

Massive MIMO meets Decision Fusion: Decode-and-Fuse vs. Decode-then-Fuse

Domenico CIUONZO[†], Pierluigi SALVO ROSSI^{†‡}, Subhrakanti DEY[§]

[†]Department of Industrial and Information Engineering, Second University of Naples, Italy.

[‡]Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Norway.

[§]Department of Engineering Sciences, Uppsala University, Sweden.

email: {salvorossi, domenico.ciuonzo}@ieee.org; Subhra.Dey@signal.uu.se

Abstract—We study channel-aware decision fusion over a multiple-input multiple-output (MIMO) channel in the large-array regime at the decision-fusion center (DFC). Inhomogeneous large-scale fading between the sensors and the DFC is considered in addition to the small-scale fading, and pilot-based channel estimation is performed at the DFC. Linear processing techniques are analyzed in order to design low-complexity alternatives to the optimum log-likelihood ratio test (LLRT). Performance evaluation based on Monte Carlo simulations are presented.

I. INTRODUCTION

Distributed detection through wireless sensor networks (WSNs) has been largely studied with reference to the parallel access channel (PAC) [1], [2], i.e. sensors exploit (dedicated) orthogonal channels to inform the decision-fusion center (DFC). Near-optimal fusion rules with full channel state information (CSI) at the receiver have been proposed in [3].

Recent works [4]–[7] extended the analysis to the multiple access channel (MAC) by exploiting appropriately the interfering nature of the wireless medium. Looking at the system as a “virtual” multi-input multi-output (MIMO), array processing techniques at the DFC have been investigated and compared in terms of performance, complexity, and knowledge requirements [8]–[10].

Massive MIMO is an emerging technology for communications where receive arrays with a few hundred elements serve many tens of terminals simultaneously [11]. The potential benefits are capacity increase, deep-fading mitigation, MAC-layer simplification, increased interference rejection. From a mathematical point of view, large-size arrays lead to averaging small-scale fading out, orthogonality of random channel vectors between the users and the base station, reduction of the transmitted energy. Focusing on the tradeoff between energy and spectral efficiency, performance of linear detectors for uplink communications in presence of imperfect CSI have been analyzed in [12].

The advantages offered by massive MIMO may be beneficial for DF in WSNs over MAC. In this paper we propose an architecture which, in addition to be spectral efficient, will mitigate severe energy constraints given by inexpensive sensor nodes (i.e. extended battery life) and allow low-complexity (but near-optimal) fusion rules at the DFC. Pilot-based channel estimation is performed in order to employ channel-aware decision fusion at the DFC.

The outline of the paper is the following: Sec. II presents the system model; Secs. III and IV describe the channel estimation and the decision fusion steps, respectively; Sec. V compares the performance of simulated systems; some final remarks are given in Sec. VI.

II. SYSTEM MODEL

We consider K sensors sensing a binary source, each taking autonomously a local decision. The two hypotheses are denoted \mathcal{H}_0 and \mathcal{H}_1 and may represent the absence and the presence of a specific target of interest, respectively.

We assume that the local sensing and decision process is fully described by the local probability of false alarm (p_f) and the local probability of detection (p_d), both assumed to be stationary, identical, and conditionally independent given the specific hypothesis. The K sensors, each with one single transmit antenna, communicate simultaneously their decision to a DFC equipped with N receive antennas ($N \gg K$, i.e. large-size array configuration is assumed) whose aim is to provide a robust decision on the basis of the multiple received information. All the sensors employ binary phase shift keying (BPSK) modulation with identical parameters (transmission pulse, carrier frequency, etc.). Also, we assume that the system is fully synchronized.

