

Massive MIMO Channel-Aware Decision Fusion

Domenico Ciuonzo, *Member, IEEE*, Pierluigi Salvo Rossi, *Senior Member, IEEE*, and Subhrakanti Dey, *Senior Member, IEEE*

Abstract—In this paper, we provide a study of channel-aware decision fusion (DF) over a “virtual” multiple-input multiple-output (MIMO) channel in the large-array regime at the DF center (DFC). The considered scenario takes into account channel estimation and inhomogeneous large-scale fading between the sensors and the DFC. The aim is the development of (widely) linear fusion rules, as opposed to the unsuitable optimum log-likelihood ratio (LLR). The proposed rules can effectively benefit from performance improvement via a large array, differently from existing suboptimal alternatives. Performance evaluation, along with theoretical achievable performance and complexity analysis, is presented. Simulation results are provided to confirm the findings. Analogies and differences with uplink communication in a multiuser (massive) MIMO scenario are underlined.

Index Terms—Decision fusion, distributed detection, large-scale MIMO, wireless sensor networks.

I. INTRODUCTION

A. Motivation

DECISION FUSION (DF) in a wireless sensor network (WSN) consists in transmitting local decisions about an observed phenomenon from sensors to a DF center (DFC) for a global decision, with the intent of surveillance and/or situation awareness purposes [2]. Typically all the studies had been focused on a parallel access channels (PACs) with instantaneous or statistical channel-state information (CSI), although some recent works extended to the case of multiple access channels (MACs).

Manuscript received September 29, 2014; revised November 19, 2014; accepted November 23, 2014. Date of publication December 02, 2014; date of current version December 22, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean Pierre Delmas. This work has been partially supported by the ERCIM within the Alain Bensoussan fellowship programme and by the Faculty of Information Technology, Mathematics and Electrical Engineering of the Norwegian University of Science and Technology, Trondheim, Norway, within the project CAMOS. Part of this paper was presented at the Eighth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), A Coruña, Spain, June 2014 (see [1]).

D. Ciuonzo is with University of Naples “Federico II”, DIETI, 80125 Naples, Italy (e-mail: domenico.ciuonzo@ieee.org).

P. Salvo Rossi is with the Department of Industrial and Information Engineering, Second University of Naples, Aversa (CE) 81031, Italy, and also with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway (e-mail: salvorossi@ieee.org).

S. Dey is with the Division of Signals and Systems, Uppsala University, Uppsala, SE 75121, Sweden (e-mail: Subhra.Dey@signal.uu.se).

This paper has supplementary downloadable multimedia material available at <http://ieeexplore.ieee.org> provided by the authors. This includes a pdf file, which contains supplementary proofs and derivations. This material is 216 KB in size.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2014.2376886

Recently, DF over MACs is becoming increasingly attractive, since it exploits both the interfering nature of the broadcast wireless medium (for spectral-efficiency purposes) and the correlated nature of decisions, all regarding the same unknown event being observed. Also, deep fading scenarios are typically overcome with the use of multiple antennas at the DFC, thus determining a “virtual” multiple-input multiple-output (MIMO) channel. The latter architectural choice only demands further complexity at the DFC side and does not affect simplicity of sensors implementation. Indeed, the main difficulty consists in the design of efficient (in terms of complexity and performance) fusion rules at the DFC, since complexity of the optimal fusion rule explodes because of the interference created by the simultaneous transmission of each sensor decision. We remark that the mentioned complexity increase is only due to the MAC choice and does not depend significantly on the number of DFC antennas.

Massive MIMO is an emerging technology for communication applications where antenna arrays with a few hundred elements are placed at base stations, serving many tens of terminals simultaneously [3], [4]. Its potential benefits consist in: (i) significant increase of the capacity and the radiated energy efficiency, (ii) reduced latency through deep-fading mitigation, (iii) simplification of the multiple-access layer and (iv) increased robustness to unintended man-made interference and to intentional jamming [4]. From a mathematical point of view, there are three main advantages in using a large antenna array. Firstly, the effect of small-scale fading is averaged out. Secondly, the random channel vectors between the users and the base station become pairwise orthogonal as the array size grows. Finally, massive MIMO systems allow for reduction of transmitted energy/power.

In our opinion, the advantages offered by massive MIMO may be beneficial in the context of WSNs, especially for DF over MACs. The DFC may be equipped with a large antenna array, thus giving rise to a “virtual” massive MIMO, similarly to the uplink of a multiuser MIMO communication. The proposed architecture, other than enjoying increased spectral efficiency, will mitigate severe energy constraints given by inexpensive sensor nodes (thus providing extended battery life) and facilitate low-complexity (but near-optimal) fusion rules design at the DFC. It is worth remarking that the proposed study relies on the assumption that the number of antennas at the DFC is much larger than the number of transmitting sensors. The latter setup may be represented, analogously as for uplink communications, by a WSNs with many sensors reporting their decision to a DFC equipped with much more antennas. Alternatively, this may be also verified in the case of massively deployed sensor nodes, through the use of appropriate sensor subset selection approaches [5].

B. Related Literature on Decision Fusion

Negative characteristics of the optimal fusion rule, i.e., the log-likelihood ratio (LLR) test, such as numerical instability and strong requirements on system knowledge, motivated the analysis of sub-optimal alternatives such as maximum ratio combining (MRC), equal gain combining and Chair-Varshney maximum likelihood over PACs [2], [6]. Max-Log rule has been studied in [7] and shown to outperform other sub-optimal rules though exhibiting higher requirements on system knowledge.

Distributed detection over MACs was first studied in [8], where perfect compensation of the fading coefficients is assumed for each sensor. Non-coherent modulation and censoring over PACs and MACs were analyzed in [9] with emphasis on processing gain and combining loss. The same scenario was studied in [10], focusing on the error exponents (obtained through the large deviation principle) and the design of energy-efficient modulations for Rayleigh and Rice fading. Optimality of received-energy statistic in Rayleigh fading scenario was demonstrated for diversity MACs with non-identical sensors in [11]. Efficient DF over MACs only with knowledge of the instantaneous channel gains and with the help of power-control and phase-shifting techniques was studied in [12]. Techniques borrowed from direct-sequence spread-spectrum systems were combined with on-off keying (OOK) modulation and censoring for DF in scenarios with statistical CSI [13].

DF over a (virtual) MIMO was first studied in [14], with focus on power-allocation design based on instantaneous CSI, under the framework of J-divergence. Distributed detection with ultra-wideband sensors over MAC was then studied in [15]. The same model was adopted to study data fusion over MIMO channels with amplify-and-forward sensors in [16], [17]. A recent theoretical study on data fusion with amplify and forward sensors and a large-array at the DFC is presented in [18]. Design of several sub-optimal fusion rules for the MIMO scenario was proposed in [19] and a theoretical study on MRC was presented in [20]. Various sub-optimal fusion rules (with reduced system knowledge) for channel-aware DF in the MIMO scenario with instantaneous CSI have been proposed in [19], where “decode-and-fuse” and “decode-then-fuse” approaches are compared through simulation results in the case of few antennas at the DFC. It is worth noticing that in MIMO-DF scenario the LLR is not a viable solution, since it suffers from the exponential growth of the computational complexity with respect to (w.r.t.) the number of sensors and a strong requirement on system knowledge.

C. Related Literature on Massive MIMO

The pioneering work of Marzetta [21] showed that the use of unlimited number of antennas at the base station provides novel interesting phenomena in the context of cellular systems: the detrimental effects of uncorrelated noise and fast fading vanish, the throughput is independent of the size of the cells, the spectral efficiency is independent of the bandwidth, the required transmitted energy per bit decreases, and the simplest sort of precoders and detectors are permitted. Additionally, a novel phe-

nomenon due to inter-cell interference (namely pilot contamination) emerges, and many works have focused on such an issue [22].

In the context of very large size MIMO systems, unified performance analysis for both uplink and downlink of non-cooperative multi-cell systems with linear precoders and detectors is proposed in [23], where the loss of realistic systems due to finite number of antennas w.r.t. asymptotic theoretical performance is evaluated. Performance of linear detectors for the uplink in presence of imperfect CSI is analyzed in [24] focusing on the tradeoff between energy and spectral efficiency, while analytic results for the data rate, symbol error rate and outage probability (in the case of perfect CSI and zero-forcing (ZF) receiver) are derived both for single- and multi-cell systems [25], [26]. A dual analysis for the downlink in the case of conjugate and ZF beamformers is given in [27].

D. Main Results and Paper Organization

The main contributions of this manuscript are related to channel-aware DF over MIMO with instantaneous CSI and more specifically are:

- The large-array regime at the DFC is analyzed, *to the best of our knowledge*, for the first time. The aim is the exploitation of the (approximate) orthogonality of interfering sensors decisions observed from a DFC which employs a massive array, with the intent of its processing simplification. Differently from [19], the considered model takes also into account path loss, shadowing and uncertainty arising from estimated CSI;
- Sub-optimal fusion rules with reduced complexity are derived, consisting of (i) linear-filters plus a soft/hard input fusion scheme and (ii) deflection-maximizer widely-linear (WL) rules, which are compared to existing alternatives, namely the standard MRC. We underline that the derivation of the first class of rules heavily relies on the advantages offered by large-arrays, and exploiting specific conditions (i.e., the favorable propagation discussed in Section II.C) and approximations we are able to obtain a result resembling [7] in a different context. Also, a modified “large-MIMO” version of MRC is developed, which can truly exploit linear SNR increase with array size;
- Closed-form performance, in terms of system false-alarm and detection probabilities, is derived when a large-array is employed (analogously as in [24]), through the *saddle-point approximation* for linear filters and *exactly* for WL rules and (modified) MRC. The proposed analysis underlines the “energy efficiency” of the present approach: indeed, the transmitted energy of each sensor can be reduced by a factor given by the number of antennas in the perfect CSI case (resp. the square root of the number of antennas, in the estimated CSI case) without leading to zero performance. This holds true even in the case of (linear) sub-optimal fusion rules;
- A detailed complexity analysis is presented for all the presented rules, along with efficient approximations which (possibly) avoid time-consuming matrix inversion, following the same lines of research as in [28], [29];

- Extensive simulation studies are performed to compare the considered rules and verify the theoretical findings. Also, analogies with the uplink in multi-user massive MIMO scenario are highlighted throughout the paper.

The manuscript is organized as follows: Section II introduces the system model; in Section III we derive the proposed fusion rules; Section IV presents a corresponding theoretical analysis while Section V contains a discussion on computational complexity; the results are verified in Section VI via simulations; finally, in Section VII we draw some concluding remarks; proofs and derivations are deferred to the Appendices.

II. SYSTEM MODEL

A. WSN Model

In this section we briefly describe the system model¹, illustrated in Fig. 1. We consider a decentralized binary hypothesis test, where K sensors are used to discriminate between the hypotheses of the set $\mathcal{H} \triangleq \{\mathcal{H}_0, \mathcal{H}_1\}$. For example \mathcal{H}_0 and \mathcal{H}_1 may represent the absence and the presence of a specific target of interest, respectively. The k th sensor, $k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$, takes a binary local decision $b_k \in \mathcal{H}$ about the observed phenomenon on the basis of its own measurements. Until otherwise stated, here we do not make any conditional (given $\mathcal{H}_i \in \mathcal{H}$) mutual independence assumption on b_k . Each decision b_k is mapped to a symbol $x_k \in \mathcal{X} = \{-1, +1\}$ representing a binary phase-shift keying (BPSK) modulation²: without loss of generality (w.l.o.g.) we assume that $b_k = \mathcal{H}_i$ maps into $x_k = (2i-1)$, $i \in \{0, 1\}$. The quality of the k th sensor decisions is characterized by the conditional probabilities $P(x_k | \mathcal{H}_i)$. More specifically, we denote $P_{D,k} \triangleq P(x_k = 1 | \mathcal{H}_1)$ and $P_{F,k} \triangleq P(x_k = 1 | \mathcal{H}_0)$ the probability of detection and false alarm of the k th sensor, respectively.

Sensors communicate with a DFC equipped with N receive antennas over a wireless flat-fading MAC in order to exploit diversity and combat signal attenuation due to small-scale fading; this setup determines a distributed or “virtual” MIMO channel [14], [19]. A large-array configuration is considered in

¹Notation—Lower-case (resp. Upper-case) bold letters denote vectors (resp. matrices), with a_n (resp. $a_{n,m}$) being the n th (resp. the (n, m) th) element of \mathbf{a} (resp. \mathbf{A}); upper-case calligraphic letters denote finite sets, with \mathcal{A}^k representing the k -ary Cartesian power of \mathcal{A} ; $\mathbf{O}_{N \times K}$ (resp. \mathbf{I}_N) denotes the $N \times K$ (resp. $N \times N$) null (resp. identity) matrix; $\mathbf{0}_N$ (resp. $\mathbf{1}_N$) denotes the null (resp. ones) vector of length N ; $\mathbb{E}\{\cdot\}$, $\text{var}\{\cdot\}$, $(\cdot)^t$, $(\cdot)^\dagger$, $\Re(\cdot)$, $\Im(\cdot)$ and $\|\cdot\|$ denote expectation, variance, transpose, conjugate transpose, real part, imaginary part and Euclidean norm operators, respectively; $\text{diag}(\mathbf{A})$ denotes the diagonal matrix extracted from \mathbf{A} ; $\underline{\mathbf{a}}$ (resp. $\underline{\mathbf{A}}$) denotes the augmented vector (resp. matrix) of \mathbf{a} (resp. \mathbf{A}), that is $\underline{\mathbf{a}} \triangleq [\mathbf{a}^t \ \mathbf{a}^\dagger]^t$ (resp. $\underline{\mathbf{A}} \triangleq [\mathbf{A}^t \ \mathbf{A}^\dagger]^t$); $P(\cdot)$ and $p(\cdot)$ denote probability mass functions (pmf) and probability density functions (pdf), while $P(\cdot | \cdot)$ and $p(\cdot | \cdot)$ their corresponding conditional counterparts; $\Sigma_{\mathbf{x}}$ (resp. $\bar{\Sigma}_{\mathbf{x}}$) denotes the covariance (resp. the complementary covariance) matrix of the complex-valued random vector \mathbf{x} ; $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \Sigma)$ (resp. $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$) denotes a proper complex (resp. real) normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ , while $\mathcal{N}_{\mathbb{U}}(\cdot)$ and $\mathcal{Q}(\cdot)$ denote the pdf and the complementary cumulative distribution function of a standard normal random variable; finally the symbols \propto , \rightarrow , \xrightarrow{d} and \sim (resp. $\stackrel{\sim}{\sim}$) mean “proportional to”, “tends to”, “tends in distribution to” and “distributed as” (resp. “approximately distributed as”), respectively.

²In the case of an absence/presence task, where \mathcal{H}_0 is much more probable, OOK can be employed for energy-efficiency purposes. Hereinafter we will only consider BPSK; the extension of the presented results to OOK will be object of future work.

