20

Event-Based Statistical Signal Processing

Yasin Yılmaz
University of Michigan
Ann Arbor, MI, USA

George V. Moustakides
University of Patras
Rio, Greece

Xiaodong Wang
Columbia University
New York, NY, USA

Alfred O. Hero
University of Michigan
Ann Arbor, MI, USA

CONTENTS

ABSTRACT In traditional time-based sampling, the sampling mechanism is triggered by predetermined sampling times, which are mostly uniformly spaced (i.e., periodic). Alternatively, in event-based sampling, some predefined events on the signal to be sampled trigger the sampling mechanism; that is, sampling times are determined by the signal and the event space. Such an alternative mechanism, setting the sampling times free, can enable simple (e.g., binary) representations in the event space. In real-time applications, the induced sampling times can be easily traced and reported with high accuracy, whereas the amplitude of a time-triggered sample needs high data rates for high accuracy.

In this chapter, for some statistical signal processing problems, namely detection (i.e., binary hypothesis testing) and parameter estimation, in resource-constrained distributed systems (e.g., wireless sensor networks), we show how to make use of the time dimension for data/information fusion, which is not possible through the traditional fixed-time sampling.

20.1 Introduction

Event-based paradigm is an alternative to conventional time-driven systems in control [2,13,28] and signal processing [37,43,61]. Event-based methods are *adaptive* to the observed entities, as opposed to the time-driven techniques. In signal processing, they are used for data compression [37], analog-to-digital (A/D) conversion [23,30,61], data transmission [42,43,55], imaging applications [11,26,29], detection [17,24,73], and estimation [16,75]. We also see a natural example in biological sensing systems. In many multicellular organisms, including plants, insects, reptiles, and mammals, the all-or-none principle, according to which neurons fire, that is, transmit electrical signals, is an event-based technique [19].

In signal processing applications, event-based paradigm is mainly used as a means of *nonuniform sampling*. In conventional uniform sampling, the sampling frequency is, in general, selected based on the highest expected spectral frequency. When the lower frequency content in the input signal is dominant (e.g., long periods of small change), such high-frequency sampling wastes considerable power. For many emerging applications that rely on scarce energy resources (e.g., wireless sensor networks), a promising alternative is event-based sampling, in which a sample is taken when a significant event occurs in the signal. Several closely related signal-dependent sampling techniques have been proposed, for example, level-crossing sampling [24], Lebesgue sampling [17], send-on-delta [42], time-encoding machine [29], and level-triggered sampling [73]. In these event-based sampling methods, samples are taken based on the signal amplitude instead of time, as opposed to the conventional uniform sampling. Analogous to the comparison between the Riemann and Lebesgue integrals, the amplitude-driven and conventional time-driven sampling techniques are also called Lebesgue sampling and Riemann sampling, respectively [2]. As a result, the signal is encoded in the sampling times, whereas in uniform sampling the sample amplitudes encode the signal. This yields a significant advantage in real-time applications, in which sampling times can be tracked via simple one-bit signaling. Specifically, event-based sampling, through one-bit representations of the samples, enables high-resolution recovery, which requires many bits per sample in uniform sampling. In other words, event-based sampling can save energy and bandwidth (if samples are transmitted to a receiver) in real-time applications in terms of encoding samples.

20.1.1 Event-Based Sampling

In level-crossing sampling, which is mostly used for A/D conversion [23,50,61], in general, uniform sampling levels in the amplitude domain are used, as shown in Figure 20.1. A/D converters based on level-crossing sampling are free of a sampling clock, which is a primary energy consumer in traditional A/D converters [61]. A version of level-crossing sampling that ignores successive crossings of the same level is used to reduce the sampling rate, especially for noisy signals [28]. This technique is called level-crossing sampling with hysteresis (LCSH) due to the hysteretic quantizer it leads to (see Figure 20.1).

Time-encoding machine is a broad event-based sampling concept, in which the signal is compared with a reference signal and sampled at the crossings [21,29]. The reference signal is possibly updated at the sampling instants (Figure 20.2). Motivated by the integrate-and-fire neuron model, a mathematical model for nerve cells, in some time-encoding machines, the signal is first integrated and then sampled. The asynchronous delta–sigma modulator, a nonlinear modulation scheme mainly used for A/D conversion, is an instance of integrate-and-fire time-encoding machine [30]. The ON–OFF time-encoding machine, which models the

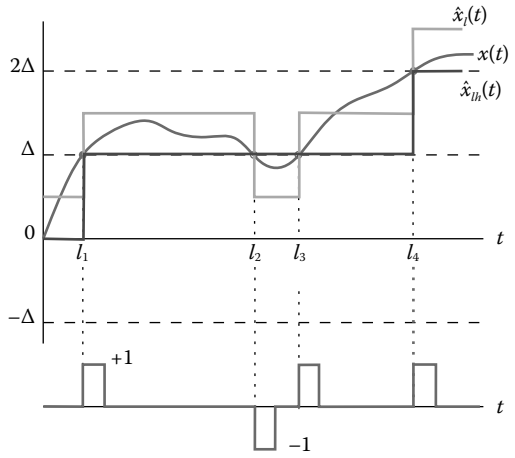

FIGURE 20.1

Level-crossing sampling with uniform sampling levels results in nonuniform sampling times l_{1-4} and the quantized signal $\hat{x}_l(t)$. If the repeated crossings at l_2 and l_3 are discarded, $\hat{x}_{lh}(t)$ is produced by a hysteretic quantizer. One-bit encoding of the samples is shown below.

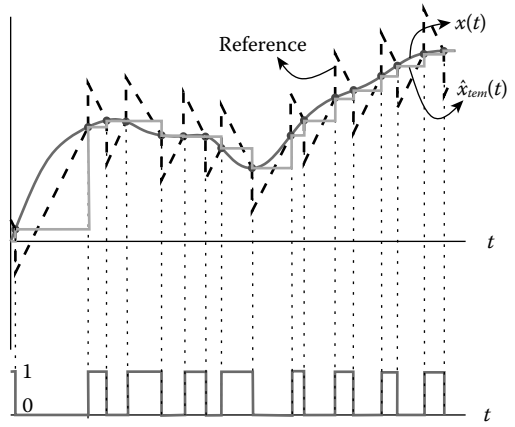

FIGURE 20.2

Reference signal representation of a time-encoding machine and the piecewise constant signal $\hat{x}_{tem}(t)$ resulting from the samples. At the sampling times, the reference signal switches the pair of offset and slope between $(-\delta, b)$ and $(\delta, -b)$, as in [30]. One-bit encoding of the samples is also shown below.

ON and OFF bipolar cells in the retina [39], uses two reference signals to capture the positive and negative changes in the signal. The ON–OFF time-encoding machine without integration coincides with the LCSH [29]. Hardware applications of ON–OFF time-encoding machines are seen in neuromorphic engineering [33,77] and brain–machine interfaces [3,49].

The theory of signal reconstruction from nonuniform samples applies to event-based sampling [61]. Exact reconstruction is possible if the average sampling rate is above the Nyquist rate (i.e., twice the bandwidth of the signal) [4,30]. Various reconstruction methods have been proposed in [4,30,38,70]. The similarity measures for sequences of level-crossing samples have been discussed, and an appropriate measure has been identified in [44].

20.1.2 Decentralized Data Collection

Decentralized data collection in resource-constrained networked systems (e.g., wireless sensor networks) is another fundamental application (in addition to A/D conversion) of sampling. In such systems, the central processor does not have access to all observations in the system due to physical constraints, such as energy and communication (i.e., bandwidth) constraints. Hence, the choice of sampling technique is of great importance to obtain a good summary of observations at the central processor. Using an adaptive sampling scheme, only the informative observations can be transmitted to the central processor. This potentially provides a better summary than the conventional (nonadaptive) sampling scheme that satisfies the same physical constraints. As a toy example to adaptive transmission, consider a bucket carrying water to a pool from a tap with varying flow. After the same number of carriages, say ten, the scheme that empties the bucket only when it is filled (i.e., adaptive to the water flow) exactly carries ten buckets of water to the pool, whereas the scheme that periodically empties the bucket (i.e., nonadaptive), in general, carries less water.

Based on such an adaptive scheme, the send-on-delta concept, for decentralized acquisition of continuous-time band-limited signals, samples and transmits only when the observed signal changes by $\pm\Delta$ since the last sampling time [42,55]. In other words, instead of transmitting at deterministic time instants, it waits for the event of $\pm\Delta$ change in the signal amplitude to sample and transmit. Although the change here is with respect to the last sample value, which is in general different from the last sampling level in level-crossing sampling, they coincide for continuous-time band-limited signals. Hence, for continuous-time band-limited signals, send-on-delta sampling is identical to LCSH (Figure 20.1). For systems in which the accumulated, instead of the current, absolute error (similar to the mean absolute error) is used as the performance criterion, an extension of the send-on-delta concept, called the integral send-on-delta, has been proposed [43]. This extension is similar to the integrate-and-fire time-encoding machine. Specifically, a Δ increase in the integral of absolute error triggers sampling (and transmission).

In essence, event-based processing aims to simplify the signal representation by mapping the real-valued amplitude, which requires infinite number of bits after

the conventional time-driven sampling, to a digital value in the event space, which needs only a few bits. In most event-based techniques, including the ones discussed above, a single bit encodes the event type when an event occurs (e.g., $\pm\Delta$ change, upward/downward level crossing, reference signal crossing). In decentralized data collection, this single-bit quantization in the event space constitutes a great advantage over the infinite-bit representation of a sample taken at a deterministic time. Moreover, to save further energy and bandwidth, the number of samples (i.e., event occurrences) can be significantly reduced by increasing Δ. That is, a large enough Δ value in send-on-delta sparsifies the signal with binary nonzero values, which is ideal for decentralized data collection. On the contrary, the resolution of observation summary at the central processor decreases with increasing Δ, showing the expected trade-off between performance and consumption of physical resources (i.e., energy and bandwidth). In real-time reporting of the sparse signal in the event space to the central processor, only the nonzero values are sampled and transmitted when encountered*. Since event-based processing techniques, in general, first quantize the signal in terms of the events of interest, and then sample the quantized signal, they apply a *quantize-and-sample* strategy, instead of the sample-and-quantize strategy followed by the conventional time-driven processing techniques.

20.1.3 Decentralized Statistical Signal Processing

If, in a decentralized system, data are collected for a specific purpose (e.g., hypothesis testing, parameter estimation), then we should locally process raw observations as much as possible before transmitting to minimize processing losses at the central processor, in addition to the transmission losses due to physical constraints. For instance, in hypothesis testing, each node in the network can first compute and then report the log-likelihood ratio (LLR) of its observations, which is the sufficient statistic. Assuming independence of observations across nodes, the central processor can simply sum the reported LLRs and decide accordingly without further processing. On the contrary, if each node transmits its raw observations in a decentralized fashion, the central processor needs to process the lossy data to approximate LLR, which is in general a nonlinear function. The LLR approximation in the latter report-and-process strategy is clearly worse than the one in the former process-and-report strategy.

*Unlike compressive sensing, the binary nonzero values are simply reported in real time without any need for offline computation.

An event-based sampling technique, called level-triggered sampling, has been proposed to report the corresponding sufficient statistic in binary hypothesis testing [17] and parameter estimation [16] for continuous-time band-limited observations. The operation of level-triggered sampling is identical to that of send-on-delta sampling (i.e., $\pm\Delta$ changes in the local sufficient statistic since the last sampling time triggers a new sample), but it is motivated by the sequential probability ratio test (SPRT), the optimum sequential detector (i.e., binary hypothesis test) for independent and identically distributed (iid) observations. Without a link to event-based sampling, it was first proposed in [27] as a repeated SPRT procedure for discrete-time observations. In particular, when its local LLR exits the interval $(-\Delta, \Delta)$, each node makes a decision: null hypothesis H_0 if it is less than or equal to $-\Delta$, and alternative hypothesis H_1 if it is greater than or equal to Δ. Then another cycle of SPRT starts with new observations. The central processor, called the fusion center, also runs SPRT by computing the joint LLR of such local decisions.

Due to the numerous advantages of digital signal processing (DSP) and digital communications over their analog counterparts, a vast majority of the existing hardware work with discrete-time signals. Although there is a significant interest in building a new DSP theory based on event-based sampling [30,44,60,64], such a theory is not mature yet, and thus it is expected that the conventional A/D converters, based on uniform sampling, will continue to dominate in the near future. Since digital communications provide reliable and efficient information transmission, with the support of inexpensive electronics, it is ubiquitous nowadays [25, page 23]. Hence, even if we perform analog signal processing and then event-based sampling on the observed continuous-time signal, we will most likely later on need to quantize time (i.e., uniformly sample the resulting continuous-time signal) for communication purposes. In that case, we should rather apply event-based sampling to uniformly sampled discrete-time observations at the nodes. This also results in a compound architecture which can perform time-driven, as well as event-driven, tasks [42,46]. As a result, level-triggered sampling with discrete-time observations (see Figure 20.3) has been considered for statistical signal processing applications [32,73–76].

In level-triggered sampling, a serious complication arises with discrete-time observations: when a sample is taken, the change since the last sampling time, in general, exceeds Δ or $-\Delta$ due to the jumps in the discrete-time signal, known as the *overshoot* problem (see Figure 20.3). Note from Figure 20.3 that the sampling thresholds are now signal dependent, as opposed to level-crossing sampling (with hysteresis), shown in

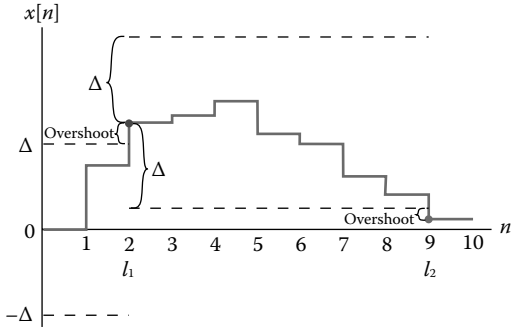

FIGURE 20.3
Level-triggered sampling with discrete-time observations.

Figure 20.1. Overshoots disturb the binary quantized values in terms of Δ change (i.e., the one-bit encoding in Figure 20.1) since now fractional Δ changes are possible. As a result, when sampled, such fractional values in the event space cannot be exactly encoded into a single bit. Overshoots are also observed with continuous-time band-unlimited signals (i.e., that have jumps), which are observed in practice due to noise. Hence, for practical purposes, we need to deal with the overshoot problem.

In level-triggered sampling, the overshoot problem is handled in several ways. In the first method, a single-bit encoding is still used with quantization levels that include average overshoot values in the event space. That is, for positive ($\geq\Delta$)/negative ($\leq-\Delta$) changes, the transmitted $+1/-1$ bit represents a fractional change $\bar{\theta} > \Delta/\underline{\theta} < -\Delta$, where $\bar{\theta} - \Delta/\underline{\theta} + \Delta$ compensates for the average overshoot above Δ/below $-\Delta$. Examples of this overshoot compensation method are seen in the decentralized detectors of [27,74], and also in Section 20.2.2, in which the LLR of each received bit at the fusion center (FC) is computed. There are two other overshoot compensation methods in the literature, both of which quantize each overshoot value. In [73] and [75], for detection and estimation purposes, respectively, each quantized value is transmitted in a few bits via separate pulses, in addition to the single bit representing the sign of the Δ change. On the contrary, in [76], and also in Section 20.3.3.2, pulse-position modulation (PPM) is used to transmit each quantized overshoot value. Specifically, the unit time interval is divided into a number of subintervals, and a short pulse is transmitted for the sign bit at the time slot that corresponds to the overshoot value. Consequently, to transmit each quantized overshoot value, more energy is used in the former method, whereas more bandwidth is required in the latter.

In the literature, level-triggered sampling has been utilized to effectively transmit the sufficient local statistics in decentralized systems for several applications, such as spectrum sensing in cognitive radio networks [73], target detection in wireless sensor networks [76], joint spectrum sensing and channel estimation in cognitive radio networks [71], security in multiagent reputation systems [32], and power quality monitoring in power grids [31].

20.1.4 Outline

In this chapter, we analyze the use of event-based sampling as a means of information transmission for decentralized detection and estimation. We start with the decentralized detection problem in Section 20.2. Two challenges, namely noisy transmission channels and multimodal information sources, have been addressed via level-triggered and level-crossing sampling in Sections 20.2.2 and 20.2.3, respectively.

Then, in Section 20.3, we treat the sequential estimation of linear regression parameters under a decentralized setup. Using a variant of level-triggered sampling, we design a decentralized estimator that achieves a close-to-optimum average stopping time performance and linearly scales with the number of parameters while satisfying stringent energy and computation constraints.

Throughout the chapter, we represent scalars with lower-case letters, vectors with bold lower-case letters, and matrices with bold upper-case letters.

20.2 Decentralized Detection

We first consider the decentralized detection (i.e., hypothesis testing) problem, in which a number of distributed nodes (e.g., sensors), under energy and bandwidth constraints, sequentially report a summary of their discrete-time observations to an FC, which makes a decision as soon as possible satisfying some performance constraints.

20.2.1 Background

Existing works on decentralized detection mostly consider the fixed-sample-size approach, in which the FC makes a decision at a deterministic time using a fixed number of samples from nodes (e.g., [57,59,67]). The sequential detection approach, in which the FC at each time chooses either to continue receiving new samples or to stop and make a decision, is also of significant interest (e.g., [8,40,62]). In [17,27,73,74,76], SPRT is used both at the nodes and the FC. SPRT is the optimum sequential detector for iid observations in terms of minimizing the average sample number among all sequential tests satisfying the same error probability constraints [65]. Compared with the best fixed-sample-size detector, SPRT requires, on average, four times less samples

for the same level of confidence for Gaussian signals [47, page 109].

Under stringent energy and bandwidth constraints where nodes can only infrequently transmit a single bit (which can be considered as a local decision), the optimum local decision function is the likelihood ratio test (LRT), which is nothing but a one-bit quantization of LLR, for a fixed *decision fusion* rule under the fixed-sample-size setup [57]. Similarly, the optimum fusion rule at the FC is also an LRT under the Bayesian [6] and Neyman–Pearson [58] criteria. Since SPRT, which is also a one-bit quantization of LLR with the deadband $(-\Delta, \Delta)$, is the sequential counterpart of LRT, these results readily extend to the sequential setup as a double-SPRT scheme [27].

Under relaxed resource constraints, the optimum local scheme is a multibit quantization of LLR [66], which is the necessary and sufficient statistic for the detection problem, while the optimum *data fusion* detector at the FC is still an LRT under the fixed-sample-size setup. Thanks to the event-based nature of SPRT, even its single-bit decision provides data fusion capabilities. More specifically, when it makes a decision, we know that LLR $\geq\Delta$ if H_1 is selected, or LLR $\leq-\Delta$ if H_0 is selected. For continuous-time bandlimited observations, we have a full precision, that is, LLR $=\Delta$ or LLR $=-\Delta$ depending on the decision, which requires infinite number of bits with LRT under the fixed-sample-size setup. The repeated SPRT structure of level-triggered sampling enables LLR tracking, that is, sequential data fusion [17,74]. For discrete-time observations, the single-bit decision at each SPRT step (i.e., one-bit representation of a level-triggered sample as in Figure 20.3) may provide high-precision LLR tracking if overshoots are small compared with Δ. Otherwise, under relaxed resource constraints, each overshoot can be quantized into additional bits [73,76], resulting in a multibit quantization of the changes in LLR with the deadband $(-\Delta, \Delta)$, analogous to the multibit LLR quantization under the fixed-sample-size setup [66].

