \psi_h\|^2$, and from the noise error, $\text{MSE}_2 = \frac{1}{K_d V}\mathbb{E}\|w_h\|^2$; configured with a pilot length sequence of $N_t = 64$ symbols and a carrier offset range of $d\xi_{ppm} = 1$ ppm.

Channel MSE: the first result (Fig. 3) shows that allocating more energy to the training has an opposite effect on the two estimators. While the MSE performance of the SP-ORCL and SP-OMP are improved by the longer measurement time, driven by a more accurate selection of $\mathcal{S}$, the LLSE suffers from a modeling mismatch that remains constant (Fig. 4) and dominates the MSE when ρ increases. The effort made to identify the parameters pays off when $d\xi_{ppm}$ is large, and contributes to one decade gain between SP-OMP and LS-INTP at $\rho = 10$ dB (Fig. 5). SP-ORCL and SP-OMP perform similarly, confirming Proposition 4; furthermore, both are clearly less sensitive to changes in $d\xi_{ppm}$ compared to the LLSE, with the former bounding SP-OMP across the entire

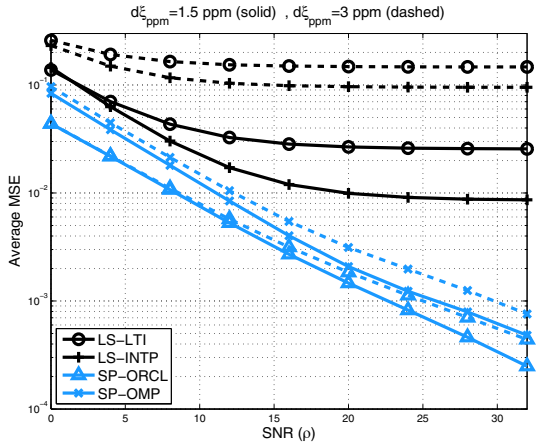

Fig. 5. Channel MSE comparison between the LLSE (LS-LTI and LS-INTP) (black) and the sparse estimator (SP-ORCL and SP-OMP) (blue) at $d\xi_{ppm} = 1.5$ ppm (solid) and $d\xi_{ppm} = 3$ ppm (dashed).

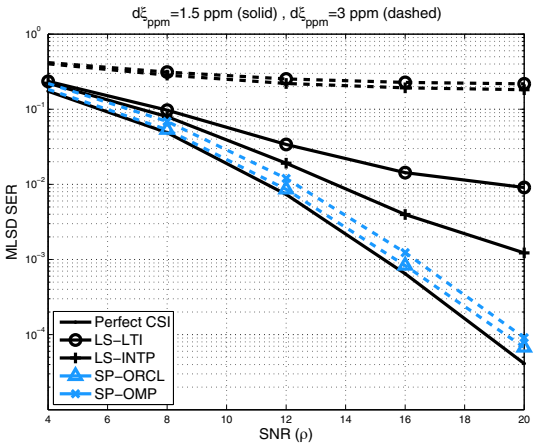

Fig. 6. MLSD SER performance $\bar{P}_e(\rho)$ (3.2), with $\bar{P}_e(h_d, \psi_h; \rho)$ given by (3.4) and channel estimates from the LLSE (LS-LTI and LS-INTP) (black) and the sparse estimator (SP-OMP and SP-ORCL) (blue) at $d\xi_{ppm} = 1.5$ ppm (solid) and $d\xi_{ppm} = 3$ ppm (dashed); (SP-ORCL and SP-OMP curves at $d\xi_{ppm} = 1.5$ ppm are not displayed).

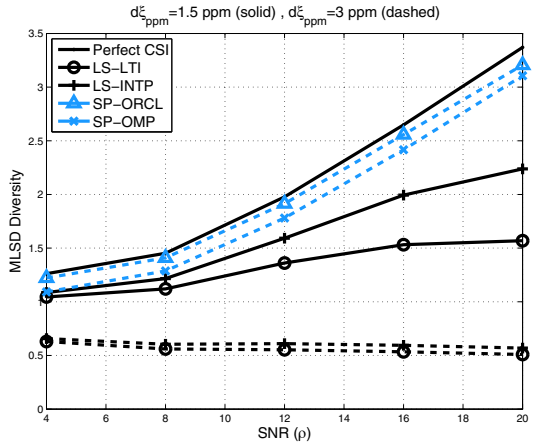

Fig. 7. MLSD Diversity $\eta(\rho)$ (4.1) with $\bar{P}_e(h_d, \psi_h; \rho)$ given by (3.4) and channel estimates from the LLSE (LS-INT and LS-INTP) (black) and the sparse estimator (SP-ORCL and SP-OMP) (blue) at $d\xi_{ppm} = 1.5$ ppm (solid) and $d\xi_{ppm} = 3$ ppm (dashed); (SP-ORCL and SP-OMP curves at $d\xi_{ppm} = 1.5$ ppm are not displayed).

