AN ENERGY-AWARE FRAMEWORK FOR CASCADED DETECTION ALGORITHMS

BY

DAVID MOSES JUN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Adviser:

Professor Douglas L. Jones

ABSTRACT

Low-power, scalable detection systems require aggressive techniques to achieve energy efficiency. Algorithmic methods that can reduce energy consumption by compromising performance are known as being energy-aware.

The cascade architecture is known for being energy-efficient, but without proper operation can end up being energy-inefficient in practice. In this thesis, we propose a framework that imposes energy-awareness on cascaded detection algorithms, which enforces proper operation of the cascade to maximize detection performance for a given energy budget. This is achieved by solving our proposed energy-constrained version of the Neyman-Pearson detection criterion, resulting in detector thresholds that can be updated to dynamically adjust to time-varying system resources and requirements.

Sufficient conditions for a global solution for a cascade of an arbitrary number of detectors are given. Explicit solutions are derived for a two-stage cascade. Applied to a canonical detection problem, simulations show that our energy-aware cascaded detectors outperform an energy-aware detection algorithm based on incremental refinement, an existing alternate approach to developing energy-aware algorithms. Combining our framework with incremental refinement reveals a promising approach to developing energy-aware energy-efficient detection systems.

To my parents, for their love and support

ACKNOWLEDGMENTS

I would like to express my gratitude to my adviser, Professor Douglas L. Jones, for his support and guidance. His interest in a wide array of problems has motivated me to think outside the box and approach seemingly unconventional problems that have a practical impact.

I would also like to thank Dr. Sriram Narayanan, who provided invaluable advice and mentorship when he was still a graduate student here at the University of Illinois. I would also like to thank Long Le, who implemented a real-time system demonstrating the work developed in this thesis.

This work was supported in part by the Multiscale Systems Center (MuSyC), one of six research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS vii

CHAPTER 1 INTRODUCTION 1
1.1 Motivation 1
1.2 Related Work 1
1.3 Our Contributions 3
1.4 Outline of Thesis 4

CHAPTER 2 PRELIMINARIES 5
2.1 Detection Theory Fundamentals 5
2.2 Cascade Architecture 9
2.3 Detection Performance of Cascaded Detectors 10

CHAPTER 3 OPTIMAL ENERGY-AWARE OPERATION OF CASCADED DETECTORS 12

CHAPTER 4 OPTIMIZATION 16
4.1 Proposed Energy Model for Cascaded Detectors 16
4.2 Log and Variable Transformation 18
4.3 Approximation: Rare-Event Detection 19
4.4 Convexity 20
4.5 Lagrangian Approach: A Geometric Interpretation 21
4.6 Special Case: $M = 2$ 26
4.7 Limitations and Discussion 29

CHAPTER 5 CASE STUDY: SINUSOID IN NOISE 31
5.1 Problem Statement 31
5.2 Cascade of Two Detectors 32
5.3 Baseline: Incremental FFT Detector 36
5.4 Simulation Setup 38
5.5 Optimization 39
5.6 Results 40

CHAPTER 6 CONCLUSION 43

APPENDIX A DERIVATION OF ENERGY DETECTOR STATISTICS 45
A.1 Mean and Variance of T_1 under H_0 45
A.2 Mean and Variance of T_1 under H_1 46

APPENDIX B DERIVATION OF FFT DETECTOR STATISTICS . 48
B.1 Mean and Variance of $X(k)$ under H_0 48
B.2 Mean and Variance of $X(k)$ under H_1 49
B.3 Noncentrality Parameter under H_1 50

REFERENCES 51

LIST OF ABBREVIATIONS

ROC	Receiver Operating Characteristic
EANP	Energy-Aware Neyman-Pearson
EABR	Energy-Aware Bayes Risk
FONC	First-Order Necessary Conditions
WGN	White Gaussian Noise
$\mathrm{SNR}_{\mathrm{in}}$	Input Signal-to-Noise Ratio
GLRT	Generalized Likelihood-Ratio Test
MLE	Maximum-Likelihood Estimate
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
CDF	Cumulative Distribution Function
CFAR	Constant False-Alarm Rate
PCG	Proper Complex Gaussian

CHAPTER 1

INTRODUCTION

1.1 Motivation

Detection applications in the low-power domain, such as sensor nodes, portable biomedical monitors, speech, vehicle and wildlife monitoring systems, are an important class of applications that need to be energy-efficient. Design of detection systems poses a particular challenge because the system must continuously monitor the environment. This warrants the use of aggressive power-reduction techniques such as approximate signal processing, which have been applied at the algorithmic level to reduce energy consumption by compromising task performance [1]. Algorithms with the ability to make this run-time energy/performance trade-off are known as being *energy-aware* [2]. The problem addressed in this paper is to increase the energy efficiency of detection systems through energy-aware algorithms.

1.2 Related Work

Efforts to develop energy-aware algorithms have gone into identifying and designing algorithms possessing the incremental refinement property [3]. The basic idea is to identify computations in the algorithm that can be terminated early to provide graceful degradation in task performance. This principle has been applied in many signal processing applications including signal detection

[4], FIR filtering [5], beamforming [2], and image processing [6].

Energy-efficient multimodal sensor nodes detecting the presence of vehicles have exploited the idea of "passive vigilence" [7] by implementing a tiered wake-up network [8]. On a single node, a cheap sensor, which is always running, is used to trigger more energy-intensive sensors to wake up and take measurements. This approach reduces energy consumption by limiting the amount of time an expensive sensor is actually taking measurements. The detection algorithm used on these sensor nodes is limited to a decision tree. The CART algorithm [8], which is used to construct the decision tree, is modified to consider the energy consumption of each sensor. A decision tree may be sufficient for simple detection tasks, but does not generalize to detection applications that require more sophisticated processing.

A particular form of the tiered wake-up network known as the *cascade architecture* is widely utilized in real-time detection applications. For example, the IEEE 802.22 standard, which is being developed for cognitive radio, envisions spectrum sensing as being based on a two-stage cascade architecture [9]. In [10], the result of a voice activity detector (implemented in hardware) triggered whether or not to turn on a microprocessor used for speech signal processing; this scheme effectively reduced the stand-by power of the microprocessor using a two-stage cascade. A similar strategy is used in [11] for a surveillance application, where the decision of a "wake-up" detector is used to arouse the sensor node to full functionality. Although these systems are energy-efficient, they are not energy-aware in that there is no systematic method to trade off performance for energy consumption.

The cascade architecture has also been successfully used in real-time object detection [12]. The Viola-Jones object-detection algorithm implements a cascade of a large number of weak classifiers. The purpose of the cascade

is to allow uninteresting regions of an image to be quickly discarded while spending more computation time on promising regions. Hence, the cascade serves to direct attention and processing power to regions of an image that are likely to be an object of interest. Although a lot of work in the past decade jointly considered energy consumption and system performance, most of these efforts have focused on cost-sensitive learning for the construction of the *classifiers* in the cascade (see [13] for a discussion and references). On the other hand, little attention has been given to the system design aspect of optimizing the final cascade performance [14]. Only very recently have we seen efforts to consider energy costs while optimizing the final cascade performance [13].

1.3 Our Contributions

The main contribution of this thesis is to develop a general framework to impose energy-awareness on cascaded detection algorithms. We formulate an energy-aware detection criterion such that solving the optimization problem results in the *energy-optimal operating point* of the cascaded detectors.

In order to demonstrate the efficacy of our framework, we work through a canonical detection application, designing and optimizing the operation of a cascade of detectors. Our resulting cascaded detectors are compared to an energy-aware detection algorithm based on the incremental refinement property [4].

1.4 Outline of Thesis

The remainder of this thesis is organized as follows: Chapter 2 reviews basic concepts and terminology in detection theory and defines the cascade architecture for detection applications. Chapter 3 motivates the need for energy-efficient detection and formally defines the problem, which is then solved in Chapter 4 for the general case of a cascade of M detectors. Concrete results are given for the specific case of $M = 2$, which are used in Chapter 5 to illustrate the efficacy of our approach through a canonical detection problem. Concluding remarks are given in Chapter 6.

CHAPTER 2

PRELIMINARIES

The likelihood ratio test, receiver operating characteristic (ROC) curve, and detection criteria are concepts from detection theory that are useful for the development of this thesis. In this chapter, we define the cascade architecture for detection, and derive its detection performance in terms of the true-detection and false-alarm rates.

2.1 Detection Theory Fundamentals

In this section, we provide a concise summary of the necessary concepts and notation from detection theory. For a more in-depth and complete study, the reader is encouraged to refer to [15].

The goal of detection theory is to provide a systematic framework for making an optimal decision between two competing hypotheses. The assumption is that there is a true correct decision, and noisy observations are available to aid in the decision making.

Detection theory addresses the issues of (1) optimal and thus (2) systematic detection. It establishes detection criteria which quantitatively measure the optimality of detection. It is shown that generally, the optimal decision rule takes the form of a likelihood ratio test (or tests if there are more than two competing hypotheses).

In this thesis, we confine ourselves to binary hypothesis testing. In other

words, there are two competing hypotheses from which we must decide: H_0 is known as the null hypothesis and corresponds to the absence of the signal of interest. H_1 corresponds to the hypothesis that the signal of interest is indeed present. The likelihood ratio test, then, is expressed as

$$T \triangleq \frac{\Pr[x|H_1 \text{ true}]}{\Pr[x|H_0 \text{ true}]} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{2.1}$$

where x is a single sample or collection of noisy observations. T is called the likelihood ratio and the test then is to compare T to some threshold τ such that if $T \geq \tau$, we decide that H_1 is true. If $T < \tau$, we decide that H_0 is true.

In order to actually determine T, we need an expression for the likelihood ratio. This is derived by assuming a signal and noise model under both hypotheses, which subsequently determines a conditional distribution on x, which is used to evaluate the likelihood of the observation.

There are two types of errors associated with binary hypothesis testing. The first is known as a missed detection, P_M, and corresponds to $T < \tau$ when in fact H_1 is true. The second error is known as a false alarm, P_{FA}, and corresponds to $T \geq \tau$ when in fact H_0 is true. Mathematically,

$$1 - P_M = 1 - \Pr\left[T < \tau | H_1 \text{ true}\right] = \Pr\left[T \geq \tau | H_1 \text{ true}\right] \tag{2.2}$$

$$P_{FA} = \Pr\left[T > \tau | H_0 \text{ true}\right] \tag{2.3}$$

where $1 - P_M$ is also known as the correct detection rate, P_D, and the probability measures are conditioned on a particular hypothesis being true.

Measuring these two errors for a particular detector completely characterizes the detector's performance, which is graphically expressed by the ROC curve, which plots P_D versus P_{FA} and is illustrated by the curves in Fig. 2.1.

