# HMM-Based Decision Fusion in Wireless Sensor Networks With Noncoherent Multiple Access

Pierluigi Salvo Rossi, *Senior Member, IEEE*, Domenico Ciuonzo, *Member, IEEE*, and Torbjörn Ekman, *Member, IEEE*

***Abstract*—We develop a novel decision fusion (DF) approach which exploits time-correlation of the unknown binary source under observation through a wireless sensor network (WSN) reporting local decisions to a fusion center (FC) over interfering Rayleigh fading channels. The system is modeled via a hidden Markov model (HMM): both learning and detection phases are developed. The learning phase is blind, i.e. it requires only a set of observations without knowledge of the corresponding source states. Remarkably, the approach allows the FC to take decisions without knowledge of the local sensor performance. Numerical results confirm the effectiveness of the proposed approach.**

***Index Terms*—Decision fusion (DF), hidden Markov model (HMM), multiple-access channel (MAC), wireless sensor network (WSN).**

## I. Introduction

Wireless sensor networks (WSN) are often employed in distributed detection tasks. Focusing on centralized architectures [1], sensors transmit local decisions to a fusion center (FC) which takes a decision by combining the received information. Moving from parallel-access channel (PAC) to multiple-access channel (MAC), the interfering nature of the wireless medium was recently exploited [2], [3]. Multiantenna processing at the FC was investigated in [4], and energy detection was studied in *non-line-of-sight* fading [5], [6]. Other works replace uncoded transmission with use of channel coding at sensor location and iterative decoding at the FC [7].

Unfortunately, prior knowledge of local performance (i.e. sensors probabilities of false alarm and detection) at the FC is often unpractical [8]. Estimation of local performance in decision fusion (DF) over PAC was considered in [9]–[11].

In this letter we consider a blind approach for DF in noncoherent MAC scenarios where the source-state sequence follows a binary Markov chain. More specifically, it requires an unsupervised training phase which is based on measurements only (no need for the corresponding source states). The proposed approach exploits the framework of hidden Markov models (HMMs), works with a reduced-order model, and does not require knowledge of sensor performance. The main contributions are: (i) to propose two versions of maximum likelihood (ML) algorithms for DF; (ii) to compare them with the single-observation energy detector (ED), i.e. the classical approach [2], [3], [5]; (iii) to link DF and HMM frameworks thus opening to solutions from machine-learning area.

The outline of the letter is the following: Section II gives some preliminaries on HMMs; Section III describes the system model under investigation; Section IV explains how to exploit results from HMMs for DF in WSNs; numerical simulations to confirm our claims are provided and discussed in Section V.

*Notation:* Lower (resp. upper) bold letters denote vectors (resp. matrices), with $a_n$ (resp. $A_{n,m}$) being the $n$th (resp. $(n,m)$th) entry of $\boldsymbol{a}$ (resp. $\boldsymbol{A}$); $\delta_{n,m}$ is the Kronecker delta; $\mathbb{E}\{\cdot\}$, $(\cdot)^t$, $(\cdot)^*$ denote expectation, transpose, and conjugate operators; $\Pr(\mathcal{A})$ and $p(a)$ denote the probability of the event $\mathcal{A}$ and the probability density function (PDF) of the random variable $a$; $|a|$ is the modulus of $a$; $\mathcal{U}(a,b)$ denotes a (real-valued) uniform distribution over $(a,b)$; $\mathcal{N}_{\mathbb{C}}(\mu,\sigma^2)$ denotes a (complex-valued) proper normal distribution with mean $\mu$ and variance $\sigma^2$; $\sim$ means "distributed as".

## II. HMM Learning and Reconstruction

An HMM with complex Gaussian mixture (GM) PDFs [12] is a stochastic model made of a (hidden) state variable $s_n \in \{0,\ldots,J-1\}$ and an observation variable $y_n \in \mathcal{C}$, where $n$ denotes discrete time. State dynamic behavior is ruled, with a Markovian assumption, by the transition matrix $\boldsymbol{A}$, i.e. $A_{i,j} = \Pr(s_n = j | s_{n-1} = i)$. The observation is modeled through conditionally independent (given the state) complex GMs with $M$ components, i.e. $y_n | \{s_n = i\} \sim \sum_{m=1}^{M} \theta_{i,m} \mathcal{N}_{\mathbb{C}}(\mu_{i,m}, \sigma^2_{i,m})$ with conditional PDF $b_i(y) = \sum_{m=1}^{M} \frac{\theta_{i,m}}{\pi\sigma^2_{i,m}} \exp\left(-\frac{|y-\mu_{i,m}|^2}{\sigma^2_{i,m}}\right)$. The HMM is completely specified by the set of parameters $\mathcal{H} = \{\boldsymbol{A}; \{\theta_{i,m}; \mu_{i,m}; \sigma^2_{i,m}\}_{m=1}^{M}{}_{i=0}^{J-1}\}$. The observation PDF is $p(y) = \sum_{i=0}^{J-1} \pi_i b_i(y)$, where $\{\pi_0,\ldots,\pi_{J-1}\}$ is the steady-state probability distribution, i.e. $\pi_i = \lim_{n\to\infty} \Pr(s_n = i)$.

