Multiple–Symbol Differential Decision Fusion for Mobile Wireless Sensor Networks

A. Lei and R. Schober

Department of Electrical and Computer Engineering
The University of British Columbia
2356 Main Mall, Vancouver, BC, V6T 1Z4, Canada
Phone: +604 - 822 - 3515, Fax: +604 - 822 - 5949
E-mail: {alei, rschober}@ece.ubc.ca

Abstract

We consider the problem of decision fusion in mobile wireless sensor networks where the channels between the sensors and the fusion center are time–variant. We assume that the sensors make independent local decisions on the $M$ hypotheses under test and report these decisions to the fusion center using differential phase–shift keying (DPSK), so as to avoid the channel estimation overhead entailed by coherent decision fusion. For this setup we derive the optimal and three low–complexity, suboptimal fusion rules which do not require knowledge of the instantaneous fading gains. Since all these fusion rules exploit an observation window of at least two symbol intervals, we refer to them collectively as multiple–symbol differential (MSD) fusion rules. For binary hypothesis testing, we derive performance bounds for the optimal fusion rule and exact or approximate analytical expressions for the probabilities of false alarm and detection for all three suboptimal fusion rules. Simulation and analytical results confirm the excellent performance of the proposed MSD fusion rules and show that in fast fading channels significant performance gains can be achieved by increasing the observation window to more than two symbol intervals.

---

1This work was submitted in part to the IEEE Global Telecommunications Conference (Gobecom), 2009.
1 Introduction

Decentralized detection is an important task in wireless sensor networks (WSNs) [1–4]. To limit complexity, the sensors usually make independent decisions based on their respective observations and forward these decisions over the wireless channel to a fusion center which forms a final decision on the hypothesis under test. Most of the existing literature on the decentralized detection problem assumes ideal error–free communication between the sensors and the fusion center. While this is a reasonable assumption for wired sensors, it may lead to significant performance degradations if wireless sensors are employed. Therefore, the problem of fusing sensor decisions transmitted over noisy fading channels has received considerable interest recently. For example, channel aware decision fusion for phase–coherent WSNs employing phase–shift keying (PSK) modulation was investigated in [5, 6]. In [7], channel statistics based fusion rules for WSNs employing on/off keying (OOK) modulation were considered. The impact of fading on the performance of power constrained WSNs was studied in [8]. In [9], the performance of type–based multiple access strategies for fading WSNs was analyzed. Furthermore, the problem of optimal power scheduling and decision fusion in fading WSNs with amplify–and–forward processing at the sensors was considered in [10]. Most recently, the impact of channel errors on decentralized detection was studied for PSK, OOK, and frequency–shift keying (FSK) modulation in [11].

Interestingly, existing work on decision fusion for noisy fading channels has mainly considered coherent (e.g. PSK) and noncoherent (e.g. OOK, FSK) modulation schemes. While the former are suitable for static fading channels, the latter are appropriate for extremely fast fading channels, where the fading gain changes from symbol to symbol due to e.g. fast frequency hopping. However, for applications where the fading gains change slowly over time due to the mobility of the sensors and/or fusion center, noncoherent modulation may not be a preferred choice due to the inherent loss in power efficiency compared to coherent modulation. On the other hand, coherent modulation requires the insertion of pilot symbols for channel estimation which reduces spectral efficiency and complicates system design. Thus, for conventional point–to–point communication systems differential PSK (DPSK) is often preferred for signaling over time–varying fading channels [12]. While DPSK does not require instantaneous channel state information (CSI) for detection, the performance loss compared to coherent PSK can be mitigated by multiple–symbol differential detection (MSDD) if statistical CSI is available at the receiver [13–15]. This motivates the investigation of DPSK for transmission in WSNs and the design of corresponding fusion rules.
In this paper, we consider the decentralized $M$–ary hypothesis testing problem in time–variant fading channels. We assume that the sensors employ $M$–DPSK to report their local decisions to the fusion center and derive corresponding multiple–symbol differential (MSD) fusion rules. Since the complexity of the optimal fusion rule is exponential in both the number of sensors and the observation window size used for MSD decision fusion, we propose three suboptimal fusion rules with significantly lower complexity and good performance. All considered fusion rules only require statistical CSI but not any knowledge about the instantaneous channel gains. For the special case of binary hypothesis testing ($M = 2$), we provide performance bounds for the optimal fusion rule, and exact or approximate analytical expressions for the probabilities of false alarm and detection for the suboptimal fusion rules. Our analytical and simulation results show that significant performance gains can be achieved by increasing the observation window size of the MSD fusion rules to more than two symbols. In particular, the performance of coherent detection with perfect knowledge of the channel gains can be approached for large enough observation window sizes.

This paper is organized as follows. In Section 2, we introduce the system model. The optimal and suboptimal fusion rules are derived in Section 3, and their performance is analyzed in Section 4. In Section 5, simulation and numerical results are presented, and conclusions are drawn in Section 6.

**Notation:** In this paper, bold upper case and lower case letters denote matrices and vectors, respectively. $[\cdot]^T$, $[\cdot]^H$, $(\cdot)^*$, $\Re\{\cdot\}$, and $\mathcal{E}\{\cdot\}$ denote transposition, Hermitian transposition, complex conjugation, the real part of a complex number, and statistical expectation, respectively. $\delta[\cdot]$ and $\bar{u}[\cdot]$ refer to the discrete–time Delta and unit step functions, respectively. In addition, $[X]_{i,j}$, $\det(\cdot)$, $I_X$, $\otimes$, and $\text{diag}\{X_1, X_2, \ldots, X_X\}$ stand for the element of matrix $X$ in row $i$ and column $j$, the determinant of a matrix, the $X \times X$ identity matrix, the Kronecker product, and a block diagonal matrix with matrices $X_1, X_2, \ldots, X_X$ on its main diagonal, respectively. Finally, $P(\cdot)$ and $p(\cdot)$ are used to denote probabilities and probability density functions (pdf), respectively. In particular, $P(A|B)$ and $p(a|b)$ denote the probability of event $A$ conditioned on event $B$ and the pdf of random variable $a$ conditioned on random variable $b$, respectively.

## 2 System Model

In this paper, we consider the distributed multiple hypothesis testing problem where a set $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$ of $K$ sensors are used to decide which one out of $M$ possible hypotheses $H_i$, 

...
i ∈ \mathcal{M}, \mathcal{M} \triangleq \{0, 1, \ldots, M - 1\}, is present. The \textit{a priori} probability of hypothesis \(H_i\) is denoted by \(P(H_i)\), \(i \in \mathcal{M}\). Fig. 1 illustrates the system model which will be discussed in detail in the following subsections.

### 2.1 Processing at Sensors

At time \(n \in \mathbb{Z}\) each sensor \(k \in \mathcal{K}\) makes an \(M\)–ary decision \(u_k[n]\) based on its own noisy observation \(x_k[n]\). We assume that the \(K\) observations \(x_k[n]\), \(k \in \mathcal{K}\), are independent of each other, conditioned on the \(M\) different hypotheses. The sensors map their local decisions to \(M\)–ary PSK (\(M\)–PSK) symbols \(a_k[n] \in \{w_i|i \in \mathcal{M}\}\), \(w_i \triangleq e^{j2\pi i/M}\), such that hypothesis \(H_i\) corresponds to the PSK symbol \(w_i\). The differential phase symbols \(a_k[n]\) are differentially encoded before transmission over the wireless channel to obtain the absolute phase symbols

\[
s_k[n] = a_k[n]s_k[n - 1],
\]

where \(s_k[n] \in \{w_i|i \in \mathcal{M}\}\). This differential encoding operation facilitates detection without CSI at the receiver which is particularly useful for transmission over time–variant fading channels [12]. In the context of WSNs, such time–variant channels may arise for example in vehicular WSNs with mobile sensors and/or mobile fusion centers [16], battlefield surveillance [17], or collaborative spectrum sensing with mobile nodes [18]. To keep our model general, we quantify the quality of the local decisions made by the sensors in terms of conditional probabilities \(P_k(a_k[n] = w_j|H_i), i \in \mathcal{M}, j \in \mathcal{M}, k \in \mathcal{K}\).