There are a few observations worth noting about the ROC curve:

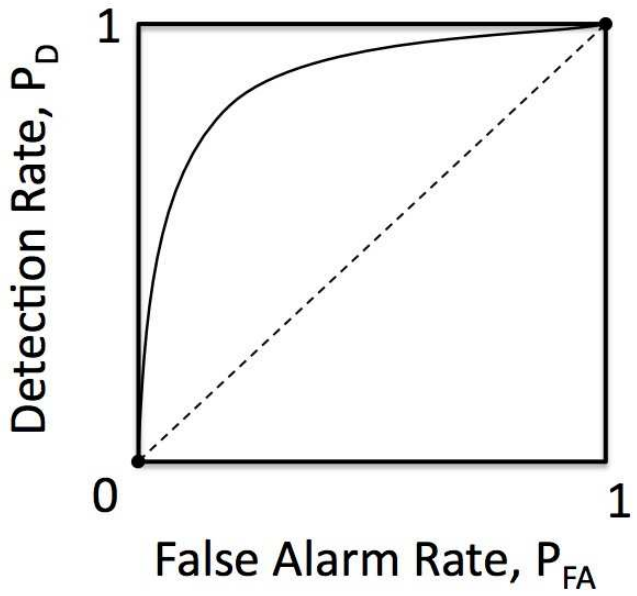


Figure 2.1: Example ROC curve.

1. The points (0,0) and (1,1) are always the endpoints of the curve. (1,1) means that $P_D = 1$ and $P_{FA} = 1$ and corresponds to setting $\tau = -\infty$ such that H_1 is always declared true regardless of the actual statistic T. Similarly, (0,0) corresponds to $\tau = \infty$.

2. There exists some threshold $\tau \in (-\infty, \infty)$ corresponding to each interior point of the ROC curve. The threshold to use for a particular application is determined by optimizing a detection criterion.

3. Assuming a different signal and noise model results in a different likelihood ratio, which results in a different ROC curve. A uniformly most powerful (UMP) likelihood ratio test has an ROC curve that lies "above" all other ROC curves.

4. All useful ROC curves lie above the curve $P_D = P_{FA}$, or the dashed line in Fig. 2.1. This dashed line corresponds to the detection performance

of flipping a (weighted) coin, which ignores all observations and thus represents the "worst" reasonable detector.

5. All ROC curves can be made concave. If they were not, randomization could be used to make them concave and therefore achieve better performance. For a proof, see [15].

There are two classical metrics used to determine the optimal operating point of a detector (see item (2) on page 7). The Neyman-Pearson criterion constrains the false-alarm rate and is used often in practice when no additional information such as costs and priors are available or relevant to the detection problem at hand. It can be posed as the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & P_D \\ \text{subject to} \quad & P_{FA} \leq \gamma \end{aligned}$$

where $\gamma \in [0, 1]$ is the false-alarm constraint. If costs and priors are available, the Bayes risk criterion can be used to systematically trade off between the two types of errors, $P_M = 1 - P_D$ and P_{FA}.

More rigorously, assume there exists a cost function C_{ij} with $0 \leq i, j \leq 1$, where C_{ij} represents the cost of deciding H_i when H_j is true. Furthermore, define the prior probability of hypothesis H_i being true as π_i for $i \in \{0, 1\}$. Then the average Bayes risk is given as

$$\begin{aligned} R &= \pi_0 \cdot (C_{00}\,(1 - P_{FA}) + C_{10} P_{FA}) + \pi_1 \cdot (C_{01} P_M + C_{11}(1 - P_M)) \\ &= \underbrace{\pi_0 C_{00} + \pi_1 C_{11}}_{\text{constant}} + \pi_0 C_0' P_{FA} + \pi_1 C_1' P_M \end{aligned} \tag{2.4}$$

where C_0' can be considered as the net cost incurred for making a false alarm,

and C_1' is the net cost incurred for missing a detection. In general, $C_0', C_1' \geq 0$ because it is more costly to make an error than it is to make the right decision. The Bayes risk criterion, then, is to minimize the average Bayes risk. From (2.4), this can be stated as

$$\text{minimize} \quad \pi_0 C_0' P_{FA} + \pi_1 C_1' P_M \tag{2.5}$$

which can be seen as determining the optimal trade-off between P_M and P_{FA}.

2.2 Cascade Architecture

A block diagram of the cascade architecture for detection algorithms is given in Fig. 2.2. Each detector represents a different likelihood ratio test. This architecture is motivated by the fact that we can have a range of simple to complex signal models, which will generally result in likelihood ratio tests with simple to complex computational complexity, and weak to powerful detection performance, respectively. Hence, we can obtain energy savings if we run an energy-efficient detector to monitor the environment, which then triggers more energy-intensive detectors only when an event of potential interest occurs.

This general strategy is known as "passive vigilance" [7] and is most beneficial when the probability of the event of interest occurring is low. We denote this probability measure as $\pi_1 \triangleq \Pr[H_1 \text{ true}]$, which is the prior probability of the signal being present; π_0 is defined as the prior probability of the signal being absent, and it follows that $\pi_0 = 1 - \pi_1$. We see from Fig. 2.2 that Detector 1 deciding H_1 triggers Detector 2 to make a possibly new observation and from that, a decision, and so on. If at any point in the cascade a

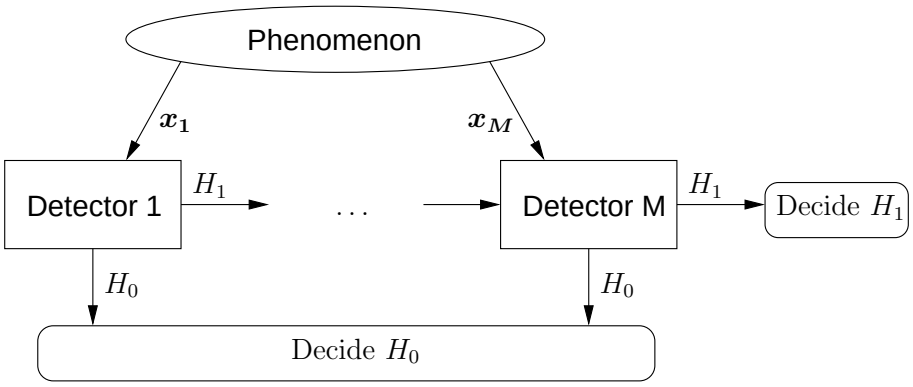


Figure 2.2: Block diagram of cascaded detectors.

detector decides H_0, the decision process ends and H_0 is declared.

In this abstract architecture, the structure of the observations $x_1, \ldots x_M$ is quite general. For example, they can be the same data (i.e. $x_1 = \cdots = x_M$), they can be data from different modalities from the same time frame (i.e. x_1 are audio samples and x_M are video frames), or they can be observations from different time frames. This last example, where detection is deferred over multiple sequences of observations, is reminiscent of (but not equivalent to) sequential hypothesis testing [16].

2.3 Detection Performance of Cascaded Detectors

As reviewed in Section 2.1, the detection performance of a detector is completely summarized by the false-alarm rate P_{FA} and true-detection rate P_D. We can extend the notion of the detection performance of the cascade as a *system*, by defining the system false-alarm rate P_{FA}^{sys}, and the system true-detection rate P_D^{sys}. As discussed in Section 2.2, in the cascade architecture, a final decision of H_1 is only made when all of the detectors in the cascade

choose H_1. Therefore,

$$P_{FA}^{sys} = \Pr\left[T_1 \geq \tau_1, \ldots, T_M \geq \tau_M | H_0 \text{ true}\right] \tag{2.6}$$

$$P_{D}^{sys} = \Pr\left[T_1 \geq \tau_1, \ldots, T_M \geq \tau_M | H_1 \text{ true}\right] \tag{2.7}$$

where the *system* detection performance is the joint probability measure of the summary statistics for all of the detectors being greater than the associated thresholds.

Using Bayes' rule, the system performance can be expressed as

$$P_{FA}^{sys} = \prod_{i=1}^{M} P_{FA_i}(\tau_i | \tau_1, \ldots, \tau_{i-1}) \tag{2.8}$$

$$P_{D}^{sys} = \prod_{i=1}^{M} P_{D_i}(\tau_i | \tau_1, \ldots, \tau_{i-1}) \tag{2.9}$$

where $P_{FA_i}(\tau_i|\tau_1, \ldots, \tau_{i-1}) \triangleq \Pr\left[T_i \geq \tau_i | T_1 \geq \tau_1, \ldots, T_{i-1} \geq \tau_{i-1}, H_0 \text{ true}\right]$, which is the *conditional* false-alarm rate of the i^{th} detector; $P_{D_i}(\tau_i|\tau_1, \ldots, \tau_{i-1}) \triangleq \Pr\left[T_i \geq \tau_i | T_1 \geq \tau_1, \ldots, T_{i-1} \geq \tau_{i-1}, H_1 \text{ true}\right]$. This dependence will hold if the observations for the detectors (i.e. $x_1, \ldots, x_M$) are conditionally correlated, which will generally be true if $x_1 = \cdots = x_M$. On the other hand, in a cascade with multimodal detectors, it is not unreasonable that conditioned on the hypothesis, the observation noise from detector to detector is independent.

CHAPTER 3

OPTIMAL ENERGY-AWARE OPERATION OF CASCADED DETECTORS

We propose a new criterion for optimal detection in order to account for energy consumption. We narrow down the general problem of designing an optimal cascade of detectors to one of operating a cascade in an optimal manner. We identify the thresholds in the likelihood ratio tests as being an appropriate optimization variable and state the resulting optimization problem to be solved.

At a high level, we would like to maximize the detection performance of the cascade while minimizing the energy consumed by the cascade. We formulate this problem as

$$\begin{aligned} \text{maximize} \quad & P_D^{\text{sys}} \\ \text{subject to} \quad & P_{FA}^{\text{sys}} \leq \gamma \\ & EC \leq \beta \end{aligned}$$

where EC is the energy consumption of the cascade and γ and β are specified constraints. We call this the Energy-Aware Neyman-Pearson (EANP) detection criterion for obvious reasons.

We could also adapt the Bayes risk criterion from (2.5) to be

$$\begin{aligned} \text{minimize} \quad & \pi_0 C_0' P_{FA}^{\text{sys}} + \pi_1 C_1' P_M^{\text{sys}} \\ \text{subject to} \quad & EC \leq \beta \end{aligned}$$

with π_i and C_i' being the prior probability and cost of error, respectively, for hypothesis H_i. We call this the Energy-Aware Bayes Risk (EABR) detection criterion.

Solving either of these problems in its most general form would require us to optimize over the form and the number of detectors M, the likelihood ratios $T_1, \ldots, T_M$, and the thresholds $\tau_1, \ldots, \tau_M$, which as pointed out in [14], is an extremely difficult problem. A suboptimal approach is to split this into two subproblems:

1. An optimization over the *design* of the cascade, which includes choosing the number of detectors M and the likelihood ratios $T_1, \ldots, T_M$.