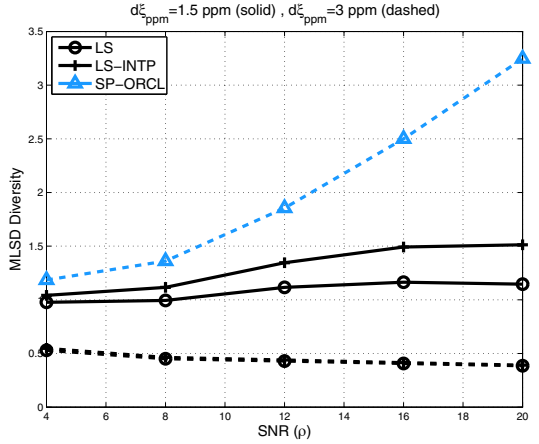

Fig. 8. MLSD Diversity $\eta(\rho)$ evaluated from the theoretical bound, (4.3), with channel estimates from the LLSE (LS-INT and LS-INTP) (black) and the sparse estimator (SP-ORCL) (blue) at $d\xi_{ppm} = 1.5$ ppm (solid) and $d\xi_{ppm} = 3$ ppm (dashed); (SP-ORCL curve at $d\xi_{ppm} = 1.5$ ppm is not displayed).

SNR range.

MLSD SER and Diversity: the ability of SP-ORCL (and SP-OMP) to track large offsets, like $d\xi_{ppm} = 3$ ppm, results in a small SER loss (< 1 dB at $\bar{P}_e(\rho) = 10^{-3}$) from the ideal curve (Fig. 6) leading to almost full diversity, $\eta^{(MLSD\text{-}CSI)} - \eta^{(MLSD)} \approx 0.3$ (Fig. 7). We observe that at $d\xi_{ppm} = 3$ ppm, the LLSE scheme fails to provide an acceptable performance, even with a more complex decoding strategy. Lastly, we observe that the diversity extracted from the empirical error rate (Fig. 7) is bounded by the theoretical expression (Fig. 8) as suggested by Prop. 3.

LMMSE SER and Diversity: The LMMSE receiver underperforms the MLSD by loosing up to 20% of the available diversity, $\eta^{(MLSD\text{-}CSI)} - \eta^{(LMMSE\text{-}CSI)} \approx 0.6$ at $d\xi_{ppm} = 1.5$ ppm (Fig. 9 and 10), due to the time variations altering the TR-STBC structure of $\Phi(h_d)$. Nevertheless, the sparse estimator harvests almost all the available diversity, in fact SP-ORCL only looses $\eta^{(LMMSE\text{-}CSI)} - \eta^{(LMMSE)} \approx 0.2$ at $\rho = 20$ dB. The approximated performance of the LMMSE relies on the approximated C_ζ and C_u (3.9), whose accuracy is tested in Fig. 11, and on the Gaussian approximation of the interference, tested in Fig. 12.

7. Conclusions

Our work shows that, even in the presence of substantial time and carrier asynchrony among relays and time varying channel conditions, it is possible to harvest diversity via cooperation by designing robust receiver strategies. In other words, the collision model is excessively pessimistic, especially considering state of the art synchronization methods, such as CCS. At more moderate offsets, even the simplest architecture (LLSE and linear equalization) gains a good portion of the expected diversity gain. However, the less sophisticated LLSE channel estimator has a steep drop in performance as the carrier offset increases ($d\xi_{ppm} = 3$ ppm), which is not unlikely in commercial radios, and the decoding architecture (MLSD or linear equalizer) makes little difference

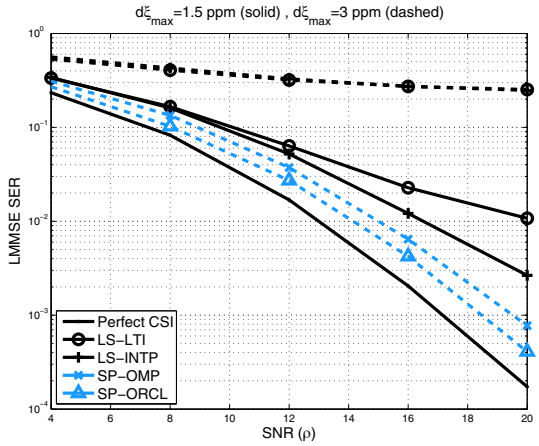

Fig. 9. LMMSE SER performance $\bar{P}_e(\rho)$ (3.2), with $\bar{P}_e(h_d, \psi_h; \rho)$ given by (3.8) and channel estimates from the LLSE (LS-LTI and LS-INTP) (black) and the sparse estimator (SP-OMP and SP-ORCL) (blue) at $d\xi_{ppm} = 1.5$ ppm (solid) and $d\xi_{ppm} = 3$ ppm (dashed).