### 2.2 Channel Model

The sensors communicate with the fusion center over orthogonal flat fading channels using e.g. a time–division multiple access (TDMA) protocol. The received signal from sensor \(k\) at time \(n\) is given by

\[
y_k[n] = \sqrt{P_K}h_k[n]s_k[n] + n_k[n],
\]

where \(P_K \triangleq P/K\) with total transmitted power \(P\), and \(h_k[n]\) and \(n_k[n]\) denote the fading gain and zero–mean complex–valued additive white Gaussian noise (AWGN), respectively. The noise is independent, identically distributed (i.i.d.) with respect to both the sensors, \(k\), and time, \(n\), and has variance \(\sigma_n^2 \triangleq \mathbb{E}\{|n_k[n]|^2\}\). We assume independent, non–identically distributed (i.n.d.) Rayleigh fading with fading gain variances \(\sigma_k^2 \triangleq \mathbb{E}\{|h_k[n]|^2\}, k \in \mathcal{K}\). For the temporal correlation of the
fading gains, we adopt Clarke’s model with 
\[ \varphi_{hh,k}(\lambda) \triangleq \mathcal{E}\{h_k[n + \lambda]h_k^*[n]\} = \sigma_k^2 J_0(2\pi B_k T \lambda), \]
where \( B_k \) denotes the Doppler shift of sensor \( k \) and \( T \) denotes the time interval between two observations \( y_k[n] \) and \( y_k[n+1] \). Note that if the sensors use TDMA to report their observations in a round–robin fashion to the fusion center, \( T \) is equal to \( T = K T_s \), where \( T_s \) is the symbol duration. It is also interesting to observe that the effective Doppler shift \( B_k T \) increases with decreasing data rate since \( T \) increases with decreasing data rate.

2.3 Fusion Center Processing

Since the differential encoding operation in (1) introduces memory, symbol–by–symbol information fusion is not optimum. Instead, results from point–to–point communication systems suggest that the received signals should be processed on a block–by–block basis [13–15]. If blocks of received signals are properly processed, performance improves as the block size \( N \geq 2 \) increases and approaches the performance of coherent detection for \( N \to \infty \) [13, 19]. Here, we adopt the same philosophy for information fusion and process blocks of \( N \) received signals \( y_k \triangleq [y_k[n - N + 1] y_k[n - N + 2] \ldots y_k[n]]^T \) corresponding to blocks of \( N - 1 \) differential symbols \( a_k \triangleq [a_k[n - N + 2] a_k[n - N + 3] \ldots a_k[n]]^T, k \in K \). Based on these blocks of received signals any one of the \( N - 1 \) differential symbols in \( a_k \) can be detected and the corresponding MSD fusion rules will be discussed in the next section.

3 Multiple–Symbol Differential Decision Fusion

In this section, we will derive the optimal and several suboptimal fusion rules for the system model introduced in Section 2. For derivation of the considered fusion rules, we assume that the fusion center has knowledge of both the statistical properties of the channel and the performance indices \( P_k(a_k[n] = w_j|H_i), i \in M, j \in M, \) of the sensors \( k \in K \). However, as will become clear in the following, for some of the considered fusion rules one or both of these conditions can be relaxed. To simplify our notation, we will address in the following the \( \nu \)th element of vectors \( y_k \) and \( a_k \) by \( y_k(\nu), 1 \leq \nu \leq N, \) and \( a_k(\nu), 1 \leq \nu \leq N - 1, \) respectively. We denote the index of the differential symbol considered for detection by \( \nu_0, \nu_0 \in \{1, 2, \ldots, N - 1\} \). To simplify the notation further, we will drop the index of the differential symbol considered for detection wherever possible and denote it by \( a_k = a_k(\nu_0) \).
3.1 Optimal Fusion Rule

The optimal fusion rule based on the observations \( y \triangleq [y_1^T \ y_2^T \ \ldots \ y_K]^T \) can be formulated as

\[
H_i = \arg\max_{H_i, i \in M} \{ \log(P(H_i|y)) + \alpha_i \},
\]

(3)

where \( \alpha_i \) is a bias term which allows the prioritization of certain hypotheses. A bias may be useful for example in applications such as spectrum sensing for cognitive radio where a missed detection is less desirable than a false alarm. Since we assume that fading and noise are independent across different sensors, the conditional probability \( P(H_i|y) \) can be rewritten as

\[
P(H_i|y) = \frac{p(y|H_i)P(H_i)}{p(y)} = \prod_{k=1}^{K} \frac{p_k(y_k|H_i)P(H_i)}{p_k(y_k)}.
\]

(4)

Furthermore, the conditional pdf \( p_k(y_k|H_i) \) of sensor \( k \) can be expanded as

\[
p_k(y_k|H_i) = \sum_{j=0}^{M-1} p_k(y_k|a_k = w_j)P_k(a_k = w_j|H_i) = \frac{1}{M^{N-2}} \sum_{j=0}^{M-1} \sum_{a_k \in A_j} p_k(y_k|a_k)P_k(a_k = w_j|H_i),
\]

(5)

where \( A_j \) contains all \( M^{N-2} \) possible vectors \( a_k \) with \( a_k = w_j \) and the conditional pdf \( p_k(y_k|a_k) \) is given by [15]

\[
p_k(y_k|a_k) = \frac{1}{\pi^N \det(R_k)} \exp \left(-r_k^H R_k^{-1} r_k\right).
\]

(6)

Here, \( r_k \triangleq [r_k[n-N+1] \ r_k[n-N+2] \ \ldots \ r_k[n]]^T \) with \( r_k[n] \triangleq y_k[n]s_k^*[n] \) and \( R_k \triangleq \mathcal{E}\{r_k r_k^H\} = \bar{P}_k R_{hh,k} + \sigma_n^2 I_N \), where \( [R_{hh,k}]_{i,j} = \varphi_{hh,k}[i-j] \). Combining (3)–(6) and omitting all irrelevant terms yields the optimal MSD fusion rule

\[
H_i = \arg\max_{H_i, i \in M} \left\{ \sum_{k=1}^{K} \log \left( \sum_{j=0}^{M-1} \sum_{a_k \in A_j} p_k(y_k|a_k)P_k(a_k = w_j|H_i) \right) + \beta_i \right\}
\]

\[
= \arg\max_{H_i, i \in M} \left\{ \sum_{k=1}^{K} \log \left( \sum_{j=0}^{M-1} \sum_{a_k \in A_j} \exp \left( 2\Re \left\{ \sum_{\mu=1}^{N} \sum_{\nu=1}^{N} t_{\mu \nu}^k y_k(\mu)^* y_k^*(\nu) \right\} \sum_{\xi=\mu}^{\nu-1} a_k(\xi) \right) \right) P_k(a_k = w_j|H_i) + \beta_i \right\},
\]

(7)

where \( t_{\mu \nu}^k \triangleq -[R_k^{-1}]_{\mu,\nu} \) and \( \beta_i \triangleq \alpha_i + \log(P(H_i)) \) denotes the new bias term.

**Discussion:** Despite its optimal performance, the MSD fusion rule in (7) has several shortcomings: (a) The complexity of the fusion rule in (7) is exponential in both \( K \) and \( N \). (b)
Because of the large dynamic range of the exponential functions in (7), especially for high channel SNRs (i.e., $\frac{P_K \sigma_K^2}{\sigma_n^2} \gg 1$), the optimal fusion rule causes numerical problems, especially in fixed point implementations. (c) The optimal fusion rule requires statistical CSI (in form of $t^{k}_{\mu\nu}$) and knowledge of the sensor performance (in form of $P_k(a_k = w_j | H_i)$). We note that the coefficients $t^{k}_{\mu\nu}$ are related to the coefficients of a linear predictor for the process $r_k[n]$ and can be efficiently computed using adaptive algorithms [20, 21]. The above-listed drawbacks of the optimal fusion rule motivate the search for suboptimal fusion rules, which overcome these problems but still provide good performance.