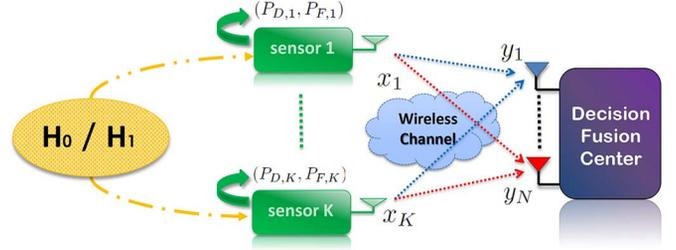


Fig. 1. DF model in presence of a (virtual) MIMO channel.

this paper, that is $N \gg K$; nonetheless we will emphasize the results which still apply in the generic MIMO-DF setup. Also, perfect synchronization³, as in [8], [11], [14], [19], is assumed at the DFC.

We denote: y_n the signal at the n th receive antenna of the DFC after matched filtering and sampling; $g_{n,k}$ the fading coefficient between the k th sensor and the n th receive antenna of the DFC; w_n the additive white Gaussian noise at the n th receive antenna of the DFC. The vector model at the DFC is:

$$\mathbf{y} = \sqrt{\rho} \mathbf{G} \mathbf{x} + \mathbf{w} \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^N$, $\mathbf{G} \in \mathbb{C}^{N \times K}$, $\mathbf{x} \in \mathcal{X}^K$, $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma_w^2 \mathbf{I}_N)$ are the received-signal vector, the channel matrix, the transmitted-signal vector and the noise vector, respectively. In (1) the constant ρ is used to control the energy spent from the generic sensor during the reporting phase. The matrix \mathbf{G} models independent fast fading, geometric attenuation and log-normal shadow fading. The generic coefficient $g_{n,k}$ is then expressed as:

$$g_{n,k} = \sqrt{\beta_k} h_{n,k}, \quad n = 1, \dots, N \quad (2)$$

where the known (since it is assumed constant over many coherence intervals) term $\sqrt{\beta_k}$ models the geometric attenuation and shadow fading⁴ and $h_{n,k}$ denotes the fast fading coefficient between the k th sensor and the n th receive antenna. All the coefficients $h_{n,k}$ are assumed independent and identically distributed (i.i.d.) random variables, with $h_{n,k} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. Based on these assumptions we have

$$\mathbf{G} = \mathbf{H} \mathbf{D}^{1/2}, \quad (3)$$

with $\mathbf{H} \in \mathbb{C}^{N \times K}$ denoting the matrix of fast fading coefficients and $\mathbf{D} \in \mathbb{C}^{K \times K}$ a diagonal matrix where $d_{k,k} = \beta_k$.

B. Channel Training and Estimation

We consider here the general case where the DFC has available estimated instantaneous CSI. We assume that a part of the coherence interval is used for channel training. Let be $\tau \in \mathbb{N}$ the number of symbols used as pilots and $(\tau_c - \tau)$ symbols are used for DF task, where τ_c is the number of symbols within the channel coherence interval. During the training phase, all sensors simultaneously transmit mutually orthogonal pilot sequences of length τ . The pilot sequences are collected in a matrix $(\sqrt{\tau \rho_p} \Phi) \in \mathbb{C}^{\tau \times K}$ (with $\tau \geq K$), where ρ_p controls pilot

³Multiple antennas at the DFC do not make these assumptions harder to verify w.r.t. a single-antenna MAC.

⁴We assume that β_k is constant over n . This assumption is justified since the sensor-DFC distance is typically much higher w.r.t. the inter-antenna distance.

energy, such that $\Phi^\dagger \Phi = \mathbf{I}_K$. Then, the received signal matrix model of the training phase is:

$$\mathbf{Y}_P = \sqrt{\tau \rho_P} \mathbf{G} \Phi^t + \mathbf{W}_P \quad (4)$$

where $\mathbf{Y}_P \in \mathbb{C}^{N \times \tau}$ and $\mathbf{W}_P \in \mathbb{C}^{N \times \tau}$ has i.i.d. elements $w_{P,i,j} \sim \mathcal{N}_\mathbb{C}(0, \sigma_w^2)$. Given the model in (4), the minimum mean-square error (MMSE) estimator of \mathbf{G} given \mathbf{Y}_P is [30]

$$\hat{\mathbf{G}} = \mathbf{Y}_P \Phi^* \frac{1}{\sqrt{\tau \rho_P}} \tilde{\mathbf{D}} \quad (5)$$

where $\tilde{\mathbf{D}} \triangleq (\frac{\sigma_w^2}{\tau \rho_P} \mathbf{D}^{-1} + \mathbf{I}_K)^{-1}$ and can be expressed, after some standard manipulations, in the alternative form

$$\hat{\mathbf{G}} = \left(\mathbf{G} + \frac{1}{\sqrt{\tau \rho_P}} \mathbf{W}_P \Phi^* \right) \tilde{\mathbf{D}}. \quad (6)$$

Also, we define the error matrix as $\Xi \triangleq (\hat{\mathbf{G}} - \mathbf{G})$. It can be shown that $\hat{\mathbf{G}}$ is independent on Ξ (also the columns of both $\hat{\mathbf{G}}$ and Ξ are mutually independent) and that both k th columns of the aforementioned matrices (denoted with $\hat{\mathbf{g}}_k$ and ξ_k , respectively) are proper complex-valued Gaussian vectors with the following moments⁵:

$$\mathbb{E}\{\hat{\mathbf{g}}_k\} = \mathbb{E}\{\xi_k\} = \mathbf{0}_N; \quad (7)$$

$$\mathbb{E}\{\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^\dagger\} = d_{\hat{g},k} \mathbf{I}_N, \quad d_{\hat{g},k} \triangleq \frac{(\tau \rho_P) \beta_k^2}{(\tau \rho_P) \beta_k + \sigma_w^2}; \quad (8)$$

$$\mathbb{E}\{\xi_k \xi_k^\dagger\} = d_{e,k} \mathbf{I}_N, \quad d_{e,k} \triangleq \frac{\sigma_w^2 \beta_k}{(\tau \rho_P) \beta_k + \sigma_w^2}. \quad (9)$$

We remark that as $\tau \rightarrow +\infty$, the perfect instantaneous CSI case is approached.

Finally, it is worth noticing that the overall energy spent by the k th sensor is given by (ϕ_k denotes the k th column of Φ)

$$\begin{aligned} \mathcal{E}_k &\triangleq (\tau_c - \tau) \rho \mathbb{E}\{x_k^2\} + \tau \rho_P \mathbb{E}\{\|\phi_k\|^2\} \\ &= (\tau_c - \tau) \rho + \tau \rho_P \end{aligned} \quad (10)$$

which accounts for both training and transmitting phases, while the energy spent from the whole WSN is $\mathcal{E} = K \cdot \mathcal{E}_k$.

C. Favorable Propagation

From Section II.B we know that $\hat{\mathbf{g}}_k \sim \mathcal{N}_\mathbb{C}(\mathbf{0}_N, d_{\hat{g},k} \mathbf{I}_N)$. Also, the vectors $\hat{\mathbf{g}}_k$, $k \in \mathcal{K}$, are mutually independent. Thus, the so-called *favorable propagation conditions* hold [24]:

$$\frac{1}{N} \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} \approx \mathbf{D}_{\hat{g}}, \quad N \gg K; \quad (11)$$

where $\mathbf{D}_{\hat{g}}$ is a diagonal matrix whose k th element equals $d_{\hat{g},k}$ (cf. (8)). Recent experimental results [3] have shown that practical large arrays approximate well the aforementioned propagation assumption.

⁵Independence of Ξ and $\hat{\mathbf{G}}$ follows since: (i) both matrices are Gaussian distributed (since they are linearly dependent on \mathbf{W}_P); (ii) their columns are zero-mean and (iii) $\mathbb{E}\{\hat{\mathbf{g}}_k \xi_k^\dagger\} = \mathbf{0}_{N \times N}$.

III. FUSION RULES

A. Optimum Rule

The optimal test [31] for this problem⁶ is formulated as

$$\left\{ \Lambda_{\text{opt}} \triangleq \ln \left[\frac{p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_1)}{p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_0)} \right] \right\}_{\hat{\mathcal{H}} = \mathcal{H}_0}^{\hat{\mathcal{H}} = \mathcal{H}_1} \underset{\gamma}{\geq} \gamma \quad (12)$$

where $\hat{\mathcal{H}}$, Λ_{opt} and γ denote the estimated hypothesis, the LLR and the threshold which the LLR is compared to. The threshold γ can be determined to assure a fixed system false-alarm rate (Neyman-Pearson approach), or can be chosen to minimize the probability of error (Bayesian approach) [31]. Exploiting the independence⁷ of \mathbf{y} from \mathcal{H}_i , given \mathbf{x} , an explicit expression of the LLR in (12) is obtained as

$$\begin{aligned} \Lambda_{\text{opt}} &= \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y} | \hat{\mathbf{G}}, \mathbf{x}) P(\mathbf{x} | \mathcal{H}_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y} | \hat{\mathbf{G}}, \mathbf{x}) P(\mathbf{x} | \mathcal{H}_0)} \right] \\ &= \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y} - \sqrt{\rho} \hat{\mathbf{G}} \mathbf{x}\|^2}{\sigma_e^2}\right) P(\mathbf{x} | \mathcal{H}_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y} - \sqrt{\rho} \hat{\mathbf{G}} \mathbf{x}\|^2}{\sigma_e^2}\right) P(\mathbf{x} | \mathcal{H}_0)} \right] \end{aligned} \quad (13)$$

where we have replaced $\mathbf{G} = (\hat{\mathbf{G}} - \Xi)$ in (1) and defined $\sigma_e^2 \triangleq (\rho \sum_{k=1}^K d_{e,k} + \sigma_w^2)$. Unfortunately, the optimal rule in (13) presents several difficulties in its implementation, such as: (i) availability of $\hat{\mathbf{G}}$, $P(\mathbf{x} | \mathcal{H}_i)$ and σ_w^2 and (ii) numerical instability of the expression, due to the presence of exponential functions with large dynamics [7], [19]. More importantly, the exponential growth of the complexity with K is prohibitive for a practical design. Design of sub-optimal DF rules with simpler implementation and (possibly) reduced system knowledge is then extremely desirable. A first study in the latter direction is provided in [19] where sub-optimal fusion rules in the standard MIMO-DF scenario (i.e., with only a few antennas at the DFC) were presented and compared. Also, even though in [19] it was shown that Max-Log rule has the closest performance to the LLR, we discard it from the comparison, since in a WSN with a medium to high number of sensors its exponential complexity remains impractical.

B. MIMO Linear Filters Plus Fusion

Large-Array pdfs of Linear Filters: Here we show how Λ_{opt} is well-approximated by a nice two-step architecture when N grows large. More specifically, the mentioned scheme is composed of a first step (a linear filter) which recovers the soft decision from each sensor and a second step involved in fusing them and obtaining a final decision, as shown in Fig. 2 (details will be clarified hereinafter). First, consider the model in (1) and

⁶In this paper we consider the optimal test conditioned on the estimated matrix $\hat{\mathbf{G}}$. However, the optimal statistic in absolute sense would be one jointly processing $\{\mathbf{y}, \mathbf{Y}_P\}$, as studied in [32]. Unfortunately, the complexity for processing such statistic is impractical, then the need for separation between the ‘‘channel-estimation’’ and the ‘‘channel-aware fusion’’ blocks.

⁷Indeed the directed triple formed by hypothesis, the transmitted-signal vector and the received-signal vector satisfies the Markov property.

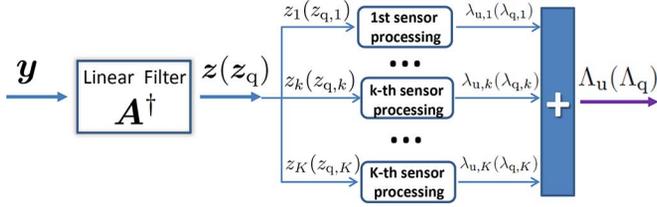


Fig. 2. Large-array architecture for conditionally independent decisions. Max-Log based (resp. CV based) fusion refers to symbols outside (inside) the parentheses.

the linear processing $\mathbf{z} \triangleq \mathbf{A}^\dagger \mathbf{y}$, closely related to matched filter (MF), ZF and MMSE processing, that is:

$$\mathbf{A} \triangleq \begin{cases} \hat{\mathbf{G}} & \text{MF} \\ \hat{\mathbf{G}}(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}})^{-1} & \text{ZF} \\ \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \mathbf{I}_K \right)^{-1} & \text{MMSE} \end{cases} \quad (14)$$

We observe that compression given by \mathbf{z} (since we are assuming $N \gg K$) does not entail loss of detection power for the considered hypothesis testing, as proved by the following lemma.

Lemma 1: The vector \mathbf{z} is a sufficient statistic for the considered hypothesis testing, in the cases of MF, ZF (iff $K \leq N$) and MMSE processing.

Proof: The proof is given in Appendix A. ■

It is also straightforward to show that $\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x}$ is Gaussian-distributed as shown in (15) at the bottom of the page.

Exploiting favorable propagation conditions in (11), $\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x}$ is approximately distributed ($N \gg K$) as follows:

$$\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x} \stackrel{a}{\sim} \begin{cases} \mathcal{N}_{\mathbb{C}}(N\mathbf{D}_{\hat{g}}, \sqrt{\rho}\mathbf{x}, \sigma_e^2 N\mathbf{D}_{\hat{g}}) & \text{MF} \\ \mathcal{N}_{\mathbb{C}}\left(\sqrt{\rho}\mathbf{x}, \frac{\sigma_e^2}{N}\mathbf{D}_{\hat{g}}^{-1}\right) & \text{ZF/MMSE} \end{cases} \quad (16)$$

It is apparent from (16) that all the filters show the desirable independence property $p(\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x}) \approx \prod_{k=1}^K \varphi(z_k | x_k)$, where

$$\varphi(z_k | x_k) \triangleq \begin{cases} \mathcal{N}_{\mathbb{C}}(Nd_{\hat{g},k}, \sqrt{\rho}x_k, \sigma_e^2 Nd_{\hat{g},k}) & \text{MF} \\ \mathcal{N}_{\mathbb{C}}\left(\sqrt{\rho}x_k, \frac{\sigma_e^2}{Nd_{\hat{g},k}}\right) & \text{ZF/MMSE} \end{cases}, \quad (17)$$

in the large-array regime. We will now consider a fusion step which is based either on the complex-valued filter output (i.e., \mathbf{z}) or the estimated (BPSK) transmitted vector, i.e., $\mathbf{z}_q \triangleq \text{sign}(\Re\{\mathbf{z}\})$ ($\mathbf{z}_q \in \mathcal{X}^K$), where:

$$\bar{\varphi}(z_{q,k} | x_k) \triangleq \begin{cases} P_{e,k} & z_{q,k} \neq x_k \\ 1 - P_{e,k} & z_{q,k} = x_k \end{cases}; \quad (18)$$

and we have denoted $P_{e,k} \triangleq \mathcal{Q}\left(\frac{\sqrt{2\rho Nd_{\hat{g},k}}}{\sigma_e}\right)$. The latter is referred to as a *decode-then-fuse* approach [19]; obviously as N grows large, also $P(\mathbf{z}_q | \hat{\mathbf{G}}, \mathbf{x}) \approx \prod_{k=1}^K \bar{\varphi}(z_{q,k} | x_k)$ holds.

