Delft University of Technology

Latency Analysis for Sequential Detection in Low-Complexity Binary Radio Systems

Stein, Manuel S.; Fauß , Micheal

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Latency Analysis for Sequential Detection in Low-Complexity Binary Radio Systems

Manuel S. Stein and Michael Fauß

Abstract—We consider the problem of making a quick decision in favor of one of two possible physical signal models while the numerical measurements are acquired by sensing devices featuring minimal digitization complexity. Therefore, the digital data streams available for statistical processing are binary and exhibit temporal and spatial dependencies. To handle the intractable multivariate binary data model, we first consider sequential tests for exponential family distributions. Within this generic probabilistic framework, we identify adaptive approximations for the log-likelihood ratio and the Kullback-Leibler divergence. The results allow designing sequential detectors for binary radio systems and analyzing their average run-time along classical arguments of Wald. In particular, the derived tests exploit the spatio-temporal correlation structure of the analog sensor signals engraved into the binary measurements. As an application, we consider the specification of binary sensing architectures for cognitive radio and GNSS spectrum monitoring where our results characterize the sequential detection latency as a function of the temporal oversampling and the number of antennas. Finally, we evaluate the efficiency of the proposed algorithms and illustrate the accuracy of our analysis via Monte-Carlo simulations.

Index Terms—cognitive radio, exponential family, GNSS, quantization, sequential detection, spectrum monitoring, 1-bit ADC

I. INTRODUCTION

The design of future sensor systems represents a challenge. For applications in the Internet of Things (IoT), the focus is on further miniaturization. Thus, small circuit size, low production cost, and low energy consumption are essential requirements. In contrast, for safety-critical applications, sensing accuracy and detection reliability are of utmost importance. In any case, optimal system design either aims at achieving minimum complexity at a specified performance level, or at converting the available resources into maximum performance. Advancing sensor technology, therefore, requires a thorough understanding of Pareto-optimal architectures, i.e., sensing systems for which it is impossible to improve on complexity or performance without sacrificing the other measure.

A difficulty is that terms like complexity and performance are fuzzy without restriction to a specific perspective. In an R&D environment, analog front-end engineers tend to equate complexity with the circuit area or the dissipated power, and performance with the degree of linearity. On the other hand, software engineers instead associate complexity with the computing effort and the size of the occupied memory, while performance for them is linked to the fast and correct response

of the digital units to specific input data. In hardware-aware statistical signal processing, the understanding is emerging that a holistic approach is required when designing advanced sensor systems. In particular, this means that the physical phenomenon, the design of the analog front-end, and digital processing should be considered as a single joint problem.

For this purpose, it is helpful to reduce the system task to its fundamental building blocks. Elementary for sensing are parameter estimation and signal detection. While estimation aims at inferring data model parameters within an open set from noisy measurements, detection covers cases with discrete parameter space. To characterize performance, the accuracy in determining the parameters is central within estimation. Detection theory focuses on the reliability in discriminating between the possible data models. An advantage of the parametric interpretation within these disciplines is that there is an understanding of optimal procedures. Also, analytic expressions are available to characterize the achievable accuracy or reliability. Hence, for optimal system design, it is desirable to formulate a suitable probabilistic measurement data model as a function of the acquisition apparatus and the physical effects acting on it. Together with technology cost models, physical sensor layouts can be determined, which offer a favorable complexity-performance trade-off. Based on these considerations, system engineers can give detailed design recommendations to analog front-end engineers and digital software engineers.

A. Motivation

The purpose of this article is to highlight the opportunities and challenges of a hardware-aware system design approach. We limit ourselves to sensing systems where the analog-to-digital (A/D) converters only forward the information concerning the sign of the analog sensor amplitude to the digital processing units. The reasoning is that amplitude resolution increases A/D complexity exponentially. Each additional bit at least doubles the demand on A/D resources [1]. Thus, from a complexity-aware perspective, digitization with more than 1-bit amplitude resolution appears unfavorable, in particular, for applications where most bits represent uninformative noise. Furthermore, binary digitization is beneficial for low-level processing, where data can be handled with low logic complexity. However, these savings are obtained by accepting a considerable loss of information during signal acquisition. Fortunately, a thorough analysis shows that probabilistic modeling of the transition from a physical phenomenon to binary measurements, front-end optimization, and likelihood-oriented data processing can compensate for the effect of digitization-induced distortion on the final inference solution, see, e.g., [2].

While the potential of binary sampling has been extensively studied, e.g., regarding wireless communication capabilities [3]–[10], signal reconstruction error [11]–[13], estimation sensitivity [14]–[19], and detection reliability [20]–[22], here we focus on the analysis of the sensing latency [23].

To this end, the binary data stream is assumed to be processed in short spatio-temporal blocks to quickly detect which of two possible physical scenarios is generating the measurements. The reliability level to be achieved is predetermined. Leaning on concepts from sequential analysis [24], [25], we characterize the average number of blocks required for reliable detection when binary sensors are employed. A challenge that we address is the fact that the probabilistic data models characterizing multivariate binary measurements are, in general, intractable. A workaround, aiming at approximating likelihood-based sequential testing, enables analyzing the detection latency as a function of the binary sensing architecture and the two underlying physical models.

Applications related to mobile communication and satellite-based synchronization systems are outlined. On the one hand, we consider cognitive radio where a secondary user observes the activity of a primary user [26]. Once the primary wireless transmitter is inactive, the secondary system uses the communication channel. Thus, without limiting the functionality of the primary user or occupying additional bands, spectral resources can be used for wireless services. Since mobile radio systems necessitate a cost-efficient and miniaturized design, reliable operation at low complexity is crucial. On the other hand, we treat the monitoring of safety-critical global navigation satellite system (GNSS) frequencies [27]. Due to the low power and the distance of the GNSS transmitters, terrestrial radio interference poses a challenge. Small jamming devices, which protect the privacy of individuals, can impair the functionality of critical infrastructure (e.g., financial markets, power networks, airports, communication networks), which ensures the safety and well-being of many people [28]. For monitoring the GNSS spectrum in the vicinity of safety-critical receivers, high performance is of utmost importance.

Both applications have in common that during system development, a favorable complexity-performance trade-off has to be identified. In addition to reaching reliability, minimizing latency plays a decisive role. A cognitive system with a small decision delay enables using spectral resources efficiently. Monitoring sensors that quickly detect interference can initiate measures for suppression or declare malfunctioning in time. For such architectures, we perform hardware-aware modeling of the analog front-ends and the resulting binary measurements. Our findings characterize how the temporal oversampling and the number of sensors affect the sequential detection latency. Additionally, we use A/D cost models to identify cases in which it might be advantageous to deploy higher A/D resolution. We close the discourse by evaluating the developed sequential algorithms and illustrating the accuracy of our latency analysis utilizing synthetically generated data streams from exemplary binary radio systems.

B. Related Work

Quantized sequential decision-making has been studied predominantly for sensor networks where measurement nodes forward compressed statistics of their observations to minimize the communication overhead [29]–[35]. In such a setup, it is usually assumed that the sensing nodes have access to the unquantized digital observations and compress them, for example, by quantizing the likelihood ratio. Here we discuss sequential detection when employing binary sensing front-ends. Therefore, the nodes do not have access to the high-resolution observations, and the centralized decision is exclusively based on the hard-limited measurements of all sensing nodes. This case is less well studied, and the focus of existing works is on the design of optimal quantizers [23], [36]–[38]. Note that works studying the sensing latency in centralized sequential detection with quantized measurements usually consider data models with deterministic signals in white noise, see, e.g., [23], [39], [40].

C. Contribution

In contrast, we consider quantizers of minimal complexity. Therefore, the analog signals are converted into binary data by symmetric hard-limiting. Furthermore, we assume random signals with temporal and spatial dependencies. Sequential analysis then exhibits particular challenges which, to the best of our knowledge, have not been addressed yet. Due to the highly nonlinear signal acquisition, for rigorous analysis, one cannot rely on linear system theory and Gaussian statistics when modeling the sensor data. Such methods only provide accurate results when a sufficiently high A/D resolution is deployed. On the other hand, with likelihood-based approaches, see, e.g., [41]–[43], the analysis becomes challenging, when considering correlation at the hard-limiter input. For the resulting multivariate binary data, the mass function and the sufficient statistics, in general, grow exponentially with the number of variables [44]. Furthermore, the unknown orthant probabilities [45] with more than four variables (an open problem in statistics) hinder access to the likelihood. We tackle this by exploiting the properties of the exponential family. Such a generic probabilistic perspective onto hardware-aware signal processing systems [46] provides an approximate log-likelihood ratio (ALLR) and, therefore, enables performing a sequential probability ratio test (SPRT) for a broad class of data models without direct access to the likelihood ratio. Additionally, the approach provides tuneable approximations for the Kullback-Leibler divergence characterizing the average sampling number (ASN) of the approximate SPRT (ASPRT). Further, using an auxiliary distribution of reduced statistical complexity, we ensure computational tractability. By studying the efficiency of binary radio sensor layouts via our results, we show that with temporal and spatial oversampling, these systems master challenging sequential detection tasks. In particular, the possibility to deploy more antennas enables sensing with a superior complexity-performance trade-off. Note that this article is an extension of our special session contribution [47], where we discussed preliminary results.

II. PROBLEM FORMULATION

A. System Model - Signal Acquisition

We consider $M \in \mathbb{N}$ analog sensor outputs, modeled as real-valued time-continuous functions

$$\boldsymbol{y}(t) = \boldsymbol{s}(t) + \boldsymbol{\eta}(t). \tag{1}$$

The analog measurements $\boldsymbol{y}(t) \in \mathbb{R}^M, t \in \mathbb{R}$, are the superposition of a source component $\boldsymbol{s}(t) \in \mathbb{R}^M$ and additive independent measurement noise $\boldsymbol{\eta}(t) \in \mathbb{R}^M$. Both signals are modeled as band-limited wide-sense stationary Gaussian processes with zero mean. The analog signals (1) are synchronously discretized in batches of $K \in \mathbb{N}$ equidistant sampling points in time. With infinite digital amplitude resolution, the nth observation instance provides a space-time sample

$$\boldsymbol{y}_n = \begin{bmatrix} \boldsymbol{y}_n^{\mathrm{T}}[1] & \boldsymbol{y}_n^{\mathrm{T}}[2] & \dots & \boldsymbol{y}_n^{\mathrm{T}}[K] \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{MK}, \tag{2}$$

where, with the sampling interval $T \in \mathbb{R}$, we write

$$\boldsymbol{y}_n[k] = \boldsymbol{y}\big((n-1)KT + (k-1)T\big), \quad k = 1, \dots, K. \tag{3}$$

Note, that the sampling duration for each block (2) is $T_o = KT$. We consider the dependencies between consecutive samples as negligible, such that, for each $n \in \mathbb{N}$, (2) can be considered to be independent. Due to stationarity and Gaussianity, the spatio-temporal data (2) follows the model

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{\eta}, \quad \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{\eta} \in \mathbb{R}^{MK}, \quad \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{y}}(\boldsymbol{\theta})), \tag{4}$$

where the space-time covariance matrix

$$\begin{aligned} \boldsymbol{R}_{\boldsymbol{y}}(\boldsymbol{\theta}) &= \mathrm{E}_{\boldsymbol{y};\boldsymbol{\theta}} \left[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}} \right] \\ &= \boldsymbol{R}_{\boldsymbol{s}}(\boldsymbol{\theta}) + \boldsymbol{R}_{\boldsymbol{\eta}}, \quad \boldsymbol{R}_{\boldsymbol{y}}(\boldsymbol{\theta}) \in \mathbb{R}^{MK \times MK} \end{aligned} \tag{5}$$

is a superposition of the source covariance $\boldsymbol{R}_{\boldsymbol{s}}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{s};\boldsymbol{\theta}} \left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}} \right]$ and the noise covariance $\boldsymbol{R}_{\boldsymbol{\eta}} = \mathrm{E}_{\boldsymbol{\eta}} \left[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}} \right]$. Note that the source covariance is a function of the parameters $\boldsymbol{\theta} \in \mathbb{R}^D$, while the noise covariance is constant.

