Theoretical Bounds in Decentralized Hypothesis Testing

Gökhan Gül

***Abstract*—Three fundamental problems are addressed for distributed detection networks regarding the maximum of performance/detection loss. The losses obtained are, first, due to the choice of decision rule in parallel sensor networks (general-case vs identical decisions), second, due to the choice of network architecture (serial vs parallel), and third, due to the choice of quantization rule (centralized vs decentralized). Previous results, if available, for all these three problems are restricted to the statement that the loss is "small" over some specific examples. The key principles underlying this study are delineated as follows. First, there is a surjection from all simple hypothesis tests to the receiver operating characteristic (ROC) curve. Second, the ROC can be well modeled with linear splines. Third, considering splines with only a finite number of line segments, in fact, on the order of the total number of sensors, is sufficient to determine the maximum loss. Leveraging these principles, infinite-dimensional optimization problems are reduced to their finite-dimensional equivalent forms. The equivalent problems are then numerically solved to obtain the theoretical bounds.**

***Index Terms*—Distributed detection, data fusion, sensor networks.**

I. Introduction

In distributed networks, data is generated at geographically dispersed sensors, with local processing employed to reduce the communication bandwidth required between sensors and the fusion center (see Fig. 1). Although parallel networks attract considerable attention from researchers, sensors are typically battery-powered, imposing limitations on energy and communication range [1]. The power consumption of sensors increases dramatically when they are located far from the fusion center, which significantly shortens the lifetime of the wireless sensor network. Additionally, distant sensors contribute minimally to the final decision at the fusion center and are more susceptible to false judgments in the presence of background noise [2]. In contrast, serial networks offer a power-efficient alternative, as the communication between adjacent sensors incurs negligible power decay, making them safe in terms of power consumption. However, serial networks are vulnerable to link failures [3].

Likelihood ratio test is recognized as optimal for both serial and parallel networks, in both Bayesian as well as in Neyman-Pearson sense [4], [5], [6], [7], with the potential necessity for randomization in the latter. While optimal tests have been established for serial and parallel networks, no general bounds have been previously formulated. Furthermore, the investigation into the maximum loss of detection performance attributable to quantization in both serial and parallel networks has been absent until now. The closest study to this article is by [8], where bounds are derived under the assumption of minimax robustness in sensor decisions. The theory presented in this paper is valid for independent sensor observations, ideal communication channels and reliable sensors and it can be extended to the cases, where these assumptions do not necessarily hold [9], [10], [11], [12]. The subsequent section will provide a summary of the most pertinent existing research.

A. Related Work

1) General-Case vs Identical Decisions in Parallel Networks: For independent and identically distributed (i.i.d.) observations there are various counterexamples showing that identical decision rules for each sensor are not always optimal. In [6] an example was given which has a cost function that introduces a large penalty if both sensors send the same message and the wrong decision is made by the fusion center. The asymmetry of the optimal decision rules for these two sensors could be ascribed to this particular example and does not prove that asymmetrical decision rules may be optimal for our cost function, i.e. the probability of error. In [13], [14] a counterexample is reported which involves discrete density functions for the error probability cost function. In [15] non-identical decisions are shown to improve the ROC over continuous density functions. Similarly [16] deals with continuous densities that are nearly discrete concentrated at two points, where three detectors are used. All these counterexamples seem to occur in situations in which the probability mass of the likelihood ratio at the sensor inputs are concentrated at and around a single point. It is also known that identical decision rules are asymptotically optimal, meaning that there is no loss of optimality in constraining the sensors to use the same decision rule if the total number of sensors in the network tends to infinity [13], [14]. In general necessary and/or sufficient conditions to ensure optimality of identical decisions have not yet been found [15].

Received 5 July 2024; revised 3 December 2024; accepted 28 January 2025. Date of publication 13 February 2025; date of current version 4 March 2025. The associate editor coordinating the review of this article and approving it for publication was Kobi Cohen.

The author is with the Preventive Cardiology and Preventive Medicine, Department of Cardiology, University Medical Center Mainz, Johannes Gutenberg University Mainz, 55131 Mainz, Germany, and also with the Clinical Epidemiology and Systems Medicine, Center for Thrombosis and Hemostasis, University Medical Center Mainz, Johannes Gutenberg University Mainz, 55131 Mainz, Germany (e-mail: goekhan.guel@unimedizin-mainz.de).

Digital Object Identifier 10.1109/TSP.2025.3541569

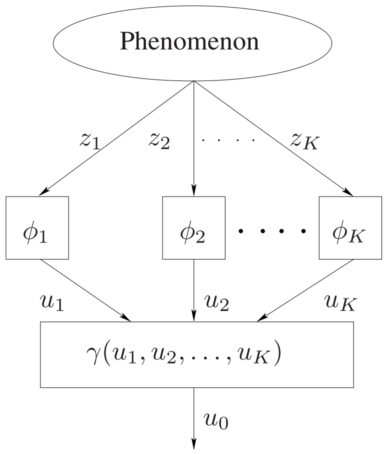

Fig. 1. Parallel network with K decision makers, each represented by the decision rule ϕ_k, and a fusion center associated with the fusion rule γ.

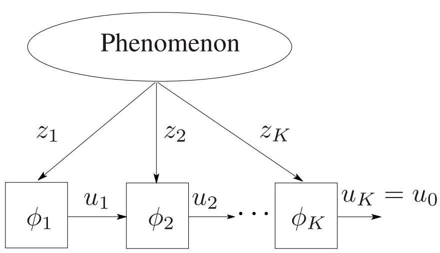

Fig. 2. Serial network with K decision makers, each represented by the decision rule ϕ_k.

However, for the signal detection problem with two sensors, where the known signal is embedded in Gaussian noise, there is no loss in optimality if both sensors use identical decision rules [17].

2) Serial vs Parallel Sensor Networks: When two sensors are present, it is established that, the serial network is at least as good as the parallel network [4]. However, for larger networks, either the serial or parallel configuration may exhibit superior performance. Conversely, in the asymptotic scenario, where the number of sensors approaches infinity, the probability of error for the parallel network tends to zero exponentially for any reasonable set of decision rules, a direct consequence of Chernoff's theorem [18]. Nonetheless, the convergence rate for the serial network lags behind that of the parallel network [3], prompting speculation that, for certain total numbers of sensors K, the parallel topology may outperform the serial topology, with such K expected to be relatively small [5]. The prerequisite for the probability of error to approach zero asymptotically is that the log-likelihood ratio of each sensor observation is unbounded [19], [20]. Moreover, for M-valued quantization, conveyed to K nearest neighbors in sequential order, asymptotic learning is feasible [19], [21]. This paper focuses on binary quantization for a finite number of sensors, where for the serial network the decision is exclusively communicated to the nearest neighbor as shown in Fig. 2.

3) Centralized vs Decentralized Sensor Networks: The performance of both parallel and serial networks is sub-optimal in comparison to a completely centralized processing scheme due to loss of information (quantization) at the local sensors. The performance of the distributed detection network can quantitatively be compared to that of the centralized network [6]. Furthermore, there are examples from algorithm development where such a difference can be observed [22], [23], [24]. No general bounds have been attempted to be derived prior to this work.

B. Contributions

In this paper the following three fundamental problems are addressed:

1) What is the maximum loss of detection accuracy across all simple binary hypothesis tests for the parallel network topology (see Fig. 1) when all sensors are constrained to provide identical decisions, compared to the general case with no such constraint?
2) What is the maximum loss of detection accuracy across all simple binary hypothesis tests when comparing serial (see Fig. 2) and parallel networks with the same number of sensors?
3) What is the maximum loss of detection accuracy across all simple binary hypothesis tests due to quantization in both serial networks as well as in parallel networks, compared to centralized detection, i.e. $\phi_k(y_k) = y_k$ for all k in Fig. 1?

A purely theoretical or purely algorithmic approach is insufficient to tackle these problems. Therefore, first a theoretical foundation is established by reducing infinite-dimensional problems to their equivalent finite-dimensional forms. These equivalent problems are then solved using non-convex optimization algorithms. The contributions of this paper addressing the aforementioned problems are summarized as follows:

1) **Scalable theoretical framework**: The proposed theory enables solving all three problems for any network size K, provided sufficient (finite) memory and computational resources are available.
2) **Bounds for specific configurations**: For problems involving serial networks, theoretical bounds are computed for $K \in \{2, 3, 4\}$. For problems excluding serial networks, bounds are obtained for $K \in \{2, \ldots, 8\}$. These limitations arise due to the inherently exponential growth in computational complexity of serial and parallel sensor networks as K increases [5], [25].
3) **Key insights from simulations:** Simulations demonstrate that quantization does not necessarily result in performance loss. Furthermore, the trivial points $(0, 1)$ and $(1, 0)$ (TPs) on the ROC curve are shown to be unconditionally necessary to achieve the minimum error probability in serial networks.