Fusion With Conditionally Independent Decisions: Asymptotic large-array pdfs in (16) can be effectively exploited as follows. Here, we assume conditionally independent decisions, i.e., $P(\mathbf{x} | \mathcal{H}_i) = \prod_{k=1}^K P(x_k | \mathcal{H}_i)$; this assumption is frequently used in DF and is used either as a simplifying assumption or a requirement to be satisfied by placing appropriate requirements on the WSN [33]. Then, we adopt the *Max-Log* approximation [34] which leads to the simple rule $\Lambda_u \triangleq \sum_{k=1}^K \lambda_{u,k}$, where $\lambda_{u,k}$ is obtained with the following linear structure with saturation:

$$\lambda_{u,k} \triangleq \begin{cases} c_{1,k}, & f_k < a_{1,k} \\ 2f_k + c_{2,k}, & a_{1,k} \leq f_k < a_{2,k}; \\ c_{3,k}, & f_k \geq a_{2,k} \end{cases} \quad (19)$$

where $c_{1,k} \triangleq \ln\left[\frac{1-P_{D,k}}{1-P_{F,k}}\right]$, $c_{2,k} \triangleq \ln\left[\frac{P_{D,k}}{1-P_{F,k}}\right]$, $c_{3,k} \triangleq \ln\left[\frac{P_{D,k}}{P_{F,k}}\right]$, $a_{1,k} \triangleq \frac{1}{2} \ln\left[\frac{1-P_{D,k}}{P_{D,k}}\right]$ and $a_{2,k} \triangleq \frac{1}{2} \ln\left[\frac{1-P_{F,k}}{P_{F,k}}\right]$. Furthermore, f_k is easily obtained from z_k as follows:

$$f_k \triangleq \begin{cases} \frac{2\sqrt{\rho}\Re\{z_k\}}{\sigma_e^2} & \text{MF} \\ \frac{2\sqrt{\rho}d_{\hat{g},k}N\Re\{z_k\}}{\sigma_e^2} & \text{ZF/MMSE} \end{cases}. \quad (20)$$

The derivation is reported in Appendix B.

On the other hand, when only the quantized vector output \mathbf{z}_q is available, we exploit the *Chair-Varshney rule* [6] (i.e., the LLR $\ln\left[\frac{P(\mathbf{z}_q | \hat{\mathbf{G}}, \mathcal{H}_1)}{P(\mathbf{z}_q | \hat{\mathbf{G}}, \mathcal{H}_0)}\right]$, obtained by exploiting large array pdfs in (16) and assuming $P(\mathbf{x} | \mathcal{H}_i) = \prod_{k=1}^K P(x_k | \mathcal{H}_i)$), in order to obtain the fusion statistic $\Lambda_q \triangleq \sum_{k=1}^K \lambda_{q,k}$, where

$$\lambda_{q,k} \triangleq \hat{b}_k \ln \left[\frac{\alpha_k(P_{D,k})}{\alpha_k(P_{F,k})} \right] + (1 - \hat{b}_k) \ln \left[\frac{\eta_k(P_{D,k})}{\eta_k(P_{F,k})} \right] \quad (21)$$

with $\hat{b}_k \triangleq \frac{(1+z_{q,k})}{2}$, $\alpha_k(P) \triangleq (P \cdot (1 - P_{e,k}) + (1 - P) \cdot P_{e,k})$ and $\eta_k(P) \triangleq (1 - \alpha_k(P))$. Also this derivation is confined to Appendix B. We remark that (21) differs from CV-based rules used in [19] through the addition of correction terms which takes into account the non-ideality of the equivalent communication channel. This is the reason for the loss w.r.t. the Max-Log is less than expected, as shown in Section VI. In Fig. 2 we illustrate a schematic structure of both proposed fusion rules.

Extension to Dependent Decisions: Here we discuss how the presented approach based on linear MIMO filters can be extended to the case $P(\mathbf{x} | \mathcal{H}_i) \neq \prod_{k=1}^K P(x_k | \mathcal{H}_i)$. Starting from

$$\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x} \sim \begin{cases} \mathcal{N}_{\mathbb{C}}\left(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} \sqrt{\rho} \mathbf{x}, \sigma_e^2 \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}}\right) & \text{for MF} \\ \mathcal{N}_{\mathbb{C}}\left(\sqrt{\rho} \mathbf{x}, \sigma_e^2 (\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}})^{-1}\right) & \text{for ZF} \\ \mathcal{N}_{\mathbb{C}}\left(\left(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \mathbf{I}_K\right)^{-1} \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} \sqrt{\rho} \mathbf{x}, \sigma_e^2 \left(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \mathbf{I}_K\right)^{-1} \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \mathbf{I}_K\right)^{-1}\right) & \text{for MMSE} \end{cases} \quad (15)$$

the results in (16) and (17), as N grows, we have $p(\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x}) \approx \prod_{k=1}^K \varphi(z_k | x_k)$. Thus the approximation

$$p(\mathbf{z} | \hat{\mathbf{G}}, \mathcal{H}_i) \approx \sum_{\mathbf{x} \in \mathcal{X}^K} \prod_{k=1}^K \varphi(z_k | x_k) P(\mathbf{x} | \mathcal{H}_i), \quad (22)$$

holds for conditionally dependent decisions. Equation (22) represents a DF model where conditionally dependent decisions are sent over PAC. However, although each sensor decision is “recovered” at the DFC, growth of complexity is still exponential with K .

Therefore, several approaches have been proposed in the literature in order to deal with exponential complexity arising from a non-factorizable $P(\mathbf{x} | \mathcal{H}_i)$, such as: (i) neglecting (only at design stage) the conditional dependence of decisions⁸, i.e., $P(\mathbf{x} | \mathcal{H}_i) \approx \prod_{k=1}^K P(x_k | \mathcal{H}_i)$ (and thus exploiting the results obtained in the conditionally independent scenario) or (ii) (partially) exploiting the conditional dependence via efficient models, such as in [36] via *Copula theory*. This topic falls outside the scope of this paper and we refer the interested reader to the survey (and references therein) [37].

C. Widely-Linear (WL) Rules

A second approach consists in adopting a WL statistic (which is then compared to γ , as for Λ_{opt} in (12)), that is $\Lambda_{\text{WL}} \triangleq \underline{\mathbf{a}}^\dagger \underline{\mathbf{y}}$. The WL approach is motivated by reduced complexity and $\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i$ being an *improper*⁹ complex-valued random vector, that is $\underline{\Sigma}_{\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i} \neq \mathbf{O}_{N \times N}$.

Additionally, several optimization metrics may be considered for obtaining $\underline{\mathbf{a}}$. The best choice (in a Neyman-Pearson sense) would be searching for the WL rule maximizing the global detection probability subject to a global false-alarm rate constraint, as proposed in [38] for a distributed detection problem. Unfortunately, the optimized $\underline{\mathbf{a}}$ presents the following drawbacks: (i) it is not in closed-form, (ii) it requires a non-trivial optimization and (iii) it depends on the prescribed false-alarm constraint. Additionally, the problem under investigation is not a multivariate Gauss-Gauss test (i.e., $\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i \sim \mathcal{N}_{\mathbb{C}}(\underline{\boldsymbol{\mu}}_i, \underline{\boldsymbol{\Sigma}}_i)$) but a test between mixtures of complex GMs (cf. (13)). This would further complicate the optimization problem tackled in [38].

Differently, in this paper we choose $\underline{\mathbf{a}}$ as the maximizer of the so-called deflection measure [39], [40], that is:

$$\underline{\mathbf{a}}_{\text{WL},i} \triangleq \arg \max_{\underline{\mathbf{a}}: \|\underline{\mathbf{a}}\|^2=1} D_i(\underline{\mathbf{a}}) \quad (23)$$

where

$$D_i(\underline{\mathbf{a}}) \triangleq \frac{(\mathbb{E}\{\Lambda_{\text{WL}} | \mathcal{H}_1\} - \mathbb{E}\{\Lambda_{\text{WL}} | \mathcal{H}_0\})^2}{\text{var}\{\Lambda_{\text{WL}} | \mathcal{H}_i\}}$$

$D_0(\underline{\mathbf{a}})$ and $D_1(\underline{\mathbf{a}})$ correspond to the normal [39] and modified [40] deflections, respectively. Hereinafter we will denote $\Lambda_{\text{WL},i} \triangleq (\underline{\mathbf{a}}_{\text{WL},i}^\dagger \underline{\mathbf{y}})$ the WL statistic maximizing $D_i(\underline{\mathbf{a}})$. The explicit expression for $\underline{\mathbf{a}}_{\text{WL},i}$ is given in the following proposition.

⁸We recall that the approximation $P(\mathbf{x} | \mathcal{H}_i) \approx \prod_{k=1}^K P(x_k | \mathcal{H}_i)$ is the Kullback-Leibler minimizer between the joint pdf $P(\mathbf{x} | \mathcal{H}_i)$ and a desirable factorizable model [35].

⁹The proof is given as supplementary material of this manuscript.

Proposition 1: The vector $\underline{\mathbf{a}}_{\text{WL},i}$, being the optimal solution of the optimization problem in (23), is given by:

$$\underline{\mathbf{a}}_{\text{WL},i} = \left(\underline{\boldsymbol{\Sigma}}_{\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i}^{-1} \hat{\mathbf{G}} \underline{\boldsymbol{\mu}}_{1,0} \right) \times \left\| \underline{\boldsymbol{\Sigma}}_{\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i}^{-1} \hat{\mathbf{G}} \underline{\boldsymbol{\mu}}_{1,0} \right\|^{-1} \quad (24)$$

where:

$$\underline{\boldsymbol{\mu}}_{1,0} \triangleq 2[(P_{D,1} - P_{F,1}) \quad \cdots \quad (P_{D,K} - P_{F,K})]^\dagger; \quad (25)$$

$$\underline{\boldsymbol{\Sigma}}_{\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i} = \left(\rho \hat{\mathbf{G}} \underline{\boldsymbol{\Sigma}}_{\mathbf{x} | \mathcal{H}_i} \hat{\mathbf{G}}^\dagger + \sigma_e^2 \mathbf{I}_{2N} \right). \quad (26)$$

Proof: The proof is given as supplementary material. ■

Maximization of deflection measures is widely used in the design of (widely) linear rules for DF, since $\underline{\mathbf{a}}_{\text{WL},i}$ is always in closed-form and also literature has shown acceptable performance loss w.r.t. the LLR in analogous DF setups [40], [41]. Additionally, WL rules take into account the general case of conditionally dependent decisions, while requiring *only* the first and second order moments of $\mathbf{x} | \mathcal{H}_i$.

Finally, it is worth remarking that deflection-optimization is optimal only for a mean-shifted Gauss-Gauss hypothesis testing (i.e., $\underline{\mathbf{y}} | \hat{\mathbf{G}}, \mathcal{H}_i \sim \mathcal{N}_{\mathbb{C}}(\underline{\boldsymbol{\mu}}_i, \underline{\boldsymbol{\Sigma}})$) [31], where normal and modified deflections coincide and they both represent the SNR of the statistic under Neyman-Pearson framework (remarkably in the latter case the LLR is a linear function of measurements). Therefore, we cannot claim any (asymptotic) optimality property for the proposed WL rules.

D. (Modified) MRC

The LLR in (13) can be simplified under the assumption of perfect sensors [7], [13], [19], [20], i.e., $P(\mathbf{x} = \mathbf{1}_K | \mathcal{H}_1) = P(\mathbf{x} = -\mathbf{1}_K | \mathcal{H}_0) = 1$. In this case $\mathbf{x} \in \{\mathbf{1}_K, -\mathbf{1}_K\}$ and (13) reduces to [19]:

$$\ln \left[\frac{\exp\left(-\frac{\|\underline{\mathbf{y}} - \sqrt{\rho} \hat{\mathbf{G}} \mathbf{1}_K\|^2}{\sigma_e^2}\right)}{\exp\left(-\frac{\|\underline{\mathbf{y}} + \sqrt{\rho} \hat{\mathbf{G}} \mathbf{1}_K\|^2}{\sigma_e^2}\right)} \right] \propto \Re(\underline{\mathbf{a}}_{\text{MRC}}^\dagger \underline{\mathbf{y}}) \triangleq \Lambda_{\text{MRC}} \quad (27)$$

where $\underline{\mathbf{a}}_{\text{MRC}} \triangleq (\hat{\mathbf{G}} \mathbf{1}_K)$ and terms independent from $\underline{\mathbf{y}}$ have been incorporated in γ as in (12). It is worth noticing that the MRC is a *sub-optimal* rule since, in the practice, the sensor local decisions are far from being perfect. However, it has been proved in [19] that MRC is the low-SNR approximation of the optimum test in (13) when local performances of sensors are identical and $\hat{\mathbf{G}} = \mathbf{G} = \mathbf{H}$. Also, it is interesting to note that the MRC in the case of estimated CSI is simply obtained by replacing \mathbf{G} in the MRC formula with perfect CSI [20] with its estimate $\hat{\mathbf{G}}$, i.e., knowledge of error statistics is not needed.

Additionally, in order to exploit the linear SNR increase with N , which would inevitably make the fusion process mainly dependent on the “sensing” errors (and consequently MRC rule becomes clearly inappropriate, since its design is unaware of sensing errors), we devise an alternative form of MRC, denoted as *modified MRC* (mMRC), which is given by:

$$\Lambda_{\text{mMRC}} \triangleq \Re(\underline{\mathbf{a}}_{\text{mMRC}}^\dagger \underline{\mathbf{y}}) \quad (28)$$

where $\mathbf{a}_{\text{mMRC}} \triangleq (\hat{\mathbf{G}}\mathbf{D}_g^{-1}\mathbf{1}_K)$. It can be observed that mMRC applies a sort of “static” zero-forcing (via \mathbf{D}_g^{-1}) in order to remove rule dependence on large-scale fading coefficients (which becomes detrimental as the SNR grows). Also, it can be shown that (28) asymptotically approaches (as N grows large) the performance of the counting rule, which has been shown to be robust even with heterogeneous sensors [42]. In fact, exploiting (11) as N grows, we have:

$$(\Lambda_{\text{mMRC}}/N) \approx \sqrt{\rho} \sum_{k=1}^K x_k + \Re\{w_m\} \quad (29)$$

where $w_m \sim \mathcal{N}_{\mathbb{C}}(0, \frac{\sigma_w^2}{N} \sum_{k=1}^K d_{g,k}^{-1})$. We will show in Section VI that mMRC significantly outperforms MRC in the large-array regime, since it enjoys the same asymptotic form as the standard MRC in the scenario $\beta_k = \beta$, $k \in \mathcal{K}$ (cf. [20]).

Remarks on Terminology: It is worth noticing that the terminology in this paper slightly differs from that used in massive MIMO literature. In fact, we refer to MRC as the sub-optimal (fusion) rule which assumes $\mathbf{x} \in \{-\mathbf{1}_K, \mathbf{1}_K\}$ and thus resembles a MIMO-MRC combiner [43]. Differently, the term matched filter (MF) is here used to denote the classical linear filter operating on \mathbf{x} studied in Section III.B. Thus, in view of the mentioned reasons, the terms will refer to different fusion rules and are not to be seen as synonymous.