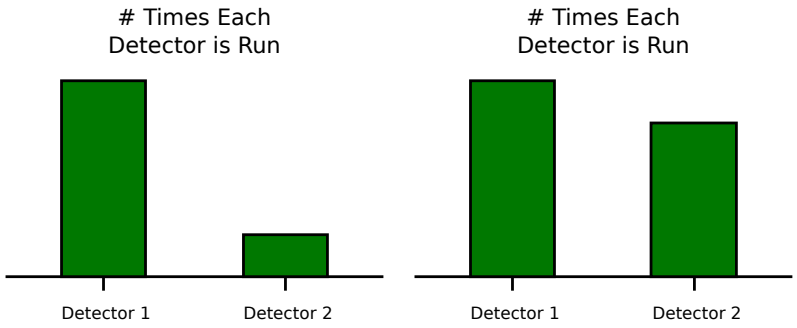
2. An optimization over the *operation* of the cascade, which includes choosing the appropriate thresholds $\tau_1, \ldots, \tau_M$ for each of the likelihood ratio tests.

The first problem is application-dependent and in many cases, improving detection algorithms is the main topic of many research communities. The second problem is important when the system is in actual operation and battery life or energy consumption becomes a fundamental constraint.

In this thesis, we focus on the second subproblem of optimizing the operation of the cascade; in particular, we consider energy-efficient operation, exploring the problem of making the optimal trade-off between energy consumption and system performance. This problem has been addressed in a limited sense in [14], where the author focused on real-time object detection using the Viola-Jones algorithm to solve the first subproblem, and presented an energy-agnostic method to determine the thresholds of the detectors in cascade. We are not aware of work that explicitly considers the energy consumption (i.e. the EANP or EABR detection criterion, or variants thereof)

of a general cascade of detectors.

To motivate why it is important to consider energy consumption in optimizing the *operation* of a cascade of detectors, consider the illustration in Fig. 3.1. Assume that for a particular detection problem, we found an



(a) Energy-efficient operation (b) Energy-inefficient operation

Figure 3.1: Illustrating two ways to operate a cascade of detectors.

energy-efficient design of a cascade of two detectors (i.e. solved the first subproblem). Ideally, we would like the first detector to trigger only when the signal is actually present. If this happens and the presence of the signal is rare, we only run the energy-intensive detector when it is absolutely necessary, as illustrated in Fig. 3.1(a). This leads to minimum energy consumption by the cascade while maintaining high detection performance.

On the other hand, if τ_1 is set very low, the first detector will exhibit behavior that can colloquially be described as "paranoid" or "trigger-happy", leading to unnecessary usage of the second detector, as illustrated in Fig. 3.1(b). This will lead to high energy consumption and negate the benefits of utilizing a cascade architecture for detection.

Hence, choosing the detector thresholds to optimally balance detection performance with energy consumption is an important step in ensuring energy-efficient operation of a cascade of detectors. For the remainder of this thesis,

we consider the following problem: given a cascade of M detectors,

$$\begin{aligned} \max_{\tau_1,\ldots,\tau_M} \quad & P_D^{sys}\left(\tau_1,\ldots,\tau_M\right) \\ \text{subject to} \quad & P_{FA}^{sys}\left(\tau_1,\ldots,\tau_M\right) \leq \gamma \\ & EC\left(\tau_1,\ldots,\tau_M\right) \leq \beta \end{aligned} \tag{3.1}$$

CHAPTER 4

OPTIMIZATION

In this chapter, we derive an expression for the energy-consumption constraint found in the EANP detection criterion and show how the detector thresholds can be used as an energy/performance "knob". Subsequently, we solve the EANP criterion to find the optimal trade-off between energy consumption and detection performance. We show convexity of the problem by variable transformation and appropriate problem relaxation. Further, we provide an intuitively pleasing geometric interpretation of the necessary conditions for optimality, from which the result in [14] follows as a special case. Finally, we provide sufficient conditions for a closed-form solution for the special but important case of a cascade of two detectors.

4.1 Proposed Energy Model for Cascaded Detectors

In this section, we develop a model of the energy consumed in the operation of a cascade of detectors. In particular, because limiting energy consumption will admittedly affect detection performance, we develop an energy model in terms of the detection performance.

As we are concerned with maximizing system lifetime, we are interested in the *average* energy consumed by the cascaded detectors. Denoting this as

EC,

$$EC = \sum_{i=1}^{M} \alpha_i \cdot c_i \tag{4.1}$$

where c_i is the energy consumed by running Detector i, and α_i is the associated activity factor. In other words, the average energy consumed by Detector i is the cost of running the detector, weighted by the probability of actually running it.

The energy consumed from running an algorithm can be physically measured [17] and is associated with the computational, storage, and communication requirements of the algorithm [1]. In our model, we assume that the energy consumed from running an algorithm is a known, fixed cost.

The underlying principle behind our energy-consumption model is that the number of times Detector i is run, and hence the activity factor α_i, is determined by the decisions made by the previous $i-1$ detectors, which is controlled by the detectors' thresholds. More succinctly,

$$\begin{aligned}
\alpha_i &= \Pr\left[\text{run Detector } i\right] = \Pr\left[T_1 \geq \tau_1, \ldots, T_{i-1} \geq \tau_{i-1}\right] \\
&= \sum_{k=0}^{1} \Pr\left[T_1 \geq \tau_1, \ldots, T_{i-1} \geq \tau_{i-1}, H_k \text{ true}\right] \\
&= \pi_0 \cdot \prod_{j=1}^{i-1} P_{FA_j}(\tau_j | \tau_1, \ldots, \tau_{j-1}) + \pi_1 \cdot \prod_{j=1}^{i-1} P_{D_j}(\tau_j | \tau_1, \ldots, \tau_{j-1})
\end{aligned} \tag{4.2}$$

where the first line follows from the cascade architecture and the second line from the law of total probability. The third line uses Bayes' rule to express the activity factor as a function of the hypothesis priors and the conditional detection performance of the detectors in the cascade.

In summary, the energy consumed by the cascaded detectors can be ex-

pressed as

$$EC(\tau_1, \ldots, \tau_{M-1}) =$$
$$c_1 + \pi_0 \cdot \sum_{i=2}^{M} c_i \cdot \prod_{j=1}^{i-1} P_{FA_j}(\tau_j|\tau_1, \ldots, \tau_{j-1}) + \pi_1 \cdot \sum_{i=2}^{M} c_i \prod_{j=1}^{i-1} P_{D_j}(\tau_j|\tau_1, \ldots, \tau_{j-1}) \tag{4.3}$$

where $\alpha_1 = 1$ because the first detector in the cascade is always run, and the energy consumption of the cascade is *not* a function of the threshold of the final detector.

4.2 Log and Variable Transformation

We can reformulate the optimization problem as a maximization over the detector false-alarm rates [14]. Doing so converts the problem into a resource-allocation problem, where the false-alarm rate can be thought of as the resource to be allocated.

Let us define a new notation for the detector performance: let $f_i = P_{FA_i}$ such that the system false-alarm rate is $f_{\text{sys}} = \prod_{i=1}^{M} f_i$. Defining $lf_i = \log(f_i)$, we get that $lf_{sys} = \log(f_{\text{sys}}) = \sum_{i=1}^{M} lf_i$. Similarly, we can define $lh_{\text{sys}} = \sum_{i=1}^{M} lh_i$, where lh_i as a function of $(lf_1, \ldots, lf_i)$ is simply the log-log ROC curve if the cascade were to end at Detector i.

Making this log and variable transformation simplifies the problem in two ways. (1) The log transformation makes the objective function and false-alarm constraint additive, and (2) solving the optimization problem in terms of the false-alarm rates instead of the thresholds works around the problem that it is not reasonable to assume that the true-detection rate is a convex (or concave) function of the thresholds. On the other hand, from what we

know about ROC curves, it is not so unreasonable to assume that the true-detection rate is a concave function of the false-alarm rate.

As a result, our optimization problem can be formulated as follows:

$$\begin{aligned} \max_{lf_1,\ldots,lf_M} \quad & \sum_{i=1}^{M} lh_i\left(lf_1,\ldots,lf_i\right) \\ \text{subject to} \quad & \sum_{i=1}^{M} lf_i \le \log(\gamma) \\ & EC\left(lf_1,\ldots,lf_{M-1}\right) \le \beta \\ & lf_i \le 0 \qquad i = 1,\ldots,M \end{aligned} \tag{4.4}$$

where the set of constraints represented by the third equation are added to ensure that the false-alarm rates do not exceed 1.

4.3 Approximation: Rare-Event Detection

The use of the cascade architecture is energy-efficient in rare-event detection applications. As an obvious counterexample, if the event of interest was always present, then in order for good detection performance, every detector in the cascade would always need to be run, voiding any potential energy savings; utilizing a cascade architecture in this scenario would not be appropriate.

If we indeed assume that the event of interest is very rare (i.e. $\pi_1 \approx 0$), we can approximate the energy consumption to be

$$\tilde{EC}(lf_1,\ldots,lf_{M-1}) \triangleq c_1 + \sum_{i=2}^{M} c_i \cdot \exp\left(\sum_{j=1}^{i-1} lf_j\right) \tag{4.5}$$

To give some intuition to (4.5), consider the case when $M = 3$:

$$\tilde{EC} = \underbrace{c_1}_{\text{energy consumed by Det. 1}} + \underbrace{c_2 e^{lf_1}}_{\text{energy consumed by Det. 2}} + \underbrace{c_3 e^{lf_1+lf_2}}_{\text{energy consumed by Det. 3}}$$

Each term in the summation corresponds to the actual energy consumed by a particular detector when the event of interest is not present (i.e. hypothesis H_0 is true). If the event never occurs, then it would not particularly matter how we allocate lf_1 and lf_2, so long as we stay within the energy budget. On the other hand, given that the event does occur, the true-detection rate is fundamentally linked to the false-alarm rate. We show in the next section that under suitable conditions, there is a unique allocation of the false-alarm rate that maximizes the true-detection rate, while satisfying the energy budget.

4.4 Convexity

Theorem 4.4.1 *If $lh_i(lf_1, \ldots, lf_i)$ is a concave function for all i, then*

$$\begin{aligned} \max_{lf_1,\ldots,lf_M} \quad & \sum_{i=1}^{M} lh_i(lf_1, \ldots, lf_i) \\ \textit{subject to} \quad & \sum_{i=1}^{M} lf_i \leq \log(\gamma) \\ & \tilde{EC}(lf_1, \ldots, lf_{M-1}) \leq \beta \\ & lf_i \leq 0 \qquad i = 1, \ldots, M \end{aligned} \tag{4.6}$$

is a convex optimization problem.

Proof Because the non-negative sum of concave functions is concave, $\sum lh_i$ is concave. The first constraint is linear in the variables we are optimizing

over, and therefore convex. The second constraint (4.5) is also convex because the exponential of a linear function is convex, and the non-negative linear combination of convex functions is convex. Finally, the M constraints $lf_i \leq 0$ are trivially linear and therefore convex. ∎

Hence, under the assumption that the log-log conditional ROC curves are all concave, we are guaranteed an operating point that achieves the global maximum, giving us the optimal cascade performance for the given energy consumption budget.

