Minimax Robust Decentralized Hypothesis Testing for Parallel Sensor Networks

Gökhan Gül, *Member, IEEE*

***Abstract*—Decentralized detection is studied for parallel-access sensor networks, where sensor statistics are not known completely and are assumed to follow distribution functions which belong to known uncertainty classes. It is shown that there exist no minimax robust tests over the deterministic decision rules for the uncertainty classes built with respect to the Kullback-Leibler (KL)-divergence. For the KL-divergence as well as for some other uncertainty classes, such as the α-divergences, the joint stochastic boundedness property, which is the fundamental rule to prove minimax robustness, fails to hold. This raises a natural question whether a solution to minimax robust decentralized detection problem can be given if the uncertainty classes do not own this property. An answer to this question has been shown to be positive, which leads to a generalization of an existing work. Moreover, it is shown that for Huber's extended uncertainty classes quantization functions at the sensors are not required to be monotone in order to claim minimax robustness. A possible generalization of the theory to minimax- and Neyman-Pearson formulations, repeated observations, imperfect reporting channels and different network topologies have been discussed. Simulation examples are provided considering clipped- and censored likelihood ratio tests.**

***Index Terms*—Robustness, decentralized detection, data fusion, sensor networks, minimax robust hypothesis testing.**

I. Introduction

In SIMPLE binary hypothesis testing the design of optimum decision rules requires the exact knowledge of the conditional probability distributions under each hypothesis [1]. However, in practice, complete knowledge of the observation statistics is often not available, such as occurs with the presence of outliers or due to model mismatch. In these cases, a reasonable approach is to represent each hypothesis by a set or class of distributions and determine the optimum decision rule via maximizing the worst case performance. The tests optimizing the worst case performance are called the minimax robust tests, the worst case distributions are called the least favorable distributions (LFDs) and the related set of distributions are called the uncertainty sets or more commonly uncertainty classes. The minimax robust tests have a nice property of guaranteeing a certain level of detection performance irrespective of the actual state of the observation statistics. Because of this property, they are often essential for the design of systems that have to function reliably in harsh environments or in environments which cannot be modeled accurately [2].

A. Related Work

The first and probably the most fundamental work in robust hypothesis testing was developed by Huber in 1965 [3]. Huber showed that the minimax robust test for the ϵ-contaminated classes of distributions and the uncertainty classes with respect to the total variation distance were clipped likelihood ratio tests, where the likelihood ratio was obtained by the ratio of so called least favorable distributions from the respective uncertainty classes. In his follow-up work, Huber extended the number of class of distributions to five, from where the same conclusions could be made [4]. The most general classes of distributions for which the clipped likelihood ratio is the minimax robust test are the two alternating capacities [5].

Huber's classes of distributions are well known to be to able to model uncertainties due to outliers. However, modeling errors, which are the other source of uncertainty in signal processing applications, cannot be well modeled by using Huber's techniques [2]. Dabak and Johnson, for the asymptotic case [6], and later Levy for the single sample case [7] suggested that for modeling errors instead of Huber's uncertainty classes the subsets of topological spaces which are created with respect to smooth distances -such as the KL-divergence- are more suitable. The results of [7] are applicable if the nominal density functions under each hypothesis are symmetric, the robustness parameters are equal and the nominal likelihood ratio function is monotone. In [8], the results of [7] are generalized to the case, where no assumption was necessary to be imposed on the choice of nominal distributions. The most general classes of distributions for modeling errors are derived by Gül and Zoubir [9] considering the set of distances -α-divergences- and removing the constraint of equal a-priori probabilities imposed in [7] and [8].

Robust hypothesis testing extended to multiple sensors provides not only reliability but also high detection accuracy [10]. The earliest study in this field was conducted by Geroniotis, who considered a distributed detection network without a fusion center (DDN-WoF) for a fixed sample size and a sequential discrete time robust detection for two sensors [11]. In [12], Geraniotis and Chau studied the robustness of distributed detection network with a fusion center (DDN-WF) and sequential data fusion where the emphasis was on the selection of robust fusion rules. In their recent work [13], Geraniotis and Chau generalized most of their results presented in [12].

Manuscript received September 10, 2019; revised April 27, 2020; accepted August 19, 2020. Date of publication October 2, 2020; date of current version December 21, 2020.

The author is with the Fraunhofer Institute for Microengineering and Microsystems (IMM), 55129 Mainz, Germany (e-mail: goekhan.guel@imm.fraunhofer.de).

Communicated by O. Johnson, Associate Editor for Probability and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2020.3028451.

Digital Object Identifier 10.1109/TIT.2020.3028451

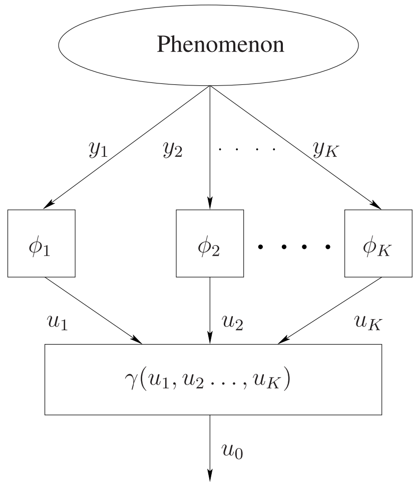

Fig. 1. Distributed detection network with K decision makers, each represented by the decision rule ϕ_k, and a fusion center associated with the fusion rule γ.

All Huber's classes of distributions satisfy joint stochastic boundedness property. Based on this observation it was proven in [14] that for jointly stochastically bounded classes of distributions, there exist least favorable distributions for DDN-WF if the individual sensors employ robust tests. Moreover, the authors formalized necessary conditions that need to be satisfied by the cost assignment procedure for DDN-WoF. The results derived in [14] are currently the state-of-the-art and generalize the DDN-WoF-results of Geroniotis [11] to a network of more than two sensors and to more general cost functions. Furthermore, the results of [14] also generalize the DDN-WF-results of Geraniotis and Chau [12], [13] to non-Bayesian formulation, non-binary decisions, non-identical sensor decisions and non-asymptotic case, both in terms of the number of sensors as well as the number of observations.

All aforementioned robust decentralized detection schemes are based on uncertainty models, which do not assume any specific shape for the actual distributions lying in the uncertainty classes. If such an assumption can be made, uncertainty classes can also be constructed parametrically and the generalized likelihood ratio test can be used to solve the resulting composite hypothesis testing problem [15]. In general, uncertainty models can also be combined with different network topologies, for instance tandem sensor networks where asymptotic analysis is of interest [16], [17]. Other variants of this work focus on the application of the earlier results to scenarios with constraints such as power [18], communication rate [19], or local optimality [20].

B. Summary of the Paper and Its Contributions

In this paper, binary minimax decentralized detection is studied for parallel sensor networks which consist of a finite number of sensors and a fusion center as illustrated in Figure 1. Each sensor in the sensor network collects a finite number of samples characterizing either the null or the alternative hypothesis and gives a decision which is possibly multilevel. The distribution of data samples for every sensor is not known exactly and characterized by uncertainty classes. The motivation of this paper is due to Huber and Strassen [5]. They state that if the classes of distributions are constructed such that every distribution in the uncertainty class is absolutely continuous with respect to a dominating measure and the domain of the uncertainty classes are uncountably infinite, the stochastic boundedness property may fail. This property specifies minimax robustness in all Huber's papers [3]–[5] and it is a necessary condition for the design of minimax robust decentralized detection in [14]. However, it is not known whether minimax robust decentralized detection is possible if at least one, some, or all sensors of the sensor network do not own this property. A positive answer to these questions implies that the minimax robust schemes which do not own stochastic boundedness property, e.g., [7], [8] can be extended to multiple sensors in a straightforward manner. In this regard, the following contributions are made in connection with their relation to prior works.

1) It is proven explicitly that there exist no minimax robust tests over deterministic decision rules for the uncertainty classes constructed with respect to the Kullback-Leibler divergence. Such tests, however, exist over the randomized decision rules, see [8].
2) It is proven that even if the joint stochastic property fails for every sensor in the sensor network, it is still possible to design a minimax robust decentralized detection scheme. This can be done without resorting to monotone sensor quantization rules (functions), which is another necessary condition in [14]. Therefore, the proposed scheme generalizes and includes the work of Veeravalli *et al.* [14] as a special case.
3) The results are generalized to minimax- and Neyman-Pearson formulations, repeated observations, imperfect reporting channels and different network topologies.

Based on these contributions it is now possible to link minimax robust hypothesis testing for modeling errors [7], [8] with decentralized detection. This was previously possible only for uncertainty models, which are suitable for outliers [14].

C. Outline of the Paper

The organization of this paper is as follows. In Section II, the existence of minimax robust tests are studied. In Section III, the problem definition is made. In Section IV the theory behind the solution of minimax robust decentralized detection problem is introduced. In Section V uncertainty classes are defined explicitly and their connection with minimax robust decentralized detection is summarized. In Section VI generalization of the theory to minimax and Neyman-Pearson formulations, repeated observations, imperfect reporting channels and other network topologies are discussed. In Section VII two specific examples are given and finally in Section VIII the paper is concluded.

