UNIVERSIDAD CARLOS III DE MADRID
Departamento de Teoría de la Señal y Comunicaciones

DOCTORAL THESIS

NEYMAN-PEARSON DETECTION IN SENSOR NETWORKS WITH DEPENDENT OBSERVATIONS

Author: JORGE PLATA CHAVES
Supervised by: MARCELINO LÁZARO
ANTONIO ARTÉS RODRÍGUEZ

June 2012

Tesis Doctoral: NEYMAN-PEARSON DETECTION IN SENSOR NETWORKS WITH DEPENDENT OBSERVATIONS

Autor: Jorge Plata Chaves

Directores: Dr. Marcelino Lázaro
Dr. Antonio Artés Rodríguez

Fecha:

Tribunal

Presidente: Dr. Ignacio Santamaría Caballero

Vocal 1: Dr. Kostas Berberidis

Vocal 2: Dr. Albert Guillén i Fàbregas

Vocal 3: Dr. Baltasar Beferull Lozano

Secretario: Dr. Joaquín Míguez Arenas

Agradecimientos y reflexiones

" El trabajo puede con todo."
Virgilio
"Nada es imposible si crees en quien eres y en quien siempre has sido."
Mongrell Cernuda

Es raro . . . cuando empiezas a escribir esta parte hay una mezcla enorme de sentimientos. Uno de ellos es el de nostalgia. Nostalgia por todos los años que he pasado en la universidad, por cada uno de los recuerdos de todo tipo que he vivido desde que comencé la carrera hasta ahora que estoy acabando la tesis. Creedme estos recuerdos son de "todo tipo". Es inevitable pensar que se queda atrás una etapa que no sólo me ha enseñado conocimientos técnicos. Además de esto, me ha dado algo muy importante, ha definido parte de lo que actualmente soy en el presente. Es aquí donde aparece el sentimiento de felicidad que me invade ahora mismo. Digo que es aquí porque precisamente este sentimiento resume lo que actualmente soy. Puedo decir que soy feliz porque miro mi pasado, mi presente y mi futuro, y la verdad, no puedo pedir nada, absolutamente nada a este mundo. No sé si ésta es la correcta definición de felicidad, pero si sé que el decir esto es el mejor agradecimiento que en unas pocas líneas puedo dar a la gente que me quiere y ha cuidado de mí. No obstante, sin dejar de agradecer a los que no nombre, siento que debo dedicar unas líneas a aquellas personas cuyo apoyo ha sido imprescindible a la hora de poder llevar a cabo esta tesis.

Como no podía ser de otra forma quiero agradecer a mi familia el apoyo incondicional que me han dado desde que conocí este mundo. Escribo esta frase de agradecimiento y me doy cuenta de lo pequeña que es en comparación con todo lo que debo a mis padres, Antonio y Pascuala, y a mi hermano Agustín. También en esta familia te incluyo a ti Sultán!!! Donde quiera que estés has sido y siempre serás uno más. Es absurdo que diga más palabras sobre vosotros, simplemente os quiero!!

Ahora quiero agradecer a las otras dos personas que más de cerca han vivido el trabajo de esta tesis, es decir, mis dos directores de tesis, Antonio y Marce. Eliminando los formalismos, quiero agradeceros vuestras explicaciones y vuestro buen hacer en cada una de las reuniones que hemos mantenido. Sin embargo, quiero hacer especial hincapié en algo que para mí ha sido muy importante. Esto es la enorme confianza que los dos habéis

depositado en mí, no sólo desde el inicio de esta tesis, sino también desde el comienzo de mi proyecto fin de carrera. Soy muy consciente de esa confianza y el enorme número de fabulosas oportunidades que me habéis brindado desde el inicio de mi proyecto fin de carrera. Profesionalmente hablando, habéis sido capaces de crear un investigador apasionado que disfruta con su trabajo. Para ello sólo necesitasteis a un joven alumno con un pequeño interés en aquellas cosas que enseñan en la carrera sobre el tratamiento y procesado de la señal. Eso resume lo bien que lo habéis hecho. Personalmente hablando, tenéis un gran amigo y os considero dos grandes amigos.

Y ahora toca hablar de un largo número de amigos. Para hacerlo rápido voy hacer la siguiente lista en la que todas las categorías están englobadas bajo el mejor concepto de amigo ...

- **Primos**. Gracias a Jesús, Elena, Alicia, Eva, Marisa, Pedro, Mari Jose, Carmelo, Paqui y Javi. Aquí permitir que de un gracias especial a mi primo Xule por las maravillosas fiestas y aventuras que hemos vivido y nos seguiremos pegando juntos!!.

- **Amigos de infancia**. Aquí no puede ser de otra forma ... tendré que usar los motes. Gracias a Coper, Alikate, Panadero, Luis Mi, Xule 2, Claudio, Pedro (Mr. Cachas) y Tania. En esta ocasión, por esos partidos de fútbol y esas escapadas regionales por España un gracias especial a Xule 2 y a Pedro, mi mejor amigo junto con mi primo Xule.

- **Amigos en general (universidad, fútbol etc.)**. Aquí la lista es enorme. Aún así quiero agradecer a todos los que recuerde porque aunque sean muchos no por ello son menos importantes. De hecho algunos de ellos son muy importantes. En particular quiero agradecer a Blanca, Fran, Sandra, Raul, Luca, Camilo, Marina, Isabel, Rafa, Pablo, Soufi, Efraín, Luis, Alberth, Darwin, Aurora, Wilton, Siana, David Luengo, Angel Bravo, Jesse, Niamh, Cristina, Concha, Katrin, Jair, Edu, Paloma, Miguel, Eugenia, Jesus, Bernd, Marta, Ricardo el Gallego, Ana, Alfredo, Fernandinho, Ramón, David Delgado, Gustavo, Julián y un largo etc. Entre estas personas debo un gracias súper especial a:

  - Siana. Eres simplemente la mejor "novocia". Nos quedan y deseo que sigamos teniendo muchas aventuras juntos.

- Luca. Eres un grande. La gente con tu espontaneidad me hace reír todos los días y hace que el trabajo ni si quiera parezca trabajo. Viva tu jaleo y tus bailes trabajando.
- Pablo. Por ser ese amigo con el que uno sabe que siempre se puede contar. Amigos como tú todo el mundo necesita.
- Blanca. Por enseñarme que se pueden tener amigas y por esas maravillosas conversaciones juntos.
- Isa y Fran. Hasta en los agradecimientos aparecéis juntos. Gracias por el viaje juntos que nos hemos hecho a Granada así como los momentos Sotana en los jueves noche. Isa, madre mía lo que tuviste que aguantar en el inolvidable viaje a Dinamarca.
- Camilo. Entre muchas otras cosas gracias por las charlas del "Information Theory Center" que siempre hacemos los viernes por la tarde. Espero poder visitarte algún día en Colombia.
- Wilton. A tí solo te diré ... "el vínculo".
- Niamh. Gracias por esas fiestas de San Patricio y descubrir el irlandés.

Por último, después de todo esto y sólo por si no ha quedado claro el mensaje ... GRACIAS!! MUCHAS GRACIAS A TODOS!!

Abstract

In this thesis, within the context of sensor networks, we are interested in the distributed detection problem under the Neyman-Pearson formulation and conditionally dependent sensor observations. In order to exploit all the detection potential of the network, the literature on this issue has faced optimal distributed detection problems, where optimality usually consists in properly designing the parameters of the network with the aim of minimizing some cost function related to the overall detection performance of the network. However, this problem of optimization has usually constraints regarding the possible physical and design parameters that we can choose when maximizing the detection performance of the network. In many applications, some physical and design parameters, for instance the network architecture or the local processing scheme of the sensor observations, are either strongly constrained to a set of possible design alternatives or either cannot be design variables in our problem of optimization. Despite the fact that those parameters can be related to the overall performance of the network, the previous constraints might be imposed by factors such as the environment where the network has to be deployed, the energy budget of the system or the processing capabilities of the available sensors. Consequently, it is necessary to characterize optimal decentralized detection systems with various architectures, different observation processes and different local processing schemes.

The mayor part of the works addressing the characterization of distributed detection systems have assumed settings where, under each one of the possible states of our phenomenon of interest, the observations are independent across the sensors. However, there are many practical scenarios where the conditional independence assumption is violated because of the presence of different spatial correlation sources. In spite of this, very few works have faced the aforementioned characterizations under the same variety of settings as under the conditional independence assumption. Actually, when the strategy of the network is not an optimization parameter, under the assumption of conditionally dependent observations the existing literature has only obtained asymptotic characterizations of the detection performance associated with parallel networks whose local processing rules are based on amplify-and-relay schemes.

Motivated by this last fact, in this thesis, under the Neyman-Pearson formulation, we undertake the characterization of distributed detection systems with dependent observations, various network architectures and binary quantization rules at the sensors. In particular, considering a parallel network randomly deployed along a straight line, we derive a closed-form error exponent for the Neyman-Pearson fusion of Markov local decisions when the involved fusion center only knows the distribution of the sensor spacings. After studying some analytical properties of the derived error exponent, we carry out evaluations of the closed-form expression in order to assess which kind of trends of detection performance can appear with increasing dependency and under two well-known models of the sensor spacing. These models are equispaced sensors with failures and exponentially spaced sensors with failures. Later, the previous results are extended to a two-dimensional parallel network that, formed by a set of local detectors equally spaced on a rectangular lattice, performs a Neyman-Pearson test discriminating between two different two-dimensional Markov causal fields defined on a binary state space. Next, under conditionally dependent observations and under the Neyman-Pearson set up, this thesis dissertation focuses on the characterization of the detection performance of optimal tandem networks with binary communications between the fusion units. We do so by deriving conditions under which, in an optimal tandem network with an arbitrary constraint on the overall probability of false alarm, the probability of misdetection of the system, i.e. at the last fusion node of the network, converges to zero as the number of fusion stages approaches infinity. Finally, after extending this result under the Bayesian set up, we provide two examples where these conditions are applied in order to characterize the detection performance of the network. From these examples we illustrate different dependence scenarios where an optimal tandem network can or cannot achieve asymptotic perfect detection under either the Bayesian set up or the Neyman-Pearson formulation.

Resumen

En esta tesis, dentro del contexto de las redes de sensores, estamos interesados en el problema de detección distribuida bajo la formulación de Neyman-Pearson y observaciones condicionalmente dependientes. Con objeto de explotar el potencial de detección de la red, la literatura sobre este tema se ha enfrentado a problemas de detección distribuida óptima, donde la optimalidad normalmente hace referencia al diseño adecuado de diferentes parámetros de la red con el objeto de minimizar alguna función de coste relacionada con las prestaciones globales de detección. Sin embargo, este problema de optimización tiene normalmente restricciones asociadas con los posibles parámetros físicos y de diseño de la red que pueden ser seleccionados a la hora de maximizar las prestaciones de detección de la misma. En muchas aplicaciones algunos parámetros físicos y de diseño, como por ejemplo la arquitectura de la red o los esquemas de procesado local de las observaciones de los sensores, bien están fuertemente restringidos a un conjunto de posibles alternativas de diseño, o bien no pueden ser variables de diseño en nuestro problema de optimización. A pesar de que estos parámetros pueden estar relacionados con las prestaciones de detección de la red, las anteriores restricciones podrían estar impuestas por factores tales como el entorno en el que la red se despliega, el presupuesto de energía disponible de la red o las capacidades de procesado de los sensores. Consecuentemente, es necesario caracterizar sistemas de detección distribuidos óptimos con varias arquitecturas, diferentes procesos de observación y diferentes esquemas de procesado local.

La mayor parte de los trabajos tratando la caracterización de sistemas de detección distribuida han asumido escenarios en los que, bajo cada uno de los posibles estados del fenómeno de interés, las observaciones son independientes de un sensor a otro. Sin embargo, hay muchos escenarios prácticos donde la asumpción de independencia condicional es violada como consecuencia de la presencia de diferentes fuentes de correlación. A pesar de esto, muy pocos trabajos han tratado las anteriores caracterizaciones bajo la misma variedad de escenarios que bajo la asunción de independencia condicional. De hecho, cuando la estrategia de la red no es un parámetro a optimizar, bajo la asunción de observaciones condicionalmente dependientes la literatura existente solo ha obtenido caracterizaciones asintóticas de las prestaciones de detección asociadas con redes paralelas cuyas reglas de procesado local se

basan en esquemas de amplificación y retransmisión.

Motivado por este último hecho, en esta tesis, bajo la formulación de Neyman-Pearson, llevamos a cabo la caracterización de sistemas de detección distribuida con observaciones dependientes, varias arquitecturas de red y reglas de cuantificación binaria en los sensores. En particular, considerando una red paralela desplegada aleatoriamente a lo largo de una línea recta, bajo la formulación de Neyman-Pearson derivamos una expresión cerrada del exponente de error asociado a la fusión de decisiones locales Makovianas cuando, con respecto a los espaciados entre sensores, sólo se conoce su distribución. Después de analizar algunas propiedades analíticas del derivado exponente de error, llevamos a cabo evaluaciones de su expresión cerrada con el objeto de determinar las diferentes tendencias de detección que pueden aparecer con dependencia creciente y bajo dos modelos de espaciado entre sensores muy conocidos. Estos dos modelos son sensores equiespaciados con fallos y sensores exponencialmente espaciados con fallos. Más tarde, los anteriores resultados son extendidos a una red paralela bidimensional que, formada por un conjunto de dispositivos equiespaciados sobre una rejilla rectangular, lleva a cabo un test de Neyman-Pearson para discriminar entre dos diferentes campos aleatorios causales de Markov definidos en un espacio de estados binario. Seguidamente, bajo observaciones condicionalmente dependientes y bajo la formulación de Neyman-Pearson, esta tesis se centra en la caracterización de las prestaciones de detección asociada a redes tándem óptimas con comunicación binaria entre los nodos de fusión. Para hacer eso, derivamos condiciones bajo las cuales, en una red tándem óptima con una arbitraria restricción en la probabilidad de falsa alarma global, la probabilidad de pérdida de la red, es decir la asociada último nodo de fusión, converge a cero según el número de etapas de fusión tiende a infinito. Finalmente, después de extender este resultado bajo la formulación bayesiana, proporcionamos dos ejemplos donde estas condiciones son aplicadas para caracterizar las prestaciones de detección de la red. A partir de estos ejemplos ilustramos diferentes escenarios de dependencia en los que una red tándem óptima puede o no lograr detección asintóticamente perfecta tanto bajo la formulación bayesiana como bajo la formulación de Neyman-Pearson.

Contents

Chapter 1

Introduction

1.1 Background and motivation

Recent progress in micro-electro-mechanical systems has allowed the development of devices integrating sensing, processing and wireless communication capabilities. These devices, named wireless sensors, are typically inexpensive, self-powered devices (by using batteries or devices allowing to capture energy from the environment), and have limited communication capability. They are deployed in large numbers over a possible large geographical area to implement dense networks that, in a distributed fashion, support a given application. Most of these applications involve canonical problems such as the detection, localization or tracking of a target. In this work we are interested in the distributed detection problem. The framework of detection theory was firstly extended to a decentralized setting by [Tenney and Sandell, 1981]. Evidently, the performance of a distributed system is suboptimal in comparison with its centralized counterpart, as information may be lost in local processing and transmission. Nonetheless, factors such as cost, communication bandwidth, and reliability have motivated the massive use of distributed systems. At the same time, this last fact has raised a growing attention regarding the design and analysis of distributed detection systems operating under different stringent resource constraints.

With the aim of exploiting all the detection potential of the network, the literature on this issue has faced optimal distributed detection problems, where optimality usually consists in properly designing the parameters of

the network in order to minimize some cost function related to the overall detection performance of the network. However, not all the possible physical and design parameters can be chosen in order to solve those optimization problems. In many applications, some of them, such as the network architecture or the local processing scheme of the sensor observations just to cite a couple of examples, are either constrained to a set of possible alternatives or to only one option. This might occur because of the environment where the network is deployed, the energy budget of the system or the processing capabilities of the devices among many other possible factors. Consequently, there is a great interest on characterizing optimal decentralized detection systems with various architectures, different local processing schemes and different observation processes that can or cannot present sources of spatial correlation among the different sensor observations. This characterization usually involves the derivation of optimal data fusion rules and optimal local processing schemes according to a specific criterion and under different design constraints. Likewise, it can also involve the development of design tools that link the detection performance of the network with its different features such as its node deployment or the kind of channels (parallel access channels or multiple access channels) employed for the communication between the sensors and the fusion node. Precisely, related to this last problem and focusing on the parallel and serial architecture, the main issue addressed in this thesis dissertation consists in characterizing the Neyman-Pearson detection performance of distributed systems based on conditionally dependent observations and binary quantization rules at the sensors.

The background associated with the problem faced in this thesis dissertation is initiated with the study of distributed detection systems under the assumption of conditionally independent sensor observations. Under this assumption and under a specific formulation, Bayesian or Neyman-Pearson, several works focused on the derivation of local processing schemes and fusion rules for optimal sensor networks. Some of them considered systems with a parallel architecture formed by a large number of devices and a data fusion center that, regarding a local and non-cooperative summary of the sensor observations, makes a global inference about the occurrence of an event of interest. Regarding the local quantization functions applied by the sensors under this kind of topology, [Tsitsiklis, 1993] showed that when the number of sensors goes to infinity and the observations taken by them are conditionally independent given any

hypothesis, the optimal binary decentralized detection is achieved by identical local detection rules. Later, in [Chamberland and Veeravalli, 2003] the authors extended this last result in order to show that identical transmitters are also optimal when an arbitrarily large parallel network has to detect deterministic signals in additive Gaussian noise. Additionally, in [Chamberland and Veeravalli, 2004], under the mentioned assumption of conditional independence, they also proved that the gain offered by having more sensors exceeds the benefits of getting detailed information from each sensor. In connection with the design of the data fusion rules that maximize the detection performance of the network, [Varshney, 1986] in the Bayesian set up and [Thomopoulos et al., 1987] under the Neyman-Pearson formulation obtained the optimal fusion rule when the sensor observations are conditionally independent under each hypothesis of the test.

The literature addressing the design of sensor networks other than the parallel configuration is less extensive. Nevertheless, due to different reasons such as the coverage of the sensors or the energy budget of the network among others, the study of networks with different architectures has also been of great interest. Focusing on distributed detection systems with serial architecture, as it happened for the analysis and design of parallel configurations, their study was initiated with the derivation of the optimal fusion rules performed at each fusion stage under the Bayesian set up and the Neyman-Pearson formulation. According to [Ekchian and Tenney, 1982] and [Viswanathan et al., 1988], under independent sensor observations given any hypothesis the optimal decision rule for each fusion stage matches a likelihood ratio test when a Bayesian set up and a Neyman-Pearson formulation is considered respectively. By formulating the distributed detection problem as a deterministic, multistage, nonlinear optimal control problem, this same result is obtained in [Tang et al., 1991] for both the Bayesian and the Neyman-Pearson set ups.

After deriving the optimal local processing schemes and the optimal data fusion rules under a specific formulation, in several works the corresponding detection performance is characterized in terms of different physical and design parameters of the network. Among others, considering a parallel architecture and a large number of sensor observations, works such as [Artés-Rodríguez et al., 2005], [Appadwedula et al., 2005], [Lázaro et al., 2005], [Artés-Rodríguez and Lázaro, 2007], [Tay et al., 2007] and [Lázaro et al., 2009] carried out the aforementioned characterization for systems operating under different censoring schemes that, avoiding specific

communications between some sensors and the fusion center, try to preserve the energy budget of the network. Alternatively, instead of considering a censoring scheme, [Artés-Rodríguez, 2004] and [Artés-Rodríguez et al., 2004] are based on a local parametric model of the detection probability of the sensors in order to determine how the sensors should be configured to be efficient in terms of energy and achieve the best local discrimination between the two hypothesis at the same time. Additionally, under the parallel architecture other studies, for instance [Huber, 1965] and [Marano et al., 2009], conducted research on the effect of byzantine sensors on the detection performance of sensor networks. When the number of sensor observations approaches infinity, it has been also of great interest to compare the detection performance achieved by distributed detection systems with or without feedback. [Zoumpoulis et al., 2010], [Tay, 2011] and [Tay and Tsitsiklis, 2011] are some examples of this kind of study. Analogously, in [Tay et al., 2008a], [Tay et al., 2008b], [Tay et al., 2009] and [Zhang et al., 2012] the previous comparison is also carried among different tree architectures and under the same asymptotic regime. In the case that the network architecture is not a design parameter, for systems other than the parallel configuration there are also works addressing similar design matters. For instance, the authors in [Papastavrou and Athans, 1992] provide the conditions that have to be satisfied by the observation model in order to make a serial system achieve zero probability of error in the aforementioned asymptotic regime. Moreover, complementing the previous work, [Tang et al., 1991] used optimal control theory in order to study the best placement for a node in a serial system with binary communication between the fusion units.

Although there are many results on the topic of decentralized detection, a few studies try to solve the cited problem without assuming conditional independence among local sensor observations given any hypothesis. In many practical scenarios the sensor observations are dependent under both hypothesis as a result of the presence of correlation sources in signals such as the noise or the one associated with the event of interest in the detection problem. Since the correlation sources are usually spatial, when they are present in the detection scenario, the conditional independence assumption is more and more violated as the density of the network increases, which usually happens in asymptotic characterizations for networks deployed over a region with a specific area and formed by an increasingly number of sensors.

Some of the first results removing the independence assumption designed

the local processing schemes and the data fusion rules of optimal sensor networks. In a scenario where the sensors send conditionally dependent local decisions to the fusion center, [Drakopoulos and Lee, 1991] showed that the optimal fusion rule under the Neyman-Pearson criterion consists in the joint likelihood ratio of the messages received by the data fusion center. In [Kam et al., 1992] the authors generalized the structure of the optimal data fusion rule when a Bayes hypothesis test is performed, and the sensor observations are dependent. Likewise, holding the conditional dependence among sensor observations the optimal local processing schemes were derived in [Hoballah and Varshney, 1989] under the Bayesian set up. The extension of this result to the Neyman-Pearson formulation can be found in [Blum, 1996] and [Yan and Blum, 2000]. More recently, some researchers addressed the previous design for sensor networks with tree architectures. For instance, in [Xiang and Wang, 2006] and [Yan and Blum, 2001] the authors derived the optimal fusion rules for tandem networks performing a Bayesian and a Neyman-Pearson test respectively.

In scenarios where the sensor observations are conditionally dependent given the true hypothesis, the optimal local processing scheme of a given sensor and the optimal fusion rule have intricate forms that depend on the decision rules of other sensors. In [Willett et al., 2000] the authors presented a thorough analysis for the binary quantization of a pair of dependent Gaussian random variables. Using that analysis they indicated that, in spite of the simplicity of the studied setting, the partition of the observation space may be hard to determine for an optimal detector. Actually, understanding and determining the partition of the sensor observation space led by the optimal decision and fusion rules is still an open problem that has been recently addressed in [Chen et al., 2011]. Due to this last fact, when characterizing the detection performance of sensor networks with conditionally dependent observations, the major part of the works assume that either the fusion rule or the local processing schemes are not optimal under the considered formulation.

In order to face the analysis of the detection performance of systems with dependent observations, under both the Bayesian set up and the Neyman-Pearson formulation the existing literature has usually considered parallel networks with optimal fusion rules and suboptimal local processing schemes. In particular, when addressing different design issues under different settings several works assumed amplify-and-relay schemes of continuous sensor observations whose correlation structure

usually follows well-known models of dependence such as one-dimensional (1-D) and two-dimensional (2-D) Markov random fields. Among them, using tools from large deviation theory and information theory provided in [Dembo and Zeitouni, 1998] and [Cover and Thomas, 2006] respectively, there are results regarding the effect of different physical and design parameters of the network on its detection performance. For 1-D sensor networks and based on closed-form expressions of error exponents most of these results are concerned with parameters such as the sensor density. See for instance [Sung et al., 2006], [Chamberland and Veeravalli, 2006] and [Misra and Tong, 2008]. Complementing the previous works, in [Li and Dai, 2007] and [Sung et al., 2008b] the parameters of interest are the sensor configuration as well as the kind of channel between each sensor and the fusion center respectively. Furthermore, in [Sung et al., 2008a], [Plata-Chaves et al., 2008], [Anandkumar et al., 2009], [Sung et al., 2009] and [Plata-Chaves, 2009] some of the previous studies were also extended to 2-D settings. However, due to the additional complexity appearing in this kind of scenarios, the results are based on implicit expressions of the error exponents or other well-known measures of dissimilarity.

In many applications, the sensor network is intended to have a long lifetime. In this case, energy consumption becomes a crucial aspect in networks with self-powered sensors because it limits the lifetime of the network and seriously affects the maintenance cost (to replace batteries in a large network, when possible, can be a high demanding task). Likewise, there are many practical scenarios where the amount of information forwarded from each sensor node to the fusion center is usually limited by the capacity of the channels between the sensors and the fusion center. When facing all these resource constraints under continuous observations, independently of the presence or absence of a correlation structure in the observation process, the quantization of the sensor observations might be more adequate than amplify-and-relay schemes. At the same time, it would be useful the use of network topologies, such as the other architectures different from the parallel, that allow for shorter-range communications. Nevertheless, these resource constraints are not the only reason to implement quantization rules at the sensors or to use different architectures. Sometimes, this may occur because of the discrete nature of the observation process, the limited cover and processing capabilities of cheap sensors and the features of area of the network deployment. Despite of everything, under conditionally

dependent observations, very few works have addressed the characterization of the detection performance associated to distributed systems with various architectures and local quantization rules at the sensors. For instance, one of the few results analyzing this kind of distributed detection systems is [Villard and Bianchi, 2011]. In this publication, considering a parallel architecture, the authors investigate the detection loss caused by the quantization of the sensor observations under the Neyman-Pearson formulation and under high-rate quantization regime i.e., when the number of quantization levels tends to infinity. However, motivating the research line followed in this thesis, so far, under conditionally dependent observations, there is no kind of characterizations of the detection performance associated to distributed systems with different architectures and a finite number of quantization levels at the sensors. Specifically, considering that the local processing scheme is a binary quantization of the sensor observations, this thesis addresses the subsequent problems:

- Derivation of analytically tractable expressions that, for parallel networks randomly deployed along a straight line, allow to assess which kind of trends of detection performance can appear with increasing dependency during the Neyman-Pearson fusion of Markov local decisions.

- Extension of the previous results to a 2-D parallel network that, being formed by a set of devices equally spaced on a rectangular lattice, discriminates between two different 2-D Markov causal fields defined on a binary state space.

- Derivation of conditions under which the fusion nodes of an optimal tandem network achieve zero probability of error as the number of fusion stages approaches infinity and when either a Bayesian set up or a Neyman-Pearson formulation is considered for an arbitrary correlation model.

1.2 Outline and contributions

This dissertation is organized as follows. Chapter 2 describes the state of the art associated with the data fusion problem in the framework of distributed hypothesis testing. Additionally, it reviews the existing tools

used to study the performance of distributed detection systems. Among other design and physical parameters of a 1-D sensor network, supported by large deviation theory Chapter 3 studies the effect of the correlation on the Neyman-Pearson fusion of dependent local decision. Next, in Chapter 4 these same results are also extended to a 2-D parallel network. Afterwards, Chapter 5 derives necessary and sufficient conditions that, under both the Bayesian set up and the Neyman-Pearson formulation, make a serial network converge to deciding the right hypothesis with arbitrary low probabilities of error. Finally, Chapter 6 concludes with a summary and discussion of our results as well as a description of the possible future research lines.

In the following we summarize the main contributions of this dissertation.

Chapter 3. In this part, published in [Plata-Chaves and Lázaro, 2009] and [Plata-Chaves and Lázaro, 2011], we consider a distributed detection system formed by a large number of local detectors and a data fusion center that, under the Neyman-Pearson formulation, fuses the binary quantizations of the sensor observations. In the analyzed two-stage detection system the local decisions are taken with no kind of cooperation among the devices and they are transmitted to the fusion center over an error free parallel access channel. In addition, the sensors are randomly deployed along a straight line, and the corresponding sensor spacings are independently drawn from a common probability density function. For both hypothesis, H_0 and H_1, depending on the correlation structure of the observed phenomenon the local decisions might be dependent. In the case of being dependent, their correlation structure is modelled with a 1-D Markov random field with nearest-neighbour dependency and binary state space. Under this scenario, we firstly derive a closed-form error exponent for the Neyman-Pearson fusion of the local decisions when the involved data fusion center only knows the distribution of the sensor spacings. Secondly, based on a single parameter that captures the mean correlation strength among the local decisions, some analytical properties of the error exponent are investigated. Finally, we develop a physical model for the conditional probabilities of the Markov random fields that might be present under each hypothesis. Using this model we characterize the error exponent for two well-known models of the sensor spacing: i) equispaced sensors with failures, and ii) exponentially spaced sensors with failures.

Chapter 4. In this part of the dissertation, whose results have been published in [Plata-Chaves and Lázaro, 2010], we consider the same distributed detection system as in the previous chapter. However, this time

the devices are located on a rectangular lattice so that sensors belonging to a specific row or column are equally spaced. Additionally, for each hypothesis H_0 and H_1, the correlation structure of the local decisions is modelled with a 2-D random causal field where the rows and columns are outcomes of the same first-order binary Markov chain. Similarly to the previous chapter, under this scenario we derive a closed-form error exponent for the Neyman-Pearson fusion of the local decisions. After studying some analytical properties of this error exponent, the transition probabilities defining the assumed 2-D random causal field are associated to a specific physical model in order to see the effect of different parameters of the network on its overall detection performance. As we do in Chapter 3, among the different parameters of the network we pay special attention to the mean correlation strength among neighbour local decisions.

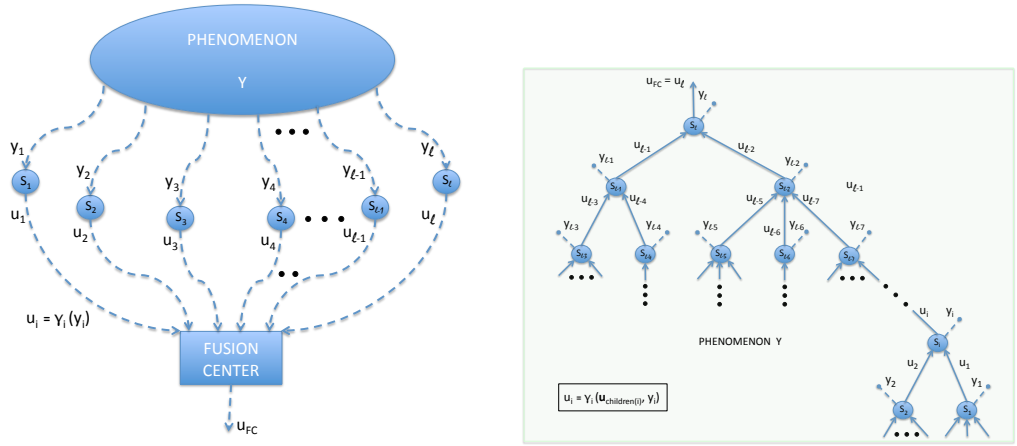
Chapter 5. In this final part, whose results have been published in [Plata-Chaves et al., 2011a] and [Plata-Chaves et al., 2012b], we consider a sensor network with tandem architecture. When solving a specific distributed detection problem where the observations might be dependent under each one of the two possible hypothesis, H_0 or H_1, we assume that, based on the available information, each fusion stage provides a binary message about the presence or absence of the event of interest. Under this scenario and under the Neyman-Pearson formulation, we firstly derive necessary and sufficient conditions ensuring that, for any arbitrary constraint on the overall probability of false alarm and an arbitrary correlation model, the corresponding probability of misdetection goes to zero as the number of fusion stages approaches infinity. Afterwards, considering the same scenario these conditions are extended to optimal tandem networks performing a Bayesian hypothesis test. Finally, we provide some illustrative examples where the application of the previous necessary and sufficient conditions is shown. With these examples we illustrate different dependence scenarios where a tandem network can or cannot achieve asymptotic perfect learning under either the Bayesian set up or the Neyman-Pearson formulation.

Chapter 2

Hypothesis testing and data fusion

2.1 Introduction

Based on observations regarding a specific phenomenon we frequently have to select a course of action from a set of possible options. This problem can be solved by one single decision maker in a centralized way. However, because of different constraints, this detection task sometimes has to be performed in a distributed fashion. In a distributed detection system several devices, deployed in a specific region, generate summaries of their locally available observations. Later, dealing with a data fusion problem the aforementioned summaries are communicated and fused along the network in order to produce a final decision regarding the state of the phenomenon of the network. Clearly, the propagation of the summaries of the sensor observations depends on the architecture of the distributed detection system. In the case that the system has a parallel architecture as the one shown in Figure 2.1(a), the summaries of the sensor observations are transmitted to an entity, called data fusion center. Based on the summaries of the sensor observations this fusion center provides a global decision regarding the best course of action that has to be selected according to some specific criteria. In other topologies the fusion of all the summaries of the sensor observations is carried out cooperatively by all the devices of the network. For instance, when a tree structure is considered (see Figure 2.1(b)), several partial (local) decisions are taken hierarchically by different entities at different levels of



(a) Parallel architecture (b) Tree architecture

Figure 2.1: *Diagrams of distributed detection systems with different architectures.*

the network These local decisions are based on the fusion of their locally available observations as well as the decisions taken by their successors in the tree. This way, the final decision of the system matches the course of action chosen by the root of the tree.

Due to the fact that, independently of the network topology, the in-network processing strategies involve detection steps under a parallel architecture, we are going to consider the topology shown in Figure 2.1(a) when describing the state of the art associated with the data fusion problem in the framework of distributed hypothesis testing. Specifically, focused on parallel fusion networks where the devices take binary local decisions when fusing its available local information, this chapter initially covers the foundational work that, in a distributed setting, established the optimal distributed schemes under the subsequent different formulations

- Bayesian detection.
- Neyman-Pearson detection.
- Sequential detection.
- Composite hypothesis testing.

Afterwards, in the spirit of the research topic associated with this thesis dissertation, the chapter concludes with a brief review of the different analytic tools used in the characterization of optimal parallel networks solving parametric detection problems with a fixed sample size.

2.2 Basic model and notation

In this section we firstly introduce the basic model used during this chapter when describing the state of the art associated with the data fusion problem in the framework of distributed hypothesis testing. At the same time, we will start to establish some notation that will be useful for the rest of the chapters of this thesis dissertation. In general terms, this notation will be standard. Scalar magnitudes are denoted using regular face letters, e.g., x, while matrices and vectors are written as bold-face upper-case and lower-case letters, respectively, e.g., matrix $\boldsymbol{X}$ and vector $\boldsymbol{x}$. All vectors are column vectors. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$ indicate the transpose and Hermitian transpose, respectively. Additionally, $\boldsymbol{A}(m,n)$ denotes the element of the m-th row and n-th column in the matrix $\boldsymbol{A}$. Sets are denoted using calligraphic letters, e.g. $\mathcal{Q}$, and random variables are in capital letters, e.g. X. The mean and the variance of a random variable X are written as $E\{X\}$ and $Var\{X\}$, respectively. When there is potential ambiguity in the argument of the expectation or the variance operator, $E_{X|y}\{\cdot\}$ and $Var_{X|y}\{\cdot\}$ denote the expectation and the variance with respect to (w.r.t.) a random variable X when the random variable $Y = y$ respectively. For instance, $E_{X|H_k}\{\cdot\}$ and $Var_{X|H_k}\{\cdot\}$ indicate the expectation and the variance given the hypothesis H_k as the true hypothesis. Moreover, $X \sim f_X(x)$ denotes that the continuous random variable X is distributed according to the probability density function (p.d.f.) $f_X(\cdot)$ whereas $Y \sim P_Y(y)$ means that the discrete random variable Y has $P_Y(\cdot)$ as probability mass function (p.m.f.). Finally, when denoting the probability that a discrete random variable, X, whose p.m.f. equals $P_X(x)$, takes on the value x, we will simply write $P(X = x)$ for the sake of simplicity.

Regarding the model used along this chapter we consider a parallel fusion network as the one shown in Figure 2.1(a). This distributed detection system is formed by a set of ℓ devices, also called sensors, deployed over a two-dimensional area, $\mathcal{D} \subseteq \mathbb{R}^2$. Assuming that, for $k \in \{0,1\}$, $i \in \{1,2,\ldots,\ell\}$ and $\boldsymbol{y} = [y_1, y_2, \ldots, y_\ell]^T$, the marginal and joint p.d.f. of the

sensor observations under H_k is continuous and given by $f_{Y_i|H}(y_i|H_k)$ and $f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)$ respectively, each sensor, placed at the coordinates $\boldsymbol{x}_i \in \mathcal{D}$, initially performs a binary quantization of its own observations, $\boldsymbol{y}_i \in \mathcal{Y}_i$. To do that, it applies the following quantization function

$$u_i = \gamma_i(\boldsymbol{y}_i): \quad \mathcal{Y}_i \longrightarrow \mathcal{U} \tag{2.1}$$

where $\mathcal{U} \in \{0, 1\}$ and

$$u_i = \begin{cases} 0 & \text{if } H_0 \text{ is locally decided,} \\ 1 & \text{if } H_1 \text{ is locally decided.} \end{cases} \tag{2.2}$$

This way, for the detection process undertaken by the i-th device

$$P_M(i) = P(U_i = 0|H_1) \tag{2.3}$$

denotes the probability of misdetection,

$$P_D(i) = 1 - P_M(i), \tag{2.4}$$

is the probability of detection and

$$P_{_{FA}}(i) = P(U_i = 1|H_0) \tag{2.5}$$

is equal to the corresponding probability of false alarm. Note that the previous probabilities characterize the performance of the first detection stage of the network. In order to undertake the second detection step the set of binary local decisions taken by all the devices of the network, $\boldsymbol{u} = [u_1, u_2, \ldots, u_\ell]^T$, is transmitted to an entity over a set of parallel access channels. Afterwards this central entity, called fusion center, fuses all the local decisions and makes a global inference about what hypothesis is present. In particular, completing the detection task of the network the fusion rule applied by the fusion center is defined as follows

$$u_{FC} = \gamma_{FC}(\boldsymbol{u}): \quad \mathcal{U}^\ell \longrightarrow \mathcal{U} \tag{2.6}$$

where $\mathcal{U}^\ell$ is the cartesian product of the set $\mathcal{U}$ with itself ℓ times and where

$$u_{FC} = \begin{cases} 0 & \text{if } H_0 \text{ is decided,} \\ 1 & \text{if } H_1 \text{ is decided.} \end{cases} \tag{2.7}$$

Consequently,

$$P_M = P(U_{FC} = 0|H_1) \tag{2.8}$$

denotes the overall probability of misdetection of the system,

$$P_D = 1 - P_M, \tag{2.9}$$

is equal to the corresponding overall probability of detection and

$$P_{_{FA}} = P(U_{FC} = 1|H_0) \tag{2.10}$$

is the overall probability of false alarm of the network. Obviously, for any given decision problem, we can select different strategies formed by a set of local decisions rules and a fusion rule

$$\Gamma = \{\gamma_{FC}, \{\gamma_i\}_{j=1}^{\ell}\}. \tag{2.11}$$

We can easily verify that each strategy, Γ, will yield specific values for P_M and $P_{_{FA}}$. However, we would like to choose some strategy, Γ^*, that is optimal in some sense. There are several useful definitions of optimality for such problems. Precisely, the main issue of this chapter is the design of the strategy under some of these useful definitions of optimality.

2.3 Some important properties of the fusion rules

For the setting considered in the previous section, when designing the fusion rule we have to take into account that it is essentially a logical function with ℓ binary inputs and one binary output. In general, we can easily realize that there are 2^{2^ℓ} possible fusion rules when there are ℓ binary local decisions. For instance, as it is listed in Table 2.1, there are 16 possible fusion rules for combining two binary local decisions. In this data fusion problem two commonly used fusion rules are the AND rule and the OR rule. In the AND rule, indexed as f_2 in Table 2.1, u_{FC} is equal to one only when both of the binary local decisions are one, i.e.

$$u_{FC} = \begin{cases} 1 & \text{if } u_1 = 1 \text{ and } u_2 = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

Input		Output u_{FC}															
u_1	u_2	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Table 2.1: *Possible fusion rules for the fusion of two binary decisions.*

In the case of the OR rule, f_8 in Table 2.1, u_{FC} equals one if at least one of the binary inputs is equal to one, i.e.,

$$u_{FC} = \begin{cases} 0 & \text{if } u_1 = 0 \text{ and } u_2 = 0, \\ 1 & \text{otherwise.} \end{cases} \tag{2.13}$$

At the same time, in the considered fusion problem there are also some functions that may not be suitable as fusion rules. Realize that the all zero function, f_1, and the all one function, f_{16}, totally disregard the two inputs and, therefore, may not be adequate in many scenarios. Similarly, some other fusion rules that may be unsuitable totally disregard one of the inputs. This is the case of f_4 which totally disregards the binary input u_2.

At first sight, from the previous example we can verify that the possible number of fusion rules is very high. However, simplifying the search of the fusion that has to be applied there are many useful definitions of optimality discarding functions $\gamma_{FC}(\boldsymbol{u})$ that do not satisfy the subsequent property.