4.5 Lagrangian Approach: A Geometric Interpretation

For any $\lambda_0, \ldots, \lambda_M, \mu \in \mathbb{R}_+$, define the following:

$$L \triangleq lh_{\text{sys}} - \lambda_0 \left(\sum_{i=1}^{M} lf_i \right) - \mu \cdot \tilde{EC} - \sum_{i=1}^{M} \lambda_i \cdot lf_i \tag{4.7}$$

$$= lh_{\text{sys}} - \sum_{i=1}^{M} \lambda_i' \cdot lf_i - \mu \cdot \tilde{EC} \tag{4.8}$$

where $\lambda_i' = \lambda_0 + \lambda_i$. Here, L represents the trade-off between the objective function and a weighted combination of the constraints, where the weights are given by $\lambda_0, \ldots, \lambda_M$ and μ. Then, as proved in [18], for any $\lambda_1', \ldots, \lambda_M', \mu \in \mathbb{R}_+$, the solution $(lf_1^*, \ldots, lf_M^*)$ to the following unconstrained problem

$$\arg\max_{lf_i} \quad lh_{\text{sys}} - \sum_{i=1}^{M} \lambda_i' \cdot lf_i - \mu \cdot \tilde{EC} \tag{4.9}$$

is also the solution to the constrained problem (4.6) with $\log \gamma = \sum lf_i^*$ and $\beta = \tilde{EC}(lf_1^*, \ldots, lf_{M-1}^*)$. This transformation to the unconstrained optimization problem is known as the Lagrangian multiplier method [19]; L is

referred to as the Lagrangian, and $\lambda_1', \ldots, \lambda_M', \mu$ are the Lagrangian multipliers.

By the first-order necessary conditions (FONC) for optimality [19],

$$\nabla L = \begin{bmatrix} \frac{\partial lh_{\text{sys}}}{\partial lf_1} - \lambda_1' - \mu \cdot \frac{\partial \tilde{EC}}{\partial lf_1} \\ \vdots \\ \frac{\partial lh_{\text{sys}}}{\partial lf_M} - \lambda_M' \end{bmatrix} = 0 \tag{4.10}$$

Define $\tilde{EC}_k$ to be

$$\tilde{EC}_k(lf_1, \ldots, lf_{k-1}) \triangleq c_1 + \sum_{i=2}^{k} c_i \cdot \exp\left(\sum_{j=1}^{i-1} lf_j\right) \tag{4.11}$$

Note that $\tilde{EC}_1 = c_1$ and $\tilde{EC}_M = \tilde{EC}$. Hence, $\tilde{EC}_k$ represents the energy consumed up to and including Detector k. Then, the FONC can be equivalently stated as the following set of M equations:

$$\frac{\partial lh_{\text{sys}}}{\partial lf_i} = \lambda_i' + \mu \cdot \left(\beta - \tilde{EC}_i\right) \tag{4.12}$$

along with the conditions that $\sum lf_i = \log(\gamma)$ and $\tilde{EC} = \beta$. Observe that $\left(\beta - \tilde{EC}_k\right)$ is the energy consumed by subsequent detectors.

As an example, for $M = 3$,

$$\frac{\partial lh_{\text{sys}}}{\partial lf_1} = \lambda_1' + \mu(c_2 e^{lf_1} + c_3 e^{lf_1 + lf_2}) \tag{4.13}$$

$$\frac{\partial lh_{\text{sys}}}{\partial lf_2} = \lambda_2' + \mu c_3 e^{lf_1 + lf_2} \tag{4.14}$$

$$\frac{\partial lh_{\text{sys}}}{\partial lf_3} = \lambda_3' \tag{4.15}$$

along with the conditions that $lf_1 + lf_2 + lf_3 = \log(\gamma)$ and $c_1 + c_2 e^{lf_1} + c_3 e^{lf_1 + lf_2} = \beta$.

Geometrically, a necessary condition for optimal operation of the cascade is that for Detector i, the slope of the system log-log ROC curve is equal to λ_i offset by a term proportional to the energy that is spent by all subsequent detectors. If $lf_i < 0$ for all i, then the last set of constraints in (4.6) are not tight, and so by complementary slackness [19], $\lambda_i' = \lambda_0$ for all i. The condition that $lf_i = 0$ implies that the false-alarm rate for Detector i is set to be 1, or all observations are declared a detection. If all observations are passed onto the next detector, then this behavior is equivalent to skipping the detector and represents a form of degenerate operation, which may be induced by a poor design of the cascade or unreasonable system constraints.

Optimality conditions for Neyman-Pearson criterion

If we exclude degenerate operation of the cascade (i.e. $\lambda_i' = \lambda_0$) and consider the Neyman-Pearson criterion where the energy constraint is ignored (i.e. $\mu = 0$), then conditions (4.12) imply that at the optimal operating point, the slope of the ROC curves are equal. To interpret this in terms of resource allocation, this condition equivalently implies that the optimal allocation of the false-alarm rate "budget" is such that the weaker detectors are allocated a higher false-alarm rate, as illustrated in Fig. 4.1. Intuitively, this makes sense as weaker detectors are usually placed first in the cascade and act as a trigger for the better, but more expensive, detectors. Their ideal role then is to maintain high true detections while reducing the number of false alarms. Based on the properties of the ROC curve, for a weak detector, we can only maintain a high true-detection rate if we allow more false alarms. Note that this is the same result as that which was derived in [14] for the original Neyman-Pearson criterion (i.e. $\mu = 0$).

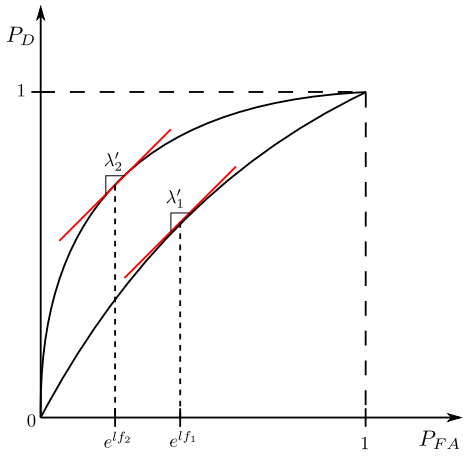


Figure 4.1: Geometric argument as to why a weaker detector is allocated a higher false-alarm rate.

Optimality conditions for EANP criterion

Now consider when $\mu > 0$. Then, the slopes of the early, weaker detectors are greater than the slopes of the subsequent detectors. This forces a reduction in the false-alarm rate that is allocated to the weaker detector. In fact, we see that the greater the energy required by subsequent detectors, the smaller the allocation that is given to the early detectors. Again, this is intuitively pleasing because if an energy budget is reduced, then the early detectors will be forced to be more frugal with the detections they declare; since we are assuming rare-event detection, a small false-alarm rate translates into energy-frugal behavior.

Optimality conditions for independent detectors

Consider the special case when the detector statistics $\{T_i\}_{i=1}^M$ are independent, conditioned on the hypothesis. That is,

$$\begin{aligned} P_{FA}^{\text{sys}} &= \Pr\left[T_1 > \tau_1, \ldots, T_M > \tau_M | H_0 \text{ true}\right] \\ &= \prod_{i=1}^{M} \Pr\left[T_i > \tau_i | H_0 \text{ true}\right] \\ P_{D}^{\text{sys}} &= \Pr\left[T_1 > \tau_1, \ldots, T_M > \tau_M | H_1 \text{ true}\right] \\ &= \prod_{i=1}^{M} \Pr\left[T_i > \tau_i | H_1 \text{ true}\right] \end{aligned}$$

Then,

$$lh_{\text{sys}}(lf_1, \ldots, lf_M) = \sum_{i=1}^{M} lh_i(lf_i) \tag{4.16}$$

where the conditional true-detection rate lh_i is a function of lf_i only, implying that $lh_i(lf_i)$ is actually the *unconditional* log-log ROC curve. A practical application where this independence assumption is reasonable is in multimodal detection (i.e. the M detectors process data observed from M different sensing modalities). Then, from (4.16), it follows that (4.12) simplifies to

$$\frac{\partial lh_i}{\partial lf_i} = \lambda_0 + \mu \cdot \left(\beta - \tilde{EC}_i\right) \tag{4.17}$$

where the left-hand side is the slope measured on the *unconditional* ROC curve of Detector i.

If we consider the Neyman-Pearson criterion (i.e. $\mu = 0$), (4.17) simplifies further and decouples into M independent conditions; the geometric interpretation for optimality is that each detector is operating on its uncondi-

tional ROC curve such that the slopes are equal, determined by the Lagrange multiplier, λ_0. The practical implication is that the optimization problem, originally in a simplex of dimension M, is reduced to M one-dimensional problems.

If we consider the EANP criterion, we see that even when the performance of one detector does not affect another, the energy consumption couples the detectors together. Hence, the false-alarm rate budget cannot be optimally allocated without jointly considering the energy consumption of all of the detectors.

4.6 Special Case: $M = 2$

For a cascade of two detectors, the constraint set in (4.6) defines the following feasible set:

$$\begin{aligned} lf_1 + lf_2 &\leq \log \gamma \\ c_1 + c_2 \cdot e^{lf_1} &\leq \beta \end{aligned}$$

Rearranging the equations leads to the following conditions for feasibility:

$$lf_1 \leq \log\left(\frac{\beta - c_1}{c_2}\right) = \log \beta' \tag{4.18}$$

$$lf_2 \leq \log \gamma - lf_1 \tag{4.19}$$

where β' is a constant defined to be $(\beta - c_1)/c_2$. These constraints, along with the constraints that $lf_1 \leq 0$ and $lf_2 \leq 0$, define the feasible region (shown in Fig. 4.2). Note that in this particular illustration, $lf_1 \leq 0$ is an inactive constraint.

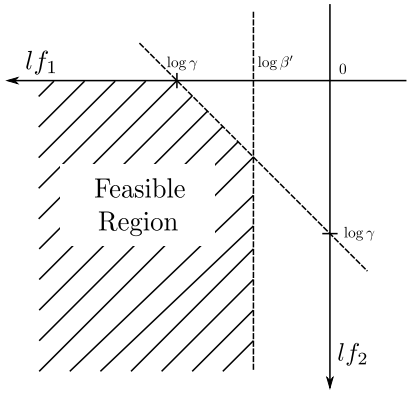


Figure 4.2: Feasible region given by constraints (4.18) and (4.19).

Lemma 4.6.1 *For a fixed lf_1, the lf_2 that maximizes (4.6) will be on the boundary of the feasible set.*

Proof If lf_1 is fixed, the objective function, given as $lh(lf_1, lf_2) = lh_1(lf_1) + lh_2(lf_1, lf_2)$, is constant in the first term; the second term is defined to be the conditional ROC curve of the second detector. With lf_1 fixed, the distribution of the second detector's test statistics is fixed. Hence, in accordance with the properties of ROC curves, lh_2 is a monotonically nondecreasing function of lf_2. Assume for lf_1 fixed that the optimal lf_2, denoted as lf_2^*, is in the interior of the feasible set. Then, there exists a $\delta > 0$ such that $lf_2^* + \delta$ is still feasible. But $lh_2(lf_2^* + \delta) \geq lh_2(lf_2^*)$ by the monotonicity property of lh_2. If they are equal, then the optimal solution is not unique. If $lh_2(lf_2^* + \delta) > lh_2(lf_2^*)$, then lf_2^* cannot be optimal. In either case, it follows that the solution is on the boundary. ∎

As a result of Lemma 4.6.1, we see from Fig. 4.2 that if $lf_1^* = \log\beta'$, then $lf_2^* = \log(\gamma) - \log\beta' = \log\frac{\gamma}{\beta'}$, which is at the intersection of the two active constraints. We denote this as a Type I solution.