The conventional approach to decentralized detection, assuming ideal transmission channels, addresses only the noise that contaminates the observations at nodes (e.g., [17,57]). Nevertheless, in practice, the channels between nodes and the FC are noisy. Following the conventional approach, at the FC, first a communication block recovers the transmitted information bits, and then an independent signal processing block performs detection using the recovered bits. Such an independent two-step procedure inflicts performance loss due to the data-processing inequality [9]. For optimum performance, without a communication block, the received signals should be processed in a channel-aware manner [7,34].

In this section, we first design in Section 20.2.2 channel-aware decentralized detection schemes based on level-triggered sampling for different noisy channel models. We then show in Section 20.2.3 how to fuse multimodal data from disparate sources for decentralized detection.

20.2.2 Channel-Aware Decentralized Detection

Consider a network of K distributed nodes (e.g., a wireless sensor network) and an FC, which can be one of the nodes or a dedicated processor (Figure 20.4). Each node k computes the LLR $L_k[n]$ of discrete-time signal $x_k[n]$ it observes, and sends the level-triggered LLR samples to the FC, which fuses the received samples and sequentially decides between two hypotheses, H_0 and H_1.

Assuming iid observations $\{x_k[n]\}_n$ across time, and independence across nodes, the local LLR at node k and the global LLR are given by

$$\begin{aligned} L_k[n] &= \log \frac{f_{k,1}(x_k[1],\ldots,x_k[n])}{f_{k,0}(x_k[1],\ldots,x_k[n])} = \sum_{m=1}^{n} \log \frac{f_{k,1}(x_k[m])}{f_{k,0}(x_k[m])} \\ &= \sum_{m=1}^{n} l_k[m] = L_k[n-1] + l_k[n], \\ L[n] &= \sum_{k=1}^{K} L_k[n], \end{aligned}$$

respectively, where $f_{k,j}, j=0,1$ is the probability density/mass function of the observed signal at node k under H_j, and $l_k[n]$ is the LLR of $x_k[n]$.

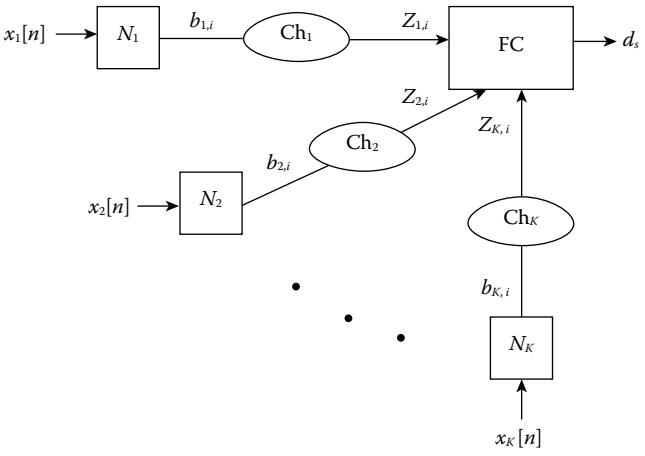

FIGURE 20.4

A network of K nodes and a fusion center (FC). Each node k processes its observations $\{x_k[n]\}_n$ and transmits information bits $\{b_{k,i}\}_i$. The FC then, upon receiving the signals $\{z_{k,i}\}$, makes a detection decision d_S at the random time S.

20.2.2.1 Procedure at Nodes

Each node k samples $L_k[n]$ via level-triggered sampling at a sequence of random times $\{t_{k,i}\}_i$ that are determined by $L_k[n]$ itself. Specifically, the ith sample is taken when the LLR change $L_k[n] - L_k[t_{k,i-1}]$ since the last sampling time $t_{k,i-1}$ exceeds a constant Δ in absolute value, that is,

$$t_{k,i} \triangleq \min\left\{n > t_{k,i-1} : L_k[n] - L_k[t_{k,i-1}] \notin (-\Delta, \Delta)\right\}, \quad t_{k,0} = 0,\ L_k[0] = 0. \tag{20.1}$$

It has been shown in [73, section IV-B] that Δ can be determined by

$$\Delta \tanh\left(\frac{\Delta}{2}\right) = \frac{1}{R}\sum_{k=1}^{K} |\mathsf{E}_j[L_k[1]]|, \tag{20.2}$$

to ensure that the FC receives messages with an average rate of R messages per unit time interval.

Let $\lambda_{k,i}$ denote the LLR change during the ith sampling interval, $(t_{k,i-1}, t_{k,i}]$, that is,

$$\lambda_{k,i} \triangleq L_k[t_{k,i}] - L_k[t_{k,i-1}] = \sum_{n=t_{k,i-1}+1}^{t_{k,i}} l_k[n].$$

Immediately after sampling at $t_{k,i}$, as shown in Figure 20.4, an information bit $b_{k,i}$ indicating the threshold crossed by $\lambda_{k,i}$ is transmitted to the FC, that is,

$$b_{k,i} \triangleq \text{sign}(\lambda_{k,i}). \tag{20.3}$$

20.2.2.2 Procedure at the FC

Let us now analyze the received signal $z_{k,i}$ at the FC corresponding to the transmitted bit $b_{k,i}$ (see Figure 20.4). The FC computes the LLR

$$\tilde{\lambda}_{k,i} \triangleq \log\frac{g_{k,1}(z_{k,i})}{g_{k,0}(z_{k,i})}, \tag{20.4}$$

of each received signal $z_{k,i}$ and approximates the global LLR, $L[n]$, as

$$\tilde{L}[n] \triangleq \sum_{k=1}^{K}\sum_{i=1}^{J_{k,n}} \tilde{\lambda}_{k,i},$$

where $J_{k,n}$ is the total number of LLR messages received from node k until time n, and $g_{k,j}$, $j = 0, 1$, is the pdf of $z_{k,i}$ under H_j.

In fact, the FC recursively updates $\tilde{L}[n]$ whenever it receives an LLR message from any node. In particular, suppose that the mth LLR message $\tilde{\lambda}_m$ from any sensor is received at time t_m. Then at t_m, the FC performs the following update:

$$\tilde{L}[t_m] = \tilde{L}[t_{m-1}] + \tilde{\lambda}_m,$$

and uses $\tilde{L}[t_m]$ in an SPRT procedure with two thresholds A and $-B$, and the following decision rule

$$d_{t_m} \triangleq \begin{cases} \mathsf{H}_1, & \text{if } \tilde{L}[t_m] \geq A, \\ \mathsf{H}_0, & \text{if } \tilde{L}[t_m] \leq -B, \\ \text{wait for } \tilde{\lambda}_{m+1}, & \text{if } \tilde{L}[t_m] \in (-B, A). \end{cases}$$

The thresholds $(A, B > 0)$ are selected to satisfy the error probability constraints

$$\mathsf{P}_0(d_S = \mathsf{H}_1) \leq \alpha \ \text{ and } \ \mathsf{P}_1(d_S = \mathsf{H}_0) \leq \beta, \tag{20.5}$$

with equalities, where P_j, $j = 0, 1$, denotes the probability under H_j, α and β are the error probability bounds given to us, and

$$S \triangleq \min\{n > 0 : \tilde{L}[n] \notin (-B, A)\}, \tag{20.6}$$

is the decision time.

Comparing (20.1) with (20.6), we see that each node, in fact, applies a local SPRT with thresholds Δ and $-\Delta$ within each sampling interval. At node k, the ith local SPRT starts at time $t_{k,i-1} + 1$ and ends at time $t_{k,i}$ when the local test statistic $\lambda_{k,i}$ exceeds either Δ or $-\Delta$. This local hypothesis testing produces a local decision represented by the information bit $b_{k,i}$ in (20.3), and induces the local error probabilities

$$\alpha_k \triangleq \mathsf{P}_0(b_{k,i} = 1) \ \text{ and } \ \beta_k \triangleq \mathsf{P}_1(b_{k,i} = -1). \tag{20.7}$$

We next discuss how to compute $\tilde{\lambda}_{k,i}$, the LLR of received signal $z_{k,i}$, given by (20.4), under ideal and noisy channels.

20.2.2.3 Ideal Channels

Lemma 20.1

Assuming ideal channels between nodes and the FC, that is, $z_{k,i} = b_{k,i}$, we have

$$\tilde{\lambda}_{k,i} = \begin{cases} \log\frac{\mathsf{P}_1(b_{k,i}=1)}{\mathsf{P}_0(b_{k,i}=1)} = \log\frac{1-\beta_k}{\alpha_k} \geq \Delta, & \text{if } b_{k,i} = 1, \\ \log\frac{\mathsf{P}_1(b_{k,i}=-1)}{\mathsf{P}_0(b_{k,i}=-1)} = \log\frac{\beta_k}{1-\alpha_k} \leq -\Delta, & \text{if } b_{k,i} = -1. \end{cases} \tag{20.8}$$

PROOF The equalities follow from (20.7). The inequalities can be obtained by applying a change of measure. To show the first one, we write

$$\alpha_k = \mathsf{P}_0(\lambda_{k,i} \geq \Delta) = \mathsf{E}_0[\mathbb{1}_{\{\lambda_{k,i} \geq \Delta\}}], \tag{20.9}$$

where E_j is the expectation under H_j, $j = 0, 1$ and $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Note that

$$e^{-\lambda_{k,i}} = \frac{f_{k,0}(x_k[t_{k,i-1}+1], \ldots, x_k[t_{k,i}])}{f_{k,1}(x_k[t_{k,i-1}+1], \ldots, x_k[t_{k,i}])},$$

can be used to compute the expectation integral in terms of $f_{k,1}$ instead of $f_{k,0}$, that is, to change the probability measure under which the expectation is taken from $f_{k,0}$ to $f_{k,1}$. Hence,

$$\begin{aligned}\alpha_k &= \mathsf{E}_1[e^{-\lambda_{k,i}}\mathbb{1}_{\{\lambda_{k,i} \geq \Delta\}}] \\ &\leq e^{-\Delta}\mathsf{E}_1[\mathbb{1}_{\{\lambda_{k,i} \geq \Delta\}}] = e^{-\Delta}\mathsf{P}_1(\lambda_{k,i} \geq \Delta) \\ &= e^{-\Delta}(1-\beta_k),\end{aligned}$$

giving us the first inequality in (20.8). The second inequality follows similarly. ■

We see from Lemma 20.1 that the FC, assuming ideal channels, can compute $\tilde{\lambda}_{k,i}$, the LLR of the sign bit $b_{k,i}$ if the local error probabilities α_k and β_k are available. It is also seen that $\tilde{\lambda}_{k,i}$ is, in magnitude, larger than the corresponding sampling threshold, and thus includes a constant compensation for the random overshoot of $\lambda_{k,i}$ above Δ or below $-\Delta$. The relationship of this constant compensation to the average overshoot, and the *order-1* asymptotic optimality it achieves are established in [17].

In the no-overshoot case, as with continuous-time band-limited observations, the inequalities in (20.8) become equalities since in (20.9) we can write $\alpha_k = \mathsf{P}_0(\lambda_{k,i} = \Delta)$. This shows that the LLR update in (20.8) adapts well to the no-overshoot case, in which the LLR change that triggers sampling is either Δ or $-\Delta$.

Theorem 20.1: [17, Theorem 2]

Consider the asymptotic regime in which the target error probabilities $\alpha, \beta \to 0$ at the same rate. If the sampling threshold $\Delta \to \infty$ is slower than $|\log \alpha|$, then, under ideal channels, the decentralized detector which uses the LLR update given by (20.8) for each level-triggered sample is order-1 asymptotically optimum, that is,

$$\frac{\mathsf{E}_j[S]}{\mathsf{E}_j[S_o]} = 1 + o(1),\ j = 0, 1, \tag{20.10}$$

where S_o is the decision time of the optimum (centralized) sequential detector, SPRT, satisfying the error probability bounds α and β [cf. (20.5)].

For the proof and more details on the result, see [17, Theorem 2] and the discussion therein. Using the traditional uniform sampler followed by a quantizer, a similar order-1 asymptotic optimality result cannot be obtained by controlling the sampling period with a constant number of quantization bits [73, section IV-B]. The significant performance gain of level-triggered sampling against uniform sampling is also shown numerically in [73, section V].

Order-1 is the most frequent type of asymptotic optimality encountered in the literature, but it is also the weakest. Note that in order-1 asymptotic optimality, although the average decision time ratio converges to 1, the difference $\mathsf{E}_j[S] - \mathsf{E}_j[S_o]$ may be unbounded. Therefore, stronger types of asymptotic optimality are defined. The difference remains bounded (i.e., $\mathsf{E}_j[S] - \mathsf{E}_j[S_o] = O(1)$) in order-2 and diminishes (i.e., $\mathsf{E}_j[S] - \mathsf{E}_j[S_o] = o(1)$) in order-3. The latter is extremely rare in the literature, and the schemes of that type are considered optimum per se for practical purposes.

20.2.2.4 Noisy Channels

In the presence of noisy channels, one subtle issue is that since the sensors asynchronously sample and transmit the local LLR, the FC needs to first reliably detect the sampling time to update the global LLR. We first assume that the sampling time is reliably detected and focus on deriving the LLR update at the FC. We discuss the issue of sampling time detection later on.

In computing the LLR $\tilde{\lambda}_{k,i}$ of the received signal $z_{k,i}$, we make use of the local sensor error probabilities α_k, β_k, and the channel parameters that characterize the statistical property of the channel.

20.2.2.4.1 Binary Erasure Channels

We first consider binary erasure channels (BECs) between sensors and the FC with erasure probabilities ϵ_k, $k = 1, \ldots, K$. Under BEC, a transmitted bit $b_{k,i}$ is lost with probability ϵ_k, and it is correctly received at the FC (i.e., $z_{k,i} = b_{k,i}$) with probability $1 - \epsilon_k$.

Lemma 20.2

Under BEC with erasure probability ϵ_k, the LLR of $z_{k,i}$ is given by

$$\tilde{\lambda}_{k,i} = \begin{cases} \log \frac{\mathsf{P}_1(z_{k,i}=1)}{\mathsf{P}_0(z_{k,i}=1)} = \log \frac{1-\beta_k}{\alpha_k}, & \text{if } z_{k,i} = 1, \\ \log \frac{\mathsf{P}_1(z_{k,i}=-1)}{\mathsf{P}_0(z_{k,i}=-1)} = \log \frac{\beta_k}{1-\alpha_k}, & \text{if } z_{k,i} = -1. \end{cases} \tag{20.11}$$

PROOF We have $z_{k,i} = b$, $b = \pm 1$, with probability $1 - \epsilon_k$ only when $b_{k,i} = b$. Hence,

$$\mathsf{P}_j(z_{k,i} = b) = \mathsf{P}_j(b_{k,i} = b)(1 - \epsilon_k),\ j = 0, 1.$$

In the LLR expression, the $1 - \epsilon_k$ terms on the numerator and denominator cancel out, giving the result in (20.11). ■

Note that under BEC, the channel parameter ϵ_k is not needed when computing the LLR $\tilde{\lambda}_{k,i}$. Note also that in this case, a received bit bears the same amount of LLR information as in the ideal channel case [cf. (20.8)], although a transmitted bit is not always received. Hence, the channel-aware approach coincides with the conventional approach which relies solely on the received signal. Although the LLR updates in (20.8) and (20.11) are identical, the fusion rules under BEC and ideal channels are not. This is because under BEC, the decision thresholds A and B in (20.6), due to the information loss, are in general different from those in the ideal channel case.

20.2.2.4.2 Binary Symmetric Channels

Next, we consider binary symmetric channels (BSCs) with crossover probabilities ϵ_k between sensors and the FC. Under BSC, the transmitted bit $b_{k,i}$ is flipped (i.e., $z_{k,i} = -b_{k,i}$) with probability ϵ_k, and it is correctly received (i.e., $z_{k,i} = b_{k,i}$) with probability $1 - \epsilon_k$.

Lemma 20.3

Under BSC with crossover probability ϵ_k, the LLR of $z_{k,i}$ can be computed as

$$\tilde{\lambda}_{k,i} = \begin{cases} \log \frac{1-\hat{\beta}_k}{\hat{\alpha}_k}, & \text{if } z_{k,i} = 1, \\ \log \frac{\hat{\beta}_k}{1-\hat{\alpha}_k}, & \text{if } z_{k,i} = -1, \end{cases} \tag{20.12}$$

where $\hat{\alpha}_k = \alpha_k(1 - 2\epsilon_k) + \epsilon_k$ and $\hat{\beta}_k = \beta_k(1 - 2\epsilon_k) + \epsilon_k$.

PROOF Due to the nonzero probability of receiving a wrong bit, we now have

$$\begin{aligned} \mathsf{P}_j(z_{k,i} = b) &= \mathsf{P}(z_{k,i} = b|b_{k,i} = b)\mathsf{P}_j(b_{k,i} = b) \\ &\quad + \mathsf{P}(z_{k,i} = b|b_{k,i} = -b)\mathsf{P}_j(b_{k,i} = -b), \\ \text{e.g., } \mathsf{P}_0(z_{k,i} = 1) &= (1 - \epsilon_k)\alpha_k + \epsilon_k(1 - \alpha_k), \end{aligned}$$

$j = 0, 1$, $b = \pm 1$. Defining $\hat{\alpha}_k = \alpha_k(1 - 2\epsilon_k) + \epsilon_k$ and $\hat{\beta}_k = \beta_k(1 - 2\epsilon_k) + \epsilon_k$, we obtain the LLR expression given in (20.12). ■

Note that for $\alpha_k < 0.5$, $\beta_k < 0.5$, $\forall k$, which we assume true for $\Delta > 0$,

$$\hat{\alpha}_k = \alpha_k + \epsilon_k(1 - 2\alpha_k) > \alpha_k,$$

and similarly $\hat{\beta}_k > \beta_k$. Thus, $|\tilde{\lambda}_{k,i}^{BSC}| < |\tilde{\lambda}_{k,i}^{BEC}|$ from which we expect a higher performance loss under BSC than the one under BEC. Finally, note also that, unlike the BEC case, under BSC the FC needs to know the channel parameters $\{\epsilon_k\}$ to operate in a channel-aware manner.

20.2.2.4.3 Additive White Gaussian Noise Channels

Now, assume that the channel between each sensor and the FC is an additive white Gaussian noise (AWGN) channel. The received signal at the FC is given by

$$z_{k,i} = y_{k,i} + w_{k,i}, \tag{20.13}$$

where $w_{k,i} \sim \mathcal{N}_c(0, \sigma_k^2)$ is the complex white Gaussian noise, and $y_{k,i}$ is the transmitted signal at sampling time $t_{k,i}$, given by

$$y_{k,i} = \begin{cases} a, & \text{if } \lambda_{k,i} \geq \Delta, \\ b, & \text{if } \lambda_{k,i} \leq -\Delta, \end{cases} \tag{20.14}$$

where the transmission levels a and b are complex in general.

Lemma 20.4

Under the AWGN channel model in (20.13), the LLR of $z_{k,i}$ is given by

$$\tilde{\lambda}_{k,i} = \log \frac{(1 - \beta_k)e^{-c_{k,i}} + \beta_k e^{-d_{k,i}}}{\alpha_k e^{-c_{k,i}} + (1 - \alpha_k)e^{-d_{k,i}}}, \tag{20.15}$$

where $c_{k,i} = \frac{|z_{k,i}-a|^2}{\sigma_k^2}$ and $d_{k,i} = \frac{|z_{k,i}-b|^2}{\sigma_k^2}$.