The organization of this paper is as follows: In the next section, linear splines are introduced to model the receiver operating characteristic (ROC) curves. In Section III the maximum loss of detection performance in parallel sensor networks due to the restriction of sensor decisions being identical is examined. In Section IV the maximum loss resulting from the choice of network topology (serial vs. parallel) is explored. In Section V, the maximum loss due to quantization in both serial and parallel networks is derived in comparison to centralized detection. For each problem, the desired theoretical bounds are numerically obtained by solving a non-convex optimization problem. Finally in Section VI the paper is concluded.

II. Receiver Operating Characteristics With Linear Splines

Let P_0 and P_1 be two distinct probability distributions and p_0 and p_1 be the corresponding density functions, respectively, defined on Ω. Consider the two simple hypotheses

$$\mathcal{H}_0 : Z_k \sim P_0$$
$$\mathcal{H}_1 : Z_k \sim P_1$$

which are tested by a collection of decision makers (sensors), $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ and possibly by a fusion center γ. The random variables (r.v.s) Z_k corresponding to the observations $z_k \in \Omega$ are assumed to be mutually independent and identically distributed. Without loss of generality, as will be shown later, we have $\Omega = [0, 1]$. The decision for each sensor is denoted by $u_k \in \{0, 1\}$, again associated with r.v.s. U_k, and the output of the fusion center is denoted by u_0. Parallel and serial sensor networks are to be studied in this work as illustrated by Figs. 1 and 2. Two special cases can be observed over the parallel network, where the first case corresponds to identical decisions scenario, i.e. $\phi_i(Z_i) = \phi_j(Z_j)$ for all pairs (i, j), and the second scenario corresponds to the centralized detection i.e. $U_k = Z_k$ for all k. While the first scenario considerably reduces the complexity of optimization of the sensor network despite possible performance drop, the second scenario leads to the best achievable detection performance despite the loss of decentralization. In some sense both identical decisions scenario as well as decentralization comes at a cost of some performance drop. In addition to these losses, it is also of our interest to compare the serial networks to parallel networks with the same number of sensors. Let $P_F = \mathbb{E}_{P_0}[U_0]$ and $P_M = \mathbb{E}_{P_1}[1 - U_0]$ be the false alarm and miss detection probabilities, respectively. Then, the Bayesian cost (minimum error probability)

$$\begin{aligned} P_E(K; P_0, P_1, \boldsymbol{\phi}, \gamma) = P(\mathcal{H}_0) P_F(K; P_0, \boldsymbol{\phi}, \gamma) \\ + P(\mathcal{H}_1) P_M(K; P_1, \boldsymbol{\phi}, \gamma) \end{aligned} \quad (1)$$

is considered for performance evaluation, i.e., with $P(\mathcal{H}_0) = 1/2$, and generalizations to other cost functions is straightforward. The key challenge in this work lies in establishing rigorous performance bounds, specifically in determining the maximum loss of detection performance across all simple hypothesis tests for the aforementioned problems.

Let $\hat{P}_0 = 1 - P_0$ denote the complementary cumulative distribution function. Hence, the receiver operating characteristic (ROC) curve for the kth sensor can be defined as

$$R(z_k) = P_1(\hat{P}_0^{-1}(z_k)), \forall z_k \in [0, 1].$$

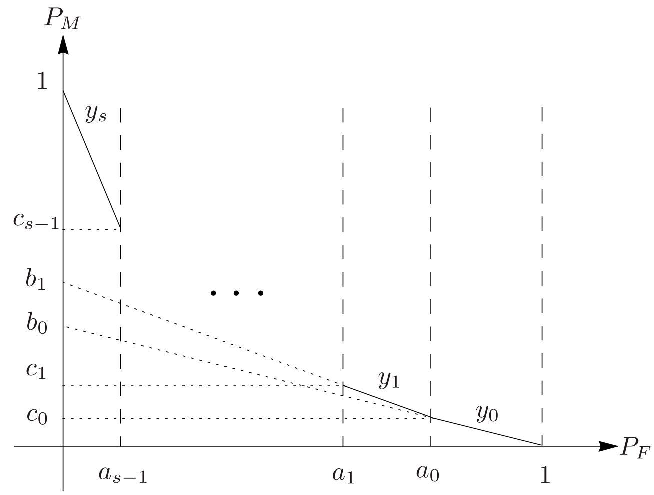

Fig. 3. A sketch of receiver operating characteristic curves given by Theorem II.1.

Remark II.1: R contains all the necessary and sufficient information for the parallel and the serial networks in order to obtain performance bounds. This is due to the fact that the likelihood ratio test is optimal for both networks and the resulting (P_F, P_M) pairs are given by R. Hence, the problem of optimization over all hypothesis tests can be reduced to optimization over all ROCs.

Next, it will be shown that linear splines are sufficient in order to model the set of all ROCs. Consider the following theorem.

Theorem II.1: Let $y : [0, 1] \mapsto [0, 1]$ be a linear spline composed of $s + 1$ line segments. Let furthermore $\boldsymbol{a} = (a_s, \ldots, a_0, a_{-1})$ denote the knot vector with $a_{-1} > a_0 > \cdots > a_s$, where $a_{-1} = 1$ and $a_s = 0$, and $\boldsymbol{b} = (b_{-1}, b_0, \ldots, b_s)$ denote another vector, where $b_s > b_{s-1} > \ldots > b_{-1}$ with $b_s = 1$ and $b_{-1} = 0$. Then, each segment can be denoted by $y_n = r_n x + b_n$, and $r_0 > r_1 > \ldots > r_s$ due to convexity, and the linear spline can be given as

$$y(x; s, a_n, b_n) = \sum_{n=0}^{s} y_n(x; r_n, b_n) \mathbf{1}_{\{a_n \le x \le a_{n-1}\}}(x),$$

where

$$r_n = r_{n-1} + \frac{b_{n-1} - b_n}{a_{n-1}}, \quad (2)$$

with $r_0 = -b_0$, see Corollary II.3. ∎

Proof: A short proof is given in Appendix A. ∎

A visual sketch of the spline model is illustrated in Fig. 3. Sufficiency of Theorem II.1 to model the ROC can be stated by the following corollary.

Corollary II.2: Let $R \in \mathscr{R}$ be any ROC curve, where $\mathscr{R}$ is the set of all ROC curves on $[0, 1]$. Then,

$$\lim_{s \to \infty} \min_{\boldsymbol{a}, \boldsymbol{b}} \int |y(x; s, \boldsymbol{a}, \boldsymbol{b}) - R(x)| dx = 0,$$

where $|\cdot|$ denotes the L_1 norm.

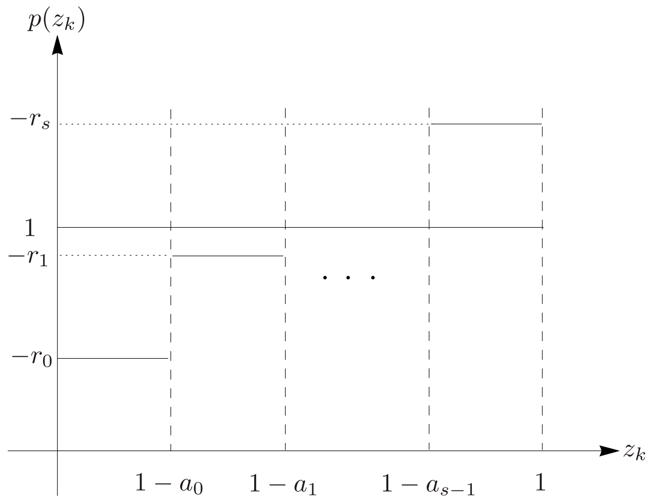

Fig. 4. A sketch of probability density functions given by (3).

Proof: R is continuously differentiable on $[0,1]$, see [8], hence rectifiable [26]. Therefore, its exact length can be measured with infinite number of polygons, i.e., the approximation error goes to zero. ■

Whenever the spline approximation is fixed, false alarm and miss detection probabilities can be found as follows.

Corollary II.3: Miss detection probability c_n corresponding to the false alarm probability a_n, cf. Fig. 3, can be calculated recursively by

$$c_n = \frac{a_n}{a_{n-1}} c_{n-1} + b_n \left(\frac{a_{n-1} - a_n}{a_{n-1}} \right).$$

Proof: A short proof is provided in Appendix B. ■

Similarly, the corresponding density functions can be obtained as in the following corollary.