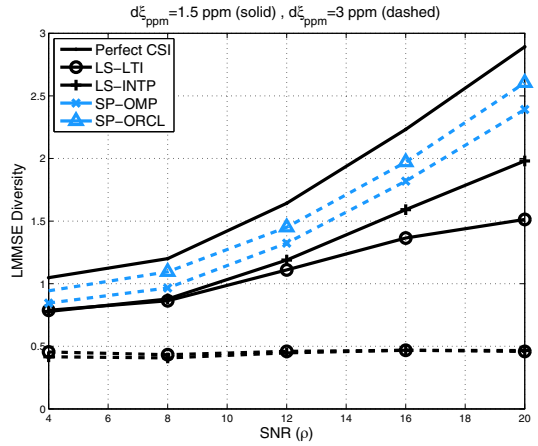

Fig. 10. LMMSE Diversity $\eta(\rho)$ (4.1) with $\bar{P}_e(h_d, \psi_h; \rho)$ given by (3.8) and channel estimates from the LLSE (LS-INT and LS-INTP) (black) and the sparse estimator (SP-ORCL and SP-OMP) (blue) at $d\xi_{ppm} = 1.5$ ppm (solid) and $d\xi_{ppm} = 3$ ppm (dashed).

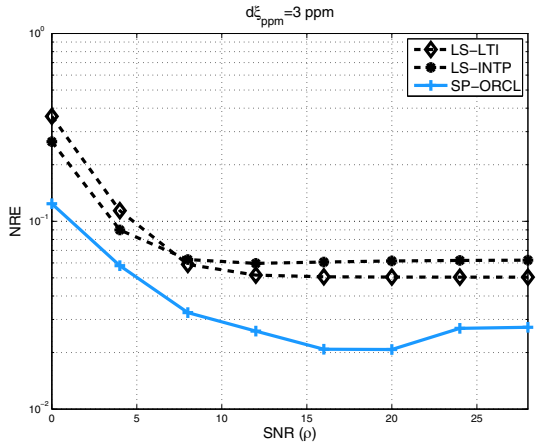

Fig. 11. Normalized residual error NRE $= \frac{1}{K_d V} \frac{\mathbb{E}\|C_u - (z_d - \Gamma d)(z_d - \Gamma d)^H\|_F}{\mathbb{E}\|(z_d - \Gamma d)(z_d - \Gamma d)^H\|_F}$ of the approximated covariance C_u (3.9) measured on a single realization of $\{h_d, \psi_h\}$, with ψ_h given by LLSE (LS-LTI and LS-INTP) (dashed black) and the sparse estimator (SP-ORCL) (solid blue); averaged over noise $\{w_d, w_h\}$ and data d at $d\xi_{ppm} = 3$ ppm. Lower carrier offsets follow a similar trend (not shown).

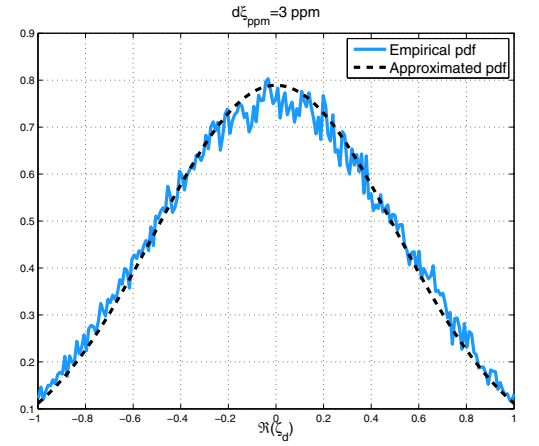

Fig. 12. Comparison of the empirical distribution (pdf) of a single element $\Re\{\{\zeta_d\}_1\}$ (A.12) (solid blue), against $\mathcal{CN}(0, \{C_\zeta\}_{1,1})$ (A.13) (dashed black); with the channel ($d\xi_{ppm} = 3$ ppm) being estimated by the LS-LTI.

when paired with it. The CCS estimator seems to decrease the sensitivity of the receiver to carrier offset considerably, allowing the robust decoding of cooperative radios even in relatively taxing time-varying multi-path conditions, creating a *collision free* environment for cooperative relays.

APPENDIX

A. Proof of Proposition 1

To bound $\bar{P}_e(h_d, \psi_h; \rho)$ we first expand the pairwise error event (PEP) with the error model $\hat{h}_d$ (3.1) and y_d (2.3) and apply an high SNR approximation. For two vectors $\{d_i, d_j\} \in \mathbb{C}^{N_d}$, the conditional PEP is $\mathbb{P}\big(d_i \to d_j \,|\, d_i, h_d, \psi_h\big) = \mathbb{P}\big(\|y_d - \Phi(\hat{h}_d)d_i\|^2 \geq \|y_d - \Phi(\hat{h}_d)d_j\|^2 \,|\, d_i, h_d, \psi_h\big)$. After substituting $\hat{h}_d$ and y_d, the random event is reformulated as

$$\|(\Phi(h_d) - \Phi(\psi_h))d_i + M_i v\|^2 \geq \|\Phi(h_d)d_i - \Phi(\psi_h)d_j + M_j v\|^2 \tag{A.1}$$

where $M_i \triangleq (-\Phi'(d_i)A_h, I_{K_d}) \in \mathbb{C}^{K_d \times (K_t + K_d)}$ and $M_j \triangleq (-\Phi'(d_j)A_h, I_{K_d}) \in \mathbb{C}^{K_d \times (K_t + K_d)}$ and with a Gaussian noise $v \triangleq \text{vec}(v_t, v_d) \sim \mathcal{CN}(0, N_0 I_{(K_t + K_d)})$ and the matrix $A_h \in \mathbb{C}^{K_d LV \times K_t}$. By the linear property of $\Phi(\cdot)$, we have $\Phi(h_d - \psi_h) = \Phi(h_d) - \Phi(\psi_h) \in \mathbb{C}^{K_d \times N_d}$, and,