D. Notations

The following notations are applied throughout the paper. Upper case symbols are used for probability distributions and random variables, and the lower case symbols of the probability distributions and random variables denote the density functions and observations, respectively. Boldface symbols are used for vectors of random variables, or sets of- distributions or functions. The hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are associated with the nominal probability measures P_0 and P_1, whereas the corresponding actual distributions are denoted by Q_0 and Q_1. The sets of probability distributions are denoted by $\mathscr{P}_0$ and $\mathscr{P}_1$. Every probability measure, e.g., $P[\cdot]$, is associated with its distribution function $P(\cdot)$ i.e., $P(y) = P[Y \leq y]$ for the random variable (r.v.) Y and the observation y. The notation $\mathcal{N}(\mu, \sigma^2)$ is used for the Gaussian distribution with mean μ and variance σ^2, whereas $\mathcal{G}(\kappa, \mu, \sigma)$ stands for the generalized Gaussian distribution with shape κ, location μ and scale σ. The notation $(\cdot)^{-1}$ is used for the inverse function and $\hat{(\cdot)}$ indicates the least favorable distribution, e.g., $\hat{U}_k^j$ implies that the r.v. U_k follows the LFD $\hat{Q}_j \in \mathscr{P}_j$.

II. Existence of Minimax Robust Tests

In this section basic theory to be used in the next sections will be introduced and the existence and non-existence of minimax robust tests will be examplified. Definitions related to Figure 1 will be detailed in the next section. Consider the following remark, lemmas and the definition.

Remark II.1: Let X and Y be two random variables defined on the same measurable space $(\Omega, \mathcal{A})$, having cumulative distribution functions P_X and P_Y, respectively. X is called stochastically larger than Y, i.e., $X \succeq Y$, if $P_Y(x) \geq P_X(x)$ for all x.

Lemma II.1: $X \succeq Y$ iff $\upsilon(X) \succeq \upsilon(Y)$ for every non-decreasing υ.

Proof of Lemma II.1 is simple and can be found, for example, in [21].

Lemma II.2: Let X_1, X_2, Y_1 and Y_2 be four random variables on $(\Omega, \mathcal{A})$, out of which X_1 and X_2, and Y_1 and Y_2 are independent. If $X_1 \succeq Y_1$ and $X_2 \succeq Y_2$, then $X_1 + X_2 \succeq Y_1 + Y_2$.

Proof: From Remark II.1, we have $P_{Y_1}(x) \geq P_{X_1}(x)$ and $P_{Y_2}(x) \geq P_{X_2}(x)$ for all x. Hence,

$$\begin{aligned} P_{Y_1+Y_2}(z) &= \int_{-\infty}^{+\infty} P_{Y_1}(z-x) dP_{Y_2}(x) \\ &\geq \int_{-\infty}^{+\infty} P_{X_1}(z-x) dP_{Y_2}(x) \\ &= \iint_{x+y\leq z} dP_{X_1}(x) dP_{Y_2}(y) \\ &= \int_{-\infty}^{+\infty} P_{Y_2}(z-y) dP_{X_1}(y) \\ &\geq \int_{-\infty}^{+\infty} P_{X_2}(z-y) dP_{X_1}(y) = P_{X_1+X_2}(z). \quad (1) \end{aligned}$$

■

Definition II.1 (Joint Stochastic Boundedness): A pair of class of distributions $(\mathscr{P}_0, \mathscr{P}_1)$ defined on a measurable space $(\Omega, \mathcal{A})$ are called to be jointly stochastically bounded by $(\hat{Q}_0, \hat{Q}_1)$, if a pair of distributions $(\hat{Q}_0, \hat{Q}_1)$ exists such that

$$Q_0[\hat{l}(Y) \leq t] \geq \hat{Q}_0[\hat{l}(Y) \leq t], \ \ \forall t \in \mathbb{R}_{\geq 0}, \forall Q_0 \in \mathscr{P}_0, \quad (2)$$
$$Q_1[\hat{l}(Y) \leq t] \leq \hat{Q}_1[\hat{l}(Y) \leq t], \ \ \forall t \in \mathbb{R}_{\geq 0}, \forall Q_1 \in \mathscr{P}_1, \quad (3)$$

where $\hat{l} = d\hat{Q}_1/d\hat{Q}_0$ is the robust likelihood ratio function.

Minimax robust tests over deterministic decision rules may or may not exist. In the sequel, two examples of uncertainty classes are provided, where the stochastic boundedness property holds and fails, respectively. For both examples, every distribution in the uncertainty classes is absolutely continuous with respect to the related nominal measure c.f. [5, p. 261].

Example II.3: Let the uncertainty classes be

$$\begin{aligned} \mathscr{P}_0 &= \{Q_0 : Q_0 = \mathcal{G}(\kappa, \mu_0, \sigma),\ \mu_0 \in [t_l^0, t_u^0]\}, \\ \mathscr{P}_1 &= \{Q_1 : Q_1 = \mathcal{G}(\kappa, , \mu_1, \sigma),\ \mu_1 \in [t_l^1, t_u^1]\}, \end{aligned} \quad (4)$$

where $\mathcal{G}$ is the generalized Gaussian distribution with shape $\kappa > 0$, location $\mu_j \in \mathbb{R}$ and scale $\sigma > 0$ for $t_l^1 > t_u^0$. Then, there exist least favorable distributions $\hat{Q}_0 = \mathcal{G}(\kappa, t_u^0, \sigma)$ and $\hat{Q}_1 = \mathcal{G}(\kappa, t_l^1, \sigma)$ which satisfy the joint stochastic boundedness property given by Definition II.1.

Proof: In order to prove (2), it is sufficient to show that $Q_0[\log \hat{l}(Y) \leq t]$ is decreasing in μ_0 for every t. The probability density function of the generalized Gaussian distribution is given by

$$q_m(y) = \frac{\kappa^{-1/\kappa}}{2\sigma\Gamma(1+\frac{1}{\kappa})} \exp(-(|y-\mu_m|/\sigma)^\kappa/\kappa),\ m \in \{0,1\}. \quad (5)$$

The corresponding log-likelihood ratio function

$$t = \log \hat{l}(y) = \frac{1}{\kappa\sigma^\kappa}\left(|y - t_u^0|^\kappa - |y - t_l^1|^\kappa\right) \quad (6)$$

is strictly increasing in y. Hence, for $y(t) = \hat{l}^{-1}(\exp(t))$, we have

$$\begin{aligned} f(\mu_0, t) &= Q_0[\log \hat{l}(Y) \leq t] = Q_0[Y \leq y(t)] \\ &= Q_0(y(t)) = \mathbf{1}_{y(t)\geq\mu_0}(y(t)) \\ &\quad + \operatorname{sign}(\mu_0 - y(t)) \frac{1}{2} \Gamma_r\left(\frac{1}{\kappa}, \frac{\left(\frac{|\mu_0 - y(t)|}{\sigma}\right)^\kappa}{\kappa}\right), \end{aligned} \quad (7)$$

where $\mathbf{1}_{(\cdot)}$ is the indicator function and $\Gamma_r(\cdot)$ is the regularized gamma function. Since

$$\frac{\partial f(\mu_0, t)}{\partial \mu_0} = -\frac{\exp\left(\frac{-\left(\frac{|y(t)-\mu_0|}{\sigma}\right)^\kappa}{\kappa}\right)\kappa\left(\frac{\left(\frac{|y(t)-\mu_0|}{\sigma}\right)^\kappa}{\kappa}\right)^{\frac{1}{\kappa}}}{2\Gamma\left(\frac{1}{\kappa}\right)|y(t)-\mu_0|} < 0 \quad (8)$$

for every t, the inequality (2) holds. The proof for (3) is similar and is omitted. ■

Example II.4: The second example will be stated with the following proposition.

Proposition II.5: Let the uncertainty classes be

$$\mathscr{P}_j = \{Q_j : D(Q_j, P_j) \leq \epsilon_j\}, \quad j \in \{0,1\}, \quad (9)$$

where

$$D(Q_j, P_j) = \int_\Omega \ln(dQ_j/dP_j) \mathrm{d}Q_j, \quad j \in \{0,1\}$$

is the KL-divergence. Then, there exists no pair of LFDs $(\hat{Q}_0, \hat{Q}_1)$ which satisfy the joint stochastic boundedness property.

A proof of Proposition II.5 is given in Appendix A.

Remark II.2: The proposition is valid for any choice of (ϵ_0, ϵ_1) so long as a minimax robust test exists, i.e., the uncertainty classes are disjoint. Existence of a minimax robust test can be guaranteed if the robustness parameters are chosen sufficiently small, depending on the chosen distinct nominal distributions. A sketchy explanation of the choice of nominals and the maximum allowable robustness parameters can be seen in [6, Fig. 3], [22]. It is worth noticing that a minimax robust test for the KL-divergence exists over randomized decision rules and the corresponding test is unique [8]. Since this test is not equivalent to any deterministic likelihood ratio test, it does not satisfy (2) and (3) either.