Corollary II.4: For the given spline approximation, there exists a pair of distributions given by

$$\begin{aligned} p_0(z_k) &= \mathbf{1}_{\{0 \le z_k \le 1\}}(z_k) \\ p_1(z_k) &= -\sum_{n=0}^{s} r_n \mathbf{1}_{\{1-a_{n-1} \le z_k \le 1-a_n\}}(z_k) \end{aligned} \tag{3}$$

or

$$\begin{aligned} p_0(z_k) &= -\sum_{n=0}^{s} \frac{1}{r_n} \mathbf{1}_{\{1-a_{n-1} \le z_k \le 1-a_n\}}(z_k) \\ p_1(z_k) &= \mathbf{1}_{\{0 \le z_k \le 1\}}(z_k) \end{aligned}$$

Proof: The proof is based on the construction of the probability density functions (p.d.f.)s from the ROC in terms of the linear spline approximation, see [8, Proof of Lemma IV.1]. ■

A sketch of the p.d.f.s given by (3) are illustrated in Fig. 4. An important question that needs to be answered is whether the spline approximation leads to mathematically tractable solutions because without further clarification we may still be facing an infinite dimensional optimization problem. Luckily the sufficient number of parameters to be considered is at the order of the total number of sensors in the network as will be stated with the following lemma.

Lemma II.5: For K sensors, it is sufficient to consider $K+1$ segments for parallel- and $2K$ segments for serial networks to reach the globally optimum solution.

Proof: Suppose that an R provides minimum error probability for a network of K sensors with a parallel topology. Then, there are K possibly different operating points (excluding the TPs) on R. For any R, the spline approximation y with $K+1$ segments provides K possibly distinct points (connection points of segments) which are exactly the same with those chosen on y via determining the necessary parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The same idea applies to the serial networks, wherein for K sensors, there are $2K-1$ possibly different thresholds (see Section IV), hence the operating points. ■

The theory developed in this section will be used in the next sections in order to obtain performance bounds, i.e. answers to the questions stated in introduction.

III. General-Case vs Identical Sensor Decisions

Consider a sensor network with a parallel topology, as illustrated in Fig. 1. Here, even for independent observations, determining the optimal sensor decision rules has an exponential computational complexity in the total number of sensors K [5]. To simplify this problem, it may be assumed that the decisions of all sensors are identical. While this assumption can greatly reduce the problem's complexity, it may also lead to a performance loss. Contrary to common intuition, this performance loss can occur even for i.i.d. observations, although the reported losses are typically small [15]. In this section, the maximum of this performance loss (identical decisions vs. general case) will be determined for i.i.d. observations across all simple hypothesis tests.

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)$ be the sensor thresholds and λ_0 be the threshold of the fusion center. Since identical decisions $\phi_i = \phi_j, \forall i, j$ is compared to the most general-case, and the likelihood ratio test is known to be the optimum strategy to minimize the error probability [5, Proposition 2.4, Page 13, Eq. 2.15], [27], one can write the difference in error probability between the general-case and the identical sensor decisions as

$$\begin{aligned} W^{\tilde{p}p}(K) = \max_{P_0, P_1} \Bigg[&\min_{\substack{\lambda_0, \lambda_j = \lambda_k \\ j \ge 1, k \ge 1}} P_E\left(K; P_0, P_1, \boldsymbol{\phi}(\boldsymbol{\lambda}), \gamma(\lambda_0)\right) \\ &- \min_{\lambda_0, \boldsymbol{\lambda}} P_E\left(K; P_0, P_1, \boldsymbol{\phi}(\boldsymbol{\lambda}), \gamma(\lambda_0)\right) \Bigg], \end{aligned} \tag{4}$$

where each ϕ_k, $k \ge 1$ is the likelihood ratio test comparing the likelihood ratio to a threshold λ_k and γ is the likelihood ratio test of the fusion center considering the threshold λ_0. In the next step, (4) will be reformulated over the operating points of the ROC, removing the dependencies on $\boldsymbol{\phi}$ and γ. This can be done because for any $\boldsymbol{\phi}$, there are corresponding false alarm and miss detection probabilities on the ROC which can be given as

$$\begin{aligned} \alpha_k &= P_0(U_k = 1), \\ \beta_k &= P_1(U_k = 0), \end{aligned} \tag{5}$$

and the error minimizing γ leads to a $P_E(\cdot)$, which is a known function of α_k and β_k (see (7)). Let

$$p_0^K(\boldsymbol{u}) = \prod_{k=1}^{K} P_0(U_k = u_k),$$
$$p_1^K(\boldsymbol{u}) = \prod_{k=1}^{K} P_1(U_k = u_k), \tag{6}$$

denote the joint distributions, where $\boldsymbol{u} = (u_1, \ldots, u_K)$. Then, the optimum fusion rule for the parallel sensor network, which minimizes the error probability can be given as

$$P_E(K; p_0^K, p_1^K) = \frac{1}{2} \sum_{\boldsymbol{u}} \min\left(p_0^K(\boldsymbol{u}), p_1^K(\boldsymbol{u})\right), \tag{7}$$

where $P_E(\cdot)$ is reformulated in three variables as it can now be written only in terms of joint distributions and K [23], [24]. Since U_ks are Bernoulli distributed (see (5) and (6)), p_0^K and p_1^K can completely be written in terms of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$, respectively. Hence, considering Remark II.1 and (7), (4) can be reformulated as

$$W^{\tilde{p}p}(K) = \max_{R \in \mathscr{R}} \left[\min_{\substack{\alpha_j = \alpha_k \\ \beta_j = \beta_k}} P_E(K; \boldsymbol{\alpha}, \boldsymbol{\beta}) - \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} P_E(K; \boldsymbol{\alpha}, \boldsymbol{\beta}) \right], \tag{8}$$

where $(\alpha_k, \beta_k) \in R$ for all k[1]. Although significant simplification is possible analytically for the P_E of identical sensor decisions, i.e. for $\alpha_j = \alpha_k$ and $\beta_j = \beta_k$ [3], it is not possible for the general-case. This makes a single analytic solution parameterized by K impossible. Therefore, each K should be evaluated separately.

The objective in the following is to reformulate (8) entirely in terms of explicitly defined functions, $\boldsymbol{h}_0^K$ and $\boldsymbol{h}_1^K$, which replace P_E (see Theorem III.1 and (9)), and the parameter vectors of the spline approximation, $\boldsymbol{a}$ and $\boldsymbol{b}$, which replace $R \in \mathscr{R}$ (see Theorem III.2). This reformulation reduces the problem to an optimization task defined solely over the spline parameters.

Theorem III.1: Let A be the binary representation of the sequence of numbers $(0, 1, \ldots, 2^K)$, e.g., $A = (00, 01, 10, 11)$ for $K = 2$, where each bit is indexed by $k \geq 1$ in an increasing order from left to right. Let $\boldsymbol{h}_0^K$ be the vector defined over A by replacing each 0 by $1 - \alpha_k$ and each 1 by α_k, e.g. for $K = 2$,

$$\boldsymbol{h}_0^K = ((1-\alpha_1)(1-\alpha_2), (1-\alpha_1)\alpha_2, \alpha_1(1-\alpha_2), \alpha_1\alpha_2).$$

Furthermore let $\boldsymbol{h}_1^K$ similarly be defined by replacing each 0 by β_k and each 1 by $1 - \beta_k$. Then,

$$P_E(K; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\left(1 - \frac{\left|\boldsymbol{h}_0^K(\boldsymbol{\alpha}) - \boldsymbol{h}_1^K(\boldsymbol{\beta})\right|}{2}\right),$$

where $|\cdot|$ denotes the L_1 norm.

Proof: A proof of Theorem III.1 is given in Appendix C. ■

[1]Throughout the paper the simplified form of the error probability $P_E(K; \boldsymbol{\alpha}, \boldsymbol{\beta})$ will be considered in the derivations and analysis.

Using Theorem III.1, (8) can be rewritten as

$$W^{\tilde{p}p}(K) = \frac{1}{2} \max_{R \in \mathscr{R}} \left[\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \frac{\left|\boldsymbol{h}_0^K(\boldsymbol{\alpha}) - \boldsymbol{h}_1^K(\boldsymbol{\beta})\right|}{2} - \max_{\substack{\alpha_j = \alpha_k \\ \beta_j = \beta_k}} \frac{\left|\boldsymbol{h}_0^K(\boldsymbol{\alpha}) - \boldsymbol{h}_1^K(\boldsymbol{\beta})\right|}{2} \right]. \tag{9}$$

In the next step, the spline model given by Theorem II.1 is used to define $\boldsymbol{\beta}$ in terms of $\boldsymbol{\alpha}$. Together with Lemma II.5 the optimization problem can be simplified as in the following.