### A. Learning the Model Parameters

The Expectation-Maximization (EM) algorithm is an iterative procedure allowing for ML estimation of the model parameters ($\mathcal{H}$) from one (or more) training sequence of observations $\{y_0,\ldots y_{N-1}\}$. In the case of HMMs, it is known as the Baum-Welch (BW) algorithm [12] and is based on the forward and backward variables, computed recursively through

$$\alpha_n(j) = \sum_{i=0}^{J-1} \alpha_{n-1}(i) A_{i,j} b_j(y_n), \tag{1}$$

$$\beta_n(i) = \sum_{j=0}^{J-1} A_{i,j} b_j(y_{n+1}) \beta_{n+1}(j), \tag{2}$$

being $\alpha_0(i) = \delta_{0,i}$ and $\beta_{N-1}(j) = 1$. Also, we define

$$\gamma_n(i;m) = \frac{\alpha_n(i)\beta_n(i)}{\sum_{j=0}^{J-1} \alpha_n(j)\beta_n(j)} \frac{\theta_{i,m} \frac{\partial}{\partial\theta_{i,m}} b_i(y_n)}{b_i(y_n)}. \tag{3}$$

Manuscript received December 18, 2014; accepted March 11, 2015. Date of publication March 16, 2015; date of current version May 7, 2015. This work was supported by the IME Faculty at NTNU with project CAMOS. The associate editor coordinating the review of this paper and approving it for publication was C. Zhong.

P. Salvo Rossi and T. Ekman are with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, 7491 Trondheim, Norway (e-mail: salvorossi@iet.ntnu.no; torbjorn.ekman@iet.ntnu.no).

D. Ciuonzo is with the Department of Electrical Engineering and Information Technology, University of Naples "Federico II", 80138 Naples, Italy (e-mail: domenico.ciuonzo@ieee.org).

Digital Object Identifier 10.1109/LCOMM.2015.2413407

More specifically, the BW algorithm is an iterative procedure based on the following updates

$$\hat{A}_{i,j} = \frac{\sum_{n=0}^{N-2} \alpha_n(i) A_{i,j} b_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{n=0}^{N-2} \alpha_n(i)\beta_n(i)}, \tag{4}$$

$$\hat{\theta}_{i,m} = \frac{\sum_{n=0}^{N-1} \gamma_n(i;m)}{\sum_{n=0}^{N-1}\sum_{m=1}^{M} \gamma_n(i;m)}, \tag{5}$$

$$\hat{\mu}_{i,m} = \frac{\sum_{n=0}^{N-1} \gamma_n(i;m) y_n}{\sum_{n=0}^{N-1} \gamma_n(i;m)}, \tag{6}$$

$$\hat{\sigma}^2_{i,m} = \frac{\sum_{n=0}^{N-1} \gamma_n(i;m)(y_n - \mu_{i,m})^2}{\sum_{n=0}^{N-1} \gamma_n(i;m)} . \tag{7}$$

The termination of the procedure is based on the evaluation of the likelihood of the training sequence given the model parameters,[1] given as $p(\{y_0, \ldots y_{N-1}\}|\mathcal{H}) = \sum_{i=1}^{J-1} \alpha_n(i)\beta_n(i)$.

### B. State Reconstruction

For given test sequence of observations $\{y_0, \ldots y_{W-1}\}$ and learned model ($\mathcal{H}$), different approaches can be used for inferring the corresponding state sequence $\{\hat{s}_0, \ldots \hat{s}_{W-1}\}$, depending on various performance metric. The two most popular approaches are [12]: (i) finding the *most-likely sequence* of states; (ii) finding the sequence of *most-likely states*.