We denote $x_k \in \mathcal{X} = \{-1, +1\}$ the symbol transmitted by the k th sensor encoding its local decision (we assume -1 for \mathcal{H}_0 and $+1$ for \mathcal{H}_1); $G_{n,k}$ the fading channel coefficient on the link between the k th sensor and n th receive antenna of the DFC; y_n the signal received by the n th antenna of the DFC; and $w_n \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_w^2)$ the additive white Gaussian noise at the n th receive antenna of the DFC. Channel coefficients are expressed as

$$G_{n,k} = \sqrt{d_k} H_{n,k}, \quad (1)$$

⁰Notation – Lower-case bold letters denote vectors, with a_n being the n th element of \mathbf{a} ; upper-case bold letters denote matrices, with $A_{n,m}$ denoting the (n, m) th element of \mathbf{A} ; $\Re(a)$ denotes the real part of a ; \mathbf{I}_N denotes the $N \times N$ identity matrix; $\mathbf{0}_N$ denotes the N -length vector whose elements are 0; $\text{diag}(\mathbf{a})$ denotes a diagonal matrix with \mathbf{a} on the main diagonal; $(\cdot)^*$, $(\cdot)^t$, $(\cdot)^\dagger$, and $\|\cdot\|$ denote conjugate, transpose, conjugate transpose, and Frobenius norm operators; $\Pr(\mathcal{A})$ denotes the probability of the event \mathcal{A} ; $p(a)$ denotes the probability density function (pdf) of the random variable a ; $\Re(a)$, $\Im(a)$, and $|a|$ denote the real part, the imaginary part, and the modulus of a , respectively; A^n denotes the n th Cartesian power of the set \mathcal{A} ; $\mathcal{N}(\mu, \sigma^2)$ denotes a (real scalar) normal distribution with mean μ and variance σ^2 ; $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a circular proper normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathcal{Q}(\cdot)$ denotes the tail distribution of a standard normal random variable; \sim means “distributed as”.

⁰This work has been partially funded by ERCIM and by CAMOS.

where $\sqrt{d_k}$ models known geometric attenuation and shadowing, while $H_{n,k} \sim \mathcal{N}_C(0,1)$ is the unknown independent fast fading. Denoting $\mathbf{D} = \text{diag}(d_1, \dots, d_K)$, then $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$ and the discrete-time model for the received signal $\mathbf{y} = (y_1, \dots, y_N)^t$ is

$$\mathbf{y} = \sqrt{\rho}\mathbf{G}\mathbf{x} + \mathbf{w}. \quad (2)$$

where $\mathbf{x} = (x_1, \dots, x_K)^t$ denotes the transmitted vector (of local decisions), $\mathbf{w} = (w_1, \dots, w_N)^t$ is the noise vector, and ρ controls the energy spent during the detection phase.

III. CHANNEL ESTIMATION

We assume that $\tilde{\tau} \geq K$ symbols are used as pilots for channel estimation at the DFC, and $\tau_c - \tilde{\tau}$ symbols are used for target detection, 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 $\tilde{\tau}$. Pilot sequences are collected in a matrix $\sqrt{\tilde{\tau}\tilde{\rho}}\tilde{\Phi}$, where $\tilde{\rho}$ controls pilot energy, such that $\tilde{\Phi}^\dagger\tilde{\Phi} = \mathbf{I}_K$.

The received signal model for the training phase is:

$$\tilde{\mathbf{Y}} = \sqrt{\tilde{\tau}\tilde{\rho}}\mathbf{G}\tilde{\Phi}^t + \tilde{\mathbf{W}}, \quad (3)$$

where $\tilde{W}_{n,k} \sim \mathcal{N}_C(0, \sigma_w^2)$. The corresponding minimum mean-square error (MMSE) estimator of \mathbf{G} given $\tilde{\mathbf{Y}}$ is [13]

$$\hat{\mathbf{G}} = \frac{1}{\sqrt{\tilde{\tau}\tilde{\rho}}}\tilde{\mathbf{Y}}\tilde{\Phi}^* \underbrace{\left(\frac{\sigma_w^2}{\sqrt{\tilde{\tau}\tilde{\rho}}}\mathbf{D}^{-1} + \mathbf{I}_K\right)^{-1}}_{\tilde{\mathbf{D}}} \quad (4)$$

$$= \left(\mathbf{G} + \frac{1}{\sqrt{\tilde{\tau}\tilde{\rho}}}\tilde{\mathbf{W}}\tilde{\Phi}^*\right)\tilde{\mathbf{D}}. \quad (5)$$

