Bayesian Design of Tandem Networks for Distributed Detection With Multi-bit Sensor Decisions

Alla Tarighati, Student Member, IEEE, and Joakim Jaldén, Senior Member, IEEE,

Abstract—We consider the problem of decentralized hypothesis testing under communication constraints in a topology where several peripheral nodes are arranged in tandem. Each node receives an observation and transmits a message to its successor, and the last node then decides which hypothesis is true. We assume that the observations at different nodes are, conditioned on the true hypothesis, independent and the channel between any two successive nodes is considered error-free but rate-constrained. We propose a cyclic design algorithm for the design of nodes using the minimum expected error probability as a design criterion in a person-by-person methodology, where the number of communicated messages is not necessarily equal to the number of hypotheses. The number of peripheral nodes in the proposed method is in principle arbitrary and the information rate constraints are satisfied by quantizing the input of each node. The performance of the proposed method for different information rate constraints, in a binary hypothesis test, is compared to the optimum rate-one solution due to Swaszek’s and a method proposed by Cover, and it is shown numerically that increasing the channel rate can significantly enhance the performance of the tandem network.

I. INTRODUCTION

DISTRIBUTED signal processing systems have been of reliability, survivability and reduced communication bandwidth requirements received a significant interest in the past. In the context of distributed detection, considerable progress was made during the past few decades, see [1]–[3] and references therein. Observations are in a distributed, or decentralized, hypothesis testing system made at spatially separated sensors. If the sensors are able to communicate all their data to a central processor there is no fundamental difference from a centralized hypothesis test where the optimal solution is given by threshold tests on the likelihood ratios computed from the complete set of observations. On the other hand, if there are communication constraints on the channels between the sensors, some preliminary processing of the data need to be carried out at each sensor and a compressed, or quantized, version of the received data is then instead given as the sensor output. The output of each sensor is then, according to the network arrangement, sent to either another sensor or to a fusion center (FC), which makes the final decision in favor of one of the hypotheses. In the context of distributed detection each sensor is thus an intelligent unit, and is therefore often referred to as a decision maker (or DM) [2], [3]. The goal of this paper is to introduce a general methodology for the design of the DMs in tandem networks for M-ary hypothesis testing.

The optimal design of the DMs in a tandem network was previously studied in [4]–[6] under the assumption that the observations at the sensors were conditionally independent. This scenario has also recently been generalized in [7] to the case of conditionally dependent observations. Common to [4]–[7] are that the channels between the DMs are considered rate-constrained but error-free. While [4]–[6] considered binary hypothesis testing and binary messages between the DMs, [7] relaxed this assumption and considered general M-ary hypothesis testing with M-valued messages for M ≥ 2. We shall herein consider M-ary hypothesis testing and conditionally independent sensor observations, but will generally allow for higher communication rates than what is provided by M-valued messages.

With respect to the optimal performance limits of tandem networks, it was shown in [1], [8] that for distributed networks with two DMs the optimal tandem network performs at least as well as the optimal parallel network. However when the number of DMs increases parallel networks performs better than serial networks, and for any given distributed detection problem with i.i.d. observations there exists a number of DMs at which the parallel network becomes better [1]. In the case of a parallel topology with any logical decision functions, the error probability goes to zero very quickly as the number of DMs increases. This does however not hold in general for the tandem topology. It was in fact shown in [9] that the rate of error probability decay of the tandem network is always sub-exponential in the total number of DMs, while the error probability decay of a parallel network is exponential in the total number of DMs [10].

The asymptotic performance of parallel and tandem networks has attracted a lot of interest over the past years [8]–[12]. It was for instance shown in [8], [11] that when the DMs are allowed to send M-valued messages for M-ary hypothesis testing, a necessary and sufficient conditions for the probability of error to asymptotically go to zero is that the log-likelihood ratio of the observation at each DM, between any two arbitrary hypotheses, is unbounded in magnitude. In other words, in the general case with potentially bounded log-likelihood ratios (strictly) more messages than hypotheses are needed to drive the error to zero. In the case of binary hypothesis testing (M = 2) and for bounded log-likelihood ratios, Cover [11] proposed an algorithm with a four-valued
message which achieves zero-limiting probability of error under each hypothesis. This idea was later generalized by Koplowitz [12] to show that \((M + 1)\)-valued messages are sufficient for achieving zero-limiting probability of error in \(M\)-ary hypothesis testing, even if the log-likelihood ratio is bounded.