PROOF The distribution of the received signal given $y_{k,i}$ is $z_{k,i} \sim \mathcal{N}_c(y_{k,i}, \sigma_k^2)$. The probability density function of $z_{k,i}$ under H_j is then given by

$$\begin{aligned} g_{k,j}(z_{k,i}) &= g_{k,j}(z_{k,i}|y_{k,i} = a)\mathsf{P}_j(y_{k,i} = a) \\ &\quad + g_{k,j}(z_{k,i}|y_{k,i} = b)\mathsf{P}_j(y_{k,i} = b), \\ \text{e.g., } g_{k,1}(z_{k,i}) &= \frac{(1 - \beta_k)e^{-\frac{|z_{k,i}-a|^2}{\sigma_k^2}} + \beta_k e^{-\frac{|z_{k,i}-b|^2}{\sigma_k^2}}}{\pi\sigma_k^2}. \end{aligned} \tag{20.16}$$

Defining $c_{k,i} \triangleq \frac{|z_{k,i}-a|^2}{\sigma_k^2}$ and $d_{k,i} \triangleq \frac{|z_{k,i}-b|^2}{\sigma_k^2}$, and substituting $g_{k,0}(z_{k,i})$ and $g_{k,1}(z_{k,i})$ into $\tilde{\lambda}_{k,i} = \log \frac{g_{k,1}(z_{k,i})}{g_{k,0}(z_{k,i})}$, we obtain (20.15). ■

If the transmission levels a and b are well separated, and the signal-to-noise ratio $\frac{|y_{k,i}|}{|w_{k,i}|}$ is high enough, then

$$\tilde{\lambda}_{k,i} \approx \begin{cases} \log \frac{1-\beta_k}{\alpha_k}, & \text{if } y_{k,i} = a, \\ \log \frac{\beta_k}{1-\alpha_k}, & \text{if } y_{k,i} = b, \end{cases}$$

resembling the ideal channel case, given by (20.8). Due to the energy constraints at nodes, assume a maximum transmission power P^2. In accordance with the above observation, it is shown in [74, Section V-C] that the antipodal signaling (e.g., $a = P$ and $b = -P$) is optimum.

20.2.2.4.4 Rayleigh Fading Channels

Assuming a Rayleigh fading channel model, the received signal is given by

$$z_{k,i} = h_{k,i} y_{k,i} + w_{k,i}, \tag{20.17}$$

where $h_{k,i} \sim \mathcal{N}_c(0, \sigma_{h,k}^2)$, $y_{k,i}$, and $w_{k,i}$ are as before.

Lemma 20.5

Under the Rayleigh fading channel model in (20.17), the LLR of $z_{k,i}$ is given by

$$\tilde{\lambda}_{k,i} = \log \frac{\frac{1-\beta_k}{\sigma_{a,k}^2} e^{-c_{k,i}} + \frac{\beta_k}{\sigma_{b,k}^2} e^{-d_{k,i}}}{\frac{\alpha_k}{\sigma_{a,k}^2} e^{-c_{k,i}} + \frac{1-\alpha_k}{\sigma_{b,k}^2} e^{-d_{k,i}}}, \tag{20.18}$$

where $c_{k,i} = \frac{|z_{k,i}|^2}{\sigma_{a,k}^2}$, $d_{k,i} = \frac{|z_{k,i}|^2}{\sigma_{b,k}^2}$, $\sigma_{a,k}^2 = |a|^2 \sigma_{h,k}^2 + \sigma_k^2$, and $\sigma_{b,k}^2 = |b|^2 \sigma_{h,k}^2 + \sigma_k^2$.

PROOF Given $y_{k,i}$, we have $z_{k,i} \sim \mathcal{N}_c(0, |y_{k,i}|^2 \sigma_{h,k}^2 + \sigma_k^2)$. Similar to (20.16), we can write

$$\begin{aligned} g_{k,1}(z_{k,i}) &= \frac{1-\beta_k}{\pi \sigma_{a,k}^2} e^{-c_{k,i}} + \frac{\beta_k}{\pi \sigma_{b,k}^2} e^{-d_{k,i}}, \\ g_{k,0}(z_{k,i}) &= \frac{\alpha_k}{\pi \sigma_{a,k}^2} e^{-c_{k,i}} + \frac{1-\alpha_k}{\pi \sigma_{b,k}^2} e^{-d_{k,i}}, \end{aligned} \tag{20.19}$$

where $c_{k,i} \triangleq \frac{|z_{k,i}|^2}{\sigma_{a,k}^2}$, $d_{k,i} \triangleq \frac{|z_{k,i}|^2}{\sigma_{b,k}^2}$, $\sigma_{a,k}^2 \triangleq |a|^2 \sigma_{h,k}^2 + \sigma_k^2$, and $\sigma_{b,k}^2 \triangleq |b|^2 \sigma_{h,k}^2 + \sigma_k^2$. Substituting $g_{k,0}(z_{k,i})$ and $g_{k,1}(z_{k,i})$ into $\tilde{\lambda}_{k,i} = \log \frac{g_{k,1}(z_{k,i})}{g_{k,0}(z_{k,i})}$, we obtain (20.18). ■

In this case, different messages a and b are expressed only in the variance of $z_{k,i}$. Hence, with antipodal signaling, they become indistinguishable (i.e., $\sigma_{a,k}^2 = \sigma_{b,k}^2$) and as a result $\tilde{\lambda}_{k,i} = 0$. This suggests that we should separate $|a|$ and $|b|$ as much as possible to decrease the uncertainty at the FC, and in turn to decrease the loss in the LLR update $\tilde{\lambda}_{k,i}$ with respect to the ideal channel case. Assuming a minimum transmission power Q^2 to ensure reliable detection of an incoming signal at the FC, in addition to the maximum transmission power P^2 due to the energy constraints, it is numerically shown in [74, Section V-D] that the optimum signaling scheme corresponds to either $|a| = P$, $|b| = Q$ or $|a| = Q$, $|b| = P$.

20.2.2.4.5 Rician Fading Channels

For Rician fading channels, we have $h_{k,i} \sim \mathcal{N}_c(\mu_k, \sigma_{h,k}^2)$ in (20.17).

Lemma 20.6

With Rician fading channels, $\tilde{\lambda}_{k,i}$ is given by (20.18), where $c_{k,i} = \frac{|z_{k,i} - a\mu_k|^2}{\sigma_{a,k}^2}$, $d_{k,i} = \frac{|z_{k,i} - b\mu_k|^2}{\sigma_{b,k}^2}$, $\sigma_{a,k}^2 = |a|^2 \sigma_{h,k}^2 + \sigma_k^2$, and $\sigma_{b,k}^2 = |b|^2 \sigma_{h,k}^2 + \sigma_k^2$.

PROOF Given $y_{k,i}$, the received signal is distributed as $z_{k,i} \sim \mathcal{N}_c(\mu_k y_{k,i}, |y_{k,i}|^2 \sigma_{h,k}^2 + \sigma_k^2)$. The likelihoods $g_{k,1}(z_{k,i})$ and $g_{k,0}(z_{k,i})$ are then written as in (20.19) with $\sigma_{a,k}^2 = |a|^2 \sigma_{h,k}^2 + \sigma_k^2$, $\sigma_{b,k}^2 = |b|^2 \sigma_{h,k}^2 + \sigma_k^2$, and the new definitions $c_{k,i} = \frac{|z_{k,i} - a\mu_k|^2}{\sigma_{a,k}^2}$, $d_{k,i} = \frac{|z_{k,i} - b\mu_k|^2}{\sigma_{b,k}^2}$. Finally, the LLR is given by (20.18). ■

The Rician model covers the previous two continuous channel models. Particularly, the $\sigma_{h,k}^2 = 0$ case corresponds to the AWGN model, and the $\mu_k = 0$ case corresponds to the Rayleigh model. It is numerically shown in [74, Section V-E] that depending on the values of parameters $(\mu_k, \sigma_{h,k}^2)$, either the antipodal signaling of the AWGN case or the ON–OFF type signaling of the Rayleigh case is optimum.

20.2.2.4.6 Discussions

Considering the unreliable detection of sampling times under continuous channels, we should ideally integrate this uncertainty into the fusion rule of the FC. In other words, at the FC, the LLR of received signal

$$z_k[n] = h_k[n] y_k[n] + w_k[n],$$

instead of $z_{k,i}$ given in (20.17), should be computed at each time instant n if the sampling time of node k cannot be reliably detected. In the LLR computations of Lemmas 20.4 and 20.5, the prior probabilities $\mathsf{P}_j(y_{k,i} = a)$ and $\mathsf{P}_j(y_{k,i} = b)$ are used. These probabilities are in fact conditioned on the sampling time $t_{k,i}$. Here, we need the unconditioned prior probabilities of the signal $y_k[n]$ which at each time n takes a value of a or b or 0, that is,

$$y_k[n] = \begin{cases} a & \text{if } L_k[n] - L_k[t_{k,i-1}] \geq \Delta, \\ b & \text{if } L_k[n] - L_k[t_{k,i-1}] \leq -\Delta, \\ 0 & \text{if } L_k[n] - L_k[t_{k,i-1}] \in (-\Delta, \Delta), \end{cases}$$

instead of $y_{k,i}$ given in (20.14).

Then, the LLR of $z_k[n]$ is given by

$$\begin{aligned}
\tilde{\lambda}_k[n] &= \log \frac{g_{k,1}(z_k[n])}{g_{k,0}(z_k[n])},\\
g_{k,1}(z_k[n]) &= \Big[g_{k,1}(z_k[n]|y_k[n]=a)(1-\beta_k)\\
&\quad + g_{k,1}(z_k[n]|y_k[n]=b)\beta_k\Big]\mathsf{P}_1(y_k[n]\neq 0)\\
&\quad + g_{k,1}(z_k[n]|y_k[n]=0)\,\mathsf{P}_1(y_k[n]=0),\\
g_{k,0}(z_k[n]) &= \Big[g_{k,0}(z_k[n]|y_k[n]=a)\alpha_k\\
&\quad + g_{k,0}(z_k[n]|y_k[n]=b)(1-\alpha_k)\Big]\\
&\quad \times \mathsf{P}_0(y_k[n]\neq 0)\\
&\quad + g_{k,0}(z_k[n]|y_k[n]=0)\,\mathsf{P}_0(y_k[n]=0),
\end{aligned}$$

where $g_{k,j}(z_k[n]|y_k[n])$ is determined by the channel model. Since the FC has no prior information on the sampling times of nodes, the probability of sampling, that is, $\mathsf{P}_j(y_k[n]\neq 0)$, can be shown to be $\frac{1}{\mathsf{E}_j[\tau_{k,i}]}$, where $\mathsf{E}_j[\tau_{k,i}]$ is the average sampling interval of node k under H_j, $j=0,1$.

Alternatively, a two-step procedure can be applied by first detecting a message and then using the LLR updates previously derived in Lemmas 20.4–20.6. Since it is known that most of the time $\tilde{\lambda}_k[n]$ is uninformative, corresponding to the no message case, a simple thresholding can be applied to perform LLR update only when it is informative. The thresholding step is in fact a Neyman–Pearson test (i.e., LRT) between the presence and absence of a message signal. The threshold can be adjusted to control the false alarm (i.e., type-I error) and misdetection (i.e., type-II error) probabilities. Setting the threshold sufficiently high, we can obtain a negligible false alarm probability, leaving us with the misdetection probability. Thus, if an LLR survives after thresholding, in the second step it is recomputed as in the channel-aware fusion rules obtained in Lemmas 20.4–20.6.

An information-theoretic analysis for the decentralized detectors in Sections 20.2.2.3 and 20.2.2.4 can be found in [74]. Specifically, using renewal processes, closed-form expressions for average decision time are derived under both the nonasymptotic and asymptotic regimes.

20.2.3 Multimodal Decentralized Detection

In monitoring of complex systems, multimodal data, such as sensor measurements, images, and texts, are collected from disparate sources. The emerging concepts of Internet of Things (IoT) and Cyber-Physical Systems (CPS) show that there is an increasing interest in connecting more and more devices with various sensing capabilities [41]. The envisioned future power grid, called Smart Grid, is a good example for such heterogeneous networks. Monitoring and managing wide-area smart grids require the integration of multimodal data from electricity consumers, such as smart home and smart city systems, as well as various electricity generators (e.g., wind, solar, coal, nuclear) and sensing devices across the grid [15].

Multisensor surveillance (e.g., for military or environmental purposes) is another application in which multimodal data from a large number of sensors (e.g., acoustic, seismic, infrared, optical, magnetic, temperature) are fused for a common statistical task [18]. An interesting multidisciplinary example is nuclear facility monitoring for treaty verification. From a data-processing perspective, using a variety of disparate information sources, such as electricity consumption, satellite images, radiation emissions, seismic vibrations, shipping manifests, and intelligence data, a nuclear facility can be monitored to detect anomalous events that violate a nuclear treaty.

Information-theoretic and machine-learning approaches to similar problems can be found in [18] and [56], respectively. We here follow a Bayesian probabilistic approach to the multimodal detection problem.

20.2.3.1 Latent Variable Model

Consider a system of K information sources (i.e., a network of K nodes). From each source k, a discrete-time signal $x_k[n], n\in\mathbb{N}$, is observed, which follows the probability distribution $\mathcal{D}_k(\boldsymbol{\theta}_k)$ with the parameter vector $\boldsymbol{\theta}_k$, $k=1,\ldots,K$. Given $\boldsymbol{\theta}_k$, the temporal observations $\{x_k[n]\}_n$ from source k are assumed iid. Some information sources may be of same modality.

A latent variable vector $\boldsymbol{\phi}$ is assumed to correlate information sources by controlling their parameters (Figure 20.5). Then, the joint distribution of all observations collected until time N can be written as

$$\begin{aligned}
&f\left(\{x_k[n]\}_{k=1,n=1}^{K,N}\right)\\
&\quad=\int_{\chi_\phi}\int_{\chi_1}\cdots\int_{\chi_K} f\Big(\{x_k[n]\}_{k,n}\big|\{\boldsymbol{\theta}_k\},\boldsymbol{\phi}\Big)\\
&\qquad\times f\Big(\{\boldsymbol{\theta}_k\}\big|\boldsymbol{\phi}\Big)\,f\Big(\boldsymbol{\phi}\Big)\,\mathrm{d}\boldsymbol{\theta}_1\cdots\mathrm{d}\boldsymbol{\theta}_K\,\mathrm{d}\boldsymbol{\phi},
\end{aligned}$$

where χ_ϕ and χ_k are the supports of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}_k$, $k=1,\ldots,K$. Assuming $\{\boldsymbol{\theta}_k\}$ are independent,

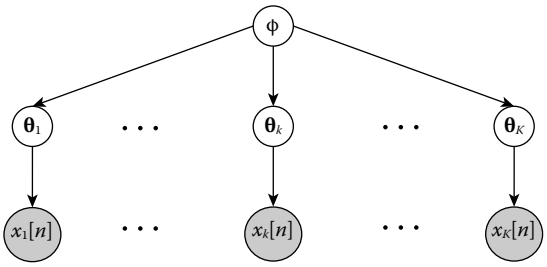

FIGURE 20.5

A Bayesian network of K information sources linked through the latent variable vector $\boldsymbol{\phi}$. The probability distribution of the observation $x_k[n]$ is parameterized by the random vector $\boldsymbol{\theta}_k$, whose distribution is determined by $\boldsymbol{\phi}$, which is also random. The observed variables are represented by filled circles.

given $\boldsymbol{\phi}$ we have

$$\begin{aligned} & f\Big(\{x_k[n]\}_{k,n}\Big) \\ &= \int_{\chi_\phi} \Bigg\{ \int_{\chi_1} \prod_{n=1}^{N} f\Big(x_1[n]\big|\boldsymbol{\theta}_1\Big)\, f\Big(\boldsymbol{\theta}_1\big|\boldsymbol{\phi}\Big)\, \mathrm{d}\boldsymbol{\theta}_1 \\ &\quad \times \cdots \int_{\chi_K} \prod_{n=1}^{N} f\Big(x_K[n]\big|\boldsymbol{\theta}_K\Big)\, f\Big(\boldsymbol{\theta}_K\big|\boldsymbol{\phi}\Big)\, \mathrm{d}\boldsymbol{\theta}_K \Bigg\} f\Big(\boldsymbol{\phi}\Big)\, \mathrm{d}\boldsymbol{\phi} \\ &= \int_{\chi_\phi} f\Big(\{x_1[n]\}\big|\boldsymbol{\phi}\Big) \cdots f\Big(\{x_K[n]\}\big|\boldsymbol{\phi}\Big)\, f\Big(\boldsymbol{\phi}\Big)\, \mathrm{d}\boldsymbol{\phi}, \end{aligned} \tag{20.20}$$

where $f\Big(x_k[n]\big|\boldsymbol{\theta}_k\Big)$, $k = 1, \ldots, K$, is the probability density/mass function of the distribution $\mathcal{D}_k(\boldsymbol{\theta}_k)$. If $f\Big(\boldsymbol{\theta}_k\big|\boldsymbol{\phi}\Big)$ corresponds to the conjugate prior distribution for $\mathcal{D}_k(\boldsymbol{\theta}_k)$, then $f\Big(\{x_k[n]\}_n\big|\boldsymbol{\phi}\Big)$ can be written in closed form.

20.2.3.2 Hypothesis Testing

If the latent variable vector $\boldsymbol{\phi}$ is deterministically specified under both hypotheses, that is,

$$\begin{aligned} \mathsf{H}_0 &: \boldsymbol{\phi} = \boldsymbol{\phi}_0, \\ \mathsf{H}_1 &: \boldsymbol{\phi} = \boldsymbol{\phi}_1, \end{aligned} \tag{20.21}$$

then observations $\{x_k[n]\}_k$ from different sources are independent under H_j, $j = 0, 1$, since $\{\boldsymbol{\theta}_k\}$ are assumed independent given $\boldsymbol{\phi}$. In that case, the global likelihood under H_j is given by (20.20) without the integral over $\boldsymbol{\phi}$, that is,

$$f_j(\{x_k[n]\}_{k,n}) = \prod_{k=1}^{K} f\Big(\{x_k[n]\}_n\big|\boldsymbol{\phi} = \boldsymbol{\phi}_j\Big).$$

Using $f_1(\{x_k[n]\}_{k,n})$ and $f_0(\{x_k[n]\}_{k,n})$, the global LLR at time N is written as

$$L[N] = \sum_{k=1}^{K} \log \frac{f\Big(\{x_k[n]\}_{n=1}^{N}\big|\boldsymbol{\phi} = \boldsymbol{\phi}_1\Big)}{f\Big(\{x_k[n]\}_{n=1}^{N}\big|\boldsymbol{\phi} = \boldsymbol{\phi}_0\Big)} = \sum_{k=1}^{K} L_k[N]. \tag{20.22}$$

For sequential detection, SPRT can be applied by comparing $L[n]$ at each time to two thresholds A and $-B$. The sequential test continues until the stopping time

$$S = \min\{n \in \mathbb{N} : L[n] \notin (-B, A)\}, \tag{20.23}$$

and makes the decision

$$d_S = \begin{cases} \mathsf{H}_1, & \text{if } L[S] \geq A, \\ \mathsf{H}_0, & \text{if } L[S] \leq -B, \end{cases} \tag{20.24}$$

at time S.

In a decentralized system, where all observations cannot be made available to the FC due to resource constraints, each node k (corresponding to information source k) can compute its LLR $L_k[n]$ and transmit event-based samples of it to the FC, as will be described in Section 20.2.3.4. Then, summing the LLR messages from nodes, the FC computes the approximate global LLR $\tilde{L}[n]$ and uses it in the SPRT procedure similar to (20.23) and (20.24).