IV. LARGE-ARRAY PERFORMANCE ANALYSIS

A. Performance Measures

In this manuscript we compare the performance of the proposed rules both in terms of instantaneous channel (IC) system false alarm and detection probabilities, defined as

$$P_{F_0}(\gamma, \hat{\mathbf{G}}) \triangleq \Pr\{\Lambda > \gamma \mid \hat{\mathbf{G}}, \mathcal{H}_0\}, \quad (30)$$

$$P_{D_0}(\gamma, \hat{\mathbf{G}}) \triangleq \Pr\{\Lambda > \gamma \mid \hat{\mathbf{G}}, \mathcal{H}_1\}, \quad (31)$$

and the corresponding channel-averaged (CA) counterparts

$$P_{F_0}(\gamma) \triangleq \mathbb{E}_{\hat{\mathbf{G}}}\{P_{F_0}(\gamma, \hat{\mathbf{G}})\} = \Pr\{\Lambda > \gamma \mid \mathcal{H}_0\}, \quad (32)$$

$$P_{D_0}(\gamma) \triangleq \mathbb{E}_{\hat{\mathbf{G}}}\{P_{D_0}(\gamma, \hat{\mathbf{G}})\} = \Pr\{\Lambda > \gamma \mid \mathcal{H}_1\}, \quad (33)$$

where Λ is the generic statistic employed at the DFC.

Hereinafter we will analyze the asymptotic performance (i.e., under the limit $N \rightarrow +\infty$) and we will show that the IC and CA system probabilities approach the same value in the two cases being considered. The two scenarios analyzed in this paper are the same as in [21] and consist of: (a) either taking the limit $N \rightarrow +\infty$ and considering an energy cut in both training and reporting phases as $\rho = \frac{\bar{\rho}}{\sqrt{N}}$ and $\rho_P = \frac{\bar{\rho}_P}{\sqrt{N}}$ or (b) taking the limit $N \rightarrow +\infty$ and considering an energy cut only in the reporting phase as $\rho = \frac{\bar{\rho}}{N}$. Scenario (a) refers to a setup with a fast-fading channel, where the energy spent in the training phase is not negligible, while scenario (b) refers to a case where the energy used for training is negligible and thus transmit energy can be scaled down *aggressively* [21].

Finally, we remark that both scenarios are of clear significance for WSN applications, since the results presented hereinafter are achieved in a regime where each sensor keeps re-

ducing the energy spent as the number of antennas at the DFC grows. Therefore each sensor behaves “efficiently” (in terms of consumed energy) and consequently prolongs its expected battery life.

B. Large MIMO Linear Filters (Cond. Indep. Decisions)

Here we derive asymptotic performance of MIMO linear filters in the case of conditionally independent decisions; the scenario with conditionally dependent decisions is more cumbersome and will be tackled elsewhere. First, it can be readily shown that MF/ZF/MMSE performance coincide in both scenarios. Also, this clearly holds when either Max-Log or CV is used in the fusion step. Indeed, asymptotically (20) and (18) yield

$$f_k \mid x_k \sim \mathcal{N}(m_k x_k, m_k), \quad P_{e,k} = \mathcal{Q}(\sqrt{m_k}), \quad (34)$$

where $m_k \triangleq \frac{2\bar{\rho}\bar{\rho}_P\tau\beta_k^2}{\sigma_w^4}$ in scenario (a), while in scenario (b) $m_k \triangleq \frac{2\bar{\rho}d_{g,k}}{\sigma_w^2}$ holds.

Also, it is apparent that false-alarm and system detection probabilities $P_{F_0, \text{th}}(\gamma) \triangleq \lim_{N \rightarrow +\infty} [P_{F_0}(\gamma, \hat{\mathbf{G}})]$ and $P_{D_0, \text{th}}(\gamma) \triangleq \lim_{N \rightarrow +\infty} [P_{D_0}(\gamma, \hat{\mathbf{G}})]$ in scenarios (a) and (b) have not a tractable form. Hence, we pass through the evaluation of corresponding cumulant generating functions (CGFs), denoted here as $\mathcal{C}(s \mid \mathcal{H}_i)$, which have an easier form. In fact, as $N \rightarrow +\infty$ the equivalent channel for DF model is a PAC and thus $\mathcal{C}(s \mid \mathcal{H}_i)$ is simply given by the sum of the individual contributions related to each sensor. Finally, $P_{F_0, \text{th}}(\gamma)$ and $P_{D_0, \text{th}}(\gamma)$ can be obtained (approximately) starting from the corresponding CGFs with the use of the *saddlepoint-approach* [44].

W.l.o.g. we derive hereinafter the CGF with reference to a generic m_k (performance in scenarios (a) and (b) are obtained by using corresponding expressions reported after (34)); also, for simplicity we will use the notation $(P_{1,k}, P_{0,k}) = (P_{D,k}, P_{F,k})$. First, we express the CGF as:

$$\mathcal{C}(s \mid \mathcal{H}_i) = \sum_{k=1}^K \ln[\mathcal{G}_k(s \mid \mathcal{H}_i)] \quad (35)$$

where $\mathcal{G}_k(s \mid \mathcal{H}_i)$ is the moment generating function (MGF) of the corresponding contribution (that is, $\lambda_{u,k}$ or $\lambda_{q,k}$). In the case of CV-based fusion this MGF is:

$$\mathcal{G}_k^{\text{CV}}(s \mid \mathcal{H}_i) = \exp(s\xi_k)\alpha_k(P_{i,k}) + \exp(s\psi_k)\eta_k(P_{i,k}) \quad (36)$$

On the other hand, in the case of Max-Log the MGF is more complicated:

$$\mathcal{G}_k^{\text{mlog}}(s \mid \mathcal{H}_i) = \sum_{x_k \in \mathcal{X}} \mathcal{G}_k^{\text{mlog}}(s \mid x_k)(P_{i,k})^{b_k}(1 - P_{i,k})^{(1-b_k)} \quad (37)$$

where $b_k \triangleq (\frac{1+x_k}{2})$ and $\mathcal{G}_k^{\text{mlog}}(s \mid x_k)$ is given as

$$\begin{aligned} \mathcal{G}_k^{\text{mlog}}(s \mid x_k) &= \rho_k \exp(c_{1,k}s) + \vartheta_k \exp(c_{3,k}s) \\ &+ \exp(c_{2,k}s + 2m_k(x_k s + s^2)) \\ &\times [\mathcal{Q}(\tilde{a}_{1,k} - 2\sqrt{m_k}s) - \mathcal{Q}(\tilde{a}_{2,k} - 2\sqrt{m_k}s)]. \end{aligned} \quad (38)$$

Also, in (38) we have denoted $\tilde{a}_{1,k} \triangleq \left(\frac{a_{1,k} - m_k x_k}{\sqrt{m_k}}\right)$, $\tilde{a}_{2,k} \triangleq \left(\frac{a_{2,k} - m_k x_k}{\sqrt{m_k}}\right)$, $\rho_k \triangleq [1 - \mathcal{Q}(\tilde{a}_{1,k})]$ and $\vartheta_k \triangleq \mathcal{Q}(\tilde{a}_{2,k})$, respectively.

Once the CGFs $\mathcal{C}(s | \mathcal{H}_i)$ ($\mathcal{H}_i \in \mathcal{H}$) are obtained, the probabilities $P_{F_0, \text{th}}(\gamma)$ and $P_{D_0, \text{th}}(\gamma)$ can be approximated respectively as [44]:

$$P_{F_0, \text{th}}(\gamma) \approx \mathcal{Q}(r_0[\hat{s}_0(\gamma)]) + \mathcal{N}_U(r_0[\hat{s}_0(\gamma)]) \times \left[\frac{1}{g_0[\hat{s}_0(\gamma)]} - \frac{1}{r_0[\hat{s}_0(\gamma)]} \right] \quad (39)$$

$$[1 - P_{D_0, \text{th}}(\gamma)] \approx \mathcal{Q}(-r_1[\hat{s}_1(\gamma)]) - \mathcal{N}_U(r_1[\hat{s}_1(\gamma)]) \times \left[\frac{1}{g_1[\hat{s}_1(\gamma)]} - \frac{1}{r_1[\hat{s}_1(\gamma)]} \right] \quad (40)$$

where we have denoted

$$r_i[s] \triangleq \text{sign}(s) \sqrt{2 \left[s \frac{\partial \mathcal{C}(s | \mathcal{H}_i)}{\partial s} - \mathcal{C}(s | \mathcal{H}_i) \right]} \quad (41)$$

$$g_i[s] \triangleq s \sqrt{\frac{\partial^2 \mathcal{C}(s | \mathcal{H}_i)}{\partial s^2}} \quad (42)$$

and $\hat{s}_i(\gamma)$ is the unique solution of the so-called saddlepoint equation¹⁰, that is $\frac{\partial \mathcal{C}(s | \mathcal{H}_i)}{\partial s} = \gamma$.

For example, $P_{F_0, \text{th}}(\gamma)$ approximation is obtained through the following steps: (i) for a given γ , $\frac{\partial \mathcal{C}(s | \mathcal{H}_0)}{\partial s} = \gamma$ is solved w.r.t. s and the solution $\hat{s}_0(\gamma)$ is obtained; (ii) the terms $r_0[s]$ and $g_0[s]$ in (41)–(42) are evaluated at $s = \hat{s}_0(\gamma)$; (iii) $g_0[\hat{s}_0(\gamma)]$ and $r_0[\hat{s}_0(\gamma)]$ are substituted in (39) for final evaluation.

C. WL Rules

Here, we first derive the exact IC system probabilities for WL rules and then we consider the two asymptotic scenarios described in Section IV.A. Since $\Lambda_{\text{WL}, i}$ arises from WL processing of \mathbf{y} , we have $\Lambda_{\text{WL}, i} | \hat{\mathbf{G}}, \mathbf{x} \sim \mathcal{N}(\mathbb{E}\{\Lambda_{\text{WL}, i} | \hat{\mathbf{G}}, \mathbf{x}\}, \text{var}\{\Lambda_{\text{WL}, i} | \hat{\mathbf{G}}, \mathbf{x}\})$, where:

$$\mathbb{E}\{\Lambda_{\text{WL}, i} | \hat{\mathbf{G}}, \mathbf{x}\} = \frac{\sqrt{\rho} \boldsymbol{\mu}_{1,0}^t \hat{\mathbf{G}}^\dagger \boldsymbol{\Sigma}^{-1} \mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i \hat{\mathbf{G}} \mathbf{x}}{\left\| \boldsymbol{\Sigma}^{-1} \mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i \hat{\mathbf{G}} \boldsymbol{\mu}_{1,0} \right\|}; \quad (43)$$

$$\text{var}\{\Lambda_{\text{WL}, i} | \hat{\mathbf{G}}, \mathbf{x}\} = \sigma_e^2. \quad (44)$$

It is apparent from (44) that $\text{var}\{\Lambda_{\text{WL}, i} | \hat{\mathbf{G}}, \mathbf{x}\}$ does not depend on \mathbf{x} . Thus, w.l.o.g. we can define $\Lambda_{\text{WL}, i}^\circ \triangleq \frac{\Lambda_{\text{WL}, i}}{\sigma_e}$ and evaluate the performance¹¹ in terms of $\Lambda_{\text{WL}, i}^\circ$. Based on (43) and (44), $\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathcal{H}_j$ is distributed as follows:

$$\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathcal{H}_j \sim \sum_{\mathbf{x} \in \mathcal{X}^K} P(\mathbf{x} | \mathcal{H}_j) \mathcal{N}(\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}, 1); \quad (45)$$

where

$$\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\} = \frac{\sqrt{\rho} \boldsymbol{\mu}_{1,0}^t \hat{\mathbf{G}}^\dagger \boldsymbol{\Sigma}^{-1} \mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i \hat{\mathbf{G}} \mathbf{x}}{\sigma_e \left\| \boldsymbol{\Sigma}^{-1} \mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i \hat{\mathbf{G}} \boldsymbol{\mu}_{1,0} \right\|}. \quad (46)$$

Equation (45) denotes a mixture of 2^K one-dimensional real-valued Gaussians, all depending on $\hat{\mathbf{G}}$ (which is random)

¹⁰Explicit expressions of first/second derivative of $\mathcal{C}(s | \mathcal{H}_i)$, for both CV and Max-Log based rules, are given as supplementary material.

¹¹Indeed $\Lambda_{\text{WL}, i}^\circ$ is statistically equivalent to $\Lambda_{\text{WL}, i}$, since a positive scaling factor (independent on \mathbf{y}) does not alter rule performance.

through their mean. A direct application of (45) and (31) leads to the following exact expression for $P_{D_0}(\gamma, \hat{\mathbf{G}})$ of $\Lambda_{\text{WL}, i}^\circ$ (identical steps apply for evaluation of $P_{F_0}(\gamma, \hat{\mathbf{G}})$):

$$P_{D_0}(\gamma, \hat{\mathbf{G}}) = \sum_{\mathbf{x} \in \mathcal{X}^K} P(\mathbf{x} | \mathcal{H}_1) \mathcal{Q}\left(\gamma - \mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}\right). \quad (47)$$

Such expression can now be evaluated in the asymptotic scenarios (a) and (b) described in Section IV.A. This is accomplished by simply taking the limit $N \rightarrow +\infty$ of $\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}$ under the assumptions determining scenarios (a) and (b), respectively. The reason is the dependence of $P_{D_0}(\gamma, \hat{\mathbf{G}})$ on N is only through the means of the mixture components. Hence, for scenario (a) we obtain

$$\lim_{N \rightarrow +\infty} \left(\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\} \right)_a = \frac{\sqrt{2\tau \bar{\rho} \bar{\rho}_P} \boldsymbol{\mu}_{1,0}^t \bar{\mathbf{V}}_i \mathbf{D}^2 \mathbf{x}}{\sigma_w^2 \sqrt{\boldsymbol{\mu}_{1,0}^t \bar{\mathbf{V}}_i \mathbf{D}^2 \bar{\mathbf{V}}_i \boldsymbol{\mu}_{1,0}}} \quad (48)$$

where $\bar{\mathbf{S}}_i \triangleq \left(\frac{\sigma_w^2}{\bar{\rho}} \boldsymbol{\Sigma}_{\mathbf{x} | \mathcal{H}_i}^{-1} + \frac{2\tau \bar{\rho}_P}{\sigma_w^2} \mathbf{D}^2\right)$ and $\bar{\mathbf{V}}_i \triangleq (\mathbf{I}_K - \frac{2\tau \bar{\rho}_P}{\sigma_w^2} \mathbf{D}^2 \bar{\mathbf{S}}_i^{-1})$. On the other hand, in the asymptotic scenario (b) $\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}$ converges to:

$$\lim_{N \rightarrow +\infty} \left(\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\} \right)_b = \frac{\sqrt{2\bar{\rho}} \boldsymbol{\mu}_{1,0}^t \check{\mathbf{V}}_i \mathbf{D}_g \mathbf{x}}{\sigma_w \sqrt{\boldsymbol{\mu}_{1,0}^t \check{\mathbf{V}}_i \mathbf{D}_g \check{\mathbf{V}}_i \boldsymbol{\mu}_{1,0}}} \quad (49)$$

where $\check{\mathbf{S}}_i \triangleq \left(\frac{\sigma_w^2}{\bar{\rho}} \boldsymbol{\Sigma}_{\mathbf{x} | \mathcal{H}_i}^{-1} + 2\mathbf{D}_g\right)$ and $\check{\mathbf{V}}_i \triangleq (\mathbf{I}_K - 2\mathbf{D}_g \check{\mathbf{S}}_i^{-1})$. Derivation of both (48) and (49) is given as supplementary material. Based on the above results it is apparent that, in both considered scenarios, (47) becomes a real-valued Gaussian mixture with 2^K components that *can be evaluated offline*.