Theorem III.2: Let $\tilde{\boldsymbol{a}} = (a_{n_1}, \ldots, a_{n_K})$, $\tilde{\boldsymbol{r}} = (r_{n_1}, \ldots, r_{n_K})$, $\tilde{\boldsymbol{b}} = (b_{n_1}, \ldots, b_{n_K})$ with the indices $n_k \in \{0, \ldots, K-1\}$, where $\tilde{\boldsymbol{n}} = (n_1, \ldots, n_K)$. Furthermore, let $\mathbf{diag}(\cdot)$ denote the square diagonal matrix and let

$$D_p(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) = \frac{\left|\boldsymbol{h}_0^K(\tilde{\boldsymbol{a}}) - \boldsymbol{h}_1^K(\tilde{\boldsymbol{r}}\mathbf{diag}(\tilde{\boldsymbol{a}}) + \tilde{\boldsymbol{b}})\right|}{2}.$$

Then, (9) can be rewritten as

$$W^{\tilde{p}p}(K) = \frac{1}{2} \max_{\boldsymbol{a}, \boldsymbol{b}} \left[\max_{\tilde{\boldsymbol{n}}} D_p(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) - \max_{n_j = n_k} D_p(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) \right]. \tag{10}$$

Proof: A proof of Theorem III.2 is given in Appendix D. ■

Remark III.1: In order to facilitate parallel computing it maybe preferable to perform the inner maximizations over two vectors, one of which is built by evaluating D_p for each $\tilde{\boldsymbol{n}}$ (K^K cases in total), and the other for $n_j = n_k$ (K cases in total), in advance. Also note that D_p is dependent on $\tilde{\boldsymbol{r}}$ only through $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ by making use of (2).

A. Numerical Results

The problem of determining $W^{\tilde{p}p}$ is inherently non-convex, generally classified as NP-Hard, and algorithms with theoretically guaranteed solutions are in general not available [28]. This conclusion also applies to the optimization problems discussed in the subsequent sections. In order to solve (10), first it may be preferable to build two different vectors, which depend only on $\boldsymbol{a}$ and $\boldsymbol{b}$ by evaluating D_p for all possible $\tilde{\boldsymbol{n}}$ as mentioned in Remark III.1. This reduces the inner maximizations in (10) to finding the maximum of a vector. Next, the outer maximization needs to be evaluated with a non-linear optimization algorithm. In this work the Nelder-Mead algorithm is considered [28, p. 238]. The algorithm is executed up to 10.000 times for randomly chosen seeds to ensure that the solution found is at worst near-global optimum. When dimensionality of the problem i.e., the size of $\boldsymbol{a}$ and $\boldsymbol{b}$ increases, this approach becomes slow and inefficient for finding near-optimal solutions. To mitigate this, Nelder-Mead algorithm is run up to 400 times with different random seeds. The best three solutions are then selected to narrow down the constraint space of the parameters, see Theorem II.1. The same procedure is repeated until no better results are obtained. The error probabilities of the parallel network for the general-case P_E^p, the identical decision case $P_E^{\tilde{p}}$ and the difference in error probabilities $W^{\tilde{p}p}$ are determined using the

TABLE I
COMPARISON OF ERROR PROBABILITIES OF IDENTICAL SENSOR DECISIONS WITH THAT OF THE GENERAL-CASE IN PARALLEL SENSOR NETWORKS

K	2	3	4	5	6	7	8
$P_E^{\tilde{p}}$	0.2224	0.1867	0.1859	0.1807	0.1803	0.1867	0.1887
P_E^{p}	0.1853	0.1509	0.1499	0.1481	0.1501	0.1580	0.1608
$W^{\tilde{p}p}$	0.0370	0.0358	0.0360	0.0326	0.0302	0.0287	0.0279

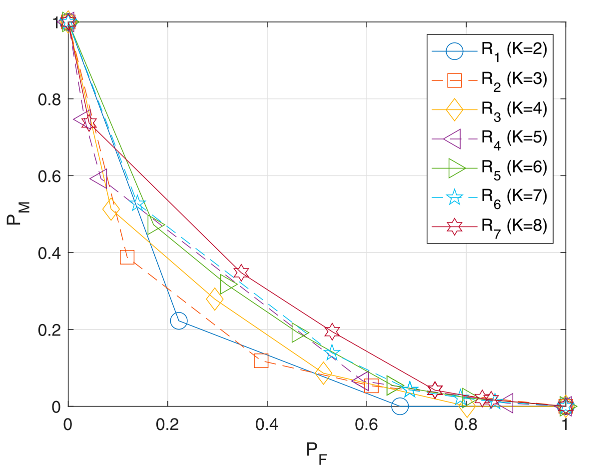

Fig. 5. The ROC curves leading to the maximum loss of detection performance between the general-case and identical sensor decisions in parallel sensor networks.

considered algorithms and presented in Table I. It is observed that the loss function $W^{\tilde{p}p}$ tends to decrease as the total number of sensors increases. The parameters leading to these results are illustrated by the ROC curves in Fig. 5.

B. Discussion

The results tabulated in Table I and depicted in Fig. 5 align with both the proposed and the asymptotic theory. This alignment is evident as it remains feasible to select a specific ROC and achieve any desired value of the error probability $P_E < 1/2$, regardless of the number of sensors within the sensor network. In Fig. 5 and also in the next sections in Figs. 7 and 8, it can be seen that as K increases ROC curves gradually come closer to the line $P_M = 1 - P_F$. This may be interpreted as the consequence of two ideas given below.

1) Increasing K can be considered a condition that eventually drives both P_E^p and $P_E^{\tilde{p}}$ to 0, thereby compelling $W^{\tilde{p}p}$ to be small. However, increasing K also augments the total number of points on the ROC curve, which in turn enhances the diversity of solutions. As the diversity of solutions leading to P_E^p is exponentially greater than those concerning $P_E^{\tilde{p}}$, the likelihood of encountering cases where P_E^p is significantly lower than $P_E^{\tilde{p}}$ becomes higher. This suggests the existence of solutions with potentially large $W^{\tilde{p}p}$.
2) It can be argued that eventually both P_E^p and $P_E^{\tilde{p}}$ are likely to decrease as K increases, even if the optimization process can select any ROC. This expectation arises from the necessity that, as K grows, to maintain a large $W^{\tilde{p}p}$, the ROC curve should be sufficiently close to $P_M = 1 - P_F$. However, choosing such an ROC may also not help to increase $W^{\tilde{p}p}$, if K is sufficiently large. This is because along the line $P_M = 1 - P_F$ we have $P_E^p = P_E^{\tilde{p}} = 0.5$ across all sensors.

IV. PARALLEL VS SERIAL SENSOR NETWORKS

Consider a serial network as illustrated by Fig. 2, wherein each sensor (associated to a decision rule ϕ_k) makes an observation z_k of its own, gives a certain decision $u_k = \phi_k(z_k)$ and passes it to its neighbor. The final decision is given by the last (Kth) sensor. Let $l = p_1/p_0$ denote the likelihood ratio function. Then, as mentioned before the likelihood ratio test

$$\begin{cases} l \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} t_1 & k=1 \\ l \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} t_k^{u_{k-1}} & k>1 \end{cases}, \quad t_k^{u_{k-1}} = \frac{P_0(U_{k-1}=u_{k-1})}{P_1(U_{k-1}=u_{k-1})} \tag{11}$$

is optimal in order to minimize the error probability, where $t_k^{u_{k-1}}$ is the threshold depending on the decision $u_{k-1} \in \{0,1\}$ of the previous ($(k-1)$th) sensor [4]. False alarm and miss detection probabilities yielding from the above given test are

$$\begin{aligned} P_F(k;P_0) &= P_0(U_{k-1}=0)P_0(l \geq t_k^0) \\ &\quad + P_0(U_{k-1}=1)P_0(l \geq t_k^1), \\ P_M(k;P_1) &= P_1(U_{k-1}=0)P_1(l < t_k^0) \\ &\quad + P_1(U_{k-1}=1)P_1(l < t_k^1). \end{aligned}$$

Let $\bar{\boldsymbol{\alpha}}_k = (\alpha_1, \alpha_2^0, \alpha_2^1, \dots, \alpha_k^1)$, $\bar{\boldsymbol{\beta}}_k = (\beta_1, \beta_2^0, \beta_2^1, \dots, \beta_k^1)$, where

$$\alpha_k^0 = P_0(l \geq t_k^0), \quad \alpha_k^1 = P_0(l \geq t_k^1),$$
$$\beta_k^0 = P_1(l < t_k^0), \quad \beta_k^1 = P_1(l < t_k^1).$$

Then, for the kth sensor, false alarm and miss detection probabilities can recursively be calculated as

$$\begin{aligned} P_F(k;\bar{\boldsymbol{\alpha}}_k) &= \alpha_k^0 + P_F(k-1;\bar{\boldsymbol{\alpha}}_{k-1})(\alpha_k^1 - \alpha_k^0), \quad k \geq 2, \\ P_M(k;\bar{\boldsymbol{\beta}}_k) &= \beta_k^1 + P_M(k-1;\bar{\boldsymbol{\beta}}_{k-1})(\beta_k^0 - \beta_k^1), \quad k \geq 2. \end{aligned}$$

where $P_F(1,\bar{\boldsymbol{\alpha}}_1) = \alpha_1$ and $P_M(1,\bar{\boldsymbol{\beta}}_1) = \beta_1$ as defined before. Hence, the overall error probability can be calculated as

$$P_E^s(K;\bar{\boldsymbol{\alpha}}_K,\bar{\boldsymbol{\beta}}_K) = \frac{P_F(K;\bar{\boldsymbol{\alpha}}_K) + P_M(K;\bar{\boldsymbol{\beta}}_K)}{2}.$$

A. Maximum of Performance Loss Between Serial and Parallel Networks

For the serial networks, there are a total of $2K-1$ thresholds to be determined, corresponding to $2K-1$ potentially different operating points on the ROC curve. This implies that the ROC, modeled by a linear spline approximation, must have at least $2K$ segments to accurately determine the minimum error probability. According to Lemma II.5, it also requires no more than $2K$ segments. In the following theorem, the maximum of the difference between the error probabilities of serial and parallel networks will be formulated by fixing R to have $2K$ segments for both networks. Note that a higher-order spline approximation can be reduced to any lower order through appropriate parameter selection.