In many cases, it may not be possible to deterministically specify $\boldsymbol{\phi}$ under the hypotheses, but a statistical description may be available, that is,

$$\begin{aligned} \mathsf{H}_0 &: \boldsymbol{\phi} \sim \mathcal{D}_{\phi,0}(\boldsymbol{\theta}_{\phi,0}), \\ \mathsf{H}_1 &: \boldsymbol{\phi} \sim \mathcal{D}_{\phi,1}(\boldsymbol{\theta}_{\phi,1}). \end{aligned} \tag{20.25}$$

In such a case, to compute the likelihood under H_j, we need to integrate over $\boldsymbol{\phi}$ as shown in (20.20). Hence, in general, the global LLR

$$\begin{aligned} & L[N] \\ &= \log \frac{\int_{\chi_\phi} f\Big(\{x_1[n]\}\big|\boldsymbol{\phi}\Big) \cdots f\Big(\{x_K[n]\}\big|\boldsymbol{\phi}\Big)\, f_1\Big(\boldsymbol{\phi}\Big)\, \mathrm{d}\boldsymbol{\phi}}{\int_{\chi_\phi} f\Big(\{x_1[n]\}\big|\boldsymbol{\phi}\Big) \cdots f\Big(\{x_K[n]\}\big|\boldsymbol{\phi}\Big)\, f_0\Big(\boldsymbol{\phi}\Big)\, \mathrm{d}\boldsymbol{\phi}}, \end{aligned} \tag{20.26}$$

does not have a closed-form expression. However, for a reasonable number of latent variables (i.e., entries of $\boldsymbol{\phi}$), effective numerical computation may be possible through Monte Carlo simulations. Once $L[n]$ is numerically computed, SPRT can be applied as in (20.23) and (20.24).

For decentralized detection, each node k can now compute the functions of $\{x_k[n]\}_n$ included in $f\Big(\{x_k[n]\}_n\big|\boldsymbol{\phi}\Big)$ (see the example below), which has a

closed-form expression thanks to the assumed conjugate prior on the parameter vector $\boldsymbol{\theta}_k$ [see (20.20)], and send event-based samples to the FC. Upon receiving such messages, the FC computes approximations to those functions; uses them in (20.26) to compute $\tilde{L}[N]$, an approximate global LLR; and applies the SPRT procedure using $\tilde{L}[N]$. Details will be provided in Section 20.2.3.4.

20.2.3.3 Example

As an example to the multimodal detection scheme presented in this section, consider a system with three types of information sources: Gaussian source (e.g., real-valued physical measurements), Poisson source (e.g., event occurrences), and multinomial source (e.g., texts). We aim to find the closed-form expression of the sufficient statistic $f\left(\{x[n]\}\middle|\boldsymbol{\phi}\right)$ for each modality. Let the discrete-time signals

$$\begin{aligned} &x_g[n] \sim \mathcal{N}(\mu_\phi, \sigma^2),\ \ x_p[n] \sim \text{Pois}(\lambda_\phi), \\ &\boldsymbol{x}_m[n] \sim \text{Mult}(1, \boldsymbol{p}_\phi),\ n \in \mathbb{N}, \end{aligned} \tag{20.27}$$

denote the Gaussian, Poisson, and multinomial observations, respectively (see Figure 20.6). Multinomial distribution with a single trial and category probabilities $\boldsymbol{p}_\phi = [p_{\phi,1}, \ldots, p_{\phi,M}]$ is used for $\boldsymbol{x}_m[n]$, whose realization is a binary vector with an entry 1 at the index corresponding to the category observed at time n, and 0 at the others. The Poisson observation $x_p[n]$ denotes the number of occurrences for an event of interest in a unit time interval, where λ_ϕ is the average rate of event occurrences.

Among the parameters, only the variance σ^2 of the Gaussian model is assumed known. We assume conjugate prior distributions for the unknown parameters. Specifically, we assume a Gaussian prior on the mean μ_ϕ of the Gaussian model, a gamma prior on the rate parameter λ_ϕ of the Poisson model, and a Dirichlet prior on the probability vector $\boldsymbol{p}_\phi$ of the multinomial model, that is,

$$\mu_\phi \sim \mathcal{N}(\bar{\mu}_\phi, \bar{\sigma}_\phi^2),\ \ \lambda_\phi \sim \Gamma(\alpha_\phi, \beta_\phi),\ \ \boldsymbol{p}_\phi \sim \text{Dir}(\boldsymbol{\gamma}_\phi), \tag{20.28}$$

where the hyperparameters $\bar{\mu}_\phi, \bar{\sigma}_\phi^2, \alpha_\phi, \beta_\phi$, and $\boldsymbol{\gamma}_\phi$ are completely specified by the latent variable vector $\boldsymbol{\phi}$.

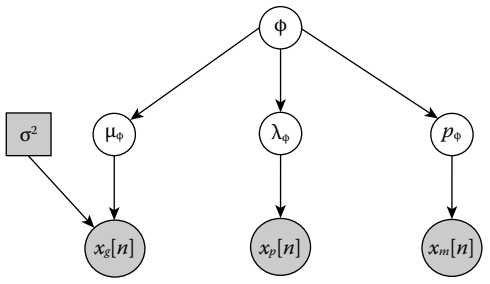

FIGURE 20.6
The Bayesian network considered in the example. The variance of the Gaussian source, which is a known constant, is represented by a filled square.

Lemma 20.7

For the example given in (20.27) and (20.28), the joint distribution of observations from each source conditioned on $\boldsymbol{\phi}$ is given by

$$\begin{aligned} &f\left(\{x_g[n]\}_{n=1}^N\middle|\boldsymbol{\phi}\right) \\ &= \frac{\exp\left(-\frac{\sum_{n=1}^N x_g[n]^2}{2\sigma^2} - \frac{\bar{\mu}_\phi^2}{2\bar{\sigma}_\phi^2} + \frac{\left(\frac{\sum_{n=1}^N x_g[n]}{\sigma^2} + \frac{\bar{\mu}_\phi}{\bar{\sigma}_\phi^2}\right)^2}{2\left(\frac{N}{\sigma^2} + \frac{1}{\bar{\sigma}_\phi^2}\right)}\right)}{(2\pi)^{N/2}\sigma^N\bar{\sigma}_\phi\sqrt{\frac{N}{\sigma^2} + \frac{1}{\bar{\sigma}_\phi^2}}}, \end{aligned} \tag{20.29}$$

$$\begin{aligned} &f\left(\{x_p[n]\}_{n=1}^N\middle|\boldsymbol{\phi}\right) \\ &= \frac{\Gamma\left(\alpha_\phi + \sum_{n=1}^N x_p[n]\right)}{\Gamma(\alpha_\phi)\prod_{n=1}^N x_p[n]!}\frac{\beta_\phi^{\alpha_\phi}}{(\beta_\phi + N)^{\alpha_\phi + \sum_{n=1} x_p[n]}}, \end{aligned} \tag{20.30}$$

$$\begin{aligned} &f\left(\{\boldsymbol{x}_m[n]\}_{n=1}^N\middle|\boldsymbol{\phi}\right) \\ &= \frac{\Gamma\left(\sum_{i=1}^M \gamma_{\phi,i}\right)}{\Gamma\left(\sum_{i=1}^M\left(\gamma_{\phi,i} + \sum_{n=1}^N x_{m,i}[n]\right)\right)} \\ &\quad\times\prod_{i=1}^M\frac{\Gamma\left(\gamma_{\phi,i} + \sum_{n=1}^N x_{m,i}[n]\right)}{\Gamma\left(\gamma_{\phi,i}\right)}, \end{aligned} \tag{20.31}$$

where $\Gamma(\cdot)$ is the gamma function.

PROOF Given $\boldsymbol{\phi}$, $\{x_g[n]\}$ are iid with $\mathcal{N}(\mu_\phi, \sigma^2)$, where $\mu_\phi \sim \mathcal{N}(\bar{\mu}_\phi, \bar{\sigma}_\phi^2)$, hence

$$f\left(\{x_g[n]\}, \mu_\phi\right) = \frac{\exp\left(-\frac{\sum_{n=1}^N (x_g[n]-\mu_\phi)^2}{2\sigma^2} - \frac{(\mu_\phi - \bar{\mu}_\phi)^2}{2\bar{\sigma}_\phi^2}\right)}{(2\pi)^{\frac{N+1}{2}}\sigma^N\bar{\sigma}_\phi}.$$

After some manipulations, we can show that

$$f\left(\{x_g[n]\},\mu_\phi\right) = \frac{\exp\left(-\frac{\sum_{n=1}^N x_g[n]^2}{2\sigma^2}-\frac{\bar{\mu}_\phi^2}{2\bar{\sigma}_\phi^2}+\frac{\left(\frac{\sum_{n=1}^N x_g[n]}{\sigma^2}+\frac{\bar{\mu}_\phi}{\bar{\sigma}_\phi^2}\right)^2}{2\left(\frac{N}{\sigma^2}+\frac{1}{\bar{\sigma}_\phi^2}\right)}\right)}{(2\pi)^{N/2}\sigma^N\bar{\sigma}_\phi\sqrt{\frac{N}{\sigma^2}+\frac{1}{\bar{\sigma}_\phi^2}}} \times \underbrace{\sqrt{\frac{\frac{1}{\bar{\sigma}_\phi^2}+\frac{N}{\sigma^2}}{2\pi}}\exp\left(-\frac{\frac{1}{\bar{\sigma}_\phi^2}+\frac{N}{\sigma^2}}{2}\left(\mu_\phi-\frac{\frac{\bar{\mu}_\phi}{\bar{\sigma}_\phi^2}+\frac{\sum_{n=1}^N x_g[n]}{\sigma^2}}{\frac{1}{\bar{\sigma}_\phi^2}+\frac{N}{\sigma^2}}\right)^2\right)}_{f\left(\mu_\phi|\{x_g[n]\}\right)},$$

where from the conjugate prior property, it is known that the posterior distribution of μ_ϕ is also Gaussian with mean $\frac{\frac{\bar{\mu}_\phi}{\bar{\sigma}_\phi^2}+\frac{\sum_{n=1}^N x_g[n]}{\sigma^2}}{\frac{1}{\bar{\sigma}_\phi^2}+\frac{N}{\sigma^2}}$ and variance $\frac{1}{\bar{\sigma}_\phi^2}+\frac{N}{\sigma^2}$. Hence, the result in (20.29) follows. Note that $f\left(\{x_g[n]\}_{n=1}^N|\boldsymbol{\phi}\right)$ is a multivariate Gaussian distribution, where all entries of the mean vector are $\bar{\mu}_\phi$, the diagonal entries of covariance matrix are $\bar{\sigma}_\phi^2+\sigma^2$, and the off-diagonals are $\bar{\sigma}_\phi^2$.

Similarly, for Poisson observations, we write

$$f\left(\{x_p[n]\},\lambda_\phi\right) = \prod_{n=1}^N \frac{\lambda_\phi^{x_p[n]}e^{-\lambda_\phi}}{x_p[n]!}\frac{\beta_\phi^{\alpha_\phi}}{\Gamma(\alpha_\phi)}\lambda_\phi^{\alpha_\phi-1}e^{-\beta_\phi\lambda_\phi},$$

since $\{x_p[n]\}$ are iid with $\text{Pois}(\lambda_\phi)$ given λ_ϕ, and the prior is $\Gamma(\alpha_\phi,\beta_\phi)$. The posterior distribution is known to be $\Gamma(\alpha_\phi+\sum_{n=1}^N x_p[n],\beta_\phi+N)$; hence,

$$f\left(\{x_p[n]\},\lambda_\phi\right) = \underbrace{\frac{(\beta_\phi+N)^{\alpha_\phi+\sum_{n=1}^N x_p[n]}}{\Gamma\left(\alpha_\phi+\sum_{n=1}^N x_p[n]\right)}\lambda_\phi^{\alpha_\phi+\sum_{n=1}^N x_p[n]-1}e^{-(\beta_\phi+N)\lambda_\phi}}_{f\left(\lambda_\phi|\{x_p[n]\}\right)} \times \underbrace{\frac{\Gamma\left(\alpha_\phi+\sum_{n=1}^N x_p[n]\right)}{\Gamma(\alpha_\phi)\prod_{n=1}^N x_p[n]!}\frac{\beta_\phi^{\alpha_\phi}}{(\beta_\phi+N)^{\alpha_\phi+\sum_{n=1}^N x_p[n]}}}_{f\left(\{x_p[n]\}\right)},$$

proving (20.30).

Finally, for the multinomial observations, $\{\boldsymbol{x}_m[n]\}$ are iid with the probability vector $\boldsymbol{p}_\phi$; the prior is $\text{Dir}(\boldsymbol{\gamma}_\phi)$; and the posterior is $\text{Dir}(\boldsymbol{\gamma}_\phi+\sum_{n=1}^N \boldsymbol{x}_m[n])$; hence,

$$\begin{aligned} f\left(\{\boldsymbol{x}_m[n]\},\boldsymbol{p}_\phi\right) &= \prod_{i=1}^M p_{\phi,i}^{\sum_{n=1}^N x_{m,i}[n]}\frac{\Gamma\left(\sum_{i=1}^M\gamma_{\phi,i}\right)}{\prod_{i=1}^M\Gamma\left(\gamma_{\phi,i}\right)}\prod_{i=1}^M p_{\phi,i}^{\gamma_{\phi,i}-1} \\ &= \frac{\Gamma\left(\sum_{i=1}^M\left(\gamma_{\phi,i}+\sum_{n=1}^N x_{m,i}[n]\right)\right)}{\prod_{i=1}^M\Gamma\left(\gamma_{\phi,i}+\sum_{n=1}^N x_{m,i}[n]\right)}\prod_{i=1}^M p_{\phi,i}^{\gamma_{\phi,i}+\sum_{n=1}^N x_{m,i}[n]-1} \\ &\quad\times \underbrace{\frac{\Gamma\left(\sum_{i=1}^M\gamma_{\phi,i}\right)}{\Gamma\left(\sum_{i=1}^M\left(\gamma_{\phi,i}+\sum_{n=1}^N x_{m,i}[n]\right)\right)} \times \prod_{i=1}^M\frac{\Gamma\left(\gamma_{\phi,i}+\sum_{n=1}^N x_{m,i}[n]\right)}{\Gamma\left(\gamma_{\phi,i}\right)}}_{f\left(\{\boldsymbol{x}_m[n]\}|\boldsymbol{\phi}\right)}, \end{aligned}$$

concluding the proof. ■

In testing hypotheses that deterministically specify $\boldsymbol{\phi}$ as in (20.21), the local LLR for each modality can be computed using Lemma 20.7, for example,

$$L_g[N] = \log\frac{f\left(\{x_g[n]\}_{n=1}^N|\boldsymbol{\phi}=\boldsymbol{\phi}_1\right)}{f\left(\{x_g[n]\}_{n=1}^N|\boldsymbol{\phi}=\boldsymbol{\phi}_0\right)}.$$

Then, under a centralized setup, SPRT can applied as in (20.23) and (20.24) using the global LLR

$$L[N] = L_g[N]+L_p[N]+L_m[N], \tag{20.32}$$

or under a decentralized setup, each node reports event-based samples of its local LLR, and the FC applies SPRT using $\tilde{L}[N]=\tilde{L}_g[N]+\tilde{L}_p[N]+\tilde{L}_m[N]$.

On the contrary, while testing hypotheses that statistically specify $\boldsymbol{\phi}$ as in (20.25), the global LLR is computed using the results of Lemma 20.7 in (20.26). In this case, for decentralized detection, each node can only compute the functions of its observations that appear in the conditional joint distributions, given by (20.29)–(20.31), and do not depend on $\boldsymbol{\phi}$.

For example, the Gaussian node from (20.29) can compute

$$\frac{\sum_{n=1}^N x_g[n]^2}{2\sigma^2} \text{ and } \frac{\sum_{n=1}^N x_g[n]}{\sigma^2}, \tag{20.33}$$

and send their event-based samples to the FC, which can effectively recover such samples as will be shown next, and uses them in (20.29). Although, in this case, event-based sampling is used to transmit only some simple functions of the observations, which needs further

processing at the FC, the advantages of using event-based sampling on the functions of $x_g[n]$, instead of conventional uniform sampling on $x_g[n]$ itself, is still significant. First, the error induced by using the recovered functions in the highly nonlinear expression of (20.29) is smaller than that results from using the recovered observations in (20.29) because transmission loss grows with the processing at the FC. Second, the transmission rate can be considerably lower than that of uniform sampling because only the important changes in functions are reported, censoring the uninformative observations.

The Poisson and multinomial processes are inherently event-based as each observation $x_p[n]/\boldsymbol{x}_m[n]$ marks an event occurrence after a random (e.g., exponentially distributed for Poisson process) waiting time since $x_p[n-1]/\boldsymbol{x}_m[n-1]$. Therefore, each new observation $x_p[n]/\boldsymbol{x}_m[n]$ is reported to the FC. Moreover, they take integer values ($\boldsymbol{x}_m[n]$ can be represented by the index of nonzero element); thus, no quantization error takes place.

20.2.3.4 Decentralized Implementation

Due to the nonlinear processing of recovered messages at the FC [cf. (20.33)], in the decentralized testing of hypotheses with statistical descriptions of $\boldsymbol{\phi}$, we should more carefully take care of the overshoot problem.

In level-triggered sampling, the change in the signal is measured with respect to the signal value at the most recent sampling time, which possibly includes an overshoot and hence is not perfectly available to the FC even if a multibit scheme is used to quantize the overshoot. Therefore, the past quantization errors, as well as the current one, cumulatively decrease the precision of recovered signal at the FC. The accumulation of quantization errors may not be of practical interest if the individual errors are small (i.e., sufficiently large number of bits are used for quantization and/or the jumps in the signal are sufficiently small) and stay small after the processing at the FC, and the FC makes a quick decision (i.e., the constraints on detection error probabilities are not very stringent). However, causing an avalanche effect, it causes a significant problem for the asymptotic decision time performance of the decentralized detector (e.g., in a regime of large decision times due to stringent error probability constraints) even if the individual errors at the FC are small.

In [31], using of fixed reference levels is proposed to improve the asymptotic performance of level-triggered sampling, which corresponds to LCSH (see Figure 20.7). Since LCSH handles the overshoot problem better than level-triggered sampling, it suits better to the case in (20.33) where the FC performs nonlinear processing on the recovered signal. We here show that it also achieves a better asymptotic performance at the expense of much more complicated nonasymptotic performance analysis. Furthermore, we consider multiple crossings of sampling levels due to large jumps in the signal (Figure 20.7).

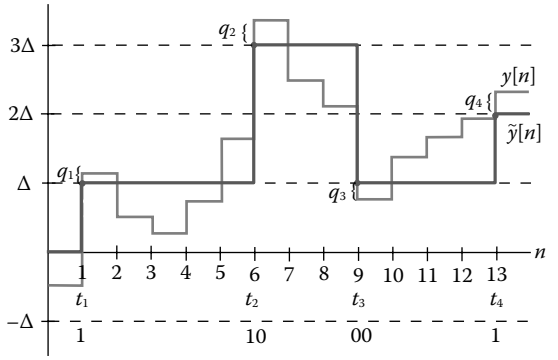

FIGURE 20.7

Level-crossing sampling with hysteresis applied to $y[n]$. The recovered signal $\tilde{y}[n]$ at the FC and the transmitted bits are shown. Multiple crossings are handled by transmitting additional bits. Overshoots $\{q_i\}$ take effect individually (no overshoot accumulation).