For tandem networks of fixed size, Papastavrou and Athans [8] proposed a simple but suboptimal scheme for the network design in which each DM is optimized for locally minimal error probability at its output, instead of for globally optimal performance. In the particular scheme of [8], a necessary and sufficient conditions to achieve zero-limiting probability of error is also that the log-likelihood ratio of the observation of each DM be unbounded from both above and below. However, a side effect of optimizing the performance (i.e., minimizing the error probability) locally at the output of each DM is that the messages are then again constrained to be \(M\)-valued for the \(M\)-ary hypothesis test as a one-to-one relation between the DM output messages and the hypotheses is needed in definition of the local probability of error. Thus, the problem of designing the DMs in a tandem network for arbitrary-valued messages remains largely open [5], even though it is known that increasing the number of communication messages can improve the performance of a network of sensors arranged in parallel [13]. The latter point was, e.g., exemplified in [14] where it was shown that allowing the first sensor to communication two-bit messages instead of one-bit messages could significantly improve the performance of a two-sensor network for binary hypothesis testing.

Motivated by the above, the main contribution of this paper is to introduce a numerical methodology for designing an \(N\)-node tandem network of DMs with arbitrary-valued messages. As in [2], [15], the objective is to design the decision rules at the DMs so as to minimize the overall average cost of making the last decision under the assumption that the observations are conditionally independent. To this end, we propose person-by-person optimization of each DM. However, to arrive at a tractable performance metric for the design (optimization) of each individual DM we design each DM jointly with the FC (fusion center), i.e., the DM is optimized under the assumption that the FC always employs the (optimal) maximum a-posteriori (MAP) rule applied to whatever input it receives. This obviates the need for the number of messages at the output of the DM to be equal to the number of hypotheses, making the proposed method more generally applicable than prior work. Each DM is then also (internally) optimized with respect to the so-obtained metric using a person-by-person method applied to the individual input to output assignments. We finally show that the proposed algorithm is computationally efficient; its complexity is linear in the number of DMs. Although the proposed design is not globally optimal, we show good performance with respect to existing benchmark solutions through numerical examples.

The outline of this paper is as follows. In Section II we describe the structure of the tandem network and formulate the problem. In Section III we introduce a restricted network model, describe how it can be connected to the tandem network, and present the proposed design method. Numerical examples are given in Section IV and Section V concludes the paper.

II. PROBLEM STATEMENT

We consider a Bayesian decentralized hypothesis testing system with \(N\) sensors in a tandem network as shown in Fig. 1. The sensors, or decision makers (DM), observe the same phenomenon \(H\). DM \(l\), using its own observation \(x_l \in X_l\) and the output \(u_{l-1} \in M_{l-1}\) of its predecessor makes a decision \(u_l \in M_l\) and sends it to its successor DM \((l + 1)\). The exception to this rule is DM 1 which using only its own observation \(x_1 \in X_1\) makes a decision \(u_1 \in M_1\). Throughout this work, the set of possible observations \(X_l\) and the set of possible message \(M_l\) are assumed to be discrete for \(l = 1, \ldots, N\). Although we restrict our attention to discrete observation spaces, \(X_l\) could be used to approximate observations in a continuous space using fine-grained binning as in [16], [17], where each bin, or interval, in the continuous observation space can then be represented by an index \(x_l\) from the discrete set \(X_l\).

The channel between DM \(l\) and its successor DM \((l + 1)\) is an error-free and rate-constrained channel of rate \(R_l = \log_2 ||M_l||\) bits where \(||M_l||\) denotes the cardinality of \(M_l\). DM \(l\) \((l > 1)\) can be viewed as a quantizer that maps its input vector \((x_l, u_{l-1})\) to an output value (message) \(u_l\) using a decision function \(\gamma_l : X_l \times M_{l-1} \rightarrow M_l\), i.e.,

\[
\gamma_l(x_l, u_{l-1}) = u_l \quad l = 2, \ldots, N.
\]

DM 1 only uses its direct observation \(x_1\) to make the decision \(u_1\) using a decision function \(\gamma_1 : X_1 \rightarrow M_1\), i.e.,

\[
\gamma_1(x_1) = u_1.
\]

Each decision function \(\gamma_l\) can also be viewed as an index assignment which assigns an index \(u_l\) to each input vector.
Our objective is to derive decision functions of DM and where

\[ P(x_l|H_j), \quad j = 1, 2, \ldots, M. \]

In this paper, as in [2], [15] referred to in the introduction, the key point is that under the person-by-person methodology when jointly designing DM l, using its observation \( y_i \), produces a message \( u_l \) from the discrete index space \( M_l \) and sends this message to DM N through a discrete channel. DM N, as FC of the network, using the received message \( u_{N-1} \in M_{N-1} \) and its own observation \( y_N \), makes the global decision \( u_N \in \{1, 2, \ldots, M\} \) for an M-ary hypothesis testing problem. The channel between the DMs is a discrete channel which maps the index \( u_l \) to \( u_{N-1} \) with a known transition probability \( P(u_{N-1}|u_l, H_j) \) which depends on the hypothesis \( H \).