Theorem IV.1: Let $\bar{\boldsymbol{a}} = (a_{\bar{n}_1}, \ldots, a_{\bar{n}_{2K-1}})$, $\bar{\boldsymbol{r}} = (r_{\bar{n}_1}, \ldots, r_{\bar{n}_{2K-1}})$ and $\bar{\boldsymbol{b}} = (b_{\bar{n}_1}, \ldots, b_{\bar{n}_{2K-1}})$ with the indices $\bar{n}_k \in \{-1, \ldots, 2K-1\}$, where $\bar{\boldsymbol{n}} = (\bar{n}_1, \ldots, \bar{n}_{2K-1})$. Furthermore, let

$$D_s(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{n}}) = 1 - 2P_E^s(K, \bar{\boldsymbol{a}}, \bar{\boldsymbol{r}}\mathbf{diag}(\bar{\boldsymbol{a}}) + \bar{\boldsymbol{b}}). \tag{12}$$

Then, the maximum of the difference between the error probabilities of the serial and parallel networks[2] can be given as

$$\begin{aligned} W^{sp}(K) &= \frac{1}{2} \max_{\boldsymbol{a},\boldsymbol{b}} \left[\max_{\tilde{\boldsymbol{n}}} D_p(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) - \max_{\bar{\boldsymbol{n}}} D_s(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{n}}) \right] \\ &\qquad s.t.\ \bar{n}_{2k-1} > \bar{n}_{2k-2} \\ W^{ps}(K) &= \frac{1}{2} \max_{\boldsymbol{a},\boldsymbol{b}} \left[\max_{\bar{\boldsymbol{n}}} D_s(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{n}}) - \max_{\tilde{\boldsymbol{n}}} D_p(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) \right] \\ &\qquad s.t.\ \bar{n}_{2k-1} > \bar{n}_{2k-2} \end{aligned} \tag{13}$$

where for $\tilde{\boldsymbol{n}} = (n_1, \ldots, n_K)$ we now have $n_k \in \{0, \ldots, 2K-2\}$ in order to allow the optimization to be performed for the parallel network over the same set of ROC curves.

Proof: A proof of Theorem IV.1 is given in Appendix E. ∎

Remark IV.1: To accurately determine W^{sp} and avoid the solver producing invalid results, it is necessary to consider $n_k \in \{-1, \ldots, 2K-1\}$. Although considering $(0,1)$ and $(1,0)$ (TPs) on R is equivalent to flipping a coin and they can never decrease P_E^p, interestingly, they do decrease P_E^s for some special cases of ROCs. An example will be provided in the next section to illustrate this phenomenon.

Remark IV.2: In the literature, it was conjectured that there may exist a K for which we have $W^{ps}(k) = 0$ for all $k > K$ [5, p. 338]. That is, the parallel network becomes better than the serial network in terms of minimum error probability over all simple hypothesis tests. Since this problem is exponential in complexity, it may be preferable to consider $K+1$ segments first instead of $2K$ segments. If such a K cannot be found by considering $K+1$ segments with an amount of computational power available, it is impossible to be found by $2K$ segments. Because the value of W^{ps} will always be smaller considering $K+1$ segments, as $K+1$ segments are sufficient to minimize the error probability of the parallel network but not necessarily that of the serial network.

B. Numerical Results

The maximum difference between the error probabilities of serial and parallel networks was determined by solving the optimization problems given in (13) using the same algorithm described in the previous section. The results are tabulated in Tables II and III, while the parameters solving the corresponding optimization problems are illustrated in Fig. 6. The tables indicate that W^{sp} is typically larger than W^{ps} for $K = 3$ and $K = 4$. For $K = 2$, it is already known, as shown in [5], the serial network is at least as good as the parallel network, which is consistent with the numerical result $W^{sp}(2) = 0$ found in this work.

[2] W^{sp} stands for the loss of detection performance, which is calculated by subtracting the error probability of the parallel network (p) from that of the serial network (s) and W^{ps} is defined similarly.

TABLE II
COMPARISON OF ERROR PROBABILITIES OF SERIAL TO PARALLEL NETWORKS

K	2	3	4
P_E^s	0.2942	0.1632	0.1703
P_E^p	0.2942	0.1151	0.1190
W^{sp}	0	0.0481	0.0513

TABLE III
COMPARISON OF ERROR PROBABILITIES OF PARALLEL TO SERIAL NETWORKS

K	2	3	4
P_E^p	0.1913	0.2174	0.2143
P_E^s	0.1462	0.1699	0.1663
W^{ps}	0.0451	0.0475	0.0480

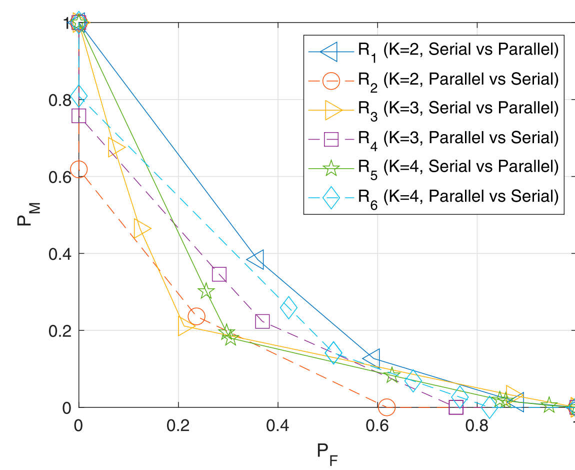

Fig. 6. The ROC curves leading to the maximum loss of detection performance between parallel and serial sensor networks.

Referring to Remark IV.2, we are unable to conclusively determine whether the parallel or serial network is superior, i.e., if there exists a K for which $W^{ps}(k) = 0$ or $W^{sp}(k) = 0$ for every $k > K$. The primary challenge in addressing this problem is the exponentially increasing computational complexity and memory requirements. Furthermore, even with substantial computational resources, identifying such a value for K may remain elusive, leaving this question potentially unresolved.

Given these computational challenges, the analysis did not extend beyond $K = 4$ sensors because, for the serial network, it is necessary to consider TPs to achieve a globally optimal solution. This requires the ROC curve to have at least $2K = 8$ segments, resulting in vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of dimension $2K + 1 = 9$. To underscore the complexity of the problem, it is worth noting that the program code written for $K = 4$ sensors to find W^{sp} occupies more than 750 megabytes of hard drive space. For the serial network, more than $317{,}000$ distinct functions of variable length must be evaluated at each iteration, requiring up to 64 gigabytes of memory and a solver capable of addressing a $4K - 2 = 14$-dimensional non-convex optimization problem (since TPs are constants and only considered in the inner maximization). The case $K=4$ for serial networks represents one of the most computationally demanding scenarios in this study, requiring significant but manageable hardware and processing resources. These requirements serve as a practical benchmark, indicating that the results presented in this paper can be reproduced using a modern consumer-level PC equipped with sufficient memory, storage, and processing power.