III. Problem Definition

Consider a decentralized detection network with a parallel topology as shown in Figure 1. There are K decision makers (physically sensors), each represented by the decision rules $\phi_k : \Omega_k \to S_k \subset \mathbb{N}$ observing a certain phenomenon, and a fusion center. Every random variable Y_k corresponding to the observation y_k takes values on a measurable space $(\Omega_k, \mathcal{A}_k)$, where Ω_k is a continuous set, e.g. $\Omega_1 = \mathbb{R}^n$, $n \in \mathbb{N}$, and $\mathcal{A}_k$ is the Borel σ-algebra defined on Ω_k. The random variables $Y_1, \ldots, Y_K$ are assumed to be independent under each hypothesis, but not necessarily identical. Given an observation y_k, each sensor transmits (error free) its own decision $u_k = \phi_k(y_k)$ to the fusion center. The fusion center represented by the fusion rule $\gamma \in \Gamma$, where Γ is the set of all fusion rules on $(S_1, \ldots, S_K)$, then makes the final binary decision u_0 based on all decisions $u_1, \ldots, u_K$ that are received. The technical details related to the random variables Y_k and U_k corresponding to the observations y_k and decisions u_k, respectively, as well as to the decision rules, which are shown in Figure 1, are detailed below.

- Under each hypothesis $\mathcal{H}_j$, the random variables Y_k^j and $U_k^j = \phi_k(Y_k^j)$ follow the distributions $Q_j^{Y_k^j}$ and $Q_j^{U_k^j}$ having the density functions $q_j^{Y_k^j}$ and $q_j^{U_k^j}$, respectively. The distributions $Q_j^{Y_k^j}$ belong to the uncertainty classes $\mathscr{P}_j^{Y_k^j}$. In order to avoid cumbersome notation, the distributions and densities will be denoted by Q_0^k, Q_1^k, q_0^k, q_1^k, and the uncertainty classes by $\mathscr{P}_0^k$ and $\mathscr{P}_1^k$ omitting the random variables in superscripts.
- Similarly, the distributions $\mathbf{Q}_0 = (Q_0^1, \ldots, Q_0^K)$ and $\mathbf{Q}_1 = (Q_1^1, \ldots, Q_1^K)$ belong to the product uncertainty classes $\mathscr{P}_0 = \mathscr{P}_0^1 \times \ldots \times \mathscr{P}_0^K$ and $\mathscr{P}_1 = \mathscr{P}_1^1 \times \ldots \times \mathscr{P}_1^K$, respectively.
- $\mathbf{Y}^j = (Y_1^j, \ldots, Y_K^j)$ and $\mathbf{U}^j = (U_1^j, \ldots, U_K^j)$ are the multivariate random variables under the hypothesis $\mathcal{H}_j$, and $\mathbf{Y}$ and $\mathbf{U}$ are defined similarly without the index j.
- The stochastically larger sign $\succeq$ is extended to vector notation $\boldsymbol{\succeq}$, i.e.,
$$\hat{\mathbf{U}}^j \boldsymbol{\succeq} \mathbf{U}^j \Longrightarrow \hat{U}_k^j \succeq U_k^j, \quad \forall k,$$
where $\hat{(\cdot)}$ indicates the LFDs, e.g., $\hat{U}_k^j$ is the random variable U_k which follows $\hat{Q}_j$.
- The vector notation is also applied to the collection of decision rules $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$, where every decision rule ϕ_k is an element of the set of all decision rules Δ_k on Ω_k, hence, $\boldsymbol{\phi} \in \boldsymbol{\Delta} = \Delta_1 \times \ldots \times \Delta_K$.

Let the false alarm and miss detection probabilities be defined as $P_F = \mathbb{E}_{\mathbf{Q}_0}[\gamma]$ and $P_M = \mathbb{E}_{\mathbf{Q}_1}[1-\gamma]$. Then, the minimum error probability can be written as

$$\begin{aligned}
&P_E(\mathbf{Q}_0, \mathbf{Q}_1, \boldsymbol{\phi}, \gamma) \\
&\quad = P(\mathcal{H}_0) P_F(\mathbf{Q}_0, \boldsymbol{\phi}, \gamma) + P(\mathcal{H}_1) P_M(\mathbf{Q}_1, \boldsymbol{\phi}, \gamma) \\
&\quad = P(\mathcal{H}_0) \mathbb{E}_{\mathbf{Q}_0}\left[\gamma(\boldsymbol{\phi}(\boldsymbol{Y}))\right] \\
&\quad + (1 - P(\mathcal{H}_0)) \mathbb{E}_{\mathbf{Q}_1}\left[1 - \gamma(\boldsymbol{\phi}(\boldsymbol{Y}))\right]. \qquad (10)
\end{aligned}$$

Accordingly, a solution to the following problem is sought.

Problem III.1 (Original minimax decentralized detection problem):

$$\{\hat{\mathbf{Q}}_0, \hat{\mathbf{Q}}_1, \hat{\boldsymbol{\phi}}, \hat{\gamma}\} = \arg \inf_{\boldsymbol{\phi} \in \boldsymbol{\Delta}, \gamma \in \Gamma} \sup_{(\mathbf{Q}_0, \mathbf{Q}_1) \in \mathscr{P}_0 \times \mathscr{P}_1} P_E(\mathbf{Q}_0, \mathbf{Q}_1, \boldsymbol{\phi}, \gamma). \quad (11)$$

where $\hat{\mathbf{Q}}_0$ and $\hat{\mathbf{Q}}_1$ are the P_E maximizing distributions, and $\hat{\boldsymbol{\phi}}$ and $\hat{\gamma}$ are the P_E minimizing decision and fusion rules, respectively.

Let $\hat{\mathbf{Q}}_0^* \in \mathscr{P}_0$ and $\hat{\mathbf{Q}}_1^* \in \mathscr{P}_1$ be two distributions, which are not necessarily the same distributions with $\hat{\mathbf{Q}}_0$ and $\hat{\mathbf{Q}}_1$, respectively. Then, instead of solving the original minimax decentralized detection problem, it is much easier and hence more preferable to solve the following simple decentralized detection problem, in case they are equivalent.

Problem III.2 (Simple Decentralized Detection Problem):

$$\{\hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*\} = \arg \inf_{\boldsymbol{\phi} \in \boldsymbol{\Delta}, \gamma \in \Gamma} P_E(\hat{\mathbf{Q}}_0^*, \hat{\mathbf{Q}}_1^*, \boldsymbol{\phi}, \gamma). \qquad (12)$$

where $\hat{\boldsymbol{\phi}}^*$ and $\hat{\gamma}^*$ are the P_E minimizing decision and fusion rules for the distributions $\hat{\mathbf{Q}}_0^*$ and $\hat{\mathbf{Q}}_1^*$.

The results of Problem III.1 and Problem III.2 are equivalent if $\hat{\mathbf{Q}}_0^*$ and $\hat{\mathbf{Q}}_1^*$ satisfy the worst case P_F and P_M conditions. This will be stated with the following proposition.

Proposition III.1: It is true that $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}^*$ and $\hat{\gamma} = \hat{\gamma}^*$, hence, $\hat{\mathbf{Q}}_0 = \hat{\mathbf{Q}}_0^*$ and $\hat{\mathbf{Q}}_1 = \hat{\mathbf{Q}}_1^*$ if

$$\begin{aligned}
&P_F(\hat{\mathbf{Q}}_0^*, \hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*) \geq P_F(\mathbf{Q}_0, \hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*), \quad \forall \mathbf{Q}_0 \in \mathscr{P}_0, \\
&P_M(\hat{\mathbf{Q}}_1^*, \hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*) \geq P_M(\mathbf{Q}_1, \hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*), \quad \forall \mathbf{Q}_1 \in \mathscr{P}_1. \qquad (13)
\end{aligned}$$

A proof of Proposition III.1 is given in Appendix B.

Remark III.1: Instead of solving Problem III.1, by using the Proposition III.1, one can equivalently find $\hat{\mathbf{Q}}_0$ and $\hat{\mathbf{Q}}_1$, and P_E minimizing decision and fusion rules, $\hat{\boldsymbol{\phi}}$ and $\hat{\gamma}$, such that the inequalities in (13) are satisfied. This strategy will be followed in the next section.

IV. Minimax Robust Decentralized Detection

Error minimizing decision rules $\hat{\boldsymbol{\phi}}$ and the fusion rule $\hat{\gamma}$ are known to be monotone likelihood ratio tests (MLRTs) [23]. The conditions that need to be satisfied for (13) to hold are twofold.

1) Conditions defined on $\mathbf{U}$ and from $\mathbf{U}$ to U_0 via the fusion rule γ.
2) Conditions defined from $\mathbf{Y}$ to $\mathbf{U}$ via ϕ_k such that the conditions defined in 1) hold.

The following theorem details 1), whereas the next two theorems suggest two possible solutions for 2).

Theorem IV.1: The inequalities defined by (13) hold if $\hat{\boldsymbol{\phi}}$ results in

1) $\hat{\mathbf{U}}^0 \succeq \mathbf{U}^0$ and $\mathbf{U}^1 \succeq \hat{\mathbf{U}}^1$,
2) Monotone non-decreasing likelihood ratio function $\hat{l}_k = \hat{q}_1^{U_k^1}/\hat{q}_0^{U_k^0}$ for every k.