### 3.2 Chair–Varshney (CV) Fusion Rule

The complexity of the optimal fusion rule can be tremendously reduced by assuming that the double sum on the right-hand side (RHS) of (5) is dominated by the maximum-likelihood (ML) vector $\hat{a}_k \triangleq [\hat{a}_k(1) \ldots \hat{a}_k(N-1)]^T$ which maximizes $p_k(y_k | a_k)$, i.e., $p_k(y_k | \hat{a}_k) \gg p_k(y_k | a_k)$, $a_k \neq \hat{a}_k$, $k \in \mathcal{K}$. This is a valid assumption for high channel SNR. In this case, the optimal fusion rule can be simplified to

$$H_i = \arg\max_{H_i, i \in M} \left\{ \sum_{k=1}^{K} \log(p_k(y_k | \hat{a}_k)P_k(\hat{a}_k | H_i)) + \beta_i \right\},$$

(8)

where $\hat{a}_k = \hat{a}_k(\nu_0)$ denotes the element of $\hat{a}_k$ which is considered for detection. We note that the ML vectors $\hat{a}_k$, $k \in \mathcal{K}$, can be efficiently obtained from $y_k$ by applying the multiple-symbol differential sphere decoding (MSDSD) algorithm in [22, Fig. 1]. For binary hypothesis testing ($M = 2$), (8) can be expressed as a likelihood ratio

$$\Lambda_{cv} = \sum_{\hat{a}_k=1}^{K} \frac{\log P_k(\hat{a}_k | H_1)}{P_k(\hat{a}_k | H_0)} + \sum_{\hat{a}_k=-1}^{K} \frac{\log P_k(\hat{a}_k | H_0)}{P_k(\hat{a}_k | H_0)},$$

(9)

and we decide in favor of $H_1$ if $\Lambda_{cv}$ exceeds threshold $\gamma_0 \triangleq \beta_0 - \beta_1$ and for $H_0$ otherwise. Thus, (8) and (9) can be regarded as the MSD version of the familiar CV fusion rule [4].

**Discussion:** The complexity of the suboptimal fusion rules in (8) and (9) grows only linearly in the number of sensors $K$. Furthermore, for sufficiently high channel SNR the average complexity of MSDSD is polynomial in $N$ [22], and thus, the complexity of the proposed fusion rule is also polynomial in $N$. Similar to the optimal fusion rule, knowledge of the sensor performance and, for $N > 2$, also statistical CSI are required for the CV fusion rule. For $N = 2$, based on (8) it can be shown that statistical CSI is not required if the channels are i.i.d. (i.e., $R_k = R, \forall k$).
3.3 Fusion Rule for Ideal Local Sensors (ILS)

For derivation of the CV fusion rule it was implicitly assumed that the uncertainty about the hypothesis at the fusion center originates only from the local sensor decisions, whereas the channel between the sensors and the fusion center was assumed ideal. The other extreme case is when we assume that the local sensor decisions are ideal, i.e., \( P_k(a_k = w_i | H_i) = 1 \) and \( P_k(a_k = w_j | H_i) = 0 \) for \( j \neq i \), and the uncertainty at the fusion center is due to the noisy transmission channel only. In this case, \( a_k = a, \forall k \in K \), is valid and the optimal ML block decision rule for \( a \) is given by

\[
\hat{a} = \arg\max_a \left\{ \sum_{k=1}^{K} \log(p_k(y_k|a)) + \beta_i \right\},
\]

where the bias \( \beta_i \) is determined by the trial symbol \( a = a(\nu_0) = w_i, i \in M \), and the hypothesis estimate \( H_\hat{i} \) can be directly obtained from the relevant element \( \hat{a} = \hat{a}(\nu_0) = w_\hat{i} \) of \( \hat{a} \). For the binary case, it is convenient to express (10) in terms of a likelihood ratio

\[
\Lambda_{ILS} = \sum_{k=1}^{K} \log \left( \frac{p_k(y_k|\hat{a}^1)}{p_k(y_k|\hat{a}^0)} \right),
\]

where \( \hat{a}^j \) is that vector \( a \in A_j \) which maximizes \( \sum_{k=1}^{K} \log(p_k(y_k|a)) \). In particular, \( H_\hat{i} = H_1 \) is chosen if \( \Lambda_{ILS} > \gamma_0 = \beta_0 - \beta_1 \), and \( H_\hat{i} = H_0 \) otherwise. The computational complexity of the fusion rules in (10) and (11) is only linear in \( K \) but still exponential in \( N \) if a brute force search over all possible \( a \) is conducted. Similar to the CV fusion rule, the application of sphere decoding is the key to reducing complexity further. For this purpose, we rewrite (10) as

\[
\hat{s} = \arg\min_{s, s(N)=1} \left\{ \sum_{k=1}^{K} s^H U_k^H U_k s - \beta_i \right\},
\]

where \( U_k \triangleq (L_k^H \text{diag}(y_k))^* \) is an upper triangular matrix and \( L_k \) is a lower triangular matrix obtained from the Cholesky factorization of \( R_k^{-1} \triangleq L_k L_k^H \). \( s \triangleq [s(1) s(2) \ldots s(N)]^T \) contains the absolute phase symbols from which the elements of \( a \) are obtained as \( a(j) = s(j+1)s^*(j) \). Because of the phase ambiguity inherent to (12), we can set \( s(N) = 1 \) without loss of generality. The sum over \( k \) and the bias term \( \beta_i \) in (12) make a direct application of the MSDSD algorithm in [22] impossible. However, as will be explained in the following, a modified version of MSDSD can be used to solve (12) efficiently provided that \( \beta_i \leq 0, i \in M \). The latter condition can always be fulfilled by properly choosing \( \alpha_i, i \in M \).
The modified MSDSD only examines candidate vectors that meet
\[
\sum_{k=1}^{K} s^H U_k^H U_k s - \beta_i \leq R^2,
\] (13)
where \( R \) is a pre-defined “radius”. Assuming we have found (preliminary) decisions \( \hat{s}(l) \) for the last \( N - l \) components \( s(l) \), \( \nu + 1 \leq l \leq N \), we can define an equivalent squared length
\[
d_{\nu+1}^2 = \sum_{l=\nu+1}^{N} \sum_{k=1}^{K} \left| \sum_{\mu=\nu}^{N} u_{k\mu}^l \hat{s}(\mu) \right|^2 - \beta_{\nu} \delta[\nu_0 - \nu - 1],
\] (14)
where \( u_{k\mu}^l \triangleq [U_k]_{\nu,\mu} \) and \( \beta_{\nu} \) is obtained from \( \hat{s}(\nu_0 + 1)\hat{s}^*(\nu_0) = \hat{a} = w_i \). Comparing (13) and (14), possible values for \( s(\nu) \) have to satisfy
\[
d_{\nu}^2 = \sum_{k=1}^{K} \left| u_{\nu\nu}^k s(\nu) + \sum_{\mu=\nu+1}^{N} u_{k\mu}^\nu \hat{s}(\mu) \right|^2 - \beta_i \delta[\nu_0 - \nu] + d_{\nu+1}^2 \leq R^2,
\] (15)
where \( \beta_i \) is determined by \( \hat{s}(\nu_0 + 1)\hat{s}^*(\nu_0) = a = w_i \). Once a valid vector \( \hat{s} \) is found, i.e., \( \nu = 1 \) is reached, the radius \( R \) is dynamically updated by \( R := d_1 \), and the search is repeated starting with \( \nu = 2 \) and the new radius \( R \). If condition (15) cannot be met for some index \( \nu \), \( \nu \) is incremented and another value of \( s(\nu) \) is tested. Based on (13)–(15) the modified MSDSD algorithm given by Algorithms 1 and 2 can be obtained. The search strategy for the proposed modified MSDSD algorithm is the Schnoor–Euchner (SE) strategy. Consequently, with an initial radius of \( R \to \infty \), the algorithm always finds a solution to (10).

**Discussion:** The complexity of the ILS fusion rule is linear in \( K \) and for high channel SNR polynomial in \( N \). While statistical CSI is still required for \( N > 2 \), the local sensor performance \( P_k(a_k = w_j|H_i) \), \( i \in \mathcal{M}, j \in \mathcal{M} \), does not have to be known at the fusion center. Of course, the price to be paid for this advantage is a loss in performance compared to the optimal fusion rule. For \( N = 2 \), it can be shown based on (10) that statistical CSI is not required if the channels are i.i.d.