Definition I (Monotonic fusion rule)**.** *Let $\mathcal{S}_1(m)$ be the set of m binary local decisions that are equal to one and $\mathcal{S}_0(\ell - m)$ be the set of remaining $\ell - m$ local decisions that are equal to zero. Given $\mathcal{S}_1(m)$ and $\mathcal{S}_0(\ell - m)$, assume that the global decision is $u_{FC} = 1$. Now, consider that $\mathcal{S}_1(m')$, with $m' > m$, is another set of m' positive local decisions containing the set $\mathcal{S}_1(m)$. Let $\mathcal{S}_0(\ell - m')$ be the corresponding set of local decisions equal to zero. Then, a fusion rule is monotonic if and only if the global decision, u_{FC}, equals one for all possible sets $\mathcal{S}_1(m')$ satisfying the previous condition.*

In [Varshney, 1997] we can find several definitions of optimality that, under the assumption of independent sensor observations, satisfy the above property and that, therefore, drastically reduce the number of fusion rules that have to be considered in the design problem. Some illustration of this

drastic reduction can be found in the example considered in this section. Looking at Table 2.1 we can easily see that, under definitions of optimality discarding non-monotonic fusion rules, only the functions f_1, f_2, f_4, f_6, f_8 and f_{16} could be taken into account when fusing two binary inputs.

In addition to the monotonocity property, other further considerations and constraints often result in further reductions of the number of fusion rules to be taken into account. One of the most typical additional considerations is the one given in the subsequent corollary.

Corollary I. *For a fusion center with ℓ binary inputs and any fixed monotonic fusion rule, the corresponding probability of detection P_D is an increasing function of the probabilities of detection associated with each one of its binary inputs, i.e. $\{P_D(i)\}_{i=1}^{\ell}$.*

It can be easily seen that the previous corollary is direct consequence of the monotonocity property appearing in the fusion rules derived under some definitions of optimality. One instance of this last fact can be found in [Thomopoulos et al., 1989] when the local decisions are conditionally independent given the true hypothesis. Nevertheless, we do not have to forget that, depending on the considered setting, there could be other constraints reducing the number of suitable fusion rules. To see that and conclude this section, assume that, in the problem of fusing two binary inputs, both local detectors have the same performance in terms of probability of detection and probability of false alarm, i.e. $P_D(1) = P_D(2)$ and $P_{FA}(1) = P_{FA}(2)$. Besides considering a definition of optimality that discards non-monotonic fusion rules, also assume that our cost function symmetrically penalizes errors and hits. In other words, there is the same cost assigned if H_1 is decided when H_0 is true and vice-versa. Analogously, the cost is the same when the null or the alternative hypothesis is correctly detected. Under all these constraints, we can realize that the output of the optimal fusion rule should be the same for both $u_1 = 0$, $u_2 = 1$ and $u_1 = 1$, $u_2 = 0$. This means that the monotonic fusion rules f_4 and f_6 are no longer suitable. Thus, out of the 16 possible fusion rules, only f_1, f_2, f_8 and f_{16} may be suitable in the setting that we have considered along this section.

2.4 Bayesian distributed detection

In this section, as it was carried out in [Hoballah and Varshney, 1989], we deal with the derivation of the optimal strategy that has to be applied by a parallel fusion network performing binary hypothesis testing under the Bayesian formulation. This means that we are looking for the strategy, Γ, that minimizes the average cost of the overall system operation, $\Re(\Gamma)$. If we consider the setting described in Section 2.2 we can easily see that four possible courses of action can be chosen by the sensor network. Thus, assigning costs to each one of these possibilities the Bayes risk function, $\Re(\Gamma)$, we wish to minimize is given by

$$\Re(\Gamma) = \sum_{k=0}^{1} \sum_{i=0}^{1} C_{ik}\, P(H_k)\, P(\text{Decide } H_i | H_k \text{ is true}) \tag{2.14}$$

where C_{ik} is a positive constant denoting the cost of global decision being H_i when H_k is present and $P(H_k)$ equals the prior probability of hypothesis H_k, with $i, k \in \{0, 1\}$. If we make the usual assumption that making a wrong decision is more costly than making a correct decision, i.e. $C_{10} > C_{00}$ and $C_{01} > C_{11}$, the design of the considered distributed detection system under the Bayesian formulation can be casted as follows

$$\begin{aligned} \Gamma^* &= \underset{\Gamma}{\operatorname{argmin}} \{\Re(\Gamma)\} \\ &= \underset{\Gamma}{\operatorname{argmin}} \{C_F\, P_{_{FA}} - C_D\, P_D + C\} \end{aligned} \tag{2.15}$$

where

$$C_F = P(H_0)\ (C_{10} - C_{00}) > 0 \tag{2.16}$$

$$C_D = (1 - P(H_0))\ (C_{01} - C_{11}) > 0 \tag{2.17}$$

and

$$C = C_{01}\ (1 - P(H_0)) + C_{00}\, P(H_0) > 0. \tag{2.18}$$

For the system optimization problem stated in (2.15)-(2.18), after omitting the dependence of $\Re$ on Γ for notational simplicity, a person-by-person optimization (PBPO) methodology is usually adopted. This

optimization procedure gives necessary but not, in general, sufficient conditions to determine the globally optimal solution of the considered optimization problem. Usually, the aforementioned necessary conditions are provided by means of a set of coupled equations whose simultaneous solution yields the desired PBPO solution. In order to obtain that set of equations the PBPO procedure focuses on the optimization of one of the optimization variables under the assumption that the rest of the variables have been optimized and remain fixed. In particular, when solving (2.15) we have to view the decentralized detection system as a team whose members are the ℓ local detectors and the data fusion center. Afterwards, we have to optimize the decision rule associated with one of those team members under the assumption that the rest of decision rules have been already designed.

Initially under the previously described PBPO methodology we are going to focus on the derivation of the local decision rule associated with i-th detector that minimizes $\Re$. Consequently, we are going to determine the decision rule associated with the i-th local detector when the fusion center and the remaining $\ell - 1$ local detectors have been designed. Toward this goal, under the previous assumption and taking into account the subsequent expressions for the overall probability of false alarm

$$P_{_{FA}} = \sum_{\boldsymbol{u}} P(U_{FC} = 1|\boldsymbol{u})\, P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0) \tag{2.19}$$

and the overall probability of detection

$$P_D = \sum_{\boldsymbol{u}} P(U_{FC} = 1|\boldsymbol{u})\, P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1) \tag{2.20}$$

with $\boldsymbol{U} = [U_1, U_2, \ldots, U_\ell]^T$ we firstly expand $\Re$ in terms of the local decision taken by the i-th local detector

$$\begin{aligned} \Re = C &+ \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i1}) \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_1)\right] \\ &+ \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i0}) \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i0}|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i0}|H_1)\right] \end{aligned} \tag{2.21}$$

where $i \in \{1, 2, \ldots, \ell\}$,

$$\boldsymbol{u}^i = [u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_\ell]^T \tag{2.22}$$

and

$$\boldsymbol{u}^{ik} = [u_1, \ldots, u_{i-1}, u_i = k, u_{i+1}, \ldots, u_\ell]^T \tag{2.23}$$

with $k \in \{0, 1\}$. Next, by noting that, for $k \in \{0, 1\}$,

$$P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i0}|H_k) = P_{\boldsymbol{U}^i|H}(\boldsymbol{u}^i|H_k) - P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_k) \tag{2.24}$$

we can express $\Re$ as

$$\begin{aligned}
\Re = C &+ \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i1}) \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_1)\right] \\
&+ \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i0}) \left[C_F\, P_{\boldsymbol{U}^i|H}(\boldsymbol{u}^i|H_0) - C_D\, P_{\boldsymbol{U}^i|H}(\boldsymbol{u}^i|H_1)\right] \\
&+ \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i0}) \left[C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_1) - C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_0)\right] \\
= C_i &+ \sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_1)\right]
\end{aligned} \tag{2.25}$$

where

$$C_i = C + \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i0}) \left[C_F\, P_{\boldsymbol{U}^i|H}(\boldsymbol{u}^i|H_0) - C_D\, P_{\boldsymbol{U}^i|H}(\boldsymbol{u}^i|H_1)\right] \tag{2.26}$$

and

$$A(\boldsymbol{u}^i) = P(U_{FC} = 1|\boldsymbol{u}^{i1}) - P(U_{FC} = 1|\boldsymbol{u}^{i0}). \tag{2.27}$$

Since the decision of each detector, defined in (2.1), depends only on its own observation we have that

$$P_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}|\boldsymbol{y}) = \prod_{i=1}^{\ell} P_{U_i|Y_i}(u_i|y_i) \tag{2.28}$$

and

$$P_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}^{ik}|\boldsymbol{y}) = P(U_i = k|y_i)\, P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i) \tag{2.29}$$

for $k \in \{0,1\}$. Subsequently, if we remember that $f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)$ denotes the joint p.d.f. of the sensor observations $\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_\ell]$ under hypothesis H_k with $k \in \{0,1\}$, after defining

$$\boldsymbol{y}^i = [y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_\ell]^T \tag{2.30}$$

we can express

$$\begin{aligned} P_{\boldsymbol{U}|H}(\boldsymbol{u}^{ik}|H_k) &= \int_{\mathcal{Y}^\ell} P_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}^{ik}|\boldsymbol{y})\, f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)\mathrm{d}\boldsymbol{y} \\ &= \int_{\mathcal{Y}^\ell} P(U_i = k|y_i)\, P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i)\, f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)\mathrm{d}\boldsymbol{y} \end{aligned} \tag{2.31}$$

where the last equality follows from (2.29) and where

$$\mathcal{Y}^\ell = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \ldots \times \mathcal{Y}_\ell. \tag{2.32}$$

Hence, the Bayes risk function, $\Re$, that we are optimizing w.r.t. $\gamma_i(y_i)$ can be written as follows

$$\begin{aligned} \Re &= C_i + \sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) C_F \int_{\mathcal{Y}^\ell} P(U_i = k|y_i)\, P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i)\, f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_0)\mathrm{d}\boldsymbol{y} \\ &\quad - \sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) C_D \int_{\mathcal{Y}^\ell} P(U_i = k|y_i)\, P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i)\, f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_1)\mathrm{d}\boldsymbol{y} \\ &= C_i + \int_{\mathcal{Y}_i} \mathrm{d}y_i\, P(U_i = 1|y_i) \\ &\quad \times \left\{ \sum_{\boldsymbol{u}^i} \int_{\mathcal{Y}^{\ell-1}_{[i]}} \mathrm{d}\boldsymbol{y}^i A(\boldsymbol{u}^i) P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i) \left[C_F f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_0) - C_D f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_1) \right] \right\} \end{aligned} \tag{2.33}$$

where

$$\mathcal{Y}^{\ell-1}_{[i]} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \ldots \mathcal{Y}_{i-1} \times \mathcal{Y}_i \times \ldots \times \mathcal{Y}_\ell. \tag{2.34}$$

At this point, realize that C_i is a constant as far as the i-th local detector is concerned. At the same time, we can also note that, if the fusion center as well as the rest of sensors are fixed, the Bayes risk is minimized when the i-th decision maker applies the subsequent decision rule

$$P(U_i = 1|y_i) = \begin{cases} 1 & \text{if } D(i) \leq 0, \\ 0 & \text{otherwise} \end{cases} \tag{2.35}$$

where

$$D(i) = \sum_{\boldsymbol{u}^i} \int_{\mathcal{Y}_{[i]}^{\ell-1}} \mathrm{d}\boldsymbol{y}^i A(\boldsymbol{u}^i) P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i) \left[C_F f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_0) - C_D f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_1) \right]. \tag{2.36}$$

To conclude the derivation of $\gamma_i(y_i)$, by noting that, for $k \in \{0, 1\}$,

$$f_{\boldsymbol{Y}|H}(\boldsymbol{y}^i|H_k) = f_{\boldsymbol{Y}^i|Y_i,H}(\boldsymbol{y}^i|y_i, H_0)\, f_{Y_i|H}(y_i|H_k) \tag{2.37}$$

from (2.35) and (2.36) we have that, for $i \in \{1, 2, \ldots, \ell\}$, the decision rule $\gamma_i(y_i)$ minimizing the Bayes risk $\Re$ is the one shown in the subsequent expression

$$\begin{aligned} u_i &= \gamma_i(y_i) \\ &= \frac{f_{Y_i|H}(y_i|H_1)}{f_{Y_i|H}(y_i|H_0)} \underset{0}{\overset{1}{\gtrless}} \frac{\sum_{\boldsymbol{u}^i} \int_{\mathcal{Y}_{[i]}^{\ell-1}} A(\boldsymbol{u}^i) C_F P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i) f_{\boldsymbol{Y}^i|Y_i,H}(\boldsymbol{y}^i|y_i, H_0) \mathrm{d}\boldsymbol{y}^i}{\sum_{\boldsymbol{u}^i} \int_{\mathcal{Y}_{[i]}^{\ell-1}} A(\boldsymbol{u}^i) C_D P_{\boldsymbol{U}^i|\boldsymbol{Y}^i}(\boldsymbol{u}^i|\boldsymbol{y}^i) f_{\boldsymbol{Y}^i|Y_i,H}(\boldsymbol{y}^i|y_i, H_1) \mathrm{d}\boldsymbol{y}^i} \end{aligned} \tag{2.38}$$

or equivalently

$$u_i = \gamma_i(y_i) = \frac{f_{Y_i|H}(y_i|H_1)}{f_{Y_i|H}(y_i|H_0)} \underset{0}{\overset{1}{\gtrless}} \frac{\sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) C_F P_{\boldsymbol{U}^i|Y_i,H}(\boldsymbol{u}^i|y_i, H_0)}{\sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) C_D P_{\boldsymbol{U}^i|Y_i,H}(\boldsymbol{u}^i|y_i, H_1)}. \tag{2.39}$$

From the previous expression we can see that the local decision rule $\gamma_i(y_i)$ minimizing $\Re$ under the PBPO methodology is not, in general, a threshold test. Realize that the right-hand side (r.h.s.) of (2.39) is dependent on the sensor observation performed by the i-th local detector. However, if the sensor observations are assumed to be conditionally independent given each one of the two possible hypothesis the aforementioned dependence disappears and the local decision rule expressed in (2.37) reduces to the following threshold test

$$\begin{aligned} u_i &= \gamma_i(y_i) \\ &= \frac{f_{Y_i|H}(y_i|H_1)}{f_{Y_i|H}(y_i|H_0)} \underset{0}{\overset{1}{\gtrless}} \frac{\sum_{\boldsymbol{u}^i} C_F A(\boldsymbol{u}^i) \prod_{j=1, j\neq i}^{\ell} P_{U_j|H}(u_j|H_0)}{\sum_{\boldsymbol{u}^i} C_D A(\boldsymbol{u}^i) \prod_{j=1, j\neq i}^{\ell} P_{U_j|H}(u_j|H_1)}. \end{aligned} \tag{2.40}$$

Note that the previous simplification does not mean a reduction in the number of coupled non-linear equations to be solved when implementing the

local decision rules. In particular, it only means a considerable reduction of the computational difficulty that appears when we have to solve the aforementioned coupled equations.

To complete the design of the parallel fusion network under the Bayesian set up we now focus on the derivation fusion rule $\gamma_{FC}(\boldsymbol{u})$ that minimizes $\Re$ when a PBPO methodology is considered. Toward this goal, we first have to remember that the local detectors take their local decisions based on their own observations and with no kind of cooperation. These observations are precisely the inputs used by the data fusion in order to give a global inference about what hypothesis is present. Next, taking this last fact into account, with the aim of finding the fusion rule $u_{FC} = \gamma_{FC}(\boldsymbol{u}^*)$ minimizing the Bayes risk function given one of the 2^ℓ possible values of $\boldsymbol{u}$, we substitute (2.19) and (2.20) into (2.15) and express the Bayes risk function as follows

$$\Re = P(U_{FC} = 1|\boldsymbol{u}^*) \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_1)\right] + K(\boldsymbol{u}^*) \quad (2.41)$$

where

$$K(\boldsymbol{u}^*) = C + \sum_{\boldsymbol{u};\boldsymbol{u}\neq\boldsymbol{u}^*} P(U_{FC} = 1|\boldsymbol{u}) \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)\right] \quad (2.42)$$

with C given in (2.18) and $\boldsymbol{u}^*$ denoting the considered vector of binary decisions taken by the ℓ local detectors of the network. Looking at (2.42) it is clear that $K(\boldsymbol{u}^*)$ is constant once $\boldsymbol{u}^*$ is fixed. Therefore, for a given $\boldsymbol{u}^*$ the Bayes risk function $\Re$ expressed in (2.41) is minimized when the following fusion rule is employed

$$P(U_{FC} = 1|\boldsymbol{u}^*) = \begin{cases} 1 & \text{if } \left[C_F\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_0) - C_D\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_1)\right] \leq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2.43)$$

Equivalently, if we use the fact that $C_F > 0$ and that $C_D > 0$, the fusion rule $\gamma_{FC}(\boldsymbol{u}^*)$ minimizing $\Re$ may also be written as follows

$$u_{FC} = \gamma_{FC}(\boldsymbol{u}^*) = \frac{P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_1)}{P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_0)} \underset{u_{FC}=0}{\overset{u_{FC}=1}{\gtrless}} \frac{C_F}{C_D}. \quad (2.44)$$

In comparison with the local decision rules provided in (2.39), the previous expression shows that, even when the sensor observations are conditionally

dependent given one of the two hypothesis, a threshold test is the optimal fusion rule, $\gamma_{FC}(\boldsymbol{u}^*)$, minimizing $\Re$ under the PBPO methodology. In this case, concluding the Bayesian design of a parallel fusion network we can check that the conditional independence among the sensor observations only allows us to implement the fusion rule as the following weighted sum of the local decisions

$$\begin{aligned} u_{FC} &= \gamma_{FC}(\boldsymbol{u}^*) \\ &= \sum_{i=1}^{\ell} \ln\left(\frac{P_D(i)\ (1-P_{_{FA}}(i))}{(1-P_D(i))\ P_{_{FA}}(i)}\right) u_i \underset{u_{FC}=0}{\overset{u_{FC}=1}{\gtrless}} \ln\left(\frac{C_F}{C_D}\prod_{i=1}^{\ell}\frac{1-P_{_{FA}}(i)}{1-P_D(i)}\right). \end{aligned} \tag{2.45}$$

2.5 Neyman-Pearson distributed detection

After performing the design of the optimal Bayesian strategy in a parallel fusion network we now consider the same problem under the Neyman-Pearson formulation. In this case, without requiring the knowledge of the prior probabilities associated with each one of the possible hypothesis, we are looking for the strategy Γ that maximizes the overall probability of detection of the system subject to (s.t.) a specific constraint on the corresponding probability of false alarm. Specifically, taking into account the aforementioned Neyman-Pearson criterion, our problem can be formulated as follows

$$\Gamma^* = \underset{\Gamma \text{ s.t. } P_{_{FA}}(\Gamma)\leq\alpha\in[0,1]}{\operatorname{argmax}} \{P_D(\Gamma)\} \tag{2.46}$$

where $P_D(\Gamma)$ and $P_{_{FA}}(\Gamma)$ denote the probability of detection and the probability of false alarm at the fusion center under the strategy Γ respectively, i.e.

$$P_D(\Gamma) = P(U_{FC}=1|\Gamma, H_1) \tag{2.47}$$

and

$$P_{_{FA}}(\Gamma) = P(U_{FC}=1|\Gamma, H_0). \tag{2.48}$$

We consider again the PBPO methodology to solve the system optimization problem stated in (2.46). This time, we are firstly going to

consider the design of the optimal fusion rule. When facing that design problem we can easily verify that the Neyman-Pearson lemma, described in [Blahut, 1987], gives the optimal fusion rule under the PBPO methodology, in other words, when all the sensor detectors have already been designed. Specifically, from Neyman-Pearson lemma we have that, when the decisions taken by the sensors of the parallel network are $\boldsymbol{u}^*$, the optimal fusion rule is given by

$$\begin{aligned} u_{FC} &= \gamma_{FC}(\boldsymbol{u}^*) \\ &= P(U_{FC}=1|\boldsymbol{u}^*) = \begin{cases} 1 & \text{if } P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_1) > \lambda_{FC} P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_0), \\ \xi(\boldsymbol{u}^*) & \text{if } P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_1) = \lambda_{FC} P_{\boldsymbol{U}|H}(\boldsymbol{u}^*|H_0), \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{2.49}$$

with the threshold, $\lambda_{FC} < \infty$, and the randomization function, $\xi(\boldsymbol{u}^*) \in (0,1)$, chosen in order to satisfy that $P_{_{FA}}(\Gamma) \leq \alpha$. Except for the presence of a randomization function $\xi(\boldsymbol{u}^*)$ as well as the way the threshold λ_{FC} is selected we can note that this fusion rule takes the same form as the one derived in the previous section under the Bayesian set up. This last fact can also be seen when the sensor observations are conditionally independent given each one of the two possible hypothesis. To be more precise, as it happens under the Bayesian set up, if we take logarithms, the conditional independence among the sensor observations allows to implement the fusion rule as the subsequent weighted sum of the binary local decisions taken by the devices of the network

$$\begin{aligned} u_{FC} &= \gamma_{FC}(\boldsymbol{u}^*) \\ &= \begin{cases} 1 & \text{if } \sum_{i=1}^{\ell} \ln\left(\frac{P_D(i)\left(1-P_{_{FA}}(i)\right)}{(1-P_D(i))\,P_{_{FA}}(i)}\right) u_i > \ln\left(\lambda_{FC}\,\prod_{i=1}^{\ell}\frac{1-P_{_{FA}}(i)}{1-P_D(i)}\right), \\ \xi(\boldsymbol{u}^*) & \text{if } \sum_{i=1}^{\ell} \ln\left(\frac{P_D(i)\left(1-P_{_{FA}}(i)\right)}{(1-P_D(i))\,P_{_{FA}}(i)}\right) u_i = \ln\left(\lambda_{FC}\,\prod_{i=1}^{\ell}\frac{1-P_{_{FA}}(i)}{1-P_D(i)}\right), \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2.50}$$

Once we have undertaken the design of the optimal fusion rule under the Neyman-Pearson formulation we are now going to consider the design of the optimal sensor detectors under the PBPO methodology. Consequently, when providing conditions on the optimal decision rule performed by the i-th sensor we are going to assume that the fusion

rule and the rest of decision rules have already been designed. Under this assumption several approaches have been considered in the literature. Initially, several works attempted to solve (2.46) by using the Lagrange multiplier technique, in other words, maximizing $P_D(\Gamma) - \lambda\, P_{_{FA}}(\Gamma)$ without constraints. However, in [Tsitsiklis, 1993] the author demonstrates that this approach, which was taken in [Srinivasan, 1986a] and [Srinivasan, 1986b], fails if the overall receiving and operating curve (ROC) associated with the fusion center is not concave. This is significant since, as it is shown in [Cherikh, 1989], [Willett and Warren, 1992], [Yan and Blum, 2000] and [Yan and Blum, 2001] by means of different examples, it is not true that the overall ROC curve must be concave. In order to overcome this problem, in [Blum, 1996] it is shown that under some mild assumptions and the PBPO methodology the conditions on the optimal sensor detectors can be derived regardless the concavity of the overall ROC. Precisely, this approach, which relies on the proof used in [Blahut, 1987] when deriving the Neyman-Pearson Lemma for a centralized system, is the one that we are going to follow in this section.

With the aim of deriving the i-th local decision rule that, under the PBPO methodology, solves the Neyman-Pearson problem stated in (2.46) we are going to assume that the marginal p.d.f., $f_{Y_i|H}(y_i|H_k)$, has no point masses for all $k \in \{0,1\}$. Additionally, consider that the i-th sensor performs the subsequent binary quantization rule of its own observations

$$u_i = \gamma_i(y_i) = P(U_i = 1|y_i) = \begin{cases} 1 & \text{if } D_{1i}(y_i) \geq \lambda_i D_{0i}(y_i), \\ 0 & \text{if } D_{1i}(y_i) < \lambda_i D_{0i}(y_i), \end{cases} \tag{2.51}$$

where, for $i \in \{1, 2, \ldots, \ell\}$ and $k \in \{0,1\}$,

$$D_{ki}(y_i) = f_{Y_i|H}(y_i|H_k) \sum_{\boldsymbol{u}^i} P_{\boldsymbol{U}^i|Y_i,H}(\boldsymbol{u}^i|y_i, H_k)\, A(\boldsymbol{u}^i) \tag{2.52}$$

with $\boldsymbol{u}^i$ and $A(\boldsymbol{u}^i)$ defined in (2.22) and (2.27) respectively. Due to the previous definition of $\gamma_i(y_i)$ we know that

$$(\gamma_i(y_i) - \gamma_i'(y_i)) \cdot (D_{1i}(y_i) - \lambda_i\, D_{0i}(y_i)) \geq 0 \tag{2.53}$$

for all y_i and for any decision rule $\gamma_i'(y_i) \in [0,1]$. Hence, if we define $\mathcal{Y}_i$ as the subset of all y_i for which $f_{Y_i|H}(y_i|H_k) > 0$ for either $k = 0$ or $k = 1$, then

$$\int_{\mathcal{Y}_i} (\gamma_i(y_i) - \gamma_i'(y_i)) \cdot (D_{1i}(y_i) - \lambda_i\, D_{0i}(y_i))\, \mathrm{d}y_i \geq 0. \tag{2.54}$$

If we now expand the previous equation we are able to obtain

$$\begin{aligned}&\int_{\mathcal{Y}_i} D_{1i}(y_i)\gamma_i(y_i)\mathrm{d}y_i - \int_{\mathcal{Y}_i} D_{1i}(y_i)\gamma_i'(y_i)\mathrm{d}y_i \\ &\qquad \geq \lambda_i \left(\int_{\mathcal{Y}_i} D_{0i}(y_i)\gamma_i(y_i)\mathrm{d}y_i - \int_{\mathcal{Y}_i} D_{0i}(y_i)\gamma_i'(y_i)\mathrm{d}y_i \right)\end{aligned} \tag{2.55}$$

At this point, we undertake a reduction of each one of the terms appearing in (2.55). First, using the definition provided in (2.52) we have that, for $k \in \{0, 1\}$,

$$\begin{aligned}&\int_{\mathcal{Y}_i} D_{ki}(y_i)\gamma_i(y_i)\mathrm{d}y_i \\ &\quad = \sum_{\boldsymbol{u}^i} \left\{ A(\boldsymbol{u}^i) \int_{\mathcal{Y}_i} \gamma_i(y_i)\, f_{Y_i|H}(y_i|H_k)\, P_{\boldsymbol{U}^i|Y_i,H}(\boldsymbol{u}^i|y_i, H_k)\mathrm{d}y_i \right\} \\ &\quad = \sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i)\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_k) \\ &\quad \overset{(a)}{=} -Q_{ki} + \sum_{\boldsymbol{u}^i} \big[P(U_{FC} = 1|\boldsymbol{u}^{i1})\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i1}|H_k) \\ &\qquad\qquad + P(U_{FC} = 1|\boldsymbol{u}^{i0})\, P_{\boldsymbol{U}|H}(\boldsymbol{u}^{i0}|H_k) \big] \\ &\quad = -Q_{ki} + P(U_{FC} = 1|\gamma_i(y_i), H_k)\end{aligned} \tag{2.56}$$

where (a) follows from (2.24) and where

$$Q_{ki} = \sum_{\boldsymbol{u}^i} P(U_{FC} = 1|\boldsymbol{u}^{i0})\, P_{\boldsymbol{U}^i|H}(\boldsymbol{u}^i|H_k). \tag{2.57}$$

Similarly, following the same steps as in (2.56) we obtain this equality

$$\int_{\mathcal{Y}_i} D_{ki}(y_i)\gamma_i'(y_i)\mathrm{d}y_i = -Q_{ki} + P(U_{FC} = 1|\gamma_i'(y_i), H_k). \tag{2.58}$$

Thus, by noting that the terms Q_{0i} and Q_{1i} will cancel out in the r.h.s. and the left-hand side (l.h.s.) of (2.55) respectively, from (2.56)-(2.58) the aforementioned inequality can be rewritten as follows

$$P_D(\gamma_i(y_i)) - P_D(\gamma_i'(y_i)) \geq \lambda_i \left(P_{_{FA}}(\gamma_i(y_i)) - P_{_{FA}}(\gamma_i'(y_i)) \right) \tag{2.59}$$

if we use the subsequent definition

$$P(U_{FC}=1|\zeta_i(y_i),H_k)=\begin{cases} P_D(\zeta_i(y_i)) & \text{if } k=1, \\ P_{_{FA}}(\zeta_i(y_i)) & \text{if } k=0, \end{cases} \tag{2.60}$$

with $P_D(\zeta_i(y_i))$ and $P_{_{FA}}(\zeta_i(y_i))$ denoting the overall probability of detection and the overall probability of false alarm at the fusion center when the decision rule $\zeta_i(y_i)$ is applied by the i-th sensor respectively. Due to the fact that we require $P_{_{FA}}(\gamma_i(y_i))=\alpha$, then for any decision rule $\gamma_i'(y_i)$ satisfying the same constraint, i.e. $P_{_{FA}}(\gamma_i'(y_i))=\alpha$, (2.59) shows

$$P_D(\gamma_i(y_i)) \geq P_D(\gamma_i'(y_i)). \tag{2.61}$$

This means that, under the PBPO methodology, the local decision rule provided in (2.51) solves (2.46) as long as we can always pick a λ_i yielding $P_{_{FA}}(\gamma_i(y_i))=\alpha$ when there exists some sensor test achieving the required false alarm probability α for the given fusion rule and the given set of sensor processor rules.

Due to the fact that

$$\begin{aligned} P_{_{FA}}=\alpha =& P_{_{FA}}(i)\left(P(U_{FC}=1|u_i=1,H_0)-P(U_{FC}=1|u_i=0,H_0)\right) \\ &+P(U_{FC}=1|u_i=0,H_0) \end{aligned} \tag{2.62}$$

by noting that the quantities $P(U_{FC}=1|u_i=1,H_0)$ and $P(U_{FC}=1|u_i=0,H_0)$ are dictated by the already designed fusion rule and the other sensor rules, we check that, for any sensor test applied by the i-th sensor, i.e. for any $P_{_{FA}}(i)\in[0,1]$, the overall probability of false alarm, $P_{_{FA}}$, is between $P(U_{FC}=1|u_i=1,H_0)$ and $P(U_{FC}=1|u_i=0,H_0)$. Therefore, in order to show that (2.51) solves (2.46) under the PBPO methodology, bearing in mind (2.61) we only need to verify that for any of the aforementioned possible values of $P_{_{FA}}$ we can obtain any $0\leq P_{_{FA}}(i)\leq 1$ when (2.51) is performed at the i-th sensor. To do that, as it frequently holds in scenarios where the conditional p.d.f. of the sensor observations is continuous under H_0 and H_1, we only have to assume that the ratio $D_{1i}(y_i)/D_{0i}(y_i)$ is a continuous scalar random variable whose p.d.f. has no point masses of probability under either hypothesis, H_0 and H_1. Note that, under this assumption, if we define

$$a(\lambda_i)=P\left(\frac{D_{1i}(y_i)}{D_{0i}(y_i)}>\lambda_i|H_0\right) \tag{2.63}$$

$1 - a(\lambda_i)$ is the cumulative distribution function of a continuous random variable, $D_{1i}(y_i)/D_{0i}(y_i)$. Hence, since $a(\lambda_i)$ is non-increasing and continuous with $a(-\infty) = 1$ and $a(\infty) = 0$, we have that, for any of the possible values of P_{FA}, we can get $0 \leq P_{FA}(i) \leq 1$ when the decision rule (2.51) is performed at the i-th sensor.

Till now we have shown that the local decision rule given in (2.51) is a PBPO solution of the Neyman-Pearson problem stated in (2.46). However, we do not know if that solution in unique or not. In order to know more about this issue we can consider some decision rule $\gamma_i''(y_i)$ that solves (2.46) under the PBPO methodology. Taking into account (2.61) we have that this local decision rule satisfies the subsequent equality

$$P_D(\gamma_i(y_i)) = P_D(\gamma_i''(y_i)). \tag{2.64}$$

Next, since $P_{FA}(\gamma_i(y_i)) = P_{FA}(\gamma_i''(y_i)) = \alpha$ the following equation obtained from (2.64) is still valid

$$\begin{aligned} P_D(\gamma_i(y_i)) - Q_{1i} - \lambda_i \left(P_{FA}(\gamma_i(y_i)) - Q_{0i}\right) \\ = P_D(\gamma_i''(y_i)) - Q_{1i} - \lambda_i \left(P_{FA}(\gamma_i''(y_i)) - Q_{0i}\right). \end{aligned} \tag{2.65}$$

If we now recall (2.56)-(2.58) and (2.60), the previous equation can be rewritten as

$$\int_{\mathcal{Y}_i} (\gamma_i(y_i) - \gamma_i'(y_i)) \cdot (D_{1i}(y_i) - \lambda_i D_{0i}(y_i)) \, \mathrm{d}y_i = 0. \tag{2.66}$$

Due to the fact that (2.54) holds for any local decision rule, we have that the integrand in (2.66) has to be equal to zero for all $y_i \in \mathcal{Y}_i$. This last means that $\gamma_i(y_i)$ and $\gamma_i''(y_i)$ can differ only for $y_i \in \mathcal{Y}_i$ such that $D_{1i}(y_i) - \lambda_i D_{0i}(y_i) = 0$. Consequently, we can conclude that, as long as y_i and $D_{1i}(y_i)/D_{0i}(y_i)$ are continuous random variables whose marginal probability density functions have no point masses of probability, the PBPO solutions of (2.46) have the form expressed in (2.51) except possibly on a set having zero probability under H_0 and H_1.

As it happens under a Bayesian set up, when the sensor observations are conditionally independent given H_0 or H_1, the PBPO solution of (2.46) associated with the i-th detector reduces to the following likelihood ratio test

$$\begin{aligned} u_i &= \gamma_i(y_i) \\ &= \frac{f_{Y_i|H}(y_i|H_1)}{f_{Y_i|H}(y_i|H_0)} \underset{0}{\overset{1}{\gtrless}} \lambda_i \frac{\sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) \prod_{j=1, j\neq i}^{\ell} P_{U_j|H}(u_j|H_0)}{\sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) \prod_{j=1, j\neq i}^{\ell} P_{U_j|H}(u_j|H_1)}. \end{aligned} \tag{2.67}$$

Looking at (2.67) we can realize that the previous simplification does not mean a reduction in the number of coupled non-linear equations associated with the local decision rules obtained as PBPO solutions of (2.46). In particular, as we mentioned when we considered the Bayesian set up, the conditional independence among the sensor observations only means a considerable reduction of the computational difficulty that appears when solving the aforementioned coupled equations. This shows one of the many connections that exist between the Neyman-Pearson optimal strategy and the corresponding one under the Bayesian formulation. Actually, the only differences in the form of both strategies are that the Neyman-Pearson strategy may need randomization at the fusion center and that each device of the network uses a different threshold λ_i. Nevertheless, except the fact that the thresholds $\{\lambda_i\}_{i=1}^{\ell}$ and λ_{FC} are chosen in order to satisfy a constraint on the overall probability of false alarm, in the literature we can find several conditions that make the randomized fusion rules be suboptimal and that make all the thresholds be identical. On the one hand, in [Willett and Warren, 1992] it was shown that, as long as likelihood ratios of the unquantized observations are independent and contain no point masses of probability, the optimal Neyman-Pearson strategy does not randomize when the sensor observations are independent upon the hypothesis. This is so despite the fact that the data to be fused can be considered discrete. On the other hand, in [Yan and Blum, 2000] and [Yan and Blum, 2001] the authors show that, under the assumptions made in the derivation of the PBPO solution of (2.46), the optimal scheme has to have identical thresholds, i.e. $\lambda_{FC} = \lambda_1 = \lambda_2 = \ldots = \lambda_{\ell-1} = \lambda_\ell$, whenever the p.d.f. of $D_{1i}(y_i)/D_{0i}(y_i)$ under H_j is greater than zero for $0 < D_{1i}(y_i)/D_{0i}(y_i) < \infty$ and $j \in \{0, 1\}$. Thus, as it is pointed out in [Yan and Blum, 2000], when designing and implementing algorithms that provide the PBPO solution of (2.46), it should be taken into account that some conditions allow us to write in the same form the optimal strategies derived under the Bayesian set up and under the Neyman-Pearson formulation. This means that those algorithms can have many common features.

2.6 Distributed sequential detection

This section considers the derivation of the optimal strategy that solves a sequential detection problem in a parallel fusion network as the one described

in Section 2.2. Unlike the fixed-sample-size detection problems studied in the last two sections, observations are collected sequentially so that more information is available as time progresses. Under this setting, it is clear that the optimal strategy has to allow the system to process the observations sequentially and make a final decision as soon as we are able to guarantee the constraints associated with the detection performance. In this way, since the main aim is to take additional observations only if they are necessary, on average we need fewer observations to achieve the same probability of error as a distributed detection system solving a fixed-sample-size test. Nevertheless, as we will be able to verify in the optimal strategies, the aforementioned advantage is attained at the expense of an additional computation that is affordable under the restrictive assumption of independent and identically distributed (i.i.d.) sensor observations given each one of the two possible hypothesis.

As it will be assumed throughout this section, when sensor observations are i.i.d. under both hypothesis, two distinct formulations are possible during the design of a parallel fusion network that solves a sequential detection problem. In one case, first each sensor performs a sequential test on its own observations and arrives at a final local decision. Subsequently, in order to make a global decision about what hypothesis is present, the local decisions are fused at a site possibly remote to all local detectors. Thus, unlike fixed-sample-size detection problems, in this context the different fusion schemes have to take into account that local decisions may arrive at the fusion center at different times depending on the stopping times of the local detectors. In the other case, at each time instant each sensor i, with $i \in \{1, 2, \ldots, \ell\}$, uses the rule $\gamma_i^n(\cdot)$ in order to perform a binary quantization of its own observation. The resulting local decision is transmitted to the fusion center over an error free channel. This entails that each local detector sends a sequence of binary messages to the fusion center. In order to determine the true hypothesis, the fusion center carries out a sequential test where ℓ binary local decisions are processed at each time instant.

Following the same spirit as in the preceding sections, we consider the latter formulation. In other words, we face the design of a distributed detection system where the fusion center performs a test under a specific set up, in this case, under the sequential formulation. More formally, our problem is stated as deriving the optimal strategy associated with a parallel fusion network where, at each time instant n, the fusion center may choose to make a final decision on a stopping time τ or continue to take a new set

of ℓ local decisions, $\boldsymbol{U}^n = [U_1^n, U_2^n, \ldots, U_\ell^n]^T$, with u_i^n equal to the binary decision taken by the i-th detector at time n. Hence, considering a Bayesian formulation, if the final decision of the distributed detection system is given by

$$u_{FC} = \gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau}) \in \{0, 1\} \tag{2.68}$$

with

$$\boldsymbol{U}^n = [\gamma_1^n(\cdot), \ldots, \gamma_\ell^n(\cdot)]^T = \boldsymbol{\gamma}^n \tag{2.69}$$

and

$$\boldsymbol{U}^{n_a:n_b} = \left[\boldsymbol{\gamma}^{n_a}, \boldsymbol{\gamma}^{n_a+1}, \ldots, \boldsymbol{\gamma}^{n_b}\right]^T = \boldsymbol{\gamma}^{n_a:n_b} \tag{2.70}$$

for $n_a, n_b \in \{1, 2, \ldots, \tau\}$ and $n_a < n_b$, the problem that we wish to solve is

$$\Gamma^* = \underset{\Gamma, \tau > 0}{\operatorname{argmin}} \, E\left\{J(\gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau}), H)\right\} \tag{2.71}$$

where

$$\Gamma = [\gamma_{FC}^1(\boldsymbol{\gamma}^1), \gamma_{FC}^2(\boldsymbol{\gamma}^{1:2}), \ldots, \gamma_{FC}^{\tau}(\boldsymbol{\gamma}^{1:\tau}), \boldsymbol{\gamma}^{1:\tau}] \tag{2.72}$$

and

$$J(\gamma_{FC}^{\tau}(\boldsymbol{\gamma}^{1:\tau}) = i, H = H_k) = C\,\tau + C_{ik} \tag{2.73}$$

with C equal to positive cost associated with each time step used by the fusion center to take a set of ℓ local decisions and with C_{ik} denoting the cost of the final global decision, at time τ, being H_i when $H = H_k$ for $i, k \in \{0, 1\}$.