V. Centralized vs Decentralized Detection

Consider a distributed detection network given by Fig. 1, where each sensor relays its observation to the fusion center without any processing, or equivalently a fusion center which gives a decision based on K observations. This scheme corresponds to centralized detection, and the optimum decision making based on K observations, in both Bayesian and Neyman-Pearson senses, is the (log)-likelihood ratio test

$$\sum_{k=1}^{K} \log \frac{p_1(z_k)}{p_0(z_k)} = \sum_{k=1}^{K} \theta_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_0,$$

where $\lambda_0 = 0$ is chosen for the minimum error probability, (1), [3]. Consider the spline approximation given by Theorem II.1 and the corresponding density functions given by (3). Then, the density functions corresponding to $\theta_k = \log(p_1(z_k)/p_0(z_k))$ under the null and alternative hypothesis can respectively be written as

$$f_0(\theta_k) = \sum_{i=0}^{s} (a_{i-1} - a_i)\delta(\theta_k - \log(-r_i)),$$
$$f_1(\theta_k) = \sum_{i=0}^{s} -r_i(a_{i-1} - a_i)\delta(\theta_k - \log(-r_i)),$$

where δ is the dirac delta function. Since the test statistic is the sum of K independent r.v.s, the density corresponding to the test statistic under $\mathcal{H}_m$ is the K-fold convolution of f_m, $m \in \{0,1\}$ with itself. Noticing that there are at most $s+1$ non-zero elements (point masses) of f_0 and f_1 on $\mathbb{R}$, it is more convenient to transform and evaluate the problem at the discrete domain, where f_0 and f_1 take discrete values of the corresponding point masses at $\theta_{i_k} = \log(-r_{i_k})$ and δ is now the Kronecker delta function. As such, K-fold convolution of f_m can be calculated as

$$v_m^K(\kappa) = \sum_{i_1=0}^{s} \cdots \sum_{i_K=0}^{s} f_m(\theta_{i_1}) \cdots f_m(\theta_{i_K}) \delta\left(\sum_{k=1}^{K} \theta_{i_k} - \kappa\right),$$

and for $N = \dim(v_m^K)$ the minimum error probability can be found by

$$P_E^c(K; \boldsymbol{a}, \boldsymbol{b}) = \frac{1}{2} \sum_{n=1}^{N} \min\left(v_0^K(n), v_1^K(n)\right), \tag{14}$$

using an order preserving bijective mapping $\kappa \mapsto n$. Let $\boldsymbol{v}_0^K = (v_0^K(1), \ldots, v_0^K(N))$ and $\boldsymbol{v}_1^K = (v_1^K(1), \ldots, v_1^K(N))$. Then, we can write (14), cf. (7) and Theorem III.1, as

$$D_c(\boldsymbol{a}, \boldsymbol{b}) = 1 - 2P_E^c(K; \boldsymbol{a}, \boldsymbol{b}) = \frac{|\boldsymbol{v}_0^K - \boldsymbol{v}_1^K|}{2}$$

Remark V.1: For the centralized detection we do not have arguments of sufficiency such as given by Lemma II.5. Therefore, it is not clear how many segments would be sufficient to reach the maximum of the difference in error probability between centralized and serial or parallel networks. A general argument is that larger than some s, increasing the number of segments would either not increase the loss or the contribution of adding one more segment would be insignificant and can be omitted without loss of generality.

A. Centralized vs. Parallel Sensor Networks

Consider a decentralized detection network as in Fig. 1, where in one case there is (one-bit) quantization and in the other case, there is no quantization (equivalent to centralized detection). The maximum of the loss in detection performance between these two cases over all simple hypothesis tests can be formulated as

$$W^{pc}(K) = \frac{1}{2} \max_{\boldsymbol{a},\boldsymbol{b}} \left[D_c(\boldsymbol{a}, \boldsymbol{b}) - \max_{\boldsymbol{n}} D_p(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}})\right], \tag{15}$$

where for $\tilde{\boldsymbol{n}} = (n_1, \ldots, n_K)$ we now have $n_k \in \{0, \ldots, s\}$ for which s is large enough. Notice that in (15), D_c is defined on the same ROC model of D_p.

Remark V.2: There are many non-trivial examples, where $W^{pc}(K) = 0$. This means that the quantization does not cause any loss in detection performance. Here, the conditions, for which loss free detection is possible, will not be established. However, some examples will be provided in the next section for illustration.

B. Centralized vs. Serial Sensor Networks

Similar to the previous section, consider a serial network, as illustrated by Fig. 2, where in one case there is (one-bit) quantization and in the other case, there is no quantization (equivalent to centralized detection) and each observation is relayed to the next sensor. The maximum of the loss in detection performance (equivalently in error probability) between serial network and centralized network can be given as

$$W^{sc}(K) = \frac{1}{2} \max_{\boldsymbol{a},\boldsymbol{b}} \left[D_c(\boldsymbol{a}, \boldsymbol{b}) - \max_{\bar{\boldsymbol{n}}} D_s(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{n}})\right], \tag{16}$$

where for $\bar{\boldsymbol{n}} = (n_1, \ldots, n_{2K-1})$ we now have $n_k \in \{-1, \ldots, s\}$ for which s is large enough. Notice that here D_c is defined on the same ROC model of D_s, and in fact more than $s > 2K$ segments may be necessary for the optimization as stated in the previous section. Moreover, due to the nature of the serial network, again TPs should be included into the optimization.

C. Numerical Results

The maximums of the differences in error probabilities between centralized networks and their corresponding decentralized versions, as defined in (15) and (16), are calculated for each K and are tabulated in Tables IV and V. The ROC curves that solve the underlying optimization problems are illustrated in

TABLE IV
COMPARISON OF ERROR PROBABILITIES OF PARALLEL NETWORKS TO CENTRALIZED NETWORKS

K	2	3	4	5	6	7	8
P_E^p	0.2057	0.2303	0.2263	0.2333	0.2345	0.2327	0.2286
P_E^c	0.1422	0.1626	0.1450	0.1436	0.1391	0.1325	0.1249
W^{pc}	0.0635	0.0677	0.0812	0.0897	0.0954	0.1002	0.1037

TABLE V
COMPARISON OF ERROR PROBABILITIES OF SERIAL NETWORKS TO CENTRALIZED NETWORKS

K	2	3	4
P_E^s	0.1909	0.1845	0.1954
P_E^c	0.1459	0.1231	0.1269
W^{sc}	0.0451	0.0615	0.0685

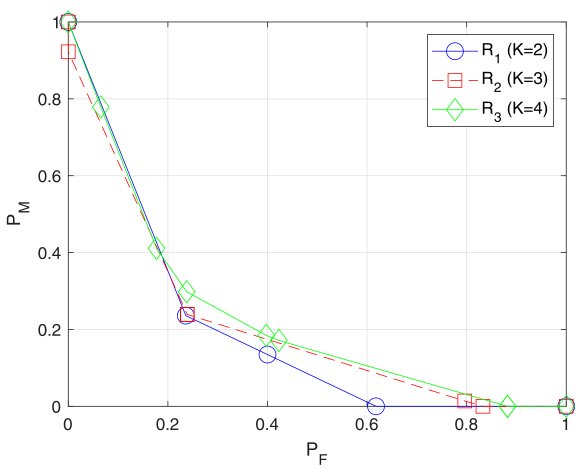

Fig. 7. The ROC curves leading to the maximum loss of detection performance due to quantization in serial sensor networks in comparison to centralized detection.

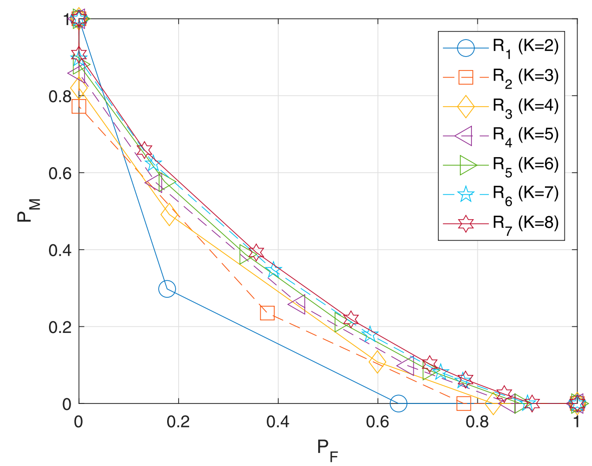

Fig. 8. The ROC curves leading to the maximum loss of detection performance due to quantization in parallel sensor networks in comparison to centralized detection.

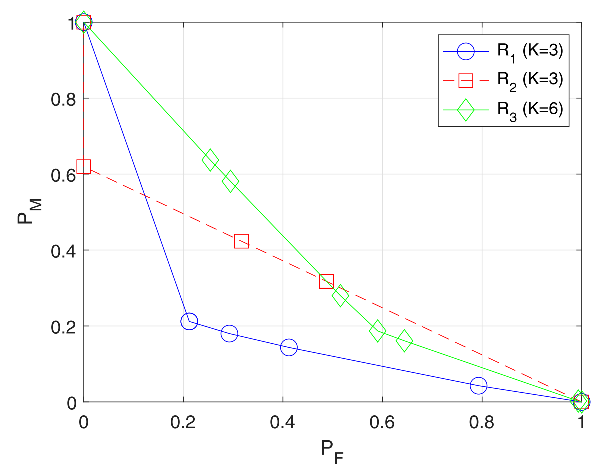

Fig. 9. The ROC curves R_1, R_2 and R_3 for special cases.

Figs. 7 and 8. For both cases, the difference in error probability increases as K increases, but the rate of increase decreases. Two distinct and somewhat counterintuitive phenomena were observed in the simulations.

Firstly, one might conjecture that the TPs should never be selected by sensors in either serial or parallel networks. These points correspond to always accepting one of the hypotheses, $\mathcal{H}_0$ or $\mathcal{H}_1$, for a single sensor ($K = 1$), resulting in an error probability of $P_E = 1/2$. While this conjecture holds for parallel networks — since the corresponding sensor could simply be excluded — it does not necessarily apply to serial networks. In serial configurations, the upper or lower thresholds of the sensors may necessitate sampling the ROC curve at the TPs (i.e., thresholds taking values of 0 or ∞) to achieve the minimum error probability.