Under the person-by-person methodology, the design of DM \( l (1 \leq l < N) \) in the original tandem network of Fig. 1 is analogous to the design of DM \( l \) in a particular instance of the restricted model in Fig. 2. To see this, let

\[ \gamma_l \triangleq \gamma(\gamma_{l-1}, H_j) \]

be the complete observation of DM \( l - 1 \) combining the direct observation of DM \( l \) in the original network and the input from DM \( l - 1 \) – and let \( y_N \triangleq y_{N-1} \). The conditional PMFs of the inputs to DM \( l \) and DM \( N \) are due to the independence of \( x_l \) and \( u_{l-1} \) given by

\[ P_j(y_l) = \begin{cases} P_j(x_l) & \text{if } l = 1 \\ P_j(x_l)P_j(u_{l-1}) & \text{if } 1 < l < N \end{cases} \]

\[ P_j(y_N) = P_j(x_N). \]

The transition probability \( P(u_{N-1}|u_l, H_j) \) is simply the transition probability from \( u_l \) to \( u_{N-1} \) in the original network. The key point is that under the person-by-person design methodology when jointly designing DM \( l \) and DM \( N \), DM \( l \) to DM \( l - 1 \) and DM \( l + 1 \) to DM \( N - 1 \) remain fixed and so does the observation \( u_{N-1} \) conditioned on the hypothesis, are independent. We also assume that the observations at the DMs, conditioned on the hypothesis, are independent, which is also the fusion center of the network.

We assume that the observations at the DMs, conditioned on the hypothesis, are independent, which implies that \( x_l \) and \( u_{l-1} \), conditioned on the hypothesis, are independent. We also assume that the observation \( x_l \) of DM \( l \) is a random variable with known conditional probability mass functions (PMF) \( P(x_l|H_j), \quad j = 1, 2, \ldots, M. \)

For a fixed set of decision functions \( \gamma_l \) to \( \gamma_{N-1} \), the optimal decision rule for the FC is the maximum a-posteriori (MAP) rule. For this reason, and since the MAP rule allows for a tractable implementation in a single sensor scenario, we will assume that the FC always uses the MAP rule in order to make the global decision \( u_N \) given its input \( z \triangleq (x_N, u_{N-1}) \). Given \( z \), the FC thus decides on \( H_m \) if

\[ \pi_m P(z|H_m) = \max_j \{ \pi_j P(z|H_j) \} \]

where \( \pi_j \triangleq P(H_j) \) is the a-priori probability of hypothesis \( H_j \) and where \( j = \{1, 2, \ldots, M\} \) for the M-ary hypothesis testing problem. The expected minimal error probability in estimating \( H \) given an observation \( z \) from the complete observation set \( \mathcal{Z} \triangleq \mathcal{X}_N \times \mathcal{M}_{N-1} \) is

\[ P_{\text{E}} = 1 - \sum_{z \in \mathcal{Z}} \max_j \{ \pi_j P(z|H_j) \}. \]

Our objective is to derive decision functions of DM \( l \) through \( N - 1 \) that attempts to minimize the expression in (2), so as to minimize the global error probability.

**III. DM DESIGN THROUGH A RESTRICTED MODEL**

In this section we will show that under the person-by-person methodology, the design of each DM in the tandem network shown in Fig. 1 is analogous to the design of a DM (labelled DM \( l \) for notational consistency) in a restricted model as shown in Fig. 2 where DM \( N \) in both networks use the MAP rule [cf. (1)] as the fusion function. Then, using the restricted model, we introduce a computationally efficient algorithm for the design of the DMs.
The Markov chain property implies that the transition probability
from $u_1$ to $u_{N-1}$ in the original tandem network is given by
\[ P_j(u_{N-1}|u_1) = \sum_{u_2} \ldots \sum_{u_{N-2}} P_j(u_{N-1},u_{N-2},\ldots,u_1|u_1) \]
\[ = \sum_{u_2} \ldots \sum_{u_{N-2}} \prod_{i=l+1}^{N-1} P_j(u_i|u_{i-l},u_l) \]
\[ = \sum_{u_2} \ldots \sum_{u_{N-2}} \prod_{i=l+1}^{N-1} P_j(u_i|u_{i-1}). \]  
(6)