As we proceeded in the previous sections we use the PBPO methodology with the aim of providing necessary conditions that have to be satisfied by the strategy solving (2.71). Initially, focusing on the derivation of the local decision rules at a specific time instant n_0, i.e. $\boldsymbol{\gamma}^{n_0}$, we can verify that the solution to (2.71) may vary with the information available at the local detectors, in other words, the arguments of the functions $\gamma_i^{n_0}(\cdot)$ with $i \in \{1, 2, \ldots, \ell\}$. Due to this last fact, before solving (2.71) we have to know the amount of available memory at the sensors as well as the presence or absence of feedback channels in the distributed detection system. We can easily realize that, depending on those features, the binary local decision, $u_i^{n_0}$,

taken by the i-th sensor at the time instant n_0 can be a function of any subset of the union of its current observation $y_i^{n_0}$, all its past observations $\{y_i^n\}_{n=1}^{n_0-1}$, all its past local decisions $\{u_i^n\}_{n=1}^{n_0-1}$ and all the past decisions taken by the rest of local detectors $\{u_1^n, u_2^n, \ldots, u_{i-1}^n, u_{i+1}^n, \ldots, u_{\ell-1}^n, u_\ell^n\}_{n=1}^{n_0-1}$. In this section, in order to have the same network architecture as the one considered along the current chapter, we are going to assume a system with neither feedback from the fusion center nor local memory at the sensors. Equivalently, we will assume that, for any sensor i with $i \in \{1, 2, \ldots, \ell\}$, the corresponding binary decision u_i^n at time instant n only depends on its own current observation y_i^n, i.e.

$$u_i^n = \gamma_i^n \left(y_i^n \right). \tag{2.74}$$

Consequently, if we suppose that the fusion center and all the local detectors other than the i-th sensor at time n_0 have already been designed, for that time instant the objective function appearing in (2.71) can be written as follows when the sensor observations are conditionally independent across time as well as from sensor to sensor

$$E_{Y_i^{n_0}} \left\{ \underset{u_i^{n_0}}{\operatorname{argmin}} \, E_{H|y_i^{n_0}} \left\{ \Xi_i^{n_0}(u_i^{n_0}, H) \right\} \right\} \tag{2.75}$$

where

$$\Xi_i^{n_0}(u_i^{n_0}, H) = E_{\boldsymbol{U}(n_0,i)} \left\{ J(\gamma_{FC}^\tau(\boldsymbol{\gamma}^{1:\tau}), H) \right\} \tag{2.76}$$

and

$$\boldsymbol{U}(n_0, i) = \boldsymbol{U}^{1:n_0-1}, \boldsymbol{U}^{n_0,i}, \boldsymbol{U}^{n_0+1:\tau} \tag{2.77}$$

with

$$\boldsymbol{U}^{n_0,i} = \left[U_1^{n_0}, U_2^{n_0}, \ldots, U_{i-1}^{n_0}, U_{i+1}^{n_0}, \ldots, U_\ell^{n_0} \right]^T. \tag{2.78}$$

Developing the innermost expectation appearing in (2.75) we have that, under the PBPO methodology, (2.71) is minimized when the i-th detector applies the subsequent decision rule at time n_0

$$u_i^{n_0} = \gamma_i^{n_0}(y_i^{n_0}) = \begin{cases} 0 & D_i^{n_0}(y_i^{n_0}) \leq 0, \\ 1 & \text{otherwise} \end{cases} \tag{2.79}$$

with

$$\begin{aligned} D_i^{n_0}(y_i^{n_0}) = \Xi_i^{n_0}(0,H_0)P(H_0|y_i^{n_0}) + \Xi_i^{n_0}(0,H_1)P(H_1|y_i^{n_0}) \\ - \Xi_i^{n_0}(1,H_0)P(H_0|y_i^{n_0}) - \Xi_i^{n_0}(1,H_1)P(H_1|y_i^{n_0}). \end{aligned} \tag{2.80}$$

At this point, taking into account

$$P(H_k|y_i^{n_0}) = \frac{f_{Y_i^{n_0}|H}(y_i^{n_0}|H_k)\, P(H_k)}{f_{Y_i^{n_0}}(y_i^{n_0})} \tag{2.81}$$

with $k \in \{0,1\}$ and $P(H_k)$ denoting the prior probability of hypothesis H_k, if we note that $\Xi(1,H_0) > \Xi(0,H_0)$ and that $\Xi(0,H_1) > \Xi(1,H_1)$ when $C_{10} > C_{00}$ and $C_{01} > C_{11}$, we can express (2.79) as the subsequent decision rule

$$u_i^{n_0} = \gamma_i^{n_0}(y_i^{n_0}) = \begin{cases} 0 & \frac{f_{Y_i^{n_0}|H}(y_i^{n_0}|H_1)}{f_{Y_i^{n_0}|H}(y_i^{n_0}|H_0)} \leq \frac{P(H_0)}{P(H_1)}\, \frac{\Xi_i^{n_0}(1,H_0)-\Xi_i^{n_0}(0,H_0)}{\Xi_i^{n_0}(0,H_1)-\Xi_i^{n_0}(1,H_1)}, \\ 1 & \text{otherwise.} \end{cases} \tag{2.82}$$

Next, we focus on the derivation of the fusion rule $\gamma_{FC}^{n_0}(\boldsymbol{u})$ that, at time $n = n_0$, solves (2.71) under the PBPO methodology. To do that, it suffices to realize that, when all the local detectors have already been designed, the joint conditional density $p_{\boldsymbol{U}^{1:\tau}|H}(\boldsymbol{u}^{1:\tau}|H_k)$ can be determined for any $k \in \{0,1\}$. Hence, the fusion center is faced with a classical centralized sequential detection problem that, from (2.71), can be casted as the subsequent optimization problem

$$\begin{aligned} \boldsymbol{\gamma}_{FC}^*(\tau) &= \underset{\boldsymbol{\gamma}_{FC}(\tau),\tau>0}{\operatorname{argmin}}\; E\left\{J(\gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau}),H)\right\} \\ &= \underset{\boldsymbol{\gamma}_{FC}(\tau),\tau>0}{\operatorname{argmin}}\; \left\{E\left\{J(\gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau}),H_0)\right\}\,(1-P(H_1))\right. \\ &\qquad \left. + E\left\{J(\gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau}),H_1)\right\}\,P(H_1)\right\}. \end{aligned} \tag{2.83}$$

with

$$\boldsymbol{\gamma}_{FC}(\tau) = \left[\gamma_{FC}^1(\boldsymbol{U}^1), \gamma_{FC}^2(\boldsymbol{U}^{1:2}), \ldots, \gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau})\right]^T. \tag{2.84}$$

Taking into account the definition of $J(\gamma_{FC}^{\tau}(\boldsymbol{U}^{1:\tau}),H)$ given in (2.73) we can initially check that, once the local detectors have been fixed, the objective function of the previous optimization problem basically depends on the

stopping time τ as well as the value of the global decision, U_{FC}, at that time. This implies that the solution of (2.83) does not only need a fusion rule that gives a global decision from the local decisions received by the fusion center up to the stopping time. Actually, in addition to that fusion rule, the solution of (2.83) also needs a stopping rule that, based on the local decisions received up to each time instant $n \leq \tau$, indicates what is the optimal stopping time τ, i.e. the stopping time associated with the sequential fusion rule $\boldsymbol{\gamma}_{FC}^{*}(\tau)$ minimizing (2.83).

From the classical detection theory associated with the sequential test, it can be straightforwardly checked that, for an arbitrary number of sets of ℓ independent local decisions, at time $n_0 \leq \tau$ the optimal fusion test can be expressed as

$$\gamma_{FC}(\boldsymbol{u}^{1:n_0}) = \begin{cases} 0 & \text{if } \Lambda(\boldsymbol{u}^{n_0}) + \Lambda(\boldsymbol{u}^{1:n_0-1}) \leq \log\left(\underline{P}_L\right), \\ 1 & \text{if } \Lambda(\boldsymbol{u}^{n_0}) + \Lambda(\boldsymbol{u}^{1:n_0-1}) \geq \log\left(\overline{P}_U\right) \end{cases} \tag{2.85}$$

with the thresholds $\underline{P}_L \in (0,1)$ and $\overline{P}_U \in (0,1)$, and with

$$\Lambda(\boldsymbol{u}^n) = \ln\left(\frac{p_{\boldsymbol{U}^n|H}(\boldsymbol{u}^n|H_1)}{p_{\boldsymbol{U}^n|H}(\boldsymbol{u}^n|H_0)}\right) = \sum_{i=1}^{\ell} \ln\left(\frac{p_{U_i^n|H}(u_i^n|H_1)}{p_{U_i^n|H}(u_i^n|H_0)}\right), \tag{2.86}$$

Otherwise, the test is continued by taking the set of ℓ local decisions $\boldsymbol{u}^{n_0+1}$.

As it is illustrated in Figure 2.2, the Bayes sequential test, also called *Sequential Probability Ratio Test* (SPRT), takes samples until the accumulated sum of the log-likelihoods

$$\Lambda(\boldsymbol{u}^{1:n_0}) = \sum_{n=1}^{n_0} \Lambda(\boldsymbol{u}^n) \tag{2.87}$$

with $n_0 \in \{1, 2, \ldots, \tau\}$, falls outside the interval $\left(\log\left(\underline{P}_L\right), \log\left(\overline{P}_U\right)\right)$. Then, the aforementioned test decides on H_0 or H_1 depending on whether $\Lambda(\boldsymbol{u}^{1:n_0})$ falls bellow the threshold $\log\left(\underline{P}_L\right)$ or above the threshold $\log\left(\overline{P}_L\right)$ respectively. Therefore, due to the fact that the derived fusion rule converges almost surely to 1 under H_1 and to 0 under H_0, any level of performance can be achieved if we provide the right thresholds $\log\left(\underline{P}_L\right)$ and $\log\left(\overline{P}_U\right)$. Precisely, this calculation is the second aspect we have to deal with when implementing the optimal sequential test provided in (2.85) and (2.86). Unfortunately, unlike the computation of the posterior $P(H_1|\boldsymbol{u}^{1:n})$, an exact

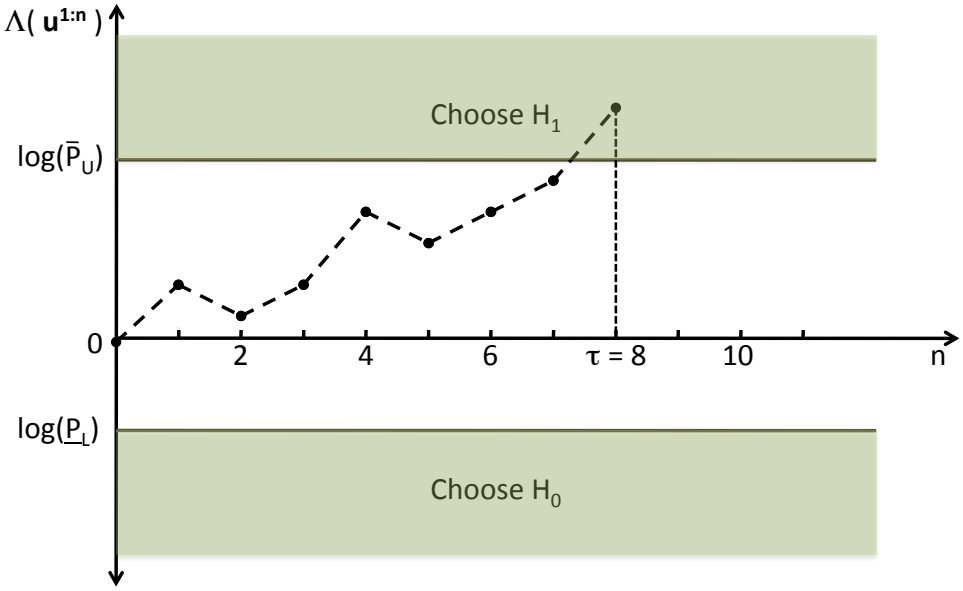


Figure 2.2: *Depiction of a realization of a Bayes sequential test with stopping time* $\tau = 8$.

calculation of $\log(\underline{P}_L)$ and $\log(\overline{P}_L)$ can only be carried out in some special cases. Typically, they are selected according to the well-known *Wald's approximations* [Wald, 1947]. Considering that, at the decision stage, the log-likelihood ratio $\Lambda(\boldsymbol{u}^{1:\tau})$ is likely to exceed the threshold only by a small amount, these approximations establish that the thresholds appearing in (2.85) can be approximated by

$$\log(\underline{P}_L) \approx \log\left(\frac{P_M}{1-P_{FA}}\right) \tag{2.88}$$

and

$$\log(\overline{P}_U) \approx \log\left(\frac{1-P_M}{P_{FA}}\right). \tag{2.89}$$

At this point, we already know that the local processing scheme and the fusion rule provided in (2.82) and (2.85) constitute the PBPO solution of (2.71) when the sensor observations are conditionally independent and when the local decision $u_i^{n_0}$ only depends on $y_i^{n_0}$. As it happened under

previous formulations, we can see that the thresholds appearing in those PBPO rules are coupled. However, unlike the fixed-sample-size detection problems, a different set of thresholds has to be calculated at each time instant $n_0 \leq \tau$. This highlights the extra computational burden associated with sequential detection schemes. Nevertheless, recalling the Wald-Wolfowitz theorem [Poor, 1994, Section III.D] we do not have to forget that, at the expense of that extra computational cost, the average sample size of a distributed sequential scheme is no larger than the sample size of a fixed-sample-size strategy with the same performance.

In many occasions, due to the extra computational burden associated with the sequential schemes, when deciding whether implementing this kind scheme or a fixed-sample-size strategy it is interesting to know the ratio between the average sample size of both detectors under the same detection performance. This motivates the computation of closed-form expressions of the average sample size associated with each one of the possible schemes. Note that, in fixed-sample-size detection problems, the number of observations is a fixed constant, and therefore, the average sample size is equal to the number of observations required to achieve the desired level of performance. On the contrary, from the fusion rule provided in (2.85) realize that the number of sensor observations required for terminating the test is random. This implies that the computation of a closed-form expression for the average sample size is not so straightforward. In publications such as [Tartakovsky, 1998] and [Tartakovsky and Veeravalli, 2005] recent advances in that direction are available for scenarios where, as it happens in the current section, the observations might be non-identically distributed given each one of the two possible hypothesis. In our setting it can be noted that $\boldsymbol{U}^n$ and $\boldsymbol{U}^{n'}$, with $n, n' \in \{1, 2 \ldots, \tau\}$ and $n \neq n'$, may not be identically distributed depending on the value of the thresholds solving the coupled equations obtained under the PBPO methodology and expressed in (2.82). This notwithstanding, in order to give a first insight into the form of the expected sample size, we will analyze the analytical expression obtained under the assumption that the elements of the sequence $\{\boldsymbol{U}^n\}_{n=1}^{\tau}$ are conditionally i.i.d. given each one of the two possible hypothesis.

Assuming that $\{\boldsymbol{U}^n\}_{n=1}^{\tau}$ is a sequence of i.i.d. random variables and that the test is at the τ-th decision stage, i.e. the decision stage where the fusion center stops taking more sets of ℓ local decisions, supported by [Varshney, 1997, section 2.5] we can straightforwardly check that the

expected sample size under H_k is given by

$$E\{\tau|H_k\} = \frac{E_{\boldsymbol{U}^{1:\tau}|H_k}\{\Lambda(\boldsymbol{u}^{1:\tau})\}}{E_{\boldsymbol{U}^{1}|H_k}\{\Lambda(\boldsymbol{u}^{1})\}} \tag{2.90}$$

for $k \in \{0,1\}$. According to the previous expression, under a specific hypothesis the expected sample size , $E\{\tau|H_k\}$, depends on two quantities or two expectations. On the one hand, due to the conditional independence assumption one of them can be expressed as

$$E_{\boldsymbol{U}^{1}|H_k}\left\{\Lambda(\boldsymbol{u}^{1})\right\} = \sum_{i=1}^{\ell} E_{U_i^1|H_k}\left\{\Lambda(u_i^1)\right\} \tag{2.91}$$

where

$$E_{U_i^1|H_k}\left\{\Lambda(u_i^1)\right\} = \begin{cases} (1-P^1_{_{FA}}(i))\log\left(\frac{1-P^1_D(i)}{1-P^1_{_{FA}}(i)}\right) + P^1_{_{FA}}(i)\log\left(\frac{P^1_D(i)}{P^1_{_{FA}}(i)}\right) & \text{if } k=0, \\ (1-P^1_D(i))\log\left(\frac{1-P^1_D(i)}{1-P^1_{_{FA}}(i)}\right) + P^1_D(i)\log\left(\frac{P^1_D(i)}{P^1_{_{FA}}(i)}\right) & \text{if } k=1, \end{cases} \tag{2.92}$$

with

$$P^1_D(i) = P(U_i^1 = 1|H_1) \tag{2.93}$$

and

$$P^1_{_{FA}}(i) = P(U_i^1 = 1|H_0). \tag{2.94}$$

On the other hand, as we considered when deriving *Wald's approximations*, the other expectation appearing in (2.90) can be approximated under the assumption that $\Lambda(\boldsymbol{U}^{1:\tau})$ is equal to $\log(\underline{P}_L)$ or $\log\left(\overline{P}_U\right)$. In this way, knowing that

$$P\left(\Lambda(\boldsymbol{U}^{1:\tau}) \geq \log\left(\overline{P}_U\right)\middle|\, H_0\right) = P_{_{FA}} \tag{2.95}$$

and that

$$P\left(\Lambda(\boldsymbol{U}^{1:\tau}) \leq \log\left(\underline{P}_L\right)\middle|\, H_1\right) = P_M \tag{2.96}$$

we have that, for $k \in \{0, 1\}$, the expected value of $\Lambda(\boldsymbol{U}^{1:\tau})$ under H_k is given by

$$E_{\Lambda(\boldsymbol{U}^{1:\tau})|H_k}\left\{\Lambda(\boldsymbol{u}^{1:\tau})\right\} = \begin{cases} P_{_{FA}} \log\left(\overline{P}_U\right) + (1 - P_{_{FA}}) \log\left(\underline{P}_L\right) & \text{if } k = 0, \\ (1 - P_M) \log\left(\overline{P}_U\right) + P_M \log\left(\underline{P}_L\right) & \text{if } k = 1. \end{cases} \tag{2.97}$$

As it could be expected, from the previous expressions we are now able to realize that, under i.i.d. local decisions, the conditional average sample size of our distributed detection strategy given H_k, i.e. $E\{\tau|H_k\}$, directly depends on the performance of the local detectors as well as the overall detection constraints imposed on the fusion center. Taking into account that all these detection and false alarm probabilities are given by the PBPO solution of (2.71), at the end we have that the average sample size depends on the distribution of the observations under H_0 and H_1. Due to this last fact, concerns regarding the termination of the proposed distributed detection scheme can arise. In order to address this issue and conclude this section, we can use the result shown by [Wald, 1947]. From the proofs undertaken in the aforementioned work we can easily check that, under mild conditions, the derived distributed detection strategy terminates with probability equal to one. However, this does not mean that sample size is bounded. Although the Wald-Wolfowitz theorem states that the average sample size is minimized in a SPRT, we can check that, if the observed data is ambiguous, a distributed scheme based on sequential detection can occasionally run for a large number of samples. This fact may not be practical for many applications and could make us abandon the sequential schemes in favor of fixed-sample-size strategies. Fortunately, this practical difficulty can be overcome quite easily by modifying the SPRT performed at the fusion center. That modification would make the fusion center to stop taking local decisions and undertake a hard (single-threshold) decision after some maximum number of decision stages. Even though this scheme, called *truncated* SPRT, is suboptimal, it retains favorable properties of the SPRTs provided that the truncated point is not taken too small.

2.7 Composite distributed hypothesis testing

The optimal detection strategies designed in the previous sections solve simple hypothesis testing problems where the sensor observations distribute

according to a single distribution, $f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)$, under hypothesis H_k with $k \in \{0,1\}$. However, in many hypothesis testing problems there are many possible distributions that can occur under each one of the hypothesis. This kind of detection problems are known as composite hypothesis tests and might arise in different applications where the statistical properties of the observation model are non-stationary. For instance, in radar detection, the returned signal, if present, has unknown parameters such as its exact time of arrival and its Doppler shift. Additionally, it could happen that the noise and clutter background might be varying across time and from sensor to sensor. Thus, when discriminating between the null and the alternative hypothesis, H_0 and H_1 respectively, we require detection schemes that take into account and handle the statistical nonstationarity of the observation model. Precisely, based on a centralized setting where the sensor observations, $\boldsymbol{y}$, are directly transmitted to the fusion center, the scope of this section is a review of distributed detection schemes implemented by parallel networks that solve different composite hypothesis testing problems. To do that, for the sake of clarity the composite hypothesis testing problems will be presented under a centralized setting where the sensor observations, $\boldsymbol{y}$, are directly transmitted to the fusion center over a bank of parallel access channels. Afterwards, we will comment the steps that have to be undertaken in order to obtain the optimal distributed detection schemes.

To model the most general type of composite hypothesis testing problem, we are going to consider a family of probability density functions, Υ, indexed by a vector of parameters, $\boldsymbol{\theta}$, taking values in a parameter set, $\mathcal{T}$, that represents the set of all possible states of nature. Consequently, we have a family $\left\{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}); \boldsymbol{\theta} \in \mathcal{T}\right\}$ where $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$ equals the joint p.d.f. of the sensor observations when $\boldsymbol{\theta}$ is the true parameter value. At this point, we can easily see that $\mathcal{T} \in \{0,1\}$ in a simple hypothesis testing problem. On the contrary, in a more general setting we might have that the parameter space, $\mathcal{T}$, is equal to the union of two disjoint parameter sets $\mathcal{T}_0$ and $\mathcal{T}_1$, each one of them denoting the ranges of the unknown vector of parameters under each one of the two hypothesis.

Completing the model of the considered problem we need to assume that the unknown vector of parameters is a random variable, $\boldsymbol{\Theta}$, that takes on the value in $\mathcal{T}$ according to a prior distribution $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. If we now focus on the Bayesian formulation of the composite hypothesis testing problem, we firstly have to assign costs to our decisions through a cost function $C(i, \boldsymbol{\theta})$ where $C(i, \boldsymbol{\theta})$ is the cost of deciding H_i when the sensor observations are generated

from $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$ with $i \in \{0,1\}$ and $\boldsymbol{\theta} \in \mathcal{T}$. This way, after assuming that $C(i,\boldsymbol{\theta})$ is non-negative and bounded for all $i \in \{0,1\}$ and $\boldsymbol{\theta} \in \mathcal{T}$, for a given specific fusion rule $\gamma_{FC}(\boldsymbol{y})$ we can also define conditional risks analogous to those of the simple hypothesis test

$$\Re_{\boldsymbol{\theta}}(\gamma_{FC}) = E_{\boldsymbol{Y}|\boldsymbol{\theta}}\left\{C\left(\gamma_{FC}(\boldsymbol{Y}),\boldsymbol{\theta}\right)\right\} \tag{2.98}$$

where $E_{\boldsymbol{Y}|\boldsymbol{\theta}}$ denotes the expectation w.r.t. the joint p.d.f. $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$. Similarly, the Bayes risk associated with the considered fusion rule is given by

$$\Re(\gamma_{FC}) = E\left\{R_{\boldsymbol{\Theta}}(\gamma_{FC})\right\} \tag{2.99}$$

where the expectation is now taken w.r.t. the prior distribution $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. Consequently, it is clear that, under the Bayesian set up, the optimal fusion for the composite hypothesis testing problem rule is the decision rule solving the subsequent optimization problem

$$\gamma_{FC}^{*}(\boldsymbol{y}) = \underset{\gamma_{FC}}{\operatorname{argmin}}\left\{\Re(\gamma_{FC})\right\}. \tag{2.100}$$

Due to the fact that, from the property of iterated expectations $E\{g(X,Y)\} = E_Y\{E_{X|Y}\{g(X,Y)\}\} = E_X\{E_{Y|X}\{g(X,Y)\}\}$,

$$\Re(\gamma_{FC}) = E_{\boldsymbol{Y}}\left\{E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left\{C\left(\gamma_{FC}(\boldsymbol{y}),\boldsymbol{\Theta}\right)\right\}\right\}, \tag{2.101}$$

using (2.100) we have that $\gamma_{FC}^{*}(\boldsymbol{y})$ is the decision rule that, for each $\boldsymbol{y} \in \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_\ell$, minimizes the posterior cost

$$E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left\{C\left(\gamma_{FC}(\boldsymbol{y}),\boldsymbol{\Theta}\right)\right\} \tag{2.102}$$

By noting that $\gamma_{FC}(\boldsymbol{Y}) \in \{0,1\}$ we can easily see that, for the composite hypothesis testing problem, the optimal Bayes rule is expressed as follows

$$\gamma_{FC}^{*}(\boldsymbol{y}) = \begin{cases} 1 & \text{if } E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left\{C\left(1,\boldsymbol{\Theta}\right)\right\} \leq E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left\{C\left(0,\boldsymbol{\Theta}\right)\right\}, \\ 0 & \text{if } E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left\{C\left(1,\boldsymbol{\Theta}\right)\right\} > E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left\{C\left(0,\boldsymbol{\Theta}\right)\right\}. \end{cases} \tag{2.103}$$

An analysis of the previous expression reveals that, for the composite hypothesis testing problem, the optimal Bayes rule, $\gamma_{FC}^{*}(\boldsymbol{y})$ chooses the hypothesis with the least average cost given the sensor observations, $\boldsymbol{y}$. If $\mathcal{T} \in \{0,1\}$, as it can be expected, realize that the optimal fusion rule, $\gamma_{FC}^{*}(\boldsymbol{y})$,

given in (2.103) matches the optimal Bayes rule for a simple hypothesis testing problem. Nonetheless, for many problems of interest the parameter space, $\mathcal{T}$ is usually decomposed into two disjoint sets $\mathcal{T}_0$ and $\mathcal{T}_1$, each one associated with hypothesis H_0 and H_1 respectively. In this case, if the costs are also uniform over the aforementioned sets $\mathcal{T}_0$ and $\mathcal{T}_1$, i.e.

$$C(i, \boldsymbol{\theta}) = C_{i,k}, \qquad \boldsymbol{\theta} \in \mathcal{T}_k \tag{2.104}$$

with $k \in \{0, 1\}$, we can easily see that, under the assumptions $C_{11} < C_{0,1}$ and $C_{00} < C_{10}$, the optimal Bayes rule for the composite hypothesis testing problem reduces to

$$u_{FC} = \gamma^*_{FC}(\boldsymbol{y}) = \frac{P(\boldsymbol{\Theta} \in \mathcal{T}_1 | \boldsymbol{Y} = \boldsymbol{y})}{P(\boldsymbol{\Theta} \in \mathcal{T}_0 | \boldsymbol{Y} = \boldsymbol{y})} \underset{u_{FC}=0}{\overset{u_{FC}=1}{\gtrless}} \frac{C_{10} - C_{00}}{C_{01} - C_{11}}. \tag{2.105}$$

where, for $k \in \{0, 1\}$, $P(\boldsymbol{\Theta} \in \mathcal{T}_k | \boldsymbol{Y} = \boldsymbol{y})$ denotes the conditional probability that $\boldsymbol{\Theta}$ belongs to $\mathcal{T}_k$ given that $\boldsymbol{Y} = \boldsymbol{y}$. Equivalently, if we take into account Bayes' formula

$$P(\boldsymbol{\Theta} \in \mathcal{T}_k | \boldsymbol{Y} = \boldsymbol{y}) = \frac{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y} | \boldsymbol{\theta} \in \mathcal{T}_k)\, P(\boldsymbol{\Theta} \in \mathcal{T}_k)}{f_{\boldsymbol{Y}}(\boldsymbol{y})} \tag{2.106}$$

where

$$f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y} | \boldsymbol{\theta} \in \mathcal{T}_k) = \int_{\mathcal{T}} f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y} | \boldsymbol{\theta})\, f_{\boldsymbol{\Theta}|H_k}(\boldsymbol{\theta} | H_k) \mathrm{d}\boldsymbol{\theta} \tag{2.107}$$

and

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{k=0}^{1} f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y} | \boldsymbol{\theta} \in \mathcal{T}_k)\, P(\boldsymbol{\Theta} \in \mathcal{T}_k) \tag{2.108}$$

with

$$f_{\boldsymbol{\Theta}|H_k}(\boldsymbol{\theta} | H_k) \triangleq \begin{cases} 0 & \text{if } \boldsymbol{\theta} \notin \mathcal{T}_k, \\ \frac{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{P(\boldsymbol{\Theta} \in \mathcal{T}_k)} & \text{if } \boldsymbol{\theta} \in \mathcal{T}_k \end{cases} \tag{2.109}$$

and

$$P(\boldsymbol{\Theta} \in \mathcal{T}_j) = \int_{\mathcal{T}_k} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{2.110}$$

for $k \in \{0,1\}$, the optimal Bayes rule, $\gamma^*_{FC}(\boldsymbol{y})$, provided in (2.105) can be rewritten as

$$u_{FC} = \gamma^*_{FC}(\boldsymbol{y}) = \frac{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta} \in \mathcal{T}_1)}{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta} \in \mathcal{T}_0)} \underset{u_{FC}=0}{\overset{u_{FC}=1}{\gtrless}} \frac{P(\boldsymbol{\Theta} \in \mathcal{T}_0)\,(C_{10} - C_{00})}{P(\boldsymbol{\Theta} \in \mathcal{T}_1)\,(C_{01} - C_{11})}. \tag{2.111}$$

At this point, from the previous derivation we can easily see that, under any specific fixed-sample-size formulation such as the Bayesian or the Neyman-Pearson set ups, the optimal fusion rule for composite hypothesis testing problems has the same form as the one associated with the corresponding simple hypothesis test. The only difference is that the optimal fusion rules are based on the conditional joint probability density functions $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta} \in \mathcal{T}_k)$ provided in (2.107) for $k \in \{0,1\}$. Consequently, replacing $f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)$ by $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta} \in \mathcal{T}_k)$ with $k \in \{0,1\}$ in the development of the preceding sections, we can straightforwardly design the optimal strategy implemented by a distributed detection system solving a composite hypothesis testing problem under a fixed-sample-size formulation.

Looking at (2.107)-(2.110) we can note that, for $k \in \{0,1\}$, the probability density functions $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta} \in \mathcal{T}_k)$, with $k \in \{0,1\}$ depend on the conditional p.d.f. $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$ and the prior distribution $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. This entails that we require the knowledge of the two aforementioned distributions when solving composite hypothesis testing problems with a fixed sample size. However, there might be composite hypothesis testing problems in which we do not have a prior distribution for the vector of parameters. In that case, the development of hypothesis tests that satisfy precise analytical definitions of optimality is very often an illusive task. One way of defining optimality in such problems is a generalization of the Neyman-Pearson criterion described in Section 2.5. In particular, assume that, as before, the parameter space, $\mathcal{T}$, can be decomposed into two disjoint sets $\mathcal{T}_0$ and $\mathcal{T}_1$. Then, an ideal decision rule $\gamma_{FC}(\boldsymbol{y})$ would be the one that, for every $\boldsymbol{\theta} \in \mathcal{T}_1$, maximizes the conditional probability of detection given $\boldsymbol{\theta}$, $P_D(\boldsymbol{\theta})$ s.t. the following constraint on the corresponding conditional probability of false alarm

$$P_{_{FA}}(\boldsymbol{\theta}) \leq \alpha \in (0,1) \tag{2.112}$$

with $\boldsymbol{\theta} \in \mathcal{T}_0$. Such a test is known as an uniformly most powerful (UMP) test of level α.

Unfortunately, although UMP tests are very desirable, they exist only under very special circumstances. To see this, consider the situation in which

$\mathcal{T}_1 = \mathcal{T} \backslash \boldsymbol{\theta}_0$ and the null hypothesis is simple with $\mathcal{T}_0 = \boldsymbol{\theta}_0$. Assuming that the conditional p.d.f. of the sensor observations given $\boldsymbol{\theta}$ is given by $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \in \mathcal{T}$, from the Neyman-Pearson's Lemma detailed in [Blahut, 1987] we can verify that, for each $\boldsymbol{\theta} \in \mathcal{T}_1$, the unique and most powerful α-level test for the detection problem

$$\begin{array}{ll} H_0: & \boldsymbol{y} \sim f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0) \\ H_1: & \boldsymbol{y} \sim f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}) \text{ with } \boldsymbol{\theta} \in \mathcal{T}_1 \end{array} \tag{2.113}$$

has the following acceptance region for H_1

$$\mathcal{V}_\ell^{(1)}(\boldsymbol{\theta}, \eta) = \left\{ \boldsymbol{y} \in \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_\ell : f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}) > \eta \, f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0) \right\} \tag{2.114}$$

with η equal to the smallest number such that

$$P_{_{FA}}(\boldsymbol{\theta}) = \int_{\mathcal{V}_\ell^{(1)}(\boldsymbol{\theta},\eta)} f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{y} \leq \alpha \in (0,1). \tag{2.115}$$

If we choose two elements $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$ of $\mathcal{T}_1$, this implies that, unless $\mathcal{V}_\ell^{(1)}(\boldsymbol{\theta}', \eta) = \mathcal{V}_\ell^{(1)}(\boldsymbol{\theta}'', \eta)$, the test with critical region $\mathcal{V}_\ell^{(1)}(\boldsymbol{\theta}', \eta)$ will have smaller power in testing H_0 versus $\boldsymbol{Y} \sim f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}'')$ than does the test with critical region $\mathcal{V}_\ell^{(1)}(\boldsymbol{\theta}'', \eta)$ and viceversa. Thus, it can be concluded that an UMP test exists for (2.113) if and only if the acceptance region for H_1, $\mathcal{V}_\ell^{(1)}(\boldsymbol{\theta}, \tau)$, is the same for all $\boldsymbol{\theta} \in \mathcal{T}_1$. Note that this condition does not happen in many detection scenarios. An example of this can be found in [Poor, 1994, section II.E].

In the absence of an UMP test for a specific composite hypothesis testing problem where the prior distribution $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is not known, we have to resort to suboptimal detection schemes. To see a review of these scheme see [Blum et al., 1997] and the references therein. Among them, we will focus on two of the most used techniques that, despite being suboptimal, can achieve optimality under some specific scenarios. In particular, we will cover the distributed detection schemes based on the subsequent suboptimal criteria:

- Generalized likelihood ratio test (GLRT).

- Locally optimum test.

2.7.1 Generalized likelihood ratio test

From the Neyman-Pearson's Lemma recall that, for each $\boldsymbol{\theta}_0 \in \mathcal{T}_0$ and $\boldsymbol{\theta}_1 \in \mathcal{T}_1$, the unique and most powerful α-level test for the composite hypothesis testing problem

$$\begin{array}{ll} H_0: & \boldsymbol{y} \sim f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}) \text{ with } \boldsymbol{\theta} \in \mathcal{T}_0 \\ H_1: & \boldsymbol{y} \sim f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}) \text{ with } \boldsymbol{\theta} \in \mathcal{T}_1 \end{array} \tag{2.116}$$

is given by

$$\gamma_{FC}^{*}(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \frac{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0)} > \eta, \\ \gamma & \text{if } \frac{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0)} = \eta, \\ 0 & \text{otherwise,} \end{cases} \tag{2.117}$$

where

$$\mathcal{V}_{\ell}^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \eta) = \left\{ \boldsymbol{y} : \frac{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0)} > \eta \right\}, \tag{2.118}$$

η is the smallest number such that

$$\varsigma = \sum_{\boldsymbol{u} \in \mathcal{V}_{\ell}^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \eta)} f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0) \leq \alpha \in (0,1) \tag{2.119}$$

and $\gamma \in \{0,1\}$ is a randomization function that takes on the value 1 with the subsequent probability

$$\nu = \begin{cases} 1 & \text{if } \vartheta = 0, \\ \frac{\alpha - \varsigma}{\vartheta} & \text{otherwise} \end{cases} \tag{2.120}$$

with

$$\vartheta = P\left(\frac{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_1)}{f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0)} = \eta \middle| H_0 \right). \tag{2.121}$$

Consequently, when solving the composite hypothesis testing problem provided in (2.116) no fusion rule can be better than a hypothetical test in which the fusion center first estimates $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ and then applies the randomized likelihood ratio test given in (2.117). Because of this, unless an UMP test can be derived, and therefore, the likelihood ratio test provided in (2.117) can be completely characterized without the knowledge of $\boldsymbol{\theta}_k$ with

$k \in \{0,1\}$, a logical procedure could be estimate $\boldsymbol{\theta}_1$ assuming that H_1 is true, then estimate $\boldsymbol{\theta}_0$ assuming that H_0 is true, and use these estimates in (2.117) as if they were correct. Precisely, if maximum likelihood estimates are used for $\boldsymbol{\theta}_k$ with $k \in \{0,1\}$, the previous procedure yields the generalized likelihood ratio test. Formally, the GLRT for (2.116) is expressed as

$$\gamma_{FC}(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \Lambda_{\text{GLRT}}(\boldsymbol{y}) > \eta, \\ \gamma & \text{if } \Lambda_{\text{GLRT}}(\boldsymbol{y}) = \eta, \\ 0 & \text{otherwise,} \end{cases} \tag{2.122}$$

where

$$\Lambda_{\text{GLRT}}(\boldsymbol{y}) = \frac{\max\limits_{\boldsymbol{\theta} \in \mathcal{T}_1} \left\{ f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_1) \right\}}{\max\limits_{\boldsymbol{\theta} \in \mathcal{T}_0} \left\{ f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_0) \right\}} \tag{2.123}$$

with

$$\mathcal{V}_{\ell}^{(1)}(\eta) = \{\boldsymbol{y} : \Lambda_{\text{GLRT}}(\boldsymbol{y}) > \eta\}, \tag{2.124}$$

η denoting the smallest number such that

$$\varsigma = \sum_{\boldsymbol{u} \in \mathcal{V}_{\ell}^{(1)}(\eta)} \max_{\boldsymbol{\theta} \in \mathcal{T}_0} \left\{ f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}) \right\} \leq \alpha \in (0,1) \tag{2.125}$$

and $\gamma \in \{0,1\}$ equal to a randomization function that takes on the value 1 according to the probability provided in (2.120) with

$$\vartheta = P\left(\Lambda_{\text{GLRT}}(\boldsymbol{y}) = \eta | \, H_0\right). \tag{2.126}$$

Analyzing (2.117) we can check that, from the fusion rule given in (2.122), the GLRT for (2.116) is obtained by replacing

$$f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}_k)$$

with

$$\max_{\boldsymbol{\theta} \in \mathcal{T}_k} \left\{ f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta}) \right\}$$

for all $k \in \{0,1\}$. This entails two aspects. On the one hand, using this approach we can straightforwardly design distributed detection systems that, under a formulation with a fixed sample size, solve composite hypothesis testing problems in which an UMP test does not exist and the prior distribution $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is not known. On the other hand, to conclude this

subsection we can also note that a GLRT is not optimal as long as the estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are not perfect. However, as a result of the consistency property shown by the maximum likelihood estimates under independent sensor observations [Van Trees, 2001, Section 2.4.2], the GLRT is known to perform well in practice and it has well established optimality in an asymptotic regime where the number of observations, ℓ, approaches infinity (see [Zeitouni et al., 1992], [Dembo and Zeitouni, 1998] and [Levy, 2008]). This is the main reason why this approach has been so used in sensor networks formed by a large number of devices and solving a composite hypothesis testing problem.

2.7.2 Locally optimum distributed detection

In many situations of interest, the parameter set $\mathcal{T}$ is of the form $\mathcal{T}_0 = \{\boldsymbol{\theta}_0\}$ and $\mathcal{T}_1 = \mathcal{T} \setminus \{\boldsymbol{\theta}_0\}$; so that we have the subsequent composite hypothesis testing problem

$$\begin{array}{ll} H_0: & \boldsymbol{y} \sim f_{\boldsymbol{Y}|\Theta}(\boldsymbol{y}|\boldsymbol{\theta}_0) \\ H_1: & \boldsymbol{y} \sim f_{\boldsymbol{Y}|\Theta}(\boldsymbol{y}|\boldsymbol{\theta}) \text{ with } \theta \in \mathcal{T} \setminus \{\boldsymbol{\theta}_0\}. \end{array} \tag{2.127}$$

This kind of problems usually occurs when, in a noisy environment, the parameters $\boldsymbol{\theta}$ involved in the observation process are known under H_0 and unknown under H_1. From now and on and for the sake of simplicity, as it happens when detecting a known or a random signal whose amplitude is parametrized by the scalar θ, we will consider that $\theta_0 = 0$. For the more general case where the vector of parameters $\boldsymbol{\theta}_0 \neq \mathbf{0}$, see [He and Blum, 2011] and the references therein.

In the above hypothesis test, the detection probability of a strategy Γ will generally depend on the parameter θ, so the probability of detection at the fusion center is denoted as $P_D(\Gamma, \theta)$ for that particular strategy. As we have illustrated along this section, one cannot find an UMP test that will maximize $P_D(\Gamma, \theta)$ for all possible values of θ. In these cases, one can search for a locally optimum scheme [Kassam and Thomas, 1988] that, under $P_{FA}(\Gamma) = P_D(\Gamma, 0) \leq \alpha \in (0, 1)$, maximizes

$$\left. \frac{\partial^m P_D(\Gamma, \theta)}{\mathrm{d}\theta^m} \right|_{\theta=0} \tag{2.128}$$

with m equal to the lowest order for which the previous derivative is not identically zero. Let m be defined in this manner for the rest of this subsection.