$$\begin{aligned} &2\Re\Big\{\big((M_i - M_j)^H \Phi(h_d - \psi_h)d_i + M_j^H \Phi(\psi_h)(d_i - d_j)\big)^H v\Big\} \\ &\geq \|\Phi(h_d - \psi_h)d_i - \Phi(\psi_h)(d_i - d_j)\|^2 - \|\Phi(h_d - \psi_h)d_i\|^2 \\ &+ \mathcal{O}(\|v\|^2) \end{aligned} \tag{A.2}$$

To approximate the noise variance in (A.2) we assume good correlation properties on the training design so that $A_h A_h^H << I_{K_d LV}$, therefore

$$\begin{aligned} &M_i M_i^H = \Phi'(d_i) A_h A_h^H (\Phi'(d_i))^H + I_{K_d} \approx I_{K_d}, \\ &M_j M_j^H \approx I_{K_d} \quad \text{and} \quad (M_i - M_j)(M_i - M_j)^H \approx 0_{K_d} \end{aligned} \tag{A.3}$$

For a single symbol in error, $(d_i - d_j) = \delta_{min} e_m$ the variance

is approximated by

$$\|(M_i - M_j)^H \Phi(h_d - \psi_h) d_i + M_j^H \Phi(\psi_h)(d_i - d_j)\|^2 \approx |\delta_{min}|^2 \|\phi_m(\psi_h)\|^2 \quad \text{(A.4)}$$

with the m-th column $\phi_m(\psi_h) \triangleq \{\Phi(\psi_h)\}_m \triangleq \Phi(\psi_h) e_m \in \mathbb{C}^{K_d}$. The right-end side term in (A.2) can by further simplified by

$$\|\Phi(h_d - \psi_h) d_i + \Phi(\psi_h)(d_i - d_j)\|^2 - \|\Phi(h_d - \psi_h) d_i\|^2 = |\delta_{min}|^2 \|\phi_m(\psi_h)\|^2 - 2\Re\{\, \delta_{min} d_i^H \Phi^H(h_d - \psi_h)\phi_m(\psi_h)\} \quad \text{(A.5)}$$

Given a channel realization $\{h_d, \psi_h\}$ and, since d_i is zero-mean and has a discrete uniform distribution, we approximate $\Phi(h_d - \psi_h) d_i \sim \mathcal{CN}(0, \mathcal{E}_s \Phi(h_d - \psi_h)\Phi^H(h_d - \psi_h))$, therefore

$$2\Re\{\, \delta_{min} d_i^H \Phi^H(h_d - \psi_h)\phi_m(\psi_h)\,\} \sim \mathcal{CN}(0, 2\mathcal{E}_s |\delta_{min}|^2 \|\Phi^H(h_d - \psi_h)\phi_m(\psi_h)\|^2) \quad \text{(A.6)}$$

From (A.5) and (A.6) the updated pairwise is

$$2\Re\{\, \delta_{min} d_i^H \Phi^H(h_d - \psi_h)\phi_m(\psi_h)\,\} + 2\Re\{\, |\delta_{min}| \|\phi_m(\psi_h)\| \{v\}_1\,\} \geq |\delta_{min}|^2 \|\phi_m(\psi_h)\|^2 \quad \text{(A.7)}$$

After restricting the average over d_i to the $\bar{N}_e$ average number nearest neighbors of d_j we have $\bar{P}_e(h_d, \psi_h; \rho) \lessapprox (\bar{N}_e/N_d) \sum_{m=1}^{N_d} \mathbb{E}_{d_i}\{\mathbb{P}(d_i \to (d_i - \delta_{min} e_m) \,|\, d_i, h_d, \psi_h)\}$ and (3.4).

B. Proof of Proposition 2

At first we expand the expression of the equalized symbols by substituting y_d in z_d, then proceed by isolating all the noise terms from the data symbols and by discarding those of second order. Thus, we let $\{A, E\} \in \mathbb{C}^{K \times N}$ and assume $E^H E << A^H A$, then [60]

$$(A+E)^\dagger \approx (A^H A + E^H A + A^H E)^{-1}(A+E)^H \approx A^\dagger - A^\dagger E A^\dagger + (A^H A)^{-1} E^H (I - A A^\dagger)$$