Equivalently, in matrix form if we define $P_{j,N-1}^{l}(m,n) \triangleq P_j(u_{N-1}|u_l = m,u_{l-1} = n)$, (6) implies
\[ P_{j,N-1}^{l} = P_{j}^{N-1} \times \ldots \times P_{j}^{l+2} \times P_{j}^{l+1}. \]  
(7)

Thus, using (7) we can replace all the DMs between DMs $l$ and $N$ by a single hypothesis dependent transition probability given by $P_{j,N-1}^{l}$ when designing DM $l$. Once the transition probability matrix $P_{j,N-1}^{l}$ is found, the probability masses of the messages of DM $N-1$, $P_j(u_{N-1})$, can be easily found from the probability $P_j(u_l)$ of the messages of DM $l$. The complete set of transition probability matrices $P_{j,N-1}^{l}$, $1 \leq l \leq N-2$ can also be found efficiently (with linear complexity in $N$) by a recursion with decreasing index $l$, by noting that (7) implies
\[ P_{j,N-1}^{l} = P_{j}^{l+1} \times P_{j}^{l+1} \times P_{j}^{l+1}. \]  
(8)

where $P_{j,N-1}^{l} = I_{||M_l||}$ by definition, and then stored for the forward design of $\gamma_l$ for $l = 1,\ldots,N-1$ in one pass of the iterative design algorithm.

By defining the probability mass vector of the messages at the output of DM $k$ as
\[ q_j^k \triangleq [P_j(u_k = 1),\ldots,P_j(u_k = ||M_k||)]^T, \]  
(9)
the Markov chain property implies [20]
\[ q_j^{N-1} = P_{j,N-1}^{l} \times q_j^{l}. \]  
(10)

**Algorithm 1** Algorithm for designing DMs in the tandem network Fig. 1

1: **Input:** Initialized $\gamma_1$, $l = 1,\ldots,N-1$
2: **Output:** Updated $\gamma_l$, $l = 1,\ldots,N-1$
3: **Initialize:**
\[ P_{j,N-1}^{l} \triangleq I_{||M_N||}, j = 1,\ldots,M \]
4: **repeat**
5: for $l = N - 2 : 1$
6: find $P_{j}^{l+1}$ using (5), $j = 1,\ldots,M$
7: $P_{j}^{N-1} \leftarrow P_{j}^{l} \times P_{j}^{l+1}$, $j = 1,\ldots,M$
8: **end for**
9: for $l = 1 : N - 1$
10: optimize $\gamma_l$ using restricted model
11: if $l = 1$ then
12: find $q_j^l$ using (11), $j = 1,\ldots,M$
13: else
14: update $P_{j}^{l}$, $j = 1,\ldots,M$
15: $q_j^l \leftarrow P_{j}^{l} \times q_j^{l-1}$, $j = 1,\ldots,M$
16: **end if**
17: **end for**
18: **until** a stopping condition is fulfilled

Each element of $q_j^l$ can given $\gamma_l$ (in principle) be found as
\[ P_j(u_l) = \sum_{y \in \gamma_l^{-1}(u_l)} P_j(y), \]
where $P_j(y)$ is defined in (4). When $l = 1$, $P_j(y)$ is simply equal to $P_j(x_1)$ where $x_1$ is the first direct observation in the original network [cf. (5)], while $P_j(y)$ for $l > 1$ also depends on $P_j(u_{l-1})$, or equivalently, $q_j^{l-1}$ for $j = 1,\ldots,M$. The latter probability mass vector can however also be obtained recursively by noting that $q_j^k = P_j \times q_j^{k-1}$ for $k = 2,\ldots,l-1$ and that (for $k = 1$)
\[ q_j^1 = \sum_{x_1 \in \gamma_1^{-1}(m)} P_j(x_1). \]  
(11)

where $1 \leq m \leq ||M_1||$. Inserting $P_j(u_{l-1})$ into (4) gives $P_j(y)$ which together with $P_j(u_k|u_{k-1})$ completely defines the restricted model for the design of DM $l$.