Locally optimum strategies tend to outperform GLRT for cases where the unknown parameter θ is close to $\theta_0 = 0$. This occurs because the maximization of the quantity given in (2.128) subject to $P_{_{FA}}(\Gamma) = P_D(\Gamma, 0) \leq \alpha \in (0, 1)$ matches the optimal Neyman-Pearson formulation that, under the previous upper bound on the probability of false alarm, maximizes the m-th order Taylor series of $P_D(\Gamma, \theta)$ about $\theta_0 = 0$. At the same time, from this last fact we can easily verify that the derivation of the locally optimum strategy does not differ a lot from the optimal Neyman-Pearson strategy. In particular, under the PBPO methodology and following a similar development as in Section 2.5, the fusion rule for an input vector of local decisions $\boldsymbol{u}^*$ is given by

$$
\begin{aligned}
u_{FC} &= \gamma_{FC}(\boldsymbol{u}^*) \\
&= P(U_{FC} = 1|\boldsymbol{u}^*) = \begin{cases} 1 & \text{if } \left.\frac{\partial^m P_{\boldsymbol{U}|\Theta}(\boldsymbol{u}^*|\theta)}{\mathrm{d}\theta^m}\right|_{\theta=0} > \lambda_{FC} P_{\boldsymbol{U}|\Theta}(\boldsymbol{u}^*|0), \\ \xi(\boldsymbol{u}^*) & \text{if } \left.\frac{\partial^m P_{\boldsymbol{U}|\Theta}(\boldsymbol{u}^*|\theta)}{\mathrm{d}\theta^m}\right|_{\theta=0} = \lambda_{FC} P_{\boldsymbol{U}|\Theta}(\boldsymbol{u}^*|0), \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
\tag{2.129}
$$

with the threshold, $\lambda_{FC} < \infty$, and the randomization function, $\xi(\boldsymbol{u}^*) \in (0, 1)$, chosen in order to satisfy that $P_{_{FA}}(\Gamma) \leq \alpha$. Likewise, from the results provided in [Blum, 1996] we can see that, under the PBPO methodology and conditionally dependent sensor observations, the locally optimum decision rule for the i-th sensor is given by

$$
u_i = \gamma_i(y_i) = P(U_i = 1|y_i) = \begin{cases} 1 & \text{if } D_{1i}(y_i) \geq \lambda_i D_{0i}(y_i), \\ 0 & \text{if } D_{1i}(y_i) < \lambda_i D_{0i}(y_i) \end{cases} \tag{2.130}
$$

where, for $i \in \{1, 2, \ldots, \ell\}$,

$$
D_{0i}(y_i) = \sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) \ \left[P_{\boldsymbol{U}^i|Y_i,\Theta}(\boldsymbol{u}^i|y_i, 0)\, f_{Y_i|\Theta}(y_i|0)\right] \tag{2.131}
$$

and

$$
D_{1i}(y_i) = \sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i) \ \left(\frac{\partial^m}{\mathrm{d}\theta^m}\left[P_{\boldsymbol{U}^i|Y_i,\Theta}(\boldsymbol{u}^i|y_i, \theta)\, f_{Y_i|\Theta}(y_i|\theta)\right]\right)\Bigg|_{\theta=0} \tag{2.132}
$$

with $\boldsymbol{u}^i$ and $A(\boldsymbol{u}^i)$ defined in (2.22) and (2.27) respectively. Besides assuming that $f_{Y_i|\Theta,H}(y_i|\theta)$ and the p.d.f. of $D_{1i}(y_i)/D_{0i}(y_i)$ do not contain any point masses of probability for any θ where $D_{0i}(y_i)$ and $D_{1i}(y_i)$ are defined, the derivation of the previous decision rule requires that

$$P_{\boldsymbol{U}^i|Y_i,\Theta}(\boldsymbol{u}^i|y_i,\theta)\, f_{Y_i|\Theta}(y_i|\theta)$$

is sufficiently smooth so that we can interchange the integral and the derivative in

$$\begin{aligned}&\int_{\mathcal{Y}_i} D_{1i}(y_i)\gamma_i(y_i)\mathrm{d}y_i \\ &\quad = \left[\frac{\partial^m}{\mathrm{d}\theta^m}\left(\int_{\mathcal{Y}_i}\gamma_i(y_i)\sum_{\boldsymbol{u}^i}\left\{\left[P_{\boldsymbol{U}^i|Y_i,\Theta}(\boldsymbol{u}^i|y_i,\theta)\, f_{Y_i|\Theta}(y_i|\theta)\right]\; A(\boldsymbol{u}^i)\right\}\mathrm{d}y_i\right)\right]\Bigg|_{\theta=0}\end{aligned} \tag{2.133}$$

for any γ_i. Apart from these assumptions, as it happens for the fusion rule, the derivation of the sensor decision rule given in (2.130) is analogous to the one performed under the PBPO methodology and under the Neyman-Pearson formulation. Therefore, for the sake of brevity it has also been omitted.

To conclude we show one more connection between the Neyman-Pearson formulation and the locally optimum set ups. Under conditionally independent sensor observations, we can check that the locally optimum sensor decision rule given in (2.130)-(2.132) can be simplified to likelihood ratio tests with the subsequent form

$$u_i = \gamma_i(y_i) = P(U_i = 1|y_i) = \begin{cases} 1 & \text{if } L_i(y_i) \geq \tilde{\lambda}_i, \\ 0 & \text{if } L_i(y_i) < \tilde{\lambda}_i \end{cases} \tag{2.134}$$

where, for $i \in \{1, 2, \ldots, \ell\}$,

$$L_i(y_i) = \frac{\left.\frac{\partial^m f_{Y_i|\Theta}(y_i|\theta)}{\mathrm{d}\theta^m}\right|_{\theta=0}}{f_{Y_i|\Theta}(y_i|0)} \tag{2.135}$$

and

$$\tilde{\lambda}_i = \lambda_i - \frac{\left.\sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i)\; \frac{\partial^m P_{\boldsymbol{U}^i|Y_i,\Theta}(\boldsymbol{u}^i|y_i,\theta)}{\mathrm{d}\theta^m}\right|_{\theta=0}}{\sum_{\boldsymbol{u}^i} A(\boldsymbol{u}^i)\, P_{\boldsymbol{U}^i|Y_i,\Theta}(\boldsymbol{u}^i|y_i,0)} \tag{2.136}$$

with $\boldsymbol{u}^i$ and $A(\boldsymbol{u}^i)$ defined in (2.22) and (2.27) respectively. Thus, as it happened with other formulations, the conditional independence among the sensor observations does not imply a reduction in the number of coupled non-linear equations that, provided by (2.129)-(2.132), have to be solved when implementing the locally optimum strategy. Instead, the conditional independence assumption means a considerable a reduction of the computational difficulty that appears when solving the aforementioned coupled equations.

2.8 Performance measures for distributed detection systems

In the previous sections we have undertaken the derivation of optimal schemes that have to be implemented by two-stage distributed detection systems under different formulations. Recall that, under the different formulations, optimality has usually implied the minimization of some cost function related to the overall detection performance among other physical and design parameters of the network. If we focus on the parallel fusion network considered in this chapter, we can check that the overall detection performance is given by the different error probabilities at the fusion center, i.e. the probability of misdetection, P_M, and the probability of false alarm, $P_{_{FA}}$, defined in (2.8) and (2.10) respectively. Consequently, given a specific formulation, if we want to analyze and characterize the detection performance of the network as a function of physical and design parameters such as the censoring scheme or the sensor spacing, we have to evaluate or obtain analytically tractable expressions adopted by those probabilities under the considered formulation. In this section, according to the research topic of this thesis dissertation, we will address that evaluation or derivation when, under the Bayesian or the Neyman-Pearson formulation, the considered parallel network implements the optimal strategy solving a binary and simple hypothesis testing problem with a fixed sample size. Nevertheless, from the results of the previous section it should be noted that this task will require a similar approach when we consider composite hypothesis testing problems with a fixed sample size.

Assuming that, when the number of local detectors, ℓ, is very large, there

exists a value of τ such that

$$\sum_{\boldsymbol{u}\in\mathcal{V}_\ell^1(\tau)} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0) \approx \alpha \in (0,1) \tag{2.137}$$

with

$$\mathcal{V}_\ell^1(\tau) = \left\{ \boldsymbol{u} : \frac{\Lambda(\boldsymbol{u})}{\ell} > \tau_\ell \right\}, \tag{2.138}$$

$$\Lambda(\boldsymbol{u}) = \ln\left(\frac{P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}{P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}\right) \tag{2.139}$$

and

$$\tau_\ell = \frac{\tau}{\ell}. \tag{2.140}$$

At this point, we can recall that, for both the Bayesian and Neyman-Pearson formulations, the probability of misdetection at the fusion center is

$$P_M = 1 - \sum_{\boldsymbol{u}\in\mathcal{V}_\ell^1(\tau)} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1) = P\left(\Lambda(\boldsymbol{U}) \leq \tau \middle| H_1\right) = P\left(\frac{\Lambda(\boldsymbol{U})}{\ell} \leq \tau_\ell \middle| H_1\right), \tag{2.141}$$

and the corresponding probability of false alarm is obtained as

$$P_{_{FA}} = \sum_{\boldsymbol{u}\in\mathcal{V}_\ell^1(\tau)} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0) = P\left(\Lambda(\boldsymbol{U}) > \tau \middle| H_0\right) = P\left(\frac{\Lambda(\boldsymbol{U})}{\ell} > \tau_\ell \middle| H_0\right) \tag{2.142}$$

where τ is a bounded constant chosen in a different way under the Bayesian set up and under the Neyman-Pearson formulation. From (2.141) and (2.142) and the optimal strategies derived under the two considered formulations, we can check that the derivation of closed-form expressions for P_M and $P_{_{FA}}$ is not analytical feasible when the number of local detectors ℓ is large. What is more, the evaluation of the two previous probabilities is combinatorial in ℓ, and therefore, it involves a high computational cost when we are considering sensor networks with a moderate size. For instance, note that the summations appearing in (2.141) and (2.142) have 2^{100} terms if the considered

sensor network is formed by $\ell = 100$ local detectors. Because of this, if we take into account that the global performance of parallel network is given by

$$P_e = P(H_0)P_{_{FA}} + (1 - P(H_0))\, P_M \tag{2.143}$$

under the Bayesian set up, and by P_M when $P_{_{FA}} \leq \alpha$ under the Neyman-Pearson formulation, some performance measures related to the aforementioned probabilities are required during the design or analysis of the network as a function of its physical and design parameters. In particular, we will focus on the subsequent methodologies:

- Normal approximation.
- Saddlepoint approximation.
- Error exponents.

2.8.1 Normal approximation

Possibly, the approximation covered in this subsection is the most simple but, at the same time, it is the most limited regarding the number of scenarios where it can be applied. Assuming that the sensor observations are conditionally i.i.d. and that the local detectors use the same binary quantization rule, it is based on the application of the central limit theorem in order to approximate the p.m.f. of $\Lambda(\boldsymbol{U})$ under H_0 and H_1. In particular, the central limit theorem [Papoulis and Pillai, 2002, Section 8.4] states that, under certain general conditions, the distribution function, $F(s)$, of a sum of ℓ i.i.d. random variables,

$$S = \sum_{i=1}^{\ell} X_i, \tag{2.144}$$

approaches the normal distribution with the same mean and variance as ℓ increases.

Take into account that, when the local decisions are i.i.d. under a specific hypothesis,

$$\frac{1}{\ell}\Lambda(\boldsymbol{U}) = \frac{1}{\ell}\sum_{i=1}^{\ell} \Lambda_1(u_i) \tag{2.145}$$

with

$$\Lambda_1(U_i) = \ln\left(\frac{P_{U_i|H}(u_i|H_1)}{P_{U_i|H}(u_i|H_0)}\right). \tag{2.146}$$

Applying the central limit theorem, this entails that, if ℓ is sufficiently large, the distribution of $\frac{1}{\ell}\Lambda(\boldsymbol{U})$ under H_0 can be approximated by a normal distribution whose mean is equal to

$$\begin{aligned} E_{H_0}\left\{\Lambda_1(U_1)\right\} &= -D(B(P_{_{FA}}(1))||B(P_D(1))) \\ &= -\left(1-P_{_{FA}}(1)\right)\ln\left(\frac{1-P_{_{FA}}(1)}{1-P_D(1)}\right) - P_{_{FA}}(1)\ln\left(\frac{P_{_{FA}}(1)}{P_D(1)}\right), \end{aligned} \tag{2.147}$$

and whose variance is equal to

$$Var_{H_0}\left\{\Lambda_1(U_i)\right\} = E_{H_0}\left\{\Lambda_1(U_1)^2\right\} - D^2(B(P_{_{FA}}(1))||B(P_D(1))) \tag{2.148}$$

where

$$E_{H_0}\left\{\Lambda_1(U_1)^2\right\} = \left(1-P_{_{FA}}(1)\right)\ln^2\left(\frac{1-P_D(1)}{1-P_{_{FA}}(1)}\right) + P_{_{FA}}(1)\ln^2\left(\frac{P_D(1)}{P_{_{FA}}(1)}\right) \tag{2.149}$$

and where $D(B(P_{_{FA}}(1))||B(P_D(1)))$ denotes the Kullback-Leibler divergence between $B(P_{_{FA}}(1))$ and $B(P_D(1))$ with $B(p)$ equal to a Bernoulli random variable whose probability of success equals p. Analogously, if ℓ is sufficiently large the distribution of $\frac{1}{\ell}\Lambda(\boldsymbol{U})$ under H_1 can be approximated by a normal distribution whose mean is equal to

$$\begin{aligned} E_{H_1}\left\{\Lambda_1(U_1)\right\} &= D(B(P_D(1))||B(P_{_{FA}}(1))) \\ &= \left(1-P_D(1)\right)\ln\left(\frac{1-P_D(1)}{1-P_{_{FA}}(1)}\right) + P_D(1)\ln\left(\frac{P_D(1)}{P_{_{FA}}(1)}\right), \end{aligned} \tag{2.150}$$

and whose variance is equal to

$$Var_{H_1}\left\{\Lambda_1(U_i)\right\} = E_{H_1}\left\{\Lambda_1(U_1)^2\right\} - D^2(B(P_D(1))||B(P_{_{FA}}(1))) \tag{2.151}$$

where

$$E_{H_1}\left\{\Lambda_1(U_1)^2\right\} = \left(1-P_D(1)\right)\ln^2\left(\frac{1-P_D(1)}{1-P_{_{FA}}(1)}\right) + P_D(1)\ln^2\left(\frac{P_D(1)}{P_{_{FA}}(1)}\right). \tag{2.152}$$

Consequently, from the central limit theorem the probability of misdetection and the probability of false alarm at the fusion center can be approximated as

$$P_M \approx 1 - Q\left(\frac{\tau_\ell - E_{H_1}\{\Lambda_1(U_1)\}}{\sqrt{Var_{H_1}\{\Lambda_1(U_1)\}}}\sqrt{\ell}\right) \tag{2.153}$$

and

$$P_{_{FA}} \approx Q\left(\frac{\tau_\ell - E_{H_0}\{\Lambda_1(U_1)\}}{\sqrt{Var_{H_0}\{\Lambda_1(U_1)\}}}\sqrt{\ell}\right) \tag{2.154}$$

respectively, where

$$Q(x_0) = \int_{x_0}^{\infty} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}}\, \mathrm{d}x. \tag{2.155}$$

By realizing that the local decisions probabilities can be straightforwardly computed, we can now check that the normal approximations given in (2.153) and (2.154) are easy to compute for a parallel network with different sizes. In fact, under the Bayesian and the Neyman-Pearson formulation this allows us the characterization and the design of the global detection performance w.r.t. different features of a large wireless network. However, first of all we do not have to forget that, with the normal approximations, that design and analysis can only be addressed in scenarios where the local decisions are i.i.d. under each one of the two possible hypothesis. This prevents the application of the normal approximation to many scenarios where the local decisions are not identically distributed although they are conditionally independent. Note that if the local detectors are not identical, the local decisions are not going to be identically distributed although the sensor observations are i.i.d. under both hypothesis. Unfortunately, this is not the only drawback associated with the normal approximation. Although we are considering a scenario where the binary local decisions are i.i.d., the approximations given in (2.153) and (2.154) may not provide acceptable accuracy, especially when the value of the threshold is in a tail region that is several standard deviations far from the mean of the decision variable, $\Lambda(\boldsymbol{U})$. If we note that, for all $k \in \{0, 1\}$,

$$E_{H_k}\left\{\frac{\Lambda(\boldsymbol{U})}{\ell}\right\} = E_{H_k}\{\Lambda_1(U_1)\} \tag{2.156}$$

and

$$Var_{H_k}\left\{\frac{\Lambda(\boldsymbol{U})}{\ell}\right\} = \frac{Var_{H_k}\left\{\Lambda_1(U_1)\right\}}{\ell}, \tag{2.157}$$

we can see that the larger is ℓ, the larger is the distance between $E_{H_k}\left\{\frac{\Lambda(\boldsymbol{U})}{\ell}\right\}$ and τ_ℓ in terms of $Var_{H_k}\left\{\frac{\Lambda(\boldsymbol{U})}{\ell}\right\}$. Despite the asymptotic convergence of the central limit theorem w.r.t. ℓ, this entails that, when ℓ is large, the normal approximation of P_M and $P_{_{FA}}$ may not provide adequate accuracy. In those cases other approaches as the ones presented in the subsequent sections have to be taken.

2.8.2 Saddlepoint approximation

Under the assumption of i.i.d. local decisions, the saddlepoint approximation allows the analysis and design of networks of arbitrary size by considering a large deviation theory based approximation to the error probabilities provided in (2.141) and (2.142). However, to overcome the lack of accuracy shown by the normal approximation when the fusion threshold is several standard deviations far from the mean of $\Lambda(\boldsymbol{U})$, this technique applies the normal approximation after relating the original density of $\Lambda(\boldsymbol{U})$ to a new so-called tilted density centered at the value of the fusion threshold. Due to the fact that the saddlepoint approximation for the tail probabilities of sum of continuos random variables is still valid when, as it happens in our setting, the random variables are discrete (see [Booth et al., 1994]), we will highlight results from the continuous saddlepoint theory that are relevant to the problem at hand.

In saddlepoint techniques, the original density of $\Lambda(\boldsymbol{U})$ under H_k, denoted as $f_\Lambda(s|H_k) = f_{\Lambda(\boldsymbol{U})|H}(s|H_k)$ with $k \in \{0,1\}$, is embedded in a conjugate exponential family

$$f_{\widehat{\Lambda}}(s,\hat{\theta}_k|H_k) = e^{s\hat{\theta}_k - M_k(\hat{\theta}_k)}\, f_\Lambda(s|H_k) \tag{2.158}$$

where $\widehat{\Lambda}(\boldsymbol{U})$ is the transformed random variable, $\hat{\theta}_k$ is a fixed parameter, and $M_k(\theta)$ is the cumulant generating function (CGF) of $\Lambda(\boldsymbol{U})$, which, under the assumption of i.i.d. local decisions, is given by

$$M_k(\theta) = \ell\, \ln\left(G_k(\theta)\right) \tag{2.159}$$

with $G_k(\theta)$ equal to the moment generating function (MGF) of $\Lambda_1(U_1)$ under H_k defined by

$$\begin{aligned} G_k(\theta) = &(1 - P(U_1 = 1|H_k)) \left[\frac{P(U_1 = 0|H_1)}{P(U_1 = 0|H_0)}\right]^\theta \\ &+ P(U_1 = 1|H_k) \left[\frac{P(U_1 = 1|H_1)}{P(U_1 = 1|H_0)}\right]^\theta. \end{aligned} \tag{2.160}$$

The transformation performed in (2.158) is often called exponential tilting. Its objective is to shift the mean of the transformed random variable, $\widehat{\Lambda}(\boldsymbol{U})$, so that it becomes close to the evaluation point s. This way, we are able to ensure that the normal approximation can be applied safely when estimating the density of $\widehat{\Lambda}(\boldsymbol{U})$ at s. In particular, note that the CGF of $\widehat{\Lambda}(\boldsymbol{U})$ under H_k is $M_{\widehat{\Lambda},k}(\theta) = M_k(\theta+\hat{\theta}_k)+M_k(\hat{\theta}_k)$. Taking the first derivative w.r.t. θ and evaluating it at $\theta = 0$, we get the mean of the transformed random variable, i.e.

$$E_{H_k}\left\{\widehat{\Lambda}(\boldsymbol{U})\right\} = \frac{\partial M_k(\hat{\theta}_k)}{\mathrm{d}\theta} \tag{2.161}$$

where, for $k \in \{0, 1\}$,

$$\frac{\partial M_k(\theta)}{\mathrm{d}\theta} = \ell W_{1,k}(\theta)/G_k(\theta) \tag{2.162}$$

with

$$\begin{aligned} W_{j,k}(\theta) = &(1 - P(U_1 = 1|H_k)) \left[\frac{P(U_1 = 0|H_1)}{P(U_1 = 0|H_0)}\right]^\theta (\Lambda_1(0))^j \\ &+ P(U_1 = 1|H_k) \left[\frac{P(U_1 = 1|H_1)}{P(U_1 = 1|H_0)}\right]^\theta (\Lambda_1(1))^j \end{aligned} \tag{2.163}$$

and $\Lambda_1(U)$ defined in (2.146) for $U \in \{0, 1\}$. Consequently, the mean of the transformed random variable, $\widehat{\Lambda}(\boldsymbol{U})$ can be made precisely equal to s if we find $\hat{\theta}_k$ such that

$$\frac{\partial M_k(\hat{\theta}_k)}{\mathrm{d}\theta} = s. \tag{2.164}$$

At the same time, the variance of $\widehat{\Lambda}(\boldsymbol{U})$ is obtained from the second cumulant, which is equal to the second derivative of $M_k(\theta)$ at $\theta = 0$, i.e.

$$Var_{H_k}\left\{\widehat{\Lambda}(\boldsymbol{U})\right\} = \frac{\partial^2 M_k(\hat{\theta}_k)}{\mathrm{d}\theta^2} \tag{2.165}$$

where

$$\frac{\partial^2 M_k(\theta)}{\mathrm{d}\theta^2} = \frac{\ell\left[G_k(\theta)W_{2,k}(\theta) - W_{1,k}^2(\theta)\right]}{G_k^2(\theta)}. \tag{2.166}$$

Since we set s such that it satisfies (2.164), we can accurately approximate the density of $\widehat{\Lambda}(\boldsymbol{U})$ at s using the normal approximation as follows:

$$f_{\widehat{\Lambda}}(s, \hat{\theta}_k|H_k) \approx \frac{1}{\sqrt{2\pi\frac{\partial^2 M_k(\hat{\theta}_k(s))}{\mathrm{d}\theta^2}}} \tag{2.167}$$

where $\hat{\theta}_k(s)$ is the saddlepoint at s, which equals the value of $\hat{\theta}_k$ solving (2.164). Dividing by $e^{s\,\hat{\theta}_k(s) - M_k(\hat{\theta}_k(s))}$, and therefore, carrying out the reverse transformation in (2.158), the previous expression yields the so-called saddlepoint approximation of the density of the original variable $\Lambda(\boldsymbol{U})$ at s

$$f_\Lambda(s|H_k) \approx \tilde{f}_\Lambda(s|H_k) = \frac{e^{M_k(\hat{\theta}_k(s)) - s\,\hat{\theta}_k(s)}}{\sqrt{2\pi\frac{\partial^2 M_k(\hat{\theta}_k(s))}{\mathrm{d}\theta^2}}} = \frac{\hat{\theta}_k(s)}{(2\pi)^{1/2}\, g_k\left[\hat{\theta}_k(s)\right]}\, e^{-\frac{r_k^2[\hat{\theta}_k(s)]}{2}} \tag{2.168}$$

where

$$g_k(\theta) = \theta\sqrt{\frac{\partial^2 M_k(\theta)}{\mathrm{d}\theta^2}} \tag{2.169}$$

and

$$r_k(\theta) = \mathrm{sgn}(\theta)\sqrt{2\left[\theta\,\frac{\partial M_k(\theta)}{\mathrm{d}\theta} - M_k(\theta)\right]} \tag{2.170}$$

with $\mathrm{sgn}(\cdot)$ denoting the sign operator.

Next, from the saddlepoint approximation provided in (2.168) we will derive the tail probabilities of $\Lambda(\boldsymbol{U})$ yielding the probability of misdetection,

P_M, and the probability of false alarm, $P_{_{FA}}$ at the fusion center. To do that, since P_M requires an analogous development, we only focus on $P_{_{FA}}$ for the sake of brevity. According to (2.141), the probability of false alarm at the fusion center can be found by integrating the saddlepoint approximation $\tilde{f}_\Lambda(s|H_0)$, i.e.

$$P_{_{FA}} = P\left(\Lambda(\boldsymbol{U}) > \tau |\, H_0\right) \approx \int_\tau^\infty \tilde{f}_\Lambda(s|H_0)\,\mathrm{d}s = \int_\tau^\infty \frac{e^{M_0(\hat{\theta}_0(s)) - s\,\hat{\theta}_0(s)}}{\sqrt{2\pi \frac{\partial^2 M_0(\hat{\theta}_0(s))}{\mathrm{d}\theta^2}}}\,\mathrm{d}s \tag{2.171}$$

where the last equality follows from (2.168). If we now apply the change of variables $s = \frac{\partial M_0(\hat{\theta}_0)}{\mathrm{d}\theta}$, the previous expression results into

$$P_{_{FA}} \approx \int_{\hat{\theta}_0(\tau)}^\infty \frac{1}{\sqrt{2\,\pi}} \sqrt{\frac{\partial^2 M_0(\theta)}{\mathrm{d}\theta^2}}\, e^{M_0(\theta) - \theta\, \frac{\partial M_0(\theta)}{\mathrm{d}\theta}}\,\mathrm{d}\theta \tag{2.172}$$

where $\hat{\theta}_0(\tau)$ equals the saddlepoint at τ obtained by solving

$$\frac{\partial M_0(\hat{\theta}_0)}{\mathrm{d}\theta} = \tau. \tag{2.173}$$

Adding and subtracting $1/r_0(\theta)$ and using the definition given in (2.170) for $k = 0$ as change of variable, (2.173) can be written as

$$\begin{aligned} P_{_{FA}} &\approx \int_{\hat{\theta}_0(\tau)}^\infty \frac{1}{\sqrt{2\,\pi}}\, \theta\, \frac{\partial^2 M_0(\theta)}{\mathrm{d}\theta^2} \left[\frac{1}{r_0(\theta)} - \frac{1}{r_0(\theta)} + \frac{1}{\theta \sqrt{\frac{\partial^2 M_0(\theta)}{\mathrm{d}\theta^2}}}\right] e^{-\frac{r_0(\theta)^2}{2}}\,\mathrm{d}\theta \\ &= \int_{\hat{\theta}_0(\tau)}^\infty \frac{1}{\sqrt{2\,\pi}}\, \theta\, \frac{\partial^2 M_0(\theta)}{\mathrm{d}\theta^2} \left[\frac{1}{g_0(\theta)} - \frac{1}{r_0(\theta)}\right] e^{-\frac{r_0(\theta)^2}{2}}\,\mathrm{d}\theta + \int_{r_0\left[\hat{\theta}_0(\tau)\right]}^\infty \frac{e^{-\frac{r^2}{2}}}{\sqrt{2\,\pi}}\,\mathrm{d}r \end{aligned} \tag{2.174}$$

where $g_0(\theta)$ is defined in (2.169). The first integration is the normal right tail probability while the second term in (2.174) can be integrated by parts with

$$\mathrm{d}V = \theta\, \frac{\partial^2 M_0(\theta)}{\mathrm{d}\theta^2}\, e^{-\frac{r_0(\theta)^2}{2}} \tag{2.175}$$

and

$$U = \frac{1}{g_0(\theta)} - \frac{1}{r_0(\theta)}. \tag{2.176}$$

This way, taking into account that an analogous development could be undertaken for the overall probability of misdetection P_M, the simplified approximation formulas for the probabilities of error at the fusion center are

$$\begin{aligned} P_{_{FA}} &= P\left(\Lambda(\boldsymbol{U}) > \tau \middle| H_0\right) \\ &\approx Q\left(r_0\left[\hat{\theta}_0(\tau)\right]\right) + \frac{e^{-\frac{r_0^2[\hat{\theta}_0(\tau)]}{2}}}{\sqrt{2\,\pi}} \left(\frac{1}{g_0\left[\hat{\theta}_0(\tau)\right]} - \frac{1}{r_0\left[\hat{\theta}_0(\tau)\right]}\right) \end{aligned} \tag{2.177}$$

and

$$\begin{aligned} P_M &= P\left(\Lambda(\boldsymbol{U}) \leq \tau \middle| H_1\right) \\ &\approx Q\left(-r_1\left[\hat{\theta}_1(\tau)\right]\right) - \frac{e^{-\frac{r_1^2[\hat{\theta}_1(\tau)]}{2}}}{\sqrt{2\,\pi}} \left(\frac{1}{g_1\left[\hat{\theta}_1(\tau)\right]} - \frac{1}{r_1\left[\hat{\theta}_1(\tau)\right]}\right). \end{aligned} \tag{2.178}$$

with $Q(\cdot)$ defined in (2.155).

An analysis of (2.177) and (2.178) reveals that, from a computational point of of view, the saddlepoint approximation is much less costly than the direct evaluation of the error probabilities. Note that the saddlepoint approximation only requires computing few simple expressions besides finding the saddlepoint $\hat{\theta}(\tau)$ with $k \in \{0, 1\}$. Due to the fact that $\frac{\partial M_k(\theta)}{\mathrm{d}\theta}$ is strictly increasing in θ and that, in our setting,

$$\lim_{\theta \to -\infty} \frac{\partial M_k(\theta)}{\mathrm{d}\theta} = \ell\,\Lambda_1(0), \tag{2.179}$$

and

$$\lim_{\theta \to \infty} \frac{\partial M_k(\theta)}{\mathrm{d}\theta} = \ell\,\Lambda_1(1), \tag{2.180}$$

the existence and uniqueness of the aforementioned saddlepoint can be ensured as long as $\tau \in (\ell\,\Lambda_1(0), \ell\,\Lambda_1(1))$ and can be found with numerical univariate techniques such as the Newton-Raphson method. This last fact shows that the saddlepoint approximation is simple to compute besides being

accurate. What is more, although it is based on asymptotic expansions, as it is illustrated in [Aldosari and Moura, 2007] the saddlepoint approximation is also highly accurate even for parallel networks of few sensors. Nevertheless, we have to keep in mind that it can only be applied to scenarios where the binary local decisions are conditionally i.i.d. under each one of the two possible hypothesis. For instance, in [Aldosari and Moura, 2007] some extensions to a setting where the binary local decisions are non identically distributed have been undertaken. However, as far as the authors are concerned, all of these extensions require that not all the involved random variables are non-identically distributed. In that case and in other scenarios where the binary local decisions might also be conditionally dependent, we have to resort to other analytical tools such as the error exponents.

2.8.3 Error exponent

When considering scenarios where the binary local decisions are conditionally dependent and/or non-identically distributed, the literature usually avoids the direct computation of the performance probabilities. Additionally, as we explained in the previous subsections, under the previous conditions the normal and the saddlepoint approximations cannot be applied to approximate the aforementioned probabilities. Instead, the performance measure can be obtained by following an approach that does not have a high computational cost and that consists in evaluating their exponential rate of decay in an asymptotic regime where the number ℓ of local decisions approaches infinity. These exponential rates of decay for the overall probabilities of error are the so-called error exponents.

Although the error exponents do not help with evaluating the receiver operating characteristic curve (ROC) or designing the optimal fusion rule under both the Bayesian and the Neyman-Pearson set ups, their closed-form expression allows to adapt the network parameters with the aim of minimizing the error probabilities. However, before that and given a specific detection scenario, it is necessary to derive them and, at the same time, check their existence. To do that, the Gärtner-Ellis theorem has to be employed.

Theorem I (Gärtner-Ellis theorem [Dembo and Zeitouni, 1998])**.** *Let* $\{\boldsymbol{\Lambda}_\ell\} \in \mathbb{R}^m$ *be a sequence of random variables and define*

$$I^{(\ell)}(\boldsymbol{\theta}) = \ln\left(E\left\{\exp\left(\boldsymbol{\theta}^T \boldsymbol{\Lambda}_\ell\right)\right\}\right). \tag{2.181}$$

Suppose that for each $\boldsymbol{\theta} \in \mathbb{R}^m$*, the logarithmic moment generating function, defined as*

$$I(\boldsymbol{\theta}) = \lim_{\ell \to \infty} \frac{I^{(\ell)}(\ell\, \boldsymbol{\theta})}{\ell} \tag{2.182}$$

exists as an extended real number. If $I(\cdot)$ *is an essentially smooth and lower-semicontinuous function,* $\{\boldsymbol{\Lambda}_\ell\}$ *satisfies the large deviation principle with rate transform*

$$I^*(\boldsymbol{x}) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^m} \left\{ \boldsymbol{\theta}^T \boldsymbol{x} - I(\boldsymbol{\theta}) \right\}. \tag{2.183}$$

That is, for any measurable set $\mathcal{B}$

$$\begin{aligned} -\inf_{\boldsymbol{x} \in \mathrm{int}(\mathcal{B})} I^*(\boldsymbol{x}) &\leq \liminf_{\ell \to \infty} \frac{\ln\left(P\left(\boldsymbol{\Lambda}_\ell \in \mathcal{B}\right)\right)}{\ell} \\ &\leq \limsup_{\ell \to \infty} \frac{\ln\left(P\left(\boldsymbol{\Lambda}_\ell \in \mathcal{B}\right)\right)}{\ell} \leq -\inf_{\boldsymbol{x} \in \mathrm{clos}(\mathcal{B})} I^*(\boldsymbol{x}) \end{aligned} \tag{2.184}$$

where $\mathrm{int}(\mathcal{B})$ *denotes the interior of* $\mathcal{B}$ *and* $\mathrm{clos}(\mathcal{B})$ *the closure of* $\mathcal{B}$*.*

In an implicit form the previous theorem basically provides the asymptotic behavior of a sequence of non-i.i.d. random variables $\{\boldsymbol{\Lambda}_\ell\} \in \mathbb{R}^m$. Toward this goal, unlike the approaches described in the last two subsections, assumptions about the dependency structure of the sequence $\{\boldsymbol{\Lambda}_\ell\}$ are dispensed. Instead, the proof of the previous theorem, which has been omitted for the sake of brevity, is supported by assumptions on a sequence of moment generating functions, or more specifically, the logarithmic moment generating function defined in (2.182). Note that, if we define

$$\Lambda_\ell = \frac{1}{\ell} \Lambda(\boldsymbol{U}) \tag{2.185}$$

with $\boldsymbol{U} = [U_1, U_2, \ldots, U_\ell]^T$, most of the detection scenarios satisfy those assumptions. In addition to that, the sets of interest in hypothesis testing mostly hold the continuity property

$$\inf_{x \in \mathrm{int}(\mathcal{B})} I^*(x) = \inf_{x \in \mathrm{clos}(\mathcal{B})} I^*(x), \tag{2.186}$$

which implies

$$-\lim_{\ell\to\infty}\frac{\ln\left(P\left(\Lambda_\ell\in\mathcal{B}\right)\right)}{\ell}=\inf_{x\in\mathcal{B}}I^*(x) \tag{2.187}$$

This is precisely the implicit form that, provided by the Gärtner-Ellis theorem, yields the exponential rate of decay, or error exponent, associated with each one of the two error probabilities at the fusion center. In particular, if we focus on the overall probability of error when H_k is true with $k\in\{0,1\}$, by realizing that $I^*(x)$ is a convex function that achieves its minimum at

$$\lim_{\ell\to\infty}\frac{E\left\{\Lambda_\ell\right\}}{\ell}, \tag{2.188}$$

as long as $\tau_\ell\in(\bar{l}_0,\bar{l}_1)$ with $l_0\leq 0\leq l_1$ and

$$\bar{l}_k=\lim_{\ell\to\infty}\frac{E_{H_k}\left\{\Lambda(\boldsymbol{U})\right\}}{\ell} \tag{2.189}$$

under both the Bayesian and the Neyman-Pearson set ups the error exponent is given by

$$\begin{aligned}&-\lim_{\ell\to\infty}\frac{\ln\left(P\left(U_{FC}\neq H_k\right)\right)}{\ell}\\&\qquad=\begin{cases}\lim_{\ell\to\infty}\frac{-\ln(P(\Lambda_\ell\in(\tau_\ell,\infty)|H_0))}{\ell}=I_0^*(\tau_\ell) & \text{if } k=0\\ \lim_{\ell\to\infty}\frac{-\ln(P(\Lambda_\ell\in(-\infty,\tau_\ell]|H_1))}{\ell}=I_1^*(\tau_\ell) & \text{if } k=1\end{cases}\end{aligned} \tag{2.190}$$

where

$$P\left(\Lambda_\ell\in(\tau_\ell,\infty)\middle|\,H_0\right)=P_{_{FA}}, \tag{2.191}$$

$$P\left(\Lambda_\ell\in(-\infty,\tau_\ell]\middle|\,H_1\right)=P_M \tag{2.192}$$

and

$$I_k^*(x)=\sup_{\theta\in\mathbb{R}}\left\{\theta\,x-I_k(\theta)\right\} \tag{2.193}$$

with

$$\begin{aligned}I_k(\theta)&=\lim_{\ell\to\infty}\frac{I_k^{(\ell)}(\ell\,\theta)}{\ell}=\lim_{\ell\to\infty}\frac{\ln\left(E_{H_k}\left\{\exp\left(\ell\,\theta\,\Lambda_\ell\right)\right\}\right)}{\ell}\\&=\lim_{\ell\to\infty}\frac{\ln\left(E_{H_k}\left\{\exp\left(\theta\,\Lambda(\boldsymbol{U})\right)\right\}\right)}{\ell}=\lim_{\ell\to\infty}\frac{\ln\left(E_{H_k}\left\{\left(\frac{P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}{P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}\right)^{\theta}\right\}\right)}{\ell}.\end{aligned} \tag{2.194}$$

Due to the fact that $I^*(x)$ equals zero at its minimum, we can easily check that

$$\lim_{\ell\to\infty} \frac{\ln\left(P_{_{FA}}\right)}{\ell} = 0 \tag{2.195}$$

if $\tau_l \leq \bar{l}_0$ and that

$$\lim_{\ell\to\infty} \frac{\ln\left(P_{M}\right)}{\ell} = 0 \tag{2.196}$$

if $\tau_l \geq \bar{l}_1$. However, this kind of choices of τ_ℓ are not frequent in many detection applications. For instance, if $P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k)$ is a symmetric p.m.f. for all $k \in \{0,1\}$, the previous choices of τ_ℓ would be possible whenever we want to ensure that $P_{_{FA}} \geq 0.5$ and that $P_M \geq 0.5$ at the fusion center. Because of this, from now and on we will assume that $\tau_\ell \in (\bar{l}_0, \bar{l}_1)$.

As we have just shown, the Gärtner-Ellis theorem provides the error exponent for the probability of misdetection and the probability of false alarm at the fusion center and under both the Bayesian and the Neyman-Pearson set ups. Indeed, the use of the Gärtner-Ellis theorem is wider and it is not only restricted to hypothesis testing problems. Realize that it can be applied to whatever sequence of random variables as long as the corresponding sequence of moment generating functions satisfy the conditions stated by the theorem. Because of this, taking full advantage of the particular features associated with each one of the two considered formulations, several works have undertaken the extension and particularization of the Gärtner-Ellis theorem under the Bayesian set up and under the Neyman-Pearson formulation.

Possibly, the more straightforward particularization of the Gärtner-Ellis theorem to a hypothesis testing problem is the so-called Chernoff information generalization, which it is derived for fusion rules performing an optimal Bayesian test on non-i.i.d. sensor observations. After noting that, under the Bayesian set up,

$$\lim_{\ell\to\infty} \tau_\ell = \lim_{\ell\to\infty} \frac{\tau}{\ell} = \lim_{\ell\to\infty} \frac{C_F}{C_D \ell} = 0 \tag{2.197}$$

where the last equality follows from (2.45), since

$$\begin{aligned} I_0^*(0) &= \sup_{\theta\in\mathbb{R}} \{-I_0(\theta)\} = -\inf_{\theta\in\mathbb{R}} \{I_0(\theta)\} = -\inf_{\theta\in[0,1]} \{I_0(\theta)\} \\ &= -\inf_{\theta\in[0,1]} \{I_1(\theta)\} = -\inf_{\theta\in\mathbb{R}} \{I_1(\theta)\} = \sup_{\theta\in\mathbb{R}} \{-I_1(\theta)\} = I_1^*(0) \end{aligned} \tag{2.198}$$

with

$$I_0(\theta) = \lim_{\ell\to\infty} \frac{\ln\left(\sum_{\boldsymbol{u}\in\mathcal{U}^\ell} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)^\theta P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)^{1-\theta}\right)}{\ell} \tag{2.199}$$

and

$$I_1(\theta) = \lim_{\ell\to\infty} \frac{\ln\left(\sum_{\boldsymbol{u}\in\mathcal{U}^\ell} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)^{1+\theta} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)^{-\theta}\right)}{\ell} \tag{2.200}$$

we can straightforwardly prove the Chernoff information generalization from the Gärtner-Ellis theorem. In particular, the Chernoff information generalization is stated as follows.