The former maximizes $p(s_0, \ldots, s_{W-1}, y_0, \ldots, y_{W-1}|\mathcal{H})$, i.e. a joint distribution w.r.t. the states. The solution is provided by the Viterbi algorithm (VA), which is based on

$$\xi_n(j) = \max_i \xi_{n-1}(i) A_{i,j} b_j(y_n), \tag{8}$$

$$\psi_n(j) = \arg\max_i \xi_{n-1}(i) A_{i,j}, \tag{9}$$

where $\xi_0(i) = \pi_i b_i(y_0)$ and $\psi_0(i) = 0$, plus path backtracking $\hat{s}_n = \psi_{n+1}(\hat{s}_{n+1})$ with $\hat{s}_{W-1} = \arg\max_i \xi_{W-1}(i)$.

The latter maximizes $p(s_n, y_0, \ldots, y_{W-1}|\mathcal{H})$ with $n = 0, \ldots, W-1$, i.e. a marginal distribution w.r.t. the states. The solution is provided by the Forward-Backward (FB) algorithm, which is based on the following decision rule

$$\hat{s}_n = \arg\max_i \alpha_n(i)\beta_n(i), \tag{10}$$

with initialization $\alpha_0(i) = \pi_i$, different than BW algorithm.

## III. System Model

We consider $K$ sensors sensing a binary source which is modeled as a Markov chain. The source state at discrete-time $n$ is denoted $s[n] \in \{0, 1\}$. We assume that the local sensing/decision process is stationary, homogeneous, and fully described by the (conditionally independent given the hypothesis) local probabilities of false alarm ($p_0$) and detection ($p_1$).

Sensors, each with one single transmit antenna, communicate simultaneously their decision to a FC whose aim is to provide a robust decision on the basis of the multiple received information. All the sensors employ the same binary modulation with identical parameters (transmission pulse, carrier frequency, etc.), and for energy saving purposes we consider on-off keying (OOK) modulation (i.e. the sensor is silent if detects $s[n] = 0$). Identical statistical behavior is assumed for each link from the generic sensor to the FC.[2]

We assume that the system is synchronized. The analysis of synchronization errors is beyond the scope of this work.

Referring to discrete-time $n$, we denote $x_k[n] \in \{0, 1\}$ the symbol transmitted by the $k$th sensor encoding its local decision about the source state $s[n]$; $h_k[n] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_h^2)$ the fading channel coefficient on the link between the $k$th sensor and FC; $y[n]$ the signal received by the FC; and $w[n] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_w^2)$ the additive white Gaussian noise at the FC. Also, local performance corresponds to $p_i = \Pr(x[n] = 1|s[n] = i)$. The discrete-time model for the received signal is then

$$y[n] = \sum_{k=1}^{K} h_k[n] x_k[n] + w[n] . \tag{11}$$

The source is completely described by the steady-state probability and the memory, respectively defined as

$$\pi_0 = \Pr(s[n] = 0) = \frac{A_{1,0}}{A_{0,1} + A_{1,0}}, \tag{12}$$

$$\rho = \frac{\Pr(s[n] = 0|s[n-1] = 0)}{\Pr(s[n] = 0|s[n-1] = 1)} = \frac{1 - A_{0,1}}{A_{1,0}} . \tag{13}$$

The larger $\rho$, the stronger the memory ($\rho = 1$ denotes a memoryless source). If not necessary, we drop the index $n$.

Denoting $\phi_{i,a} = \Pr(a|s = i) = \binom{K}{a} p_i^a (1 - p_i)^{K-a}$, where $a = \sum_{k=1}^{K} x_k$ is the number of active sensors, the received signal follows the conditional GM PDF

$$p(y|s = i) = \sum_{a=0}^{K} \Pr(a|s = i) p(y|a) = \sum_{a=0}^{K} \frac{\phi_{i,a}}{\pi\left(a\sigma_h^2 + \sigma_w^2\right)} \exp\left(-\frac{|y|^2}{a\sigma_h^2 + \sigma_w^2}\right) . \tag{14}$$

## IV. HMM-Based Decision Fusion

From Eq. (14), it is apparent how the system described in Section III corresponds to an HMM with $J = 2$ states and complex GM PDFs with $K + 1$ zero-mean components. However, we will show that fewer parameters are needed to model accurately the system. i.e. an under-determined HMM is still able to capture effectively the stochastic behavior of the WSN. We investigate on the suitability of using techniques from HMM for training and source monitoring. Differently from classical approaches for DF in WSNs, which typically require knowledge of local performance, SNR level, and statistical/instantaneous CSI, we use: (i) a training sequence to adjust

[1] Being iterative, the algorithm suffers from local extremes. When necessary, trainings with different initial conditions provide the global solution. Also, the implementation requires proper scaling to avoid underflow [12].