Proof: Since $U_1,\dots,U_K$ are all mutually independent random variables, the optimum fusion rule $\hat{\gamma}$ at the fusion center is to make a decision based on

$$\prod_{k=1}^{K}\frac{\hat{q}_1^k(U_k)}{\hat{q}_0^k(U_k)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t,$$

which is equivalent to

$$\log\prod_{k=1}^{K}\frac{\hat{q}_1^k(U_k)}{\hat{q}_0^k(U_k)} = \sum_{k=1}^{K}\log\hat{l}_k(U_k) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \log t. \tag{14}$$

From condition 2), recall that $\hat{l}_k$ is monotone non-decreasing, $\log\hat{l}_k$ is also monotone non-decreasing for all k. Using Lemma II.1 in condition 1) with $v=\log\hat{l}_k$, all summands in (14) satisfy

$$\begin{aligned}\log\hat{l}_k(\hat{U}_k^0) \succeq \log\hat{l}_k\left(U_k^0\right), \quad \forall k,\forall Q_0^k\in\mathscr{P}_0^k,\\ \log\hat{l}_k\left(U_k^1\right) \succeq \log\hat{l}_k(\hat{U}_k^1), \quad \forall k,\forall Q_1^k\in\mathscr{P}_1^k.\end{aligned} \tag{15}$$

Accordingly, by applying Lemma II.2 to both inequalities in (15) inductively, i.e., to the pairs of random variables iteratively, leads to

$$\begin{aligned}\sum_{k=1}^{K}\log\hat{l}_k(\hat{U}_k^0) \succeq \sum_{k=1}^{K}\log\hat{l}_k\left(U_k^0\right), \quad \forall Q_0^k\in\mathscr{P}_0^k,\\ \sum_{k=1}^{K}\log\hat{l}_k\left(U_k^1\right) \succeq \sum_{k=1}^{K}\log\hat{l}_k(\hat{U}_k^1), \quad \forall Q_1^k\in\mathscr{P}_1^k.\end{aligned} \tag{16}$$

Let $\hat{Q}_j$ and Q_j be the probability distributions of the random variable $Z=\sum_{k=1}^{K}\log\hat{l}_k\left(U_k\right)$, when U_k is distributed as $\hat{Q}_j^k$ and Q_j^k, respectively. Then, by Remark II.1, the stochastic ordering conditions given by (16) can equivalently be written as

$$\begin{aligned}\hat{Q}_0\left[Z>t\right] \geq Q_0\left[Z>t\right], \quad \forall t,\forall Q_0,\\ \hat{Q}_1\left[Z\leq t\right] \geq Q_1\left[Z\leq t\right], \quad \forall t,\forall Q_1.\end{aligned} \tag{17}$$

The inequalities in (17) imply the assertion, hence, the proof is complete. ■

Sufficient conditions among the random variables $U_1,\dots,U_K$ as well as from $\mathbf{U}$ to U_0 have been established with Theorem IV.1. Next, by Theorem IV.2 and Theorem IV.4, the sufficient conditions from $\mathbf{Y}$ to $\mathbf{U}$ will be stated with a suitable choice of decision rules $\boldsymbol{\phi}_k$.

Theorem IV.2: Let each decision rule ϕ_k be a monotone likelihood ratio test

$$\begin{aligned}U_k &= \phi_k(X_k)\\ &= \begin{cases}0, & X_k<t_0^k\\ d, & t_{d-1}^k\leq X_k<t_d^k\ ,\ d\in\{1,\dots,D_k-1\}.\\ D_k & X_k>t_{D_k-1}\end{cases}\end{aligned} \tag{18}$$

where $X_k=\hat{l}_k(Y_k)$ and t_d^k are some constants. Furthermore, let us assume that

$$\hat{X}_k^0 \succeq X_k^0 \quad\text{and}\quad X_k^1\succeq\hat{X}_k^1, \quad \forall k,\forall Q_j^k\in\mathscr{P}_j^k. \tag{19}$$

Then, the two conditions described in Theorem IV.1 hold and all conclusions therein follow.

Proof: For the monotone non-decreasing function ϕ_k (18) and the r.v.s defined by (19), Lemma II.1 implies

$$\hat{U}_k^0\succeq U_k^0 \quad\text{and}\quad U_k^1\succeq\hat{U}_k^1, \quad \forall k,\forall Q_j^k\in\mathscr{P}_j^k.$$

The function $\hat{l}_k=\hat{q}_1^{U_k^1}/\hat{q}_0^{U_k^0}$ is a monotone non-decreasing function for all k as

$$\begin{aligned}\hat{l}_k(d) &= \frac{\hat{Q}_1^k[t_{d-1}\leq X_k<t_d]}{\hat{Q}_0^k[t_{d-1}\leq X_k<t_d]} \leq \frac{\hat{Q}_1^k[t_d\leq X_k<t_{d+1}]}{\hat{Q}_0^k[t_d\leq X_k<t_{d+1}]}\\ &= \hat{l}_k(d+1)\end{aligned}$$

holds for all d since

$$\begin{aligned}&\frac{\hat{Q}_1^k[t_{d-1}\leq X_k<t_d]}{\hat{Q}_0^k[t_{d-1}\leq X_k<t_d]} =\\ &\frac{1}{\hat{Q}_0^k[t_{d-1}\leq X_k<t_d]}\int_{\{t_{d-1}\leq X_k<t_d\}}\mathrm{d}\hat{Q}_1^k =\\ &\frac{1}{\hat{Q}_0^k[t_{d-1}\leq X_k<t_d]}\int_{\{t_{d-1}\leq X_k<t_d\}}X_k\mathrm{d}\hat{Q}_0^k =\\ &\mathbb{E}_{\hat{Q}_0^k}[X_k|t_{d-1}\leq X_k<t_d]\end{aligned}$$

implies $t_{d-1}\leq\hat{l}_k(d)<t_d$ and $t_d\leq\hat{l}_k(d+1)<t_{d+1}$. The result also applies to the end points, i.e., $\hat{l}_k(0)$ and $\hat{l}_k(D_k)$, considering the intervals $(0,t_0^k)$ and (t_{D_k-1},∞), respectively. ■

In the following corollary, the results of Theorem IV.2 are extended to the case, where the decision rules ϕ_k are not necessarily monotone. In order to have this property, however, the fusion center must apply a well defined permutation function to the received decisions.

Corollary IV.3: Let $\phi_k: X_k\mapsto U_k$ be any bijective mapping (not necessarily monotone) from the intervals of X_k to possibly multilevel discrete decisions U_k. Then, there exists a permutation mapping $\boldsymbol{\pi}=\{\pi_1,\dots,\pi_K\}$ applied by the fusion center such that the two conditions described in Theorem IV.1 hold and all conclusions therein follow.

Proof: For each decision maker ϕ_k, bijective mapping from the intervals of X_k to U_k makes $\hat{l}_k$ not necessarily monotone. However, this is reversible at the fusion center by applying a well defined permutation function $\pi_k: S_k\to S_k$ to U_k such that $\hat{l}_k$ becomes monotone. Since the same procedure is applied to all decision makers by $\boldsymbol{\pi}$, the overall likelihood ratio test performed by the fusion center is equivalent to the case, where $\hat{l}_k$ is monotone. Hence, Theorem IV.2 and accordingly Theorem IV.1 follow. ■

The mapping described by ϕ_k can also be deduced from [23, p. 310]. Notice that the fusion center must know which decisions correspond to which decision makers to be able to perform the required permutations.

Theorem IV.2 provides sufficient conditions on $\boldsymbol{\phi}$ for minimax robust decentralized detection. These conditions require joint stochastic boundedness property reformulated by (19). Second possible design of $\boldsymbol{\phi}$, which does not require the joint stochastic boundedness property is given by the following theorem.

Theorem IV.4: Let $\phi_k : \Omega_k \to \{0,1\}$ be a (random) mapping which results in

$$\hat{U}_k^0 \succeq U_k^0 \quad \text{and} \quad U_k^1 \succeq \hat{U}_k^1 \quad \forall k, \forall Q_j^k \in \mathscr{P}_j^k, \tag{20}$$

satisfying the condition $\hat{Q}_1^k(U_k = 0) + \hat{Q}_0^k(U_k = 1) \le 1$. Then, all conclusions of Theorem IV.1 follow.

Proof: By the definition of ϕ_k the first condition in Theorem IV.1 is immediately satisfied. What remains to be shown is that $\hat{l}_k$ is equal to a non-decreasing (discrete) function. This condition is true because for all k

$$\hat{Q}_1^k(U_k = 0) \le 1 - \hat{Q}_0^k(U_k = 1),$$
$$\hat{Q}_0^k(U_k = 1) \le 1 - \hat{Q}_1^k(U_k = 0),$$

implies

$$\hat{Q}_0^k(U_k\!=\!1)\hat{Q}_1^k(U_k\!=\!0)\!\le\!(1\!-\!\hat{Q}_0^k(U_k = 1))(1\!-\!\hat{Q}_1^k(U_k = 0)),$$

which is

$$\hat{l}_k(1) = \frac{1 - \hat{Q}_1^k(U_k = 0)}{\hat{Q}_0^k(U_k = 1)} \ge \frac{\hat{Q}_1^k(U_k = 0)}{1 - \hat{Q}_0^k(U_k = 1)} = \hat{l}_k(0).$$

■

Both Theorem IV.2 and Theorem IV.4 imply Theorem IV.1. From Theorem IV.1 to the inequalities given by (13), what remains to be shown is that among all possible $\boldsymbol{\phi} \in \boldsymbol{\Delta}$, $\hat{\boldsymbol{\phi}}$ minimizes the overall error probability P_E, since by Theorem IV.1, $\hat{\gamma}$ is already chosen to be the likelihood ratio test, which is known to minimize P_E [23], cf. Remark III.1.