Theorem II (Chernoff information generalization)**.** *Let $\boldsymbol{U} = [U_1, \ldots, U_\ell]^T$ be a sequence of ℓ random variables jointly distributed according $P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)$ under H_0 and according to $P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)$ under H_1. Also let $\mathcal{V}_\ell^1(\tau)$ be an acceptance region for hypothesis H_1 such that the probabilities of error for the hypothesis testing problem is $P_e(\ell) = P(H_0)\, P_{FA} + (1 - P(H_0))\, P_M$ with*

$$P_{_{FA}} = \sum_{\boldsymbol{u}\in\mathcal{V}_\ell^1} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0), \tag{2.201}$$

$$P_M = 1 - \sum_{\boldsymbol{u}\in\mathcal{V}_\ell^1} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1) \tag{2.202}$$

and $P(H_0)$ denoting the prior probability of H_0. Also define $P_e^* = \min_{\mathcal{V}_\ell^1(\tau)} P_e$, which is given by the optimal Bayes fusion rule. Then, the best exponential rate of decay in P_e^* is given by

$$\begin{aligned} K &= \lim_{\ell\to\infty} -\frac{1}{\ell}\ln(P_e^*) \\ &= -\min_{0\le\theta\le 1} \lim_{\ell\to\infty} \frac{\ln\left(\sum_{\boldsymbol{u}\in\mathcal{U}^\ell} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)^\theta P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)^{1-\theta}\right)}{\ell}. \end{aligned} \tag{2.203}$$

As it happens under the Bayesian set up, the Gärtner-Ellis theorem admits a particularization to the Neyman-Pearson formulation. From (2.190) we can verify that the Gärtner-Ellis theorem provides the error exponent for the overall probability of misdetection by means of an optimization whose

solution usually requires numerical methods as well as the integration of the spectrum of the observation process. Most of the results following this approach have derived the aforementioned error exponent when the log-likelihood ratio test (LLRT) is based on a fixed threshold. This results in rates of decay for P_M that are functions of the threshold. However, for ergodic cases the normalized log-likelihood ratio Λ_ℓ, defined in (2.185), converges to a constant under the null hypothesis

$$\bar{l}_0 = \lim_{\ell\to\infty} \frac{1}{\ell} \ln\left(\frac{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}\right) \quad \text{(a.s. in } H_0\text{)}, \tag{2.204}$$

provided that the previous limit exists and where the notation (a.s. in H_0) means that the limit has to be taken in the almost sure sense under the distribution present under H_0. Consequently, in this case the overall probability of false alarm also decays exponentially for a fixed threshold. This last fact entails that a detector with a fixed threshold is not optimal in the Neyman-Pearson sense since it does not use the constraint of false alarm fully. In particular, for the optimal Neyman-Pearson detector the threshold is a function of the sample size, ℓ, satisfying

$$\lim_{\ell\to\infty} \tau_\ell = \lim_{\ell\to\infty} \frac{\tau}{\ell} = \bar{l}_0 \tag{2.205}$$

in order that $P_{FA} \leq \alpha \in (0,1)$. If we now write P_M as follows

$$P_M = P\left(\Lambda_\ell - \tau_\ell \leq 0 \middle| H_1\right) = P\left(\widetilde{\Lambda_\ell} \leq 0 \middle| H_1\right) \tag{2.206}$$

where

$$\widetilde{\Lambda_\ell} = \Lambda_\ell - \tau_\ell \tag{2.207}$$

with Λ_ℓ defined in (2.185), the application of the Gärtner-Ellis theorem to the sequence $\{\widetilde{\Lambda_\ell}\}$ under H_1 yields

$$\begin{aligned} K &= \lim_{\ell\to\infty} -\frac{1}{\ell}\ln(P_M) = -\inf_{x\in\in(-\infty,0]} I_1^*(x) = I_1^*(0) \\ &= -\inf_{\theta\in\mathbb{R}} \{I_1(\theta)\} \end{aligned} \tag{2.208}$$

with

$$I_1(\theta) = -\theta\,\bar{l}_0 + \lim_{\ell\to\infty} \frac{\ln\left(\sum_{\boldsymbol{u}\in\mathcal{U}^\ell} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)^{1+\theta} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)^{-\theta}\right)}{\ell}. \tag{2.209}$$

Under proper mathematical conditions we can verify that, when $\theta = -1$, we attain the unique solution of the optimization problem expressed in (2.208) and (2.209). Since we still take full advantage of the constraint $P_{_{FA}} \leq \alpha \in (0,1)$, when $\theta = -1$ we obtain the particularization of the Gärtner-Ellis theorem under the Neyman-Pearson formulation. This particularization is the so-called Stein's lemma generalization and is stated as follows.

Theorem III (Stein's lemma generalization [Vajda, 1989])**.** *Let* $\boldsymbol{U} = [U_1, \ldots, U_\ell]^T$ *be a sequence of* ℓ *random variables jointly distributed according* $P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)$ *under* H_0 *and according to* $P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)$ *under* H_1*. Also let* $\mathcal{V}_\ell^1(\tau)$ *be an acceptance region for hypothesis* H_1 *such that the probability of false alarm for the hypothesis testing problem is*

$$P_{_{FA}} = \sum_{\boldsymbol{u} \in \mathcal{V}_\ell^1} P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0). \tag{2.210}$$

For $0 < \alpha < 1$*, define*

$$P_M^\alpha = \min_{\mathcal{V}_\ell^1(\tau) \subseteq \mathcal{U}^\ell;\, P_{_{FA}} \leq \alpha} \{P_M\} \tag{2.211}$$

then, the exponential rate of decay in P_M^α *is given by the asymptotic Kullback-Leibler rate*

$$K = \lim_{\ell \to \infty} -\frac{1}{\ell} \log(P_M^\alpha) = -\lim_{\ell \to \infty} \frac{1}{\ell} \ln \left(\frac{P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}{P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)} \right) \quad \textit{(a.s. } H_0\textit{)}. \tag{2.212}$$

At this point, we want to remark that, under i.i.d. local decisions, we can perform further simplifications of the implicit expressions of the error exponents provided by both the Chernoff information generalization and the Stein's lemma generalization. On the one hand, under i.i.d. binary local decisions and under the Bayesian set up the error exponent for the probability of error of the optimal fusion rule, P_e^*, is given by the Chernoff information [Cover and Thomas, 2006]

$$\begin{aligned} K &= \lim_{\ell \to \infty} -\frac{1}{\ell} \ln(P_e^*) \\ &= -\min_{0 \leq \theta \leq 1} \ln \left(\sum_{u_1=0}^{1} P_{U_1|H}(u_1|H_1)^\theta P_{U_1|H}(u_1|H_0)^{1-\theta} \right). \end{aligned} \tag{2.213}$$

On the other hand, subject to $P_{_{FA}} \leq \alpha \in (0,1)$ and under i.i.d. local decisions, the Stein's lemma generalization collapses to the Stein's lemma given in [Cover and Thomas, 2006], which gives the best error exponent for the probability of misdetection of the optimal fusion rule, P_M^{α}, under the Neyman-Pearson formulation

$$\begin{aligned} K &= \lim_{\ell \to \infty} -\frac{1}{\ell} \log(P_M^{\alpha}) \\ &= D(P_{U_1|H}(u_1|H_0)||P_{U_1|H}(u_1|H_1)) = D(B(P_{_{FA}}(1))||B(P_D(1))) \end{aligned} \tag{2.214}$$

where $D(P_{U_1|H}(u_1|H_0)||P_{U_1|H}(u_1|H_1))$ denotes the Kullback-Leibler divergence between $P_{U_1|H}(u_1|H_0)$ and $P_{U_1|H}(u_1|H_1)$ and where the last equality follows because, when the binary local decisions are i.i.d under H_k with $k \in \{0,1\}$,

$$P_{U_1|H}(u_1|H_k) = P_{U_1|H}(u_1|H_k) = B(P(U_1 = 1|H_k)) \tag{2.215}$$

with $i \in \{1, 2, \ldots, \ell\}$ and $B(p)$ denoting a Bernoulli random variables whose probability of success equals p. To conclude the subsection, do not forget that, analyzing and tuning the performance of a parallel network, under the assumption of conditionally i.i.d. local decisions, the saddlepoint approximation is a more powerful tool than the error exponent. This is due to the fact that, unlike the saddlepoint approximations, which are also easy to compute, the error exponents are not approximations of the probabilities of error at the fusion center. Nevertheless, under both non-i.i.d. local decisions or i.i.d. local decisions they are an important measure of performance. Since they give the exponential rate of decay for the error probabilities at the fusion center, in an asymptotic regime where the number of observations approaches infinity the error exponents give a rough estimate of the number of samples required for a given detector performance. Consequently, realize that faster decay rate implies that fewer samples are needed for a given probability of misdetection or probability of error, or that better performance can be obtained with a given number of samples.

Chapter 3

Neyman-Pearson fusion of dependent local decisions in a 1-D network

3.1 Introduction

Supported by different theoretical areas such as the information theory and the large deviation theory, the latest results on distributed detection systems with parallel architecture and dependent observations provide tools that, by means of physical and design parameters, characterize the detection performance of networks with optimal fusion rules and dependent observations. Some works that followed this philosophy, for instance [Sung et al., 2006] and [Chamberland and Veeravalli, 2006], analyzed a scenario where the devices are equally spaced along a straight line, the amplify-and-relay scheme is chosen as the local processing strategy of the sensor observations, and the sensors send their locally processed measures to the fusion center through a bank of independent parallel channels. On the one hand, in the Bayesian set up the authors in [Chamberland and Veeravalli, 2006] used the Gärtner-Ellis theorem provided in [Dembo and Zeitouni, 1998] to study the impact of sensor density on two detection problems: one consists in detecting two deterministic signals in first-order Gauss-Markov noise, and the other is the detection of a first-order Gauss-Markov signal in zero mean additive white Gaussian noise (AWGN). On the other hand, [Sung et al., 2006] obtained a closed-

form expression of the asymptotic Kullback-Leibler rate by means of the innovations approach described in [Kailath et al., 2000] as well as the connection between the asymptotic behaviour of the optimal Neyman-Pearson detector and that of the Kalman filter (see [Helstrom, 1994]). This closed-form expression allowed them to study the effect of sensor spacing on the Neyman-Pearson detection of a first-order Gauss-Markov process in AWGN. Later, [Li and Dai, 2007], [Sung et al., 2008b] and [Misra and Tong, 2008] extended the work of [Sung et al., 2006] and [Chamberland and Veeravalli, 2006] in order to give some insights into several design matters related to 1-D sensor networks that perform a distributed detection task. Under the same scenario as in [Sung et al., 2006], the authors in [Li and Dai, 2007] utilized asymptotic properties of Toeplitz matrices given in [Grenander and Szegö, 1984] in order to compare the detection performance when the devices send their observations through a bank of parallel independent channels, or through an unique multiple access channel. [Sung et al., 2008b] extended their work in [Sung et al., 2006] in order to study the optimal sensor configuration along the straight line where the devices are deployed. For a 1-D sensor network where the devices are randomly spaced and the distances between two neighbour nodes are known by the fusion center, [Misra and Tong, 2008] derived a closed-form expression of the Neyman-Pearson error exponent associated with the discrimination of two first-order Gauss-Markov processes.

So far, there are very few works that characterize networks with optimal fusion rules performed on dependent local decisions Some of the most recent results regarding this problem are [Villard et al., 2010] and [Villard and Bianchi, 2011]. Extending [Villard et al., 2010] the same authors in [Villard and Bianchi, 2011] investigated the effect of quantization on the performance of a Neyman-Pearson fusion rule undertaken in a scenario where the sensor observations are dependent. More precisely, in a high-rate quantization regime i.e., when the number of quantization levels tends to infinity, the authors derived a compact expression of the error exponent loss induced by quantization. Using this compact expression they showed that, as the number of quantization levels tends to infinity, the error exponent for quantized observations converges to the one associated with a scenario where the fusion center receives unquantized measurements. However, no work obtains analytically tractable expressions that allow to assess which kind of trends of detection performance can appear with increasing dependency in systems where the local quantizers have a finite number of quantization

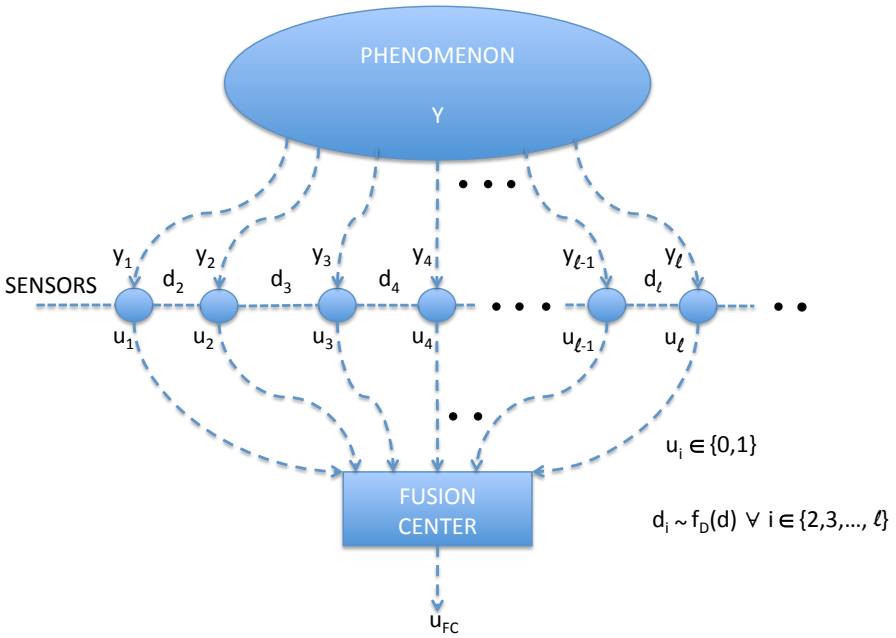


Figure 3.1: *Block diagram of a 1-D parallel fusion network.*

levels. Motivated by this last fact and taking into account the relevance of the distributed detection systems with binary local processing schemes in the literature of sensor networks, in this chapter we derive a closed-form expression for the error exponent associated with the Neyman-Pearson fusion rule performed in the 1-D sensor network of Figure 3.1. Firstly, this closed-form expression is derived for a scenario where the devices are randomly deployed along a straight line and where the data fusion center only knows the common p.d.f. from which each sensor spacing is drawn independently. After analyzing the different properties of the error exponent, this one is simplified for specific cases of the p.d.f. that have already been assumed in [Misra and Tong, 2008]. In particular, the cases that have been analyzed are equispaced sensors with failure rate q and exponentially spaced sensors with placement density λ and failure rate q. This way, we provide an amenable tool that links the detection performance of a 1-D version of the sensor network described in [Drakopoulos and Lee, 1991] with its physical and design parameters.

The rest of this chapter is organized as follows. Section 3.2 is devoted to

the problem statement. In Section 3.3 we perform a large system analysis that firstly provides a closed-form expression for the error exponent for the Neyman-Pearson fusion rule performed in the sensor network of Figure 3.1. Afterwards, using a parameter that captures the mean correlation among the random variables involved in the fusion rule, the analysis continues investigating some properties of the aforementioned error exponent. For two different models of the sensor spacing, Section 3.4 undertakes evaluations of the derived error exponent as well as several numerical experiments in order to study the effect of different physical and design parameters of the network on its detection performance. Next, Section 3.5 summarizes the results of this chapter. Finally, as a matter of organization style, we defer all proofs of this chapter to a pair of appendices.

3.2 Problem statement

In this chapter, we consider the design of a sensor network as the one shown in Figure 3.1. This network is formed by a data fusion center that receives a large number, ℓ, of binary local decisions from sensors randomly deployed along a straight line according to a random sequence of spacings, $\{d_i\}_{i=2}^{\ell}$, i.i.d. according to an arbitrary p.d.f., $f_D(d)$, defined on $\mathcal{D}$, which is known at the fusion center. With the aim of deciding what state of the phenomenon is present, H_0 or H_1, the distributed detection system undertakes the following steps. Firstly, each device performs a local observation of the environment, y_i. Secondly, it applies a binary detection rule to it, $\gamma_i(y_i)$, not necessarily based on a LLRT. This way, each device makes a local decision, $u_i = \gamma_i(y_i)$, regarding the presence or absence of the phenomenon we want to detect. Thirdly, the binary local quantizations of the sensor observations are transmitted to the fusion center over error free parallel access channels. Finally, based on the ℓ local decisions taken by the devices of the network, the fusion center makes a global decision, u_{FC}, under the Neyman-Pearson formulation. In this detection process, we assume that the correlation structure of the local decisions is hypothesis dependent and modelled with a 1-D Markov random field (MRF) with nearest-neighbour dependency and binary state space. This way, given the elements of the

sequence $\{d_i\}_{i=2}^{\ell}$ whose joint p.d.f. is expressed as

$$f_{\boldsymbol{D}}(\boldsymbol{d}) = \prod_{i=2}^{\ell} f_D(d_i) = f_D(d)^{\ell-1}, \tag{3.1}$$

the subsequent inference problem is established at the fusion center under the Neyman-Pearson formulation

$$\left.\begin{array}{ll} H_k\text{:} & \boldsymbol{U} \sim \text{1-D MRF with nearest-neighbour dependency} \\ & \quad \text{and non-homogeneous probabilities, } \{p_k^t(d_i, d_{i+1}, \mathcal{P}_k)\}_{i=2}^{\ell-1} \end{array}\right\} \tag{3.2}$$

where, for $k \in \{0, 1\}$ and $t \in \{0, 1, 1', 2\}$,

- $\mathbf{U} = [U_1, U_2, \ldots, U_\ell]^T$, defined on $\mathcal{U}^\ell$ with $\mathcal{U} = \{0, 1\}$, denotes the ℓ binary decisions received at the fusion center and locally performed by the devices that form the network.

- $\mathcal{P}_k$ are physical and design parameters of the network that, being known by the data fusion center, are arguments of the conditional probabilities, $\{p_k^t(d_i, d_{i+1}, \mathcal{P}_k)\}_{i=2}^{\ell-1}$ (e.g. the employed local decision rules or the physical properties of the environment where the network is deployed). However, in this chapter from now on we omit this dependency in order to keep the notation simple.

- $\{p_k^t(d_i, d_{i+1})\}_{i=2}^{\ell-1}$ denotes the set that contains all the one-point conditional probabilities of the MRF present under H_k. For each one of the values of t its elements take the subsequent expressions

$$p_k^0(d_i, d_{i+1}) = P(U_i = 1 | U_{i-1} = U_{i+1} = 1, d_i, d_{i+1}, H_k), \tag{3.3}$$

$$p_k^1(d_i, d_{i+1}) = P(U_i = 1 | U_{i-1} = 1, U_{i+1} = 0, d_i, d_{i+1}, H_k), \tag{3.4}$$

$$p_k^{1'}(d_i, d_{i+1}) = P(U_i = 1 | U_{i-1} = 0, U_{i+1} = 1, d_i, d_{i+1}, H_k), \tag{3.5}$$

$$p_k^2(d_i, d_{i+1}) = P(U_i = 1 | U_{i-1} = 0, U_{i+1} = 0, d_i, d_{i+1}, H_k) \tag{3.6}$$

  when $i \in \{2, 3, \ldots, \ell - 1\}$. For a specific sensor i we consider that the non-homogeneity of the previous probabilities is exclusively caused by the spacings with the nearest neighbour nodes, i.e. d_i and d_{i+1}.

Note that sensor spacings have the same distribution under each one of the two hypothesis. Therefore, the distribution of $\boldsymbol{U}$ conditioned on $\{d_i\}_{i=2}^{\ell}$ under each hypothesis is relevant for inference. Taking into account that, under both hypothesis, the data fusion center only knows the statistics of the aforementioned distances, in order to ensure that $P_{_{FA}} \leq \alpha \in (0,1)$ the optimal fusion rule for (3.2) is given by the subsequent Neyman-Pearson detector

$$u_{_{FC}} = \begin{cases} 1 & \text{if } \ln\left(\frac{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}\right) < \tau, \\ \gamma & \text{if } \ln\left(\frac{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}\right) = \tau, \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

where, for $k \in \{0,1\}$ and defining the acceptance region for H_1

$$\mathcal{V}_{\ell}^{1}(\tau) = \left\{ \boldsymbol{u} : \ln\left(\frac{\overline{P}_{\boldsymbol{U}|H_0}(\boldsymbol{u}|H_0)}{\overline{P}_{\boldsymbol{U}|H_1}(\boldsymbol{u}|H_1)}\right) < \tau \right\}, \tag{3.8}$$

we have that

$$\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k) = \int_{\mathcal{D}^{\ell-1}} P_{\boldsymbol{U}|\boldsymbol{D},H}(\boldsymbol{u}|\boldsymbol{d},H_k) \cdot f_{\boldsymbol{D}}(\boldsymbol{d}) \cdot \mathrm{d}\boldsymbol{d}, \tag{3.9}$$

τ is the smallest number such that

$$\varsigma = \sum_{\boldsymbol{u} \in \mathcal{V}_{\ell}^{1}(\tau)} \overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0) \leq \alpha \tag{3.10}$$

and $\gamma \in \{0,1\}$ is a randomization function that takes on the value 1 with the subsequent probability

$$\nu = \begin{cases} 1 & \text{if } \vartheta = 0, \\ \frac{\alpha - \varsigma}{\vartheta} & \text{otherwise} \end{cases} \tag{3.11}$$

with

$$\vartheta = P\left(\ln\left(\frac{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)}\right) = \tau \,\middle|\, H_0 \right). \tag{3.12}$$

As it is stated in [Wainwright and Jordan, 2008], graphical models provide the formalism to capture complex dependencies among random

variables, and to represent them by means of different types of dependency graphs where the nodes denote the involved random variables, and where the edges capture the dependencies among them. Examples include Markov models or Markov random fields. For directed acyclic graphical models, such as a Markov chain, distributions of the random variables associated with each one of the vertex of the graph can be factorized in terms of the conditional distributions of each random variable given its parents, i.e. given the random variables directly linked with its incoming edges in the graph. For undirected graphical models, such as a MRF, distributions can be factorized by a set of compatibility functions, each one defined over one of the cliques of the graph. Taking into account that a clique is a subset of random variables whose corresponding nodes are fully connected in the graph, a compatibility function is a positive local quantity defined only for elements within the clique and not necessarily expressed in terms of marginal or conditional distributions, as it happens with the factorization terms of directed graphs. Typically, for undirected graphical models the factorization of the distributions is performed over the set of all maximal cliques of the graph, which is the set of cliques that are not properly contained within any other clique.

In Figure 3.2 we now show the dependency graph for the MRF of hypothesis H_k. Since for this 1-D graph the set of maximal cliques is the set consisting in all pairs of consecutive vertex, from the Hammersley-Clifford theorem given in [Lauritzen, 1996] we have that

$$\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k) = \frac{1}{Z_k} \prod_{i=2}^{\ell} \overline{\psi}(u_i, u_{i-1}|H_k) \tag{3.13}$$

with $k \in \{0, 1\}$, Z_k denoting a bounded constant chosen to ensure that the distribution is normalized,

$$0 < Z_k = \sum_{\boldsymbol{U} \in \mathcal{U}^{\ell}} \prod_{i=2}^{\ell} \overline{\psi}(u_i, u_{i-1}|H_k) < \infty, \tag{3.14}$$

and

$$\prod_{i=2}^{\ell} \overline{\psi}(u_i, u_{i-1}|H_k) = \int_{\mathcal{D}^{\ell-1}} \prod_{i=2}^{\ell} \psi(u_i, u_{i-1}|\boldsymbol{d}, H_k)\, f_{\boldsymbol{D}}(\boldsymbol{d})\, \mathrm{d}\boldsymbol{d} \tag{3.15}$$

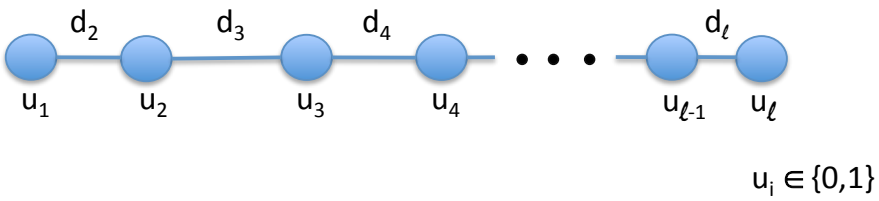


Figure 3.2: *Dependency graph of the local decisions under the hypothesis H_k with $k \in \{0, 1\}$.*

where $\psi(u_i, u_{i-1}|\boldsymbol{d}, H_k)$ equals the compatibility function associated with the i-th maximal clique that appears in the dependency graph of $\boldsymbol{U}$ given $\boldsymbol{D}$ and H_k (see Figure 3.2).

3.3 Large system analysis

Here we provide tools that allow the design and analysis of sensor networks as the one described in the previous section. Firstly, based on information theoretic results we derive a design tool that links the detection performance of (3.7) with different physical and design parameters of the network. Secondly, paying special attention to the effect of the network density on its detection performance, we obtain a single parameter that measures the mean degree of correlation among the neighbour local decisions involved in the inference problem of Section 3.2. Finally, we investigate some analytical properties of the derived error exponent when the dependence among the local decisions results in different boundary values of the aforementioned correlation parameter.

3.3.1 Error exponent.

In order to characterize the detection performance of the Neyman-Pearson test shown in (3.7), as it is explained in Subsection 2.8.3 the common procedure is to derive a closed-form expression of its corresponding probability of detection

$$P_D = \sum_{\boldsymbol{u} \in \mathcal{V}_\ell^1(\tau)} \overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1) + \nu \cdot P\left(\ln\left(\frac{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)} \right) = \tau \middle| H_1 \right). \quad (3.16)$$

However, this kind of solution is not feasible because of several practical considerations. Among them, we can remark that the evaluation of this performance metric might need numerical computations that make the derived tool be non-amenable when the available number of binary local decisions, ℓ, is very large. Thus, in this chapter our approach consists in obtaining the mentioned tool by means of more easy-to-use measures of performance related to P_D. In particular, in the spirit of other publications of the literature, focusing on the distributed detection problem of Section 3.2 our objective is to derive a closed-form expression of the exponential rate of decay in $P_M = 1 - P_D$ as ℓ approaches infinity.

Given the problem statement of Section 3.2, the error exponent of the Neyman-Pearson detector (3.7) equals the almost sure limit of the asymptotic Kullback-Leibler rate under $\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)$ (see Section 2.8), i.e.,

$$K = \lim_{\ell \to \infty} \frac{1}{\ell} \ln \left(\frac{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)}{\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)} \right) \quad \text{(a.s. in } H_0) \tag{3.17}$$

provided that the limit exists and where the notation (a.s. in H_0) means that the limit has to be taken in the almost sure sense under the distribution present under H_0. When obtaining K, if we look at (3.13), we firstly need to define the compatibility functions, $\overline{\psi}(u_i, u_{i-1}|H_k)$, through the conditional probabilities associated with the dependence model assumed under H_k. From (3.15) this problem results in writing $\psi(u_i, u_{i-1}|\boldsymbol{d}, H_k)$ in terms of $\{p_k^t(d_i, d_{i+1})\}_{i=2}^{\ell-1}$ with $t \in \{0, 1, 1^{\shortmid}, 2\}$. Using a straightforward extension of the results derived by Spitzer in [Spitzer, 1971] for homogeneous MRF, it can be easily proved that, in a 1-D scenario where the random variables are defined on a binary state space, a non-homogeneous MRF with nearest-neighbour dependency has a biunique correspondence with a non-homogeneous first-order Markov chain. Consequently, if the transition probabilities related to the i-th step of a non-homogeneous Markov chain are denoted as

$$p_k^{(0,1)}(d_i) = P(U_i = 1|U_{i-1} = 0, d_i, H_k) \tag{3.18}$$

and

$$p_k^{(1,0)}(d_i) = P(U_i = 0|U_{i-1} = 1, d_i, H_k), \tag{3.19}$$

the conditional probabilities of the corresponding non-homogeneous MRF

are

$$p_k^0(d_i, d_{i+1}) = \frac{(1-p_k^{(1,0)}(d_i))(1-p_k^{(1,0)}(d_{i+1}))}{(1-p_k^{(1,0)}(d_i))(1-p_k^{(1,0)}(d_{i+1})) + p_k^{(1,0)}(d_i)p_k^{(0,1)}(d_{i+1})}, \tag{3.20}$$

$$p_k^1(d_i, d_{i+1}) = \frac{(1-p_k^{(1,0)}(d_i))p_k^{(1,0)}(d_{i+1})}{(1-p_k^{(1,0)}(d_i))p_k^{(1,0)}(d_{i+1}) + p_k^{(1,0)}(d_i)(1-p_k^{(0,1)}(d_{i+1}))}, \tag{3.21}$$

$$p_k^{1'}(d_i, d_{i+1}) = \frac{p_k^{(0,1)}(d_i)(1-p_k^{(1,0)}(d_{i+1}))}{p_k^{(0,1)}(d_i)(1-p_k^{(1,0)}(d_{i+1})) + (1-p_k^{(0,1)}(d_i))p_k^{(0,1)}(d_{i+1})} \tag{3.22}$$

and

$$p_k^2(d_i, d_{i+1}) = \frac{p_k^{(0,1)}(d_i)p_k^{(1,0)}(d_{i+1})}{p_k^{(0,1)}(d_i)p_k^{(1,0)}(d_{i+1}) + (1-p_k^{(0,1)}(d_i))(1-p_k^{(0,1)}(d_{i+1}))} \tag{3.23}$$

for $i \in \{2, 3, \ldots, \ell-1\}$. Hence, besides the p.m.f. of the first local decision, $P_{U_1|H}(u_1|H_k)$, as long as $\ell \geq 2$ the MRF associated with the hypothesis H_k is uniquely determined by $\{p_k^{(0,1)}(d_i), p_k^{(1,0)}(d_i)\}_{i=2}^{\ell}$, or equivalently by $p_k^{1'}(d_2, d_3)$, $p_k^2(d_2, d_3)$ and $\{p_k^t(d_i, d_{i+1})\}_{i=2}^{\ell-1}$ with $t \in \{0, 1\}$. Note that the rest of the conditional probabilities, i.e. $\{p_k^t(d_i, d_{i+1})\}_{i=3}^{\ell-1}$ with $t \in \{1', 2\}$, can be obtained through (3.67) and (3.68) shown in Appendix 3.A and obtained after performing some algebraic manipulations on (3.20)-(3.23). At the same time, as a result of the previous biunique correspondence, for $i \in \{2, 3, \ldots, \ell\}$ we have that $\psi(u_i, u_{i-1}|\boldsymbol{d}, H_k)$ matches the conditional p.m.f. of U_i given $\boldsymbol{D}$ and U_{i-1} in a non-homogenous Markov chain with transition probabilities satisfying (3.20)-(3.23), i.e.

$$\begin{aligned}\psi(u_i, u_{i-1}|\boldsymbol{d}, H_k) &= P_{U_i|U_{i-1},\boldsymbol{D},H}(u_i|u_{i-1}, \boldsymbol{d}, H_k) = P_{U_i|U_{i-1},D_i,H}(u_i|u_{i-1}, d_i, H_k) \\ &= p_k^{(0,1)}(d_i)\delta[u_i - 1]\delta[u_{i-1}] + p_k^{(1,0)}(d_i)\delta[u_i]\delta[u_{i-1} - 1] \\ &\quad + (1-p_k^{(1,0)}(d_i))\delta[u_i - 1]\delta[u_{i-1} - 1] \\ &\quad + (1-p_k^{(0,1)}(d_i))\delta[u_i]\delta[u_{i-1}]\end{aligned} \tag{3.24}$$

with $i \in \{2, 3, \ldots, \ell\}$, $k \in \{0, 1\}$ and $\delta[n]$ denoting the discrete time Kronecker delta function. Due to the fact that the local decision U_i is conditionally independent of the rest of the sensor spacings given D_i and U_{i-1}, the compatibility function $\psi(u_i, u_{i-1}|\boldsymbol{d}, H_k)$ only depends on D_i, U_{i-1}

and H_k. From (3.14) and (3.15), (3.24) implies that $Z_k = 1$. Additionally, taking into account that the sensor spacing are i.i.d. according to $f_D(d)$, substituting (3.24) in (3.15) yields

$$\begin{aligned}\overline{\psi}(u_i, u_{i-1}|H_k) &= \int_{\mathcal{D}} P_{U_i|U_{i-1},D_i,H}(u_i|u_{i-1}, d_i, H_k)\, f_D(d_i)\, \mathrm{d}d_i \\ &= (1-\overline{b}_k)\delta[u_i - 1]\delta[u_{i-1} - 1] + (1-\overline{a}_k)\delta[u_i]\delta[u_{i-1}] \\ &\quad + \overline{a}_k\delta[u_i - 1]\delta[u_{i-1}] + \overline{b}_k\delta[u_i]\delta[u_{i-1} - 1]\end{aligned} \tag{3.25}$$

where

$$\overline{a}_k = \overline{P}(U_i = 1|U_{i-1} = 0, H_k) = \int_{\mathcal{D}} p_k^{(0,1)}(d_i)\, f_D(d_i)\, \mathrm{d}d_i, \tag{3.26}$$

$$\overline{b}_k = \overline{P}(U_i = 0|U_{i-1} = 1, H_k) = \int_{\mathcal{D}} p_k^{(1,0)}(d_i)\, f_D(d_i)\, \mathrm{d}d_i \tag{3.27}$$

are parameters linked with the conditional probabilities responsible for the correlation structure of the local decisions under the hypothesis H_k. Therefore, as a consequence of assuming that the non-homogeneity of the random field associated with $P_{\boldsymbol{U}|\boldsymbol{D},H}(\boldsymbol{u}|\boldsymbol{d}, H_k)$ is exclusively caused by the set of distances $\{d_i\}_{i=2}^{\ell}$, from (3.13) we have that $\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k)$ is the joint p.m.f. of a homogeneous 1-D MRF uniquely determined by the conditional probabilities

$$\overline{p}_k^0 = \overline{P}(U_i = 1|U_{i-1} = U_{i+1} = 1, H_k), \tag{3.28}$$

$$\overline{p}_k^1 = \overline{P}(U_i = 1|U_{i-1} = 1, U_{i+1} = 0, H_k). \tag{3.29}$$

In this way, we are able to arrive at the following theorem.

Theorem 1. *Suppose that, with* $k \in \{0, 1\}$*,*

- $P_{U_1|H}(u_1|H_0)$ *is absolutely continuous w.r.t.* $P_{U_1|H}(u_1|H_1)$*, i.e.,*

$$P_{U_1|H}(u_1|H_0) \ll P_{U_1|H}(u_1|H_1).$$

- $0 < p_k^t(d_i, d_{i+1}) < 1$ *for* $t \in \{0, 1\}$ *and* $i \in \{2, 3, \ldots, \ell\}$.

- $0 < p_k^t(d_2, d_3) < 1$ *for* $t \in \{1', 2\}$.

Then, given a fixed constraint $P_{_{FA}} \leq \alpha \in (0,1)$*, and a discrete or continuous arbitrary p.d.f.,* $f_D(d)$*, defined on* $\mathcal{D}$*, the best Neyman-Pearson error exponent for the distributed detection problem given by* (3.7) *is*

$$\begin{aligned} K &= \mu_0(0)\,\overline{a}_0 \ln\left(\frac{\overline{a}_0}{\overline{a}_1}\right) + (1-\mu_0(0))\,(1-\overline{b}_0)\ln\left(\frac{1-\overline{b}_0}{1-\overline{b}_1}\right) \\ &\quad + \mu_0(0)(1-\overline{a}_0)\ln\left(\frac{1-\overline{a}_0}{1-\overline{a}_1}\right) + (1-\mu_0(0))\,\overline{b}_0 \ln\left(\frac{\overline{b}_0}{\overline{b}_1}\right) \\ &= D(\overline{P}(u_i|u_{i-1},H_0)||\overline{P}(u_i|u_{i-1},H_1)) \end{aligned} \tag{3.30}$$

where

- $\overline{a}_k$ *and* $\overline{b}_k$ *are obtained through* (3.26) *and* (3.27) *respectively.*
- $p_k^{(0,1)}(d_i)$ *and* $p_k^{(1,0)}(d_i)$ *satisfy* (3.20)-(3.23)*.*
- $\mu_0(0)$ *is defined as*

$$\mu_0(0) = \overline{P}(U=0|H_0) = \frac{\overline{b}_0}{\overline{b}_0 + \overline{a}_0} \tag{3.31}$$

  and denotes the unique stationary probability of deciding the null hypothesis when the MRF associated with H_0 *is present.*

- $D(\overline{P}(u_i|u_{i-1},H_0)||\overline{P}(u_i|u_{i-1},H_1))$ *corresponds with the conditional Kullback-Leibler divergence between* $\overline{P}_{U_i|U_{i-1},H}(u_i|u_{i-1},H_0)$ *and* $\overline{P}_{U_i|U_{i-1},H}(u_i|u_{i-1},H_1)$ *in the stationary regime.*

Proof. See the Appendix 3.A. □

3.3.2 Correlation parameter.

Among the different physical and design parameters of the network we are specially interested in studying the detection performance of the network versus the dependence among the local quantizations performed by the sensors of the network. However, there is not any single index that measures the dependence among the random variables that form a MRF. Motivated by this last fact, we firstly develop the correlation index defined in [Drakopoulos and Lee, 1991] for the case of two binary local decisions,

U_{i-1} and U_i, corresponding to the steps $i-1$ and i of a homogeneous binary MRF where the joint p.m.f. is characterized through (3.13) and (3.25)-(3.27)

$$\begin{aligned}\overline{\rho}_k^{\,i-1,i} &= \frac{E_{\boldsymbol{D}}\{E_{U_i,U_{i-1}|\boldsymbol{D},H_k}\{U_iU_{i-1}\}\}}{E_{\boldsymbol{D}}\{E_{U_{i-1}|\boldsymbol{D},H_k}\{U_{i-1}\}\}\left(1-E_{\boldsymbol{D}}\{E_{U_i|\boldsymbol{D},H_k}\{U_i\}\}\right)} \\ &\quad - \frac{E_{\boldsymbol{D}}\{E_{U_i|\boldsymbol{D},H_k}\{U_i\}\}\cdot E_{\boldsymbol{D}}\{E_{U_j|\boldsymbol{D},H_k}\{U_j\}\}}{E_{\boldsymbol{D}}\{E_{U_{i-1}|\boldsymbol{D},H_k}\{U_{i-1}\}\}\left(1-E_{\boldsymbol{D}}\{E_{U_i|\boldsymbol{D},H_k}\{U_i\}\}\right)} \\ &= \frac{(1-\overline{b}_k)\cdot\overline{P}_{U_{i-1}|H}(u_{i-1}=1|H_k)}{\overline{P}_{U_{i-1}|H}(u_{i-1}=1|H_k)\cdot\left(1-\overline{P}_{U_i|H}(u_i=1|H_k)\right)} \\ &\quad - \frac{\overline{P}_{U_{i-1}|H}(u_{i-1}=1|H_k)\cdot\overline{P}_{U_i|H}(u_i=1|H_k)}{\overline{P}_{U_{i-1}|H}(u_{i-1}=1|H_k)\cdot\left(1-\overline{P}_{U_i|H}(u_i=1|H_k)\right)} \\ &= \frac{(1-\overline{b}_k)-\overline{P}_{U_i|H}(u_i=1|H_k)}{1-\overline{P}_{U_i|H}(u_i=1|H_k)}.\end{aligned} \tag{3.32}$$

Now, we use $\overline{\rho}_k^{\,i-1,i}$ in order to derive a parameter that captures the mean correlation strength among the neighbour local decisions involved in the distributed detection problem (3.7). In particular, considering that the local decisions might be dependent under both hypothesis, that their joint p.m.f. under H_k is equal to $\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k)$, and that the number of sensors that form the network is very large, we obtain

$$\overline{\rho} = \begin{cases} P(H_0)\overline{\rho}_0 + P(H_1)\overline{\rho}_1 & \text{When the local decisions} \\ & \text{are dependent under } H_0 \text{ and } H_1. \\ \overline{\rho}_k & \text{When the local decisions} \\ & \text{are only dependent under } H_k. \end{cases} \tag{3.33}$$

where, for $k \in \{0,1\}$, $P(H_k) \in (0,1)$ equals the prior probability of hypothesis H_k, and $\overline{\rho}_k$ is the mean correlation between the binary local decisions of two consecutive sensors under H_k

$$\begin{aligned}\overline{\rho}_k &= \lim_{\ell\to\infty}\frac{1}{\ell}\sum_{i=2}^{\ell}\overline{\rho}_k^{\,i-1,i} = \lim_{\ell\to\infty}\frac{1}{\ell}\sum_{i=2}^{\ell}\frac{(1-\overline{b}_k)-\overline{P}_{U_i|H}(u_i=1|H_k)}{1-\overline{P}_{U_i|H}(u_i=1|H_k)} \\ &\overset{(a)}{=} \lim_{\ell\to\infty}\frac{1}{\ell}\sum_{i=2}^{i_0}\frac{(1-\overline{b}_k)-\overline{P}_{U_i|H}(u_i=1|H_k)}{1-\overline{P}_{U_i|H}(u_i=1|H_k)} \\ &\quad + \lim_{\ell\to\infty}\frac{\ell-i_0}{\ell}\frac{(1-\overline{b}_k)+\mu_k(0)-1}{\mu_k(0)}\end{aligned}$$

$$\overset{(b)}{=} \frac{(1-\overline{b}_k)+\mu_k(0)-1}{\mu_k(0)} = 1-\overline{a}_k-\overline{b}_k \in [0,1]. \tag{3.34}$$

In the previous equation (a) follows since the marginal p.m.f. of the i-th local decision conditioned on H_k, $\overline{P}_{U_i|H}(u_i|H_k)$, matches the unique and stationary p.m.f. of the underlying MRF as long as it is regular and $i > i_0$ for some bounded $i_0 < \infty$ (see [Kemeny and Snell, 1976]). Moreover, taking into account that $i_0 < \infty$ as well as the regularity of the corresponding MRF, we have that

$$\sum_{i=2}^{i_0} \frac{(1-\overline{b}_k)-\overline{P}_{U_i|H}(u_i=1|H_k)}{1-\overline{P}_{U_i|H}(u_i=1|H_k)} < \infty. \tag{3.35}$$

Thus, (b) occurs because the first term in the r.h.s. of (a) vanishes as ℓ goes to infinity.