Sampling a signal $y[n]$ via LCSH with level spacing Δ, a sample is taken whenever an upper or lower sampling level is crossed, as shown in Figure 20.7. Specifically, the ith sample is taken at time

$$t_i \triangleq \{n > t_{i-1} : |y[n] - \psi_{i-1}\Delta| \geq \Delta\}, \tag{20.34}$$

where ψ_{i-1} is the sampling level in terms of Δ that was most recently crossed. In general, $y[t_i]$ may cross multiple levels, that is, the number of level crossings

$$\eta_i \triangleq \left\lfloor \frac{|y[t_i] - \psi_{i-1}\Delta|}{\Delta} \right\rfloor \geq 1. \tag{20.35}$$

In addition to the sign bit

$$b_{i,1} = \text{sign}(y[t_i] - \psi_{i-1}\Delta), \tag{20.36}$$

which encodes the first crossing with its direction, we send

$$r \triangleq \left\lceil \frac{\eta_i - 1}{2} \right\rceil, \tag{20.37}$$

more bits $b_{i,2}, \ldots, b_{i,r+1}$, where each following bit 1/0 represents a double/single crossing. For instance, the bit sequence 0110, where the first 0 denotes downward crossing (i.e., $b_{i,1} = -1$), is sent for $-7\Delta < y[t_i] - \psi_{i-1}\Delta \leq -6\Delta$.

In that way, the FC can obtain η_i from received bits and keep track of the most recently crossed level as

$$\psi_i = \psi_{i-1} + b_{i,1}\eta_i. \tag{20.38}$$

It approximates $y[n]$ with

$$\tilde{y}[n] = \psi_i\Delta, \;\; t_i \leq n < t_{i+1}. \tag{20.39}$$

As a result, only the current overshoot causes error in $\tilde{y}[n]$; that is, overshoots do not accumulate, as opposed

to level-triggered sampling. This is especially important when reporting signals that are processed further at the FC [cf. (20.33)]. It also ensures order-2 asymptotic optimality with finite number of bits per sample when used to transmit iid local LLRs from unimodal sources (i.e., the case in Theorem 20.1).

Theorem 20.2

Consider the decentralized detector that uses the LCSH-based transmission scheme given in (20.34)–(20.39) to report the local LLRs, $\{L_k[n]\}$, from iid nodes to the FC, and applies the SPRT procedure at the FC substituting the recovered global LLR, $\tilde{L}[n] = \sum_{k=1}^{K} \tilde{L}_k[n]$, in (20.23) and (20.24). It is order-2 asymptotically optimum, that is,

$$\mathsf{E}_j[S] - \mathsf{E}_j[S_o] = O(1),\ j = 0, 1,\ \text{ as } \alpha, \beta \to 0, \quad (20.40)$$

where S and S_o are the decision times of decentralized detector and the optimum (centralized) SPRT satisfying the same (type-I and type-II) error probability bounds α and β [cf. (20.5)].

PROOF Assuming finite and nonzero Kullback–Leibler (KL) information numbers $-\mathsf{E}_0[L[1]], \mathsf{E}_1[L[1]]$, for order-2 asymptotic optimality, it suffices to show that

$$\mathsf{E}_1\left[L[S]\right] - \mathsf{E}_1\left[L[S_o]\right] = O(1),\ j = 0, 1,\ \text{ as } \alpha, \beta \to 0. \quad (20.41)$$

The proof under H_0 follows similarly. Let us start by writing

$$\mathsf{E}_1\left[L[S]\right] = \mathsf{E}_1\left[\tilde{L}[S] + (L[S] - \tilde{L}[S])\right]. \quad (20.42)$$

Thanks to the multibit transmission scheme based on LCSH, no overshoot accumulation takes place, and thus the absolute errors satisfy

$$\begin{aligned} &|\tilde{L}_k[n] - L_k[n]| < \Delta,\ \forall k, n, \\ &|\tilde{L}[n] - L[n]| = \Big|\sum_{k=1}^{K} \tilde{L}_k[n] - L_k[n]\Big| \\ &\qquad \le \sum_{k=1}^{K} |\tilde{L}_k[n] - L_k[n]| < K\Delta,\ \forall n. \end{aligned} \quad (20.43)$$

The approximate LLR $\tilde{L}[S]$ at the stopping time exceeds A or $-B$ by a finite amount, that is,

$$\tilde{L}[S] < A + C \ \text{ or } \ \tilde{L}[S] > -B - C, \quad (20.44)$$

where C is a constant. Now let us analyze how the stopping threshold A behaves as $\alpha, \beta \to 0$. Start with

$$\alpha = \mathsf{P}_0(\tilde{L}[S] \ge A) = \mathsf{E}_0\left[\mathbb{1}_{\{\tilde{L}[S] \ge A\}}\right],$$

where applying a change of measure using $e^{-L[S]}$ as in Lemma 20.1 we can write

$$\begin{aligned} \alpha &= \mathsf{E}_1\left[e^{-L[S]}\mathbb{1}_{\{\tilde{L}[S] \ge A\}}\right] \\ &= \mathsf{E}_1\left[e^{-\tilde{L}[S] + \tilde{L}[S] - L[S]}\mathbb{1}_{\{\tilde{L}[S] \ge A\}}\right]. \end{aligned}$$

From (20.43),

$$\begin{aligned} \alpha &\le e^{-A + K\Delta} \\ A &\le |\log \alpha| + K\Delta. \end{aligned} \quad (20.45)$$

Combining (20.42)–(20.45), we get

$$\mathsf{E}_1\left[L[S]\right] \le |\log \alpha| + 2K\Delta + C. \quad (20.46)$$

In SPRT with discrete-time observations, due to the overshoot problem, the KL divergence at the stopping time is larger than that in the no-overshoot case [53, page 21], that is,

$$\begin{aligned} \mathsf{E}_1[L[S_o]] &\ge (1-\beta)\log\frac{1-\beta}{\alpha} + \beta\log\frac{\beta}{1-\alpha} \\ &= (1-\beta)|\log\alpha| - \beta|\log\beta| + (1-\beta)\log(1-\beta) \\ &\quad - \beta\log(1-\alpha). \end{aligned} \quad (20.47)$$

From (20.46) and (20.47),

$$\begin{aligned} &\mathsf{E}_1\left[L[S]\right] - \mathsf{E}_1[L[S_o] \\ &\quad \le 2K\Delta + C - \beta|\log\alpha| - \beta|\log\beta| \\ &\qquad + (1-\beta)\log(1-\beta) - \beta\log(1-\alpha), \end{aligned}$$

where the last three terms tend to zero as $\alpha, \beta \to 0$. Assuming α and β tend to zero at comparable rates, the term $\beta|\log\alpha|$ also tends to zero, leaving us with the constant $2K\Delta + C$. The decentralized detector applies the SPRT procedure with a summary of observations; hence, it cannot satisfy the error probability constraints with a smaller KL divergence than that of the centralized SPRT, that is, $\mathsf{E}_1\left[L[S]\right] - \mathsf{E}_1[L[S_o]] \ge 0$. This concludes the proof. ■

In fact, the proof for (20.40) holds also for the case of multimodal sources in (20.22) and (20.32), where the local LLRs are independent but not identically distributed. Since in this non-iid case SPRT may not be optimum, we cannot claim asymptotic optimality by satisfying (20.40). However, centralized SPRT still serves as a very important benchmark; hence, (20.40) is a valuable result also for the multimodal case.

The power of Theorem 20.2 lies in the fact that the LCSH-based decentralized detector achieves order-2 asymptotic optimality by using a finite (in most cases small) number of bits per sample. Order-2 asymptotic optimality resolves the overshoot problem because it

is the state-of-the-art performance in the no-overshoot case (i.e., with continuous-time band-limited observations), achieved by LCSH, which coincides with level-triggered sampling in this case. On the contrary, for order-2 asymptotic optimality with discrete-time observations, the number of bits per sample required by the level-triggered sampling-based detector tends to infinity with a reasonably low rate, $\log|\log\alpha|$ [73, Section IV-B].

In the LCSH-based detector, to avoid overshoot accumulation, the overshoot of the last sample is included toward the new sample, correlating the two samples. Consequently, samples (i.e., messages of change in the signal) that result from LCSH are neither independent nor identically distributed. As opposed to level-triggered sampling, in which samples are iid and hence form a renewal process, the statistical descriptions of samples in LCSH are quite intractable. The elegant (nonasymptotic and asymptotic) results obtained for level-triggered sampling in [74] therefore do not apply to LCSH here.

20.3 Decentralized Estimation

In this section, we are interested in sequentially estimating a vector of parameters (i.e., regression coefficients) $\boldsymbol{\theta} \in \mathbb{R}^p$ at a random stopping time S in the following linear (regression) model:

$$x[n] = \boldsymbol{h}[n]^T\boldsymbol{\theta} + w[n],\ n \in \mathbb{N}, \tag{20.48}$$

where $x[n] \in \mathbb{R}$ is the observed sample, $\boldsymbol{h}[n] \in \mathbb{R}^p$ is the vector of regressors, and $w[n] \in \mathbb{R}$ is the additive noise. We consider the general case in which $\boldsymbol{h}[n]$ is random and observed at time n, which covers the deterministic $\boldsymbol{h}[n]$ case as a special case. This linear model is commonly used in many applications. For example, in system identification, $\boldsymbol{\theta}$ is the unknown system coefficients, $\boldsymbol{h}[n]$ is the (random) input applied to the system, and $x[n]$ is the output at time n. Another example is the estimation of wireless (multiple-access) channel coefficients, in which $\boldsymbol{\theta}$ is the unknown channel coefficients, $\boldsymbol{h}[n]$ is the transmitted (random) pilot signal, $x[n]$ is the received signal, and $w[n]$ is the additive channel noise.

In (20.48), at each time n, we observe the sample $\boldsymbol{x}[n]$ and the vector $\boldsymbol{h}[n]$; hence, $\{(\boldsymbol{x}[m], \boldsymbol{h}[m])\}_{m=1}^n$ are available. We assume $\{\boldsymbol{w}[n]\}$ are i.i.d. with $\mathsf{E}[\boldsymbol{w}[n]] = 0$ and $\mathsf{Var}(\boldsymbol{w}[n]) = \sigma^2$. The least squares (LS) estimator minimizes the sum of squared errors, that is,

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} (x[n] - \boldsymbol{h}[n]^T\boldsymbol{\theta})^2, \tag{20.49}$$

and is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_N &= \left(\sum_{n=1}^{N} \boldsymbol{h}[n]\boldsymbol{h}[n]^T\right)^{-1} \sum_{n=1}^{N} \boldsymbol{h}[n]x[n] \\ &= (\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}\boldsymbol{H}_n^T\boldsymbol{x}_n,\end{aligned} \tag{20.50}$$

where $\boldsymbol{H}_n = [\boldsymbol{h}[1], \ldots, \boldsymbol{h}[n]]^T$ and $\boldsymbol{x}_n = [x[1], \ldots, x[n]]^T$. Note that spatial diversity (i.e., a vector of observations and a regressor matrix at time n) can be easily incorporated in (20.48) in the same way we deal with temporal diversity. Specifically, in (20.49) and (20.50), we would also sum over the spatial dimensions.

Under the Gaussian noise, $w[n] \sim \mathcal{N}(0, \sigma^2)$, the LS estimator coincides with the minimum variance unbiased estimator (MVUE) and achieves the CRLB, that is, $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n) = \mathrm{CRLB}_n$. To compute the CRLB, we first write, given $\boldsymbol{\theta}$ and $\boldsymbol{H}_n$, the log-likelihood of the vector $\boldsymbol{x}_n$ as

$$\begin{aligned}L_n &= \log f(\boldsymbol{x}_n|\boldsymbol{\theta}, \boldsymbol{H}_n) \\ &= -\sum_{m=1}^{n} \frac{(\boldsymbol{x}[m] - \boldsymbol{h}[m]^T\boldsymbol{\theta})^2}{2\sigma^2} - \frac{t}{2}\log(2\pi\sigma^2).\end{aligned} \tag{20.51}$$

Then, we have

$$\mathrm{CRLB}_n = \left(\mathsf{E}\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}^2} L_n \middle| \boldsymbol{H}_n\right]\right)^{-1} = \sigma^2\boldsymbol{U}_n^{-1}, \tag{20.52}$$

where $\mathsf{E}\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}^2} L_n \middle| \boldsymbol{H}_n\right]$ is the Fisher information matrix and $\boldsymbol{U}_n \triangleq \boldsymbol{H}_n^T\boldsymbol{H}_n$ is a nonsingular matrix. Since $\mathsf{E}[\boldsymbol{x}_n|\boldsymbol{H}_n] = \boldsymbol{H}_n\boldsymbol{\theta}$ and $\mathsf{Cov}(\boldsymbol{x}_n|\boldsymbol{H}_n) = \sigma^2\boldsymbol{I}$, from (20.50) we have $\mathsf{E}[\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n] = \boldsymbol{\theta}$ and $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n) = \sigma^2\boldsymbol{U}_n^{-1}$; thus, from (20.52) $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n) = \mathrm{CRLB}_n$. Note that the maximum likelihood (ML) estimator that maximizes (20.51) coincides with the LS estimator in (20.50).

In general, the LS estimator is the best linear unbiased estimator (BLUE). In other words, any linear unbiased estimator of the form $\boldsymbol{A}_n\boldsymbol{x}_n$ with $\boldsymbol{A}_n \in \mathbb{R}^{n\times t}$, where $\mathsf{E}[\boldsymbol{A}_n\boldsymbol{x}_n|\boldsymbol{H}_n] = \boldsymbol{\theta}$, has a covariance no smaller than that of the LS estimator in (20.50), that is, $\mathsf{Cov}(\boldsymbol{A}_n\boldsymbol{x}_n|\boldsymbol{H}_n) \geq \sigma^2\boldsymbol{U}_n^{-1}$ in the positive semidefinite sense. To see this result, we write $\boldsymbol{A}_n = (\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}\boldsymbol{H}_n^T + \boldsymbol{B}_n$ for some $\boldsymbol{B}_n \in \mathbb{R}^{n\times t}$, and then $\mathsf{Cov}(\boldsymbol{A}_n\boldsymbol{x}_n|\boldsymbol{H}_n) = \sigma^2\boldsymbol{U}_n^{-1} + \sigma^2\boldsymbol{B}_n\boldsymbol{B}_n^T$, where $\boldsymbol{B}_n\boldsymbol{B}_n^T$ is a positive semidefinite matrix.

The recursive least squares (RLS) algorithm enables us to compute $\hat{\boldsymbol{\theta}}_n$ in a recursive way as follows:

$$\begin{aligned}&\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n-1} + \boldsymbol{q}_n(x[n] - \boldsymbol{h}[n]^T\hat{\boldsymbol{\theta}}_{n-1}), \\ &\text{where } \boldsymbol{q}_n = \frac{\boldsymbol{P}_{n-1}\boldsymbol{h}[n]}{1 + \boldsymbol{h}[n]^T\boldsymbol{P}_{n-1}\boldsymbol{h}[n]} \\ &\text{and } \boldsymbol{P}_n = \boldsymbol{P}_{n-1} - \boldsymbol{q}_n\boldsymbol{h}[n]^T\boldsymbol{P}_{n-1},\end{aligned} \tag{20.53}$$

where $\boldsymbol{q}_n \in \mathbb{R}^p$ is a gain vector and $\boldsymbol{P}_n = \boldsymbol{U}_n^{-1}$. While applying RLS, we first initialize $\hat{\boldsymbol{\theta}}_0 = 0$ and $\boldsymbol{P}_0 = \delta^{-1}\boldsymbol{I}$,

where 0 represents a zero vector and δ is a small number, and then at each time n compute $\boldsymbol{q}_n$, $\hat{\boldsymbol{\theta}}_n$, and $\boldsymbol{P}_n$ as in (20.53).

20.3.1 Background

Energy constraints are inherent to wireless sensor networks [1]. Since data transmission is the primary source of energy consumption, it is essential to keep transmission rates low in wireless sensor networks, resulting in a *decentralized* setup. Decentralized parameter estimation is a fundamental task performed in wireless sensor networks [5,10,14,35,45,48,51,52,54,68,69,78]. In *sequential* estimation, the objective is to minimize the (average) number of observations for a given target accuracy level [36]. To that end, a sequential estimator $(S, \hat{\boldsymbol{\theta}}_S)$, as opposed to a traditional fixed-sample-size estimator, is equipped with a stopping rule which determines an appropriate time S to stop taking new observations based on the observation history. Hence, the stopping time S (i.e., the number of observations used in estimation) is a random variable. Endowed with a stopping mechanism, a sequential estimator saves not only time but also energy, both of which are critical resources. In particular, it avoids unnecessary data processing and transmission.

Decentralized parameter estimation has been mainly studied under two different network topologies. In the first one, sensors communicate to an FC that performs estimation based on the received information (e.g., [14,35,45,48,51,68]). The other commonly studied topology is called ad hoc network, in which there is no designated FC, but sensors compute their local estimators and communicate them through the network (e.g., [5,10,52,54,78]). Decentralized estimation under both network topologies is reviewed in [69]. Many existing works consider parameter estimation in linear models (e.g., [10,14,35,45,54,68]). Whereas in [5,48,51,52,69,78] a general nonlinear signal model is assumed. The majority of existing works on decentralized estimation (e.g., [10,14,35,45,48,51,52,54,68,69]) study fixed-sample-size estimation. There are a few works, such as [5,16], that consider sequential decentralized parameter estimation. Nevertheless, [5] assumes that sensors transmit real numbers, and [16] focuses on continuous-time observations, which can be seen as practical limitations.

In decentralized detection [17,73,74,76] and estimation [75], level-triggered sampling (cf. Figure 20.3), an adaptive sampling technique which infrequently transmits a few bits, for example, one bit, from sensors to the FC, has been used to achieve low-rate transmission. It has been also shown that the decentralized schemes based on level-triggered sampling significantly outperform their counterparts based on conventional uniform sampling in terms of average stopping time. We here use a form of level-triggered sampling that infrequently transmits a single pulse from sensors to the FC and, at the same time, achieves a close-to-optimum average stopping time performance [76].

The stopping capability of sequential estimators comes with the cost of sophisticated analysis. In most cases, it is not possible with discrete-time observations to find an optimum sequential estimator that attains the sequential Cramér-Rao lower bound (CRLB) if the stopping time S is adapted to the complete observation history [20]. Alternatively, in [22] and more recently in [16,75], it was proposed to restrict S to stopping times that are adapted to a specific subset of the complete observation history, which leads to simple optimum solutions. This idea of using a restricted stopping time first appeared in [22] with no optimality result. In [16], with continuous-time observations, a sequential estimator with a restricted stopping time was shown to achieve the sequential version of the CRLB for scalar parameter estimation. In [75], for scalar parameter estimation with discrete-time observations, a similar sequential estimator was shown to achieve the conditional sequential CRLB for the same restricted class of stopping times.

We deal with discrete-time observations in this section. In Section 20.3.2, the optimum sequential estimator that achieves the conditional sequential CRLB for a certain class of stopping times is discussed. We then develop in Section 20.3.3 a computation- and energy-efficient decentralized scheme based on level-triggered sampling for sequential estimation of vector parameters.

20.3.2 Optimum Sequential Estimator

In this section, we aim to find the optimal pair $(S, \hat{\boldsymbol{\theta}}_S)$ of stopping time and estimator corresponding to the optimal sequential estimator. The stopping time for a sequential estimator is determined according to a target estimation accuracy. In general, the average stopping time is minimized subject to a constraint on the estimation accuracy, which is a function of the estimator covariance, that is,

$$\min_{S, \hat{\boldsymbol{\theta}}_S} \mathsf{E}[S] \;\text{ s.t. }\; f\left(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S)\right) \leq C, \tag{20.54}$$

where $f(\cdot)$ is a function from $\mathbb{R}^{p\times p}$ to $\mathbb{R}$ and $C \in \mathbb{R}$ is the target accuracy level.