Defining the error matrix $\tilde{\mathbf{\Xi}} = \hat{\mathbf{G}} - \mathbf{G}$, it can be shown that $\tilde{\mathbf{\Xi}}$ and $\hat{\mathbf{G}}$ are statistically independent and have mutually statistically independent columns, denote $\tilde{\boldsymbol{\xi}}_k$ and $\hat{\mathbf{g}}_k$, respectively. More specifically, $\tilde{\boldsymbol{\xi}}_k \sim \mathcal{N}_C(\mathbf{0}_N, \epsilon_k^2\mathbf{I}_N)$ and $\hat{\mathbf{g}}_k \sim \mathcal{N}_C(\mathbf{0}_N, \delta_k^2\mathbf{I}_N)$, where

$$\epsilon_k^2 = \frac{\sigma_w^2 d_k}{\tilde{\tau}\tilde{\rho}d_k + \sigma_w^2}, \quad \delta_k^2 = \frac{\tilde{\tau}\tilde{\rho}d_k^2}{\tilde{\tau}\tilde{\rho}d_k + \sigma_w^2}. \quad (6)$$

It is worth noticing that perfect channel estimation is approached for large $\tilde{\tau}$.

In the following, we will make use of the following approximation, namely *favorable propagation condition*, due to the mutual statistical independence of $\hat{\mathbf{g}}_k$ and due to $N \gg K$

$$\frac{1}{N}\hat{\mathbf{G}}^\dagger\hat{\mathbf{G}} \approx \mathbf{\Delta}. \quad (7)$$

where $\mathbf{\Delta} = \text{diag}(\delta_1^2, \dots, \delta_K^2)$.

Finally, the energy spent by the generic sensor is

$$\mathcal{E}_s = \tilde{\tau}\tilde{\rho} + (\tau_c - \tilde{\tau})\rho, \quad (8)$$

while the total energy spent by the WSN is $\mathcal{E} = K\mathcal{E}_s$.

IV. DECISION FUSION

The decision is usually performed as a test comparing a signal-dependent statistic ($\lambda(\mathbf{y})$) and a fixed threshold (γ)

$$\lambda(\mathbf{y}) \underset{\hat{\mathcal{H}}=\mathcal{H}_0}{\overset{\hat{\mathcal{H}}=\mathcal{H}_1}{\geq}} \gamma, \quad (9)$$

where $\hat{\mathcal{H}}$ denotes the estimated hypothesis. Performance are evaluated in terms of global probability of false alarm (q_f) and global probability of detection (q_d), defined as follows

$$q_f = \Pr(\lambda(\mathbf{y}) > \gamma | \mathcal{H}_0), \quad q_d = \Pr(\lambda(\mathbf{y}) > \gamma | \mathcal{H}_1). \quad (10)$$

It is worth noticing that $\Pr(\lambda > \gamma | \mathcal{H}_i)$ describes both q_f and q_d (with $i = 0$ and $i = 1$, respectively). The threshold in Eq. (9) is usually selected according to Bayes or Neyman-Pearson criteria [13]. For system performance evaluation, we consider the behavior of the global probability of detection (q_d) versus the global probability of false alarm (q_f), commonly denoted *receiver operating characteristic* (ROC).