By replacing $A \triangleq \Phi(\psi_h)$ and $E \triangleq \Phi(w_h)$, then $\Phi^\dagger(\hat{h}_d) \triangleq (A+E)^\dagger$ and, after few algebraic manipulations, the symbols of interests can be isolated from the the noise as

$$z_d = \Gamma\, d + u_d + w_z \quad \text{(A.8)}$$

with an error $w_z \in \mathbb{C}^{N_d}$ such that $\|w_z\|^2 < \max(\mathcal{O}(\|w_h\|^4), \mathcal{O}(\|w_d\|^4))$ and $u_d \in \mathbb{C}^{N_d}$ given by

$$u_d = \big(\Phi^H(\psi_h)\Phi(\psi_h)\big)^{-1} \times \big(\Phi^H(\psi_h) w_d - \Phi^H(\psi_h)\Phi(w_h)\Gamma d + \Phi^H(w_h)\Omega d\big) \quad \text{(A.9)}$$

where $\Gamma(h_d, \psi_h) = I_{N_d} + \Phi^\dagger(\psi_h)\Phi(h_d - \psi_h) \in \mathbb{C}^{N_d \times N_d}$ and $\Omega(h_d, \psi_h) = (I_{K_d} - \Phi(\psi_h)\Phi^\dagger(\psi_h))\Phi(h_d - \psi_h) \in \mathbb{C}^{K_d \times N_d}$. where $\Phi(a_i) \in \mathbb{C}^{K_d \times N_d}$ and $a_i \triangleq \{A_h\}_i \in \mathbb{C}^{K_d LV}$.

To determine a statistical model for z_d, based on a maximum entropy approximation and that uniquely depends on the channel realizations, requires the first and second moments of u_d and Γd.

The mean of u_d, $m_u \in \mathbb{C}^{N_d}$, and covariance, $C_u \in \mathbb{C}^{N_d \times N_d}$, require the moments of $u_\Gamma = \Phi(w_h)\Gamma d \in \mathbb{C}^{K_d}$ and $u_\Omega = \Phi^H(w_h)\Omega d \in \mathbb{C}^{N_d}$ in (A.9). Since $\mathbb{E}\{d\} = 0$, $\mathbb{E}\{w_h\} = 0$ and $\mathbb{E}\{w_h d\} = 0$, it follows that $\mathbb{E}\{u_\Gamma\} = 0$ and $\mathbb{E}\{u_\Omega\} = 0$, with normalized covariances, $C_\Gamma \in \mathbb{C}^{K_d \times K_d}$ and $C_\Omega \in \mathbb{C}^{N_d \times N_d}$, as

$$C_\Gamma = \frac{1}{N_0}\mathbb{E}\{u_\Gamma u_\Gamma^H\} = \mathcal{E}_s \sum_{i=1}^{N_d} \Phi(\mathrm{a}_i)\Gamma\Gamma^H\Phi^H(\mathrm{a}_i)$$

$$C_\Omega = \frac{1}{N_0}\mathbb{E}\{u_\Omega u_\Omega^H\} = \mathcal{E}_s \sum_{i=1}^{N_d} \Phi^H(\mathrm{a}_i)\Omega\Omega^H\Phi(\mathrm{a}_i) \quad \text{(A.10)}$$

since $\Phi(w_h)\Gamma d \triangleq \Phi'(\Gamma d) w_h$, $\Phi^H(w_h)\Omega d \triangleq (\Phi'(\Omega d))^H w_h$ and $\mathrm{a}_i \triangleq \{A_h\}_i \in \mathbb{C}^{K_d LV}$ Similarly, the noise w_d is also zero-mean and uncorrelated with w_h and d, therefore $m_u = \mathbb{E}\{u_d\} = 0$ with normalized covariance $C_u = \frac{1}{N_0}\mathbb{E}\{u_d u_d^H\}$ whose final expression requires C_Γ and C_Ω

$$C_u = \Phi^\dagger(\psi_h)\Big(I_{N_d} + \mathcal{E}_s \sum_{i=1}^{N_d} \big(\Phi(\mathrm{a}_i)\Gamma\Gamma^H\Phi^H(\mathrm{a}_i) + (\Phi(\mathrm{a}_i)\Phi^\dagger(\psi_h))^H \Omega\Omega^H \Phi(\mathrm{a}_i)\Phi^\dagger(\psi_h)\big)\Big)\big(\Phi^\dagger(\psi_h)\big)^H \quad \text{(A.11)}$$

The other term, Γd (A.8), is noiseless but shows the presence of an inter-symbol interferer (ISI) on each symbol in z_d. To evaluate the performance of a symbol-by-symbol detection, the ISI terms are isolated as a single vector $\zeta_d(h_d, \psi_h) \in \mathbb{C}^{N_d}$, therefore $\Gamma d = d + \zeta_d$, and

$$z_d = d + \zeta_d + u_d + w_z \quad \text{(A.12)}$$

where $\zeta_d = \Phi^\dagger(\psi_h)\Phi(h_d - \psi_h) d$ is the ISI with mean $m_\zeta = \mathbb{E}\{\zeta_d\} = 0 \in \mathbb{C}^{N_d}$ and normalized covariance $C_\zeta = \frac{1}{\mathcal{E}_s}\mathbb{E}\{\zeta_d \zeta_d^H\} \in \mathbb{C}^{N_d \times N_d}$,

$$C_\zeta = \Phi^\dagger(\psi_h)\Phi(h_d - \psi_h)\Phi^H(h_d - \psi_h)\big(\Phi^\dagger(\psi_h)\big)^H. \quad \text{(A.13)}$$

To approximate the error rate, we discard the error from the second order noise terms and and adopt a maximum entropy model for the ISI plus noise term in (A.12), i.e. $(\zeta_d + u_d) \sim \mathcal{CN}(0, \mathcal{E}_s C_\zeta + N_0 C_u)$, since ζ_d and u_d are uncorrelated by the mutual independence of w_h, w_d and d. For every pair $\{d_i, d_j\}$, the SER is bounded by a symbol-by-symbol pairwise error, i.e. $\bar{P}_e(h_d, \psi_h; \rho) \leq (1/N_d) \sum_{m=1}^{N_d} \mathbb{E}_{d_i}\{\sum_{j \neq i} \mathbb{P}(\{d_i\}_m \to \{d_j\}_m | d_i, h_d, \psi_h)\}$, that, together with the Gaussian model (A.12) and the restriction on the nearest neighbors of d_j, leads to the result (3.8).