### 3.4 Max–Log Fusion Rule

For high channel SNR (i.e., \( P_k\sigma_k^2/\sigma_n^2 \gg 1, \forall k \in K \)) one of the exponentials in (7) will be dominant and the max–log approximation, which is well known from the Turbo–coding literature
[23], can be applied

\[
H_i = \arg\max_{H_i, i \in M} \left\{ \sum_{k=1}^{K} \max_{j \in M \setminus i} \left\{ \log(p_k(y_k|a_k)) + \log(P_k(a_k = w_j|H_i)) \right\} + \beta_i \right\}
\]

\[
= \arg\max_{H_i, i \in M} \left\{ \sum_{k=1}^{K} \max_{j \in M \setminus i} \left\{ \sum_{j=1}^{N} \sum_{\nu=1}^{N} \log(y_k(\mu) y_k(\nu) \prod_{\xi=\nu}^{\mu-1} a_k(\xi)) \right\} + \log(P_k(a_k = w_j|H_i)) \right\} + \beta_i \right\}
\]

(16)

The max–log fusion rule in (16) is computationally more efficient and numerically more stable than the optimal fusion rule in (7) since exponential functions are avoided in (16). However, if (16) is implemented in a straightforward fashion, its computational complexity is still exponential in \(N\), since for every test hypothesis \(H_i, i \in M\), the maximum of \(\log(p_k(y_k|a_k))\) has to be found over all \(a_k \in A_j, j \in M\). However, the max–log fusion rule can be rewritten as

\[
H_i = \arg\max_{H_i, i \in M} \left\{ \sum_{k=1}^{K} \left\{ \max_{j \in M} \left\{ \log(p_k(y_k|\hat{a}_k^j)) + \log(P_k(\hat{a}_k = w_j|H_i)) \right\} \right\} + \beta_i \right\}
\]

(17)

where \(\hat{a}_k^j\) is that \(a_k \in A_j\) which maximizes \(p_k(y_k|a_k)\). \(\hat{a}_k^j\) can be efficiently computed using sphere decoding. In particular, the MSDSD algorithm in [22, Fig. 1] can be slightly modified to account for the fact that the search is constrained to those \(a_k\) with \(a(\nu_0) = a = w_j\). For the binary case, \(M = 2\), (17) is equivalent to choosing \(H_1\) if likelihood ratio \(\Lambda_{m-log}\) exceeds threshold \(\gamma_0 = \beta_0 - \beta_1\), and \(H_0\) otherwise, where \(\Lambda_{m-log}\) is defined as

\[
\Lambda_{m-log} = \frac{\sum_{k=1}^{K} \max_{j \in M} \left\{ \log\left( \frac{p_k(y_k|\hat{a}_k^j) P_k(\hat{a}_k = w_j|H_1)}{p_k(y_k|\hat{a}_k^j) P_k(\hat{a}_k = w_j|H_0)} \right) \right\}}{\sum_{k=1}^{K} \max_{j \in M} \left\{ \log\left( \frac{p_k(y_k|\hat{a}_k^j) P_k(\hat{a}_k = w_j|H_1)}{p_k(y_k|\hat{a}_k^j) P_k(\hat{a}_k = w_j|H_0)} \right) \right\}}
\]

(18)

Discussion: The complexity of the max–log fusion rule is linear in \(K\) and for high channel SNR polynomial in \(N\). The implementation of (17) and (18) requires knowledge of the sensor performance \(P_k(a_k = w_j|H_i), i \in M, j \in M\). Furthermore, the complexity of the max–log fusion rule is higher than that of the CV and ILS fusion rules discussed in Sections 3.2 and 3.3, respectively. In particular, for the max–log fusion rule, MSDSD has to be performed \(MK\) times, whereas the CV and ILS fusion rules require only \(K\) and one MSDSD operations, respectively. In addition, in contrast to the CV and ILS fusion rules, the max–log fusion rule requires statistical CSI even for \(N = 2\) and i.i.d. channels. On the other hand, the max–log fusion rule achieves a superior performance compared to the CV and ILS fusion rules, cf. Section 5.
4 Analysis of Suboptimal Fusion Rules

An analysis of the optimal fusion rule does not seem to be possible. Therefore, we concentrate in this section on the CV, ILS, and max–log fusion rules and on general performance bounds valid for any fusion rule. To make the analysis tractable, we assume $M = 2$, i.e., $M = \{0, 1\}$, i.i.d. channels, i.e., $\sigma_k^2 = \sigma^2$, $B_k = B$, $R_k = R$, $t^k_{\mu \nu} = t_{\mu \nu}$, and $p_k(y_k|a_k) = p(y_k|a_k)$, $\forall k \in \mathcal{K}$, and identical sensors with probability of false alarm $P_f \triangleq P(a_k = -1|H_0) = P_k(a_k = -1|H_0)$ and probability of detection $P_d \triangleq P(a_k = -1|H_1) = P_k(a_k = -1|H_1)$, $\forall k \in \mathcal{K}$.

4.1 Performance Bounds

Before considering specific fusion rules, we provide two performance upper bounds valid for any fusion rule including the optimal one.

1) Bound I: For the first bound, we assume that all sensors make correct decisions and decision errors at the fusion center are due to transmission errors only, i.e., $a_k = a$, $\forall k \in \mathcal{K}$, and zero bias, i.e., $\beta_0 = \beta_1 = 0$. In this case, the sensor network is equivalent to a point–to–point transmission with $K$–fold receive diversity and conventional MSDD [14, 15] is the optimal fusion rule. Thus, the probabilities of false alarm and detection are given by

$$P_{f_0} = \text{BER}_{\nu_0} \quad \text{and} \quad P_{d_0} = 1 - \text{BER}_{\nu_0},$$

where $\text{BER}_{\nu_0}$ denotes the probability that $a = a(\nu_0)$ was transmitted and $\hat{a} \neq a$, $\hat{a} \in \{\pm 1\}$, $1 \leq \nu_0 \leq N - 1$, was detected, i.e., $\text{BER}_{\nu_0}$ is the bit error rate (BER) for 2–DPSK symbol $a(\nu_0)$ for point–to–point transmission and MSDD at the receiver. $\text{BER}_{\nu_0}$ can be lower bounded as [19]

$$\text{BER}_{\nu_0} \geq \frac{\text{PEP}_{\nu_0} + \text{PEP}_{\nu_0 + 1}}{2}, \quad 1 \leq \nu_0 \leq N - 1,$$

where the pairwise error probability (PEP), $\text{PEP}_{\nu_0}$, is the probability that vector $s$ was transmitted and $\hat{s}(\nu_0) \triangleq [s(1) \ldots s(\nu_0 - 1) \hat{s}(\nu_0) s(\nu_0 + 1) \ldots s(N)]^T$, $\hat{s}(\nu_0) \neq s(\nu_0)$, was detected. The averaging over two error events in (20) is necessary, since, because of the differential encoding, $\hat{a}(\nu_0) \neq a(\nu_0)$ may be caused by either $\hat{s}(\nu_0)$ or $\hat{s}(\nu_0 + 1)$. Note that in order to get performance upper bounds, we only count error events causing a single erroneous symbol, $\hat{s}(\nu) \neq s(\nu)$, in (20).

Taking into account the $K$–fold diversity, we obtain from [19, Eq. (12a)] for the PEP for the
problem at hand

\[
\text{PEP}_\nu = \left[ \frac{1}{2} \left( 1 - \frac{1}{\sqrt{1 + 1/\rho_\nu}} \right) \right]^K \sum_{k=0}^{K-1} \frac{1}{2^k} \binom{K + k - 1}{K - 1} \left( 1 + \frac{1}{\sqrt{1 + 2/2 + \rho_\nu}} \right)^k, \tag{21}
\]

where \( \rho_\nu \triangleq - t_{\nu\nu}(\bar{P}_K + \sigma_n^2) - 1 \), \( 1 \leq \nu \leq N \). Eqs. (19)–(21) constitute a performance upper bound for any fusion rule with noisy sensors. This bound becomes tight for optimal decision fusion if transmission errors dominate the overall performance, which is the case for example at low channel SNRs and for highly reliable local sensors.