A. Optimal Test

The log-likelihood ratio (LLR) of the received signal under the two hypotheses provides the optimal test¹

$$\begin{aligned} \lambda(\mathbf{y}) &= \log \left(\frac{p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_1)}{p(\mathbf{y} | \hat{\mathbf{G}}, \mathcal{H}_0)} \right) \\ &= \log \left(\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y} | \mathbf{x}, \hat{\mathbf{G}}) \prod_{k=1}^K \Pr(x_k | \mathcal{H}_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y} | \mathbf{x}, \hat{\mathbf{G}}) \prod_{k=1}^K \Pr(x_k | \mathcal{H}_0)} \right) \\ &= \log \left(\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y} - \sqrt{\rho}\hat{\mathbf{G}}\mathbf{x}\|^2}{\sigma_e^2}\right) \prod_{k=1}^K \Pr(x_k | \mathcal{H}_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y} - \sqrt{\rho}\hat{\mathbf{G}}\mathbf{x}\|^2}{\sigma_e^2}\right) \prod_{k=1}^K \Pr(x_k | \mathcal{H}_0)} \right), \end{aligned} \quad (11)$$

where we replaced $\mathbf{G} = \hat{\mathbf{G}} - \tilde{\mathbf{\Xi}}$ in Eq. (2) and denoted $\sigma_e^2 = \rho \sum_{k=1}^K \epsilon_k^2 + \sigma_w^2$. Also, it is worth noticing that $\Pr(x_k = +1 | \mathcal{H}_0) = p_f$ and $\Pr(x_k = +1 | \mathcal{H}_1) = p_d$.

Unfortunately, Eq. (11) is not attractive because: (i) requires availability of the CSI, of the local sensor performance, and of the signal-to-noise ratio (SNR); (ii) is numerically unstable because of exponential functions; (iii) exhibits exponential complexity with the number of sensors K . Design of sub-optimal DF rules with simpler implementation and reduced system knowledge is then extremely desirable.

B. Linear-Processing-Based Test

The optimal test can be replaced with a two-step architecture undergoing almost the same performance but much lower complexity. Starting from Eq. (2), we consider the linear processing $\mathbf{z} = \mathbf{A}^\dagger\mathbf{y}$ which denotes matched filter (MF), zero forcing (ZF) and MMSE processing, i.e.

$$\mathbf{A} = \begin{cases} \hat{\mathbf{G}} & \text{MF} \\ \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^\dagger\hat{\mathbf{G}}\right)^{-1} & \text{ZF} \\ \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^\dagger\hat{\mathbf{G}} + \frac{\sigma_e^2}{\rho}\mathbf{I}_K\right)^{-1} & \text{MMSE} \end{cases}. \quad (12)$$

¹We consider the optimal test conditioned on the estimated matrix $\hat{\mathbf{G}}$. The optimal statistic in absolute sense would be one jointly processing $\{\mathbf{y}; \tilde{\mathbf{Y}}\}$ (as in [14]) whose computational complexity is however unpractical.

It can be proven that \mathbf{z} represents a sufficient statistic for the three cases (the proof is omitted due to space limitations). Additionally, it is easily shown to be Gaussian distributed, namely $\mathbf{z}|\hat{\mathbf{G}}, \mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$. In the specific case of favorable propagation conditions

$$\boldsymbol{\mu}_z \approx \begin{cases} N\sqrt{\rho}\boldsymbol{\Delta}\mathbf{x} & \text{MF} \\ \sqrt{\rho}\mathbf{x} & \text{ZF, MMSE} \end{cases}, \quad (13)$$

$$\boldsymbol{\Sigma}_z \approx \begin{cases} N\sigma_e^2\boldsymbol{\Delta} & \text{MF} \\ \frac{\sigma_e^2}{N}\boldsymbol{\Delta}^{-1} & \text{ZF, MMSE} \end{cases}, \quad (14)$$

which reveals an appealing asymptotic (in the case of large-size array) behavior, i.e. independence among sensor contributions

$$p(\mathbf{z}|\hat{\mathbf{G}}, \mathbf{x}) \approx \prod_{k=1}^K p(z_k|x_k), \quad (15)$$

where

$$z_k|x_k \sim \begin{cases} \mathcal{N}_{\mathbb{C}}(N\sqrt{\rho}\delta_k^2 x_k, N\sigma_e^2\delta_k^2) & \text{MF} \\ \mathcal{N}_{\mathbb{C}}(\sqrt{\rho}x_k, \sigma_e^2/(N\delta_k^2)) & \text{ZF, MMSE} \end{cases}. \quad (16)$$