Before concluding the subsection, it is worth noting that the parameters $\overline{\rho}_k$, and consequently $\overline{\rho}$, satisfy some important properties given in the definition of a correlation coefficient. On the one hand, we have that $\overline{\rho}_k$ equals zero if and only if the random variables involved in (3.7) are independent under H_k, i.e.

$$\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k) = \prod_{i=1}^{\ell} \overline{P}_{U_i|H}(u_i|H_k). \tag{3.36}$$

On the other hand, $\overline{\rho}_k$ is equal to one if and only if the random variables involved in (3.7) are maximally dependent under H_k, i.e., as the definition provided in [Drakopoulos and Lee, 1991]

$$\overline{P}(U_i=1|U_j=1,H_k) = 1 \tag{3.37}$$

and

$$\overline{P}(U_j=0|U_i=0,H_k) = 1 \tag{3.38}$$

for every pair of index $i,j \in \{1,2,\ldots,\ell\}$ with $\overline{P}(U_j=1|H_k) \leq \overline{P}(U_i=1|H_k)$ without loss of generality. Both statements can be verified in one of the directions if we substitute in (3.34) the constraints that, under the different scenarios of dependence, have to be satisfied by $\overline{a}_k$ and $\overline{b}_k$ with $k \in \{0,1\}$. In the case of independent local decisions under H_k, (3.36) results in $\overline{a}_k = 1-\overline{b}_k$ if we take into account (3.25). For maximally dependent local decisions

under H_k, the aforementioned constraints are $\overline{a}_k = \overline{b}_k = 0$ if we take (3.26) and (3.27) together with (3.37) and (3.38). Meanwhile, the converse direction is shown by noting that the constraints imposed on $\overline{a}_k$ and $\overline{b}_k$ when $\overline{\rho}_k = 0$ or $\overline{\rho}_k = 1$ are the same as the ones obtained when (3.36) occurs or (3.37)-(3.38) hold, respectively.

3.3.3 Properties of the error exponent.

In this subsection we examine the closed-form error exponent, K, when some particular situations, related to the dependence among the local decisions, happen. Specifically, through Corollary 1 we provide the behaviour of K when the dependence among the local decisions corresponds with the boundary values $\overline{\rho} = 0$, $\overline{\rho} = 1$, and $\overline{\rho}_1 = 1$ s.t. $\overline{\rho}_0 = 0$.

Corollary 1. *For a continuous or discrete arbitrary p.d.f. $f_D(d)$, the error exponent K has the following properties.*

1) When the local decisions are independent under both hypothesis, K is equal to the Kullback-Leibler divergence $D(B(\xi_0)||B(\xi_1))$ where, for $k \in \{0,1\}$, $B(\xi_k)$ denotes a Bernoulli random variable with probability of success $\xi_k = P(U_i = 1|H_k) \in (0,1)$,

$$K(\overline{\rho} = 0) = (1-\xi_0)\ln\left(\frac{1-\xi_0}{1-\xi_1}\right) + \xi_0 \ln\left(\frac{\xi_0}{\xi_1}\right) = D(B(\xi_0)||B(\xi_1)). \tag{3.39}$$

2) When the local decisions tend to be maximally dependent under H_0 and H_1, K converges to zero,

$$\lim_{\overline{\rho}\to 1} K = 0. \tag{3.40}$$

3) When the local decisions tend to be maximally dependent under H_1 at the same time as they are conditionally independent given H_0, K diverges,

$$\lim_{\overline{\rho}_1\to 1} K = \infty \quad s.t.\ \overline{\rho}_0 = 0. \tag{3.41}$$

Proof. See the Appendix 3.B. □

The proof of each claim of the corollary follows from computing the corresponding limit or closed-form expression of K through the constraints that, under each dependence scenario, are imposed on the parameters $\overline{a}_k$ and $\overline{b}_k$ with $k \in \{0,1\}$. Furthermore, analyzing each claim we can remark several aspects. If the local decisions are independent under each one of the two hypothesis, the derived error exponent, K, collapses to $D(B(\xi_0)||B(\xi_1))$, i.e. the Neyman-Pearson error exponent stated by the Stein's Lemma detailed in [Cover and Thomas, 2006] when the involved random variables are i.i.d. under H_0 and H_1. This result is consistent because, as it can be seen in (3.88), when the local decisions $\{U_i\}_{i=1}^{\ell}$ are independent under the hypothesis H_k, i.e. $\overline{\rho}_k = 0$, they are identically distributed according to a Bernoulli random variable, $B(\xi_k)$, where the probability of success equals ξ_k. An explanation to the result obtained in the second claim is obvious if we note that, for maximally dependent binary local quantizations under both hypothesis, the conditional joint p.m.f. of the local decisions given U_1 are identical under both hypothesis. On the contrary, the divergence of K shown in the third claim is less intuitive. Note that, when the local decisions are maximally dependent under H_1, from the results obtained in [Drakopoulos and Lee, 1991] the support of $\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1)$ only consists of the events where all the devices decide the same hypothesis. However, under conditional independence given the hypothesis H_0, in addition to those events the support of $\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)$ is formed by the rest of the 2^{ℓ} binary ℓ-tuples. Hence, $\overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_1) \ll \overline{P}_{\boldsymbol{U}|H}(\boldsymbol{u}|H_0)$ holds, and therefore, the existence of K can not be ensured although $\overline{P}_{U_1|H}(u_1|H_0) \ll \overline{P}_{U_1|H}(u_1|H_1)$ happens.

3.4 Simulation results

For two different models of the sensor spacing, equispaced sensors with failures and exponentially spaced devices with failures, we now study the change of the detection performance of the network described in Section 3.2 when some of its physical and design parameters vary. In order to do that, under each one of the distributions of the sensor spacings we analyze the effect of these parameters on evaluations of K at the same time as we corroborate our asymptotic conclusions by means of several numerical simulations.

For both the evaluations and the numerical experiments performed in this section we have considered the subsequent isotropic exponential model for

the correlation structure among the local decisions given the hypothesis H_k

$$p_k^{(0,1)}(d_i) = \xi_k(1 - m_k e^{-\gamma_k d_i}) \tag{3.42}$$

$$p_k^{(1,0)}(d_i) = (1 - \xi_k)(1 - m_k e^{-\gamma_k d_i}) \tag{3.43}$$

where, for $k \in \{0, 1\}$ and $i \in \{2, 3, \ldots, \ell\}$,

- ξ_k is the false alarm probability or detection probability of the local detectors when independence among the local decisions is assumed under H_0 or H_1 respectively.

- γ_k is a strictly positive constant that indicates the exponential rate of growth of the transition probabilities $p_k^{(0,1)}(d_i)$ and $p_k^{(1,0)}(d_i)$ as d_i increases.

- m_k is a strictly positive constant that, taking on values less than one, controls the correlation between U_{i-1} and U_i when the distance between the corresponding devices is zero, i.e. $d_i = 0$.

Note that the physical model considered for $p_k^{(0,1)}(d_i)$ and $p_k^{(1,0)}(d_i)$ fulfills some regularity conditions that typically appear in a detection scenario. On the one hand, there is a correlation strength decay as the devices become farther apart. In particular, due to the fact that $p_k^{(0,1)}(d_i)$ and $p_k^{(1,0)}(d_i)$ are monotonically concave increasing functions of d_i, the larger is the mean distance among the devices that form the network, the smaller is the value that $\overline{\rho}$ takes on. On the other hand we have that, through $m_k \in (0, 1)$, (3.42) and (3.43) model the well-known nugget effect described in [Misra and Tong, 2008]. As imposing a minimum distance between two neighbour devices, due to the nugget effect $p_k^{(0,1)}(d_i = 0) > 0$ and $p_k^{(1,0)}(d_i = 0) > 0$ occur for all $i \in \{2, 3, \ldots, \ell\}$. In this way, bearing in mind (3.37), (3.38) as well as the behaviour of $p_k^{(0,1)}(d_i)$ and $p_k^{(1,0)}(d_i)$ w.r.t. d_i, the local decisions, U_{i-1} and U_i, are never going to be maximally dependent when $m_k < 1$, and equivalently $\overline{\rho} < 1$ always holds.

In addition to the regularity conditions mentioned in the previous paragraph, we can realize that, given (3.42) and (3.43), the local decisions U_i and U_{i-1} are only independent under H_k when the distance between the corresponding nodes equals infinity, $d_i = \infty$. More precisely, when $p_k^{(0,1)}(d_i)$ and $p_k^{(1,0)}(d_i)$ follow (3.42) and (3.43) respectively, for $d_i = \infty$ and $k \in \{0, 1\}$

$$P_{U_i|U_{i-1},H}(u_i|u_{i-1}, H_k) = P_{U_i|H}(u_i|H_k) = (1 - \xi_k)\,\delta[u_i] + \xi_k\,\delta[u_i - 1] \tag{3.44}$$

happens, and therefore, $\overline{\rho}_k^{i-1,i} = 0$ since $p_k^{(0,1)}(\infty) = 1 - p_k^{(1,0)}(\infty)$. Moreover, before evaluating (3.30) for each one of the distributions considered for the sensor spacing, realize that the initial assumptions of Theorem 1 hold when mild conditions are imposed on the correlation model described in (3.42) and (3.43). By means of (3.20)-(3.23), those conditions result in $0 < p_k^{(0,1)}(d_i), p_k^{(1,0)}(d_i) < 1$ for all $k \in \{0,1\}$ and $i \in \{2,\ldots,\ell\}$ (see Appendix 3.A). As said before, the lower bound is satisfied as long as the aforementioned regularity conditions of a detection scenario appear. Meanwhile, taking into account that $p_k^{(0,1)}(d_i)$ and $p_k^{(1,0)}(d_i)$ are monotonically concave increasing functions of d_i, the upper bound only requires that the devices do not have perfect detection features in the sense of $\xi_k \in (0,1)$ for all $k \in \{0,1\}$.

3.4.1 Equispaced sensors with failures

In this case, we consider that the devices are equally spaced and that the distance $\overline{d}$ between two neighbour sensors is known by the data fusion center. We also assume that the local decision performed by a sensor may not be received by the fusion center with probability $q \in [0,1)$ independently from sensor to sensor. In this way, as the terminology used in [Misra and Tong, 2008], there are two kinds of devices in the network. One class is formed by the operational devices, in other words, those sensors whose local decisions have been successfully received by the fusion center (in our scenario, the ℓ sensors whose decisions are available at the fusion center). The other class includes the devices that are in failure, in other words, those sensors whose local decisions have not been received by the fusion center. Note that the failure of a specific device can have different causes. One of them could be the malfunction or battery depletion at the sensor. Other reason could be lost transmissions due to errors in the communication established between the sensor and the fusion center. Another cause could even be the use of probabilistic censoring schemes in which the devices perform and send their local decisions according to a probability chosen in order to satisfy some criteria, e.g. minimize the overall energy consumption s.t. a minimum detection performance.

Taking into account that the fusion rule (3.7) is only evaluated for the ℓ local decisions that have been successfully received, the p.d.f. $f_D(d_i)$ describes the spacing among the corresponding operational devices, and

consequently takes the expression

$$f_D(d_i) = f_D(d) = \begin{cases} (1-q)\, q^{n-1} & d = n\,\overline{d} \\ 0 & d \neq n\,\overline{d} \end{cases} \tag{3.45}$$

where $n \in \{1, 2, \ldots\}$ and $i \in \{2, 3, \ldots, \ell\}$. Note that the devices are deterministically equispaced when $q = 0$, and therefore

$$f_D(d_i) = f_D(d) = \delta[d - \overline{d}]. \tag{3.46}$$

Now, if we characterize the error exponent given in (3.30) when the devices are equispaced and each one of them has a failure rate q, for a specific $k \in \{0, 1\}$ we obtain

$$\overline{a}_k = \xi_k \left[\frac{1 - e^{-\gamma_k \overline{d}}\,(q\,(1 - m_k) + m_k)}{1 - q\, e^{-\gamma_k \overline{d}}} \right] \tag{3.47}$$

and

$$\overline{b}_k = (1 - \xi_k) \left[\frac{1 - e^{-\gamma_k \overline{d}}\,(q\,(1 - m_k) + m_k)}{1 - q\, e^{-\gamma_k \overline{d}}} \right] \tag{3.48}$$

as a result of substituting (3.42), (3.43) and (3.45) in (3.26) and (3.27) respectively. Thus, the evaluation of (3.34) for (3.47) and (3.48) yields the mean correlation strength

$$\overline{\rho}_k = \frac{(1-q)\, m_k e^{-\gamma_k \overline{d}}}{1 - q\, e^{-\gamma_k \overline{d}}}. \tag{3.49}$$

Based on the characterization of K through (3.47)-(3.48) Figure 3.3 plots how the Neyman-Pearson error exponent behaves when the mean correlation strength expressed in (3.49) is varied. In this figure, theoretical curves are generated when $q = 0.3$, ξ_1 takes on the values $\{0.7, 0.8, 0.9\}$, and the local decisions are only dependent under hypothesis H_1, i.e. $\overline{\rho} = \overline{\rho}_1$. As it can be seen, for all the represented values of ξ_1, K initially decreases as $\overline{\rho}$ increases, and after a specific value $\overline{\rho}^{*}$, it increases as $\overline{\rho}$ approaches one. On the one hand, the initial decreasing behaviour is due to the fact that there is a loss of discrimination between H_0 and H_1 when the information provided by each sensor is more and more correlated without an improvement in the detection performance of the sensors. On the other hand, the increase of K

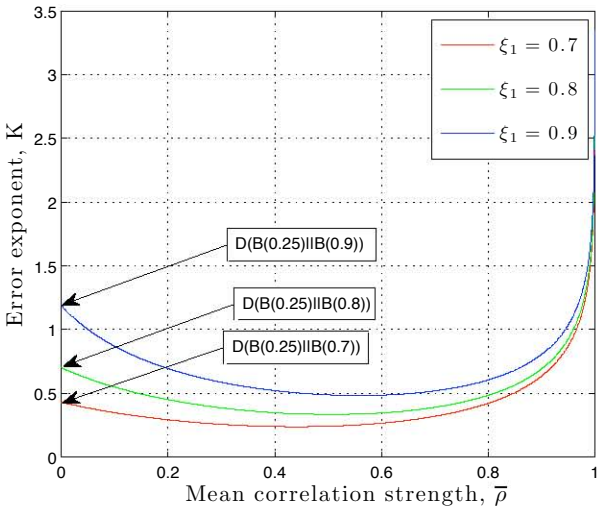


Figure 3.3: *For a 1-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of the mean correlation strength among neighbour binary decisions, $\overline{\rho}$, when they are only dependent under H_1, $\xi_1 = \{0.7, 0.8, 0.9\}$ and the corresponding devices are equally spaced with $\overline{d}$ equal to the distance between two neighbour nodes. Parameters: $\xi_0 = 0.25$, $\gamma_1 = 0.9$, $q = 0.3$, and $m_1 = 1 - 10^{-4}$.*

in $\overline{\rho}$ happens because, if $\overline{\rho}$ goes to one, under H_1 all local detectors tend to decide the same hypothesis with an increasing probability. Meanwhile, due to the fact that $\overline{\rho}_0 = 0$, when H_0 is present each sensor performs its decision independently of the rest of the network. Thus, discriminating H_1 against H_0 is easier (see [Drakopoulos and Lee, 1991]).

In order to find a closed-form expression for the mean correlation strength, $\overline{\rho}^*$, at which K achieves a minimum in a specific curve of Figure 3.3, we have performed a study of the monotonicity of K when $\overline{\rho} = \overline{\rho}_1$. After straightforward algebraic manipulations, the analysis of the monotonicity of K reveals that the derivative of the error exponent vanishes for two values of $\overline{\rho}$. However, at most one of these two points belongs to the domain where $\overline{\rho}$ is defined (see (3.34)). In the case of existing one point of $\overline{\rho}$ that belongs to the close interval $[0, 1]$, that point matches $\overline{\rho}^*$ and satisfies

$$\overline{\rho}^* = 1 - \frac{\overline{a}_1^*}{\xi_1} \tag{3.50}$$

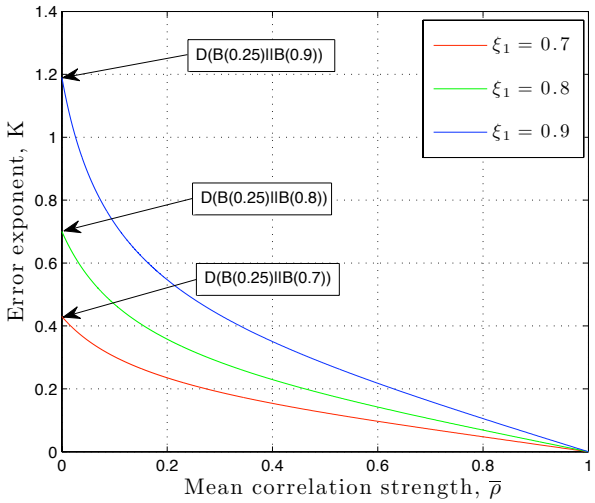


Figure 3.4: *For a 1-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of the mean correlation strength among the neighbour binary decisions, $\overline{\rho}$, when they are dependent under H_1 and H_0, $\xi_1 = \{0.7, 0.8, 0.9\}$ and the corresponding devices are equally spaced with $\overline{d}$ equal to the distance between two neighbour nodes. Parameters: $\xi_0 = 0.25$, $\gamma_0 = 2$, $\gamma_1 = 0.9$, $q = 0.3$, $P(H_k) = 0.5$ and $m_k = 1 - 10^{-4}$ with $k \in \{0, 1\}$.*

where, given the following definition,

$$\varphi(\xi_0, \xi_1) \triangleq \xi_0^4 - 4\,(1-\xi_1)\,\xi_0^3 + 2\,(2-\xi_1-2\,\xi_1^2)\,\xi_0^2 - 4\,\xi_1\,(1-\xi_1)\,\xi_0 + \xi_1^2, \tag{3.51}$$

after simple calculations $\overline{a}_1^*$ is expressed as

$$\overline{a}_1^* = \frac{2\,\xi_0\,(1-\xi_1) + \xi_1 - \xi_0^2 - \sqrt{\varphi(\xi_0, \xi_1)}}{2\,(1-\xi_1)}. \tag{3.52}$$

From (3.50)-(3.52) we can observe that $\overline{\rho}^*$ only depends on the detection features of the devices when their observations are mutually independent under both hypothesis. What is more, explaining other characteristic of the behaviour of K as a function of $\overline{\rho}$ it can be proved that $\overline{\rho}^*$ is shifted closer to one as the ratio ξ_1/ξ_0 increases.

Concluding the analysis of Figure 3.3, we can check that the results obtained in two of the three claims of Corollary 1 occur. In particular, for all

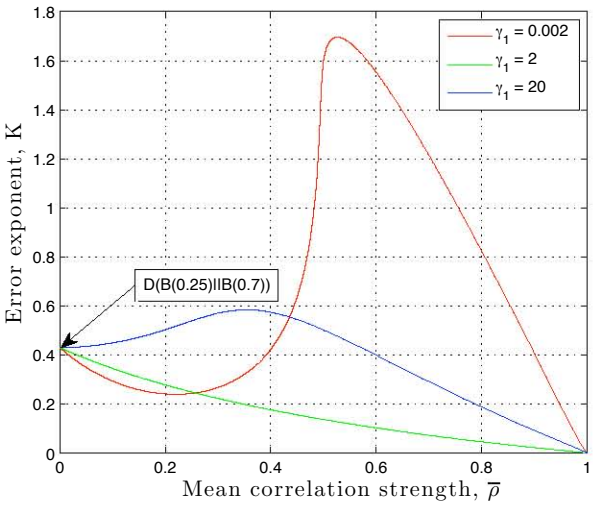


Figure 3.5: *For a 1-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of the mean correlation strength among the neighbour binary decisions, $\overline{\rho}$, when they are dependent under H_1 and H_0, $\gamma_1 = \{2 \cdot 10^{-3}, 2, 20\}$ and the corresponding devices are equally spaced with $\overline{d}$ equal to the distance between two neighbour nodes. Parameters: $\xi_0 = 0.25$, $\xi_1 = 0.7$, $\gamma_0 = 2$, $q = 0.3$, $P(H_k) = 0.5$ and $m_k = 1 - 10^{-4}$ with $k \in \{0, 1\}$.*

the plotted values of ξ_1, in Figure 3.3 the derived error exponent collapses to the Kullback-Leibler divergence, $D(B(\xi_0)||B(\xi_1))$, when the local decisions are independent under both hypothesis. Furthermore, as the third claim of Corollary 1 asserts, in Figure 3.3 we can note that K diverges when the local decisions are independent under H_0 and tend to be maximally dependent under H_1.

In Figure 3.4 we make the same analysis as in Figure 3.3 with the proviso that the local decisions are also dependent under H_0. For each one of the curves plotted in Figure 3.4 we can observe that K decreases as the correlation strength, $\overline{\rho}$, increases from the case of independent local decisions, $\overline{\rho} = 0$, to the case of maximally dependent local decisions, $\overline{\rho} = 1$. In addition, it can be seen that the amount of decrease in K becomes smaller as $\overline{\rho}$ increases. Nevertheless, the previous behaviour of K w.r.t. $\overline{\rho}$ is not common to all the possible set of values that can take the parameters involved in (3.47) and (3.48). This can be checked in Figure 3.5 where K achieves different number of maximums and minimums depending on the value that

γ_1 takes on. To be more precise, when the local decisions are dependent under both hypothesis, $\xi_0 = 0.25$, $\xi_1 = 0.7$, $\gamma_0 = 2$, $q = 0.3$, $P(H_k) = 0.5$ and $m_k = 1 - 10^{-4}$ with $k \in \{0, 1\}$, the derivative of K can be vanished for two, one or zero values of $\overline{\rho}$ if γ_1 is equal to 0.002, 20 or 2 respectively.

If the local decisions are dependent under H_0 and H_1, a study of the monotonicity of K w.r.t. $\overline{\rho}$ is not analytically tractable. However, as it happens in Figure 3.3, both in Figure 3.4 and Figure 3.5, when $\overline{\rho}$ is equal to zero, K converges to the maximal error exponent derived in the Stein's Lemma for the i.i.d. case. Moreover, as it is expected from the second claim of Corollary 1, in both figures K goes to zero as $\overline{\rho}$ approaches one. An interpretation for this convergence of K can be found by realizing that excessive dependence among the local decisions under both hypothesis makes the reading of an extra device provide the fusion center with a lot of redundant information when discriminating H_1 against H_0.

Before continuing with the analysis of more results, looking at Figures 3.3 and 3.4 it is of great value to note that, as it can be expected, the better is the detection performance of the devices that form the network, the larger is the error exponent. This behavior is consistent and can be expected if we note that the overall detection performance of the network has to improve when the probability of error of the local detectors is closer to zero under the assumption of independent sensor observations.

Next, in order to verify the behavior of the probability of miss-detection predicted by the previous evaluations of the error exponent, we provide some numerical results. Under the sensor spacing distribution (3.45) we consider scenarios where the local decisions are only dependent under H_1, as well as scenarios where a mechanism of correlation is also present under the null hypothesis. For all the values of $\overline{\rho}$ considered in the different scenarios the probability of false alarm, $P_{_{FA}}$, is set at 10^{-2} when $q = 0.3$, $\xi_1 = 0.7$ and $\xi_0 = 0.25$.

Figure 3.6 shows the overall probability of miss-detection, P_M, as a function of the number, ℓ, of local decisions successfully received by the data fusion center. As it happened for the theoretical error exponent when the local decisions are only dependent under H_1, in Figure 3.6 the simulated detection performance has an unimodal behaviour w.r.t. the mean correlation strength, $\overline{\rho}$. It can be seen that the slope for the probability of miss-detection initially decreases as $\overline{\rho}$ increases from zero to a specific value at which the system detection performance is minimum. At the same time, as long as the mean correlation strength is greater than this value, it can be also observed

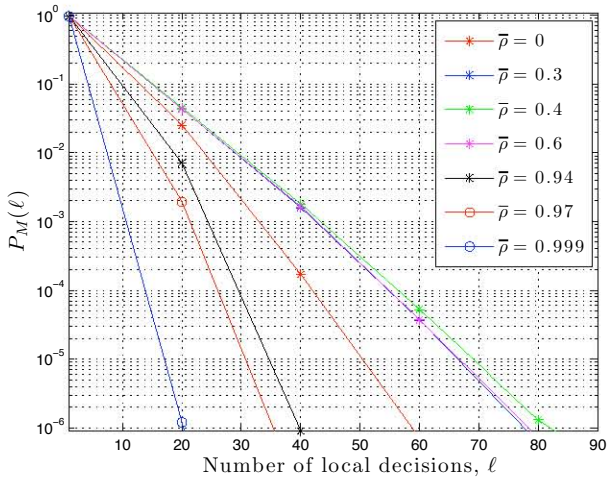
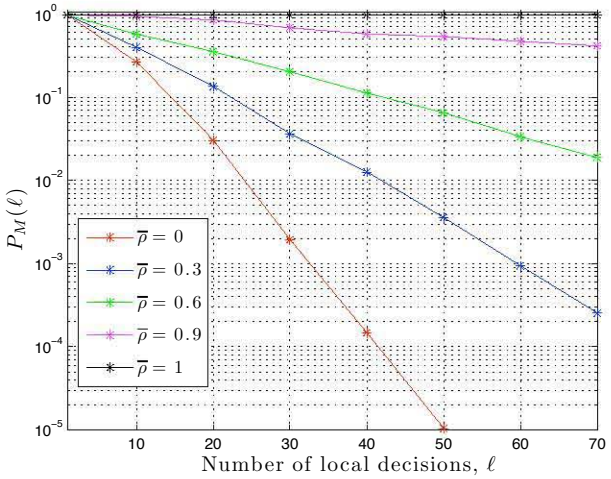


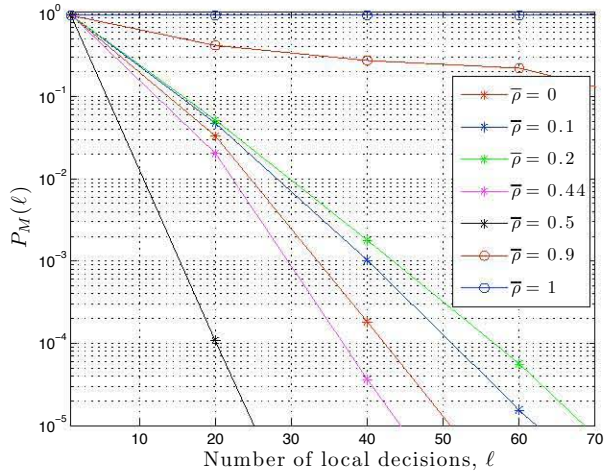
Figure 3.6: *For a 1-D parallel network, system probability of miss-detection, P_M, as a function of the number of local decisions, ℓ, when they are only dependent under H_1, $\overline{\rho} \in \{0, 0.3, 0.4, 0.6, 0.94, 0.97, 0.999\}$, and the corresponding devices are equally spaced with $q = 0.3$. Parameters: $\xi_0 = 0.1$, $\xi_1 = 0.98$, $\gamma_0 = 2$, $\gamma_1 = 0.9$, and $m_1 = 1 - 10^{-4}$ and $P(H_k) = 0.5$ with $k \in \{0, 1\}$.*

that the slope of $\log[P_M]$ increases as $\overline{\rho}$ approaches one. Not only note the similar magnitude between the corresponding evaluation of $\overline{\rho}^*$ and the mean correlation strength that makes the slope for the error decay be minimum in Figure 3.6. In addition, as it can be predicted during the analysis of the theoretical error exponent plotted in Figure 3.3 for $\xi_0 = 0.25$ and $\xi_1 = 0.7$, realize that the slopes for the error decay are similar when $\overline{\rho} = 0.3$ and $\overline{\rho} = 0.6$.

Finally, in Figure 3.7 we plot how the probability of miss-detection, P_M, changes when the number of local decisions, ℓ, increases in a scenario where the local decisions are dependent under both hypothesis. If we look at the slope shown by $\log[P_M]$ under each one of the values considered for the mean correlation strength, $\overline{\rho}$, we can again assert that the numerical results are seen to coincide with the theoretical predictions based on K. To be precise, being consistent with the third claim of Corollary 1, in Figures 3.7(a) and 3.7(b) we can note that the probability of miss-detection is sub-exponentially decaying when $\overline{\rho}$ equals one. In addition, we can observe that, depending on the values that the ratio γ_1/γ_0 takes on, there can be a



(a)



(b)

Figure 3.7: *For a 1-D parallel network, system probability of miss-detection, P_M, as a function of the number of local decisions, ℓ, when they are dependent under H_0 and H_1, the devices are equally spaced with $q = 0.3$, and $\overline{\rho}$ takes on different values. Parameters: (a) $\alpha = 10^{-2}$, $\xi_0 = 0.25$, $\xi_1 = 0.7$, $\gamma_0 = 2$, $\gamma_1 = 0.9$, $m_0 = m_1 = 1 - 10^{-4}$, $P(H_0) = P(H_1) = 0.5$, (b) $\gamma_1 = 2 \cdot 10^{-3}$ as well as the rest of the parameters used for (a).*

different number of mean correlation strengths at which the slope of $\log[P_M]$ achieves a maximum or a minimum. For instance, in Figure 3.7(a), when $\gamma_0 = 2$ and $\gamma_1 = 0.9$, we can realize that the exponential decay of the probability of miss-detection w.r.t. ℓ is maximized when $\overline{\rho} = 0$, and that the slope for the error decay is monotonically decreasing as $\overline{\rho}$ approaches one. However, agreeing with analytical evaluations of K (see Figure 3.5), the previous decreasing behaviour does not always happen in Figure 3.7(b) when $\gamma_0 = 2$ and $\gamma_1 = 2 \cdot 10^{-3}$. In this case, besides a decrease of the slope of $\log[P_M]$ as $\overline{\rho}$ increases from zero or as $\overline{\rho}$ approaches one, there is an interval of mean correlation strengths where the detection performance of the network improves as $\overline{\rho}$ rises. Due to this lack of monotonicity, the slope of $\log[P_M]$ achieves a local minimum and a global maximum for two different values of $\overline{\rho}$ whose magnitudes are again close to the ones predicted by the corresponding evaluation of K.

3.4.2 Exponentially spaced sensors with failures

In this subsection we are going to undertake the deployment of the 1-D network according to a Poisson distribution with rate λ sensors per unit of space. This way, the probability that there are ℓ sensors in a 1-D space $(d_0, d_0 + \Delta]$ is given by

$$P((N(d_0 + \Delta) - N(d_0)) = \ell) = \frac{e^{-\lambda\,\delta}\,(\lambda\,\delta)^{\ell}}{\ell!} \tag{3.53}$$

where $\ell \in \{0, 1, 2, \ldots\}$ and where $N(d_0+\Delta)-N(d_0)$ is the number of sensors deployed in the 1-D space $(d_0, d_0 + \Delta]$. This means that, if the sensors can be in failure with probability $q < 1$ independently from sensor to sensor, this time the sensor spacings are exponentially distributed with placement density $\lambda\,(1 - q)$. Thus, with $i \in \{2, 3, \ldots, \ell\}$ and $\lambda = 1/E_D\{d\} = 1/\overline{d}$,

$$f_D(d_i) = f_D(d) = \lambda\,(1 - q)\,e^{\lambda\,(1-q)\,d}; \quad d \geq 0. \tag{3.54}$$

For (3.54) and the correlation model expressed in (3.42) and (3.43), using (3.26)-(3.27) and skipping the corresponding straightforward proofs, the error exponent (3.30) is totally characterized through

$$\overline{a}_k = \xi_k \left[1 - \frac{m_k\,\lambda\,(1 - q)}{\gamma_k + \lambda\,(1 - q)}\right] \tag{3.55}$$

and

$$\overline{b}_k = (1-\xi_k)\left[1 - \frac{m_k\,\lambda\,(1-q)}{\gamma_k + \lambda\,(1-q)}\right] \tag{3.56}$$

with $k \in \{0,1\}$. Therefore, after substituting (3.55) and (3.56) in (3.34) we get

$$\overline{\rho}_k = \frac{m_k\,\lambda\,(1-q)}{\gamma_k + \lambda\,(1-q)}. \tag{3.57}$$

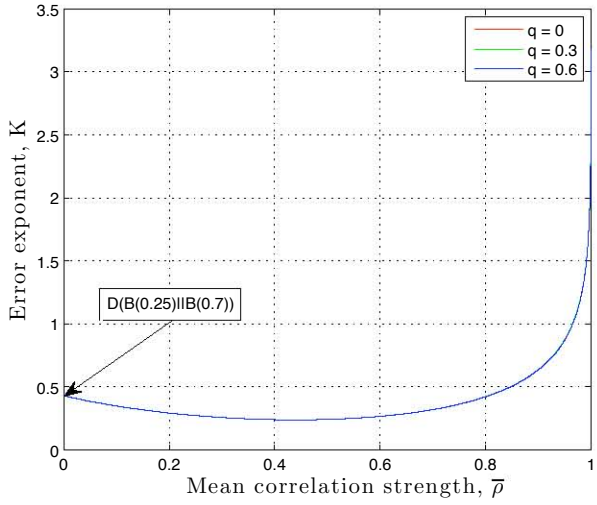
Centering on the sensor spacing model (3.54), Figure 3.8 shows how the error exponent K changes when the mean correlation strength provided in (3.57) varies. Theoretical curves of Figure 3.8(a) and (b) have been generated when q is equal to $\{0, 0.3, 0.6\}$. Again, Figure 3.8(a) plots the error exponent when the local decisions are only dependent under H_1, whereas the curves in Figure 3.8(b) have been obtained when a correlation structure is present under both hypothesis. The first aspect we can remark regarding Figure 3.8 is that all the plotted curves again confirm the results asserted by the Corollary 1 when the dependence among the local decisions matches the boundary values $\overline{\rho} = 0$, $\overline{\rho} = 1$, and $\overline{\rho}_1 = 1$ s.t. $\overline{\rho}_0 = 0$. Moreover, when we study the monotonocity of K w.r.t. $\overline{\rho}$, we obtain results that are similar to the ones derived in the corresponding analysis of the preceding subsection. In Figure 3.8(a), when the local decisions are only dependent under H_1, the error exponent achieves a minimum for a specific value, $\overline{\rho}^*$, of the mean correlation strength. Specifically, as it happened for the other sensor spacing model, an analytic study reveals that the derivative of K w.r.t. $\overline{\rho}$ at most vanishes for one point, $\overline{\rho}^*$, belonging to the domain where $\overline{\rho}$ is defined and whose closed-form expression is

$$\overline{\rho}^* = 1 - \overline{a}_1^*/\xi_1 \tag{3.58}$$

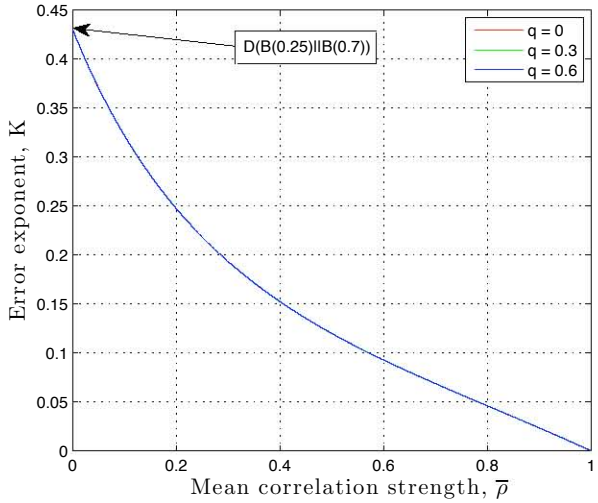
when

$$\overline{a}_1^* = \xi_1\,\frac{3\,\xi_0^2 - 2\,(1+\xi_1)\,\xi_0 + \xi_1 - \sqrt{\varphi(\xi_0,\xi_1)}}{\xi_0^2 - 2\,(1-\xi_1)\,\xi_0 + \xi_1 - 2\,\xi_1^2 - \sqrt{\varphi(\xi_0,\xi_1)}} \tag{3.59}$$

and $\varphi(\xi_0,\xi_1)$ is given by (3.51). Note that, although the definition of $\overline{a}_1^*$ differs from the one provided in (3.52), as we could notice for the previous sensor spacing distribution, $\overline{\rho}^*$ only depends on the detection features ξ_0 and ξ_1. Regarding the evolution of K w.r.t. $\overline{\rho}$ when the local decisions are dependent under both hypothesis, we again face with a problem that can



(a)



(b)

Figure 3.8: *For a 1-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of $\overline{\rho}$ when the devices are exponentially spaced with a failure rate $q \in \{0, 0.3, 0.6\}$. In (a) a correlation structure is only present under H_1 whereas in (b) the dependence among the local decisions appears in both H_0 and H_1. Parameters: (a) $\xi_0 = 0.25$, $\xi_1 = 0.7$, $\gamma_1 = 0.9$, $m_1 = 1 - 10^{-4}$, $P(H_0) = P(H_1) = 0.5$, (b) $\gamma_0 = 2$, $m_0 = 1 - 10^{-4}$ as well as the parameters used for (a). There are three curves in every figure, but all of them overlap and appear as one.*

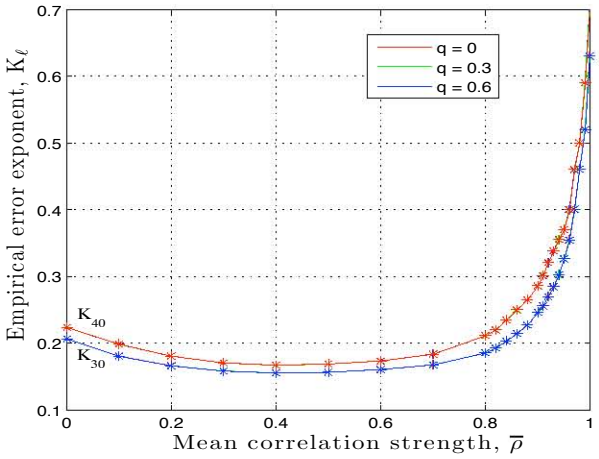
not be solved analytically. Only by means of numerical experiments and simulations similar to the ones shown in Figures 3.8(b) and 3.9, we have checked that, depending on the values taken by the parameters involved in (3.55) and (3.56), there are several mean correlation strengths, $\overline{\rho}$, at which K achieves a maximum or a minimum.

Subsequently, we study the effect of the failure rate, q, on K when the mean correlation strength, $\overline{\rho}$, is held fixed. Looking at Figures 3.8(a) and (b) we can observe that, given a specific mean correlation strength, $\overline{\rho}$, the derived error exponent, K, remains unaltered although the failure rate varies. This happens because the parameters $\overline{a}_k$ and $\overline{b}_k$, with $k \in \{0, 1\}$, do not depend on q when they are written in terms of $\overline{\rho}$ using the biunique maps established between $\overline{d}$ and $\overline{\rho}$ through (3.33) and (3.57). At the same time, we can conclude that, for the sensor spacing distribution (3.54), the dependence of K on q and $\overline{d}$ can be expressed via the mean correlation strength, $\overline{\rho}$, achieved with those parameters. To put it differently, whenever $\overline{\rho}$ remains unchanged and $f_D(d)$ equals (3.54), the detection performance of the Neyman-Pearson fusion rule (3.7) can be the same for different pair of values $(\overline{d}, q)$, or equivalently, for different design strategies. Considering the Neyman-Pearson fusion rule (3.7), we can extend this result to other sensor spacing distributions as well as other parameters that appear in a specific correlation model. Nevertheless, not all the scenarios have an associated $\overline{\rho}$ whose value describes the joint effect of several design and physical parameters of the network on K. In that case, although two different sets of physical and design parameters of the network result in the same mean correlation strength, $\overline{\rho}$, the error exponents of both scenarios do not have to be necessarily identical.