2) Bound II: For the second bound, we assume a noise–free transmission channel, i.e., the decision errors at the fusion center are caused by local decision errors at the sensors only. In this case, the CV fusion rule is optimum and the corresponding probabilities of false alarm and detection are given by [4]

\[
P_{f0} = \sum_{i=K_{\gamma_0}}^{K} \binom{K}{i} P_f^i (1 - P_f)^{K-i} \quad \text{and} \quad P_{d0} = \sum_{i=K_{\gamma_0}}^{K} \binom{K}{i} P_d^i (1 - P_d)^{K-i}, \tag{22}
\]

where \( K_{\gamma_0}, 0 \leq K_{\gamma_0} \leq K \), is a parameter that can be used to achieve a desired trade–off between \( P_{f0} \) and \( P_{d0} \). For realistic, noisy transmission channels, (22) constitutes a performance upper bound which becomes tight for high channel SNRs.

4.2 CV Fusion Rule

Considering (9) the probabilities of false alarm and detection at the fusion center can be expressed as

\[
P_{f0} = \sum_{i=K_{\gamma_0}}^{K} \binom{K}{i} P_0^i (1 - P_0)^{K-i} \quad \text{and} \quad P_{d0} = \sum_{i=K_{\gamma_0}}^{K} \binom{K}{i} P_1^i (1 - P_1)^{K-i}, \tag{23}
\]

where \( P_0 = P(H_i = H_1|H_0) \) and \( P_1 = P(H_i = H_1|H_1) \). \( P_0 \) and \( P_1 \) can be expanded as

\[
P_i = \begin{cases} 
P(H_i = H_1|a_k = -1)P(a_k = -1|H_i) + P(H_i = H_1|a_k = 1)P(a_k = 1|H_i) \\
(1 - \text{BER}_\nu)_{x_i} + \text{BER}_\nu(1 - P_{x_i}), \quad i \in \mathcal{M},
\end{cases}
\]

where \( x_0 = f \) and \( x_1 = d \), and \( \text{BER}_\nu \) is the BER of 2–DPSK for a point–to–point link without diversity and MSDD at the receiver. This BER can be approximated as [19, Table I, Eq. (12)]

\[
\text{BER}_\nu \approx \text{PEP}_\nu + \text{PEP}_{\nu+1}, \quad 1 \leq \nu_0 \leq N - 1,
\]

\[
\text{PEP}_\nu = \frac{1}{2} \left( 1 - \frac{1}{\sqrt{1 + 1/\rho_\nu}} \right). \tag{25}
\]

where \( \rho_\nu \triangleq - t_{\nu\nu}(\bar{P}_K + \sigma_n^2) - 1 \), \( 1 \leq \nu \leq N \).
For the special case of \( N = 2 \), (25) is exact, while it is an accurate approximation for \( N > 2 \) and sufficiently high channel SNRs. Using (23)–(26) the probabilities of false alarm and detection for the CV decision rule can be computed approximately (exactly) for \( N > 2 \) \((N = 2)\).

### 4.3 ILS Fusion Rule

The ILS decision in (10) is influenced by the local sensor decisions \( a_k(\nu), 1 \leq k \leq K, 1 \leq \nu \leq N - 1 \), which makes an exact analysis for \( N > 2 \) intractable and renders both approximations and bounds loose. Therefore, in this subsection, we concentrate on the case \( N = 2 \). The probabilities of false alarm and detection can be expressed as

\[
P_{f_0} = \sum_{\bar{a}} P(\hat{a} = -1|\bar{a}) P(\bar{a}|H_0) \quad \text{and} \quad P_{d_0} = \sum_{\bar{a}} P(\hat{a} = -1|\bar{a}) P(\bar{a}|H_1),
\]

where \( P(\hat{a} = -1|\bar{a}) \) denotes the probability that the fusion center detects \( \hat{a} = -1 \) given the local sensor decisions \( \bar{a} \triangleq [a_1 \ a_2 \ldots \ a_K]^T \). Furthermore, the conditional sensor decision probabilities in (27) are given by \( P(\bar{a}|H_0) = P_f^{K-k_0}(1 - P_f)^{k_0} \) and \( P(\bar{a}|H_1) = P_d^{K-k_0}(1 - P_d)^{k_0} \), where \( k_0 \) denotes the number of elements of \( \bar{a} \) that are equal to 1. Based on (11) \( P(\hat{a} = -1|\bar{a}) \) can be expressed as \( P(\hat{a} = -1|\bar{a}) = \Pr\{-\Lambda_{\text{ILS}} < -\gamma_0\} \), which leads to

\[
P(\hat{a} = -1|\bar{a}) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_{\text{ILS}}(s|\bar{a}) e^{-\gamma_0 s} \frac{ds}{s},
\]

where \( \Phi_{\text{ILS}}(s|\bar{a}) \) denotes the Laplace transform of the pdf of \(-\Lambda_{\text{ILS}}\) given \( \bar{a} \) and \( \hat{a}_k = \hat{a} \), and \( c \) is a small positive constant that lies in the region of convergence of the integral. Closer examination of (11) reveals that \( \Lambda_{\text{ILS}} \) is a quadratic form of Gaussian random variables. Consequently, after some manipulations, we obtain

\[
\Phi_{\text{ILS}}(s|\bar{a}) = 1/\det(I_{2K} + s\tilde{R}M(\bar{a})),
\]

where \( \tilde{R} \triangleq I_K \otimes R \) and \( M(\bar{a}) \triangleq \text{diag}\{M_1(a_1), \ldots, M_K(a_K)\} \) with \( M_k(a_k) \triangleq 2a_k \begin{bmatrix} 0 & t_{12} \\ t_{21} & 0 \end{bmatrix} \). We note that the integral in (28) can be numerically evaluated efficiently using e.g. Gauss–Chebyshev quadrature rules [24]. Thus, the exact probabilities of false alarm and detection for the ILS fusion rule with \( N = 2 \) can be computed using (27)–(29).
4.4 Max–Log Fusion Rule

For the max–log fusion rule, the probabilities of false alarm and detection can be expressed as

\[ P_{f_0} = \Pr\{-\Lambda_{m-\log} < -\gamma_0|H_0\} \quad \text{and} \quad P_{d_0} = \Pr\{-\Lambda_{m-\log} < -\gamma_0|H_1\}, \quad (30) \]

cf. (18). Denoting the Laplace transform of the pdf of the negative log–likelihood ratio \(-\Lambda_{m-\log}\) by \(\Phi_{m-\log}(s|H_i)\), (30) can be rewritten as

\[ P_{y_i} = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_{m-\log}(s|H_i)e^{-\gamma_0 s} \frac{ds}{s}, \quad i \in \mathcal{M}, \quad (31) \]

where \(y_0 = f_0\) and \(y_1 = d_0\). Since \(P_{f_0}\) and \(P_{d_0}\) can be obtained by numerical integration from (31) if \(\Phi_{m-\log}(s|H_i)\) is known, the remainder of this section will be devoted to the calculation of this Laplace transform. As the fading gains and noise samples in the different diversity branches are i.i.d., respectively, \(\Phi_{m-\log}(s|H_i)\) can be expressed as

\[ \Phi_{m-\log}(s|H_i) = (\Phi_z(s|H_i))^K, \quad i \in \mathcal{M}, \quad (32) \]

where \(\Phi_z(s|H_i)\) denotes the Laplace transform of the pdf of

\[ z_k \triangleq \max_{j \in \mathcal{M}} \min_{i \in \mathcal{M}} \left\{ \log \left( \frac{p(y_k|\hat{a}^0_k)P(\hat{a}_k = w_j|H_0)}{p(y_k|\hat{a}^1_k)P(\hat{a}_k = w_i|H_1)} \right) \right\}. \quad (33) \]

\(\Phi_z(s|H_i)\) can be rewritten as

\[ \Phi_z(s|H_i) = (1 - P_{\chi})\Phi_z(s|\hat{a} = -1, a = 1) + P_{\chi}\Phi_z(s|\hat{a} = -1, a = -1), \quad i \in \mathcal{M}, \quad (34) \]

where \(\Phi_z(s|\hat{a}, a)\) denotes the Laplace transform of the pdf of \(z_k\) given \(a_k = a\) and \(\hat{a}\). For calculation of \(\Phi_z(s|\hat{a}, a)\) it is useful to note that for \(M = 2\), (33) can be rewritten as

\[ z_k = \log \left( \frac{\max\{p(y_k|\hat{a}^0_k)(1 - P_f), p(y_k|\hat{a}^1_k)P_f\}}{\max\{p(y_k|\hat{a}^0_k)(1 - P_d), p(y_k|\hat{a}^1_k)P_d\}} \right). \quad (35) \]