The accuracy function f should be a monotonic function of the covariance matrix $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S)$, which is positive semidefinite, to make consistent accuracy assessments; for example, $f(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S)) > f(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_{S'}))$ for $S < S'$ since $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S) \succ \mathsf{Cov}(\hat{\boldsymbol{\theta}}_{S'})$ in the positive definite sense. Two popular and easy-to-compute choices are the trace $\mathsf{Tr}(\cdot)$, which corresponds to the mean squared error (MSE), and the Frobenius norm $\|\cdot\|_F$. Before handling the

problem in (20.54), let us explain why we are interested in restricted stopping times that are adapted to a subset of observation history.

20.3.2.1 Restricted Stopping Time

Denote $\{\mathcal{F}_n\}$ as the filtration that corresponds to the samples $\{\boldsymbol{x}[1],\ldots,\boldsymbol{x}[n]\}$ where $\mathcal{F}_n = \sigma\{\boldsymbol{x}[1],\ldots,\boldsymbol{x}[n]\}$ is the σ-algebra generated by the samples observed up to time n, that is, the accumulated history related to the observed samples, and $\mathcal{F}_0$ is the trivial σ-algebra. Similarly, we define the filtration $\{\mathcal{H}_n\}$ where $\mathcal{H}_n = \sigma\{\boldsymbol{h}[1],\ldots,\boldsymbol{h}[n]\}$ and $\mathcal{H}_0$ is again the trivial σ-algebra. It is known that, in general, *with discrete-time observations* and an unrestricted stopping time, that is $\{\mathcal{F}_n \cup \mathcal{H}_n\}$-adapted, the sequential CRLB is not attainable under any noise distribution except for the Bernoulli noise [20].

On the contrary, in the case of *continuous-time observations with continuous paths*, the sequential CRLB is attained by the LS estimator with an $\{\mathcal{H}_n\}$-adapted stopping time that depends only on $\boldsymbol{H}_S$ [16]. Moreover, in the following lemma, we show that, with discrete-time observations, the LS estimator attains the conditional sequential CRLB for the $\{\mathcal{H}_n\}$-adapted stopping times.

Lemma 20.8

With a monotonic accuracy function f and an $\{\mathcal{H}_n\}$-adapted stopping time S, we can write

$$f\left(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\mathcal{H}_S)\right) \geq f\left(\sigma^2\boldsymbol{U}_S^{-1}\right), \tag{20.55}$$

for all unbiased estimators under Gaussian noise, and for all linear unbiased estimators under non-Gaussian noise, and the LS estimator

$$\hat{\boldsymbol{\theta}}_S = \boldsymbol{U}_S^{-1}V_S,\ V_S \triangleq \boldsymbol{H}_S^T\boldsymbol{x}_S, \tag{20.56}$$

satisfies the inequality in (20.55) with equality.

PROOF Since the LS estimator, with $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\mathcal{H}_n) = \sigma^2\boldsymbol{U}_n^{-1}$, is the MVUE under Gaussian noise and the BLUE under non-Gaussian noise, we write

$$\begin{aligned}
&f\left(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\mathcal{H}_S)\right)\\
&\quad= f\left(\mathsf{E}\left[\sum_{n=1}^{\infty}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})^T\,\mathbb{1}_{\{n=S\}}\big|\boldsymbol{H}_n\right]\right)\\
&\quad= f\left(\sum_{n=1}^{\infty}\mathsf{E}\left[(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})^T\big|\boldsymbol{H}_n\right]\,\mathbb{1}_{\{n=S\}}\right) \qquad (20.57)\\
&\quad\geq f\left(\sum_{n=1}^{\infty}\sigma^2\boldsymbol{U}_n^{-1}\,\mathbb{1}_{\{n=S\}}\right) \qquad (20.58)\\
&\quad= f\left(\sigma^2\boldsymbol{U}_S^{-1}\right), \qquad (20.59)
\end{aligned}$$

for all unbiased estimators under Gaussian noise and for all linear unbiased estimators under non-Gaussian noise. The indicator function $\mathbb{1}_{\{A\}} = 1$ if A is true, and 0 otherwise. We used the facts that the event $\{S=n\}$ is $\mathcal{H}_n$-measurable and $\mathsf{E}[(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})^T|\boldsymbol{H}_n] = \mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n) \geq \sigma^2\boldsymbol{U}_n^{-1}$ to write (20.57) and (20.58), respectively. ■

20.3.2.2 Optimum Conditional Estimator

We are interested in $\{\mathcal{H}_n\}$-adapted stopping times to use the optimality property of the LS estimator in the sequential sense, shown in Lemma 20.8.

The common practice in sequential analysis minimizes the average stopping time subject to a constraint on the estimation accuracy which is a function of the estimator covariance. The optimum solution to this classical problem proves to be intractable for even moderate number of unknown parameters [72]. Hence, it is not a convenient model for decentralized estimation. Therefore, we follow an alternative approach and formulate the problem conditioned on the observed $\{\boldsymbol{h}[n]\}$ values, which yields a tractable optimum solution for any number of parameters.

In the presence of an ancillary statistic whose distribution does not depend on the parameters to be estimated, such as the regressor matrix $\boldsymbol{H}_n$, the conditional covariance $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n)$ can be used to assess the accuracy of the estimator more precisely than the (unconditional) covariance, which is in fact the mean of the former (i.e., $\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S) = \mathsf{E}[\mathsf{Cov}(\hat{\boldsymbol{\theta}}_n|\boldsymbol{H}_n)]$) [12,22]. Motivated by this fact, we propose to reformulate the problem in (20.54) conditioned on $\boldsymbol{H}_n$, that is,

$$\min_{S,\hat{\boldsymbol{\theta}}_S}\mathsf{E}[S]\ \text{ s.t. }\ f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) \leq C. \tag{20.60}$$

Note that the constraint in (20.60) is stricter than the one in (20.54) since it requires that $\hat{\boldsymbol{\theta}}_S$ satisfies the target accuracy level for each realization of $\boldsymbol{H}_S$, whereas in (20.54) it is sufficient that $\hat{\boldsymbol{\theta}}_S$ satisfies the target accuracy level on average. In other words, in (20.54), even if $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) > C$ for some realizations of $\boldsymbol{H}_S$, we can still satisfy $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S)\big) \leq C$. In fact, we can always have $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S)\big) = C$ by using a probabilistic stopping rule such that we sometimes stop above C, that is, $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) > C$, and the rest of the time at or below C, that is, $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) \leq C$. On the contrary, in (20.60) we always have $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) \leq C$; moreover, since we observe discrete-time samples, in general we have $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) < C$ for each realization of $\boldsymbol{H}_S$. Hence, the optimal objective value $\mathsf{E}[S]$ in (20.54) will, in general, be smaller than that in (20.60). Note that on the contrary, if we observed continuous-time processes with continuous paths, then we could always

have $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) = C$ for each realization of $\boldsymbol{H}_S$, and thus the optimal objective values of (20.60) and (20.54) would be the same.

Since minimizing S also minimizes $\mathsf{E}[S]$, in (20.60) we are required to find the first time that a member of our class of estimators (i.e., unbiased estimators under Gaussian noise and linear unbiased estimators under non-Gaussian noise) satisfies the constraint $f\big(\mathsf{Cov}(\hat{\boldsymbol{\theta}}_S|\boldsymbol{H}_S)\big) \leq C$, as well as the estimator that attains this earliest stopping time. From Lemma 20.8, it is seen that the LS estimator, given by (20.56), among its competitors, achieves the best accuracy level $f\big(\sigma^2\boldsymbol{U}_S^{-1}\big)$ at any stopping time S. Hence, for the conditional problem, the optimum sequential estimator is composed of the stopping time

$$S = \min\{n \in \mathbb{N} : f\left(\sigma^2\boldsymbol{U}_n^{-1}\right) \leq C\}, \tag{20.61}$$

and the LS estimator

$$\hat{\boldsymbol{\theta}}_S = \boldsymbol{U}_S^{-1}V_S, \tag{20.62}$$

which can be computed recursively as in (20.53). The recursive computation of $\boldsymbol{U}_n^{-1} = \boldsymbol{P}_n$ in the test statistic in (20.61) is also given in (20.53).

Note that for an accuracy function f such that $f(\sigma^2\boldsymbol{U}_n^{-1}) = \sigma^2 f(\boldsymbol{U}_n^{-1})$, for example, $\mathsf{Tr}(\cdot)$ and $\|\cdot\|_F$, we can use the following stopping time:

$$S = \min\{n \in \mathbb{N} : f\left(\boldsymbol{U}_n^{-1}\right) \leq C'\}, \tag{20.63}$$

where $C' = \frac{C}{\sigma^2}$ is the relative target accuracy with respect to the noise power. Hence, given C' we do not need to know the noise variance σ^2 to run the test given by (20.63). Note that $\boldsymbol{U}_n = \boldsymbol{H}_n^T\boldsymbol{H}_n$ is a nondecreasing positive semidefinite matrix, that is, $\boldsymbol{U}_n \succeq \boldsymbol{U}_{n-1}, \forall t$, in the positive semidefinite sense. Thus, from the monotonicity of f, the test statistic $f\left(\sigma^2\boldsymbol{U}_n^{-1}\right)$ is a nonincreasing scalar function of time. Specifically, for accuracy functions $\mathsf{Tr}(\cdot)$ and $\|\cdot\|_F$, we can show that if the minimum eigenvalue of $\boldsymbol{U}_n$ tends to infinity as $t \to \infty$, then the stopping time is finite, that is, $S < \infty$.

In the conditional problem, for any n, we have a simple stopping rule given in (20.63), which uses the target accuracy level $\frac{C}{\sigma^2}$ as its threshold, hence known beforehand. For the special case of scalar parameter estimation, we do not need a function f to assess the accuracy of the estimator because instead of a covariance matrix we now have a variance $\frac{\sigma^2}{u_n}$, where $u_n = \sum_{m=1}^n h_m^2$ and h_n is the scaling coefficient in (20.48). Hence, from (20.62) and (20.63), the optimum sequential estimator in the scalar case is given by

$$S = \min\left\{n \in \mathbb{N} : u_n \geq \frac{1}{C'}\right\}, \quad \hat{\theta}_S = \frac{v_S}{u_S}, \tag{20.64}$$

where $\frac{u_n}{\sigma^2}$ is the Fisher information at time n. That is, we stop the first time the gathered Fisher information exceeds the threshold $1/C$, which is known.

20.3.3 Decentralized Estimator

In this section, we propose a computation- and energy-efficient decentralized estimator based on the optimum conditional sequential estimator and level-triggered sampling. Consider a network of K distributed sensors and an FC which is responsible for determining the stopping time and computing the estimator. In practice, due to the stringent energy constraints, sensors must infrequently convey low-rate information to the FC, which is the main concern in the design of a decentralized sequential estimator.

As in (20.48), each sensor k observes

$$x_k[n] = \boldsymbol{h}_k[n]^T\boldsymbol{\theta} + w_k[n], \;\; n \in \mathbb{N}, \;\; k = 1, \ldots, K, \tag{20.65}$$

as well as the regressor vector $\boldsymbol{h}_k[n] = [h_{k,1}[n], \ldots, h_{k,p}[n]]^T$ at time n, where $\{w_k[n]\}_{k,n}$ are independent, zero-mean, that is, $\mathsf{E}[w_k[n]] = 0, \;\forall k, n$, and $\mathsf{Var}(w_k[n]) = \sigma_k^2, \;\forall n$. Then, similar to (20.50), the weighted least squares (WLS) estimator

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^K \sum_{m=1}^n \frac{\left(x_k[m] - \boldsymbol{h}_k[m]^T\boldsymbol{\theta}\right)^2}{\sigma_k^2},$$

is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_n &= \left(\sum_{k=1}^K \sum_{m=1}^n \frac{\boldsymbol{h}_k[m]\boldsymbol{h}_k[m]^T}{\sigma_k^2}\right)^{-1} \sum_{k=1}^K \sum_{m=1}^n \frac{\boldsymbol{h}_k[m]x_k[m]}{\sigma_k^2} \\ &= \bar{\boldsymbol{U}}_n^{-1}\bar{V}_n,\end{aligned} \tag{20.66}$$

where $\bar{\boldsymbol{U}}_n^k \triangleq \frac{1}{\sigma_k^2}\sum_{m=1}^n \boldsymbol{h}_k[m]\boldsymbol{h}_k[m]^T$, $\bar{V}_n^k \triangleq \frac{1}{\sigma_k^2}\sum_{m=1}^n \boldsymbol{h}_k[m]x_k[m]$, $\bar{\boldsymbol{U}}_n = \sum_{k=1}^K \bar{\boldsymbol{U}}_n^k$, and $\bar{V}_n = \sum_{k=1}^K \bar{V}_n^k$. As before, it can be shown that the WLS estimator $\hat{\boldsymbol{\theta}}_n$ in (20.66) is the BLUE under the general noise distributions. Moreover, in the Gaussian noise case, where $w_k[n] \sim \mathcal{N}(0, \sigma_k^2) \;\forall n$ for each k, $\hat{\boldsymbol{\theta}}_n$ is also the MVUE.

Following the steps in Section 20.3.2.2, it is straightforward to show that the optimum sequential estimator for the conditional problem in (20.60) is given by the stopping time

$$S = \min\left\{n \in \mathbb{N} : f\left(\bar{\boldsymbol{U}}_n^{-1}\right) \leq C\right\}, \tag{20.67}$$

and the WLS estimator $\hat{\boldsymbol{\theta}}_S$, given by (20.66). Note that $(S, \hat{\boldsymbol{\theta}}_S)$ is achievable only in the centralized case,

where all local observations until time n, that is, $\{(x_k[m], \boldsymbol{h}_k[m])\}_{k=1,m=1}^{K,n}$, are available to the FC. Local processes $\{\bar{\boldsymbol{U}}_n^k\}_{k,n}$ and $\{\bar{V}_n^k\}_{k,n}$ are used to compute the stopping time and the estimator as in (20.67) and (20.66), respectively. On the contrary, in a decentralized system, the FC can compute approximations $\widetilde{\boldsymbol{U}}_n^k$ and $\widetilde{V}_n^k$ and then use these approximations to compute the stopping time and estimator as in (20.67) and (20.66), respectively.

20.3.3.1 Linear Complexity

If each sensor k reports $\bar{\boldsymbol{U}}_n^k \in \mathbb{R}^{p\times p}$ and $\bar{V}_n^k \in \mathbb{R}^p$ to the FC in a straightforward way, then $O(p^2)$ terms need to be transmitted, which may not be practical, especially for large p, in a decentralized setup. Similarly, in the literature, the distributed implementation of the Kalman filter, which covers RLS as a special case, through its inverse covariance form, namely the information filter, requires the transmission of an $p \times p$ information matrix and an $p \times 1$ information vector (e.g., [63]).

To overcome this problem, considering $\mathsf{Tr}(\cdot)$ as the accuracy function f in (20.67), we propose to transmit only the p diagonal entries of $\bar{\boldsymbol{U}}_n^k$ for each k, yielding linear complexity $O(p)$. Using the diagonal entries of $\bar{\boldsymbol{U}}_n$, we define the diagonal matrix

$$\begin{aligned} \boldsymbol{D}_n &\triangleq \mathrm{diag}\left(d_{n,1},\ldots,d_{n,p}\right) \\ \text{where } d_{n,i} &= \sum_{k=1}^{K}\sum_{m=1}^{n}\frac{h_{k,i}[m]^2}{\sigma_k^2},\ i=1,\ldots,p. \end{aligned} \tag{20.68}$$

We further define the correlation matrix

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}, \tag{20.69}$$

$$\text{where } r_{ij} = \frac{\sum_{k=1}^{K}\frac{\mathsf{E}[h_{k,i}[n]h_{k,j}[n]]}{\sigma_k^2}}{\sqrt{\sum_{k=1}^{K}\frac{\mathsf{E}[h_{k,i}[n]^2]}{\sigma_k^2}\sum_{k=1}^{K}\frac{\mathsf{E}[h_{k,j}[n]^2]}{\sigma_k^2}}}, i,j=1,\ldots,p.$$

Proposition 20.1

For sufficiently large n, we can make the following approximations:

$$\begin{aligned} \bar{\boldsymbol{U}}_n &\approx \boldsymbol{D}_n^{1/2}\boldsymbol{R}\,\boldsymbol{D}_n^{1/2} \\ \text{and } \mathsf{Tr}\left(\bar{\boldsymbol{U}}_n^{-1}\right) &\approx \mathsf{Tr}\left(\boldsymbol{D}_n^{-1}\boldsymbol{R}^{-1}\right). \end{aligned} \tag{20.70}$$

PROOF The approximations are motivated from the special case where $\mathsf{E}[h_{k,i}[n]h_{k,j}[n]] = 0,\ \forall k,\ i,j = 1,\ldots,p,\ i \neq j$. In this case, by the *law of large numbers* for sufficiently large n, the off-diagonal elements of $\frac{\bar{\boldsymbol{U}}_n}{n}$ vanish, and thus we have $\frac{\bar{\boldsymbol{U}}_n}{n} \approx \frac{\boldsymbol{D}_n}{n}$ and $\mathsf{Tr}(\bar{\boldsymbol{U}}_n^{-1}) \approx \mathsf{Tr}(\boldsymbol{D}_n^{-1})$. For the general case where we might have $\mathsf{E}[h_{k,i}[n]h_{k,j}[n]] \neq 0$ for some k and $i \neq j$, using the diagonal matrix $\boldsymbol{D}_n$ we write

$$\mathsf{Tr}\left(\bar{\boldsymbol{U}}_n^{-1}\right) = \mathsf{Tr}\left(\left(\boldsymbol{D}_n^{1/2}\underbrace{\boldsymbol{D}_n^{-1/2}\bar{\boldsymbol{U}}_n\boldsymbol{D}_n^{-1/2}}_{\boldsymbol{R}_n}\boldsymbol{D}_n^{1/2}\right)^{-1}\right), \tag{20.71}$$

$$= \mathsf{Tr}\left(\boldsymbol{D}_n^{-1/2}\boldsymbol{R}_n^{-1}\boldsymbol{D}_n^{-1/2}\right),$$

$$= \mathsf{Tr}\left(\boldsymbol{D}_n^{-1}\boldsymbol{R}_n^{-1}\right). \tag{20.72}$$

Note that each entry $r_{n,ij}$ of the newly defined matrix $\boldsymbol{R}_n$ is a normalized version of the corresponding entry $\bar{u}_{n,ij}$ of $\bar{\boldsymbol{U}}_n$. Specifically, $r_{n,ij} = \frac{\bar{u}_{n,ij}}{\sqrt{d_{n,i}d_{n,j}}} = \frac{\bar{u}_{n,ij}}{\sqrt{\bar{u}_{n,ii}\bar{u}_{n,jj}}}$, $i,j = 1,\ldots,p$, where the last equality follows from the definition of $d_{n,i}$ in (20.68). Hence, $\boldsymbol{R}_n$ has the same structure as in (20.69) with entries

$$r_{n,ij} = \frac{\sum_{k=1}^{K}\sum_{m=1}^{n}\frac{h_{k,i}[m]h_{k,j}[m]}{\sigma_k^2}}{\sqrt{\sum_{k=1}^{K}\sum_{m=1}^{n}\frac{h_{k,i}[m]^2}{\sigma_k^2}\sum_{k=1}^{K}\sum_{m=1}^{n}\frac{h_{k,j}[m]^2}{\sigma_k^2}}},$$
$$i,j=1,\ldots,p.$$

For sufficiently large n, by the law of large numbers

$$r_{n,ij} \approx r_{ij} = \frac{\sum_{k=1}^{K}\frac{\mathsf{E}[h_{k,i}[n]h_{k,j}[n]]}{\sigma_k^2}}{\sqrt{\sum_{k=1}^{K}\frac{\mathsf{E}[h_{k,i}[n]^2]}{\sigma_k^2}\sum_{k=1}^{K}\frac{\mathsf{E}[h_{k,j}[n]^2]}{\sigma_k^2}}}, \tag{20.73}$$

and $\boldsymbol{R}_n \approx \boldsymbol{R}$, where $\boldsymbol{R}$ is given in (20.69). Hence, for sufficiently large n, we can make the approximations in (20.70) using (20.71) and (20.72). ■