[2] Symmetry assumptions (both for the local performance and the channels to the FC) are only for simplicity of the description. The methodology is suitable for asymmetric scenarios, but the mathematical description is tedious.

the parameters of an HMM through the BW algorithm; (ii) such a trained model for source monitoring through VA- or FB-based state reconstruction.

Performance is evaluated in terms of global probabilities of false alarm $(q_f)$ and detection $(q_d)$, defined as follows

$$q_f = \Pr(\hat{s}=1|s=0)\,,\; q_d = \Pr(\hat{s}=1|s=1)\ . \qquad (15)$$

A tunable threshold (selected according to Bayes or Neyman-Pearson criteria [13]) allows the FC trading-off between the global probabilities of false alarm and detection. The behavior of the global probabilities of detection $(q_d)$ vs. false alarm $(q_f)$ is denoted *receiver operating characteristic* (ROC).

### A. Learning Phase

A training sequence is used for appropriate selection of the set of parameters $\mathcal{H}$. It is worth remarking that the BW algorithm requires the observation sequence only, and does not need the corresponding state sequence, i.e. it is a blind procedure from the point of view of the source estimation.

Since the conditional means are null,[3] and also the weights and the conditional variances are dependent on a few parameters $(K, p_0, p_1, \sigma_h^2, \sigma_w^2)$, we use an HMM with zero-mean GMs with $M \ll K+1$ components. The number of hidden states is obviously set to $J=2$. Correct association of states (trained model to binary source) is forced through the starting model as described below. The number of components $M$ must be selected trading-off performance and complexity. Various techniques may be considered for choosing of $M$ (e.g. [14]), however such a discussion falls beyond the scope of this letter.

The starting model is chosen as follows: $A_{i,j}=\frac{1}{2}$, $\theta_{i,m}=\frac{1}{M}$, $\sigma_{0,m}^2 \sim \mathcal{U}(0,\tilde{\sigma}_y^2)$, $\sigma_{1,m}^2 \sim \mathcal{U}(0,4\tilde{\sigma}_y^2)$ where $\tilde{\sigma}_y^2$ denotes the empirical variance of the training sequence. Weights are set equal such that all states and components have equal chances to contribute to the model. Randomness in the variances of each component is introduced to explore the parameter space while lying within a neighborhood of the empirical variance. It is worth noticing that the variance in state $s=0$ is (on average) smaller than the variance in state $s=1$, as expected from the physical model. The corresponding source parameters for the starting model are $\pi_0 = 1/2$ and $\rho = 1$.

### B. Monitoring Phase

The VA and the FB algorithm in Section II provide the most likely state sequence and the sequence of most likely states, respectively. However, it is apparent that they do not allow any trade-off between global probabilities of false alarm and detection. The ROC for the standard VA and for the standard FB algorithm degenerates into a *single operation point*.

In order to have a proper ROC, with operating point tunable at the FC depending on a threshold $\zeta \in [0,+\infty)$, we modify both the VA and the FB algorithm replacing (in Eqs. (8) and (9) for the VA and Eq. (10) for the FB algorithm) the operators $\max(\cdot,\cdot)$ and $\arg\max(\cdot,\cdot)$ with the operators $\lambda(\cdot,\cdot;\zeta)$ and $\arg\lambda(\cdot,\cdot;\zeta)$, respectively, defined as

$$\lambda(u_0,u_1;\zeta)=\begin{cases}u_0 & \frac{u_1}{u_0}\le\zeta\\ u_1 & \frac{u_1}{u_0}>\zeta,\end{cases}\quad \arg\lambda(u_0,u_1;\zeta)=\begin{cases}0 & \frac{u_1}{u_0}\le\zeta\\ 1 & \frac{u_1}{u_0}>\zeta.\end{cases} \qquad (16)$$

It is worth noticing that $\lambda(\cdot,\cdot;\zeta=1)=\max(\cdot,\cdot)$ and $\arg\lambda(\cdot,\cdot;\zeta=1)=\arg\max(\cdot,\cdot)$, then selecting $\zeta=1$ we get the standard VA and the standard FB algorithm.