Remark IV.1: For $D_k = 1$, Theorem IV.2 (see its proof) implies Theorem IV.4. However, the converse is not true, i.e., the inequalities in (19), which are necessary for Theorem IV.2, are not necessary for Theorem IV.4. This means that for binary U_k, one can combine Theorem IV.4 with Theorem IV.1 and bypass Theorem IV.2, which also implies circumventing the stochastic boundedness property, in order to reach the inequalities given by Proposition III.1.

V. Uncertainty Classes

In the previous section uncertainty classes have not been defined explicitly. Instead, it has been assumed that there are decision rules $\boldsymbol{\phi}_k$ based on the least favorable distributions $\hat{\boldsymbol{Q}}_0$ and $\hat{\boldsymbol{Q}}_1$, which satisfy certain conditions. In this section, three different uncertainty classes will be introduced and it will be shown that the conditions defined by Theorem IV.2 and/or Theorem IV.4 are satisfied by these uncertainty classes. This eventually completes the link from the uncertainty classes to the minimax equation given by Problem III.1.

A. Huber's Extended Uncertainty Classes

Huber showed that there exists a pair of LFDs, which satisfies the joint stochastic boundedness property, i.e., (19), for binary (robust) hypothesis testing with a single decision maker, where under each hypothesis the set of distributions are represented by the uncertainty classes that include ϵ- contamination-, total variation-, Prohorov-, Kolmogorov- and Levy neighborhood as special cases [4], [24, p. 271]. The robust decision rule is obtained by quantizing the robust likelihood ratios, which are basically the clipped versions of the nominal likelihood ratios.

Assume that for a parallel decentralized detection network, all decision makers consider Huber's extended uncertainty classes. Then, since (19) holds, by quantizing the robust likelihood ratios with a monotone likelihood ratio test as given by (18), or with any bijective mapping due to Corollary IV.3, we can reach Theorem IV.2, hence to the minimax equation given by Problem III.1. This result was previously obtained by [14]. However, if the quantization is binary, i.e., U_k are binary or $D_k = 1$, by making use of Theorem IV.4, we can find a solution to Problem III.1 without any need for stochastic boundedness property, see Remark IV.1. Notice that for each decision maker the quantization thresholds are free to choose similar to the threshold of the fusion rule. Therefore, the decision rule can be chosen such that it minimizes P_E.

B. Uncertainty Classes Based on KL-Divergence

A minimax robust test based on the KL-divergence gets user defined pair of robustness parameters and a pair of nominal distributions as inputs and gives a unique pair of least favorable density functions $(\hat{q}_0, \hat{q}_1)$ and a randomized robust decision rule $\hat{\phi}$ as outputs [8]. The robustness parameters should be chosen small enough in order to ensure that the hypotheses do not overlap, i.e., a minimax robust test exists or equivalently for a single decision maker ($k = 1$),

$$\hat{Q}_1^k(U_k = 0) + \hat{Q}_0^k(U_k = 1) < 1. \tag{21}$$

From [8], we also have ($k = 1$),

$$P_F(\hat{Q}_0^k, \cdot, \hat{\phi}) \ge P_F(Q_0^k, \cdot, \hat{\phi}), \quad \forall Q_0^k \in \mathscr{P}_0^k,$$
$$P_M(\hat{Q}_1^k, \cdot, \hat{\phi}) \ge P_M(Q_1^k, \cdot, \hat{\phi}), \quad \forall Q_1^k \in \mathscr{P}_1^k. \tag{22}$$

Suppose that for a parallel decentralized detection network, all decision makers $k \in \{1, \dots, N\}$ are associated with the uncertainty classes based on the KL-divergence. Then, by (21) and the following inequalities given by (22), which imply (20), Theorem IV.4 and hence Theorem IV.1 follow. From Theorem IV.1 to a solution to Problem III.1, what remains to be shown is that the decision rules $\hat{\boldsymbol{\phi}}$ jointly minimize P_E, see Remark III.1. However, this is not necessarily true, since for the KL-divergence based uncertainty classes, the decision rule for each decision maker is unique, and designed to minimize the error probability of each decision maker P_{E_k}, not the global error probability P_E. Minimizing the error probability of each decision maker P_{E_k} for every decision maker k does not guarantee that P_E is also minimized. However, there are special cases, for which P_E is also minimized. Suppose that the uncertainty classes under each hypotheses are the same for all decision makers. Then, the corresponding robust decision rule for each decision maker will be identical. It is known that identical decision rules are not always P_E minimizer [25]. However, for the majority of decision making problems, i.e., for the choice of nominal distributions, identical decision makers are optimum and minimize P_E for some fusion rule $\hat{\gamma}$ [26]. Similarly, if no assumption is made on the choice of robustness parameters and nominal distributions, there are some detection problems for which P_E is minimized by $\hat{\boldsymbol{\phi}}$. This result together with Theorem IV.1 implies a solution to Problem III.1 and generalizes [14], which requires stochastic boundedness property. In other words, it is not possible to use the theory in [14] and come up with a solution to Problem III.1 for the KL-divergence based uncertainty classes. Because, as shown by Proposition II.5, the KL-divergence does not accept joint stochastic boundedness property, which is a precondition in [14].

C. Uncertainty Classes Based on α-Divergence

Similar to the KL-divergence, for the choice of α-divergence, the related uncertainty classes are not jointly stochastically bounded, because minimax decision rules are randomized [9]. However, a solution to Problem III.1 through Remark III.1 can again be obtained since (21) and (22) also hold for the α-divergence [9]. The advantage of α-divergence over the KL-divergence is that both the distance, symbolized by the parameter α, as well as the threshold of the nominal test are selectable for every decision maker. This provides flexibility and a more likely scenario that the designed decision rules $\boldsymbol{\phi}$ minimize not only P_{E_k}s but also P_E. Both for the KL-divergence and the $\alpha-$divergence, without imposing any additional constraints on the choice of parameters or nominal distributions, inequalities in (21) and (22) are always satisfied. Therefore, the power of the test is guaranteed to be above a certain threshold, despite the uncertainty on the sensor network.

D. Composite Uncertainty Classes

Uncertainty classes for each decision maker can be chosen arbitrarily either from Huber's extended uncertainty classes or from the uncertainty classes formed with respect to the α-divergence[1]. Based on the information from the previous sections, it can be concluded that the decentralized detection network can be minimax robust, if sensor and fusion thresholds minimize the overall error probability P_E for the least favorable distributions $\hat{\boldsymbol{Q}}_{\mathbf{0}}$ and $\hat{\boldsymbol{Q}}_{\mathbf{1}}$.

VI. GENERALIZATIONS

A. Minimax and Neyman-Pearson Formulations

Minimax and Neyman-Pearson versions of the same problem can respectively be stated as follows:

$$\inf_{\boldsymbol{\phi}\in\boldsymbol{\Phi},\gamma\in\Gamma} \max\left\{\sup_{\boldsymbol{Q}_{\mathbf{0}}\in\mathscr{P}_0} P_F(\boldsymbol{Q}_{\mathbf{0}},\boldsymbol{\phi},\gamma), \sup_{\boldsymbol{Q}_{\mathbf{1}}\in\mathscr{P}_1} P_M(\boldsymbol{Q}_{\mathbf{1}},\boldsymbol{\phi},\gamma)\right\},$$
$$\inf_{\boldsymbol{\phi}\in\boldsymbol{\Phi},\gamma\in\Gamma} \sup_{\boldsymbol{Q}_{\mathbf{1}}\in\mathscr{P}_1} P_M(\boldsymbol{Q}_{\mathbf{1}},\boldsymbol{\phi},\gamma) \text{ s.t. } \sup_{\boldsymbol{Q}_{\mathbf{0}}\in\mathscr{P}_0} P_F(\boldsymbol{Q}_{\mathbf{0}},\boldsymbol{\phi},\gamma)\leq t. \tag{23}$$

Application of exactly the same procedure defined by Proposition III.1 to the minimax formulation results in the simple version of the same problem, where $\mathscr{P}_0$ and $\mathscr{P}_1$ are replaced by the singletons $\hat{\boldsymbol{Q}}_{\mathbf{0}}$ and $\hat{\boldsymbol{Q}}_{\mathbf{1}}$. For the Neyman-Pearson formulation, depending on the choice of the threshold of the constraint, dependently randomized decision and/or fusion rules may need to be employed at the decision makers and/or at the fusion center [23], [27]. It was stated in [14] that the same simplification then applies to the Neyman-Pearson formulation as well.

Huber's uncertainty classes satisfy the joint stochastic boundedness property, which makes LFDs to be defined independent of the decision rules. Once the LFDs are found, Neyman-Pearson tests can be designed by jointly randomizing MLRTs at the decision makers. However, the same result does not apply to the α-divergence based uncertainty classes, because, the LFDs are in this case dependent on the decision rule, which is unique, and modifying the optimal decision rule results in the loss of minimax robustness [9], [22].

B. Repeated Observations

Suppose that one or more decision makers give their decisions based on a block of observations, which are not necessarily obtained from identically distributed random variables. Then, for every decision maker minimax condition holds by multiplying the robust likelihood ratio functions and comparing the result to a threshold, if Huber's uncertainty classes are considered [3, p. 1756]. However, the same conclusion cannot be made in the same way for the α-divergence based uncertainty classes [9]. Because, multiplication of the likelihood ratios removes the randomization information, which is unconditionally required for minimax robustness. An alternative way in this case could be to consider designs over multi-variate distributions [22].