Lemma 4.6.2 *If $\log\beta' < \log\gamma$, then a Type I solution does not exist and the constraint given in (4.19) is inactive. In this case, the optimal false-alarm*

rate for the second detector is given by $lf_2^* = 0$.

Proof If $\log\beta' < \log\gamma$, the intersection of the constraints (4.18) and (4.19) is not in the feasible region. Hence, a Type I solution is impossible. From Fig. 4.2, it is clear that if $\log\beta' < \log\gamma$, the constraint (4.19) is inactive. It follows directly from Lemma 4.6.1 that $lf_2^* = 0$ (i.e. the false-alarm rate of the second detector is 1). This situation is denoted as a Type II solution and arises under an extremely stringent energy-budget β. ∎

Theorem 4.6.3 *Assuming* $\log\beta' > \log\gamma$ *(i.e. no Type II solution), the global solution to the optimization problem in (4.6) is given by* $(lf_1^*, lf_2^*) = \left(\beta', \log\frac{\gamma}{\beta'}\right)$ *if the following two conditions hold:*

1. *the log-log conditional ROC curves are concave*
2. $\partial lh_{sys}/\partial lf_1 > \partial lh_{sys}/\partial lf_2 \geq 0$, *where the partial derivatives are evaluated at* (lf_1^*, lf_2^*)

Proof The a-superlevel set of lh_{sys} is defined as

$$C_a = \{(lf_1, lf_2) \mid lh_{\text{sys}}(lf_1, lf_2) \geq a\}$$

By the first condition, it follows that for all a, C_a is a convex set [20]. Define $\mathbf{lf} = [lf_1, lf_2]^T$ and $\mathbf{lf}^* = [\beta', \log\gamma/\beta']^T$. Then, by convexity of C_a, $C_{lh_{\text{sys}}(\mathbf{lf}^*)}$ is supported by the tangent hyperplane of lh_{sys}, which is given by

$$\nabla lh_{\text{sys}}^T (\mathbf{lf} - \mathbf{lf}^*) = 0 \tag{4.20}$$

where $\nabla lh_{\text{sys}}^T = [\partial lh_{\text{sys}}/\partial lf_1, \partial lh_{\text{sys}}/\partial lf_2]$ is the transpose of the gradient of lh_{sys}, evaluated at $\mathbf{lf}^*$. Expanding (4.20), the tangent line can be expressed

as

$$lf_2 = -\frac{\partial lh_{\text{sys}}/\partial lf_1}{\partial lh_{\text{sys}}/\partial lf_2} \cdot lf_1 + c$$

where c is some constant. From the second condition, it follows that the magnitude of the slope of the tangent line is greater than the magnitude of the slope of constraint (4.19). Hence, the only feasible point in $C_{lh_{\text{sys}}(\mathbf{lf}^*)}$ is $\mathbf{lf}^*$. Furthermore, because the feasible set minus $\mathbf{lf}^*$ is not in $C_{lh_{\text{sys}}(\mathbf{lf}^*)}$, lf^* achieves the global solution to (4.6). ■

Comment: If the gradient condition does not hold, then the optimal solution will lie on the boundary of condition (4.19), and condition (4.18) will be inactive. We denote this as a Type III solution.

4.7 Limitations and Discussion

There are a few observations and outstanding issues that need to be addressed. First, it is not immediately obvious whether or not Theorem 4.4.1 extends to the problem when the exact energy consumption expression is used. Regardless, if required, the relaxed problem can be solved first in hopes that the solution will be in the neighborhood of the true solution, from which a local exact method can be used to converge to the true solution. Second, log-log conditional ROC curves are not always concave; more general conditions under which this holds have not been fully investigated. Third, in the general case where the detectors are not independent, the Lagrangian first-order conditions represent a set of M coupled equations with M unknowns. If the gradient can be computed efficiently, there exist many methods that can be used to solve this problem efficiently. This leads us to

the final issue, which is that in practice, the gradient of the log-log conditional ROC curves may be hard to compute efficiently. A practical algorithm for the particular application of object detection using a cascade designed using the Viola-Jones algorithm is given in [14]. This algorithm can be directly modified to use the energy-adjusted slopes derived here.

CHAPTER 5

CASE STUDY: SINUSOID IN NOISE

In this chapter, we consider the application of detecting a complex sinusoid of unknown frequency and unknown phase in complex white Gaussian noise (WGN). This problem is relevant in many real-world applications, such as detecting communication signals (e.g. spectrum sensing in a cognitive-radio application), detecting the presence of vehicles or wildlife in acoustic applications, or fault monitoring in structures and machines. A cascade of two detectors was developed and its energy-aware operation optimized in [21]. We provide a complete explanation using the theory developed in this thesis.

5.1 Problem Statement

The problem is formulated as deciding between two alternative hypotheses:

$$\begin{aligned} H_0 &: x[n] = w[n] \qquad n = 0, 1, \ldots, N-1 \\ H_1 &: x[n] = s[n] + w[n] \quad n = 0, 1, \ldots, N-1 \end{aligned} \tag{5.1}$$

using N discrete observation samples, where

$$w[n] = w_I[n] + j w_Q[n] \tag{5.2}$$

$$s[n] = \sqrt{E} e^{j(2\pi f_0 n + \phi)} \tag{5.3}$$

and $w_I[n]$ and $w_Q[n]$ are both real-valued WGN processes with variance $N_0/4$, $f_0 = k_0/N$ is the unknown normalized frequency with frequency index $k_0 \in \{0, 1, \ldots, N-1\}$, and ϕ is the unknown phase with $0 \leq \phi \leq 2\pi$.

5.2 Cascade of Two Detectors

To detect between H_0 and H_1, we cascade the energy detector followed by the FFT detector, which is optimal [22]. The energy detector is a suboptimal detector for this problem as it does not take advantage of the sinusoidal signal model, but its advantage is its low complexity, resulting in low energy consumption.

5.2.1 Energy detector

The summary statistic for the energy detector is the sum of the received signal energy,

$$T_1 \triangleq \sum_{n=0}^{N-1} |x[n]|^2 \tag{5.4}$$

The decision rule then is given as

$$T_1 \underset{H_0}{\overset{H_1}{\gtrless}} \tau_1 \tag{5.5}$$

where we decide that a sinusoid is present (i.e. choose H_1) if T_1 is greater than some threshold τ_1. Otherwise, we choose H_0.

By the central limit theorem [22], for large N, the summary statistic T_1 is approximately Gaussian. Under the assumption that H_0 is true, the mean

and variance of T_1 are given as [22]

$$\mathbb{E}[T_1|H_0] = N\left(\frac{N_0}{2}\right) \tag{5.6}$$

$$\text{var}(T_1|H_0) = N\left(\frac{N_0}{2}\right)^2 \tag{5.7}$$

Under the assumption that H_1 is true, we have

$$\mathbb{E}[T_1|H_1] = N\left(\frac{N_0}{2} + E\right) \tag{5.8}$$

$$\text{var}(T_1|H_1) = N\left(\frac{N_0}{2} + 2E\right)\left(\frac{N_0}{2}\right) \tag{5.9}$$

with derivations for (5.6) – (5.9) provided in Appendix A.

Hence, the detection performance, characterized by the false-alarm rate P_{FA_1} and detection rate P_{D_1}, can be approximately given in terms of the Q-function:

$$P_{FA_1}(\tau) = \Pr\left[T_1 \geq \tau|H_0\right] = Q\left(\frac{\tau - \mathbb{E}[T_1|H_0]}{\sqrt{\text{var}(T_1|H_0)}}\right) \tag{5.10}$$

$$P_{D_1}(\tau) = \Pr\left[T_1 \geq \tau|H_1\right] = Q\left(\frac{\tau - \mathbb{E}[T_1|H_1]}{\sqrt{\text{var}(T_1|H_1)}}\right) \tag{5.11}$$

5.2.2 FFT detector

As the amplitude, phase, and frequency are unknown parameters in the sinusoidal model, we adopt the generalized likelihood-ratio test (GLRT) which consists of replacing the unknown parameters with the maximum-likelihood estimates (MLE) [22]. Defining the discrete Fourier transform (DFT) of the

signal at frequency index $k \in \{0, \ldots, N-1\}$ to be

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \tag{5.12}$$

the summary statistic is equivalent to taking the maximum of the squared magnitude of the DFT over all frequency indices. The decision rule is expressed as

$$T_2 \triangleq \max_k \left\{ |X(k)|^2 \right\} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_2 \tag{5.13}$$

where hypothesis H_1 is chosen if the maximum of the DFT is greater than some threshold τ_2, and H_0 is chosen otherwise. The DFT can be computed efficiently using the fast Fourier transform (FFT), and so we call this GLRT the FFT detector.

From (5.12), as $X(k)$ is a linear combination of independent Gaussian random variables, the real and imaginary components of $X(k)$, defined as $X_I(k)$ and $X_Q(k)$, respectively, are independent Gaussian random variables. For all k under H_0 and for $k \neq k_0$ under H_1,

$$X_I(k), X_Q(k) \sim \mathcal{N}\left(0, N \cdot \frac{N_0}{4}\right) \tag{5.14}$$

For $k = k_0$ under H_1, $X_I(k)$ and $X_Q(k)$ are also Gaussian, but mean-shifted to account for the signal energy in the frequency bin.

If we define $C(k) \triangleq |X(k)|^2 = X_I(k)^2 + X_Q(k)^2$, then $C(k)$ is chi-square distributed with two degrees of freedom. For all k under H_0 and for $k \neq k_0$ under H_1,

$$\frac{4}{N \cdot N_0} C(k) \sim \chi^2 \left(\kappa = 2\right) \tag{5.15}$$

where κ is the degree of freedom. For $k = k_0$ under H_1,

$$\frac{4}{N \cdot N_0} C(k) \sim \chi^2 \left(\kappa = 2, \lambda = \frac{4NE}{N_0} \right) \tag{5.16}$$

with λ defined as the noncentrality parameter. Derivations for (5.14) – (5.16) are provided in Appendix B.