1) *Decode-and-Fuse (DaF) Approach*: Replacing Eq. (15) into (11) leads to a DFC operating over a PAC, whose properties have been studied in details in [1], [3]. Additionally, using the well-known Max-Log approximation [15], we finally get the test $\lambda(\mathbf{z}) = \sum_{k=1}^K \lambda_k(z_k)$, where $\lambda_k(z_k)$ is obtained through the following saturated linear structure

$$\lambda_k(z_k) = \begin{cases} c_1 & f_k(z_k) < a_1 \\ 2f_k(z_k) + c_2 & a_1 \leq f_k(z_k) < a_2 \\ c_3 & f_k(z_k) \geq a_2 \end{cases}, \quad (17)$$

with $c_1 = \log\left(\frac{1-p_d}{1-p_f}\right)$, $c_2 = \log\left(\frac{p_d}{1-p_f}\right)$, $c_3 = \log\left(\frac{p_d}{p_f}\right)$, $a_1 = \frac{1}{2}\log\left(\frac{p_d}{1-p_d}\right)$, $a_2 = \frac{1}{2}\log\left(\frac{p_f}{1-p_f}\right)$, and

$$f_k(z_k) = \begin{cases} \frac{2\sqrt{\rho}}{\sigma_e^2} \Re\{z_k\} & \text{MF} \\ \frac{2N\sqrt{\rho}\delta_k^2}{\sigma_e^2} \Re\{z_k\} & \text{ZF, MMSE} \end{cases}. \quad (18)$$

2) *Decode-then-Fuse (DtF) Approach*: Alternatively, we consider the quantized (one-bit per component) version of \mathbf{z} , namely $\tilde{\mathbf{z}} = \text{sign}(\Re\{\mathbf{z}\})$, in which case Eq. (16) is replaced with

$$\Pr(\tilde{z}_k = x_k|x_k) = 1 - \mathcal{Q}\left(\frac{\sqrt{2N\rho\delta_k^2}}{\sigma_e}\right), \quad (19)$$

and, exploiting the *Chair-Varshney rule* [1], the test is $\lambda(\tilde{\mathbf{z}}) = \sum_{k=1}^K \lambda_k(\tilde{z}_k)$, where $\lambda_k(\tilde{z}_k)$ is obtained through

$$\lambda_k(\tilde{z}_k) = u_k \log\left(\frac{\alpha_k(p_d)}{\alpha_k(p_f)}\right) + (1 - u_k) \log\left(\frac{1 - \alpha_k(p_d)}{1 - \alpha_k(p_f)}\right), \quad (20)$$

with $\alpha_k(p) = p\left(1 - \mathcal{Q}\left(\frac{\sqrt{2N\rho\delta_k^2}}{\sigma_e}\right)\right) + (1-p)\mathcal{Q}\left(\frac{\sqrt{2N\rho\delta_k^2}}{\sigma_e}\right)$ and $u_k = (1 + \tilde{z}_k)/2$.

3) *Computational Complexity*: It is worth noticing that the computational complexity of the linear-processing-based test is mainly due to the computation of the sufficient statistic \mathbf{z} or $\tilde{\mathbf{z}}$, then being $\mathcal{O}(KN)$. This is the computation complexity for the detection phase, when CSI is assumed to be known. The complexity for CSI acquisition, in the case of ZF and MMSE, is mainly due to the matrix inversion in Eq. (12), while in the case of MF the dominant term is due to Eq. (5).