C. Imperfect Reporting Channels

It was assumed in the problem formulation that the reporting channel between the sensors and the fusion center is error free. In general the reporting channel can be modeled as a q-ary discrete channel in order to account for imperfect channel conditions. Let the communication link between the kth sensor and the fusion center is modeled by a state transition matrix $\mathbf{T}^k$, where each element of this matrix is given by

$$T^k_{\tilde{d}d} = P_k[\tilde{U}_k = \tilde{d}|U_k = d], \quad d, \tilde{d} \in \{0, \dots D_k - 1\}. \tag{24}$$

Here, $\tilde{U}_k$ are the corrupted (multilevel) decisions received by the fusion center, P_k are the conditional probabilities corresponding to the underlined events by each sensor k and $\sum_{\tilde{d}=0}^{D_k-1} T^k_{\tilde{d}d} = 1$ for any $d \in \{0, \dots D_k - 1\}$. Hence, the distribution of the received symbols by the fusion center can be calculated as

$$\hat{Q}^k_j[\tilde{U}_k = \tilde{d}] = \sum_{d=0}^{D_k-1} T^k_{\tilde{d}d}\hat{Q}^k_j[U_k = d]. \tag{25}$$

Assuming that $\mathbf{T}^k$ can perfectly be estimated by the fusion center, instead of U_k the induced random variable $\tilde{U}_k$ should satisfy Theorem IV.2 or Theorem IV.4 for minimax robustness. In general $\tilde{U}_k$ does not have to satisfy one of these theorems. However, if U_k satisfies one of these theorems, so does $\tilde{U}_k$ if the non-diagonal elements of $\mathbf{T}^k$ are small enough. Moreover, $\hat{\phi}$ should minimize the error probability of the fusion center, when U_k are replaced by $\tilde{U}_k$ in Theorem IV.1. As a special case, $\tilde{U}_k$s satisfy the conditions in Theorem IV.4 if U_ks are transmitted over binary symmetric memoryless channels.

[1]As $\alpha \to 1$, the $\alpha-$divergence tends to the KL-divergence [22].

A proof is provided in Appendix C. More details of imperfect reporting channels can be found in [15], [28]. See also [29], [30] for applications.

D. Different Network Topologies

Although it may be true that the parallel network topology has received the most attention in the literature [27], depending on the application, decentralized detection networks can be designed considering a number of different topologies, for example a serial topology, a tree topology, or an arbitrary topology [23]. For arbitrary network topologies, it is known that likelihood ratio tests are no longer optimal in general [23, p. 331]. Therefore, the results obtained for a parallel network topology cannot be generalized to arbitrary networks in a straightforward manner. Each network structure requires a new and possibly a much complicated design.

If the network topology is parallel, P_E goes to zero as the number of sensors goes to infinity. This is a consequence of Cramér's Theorem [31] for independent Bernoulli random variables U_k. If the network is serial, from [32], [33], we know that the error probability is almost surely bounded away from zero in case the underlined likelihood ratio function is bounded from above and below. Remember that Huber's clipped likelihood ratio test bounds the nominal likelihoods, therefore, a minimax robust serial network may never be asymptotically error free [17]. On the other hand, a minimax robust test based on the KL-divergence or α-divergence does not alter the boundedness properties of the nominal likelihood ratios, hence, it preserves the asymptotic property of the network.

VII. Examples

In this section two examples regarding the theoretical derivations will be given in connection with the uncertainty classes introduced in the previous sections. Let us consider the parallel decentralized detection network with a fusion center as illustrated by Figure 1. Suppose that for every decision maker ϕ_k the phenomenon is nominally characterized by the binary hypothesis test

$$\begin{aligned} \mathcal{H}_0 &: Y_k \sim \mathcal{N}(-1,1) \\ \mathcal{H}_1 &: Y_k \sim \mathcal{N}(1,1) \end{aligned} \tag{26}$$

where Y_k are mutually independent random variables following the Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$ with mean μ and variance σ^2. The actual distributions of the r.v.s Y_k, however, may depend either on the outliers that corrupt the collected data, or the secondary physical effects that go unmodeled by (26). If the nominal model is correct, optimum nominal test is an MLRT based on the nominal distributions. Else, some sort of robust hypothesis testing procedure is required for a reliable detection and this will be investigated in the next two sections.

A. Clipped Likelihood Ratio Test

Huber's extended uncertainty classes result in clipped likelihood ratio tests. Here the ϵ-contamination model is considered, due to its practical use in various applications, e.g., in cognitive radio [34] or in localization [35]. It is assumed that ten percent of the collected data contains outliers under each hypothesis. Therefore, the robustness parameters are chosen as $\epsilon_0 = \epsilon_1 = 0.1$. The corresponding minimax robust test is an MLRT based on the LFDs, which are determined by solving the related equations given in [3]. For comparison, another robust test has been designed as an MLRT for the likelihood ratio function $\bar{l} = \bar{q}_1/\bar{q}_0$, where

$$\begin{aligned} \bar{q}_0 &= \epsilon_0 p_1 + (1-\epsilon_0) p_0, \\ \bar{q}_1 &= \epsilon_1 p_0 + (1-\epsilon_1) p_1. \end{aligned} \tag{27}$$

Accordingly the following notations are adopted. The nominal test is denoted by the (n)-test, the minimax robust clipped likelihood ratio test is denoted by the (h)-test and the second robust test is denoted by the (r)-test. The data to be tested are sampled from the pair of distributions building the likelihood ratio functions of each of these three tests. The notation $|_b^a$ indicates that the test under consideration is the (b)-test while the data samples come from the pair of distributions, which yield the (a)-test. For binary decisions, all three tests are the same and have no difference in terms of robustness or performance. Therefore, all three tests are simulated for multilevel decisions, i.e., for $D_k = 3$ and $D_k = 7$. These choices of the number of quantization levels correspond to 2- and 3-bits of transmission per observation per decision maker. For n-bit quantization there are $2^n - 1$ thresholds that need to be found optimally for each sensor. Given the quantization thresholds, P_E minimizing fusion rule is a likelihood ratio test based on the distributions of the multilevel decisions U_k under each hypothesis. Determining the optimum decision and fusion rules for parallel decentralized detection is computationally intractable even for a few sensors, i.e., it is an NP-complete problem in general and an exhaustive search requires KD_k loops to be evaluated [27]. Therefore, we make the assumption that the sensor decisions are identical, i.e., $\phi_k(y_k) = \phi_i(y_k)$ for all i given k and y_k. With this assumption the computational complexity is reduced drastically and only D_k loops need to be evaluated. Notice that identical sensor decisions result in very little to no loss of detection performance [23, p. 313], [25] and they are asymptotically optimum [36]. For both 2- and 3-bit quantizations and for all three tests exhaustive search has been launched to determine the optimum (P_E minimizing) quantization thresholds and the corresponding fusion rules. This corresponds to the error probabilities, which are denoted by ${P_E}_n^n$, ${P_E}_r^r$ and ${P_E}_h^h$. Fixing the thresholds and fusion rules found, minimum error probabilities have been calculated also for the mismatch cases, i.e., ${P_E}_b^a$, for $a \not\equiv b$. In all cases the a-priori probabilities have been chosen to be equal $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$. Figures 2 and 3 illustrate the results of the tests for 2- and 3-bit quantizations, respectively, for the number of sensors varying from 2 to 51. The solid lines correspond to Huber's clipped likelihood ratio test, the dashed-lines correspond to the nominal test and the dotted lines correspond to the second robust test. In both figures, the (h)-test guarantees a certain level of detection performance, shown by ${P_E}_h^h$, irrespective of the uncertainties on the system model. The nominal test, although having the best performance for the nominal distributions have the largest degradation when the samples are actually

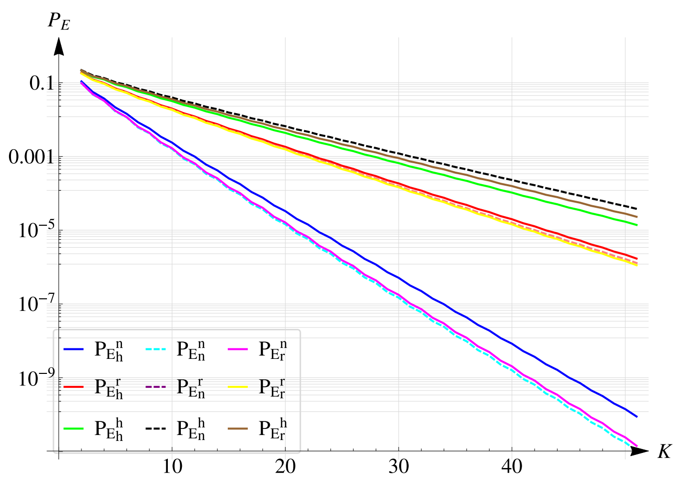

Fig. 2. Minimum error probabilities of three different tests for the data samples from three different statistics over various number of sensors, when the observations are quantized to 2-bits by all decision makers.