The false-alarm rate of the FFT detector can be derived as follows:

$$\begin{aligned} P_{FA_2}(\tau_2) &= \Pr\left[\max_k \{C(k)\} > \tau_2 | H_0 \text{ true} \right] \\ &= 1 - \Pr\left[C(0) < \tau_2, \ldots, C(N-1) < \tau_2 | H_0 \text{ true} \right] \\ &= 1 - \prod_{k=0}^{N-1} \Pr\left[C(k) < \tau_2 | H_0 \text{ true} \right] \\ &= 1 - \left(1 - \exp\left(-\frac{2\tau_2}{N \cdot N_0} \right) \right)^N \end{aligned} \tag{5.17}$$

where $1 - e^{-x/2}$ is the cumulative distribution function (CDF) of a normalized chi-square distribution with two degrees of freedom. The CDF of a normalized noncentral chi-square distribution with two degrees of freedom and parameter λ is given as [4]

$$P(x, \lambda) = 1 - \mathcal{Q}\left(\sqrt{\lambda}, \sqrt{x} \right) \tag{5.18}$$

where $\mathcal{Q}$ is Marcum's Q function. The true-detection rate is derived similarly

and is given as

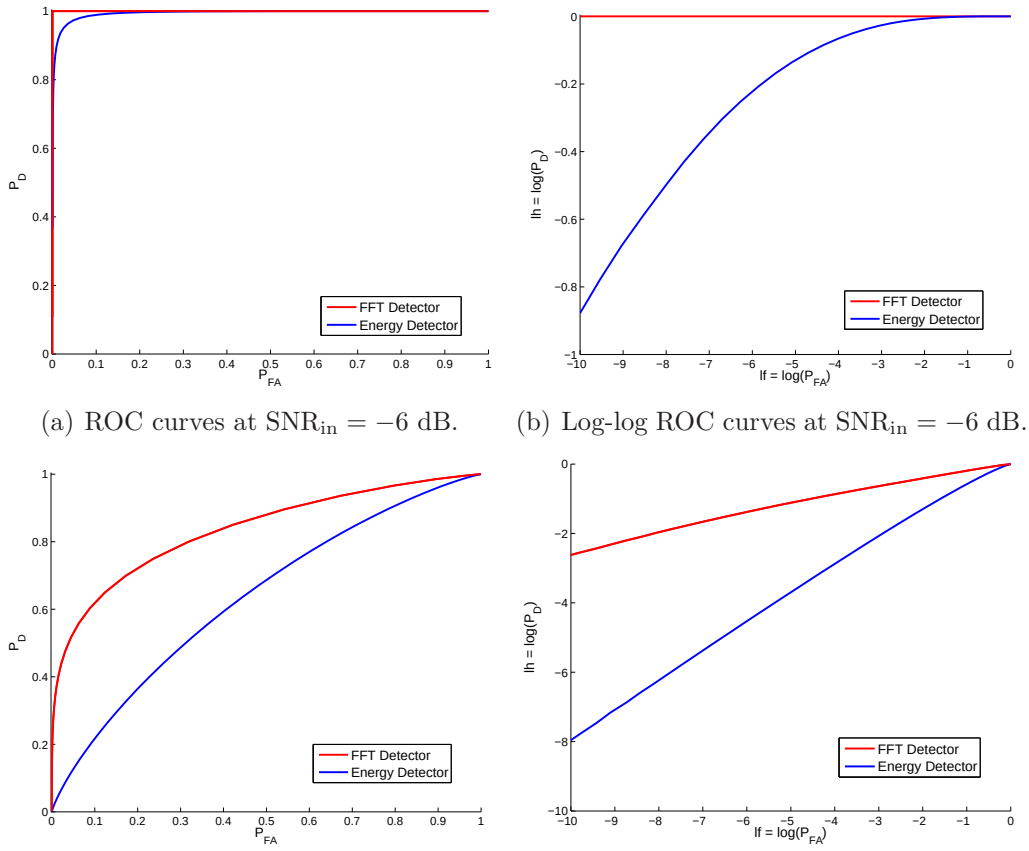
$$
\begin{aligned}
P_{D_2}(\tau_2) &= \Pr\left[\max_k \{C(k)\} > \tau_2 | H_1 \text{ true}\right] \\
&= 1 - \Pr\left[C(0) < \tau_2, \ldots, C(N-1) < \tau_2 | H_1 \text{ true}\right] \\
&= 1 - \Pr\left[C(k_0) < \tau_2 | H_1 \text{ true}\right] \times \prod_{k \neq l} \Pr\left[C(k) < \tau_2 | H_1 \text{ true}\right] \\
&= 1 - \left(1 - \mathcal{Q}\left(\sqrt{\frac{4NE}{N_0}}, \sqrt{\frac{4\tau_2}{N \cdot N_0}}\right)\right) \times \left(1 - \exp\left(-\frac{2\tau_2}{N \cdot N_0}\right)\right)^{N-1}
\end{aligned}
\tag{5.19}
$$

Figure 5.1 compares the detection performance between the energy detector and FFT detector at two different input SNRs when the block size is $N = 256$. There are a few comments to be made.

- For each detector, we plot (1) the traditional ROC curves (see Fig. 5.1(a) and 5.1(c)), and (2) the log-log ROC curves (see Fig. 5.1(b) and 5.1(d)).
- In general, for all input SNRs, the FFT detector is more powerful than the energy detector. In other words, for any false-alarm rate, the FFT detector achieves a higher true-detection rate.
- For $\text{SNR}_{\text{in}} = -6$ dB, the FFT detector attains near-perfect performance (see Fig. 5.1(a) or 5.1(b)), whereas at -15 dB, we see a trade-off between true detections and false alarms (see Fig. 5.1(c) or 5.1(d)).

5.3 Baseline: Incremental FFT Detector

In [4], the FFT detector was made energy-aware by identifying an incremental refinement property in the computation of a radix-2 FFT. Because the FFT



(a) ROC curves at $\mathrm{SNR_{in}} = -6$ dB. (b) Log-log ROC curves at $\mathrm{SNR_{in}} = -6$ dB.

(c) ROC curves at $\mathrm{SNR_{in}} = -15$ dB. (d) Log-log ROC curves at $\mathrm{SNR_{in}} = -15$ dB.

Figure 5.1: Detection performance comparison for the energy detector and FFT detector at different input SNRs with block size $N = 256$.

is computed in stages, we can stop after any stage and use (5.13) to make a decision about the hypothesis, as shown in Fig. 5.2. We define the n-*incremental detector* to be the FFT detector terminated after the n^{th} FFT stage. The $(\log_2 N)$-detector is equivalent to computing the full FFT (5.12).

Clearly, ending prematurely saves on computations but results in a degradation in detection performance. The performance analysis of the incremental detectors is similar to (5.17) and (5.19) and can be found in [4]. Figure 5.3 shows the true-detection rate as a function of the termination stage n, at $\mathrm{SNR_{in}} = -6$ dB, for a fixed false-alarm rate of $\gamma = 10^{-4}$ and an input block size of $N = 256$.

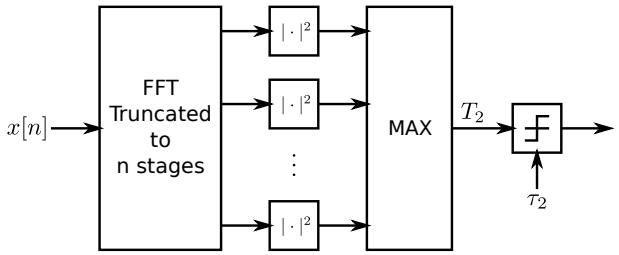


Figure 5.2: Block diagram of the n-incremental FFT detector (adapted from [4]).

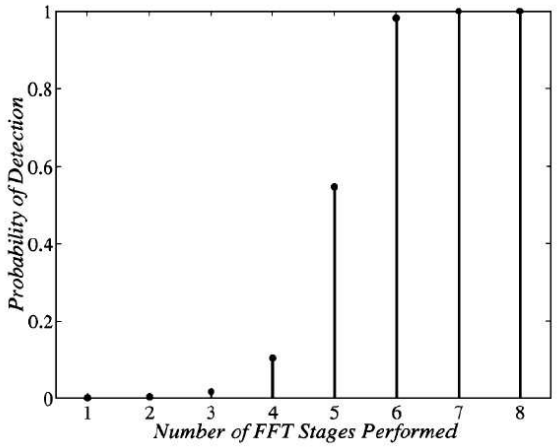


Figure 5.3: Detection performance of the energy-aware n-incremental FFT detector at $\text{SNR}_{\text{in}} = -6$ dB, for $\gamma = 10^{-4}$ and $N = 256$ (taken from [4]).

5.4 Simulation Setup

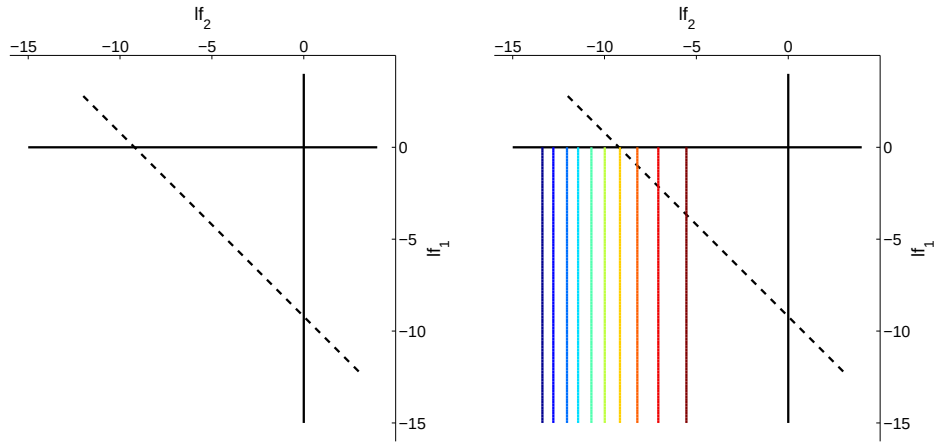
We present results using the same parameters as used in [4]. In particular, the signal power and noise variance parameters, E and N_0, are determined by assuming that $\text{SNR}_{\text{in}} = 10 \cdot \log_{10} \frac{E}{N0/2} = -6$ dB; the false-alarm constraint is set at $\gamma = 10^{-4}$, and the input block size is $N = 256$.

For our energy-consumption model, we choose typical hypothesis priors and fixed energy costs for the detectors. As discussed earlier, the cascade architecture is most beneficial in applications where the event to be detected does not occur often. Therefore, we assume priors of $\pi_1 = 10^{-1}$ and $\pi_0 = 1 - \pi_1$. We choose the energy costs $c_1 = 2N$ and $c_2 = 5N \log_2 N$ based only on the computational cost associated with each algorithm, which is proportional

to the number of MAC instructions used to form the summary statistics.

5.5 Optimization

Given the parameters in the previous section, the false-alarm constraint and domain of the objective function are shown in Fig. 5.4(a). The energy



(a) Problem domain for simulation setup. (b) Contours of the objective function.

Figure 5.4: Problem domain and objective function contours, estimated from Monte Carlo simulation.

constraint is a function of the energy budget β, which ranges from $2N$ to $N(2+5\log_2 N)$, or 512 to 10752 for our simulation setup. From Lemma 4.6.2, if $\beta < c_1 + \gamma \cdot c_2$, the optimal solution will be a Type II solution. As this represents a very stringent energy budget, in order to avoid these degenerate solutions, we only consider $\beta \in [c_1 + \gamma \cdot c_2, c_1 + c_2] = [513.024, 10752]$.