Finally, we remark that the latter equations are greatly simplified in the case of conditionally uncorrelated decisions¹², $\mathbb{E}\{x_k x_j | \mathcal{H}_j\} = \mathbb{E}\{x_k | \mathcal{H}_j\} \cdot \mathbb{E}\{x_\ell | \mathcal{H}_j\}$, $k \neq \ell$. In fact

$$\lim_{N \rightarrow +\infty} \left(\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\} \right)_a = \frac{\sqrt{2\tau \bar{\rho} \bar{\rho}_P} \sum_{k=1}^K \beta_k^2 \bar{v}_{i,k} x_k \boldsymbol{\mu}_{1,0,k}}{\sigma_w^2 \sqrt{\sum_{k=1}^K \beta_k^2 \bar{v}_{i,k}^2 \boldsymbol{\mu}_{1,0,k}^2}} \quad (50)$$

in scenario (a), while scenario (b) yields:

$$\lim_{N \rightarrow +\infty} \left(\mathbb{E}\{\Lambda_{\text{WL}, i}^\circ | \hat{\mathbf{G}}, \mathbf{x}\} \right)_b = \frac{\sqrt{2\bar{\rho}} \sum_{k=1}^K d_{\hat{g},k} \check{v}_{i,k} x_k \boldsymbol{\mu}_{1,0,k}}{\sigma_w \sqrt{\sum_{k=1}^K d_{\hat{g},k}^2 \check{v}_{i,k}^2 \boldsymbol{\mu}_{1,0,k}^2}} \quad (51)$$

where we have denoted $\bar{v}_{i,k} \triangleq \sigma_w^4 \times (\sigma_w^4 + 2\tau \bar{\rho} \bar{\rho}_P \beta_k^2 \text{var}\{x_k | \mathcal{H}_i\})^{-1}$ and $\check{v}_{i,k} \triangleq \sigma_w^2 \times (\sigma_w^2 + 2\bar{\rho} d_{\hat{g},k} \text{var}\{x_k | \mathcal{H}_i\})^{-1}$, respectively.

D. (Modified) MRC

The IC system probabilities for MRC can be obtained in closed form as follows. The derivation is analogous to the expression obtained in the simpler case $\mathbf{D} = \mathbf{I}_K$ and

¹²Even though, from a theoretical point of view, conditionally uncorrelated decisions is a weaker assumption than conditional independence, it should be understood that in many relevant cases in DF both properties are to be considered as synonymous.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE CONSIDERED RULES: N IS THE NUMBER OF ANTENNAS AT THE DFC; K IS THE NUMBER OF SENSORS;
 $\tau \geq K$ IS THE LENGTH OF PILOT SEQUENCES EMPLOYED FOR CHANNEL ESTIMATION

Fusion Rule	Complexity for each realization of \mathbf{G}	Complexity for each realization of \mathbf{x}
Optimum (LLR)	$\mathcal{O}(N\tau K)$	$\mathcal{O}(NK2^K)$
MF + Max-Log / CV	$\mathcal{O}(N\tau K)$	$\mathcal{O}(NK)$
ZF/MMSE + Max-Log / CV	$\mathcal{O}(N\tau K + NK^2 + K^3)$ (approx. inversion $\mathcal{O}(N\tau K + NK^2)$)	$\mathcal{O}(NK)$
WL rules	$\mathcal{O}(N\tau K + NK^2 + K^3)$ (approx. inversion $\mathcal{O}(N\tau K + NK^2)$)	$\mathcal{O}(N)$
(m)MRC	$\mathcal{O}(N\tau K + NK)$	$\mathcal{O}(N)$

$\hat{\mathbf{G}} = \mathbf{G}$ (i.e., no shadowing/path loss and perfect CSI) contained in [20] and it is not reported for the sake of brevity. We start by recalling that $\mathbf{y} | \hat{\mathbf{G}}, \mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\sqrt{\rho}\hat{\mathbf{G}}\mathbf{x}, \sigma_e^2 \mathbf{I}_N)$ (cf. (13)). From inspection of (27), we observe that Λ_{MRC} arises from WL processing of \mathbf{y} . As an immediate consequence, $\Lambda_{\text{MRC}} | \hat{\mathbf{G}}, \mathbf{x} \sim \mathcal{N}(\mathbb{E}\{\Lambda_{\text{MRC}} | \hat{\mathbf{G}}, \mathbf{x}\}, \text{var}\{\Lambda_{\text{MRC}} | \hat{\mathbf{G}}, \mathbf{x}\})$, where:

$$\mathbb{E}\{\Lambda_{\text{MRC}} | \hat{\mathbf{G}}, \mathbf{x}\} = \sqrt{\rho}\Re\{\mathbf{a}_{\text{MRC}}^\dagger \hat{\mathbf{G}}\mathbf{x}\}; \quad (52)$$

$$\text{var}\{\Lambda_{\text{MRC}} | \hat{\mathbf{G}}, \mathbf{x}\} = \frac{1}{2}\sigma_e^2 \|\mathbf{a}_{\text{MRC}}\|^2. \quad (53)$$

It is worth noticing that $\text{var}\{\Lambda_{\text{MRC}} | \hat{\mathbf{G}}, \mathbf{x}\}$ does not depend on \mathbf{x} . Thus, w.l.o.g. we can define $\Lambda_{\text{MRC}}^\circ \triangleq \frac{\sqrt{2}\Lambda_{\text{MRC}}}{\sigma_e \|\mathbf{a}_{\text{MRC}}\|}$ and evaluate the performance in terms of $\Lambda_{\text{MRC}}^\circ$. It is readily shown that $\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathcal{H}_j$ is distributed as

$$\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathcal{H}_j \sim \sum_{\mathbf{x} \in \mathcal{X}^K} P(\mathbf{x} | \mathcal{H}_j) \mathcal{N}(\mathbb{E}\{\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}, 1) \quad (54)$$

where $\mathbb{E}\{\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\} = \frac{\sqrt{2\rho}\Re\{\mathbf{1}_K^\dagger \hat{\mathbf{G}} \hat{\mathbf{G}}\mathbf{x}\}}{\sigma_e \sqrt{\mathbf{1}_K^\dagger \hat{\mathbf{G}} \hat{\mathbf{G}} \mathbf{1}_K}}$. Equation (54) de-

notes a mixture of 2^K one-dimensional real-valued Gaussians, all depending on $\hat{\mathbf{G}}$ (which is random) through their mean. A direct application of (54) and (31) leads to the following exact expression for $P_{D_0}(\gamma, \hat{\mathbf{G}})$ of $\Lambda_{\text{MRC}}^\circ$ (identical steps apply for evaluation of $P_{F_0}(\gamma, \hat{\mathbf{G}})$):

$$P_{D_0}(\gamma, \hat{\mathbf{G}}) = \sum_{\mathbf{x} \in \mathcal{X}^K} P(\mathbf{x} | \mathcal{H}_1) \mathcal{Q}(\gamma - \mathbb{E}\{\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}). \quad (55)$$

The expression in (55) can now be evaluated in the large-array regime for asymptotic scenarios (a) and (b), described in Section IV.A. This is accomplished by simply taking the limit $N \rightarrow +\infty$ of $\mathbb{E}\{\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\}$ under the assumptions determining scenarios (a) and (b), respectively. The reason is the dependence of $P_{D_0}(\gamma, \hat{\mathbf{G}})$ on N is only through the means of the mixture components. Thus, considering scenario (a) gives:

$$\lim_{N \rightarrow +\infty} (\mathbb{E}\{\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\})_a = \frac{\sqrt{2\tau\rho\bar{\rho}_F} \sum_{k=1}^K \beta_k^2 x_k}{\sigma_w^2 \sqrt{\sum_{k=1}^K \beta_k^2}} \quad (56)$$

Differently, in the case of scenario (b) we obtain:

$$\lim_{N \rightarrow +\infty} (\mathbb{E}\{\Lambda_{\text{MRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\})_b = \frac{\sqrt{2\bar{\rho}} \sum_{k=1}^K d_{\hat{g},k} x_k}{\sigma_w \sqrt{\sum_{k=1}^K d_{\hat{g},k}}}. \quad (57)$$

Derivation of both (56) and (57) is given as supplementary material. Based on the above results it is worth remarking that, in the asymptotic scenarios considered, (55) is a real-valued Gaussian mixture with 2^K components that *can be evaluated offline*.

By similar reasoning, we get the performance of mMRC in the asymptotic scenarios (a) and (b). Indeed, it can be shown that, after defining $\Lambda_{\text{mMRC}}^\circ \triangleq \frac{\sqrt{2}\Lambda_{\text{mMRC}}}{\sigma_e \|\mathbf{a}_{\text{mMRC}}\|}$, asymptotic scenario (a) gives

$$\lim_{N \rightarrow +\infty} (\mathbb{E}\{\Lambda_{\text{mMRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\})_a = \frac{\sqrt{2\tau\rho\bar{\rho}_F} \sum_{k=1}^K x_k}{\sigma_w^2 \sqrt{\sum_{k=1}^K \beta_k^{-2}}}, \quad (58)$$

while scenario (b) yields

$$\lim_{N \rightarrow +\infty} (\mathbb{E}\{\Lambda_{\text{mMRC}}^\circ | \hat{\mathbf{G}}, \mathbf{x}\})_b = \frac{\sqrt{2\bar{\rho}} \sum_{k=1}^K x_k}{\sigma_w \sqrt{\sum_{k=1}^K d_{\hat{g},k}^{-1}}}. \quad (59)$$

V. COMPLEXITY ANALYSIS AND COMPUTATIONALLY EFFICIENT APPROXIMATIONS

In Table I we compare the computational complexity¹³ of the proposed rules, by separating the contribution related to $\hat{\mathbf{G}}$ and the computations required whenever each new \mathbf{x} is transmitted (but differently $\hat{\mathbf{G}}$ does not change). It is worth noticing that Table I takes into account also the complexity needed for channel training (i.e., obtaining $\hat{\mathbf{G}}$), which is $\mathcal{O}(N\tau K)$ (we remark that $\tau \geq K$, cf. Section II.B).

It is apparent that the optimum rule (i.e., the LLR) is unfeasible, especially when K is very large. Differently, all the proposed rules have polynomial complexity w.r.t. both K and N , becoming linear when slowly fading scenarios are considered (i.e., when the left column of complexity contribution can be considered as a “static step”).

The computational complexities of MF, ZF and MMSE with Max-Log (resp. CV) based rules are mainly given by the computation of \mathbf{z} (resp. \mathbf{z}_q), since the complexity of the fusion step ($\mathcal{O}(K)$) is negligible, thus leading to analogous expressions as in the case of classic MIMO detection [3]. On the other hand, we

¹³Here $\mathcal{O}(\cdot)$ indicates the Landau notation, i.e., the order of complexity.

remark that the complexity of $\Lambda_{\text{WL},i}$ has been evaluated by exploiting an alternative expression of (24), which uses the *Woodbury formula* [45] applied to $\Sigma^{-1} \underline{\mathbf{y}} \hat{\mathbf{G}}_{\mathcal{H}_i}$. In fact, $\Lambda_{\text{WL},i}$ can be recast equivalently as

$$\begin{aligned} \underline{\mathbf{a}}_{\text{WL},i} &= \frac{\frac{1}{\sigma_e^2} (\mathbf{I}_{2N} - \hat{\mathbf{G}} \mathbf{S}_i^{-1} \hat{\mathbf{G}}^\dagger) \hat{\mathbf{G}} \boldsymbol{\mu}_{1,0}}{\left\| \frac{1}{\sigma_e^2} (\mathbf{I}_{2N} - \hat{\mathbf{G}} \mathbf{S}_i^{-1} \hat{\mathbf{G}}^\dagger) \hat{\mathbf{G}} \boldsymbol{\mu}_{1,0} \right\|} \\ &= \frac{\frac{1}{\sigma_e^2} \hat{\mathbf{G}} \mathbf{S}_i^{-1} (\mathbf{S}_i - \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}}) \boldsymbol{\mu}_{1,0}}{\left\| \frac{1}{\sigma_e^2} \hat{\mathbf{G}} \mathbf{S}_i^{-1} (\mathbf{S}_i - \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}}) \boldsymbol{\mu}_{1,0} \right\|} \\ &= \frac{\hat{\mathbf{G}} \mathbf{S}_i^{-1} \Sigma_{\mathbf{x}|\mathcal{H}_i}^{-1} \boldsymbol{\mu}_{1,0}}{\left\| \hat{\mathbf{G}} \mathbf{S}_i^{-1} \Sigma_{\mathbf{x}|\mathcal{H}_i}^{-1} \boldsymbol{\mu}_{1,0} \right\|} \end{aligned} \quad (60)$$

where $\mathbf{S}_i \triangleq (\frac{\sigma_e^2}{\rho} \Sigma_{\mathbf{x}|\mathcal{H}_i}^{-1} + \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}})$ and we observe that $(\mathbf{S}_i - \hat{\mathbf{G}}^\dagger \hat{\mathbf{G}}) = \frac{\sigma_e^2}{\rho} \Sigma_{\mathbf{x}|\mathcal{H}_i}^{-1}$ holds. Unfortunately, although the complexity of ZF, MMSE and WL rules is substantially reduced w.r.t. the optimum rule, the term $\mathcal{O}(K^3)$ can become quite dominant in a WSN of large size. Such a term arises from the inversion of \mathbf{A} for ZF/MMSE (cf. (14)) and \mathbf{S}_i for WL rules. Aiming at mitigating the aforementioned issue, we adopt the matrix approximation proposed in [28], [29]. More specifically, we exploit the *Neumann series* [45] with the intent of expressing the inverse of the generic matrix $\mathbf{U} \in \mathbb{C}^{J \times J}$ as

$$\mathbf{U}^{-1} = \sum_{n=0}^{+\infty} (\mathbf{X}^{-1}(\mathbf{X} - \mathbf{U}))^n \mathbf{X}^{-1}, \quad (61)$$

where (61) holds only if \mathbf{U} satisfies $\lim_{n \rightarrow +\infty} (\mathbf{I}_J - \mathbf{X}^{-1} \mathbf{U})^n = \mathbf{O}_{J \times J}$ ($\mathbf{X} \in \mathbb{C}^{J \times J}$ is referred to as a “similar” matrix). Truncating the series up to the second term leads to:

$$\mathbf{U}^{-1} \approx \mathbf{X}^{-1} + \mathbf{X}^{-1}(\mathbf{X} - \mathbf{U})\mathbf{X}^{-1}. \quad (62)$$

Equation (62) requires only $\mathcal{O}(K^2)$ operations as opposed to $\mathcal{O}(K^3)$ for the exact inversion. Hence, we use (62) by replacing matrix \mathbf{X} with $\text{diag}(\mathbf{A})$ (resp. $\text{diag}(\mathbf{S})$) for computing \mathbf{A}^{-1} (resp. \mathbf{S}^{-1}) approximately. Such choices are justified since it can be easily verified that both \mathbf{A} and \mathbf{S} satisfy the required limit condition. In fact, as N grows, they both become *diagonally dominant*, i.e., $\frac{1}{N} \mathbf{A} \approx \frac{1}{N} \text{diag}(\mathbf{A}) \approx \mathbf{D}_{\hat{\mathbf{g}}}$ (resp. $\frac{1}{N} \mathbf{S} \approx \frac{1}{N} \text{diag}(\mathbf{S}) \approx \mathbf{D}_{\hat{\mathbf{g}}}$). Performance loss arising from approximate inversion in ZF/MMSE and WL rules is assessed in Section VI via simulations.