Realizing a signal acquisition, which approximately produces a data stream according to the multivariate Gaussian model (4), in practice, requires an A/D converter with several bits digital amplitude resolution at each analog output. To minimize A/D complexity, we here assume that within the considered system, only the signs of the analog measurements (1) are transferred to the digital processing units. Such a binary signal acquisition results in the space-time observations

$$\boldsymbol{z}_n = \mathrm{sign}\,(\boldsymbol{y}_n), \tag{6}$$

where the element-wise hard-limiter $\mathrm{sign}\,(\cdot)$ is defined

$$[\boldsymbol{z}_n]_i = \begin{cases} +1 & \text{if } [\boldsymbol{y}_n]_i \geq 0, \\ -1 & \text{if } [\boldsymbol{y}_n]_i < 0. \end{cases} \tag{7}$$

Per space-time sample (2), analog-to-binary (A/B) conversion (6) can be realized by K comparator operations for each of the M analog outputs, while a b-bit receiver requires to activate

$$\mathrm{SC}^{(b)}(M, K) = MK(2^b - 1), \quad b \in \mathbb{N}, \tag{8}$$

comparators. Further, the binary data (6) can be stored on small memory, transmitted using channels with moderate capacity, and preprocessed at a high rate and low computational cost.

B. Processing Task - Sequential Decision-Making

The binary measurements (6) gathered up to the nth observation instance, are summarized

$$\boldsymbol{Z}_n = \begin{bmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \dots & \boldsymbol{z}_n \end{bmatrix}, \quad \boldsymbol{Z}_n \in \mathbb{B}^{MK \times n}. \tag{9}$$

The inference task is to use the available data $\boldsymbol{Z}_n$ to decide which of the two possible probability laws

$$\mathcal{H}_0 : \boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z}; \boldsymbol{\theta}_0) \quad \text{or} \quad \mathcal{H}_1 : \boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z}; \boldsymbol{\theta}_1) \tag{10}$$

is the model generating the output data (6). The detection is to be conducted reliably, i.e.,

$$\Pr\{\text{decision } \mathcal{H}_0 | \mathcal{H}_1\} \leq \alpha_0, \tag{11}$$

$$\Pr\{\text{decision } \mathcal{H}_1 | \mathcal{H}_0\} \leq \alpha_1. \tag{12}$$

If a decision based on $\boldsymbol{Z}_n$ would lead to a violation of these reliability constraints, the processing unit waits for the next space-time sample $\boldsymbol{z}_{n+1}$ and tries to perform the test with the augmented data stream $\boldsymbol{Z}_{n+1}$. The instance in which the detection is finally performed is denoted by $n_{\mathrm{D}} \in \mathbb{N}$. The performance of the sequential test is characterized by the average sampling number (ASN), which is defined as the expected value of n_{D} under the data-generating model $p_{\boldsymbol{z}}(\boldsymbol{z}; \boldsymbol{\theta})$, i.e.,

$$\mathrm{ASN}(\boldsymbol{\theta}) = \mathrm{E}_{n_{\mathrm{D}};\boldsymbol{\theta}} \left[n_{\mathrm{D}} \right]. \tag{13}$$

Note that the ASN characterizes expected latency, i.e., individual decisions might require less or more samples (6). A classical approach to construct a decision-making algorithm minimizing (13) is the sequential probability ratio test (SPRT) [24]. Given the data stream $\boldsymbol{Z}_n$, the log-likelihood ratio (LLR)

$$l(\boldsymbol{Z}_n) = \sum_{i=1}^{n} l(\boldsymbol{z}_i) = \sum_{i=1}^{n} \ln \frac{p_{\boldsymbol{z}}(\boldsymbol{z}_i; \boldsymbol{\theta}_1)}{p_{\boldsymbol{z}}(\boldsymbol{z}_i; \boldsymbol{\theta}_0)} \tag{14}$$

is calculated and compared against two decision thresholds. If

$$l(\boldsymbol{Z}_n) \leq \ln \frac{\alpha_1}{1 - \alpha_0} = L_0, \tag{15}$$

the test is stopped with a decision in favor of $\mathcal{H}_0$. In case

$$l(\boldsymbol{Z}_n) \geq \ln \frac{1 - \alpha_1}{\alpha_0} = L_1, \tag{16}$$

the sequential decision-making is terminated in favor of the hypothesis $\mathcal{H}_1$. Otherwise, an additional signal sample $\boldsymbol{z}_{n+1}$ is taken to continue the test with $\boldsymbol{Z}_{n+1}$. With the short notations

$$N_0 = (1 - \alpha_0) \ln \frac{\alpha_1}{1 - \alpha_0} + \alpha_0 \ln \frac{1 - \alpha_1}{\alpha_0}, \tag{17}$$

$$N_1 = \alpha_1 \ln \frac{\alpha_1}{1 - \alpha_0} + \left(1 - \alpha_1\right) \ln \frac{1 - \alpha_1}{\alpha_0}, \tag{18}$$

the ASN of the SPRT under the two possible data model hypotheses (10) is approximately [24]

$$\mathrm{ASN}_0 \approx \frac{N_0}{\mathrm{E}_{\boldsymbol{z};\boldsymbol{\theta}_0} \left[l(\boldsymbol{z}) \right]} = -\frac{N_0}{D(p_{\boldsymbol{z};\boldsymbol{\theta}_0} || p_{\boldsymbol{z};\boldsymbol{\theta}_1})} \tag{19}$$

and

$$\mathrm{ASN}_1 \approx \frac{N_1}{\mathrm{E}_{\boldsymbol{z};\boldsymbol{\theta}_1} \left[l(\boldsymbol{z}) \right]} = \frac{N_1}{D(p_{\boldsymbol{z};\boldsymbol{\theta}_1} || p_{\boldsymbol{z};\boldsymbol{\theta}_0})}, \tag{20}$$

where $D(p_{\boldsymbol{u};\boldsymbol{\theta}} || q_{\boldsymbol{w};\boldsymbol{\theta}})$ denotes the Kullback–Leibler divergence between the distributions $p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta})$ and $q_{\boldsymbol{w}}(\boldsymbol{w}; \boldsymbol{\theta})$.

C. Challenge - Data Models for Binary Measurements

While A/B conversion, according to (6), offers significant savings regarding hardware cost and power consumption, the probabilistic characterization of the binary sensor outputs forms a challenge. To obtain the exact binary likelihood ratio required in (14), the multidimensional integral

$$p_{\boldsymbol{z}}(\boldsymbol{z}; \boldsymbol{\theta}) = \int_{\boldsymbol{\mathcal{Y}}(\boldsymbol{z})} p_{\boldsymbol{y}}(\boldsymbol{y}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y} \qquad (21)$$

needs to be evaluated, where $p_{\boldsymbol{y}}(\boldsymbol{y}; \boldsymbol{\theta})$ denotes the distribution of the data at the input to (6) and

$$\boldsymbol{\mathcal{Y}}(\boldsymbol{z}) = \left\{ \boldsymbol{y} \in \mathbb{R}^{MK} \mid \boldsymbol{z} = \mathrm{sign}\,(\boldsymbol{y}) \right\}. \qquad (22)$$

The calculation of an integral like (21) can turn out to be challenging. If, like in our case, the input to (6) is multivariate Gaussian, evaluation of (21) requires the orthant probabilities, for which solutions are only known for $MK \leq 4$. Even if these probabilities were available, the memory required to store all possible values for a single hypothesis scales as $\mathcal{O}(2^{MK})$. This renders using (14) for sequential detection and analyzing the resulting ASN by (19) and (20) prohibitively complicated even for scenarios with moderately large M and K.

III. LIKELIHOOD RATIO IN THE EXPONENTIAL FAMILY

A conceptual observation that turns out useful is that distributions of multivariate binary measurements (6) can be represented within the framework of the exponential family.

A. Exponential Family Data Models

A distribution belongs to the exponential family if

$$p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta}) = \exp\left(\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{\phi}(\boldsymbol{u}) - \lambda(\boldsymbol{\theta}) + \nu(\boldsymbol{u}) \right), \qquad (23)$$

where $\boldsymbol{u} \in \mathcal{U}$ is the V-variate data with support $\mathcal{U}$, $\boldsymbol{\theta} \in \mathbb{R}^D$ the physical parameters, $\boldsymbol{\beta}(\boldsymbol{\theta})\colon \mathbb{R}^D \to \mathbb{R}^C$ the statistical weights[1], $\boldsymbol{\phi}(\boldsymbol{u})\colon \mathcal{U} \to \mathbb{R}^C$ the sufficient statistics, $\lambda(\boldsymbol{\theta})\colon \mathbb{R}^D \to \mathbb{R}$ the log-normalizer and $\nu(\boldsymbol{u})\colon \mathcal{U} \to \mathbb{R}$ the carrier measure. While the multivariate Gaussian model (4) also factorizes according to (23), the number of its sufficient statistics C scales as $\mathcal{O}(V^2)$. For multivariate binary data, in contrast, these statistics scale as $\mathcal{O}(2^V)$. This is due to the fact that in multivariate binary distributions the sufficient statistics not only comprise the pairwise products between all variables but also all higher-order products [44]. Therefore, the LLR

$$l(\boldsymbol{u}) = \ln \frac{p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta}_1)}{p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta}_0)}, \qquad (24)$$

required in (14), can be inconvenient to handle. In the following, we discuss LLR approximations

$$\tilde{l}(\boldsymbol{u}) \approx l(\boldsymbol{u}), \qquad (25)$$

[1] Usually the term "natural parameters" is used for $\boldsymbol{\beta}(\boldsymbol{\theta})$. We use "statistical weights" to emphasize that, in our engineering-oriented perspective, a probabilistic data model $p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta})$ forms the connection between an analog physical/natural phenomenon $\boldsymbol{\theta}$ and digital measurement data $\boldsymbol{u}$.

enabling to implement the test defined in (14)–(16) and assess its latency (19) and (20) via

$$\mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i}\left[l(\boldsymbol{u}) \right] \approx \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i}\left[\tilde{l}(\boldsymbol{u}) \right], \quad i = 0, 1. \qquad (26)$$

To this end, note that the exact LLR between two hypotheses within the exponential family (23) is given by

$$l(\boldsymbol{u}) = \left(\boldsymbol{\beta}(\boldsymbol{\theta}_1) - \boldsymbol{\beta}(\boldsymbol{\theta}_0) \right)^{\mathrm{T}} \boldsymbol{\phi}(\boldsymbol{u}) - \left(\lambda(\boldsymbol{\theta}_1) - \lambda(\boldsymbol{\theta}_0) \right), \quad (27)$$

such that, with the mean of the sufficient statistics

$$\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}}\left[\boldsymbol{\phi}(\boldsymbol{u}) \right] \qquad (28)$$

and their covariance matrix

$$\boldsymbol{R}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}}\left[\left(\boldsymbol{\phi}(\boldsymbol{u}) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \right)\left(\boldsymbol{\phi}(\boldsymbol{u}) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \right)^{\mathrm{T}} \right], \quad (29)$$

the mean of the LLR is given by

$$\begin{aligned} \mu_i &= \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i}\left[l(\boldsymbol{u}) \right] \\ &= \left(\boldsymbol{\beta}(\boldsymbol{\theta}_1) - \boldsymbol{\beta}(\boldsymbol{\theta}_0) \right)^{\mathrm{T}} \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_i) - \left(\lambda(\boldsymbol{\theta}_1) - \lambda(\boldsymbol{\theta}_0) \right), \quad (30) \end{aligned}$$

while the variance of (27) can be written as

$$\begin{aligned} \sigma_i^2 &= \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i}\left[\left(l(\boldsymbol{u}) - \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i}\left[l(\boldsymbol{u}) \right] \right)^2 \right] \\ &= \left(\boldsymbol{\beta}(\boldsymbol{\theta}_1) - \boldsymbol{\beta}(\boldsymbol{\theta}_0) \right)^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_i)\left(\boldsymbol{\beta}(\boldsymbol{\theta}_1) - \boldsymbol{\beta}(\boldsymbol{\theta}_0) \right). \quad (31) \end{aligned}$$