The minimum error probability of a given decision function $\gamma_l$ under MAP decoding at the FC can thus be calculated by calculating $P_j(u_l|u_{l-1})$ using (5), forming $P_j^{l}$, and computing $q_j^{N-1} = P_j^{l} \times q_j^{l-1}$ which yields $P_j(u_{N-1})$ for $j = 1,\ldots,M$; and then applying (2) with $P(z|H_j) = P_j(x_N)P_j(u_{N-1})$. This, in principle, allows for optimizing $\gamma_l$ with respect to the global error probability. Note here that both $P_j^{l} \times q_j^{l}$ and $q_j^{l-1}$ are considered fixed (and precomputed) when designing DM $l$.

Algorithm 1 summarizes the overall proposed design procedure of the tandem network, in which for the design of each DM a restricted model should be formed. In each cycle of the optimization, the DMs – from DM 1 to DM $(N-1)$ – are updated one-by-one jointly with DM $N$. After updating DM $l$ its conditional transition probability matrices $P_j^{l}$ are updated for $j = 1,\ldots,M$ and after each cycle the algorithm
does another cycle until a given stopping condition is fulfilled (e.g., maximum number of iterations). The Algorithm then terminates and the last set of decision functions is the final design.

Algorithm 1 shows how, regardless of network size, each DM in a tandem network can be designed using the restricted model with a fixed computational burden. Once an explicit design method for the design of the DMs in the restricted model is found, it can be used for the design of a tandem network with arbitrary size \(N\), at an overall complexity that grows only linearly in \(N\) (assuming a fixed number of iterations over the sequence of DMs). In the next subsection we will introduce a suboptimal, but computationally efficient, method for the design of the DMs in the restricted model.

### B. Design of DMs in the Restricted Model

From now on our focus will be on the restricted model and we derive the optimization equations for this model, since as explained above the design of DM \(l\) in the original tandem network is analogous to the design of DM \(l\) in the restricted model with hypothesis dependent transition probability matrices given by \(P_{j}^{l=1,\ldots,N-1}\) for \(j = 1, \ldots, M\). The minimal expected error probability of the restricted model, obtained by MAP decoding at DM \(N\), is given by [cf. (2)]

\[
P_E = 1 - \sum_{y_N \in \mathcal{Y}_N} \sum_{u_{N-1} \in \mathcal{M}_{N-1}} \max_j \{ P_j(y_N | P_j(u_{N-1}) \nu_j) \}.
\]

To find the index assignment of each input of DM \(l\) that minimizes the global error probability is a combinatorial problem. However, in order to arrive at a computationally efficient procedure, we propose in the following a simple, but suboptimal, method for the design of a particular DM. To do this, we adopt person-by-person optimization within each DM. In other words, the index assignment is done in a person-by-person manner in terms of the input set; an index is assigned to a specific input, while the assigned indices to the other inputs are fixed. Then the optimization formulation for the design of DM \(l\) is given as

\[
\gamma_l^+(y_l) = \arg \max_{\nu_l \in \mathcal{M}_l} \sum_{y_N \in \mathcal{Y}_N} \sum_{u_{N-1} \in \mathcal{M}_{N-1}} \max_j \{ P_j(y_N | P_j(u_{N-1}) \nu_j) \},
\]

where the index assignment \(\nu_l\) can change the probability masses in the vector \(\nu_l\) [cf. (9)] which consequently affects the PMFs \(P_j(u_{N-1})\) through the transition probability matrix \(P_{j}^{l=1,\ldots,N-1}\) according to (10). The optimizer of (13) is found by searching over all possible indices \(\nu_l \in \mathcal{M}_l\) for and every input \(y_l \in \mathcal{Y}_l\). Now let

\[
P_{j}^{l=1,\ldots,N-1} = [r_{j,1} | r_{j,2} | \ldots | r_{j,\tilde{V}}]^T,
\]

where \(\tilde{V} \equiv |\mathcal{M}_{N-1}|\), and where \(r_{j,m}\) is a column-vector containing the elements of the \(m\)th row of \(P_{j}^{l=1,\ldots,N-1}\). Then (10) implies that the \(m\)th element of \(\nu_l\), or \(P_j(u_{N-1} = m)\), is found by

\[
P_j(u_{N-1} = m) = r_{j,m}^T \times q_j^l = \langle q_j^l, r_{j,m} \rangle,
\]

where \(\langle a, b \rangle\) is the inner product of the vectors \(a\) and \(b\). Using (15) the optimizer (13) is then written as

\[
\sum_{y_N \in \mathcal{Y}_N} \sum_{u_{N-1} \in \mathcal{M}_{N-1}} \max_j \{ P_j(y_N | \nu_j, u_{N-1}) \},
\]

where we use the superscript \((y_l, \nu_l)\) for the vector \(q_j^l(y_l, \nu_l)\) to emphasize that the assigned index to \(y_l\) is \(\nu_l\), i.e., \(\gamma_l^+(y_l) = \nu_l\). In a shorthand notation,

\[
q_j^l(y_l, \nu_l) \triangleq [P_j(u_l = 1), \ldots, P_j(u_l = V)]^T_{\gamma_l^+(y_l) = \nu_l},
\]

where \(V \triangleq |\mathcal{M}_l|\).