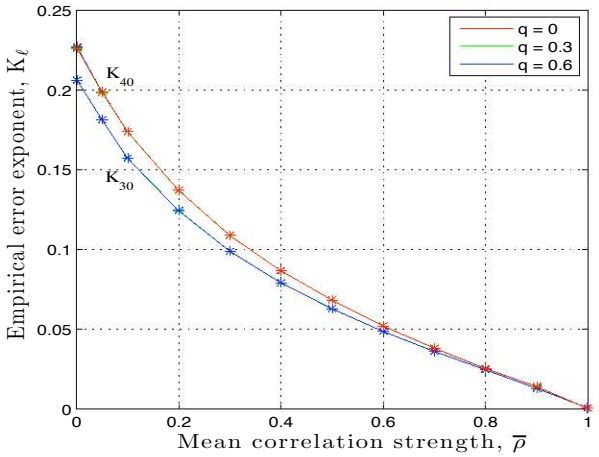
In order to end this section, we confirm the previous results by means of different numerical experiments. This time, as in [Misra and Tong, 2008] we do so by plotting the empirical error exponent

$$K_\ell = -\frac{1}{\ell} \ln (P_M) \tag{3.60}$$

as a function of the mean correlation strength, $\overline{\rho}$. In particular, for the same scenarios as in Figure 3.8, in Figure 3.9 we plot K_ℓ versus $\overline{\rho}$ when $P_{_{FA}} = 10^{-2}$ and $\ell = \{30, 40\}$. If we now analyze the numerically determined curves in both Figure 3.9(a) and Figure 3.9(b), we can verify that the behaviour predicted by K holds for K_ℓ. On the one hand, in a scenario where the local decisions are only dependent under H_1 the empirical error exponent K_ℓ



(a)



(b)

Figure 3.9: *For a 1-D parallel network, Neyman-Pearson empirical error exponent, K_ℓ, as a function of $\overline{\rho}$ when the devices are exponentially spaced with a failure rate $q \in \{0, 0.3, 0.6\}$. The numerically determined curves have been plotted for a system where $\ell \in \{30, 40\}$. In (a) a correlation structure is only present under H_1 whereas in (b) the dependence among the local decisions appears in both H_0 and H_1. Parameters: (a) $\xi_0 = 0.25$, $\xi_1 = 0.7$, $\gamma_1 = 0.9$, $m_1 = 1 - 10^{-4}$, $P(H_0) = P(H_1) = 0.5$, (b) $\gamma_0 = 2$, $m_0 = 1 - 10^{-4}$ as well as the parameters used for (a). In every figure there are three curves for every value of $\overline{\rho}$, but all of them overlap and appear as one.*

plotted in Figure 3.9(a) is seen to be unimodal in $\overline{\rho}$ independently of the failure rate q. On the other hand, in Figure 3.9(b), when $\gamma_0 = 2$, $\gamma_1 = 0.9$, $q = \{0, 0.3, 0.6\}$, and the local decisions have a correlation structure under both hypothesis, K_ℓ is monotonically decreasing in $\overline{\rho}$ and converges to zero as $\overline{\rho}$ approaches one. Looking at both figures, as predicted by our asymptotic analysis, in Figures 3.9(a) and 3.9(b) we can also realize that K_ℓ remains unaltered for different failures rates, q, as long as the resulting $\overline{\rho}$ is the same. In addition, we are able to note that the magnitude of K and K_ℓ do not coincide. However, as we can expect from the subsequent expression

$$\lim_{\ell \to \infty} K_\ell = K, \tag{3.61}$$

they are in relatively close agreement in both Figure 3.9(a) and Figure 3.9(b). Despite this last fact, if we take into account the rest of the numerical results as well as the ones obtained in similar simulations over a wide variety of parameters, we can conclude that, even for the case where ℓ is finite, K is useful when predicting the detection performance of the sensor network described in Section 3.2.

3.5 Summary

In this chapter we characterized a distributed detection system with parallel architecture and formed by a large number of devices as well as data fusion center. With no kind of cooperation and from its own observation each device performs a local decision for the underlying binary hypothesis testing problem. Afterwards, over an error free parallel access channel each sensor transmits its decision to a fusion center that makes a binary global decision, $u_{_{FC}}$, under the Neyman-Pearson formulation. We also considered that the devices are randomly deployed along a straight line, and that the corresponding sensor spacings are drawn independently from a common p.d.f. $f_D(d)$. Additionally, the local decisions were assumed to be dependent under both hypothesis because of the correlation structure of the phenomenon observed by the devices under each one of the two hypothesis. In particular, under both hypothesis the dependence among the local decisions is modelled by means of a 1-D MRF with nearest-neighbour dependency and binary state space. For this scenario, we firstly derived a closed-form error exponent for the Neyman-Pearson test performed at the fusion center when this one

only knows $f_D(d)$ regarding the sensor spacings. Afterwards, based on a single correlation index that captures the mean correlation strength among neighbour local decisions, some analytical properties of the derived error exponent were investigated. Finally, after choosing a physical model for the conditional probabilities of the MRF present under each hypothesis, we evaluated the derived error exponent for two different models of the sensor spacing: i) equispaced sensors with failures, and ii) exponentially spaced sensors with failures. In each model of the sensor spacing, by means of the corresponding evaluations of the error exponent and supported by several numerical experiments we analyzed how the error exponent behaves as different physical and design parameters of the network vary.

3.A Proof of Theorem 1

Since the error exponent for the Neyman-Pearson detector with a fixed level $\alpha \in (0,1)$ is given in implicit form by the asymptotic Kullback-Leibler rate (3.17), we focus on the calculation of this limit for the scenario described in Section 3.2. Taking into account (3.13), the limit (3.17) can be rewritten as follows

$$K = \lim_{\ell\to\infty} \frac{1}{\ell} \ln\left(\frac{Z_1}{Z_0}\right) + \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{i=2}^{\ell} \ln\left(\frac{\overline{\psi}(u_i, u_{i-1}|H_0)}{\overline{\psi}(u_i, u_{i-1}|H_1)}\right) \quad \text{(a.s. in } H_0\text{)}. \tag{3.62}$$

Using the characterization of $\overline{\psi}(u_i, u_{i-1}|H_k)$ provided in (3.25) for $k \in \{0,1\}$ and valid as long as (3.20)-(3.23) hold, we have that

$$\begin{aligned} K = {} & \ln\left(\frac{\overline{a}_0}{\overline{a}_1}\right) \lim_{\ell\to\infty} \frac{N(0,1|\boldsymbol{u}, H_0)}{\ell} \\ & + \ln\left(\frac{\overline{b}_0}{\overline{b}_1}\right) \lim_{\ell\to\infty} \frac{N(1,0|\boldsymbol{u}, H_0)}{\ell} + \ln\left(\frac{(1-\overline{a}_0)}{(1-\overline{a}_1)}\right) \lim_{\ell\to\infty} \frac{N(0,0|\boldsymbol{u}, H_0)}{\ell} \\ & + \ln\left(\frac{(1-\overline{b}_0)}{(1-\overline{b}_1)}\right) \lim_{\ell\to\infty} \frac{N(1,1|\boldsymbol{u}, H_0)}{\ell} \quad \text{(a.s. in } H_0\text{)} \end{aligned} \tag{3.63}$$

where

- $\overline{a}_k$ and $\overline{b}_k$ are obtained from (3.26) and (3.27) respectively.

- $N(u_{i-1}, u_i|\boldsymbol{u}, H_0)$ is the number of times that the pair of local decisions, (u_{i-1}, u_i), occurs in the sequence $\boldsymbol{u}$ defined on $\mathcal{U}^\ell$ and drawn from $\overline{P}_{U|H}(u|H_0)$.

Next, we focus on calculating an asymptotic closed-form expression for the empirical joint probability of the pair, (u_{i-1}, u_i), $\frac{N(u_{i-1},u_i|\boldsymbol{u},H_0)}{\ell}$. For this calculation, from [Spitzer, 1971] we firstly have to know that the homogeneous MRF associated with $\overline{P}_{U|H}(u|H_0)$ has a one-to-one map onto a homogeneous two-state Markov chain uniquely determined by $\overline{a}_0$ and $\overline{b}_0$. From this result, it is straightforward to show that

$$\lim_{\ell\to\infty} \frac{N(u_{i-1}, u_i|\boldsymbol{u}, H_0)}{\ell} = \overline{P}(U_i|U_{i-1}, H_0)\mu_0(u_{i-1}) \tag{3.64}$$

whenever the Markov chain associated with the MRF present under H_0 is regular with $\mu_0(u_{i-1})$ equal to the unique and stationary probability of state u_{i-1}. Now, if we substitute (3.31) and (3.64) into (3.63) we get the closed-form error exponent provided in Theorem 1.

Finally, besides guaranteeing the existence of the error exponent, K, we have to show the regularity of the Markov chain that has a biunique map with the MRF involved in the hypothesis H_0 of the detection process (3.7). On the one hand, the existence of K is ensured when

$$0 < \overline{a}_k, \overline{b}_k < 1 \tag{3.65}$$

happens for $k \in \{0, 1\}$. On the other hand, the regularity of the aforementioned Markov chain is held if

$$\overline{P}_{U_1|H}(u_1|H_0) \ll \overline{P}_{U_1|H}(u_1|H_1) \tag{3.66}$$

and (3.65) occurs for $k = 0$ (see [Kemeny and Snell, 1976]). Hence, proving that (3.65) holds for $k \in \{0, 1\}$, and that (3.66) occurs, we complete the proof of Theorem 1. In this case, the initial assumptions of the theorem are the starting point. Firstly, if we take into account that $\overline{P}_{U_1|H}(u_1|H_k) = P_{U_1|H}(u_1|H_k)$ for $k \in \{0, 1\}$, assuming $P_{U_1|H}(u_1|H_0) \ll P_{U_1|H}(u_1|H_1)$ makes (3.66) hold. Secondly, given the two subsequent expressions obtained from the classical probability theory

$$p_k^{1'}(d_{i+1}, d_{i+2}) = \frac{p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1})\ \left(1 - p_k^0(d_i, d_{i+1})\right)}{\varrho_k(d_i, d_{i+1}))} \tag{3.67}$$

$$p_k^2(d_{i+1}, d_{i+2}) = \frac{p_k^1(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1})\ \left(1 - p_k^0(d_i, d_{i+1})\right)}{\vartheta(d_i, d_{i+1}))} \tag{3.68}$$

with

$$\begin{aligned} \varrho_k(d_i, d_{i+1})) = p_k^0(d_i, d_{i+1})\ \left(1 - p_k^0(d_{i+1}, d_{i+2}) - p_k^1(d_i, d_{i+1})\right) \\ + p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1}) \end{aligned} \tag{3.69}$$

and

$$\begin{aligned} \vartheta(d_i, d_{i+1})) = p_k^0(d_i, d_{i+1})\ \left(1 - p_k^1(d_{i+1}, d_{i+2}) - p_k^1(d_i, d_{i+1})\right) \\ + p_k^1(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1}) \end{aligned} \tag{3.70}$$

we are going to check that $0 < p_k^{1'}(d_{i+1}, d_{i+2}), p_k^2(d_{i+1}, d_{i+2}) < 1$ if

$$0 < p_k^0(d_i, d_{i+1}), p_k^1(d_i, d_{i+1}) < 1$$

and

$$0 < p_k^0(d_{i+1}, d_{i+2}), p_k^1(d_{i+1}, d_{i+2}) < 1$$

happen for $k \in \{0, 1\}$ and $i \in \{2, 3, \ldots, \ell - 3, \ell - 2\}$. Now, we prove the lower and upper bounds of $p_k^{1'}(d_{i+1}, d_{i+2})$, while, for the sake of brevity, we will omit the corresponding proofs for $p_k^2(d_{i+1}, d_{i+2})$ since they require an analogous development focused on (3.68). Before proceeding, note that the term

$$A_1 = p_k^0(d_{i+1}, d_{i+2}) + p_k^1(d_i, d_{i+1})\left(1 - p_k^0(d_{i+1}, d_{i+2})\right) < 1 \tag{3.71}$$

since $p_k^0(d_{i+1}, d_{i+2}), p_k^1(d_i, d_{i+1}) < 1$, and because A_1 is a strictly increasing function of $p_k^1(d_i, d_{i+1})$ when $p_k^0(d_{i+1}, d_{i+2})$ is held fixed and vice versa, i.e.,

$$\frac{\partial A_1}{\partial\, p_k^1(d_i, d_{i+1})} = 1 - p_k^0(d_{i+1}, d_{i+2}) > 0, \tag{3.72}$$

$$\frac{\partial A_1}{\partial\, p_k^0(d_{i+1}, d_{i+2})} = 1 - p_k^1(d_i, d_{i+1}) > 0 \tag{3.73}$$

for $p_k^0(d_{i+1}, d_{i+2})$, $p_k^1(d_{it}, d_{i+1}) < 1$. Additionally, we have that A_1 is lower bounded by 0 since $p_k^0(d_{i+1}, d_{i+2})$, $p_k^1(d_i, d_{i+1}) > 0$ and $p_k^0(d_{i+1}, d_{i+2}) < 1$. This way, multiplying by $p_k^0(d_i, d_{i+1}) > 0$ both sides of (3.71) yields

$$\begin{aligned} p_k^0(d_i, d_{i+1}) > & p_k^0(d_i, d_{i+1})\left[p_k^1(d_i, d_{i+1})\left(1 - p_k^0(d_{i+1}, d_{i+2})\right)\right] \\ & + p_k^0(d_i, d_{i+1})\, p_k^0(d_{i+1}, d_{i+2}) \end{aligned} \tag{3.74}$$

Summing the term $p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1})$ in both sides of (3.74), after some algebraic manipulations we are able to obtain

$$\begin{aligned} &p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1}) + p_k^0(d_i, d_{i+1}) \left(1 - p_k^1(d_i, d_{i+1}) - p_k^0(d_{i+1}, d_{i+2})\right) \\ &> p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1}) \, \left(1 - p_k^0(d_i, d_{i+1})\right). \end{aligned} \tag{3.75}$$

From (3.75) together with (3.67) and (3.69) we can straightforwardly check that $p_k^{1'}(d_{i+1}, d_{i+2}) < 1$. Moreover, continuing the analysis of (3.67), it can be verified that $p_k^{1'}(d_{i+1}, d_{i+2}) > 0$ when

$$p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1}) \, \left(1 - p_k^0(d_i, d_{i+1})\right) > 0 \tag{3.76}$$

and $\varrho_k(d_i, d_{i+1})) > 0$. On the one hand, (3.76) is direct consequence of the initial assumptions

$$p_k^0(d_i, d_{i+1}), p_k^0(d_{i+1}, d_{i+2}), p_k^1(d_i, d_{i+1}) \in (0, 1).$$

On the other hand, we can show that $\varrho_k(d_i, d_{i+1})) > 0$ by realizing that the l.h.s. of (3.75) is equal to $\varrho_k(d_i, d_{i+1}))$ whereas the r.h.s. is greater than 0 as long as $p_k^0(d_{i+1}, d_{i+2})\, p_k^1(d_i, d_{i+1}) > 0$ and $p_k^0(d_i, d_{i+1}) < 1$.

Last step consists in showing that $0 < p_k^t(d_i, d_{i+1}) < 1$ implies that $0 < \overline{a}_k, \overline{b}_k < 1$ for all i belonging to the set $\{2, 3, \ldots, \ell - 1\}$, $k \in \{0, 1\}$ and $t \in \{0, 1, 1', 2\}$. If we use the inequalities obtained when imposing the upper bounds $p_k^t(d_i, d_{i+1}) < 1$ on (3.20)-(3.23)

$$p_k^{(1,0)}(d_i) p_k^{(0,1)}(d_{i+1}) > 0, \tag{3.77}$$

$$p_k^{(1,0)}(d_i)(1 - p_k^{(0,1)}(d_{i+1})) > 0, \tag{3.78}$$

$$(1 - p_k^{(0,1)}(d_i)) p_k^{(0,1)}(d_{i+1}) > 0, \tag{3.79}$$

$$(1 - p_k^{(0,1)}(d_i))(1 - p_k^{(0,1)}(d_{i+1})) > 0, \tag{3.80}$$

as well as some of the inequalities corresponding to the lower bounds, $p_k^t(d_i, d_{i+1}) > 0$,

$$(1 - p_k^{(1,0)}(d_i))(1 - p_k^{(1,0)}(d_{i+1})) > 0, \tag{3.81}$$

$$(1 - p_k^{(1,0)}(d_i)) p_k^{(1,0)}(d_{i+1}) > 0, \tag{3.82}$$

$$p_k^{(0,1)}(d_i)(1 - p_k^{(1,0)}(d_{i+1})) > 0, \tag{3.83}$$

$$p_k^{(0,1)}(d_i)p_k^{(1,0)}(d_{i+1}) > 0, \tag{3.84}$$

we can straightforwardly prove that the bounds $0 < p_k^{(0,1)}(d_i), p_k^{(1,0)}(d_i) < 1$ hold for $i \in \{1, 2, \ldots, \ell\}$. In this way and concluding the proof, from (3.26) and (3.27) we have that

$$0 < \overline{a}_k = \int_{\mathcal{D}} p_k^{(0,1)}(d_i)\, f_D(d_i)\, \mathrm{d}d_i < \int_{\mathcal{D}} 1 \cdot f_D(d_i)\, \mathrm{d}d_i = 1 \tag{3.85}$$

and

$$0 < \overline{b}_k = \int_{\mathcal{D}} p_k^{(1,0)}(d_i)\, f_D(d_i)\, \mathrm{d}d_i < \int_{\mathcal{D}} 1 \cdot f_D(d_i)\, \mathrm{d}d_i = 1 \tag{3.86}$$

when $0 < p_k^t(d_i, d_{i+1}), p_k^v(d_2, d_3) < 1$ happens for $t \in \{0, 1\}$, $i \in \{2, 3, \ldots, \ell - 1\}$ and $v \in \{1', 2\}$.

3.B Proof of Corollary 1

When the local decisions are independent under hypothesis H_k

$$p_k^0(d_i, d_{i+1}) = p_k^1(d_i, d_{i+1}) = p_k^{1'}(d_i, d_{i+1}) = p_k^2(d_i, d_{i+1}) \tag{3.87}$$

occurs with $k \in \{0, 1\}$ and $i \in \{2, 3, \ldots, \ell - 1\}$. This means that, for $i \in \{1, 2, \ldots, \ell\}$, the i-th device performs its own local decision independently of its nearest neighbour nodes. At the same time, the aforementioned independence entails that the conditional probabilities, $p_k^t(d_i, d_{i+1})$ with $t \in \{0, 1, 1', 2\}$, do not depend on the pair of distances (d_i, d_{i+1}). Because of this, using (3.20)-(3.23) and (3.26)-(3.27), the constraint obtained when substituting (3.34) in the equality $\overline{\rho}_k = 0$ can be written as follows

$$\overline{a}_k = 1 - \overline{b}_k = P(U_i = 1|H_k) = \xi_k. \tag{3.88}$$

Concluding the proof of the first claim, we can substitute (3.88) in (3.30) in order to verify that the error exponent K equals $D(B(\xi_0)||B(\xi_1))$ when the local decisions are independent, i.e. $\overline{\rho} = 0$.

Next we focus on the second claim of Corollary 1. When the local decisions are maximally dependent under H_k, for $k \in \{0, 1\}$ and $i \in \{2, 3, \ldots, \ell - 1\}$ we have that

$$p_k^0(d_i, d_{i+1}) = 1 - p_k^2(d_i, d_{i+1}) = 1, \tag{3.89}$$

or that $p_k^{(0,1)}(d_i) = p_k^{(1,0)}(d_i) = 0$ if we take (3.20) and (3.23). From (3.26) and (3.27), (3.89) results in

$$\overline{a}_k = \overline{b}_k = 0, \tag{3.90}$$

or equivalently $\overline{\rho}_k = 1$ as the definition provided in (3.34). Hence, for all $k \in \{0, 1\}$, the behaviour of K when the local decisions tend to be maximally dependent is given through the limit

$$\begin{aligned}\lim_{\overline{\rho}\to 1} K = \lim_{\overline{a}_k,\overline{b}_k\to 0} K \overset{(a)}{=} & \underbrace{\lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0\,\overline{b}_0}{\overline{a}_0+\overline{b}_0} \ln\left(\frac{\overline{a}_0}{\overline{a}_1}\right)}_{\triangleq A_1} + \underbrace{\lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0\,\overline{b}_0}{\overline{a}_0+\overline{b}_0} \ln\left(\frac{\overline{b}_0}{\overline{b}_1}\right)}_{\triangleq A_2} \\ & + \underbrace{\lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{b}_0-\overline{b}_0\,\overline{a}_0}{\overline{a}_0+\overline{b}_0} \ln\left(\frac{1-\overline{a}_0}{1-\overline{a}_1}\right)}_{\triangleq A_3} + \underbrace{\lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0-\overline{a}_0\,\overline{b}_0}{\overline{a}_0+\overline{b}_0} \ln\left(\frac{1-\overline{b}_0}{1-\overline{b}_1}\right)}_{\triangleq A_4}\end{aligned} \tag{3.91}$$

where (a) follows from substituting (3.31) in (3.30). From the closed-form error exponent provided in (3.30) we can easily show that $K \geq 0$, and consequently, we can assert that

$$\lim_{\overline{\rho}\to 1} K = \lim_{\overline{a}_k,\overline{b}_k\to 0} K = A_1 + A_2 + A_3 + A_4 \geq 0. \tag{3.92}$$

Deriving an upper bound for (3.91) we firstly have

$$\begin{aligned} A_1 &= \lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0\,\overline{b}_0}{\overline{a}_0+\overline{b}_0} \ln\left(\frac{\overline{a}_0}{\overline{a}_1}\right) = \lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0\,\overline{b}_0}{\overline{a}_0+\overline{b}_0} \left[\ln(\overline{a}_0) - \ln(\overline{a}_1)\right] \\ &\overset{(a)}{\leq} -\lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0\,\overline{b}_0}{\overline{a}_0+\overline{b}_0} \ln(\overline{a}_1) \overset{(b)}{\leq} -\lim_{\overline{a}_k,\overline{b}_k\to 0} \frac{\overline{a}_0\,\overline{b}_0}{\overline{b}_0} \ln(\overline{a}_1) \\ &= -\lim_{(\overline{a}_0,\overline{a}_1)\to(0,0)} \overline{a}_0 \ln(\overline{a}_1) \overset{(c)}{=} 0 \end{aligned} \tag{3.93}$$

where (a) and (b) follow since, whenever the local decisions are not maximally dependent under both hypothesis, $\overline{a}_0$, $\overline{b}_0$ and $\overline{a}_1$ are strictly positive and do not take on values greater than one, and (c) is easily shown if the limit $\lim_{(\overline{a}_0,\overline{a}_1)\to(0,0)} \overline{a}_0 \ln(\overline{a}_1)$ is calculated by applying l'Hôpital rule after changing $\overline{a}_0$ and $\overline{a}_1$ to polar coordinates. In addition, if we make an analogous

development we can prove that $A_2 \leq 0$. Now, considering the evaluation of A_3 we obtain

$$\begin{aligned} A_3 &= \lim_{\overline{a}_k, \overline{b}_k \to 0} \frac{\overline{b}_0 - \overline{b}_0\,\overline{a}_0}{\overline{a}_0 + \overline{b}_0} \ln\left(\frac{1-\overline{a}_0}{1-\overline{a}_1}\right) \\ &= \lim_{\overline{a}_k, \overline{b}_k \to 0} \frac{\overline{b}_0\,(1-\overline{a}_0)}{\overline{a}_0 + \overline{b}_0} \left[\ln(1-\overline{a}_0) - \ln(1-\overline{a}_1)\right] \\ &\overset{(a)}{\leq} - \lim_{\overline{a}_k, \overline{b}_k \to 0} \frac{\overline{b}_0\,(1-\overline{a}_0)}{\overline{a}_0 + \overline{b}_0} \ln(1-\overline{a}_1) \overset{(b)}{\leq} - \lim_{\overline{a}_k, \overline{b}_k \to 0} \frac{\overline{b}_0\,(1-\overline{a}_0)}{\overline{b}_0} \ln(1-\overline{a}_1) \\ &= - \lim_{(\overline{a}_0, \overline{a}_1) \to (0,0)} (1-\overline{a}_0) \ln(1-\overline{a}_1) \overset{(c)}{=} -1 \ln(1) = 0 \end{aligned} \tag{3.94}$$

where (a) and (b) follow because of the same reasons as (a) and (b) in (3.93), and (c) is due to the fact that the function $(1-\overline{a}_0)\ln(1-\overline{a}_1)$ is continuous at the point $(0,0)$. Again, using a reasoning similar to the previous one we can prove that $A_4 \leq 0$. This way, the subsequent inequality is obtained

$$\lim_{\overline{\rho} \to 1} K = \lim_{\overline{a}_k, \overline{b}_k \to 0} K = A_1 + A_2 + A_3 + A_4 \leq 0. \tag{3.95}$$

Therefore, we complete the proof of the second claim because (3.92) and (3.95) yield

$$\lim_{\overline{\rho} \to 1} K = \lim_{\overline{a}_k, \overline{b}_k \to 0} K = A_1 + A_2 + A_3 + A_4 = 0. \tag{3.96}$$

Now, consider the last claim of the Corollary 1. If the binary local quantizations are maximally dependent under H_1 and independent under H_0, (3.88) holds for $k = 0$ whereas (3.90) happens for $k = 1$. Thus, in order to know how K behaves when the local decisions are independent under the null hypothesis and they tend to be maximally dependent under the

alternative, we evaluate the limit

$$\lim_{\overline{\rho}_1 \to 1} K(\overline{\rho}_0 = 0) = \lim_{(\overline{a}_1, \overline{b}_1) \to (0,0)} K(\overline{a}_0 = 1 - \overline{b}_0 = \xi_0) \overset{(a)}{=} \underbrace{\lim_{\overline{a}_1 \to 0} \xi_0 \left(1 - \xi_0\right) \ln \left(\frac{\xi_0}{\overline{a}_1} \right)}_{\triangleq B_1}$$
$$+ \underbrace{\lim_{\overline{b}_1 \to 0} \xi_0 \left(1 - \xi_0\right) \ln \left(\frac{1 - \xi_0}{\overline{b}_1} \right)}_{\triangleq B_2} + \underbrace{\lim_{\overline{a}_1 \to 0} \left(1 - \xi_0\right)^2 \ln \left(\frac{1 - \xi_0}{1 - \overline{a}_1} \right)}_{\triangleq B_3}$$
$$+ \underbrace{\lim_{\overline{b}_1 \to 0} \xi_0^2 \ln \left(\frac{\xi_0}{1 - \overline{b}_1} \right)}_{\triangleq B_4} \tag{3.97}$$

where (a) follows after using the definition of $\mu_0(0)$ provided in (3.31). Due to the fact that $B_1 = B_2 = \infty$ and $B_3, B_4 < \infty$ for $\xi_0 \in (0, 1)$, we are able to assert that the limit (3.97) is equal to infinity. This ends the proof.

Chapter 4

Neyman-Pearson fusion of Markov local decisions in a 2-D network

4.1 Introduction

As it happened under the assumption of conditionally independent sensor observations, various works provided answers to different aspects that appear when designing distributed detection systems with dependent observations. To do that, as we commented in the introduction of the previous chapter, considering a 1-D setting many of these works addressed aspects such the impact of sensor density as well as the kind of channel between the sensors and the fusion center on the detection performance of the network. Some of these results can be found in [Chamberland and Veeravalli, 2006], [Li and Dai, 2007] and [Plata-Chaves and Lázaro, 2011].

Although there are several results for 1-D sensor networks, the literature that solves the cited problem in a 2-D setting is less extensive. Some of the existing results are [Anandkumar et al., 2009] and [Sung et al., 2009]. Under the Neyman-Pearson formulation [Anandkumar et al., 2009] derived a closed-form error exponent when a test for independence is considered, and when the correlation structure of the sensor observations is given by a nearest-neighbour Gauss-Markov random field (GMRF). In this way, the authors characterized the detection performance of the employed Neyman-Pearson fusion rule w.r.t. different design parameters of the network, e.g. the density

of the deployment or the signal to noise ratio of the observations performed by the devices. In [Sung et al., 2009] the authors investigated the amount of information obtainable from a sensor network where the devices are located on a 2-D lattice, and where under each one of the two hypothesis the observations received by the fusion center are distributed according to a 2-D hidden GMRF defined by a symmetric first-order conditional autoregression model. Toward this goal, the authors used implicit expressions of the Kullback-Leibler rate and the mutual information rate as the information measures.

However, up to now no work obtains analytically tractable expressions that allow the design of 2-D sensor networks where the devices are located on a rectangular grid and where the Neyman-Pearson fusion rule is performed on dependent quantized summaries of the sensor observations. Because of this, in the spirit of [Chamberland and Veeravalli, 2006], [Li and Dai, 2007], and [Anandkumar et al., 2009], we extend the results of the previous chapter in order to derive a closed-form error exponent for the Neyman-Pearson fusion rule performed by the fusion center of the parallel network shown in Figure 4.1. This derivation is carried out when the correlation structure of the local decisions is modelled with 2-D random fields defined on a discrete space. Specifically, among the extremely few 2-D discrete fields that are analytically tractable, we have considered a 2-D random process that, used as observation model in different applications such as image processing and coding, is constructed from a first-order binary Markov chain (see [Pickard, 1977], [Champagnat et al., 1998] and [Justesen, 2005]). This way, in this part of the dissertation, by means of the resulting closed-form expression of the error exponent we provide an amenable tool that links the detection performance of a 2-D version of the sensor network described in [Drakopoulos and Lee, 1991] with some of its physical and design features.

This chapter is organized as follows. The problem is stated in Section 4.2. In section 4.3, in the large deviation framework we present the derivation of the closed-form error exponent of the Neyman-Pearson test performed at the data fusion center of the network plotted in Figure 4.1. In order to analyze how the detection performance behaves as the density of the network changes, Section 4.4 shows several simulation results based on the derived error exponent. Next, Section 4.5 gives a summary of the results of this chapter. Finally, as a matter of organization style, we defer all proofs of this chapter to a pair of appendices.

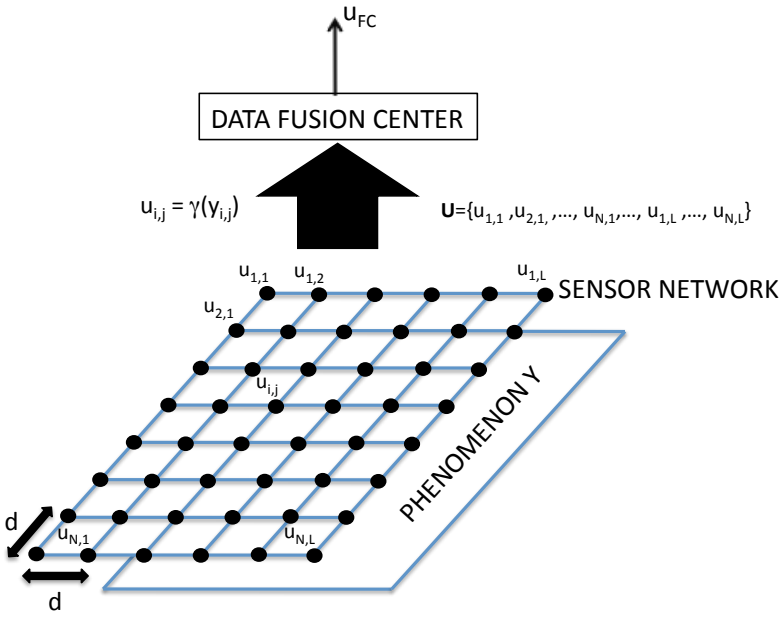


Figure 4.1: *Block diagram of a 2-D parallel fusion network.*

4.2 Problem statement

We consider a network formed by a data fusion center and NL devices located on a 2-D lattice $\mathcal{I}_{N,L}$ where, as it is shown in Figure 4.1, the sensors belonging to a specific row or column of the lattice are equally spaced. In order to decide what state of the phenomenon is present, H_0 or H_1, the distributed system undertakes the following steps. As it happened in the network considered in Chapter 3, each device firstly performs a local observation of the environment, $y_{i,j}$. Secondly, it applies a binary detection rule to it, $\gamma_{i,j}(y_{i,j})$, not necessarily based on LLRT. This way, each device makes a local decision, $u_{i,j} = \gamma_{i,j}(y_{i,j})$, regarding the state of the phenomenon of interest. Thirdly, the binary local quantizations of the sensor observations are transmitted to the fusion center over error free parallel access channels. Finally, based on the NL local decisions taken by the devices of the network, the fusion center makes a global decision, $u_{_{FC}}$, under the Neyman-Pearson formulation. Considering that the correlation structure of the local decisions is hypothesis dependent and modelled with a 2-D called Pickard field (see [Pickard, 1977]) where the rows and columns are outcomes of the same first-order binary Markov

chain, under the Neyman-Pearson formulation we have the following inference problem at the fusion center

$$\left.\begin{array}{ll} H_k: & \boldsymbol{U} \sim \text{ 2-D process where the binary local quantizations} \\ & \text{belonging to a specific row } i \in \{1,\ldots,N\} \text{ or column} \\ & j \in \{1,\ldots,L\} \text{ form the same first-order Markov} \\ & \text{chain, } M_k, \text{ with transition matrix} \\ & \qquad \boldsymbol{\Pi}_k = \begin{bmatrix} p_k^{(0,0)}(d,\mathcal{P}_k) & p_k^{(0,1)}(d,\mathcal{P}_k) \\ p_k^{(1,0)}(d,\mathcal{P}_k) & p_k^{(1,1)}(d,\mathcal{P}_k) \end{bmatrix}^T \end{array}\right\} \quad (4.1)$$

where, for $k \in \{0,1\}$,

- $\boldsymbol{U} = [U_{1,1}, U_{2,1}, \ldots, U_{N,1}, \ldots, U_{1,L}, \ldots, U_{N,L}]^T$, defined on $\mathcal{U}^{NL}$ with $\mathcal{U} = \{0,1\}$, denotes the NL local decisions performed by the NL devices that form the network.

- $p_k^{(0,1)}(d,\mathcal{P}_k)$ and $p_k^{(1,0)}(d,\mathcal{P}_k)$ are the transition probabilities of the Markov chain, M_k, that extends along the rows and columns of $\mathcal{I}_{N,L}$ under H_k, i.e. for $u_m, u_{m-1} \in \{0,1\}$

$$p_k^{(u_{m-1},u_m)}(d,\mathcal{P}_k) = P(U_m = u_m | U_{m-1} = u_{m-1}, d, \mathcal{P}_k, H_k). \quad (4.2)$$

- d is the distance between two neighbour devices that belong to the same row or column of the lattice $\mathcal{I}_{N,L}$.

- $\mathcal{P}_k$ is the set of physical and design parameters of the network that being known by the data fusion center are arguments of the transition probabilities associated with the Markov chain M_k (e.g. the employed local decision rules or the physical properties of the environment where the network is deployed).

Given the correlation structure assumed for $\boldsymbol{U}$ under both hypothesis and according to [Drakopoulos and Lee, 1991], if the spacing between the nearest neighbors, d, is known by the data fusion center, the optimal fusion rule ensuring that $P_{FA} \leq \alpha \in (0,1)$ in (4.1) is given by the subsequent

Neyman-Pearson detector

$$u_{FC} = \begin{cases} 1 & \text{if } \ln\left(\frac{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_0,H_0)}{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_1,H_1)}\right) < \tau, \\ \gamma & \text{if } \ln\left(\frac{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_0,H_0)}{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_1,H_1)}\right) = \tau, \\ 0 & \text{otherwise,} \end{cases} \tag{4.3}$$

where, for $k \in \{0,1\}$ and

$$\mathcal{V}_{NL}^{1}(\tau) = \left\{ \boldsymbol{u} : \ln\left(\frac{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_0,H_0)}{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_1,H_1)}\right) < \tau \right\}, \tag{4.4}$$

τ is the smallest number such that

$$\varsigma = \sum_{\boldsymbol{u}\in\mathcal{V}_{\ell}^{1}(\tau)} P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_0,H_0) \leq \alpha \tag{4.5}$$

and $\gamma \in \{0,1\}$ is a randomization function that takes on the value 1 with the subsequent probability

$$\nu = \begin{cases} 1 & \text{if } \vartheta = 0, \\ \frac{\alpha-\varsigma}{\vartheta} & \text{otherwise} \end{cases} \tag{4.6}$$

with

$$\vartheta = P\left(\ln\left(\frac{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_0,H_0)}{P_{\boldsymbol{U}|D,\mathcal{P},H}(\boldsymbol{u}|d,\mathcal{P}_1,H_1)}\right) = \tau \middle| H_0 \right). \tag{4.7}$$

Note that, due to the causality of the correlation model assumed under H_k, the local decision $U_{i,j}$ is independent of the rest of the binary local quantizations when some of the ones taken by the neighbour nodes are known. Specifically, given the hypothesis H_k, with $k \in \{0,1\}$, the local decisions responsible for the aforementioned conditional independence are established by the directions that the Markov chains M_k adopt when extending along the rows and columns of the 2-D lattice, $\mathcal{I}_{N,L}$. For instance, as it happens in [Justesen, 2005], if we assume that, under both hypothesis H_0 and H_1, the Markov chains extend from left to right along the rows and from top to bottom along the columns, in [Pickard, 1977] it was showed that $U_{i,j}$ is independent of the rest of the local decisions when $U_{i,j-1}$, $U_{i-1,j}$

and $U_{i-1,j-1}$ are known. Nonetheless, under a more general definition of the considered symmetric Pickard field (see [Champagnat et al., 1998]), the author in [Justesen, 2005] used a stationarity condition associated with the Markov chain M_k in order to prove that, under H_k with $k \in \{0,1\}$, $U_{i,j}$ is also conditionally independent of the rest of the binary local quantizations when $U_{i,j-1}$ and $U_{i+1,j}$ are given. Without loss of generality and from now and on, given the hypothesis H_k we are going to assume the previous conditional independence as well as the corresponding direction in the flow of the Markov chains present in each column and row of $\mathcal{I}_{N,L}$. Consequently, after omitting whatever kind of dependency on $\mathcal{P}_k$ for the sake of simplicity, the joint p.m.f. $P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d,H_k)$ involved in (4.3) can be factorized as follows

$$\begin{aligned} P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d,H_k) =& P_{U_{1,1}|H}(u_{1,1}|H_k) \times \prod_{i=2}^{N} P_{U_{i,1}|U_{i-1,1},D,H}(u_{i,1}|u_{i-1,1},d,H_k) \\ & \times \prod_{j=2}^{L} P_{U_{N,j}|U_{N,j-1},D,H}(u_{N,j}|u_{N,j-1},d,H_k) \\ & \times \prod_{i=1}^{N-1} \prod_{j=2}^{L} P_{U_{i,j}|U_{i+1,j},U_{i,j-1},D,H}(u_{i,j}|u_{i+1,j},u_{i,j-1},d,H_k) \end{aligned} \tag{4.8}$$

where, for $k \in \{0,1\}$, $P_{U_{1,1}|H}(u_{1,1}|H_k)$ equals the p.m.f. of the initial state associated with the 2-D random process present under H_k,

$$P_{U_{i,j}|U_{i-1,j},D,H}(u_{i,j}|u_{i-1,j},d,H_k) = \sum_{(r,s)\in\mathcal{U}^2} p_k^{(s,r)}(d)\delta[r-u_{i,j}]\delta[s-u_{i-1,j}] \tag{4.9}$$

with $i \in \{2,3,\ldots,N\}$ and $j \in \{1,2,\ldots,L\}$,

$$P_{U_{i,j}|U_{i,j-1},D,H}(u_{i,j}|u_{i,j-1},d,H_k) = \sum_{(r,s)\in\mathcal{U}^2} p_k^{(s,r)}(d)\delta[r-u_{i,j}]\delta[s-u_{i,j-1}] \tag{4.10}$$

with $i \in \{1,2,\ldots,N\}$ and $j \in \{2,3,\ldots,L\}$, and

$$\begin{aligned} & P_{U_{i,j}|U_{i+1,j},U_{i,j-1},D,H}(u_{i,j}|u_{i+1,j},u_{i,j-1},d,H_k) \\ & = \sum_{(t,r,s)\in\mathcal{U}^3} \frac{p_k^{(t,r)}(d)\, p_k^{(r,s)}(d)}{\hat{p}_k^{(t,s)}(d)} \delta[t-u_{i,j-1}]\delta[s-u_{i+1,j}]\delta[r-u_{i,j}] \end{aligned} \tag{4.11}$$

with $i \in \{1, 2, \ldots, N-1\}$, $j \in \{2, 3, \ldots, L\}$ and

$$\hat{p}_k^{(t,s)}(d) = \sum_{v=0}^{1} p_k^{(t,v)}(d)\, p_k^{(v,s)}(d). \tag{4.12}$$

4.3 Error exponent

Here we provide a tool that allows the design and analysis of a sensor network as the one described in the previous section. To be specific, based on information theoretic results we derive a design tool that links the detection performance of the Neyman-Pearson test (4.3) with different physical and design parameters of the network shown in Figure 4.1. In order to accomplish this aim, we would want to characterize the overall probability of detection of the network

$$\begin{aligned} P_D = & \sum_{\boldsymbol{u} \in \mathcal{V}_{NL}^1(\tau)} P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_1) \\ & + \nu \cdot P\left(\ln\left(\frac{P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_0)}{P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_1)} \right) = \tau \middle| H_1 \right) \end{aligned} \tag{4.13}$$

when a fixed constraint is imposed on the overall false alarm probability, i.e. when $P_{FA} \leq \alpha \in (0, 1)$. However, as it happened under the 1-D setting, the derivation of a closed-form expression for (4.13) is not possible. Because of this, our approach has focused on the analytic tools provided by the large deviations theory corresponding with Neyman-Pearson hypothesis tests. In particular, similarly to the previous chapter our objective is to derive a closed-form expression of the exponential rate of decay in $P_M = 1 - P_D$ as N and L approach infinity and when $P_{FA} \leq \alpha \in (0, 1)$. As it is proved in [Vajda, 1989], this problem results in calculating the almost-sure limit under H_0 of the asymptotic Kullback-Leibler rate

$$\begin{aligned} K &\triangleq \lim_{N,L\to\infty} -\frac{1}{N \cdot L} \log\left(P_M\right) \\ &= \lim_{N,L\to\infty} \frac{1}{N \cdot L} \ln\left(\frac{P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_0)}{P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_1)} \right) \text{ (a.s. in } H_0). \end{aligned} \tag{4.14}$$

In this way, we arrive at the following theorem.