B. Approximations for the Exponential Family LLR

In practice, access to the statistical weights $\boldsymbol{\beta}(\boldsymbol{\theta})$ and the log-normalizer $\lambda(\boldsymbol{\theta})$, for executing and analyzing likelihood-based tests, can be challenging to obtain. For example, for the binary output (6), the integral providing the log-normalizer

$$\lambda(\boldsymbol{\theta}) = \ln \int_{\mathcal{U}} \exp\left(\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{\phi}(\boldsymbol{u}) + \nu(\boldsymbol{u}) \right) \mathrm{d}\boldsymbol{u} \qquad (32)$$

results in a sum with 2^V terms. To obtain a representation of (27) which does not require explicit access to $\boldsymbol{\beta}(\boldsymbol{\theta})$ and $\lambda(\boldsymbol{\theta})$, we assume to have at hand (28) and (29) as functions of $\boldsymbol{\theta}$. These measures are usually easier to obtain than $\boldsymbol{\beta}(\boldsymbol{\theta})$ and $\lambda(\boldsymbol{\theta})$. To link (28) and (29) to the exponential family LLR (27), we use that all distributions (23) exhibit regularity, i.e.,

$$\mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}}\left[\frac{\partial \ln p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = \boldsymbol{0}^{\mathrm{T}}. \qquad (33)$$

Consequently, for any exponential family (23), it holds that

$$\left(\frac{\partial \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} = \left(\frac{\partial \boldsymbol{\beta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}). \qquad (34)$$

Defining an LLR linearization point

$$\tilde{\boldsymbol{\theta}}(\xi) = \xi\boldsymbol{\theta}_0 + (1 - \xi)\boldsymbol{\theta}_1, \quad \xi \in [0, 1], \qquad (35)$$

and applying the finite difference approximation (116) in the Appendix together with the regularity constraint (34), gives

$$\begin{aligned} \lambda(\boldsymbol{\theta}_1) - \lambda(\boldsymbol{\theta}_0) &\approx \frac{\partial \lambda(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) \\ &= \boldsymbol{\mu}_{\boldsymbol{\phi}}^{\mathrm{T}}(\tilde{\boldsymbol{\theta}}(\xi)) \frac{\partial \boldsymbol{\beta}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0). \quad (36) \end{aligned}$$

Further, for exponential distributions (23), one obtains [46]

$$\frac{\partial \boldsymbol{\beta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{37}$$

such that (36) can be reformulated as

$$\lambda(\boldsymbol{\theta}_1) - \lambda(\boldsymbol{\theta}_0) \approx \boldsymbol{\mu}_{\boldsymbol{\phi}}^{\mathrm{T}}(\tilde{\boldsymbol{\theta}}(\xi)) \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0). \tag{38}$$

Accordingly, the difference of the statistical weights in (27) can be approximated by

$$\begin{aligned} \boldsymbol{\beta}(\boldsymbol{\theta}_1) - \boldsymbol{\beta}(\boldsymbol{\theta}_0) &\approx \frac{\partial \boldsymbol{\beta}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) \\ &= \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0). \end{aligned} \tag{39}$$

Defining an LLR hyperplane

$$\boldsymbol{b}_{\partial}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) = \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0), \tag{40}$$

and using (38) and (39) in (27), provides the approximation

$$l(\boldsymbol{u}) \approx \boldsymbol{b}_{\partial}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \big(\boldsymbol{\phi}(\boldsymbol{u}) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \big) = \tilde{l}_{\partial}(\boldsymbol{u}; \xi). \tag{41}$$

The structure of this approximate LLR (ALLR) enables interpreting the LLR of exponential family distributions (27) as the signed distance of a suffcient statistics residual from the hyperplane (40). Note, that (40) requires access to the derivative of the mean (28) with respect to $\boldsymbol{\theta}$ evaluated at $\tilde{\boldsymbol{\theta}}(\xi)$. By using the finite difference (116) to eliminate the derivative in (40), the LLR hyperplane can also be written as

$$\boldsymbol{b}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) = \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \big(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_1) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) \big). \tag{42}$$

Therefore, an alternative to the LLR approximation (41) is

$$l(\boldsymbol{u}) \approx \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \big(\boldsymbol{\phi}(\boldsymbol{u}) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \big) = \tilde{l}(\boldsymbol{u}; \xi). \tag{43}$$

Evaluating the ALLR (43) requires access to the mean (28) with respect to $p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta}_0)$, $p_{\boldsymbol{u}}(\boldsymbol{u}; \boldsymbol{\theta}_1)$, and $p_{\boldsymbol{u}}(\boldsymbol{u}; \tilde{\boldsymbol{\theta}}(\xi))$. Further, one requires the covariance (29) with respect to $p_{\boldsymbol{u}}(\boldsymbol{u}; \tilde{\boldsymbol{\theta}}(\xi))$. Note that integrals like (21) and (32) are not required for evaluating the ALLRs (41) and (43).

IV. Approximate Tests in the Exponential Family

A. Approximate Sequential Probability Ratio Test

Defining the empirical mean of the suffcient statistics

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi}}(\boldsymbol{U}_n) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\phi}(\boldsymbol{u}_i), \tag{44}$$

and using the LLR approximation (41) or (43), for any data stream associated with an exponential family model (23), an approximate SPRT (ASPRT) can be performed by comparing

$$\begin{aligned} \tilde{l}_{\partial}(\boldsymbol{U}_n; \xi) &= \sum_{i=1}^{n} \tilde{l}_{\partial}(\boldsymbol{u}_i; \xi) \\ &= n \boldsymbol{b}_{\partial}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \big(\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi}}(\boldsymbol{U}_n) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \big) \end{aligned} \tag{45}$$

or

$$\tilde{l}(\boldsymbol{U}_n; \xi) = n \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \big(\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi}}(\boldsymbol{U}_n) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \big), \tag{46}$$

to the decision thresholds (15) and (16). Note that with such an approach (40) or (42) are determined offline such that during run-time, the sequential test consists of updating (44), taking a vector difference, and calculating an inner product. The decision-making latency (13) can be assessed using (19) and (20) with the mean of the ALLR

$$\begin{aligned} \tilde{\mu}_{\partial,i}(\xi) &= \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i} \big[\tilde{l}_{\partial}(\boldsymbol{u}; \xi) \big] \\ &= \boldsymbol{b}_{\partial}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \big(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_i) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \big), \end{aligned} \tag{47}$$

or

$$\tilde{\mu}_i(\xi) = \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \big(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_i) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \big). \tag{48}$$

Further, the variance of the ALLR is

$$\begin{aligned} \tilde{\sigma}_{\partial,i}^2(\xi) &= \mathrm{E}_{\boldsymbol{u}; \boldsymbol{\theta}_i} \left[\big(\tilde{l}_{\partial}(\boldsymbol{u}; \xi) - \tilde{\mu}_{\partial,i}(\xi) \big)^2 \right] \\ &= \boldsymbol{b}_{\partial}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \boldsymbol{R}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_i) \boldsymbol{b}_{\partial}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi), \end{aligned} \tag{49}$$

or

$$\tilde{\sigma}_i^2(\xi) = \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi) \boldsymbol{R}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_i) \boldsymbol{b}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1; \xi). \tag{50}$$

B. Tuning of the Probabilistic Linearization Model

The choice of the probabilistic linearization model $p_{\boldsymbol{u}}(\boldsymbol{u}; \tilde{\boldsymbol{\theta}}(\xi))$ impacts the accuracy of (41) and (43). While one can use $\xi = \frac{1}{2}$, we propose a heuristic method to adapt the approximation parameter ξ. The idea is as follows: Instead of using the *geometric* midpoint between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, we use the *statistical* midpoint. That is, we tune ξ such that the approximated test statistic $\tilde{l}(\boldsymbol{u})$ admits the same properties under both hypotheses. More precisely, we consider the standardized drift

$$\tilde{d}_i(\xi) = \frac{|\tilde{\mu}_i(\xi)|}{\tilde{\sigma}_i^{\rho}(\xi)}, \tag{51}$$

where $\rho > 0$ can be chosen freely. Note that (51) is closely related to the error probabilities of the underlying statistical test: for $d_i = 0$, the test does not admit a drift towards any threshold such that it decides randomly; for $d_i \to \infty$, the mean of the LLR dominates the variance making the test decide correctly for $\mathcal{H}_i$ with probability one. To balance the decision-making performances under both hypotheses, it needs to hold that $\tilde{d}_0(\xi) \approx \tilde{d}_1(\xi)$. Therefore, we define the ratio

$$\tilde{\nu}(\xi) = \frac{|\tilde{\mu}_1(\xi)| \tilde{\sigma}_0^{\rho}(\xi)}{|\tilde{\mu}_0(\xi)| \tilde{\sigma}_1^{\rho}(\xi)} \tag{52}$$

and choose the linearization parameter ξ such that the difference between the drifts is minimized

$$\xi^* = \underset{\xi \in [0;1]}{\arg\min} \big(\tilde{\nu}(\xi) - 1 \big)^2. \tag{53}$$

This approach results in high-quality LLR approximations, as illustrated by examples in Sec. V.

C. Approximations for the Kullback–Leibler Divergence

Within the exponential family (23), the results (47) and (48) imply approximations for the Kullback–Leibler divergence

$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_0}||p_{\boldsymbol{u};\boldsymbol{\theta}_1}) \approx -(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \left(\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}}$$
$$\cdot \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \left(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \right), \quad (54)$$

$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_1}||p_{\boldsymbol{u};\boldsymbol{\theta}_0}) \approx (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \left(\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi))}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}}$$
$$\cdot \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \left(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_1) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \right) \quad (55)$$

or

$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_0}||p_{\boldsymbol{u};\boldsymbol{\theta}_1}) \approx -\left(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_1) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) \right)^{\mathrm{T}}$$
$$\cdot \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \left(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \right), \quad (56)$$

$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_1}||p_{\boldsymbol{u};\boldsymbol{\theta}_0}) \approx \left(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_1) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) \right)^{\mathrm{T}}$$
$$\cdot \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\theta}}(\xi)) \left(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_1) - \boldsymbol{\mu}_{\boldsymbol{\phi}}(\tilde{\boldsymbol{\theta}}(\xi)) \right). \quad (57)$$

With the Fisher information matrix for exponential family distributions (23) being characterized by [46]

$$\boldsymbol{F}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{u};\boldsymbol{\theta}} \left[\left(\frac{\partial \ln p_{\boldsymbol{u}}(\boldsymbol{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \frac{\partial \ln p_{\boldsymbol{u}}(\boldsymbol{u};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]$$
$$= \left(\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\phi}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad (58)$$

applying forward (114) and backward (115) approximations to (54) and (55), one also obtains

$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_0}||p_{\boldsymbol{u};\boldsymbol{\theta}_1}) \approx (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \boldsymbol{F}(\tilde{\boldsymbol{\theta}}(\xi))(\tilde{\boldsymbol{\theta}}(\xi) - \boldsymbol{\theta}_0), \quad (59)$$
$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_1}||p_{\boldsymbol{u};\boldsymbol{\theta}_0}) \approx (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \boldsymbol{F}(\tilde{\boldsymbol{\theta}}(\xi))(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}(\xi)). \quad (60)$$

These Kullback–Leibler divergence approximations are reminiscent of expressions found in the literature [48, pp. 85-86]

$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_0}||p_{\boldsymbol{u};\boldsymbol{\theta}_1}) \approx -\frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \boldsymbol{F}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$$
$$= \hat{D}(p_{\boldsymbol{u};\boldsymbol{\theta}_0}||p_{\boldsymbol{u};\boldsymbol{\theta}_1}), \quad (61)$$
$$D(p_{\boldsymbol{u};\boldsymbol{\theta}_1}||p_{\boldsymbol{u};\boldsymbol{\theta}_0}) \approx \frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \boldsymbol{F}(\boldsymbol{\theta}_1)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$$
$$= \hat{D}(p_{\boldsymbol{u};\boldsymbol{\theta}_1}||p_{\boldsymbol{u};\boldsymbol{\theta}_0}), \quad (62)$$

while having a flexible probabilistic linearization point $\tilde{\boldsymbol{\theta}}(\xi)$. Note that the approximations (61) and (62) are derived in [48] for the special case of a natural exponential family, i.e., under the restriction $\boldsymbol{\beta}(\boldsymbol{\theta}) = \boldsymbol{\theta}$. A potential disadvantage of (59) and (60) is that access to the Fisher information matrix (58) of the exponential family model $p_{\boldsymbol{u}}(\boldsymbol{u};\tilde{\boldsymbol{\theta}}(\xi))$ is required.