V. SIMULATION RESULTS

We assume that \mathbf{D} is generated analogously as in [12], where sensors are deployed within a 2D ring with outer (resp. inner) radius $r_{\max} = 1$ km (resp. $r_{\min} = 0.1$ km) and the DFC is located in the center. The large-scale fading is modeled as $d_k = \nu_k (r_{\min}/r_k)^2$ where ν_k is a log-normal random variable, i.e. $10\log_{10}(\nu_k) \sim \mathcal{N}(\mu_\nu, \sigma_\nu^2)$ and r_k denotes the distance between the k th sensor and the DFC. CSI is obtained selecting $\tilde{\tau} = K$ (corresponding to the minimum length for the training sequences, i.e. the minimum the energy spent for channel training).

A WSN with $K = 10$ sensors with local performance $p_f = 0.05$ and $p_d = 0.5$ is considered. For sake of simplicity we assume $\rho = \tilde{\rho} = 1$ and $\sigma_w^2 = 1$. Performance with moderate-size array ($N = 20$) and large-size array ($N = 100$) at the DFC are compared.

The benchmark for performance evaluation is the *observation bound* [8], i.e. the performance achieved in the ideal case that the reporting channel is perfect

$$q_f = \sum_{\ell=c}^K \binom{K}{\ell} p_f^\ell (1 - p_f)^{K-\ell}, \quad (21)$$

$$q_d = \sum_{\ell=c}^K \binom{K}{\ell} p_d^\ell (1 - p_d)^{K-\ell}, \quad (22)$$

where $c \in \{0, \dots, K\}$ is a discrete threshold.

Fig. 1 shows the the global probability of detection (q_d) vs. the mean parameter of the log-normal fading (μ_ν , which can be thought as a sort of SNR), while fixing $\sigma_\nu = 2$ dB and $q_f = 0.01$. Both DaF (solid lines) and DtF (dashed lines) approaches are considered. The loss in performance of the latter with respect to the former is apparent in some cases (MF in moderate- and large-size arrays, ZF in large-size array, MMSE in moderate-size array), while in other cases both approaches perform quite similar for a large SNR range.

Fig. 2 compares the ROC curves only for the DaF approach. It is apparent how the gap from the optimal performance reduces dramatically when large-size arrays are considered, confirming the asymptotic optimality. This is more visible in Fig. 3 where the ZF linear processing seems to guarantee the fastest convergence with respect to the number of receive antennas (N).

VI. CONCLUSION

Channel-aware (through pilot-based channel estimation) decision fusion over massive MIMO has been analyzed. Linear processing techniques (MF, ZF, and MMSE) have been considered in order to design low-complexity fusion rules

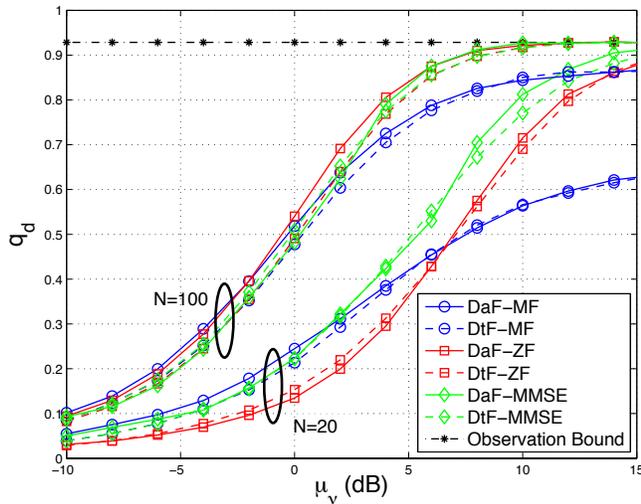


Fig. 1. q_d vs μ_ν at $q_f = 0.01$ with for DaF (solid lines) and DtF (dashed lines) approaches with $K = 10$ sensors. Log-normal fading with $\sigma_\nu = 2$ dB is assumed.