Then, assuming that the FC knows the correlation matrix $\boldsymbol{R}$, that is, $\left\{\mathsf{E}[h_{k,i}[n]h_{k,j}[n]]\right\}_{i,j,k}$* and $\{\sigma_k^2\}$ [cf. (20.69)], it can compute the approximations in (20.70) if sensors report their local processes $\left\{\boldsymbol{D}_n^k\right\}_{k,n}$ to the FC, where $\boldsymbol{D}_n = \sum_{k=1}^{K}\boldsymbol{D}_n^k$. Note that each local process $\left\{\boldsymbol{D}_n^k\right\}_n$ is p-dimensional, and its entries at time n are

*The subscripts i and j in the set notation denote $i = 1,\ldots,p$ and $j = i,\ldots,p$. In the special case where $\mathsf{E}[h_{k,i}[n]^2] = \mathsf{E}[h_{\ell,i}[n]^2]$, $k,\ell = 1,\ldots,K,\ i = 1,\ldots,p$, the correlation coefficients

$$\left\{\xi_{ij}^k = \frac{\mathsf{E}[h_{k,i}[n]h_{k,j}[n]]}{\sqrt{\mathsf{E}[h_{k,i}[n]^2]\mathsf{E}[h_{k,j}[n]^2]}} :\ i=1,\ldots,p-1,\ j=i+1,\ldots,p\right\}_k,$$

together with $\{\sigma_k^2\}$ are sufficient statistics since $r_{ij} = \frac{\sum_{k=1}^{K}\xi_{ij}^k/\sigma_k^2}{\sum_{k=1}^{K}1/\sigma_k^2}$ from (20.73).

given by $\left\{d_{n,i}^k = \sum_{m=1}^n \frac{h_{k,i}[m]^2}{\sigma_k^2}\right\}_i$ [cf. (20.68)]. Hence, we propose that each sensor k sequentially reports the local processes $\{\boldsymbol{D}_n^k\}_n$ and $\{\bar{V}_n^k\}_n$ to the FC, achieving linear complexity $O(p)$. On the other side, the FC, using the information received from sensors, computes the approximations $\{\widetilde{\boldsymbol{D}}_n\}$ and $\{\widetilde{V}_n\}$, which are then used to compute the stopping time

$$\widetilde{S} = \min\left\{n \in \mathbb{N} : \mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_n^{-1}\right) \leq \widetilde{C}\right\}, \tag{20.74}$$

and the estimator

$$\widetilde{\boldsymbol{\theta}}_{\widetilde{S}} = \widetilde{\boldsymbol{U}}_{\widetilde{S}}^{-1}\widetilde{V}_{\widetilde{S}}, \tag{20.75}$$

similar to (20.67) and (20.66), respectively. The approximations $\mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_n^{-1}\right)$ in (20.74) and $\widetilde{\boldsymbol{U}}_{\widetilde{S}}$ in (20.75) are computed using $\widetilde{\boldsymbol{D}}_n$ as in (20.70). The threshold $\widetilde{C}$ is selected through simulations to satisfy the constraint in (20.60) with equality, that is, $\mathsf{Tr}\left(\mathsf{Cov}\left(\widetilde{\boldsymbol{\theta}}_{\widetilde{S}}|\boldsymbol{H}_{\widetilde{S}}\right)\right) = C$.

20.3.3.2 Event-Based Transmission

Level-triggered sampling provides a very convenient way of information transmission in decentralized systems [17,73–76]. Specifically, decentralized methods based on level-triggered sampling, transmitting low-rate information, enable highly accurate approximations and thus high-performance schemes at the FC. They significantly outperform conventional decentralized methods, which sample local processes using the traditional uniform sampling and send the quantized versions of samples to the FC [73,75].

Existing methods employ level-triggered sampling to report a scalar local process to the FC. Using a similar procedure to report each distinct entry of $\bar{\boldsymbol{U}}_n^k$ and $\bar{V}_n^k$, we need $O(p^2)$ parallel procedures, which may be prohibitive in a decentralized setup for large p. Hence, we use the approximations introduced in the previous subsection, achieving linear complexity $O(p)$. Data transmission and thus energy consumption also scale linearly with the number of parameters, which may easily become prohibitive for a sensor with limited battery. We address this energy efficiency issue by infrequently transmitting a single pulse with very short duration, which encodes, in time, the overshoot in level-triggered sampling [76].

We will next describe the proposed decentralized estimator based on level-triggered sampling in which each sensor nonuniformly samples the local processes $\{\boldsymbol{D}_n^k\}_n$ and $\{\bar{V}_n^k\}_n$, and transmits a single pulse for each sample to the FC, and the FC computes $\{\widetilde{\boldsymbol{D}}_n\}$ and $\{\widetilde{V}_n\}$ using received information.

20.3.3.2.1 Sampling and Recovery of $\boldsymbol{D}_n^k$

Each sensor k samples each entry $d_{n,i}^k$ of $\boldsymbol{D}_n^k$ at a sequence of random times $\{s_{m,i}^k\}_{m\in\mathbb{N}}$ given by

$$s_{m,i}^k \triangleq \min\left\{n \in \mathbb{N} : d_{n,i}^k - d_{s_{m-1,i},i}^k \geq \Delta_i^k\right\}, \; s_{0,i}^k = 0, \tag{20.76}$$

where $d_{n,i}^k = \sum_{p=1}^n \frac{h_{k,i}[p]^2}{\sigma_k^2}$, $d_{0,i}^k = 0$, and $\Delta_i^k > 0$ is a constant threshold that controls the average sampling interval. Note that the sampling times $\{s_{m,i}^k\}_m$ in (20.76) are dynamically determined by the signal to be sampled, that is, realizations of $d_{n,i}^k$. Hence, they are random, whereas sampling times in the conventional uniform sampling are deterministic with a certain period. According to the sampling rule in (20.76), a sample is taken whenever the signal level $d_{n,i}^k$ increases by at least Δ_i^k since the last sampling time. Note that $d_{n,i}^k = \sum_{p=1}^n \frac{h_{k,i}[p]2}{\sigma_k^2}$ is nondecreasing in n.

After each sampling time $s_{m,i}^k$, sensor k transmits a single pulse to the FC at time

$$t_{m,i}^k \triangleq s_{m,i}^k + \delta_{m,i}^k,$$

indicating that $d_{n,i}^k$ has increased by at least Δ_i^k since the last sampling time $s_{m-1,i}^k$. The delay $\delta_{m,i}^k$ between the transmission time and the sampling time is used to linearly encode the overshoot

$$q_{m,i}^k \triangleq \left(d_{s_{m,i},i}^k - d_{s_{m-1,i},i}^k\right) - \Delta_i^k, \tag{20.77}$$

and is given by

$$\delta_{m,i}^k = \frac{q_{m,i}^k}{\phi_d} \in [0,1), \tag{20.78}$$

where ϕ_d^{-1} is the slope of the linear encoding function, as shown in Figure 20.8, known to sensors and the FC.

Assume a global clock, that is, the time index $n \in \mathbb{N}$ is the same for all sensors and the FC, meaning that the FC knows the potential sampling times. Assume further ultra-wideband (UWB) channels between sensors and the FC, in which the FC can determine the time of flight of pulses transmitted from sensors. Then, FC can measure the transmission delay $\delta_{m,i}^k$ if it is bounded by unit time, that is, $\delta_{m,i}^k \in [0,1)$. To ensure this, from (20.78), we need to have $\phi_d > q_{m,i}^k$, $\forall k,m,i$. Assuming a bound for overshoots, that is, $q_{m,i}^k < \theta_d$, $\forall k,m,i$, we can achieve this by setting $\phi_d > \theta_d$.

Consequently, the FC can uniquely decode the overshoot by computing $q_{m,i}^k = \phi_d \delta_{m,i}^k$ (cf. Figure 20.8), using

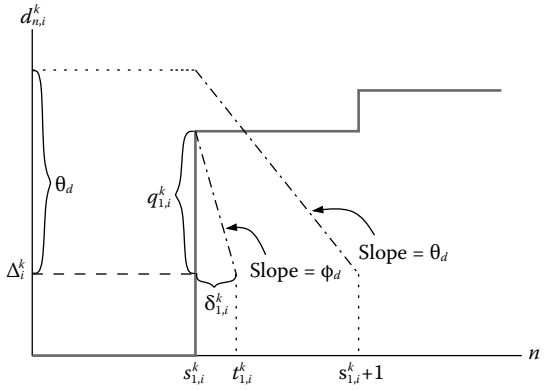

FIGURE 20.8

Illustration of sampling time $s^k_{m,i}$, transmission time $t^k_{m,i}$, transmission delay $\delta^k_{m,i}$, and overshoot $q^k_{m,i}$. We encode $q^k_{m,i} < \theta_d$ in $\delta^k_{m,i} = t^k_{m,i} - s^k_{m,i} < 1$ using the slope $\phi_d > \theta_d$.

which it can also find the increment occurred in $d^k_{n,i}$ during the interval $(s^k_{m-1,i}, s^k_{m,i}]$ as

$$d^k_{s^k_{m,i},i} - d^k_{s^k_{m-1,i},i} = \Delta^k_i + q^k_{m,i},$$

from (20.77). It is then possible to reach the signal level $d^k_{s^k_{m,i},i}$ by accumulating the increments occurred until the mth sampling time, that is,

$$d^k_{s^k_{m,i},i} = \sum_{\ell=1}^{m} \left(\Delta^k_i + q^k_{\ell,i}\right) = m\Delta^k_i + \sum_{\ell=1}^{m} q^k_{\ell,i}. \tag{20.79}$$

Using $\left\{d^k_{s^k_{m,i},i}\right\}_m$, the FC computes the staircase approximation $\widetilde{d}^k_{n,i}$ as

$$\widetilde{d}^k_{n,i} = d^k_{s^k_{m,i},i}, \quad t \in [t^k_{m,i}, t^k_{m+1,i}), \tag{20.80}$$

which is updated when a new pulse is received from sensor k, otherwise kept constant. Such approximate local signals of different sensors are next combined to obtain the approximate global signal $\widetilde{d}_{n,i}$ as

$$\widetilde{d}_{n,i} = \sum_{k=1}^{K} \widetilde{d}^k_{n,i}. \tag{20.81}$$

In practice, when the mth pulse in the global order regarding dimension i is received from sensor k_m at time $t_{m,i}$, instead of computing (20.79) through (20.81), the FC only updates $\widetilde{d}_{n,i}$ as

$$\widetilde{d}_{t_{m,i},i} = \widetilde{d}_{t_{m-1,i},i} + \Delta^{k_m}_i + q_{m,i}, \quad \widetilde{d}_{0,i} = \epsilon, \tag{20.82}$$

and keeps it constant when no pulse arrives. We initialize $\widetilde{d}_{n,i}$ to a small constant ϵ to prevent dividing by zero while computing the test statistic [cf. (20.83)].

Note that in general $\widetilde{d}_{t_{m,i},i} \neq d_{s_{m,i},i}$ unlike (20.80) since all sensors do not necessarily sample and transmit at the same time. The approximations $\left\{\widetilde{d}_{n,i}\right\}_i$ form $\widetilde{\boldsymbol{D}}_n = \mathrm{diag}(\widetilde{d}_{n,1}, \ldots, \widetilde{d}_{n,p})$, which is used in (20.74) and (20.75) to compute the stopping time and the estimator, respectively. Note that to determine the stopping time as in (20.74), we need to compute $\mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_t^{-1}\right)$ using (20.70) at times $\{t_m\}$ when a pulse is received from any sensor regarding any dimension. Fortunately, when the mth pulse in the global order is received from sensor k_m at time t_m regarding dimension i_m, we can compute $\mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_{t_m}^{-1}\right)$ recursively as follows:

$$\begin{aligned}
\mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_{t_m}^{-1}\right) &= \mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_{t_{m-1}}^{-1}\right) - \frac{\kappa_{i_m}(\Delta^{k_m}_{i_m} + q_m)}{\widetilde{d}_{t_m,i_m}\widetilde{d}_{t_{m-1},i_m}}, \\
\mathsf{Tr}\left(\widetilde{\boldsymbol{U}}_0^{-1}\right) &= \sum_{i=1}^{p} \frac{\kappa_i}{\epsilon},
\end{aligned} \tag{20.83}$$

where κ_i is the ith diagonal element of the inverse correlation matrix $\boldsymbol{R}^{-1}$, known to the FC. In (20.83), pulse arrival times are assumed to be distinct for the sake of simplicity. In case multiple pulses arrive at the same time, the update rule will be similar to (20.83) except that it will consider all new arrivals together.

20.3.3.2.2 Sampling and Recovery of $\bar{V}^k_n$

Similar to (20.76), each sensor k samples each entry $\bar{v}^k_{n,i}$ of $\bar{V}^k_n$ at a sequence of random times $\left\{\rho^k_{m,i}\right\}_m$ written as

$$\rho^k_{m,i} \triangleq \min\left\{n \in \mathbb{N} : \left|\bar{v}^k_{n,i} - \bar{v}^k_{\rho^k_{m-1,i},i}\right| \geq \gamma^k_i\right\}, \ \rho^k_{0,i} = 0, \tag{20.84}$$

where $\bar{v}^k_{n,i} = \sum_{p=1}^{n} \frac{h^k_{p,i} y^k_p}{\sigma^2_k}$ and γ^k_i is a constant threshold, available to both sensor k and the FC. See (20.2) for selecting γ^k_i. Since $\bar{v}^k_{n,i}$ is neither increasing nor decreasing, we use two thresholds γ^k_i and $-\gamma^k_i$ in the sampling rule given in (20.84).

Specifically, a sample is taken whenever $\bar{v}^k_{n,i}$ increases or decreases by at least γ^k_i since the last sampling time. Then, after a transmission delay

$$\chi^k_{m,i} = \frac{\eta^k_{m,i}}{\phi_v},$$

where $\eta^k_{m,i} \triangleq \left|\bar{v}^k_{\rho^k_{m,i},i} - \bar{v}^k_{\rho^k_{m-1,i},i}\right| - \gamma^k_i$ is the overshoot, sensor k at time

$$\tau^k_{m,i} \triangleq \rho^k_{m,i} + \chi^k_{m,i},$$

transmits a single pulse $b_{m,i}^k$ to the FC, indicating whether $\bar{v}_{n,i}^k$ has changed by at least γ_i^k or $-\gamma_i^k$ since the last sampling time $\rho_{m-1,i}^k$. We can simply write $b_{m,i}^k$ as

$$b_{m,i}^k = \text{sign}\big(\bar{v}_{\rho_{m,i}^k,i}^k - \bar{v}_{\rho_{m-1,i}^k,i}^k\big). \tag{20.85}$$

Assume again that (i) there exists a global clock among sensors and the FC, (ii) the FC determines channel delay (i.e., time of flight), and (iii) overshoots are bounded by a constant, that is, $\eta_{m,i}^k < \theta_v$, $\forall k, m, i$, and we set $\phi_v > \theta_v$. With these assumptions, we ensure that the FC can measure the transmission delay $\chi_{m,i}^k$ and accordingly decode the overshoot as $\eta_{m,i}^k = \phi_v \chi_{m,i}^k$. Then, upon receiving the mth pulse $b_{m,i}$ regarding dimension i from sensor k_m at time $\tau_{m,i}$, the FC performs the following update:

$$\widetilde{v}_{\tau_{m,i},i} = \widetilde{v}_{\tau_{m-1,i},i} + b_{m,i}\big(\gamma_i^{k_m} + \eta_{m,i}\big), \tag{20.86}$$

where $\{\widetilde{v}_{n,i}\}_i$ compose the approximation $\widetilde{V}_n = [\widetilde{v}_{n,1}, \ldots, \widetilde{v}_{n,p}]^T$. Recall that the FC employs $\widetilde{V}_n$ to compute the estimator as in (20.75).

The level-triggered sampling procedure at each sensor k for each dimension i is summarized in Algorithm 20.1. Each sensor k runs p of these procedures in parallel. The sequential estimation procedure at the FC is also summarized in Algorithm 20.2. We assumed, for the sake of clarity, that each sensor transmits pulses to the FC for each dimension through a separate channel, that is, parallel architecture. On the contrary, in practice the number of parallel channels can be decreased to two by using identical sampling thresholds Δ and γ for all sensors and for all dimensions in (20.76) and (20.84), respectively. Moreover, sensors can even employ a single channel to convey information about local processes $\{d_{n,i}^k\}$ and $\{\bar{v}_{n,i}^k\}$ by sending ternary digits to the FC. This is possible since pulses transmitted for $\{d_{n,i}^k\}$ are unsigned.

20.3.3.3 Discussions

We introduced the decentralized estimator in Section 20.3.3.2 initially for a system with infinite time precision. In practice, due to bandwidth constraints, discrete-time systems with finite precision are of interest. For example, in such systems, the overshoot $q_{m,i}^k \in \left[j\frac{\theta_d}{N}, (j+1)\frac{\theta_d}{N}\right), j = 0, 1, \ldots, N-1$, is quantized into $\hat{q}_{m,i}^k = \left(j + \frac{1}{2}\right)\frac{\theta_d}{N}$, where N is the number of quantization levels. More specifically, a pulse is transmitted at time $t_{m,i}^k = s_{m,i}^k + \frac{j+1/2}{N}$, where the transmission delay $\frac{j+1/2}{N} \in (0,1)$ encodes $\hat{q}_{m,i}^k$. This transmission scheme is called pulse position modulation (PPM).

In UWB and optical communication systems, PPM is effectively employed. In such systems, N, which denotes the precision, can be easily made large enough so that the quantization error $|\hat{q}_{m,i}^k - q_{m,i}^k|$ becomes insignificant. Compared with conventional transmission techniques which convey information by varying the power level, frequency, and/or phase of a sinusoidal wave, PPM (with UWB) is extremely energy efficient at the expense of high bandwidth usage since only a single pulse with very short duration is transmitted per sample. Hence, PPM suits well to energy-constrained sensor network systems.

20.3.3.4 Simulations

We next provide simulation results to compare the performances of the proposed scheme with linear complexity, given in Algorithms 20.1 and 20.2, the nonsimplified version of the proposed scheme with quadratic complexity and the optimal centralized scheme. A wireless sensor network with 10 identical sensors and an FC is considered to estimate a five-dimensional deterministic vector of parameters, that is, $p = 5$. We assume i.i.d. Gaussian noise with unit variance at all sensors, that is, $w_k[n] \sim \mathcal{N}(0,1), \forall k, n$. We set the correlation coefficients $\{r_{ij}\}$ [cf. (20.73)] of the vector $\boldsymbol{h}_k[n]$ to 0 and 0.5 in Figure 20.9 to test the performance of the proposed

Algorithm 20.1 The level-triggered sampling procedure at the kth sensor for the ith dimension

1: Initialization: $n \leftarrow 0$, $m \leftarrow 0$, $\ell \leftarrow 0$, $\lambda \leftarrow 0$, $\psi \leftarrow 0$
2: **while** $\lambda < \Delta_i^k$ **and** $\psi \in (-\gamma_i^k, \gamma_i^k)$ **do**
3: $\quad n \leftarrow n + 1$
4: $\quad \lambda \leftarrow \lambda + \frac{h_{k,i}[n]^2}{\sigma_k^2}$
5: $\quad \psi \leftarrow \psi + \frac{h_{k,i}[n]x_k[n]}{\sigma_k^2}$
6: **end while**
7: **if** $\lambda \geq \Delta_i^k$ {sample $d_{n,i}^k$} **then**
8: $\quad m \leftarrow m + 1$
9: $\quad s_{m,i}^k = n$
10: $\quad$ Send a pulse to the fusion center at time instant $t_{m,i}^k = s_{m,i}^k + \frac{\lambda - \Delta_i^k}{\phi_d}$
11: $\quad \lambda \leftarrow 0$
12: **end if**
13: **if** $\psi \notin (-\gamma_i^k, \gamma_i^k)$ {sample $\bar{v}_{n,i}^k$} **then**
14: $\quad \ell \leftarrow \ell + 1$
15: $\quad \rho_{\ell,i}^k = n$
16: $\quad$ Send $b_{\ell,i}^k = \text{sign}(\psi)$ to the fusion center at time instant $\tau_{\ell,i}^k = \rho_{\ell,i}^k + \frac{|\psi| - \gamma_i^k}{\phi_v}$
17: $\quad \psi \leftarrow 0$
18: **end if**
19: Stop if the fusion center instructs so; otherwise go to line 2.