It should be noted that after changing the decision function for an input \(y_l\), all the conditional probability masses (or equivalently \(q_j^l(y_l, \nu_l)\)) need to be calculated which has the potential to make the algorithm difficult to implement for larger rates. However, in the following we will show that only a couple of probability masses in each vector \(q_j^l(y_l, \nu_l)\) needs to be updated while the other probability masses remain fixed. Furthermore, we will propose an iterative algorithm for the design of DM \(l\) in the restricted model.

To this end, assume now that the decision function for a specific input \(y_l \in \mathcal{Y}_l\) is \(\nu_l \in \mathcal{M}_l\), i.e., \(\gamma_l(y_l) = \nu_l\), and the corresponding conditional PMFs are \(P_j(u_l)\), \(u_l = 1, 2, \ldots, V\), where the vector \(q_j^l(y_l)\) is defined as

\[
q_j^l(y_l) \triangleq [P_j(u_l = 1), \ldots, P_j(u_l = V)]^T_{\gamma_l(y_l) = \nu_l}.
\]

**Algorithm 2** Algorithm for designing DM \(l\) in restricted model

**Fig. 2**

1: **Input**: \(\gamma_1, P_j(y_l), P_j(y_N)\) and \(P_j(u_{N-1} | u_l)\)
2: **Output**: Updated \(\gamma_l\)
3: set \(\Delta P_E \leftarrow \infty\), a threshold \(\eta\) and find \(P_E\) using (12)
4: **while** \(\Delta P_E > \eta\) **do**
5: for \(i = 1 : \|Y_l\|\) **do**
6: \(y_l \leftarrow \gamma_l(i)\)
7: update the assigned index to input \(y_l\) using (20)
8: for \(j = 1 : M\) **do**
9: update vector \(q_j\) using (18)
10: **end for**
11: **end for**
12: find \(P_E^+\) using (12) and evaluate \(\Delta P_E = P_E - P_E^+\)
13: \(P_E \leftarrow P_E^+\)
14: **end while**
Then, the conditional PMFs when $u_l = \nu$ are
\[
P_j^+(u_l = \nu) = \sum_{y \in \mathcal{Y}_l(\nu)} P_j(y) = \sum_{y \in \mathcal{Y}_l(\nu), y \neq \nu} P_j(y) + P_j(\nu),
\]
where $\mathcal{Y}_l(\nu)$ is the set of all inputs $y$ which gives $\gamma_l(y) = \nu$ (including $y_l$). Now assume that the assigned index to input $y_l$ changes to $\nu, \nu_l \in \mathcal{M}_l$, or equivalently $\gamma_l^+(y_l) = \nu_l$, then $y_l$ does not belong to $\mathcal{Y}_l(\nu)$ anymore (it belongs to $\mathcal{Y}_l(\nu_l)$) and the new conditional PMFs when $u_l = \nu$ and $u_l = \nu_l$ are
\[
P_j^+(u_l = \nu) = \sum_{y \in \mathcal{Y}_l(\nu)} P_j(y) = P_j(u_l = \nu) - P_j(y_l),
\]
\[
P_j^+(u_l = \nu) = P_j(u_l = \nu_l) + P_j(y_l),
\]
while the other conditional probability masses remain fixed
\[
P_j^+(u_l \neq \nu, \nu) = P_j(u_l \neq \nu, \nu_l).
\]

Consequently, the vector of probability masses for the new index assignment $\mathbf{q}_{j}^{\nu,\nu_l}$ can be found from the old vector $\mathbf{q}_{j}^{\nu,\nu}$ using
\[
\mathbf{q}_{j}^{\nu,\nu_l} = \mathbf{q}_{j}^{\nu,\nu} + P_j(y_l)(\mathbf{e}_{\nu} - \mathbf{e}_{\nu_l}), \tag{18}
\]
where $\mathbf{e}_{\nu}$ is the $\nu$th basis vector in the $V$-dimensional Euclidean space.