Finally, the computational complexity of MRC (resp. mMRC) is simply given by the computation of \mathbf{a}_{MRC} (resp. $\mathbf{a}_{\text{m-MRC}}$) whenever a new $\hat{\mathbf{G}}$ is acquired ($\mathcal{O}(KN)$) and a scalar product (cf. (27)) whenever each \mathbf{x} is observed.

VI. NUMERICAL RESULTS

Simulation Parameters: We assume \mathbf{D} is generated analogously as in [24]. More specifically, we consider sensors deployed in a circular area around the DFC with radius $r_{\text{max}} = 1000$ m. Sensors are located uniformly at random and we assume that no sensor is closer to the DFC than $r_{\text{min}} = 100$ m. The large-scale fading is modelled via $\beta_k = p_k (\frac{r_{\text{min}}}{r_k})^\nu$, where p_k is a log-normal random variable, i.e., $10 \log_{10}(p_k) \sim \mathcal{N}(\mu_P, \sigma_P^2)$

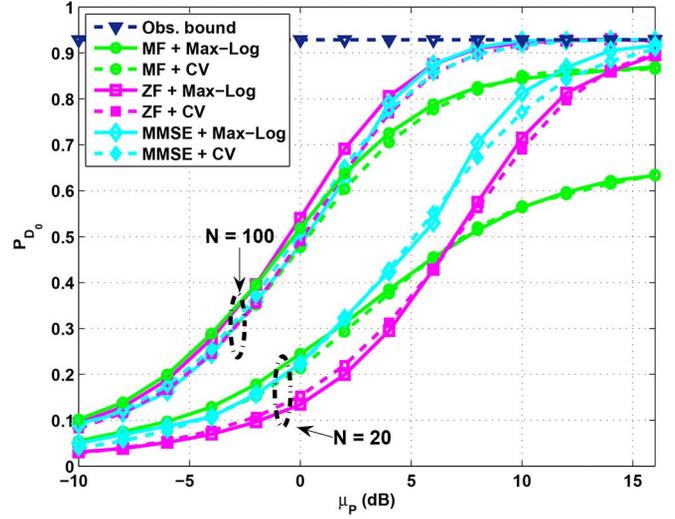


Fig. 3. P_{D_0} vs. μ_P (dB) for the massive MIMO linear rules, $P_{F_0} = 0.01$; comparison of CV and Max-Log based fusion; WSN with $K = 10$ sensors, $N \in \{20, 100\}$, $\sigma_P = 2$ (dB).

where μ_P and σ_P are the mean and standard deviation in dB, respectively. Also, r_k is the distance between the k th sensor and the DFC and ν denotes the path-loss exponent. For our simulations, we choose $\nu = 2$. Hereinafter we assume estimated CSI and we set $\tau = K$, corresponding to the minimum training-sequence length required for a meaningful MMSE estimation. This choice aims at minimizing the energy spent by each sensor for channel training. Furthermore, we assume conditionally i.i.d. decisions, that is $P(\mathbf{x} | \mathcal{H}_i) = \prod_{k=1}^K P(x_k | \mathcal{H}_i)$ with $(P_D, P_F) = (0.5, 0.05)$. Finally, we set for simplicity $\rho_P = \rho = 1$ and $\sigma_w^2 = 1$ unless differently specified.

In the following figures, for comparison purposes, we report the (upper) “observation bound” [9], i.e., the optimum performances over a noise-free channel, given by:

$$P_{D_0}^{\text{obs}}(\gamma_D) = \sum_{i=\gamma_D}^K \binom{K}{i} (P_D)^i (1 - P_D)^{K-i}, \quad (63)$$

$$P_{F_0}^{\text{obs}}(\gamma_D) = \sum_{i=\gamma_D}^K \binom{K}{i} (P_F)^i (1 - P_F)^{K-i}. \quad (64)$$

where $\gamma_D \in \{0, \dots, K\}$ denotes a discrete-valued threshold.

Max-Log vs CV Based Linear Filters: In Fig. 3 we report P_{D_0} as a function of μ_P (in dB) while fixing $\sigma_P = 2$ (dB) for MF, ZF and MMSE based either on Max-Log (solid) or CV (dashed) fusion rules (cf. (19) and (21), respectively), assuming¹⁴ $P_{F_0} = 0.01$. This is done in order to assess the performance gain given by the availability of the complex-valued output \mathbf{z} . We consider a WSN with $K = 10$ sensors and we show both scenarios with $N = 20$ (i.e., an array of moderate size) and $N = 100$ (a large array). First, it is apparent that when $N = 100$ Max-Log based rules achieve a slight performance improvement w.r.t. corresponding CV-based rules. The mentioned improvement is not always appreciated in the case $N = 20$, since at low SNR the quantization loss of \mathbf{z}_q for CV-based rules becomes less severe than imperfect match of Max-Log approximation. Therefore,

¹⁴Aiming at a fair comparison, we use *randomization* whenever the discrete nature of CV-based rules does not allow to meet the desired P_{F_0} exactly.

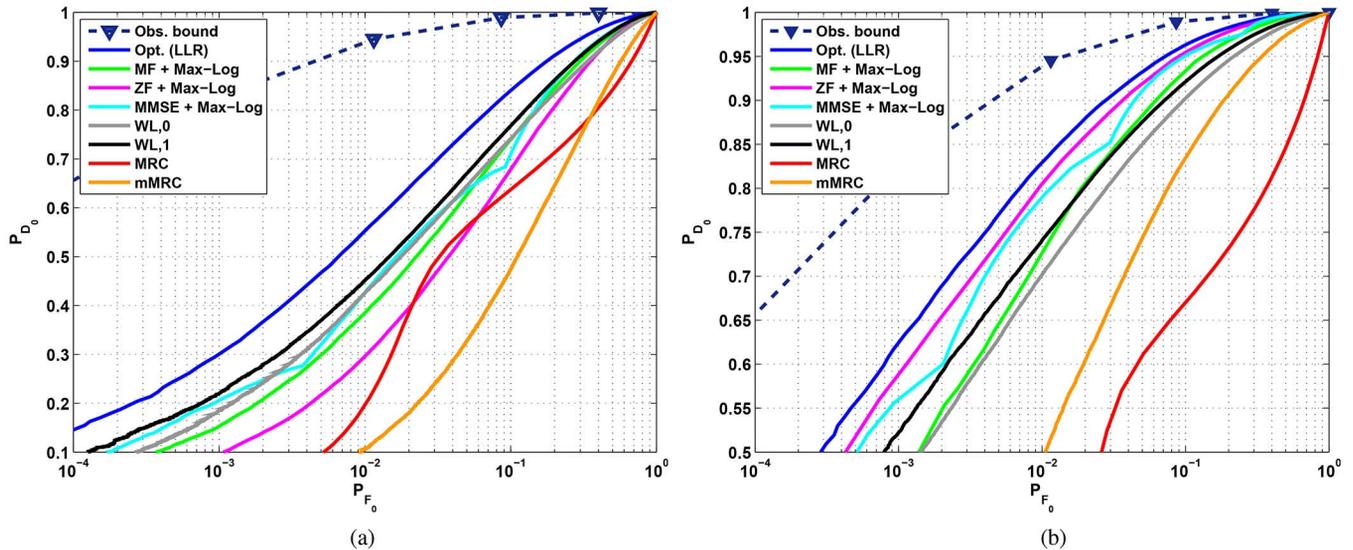


Fig. 4. P_{D_0} vs. P_{F_0} for all the presented rules; WSN with $K = 10$ sensors, $(\mu_P, \sigma_P) = (4, 2)$ (dB). (a) $N = 20$ (moderate size array). (b) $N = 100$ (large size array).

the availability of soft-output values of the linear filters is beneficial only in the large-array case. For the mentioned reasons we only keep rules based on Max-Log in what follows.

Receiver Operating Characteristic (ROC): In Figs. 4(a) and (b) we show P_{D_0} as a function of P_{F_0} for two different scenarios: (i) a scenario with a receive array of moderate size ($N = 20$, Fig. 4(a)) and (ii) a scenario with a large array ($N = 100$, Fig. 4(b)). It is apparent that when $N = 20$ the MRC remains an appealing solution, due to substantially low system knowledge requirement, while in the large array regime ($N = 100$) its performance suffers from a significant loss w.r.t. the LLR and it is outperformed by mMRC. Differently, all the fusion rules based on MIMO linear filters exploit effectively the dramatic increase of diversity and received SNR, and at $N = 100$ they approach the performance of optimum fusion rule. This is more pronounced in the case of linear-filters plus Max-Log, since they are *asymptotically optimal*¹⁵. The WL solutions perform quite well, but have a moderate loss in performance; this is due to the reduced required system knowledge (i.e., a mere second order characterization of $\mathbf{x} | \mathcal{H}_i$). Finally, it is apparent that $\Lambda_{WL,1}$ performs slightly better than $\Lambda_{WL,0}$. Unfortunately this is achieved at an increased difficulty in the acquisition of required parameters (indeed $\Sigma_{\mathbf{x} | \mathcal{H}_1}$ is usually harder to obtain, due to the less predictable statistical characterization of the unknown event observed).

P_{D_0} vs. N : Analogous considerations can be drawn from Fig. 5, where P_{D_0} as a function of N is shown, assuming $P_{F_0} = 0.01$. We plot the scenario for a WSN with $K = 10$ sensors and $\{\mu_P, \sigma_P\} = \{4, 2\}$ (dB). From inspection of the figure, we see that ZF+Max-Log and MMSE+Max-Log approach the LLR as N grows, thus confirming the theoretical findings. Differently, convergence of MF+Max-Log to the optimum performance is much slower and cannot be appreciated from the figure. Furthermore, MRC performs poorly when N is much higher than the number of sensors as opposed to mMRC, the latter being capable of effectively exploiting the linear SNR increase.

¹⁵Here the term “asymptotically optimal” refers to optimality as the number of antennas N grows large.

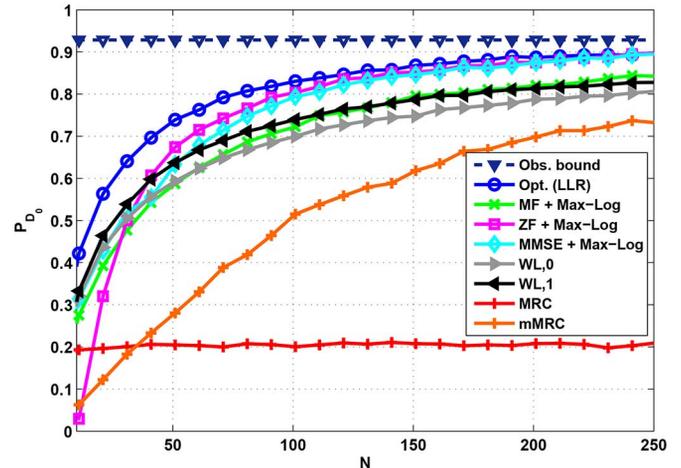


Fig. 5. P_{D_0} vs. N for all the considered rules, $P_{F_0} = 0.01$; WSN with $K = 10$ sensors, $(\mu_P, \sigma_P) = (4, 2)$ (dB).

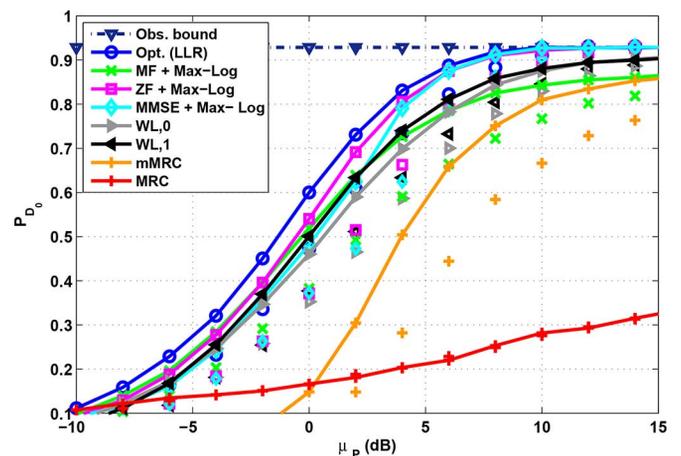


Fig. 6. P_{D_0} vs. μ_P (dB) for all the considered rules, $P_{F_0} = 0.01$; WSN with $K = 10$ sensors, $N = 50$ (markers only) vs. $N = 100$ (markers plus solid lines) antennas and $\sigma_P = 2$ (dB).

P_{D_0} vs. μ_P : In Fig. 6 we show P_{D_0} as a function of μ_P , assuming $P_{F_0} = 0.01$, in order to assess the performance improvement w.r.t. the average SNR. We consider a WSN with

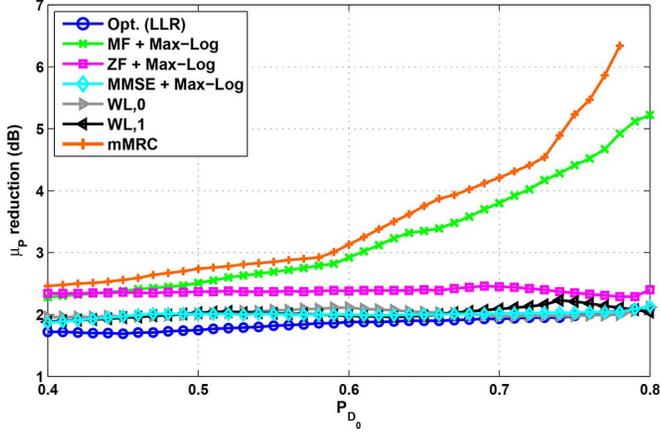


Fig. 7. μ_P reduction (dB) vs. P_{D_0} when moving from a DFC with $N = 50$ to $N = 100$ antennas; $K = 10$ sensors, $\sigma_P = 2$ (dB), $P_{F_0} = 0.01$. MRC is excluded since it does not meet prescribed target P_{D_0} .