Secondly, it might be assumed that quantization inevitably leads to information loss and, consequently, a reduction in detection performance. However, this assumption is not universally valid. Numerous non-trivial examples demonstrate that quantization in both parallel and serial networks can preserve detection performance without any loss. These phenomena are illustrated in Fig. 9, which presents three examples. In all these cases, the parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are manually selected, meaning that the outer optimization steps are not applied in the respective optimization problems.

In the first example, there are $K = 3$ sensors in both serial and parallel configurations, where each sensor samples its operating points (P_F, P_M) on R_1. The ROC curve R_1 consists of $2K = 6$ segments, the minimum required by the serial network. In this example, the parallel and serial networks result in different minimum error probabilities, with $P_E^p = P_E^c = 0.1157$ and $P_E^s = 0.1639$, respectively. However, the TPs are unconditionally required by the serial network to achieve this result. If the TPs are excluded from R_1, P_E^s increases to 0.2062. Notice that in this example, there is no loss of detection performance due to quantization for the parallel sensor network.

In the second example, the number of sensors and segments are identical to the previous example, but instead of R_1, R_2 is considered as the ROC curve. With this setup, the centralized network, parallel network, and serial network all yield the same error probability of $P_E^c = P_E^s = P_E^p = 0.1189$. However, excluding the TPs on R_2 results in P_E^s increasing to 0.2350. This shows that TPs are still necessary, even if all sensor networks result in the same error probability. Moreover, this is an example where no loss of detection performance is observed for both serial and parallel networks compared to centralized detection. Additionally, it is demonstrated that a point on R with $P_F \approx 0$ is not necessary, as in the previous example (see Appendix F), to highlight the importance of TPs and the absence of detection performance loss due to quantization.

In the final example, given by the ROC curve R_3, a parallel sensor network with $K = 6$ sensors (with $K + 1$ segments) is compared to its centralized counterpart. The minimum error probabilities for both the centralized and parallel networks are found to be the same $P_E^c = P_E^p = 0.2655$. This indicates that the previously obtained results are not restricted to $K = 3$ but also apply to larger networks, such as with $K = 6$. Details of the sampling points and error probabilities of Type I and II errors in all three examples are provided in Apppendix F.

For $K = 3$, if a parallel network achieves the minimum error probability with identical decisions where $\alpha_1 = \alpha_2 = \alpha_3 = 0$, the overall error probability of this network is given by $P_E^p = \beta_1^3/2$. A serial network can always achieve exactly the same error probability with the help of the $(1, 0)$ point, as in the example with R_2. Due to symmetry, if a parallel network achieves the minimum error probability with identical decisions where $\beta_1 = \beta_2 = \beta_3 = 0$, the overall error probability of this network is given by $P_E^p = \alpha_1^3/2$. In this case, a serial network can also achieve exactly the same error probability by considering the $(0, 1)$ point. For instance, in the example with R_2, both ϕ_2 and ϕ_3 use $(0, 1)$ and (a_4, c_4) instead of (a_4, c_4) and $(1, 0)$, as shown in Table IX in Apppendix F.

D. Discussion

Implications of the theory developed in this paper can be summarized as below

1) Decentralized detection without loss of detection performance due to quantization can be designed based on the theory presented in this paper. It turns out that lossless decentralized detection occurs only when the ROC curve has a special shape. This a-priori information may especially be useful to drive the actual ROC to one of these special curves.
2) If the distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$ are not simple, new estimations of the distributions are needed periodically. If the losses given in the paper are found not to be significant, identical sensor decisions may be preferred. Because it may be computationally heavy or power inefficient to calculate the optimum solutions (general-case) at each iteration.
3) The results given in Section IV provide guidelines i.e., maximum possible gains and losses, if the sensor network is to be redesigned with a different architecture. Similarly, Section V reveals maximum losses due to decentralization.
4) The problem studied in Section III can be recast as data transmission over a noisy binary memoryless channel or a failure detection problem as described in [13]. However, in contrary to the asymptotic results given in [13], the conclusions of this paper indicate that there is no single channel configuration or a particular device which solve the data transmission or fault detection problem, respectively, for any finite number of sensors.

VI. Conclusion

Theoretical bounds were established for three distinct decentralized detection problems, focusing on different aspects of performance loss:

- **Parallel Sensor Networks:** The maximum loss of detection performance was determined between two parallel sensor networks, where one network has no constraints and the other is restricted to identical sensor decisions.
- **Serial vs. Parallel Networks:** The maximum loss of detection performance was evaluated between serial and parallel networks, without imposing any constraints on decision-making strategies.
- **Quantization Effects:** The maximum loss of detection performance due to quantization in both parallel and serial networks was calculated by comparison to a centralized network with the same number of sensors.

To achieve the objectives, the ROC curve was modeled by a linear spline with potentially infinite number of segments. It was shown that the necessary number of segments for an optimal solution is finite and linearly dependent on the total number of sensors. It was also observed that considering only the endpoints of the linear spline segments is sufficient to find the optimal solutions.

Specific examples demonstrated that the quantization imposed by decentralization does not necessarily lead to a loss in detection performance. Additionally, it was shown that a serial network unconditionally requires the TPs on the ROC curve to achieve the minimum error probability.

The theory developed can be extended to hypothesis tests which are not necessarily simple. While error-free communication was assumed, the framework can be easily adapted to include error-prone communication channels, such as the binary symmetric channel. In this case, the statistics of the quantized data need to be adjusted by the channel's error probabilities, and the optimization process should incorporate these probabilities as additional parameters.

Appendix A
Proof of Theorem II.1

Proof: To satisfy the continuity each line segment y_n of the spline y intersects the previous segment y_{n-1} at a_{n-1}. This means $r_n a_{n-1} + b_n = r_{n-1} a_{n-1} + b_{n-1}$ giving

$$r_n = r_{n-1} + \frac{b_{n-1} - b_n}{a_{n-1}}.$$

■

Appendix B
Proof of Corollary II.3

Proof: We know that

$$c_n = r_n a_n + b_n, \tag{17}$$

and additionally by using (2), we have

$$c_n = \left(r_{n-1} + \frac{b_{n-1} - b_n}{a_{n-1}} \right) a_n + b_n. \tag{18}$$

Solving (17) and (18) (after $n := n+1$ in (18)) for r_n we get

$$r_n = \frac{c_{n+1} - c_n}{a_{n+1} - a_n} - \frac{b_n - b_{n+1}}{a_n}.$$

Writing r_n back in (17), applying $n := n-1$ and solving for c_n gives

$$c_n = \frac{a_n}{a_{n-1}} c_{n-1} + b_n \left(\frac{a_{n-1} - a_n}{a_{n-1}} \right).$$

Appendix C
Proof of Theorem III.1

Proof: Using the identity

$$\min(a,b) = \frac{a+b}{2} - \frac{|a-b|}{2}$$

we have

$$P_E(K; p_0^K, p_1^K) = \frac{1}{2} \sum_{\boldsymbol{u}} \frac{p_0^K(\boldsymbol{u}) + p_1^K(\boldsymbol{u})}{2} - \frac{|p_0^K(\boldsymbol{u}) - p_1^K(\boldsymbol{u})|}{2}$$

where the sum over the first term adds up to 1 and the second term can be rewritten in terms of $\boldsymbol{h}_0^K$ and $\boldsymbol{h}_1^K$ as in Theorem III.1. ∎

Appendix D
Proof of Theorem III.2

Proof: From Theorem II.1 and Corollary II.2 one can consider $\boldsymbol{\beta} = \tilde{\boldsymbol{r}}\mathbf{diag}(\boldsymbol{\alpha}) + \tilde{\boldsymbol{b}}$ in (9). This gives

$$W^{\tilde{p}p}(K) = \frac{1}{2} \lim_{s\to\infty} \max_{\boldsymbol{\alpha},\boldsymbol{b}} \times \left[\max_{\tilde{\boldsymbol{n}}} D_p(\boldsymbol{\alpha}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) - \max_{n_j = n_k} D_p(\boldsymbol{\alpha}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}) \right]$$

where $n_k \in \{0, \ldots, s\}$, as each (α_k, β_k) can be on any of $s+1$ line segments. Applying Lemma II.5 in (10) yields $s = K$ and each $\boldsymbol{\alpha}$ to be replaced by $\tilde{\boldsymbol{a}}$ since the operating points of R are the connection points of the spline segments, see Fig. 3. Hence, the problem can be redefined as given in (10), where we now have $n_k \in \{0, \ldots, K-1\}$, since only K connection points are required for optimization excluding the points $(0,1)$ and $(1,0)$. ∎