Figure 5.4(b) shows the contours of the objective function lh_{sys}, which is the true-detection rate of the cascade. The contours are nearly vertical lines (i.e. nearly not a function of lf_2) because as seen in Fig. 5.1(b), the FFT detector achieves near-perfect detection, regardless of the false-alarm rate. As such, $\partial lh_{\text{sys}}/\partial lf_1 > lh_{\text{sys}}/\partial lf_2$ at any feasible point and therefore from Theorem 4.6.3, the optimal solution to the problem (4.6) must be $(lf_1^*, lf_2^*) =$

$(\log \beta', \log(\gamma/\beta')$, where $\beta' = \frac{\beta - c_1}{c_2}$, the Type I solution.

In order to actually use this result to operate the cascade, we must map the optimal false-alarm rates to corresponding thresholds:

$$\exp(l f_1^*) = \beta' = \Pr(T_1 > \tau_1 | H_0 \text{ true})$$
$$\exp(l f_2^*) = \frac{\gamma}{\beta'} = \Pr(T_2 > \tau_2 | T_1 > \tau_1, H_0 \text{ true})$$

From (5.10), we can find τ_1 analytically:

$$\tau_1 = Q^{-1}\left(\frac{\beta' - \mathbb{E}[T_1|H_0]}{\sqrt{\text{var}(T_1|H_0)}}\right)$$

where Q^{-1} is the inverse Q-function. In general, the conditional false-alarm rate for the second detector will depend on τ_1 and is difficult to compute analytically. A constant false-alarm rate (CFAR) method can be used to adjust the threshold τ_2 until the desired false-alarm rate is met. This strategy will operate the cascade such that the optimal detection performance is achieved for the given energy-consumption constraint.

5.6 Results

Figure 5.5 shows the optimal detection performance for varying energy constraints. The dashed stems correspond to the detection performance of the eight different incremental detectors. In particular, the n-incremental detector has an energy cost of $5nN$, as computations are terminated after the n^{th} FFT stage. The solid black curve corresponds to the performance of our proposed energy-aware cascaded detector.

Although the incremental refinement principle transforms an algorithm to be energy-aware, Fig. 5.5 illustrates that for detection applications, exploit-

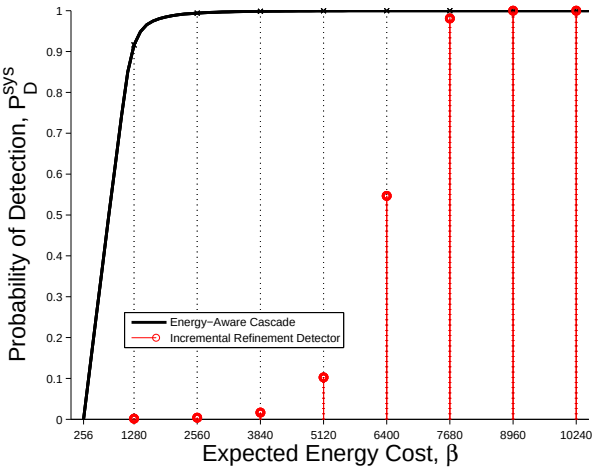


Figure 5.5: Detection probabilities at varying energy constraints with $\text{SNR}_{\text{in}} = -6$ dB, $N = 256$.

ing the fact that the system must be continuously monitoring the environment results in the optimized energy-aware cascade architecture providing a much better energy/performance trade-off. More importantly, this figure also demonstrates the performance scalability of the cascade with varying energy requirements. Unlike heuristic methods which would set the energy threshold arbitrarily, our proposed method gives a systematic method to meet energy constraints, and provides much finer granularity of control over the energy/performance trade-off.

Figure 5.6 shows the energy/performance trade-off when using the n-incremental detector as the second detector in our cascade, where $n = 6, 7, 8$. As the figure illustrates, for $\text{SNR}_{\text{in}} = -6$ dB and $N = 256$, the best detector to use depends on the particular β. The 6-incremental detector should be used until $\beta \approx 7N$, and the 7-incremental detector should be used for greater energy constraints. Surprisingly, we observe that the full 8-incremental detector will not be used in practice under these SNR and false-alarm rate conditions. This illustrates how our energy-aware cascaded detectors can be used together with incremental refinement methods, resulting in even more

energy-efficient systems.

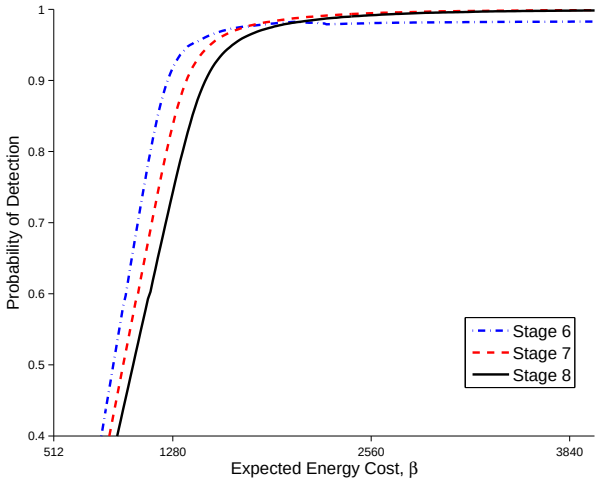


Figure 5.6: Detection probabilities at varying energy constraints using the 6-,7-,8-incremental detector as the second detector in the cascade.

CHAPTER 6

CONCLUSION

The cascade architecture for detection algorithms is an effective framework for designing energy-efficient detection applications. Controlling the thresholds of the detectors in the cascade is an efficient strategy to control the average energy consumed while operating the cascaded detectors. The derived analytic expressions for the energy consumed by the cascaded detectors can be used to maximize detection performance for a specified energy consumption constraint, which is crucial in applications where battery life is a fundamental constraint.

Assuming a rare-event scenario (which is relevant to most applications where a cascade of detectors would be used) simplifies the energy consumption model. This results in a convex optimization problem such that any local solution is also a global solution, which is the energy-optimal operating point of the cascaded detectors.

For a two-stage cascade, the generalized CFAR method we propose is a practical solution to systematically determine the energy-optimal operating point and to dynamically adjust the cascade's operation to time-varying system-level energy requirements. Simulations for detecting a sinusoid in white Gaussian noise show that imposing energy-awareness on cascaded detection algorithms greatly outperforms incremental refinement, an alternate method to design energy-aware detection algorithms.

For a cascade of more than two detectors with conditionally correlated

test statistics, we have derived necessary conditions for optimality. Practical, simple CFAR-like methods to determine this energy-optimal operating point have yet to be identified. If the detector test statistics are independent, conditioned on the hypothesis, then the problem greatly simplifies and practical, efficient methods can be developed to determine the energy-optimal operating point of the cascade. This conditional independence assumption is reasonable in multi-modal detection, which represents a wide class of applications.

APPENDIX A

DERIVATION OF ENERGY DETECTOR STATISTICS

A.1 Mean and Variance of T_1 under H_0

Under the null hypothesis H_0, $x[n] = w_n \sim c\mathcal{N}\left(0, \frac{N_0}{2}\right)$, where $w_n = w_{nI} + jw_{nQ}$ is proper complex Gaussian (PCG), we derive derive the mean and variance of the test statistic T_1 of an energy detector.

A.1.1 Mean

$$\mathbb{E}\left[T_1 \middle| H_0\right] = \mathbb{E}\left[\sum_{n=0}^{N-1} |x_n|^2\right] = \sum_{n=0}^{N-1} \mathbb{E}\left[|w_n|^2\right] = \sum_{n=0}^{N-1} \frac{N_0}{2} = N \cdot \frac{N_0}{2}$$

A.1.2 Variance

$$\begin{aligned}
\mathbb{E}\left[T_1^2 \middle| H_0\right] &= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} |x_n|^2\right)\left(\sum_{m=0}^{N-1} |x_m|^2\right)\right] = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}\left[|w_n|^2 \cdot |w_m|^2\right] \\
&= \sum_{n=0}^{N-1} \mathbb{E}\left[|w_n|^4\right] + \sum_{n=0}^{N-1}\sum_{m\neq n} \mathbb{E}\left[|w_n|^2\right] \cdot \mathbb{E}\left[|w_m|^2\right] \\
&= \sum_{n=0}^{N-1} E\left[w_{nI}^4 + w_{nQ}^4 + 2w_{nI}^2 \cdot w_{nQ}^2\right] + \sum_{n=0}^{N-1}\sum_{m\neq n} \left(\frac{N_0}{2}\right)^2 \\
&= \sum_{n=0}^{N-1} 6\left(\frac{N_0}{4}\right)^2 + 2\left(\frac{N_0}{4}\right)^2 + N(N-1)\left(\frac{N_0}{2}\right)^2 \\
&= 2N\left(\frac{N_0}{2}\right)^2 + N(N-1)\left(\frac{N_0}{2}\right)^2 = N\left(\frac{N_0}{2}\right)^2 + N^2\left(\frac{N_0}{2}\right)^2
\end{aligned}$$

where we used the fact that if $Y \sim \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}[Y^4] = 3\sigma^4$. So,

$$\text{var}(T_1|H_0) = \mathbb{E}\left[T_1^2 \middle| H_0\right] - \mathbb{E}\left[T_1 \middle| H_0\right]^2 = N \cdot \left(\frac{N_0}{2}\right)^2$$

A.2 Mean and Variance of T_1 under H_1

Under hypothesis H_1, $x_n = s_n + w_n = \sqrt{E}e^{j(2\pi f_0 n + \phi)} + w_n$, where $w_n \sim c\mathcal{N}(0, \frac{N_0}{2})$ is PCG.