D. Controlling the Statistical Complexity of the ALLR

While (41) and (43) enable running and analyzing a likelihood-oriented test via (28) and (29) without direct access to the likelihood ratio (27), they do not per se solve its intractability with a large number of sufficient statistics C

(as with multivariate binary distributions). Under such circumstances, replacing the data model by an auxiliary version [46]

$$\tilde{p}_{\boldsymbol{u}}(\boldsymbol{u};\boldsymbol{\theta}) = \exp \left(\tilde{\boldsymbol{\beta}}^{\mathrm{T}}(\boldsymbol{\theta})\tilde{\boldsymbol{\phi}}(\boldsymbol{u}) - \tilde{\lambda}(\boldsymbol{\theta}) + \tilde{\nu}(\boldsymbol{u}) \right), \quad (63)$$

allows to control the ALLR complexity. In (63), $\tilde{\boldsymbol{\phi}}(\boldsymbol{u}) : \mathcal{U} \to \mathbb{R}^{\tilde{C}}$ is a subset ($\tilde{C} < C$) of the sufficient statistics in the model (23), i.e., $\boldsymbol{\phi}(\boldsymbol{u}) = \begin{bmatrix} \tilde{\boldsymbol{\phi}}^{\mathrm{T}}(\boldsymbol{u}) & \boldsymbol{\phi}'^{\mathrm{T}}(\boldsymbol{u}) \end{bmatrix}^{\mathrm{T}}$ while mean

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{\phi}}}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{u};\boldsymbol{\theta}} \left[\tilde{\boldsymbol{\phi}}(\boldsymbol{u}) \right] \quad (64)$$

and covariance

$$\boldsymbol{R}_{\tilde{\boldsymbol{\phi}}}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{u};\boldsymbol{\theta}} \left[\left(\tilde{\boldsymbol{\phi}}(\boldsymbol{u}) - \boldsymbol{\mu}_{\tilde{\boldsymbol{\phi}}}(\boldsymbol{\theta}) \right) \left(\tilde{\boldsymbol{\phi}}(\boldsymbol{u}) - \boldsymbol{\mu}_{\tilde{\boldsymbol{\phi}}}(\boldsymbol{\theta}) \right)^{\mathrm{T}} \right] \quad (65)$$

are equivalent when formed with respect to (63) or (23). It can then be shown that [46]

$$\boldsymbol{F}(\boldsymbol{\theta}) \succeq \tilde{\boldsymbol{F}}(\boldsymbol{\theta}), \quad (66)$$

where $\tilde{\boldsymbol{F}}(\boldsymbol{\theta})$ is the Fisher information matrix (58) of the auxiliary data model (63). Due to the matrix inequality (66) and the quadratic form of (59) and (60), (63) constitutes a conservative modification of the ALLRs (41) and (43).

When evaluating (64) and (65) for the hard-limited version of (4), we use an auxiliary model (63) with statistics [46]

$$\tilde{\boldsymbol{\phi}}(\boldsymbol{z}) = \boldsymbol{\Phi} \, \mathrm{vec}\left(\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}} \right), \quad (67)$$

where $\boldsymbol{\Phi} \in [0;1]^{\tilde{C} \times (MK)^2}$ is an elimination matrix canceling the duplicate and constant statistics on $\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}$. This reduces the binary data model to a quadratic exponential binary distribution [49] where the statistical complexity is $\tilde{C} = \frac{MK}{2}(MK - 1)$. For the calculation of the mean (64) and the covariance (65) with (67), we proceed as described in [50] and use the classic arcsine law [51, p. 284] together with recent results for the quadrivariate orthant probabilities [52]. Note that, through the concept of a reduced distribution (63), with (28) and (29) for a subset of sufficient statistics at hand, the presented framework is also applicable to other sensor data models like receiver systems with higher A/D resolution.

V. ACCURACY OF THE LLR APPROXIMATION

A. Digital Data Model and Error Measures

To analyze the accuracy of (41) and (43), we consider a K-variate Gaussian model with scalar parameter $\theta \in \mathbb{R}$

$$p_{\boldsymbol{y}}(\boldsymbol{y};\theta) = \frac{\exp \left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta)\boldsymbol{y} \right)}{\sqrt{(2\pi)^K \det \left(\boldsymbol{R}_{\boldsymbol{y}}(\theta) \right)}}. \quad (68)$$

Note, that by

$$\boldsymbol{y}^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta)\boldsymbol{y} = \mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta)\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}} \right)$$
$$= \mathrm{vec}^{\mathrm{T}}\left(\boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta) \right) \mathrm{vec}\left(\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}} \right), \quad (69)$$

(68) can be factorized along (23) by using $\nu(\boldsymbol{y}) = 0$ and

$$\boldsymbol{\beta}(\theta) = -\frac{1}{2} \mathrm{vec}\left(\boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta) \right), \quad (70)$$
$$\boldsymbol{\phi}(\boldsymbol{y}) = \mathrm{vec}\left(\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}} \right), \quad (71)$$
$$\lambda(\theta) = \frac{1}{2} \ln \det \left(\boldsymbol{R}_{\boldsymbol{y}}(\theta) \right) + \frac{K}{2} \ln 2\pi. \quad (72)$$

Additionally, with (71), one obtains

$$\boldsymbol{\mu}_{\boldsymbol{\phi}}(\theta) = \text{vec}\left(\boldsymbol{R}_{\boldsymbol{y}}(\theta)\right), \tag{73}$$

$$\boldsymbol{R}_{\boldsymbol{\phi}}(\theta) = 2\left(\boldsymbol{R}_{\boldsymbol{y}}(\theta) \otimes \boldsymbol{R}_{\boldsymbol{y}}(\theta)\right). \tag{74}$$

The exact expected LLRs between two distributions (68) are

$$\mu_0 = \frac{1}{2} \ln \frac{\det\left(\boldsymbol{R}_{\boldsymbol{y}}(\theta_0)\right)}{\det\left(\boldsymbol{R}_{\boldsymbol{y}}(\theta_1)\right)} + \frac{K}{2} - \frac{1}{2} \text{tr}\left(\boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta_1)\boldsymbol{R}_{\boldsymbol{y}}(\theta_0)\right), \tag{75}$$

$$\mu_1 = \frac{1}{2} \ln \frac{\det\left(\boldsymbol{R}_{\boldsymbol{y}}(\theta_0)\right)}{\det\left(\boldsymbol{R}_{\boldsymbol{y}}(\theta_1)\right)} - \frac{K}{2} + \frac{1}{2} \text{tr}\left(\boldsymbol{R}_{\boldsymbol{y}}^{-1}(\theta_0)\boldsymbol{R}_{\boldsymbol{y}}(\theta_1)\right). \tag{76}$$

Therefore, under the Gaussian distribution (68), we can evaluate the relative approximation errors

$$\tilde{\epsilon}_i(\xi) = \frac{\left|\tilde{\mu}_i(\xi)\right| - \left|\mu_i\right|}{\left|\mu_i\right|}, \tag{77}$$

and compare them to (61) and (62) through the expression

$$\hat{\epsilon}_i = \frac{\hat{D}(p_{\boldsymbol{u};\boldsymbol{\theta}_i}||p_{\boldsymbol{u};\boldsymbol{\theta}_j}) - D(p_{\boldsymbol{u};\boldsymbol{\theta}_i}||p_{\boldsymbol{u};\boldsymbol{\theta}_j})}{D(p_{\boldsymbol{u};\boldsymbol{\theta}_i}||p_{\boldsymbol{u};\boldsymbol{\theta}_j})}, \quad i \neq j. \tag{78}$$

B. Application - Sampling Random Gaussian Processes

To connect the multivariate Gaussian model (68) to a practical sensing application, we assume that it models the digital data stream obtained by sampling (1) with $M = 1$ via an ideal A/D converter featuring ∞-bit amplitude resolution. The continuous-time Gaussian random process $y(t) \in \mathbb{R}$ is assumed to be the superposition of a band-limited source signal $\breve{s}(t) \in \mathbb{R}$ and white noise $\breve{\eta}(t)$ after preprocessing with an ideal low-pass filter $h(t; B)$ of bandwidth $B = B_y$, i.e.,

$$y(t) = h(t; B_y) * \left(\breve{s}(t) + \breve{\eta}(t)\right). \tag{79}$$

The band-limited source $\breve{s}(t)$ features power spectral density $\Psi_s(\omega) = \Psi_s, \omega \in [-\Omega_s; \Omega_s]$ with $\Omega_s \leq \Omega_y$, and $\Psi_s(\omega) = 0$ elsewhere. Consequently, the source passes unaffected

$$s(t) = h(t; B_y) * \breve{s}(t) = \breve{s}(t). \tag{80}$$

The noise $\breve{\eta}(t)$ has constant power spectral density $\Psi_\eta(\omega) = \Psi_\eta$ on $\omega \in [-\Omega_\eta; \Omega_\eta]$ with $\Omega_\eta \gg \Omega_y$. Consequently,

$$\eta(t) = h(t; B_y) * \breve{\eta}(t) \neq \breve{\eta}(t). \tag{81}$$

Note that, by the Wiener–Khintchine theorem [53], the autocorrelation function of a band-limited random process $u(t)$ (bandwidth B_u or Ω_u) with power spectral density Ψ_u is

$$r_u(t) = \text{E}_u\left[u(\tau)u(\tau - t)\right] = \frac{1}{2\pi} \int_{-\Omega_u}^{\Omega_u} \Psi_u(\omega)\, \text{e}^{\text{j}\omega t}\, \text{d}\omega$$

$$= \frac{\Psi_u}{\pi} \frac{\sin\left(2\pi B_u t\right)}{t} = 2B_u \Psi_u \text{sinc}\left(2B_u t\right). \tag{82}$$

With the low-pass filter $h(t; B_y)$ including a gain-control factor $\frac{1}{\sqrt{2B_y \Psi_\eta}}$ and sampling K times at a rate $f_{\text{T}} = \frac{1}{T} = 2B_y$

$$\boldsymbol{R}_{\boldsymbol{y}}(\theta) = \boldsymbol{R}_{\boldsymbol{s}}(\theta) + \boldsymbol{R}_{\boldsymbol{\eta}} = \frac{\theta}{\kappa}\boldsymbol{\Sigma}(\kappa) + \boldsymbol{I}, \tag{83}$$

where the signal-to-noise ratio (SNR) is

$$\theta = \text{SNR} = \frac{\Psi_s}{\Psi_\eta}. \tag{84}$$

With a source oversampling factor $\kappa \in \mathbb{R}, \kappa = \frac{B_y}{B_s} \geq 1$, the source covariance matrix has entries

$$[\boldsymbol{\Sigma}(\kappa)]_{ij} = \text{sinc}\left(\frac{|i - j|}{\kappa}\right), \quad i, j = 1, \dots, K. \tag{85}$$

Note that, in contrast to the source covariance $\boldsymbol{R}_{\boldsymbol{s}}(\theta)$, in (83) the noise covariance $\boldsymbol{R}_{\boldsymbol{\eta}}$ is the identity matrix (temporally white noise) irrespective of κ as $f_{\text{T}} = 2B_y = \kappa 2B_s$.
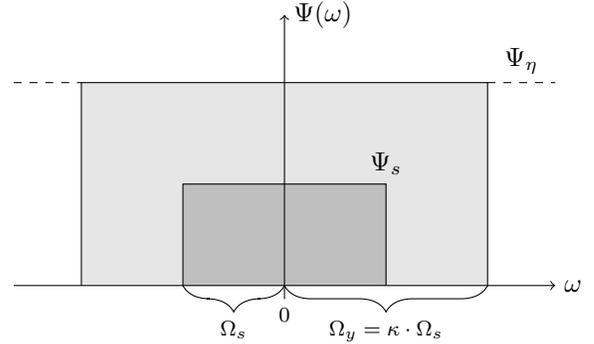

Fig. 1: Power Spectral Densities and Signal Bandwidths ($\kappa = 2$, SNR $= -3.0\,\text{dB}$)

For illustration, Fig. 1 visualizes an exemplary sensing situation with oversampling $\kappa = 2$ and SNR $= -3.0\,\text{dB}$ in the frequency domain. Note, that we define (84) independently of $h(t; B_y)$ and κ to ensure that the design of the analog preprocessing does not affect the physical SNR.