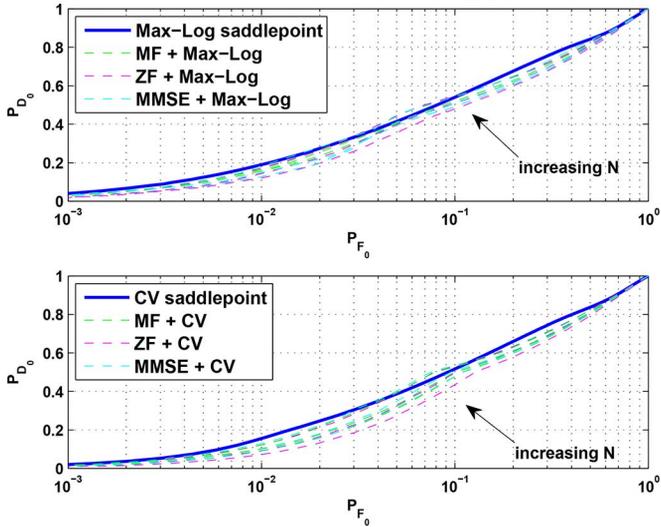


Fig. 8. P_{D_0} vs. P_{F_0} for Max-Log (top) and CV based (bottom) linear fusion rules; simulated (dashed) vs. theoretical (solid, saddlepoint approx.) ROC in scenario (a) ($\bar{\rho} = \bar{\rho}_P = 1$). WSN with $K = 10$ sensors, $N \in \{50, 100, 1000\}$ antennas at the DFC and $(\mu_P, \sigma_P) = (2, 4)$ (dB) (single instance of \mathbf{D} considered).

$K = 10$ sensors, $N \in \{50, 100\}$ antennas at the DFC and $\sigma_P = 2$ (dB). It is apparent that ZF + Max-Log and MMSE + Max-Log approach the optimum at $\mu_P \approx 4$ (dB) when $N = 100$ (resp. $\mu_P \approx 6$ (dB) when $N = 50$). MF + Max-Log has a moderate loss in performance w.r.t. the optimum, even at high SNR, due to the non-ideal separation of sensors contributions (this is a consequence of non-nulled residual interference, which causes performance loss when the so-called “open-eye” condition [46] is not verified) and it is outperformed by WL rules. As supported theoretically by [20], MRC performs poorly at high SNR and it is appealing only at very-low SNR, as opposed to mMRC. Furthermore, all the proposed rules benefit from an increase of the number of receive antennas at the DFC.

Also, in order to first investigate energy efficiency of the WSN for finite values of N , the corresponding sensor energy saving, when the array size is increased and the target performance (in terms of (P_{F_0}, P_{D_0})) is kept fixed, is analyzed. To this end, Fig. 7 shows the reduction of μ_P (quantified as the difference in dB) achieved when moving from $N = 50$ to

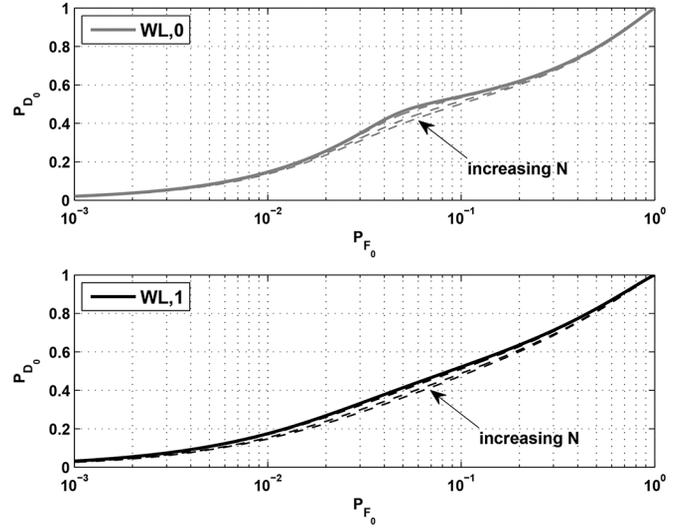


Fig. 9. P_{D_0} vs. P_{F_0} for WL, 1 (top) and WL, 0 (bottom) fusion rules; simulated (dashed) vs. theoretical (solid) ROC in scenario (a) ($\bar{\rho} = \bar{\rho}_P = 1$). WSN with $K = 10$ sensors, $N \in \{50, 100, 1000\}$ antennas at the DFC and $(\mu_P, \sigma_P) = (2, 4)$ (dB) (single instance of \mathbf{D} considered).

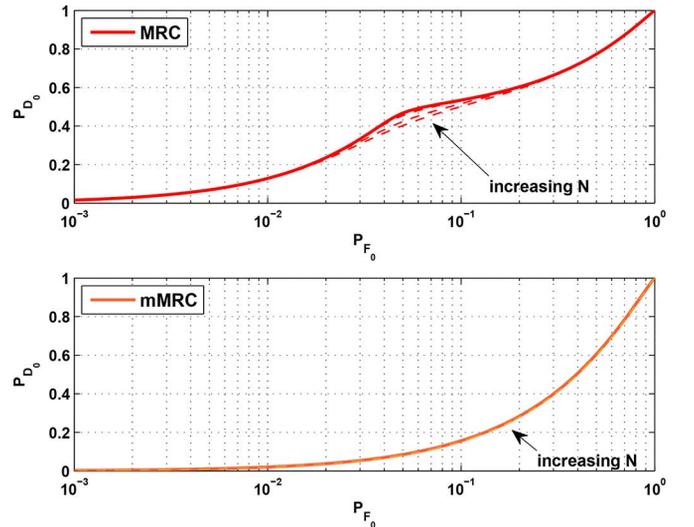


Fig. 10. P_{D_0} vs. P_{F_0} for MRC (top) and mMRC (bottom) fusion rules; simulated (dashed) vs. theoretical (solid) ROC in scenario (a) ($\bar{\rho} = \bar{\rho}_P = 1$). WSN with $K = 10$ sensors, $N \in \{50, 100, 1000\}$ antennas at the DFC and $(\mu_P, \sigma_P) = (2, 4)$ (dB) (single instance of \mathbf{D} considered).

$N = 100$ antennas, by ensuring the same prescribed system false-alarm rate (set to $P_{F_0} = 0.01$) and system detection probability ($P_{D_0} \in [0.4, 0.8]$ in the plot). It is apparent that for all the sub-optimal rules the energy reduction is remarkable. As an example, transmitted SNR of MF+Max-Log can be reduced by ≈ 3 dB when the target system performance is $(P_{F_0}, P_{D_0}) = (0.01, 0.6)$. Also, mMRC and MF+Max-Log present an increasing μ_P reduction with P_{D_0} since their performance at $N = 50$ is much worse than the case $N = 100$ in comparison with other rules, as also apparent in Fig. 6.

Large N Theoretical Performance for Max-Log/CV Linear, WL and (m)MRC Rules: In Figs. 8, 9 and 10 we show P_{D_0} vs. of P_{F_0} in order to assess the theoretical convergence of Max-Log/CV based MF/ZF/MMSE, WL and (m)MRC rules, respectively, obtained in Section IV. Here we consider performance convergence for the asymptotic scenario (a) (i.e., $N \rightarrow +\infty$,

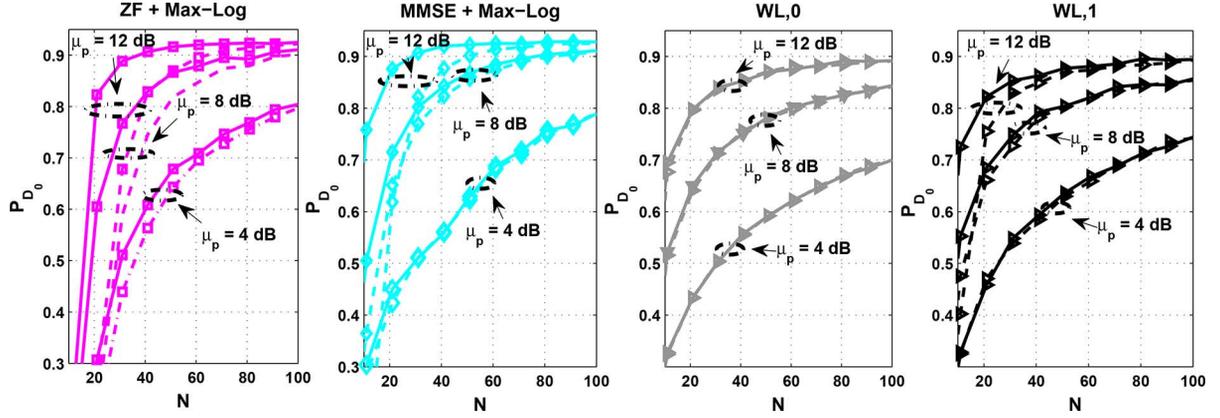


Fig. 11. P_{D_0} vs. N for ZF + Max-Log, MMSE + Max-Log and WL rules, $P_{F_0} = 0.01$; exact (solid) vs approximate (dashed) matrix inversion; WSN with $K = 10$ sensors, $\mu_p \in \{4, 8, 12\}$ (dB), $\sigma_p = 2$ (dB).

where $\rho = \frac{\bar{p}}{\sqrt{N}}$ and $\rho_p = \frac{\bar{p}_p}{\sqrt{N}}$, namely we assume an energy cut in both training and reporting phases. Both scenarios are clearly considered aiming at establishing energy efficiency of the WSN when N grows unbounded [24]. We adopt a single realization of \mathbf{D} (drawn from a log-normal with $(\mu_p, \sigma_p) = (4, 2)$ (dB)), a WSN with $K = 10$ sensors and $N \in \{50, 100, 1000\}$; finally we set $\bar{\rho} = \bar{\rho}_p = 1$. It is apparent that MC-based ROCs approach the corresponding theoretical ones as N grows, thus confirming the theoretical findings. Therefore, the mentioned theoretical performance are achieved even though the sensors are cutting their transmitted energy in a way proportional to \sqrt{N} . This underlines the energy-efficiency of the considered WSN with respect to all the fusion rules being considered. Indeed, even though they are sub-optimal (and have polynomial complexity, see Table I) they all enjoy this desirable property. It is worth remarking that the theoretical results hold exactly as N goes to infinity, except for the Max-Log and CV based linear rules, where the saddlepoint approximation is used to draw the theoretical ROCs. Nonetheless there is a satisfactory match. Finally, we notice that the absolute large-array performance of mMRC is worse than MRC, since we are actually considering a very-low SNR scenario (i.e., the transmit energy for both phases is being cut proportionally to N).

Efficient Approximate Inversion: In Fig. 11 we assess the performance loss given by the approximate inversion described in Section V for ZF+Max-Log, MMSE+Max-Log and WL rules. We show P_{D_0} vs. N , assuming $P_{F_0} = 0.01$. We consider a WSN with $K = 10$ sensors, with $\mu_p \in \{4, 8, 12\}$ (dB) and $\sigma_p = 2$ (dB). It is apparent that each approximated rule approach the corresponding exact implementation as N grows large. However, we observe that in general the lower μ_p is (which is the expected operating scenario in a WSN), the less is the performance degradation with a moderate-sized array. Furthermore, in the considered scenario (i.e., with conditionally independent decisions) the performance loss is less pronounced for MMSE+Max-Log and WL rules w.r.t. ZF+Max-Log. This is expected since in the former case the matrices $\mathbf{A}_{\text{MMSE}} = (\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \mathbf{I}_K)$ and $\mathbf{S}_i = (\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \Sigma_{\mathbf{x}}^{-1} |_{\mathcal{H}_i})$ need to be inverted, respectively, which are diagonally dominant (even for moderate N) when μ_p is low (since both the right terms are diagonal and are dominating the sum at low SNR).

VII. CONCLUSION

In this paper we studied channel-aware decision fusion over MIMO channels, in the presence of a large antenna-array at the DFC. We presented a wide choice of low-complexity sub-optimal rules which efficiently exploit large-array benefits and are able to achieve near-optimal performance. The proposed rules were numerically compared with existing alternatives, namely MRC, which was shown not to significantly exploit the increase of the array size. Additionally, a theoretical analysis of the proposed rules was provided under energy-reduction laws commonly considered in massive MIMO literature. Furthermore, a detailed discussion on computational complexity of the mentioned rules was given. Accordingly, (further) computationally-efficient versions of some of the proposed rules, via the approximation of Neumann series of a matrix inverse, were derived. Finally, the corresponding performance loss was assessed via simulations and shown to be negligible when a large-array is employed at DFC.

APPENDIX PROOF OF LEMMA 1

We start by expanding $p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i)$ as follows (cf. (13)):

$$\begin{aligned}
 p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i) &= (\pi \sigma_e^2)^{-N} \exp[-\sigma_e^{-2} \|\mathbf{y}\|^2] \\
 &\times \sum_{\mathbf{x} \in \mathcal{X}^K} \exp \left[\sigma_e^{-2} \left(-\rho \|\hat{\mathbf{G}} \mathbf{x}\|^2 + 2\sqrt{\rho} \Re\{\mathbf{x}^\dagger \hat{\mathbf{G}}^\dagger \mathbf{y}\} \right) \right] \\
 &\times P(\mathbf{x} | \mathcal{H}_i).
 \end{aligned} \tag{65}$$

Equation (65) is in the form $p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_i) = a(\mathbf{y})b(\mathbf{t}(\mathbf{y}) | \mathcal{H}_i)$, where $\mathbf{t}(\mathbf{y}) \triangleq (\hat{\mathbf{G}}^\dagger \mathbf{y})$. Thus $\mathbf{t} = \mathbf{z}_{\text{MF}}$ represents a *sufficient statistic* (independently on the relationship between K and N) [47]. Furthermore, we notice that \mathbf{z}_{ZF} (resp. \mathbf{z}_{MMSE}) is in one-to-one correspondence with \mathbf{z}_{MF} , since $(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}})$ (resp. $(\hat{\mathbf{G}}^\dagger \hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho} \mathbf{I}_K)$) is invertible (this assumption is valid iff $K \leq N$ for ZF). Finally, exploiting the property stating that *every one-to-one transformation of a sufficient statistic is itself sufficient* [47], we prove the lemma for ZF and MMSE.