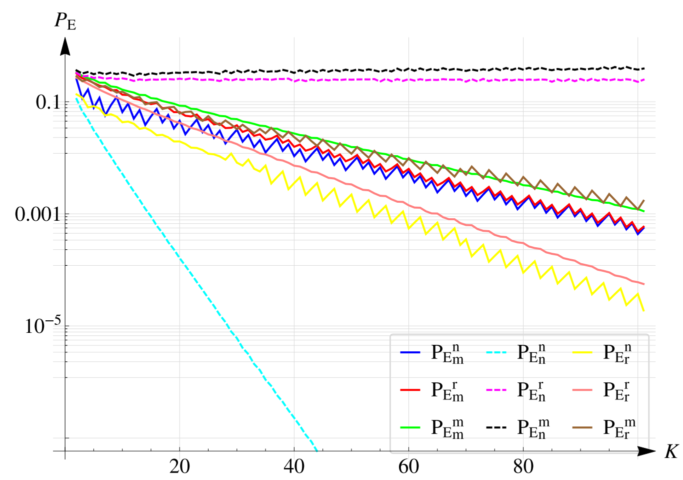

Fig. 4. Minimum error probabilities of three different tests for the data samples from three different statistics over various number of sensors, when the observations are quantized to 1-bit by all decision makers.

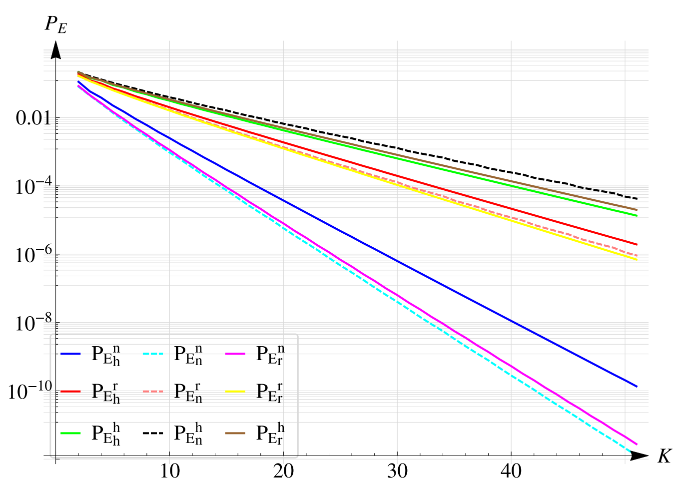

Fig. 3. Minimum error probabilities of three different tests for the data samples from three different statistics over various number of sensors, when the observations are quantized to 3-bits by all decision makers.

obtained from the LFDs of the (h)-test. The best performance guarantee property of the (h)-test is obtained at a cost of certain performance degradation in comparison to the (n)-test, if indeed the distributions of data samples are given by the nominal distributions and this degradation increases from 2-bit to 3-bit quantization.

B. Censored Randomized Likelihood Ratio Test

All minimax robust tests based on the KL-divergence [7], [8] or the α-divergence [9] (except for the limiting cases) result in censored randomized likelihood ratio tests. For simplicity the KL-divergence is considered. Suppose that the measurements show that the r.v. Y_k under $\mathcal{H}_0$ follows the nominal distribution $\mathcal{N}(-1,1)$, however under $\mathcal{H}_1$ it is not necessarily Gaussian due to secondary physical effects, and whenever it is Gaussian the mean can decrease down to 0 and the variance can increase up to $3/2$. Accordingly, alternative hypothesis should include $\mathcal{N}(0,3/2)$ and potentially all other distributions on Ω which are at least $\epsilon_1 = D(\mathcal{N}(0,3/2),\mathcal{N}(1,1))$ close to $\mathcal{N}(1,1)$, where D is the KL-divergence. For this uncertainty model the LFDs and the robust decision rule are obtained by solving the related equations in [8]. The resulting minimax robust test is denoted by the (m)-test. For comparison another robust test has been designed as an MLRT for the likelihood ratio $\bar{l} = d\bar{Q}_1/d\bar{Q}_0$, where $\bar{Q}_0 = \mathcal{N}(-1,1)$ and $\bar{Q}_1 = \mathcal{N}(0,3/2)$. As before this second robust test is denoted by the (r)-test and the nominal test is denoted by the (n)-test. All other notations, assumptions and experimental approach are inherited from the previous section. The decisions are now binary as the (m)-test is originally designed for binary quantization. Figure 4 illustrates the error probabilities of three different tests for three different data statistics and for various number of sensors. Out of three conclusions that can be made, the first is that the (m)-test guarantees a certain level of detection performance for every number of sensors, i.e., ${P_E}_m^a$, is upper-bounded for $a := m$ and it does not increase for $a := n$ or $a := r$. Secondly, it can be seen that the nominal likelihood ratio test is not robust as ${P_E}_n^r$ or ${P_E}_n^m$ do not tend to zero, no matter how many sensors there are in the sensor network. Finally, unlike for the (h)-test, the (m)-test does not minimize the overall error probability for every number of sensors, when the data samples come from the LFDs of the (m)-test. This can be seen by comparing ${P_E}_m^m$ with ${P_E}_r^m$, where the latter is expected to be larger than the former if the (m)-test indeed minimizes P_E. As explained in Section V-B, the (m)-test minimizes P_{E_k} for each sensor and this cannot necessarily be interpreted as minimizing the overall error probability P_E of the sensor network. However, asymptotically, it is possible that ${P_E}_m^m$ gets below ${P_E}_r^m$, as implied by the given example.

VIII. Conclusion

Minimax robust decentralized detection has been studied for parallel sensor networks. It was proven that the minimax robust tests designed from the KL-divergence neighborhood do not satisfy the joint stochastic boundedness property. For this case, it was not known whether the minimax optimization problem could be reduced to its simple version, which could be solved much more easily. We showed that an answer to this question was positive and this led to a generalization of an earlier work by Veeravalli *et al.* [14]. Such a generalization allows, without sacrificing from the minimax optimality, different types of minimax robust tests to be simultaneously employed by the decision makers, not only the clipped likelihood ratio tests. The theory has been developed under the assumption that the observations are independent but not necessarily identical. Additionally, multilevel quantization at decision makers was also allowed. An extension of the proposed model to minimax and Neyman-Pearson tests, repeated observations, imperfect reporting channels and different network topologies has been discussed. Two examples have been provided to show the benefits of minimax robust decentralized detection. Open problems arising from this work can be listed as follows:

- What are the minimax strategies for the sensor networks with arbitrary topologies, for which likelihood ratio test is known not to be optimum?
- How can one design minimax robust tests for decentralized sensor networks when observations are not mutually independent?
- How do minimax robust decentralized tests look like if the reporting channels are imperfect or if they introduce additional uncertainty, e.g., due to imperfect channel estimation?

APPENDIX A
PROOF OF PROPOSITION II.5

Proof: The claim can be proven by contradiction. Assume that there exists such a pair of LFDs $(\hat{Q}_0, \hat{Q}_1)$. Then, the same pair must satisfy

$$\begin{aligned}\hat{Q}_0 &= \arg\max_{Q_0\in\mathscr{P}_0} \mathbb{E}_{Q_0} \ln(d\hat{Q}_1/d\hat{Q}_0),\\ \hat{Q}_1 &= \arg\min_{Q_1\in\mathscr{P}_1} \mathbb{E}_{Q_1} \ln(d\hat{Q}_1/d\hat{Q}_0)\end{aligned} \tag{28}$$

by applying Remark II.1 and Lemma II.1 in (2) and (3). By Huber and Strassen [5, Theorem 7.1], see also [22], (28) is equivalent to

$$\begin{aligned}\hat{Q}_0 &= \arg\max_{Q_0\in\mathscr{P}_0} \mathbb{E}_{Q_0} \ln(d\hat{Q}_1/dQ_0),\\ \hat{Q}_1 &= \arg\min_{Q_1\in\mathscr{P}_1} \mathbb{E}_{Q_1} \ln(dQ_1/d\hat{Q}_0).\end{aligned} \tag{29}$$

By Dabak and Johnson [6], see also [2], the pair of distributions solving (29) are given by

$$\hat{q}_0 = \frac{{p_0}^{1-u}{p_1}^{u}}{\int_\Omega {p_0}^{1-u}{p_1}^{u}\mathrm{d}\mu}, \quad \hat{q}_1 = \frac{{p_0}^{v}{p_1}^{1-v}}{\int_\Omega {p_0}^{v}{p_1}^{1-v}\mathrm{d}\mu}, \tag{30}$$

where (p_0, p_1) and $(\hat{q}_0, \hat{q}_1)$ are the density pairs corresponding to (P_0, P_1) and $(\hat{Q}_0, \hat{Q}_1)$, respectively, μ is a suitable measure, and u and v are the parameters to be determined such that

$$D(\hat{Q}_0, P_0) = \epsilon_0, \quad D(\hat{Q}_1, P_1) = \epsilon_1. \tag{31}$$

The test based on $\hat{l} = \hat{q}_1/\hat{q}_0$ is still a nominal likelihood ratio test [2], [6], though with a modified threshold. It is known that nominal likelihood ratio tests are not minimax robust [24]. Therefore, the corresponding pair of distributions $(\hat{Q}_0, \hat{Q}_1)$ do not satisfy the joint stochastic boundedness property. Hence, no pair of distributions is jointly stochastically bounded for the KL-divergence neighborhood. ■