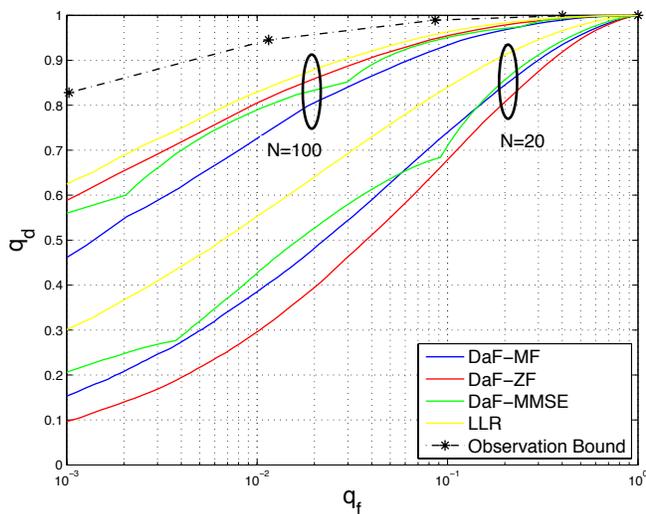


Fig. 2. q_d vs q_f at $q_f = 0.01$ for DaF approach with $K = 10$ sensors. Log-normal fading parameters are $\mu_\nu = 4$ dB and $\sigma_\nu = 2$ dB.

which asymptotically (for extremely large number of receive antennas) approach the optimum performance. DaF and DtF approaches have been compared with DaF-ZF being the most interesting in terms of complexity and performance.

REFERENCES

- [1] 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.
- [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] 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.

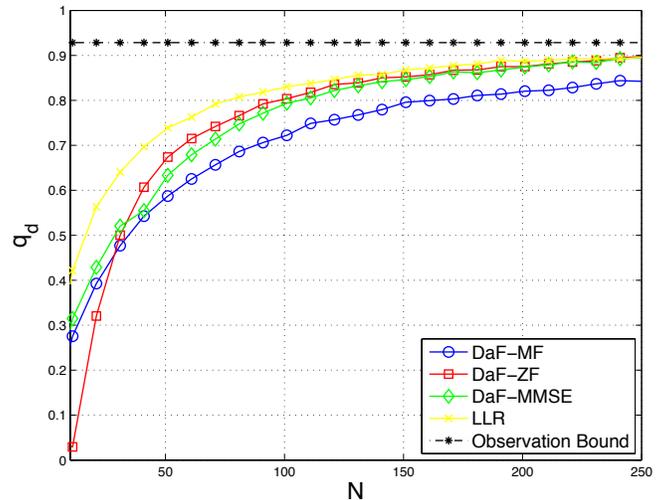


Fig. 3. q_d vs N at $q_f = 0.01$ for DaF approach with $K = 10$ sensors. Log-normal fading parameters are $\mu_\nu = 4$ dB and $\sigma_\nu = 2$ dB.

- [4] 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, May 2007.
- [5] C. R. Berger, M. Guerriero, S. Zhou, and P. Willett, "PAC vs. MAC for decentralized detection using noncoherent modulation," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3562–3575, Sep. 2009.
- [6] 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.
- [7] M. K. Banavar, A. D. Smith, C. Tepedelenioglu, 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.
- [8] 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.
- [9] D. Ciuonzo, G. Romano, and P. Salvo Rossi, "Optimality of received energy in decision fusion over a Rayleigh fading diversity MAC with non-identical sensors," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 22–27, Jan. 2013.
- [10] D. Ciuonzo, G. Romano, and P. Salvo Rossi, "Performance analysis and design of maximum ratio combining in channel-aware MIMO decision fusion," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4716–4728, Sept. 2013.
- [11] 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, Jan. 2013.
- [12] 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.
- [13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall PTR, 1998.
- [14] 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.
- [15] P. Robertson, E. Vilebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *IEEE Int. Conf. Commun. (ICC)*, vol. 2, pp. 1009–1013, Jun. 1995.