Using the definition \(y \triangleq \log(p(y_k|\hat{a}^0_k)/p(y_k|\hat{a}^1_k))\) at \(a = 1\) and assuming \(P_f < 0.5\) and \(P_d > 0.5\), we can show that (35) can be rewritten as

\[ z_k = \begin{cases} 
\beta_1, & ay < b_1 \\
ay + \beta_2, & b_1 \leq ay \leq b_2 \\
\beta_3, & ay > b_2 
\end{cases} \quad (36) \]
where \( \beta_1 \triangleq \log(P_f/P_d), \beta_2 \triangleq \log((1 - P_f)/P_d), \beta_3 \triangleq \log((1 - P_f)/(1 - P_d)), \) and \( b_1 \triangleq \log(P_f/(1 - P_f)), \) and \( b_2 \triangleq \log(P_d/(1 - P_d)). \) To arrive at (36) for \( a = -1, \) we have exploited \( \log(P(y_k|\hat{\alpha}_k^0)/P(y_k|\hat{\alpha}_k^1))|_{a=1} = -\log(P(y_k|\alpha_k^0)/P(y_k|\alpha_k^1))|_{a=-1}. \) For convenience (36) is illustrated in Fig. 2 for the case \( a = 1. \) Fig. 2 reveals that the max–log fusion rule soft–limits the log–likelihood ratios \( y \) of the individual sensors at the fusion center by taking into account the \textit{a priori} values \( P_f \) and \( P_d. \) Denoting the pdf of \( y \) by \( p_y(y) \) and exploiting (36), we can express \( \Phi_2(s|\hat{a} = -1, a) \) as

\[
\Phi_2(s|\hat{a} = -1, a) = \int_{-\infty}^{b_1} e^{-s\beta_1}p_y(ay)\,dy + \int_{b_1}^{b_2} e^{-s(y+\beta_2)}p_y(ay)\,dy + \int_{b_2}^{\infty} e^{-s\beta_3}p_y(ay)\,dy. \tag{37}
\]

For calculation of \( p_y(y), \) we distinguish in the following the cases \( N = 2 \) and \( N > 2. \)

For \( N = 2, \) assuming \( s = [s(1) 1]^T \) the only possible error event leading to \( \hat{a} = -1 \) is \( s = [-s(1) 1]^T \) and \( y \) can be expressed as \( y = -r_k^H R_k^{-1} r_k|_s + r_k^H R_k^{-1} r_k|_{\hat{s}}. \) In other words, \( y \) is simply the decision variable for conventional differential detection. Thus, the Laplace transform of \( p_y(y) \) is given by [19, Eq. (27)]

\[
\Phi_y(s) = \frac{-v_1v_2}{(s + v_1)(s - v_2)} \tag{38}
\]

where \( v_{1/2} = (\sqrt{1 + 1/\rho_0} \mp 1)/2 \) with \( \rho_0 = 1. \) From (38) we can calculate \( p_y(y) \) as

\[
p_y(y) = c_v \left(e^{-v_1y}u(y) + e^{v_2y}u(-y)\right), \tag{39}
\]

where \( c_v \triangleq v_1v_2/(v_1 + v_2). \) Combining (37) and (39) we obtain

\[
\Phi_2(s|\hat{a} = -1, a) = c_v \left(\frac{e^{-s\beta_1 + v_1b_1}}{v_j} + e^{-s\beta_2} \left(\frac{1 - e^{-(s+v_1)b_1}}{s + v_j} + \frac{1 - e^{-(s+v_2)b_2}}{s + v_i}\right) + \frac{e^{-s\beta_3 - v_1b_2}}{v_i}\right), \tag{40}
\]

where \((i, j) = (1, 2)\) and \((i, j) = (2, 1)\) for \( a = 1 \) and \( a = -1, \) respectively.

For \( N > 2 \) the problem is more difficult, since there are more than one possible error events that lead to \( \hat{a}(\nu_0) = \hat{a} = -1. \) The most likely error events are \( \hat{s}_{\nu_0} \) and \( \hat{s}_{\nu_0 + 1} \) which differ from \( s \) only in positions \( \nu_0 \) and \( \nu_0 + 1, \) respectively. The corresponding likelihood ratios are denoted by \( y_1 \triangleq \log(P(y_k|\hat{\alpha}_k^0)/P(y_k|\hat{\alpha}_k^1))|_{\hat{s}_{\nu_0}} \) and \( y_2 \triangleq \log(P(y_k|\hat{\alpha}_k^0)/P(y_k|\hat{\alpha}_k^1))|_{\hat{s}_{\nu_0 + 1}}. \) To make the problem tractable, we assume that \( \hat{s}_{\nu_0} \) and \( \hat{s}_{\nu_0 + 1} \) are the only relevant error events, i.e., we neglect all other error events, which is a valid approximation for high channel SNRs. In this case, \( y \) is given by \( y = \min\{y_1, y_2\}. \) In order to get closed–form results, we make the following
two additional approximations: (a) $y_1$ and $y_2$ are independent and (b) $y_1$ and $y_2$ are identically distributed. Both assumptions are justified for high channel SNRs. By exploiting results from order statistics [25] and [19], we obtain for the pdf of $y$

$$p_y(y) = 2p_{y_1}(y)(1 - P_{y_1}(y)),$$  \hspace{1cm} (41)

where $p_{y_1}(y) = c_v(e^{-v_1y}u(y) + e^{-v_2y}u(-y))$, cf. (39), $P_{y_1}(y) = \int_{-\infty}^{y} p_{y_1}(x) \, dx$, and $v_{1/2} = (\sqrt{1 + 1/\rho_{\nu_0}} \mp 1)/2$ with $1 \leq \nu_0 \leq N - 1$. Combining (37) and (41) leads to a closed-form expression for $\Phi_z(s|\hat{a} = -1, a)$ similar to (40). We do not provide this expression here because of space limitation.

Combining (31), (32), (34), (40), and the corresponding expression for $\Phi_z(s|\hat{a} = -1, a)$ for $N > 2$, the probabilities of false alarm and detection can be exactly (approximately) computed for $N = 2$ ($N > 2$). We note that a direct numerical integration of (31) is problematic since the inverse Laplace transform of $\Phi_m - \log(s|H_1)$ has discontinuities (reflected e.g. by the first and last term in the sum on the RHS of (40)). However, the terms corresponding to the discontinuities can be easily inverted in closed form, and the remaining terms without discontinuities can then be inverted numerically using the methods in [24].

5 Numerical and Simulation Results

In this section, we present numerical and simulation results for the proposed optimal and suboptimal fusion rules. Thereby, we will consider binary and non-binary hypothesis testing separately. For all results shown in this section, the middle symbol of the observation window is used for detection, i.e., $\nu_0 = N/2$, since this leads to the best performance.

5.1 Binary Hypothesis Testing

In this subsection, in order to confirm our simulation results with the analytical results from Section 4, we assume $M = 2$, i.i.d. Rayleigh fading channels, identical sensors, and $P(H_0) = P(H_1) = 1/2$. All curves labeled with "Theory" (for $N = 2$) and "Approximation" (for $N = 6$) were generated using the analytical methods discussed in Section 4, while the remaining curves were obtained by computer simulation.