C. Proof of Proposition 3

We first invoke Cauchy-Schwarz [60] and upperbound the denominator in (3.4) by

$$\sqrt{\rho^{-1}\,\|\phi_m(\psi_h)\|^2 + \|\Phi^H(h_d - \psi_h)\phi_m(\psi_h)\|^2} \leq \|\phi_m(\psi_h)\| \sqrt{\rho^{-1} + \|\Phi(h_d - \psi_h)\|_F^2} \quad \text{(A.14)}$$

Since $\|\phi_m(\psi_h)\|^2 = \|J_{\phi_m}\psi_h\|^2$ and $\|\Phi(h_d - \psi_h)\|_F^2 = \|J_\Phi(\psi_h - h_d)\|^2$, the local diversity, $\eta(\rho)$, is lower bounded by

$$\eta(\rho) \geq -\frac{1}{\log\rho} \log \frac{\bar{N}_e}{N_d} \sum_{m=1}^{N_d} \mathbb{E}_a\left\{ \mathrm{Q}\left(\sqrt{\frac{\gamma_{min}^2 \|J_{\phi_m}\Pi_h a\|^2}{\rho^{-1} + \|J_\Phi(\Pi_h - \Pi_d)a\|^2}}\right)\right\} \quad \text{(A.15)}$$

where $\Pi_h = A_h \Psi'(t) \Pi_t \in \mathbb{C}^{K_d LV \times \bar{\eta}}$. The bound depends on the distribution of a ratio of Gaussian norms, studied in [61], which relies on the distribution of a weighted sum of independent chi-square's [62]. Thus, to evaluate the mean, we exploit the identity, $\mathbb{E}\{Y\} = \int_0^\infty \mathbb{P}(Y > y) dy$ and the inverse of the Q-function, $\mathrm{Q}^{-1}(\cdot) \in \mathbb{R}^+$, and rewrite

$$\mathbb{E}_a \left\{ \mathrm{Q}\left(\sqrt{\frac{\gamma_{min}^2 \| J_{\phi_m} \Pi_h a \|^2}{\rho^{-1} + \| J_\Phi (\Pi_h - \Pi_d) a \|^2}} \right) \right\} = \int_0^1 \mathbb{P}\Big(\sum_{i=1}^{\bar{\eta}} \sigma_i(y;m) |\{\tilde{a}\}_i|^2 \leq \rho^{-1} \Big) dy \qquad \text{(A.16)}$$

after an eigenvalue decomposition with the ith eigenvalue, $\sigma_i(y;m) \in \mathbb{R}$, specified by (4.4), for every $i \in \{1, \ldots, \bar{\eta}\}$ with $\tilde{a} \sim \mathcal{CN}(0, (1/\bar{\eta}) I_{\bar{\eta}})$ as the equivalent channel fades whose statistics remain unaltered after the decomposition, $\tilde{a} \sim a$. Lastly, the probability in (A.16) is mapped to its distribution [61] which leads to (4.3).

D. Proof of Proposition 4

The oracle aided receiver, informed about the set $\mathcal{S}$, simplifies the minimization (5.4),

$$\hat{\alpha}_\mathcal{S} = \underset{\alpha_\mathcal{S} \in \mathbb{C}^{|\mathcal{S}|}}{\arg\min} \| y_t - \Psi'(t) G_{t,\mathcal{S}} \alpha_\mathcal{S} \|^2 \qquad \text{(A.17)}$$

where $G_{t,\mathcal{S}} \in \mathbb{C}^{K_t LV \times |\mathcal{S}|}$ is the selected subspace of G_t, $\alpha_\mathcal{S} \in \mathbb{C}^{|\mathcal{S}|}$ are the non-zero elements of α and $\hat{\alpha}_\mathcal{S} \in \mathbb{C}^{|\mathcal{S}|}$ their estimate, given by the LLSE, $\hat{\alpha}_\mathcal{S} = (\Psi'(t) G_{t,\mathcal{S}})^\dagger y_t$.