This is illustrated in Fig. 3 which shows how the probability masses change when the assigned index to input $y_l$ changes from $\nu$ to $\nu_l$. Thus after updating the assigned index to each input $y_l$ a couple of conditional PMFs, corresponding to the previous and the new assignment, needs to be updated. In other words, only a couple of conditional probability masses in the vector $\mathbf{q}_{j}^{\nu,\nu_l}$ needs to be modified using (18), while the other probability masses remain fixed.

Consider again the optimizer (16) for updating the assigned index to input $y_l$. Assume that the assigned index to input $y_l$ prior to updating it is $\nu$, i.e., $\gamma_l(y_l) = \nu$ and the corresponding vector of probability masses is $\mathbf{q}_{j}^{\nu,\nu_l}$. Using (18) the inner product $\langle \mathbf{q}_{j}^{\nu,\nu_l}, \mathbf{r}_{j,u,N-1} \rangle$ is written
\[
\langle \mathbf{q}_{j}^{\nu,\nu_l}, \mathbf{r}_{j,u,N-1} \rangle = \langle \mathbf{q}_{j}^{\nu,\nu_l}, \mathbf{r}_{j,u,N-1} \rangle + P_j(y_l)(\mathbf{e}_{\nu} - \mathbf{e}_{\nu_l}, \mathbf{r}_{j,u,N-1})
\]
\[
= \langle \mathbf{q}_{j}^{\nu,\nu_l}, \mathbf{r}_{j,u,N-1} \rangle + P_j(y_l) \Delta r_{j,u,N-1}(\nu, \nu_l),
\]
where
\[
\Delta r_{j,u,N-1}(\nu, \nu_l) = \mathbf{r}_{j,u,N-1}(\nu_l) - \mathbf{r}_{j,u,N-1}(\nu). \tag{19}
\]
The optimizer (16) can be written as
\[
\gamma_l^+(y_l) = \arg \max_{\nu, \nu_l} \sum_{y \in \mathcal{Y}_l} \sum_{u \neq \nu, \nu_l} \max_j \left\{ \pi_j P_j(y_N) \right\},
\]
\[
\left\{ \langle \mathbf{q}_{j}^{\nu,\nu_l}, \mathbf{r}_{j,u,N-1} \rangle + P_j(y_l) \Delta r_{j,u,N-1}(\nu, \nu_l) \right\}, \tag{20}
\]
where $\nu$ is the assigned index to input $y_l$ prior to updating it.

The updating rule for the design of DM $l$ in the restricted model is described in Algorithm 2. In this algorithm, after updating all the input indices $y_l$, the conditional PMFs of DM $l$ are updated and the performance improvement (the improvement in error probability $\Delta P_E$) is calculated. If it is greater than a threshold $\eta$ the algorithm does another cycle. Otherwise it terminates and the last index assignment for DM $l$ is the final index assignment.

In closing, we should mention that the optimizer (20), which is equal to $1 - P_E$, used for the design of DMs arranged in tandem, has a close relation to the true error probability at the FC [cf. (12), (13)]. While we are updating each DM, we try to minimize the error probability of the network, while the other DMs are kept fixed. The error probability is therefore decreased gradually until it converge to an optimal solution. Because of the nature of the person-by-person methodology, it (in general) does however not give a globally optimum solution and the resulting locally optimum solution is affected by the initialization of the DMs.

IV. SIMULATIONS

In this section we present some results illustrating the application of the proposed method in the design of tandem networks. We consider the case of binary hypothesis testing, i.e., $M = 2$, for independent and identically distributed observations $x_i$, $i = 1, 2, \ldots, N$. We further limit our attention to...
to the case where each real valued observation consists of an antipodal signal ±α in unit-variance additive white Gaussian noise. The observation model at each DM is:

\[ H_0 : y_i = -\alpha + n_i \]
\[ H_1 : y_i = +\alpha + n_i. \]

We also define the per channel signal-to-noise ratio (SNR) as \( \mathcal{E} \triangleq |\alpha|^2 \), and assume that the hypotheses are equally likely (\( \pi_0 = \pi_1 = 0.5 \)). Furthermore the channel rates are considered to be the same for all links and equal to \( R \) which implies the DMs output messages are from the set \( \{1, 2, \ldots, 2^R\} \).

Although the proposed design method is for discrete observation sets, it can be applied to the continuous real valued observations using the fine-grained binning idea \([17]\). To do that, the interval \([-\alpha-4, \alpha+4] \) (containing 0.9997 of the total probability mass for each DM) is represented by 128 discrete probability masses per hypothesis to form discrete observation sets from the continuous observations, i.e., \( ||x_k|| = 128 \).