In Figs. 3 and 4, we consider the error probability $P_e \triangleq P_{f_0}P(H_0) + (1 - P_{d_0})P(H_1)$ of the considered suboptimal MSD fusion rules vs. $E_b/N_0$ for $BT = 0.1$ and $BT = 0$, respectively. Here,
$E_b$ is the total received average energy per bit (from all sensors), and $N_0$ denotes the one-sided power spectral density of the underlying continuous-time noise process. The decision threshold $\gamma_0 = 0$ was used for all fusion rules and $K = 8$, $P_d = 0.8$, and $P_f = 0.01$. In addition to the suboptimal MSD fusion rules, Figs. 3 and 4 also contain the two performance upper bounds introduced in Section 4.1. Furthermore, in Fig. 3 we have also included the performance of the optimal fusion rule for $N = 2$ (the optimal fusion rule is computationally not feasible for $N = 6$) and the error probability of the coherent max–log fusion rule for DPSK, whereas in Fig. 4, we display the performance of the coherent versions of all three suboptimal MSD fusion rules. We note that these coherent fusion rules require perfect knowledge of the fading channel gains. Figs. 3 and 4 show that while the ILS fusion rule has the best performance for very low $E_b/N_0$, where transmission errors dominate the overall performance, the CV and the max–log fusion rule yield a superior performance for medium–to–high $E_b/N_0$. A comparison of Figs. 3 and 4 reveals that increasing the observation window size from $N = 2$ to $N = 6$ is more beneficial for fast fading ($BT = 0.1$) than for static fading ($BT = 0$). In the latter case, the performance gap between coherent detection and the respective MSD fusion rules is relatively small even for $N = 2$. In contrast, for $BT = 0.1$ and $N = 2$ the performance of both the CV and the max–log fusion rules is limited by the high error floor caused by the fast fading. This error floor is also not overcome by the optimal fusion rule which yields a negligible performance gain compared to the computationally simpler max–log fusion rule for $N = 2$. For $N = 6$ this error floor is mitigated and both the CV and the max–log fusion rules approach Bound II for high $E_b/N_0$, i.e., performance is limited by local sensor decision errors in this case and not by transmission errors. For the ILS fusion rule increasing the observation window to $N > 2$ is not beneficial and even leads to a loss in performance for high $E_b/N_0$ for $BT = 0$. This somewhat surprising behavior is caused by the local decision errors at the sensors, which were ignored for derivation of the ILS fusion rule. For $N = 2$, theoretical and simulation results in Figs. 3 and 4 match perfectly confirming the analysis in Section 4. As expected from the discussions in Section 4, for $N = 6$, there is a good agreement between theoretical and simulation results for the CV and the max–log fusion rules at high $E_b/N_0$ ratios, cf. Fig. 3 (for clarity of presentation the analytical curves for $N = 6$ were omitted in Fig. 4). At low $E_b/N_0$ ratios, the analytical results overestimate the actual $P_e$ since the assumptions leading to the analytical result for $N > 2$ are less justified.

In Fig. 5, we show $P_{d0}$ as a function of $E_b/N_0$ for a fixed probability of false alarm of $P_{f0} = 0.001$, which is achieved by adjusting decision threshold $\gamma_0$ accordingly. Furthermore, $K = 8$,
\( P_d = 0.7, P_f = 0.05, \) and \( BT = 0.1 \). In Fig. 5, the max–log fusion rule yields a superior performance compared to the other suboptimal MSD fusion rules but the CV and ILS fusion rules approach the max–log performance for high and low \( E_b/N_0 \), respectively. For \( N = 6 \), both the max–log and the CV fusion rules approach Bound II for high enough \( E_b/N_0 \), whereas for \( N = 2 \), these fusion rules as well as the optimal fusion rule are limited by transmission errors caused by fast fading. In contrast, the ILS fusion rule achieves a better performance for \( N = 2 \) than for \( N = 6 \). Fig. 6 shows again a good agreement between analytical and simulation results.

In Fig. 6, we show the receiver operating curve (ROC) for the considered MSD fusion rules and the coherent max–log fusion rule for DPSK. \( K = 8, P_d = 0.7, P_f = 0.05, BT = 0.1, \) and \( E_b/N_0 = 20 \) dB. Fig. 6 shows the superiority of the max–log fusion rule especially if low probabilities of false alarm are desired. Increasing \( N \) from two to six yields significant gains for both the max–log and the CV fusion rules. In fact, the max–log fusion rule with \( N = 6 \) bridges half of the performance gap between the coherent max–log fusion rule and the MSD max–log fusion rule with \( N = 2 \). On the other hand, for \( N = 2 \) the optimal fusion rule performs only slightly better than the max–log fusion rule.

Figs. 7 and 8 show the impact of the number of sensors on \( P_e \) and \( P_{d_0} \), respectively, for \( P_d = 0.7, P_f = 0.05, BT = 0.1, \) and \( E_b/N_0 = 20 \) dB. For Fig. 7, we optimized the decision threshold \( \gamma_0 \) for minimization of \( P_e \) for each fusion rule and each considered \( K \). For Fig. 8, \( \gamma_0 \) was chosen such as to guarantee \( P_{f_0} = 0.001 \). Figs. 7 and 8 indicate that the max–log fusion rule benefits more from an increasing number of sensors than the ILS and the CV fusion rules. In particular, the CV fusion rule shows a saturation effect for large \( K \) in Fig. 7. This is due to the fact that since the total \( E_b/N_0 \) of all sensors is fixed, the channel SNR per sensor decreases as \( K \) increases. Therefore, the assumption of a perfect transmission channel, which was implicitly made for derivation of the CV fusion rule, becomes less justified as \( K \) increases leading to a loss in performance.

### 5.2 Multiple Hypothesis Testing

For the multiple hypothesis testing case, we assume that the local sensor observations are given by \( x_k[n] = u_k[n] + \tilde{n}_k[n], \) \( k \in \mathcal{K} \), where \( u_k[n] \in \{-M+1, -(M-3), \ldots, M-1\} \) and \( \tilde{n}_k[n] \) is real AWGN. Throughout this subsection, we assume identical sensors, \( P(H_i) = 1/M, i \in \mathcal{M}, \) and \( M = 4 \). The sensor performance indices \( P_k(a_k[n] = w_j|H_i), i \in \mathcal{M}, j \in \mathcal{M}, k \in \mathcal{K}, \) depend
on the sensor SNR, \( \text{SNR}_s \triangleq \mathbb{E}\{|u_k[n]|^2\}/\mathbb{E}\{|\tilde{n}_k[n]|^2\} \).

In Fig. 9, we show the probability of missed detection \( P_m \triangleq \sum_{i=0}^{M-1} \sum_{i \neq i} P(H_i|H_i)P(H_i) \) as a function of the sensor SNR, \( \text{SNR}_s \), for the proposed suboptimal MSD fusion rules and the corresponding coherent fusion rules. \( K = 8, BT = 0.1, E_b/N_0 = 30 \text{ dB}, \) and the channel SNR of sensors \( k \in \{1, 2, 3, 4\} \) was 3 dB higher than that of the remaining four sensors, i.e., the fading was i.n.d. For low sensor SNRs, the CV fusion rule achieves a similar performance as the max–log fusion rule since the overall performance is dominated by the unreliable sensors. However, the CV fusion rule is not able to fully exploit the increasing reliability of the sensors when the sensor SNR improves and is ultimately limited by an error floor caused by transmission errors which are not optimally taken into account in the CV fusion rule. For highly reliable sensors the CV fusion rule is even outperformed by the ILS fusion rule whose performance steadily improves with increasing sensor SNR since the assumption on which the ILS fusion rule is based, namely error–free sensors, becomes more and more justified at high sensor SNR. Nevertheless, the max–log fusion rule yields the best performance among all considered MSD fusion rules, and closely approaches the performance of the coherent max–log fusion rule with \( N = 6 \).