The oracle minimizes a cost function given by the channel MSE, $\mathbb{E}_{v_t} \| h_d - \hat{h}_d(\mathcal{S}) \|^2$, that is

$$\begin{aligned} \mathcal{S} &= \underset{\mathcal{S}' \in \mathcal{P}}{\arg\min}\ \mathbb{E}_{v_t} \| h_d - G_{d,\mathcal{S}'} \alpha_{\mathcal{S}'} \|^2 \\ &= \underset{\mathcal{S}' \in \mathcal{P}}{\arg\min}\ \| h_d - \psi_h(\mathcal{S}') \|^2 + N_0 \,\mathrm{Trace}\left(A_h(\mathcal{S}') A_h^H(\mathcal{S}') \right) \end{aligned}$$

where $G_{d,\mathcal{S}} \in \mathbb{C}^{K_t LV \times |\mathcal{S}|}$ is the selected subspace of G_d and, by basis continuation, leads to the oracle aided estimate, $\hat{h}_d = G_{d,\mathcal{S}} \hat{\alpha}_\mathcal{S} = (\Psi'(t) G_{t,\mathcal{S}})^\dagger y_t$, and to the error model (5.6).

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I and II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, 2003.

[2] J. N. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062–3080, 2004.

[3] M. Khormuji and E. G. Larsson, "Finite-SNR analysis and optimization of decode-and-forward relaying over slow-fading channels decode-and-forward relaying over slow-fading channels," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4292–4305, 2009.

[4] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 283–289, 2006.

[5] Y. Li, "Distributed coding for cooperative wireless networks: an overview and recent advances," *IEEE Commun. Mag.*, vol. 47, no. 8, pp. 71–77, 2009.

[6] S. Yiu, R. Schober, and L. Lampe, "Distributed space-time block coding," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1195–1206, 2006.

[7] A. Chakrabarti, A. D. Baynast, A. Sabharwal, and B. Aazhang;, "Low density parity check codes for the relay channel," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 280–291, 2007.

[8] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1470–1476, 2004.

[9] R. Madan, N. Mehta, and A. Molisch, "Energy-efficient cooperative relaying over fading channels with simple relay selection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3013–3025, 2008.

[10] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Trans. Commun.*, vol. 55, no. 7, 2007.

[11] S. Simoens, O. Muñoz-Medina, J. Vidal, and A. del Coso, "On the Gaussian mimo relay channel with full channel state information," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3588–3599, 2009.

[12] W. Su, A. Sadek, and R. Liu, "Cooperative communication protocols in wireless networks: performance analysis and optimum power allocation," *Wireless Personal Commun.*, vol. 44, no. 2, pp. 181–217, 2007.

[13] C. Watterson, J. Juroshek, and W. Bensema, "Experimental confirmation of an HF channel model," *IEEE Trans. Commun. Technol.*, vol. 18, pp. 792–803, 1970.

[14] T. Nguyen, O. Berder, and O. Sentieys, "Impact of transmission synchronization error and cooperative reception techniques on the performance of cooperative MIMO systems," in *Proc. 2008 IEEE Int. Conf. Commun.*, pp. 4601–4605.

[15] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, 2005.

[16] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim, "A central limit theorem for the SINR at the LMMSE estimator output for large-dimensional signals," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5048–5063, 2009.

[17] R. Ziegler and J. M. Cioffi, "Estimation of time-varying digital radio channels," *IEEE Trans. Veh. Technol.*, vol. 41, no. 2, pp. 134–151, 1992.

[18] W. U. Bajwa, J. J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: a new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[19] J. Wexler and S. Raz, "Discrete Gabor expansion," *Signal Process.*, vol. 21, no. 3, pp. 207–220, 1990.

[20] P. Mitran, H. Ochiai, and V. Tarokh, "Space–time diversity enhancements using collaborative communications," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2041–2057, 2005.

[21] D. Gündüz and E. Erkip, "Source and channel coding for cooperative relaying," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3454–3475, 2007.

[22] L. Xiao, T. E. Fuj, J. Kliewer, and D. J. Costello, "A network coding approach to cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3714–3722, 2007.

[23] P. Dayal and M. Varanasi, "Distributed QAM-based space-time block codes for efficient cooperative multiple-access communication," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4342–4354, 2008.

[24] L. Kong, S. Ng, R. Maunder, and L. Hanzo, "Near-capacity cooperative space-time coding employing irregular design and successive relaying," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2232–2241, 2010.

[25] X. E. Li, "Space-time coded multi-transmission among distributed transmitters without perfect synchronization," *IEEE Signal Process. Lett.*, vol. 11, no. 12, pp. 948–951, 2004.

[26] S. Wei, D. L. Goeckel, and M. C. Valenti, "Asynchronous cooperative diversity," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1547–1557, 2006.

[27] Y. Li and X.-G. Xia, "A family of distributed space-time trellis codes with asynchronous cooperative diversity," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 790–800, 2007.

[28] S. We, "Diversity-multiplexing tradeoff of asynchronous cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4150–4172, 2007.

[29] A. Ibrahim and R. Liu, "Mitigating channel estimation error with timing synchronization tradeoff in cooperative communications," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 337–348, 2010.

[30] C. Berger, W. Zhaohui, H. Jianzhong, and Z. Shengli, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, pp. 164–174, 2010.

[31] W. Li and J. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Oceanic Eng.*, vol. 32, no. 4, pp. 927–939, 2007.

[32] N. Richard and U. Mitra, "Sparse channel estimation for cooeprative underwater communication: a structured multichannel approach," in *Proc. 2008 ICASSP*, pp. 5300–5303.