Both the modified VA and the modified FB algorithm are applied to consecutive non-overlapping windows of length $W$, i.e. states $\{s[(\ell-1)W],\ldots,s[\ell W-1]\}$ are estimated through the observation $\{y[(\ell-1)W],\ldots,y[\ell W-1]\}$. The impact of $W$ is evaluated in Section V, with small values to be preferred to reduce the estimation delay.

TABLE I
STATISTICS OF THE SOURCE PARAMETERS FOR THE TRAINED MODELS. THE TRUE SOURCE PARAMETERS ARE $\pi_0 = 0.9$ AND $\rho = 10$

| SNR | $\hat{\pi}_0$ (st.d.) | $\hat{\rho}$ (st.d.) |
|---|---|---|
| 5 dB | 0.851 (0.064) | 8.94 (5.26) |
| 15 dB | 0.853 (0.042) | 9.31 (4.28) |

## V. SIMULATION RESULTS AND DISCUSSION

Results refer to Monte Carlo simulations using MATLAB. We considered a binary source with $\pi_0=0.9$ and $\rho=10$, i.e. an anomaly-detection problem, and a WSN with $K=10$ sensors whose local performance are $p_0=0.05$ and $p_1=0.5$. Two scenarios were considered for the signal-to-noise ratio (SNR), defined as $\sigma_h^2/\sigma_w^2$, i.e. SNR $=5$ dB and SNR $=15$ dB.

An HMM with $M=3$ components per state was learned for each scenario, with a training sequence of $N=10^3$ observations. It is worth noticing that the real WSN presents 44 parameters and is mapped into a reduced model with 12 parameters. Trained models were tested with a test sequence of $10^5$ observations (with corresponding state sequence) from the real WSN, to compute the global performance ($q_f$ and $q_d$) for different approaches.

Estimated values of the source parameters (mean $\hat{\pi}_0$ and $\hat{\rho}$ with corresponding standard deviation over 50 repeated experiments) are shown in Table I. Simulations confirm that the training works better for higher SNR, both in terms of computational complexity (fewer iterations to achieve convergence) and performance (higher likelihood). The impact of the length of the training sequence on the performance of the trained model has not been investigated, although some preliminary results confirm that scenarios with larger SNR requires less data for satisfactory training. The excellent fitting capability is not shown for brevity: empirical PDF of the training sequences matches perfectly the analytical PDF of the trained models.

Fig. 1 shows the impact of the window length $(W)$ on the global performance when VA and FB algorithm are considered. Three different values for the threshold are shown ($\zeta=1/9$, 1, 9). It is apparent how the VA outperforms the FB algorithm in most cases, unless low SNR, large window length, and small threshold (i.e. high $q_f$ and high $q_d$) are selected (see the blue curves in Figs. 1(a) and 1(b)). This could be related to the

[3]It is worth noticing that $\mu_{i,m}=0$ and then Eq. (6) must be ignored.