C. Results - Approximation Accuracy

Fig. 2a shows the relative errors (77) and (78) for a setting with $K = 10, \kappa = 2, \theta_0 = -20\,\text{dB}$ as a function of θ_1. It can be observed that the error $|\hat{\epsilon}_i|$ increases quickly with the distance $\theta_1 - \theta_0$ and beyond $\theta_1 = -10\,\text{dB}$ exceeds $27.1\,\%$. In contrast, for all considered SNR values, the error $|\epsilon_i(\frac{1}{2})|$ is not larger than $22.3\,\%$. Using (53) with $\rho = \frac{2}{3}$, the error $|\epsilon_i(\xi^*)|$ is below $2.6\,\%$ over the entire depicted SNR range. Fig. 2b shows ξ^* resulting from (53) with $\rho = \frac{2}{3}$.

Note that a typical sequential task is to detect small SNR changes from the digital measurements. To asses the quality of (43) in such a context, using the definition $\theta_0 = \bar{\theta} - \Delta_\theta$ and $\theta_1 = \bar{\theta} + \Delta_\theta$, in Fig. 3a, we visualize the relative errors (77) and (78) for $\Delta_\theta = 1.5\,\text{dB}, K = 10, \kappa = 2$, as a function of $\bar{\theta}$. The error $|\hat{\epsilon}_i|$ exceeds $16.6\,\%$ for $\bar{\theta} \geq -5\,\text{dB}$. The relative error $|\epsilon_i(\frac{1}{2})|$ stays below $9.2\,\%$ for all SNR values, while employing (53) with $\rho = \frac{2}{3}$ results in ξ^* shown in Fig. 3b and an error $|\epsilon_i(\xi^*)|$ smaller than $0.52\,\%$. The results for the exemplary model (68) show that (43) is of high quality, in particular if the probabilistic linearization $\tilde{\theta}(\xi)$ is optimized.
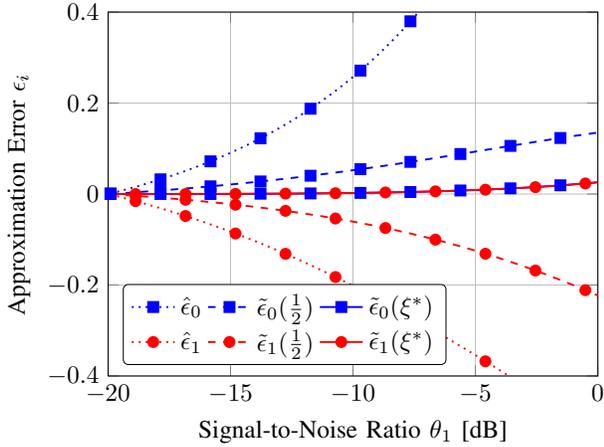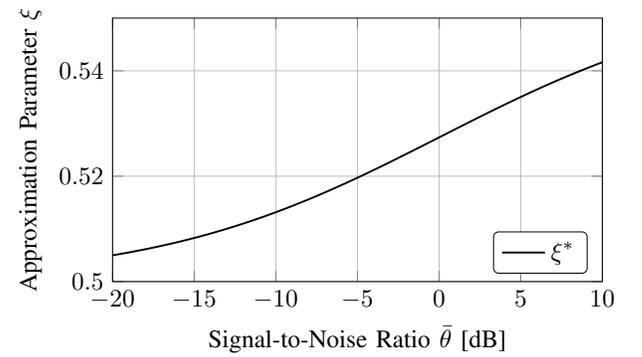
(a) Approximation Quality
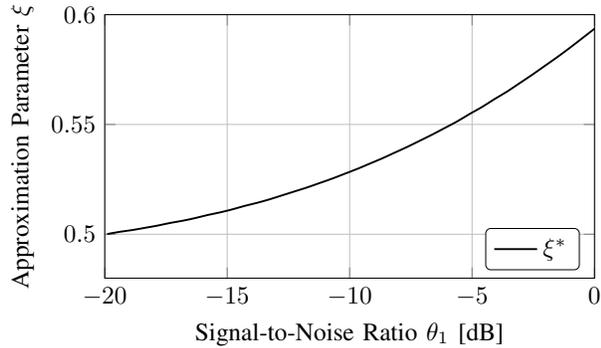
(b) Probabilistic Linearization Model

Fig. 2: ALLR - Approximation Quality and Tuning ($K = 10, \kappa = 2, \theta_0 = -20\,\text{dB}$)
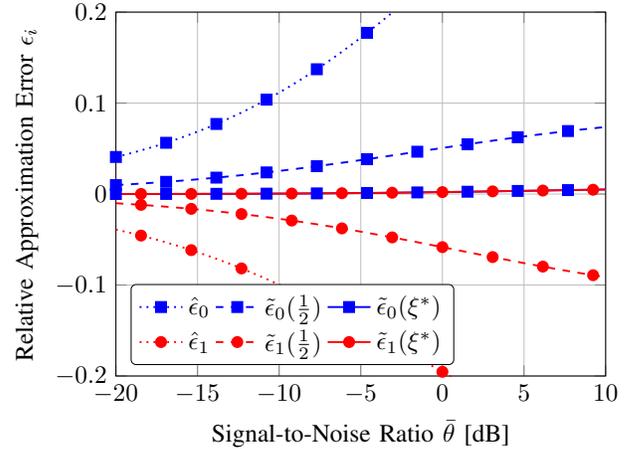
(a) Approximation Quality

(b) Probabilistic Linearization Model

Fig. 3: ALLR - Approximation Quality and Tuning ($K = 10, \kappa = 2, \Delta_\theta = 1.5\,\text{dB}$)

Fig. 4: Superheterodyne Analog Sensor Front-End

VI. APPLICATION – BINARY RADIO SYSTEM DESIGN

Next, the impact of theoretical results like (41) or (43) on technical systems is illustrated by exemplary applications in radio system engineering. To emphasize the significance of binary sensing and data processing for future wireless systems at large, we use a consumer-oriented cognitive radio application with low complexity as the main focus and a safety-critical spectrum monitoring application with a distinct emphasis on high performance. For each application, a particular analog front-end is used when modeling the sensor signals (1).

A. Low-Cost Cognitive Radio Communication Systems

First, we consider a cognitive system for mobile communication. The task of the receiver is to monitor a certain part of the spectrum and to determine whether a primary transmitter is currently utilizing the radio channel. The scenario is, therefore, similar to the example considered in Sec. V. However, when modeling the analog outputs (1), it is essential to take into account the front-end architecture. For cognitive radio, we assume a single superheterodyne receiver ($M = 1$), as depicted in Fig. 4. The difficulty in modeling its analog output lies in the fact that the receiver does not demodulate with the carrier frequency f_C of the transmitter. A local oscillator demodulates the received signal with $f_{LO} = f_C - f_{IF}$, where $f_{IF} < f_C$ is
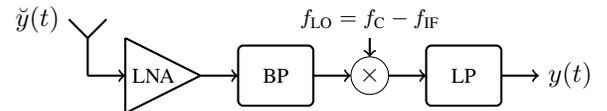
an intermediate frequency (IF). As a consequence, sampling is performed on one real-valued analog output

$$y(t) = x_I(t)\sqrt{2}\cos(\omega_{IF}t) - x_Q(t)\sqrt{2}\sin(\omega_{IF}t) + \eta(t), \quad (86)$$

where $x_I(t)$ and $x_Q(t)$ are assumed to be two jointly independent zero-mean Gaussian processes of bandwidth B_s modeling the transmitter. The independent zero-mean Gaussian process $\eta(t)$ models noise with bandwidth B_y. Sampling K times at a rate of $\frac{1}{T} = 2B_y$ and with an amplitude resolution of $b = \infty$ bits, one obtains zero-mean multivariate Gaussian data with

$$\boldsymbol{R_y}(\theta) = \frac{\theta}{\kappa}\boldsymbol{\Sigma}(\kappa) \odot 2\boldsymbol{W} + \boldsymbol{I}, \quad \boldsymbol{R_y}(\theta) \in \mathbb{R}^{K \times K}, \quad (87)$$

where θ is the physically received SNR as defined in (84), $\boldsymbol{\Sigma}(\kappa)$ the source covariance as defined in (85), and the matrix $\boldsymbol{W} \in \mathbb{R}^{K \times K}$ models the mixing effects at the intermediate frequency f_{IF}. For simplicity, we assume f_{IF} to be symmetric

with respect to the bandwidth B_y of the low-pass (LP) filter, i.e., $f_{\text{IF}} = \frac{B_y}{2}$. Thus the entries of the mixing matrix are

$$
\begin{aligned}
[\boldsymbol{W}]_{ij} &= \cos\left(\omega_{\text{IF}}T(i-1)\right)\cos\left(\omega_{\text{IF}}T(j-1)\right) \\
&\quad + \sin\left(\omega_{\text{IF}}T(i-1)\right)\sin\left(\omega_{\text{IF}}T(j-1)\right) \\
&= \cos\left(\frac{\pi}{2}(i-1)\right)\cos\left(\frac{\pi}{2}(j-1)\right) \\
&\quad + \sin\left(\frac{\pi}{2}(i-1)\right)\sin\left(\frac{\pi}{2}(j-1)\right).
\end{aligned}
\tag{88}
$$

Fig. 5 visualizes the power spectral density for a superheterodyne front-end and $\kappa = 4$, SNR $= -3.0\,\text{dB}$.
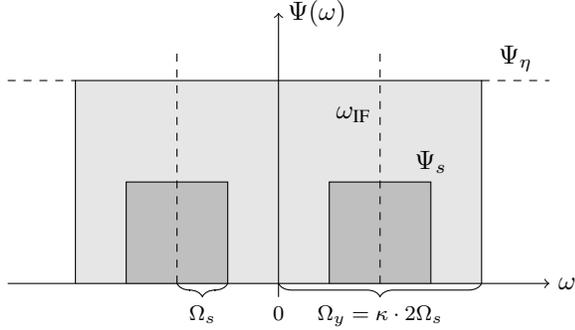
Fig. 5: Power Spectral Densities and Signal Bandwidths (Superheterodyne Receiver, $\kappa = 4$, SNR $= -3.0\,\text{dB}$)

For the considered receiver, the source bandwidth is $B_s \leq \frac{B_y}{2}$, such that $\kappa \geq 2$. For evaluation, we define

$$
\chi_i = \frac{\tilde{\mu}_i(\xi^*)}{\mu_i},
\tag{89}
$$

to compare the binary latency to the average run-time with high-resolution sampling. Here we assume a binary receiver with $K = \kappa K_0$ such that the absolute observation time $T_o = KT$ of each block stays constant when increasing κ.

Fig. 6a shows (89) for a low SNR scenario with $\bar{\theta} = -9\,\text{dB}$ and $\Delta_\theta = 1.5\,\text{dB}$. Without oversampling ($\kappa = 2$), binary sampling of the analog sensor signal (86) makes it impossible to perform the detection task. The activity of the transmitter can then only be detected by discriminating between two variance levels. The output of (6), however, is invariant to changes in its input scale. With oversampling, the presence of a band-limited source introduces correlations among the K samples (non-zero off-diagonal entries in (87)), which can be detected from the binary measurements. The ∞-bit receiver does not benefit from oversampling as the full information about the band-limited sources in (86) is already embedded in the digital measurement data obtained with $\kappa = 2$. Fig. 6b depicts the efficiency (89) at medium SNR ($\bar{\theta} = 0\,\text{dB}$, $\Delta_\theta = 1.5\,\text{dB}$). Like for the low SNR case in Fig. 6a, oversampling decreases detection latency. Fig. 6a and Fig. 6b show that, in general, a binary wireless receiver is suitable for cognitive radio. However, oversampling of the source signal is required, and a larger latency needs to be accepted in return for the simplicity of the radio front-end and the data structure.