APPENDIX

FUSION ARCHITECTURES WITH MIMO LINEAR FILTERS

(*z* available, conditionally independent decisions)—Here we derive the fusion rules based on MIMO linear filters. We first recall that, under a large array regime ($N \gg K$), we have $p(\mathbf{z} | \hat{\mathbf{G}}, \mathbf{x}) \approx \prod_{k=1}^K \varphi(z_k | x_k)$ (cf. (16)). Also, by exploiting $P(\mathbf{x} | \mathcal{H}_i) = \prod_{k=1}^K P(x_k | \mathcal{H}_i)$, $\ln[p(\mathbf{z} | \hat{\mathbf{G}}, \mathcal{H}_i)]$ is approximated as

$$\ln[p(\mathbf{z} | \hat{\mathbf{G}}, \mathcal{H}_i)] \approx \sum_{k=1}^K \ln \left[\sum_{x_k \in \mathcal{X}} \varphi(z_k | x_k) P(x_k | \mathcal{H}_i) \right] \quad (66)$$

The right hand side of (66) corresponds to a DF model operating over PAC, whose properties have been studied in detail in [6], [7]. Such a model has the advantage of being of feasible complexity w.r.t. the model under investigation (i.e., a DF model over MIMO), since its complexity is simply $\mathcal{O}(K)$. Based on (66), the LLR of \mathbf{z} is approximated as follows:

$$\begin{aligned} \Lambda_{\mathbf{z}} &\triangleq \ln \left[\frac{p(\mathbf{z} | \hat{\mathbf{G}}, \mathcal{H}_1)}{p(\mathbf{z} | \hat{\mathbf{G}}, \mathcal{H}_0)} \right] \\ &\approx \sum_{k=1}^K \ln \left[\frac{\varphi(z_k | x_k = 1) P_{D,k} + \varphi(z_k | x_k = -1) (1 - P_{D,k})}{\varphi(z_k | x_k = 1) P_{F,k} + \varphi(z_k | x_k = -1) (1 - P_{F,k})} \right] \end{aligned} \quad (67)$$

Although (67) has reduced computational complexity, it is well known that is numerically unstable, due to the presence of exponentials with large dynamics. For this reason, we adopt the near-optimum rule proposed in [7], based on *Max-Log* approximation. The latter originates from turbo-codes literature [34] and states that

$$\ln \left[\sum_{\ell=1}^L B_{\ell} \exp(A_{\ell}) \right] \approx \max_{\ell \in \{1, 2, \dots, L\}} [A_{\ell} + \ln(B_{\ell})] \quad (68)$$

where $A_{\ell} \in \mathbb{R}$ and $B_{\ell} \in \mathbb{R}^+$. The approximation in (68) is accurate when one of the terms in the sum $\sum_{\ell=1}^L B_{\ell} \exp(A_{\ell})$ dominates over the remaining terms. Second line of (67) can be put in the same form of (68) as

$$\begin{aligned} \Lambda_{\mathbf{z}} &\approx \sum_{k=1}^K \left\{ \ln [\exp(f_k) \cdot P_{D,k} + \exp(-f_k) \cdot (1 - P_{D,k})] \right. \\ &\quad \left. - \ln [\exp(f_k) \cdot P_{F,k} + \exp(-f_k) \cdot (1 - P_{F,k})] \right\} \end{aligned} \quad (69)$$

where we have defined:

$$f_k \triangleq \begin{cases} \frac{2\sqrt{\rho}\Re\{z_k\}}{\sigma_e^2} & \text{MF} \\ \frac{2\sqrt{\rho}d_{\hat{g},k}N\Re\{z_k\}}{\sigma_e^2} & \text{ZF/MMSE} \end{cases}. \quad (70)$$

Using (68) leads to the simplified expression for $\Lambda_{\mathbf{z}}$:

$$\begin{aligned} \Lambda_{\mathbf{z}} &\approx \sum_{k=1}^K \left(\max \{f_k + \ln [P_{D,k}], -f_k + \ln [1 - P_{D,k}]\} \right. \\ &\quad \left. - \max \{f_k + \ln [P_{F,k}], -f_k + \ln [1 - P_{F,k}]\} \right) \end{aligned} \quad (71)$$

Similarly to [7], (71) can be put in compact form as $\Lambda_{\mathbf{z}} \approx \sum_{k=1}^K \lambda_{u,k}$, with

$$\lambda_{u,k} \triangleq \begin{cases} c_{1,k}, & f_k < a_{1,k} \\ 2f_k + c_{2,k}, & a_{1,k} \leq f_k < a_{2,k} \\ c_{3,k}, & f_k \geq a_{2,k} \end{cases}, \quad (72)$$

where $c_{1,k} \triangleq \ln \left[\frac{1 - P_{D,k}}{1 - P_{F,k}} \right]$, $c_{2,k} \triangleq \ln \left[\frac{P_{D,k}}{1 - P_{F,k}} \right]$, $c_{3,k} \triangleq \ln \left[\frac{P_{D,k}}{P_{F,k}} \right]$, $a_{1,k} \triangleq \frac{1}{2} \ln \left[\frac{1 - P_{D,k}}{P_{D,k}} \right]$ and $a_{2,k} \triangleq \frac{1}{2} \ln \left[\frac{1 - P_{F,k}}{P_{F,k}} \right]$.

(*z_q* available, conditionally independent decisions)—Here we assume that we have available only the quantized output \mathbf{z}_q . Similarly to the case of \mathbf{z} , as N grows, we have $P(\mathbf{z}_q | \hat{\mathbf{G}}, \mathbf{x}) \approx \prod_{k=1}^K \bar{\varphi}(z_{q,k} | x_k)$ (cf. (18)). Therefore, we can use the optimum rule over PAC in the case of binary symmetric channels, namely the Chair-Varshney rule [6], whose statistic is obtained as follows:

$$\begin{aligned} &\ln \left[\frac{P(\mathbf{z}_q | \hat{\mathbf{G}}, \mathcal{H}_1)}{P(\mathbf{z}_q | \hat{\mathbf{G}}, \mathcal{H}_0)} \right] \\ &\approx \sum_{k=1}^K \ln \left[\frac{\sum_{x_k \in \mathcal{X}} \bar{\varphi}(z_{q,k} | x_k) P(x_k | \mathcal{H}_1)}{\sum_{x_k \in \mathcal{X}} \bar{\varphi}(z_{q,k} | x_k) P(x_k | \mathcal{H}_0)} \right] \\ &= \sum_{k=1}^K \left\{ \lambda_{q,k} \triangleq \ln \left[\frac{\alpha_k (P_{D,k})^{\hat{b}_k} \eta_k (P_{D,k})^{(1-\hat{b}_k)}}{\alpha_k (P_{F,k})^{\hat{b}_k} \eta_k (P_{F,k})^{(1-\hat{b}_k)}} \right] \right\} \end{aligned} \quad (73)$$

where we have defined $\alpha_k(P) \triangleq P \cdot (1 - P_{e,k}) + (1 - P) \cdot (P_{e,k})$, $\eta_k(P) \triangleq (1 - \alpha_k(P))$ and $\hat{b}_k \triangleq \left(\frac{1 + z_{q,k}}{2} \right)$. Finally, exploiting the standard properties of logarithms provides (21).

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Associate Editor and the anonymous reviewers for taking their time into reviewing this manuscript and providing comments that contributed to improve the quality and the readability of the manuscript.

REFERENCES

- [1] D. Ciunzo, P. Salvo Rossi, and S. Dey, "Massive MIMO meets decision fusion: Decode-and-fuse vs. decode-then-fuse," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, 2014, pp. 265–268.
- [2] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.
- [3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.
- [4] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [5] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
- [6] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3454–3458, Dec. 2004.
- [7] A. Lei and R. Schober, "Coherent Max-Log decision fusion in wireless sensor networks," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1327–1332, May 2010.
- [8] W. Li and H. Dai, "Distributed detection in wireless sensor networks using a multiple access channel," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 822–833, Mar. 2007.

- [9] C. R. Berger, M. Guerriero, S. Zhou, and P. K. Willett, "PAC vs. MAC for decentralized detection using noncoherent modulation," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3562–3575, Sep. 2009.
- [10] F. Li, J. S. Evans, and S. Dey, "Decision fusion over noncoherent fading multiaccess channels," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4367–4380, Sep. 2011.
- [11] D. Ciuonzo, G. Romano, and P. Salvo Rossi, "Optimality of received energy in decision fusion over Rayleigh fading diversity MAC with non-identical sensors," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 22–27, Jan. 2013.
- [12] K. Umehayashi, J. J. Lehtomaki, T. Yazawa, and Y. Suzuki, "Efficient decision fusion for cooperative spectrum sensing based on OR-rule," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2585–2595, Jul. 2012.
- [13] S. Yiu and R. Schober, "Nonorthogonal transmission and noncoherent fusion of censored decisions," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 263–273, Jan. 2009.
- [14] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4124–4140, Sep. 2008.
- [15] K. Bai and C. Tepedelenlioglu, "Distributed detection in UWB wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 804–813, Feb. 2010.
- [16] M. K. Banavar, A. D. Smith, C. Tepedelenlioglu, and A. Spanias, "On the effectiveness of multiple antennas in distributed detection over fading MACs," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1744–1752, May 2012.
- [17] I. Nevat, G. W. Peters, and I. B. Collings, "Distributed detection in sensor networks over fading channels with multiple antennas at the fusion centre," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 671–683, Feb. 2014.
- [18] F. Jiang, J. Chen, A. L. Swindlehurst, and J. A. Lopez-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," 2014, arXiv preprint arXiv:1408.2192 [Online]. Available: <http://arxiv.org/abs/1408.2192>
- [19] D. Ciuonzo, G. Romano, and P. Salvo Rossi, "Channel-aware decision fusion in distributed MIMO wireless sensor networks: Decode-and-fuse vs. decode-then-fuse," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2976–2985, Aug. 2012.
- [20] D. Ciuonzo, G. Romano, and P. Salvo Rossi, "Performance analysis of maximum ratio combining in channel-aware MIMO decision fusion," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4716–4728, Sep. 2013.
- [21] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [22] J. Jose, A. Ashikhmin, T. L. Marzetta, L. Thomas, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [23] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [24] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [25] H. Q. Ngo, M. Matthaiou, T. Q. Duong, and E. G. Larsson, "Uplink performance analysis of multiuser MU-SIMO systems with ZF receivers," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4471–4483, 2013.
- [26] M. Matthaiou, C. Zhong, M. R. McKay, and T. Ratnarajah, "Sum rate analysis of ZF receivers in distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 180–191, Feb. 2013.
- [27] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
- [28] M. Wu, B. Yin, A. Vosoughi, C. Studer, J. R. Cavallaro, and C. Dick, "Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 2155–2158.
- [29] B. Yin, M. Wu, C. Studer, J. R. Cavallaro, and C. Dick, "Implementation trade-offs for linear detection in large-scale MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2013, pp. 2679–2683.
- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall PTR, 1993, vol. 1, Estimation Theory.
- [31] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall PTR, Jan. 1998, vol. 2, Detection Theory.
- [32] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul. 2005.
- [33] J. H. Kotecha, V. Ramachandran, and A. M. Sayeed, "Distributed multi-target classification in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 703–713, Apr. 2005.
- [34] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proc. IEEE Int. Conf. Communications (ICC)*, Seattle, WA, USA, Jun. 1995, vol. 2, pp. 1009–1013.
- [35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 2006.
- [36] S. G. Iyengar, R. Niu, and P. K. Varshney, "Fusing dependent decisions for hypothesis testing with heterogeneous sensors," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4888–4897, Sep. 2012.
- [37] V. V. Veeravalli and P. K. Varshney, "Distributed inference in wireless sensor networks," *Philos. Trans. Roy. Soc. London A, Math. Phys. Eng. Sci.*, vol. 370, no. 1958, pp. 100–117, 2012.
- [38] Z. Quan, W.-K. Ma, S. Cui, and A. H. Sayed, "Optimal linear fusion for distributed detection via semidefinite programming," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2431–2436, Apr. 2010.
- [39] B. Picinbono, "On deflection as a performance criterion in detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 3, pp. 1072–1081, Jul. 1995.
- [40] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008.
- [41] K.-C. Lai, Y.-L. Yang, and J.-J. Jia, "Fusion of decisions transmitted over flat fading channels via maximizing the deflection coefficient," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 3634–3640, July 2010.
- [42] D. Ciuonzo and P. Salvo Rossi, "Decision fusion with unknown sensor detection probability," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 208–212, Feb. 2014.
- [43] D. J. Love and R. W. Heath, Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.
- [44] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: The saddlepoint approximation," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 327–340, Jan. 2007.
- [45] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [46] S. Verdú, *Multuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [47] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signal*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

Domenico Ciuonzo (S'11–M'14) was born in Aversa, Italy, on June 29th, 1985. He received the B.Sc. (*summa cum laude*), the M.Sc. (*summa cum laude*) degrees in computer engineering and the Ph.D. in electronic engineering, respectively in 2007, 2009 and 2013, from the Second University of Naples (SUN), Aversa, Italy. In 2011 he was involved in the Visiting Researcher Programme of Centre for Maritime Research and Experimentation, La Spezia, Italy, within the "Maritime Situation Awareness" project. In 2012 he was a visiting scholar at the Electrical and Computer Engineering Department of University of Connecticut (UConn), Storrs, CT, USA. In 2013–2014 he was a Postdoctoral Researcher at the Department of Industrial and Information Engineering of SUN, under the project "MICENEA", funded by MIUR within the program Futuro in Ricerca (FIRB) 2013. He is currently a Postdoctoral Researcher at DIETI, University of Naples, "Federico II", Italy. His research interests are mainly in the areas of Data Fusion, Statistical Signal Processing, Target Tracking and Wireless Sensor Networks. The quality of his reviewing activity was recognized by the Editorial Board of IEEE COMMUNICATIONS LETTERS, which nominated him "Exemplary Reviewer" for the year 2013. Dr. Ciuonzo currently serves as Associate Editor for the IEEE AEROSPACE AND ELECTRONIC SYSTEMS MAGAZINE.



Pierluigi Salvo Rossi (SM'11) was born in Naples, Italy, on April 26, 1977. He received the “Laurea” degree in telecommunications engineering (*summa cum laude*) and the Ph.D. degree in computer engineering in 2002 and 2005, respectively, both from the University of Naples “Federico II”, Naples, Italy. During his Ph.D. studies, he has been a researcher at the Inter-departmental Research Center for Signals Analysis and Synthesis (CIRASS), University of Naples “Federico II”, Naples, Italy; at the Department of Information Engineering, Second University of Naples, Aversa (CE), Italy; at the Communications and Signal Processing Laboratory (CSPL), Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, USA; and an adjunct professor at the Faculty of Engineering, Second University of Naples, Aversa (CE), Italy. From 2005 to 2008 he worked as a postdoctoral researcher at the Department of Computer Science and Systems, University of Naples “Federico II”, Naples, Italy; at the Department of Information Engineering, Second University of Naples, Aversa (CE), Italy; at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway; and visited the Department of Electrical and Information Technology, Lund University, Lund, Sweden. In 2008, he was appointed assistant professor (tenured in 2011) in telecommunications at the Department of Industrial and Information Engineering, Second University of Naples, Aversa (CE), Italy; and visited the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway and the Excellence Center for Wireless Sensor Networks (WISNET), Uppsala University, Uppsala, Sweden. Currently, he is an associate professor in signal processing at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway. He is an Editor for the IEEE COMMUNICATIONS LETTERS. His research interests fall within the areas of communications and signal processing.



Subhrakanti Dey (M'96–SM'06) was born in India in 1968. He received the B.Tech. and M.Tech. degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, in 1991 and 1993, respectively, and the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, in 1996.

He is currently a Professor with the Department of Engineering Sciences in Uppsala University, Sweden. Prior to this, he was a Professor with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, from 2000 until early 2013. From September 1995 to September 1997, and September 1998 to February 2000, he was a Postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park.

His current research interests include networked control systems, wireless communications and networks, signal processing for sensor networks, and stochastic and adaptive estimation and control.

Professor Dey currently serves on the Editorial Board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and Elsevier SYSTEMS AND CONTROL LETTERS. He was also an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING during 2007–2010 and the IEEE TRANSACTIONS ON AUTOMATIC CONTROL during 2004–2007.