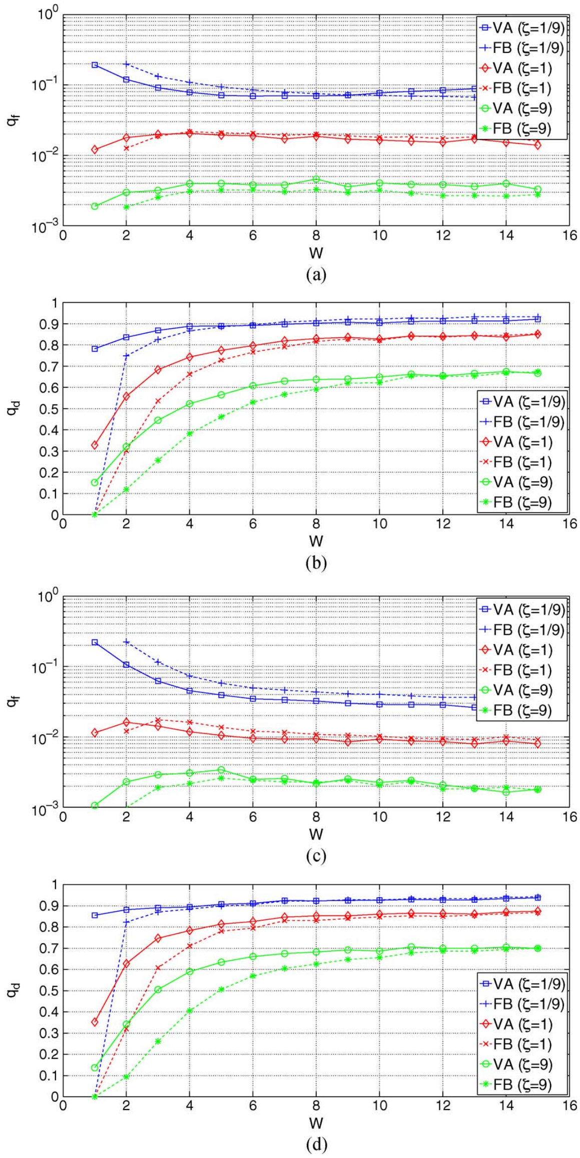


Fig. 1. Impact of the monitoring window $(W)$. (a) $q_f$ at SNR $= 5$ dB. (b) $q_d$ at SNR $= 5$ dB. (c) $q_f$ at SNR $= 15$ dB. (d) $q_d$ at SNR $= 15$ dB.

robustness of the sequence-detection approach in absence of channel state information.

Fig. 2 compares the performance in terms of ROC curves of the VA and the FB algorithm. Also, the ROC for the ED, shown to be the optimum receiver for isolated decisions (i.e. $W = 1$) over Rayleigh channels [5], are reported for comparison. Dashed and dash-dotted lines in Fig. 2(b) refer to the ROC for HMMs with $M = 2$ and $M = 4$ components, respectively, showing how $M = 3$ is a reasonable choice for the considered scenario. Performance of the standard VA and standard FB algorithm are also depicted, as well as sensor local performance: the VA outperforms the FB algorithm. Also, the performance of the VA with $W = 1$ equals those of the ED, thus the VA may be considered as a valid extension of the ED to exploit time correlation of the source.

Summarizing, HMMs represent a suitable alternative for blind DF in WSNs. Numerical results confirm the effectiveness of the proposed approach, with VA outperforming FB.

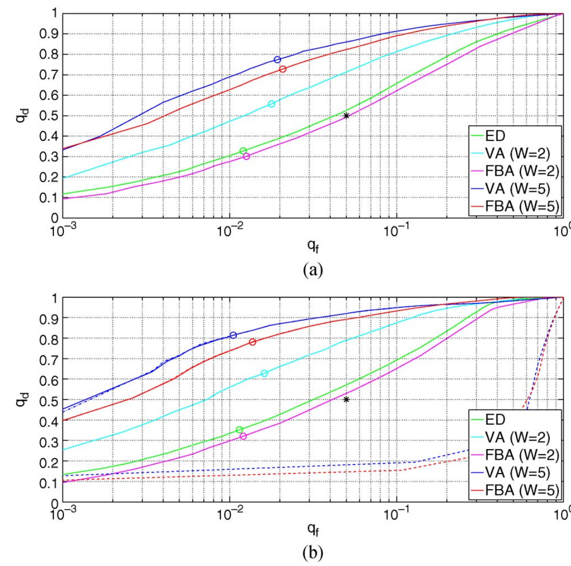


Fig. 2. ROC. Circles refer to the standard VA and standard FB algorithm. The black asterisk denotes sensor local performance. Dashed and dot-dashed lines refer to the case with $M = 2$ and $M = 5$ components. (a) SNR $= 5$ dB. (b) SNR $= 15$ dB.

## REFERENCES


[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] P. Salvo Rossi, D. Ciuonzo, K. Kansanen, and T. Ekman, "On energy detection for MIMO decision fusion in wireless sensor networks over NLOS fading," *IEEE Commun. Lett.*, vol. 19, no. 2, pp. 303–306, Feb. 2015.

[7] J. Del Ser, D. Manjarres, P. M. Crespo, S. Gil-Lopez, J. Garcia-Frias, "Iterative fusion of distributed decisions over the Gaussian multiple-access channel using concatenated BCH-LDGM codes," *EURASIP J. Wireless Commun. Netw.*, vol. 2011, pp. 1–12, Jan. 2011.

[8] 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.

[9] N. Ansari, E. Hou, B. Zhu, and J. Chen, "Adaptive fusion by reinforcement learning for distributed detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 524–531, Apr. 1996.

[10] G. Mirjalili, Z. Luo, T. Davidson, and E. Bosse, "Blind adaptive decision fusion for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 34–52, Jan. 2003.

[11] B. Liu, A. Jeremic, and K. Wong, "Optimal distributed detection of multiple hypotheses using blind algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 1, pp. 317–331, Jan. 2011.

[12] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

[14] J. Du, Y. Hu, and H. Jiang, "Boosted mixture learning of Gaussian mixture hidden Markov models based on maximum likelihood for speech recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2091–2100, Sep. 2011.