To study complexity, we define the average sampling cost

$$
\text{ASC}^{(b)}(M, K; \boldsymbol{\theta}) = \text{SC}^{(b)}(M, K) \cdot \text{ASN}^{(b)}(\boldsymbol{\theta}).
\tag{90}
$$
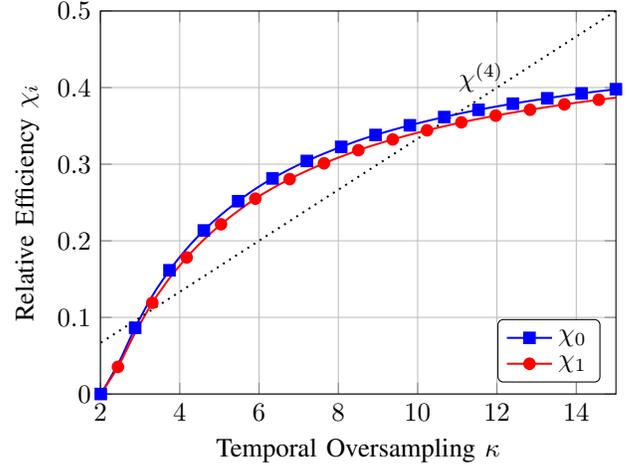
(a) Low SNR ($\bar{\theta} = -9\,\text{dB}$, $\Delta_\theta = 1.5\,\text{dB}$)
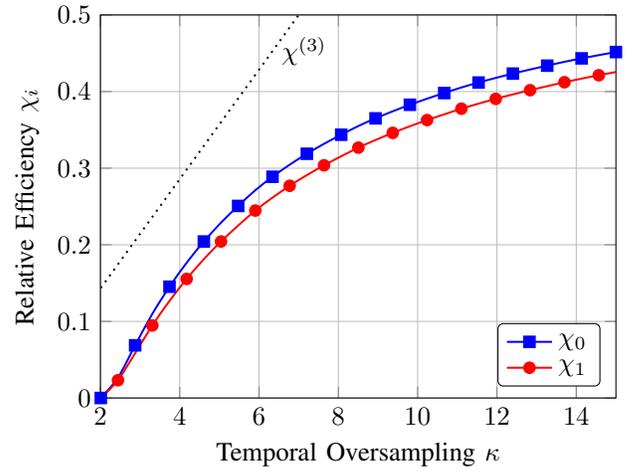
(b) Medium SNR ($\bar{\theta} = 0\,\text{dB}$, $\Delta_\theta = 1.5\,\text{dB}$)

Fig. 6: Binary Cognitive Radio (Superheterodyne Front-end, $M = 1$, $K_0 = 5$)

A binary receiver outperforms a b-bit system in terms of A/D comparator operations when

$$
\frac{\text{ASC}^{(1)}(M, K; \boldsymbol{\theta})}{\text{ASC}^{(b)}(M, K; \boldsymbol{\theta})} < 1.
\tag{91}
$$

For a cognitive radio system with the front-end depicted in Fig. 4, inequality (91) is fulfilled if

$$
\frac{\text{SC}^{(1)}(M, \kappa K_0)}{\text{SC}^{(b)}(M, 2K_0)} \frac{\text{ASN}^{(1)}(\boldsymbol{\theta})}{\text{ASN}^{(\infty)}(\boldsymbol{\theta})} < 1,
\tag{92}
$$

such that a 1-bit receiver, which exceeds the efficiency level

$$
\chi^{(b)} = \frac{\kappa}{2(2^b - 1)}
\tag{93}
$$

is superior to a b-bit system. Fig. 6a demonstrates that a binary device, in terms of (90), outperforms receivers with 4 or more bits A/D resolution. Under the benchmark (92), Fig. 6b indicates that a binary system might not be able to compete with a 3-bit receiver. Note that, concerning the 1-bit system, (92) is conservative as we approximate $\text{ASN}^{(b)} \approx \text{ASN}^{(\infty)}$ while $\text{ASN}^{(b)} > \text{ASN}^{(\infty)}$. Also, note that (92) does not account for

the reduced complexity of the analog preprocessing with A/B conversion, e.g., no automatic gain control (AGC) required.

B. High-Performance GNSS Spectrum Monitoring Systems

As a second application, we consider GNSS spectrum monitoring. The task is to detect interference in the vicinity of a satellite radio receiver synchronizing critical infrastructure (e.g., financial market with high-speed trading, supply point of an electrical network) or providing correction data to mobile GNSS receivers which have a strict reliability requirement on their real-time positioning solution (e.g., ground-based augmentation system at an airport). In case the interference on the GNSS band exceeds a certain power level, the monitor issues a warning to the GNSS receiver, which initiates measures to suppress the interference or reports a temporary malfunction.
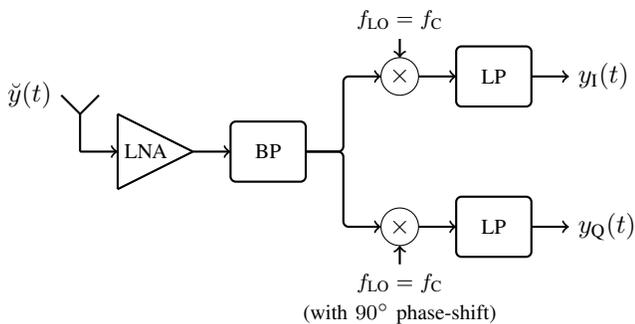

Fig. 7: Homodyne Analog Sensor Front-End

In contrast to the cognitive receiver in Sec. VI-A, for GNSS monitoring, we assume that the receiver features multiple antennas with a homodyne front-end, as depicted in Fig. 7. Homodyne front-ends demodulate the received signal within two real-valued channels (in-phase and quadrature) at carrier frequency f_C, where the quadrature oscillator features a phase-shift of $90°$ relative to the in-phase demodulator. In the signal processing and communication engineering literature, the two real-valued outputs in Fig. 7 are usually summarized in one complex-valued variable. Note, that this is a mathematical convention that serves compactness. The information embedded in the two signals, and thus the achievable performance does not change when switching between real-valued and complex-valued notation. In fact, complex-valued Gaussian models are usually limited to both variables being uncorrelated and of equivalent variance. Removing these limitations within a complex-valued framework, such that the full information carried by both components can be exploited, requires additional effort concerning notation, see, e.g., [54]. A real-valued digital data characterization, which is adjusted to the physical signal acquisition does not face such restrictions and allows analyzing unconventional front-ends, e.g., devices with more than two analog demodulation outputs [17]. Also, note that with a superheterodyne front-end layout like in Fig. 4 and ∞-bit A/D conversion, one can obtain the digital equivalents of the two analog homodyne outputs in Fig. 7 through digital processing. With low-resolution sampling, however, this is not possible. Consequently, when analyzing homodyne front-ends, we stay

in a real-valued notation to keep our framework independent from the specific front-end architecture. So, we denote the analog homodyne outputs (1) as $\boldsymbol{y}(t) = \begin{bmatrix} \boldsymbol{y}_\mathrm{I}^\mathrm{T}(t) & \boldsymbol{y}_\mathrm{Q}^\mathrm{T}(t) \end{bmatrix}^\mathrm{T}$, where $\boldsymbol{y}_\mathrm{I}(t) \in \mathbb{R}^{M_\mathrm{A}}$ and $\boldsymbol{y}_\mathrm{Q}(t) \in \mathbb{R}^{M_\mathrm{A}}$ summarize the analog in-phase and quadrature outputs of the $M_\mathrm{A} = \frac{M}{2}$ receivers. The analog signals have the structure

$$\boldsymbol{y}(t) = \theta \boldsymbol{A} \boldsymbol{x}(t) + \boldsymbol{\eta}(t), \tag{94}$$

where the independent zero-mean random Gaussian processes $\boldsymbol{x}(t) = \begin{bmatrix} x_\mathrm{I}(t) & x_\mathrm{Q}(t) \end{bmatrix}^\mathrm{T}$ model an interferer with bandwidth $B_s = B_y$ received with an SNR θ. Under the assumption that the bandwidth B_y is narrow in comparison to the carrier frequency f_C and the M_A sensors are placed as an uniform linear array (ULA) at a distance of half the carrier wavelength, the array steering matrix can be characterized by $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_\mathrm{I}^\mathrm{T} & \boldsymbol{A}_\mathrm{Q}^\mathrm{T} \end{bmatrix}^\mathrm{T}$, where $\boldsymbol{A}_\mathrm{I}, \boldsymbol{A}_\mathrm{Q} \in \mathbb{R}^{M_\mathrm{A} \times 2}$ are [50]

$$\boldsymbol{A}_\mathrm{I} = \begin{bmatrix} \cos(0) & \sin(0) \\ \cos(\pi \sin(\varphi)) & \sin(\pi \sin(\varphi)) \\ \vdots & \vdots \\ \cos((M_\mathrm{A} - 1)\pi \sin(\varphi)) & \sin((M_\mathrm{A} - 1)\pi \sin(\varphi)) \end{bmatrix} \tag{95}$$

and

$$\boldsymbol{A}_\mathrm{Q} = \begin{bmatrix} -\sin(0) & \cos(0) \\ -\sin(\pi \sin(\varphi)) & \cos(\pi \sin(\varphi)) \\ \vdots & \vdots \\ -\sin((M_\mathrm{A} - 1)\pi \sin(\varphi)) & \cos((M_\mathrm{A} - 1)\pi \sin(\varphi)) \end{bmatrix}, \tag{96}$$

while φ denotes the angle under which the source $\boldsymbol{x}(t)$ impinges onto the ULA. Assuming that the sampling rate is given by $f_\mathrm{T} = \kappa 2 B_y$, for a wireless array receiver with an ideal ∞-bit A/D conversion process, one obtains zero-mean multivariate Gaussian data (4) with covariance

$$\boldsymbol{R}_{\boldsymbol{y}}(\theta) = (\theta \boldsymbol{A}\boldsymbol{A}^\mathrm{T} + \boldsymbol{I}) \otimes \boldsymbol{\Sigma}(\kappa), \quad \boldsymbol{R}_{\boldsymbol{y}}(\theta) \in \mathbb{R}^{MK \times MK}. \tag{97}$$

In this case, the noise is not necessarily temporally white as the sampling rate can be misaligned (regarding the sampling theorem) with the analog prefilter bandwidth, i.e., $f_\mathrm{T} > 2 B_y$.

Fig. 8 shows the relative performance measure (89), where the GNSS radio spectrum is monitored by a binary array with $\bar{\theta} = -15\,\mathrm{dB}$ SNR and $\Delta_\theta = 3\,\mathrm{dB}$. No temporal oversampling is performed, i.e., $\kappa = 1$. It can be observed that the sensing efficiency of the binary array, in comparison to an ideal ∞-bit system, increases with the number of antennas.