A.2.1 Mean

$$\begin{aligned}
\mathbb{E}\left[T_1|H_1\right] &= \sum_{n=0}^{N-1} \mathbb{E}\left[|s_n + w_n|^2\right] \\
&= \sum_{n=0}^{N-1} \mathbb{E}\left[E + 2\sqrt{E} \cdot \text{Re}\{e^{j(2\pi f_0 n + \phi)} \cdot w_n^*\} + |w_n|^2\right] \\
&= N \cdot E + 2\sqrt{E} \sum_{n=0}^{N-1} \mathbb{E}\left[w_{nI} \cdot \cos(2\pi f_0 n + \phi) + w_{nQ} \cdot \sin(2\pi f_0 n + \phi)\right] \\
&\quad + N \cdot \frac{N_0}{2} \\
&= N \cdot \left(\frac{N_0}{2} + E\right)
\end{aligned}$$

A.2.2 Variance

$$\begin{aligned}
\mathbb{E}\left[T_1^2|H_1\right] &= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\mathbb{E}\left[\left(E+2\sqrt{E}\cdot\mathrm{Re}\{e^{j(2\pi f_0 n+\phi)}\cdot w_n^*\}+|w_n|^2\right)\cdot\right. \\
&\qquad\left.\left(E+2\sqrt{E}\cdot\mathrm{Re}\{e^{j(2\pi f_0 m+\phi)}\cdot w_m^*\}+|w_m|^2\right)\right] \\
&= \sum_n\sum_m E^2+2E\frac{N_0}{2}+\mathbb{E}\left[|w_n|^2\cdot|w_m|^2+\right. \\
&\qquad\left.4E\cdot\mathrm{Re}\{e^{j(2\pi f_0 n+\phi)}\}\cdot\mathrm{Re}\{e^{j(2\pi f_0 m+\phi)}\}\right] \\
&= N^2\cdot\left(E^2+2E\frac{N_0}{2}\right)+\mathbb{E}\left[T_1^2|H_0\right]+4E\sum_n\cos^2(2\pi f_0 n)\mathbb{E}\left[w_{nI}^2\right]+ \\
&\qquad\sin^2(2\pi f_0 n)\mathbb{E}\left[w_{nQ}^2\right] \\
&= N^2\cdot\left(E^2+2E\frac{N_0}{2}\right)+\mathbb{E}\left[T_1^2|H_0\right]+4EN\frac{N_0}{4}
\end{aligned}$$

So,

$$\begin{aligned}
\mathrm{var}(T_1|H_1) &= \mathbb{E}\left[T_1^2|H_1\right]-\mathbb{E}\left[T_1|H_1\right]^2 \\
&= \mathbb{E}\left[T_1^2|H_1\right]-N^2\cdot\left[\left(\frac{N_0}{2}\right)^2+E^2+2E\frac{N_0}{2}\right] \\
&= \mathbb{E}\left[T_1^2|H_0\right]+ENN_0-N^2\left(\frac{N_0}{2}\right)^2 \\
&= N\left(\frac{N_0}{2}\right)^2+N^2\left(\frac{N_0}{2}\right)^2+ENN_0-N^2\left(\frac{N_0}{2}\right)^2 \\
&= N\left(\frac{N_0}{2}\right)^2+ENN_0 \\
&= N\left(\frac{N_0}{2}+2E\right)\cdot\frac{N_0}{2}
\end{aligned}$$

APPENDIX B

DERIVATION OF FFT DETECTOR STATISTICS

Define $\mathbf{f_k} = \left[1, e^{-j2\pi k/N}, \ldots, e^{-j2\pi k(N-1)/N}\right]^\dagger$, where $\dagger$ is the Hermitian transpose operator. It can be shown that $||\mathbf{f_k}||^2 = \mathbf{f_k}^\dagger \mathbf{f_k} = N$. With this notation, the DFT can be succinctly represented as an inner product: $X(k) = \mathbf{f_k}^\dagger \mathbf{x}$, where $\mathbf{x} = [x(0), \ldots, x(N-1)]^T$ is the time-series stacked into a vector of size N.

B.1 Mean and Variance of $X(k)$ under H_0

Under H_0, $\mathbf{x} = \mathbf{w} = [w_0, \ldots, w_{N-1}]^T$, where $\mathbf{w}$ is PCG with variance $\frac{N_0}{2}\mathbf{I}$.

B.1.1 Mean

$$\mathbb{E}\left[X(k)|H_0\right] = \mathbb{E}\left[\mathbf{f_k}^\dagger \mathbf{w}\right] = 0$$

B.1.2 Variance

$$\begin{aligned}\mathbb{E}\left[|X(k)|^2 \mid H_0\right] &= \mathbb{E}\left[\mathbf{f_k}^\dagger \mathbf{w}\mathbf{w}^\dagger \mathbf{f_k}\right] = \mathbf{f_k}^\dagger \mathbb{E}\left[\mathbf{w}\mathbf{w}^\dagger\right] \mathbf{f_k} = \mathbf{f_k}^\dagger \frac{N_0}{2}\mathbf{I}\mathbf{f_k} = \frac{N_0}{2}\mathbf{f_k}^\dagger \mathbf{f_k} \\ &= N \cdot \frac{N_0}{2}\end{aligned}$$

Because the PCG property is closed under addition and the components

of $\mathbf{w}$ are PCG, $X(k)$ is PCG. As a result:

1. $X_I(k)$ and $X_Q(k)$, the real and imaginary components of $X(k)$ are independent.

2. $\mathbb{E}[X_I(k)^2] = \mathbb{E}[X_Q(k)^2] = 1/2 \cdot \mathbb{E}[X(k)^2] = N \cdot \frac{N_0}{4}$.

Thus, $C(k) \triangleq |X(k)|^2 = X_I(k)^2 + X_Q(k)^2$ is chi-square with two degrees of freedom. Normalizing by the variance of X_I and X_Q, $\frac{4}{N \cdot N_0} C(k) \sim \chi^2$.

B.2 Mean and Variance of $X(k)$ under H_1

Under H_1, let $k = k_0$. For all other integer k, the value of the FFT of the sinusoid is 0. The statistics in these bins will be the same as those under H_0.

Define $\mathbf{s} = [s_0, \ldots, s_{N-1}]^T$ such that $\mathbf{x} = \mathbf{s} + \mathbf{w}$. Then,

$$X(k_0) = \mathbf{f_k}^\dagger \mathbf{s} + \mathbf{f_k}^\dagger \mathbf{w}$$

Because we assume $\mathbf{s}$ is deterministic, $X(k_0)$ is then just a mean-shifted proper complex Gaussian random variable with variance $N \cdot \frac{N_0}{2}$, as derived in the previous section. To compute the mean shift:

$$\mathbf{f_k}^\dagger \mathbf{s} = \mathbf{f_k}^\dagger \sqrt{E} \cdot \begin{bmatrix} e^{j\phi} \\ e^{j(2\pi k_0/N + \phi)} \\ \vdots \\ e^{j(2\pi k_0 (N-1)/N + \phi)} \end{bmatrix} = \sqrt{E} e^{j\phi} \cdot \mathbf{f_k}^\dagger \begin{bmatrix} 1 \\ e^{j2\pi k_0/N} \\ \vdots \\ e^{j2\pi k_0 (N-1)/N} \end{bmatrix}$$

$$= \sqrt{E} e^{j\phi} \cdot \mathbf{f_k}^\dagger \mathbf{f_k} = N\sqrt{E} e^{j\phi} = \underbrace{N\sqrt{E}\cos(\phi)}_{\mathbb{E}[X_I(k_0)|H_1]} + j \cdot \underbrace{N\sqrt{E}\sin(\phi)}_{\mathbb{E}[X_Q(k_0)|H_1]}$$

B.3 Noncentrality Parameter under H_1

$$\begin{aligned}\lambda &\triangleq \frac{\mathbb{E}\left[X_I(k_0)|H_1\right]^2}{\mathrm{var}\left(X_I(k_0)^2|H_1\right)} + \frac{\mathbb{E}\left[X_Q(k_0)|H_1\right]^2}{\mathrm{var}\left(X_Q(k_0)^2|H_1\right)} \\ &= \frac{4}{N \cdot N_0}\left[\left(N\sqrt{E}\cos(\phi)\right)^2 + \left(N\sqrt{E}\cos(\phi)\right)^2\right] \\ &= \frac{4NE}{N_0}\end{aligned}$$

So under H_1, $\frac{4}{N \cdot N_0} C(k_0) \sim \chi^2(\lambda)$, with noncentrality parameter $\lambda = 4NE/N_0$.

REFERENCES

[1] L. Benini and G. De Micheli, "System-level power optimization: techniques and tools," in *ACM Trans. Design Automat. Embed. Syst.*, vol. 5, April 2000, pp. 115–192.

[2] A. Sinha, A. Wang, and A. Chandrakasan, "Energy scalable system design," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 10, no. 2, pp. 135–145, April 2002.

[3] S. Nawab, A. Oppenheim, A. Chandrakasan, J. Winograd, and J. Ludwig, "Approximate signal processing," *J. VLSI Signal Process. Syst.*, vol. 15, no. 1/2, pp. 177–200, 1997.

[4] J. Winograd, S. Nawab, and A. Oppenheim, "FFT-based incremental refinement of suboptimal detection," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, May 1996, pp. 2479–2482.

[5] M. Goel and N. Shanbhag, "Dynamic algorithm transformations (DAT)-a systematic approach to low-power reconfigurable signal processing," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 7, no. 4, pp. 463–476, Dec. 1999.

[6] Y. Andreopoulos and I. Patras, "Incremental refinement of image salient-point detection," *IEEE Trans. on Image Processing*, vol. 17, no. 9, pp. 1685–1699, Sept. 2008.

[7] P. Dutta, M. Grimmer, A. Arora, S. Bibyk, and D. Culler, "Design of a wireless sensor network platform for detecting rare, random, and ephemeral events," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, 2005, p. 70.

[8] A. Benbasat and J. Paradiso, "A framework for the automated generation of power-efficient classifiers for embedded sensor nodes," in *SenSys '07: Proceedings of the 5th Int. Conference on Embedded Networked Sensor Systems*, 2007, pp. 219–232.

[9] C. Cordeiro, K. Challapali, D. Birru, and N. S. Shankar, "IEEE 802.22: An introduction to the first wireless standard based on cognitive radios," *Journal of Communications*, vol. 1, no. 1, Apr. 2006.

[10] H. Noguchi, T. Takagi, M. Yoshimoto, and H. Kawaguchi, "An ultra-low-power VAD hardware implementation for intelligent ubiquitous sensor networks," in *IEEE Workshop on Signal Processing Systems*, Oct. 2009, pp. 214–219.

[11] D. H. Goldberg, A. G. Andreou, P. Julián, P. O. Pouliquen, L. Riddle, and R. Rosasco, "VLSI implementation of an energy-aware wake-up detector for an acoustic surveillance sensor network," *ACM Trans. Sen. Netw.*, vol. 2, no. 4, pp. 594–611, 2006.

[12] P. Viola and M. Jones, "Robust real-time object detection," *Int. Journal of Computer Vision*, vol. 57, pp. 137–154, 2001.

[13] V. C. Raykar, B. Krishnapuram, and S. Yu, "Designing efficient cascaded classifiers: Tradeoff between accuracy and cost," in *Proceedings of the 16th ACM SIGKDD*. ACM, 2010, pp. 853–860.

[14] H. Luo, "Optimization design of cascaded classifiers," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 480–485.

[15] B. Levy, *Principles of Signal Detection and Parameter Estimation*. New York, NY: Springer, 2008.

[16] A. Wald, *Sequential Analysis*. New York, NY: J. Wiley & Sons, 1947.

[17] A. Sinha and A. Chandrakasan, "Energy aware software," in *Thirteenth Int. Conf. on VLSI Design*, 2000, pp. 50–55.

[18] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," in *Operation Res.*, vol. 11, 1963, pp. 399–417.

[19] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, Sept. 1999.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, March 2004.

[21] D. M. Jun and D. L. Jones, "An energy-aware framework for cascaded detection algorithms," in *IEEE Workshop on Signal Processing Systems*, Oct. 2010, pp. 1–6.

[22] S. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.