APPENDIX B
PROOF OF PROPOSITION III.1

Proof: We have

$$\begin{aligned}\sup_{(\mathbf{Q}_0,\mathbf{Q}_1)\in\mathscr{P}_0\times\mathscr{P}_1} & P_E(\mathbf{Q}_0, \mathbf{Q}_1, \hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*)\\ &= P_E(\hat{\mathbf{Q}}_0^*, \hat{\mathbf{Q}}_1^*, \hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*)\\ &\le P_E(\hat{\mathbf{Q}}_0^*, \hat{\mathbf{Q}}_1^*, \boldsymbol{\phi}, \gamma)\\ &\le \sup_{(\mathbf{Q}_0,\mathbf{Q}_1)\in\mathscr{P}_0\times\mathscr{P}_1} P_E(\mathbf{Q}_0, \mathbf{Q}_1, \boldsymbol{\phi}, \gamma)\end{aligned} \tag{32}$$

for any $\boldsymbol{\phi}$ and γ, where the equality follows from (13) and the first inequality follows from (12). This implies that $\hat{\boldsymbol{\phi}}^*$ and $\hat{\gamma}^*$ jointly minimize P_E, cf. sup terms in (32). Therefore, $(\hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*)$ is also the minimizer for Problem III.1, hence $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}^*$ and $\hat{\gamma} = \hat{\gamma}^*$. Inserting $(\hat{\boldsymbol{\phi}}^*, \hat{\gamma}^*)$ in (11) and considering (13) we have $\hat{\mathbf{Q}}_0 = \hat{\mathbf{Q}}_0^*$ and $\hat{\mathbf{Q}}_1 = \hat{\mathbf{Q}}_1^*$. ■

APPENDIX C
BINARY SYMMETRIC CHANNEL

Consider the binary symmetric channel which can be modeled as

$$\tilde{U}_k = U_k + Z_k \mod 2 \tag{33}$$

where Z_k is a Bernoulli distributed random variable independent of U_k and with a success probability $P \le 1/2$. Then, the distribution (false alarm and miss detection probabilities) of $\tilde{U}_k$ can be obtained by modulo 2 convolution of the probability mass function of U_k with that of Z_k as

$$\begin{aligned}\hat{Q}_0^k(\tilde{U}_k = 1) &= P + \hat{Q}_0^k(U_k = 1) - 2P\hat{Q}_0^k(U_k = 1),\\ \hat{Q}_1^k(\tilde{U}_k = 0) &= P + \hat{Q}_1^k(U_k = 0) - 2P\hat{Q}_1^k(U_k = 0).\end{aligned} \tag{34}$$

Let $\hat{Q}_0^k(U_k = 1) + \hat{Q}_1^k(U_k = 0) = (1 - \epsilon)$, where $0 \le \epsilon \le 1$. Then,

$$\begin{aligned}\hat{Q}_0^k(\tilde{U}_k = 1) &+ \hat{Q}_1^k(\tilde{U}_k = 0)\\ &= 2P + \hat{Q}_0^k(U_k = 1)\\ &+ \hat{Q}_1^k(U_k = 0) - 2P(\hat{Q}_0^k(U_k = 1)\\ &+ \hat{Q}_1^k(U_k = 0))\\ &= 1 - \epsilon(1 - 2P) \le 1.\end{aligned} \tag{35}$$

Hence, the first condition in Theorem IV.4 holds. Let $\hat{Q}_0^k(U_k = 1) - Q_0^k(U_k = 1) = \epsilon$, where $0 \le \epsilon \le 1$. Then,

$$\begin{aligned}\hat{Q}_0^k(\tilde{U}_k = 1) &- Q_0^k(\tilde{U}_k = 1)\\ &= \hat{Q}_0^k(U_k = 1) - Q_0^k(U_k = 1)\\ &+ 2P(Q_0^k(U_k = 1) - \hat{Q}_0^k(U_k = 1))\\ &= \epsilon(1 - 2P) \ge 0, \quad \forall Q_0^k \in \mathscr{P}_0^k.\end{aligned} \tag{36}$$

Since the same equations also apply to $\hat{Q}_1^k(\tilde{U}_k = 0) - Q_1^k(\tilde{U}_k = 0)$, the second condition in Theorem IV.4 holds and the proof is complete.

REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[2] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, 1st ed. Heidelberg, Germany: Springer, 2008.

[3] P. J. Huber, “A robust version of the probability ratio test,” *Ann. Math. Statist.*, vol. 36, no. 6, pp. 1753–1758, Dec. 1965.

[4] P. J. Huber, "Robust confidence limits," *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 10, no. 4, pp. 269–278, 1968.

[5] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, vol. 1, no. 2, pp. 251–263, Mar. 1973.

[6] A. G. Dabak and D. H. Johnson, "Geometrically based robust detection," in *Proc. Conf. Inf. Sci. Syst.*, Baltimore, MD, USA, May 1994, pp. 73–77.

[7] B. C. Levy, "Robust hypothesis testing with a relative entropy tolerance," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 413–421, Jan. 2009.

[8] G. Gül and A. M. Zoubir, "Minimax robust hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5572–5587, Sep. 2017.

[9] G. Gül and A. M. Zoubir, "Robust hypothesis testing with α-divergence," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4737–4750, Sep. 2016.

[10] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sums of observations," *Ann. Math. Statist.*, vol. 23, pp. 409–507, 1952.

[11] E. Geroniotis, "Robust distributed discrete-time block and sequential detection," in *Proc. Conf. Inform. Sci. Syst.*, Baltimore, MD, USA: Johns Hopkins Univ., Mar. 1987, pp. 354–360.

[12] E. Geroniotis and Y. A. Chau, "On minimax robust data fusion," in *Proc. Conf. Inform. Sci. Syst.*, Princeton, NJ, USA: Princeton Univ., Mar. 1988, pp. 876–881.

[13] E. Geraniotis and Y. A. Chau, "Robust data fusion for multisensor detection systems," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1265–1279, Nov. 1990.

[14] V. V. Veeravalli, T. Basar, and H. V. Poor, "Minimax robust decentralized detection," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 35–40, Jan. 1994.

[15] D. Ciuonzo and P. Salvo Rossi, "Distributed detection of a non-cooperative target via generalized locally-optimum approaches," *Inf. Fusion*, vol. 36, pp. 261–274, Jul. 2017.

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

[17] J. Ho, W. P. Tay, T. Q. S. Quek, and E. K. P. Chong, "Robust decentralized detection and social learning in tandem networks," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5019–5032, Oct. 2015.

[18] J. Park, G. Shevlyakov, and K. Kim, "Robust distributed detection with total power constraint in large wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2058–2062, Jul. 2011.

[19] G. Gül and A. M. Zoubir, "Robust detection under communication constraints," in *Proc. IEEE 14th Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2013, pp. 410–414.

[20] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Decentralized detection with censoring sensors," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1362–1373, Apr. 2008.

[21] E. Wolfstetter, *Topics in Microeconomics: Industrial Organization, Auctions, and Incentives*. Cambridge, U.K.: Cambridge Univ. Press, 1999, pp. 135–148.

[22] G. Gül, *Robust and Distributed Hypothesis Testing*, 1st ed. Heidelberg, Germany: Springer, 2017.

[23] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. Greenwich, CT, USA: JAI Press, 1993, pp. 297–344.

[24] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.

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

[26] W. W. Irving and J. N. Tsitsiklis, "Some properties of optimal thresholds in decentralized detection," *IEEE Trans. Autom. Control*, vol. 39, no. 4, pp. 835–838, Apr. 1994.

[27] P. K. Varshney, *Distributed Detection and Data Fusion*, 1st ed. Syracuse, NJ, USA: Springer-Verlag, 1996.

[28] G. Fabeck and R. Mathar, "Chernoff information-based optimization of sensor networks for distributed detection," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2009, pp. 606–611.

[29] D. Ciuonzo, G. Romano, and P. Salvo Rossi, "Channel-aware decision fusion in distributed MIMO wireless sensor networks: Decode-and-fuse vs. decode-then-fuse," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2976–2985, Aug. 2012.

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

[31] H. Cramér, "Sûr un nouveau théorème-limite de la théorie des probabilités," in *Actualités Scientifiques et Industrielles* (Colloque Consacré à la Théorie des Probabilités), vol. 736. Paris, France: Hermann, 1938, pp. 5–23.

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

[33] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 765–782, Jun. 1970.

[34] F. Y. Suratman, Y. Chakhchoukh, and A. M. Zoubir, "Locally optimum detection in heavy-tailed noise for spectrum sensing in cognitive radio," in *Proc. 2nd Int. Workshop Cognit. Inf. Process.*, Jun. 2010, pp. 134–139.

[35] F. Yin and A. M. Zoubir, "Robust positioning in NLOS environments using nonparametric adaptive kernel density estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3517–3520.

[36] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, no. 2, pp. 167–182, Jun. 1988.

Gökhan Gül (Member, IEEE) received the B.Sc. degree (Hons.) in electronic engineering from Uludağ University, Bursa, Turkey, in 2005, the M.Sc. degree in digital communications from Christian Albrechts University, Kiel, Germany, in 2009, and the Dr.-Ing. degree (Hons.) from Technische Universität Darmstadt, Germany, in 2015. He wrote his master's thesis and worked as a Research Assistant with the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany, from 2008 to 2009. From 2010 to 2011, he was a Visiting Researcher with Queen's University, Belfast, U.K. In 2011, he joined the Signal Processing Group, Technische Universität Darmstadt, as a Research Associate. He is currently a Scientist with the Fraunhofer Institute for Microengineering and Microsystems IMM, where he is also working on the application of signal processing and machine learning techniques to flow cytometry. His research interests include detection and estimation theory, statistics, probability theory, signal processing, and machine learning. He was a recipient of the Second Degree Award of the first International Competition on Steganalysis (HUGO) in 2011.