With a homodyne array, the inequality (91) holds when

$$\frac{\mathrm{SC}^{(1)}(M, \kappa K_0)}{\mathrm{SC}^{(b)}(M, K_0)} \frac{\mathrm{ASN}^{(1)}(\boldsymbol{\theta})}{\mathrm{ASN}^{(\infty)}(\boldsymbol{\theta})} < 1. \tag{98}$$

Therefore, to outperform a b-bit homodyne array concerning the A/D cost measure (90), a binary system with the same number of radio antennas needs to exceed the efficiency level
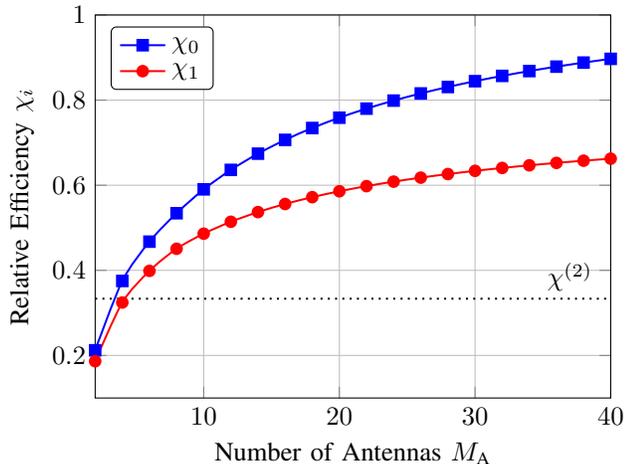
$$\chi^{(b)} = \frac{\kappa}{2^b - 1}. \tag{99}$$

Fig. 8: Binary GNSS Monitoring (Homodyne Front-end, $\varphi = 15°, \kappa = 1, K_0 = 1, \bar{\theta} = -15\,\text{dB}, \Delta_\theta = 3\,\text{dB}$)

Fig. 9: Binary GNSS Monitoring (Homodyne Front-end, $\varphi = 5°, \kappa = 1, K_0 = 1, \bar{\theta} = -15\,\text{dB}, \Delta_\theta = 3\,\text{dB}$)

The results in Fig. 8 show that binary arrays with more than four homodyne antennas outperform 2-bit systems regarding digitization complexity when performing spectrum monitoring. To assess how many additional binary sensors are required to outperform an ideal ∞-bit array with m radio sensors in terms of latency, we define the relative efficiency

$$\chi_{i,m} = \frac{\tilde{\mu}_i(\xi^*)}{\mu_i|_{M_A = m}}. \quad (100)$$

This measure is depicted in Fig. 9, where we use two ∞-bit ULAs ($m = 4$ and $m = 16$) as a performance benchmark. A binary array system with $M_A = 8$ antennas provides the same sensing latency as a ∞-bit A/D resolution system with $m = 4$ radio sensors. For outperforming an ideal ∞-bit radio system with $m = 16$, a binary array with $M_A = 40$ sensors is required. Using that (91) holds in the examined scenario if

$$\frac{\text{SC}^{(1)}(2M_A, \kappa K_0)}{\text{SC}^{(b)}(2m, K_0)} \frac{\text{ASN}^{(1)}(\boldsymbol{\theta})}{\text{ASN}^{(\infty)}(\boldsymbol{\theta})} < 1, \quad (101)$$

Fig. 9 shows that binary arrays can significantly outperform

$$\chi^{(b,m)} = \frac{M_A}{m} \frac{\kappa}{(2^b - 1)} \quad (102)$$

evaluated at $b = 2$ for the two benchmark systems.

Note that using (98) and (101) involves underestimating the latency of a 2-bit system by the ASN of ∞-bit receivers such that (99) and (102) form conservative thresholds.

VII. APPLICATION - LOW-LATENCY BINARY SENSING

Finally, we investigate the accuracy of our latency analysis via (56) and (57) by Monte-Carlo simulations of the binary radio systems considered in Sec. VI. To this end, we optimize the ALLR (43) with (53), run the ASPRT according to (46), and compare the empirical ASN with the analytic results obtained with (56) and (57). For the simulations, we run each sequential algorithm $10\,000$ times with independent observations. The target error rate, determining the thresholds (15) and (16), is set to $\alpha_0 = \alpha_1 = 0.001$ for all experiments.
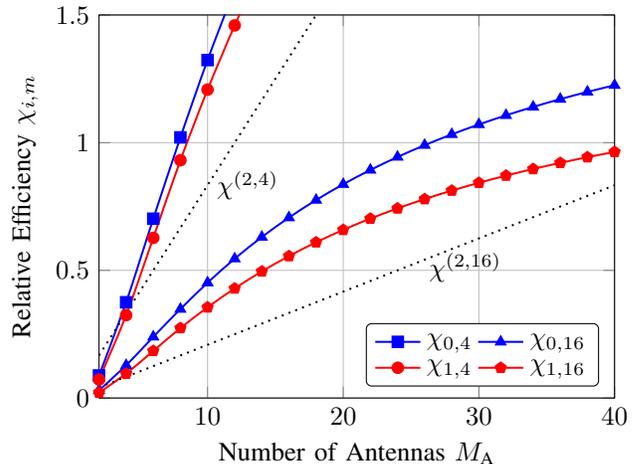
A. Low-Latency Decision-Making for Binary Cognitive Radio

For the cognitive radio setup with superheterodyne front-end (see Sec. VI-A), the simulation scenario is $M = 1, K_0 = 5, \kappa = 5.92$ with the hypotheses at a distance of $\Delta_\theta = 1.5\,\text{dB}$ from $\bar{\theta} = -9\,\text{dB}$. Table I shows a comparison between the analytical and the empirical binary receiver operating characteristics of the ASPRT. For comparison, the ∞-bit digitization case, where we employ the exact LLR (27) and the Gaussian formulas (75) and (76), is given in Table II.

	Receiver Operating Characteristics			
$\mathcal{H}_i$	α_i (ana.)	α_i (sim.)	ASN_i (ana.)	ASN_i (sim.)
$\mathcal{H}_0$	0.0010	0.0010	1361.67	1367.38
$\mathcal{H}_1$	0.0010	0.0007	1351.21	1373.34

TABLE I: Binary Cognitive Radio ($\kappa = 5.92$)

	Receiver Operating Characteristics			
$\mathcal{H}_i$	α_i (ana.)	α_i (sim.)	ASN_i (ana.)	ASN_i (sim.)
$\mathcal{H}_0$	0.0010	0.0011	365.15	368.71
$\mathcal{H}_1$	0.0010	0.0006	344.83	351.14

TABLE II: Ideal Cognitive Radio ($\kappa = 5.92$)

The empirical behavior of both tests shows a good correspondence with the analytic results. For further illustration, the empirical distribution n_D under both hypotheses and digitization approaches is depicted in Fig. 10a and Fig. 10b. The results corroborate that a binary cognitive receiver can reliably sense the activity of a weak primary user through temporal oversampling. As predicted by the analysis in Fig. 6a, the binary system requires significantly more observation time than the ideal receiver to obtain the specified reliability level.

B. Low-Latency Detection for Binary GNSS Monitoring

For the second set of simulations, we consider a binary ULA with $M_A = 8$ homodyne front-ends as analyzed in Sec. VI-B.
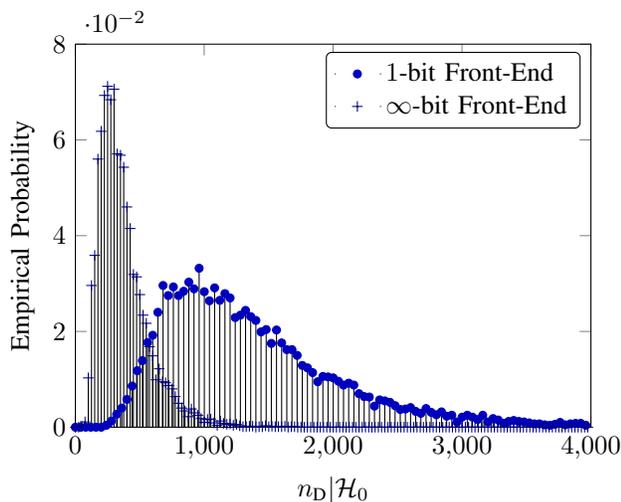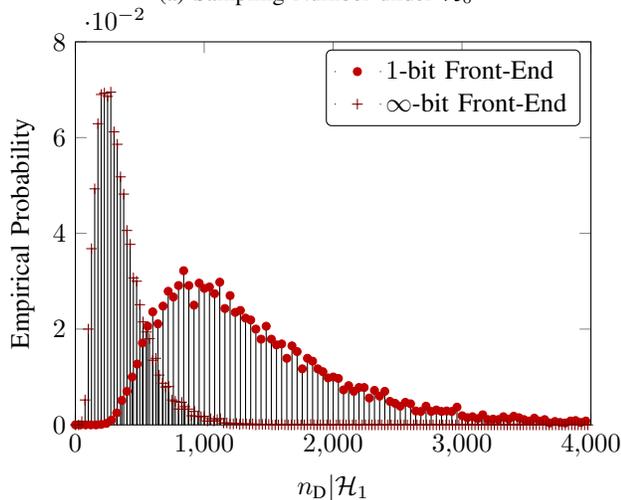
(a) Sampling Number under $\mathcal{H}_0$

(b) Sampling Number under $\mathcal{H}_1$

Fig. 10: Cognitive Radio (Superheterodyne, $M = 1, K_0 = 5, \kappa = 5.92, \bar{\theta} = -9\,\text{dB}, \Delta_\theta = 1.5\,\text{dB}$)

	Receiver Operating Characteristics			
$\mathcal{H}_i$	α_i (ana.)	α_i (sim.)	ASN_i (ana.)	ASN_i (sim.)
$\mathcal{H}_0$	0.0010	0.0002	179.31	183.48
$\mathcal{H}_1$	0.0010	0.0010	162.20	168.09

TABLE III: Binary GNSS Monitoring ($M_\text{A} = 8$)

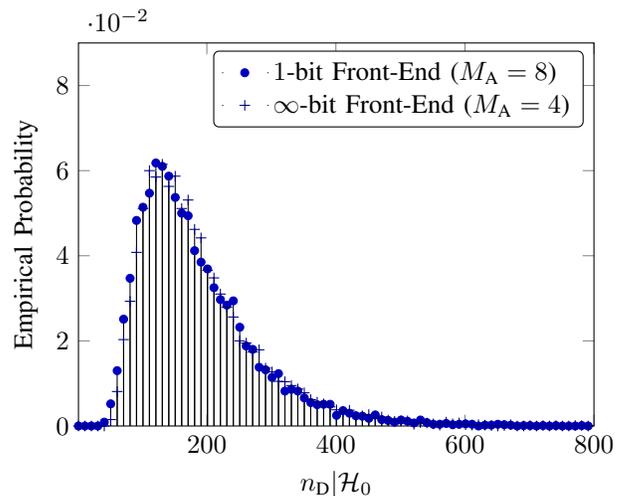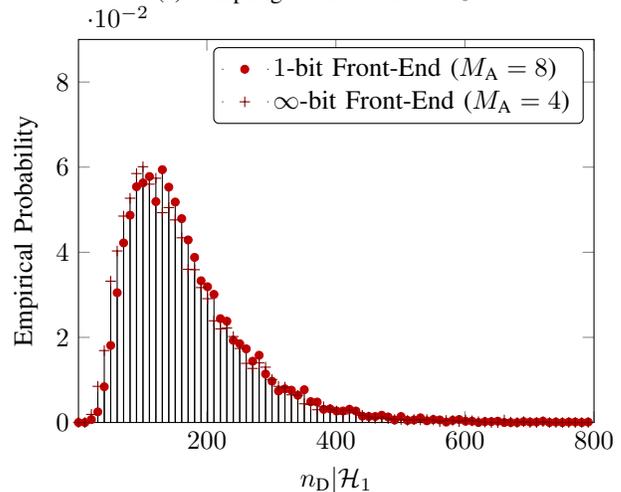|	Receiver Operating Characteristics			
$\mathcal{H}_i$	α_i (ana.)	α_i (sim.)	ASN_i (ana.)	ASN_i (sim.)
$\mathcal{H}_0$	0.0010	0.0007	182.92	184.48
$\mathcal{H}_1$	0.0010	0.0014	150.99	159.43

TABLE IV: Ideal GNSS Monitoring ($M_\text{A} = 4$)

(a) Sampling Number under $\mathcal{H}_0$

(b) Sampling Number under $\mathcal{H}_1$

Fig. 11: GNSS Monitoring (Homodyne, $\varphi = 5°, \kappa = 1, K_0 = 1, \bar{\theta} = -15\,\text{dB}, \Delta_\theta = 3\,\text{dB}$)

The direction-of-arrival is $\varphi = 5°$ and the sampling configuration $\kappa = 1, K_0 = 1$. The interference detection is performed by centering the two hypotheses around $\bar{\theta} = -15\,\text{dB}$ at a distance of $\Delta_\theta = 3\,\text{dB}$. For comparison, an ∞-bit array system with half the amount of wireless sensors ($M_\text{A} = 4$) is used. Table III and Table IV show the obtained results where the empirical outcomes match the analytical assessments.

Fig. 11a and Fig. 11b show the empirical distribution of n_D. These results confirm that double the amount of binary sensors can be sufficient to obtain a sensing apparatus equivalent to an ideal system with ∞-bit A/D resolution. Note that (102) shows that the considered binary array ($M_\text{A} = 8$) provides this excellent performance at less than $66\,\%$ of the digitization cost than any other system with $M_\text{A} = 4$ and a higher A/D resolution. Nevertheless, it requires more space, which is not reflected in our definition of digitization complexity (8).