Appendix E
Proof of Theorem IV.1

Proof: Minimization of P_E^s or P_E^p corresponds to maximization of D_s or D_p, respectively. The analyses made in Section III for the parallel networks apply directly to the serial networks. Hence, the optimization problems can be written as in (13), where the condition $\bar{n}_{2k-1} > \bar{n}_{2k-2}$, which simplifies the optimization significantly for the serial network stems from the fact that, cf. (11), we have

$$\begin{aligned} & 1 > P_F(k;\cdot) + P_M(k;\cdot) \\ & \Longrightarrow t_k^0 = \frac{1 - P_F(k-1;\cdot)}{P_M(k-1;\cdot)} > t_k^1 = \frac{P_F(k-1;\cdot)}{1 - P_M(k-1;\cdot)} \\ & \Longrightarrow \alpha_k^1 > \alpha_k^0 \text{ and } \beta_k^0 > \beta_k^1 \Longrightarrow \bar{n}_{2k-1} > \bar{n}_{2k-2}, \end{aligned}$$

and the indices $\bar{n}_k = -1$ and $\bar{n}_k = 2K-1$ are required to include $(0,1)$ and $(1,0)$ points on the ROC for the optimization. ∎

TABLE VI
Comparison of Error Probabilities of Serial, Parallel and Centralized Networks for R_1

Error Probabilities	P_F	P_M	P_E
Parallel Network	0.1157	0.1157	0.1157
Serial Network	0.2474	0.0803	0.1639
Serial Network Without TPs	0.2544	0.1580	0.2062
Centralized Network	0.1157	0.1157	0.1157

TABLE VII
Parameters of Serial, Parallel and Centralized Networks for R_1

Sensors	ϕ_1		ϕ_2				ϕ_3			
Par. Net. Par.	α_1	β_1	α_2		β_2		α_3		β_3	
Parameters	a_3	c_3	a_3		c_3		a_3		c_3	
Ser. Net. Par.	α_1	β_1	α_2^0	β_2^0	α_2^1	β_2^1	α_3^0	β_3^0	α_3^1	β_3^1
Par. W. TPs	a_3	c_3	0	1	a_3	c_3	a_4	c_4	1	0
Par. W.O. TPs	a_3	c_3	a_4	c_4	a_3	c_3	a_4	c_4	a_1	c_1

TABLE VIII
Comparison of Error Probabilities of Serial, Parallel and Centralized Networks for R_2

Error Probabilities	P_F	P_M	P_E
Parallel Network	0	0.2377	0.1189
Serial Network	0	0.2377	0.1189
Serial Network Without TPs	0	0.4701	0.2350
Centralized Network	0	0.2377	0.1189

Appendix F
Parameters and Error Probabilities of Special Cases Given in Section V-C

Error probabilities of different networks for the first example (with R_1) are given in Table VI. The parameters yielding these results are tabulated in Table VII, where $(a_1, c_1) = (0.4120, 0.2576)$, $(a_3, c_3) = (0.2420, 0.2001)$, $(a_4, c_4) = (0.2120, 0.1435)$. In Table VII, and also in Table IX the values of the parameters are listed under the given parameters. For the serial network, there are two different results, first with and the second without the TPs. Similarly, the results for the second example (with R_2) are tabulated in Tables VIII-IX, where, $(a_0, c_0) = (0.4870, 0.3178)$, $(a_4, c_4) = (0, 0.6195)$. In the third example, The error probability of the centralized detection and that of the parallel network are the same $P_E^c = P_E^p = 0.2655$, where $P_F^c = P_F^p = 0.2184$ and $P_M^c = P_M^p = 0.3126$. The

TABLE IX
PARAMETERS OF SERIAL, PARALLEL AND CENTRALIZED NETWORKS FOR R_2

Sensors	ϕ_1		ϕ_2				ϕ_3			
Par. Net. Par.	α_1	β_1	α_2		β_2		α_3		β_3	
Parameters	a_4	c_4	a_4		c_4		a_4		c_4	
Ser. Net. Par.	α_1	β_1	α_2^0	β_2^0	α_2^1	β_2^1	α_3^0	β_3^0	α_3^1	β_3^1
Par. W. TPs	a_4	c_4	a_4	c_4	1	0	a_4	c_4	1	0
Par. W.O. TPs	a_4	c_4	a_4	c_4	a_0	c_0	a_4	c_4	a_0	c_0

parallel network achieves this result with $K = 6$ sensors with identical decisions all sampling R_3 at (a_2, c_2), where $a_2 = 0.5903$ and $c_2 = 0.1870$. ■

REFERENCES

[1] J.-F. Chamberland and V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas in Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[2] J. Luo and Z. Liu, "Serial distributed detection for wireless sensor networks with sensor failure," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–9, Jul. 2017.

[3] P. K. Varshney, *Distributed Detection and Data Fusion*, 1st ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1996.

[4] R. Viswanathan, S. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, Jul. 1988.

[5] J. N. Tsitsiklis, "Decentralized detection," in *Proc. Adv. Statist. Signal Process.,* Greenwich, CT, USA: JAI Press, 1993, pp. 297–344.

[6] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

[7] S. Thomopoulos, R. Viswanathan, and D. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 5, pp. 761–765, Sep. 1989.

[8] G. Gül and A. M. Zoubir, "Theoretical bounds in minimax decentralized hypothesis testing," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 15–26, Jan. 2017.

[9] O. A. Hanna, X. Li, C. Fragouli, and S. Diggavi, "Can we break the dependency in distributed detection?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2022, pp. 2720–2725.

[10] P. A. Traganitis and G. B. Giannakis, "Bayesian constrained decision fusion," in *Proc. IEEE 22nd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2021, pp. 166–170.

[11] B. Chen, R. Jiang, T. Kasetkasem, and P. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3454–3458, Dec. 2004.

[12] D. Ciuonzo, P. S. Rossi, and S. Dey, "Massive MIMO channel-aware decision fusion," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 604–619, Feb. 2015.

[13] J. N. Tsitsiklis, "On threshold rules in decentralized detection," in *Proc. 25th IEEE Conf. Decis. Control*, 1986, pp. 232–236.

[14] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, pp. 167–182, 1988.

[15] M. Cherikh and P. B. Kantor, "Counterexamples in distributed detection," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 162–165, Jan. 1992.

[16] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 355–361, Mar. 1992.

[17] W. W. I. Jr. and J. N. Tsitsiklis, "Some properties of optimal thresholds in decentralized detection," in *Sensor Fusion IV: Control Paradigms and Data Structures*, P. S. Schenker, Ed., vol. 1611, International Society for Optics and Photonics, Bellingham, WA, USA: SPIE, 1992, pp. 434–442.

[18] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sums of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 409–507, 1952.

[19] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, no. 3, pp. 828–835, 1969.

[20] J. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," in *Proc. 29th IEEE Conf. Decis. Control*, 1990, vol.1, pp. 246–251.

[21] K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis, "On learning with finite memory," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6859–6872, 2013.

[22] G. Gül and M. Baßler, "Fast multilevel quantization for distributed detection based on gaussian approximation," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 2433–2437.

[23] G. Gül and M. Baßler, "Scalable multilevel quantization for distributed detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2021, pp. 5200–5204.

[24] C. Altay and H. Delic, "Optimal quantization intervals in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 38–48, Feb. 2016.

[25] L. K. Ekchian and R. R. Tenney, "Detection networks," in *Proc. 21st IEEE Conf. Decis. Control*, 1982, pp. 686–691.

[26] W. Rudin, *Principles of Mathematical Analysis*. New York-Toronto-London: McGraw-Hill Book Company, Inc., 1953.

[27] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors I. Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[28] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

Gökhan Gül received the B.Sc. degree (with first-class Hons.) in electronic engineering from Uludağ University, Bursa, Turkey, in 2005, and the M.Sc. degree in digital communications from Christian-Albrechts-Universität zu Kiel, Germany, in 2009. His master's thesis was completed at the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany. He received the Dr.-Ing. degree (with distinction) from the Technische Universität Darmstadt, Germany, in 2015. He is the Head of the Statistics Group with the Preventive Cardiology and Preventive Medicine, Department of Cardiology, University Medical Center Mainz, Johannes Gutenberg University Mainz, Germany. From 2010 to 2011, he was a Visiting Researcher with Queen's University Belfast, U.K., where he contributed to steganalysis research and received the second-place award in the international competition on steganalysis (HUGO), in 2011. That same year, he joined the Signal Processing Group, Technische Universität Darmstadt, Germany, as a Research Associate. His doctoral research focused on robust and distributed hypothesis testing. He has an extensive research background in statistics, probability theory, signal processing, and machine learning. He has held Research Positions with Fraunhofer IMM and Fraunhofer FHR, where he worked on the application of machine learning, deep learning, and signal processing to various domains, including flow cytometry and radar systems. His expertise includes adaptive and dynamic algorithm development for target detection, tracking, and cognition. His research interests span detection and estimation theory, robust statistics, sensor fusion, machine learning, and applied statistical methods in medical research.