[33] T. Kang and R. A. Iltis, "Iterative carrier frequency offset and channel estimation for underwater acoustic OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 9, pp. 1650–1661, 2008.

[34] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 383–395, 2007.

[35] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: leakage effects and sparsity-enhancing proc.," *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 255–271, 2010.

[36] F. Wan, W.-P. Zhu, and M. N. S. Swamy, "Semiblind sparse channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2569–2582, 2011.

[37] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, 1991.

[38] S. Boumard and A. Mainmela, "Synchronization using chirp training sequences in multipath channels," in *Proc. 2007 IEEE Globecom*, pp. 4030–4034.

[39] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: performance criterion in presence of channel estimation errors," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, 1999.

[40] J. Proakis and M. Salehi, *Digital Communication*, 5th edition. McGraw Hill, 2008.

[41] S. Rice, "Distribution of quadratic forms in normal random variables evaluation by numerical integration," *SIAM J. Scientific Computing*, vol. 1, pp. 438–448, 1980.

[42] B. Sirkeci-Mergen and A. Scaglione, "Randomized space-time coding for distributed cooperative communication," *IEEE Trans. Signal Process.*, vol. 10, pp. 4501–4506, 2007.

[43] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, 1999.

[44] E. Lindskog and A. Paulraj, "A transmit diversity scheme for channels with intersymbol interference," in *Proc. 2000 IEEE International Conf. Commun.*, vol. 1, pp. 307–311.

[45] M. Sharp, A. Scaglione, and S. Galli, "Distributed randomized space-time coding for HF transmission," in *Proc. 2006 IEEE Military Commun. Conf.*, pp. 1–6.

[46] A. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, pp. 1148–1160, 2008.

[47] K. Knoche, L. Rinas, and K. Kanimeyer, "Channel estimation with linear interpolation and decision feedback," in *Proc. 2002 IEEE Int. Symp. Spread-Specmm. Technol.*, vol. 1, pp. 54–58.

[48] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[49] E. J. Candes, "Compressive sampling," in *Proc. 2006 International Congress Mathematicians*, pp. 1–20.

[50] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[51] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, 2004.

[52] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM*, vol. 43, no. 1, pp. 129–159, 2001.

[53] A. Scaglione and X. Li, "Compressed channel sensing: is the restricted isometry property the right metric?" in *Digital Signal Processing*, submitted 2011.

[54] Z. Ben-Haim and Y. Eldar, "The Cramer-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, pp. 3384–3389, 2010.

[55] J. Haupt, J. Bajwa, W. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, 2010.

[56] T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, 2010.

[57] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[58] W. O. Alltop, "Complex sequences with low periodic correlations," *IEEE Trans. Inf. Theory*, vol. 26, pp. 350–354, 1980.

[59] E. Koski, "Use of HF for extended-range combat net radio communications," in *Proc. 2006 IRST*, pp. 38–42.

[60] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[61] T. Al-Naffouri and B. Hassibi, "On the distribution of indefinite quadratic forms in Gaussian random variables," in *Proc. 2009 ISIT*, pp. 1744–1748.

[62] D. Raphaeli, "Distribution of noncentral indefinite quadratic forms in complex normal variable," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1002–1007, 1996.

Andrea Rueetschi received a Diploma in Physical Electronics from the University of Neuchâtel (IMT-Institute of Microengineering, Neuchâtel, Switzerland) in March 1997. From 1997 until 2007 he worked as an ASIC Design Engineer for Philips Semiconductors (Zürich, Switzerland and San Jose, CA), RF Micro Devices (formerly Resonext Communications, San Jose, CA) and Marvell Semiconductors (Santa Clara, CA) developing DSP modules for wireless applications. From January 2008 he joined the graduate program at Cornell University (Ithaca, NY) and transferred to UC Davis in 2009.

Anna Scaglione (IEEE F10') received the Laurea (M.Sc. degree) in 1995 and the Ph.D. degree in 1999 from the University of Rome, "La Sapienza." She is currently Professor in Electrical and Computer Engineering at University of California at Davis, where she joined in 2008. She was previously at Cornell University, Ithaca, NY, from 2001 where became Associate Professor in 2006; prior to joining Cornell she was Assistant Professor in the year 2000-2001, at the University of New Mexico. She served as Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2002 to 2005, and serves on the Editorial Board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2008, where she is Area Editor. She has co-organized several special sessions in signal processing and communication conferences; she has served in the past as guest editor for *IEEE Communication Magazine* and the JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING and she is currently serving as guest editor for the JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, for a series of special issues on Smartgrid. She has been in the Signal Processing for Communication Committee from 2004 to 2009. She was general chair of the workshop SPAWC 2005 and keynote speaker in SPAWC 2008 and ISPLC 2010. Dr. Scaglione received the 2000 IEEE Signal Processing Transactions Best Paper Award the NSF Career Award in 2002 and she is co-recipient of the Ellersick Best Paper Award (MILCOM 2005). Her expertise is in the broad area of signal processing for communication systems and networks. Her current research focuses on communication and wireless networks, sensors' systems for monitoring, control and energy management and network science.