In Fig. 10, we compare the complexity of the considered MSD fusion rules for \( N = 6 \) as a function of \( E_b/N_0 \). \( K = 8, \) i.i.d. Rayleigh fading, and \( \text{SNR}_s = -3 \text{ dB} \) are valid. The complexity is measured in terms of the (average) number of real multiplications required per decision. The dashed lines in Fig. 10 denote the number of multiplications required by the respective sphere decoders to find the first vector \( \hat{s} \) and constitute lower bounds for the actual complexity. Note that the lower bounds for the CV and ILS fusion rules practically coincide for the considered example. Fig. 10 shows that the CV fusion rule closely approaches the corresponding lower bound. In contrast, for the ILS and max–log fusion rules there is always a considerable gap between the actual complexity and the lower bound even at high \( E_b/N_0 \). For the ILS fusion rule, this gap is due to erroneous sensor decisions as can be observed from the comparison with the (hypothetical) case of ideal local sensor decisions. For the max–log fusion rule the gap is due to the fact that the sphere decoder does not only have to find the ML vector as for the CV and ILS fusion rules but has to perform a constrained search over all \( \alpha_k \) with \( a(v_0) = w_j, j \in \mathcal{M}, \) cf. Section 3.4. Nevertheless, all three suboptimal fusion rules have a significantly lower complexity than the optimal fusion rule.
6 Conclusions

In this paper, we have considered the distributed multiple hypothesis testing problem for mobile wireless sensor networks where sensors employ DPSK to cope with time-variant fading. We have shown that since the differential modulation introduces memory, it is advantageous to consider fusion rules that base their decisions on an observation window of multiple symbol intervals. Specifically, we have derived the optimal MSD fusion rule, whose complexity is exponential in the number of sensors and the observation window size, and three suboptimal MSD fusion rules, whose complexity is linear in the number of sensors and, at high channel SNR, polynomial in the observation window size. For binary hypothesis testing, performance bounds for the optimal fusion rule have been derived, and for the suboptimal fusion rules, exact or approximate expressions for the probabilities of false alarm and detection have been provided. Our simulation and analytical results show that the CV and ILS fusion rules approach the performance of the optimal fusion rule for high and low channel SNRs, respectively. The proposed max-log fusion rule achieves a close-to-optimal performance over the entire SNR range but has a higher complexity than the CV and ILS fusion rules.

References


Figures:

Figure 1: System model for MSD decision fusion.
Algorithm 1 Pseudocode for MSDSD for ILS Fusion Rule

1: \[ s_N := 1 \text{ \hspace{10pt} \triangleright initial radius } R \text{, vector of bias terms } \beta \]
2: \[ \hat{d}_N := \sum_{k=1}^{K} u_{vN}^k \text{ \hspace{10pt} \triangleright fix last component of } s \]
3: \[ v := N - 1 \text{ \hspace{10pt} \triangleright initialize squared length } \]
4: \[ \text{for } k = 1 \text{ to } K \text{ do } q_v^k := \sum_{i=v+1}^{N} u_{i}^k s_i \text{ end for} \]
5: \[ [\text{step}_v, n_v] = \text{findBestILS}(q_v^1, \ldots, q_v^K, u_{v1}, \ldots, u_{vN}, M) \text{ \hspace{10pt} \triangleright find with first candidate for component at } v = N - 1 \]
6: \[ \text{search} := 1 \]
7: \[ \text{while search} = 1 \text{ do} \]
8: \[ \hat{d}_v^0 := \sum_{k=1}^{K} u_{vv}^k e^{\frac{2\pi}{M} \cdot \text{step}_v(n_v)} + q_v^k |^2 + d_{v+1}^2 \]
9: \[ \text{if } v = v_0 \text{ then} \]
10: \[ i := \text{mod}(M + (\text{step}_{v+1}(n_{v+1}) - \text{step}_v(n_v)), M) \]
11: \[ \hat{d}_v^0 := \hat{d}_v^0 - \beta_i \]
12: \[ \text{end if} \]
13: \[ \text{if } d_v < R \text{ and } n_v < M \text{ then} \]
14: \[ \text{search} := 0 \]
15: \[ \text{end if} \]
16: \[ \text{if } v \neq 1 \text{ then} \]
17: \[ v := v - 1 \]
18: \[ \text{for } k = 1 \text{ to } K \text{ do } q_v^k := \sum_{i=v+1}^{N} u_{i}^k s_i \text{ end for} \]
19: \[ [\text{step}_v, n_v] = \text{findBestILS}(q_v^1, \ldots, q_v^K, u_{v1}, \ldots, u_{vN}, M) \]
20: \[ \text{end if} \]
21: \[ \hat{s} := s \]
22: \[ R := d_v \]
23: \[ v := v + 1 \]
24: \[ \text{while } n_v = M \text{ and } \text{search} = 1 \text{ do} \]
25: \[ \text{if } v = N - 1 \text{ then } \text{search} := 0 \text{ else } v := v + 1 \text{ end if} \]
26: \[ \text{end while} \]
27: \[ n_v := n_v + 1 \]
28: \[ \text{end if} \]
29: \[ \text{else} \]
30: \[ \text{if } v = N - 1 \text{ then} \]
31: \[ \text{search} := 0 \]
32: \[ \text{end if} \]
33: \[ v := v + 1 \]
34: \[ \text{while } n_v = M \text{ and } \text{search} = 1 \text{ do} \]
35: \[ \text{if } v = N - 1 \text{ then } \text{search} := 0 \text{ else } v := v + 1 \text{ end if} \]
36: \[ \text{end while} \]
37: \[ \text{end if} \]
38: \[ n_v := n_v + 1 \]
39: \[ \text{end if} \]
40: \[ \text{end while} \]
41: \[ \text{end function} \]

Algorithm 2 Pseudocode for findBest used in MSDSD

1: \[ \text{function } [\text{step}_v, n_v] = \text{findBestILS}(q_v^1, \ldots, q_v^K, u_{v1}, \ldots, u_{vN}, M) \text{ \hspace{10pt} \triangleright find order of } M\text{-PSK symbols according to SE strategy} \]
2: \[ \text{for } m = 0 \text{ to } M - 1 \text{ do } d_{\text{test}} := \sum_{k=1}^{K} u_{v}^k e^{\frac{2\pi}{M} \cdot m} + q_v^k |^2 \text{ end for} \text{ \hspace{10pt} \triangleright compute all possible square length contrib.} \]
3: \[ \text{step}_v := \text{sortascending}(d_{\text{test}}) \text{ \hspace{10pt} \triangleright returns step}_v \text{ which contains values } m \text{ sorted in order of increasing value of } d_{\text{test}} \]
4: \[ n_v := 1 \]
5: \[ \text{end function} \]
Figure 2: Illustration of relationship between $z_k$ and $y$ for $a = 1$. Note that with the definitions in Section 4.4, $b_1 < 0$, $b_2 > 0$, $\beta_1 < 0$, and $\beta_3 > 0$ as long as $P_d > 0.5$ and $P_f < 0.5$.

Figure 3: Probability of error $P_e$ vs. $E_b/N_0$ for decision threshold $\gamma_0 = 0$. $K = 8$, $M = 2$, $P_d = 0.8$, $P_f = 0.01$, $BT = 0.1$, and i.i.d. Rayleigh fading.
Figure 4: Probability of error $P_e$ vs. $E_b/N_0$ for decision threshold $\gamma_0 = 0$. $K = 8$, $M = 2$, $P_d = 0.8$, $P_f = 0.01$, $BT = 0$, and i.i.d. Rayleigh fading.
Figure 5: Probability of detection $P_{d_0}$ vs. $E_b/N_0$ for a probability of false alarm of $P_{f_0} = 0.001$. $K = 8$, $M = 2$, $P_d = 0.7$, $P_f = 0.05$, $BT = 0.1$, and i.i.d. Rayleigh fading.
Figure 6: Probability of detection $P_d$ vs. probability of false alarm $P_f$, $K = 8$, $M = 2$, $P_d = 0.7$, $P_f = 0.05$, $BT = 0.1$, $E_b/N_0 = 20$ dB, and i.i.d. Rayleigh fading.
Figure 7: Probability of error $P_e$ vs. number of sensors $K$. $M = 2$, $P_d = 0.7$, $P_f = 0.05$, $BT = 0.1$, $E_b/N_0 = 20$ dB, and i.i.d. Rayleigh fading.
Figure 8: Probability of detection $P_{d_0}$ vs. number of sensors $K$ for a probability of false alarm of $P_{f_0} = 0.001$. $M = 2$, $P_d = 0.7$, $P_f = 0.05$, $BT = 0.1$, $E_b/N_0 = 20$ dB, and i.i.d. Rayleigh fading.
Figure 9: Probability of missed detection $P_m$ vs. sensor SNR, $SNR_s$. $M = 4$, $BT = 0.1$, $E_b/N_0 = 30$ dB, and i.n.d. Rayleigh fading.
Figure 10: Number of real multiplications per decision vs. $E_b/N_0$. $M = 4$, $\text{SNR}_s = -3$ dB, and i.i.d. Rayleigh fading.