Fig. 4 shows the evolution of the error probability for designed tandem networks with 1 ≤ \( N \) ≤ 20 DMs after three iterations of Algorithm \([1]\). The channel rates are equal to three bits and the per channel SNR is \( \mathcal{E} = -10 \) dB. The system is initialized in such a way that each DM, regardless of its observation, passes its input from its predecessor to its successor (i.e., \( \gamma_k(x_{k}, u_{k-1}) = u_{k-1}, 1 < k < N \) and the first DM provides its output randomly from the index set \( \{1, \ldots, 2^R\} \). Then, in the initialized network, the fusion center (DM \( N \)) using the MAP criterion to make the global decision, which gives the same error probability regardless of the number of DMs before it. Although the system is not initialized in an intelligent way, the proposed algorithm results in a significant performance improvement after the first iteration and then the algorithm shows no visible improvement after three iterations.

\(^1\)It should be mentioned that in this paper the hypothesis set \( \{H_1, \ldots, H_M\} \) is used for \( M \)-ary hypothesis testing, while for \( M = 2 \) we use the hypothesis set \( \{H_0, H_1\} \) instead of \( \{H_1, H_2\} \) to have the same notation as the classical texts on binary hypothesis testing problems.

In Fig. 5 and Fig. 6 the performance of the designed tandem network for various channel rates and number of DMs is compared to the optimum rate-one performance \([5]\) and Cover’s \([11]\) rate-one method, for \( \mathcal{E} = -10 \) dB and \( \mathcal{E} = 0 \) dB. We also include the performance of the unconstrained linear detector which is optimum for this problem when the channels are infinite-rate (\( R = \infty \)). The results in Fig. 5 and Fig. 6 are achieved after three iterations. For rate-one channels, the performance of the proposed method is indistinguishable from the optimum solution, while increasing the channels rate leads to better performance which is in harmony with the parallel network. The simulation results show that increasing the rate of the channels between the DMs can significantly improve the performance of the tandem network: for example when the channel rates are equal to \( R = 4 \) the performance of the designed network is very close to the un-constrained case, at least for \( N \) up to 20 for \( \mathcal{E} = -10 \) dB and for \( N \) up to 10 for \( \mathcal{E} = 0 \) dB.

Although the proposed method gives an applicable tool for the design of tandem networks, it should again be mentioned that, because of the nature of the person-by-person optimization, it leads to only locally optimal solutions which depend on the initialization of the DMs.

V. CONCLUSION

In the context of decentralized hypothesis testing in tandem networks, we have proposed an iterative algorithm which cyclically improves the performance of the network in terms of the minimum error probability. Introducing a restricted model, we have shown that it is possible to update the decision function of each node together with the fusion function, while all the other peripheral nodes in the network are modeled as a Markov chain.

In this paper, we have considered the hypothesis testing problem in tandem networks which is of interest since it provides a tool for the study of other more complicated
topologies, like tree topologies. It can also be relevant in topologies where a single node with \( m \) bit memory makes observation at different time periods and at each time makes a decision based on its current observation and a previous decision which is stored in memory, and updates its memory with the new decision.

In our model, we have assumed multi-bit communication between the sensors for the general \( M \)-ary hypothesis testing problem while the observations at the sensors are, conditioned on the true hypothesis, independent. The \( M \)-ary hypothesis test in tandem networks when the sensors make an \( M \)-ary decision and have conditionally dependent observations, was considered in [7]. However the problem of multi-bit communication (not necessarily \( M \)-ary) and conditionally dependent observations still remains open, and an extension to our work could include the design of tandem networks for conditionally dependent observations.

**REFERENCES**

[1] J. N. Tsitsiklis, “Decentralized detection,” in *Advances in Statistical Signal Processing*, 1993, pp. 297–344.
[2] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag New York, Inc., 1996.
[3] R. Viswanathan and P. Varshney, “Distributed detection with multiple sensors: Part I—Fundamentals,” *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.
[4] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, “Optimal serial distributed decision fusion,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, 1988.
[5] P. Swaszek, “On the performance of serial networks in distributed detection,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, 1993.
[6] Z.-B. Tang, K. Pattipati, and D. Kleinman, “Optimization of detection networks. I. Tandem structures,” *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, 1991.
[7] P. Yang, B. Chen, H. Chen, and P. Varshney, “Tandem distributed detection with conditionally dependent observations,” in *15th Int. Conf. Inf. Fusion (FUSION)*, 2012, pp. 1808–1813.
[8] I. Papastavrou and M. Athans, “Distributed detection by a large team of sensors in tandem,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, 1992.