VIII. CONCLUSION

We have discussed sequential detection with measurements from binary radio systems. Considering statistical tests in the exponential family, we circumvent the intractable distribution

model arising when hard-limiting multivariate Gaussian data. Under this general probabilistic framework, we have derived approximations for the log-likelihood ratio and the Kullback–Leibler divergence. The expressions have the advantage that they cover a broad range of different sensor data models and can be adapted to the testing scenario. This provides stronger approximations than fixed methods. For demonstrating the practical impact, we have applied these theoretical tools to the system design specification of wireless systems with binary sensor signal digitization. The results show that radio systems with low-complexity front-ends are capable of performing challenging sequential detection tasks. In particular, using a large number of binary sensors shows to be a favorable approach concerning detection latency and digitization resources. Finally, we have demonstrated that our analysis matches the empirical behavior of sequential detectors operating based on binary radio data streams. In summary, our discussion provides a versatile framework for sequential tests with hardware-aware probabilistic modeling of the sensor data streams and the rule of thumb that for reaching a favorable complexity-latency trade-off in modern sensing architectures, operating in noisy environments, one should reduce the A/D amplitude resolution to a single bit and double the number of sensor devices.

APPENDIX

For $f : \mathbb{R} \to \mathbb{R}$, infinitely differentiable at $\hat{u} \in \mathbb{R}$, with $\Delta \in \mathbb{R}, \Delta \geq 0$, by the Taylor series

$$f(\hat{u} + \Delta) = \sum_{i=0}^{\infty} \frac{\Delta^i}{i!} \frac{\partial^i f(\hat{u})}{\partial u^i}, \tag{103}$$

$$f(\hat{u} - \Delta) = \sum_{i=0}^{\infty} (-1)^i \frac{\Delta^i}{i!} \frac{\partial^i f(\hat{u})}{\partial u^i}. \tag{104}$$

A forward finite difference approximation at $\hat{u}$ results in

$$f(\hat{u} + \Delta) - f(\hat{u}) \approx \frac{\partial f(\hat{u})}{\partial u} \Delta. \tag{105}$$

With a backward finite difference approximation at $\hat{u}$

$$f(\hat{u}) - f(\hat{u} - \Delta) \approx \frac{\partial f(\hat{u})}{\partial u} \Delta. \tag{106}$$

Defining $\Delta_a, \Delta_b \in \mathbb{R}$ where $\Delta_a, \Delta_b \geq 0$, an alternative approximation at $\hat{u}$ is

$$f(\hat{u} + \Delta_a) - f(\hat{u} - \Delta_b) \approx \frac{\partial f(\hat{u})}{\partial u}(\Delta_a + \Delta_b). \tag{107}$$

Defining $u_1 \geq u_0$, with the forward approximation (105)

$$f(u_1) - f(u_0) \approx \frac{\partial f(u_0)}{\partial u}(u_1 - u_0), \tag{108}$$

and with the backward approximation (106)

$$f(u_1) - f(u_0) \approx \frac{\partial f(u_1)}{\partial u}(u_1 - u_0). \tag{109}$$

Defining $\xi \in [0; 1]$ and

$$\hat{u} = \xi u_0 + (1 - \xi) u_1, \tag{110}$$
$$\Delta_a = \xi(u_1 - u_0), \tag{111}$$
$$\Delta_b = (1 - \xi)(u_1 - u_0), \tag{112}$$

with (107)

$$f(u_1) - f(u_0) \approx \frac{\partial f(\xi u_0 + (1 - \xi) u_1)}{\partial u}(u_1 - u_0). \tag{113}$$

Extending (108) - (113) to $\boldsymbol{f} : \mathbb{R}^I \to \mathbb{R}^J$ with $I, J \in \mathbb{N}$

$$\boldsymbol{f}(\boldsymbol{u}_1) - \boldsymbol{f}(\boldsymbol{u}_0) \approx \frac{\partial \boldsymbol{f}(\boldsymbol{u}_0)}{\partial \boldsymbol{u}}(\boldsymbol{u}_1 - \boldsymbol{u}_0), \tag{114}$$

$$\boldsymbol{f}(\boldsymbol{u}_1) - \boldsymbol{f}(\boldsymbol{u}_0) \approx \frac{\partial \boldsymbol{f}(\boldsymbol{u}_1)}{\partial \boldsymbol{u}}(\boldsymbol{u}_1 - \boldsymbol{u}_0), \tag{115}$$

$$\boldsymbol{f}(\boldsymbol{u}_1) - \boldsymbol{f}(\boldsymbol{u}_0) \approx \frac{\partial \boldsymbol{f}(\xi \boldsymbol{u}_0 + (1 - \xi) \boldsymbol{u}_1)}{\partial \boldsymbol{u}}(\boldsymbol{u}_1 - \boldsymbol{u}_0). \tag{116}$$

REFERENCES

[1] B. Murmann. ADC performance survey 1997-2018. [Online]. Available: http://web.stanford.edu/%7emurmann/adcsurvey.html
[2] M. S. Stein, "Performance analysis for time-of-arrival estimation with oversampled low-complexity 1-bit A/D conversion," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 4491–4495.
[3] M. T. Ivrlac and J. A. Nossek, "Challenges in coding for quantized MIMO systems," in *IEEE Int. Symposium on Information Theory*, July 2006, pp. 2114–2118.
[4] O. Dabeer, J. Singh, and U. Madhow, "On the limits of communication performance with one-bit analog-to-digital conversion," in *IEEE Workshop on Signal Processing Advances in Wireless Communications*, July 2006, pp. 1–5.
[5] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *IEEE Int. Symposium on Information Theory*, June 2007, pp. 1286–1289.
[6] ——, "Capacity lower bound of MIMO channels with output quantization and correlated noise," July 2012, presented at IEEE Int. Symposium on Information Theory. [Online]. Available: http://mediatum.ub.tum.de/1171263
[7] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
[8] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *IEEE Int. Conference on Communication Workshop (ICCW)*, June 2015, pp. 1304–1309.
[9] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.
[10] Y. Jeon, N. Lee, S. Hong, and R. W. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit adcs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, July 2018.
[11] I. Daubechies and R. DeVore, "Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order," *Ann. Math.*, vol. 158, no. 2, pp. 679–710, 2003.
[12] U. S. Kamilov, A. Bourquard, A. Amini, and M. Unser, "One-bit measurements with adaptive thresholds," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 607–610, Oct. 2012.
[13] P. T. Boufounos, L. Jacques, F. Krahmer, and R. Saab, *Quantization and Compressive Sensing*. Cham: Springer International Publishing, 2015, pp. 193–237.
[14] A. Host-Madsen and P. Handel, "Effects of sampling and quantization on single-tone frequency estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 650–662, Mar. 2000.
[15] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks - Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.
[16] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *Int. ITG Workshop on Smart Antennas (WSA)*, Feb. 2010, pp. 143–150.
[17] M. Stein, S. Theiler, and J. A. Nossek, "Overdemodulation for high-performance receivers with low-resolution ADC," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 169–172, Apr. 2015.
[18] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[19] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, June 2018.

[20] P. Willett and P. F. Swaszek, "On the performance degradation from one-bit quantized detection," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1997–2003, Nov. 1995.

[21] D. Ciuonzo, G. Papa, G. Romano, P. Salvo Rossi, and P. Willett, "One-bit decentralized detection with a Rao test for multisensor fusion," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 861–864, Sep. 2013.

[22] M. S. Stein, "Asymptotic signal detection rates with 1-bit array measurements," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 4534–4538.

[23] S. Tantaratana and J. Thomas, "Quantization for sequential signal detection," *IEEE Trans. Commun.*, vol. 25, no. 7, pp. 696–703, July 1977.

[24] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.

[25] M. Fauß and A. M. Zoubir, "A linear programming approach to sequential hypothesis testing," *Sequential Analysis*, vol. 34, no. 2, pp. 235–263, 2015.

[26] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[27] A. Broumandan, A. Jafarnia-Jahromi, S. Daneshmand, and G. Lachapelle, "Overview of spatial processing approaches for GNSS structural interference detection and mitigation," *Proc. IEEE*, vol. 104, no. 6, pp. 1246–1257, June 2016.

[28] R. T. Ioannides, T. Pany, and G. Gibbons, "Known vulnerabilities of global navigation satellite systems, status, and potential mitigation techniques," *Proc. IEEE*, vol. 104, no. 6, pp. 1174–1194, June 2016.

[29] H. R. Hashemi and I. B. Rhodes, "Decentralized sequential detection," *IEEE Trans. Inf. Theory*, vol. 35, no. 3, pp. 509–520, May 1989.

[30] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433–442, Mar. 1993.

[31] A. M. Hussain, "Multisensor distributed sequential detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 698–708, July 1994.

[32] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2072–2089, May 2008.

[33] Y. Yilmaz, G. V. Moustakides, and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4509–4524, Sep. 2012.

[34] S. Chaudhari, J. Lundn, and V. Koivunen, "Effects of quantization and channel errors on sequential detection in cognitive radios," in *Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2012, pp. 1–6.

[35] Y. Wang and Y. Mei, "Quantization effect on the log-likelihood ratio and its application to decentralized sequential detection," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1536–1543, Mar. 2013.

[36] R. S. Blum, "Quantization in multisensor random signal detection," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 204–215, Jan. 1995.

[37] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On optimal quantization rules for some problems in sequential decentralized detection," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3285–3295, July 2008.

[38] D. Teng and E. Ertin, "Optimal quantization of likelihood for low complexity sequential testing," in *IEEE Global Conference on Signal and Information Processing*, Dec. 2013, pp. 675–678.

[39] S. Tantaratana and J. Thomas, "On sequential sign detection of a constant signal," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 304–315, May 1977.

[40] C. C. Lee and J. B. Thomas, "Sequential detection based on simple quantization," *J. Franklin Inst.*, vol. 312, pp. 119–135, 1981.

[41] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[42] S. Hong and N. Lee, "Soft-output detector for uplink MU-MIMO systems with one-bit ADCs," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 930–933, May 2018.

[43] M. S. Stein, S. Bar, J. A. Nossek, and J. Tabrikian, "Performance analysis for channel estimation with 1-bit ADC and unknown quantization threshold," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2557–2571, May 2018.

[44] B. Dai, S. Ding, and G. Wahba, "Multivariate Bernoulli distribution," *Bernoulli*, vol. 19, no. 4, pp. 1465–1483, 2013.

[45] S. S. Gupta, "Probability integrals of multivariate normal and multivariate t^1," *Ann. Math. Statist.*, vol. 34, no. 3, pp. 792–828, 09 1963.

[46] M. S. Stein, J. A. Nossek, and K. Barbé, "Fisher information lower bounds with applications in hardware-aware nonlinear signal processing," 2015. [Online]. Available: http://arxiv.org/abs/1512.03473v2

[47] M. S. Stein and M. Fauß, "In a one-bit rush: Low-latency wireless spectrum monitoring with binary sensor arrays," in *IEEE Statistical Signal Processing Workshop (SSP)*, June 2018, pp. 223–227.

[48] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman & Hall, 2014.

[49] D. R. Cox and N. Wermuth, "A note on the quadratic exponential binary distribution," *Biometrika*, vol. 81, no. 2, pp. 403–408, 1994.

[50] M. Stein, K. Barbé, and J. A. Nossek, "DOA parameter estimation with 1-bit quantization - Bounds, methods and the exponential replacement," in *Int. ITG Workshop on Smart Antennas (WSA)*, Mar. 2016, pp. 1–6.

[51] J. B. Thomas, *An introduction to statistical communication theory*. Wiley, 1969.

[52] M. Sinn and K. Keller, "Covariances of zero crossings in Gaussian processes," *Theory of Probability & Its Applications*, vol. 55, no. 3, pp. 485–504, 2011.

[53] A. Khintchine, "Korrelationstheorie der stationären stochastischen Prozesse," *Mathematische Annalen*, vol. 109, no. 1, pp. 604–615, Dec. 1934.

[54] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.