Algorithm 20.2 The sequential estimation procedure at the fusion center

Initialization: $\mathsf{Tr} \leftarrow \sum_{i=1}^{p} \frac{\kappa_i}{\epsilon}$, $m \leftarrow 1$, $\ell \leftarrow 1$, $\widetilde{d}_i \leftarrow \epsilon \ \forall i$, $\widetilde{v}_i \leftarrow 0 \ \forall i$
while $\mathsf{Tr} < \widetilde{C}$ **do**
  Wait to receive a pulse
  if mth pulse about $d_{n,i}$ arrives from sensor k at time n **then**
    $q_m = \phi_d(n - \lfloor n \rfloor)$
    $\mathsf{Tr} \leftarrow \mathsf{Tr} - \frac{\kappa_i(\Delta_i^k + q_m)}{\widetilde{d}_i(\widetilde{d}_i + \Delta_i^k + q_m)}$
    $\widetilde{d}_i = \widetilde{d}_i + \Delta_i^k + q_m$
    $m \leftarrow m + 1$
  end if
  if ℓth pulse b_ℓ about $v_{n,j}$ arrives from sensor k at time n **then**
    $\eta_\ell = \phi_v(n - \lfloor n \rfloor)$
    $\widetilde{v}_j = \widetilde{v}_j + b_\ell(\gamma_j^k + \eta_\ell)$
    $\ell \leftarrow \ell + 1$
  end if
end while
Stop at time $\widetilde{S} = n$
$\widetilde{\boldsymbol{D}} = \mathrm{diag}(\widetilde{d}_1, \ldots, \widetilde{d}_p)$, $\widetilde{\boldsymbol{U}}^{-1} = \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{R}^{-1}\widetilde{\boldsymbol{D}}^{-1/2}$, $\widetilde{V} = [\widetilde{v}_1, \ldots, \widetilde{v}_p]^T$
$\widetilde{\boldsymbol{\theta}} = \widetilde{\boldsymbol{U}}^{-1}\widetilde{V}$
Instruct sensors to stop.

scheme in the uncorrelated and correlated cases. We compare the average stopping time performance of the proposed scheme with linear complexity to those of the other two schemes for different MSE values. In Figure 20.9, the horizontal axis represents the signal-to-error ratio in decibel, where $n\mathrm{MSE} \triangleq \frac{\mathrm{MSE}}{\|\boldsymbol{\theta}\|_2^2}$, that is, the MSE normalized by the square of the Euclidean norm of the vector to be estimated.

In the uncorrelated case, where $r_{ij} = 0$, $\forall i, j$, $i \neq j$, the proposed scheme with linear complexity nearly attains the performance of the nonsimplified scheme with quadratic complexity as seen in Figure 20.9. This result is rather expected since in this case $\bar{\boldsymbol{U}}_n \approx \boldsymbol{D}_n$ for sufficiently large n, where $\bar{\boldsymbol{U}}_n$ and $\boldsymbol{D}_n$ are used to compute the stopping time and the estimator in the nonsimplified and simplified schemes, respectively. Strikingly, the decentralized schemes (simplified and nonsimplified) achieve very close performances to that of the optimal centralized scheme, which is obviously unattainable in a decentralized system, thanks to the efficient information transmission through level-triggered sampling.

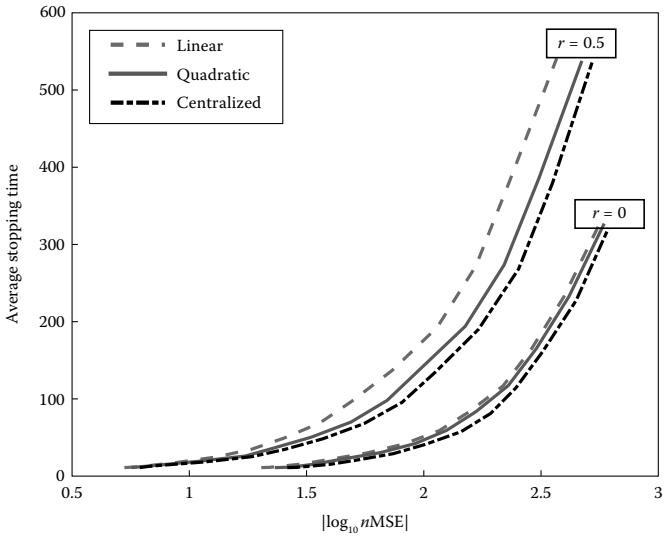

FIGURE 20.9
Average stopping time performances of the optimal centralized scheme and the decentralized schemes based on level-triggered sampling with quadratic and linear complexity versus normalized MSE values when scaling coefficients are uncorrelated, that is, $r_{ij} = 0, \forall i, j$, and correlated with $r_{ij} = 0.5, \forall i, j$.

It is seen in Figure 20.9 that the proposed simplified scheme exhibits an average stopping time performance close to those of the nonsimplified scheme and the optimal centralized scheme even when the scaling coefficients $\{h_{k,i}[n]\}_i$ are correlated with $r_{ij} = 0.5$, $\forall i, j$, $i \neq j$, justifying the simplification proposed in Section 20.3.3.1 to obtain linear complexity.

20.4 Conclusion

Event-based sampling techniques, adapting the sampling times to the signal to be sampled, provide energy- and bandwidth-efficient information transmission in resource-constrained distributed (i.e., decentralized) systems, such as wireless sensor networks. We have first designed and analyzed event-based detection schemes under challenging environments, namely noisy transmission channels between nodes and the fusion center, and multimodal observations from disparate information sources. Then, we have identified an optimum sequential estimator which lends itself to decentralized systems. For large number of unknown parameters, we have further proposed a simplified scheme with linear complexity.

Acknowledgments

This work was funded in part by the U.S. National Science Foundation under grant CIF1064575, the U.S. Office of Naval Research under grant N000141210043, the Consortium for Verification Technology under Department of Energy National Nuclear Security Administration award number DE-NA0002534, and the Army Research Office (ARO) grant number W911NF-11-1-0391.

Bibliography

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102–114, 2002.

[2] K. J. Astrom and B. M. Bernhardsson. Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In *41st IEEE Conference on Decision and Control*, Las Vegas, Nevada, volume 2, pages 2011–2016, December 2002.

[3] R. Bashirullah, J. G. Harris, J. C. Sanchez, T. Nishida, and J. C. Principe. Florida wireless implantable recording electrodes (FWIRE) for brain machine interfaces. In *IEEE International Symposium on Circuits and Systems (ISCAS 2007)*, New Orleans, LA, pages 2084–2087, May 2007.

[4] F. J. Beutler. Error-free recovery of signals from irregularly spaced samples. *SIAM Review*, 8: 328–355, 1966.

[5] V. Borkar and P. P. Varaiya. Asymptotic agreement in distributed estimation. *IEEE Transactions on Automatic Control*, 27(3):650–655, 1982.

[6] Z. Chair and P. K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, 22(1):98–101, 1986.

[7] J.-F. Chamberland and V. V. Veeravalli. Decentralized detection in sensor networks. *IEEE Transactions on Signal Processing*, 51(2):407–416, 2003.

[8] S. Chaudhari, V. Koivunen, and H. V. Poor. Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios. *IEEE Transactions on Signal Processing*, 57(7):2690–2700, 2009.

[9] B. Chen, L. Tong, and P. K. Varshney. Channel-aware distributed detection in wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):16–26, 2006.

[10] A. K. Das and M. Mesbahi. Distributed linear parameter estimation over wireless sensor networks. *IEEE Transactions on Aerospace and Electronic Systems*, 45(4):1293–1306, 2009.

[11] D. Drazen, P. Lichtsteiner, P. Hafliger, T. Delbruck, and A. Jensen. Toward real-time particle tracking using an event-based dynamic vision sensor. *Experiments in Fluids*, 51(5):1465–1469, 2011.

[12] B. Efron and D. V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65(3):457–487, 1978.

[13] P. Ellis. Extension of phase plane analysis to quantized systems. *IRE Transactions on Automatic Control*, 4(2):43–54, 1959.

[14] J. Fang and H. Li. Adaptive distributed estimation of signal power from one-bit quantized data. *IEEE Transactions on Aerospace and Electronic Systems*, 46(4):1893–1905, 2010.

[15] H. Farhangi. The path of the smart grid. *IEEE Power and Energy Magazine*, 8(1):18–28, 2010.

[16] G. Fellouris. Asymptotically optimal parameter estimation under communication constraints. *Annals of Statistics*, 40(4):2239–2265, 2012.

[17] G. Fellouris and G. V. Moustakides. Decentralized sequential hypothesis testing using asynchronous communication. *IEEE Transactions on Information Theory*, 57(1):534–548, 2011.

[18] J. W. Fisher, M. J. Wainwright, E. B. Sudderth, and A. S. Willsky. Statistical and information-theoretic methods for self-organization and fusion of multimodal, networked sensors. *The International Journal of High Performance Computing Applications*, 16(3):337–353, 2002.

[19] J. Fromm and S. Lautner. Electrical signals and their physiological significance in plants. *Plant, Cell & Environment*, 30(3):249–257, 2007.

[20] B. K. Ghosh. On the attainment of the Cramér-rao bound in the sequential case. *Sequential Analysis*, 6(3):267–288, 1987.

[21] D. Gontier and M. Vetterli. Sampling based on timing: Time encoding machines on shift-invariant subspaces. *Applied and Computational Harmonic Analysis*, 36(1):63–78, 2014.

[22] P. Grambsch. Sequential sampling based on the observed fisher information to guarantee the accuracy of the maximum likelihood estimator. *Annals of Statistics*, 11(1):68–77, 1983.

[23] K. M. Guan, S. S. Kozat, and A. C. Singer. Adaptive reference levels in a level-crossing analog-to-digital converter. *EURASIP Journal on Advances in Signal Processing*, 2008:183:1–183:11, 2008.

[24] K. M. Guan and A. C. Singer. Opportunistic sampling by level-crossing. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, HI, volume 3, pages III-1513–III-1516, April 2007.

[25] S. Haykin. *Communication Systems*, 4th edition. Wiley, New York, NY, 2001.

[26] M. Hofstatter, M. Litzenberger, D. Matolin, and C. Posch. Hardware-accelerated address-event processing for high-speed visual object recognition. In *18th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Beirut, Lebanon, pages 89–92, December 2011.

[27] A. M. Hussain. Multisensor distributed sequential detection. *IEEE Transactions on Aerospace and Electronic Systems*, 30(3):698–708, 1994.

[28] E. Kofman and J. H. Braslavsky. Level crossing sampling in feedback stabilization under data-rate constraints. In *45th IEEE Conference on Decision and Control*, San Diego, CA, pages 4423–4428, December 2006.

[29] A. A. Lazar and E. A. Pnevmatikakis. Video time encoding machines. *IEEE Transactions on Neural Networks*, 22(3):461–473, 2011.

[30] A. A. Lazar and L. T. Toth. Perfect recovery and sensitivity analysis of time encoded bandlimited signals. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51(10):2060–2073, 2004.

[31] S. Li and X. Wang. Cooperative change detection for online power quality monitoring. http://arxiv.org/abs/1412.2773.

[32] S. Li and X. Wang. Quickest attack detection in multi-agent reputation systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):653–666, 2014.

[33] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128 $\times$ 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008.

[34] B. Liu and B. Chen. Channel-optimized quantizers for decentralized detection in sensor networks. *IEEE Transactions on Information Theory*, 52(7):3349–3358, 2006.

[35] Z.-Q. Luo, G. B. Giannakis, and S. Zhang. Optimal linear decentralized estimation in a bandwidth constrained sensor network. In *International Symposium on Information Theory*, Adelaide, Australia, pages 1441–1445, September 2005.

[36] N. Mukhopadhyay, M. Ghosh, and P. K. Sen. *Sequential Estimation*. Wiley, New York, NY, 1997.

[37] J. W. Mark and T. D. Todd. A nonuniform sampling approach to data compression. *IEEE Transactions on Communications*, 29(1):24–32, 1981.

[38] F. Marvasti. *Nonuniform Sampling Theory and Practice*. Kluwer, New York, NY, 2001.

[39] R. H. Masland. The fundamental plan of the retina. *Nature Neuroscience*, 4(9):877–886, 2001.

[40] Y. Mei. Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks. *IEEE Transactions on Information Theory*, 54(5):2072–2089, 2008.

[41] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac. Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7):1497–1516, 2012.

[42] M. Miskowicz. Send-on-delta concept: An event-based data reporting strategy. *Sensors*, 6:49–63, 2006.

[43] M. Miskowicz. Asymptotic effectiveness of the event-based sampling according to the integral criterion. *Sensors*, 7:16–37, 2007.

[44] B. A. Moser and T. Natschlager. On stability of distance measures for event sequences induced by level-crossing sampling. *IEEE Transactions on Signal Processing*, 62(8):1987–1999, 2014.

[45] E. J. Msechu and G. B. Giannakis. Sensor-centric data reduction for estimation with WSNs via censoring and quantization. *IEEE Transactions on Signal Processing*, 60(1):400–414, 2012.

[46] F. De Paoli and F. Tisato. On the complementary nature of event-driven and time-driven models. *Control Engineering Practice*, 4(6):847–854, 1996.

[47] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer, New York, NY, 1994.

[48] A. Ribeiro and G. B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part II: Unknown probability density function. *IEEE Transactions on Signal Processing*, 54(7):2784–2796, 2006.

[49] J. C. Sanchez, J. C. Principe, T. Nishida, R. Bashirullah, J. G. Harris, and J. A. B. Fortes. Technology and signal processing for brain–machine interfaces. *IEEE Signal Processing Magazine*, 25(1):29–40, 2008.

[50] N. Sayiner, H. V. Sorensen, and T. R. Viswanathan. A level-crossing sampling scheme for a/d conversion. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 43(4):335–339, 1996.

[51] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo. Distributed estimation using reduced-dimensionality sensor observations. *IEEE Transactions on Signal Processing*, 55(8):4284–4299, 2007.

[52] I. D. Schizas, A. Ribeiro, and G. B. Giannakis. Consensus in ad hoc wsns with noisy links-part I: Distributed estimation of deterministic signals. *IEEE Transactions on Signal Processing*, 56(1):350–364, 2008.

[53] D. Siegmund. *Sequential Analysis, Tests and Confidence Intervals*. Springer, New York, NY, 1985.

[54] S. S. Stankovic, M. S. Stankovic, and D. M. Stipanovic. Decentralized parameter estimation by consensus based stochastic approximation. *IEEE Transactions on Automatic Control*, 56(3):531–543, 2011.

[55] Y. S. Suh. Send-on-delta sensor data transmission with a linear predictor. *Sensors*, 7(4):537–547, 2007.

[56] S. Sun. A survey of multi-view machine learning. *Neural Computing and Applications*, 23(7–8):2031–2038, 2013.

[57] R. R. Tenney and N. R. Sandell. Detection with distributed sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 17(4):501–510, 1981.

[58] S. C. A. Thomopoulos, R. Viswanathan, and D. C. Bougoulias. Optimal decision fusion in multiple sensor systems. *IEEE Transactions on Aerospace and Electronic Systems*, 23(5):644–653, 1987.

[59] J. Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals, and Systems*, 1(2):167–182, 1988.

[60] Y. Tsividis. Digital signal processing in continuous time: A possibility for avoiding aliasing and reducing quantization error. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, Montreal, Quebec, Canada, volume 2, pages II-589–II-592, May 2004.

[61] Y. Tsividis. Event-driven data acquisition and digital signal processing—A tutorial. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 57(8):577–581, 2010.

[62] V. V. Veeravalli, T. Basar, and H. V. Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Transactions on Information Theory*, 39(2):433–442, 1993.

[63] T. Vercauteren and X. Wang. Decentralized sigma-point information filters for target tracking in collaborative sensor networks. *IEEE Transactions on Signal Processing*, 53(8):2997–3009, 2005.

[64] C. Vezyrtzis and Y. Tsividis. Processing of signals using level-crossing sampling. In *IEEE International Symposium on Circuits and Systems (ISCAS 2009)*, Taipei, Taiwan, pages 2293–2296, May 2009.

[65] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–329, 1948.

[66] D. J. Warren and P. K. Willett. Optimal decentralized detection for conditionally independent sensors. In *American Control Conference*, Pittsburgh, PA, pages 1326–1329, June 1989.

[67] P. Willett, P. F. Swaszek, and R. S. Blum. The good, bad and ugly: Distributed detection of a known signal in dependent Gaussian noise. *IEEE Transactions on Signal Processing*, 48(12):3266–3279, 2000.

[68] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith. Linear coherent decentralized estimation. *IEEE Transactions on Signal Processing*, 56(2):757–770, 2008.

[69] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis. Distributed compression-estimation using wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):27–41, 2006.

[70] J. L. Yen. On nonuniform sampling of bandwidth-limited signals. *IRE Transactions on Circuit Theory*, 3(4):251–257, 1956.

[71] Y. Yılmaz, Z. Guo, and X. Wang. Sequential joint spectrum sensing and channel estimation for dynamic spectrum access. *IEEE Journal on Selected Areas in Communications*, 32(11):2000–2012, 2014.

[72] Y. Yılmaz, G. V. Moustakides, and X. Wang. Sequential and decentralized estimation of linear regression parameters in wireless sensor networks. *IEEE Transactions on Aerospace and Electronic Systems*, to be published. http://arxiv.org/abs/1301.5701.

[73] Y. Yılmaz, G. V. Moustakides, and X. Wang. Cooperative sequential spectrum sensing based on level-triggered sampling. *IEEE Transactions on Signal Processing*, 60(9):4509–4524, 2012.

[74] Y. Yılmaz, G. V. Moustakides, and X. Wang. Channel-aware decentralized detection via level-triggered sampling. *IEEE Transactions on Signal Processing*, 61(2):300–315, 2013.

[75] Y. Yılmaz and X. Wang. Sequential decentralized parameter estimation under randomly observed fisher information. *IEEE Transactions on Information Theory*, 60(2):1281–1300, 2014.

[76] Y. Yılmaz and X. Wang. Sequential distributed detection in energy-constrained wireless sensor networks. *IEEE Transactions on Signal Processing*, 62(12):3180–3193, 2014.

[77] K. A. Zaghloul and K. Boahen. Optic nerve signals in a neuromorphic chip I: Outer and inner retina models. *IEEE Transactions on Biomedical Engineering*, 51(4):657–666, 2004.

[78] T. Zhao and A. Nehorai. Distributed sequential Bayesian estimation of a diffusive source in wireless sensor networks. *IEEE Transactions on Signal Processing*, 55(4):1511–1524, 2007.