Theorem 2. *Suppose that* $0 < p_k^{(u_{m-1},u_m)}(d) < 1$ *for all* $u_m, u_{m-1}, k \in \{0,1\}$ *and that the p.m.f. of the local decision* $U_{1,1}$ *under* H_0 *is absolutely continuous w.r.t. the corresponding one under* H_1, $P_{U_{1,1}|H}(u_{1,1}|H_0) \ll P_{U_{1,1}|H}(u_{1,1}|H_1)$. *Then, given a fixed constraint* $P_{_{FA}} \leq \alpha \in (0,1)$, *the best Neyman-Pearson error exponent for the distributed detection problem given by Equation* (4.1) *is*

$$\begin{aligned} K &= \sum_{(t,r,s)\in\mathcal{U}^3} \pi_0^{(t)}(d)\, p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln\left(\frac{p_0^{(t,r)}(d)\, p_0^{(r,s)}(d)\, \hat{p}_1^{(t,s)}(d)}{p_1^{(t,r)}(d)\, p_1^{(r,s)}(d)\, \hat{p}_0^{(t,s)}(d)}\right) \qquad (4.15) \\ &= D(P(u_{i,j}|u_{i+1,j}, u_{i,j-1}, d, H_0)||P(u_{i,j}|u_{i+1,j}, u_{i,j-1}, d, H_1)) \end{aligned}$$

where

- $\mathcal{U}^3$ *equals the cartesian product of the set* $\mathcal{U} = \{0,1\}$ *with itself, three times.*
- $\pi_0^{(0)}(d) = 1 - \pi_0^{(1)}(d) = P(u = 0|H_0) = \frac{p_0^{(1,0)}(d)}{p_0^{(1,0)}(d)+p_0^{(0,1)}(d)}$ *is the unique stationary probability of deciding* H_0 *when the 2-D random process constructed from the Markov chain,* M_0, *is present.*
- $D(P(u_{i,j}|u_{i+1,j}, u_{i,j-1}, d, H_0)||P(u_{i,j}|u_{i+1,j}, u_{i,j-1}, d, H_1))$ *is the conditional Kullback-Leibler divergence of*

  $$P_{U_{i,j}|U_{i+1,j},U_{i,j-1},D,H}(u_{i,j}|u_{i+1,j}, u_{i,j-1}, d, H_0)$$

  and

  $$P_{U_{i,j}|U_{i+1,j},U_{i,j-1},D,H}(u_{i,j}|u_{i+1,j}, u_{i,j-1}, d, H_1)$$

  in the stationary regime.

Proof. See Appendix 4.A. □

Next, as it occurred in Chapter 3, we are interested in investigating some analytical properties of the derived error exponent when the dependence among the local decisions takes on some specific values. To do that, as a measure of the aforementioned dependence we use a particularization of the correlation index derived in the previous chapter (see (3.34)). To be more precise, capturing the mean correlation strength among the random variables

corresponding with neighbour steps of a first-order Markov chain with binary state space, the used index is defined as

$$\rho = \begin{cases} P(H_0)\rho_0 + P(H_1)\rho_1 & \text{When the local decisions are} \\ & \text{dependent under } H_0 \text{ and } H_1. \\ \rho_k & \text{When the local decisions are only} \\ & \text{dependent under } H_k. \end{cases} \tag{4.16}$$

where, for all $k \in \{0,1\}$, $P(H_k) \in (0,1)$ denotes the prior probability of hypothesis H_k and

$$\rho_k = 1 - p_k^{(0,1)}(d) - p_k^{(1,0)}(d) \in [0,1]. \tag{4.17}$$

Looking at (4.17), for some hypothesis, H_k, with $k \in \{0,1\}$, we can easily see that ρ_k is equal to zero when the local decisions taken by neighbour nodes are conditionally independent and that ρ_k equals one when they are maximally dependent according to the definition given in [Drakopoulos and Lee, 1991]. Precisely, combinations of these two correlation values under H_0 and H_1 describe the scenarios of interest considered in the study of K w.r.t. ρ carried out in the subsequent corollary.

Corollary 2. *The error exponent K derived in Theorem 2 satisfies the next three claims.*

1) When the local decisions are independent under H_0 and H_1, K is equal to the Kullback-Leibler divergence $D(B(\xi_0)||B(\xi_1))$ where, for $k \in \{0,1\}$, $B(\xi_k)$ denotes a Bernoulli random variable with probability of success $\xi_k = P(U_i = 1|H_k) \in (0,1)$,

$$K(\rho = 0) = (1-\xi_0)\ln\left(\frac{1-\xi_0}{1-\xi_1}\right) + \xi_0 \ln\left(\frac{\xi_0}{\xi_1}\right) = D(B(\xi_0)||B(\xi_1)). \tag{4.18}$$

2) When the local decisions tend to be maximally dependent under H_0 and H_1, K converges to zero,

$$\lim_{\rho \to 1} K = 0. \tag{4.19}$$

3) When the local decisions tend to be maximally dependent under H_1 at the same time as they are independent under H_0, K diverges,

$$\lim_{\rho_1 \to 1} K = \infty \quad s.t. \ \rho_0 = 0. \tag{4.20}$$

Proof. See Appendix 4.B.

□

Looking at the previous corollary we can remark that, in a 2-D parallel network solving (4.1), the error exponent for the optimal Neyman-Pearson rule has the same behaviour as the corresponding one provided in Theorem 1 under a 1-D setting. Specifically, focusing on the first claim we can realize that, if the local decisions are independent under H_0 and H_1, the closed-form expression, K, obtained in Theorem 2 is consistent with Stein's Lemma given in [Cover and Thomas, 2006]. Additionally from the second claim we can see that the derived error exponent provided in (4.15) converges to zero as the correlation strength among neighbor decisions approaches one. Again, as we stated for a 1-D parallel network fusing Markov local decisions (see Chapter 3), this behaviour is explained if we realize that, when the local decisions are maximally dependent under both hypothesis

$$P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_0) = P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_1). \tag{4.21}$$

Finally, we can note that, if the local decisions are maximally dependent under the 2-D observation model given in (4.1) for H_k with $k \in \{0, 1\}$, according to the results obtained in [Drakopoulos and Lee, 1991] the support of $P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_k)$ only consists of the two possible events where all the NL devices decide the same hypothesis. Taking into account that the support of $P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_k)$ is formed by all the 2^{NL} binary NL-tuples when the local decisions are conditionally independent under H_k, we can easily see check that $P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_1) \ll P_{\boldsymbol{U}|D,H}(\boldsymbol{u}|d, H_0)$ when $\rho_0 = 0$ and $\rho_1 = 1$ with ρ_k defined in (4.17). Although $P_{U_{1,1}|D,H}(u_{1,1}|d, H_0) \ll P_{U_{1,1}|D,H}(u_{1,1}|d, H_1)$ holds, concluding the current section this last fact explains the divergence of the limit computed in the third claim of Corollary 2.

4.4 Characterization of the error exponent

Through synthetic experiments based on the evaluation of K for a specific physical model of the transition probabilities, in this section we give

some insights into the behaviour of K when different design and physical parameters of the network vary. In particular, we will pay special attention to the dependency of K w.r.t. the mean correlation strength among local decisions taken by neighbour nodes (see (4.16) and (4.17)).To carry out this study we have considered the following physical model for the transition probabilities of the Markov chains, M_0 and M_1,

$$p_k^{(0,1)}(d) = 1 - p_k^{(0,0)} = \xi_k(1 - m_k e^{-\gamma_k d}) \tag{4.22}$$

and

$$p_k^{(1,0)}(d) = 1 - p_k^{(1,1)} = (1 - \xi_k)\,(1 - m_k e^{-\gamma_k d}) \tag{4.23}$$

where, for $k \in \{0, 1\}$,

- ξ_k is the probability of false alarm or detection probability of the local detectors when independence among the local decisions is assumed under H_0 or H_1 respectively.

- γ_k is a strictly positive constant that indicates the exponential rate of growth of the transition probabilities $p_k^{(0,1)}(d)$ and $p_k^{(1,0)}(d)$ as d increases.

- m_k is a strictly positive constant that, taking on values less than one, controls the correlation between the pair of local decisions $(U_{i-1,j}, U_{i,j})$ and $(U_{i,j}, U_{i,j+1})$ when the distance between the devices involved in each pair is zero, i.e $d = 0$.

Note that the physical model considered for $p_k^{(0,1)}(d)$ and $p_k^{(1,0)}(d)$ has been taken from Section 3.4. Consequently, it fulfills the same regularity conditions that typically appear in a detection scenario and that were explained in the aforementioned section.

In Figure 4.2, we plot the error exponent of Equation (4.15) as a function of the mean correlation strength provided through Equations (4.16) and (4.17). Theoretical curves of this figure have been generated when the local decisions are only dependent under H_1 and ξ_1 is equal to $\{0.8, 0.98, 0.998\}$. As it is expected from the results obtained in a 1-D setting (see Section 3.4), regardless the value of ξ_1, K initially decreases as ρ increases, and after a specific value ρ^*, it increases as ρ approaches one. The explanation to this trend is the same as the one given in Section 3.4 for both equispaced sensors and exponentially spaced sensors. On the one hand,

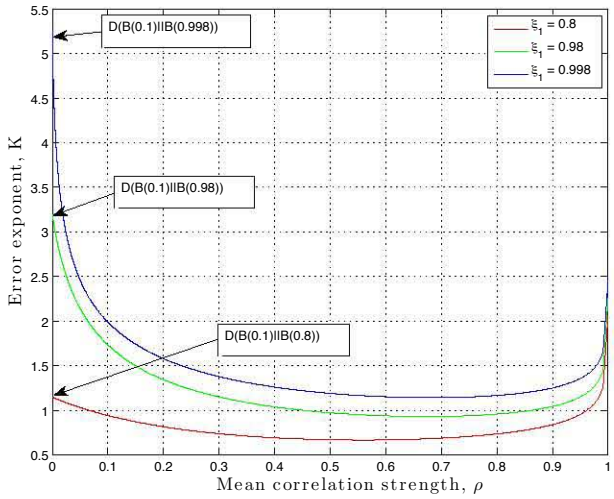


Figure 4.2: *For a 2-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of the mean correlation strength among neighbour binary decisions, ρ, when they are only dependent under H_1 and $\xi_1 = \{0.8, 0.98, 0.998\}$. Parameters: $\xi_0 = 0.1$, $\gamma_0 = 2$, $\gamma_1 = 0.9$, and $m_1 = 1 - 10^{-4}$.*

similarly to a scenario where the sensors are deployed along a straight line, the initial decrease of K in ρ occurs because discrimination between H_0 and H_1 can be more difficult when the information provided by each sensor is more and more correlated without improving the detection performance of the sensors. On the other hand, in order to explain the increasing behaviour of K w.r.t. ρ we have to recall that the correlation among the local decisions might be more informative than the independence when a correlation structure is only present under one of the two hypothesis. Note that, given a specific correlation model, the value of ρ^*, at which the dependence among the local decisions is informative, is a function of other design parameters of the network such as ξ_0 and ξ_1. For instance, as it can be checked from simulations as well as a careful analysis, under the considered correlation model the aforementioned value of ρ^* is shifted closer to one as the ratio ξ_1/ξ_0 increases.

At this point, we can check a pair of analytic results that have been proved in Corollary 2 and can also be observed in Figure 4.2. Firstly, when

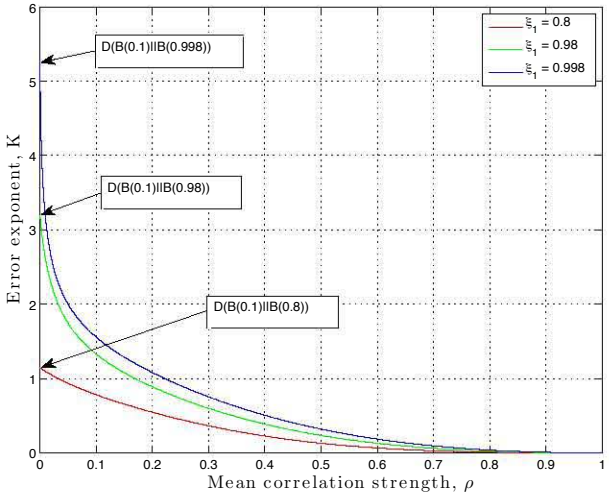


Figure 4.3: *For a 2-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of the mean correlation strength among neighbour binary decisions, ρ, when they are dependent under H_1 and H_0, and $\xi_1 = \{0.8, 0.98, 0.998\}$. Parameters: $\xi_0 = 0.1$, $\gamma_0 = 2$, $\gamma_1 = 0.9$, $P(H_k) = 0.5$ and $m_k = 1 - 10^{-4}$ with $k \in \{0, 1\}$.*

ρ equals zero, we can realize that K collapses to the subsequent Kullback-Leibler divergence, $D(B(\xi_0)||B(\xi_1))$, where $B(\xi_k)$ denotes a Bernoulli random variable with probability of success equal to ξ_k. This behaviour shows the consistency of the error exponent derived in Theorem 2 with the Neyman-Pearson error exponent stated by the Stein's Lemma detailed in [Cover and Thomas, 2006] under independent observations. Secondly, as it occurred under the 1-D scenario analyzed in Chapter 3, we can conclude the analysis of Figure 4.2 by noting that, regardless the value of ξ_1, K diverges when ρ approaches one. This behaviour corroborates the third claim provided in Corollary 2.

Next, in Figure 4.3 we make the same analysis as in Figure 4.2. Nonetheless, in this case the local decisions are also dependent under H_0. Looking at both Figures 4.2 and 4.3, we can firstly check that the derived error exponent increases as the probability of error of the local detector decreases when the rest of the physical and designed parameters of the 2-D network are held fixed. If we take into account that a larger error

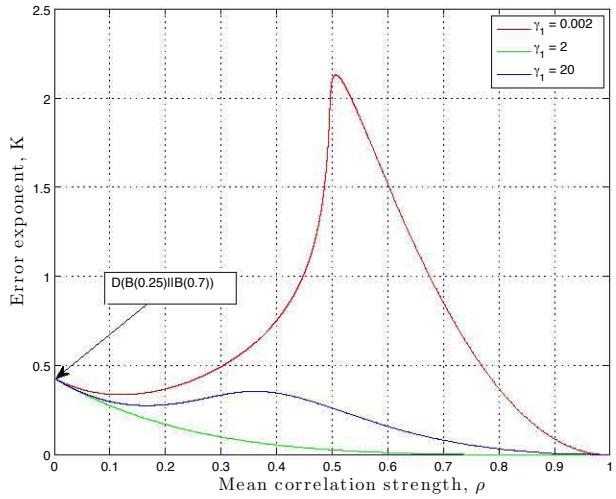


Figure 4.4: *For a 2-D parallel network, Neyman-Pearson optimal error exponent, K, as a function of the mean correlation strength among the neighbour binary decisions, $\overline{\rho}$, when they are dependent under H_1 and H_0, $\gamma_1 = \{2 \cdot 10^{-3}, 2, 20\}$. Parameters: $\xi_0 = 0.25$, $\xi_1 = 0.7$, $\gamma_0 = 2$, $P(H_k) = 0.5$ and $m_k = 1 - 10^{-4}$ with $k \in \{0, 1\}$.*

exponent entails a better detection performance for a fixed number of sensor observation, as we said in the Section 3.4 this last feature agrees with what we can intuitively expect. Continuing with the analysis of Figure 4.3 we also have that, for each one of the curves associated with a different value of ξ_1, we can again observe that K shows the same monotonocity behaviour as the one seen under the 1-D setting. Specifically, for the considered set of physical and design parameters, K is monotonically decreasing and its rate of decay is slower as ρ increases. Nevertheless, unlike the case where a correlation structure is only present under H_1, the evolution of K w.r.t. ρ does not have the previous features for all the possible values that can take the parameters appearing in (4.22) and (4.23). This is corroborated by performing simulations such as the one shown in Figure 4.4.

Finally, we analyze the convergence of K shown in Figure 4.3 and 4.4 for $\rho = 0$ and $\rho = 1$. As it can be observed in those figures when $\rho = 0$, independently of the physical model the derived error exponent reduces to $D(B(\xi_0)||B(\xi_1))$, i.e. the Stein's Lemma. Meanwhile, as it is

expected from the analytic studies undertaken in claim c) of Corollary 2, in Figures 4.3 and 4.4, when ρ goes to one, K converges to zero independently of the physical and design parameters involved in (4.22) and (4.23). As it happens under the 1-D setting, note that, if the local decisions are maximally dependent under both hypothesis, once the fusion center has read a device, new readings do not provide additional information when discriminating H_1 against H_0.

4.5 Summary

Here, we characterized the performance associated with a 2-D version of the two-stage distributed detection system studied in the previous chapter. In particular, we considered that the sensors are located on a rectangular lattice where the sensors belonging to a specific row or column are equally spaced. Additionally, we assumed that there is no kind of cooperation among the devices so that, from its own observation each device performs a local decision regarding the underlying binary hypothesis testing problem. However, due to the presence of a correlation source in the sensor observations we supposed that the local decisions sent to the fusion center might be dependent under both hypothesis. Given each one of the two possible hypothesis, this dependency among the local decisions was modelled by means of a 2-D causal field where the rows and columns are outcomes of the same first-order binary Markov chain. Under this scenario and for an arbitrary physical model that links the physical parameters of the network with the transition probabilities of the aforementioned Markov chains, we firstly derived a closed-form expression of the error exponent associated with the Neyman-Pearson test performed at the fusion center. Finally, after studying some properties of this closed-form error exponent w.r.t. the mean correlation strength among local decisions taken by neighbor sensors, we evaluated the derived error exponent under the physical model used in Chapter 3 for the transition probabilities that define the Markov observation process associated with each hypothesis. By means of this evaluation we analyzed the dependency between the detection performance of the network and its different physical and design parameters. Specially, we paid attention to the different trends of detection performance that appear when the dependence among the local decisions changes.

4.A Proof of Theorem 2

Since the error exponent for the Neyman-Pearson detector with a fixed level $\alpha \in (0,1)$ is given in implicit form by (4.14), we focus on the calculation of this limit for the scenario described in Section 4.2. Taking into account the factorization of $P_{\boldsymbol{U}|H}(\boldsymbol{u}|H_k)$ given in (4.8) as well as the characterizations provided in (4.9)-(4.12), Equation (4.14) can be written as follows

$$\begin{aligned} K = & \underbrace{\lim_{N,L\to\infty} \frac{1}{N\cdot L} \ln\left(\frac{P_{U_{1,1}|H}(u_{1,1}|H_0)}{P_{U_{1,1}|H}(u_{1,1}|H_1)}\right) \quad \text{(a.s in } H_0)}_{\triangleq D_1} \\ & + \underbrace{\sum_{r,s\in\mathcal{U}^2} \ln\left(\frac{p_0^{(s,r)}(d)}{p_1^{(s,r)}(d)}\right) \lim_{N,L\to\infty} \frac{N_r(s,r|\boldsymbol{u}_{2,1}^{N,1})}{N\cdot L}}_{\triangleq D_2} \\ & + \underbrace{\sum_{r,s\in\mathcal{U}^2} \ln\left(\frac{p_0^{(s,r)}(d)}{p_1^{(s,r)}(d)}\right) \lim_{N,L\to\infty} \frac{N_c(s,r|\boldsymbol{u}_{N,2}^{N,L})}{N\cdot L}}_{\triangleq D_3} \\ & + \underbrace{\sum_{r,s,t\in\mathcal{U}^3} \ln\left(\frac{p_0^{(t,r)}(d)\, p_0^{(r,s)}(d)\, \hat{p}_1^{(t,s)}}{p_1^{(t,r)}(d)\, p_1^{(r,s)}(d)\, \hat{p}_0^{(t,s)}}\right) \lim_{N,L\to\infty} \frac{N_0(t,r,s,|\boldsymbol{u}_{1,2}^{N-1,L})}{NL}}_{\triangleq D_4} \end{aligned} \tag{4.24}$$

where, for the sequence of local decisions $\boldsymbol{u}_{a,c}^{b,d} = \{u_{i,j} : a \le i \le b, c \le j \le d\}$,

- $N_r(u_{i-1,j}, u_{i,j}|\boldsymbol{u}_{a,c}^{b,d})$ equals the number of times that, under H_0, the pair $(u_{i-1,j}, u_{i,j})$ occurs in the sequence of local decisions $\boldsymbol{u}_{a,c}^{b,d}$ performed by the corresponding devices in the lattice.

- $N_c(u_{i,j-1}, u_{i,j}|\boldsymbol{u}_{a,c}^{b,d})$ equals the number of times that, under H_0, the pair $(u_{i,j-1}, u_{i,j})$ occurs in the sequence of local decisions $\boldsymbol{u}_{a,c}^{b,d}$ performed by the corresponding devices in the lattice.

- $N_0(u_{i,j-1}, u_{i,j}, u_{i+1,j}|\boldsymbol{u}_{a,c}^{b,d})$ equals the number of times that, under H_0, the triple $(u_{i,j-1}, u_{i,j}, u_{i+1,j})$ occurs in the lattice of local decisions $\boldsymbol{u}_{a,c}^{b,d}$ performed by the corresponding devices.

Starting from the assumptions that ensure the existence of K we calculate each one of the terms that appear in the r.h.s. of K. In particular, we are going to prove that Equation (4.24) yields (4.15) when $P_{U_{1,1}|H}(u_{1,1}|H_0) \ll P_{U_{1,1}|H}(u_{1,1}|H_1)$ and $0 < p_k^{(u_{m-1},u_m)} < 1$ holds for all $u_m, u_{m-1}, k \in \{0,1\}$. Firstly, if we take into account that $P_{U_{1,1}|H}(u_{1,1}|H_0) \ll P_{U_{1,1}|H}(u_{1,1}|H_1)$ we have that

$$\ln\left(\frac{P_{U_{1,1}|H}(u_{1,1}|H_0)}{P_{U_{1,1}|H}(u_{1,1}|H_1)}\right) < \infty \tag{4.25}$$

and therefore, we prove that D_1 vanishes as N and L go to infinity. Secondly, we derive the terms D_2 and D_3. Considering the assumptions that ensure the existence of K, we guarantee the regularity of the Markov chain, M_k, that extends along the rows and columns of $\mathcal{I}_{N,L}$ under H_k with $k \in \{0,1\}$ (see [Kemeny and Snell, 1976]). Consequently, given the hypothesis H_k and knowing that the sequences $\boldsymbol{u}_{2,1}^{N,1}$ and $\boldsymbol{u}_{N,2}^{N,L}$ are outcomes of the same first-order binary Markov chain, M_0, as it is proved in Theorem 1 under the same initial assumptions, we obtain

$$\begin{aligned}
&\sum_{r,s\in\mathcal{U}^2} \ln\left(\frac{p_0^{(s,r)}(d)}{p_1^{(s,r)}(d)}\right) \lim_{N\to\infty} \frac{N_r(s,r|\boldsymbol{u}_{2,1}^{N,1}\ )}{N} \\
&\qquad = \sum_{r,s\in\mathcal{U}^2} \ln\left(\frac{p_0^{(s,r)}(d)}{p_1^{(s,r)}(d)}\right) \lim_{L\to\infty} \frac{N_c(s,r|\boldsymbol{u}_{N,2}^{N,L}\ )}{L} \\
&\qquad = \sum_{r,s\in\mathcal{U}^2} \pi_0^{(s)}(d)\, p_0^{(s,r)}(d) \ln\left(\frac{p_0^{(s,r)}(d)}{p_1^{(s,r)}(d)}\right) < \infty
\end{aligned} \tag{4.26}$$

From (4.26) we can easily show that D_2 and D_3 are equal to zero. Finally, we evaluate D_4. For this calculation, we need to obtain an asymptotic closed-form expression for the empirical joint probability of the triple, $(u_{i,j-1}, u_{i,j}, u_{i+1,j}) \in \mathcal{U}^3$ given the hypothesis H_0 and the lattice $\boldsymbol{u}_{1,2}^{N-1,L}$. From the stationarity of the regular Markov chain, M_0, regarding the type of $(u_{i,j-1}, u_{i,j}, u_{i+1,j})$ in $\boldsymbol{u}_{1,2}^{N-1,L}$ it is straightforward to prove that

$$\begin{aligned}
&\lim_{N,L\to\infty} \frac{N_0(u_{i,j-1}, u_{i,j}, u_{i+1,j})|\boldsymbol{u}_{1,2}^{N-1,L}\ )}{N\cdot L} \\
&\qquad = \pi_0^{(u_{i,j-1})}(d)\, p_0^{(u_{i,j-1},u_{i,j})}(d)\, p_0^{(u_{i,j},u_{i+1,j})}(d)
\end{aligned} \tag{4.27}$$

if, as we assumed without loss of generality, M_0 extends from left to right and from top to bottom along the rows and columns of $\mathcal{I}_{N,L}$, respectively. Lastly, a point-wise substitution of (4.27) into the definition of D_4 results in the closed-form error exponent provided in Theorem 2. This way we end the proof.

4.B Proof of Corollary 2

The proof of each claim of the corollary is analogous to the one undertaken for the same claim in Corollary 1. In particular, we simply have to evaluate the limit or the closed-form expression of K at the constraints yielding the values of ρ_0 and ρ_1 associated with each one of the considered scenarios of dependence.

When the local decisions are independent under the hypothesis H_k

$$p_k^{(0,1)}(d) = 1 - p_k^{(1,0)}(d) \tag{4.28}$$

occurs with $k \in \{0,1\}$. This entails that, for $i \in \{1,2,\ldots,N\}$ and $j \in \{1,2,\ldots,L\}$ and independently of the parameter d, the (i,j)-th device takes its own local decision independently of its nearest neighbour nodes. Thus, next step consists in substituting (4.28) in (4.15). As a result of that substitution, we can straightforwardly check that, when the local decisions are independent ($\rho = 0$), the error exponent K equals $D(B(\xi_0)||B(\xi_1))$ with $\xi_k = p_k^{(0,1)}(d)$ for $k \in \{0,1\}$. This concludes the proof of the first claim.

Next we focus on the second claim of Corollary 1. When the local decisions are maximally dependent under H_k, for all $k \in \{0,1\}$ we have that

$$p_k^{(0,1)}(d) = p_k^{(1,0)}(d) = 0 \tag{4.29}$$

Hence, for all $k \in \{0,1\}$, taking into account (4.15) the behaviour of K when the local decisions tend to be maximally dependent is given through the limit

$$\lim_{\rho \to 1} K = \lim_{p_k^{(0,1)}(d), p_k^{(1,0)}(d) \to 0} K \tag{4.30}$$

with $k \in \{0,1\}$. From the closed-form error exponent provided in (4.15) we can easily show that $K \geq 0$. This entails that

$$\lim_{\rho \to 1} K = \lim_{p_k^{(0,1)}(d), p_k^{(1,0)}(d) \to 0} K \geq 0. \tag{4.31}$$

Equivalently, the proof of the second claim is concluded by showing that the aforementioned limit is lower than or equal to zero. Toward this goal, recalling the closed-form expression given in (4.15) we have to derive upper bounds for the limits

$$E(t,r,s) = \lim_{p_k^{(0,1)}(d),p_k^{(1,0)}(d)\to 0} \pi_0^{(t)}(d)\, p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln\left(\frac{p_0^{(t,r)}(d)\, p_0^{(r,s)}(d)\, \hat{p}_1^{(t,s)}(d)}{p_1^{(t,r)}(d)\, p_1^{(r,s)}(d)\, \hat{p}_0^{(t,s)}(d)}\right) \tag{4.32}$$

with $(t,r,s) \in \mathcal{U}^3$, $\hat{p}_k^{(t,s)}(d)$ is defined in (4.12) and

$$\pi_0^{(0)}(d) = 1 - \pi_0^{(1)}(d) = \frac{p_0^{(1,0)}(d)}{p_0^{(1,0)}(d) + p_0^{(0,1)}(d)}. \tag{4.33}$$

By noting that $p_k^{(u_{m-1},u_m)}(d) \in (0,1)$ and that

$$\hat{p}_k^{(u_{m-1},u_m)}(d) \geq p_k^{(u_{m-1},r)}(d)\, p_k^{(r,u_m)}(d) \tag{4.34}$$

for all $u_{m-1}, u_m, r, k \in \{0,1\}$, we have the subsequent upper bound for $E(t,r,s)$

$$\begin{aligned}
E(t,r,s) &\leq -2 \lim_{p_k^{(0,1)}(d),p_k^{(1,0)}(d)\to 0} p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln\left(p_1^{(t,r)}(d)\, p_1^{(r,s)}(d)\right) \\
&= -2 \lim_{p_k^{(0,1)}(d),p_k^{(1,0)}(d)\to 0} p_0^{(t,r)}(d)\, p_0^{(r,s)}(d)\, \ln\left(p_1^{(t,r)}(d)\right) \\
&\quad - 2 \lim_{p_k^{(0,1)}(d),p_k^{(1,0)}(d)\to 0} p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln\left(p_1^{(r,s)}(d)\right) \\
&\leq -2 \lim_{p_k^{(0,1)}(d),p_k^{(1,0)}(d)\to 0} p_0^{(t,r)}(d)\, \ln\left(p_1^{(t,r)}(d)\right) \\
&\quad - 2 \lim_{p_k^{(0,1)}(d),p_k^{(1,0)}(d)\to 0} p_0^{(r,s)}(d) \ln\left(p_1^{(r,s)}(d)\right)
\end{aligned} \tag{4.35}$$

for all $(t,r,s) \in \mathcal{U}^3$. At this point, we have to recall that

$$\lim_{x,y\to 0} x \ln(y) = 0. \tag{4.36}$$

Using this limit, whose solutions is obtained by applying l'Hôpital rule after changing x and y to polar coordinates, we can straightforwardly check that $E(t,r,s) \leq 0$ for all $(t,r,s) \in \mathcal{U}^3$. This last result together with (4.31) proves the second claim of the corollary, i.e

$$\lim_{\rho \to 1} K = 0. \tag{4.37}$$

Finally, we consider the last claim of the Corollary 2. If the binary local quantizations are maximally dependent under H_1 and independent under H_0, (4.28) holds for $k = 0$ whereas (4.29) happens for $k = 1$. i.e.

$$p_1^{(0,1)}(d) = p_1^{(1,0)}(d) = 0 \tag{4.38}$$

and

$$p_0^{(0,1)}(d) = 1 - p_0^{(1,0)}(d) = \xi_0. \tag{4.39}$$

Thus, with the aim of proving the third claim of the considered corollary we have to evaluate the limit

$$\begin{aligned} \lim_{\rho_1 \to 1} K(\rho_0 = 0) &= \lim_{(p_1^{(0,1)}(d), p_1^{(1,0)}(d)) \to (0,0)} K(p_0^{(0,1)}(d) = 1 - p_0^{(1,0)}(d) = \xi_0) \\ &= \sum_{(t,r,s) \in \mathcal{U}^3} E(t,r,s). \end{aligned} \tag{4.40}$$

where $E(t,r,s)$ is given by (4.32). This time, using (4.12) the general expression for each one of the eight possible evaluations of $E(t,r,s)$ can be rewritten as the sum of two different terms

$$\begin{aligned} &E(t,r,s) \\ &= \lim_{(p_1^{(0,1)}(d), p_1^{(1,0)}(d)) \to (0,0)} \pi_0^{(t)}(d)\, p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln \left(\frac{p_0^{(t,r)}(d)\, p_0^{(r,s)}(d)}{\hat{p}_0^{(t,s)}(d)} \right) \\ &\quad + \lim_{(p_1^{(0,1)}(d), p_1^{(1,0)}(d)) \to (0,0)} \pi_0^{(t)}(d)\, p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln \left(\frac{\hat{p}_1^{(t,s)}(d)}{p_1^{(t,r)}(d)\, p_1^{(r,s)}(d)} \right) \\ &= \underbrace{\pi_0^{(t)}(d)\, p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln \left(\frac{p_0^{(t,r)}(d)\, p_0^{(r,s)}(d)}{\hat{p}_0^{(t,s)}(d)} \right)}_{\triangleq E_1(t,r,s)} \end{aligned}$$

$$+\underbrace{\lim_{(p_1^{(0,1)}(d),p_1^{(1,0)}(d))\to(0,0)} \pi_0^{(t)}(d)\, p_0^{(t,r)}(d)\, p_0^{(r,s)}(d) \ln\left(1+\frac{1}{p_1^{(t,\overline{r})}(d)\, p_1^{(\overline{r},s)}(d)}\right)}_{\triangleq E_2(t,r,s)} \tag{4.41}$$

with $p_0^{(u_{m-1},u_m)}(d)$ given by (4.39) for all $u_{m-1}, u_m \in \{0,1\}$,

$$\pi_0^{(0)}(d) = 1 - \pi_0^{(1)}(d) = 1 - \xi_0, \tag{4.42}$$

$$\hat{p}_0^{(0,1)}(0) = 1 - \hat{p}_0^{(1,0)}(d) = \xi_0 \tag{4.43}$$

and

$$\overline{r} = \begin{cases} 0 & \text{if } r = 1, \\ 1 & \text{if } r = 0. \end{cases} \tag{4.44}$$

Due to the fact that $\xi_0 \in (0,1)$, from (4.39), (4.42) and (4.43) we can easily realize that $E_1(t,r,s) < \infty$ for all $(t,r,s) \in \mathcal{U}^3$. On the contrary, a straightforward analysis reveals

$$E_2(t,r,s) = \begin{cases} \infty & \text{if } s = r \text{ or } t = r, \\ E_3 & \text{otherwise.} \end{cases} \tag{4.45}$$

with $E_3 < \infty$ and $(t,r,s) \in \mathcal{U}^3$. From this last expression we are able to assert that the limit (4.41) is equal to infinity. This concludes the proof of the corollary.

Chapter 5

Neyman-Pearson detection in tandem networks with dependent observations

5.1 Introduction

Due to its reduced energy consumption and bandwidth requirements for simultaneous transmission, the serial distributed configuration with binary communication between the fusion units has been one of the architectures generating more interest among researchers of several disciplines. For instance, the one corresponding with the sequential detection problem addressed in [Cover, 1969], [Hellman and Cover, 1970] and [Koplowitz, 1975]. Note that a network where the binary global decision of the system is performed sequentially by different data fusion units arranged in tandem is equivalent to a single node with one bit of memory and performing observations at different time periods.

Initially, under the assumption of independent sensor observations the study of tandem networks has been based on the characterization of their detection performance when they are implementing a set of optimal decision rules. Firstly, due to the explosive combinatorial complexity that appears in the analysis of these optimal systems, works such as [Ekchian and Tenney, 1982] and [Reibman and Nolte, 1987] were restricted to networks with a moderate size. Later, [Tang et al., 1991] used optimal control theory in order to derive numerical algorithms that

yield the optimal strategy under the Bayesian and the Neyman-Pearson formulations. Since the complexity of the previous algorithms is linear in the number of fusion units, for tandem networks with a larger size they were able to study the best placement for a node in the system and how the nodes should weigh and combine distributed data from multiple sources. Afterwards, to further understand the effects of distribution of the data processing, in [Papastavrou and Athans, 1992] and [Tay et al., 2008b] the aforementioned characterizations and comparisons were undertaken in an asymptotic regime where the number of sensors approached infinity. Under a Bayesian set up, [Papastavrou and Athans, 1992] derived the necessary and sufficient conditions for the error probability decrease to zero as the number of fusion stages increases. Complementing these results and establishing a long-standing conjecture, the authors in [Tay et al., 2008b] showed that the rate of decay for the probability of error is always subexponential when a Bayesian or a Neyman-Pearson decentralized binary hypothesis test is performed by a sensor network with tandem architecture. This suggests that the detection performance of the tandem configuration is worse than the one associated with a parallel architecture where, as it is shown in [Tsitsiklis, 1988], the probability of error decays exponentially fast with the number of sensors. At the same time, the proof provided in [Tay et al., 2008b] has motivated the use of the tandem configuration for the study of more complicated tree architectures that reduce the energy consumption of the parallel configuration and achieve an exponential rate of decay for the probability of error. Some of the first results on this issue are [Tay et al., 2008a] and [Tay et al., 2009]. In these works Tay *et al.* performed a characterization of the asymptotic detection performance of bounded height trees in order to show that, for both the Neyman-Pearson formulation and the Bayesian set up, the error probabilities decay exponentially fast with the number of observations.

Taking into account that in many real scenarios the independence assumption among the sensor observations is violated, some researchers have considered the design of such systems when the local sensor observations are dependent given any hypothesis. So far, there are very few results regarding the design of distributed detection systems with dependent observations and a serial architecture. Under the Bayesian set up, in one of the works that addresses the design of such systems, [Xiang and Wang, 2006] provided the necessary conditions for optimal decision rules. As an extension of the results obtained in [Blum, 1996] and [Yan and Blum, 2000], the aforementioned

necessary conditions were also derived in [Yan and Blum, 2001] for a tandem network performing a Neyman-Pearson test on dependent observations. Unlike the scenario where the sensor observations are independent, the two previous results showed that, for both the Bayesian and the Neyman-Pearson formulations, the optimal fusion rules of tandem networks with dependent observations are generally not likelihood ratio tests.

Up to now, as far as the authors are concerned, research on the detection performance of optimal tandem networks with dependent observations has been lacking. Motivated by this last fact and following the trend of publications such as [Tay et al., 2008a] and [Tay et al., 2009], in this chapter we extend the results provided in [Papastavrou and Athans, 1992] in order to characterize the detection performance of the tandem networks with dependent observations. In particular, considering that the sensor observations are conditionally dependent in a network as the one shown in Figure 5.1, s.t. a given upper bound on the overall probability of false alarm we derive necessary and sufficient conditions for the probability of misdetection go to zero as the number of fusion stages increases. Afterwards we extend these conditions under the Bayesian set up. In other words, we derive the necessary and sufficient conditions that make the overall probability of error go to zero as the number of fusion nodes approaches infinity. Finally, complementing the previous results we provide several practical scenarios where the aforementioned conditions are illustrated.

The rest of this chapter is organized as follows. Section 5.2 is devoted to the problem statement. In Section 5.3 under the Neyman-Pearson formulation we obtain necessary and sufficient conditions for the overall probability of misdetection go to zero as the number of fusion stages increases. Later, in the same section the previous result is extended to the Bayesian set up. Next, using several practical scenarios Section 5.4 illustrates the conditions obtained in the previous section. After these examples, Section 5.5 summarizes the work undertaken in this chapter. Finally, as a matter of organization style, we defer two proofs of this chapter to a pair of appendices.

5.2 Problem statement

We consider the characterization of the detection performance achieved by a sensor network with the serial configuration shown in Figure 5.1. This network is formed by $\ell = N \cdot L$ devices with $N - 1$ equal to the number of

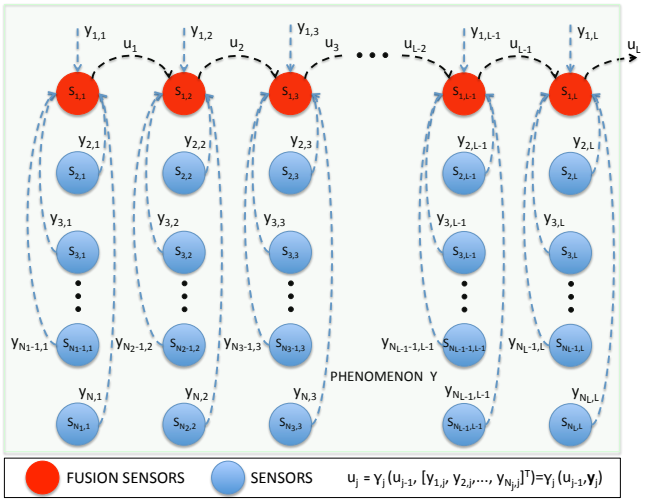


Figure 5.1: *Block diagram of a distributed detection system with tandem architecture.*

sensor devices that directly transmit their measurements to one of the $L \geq 1$ fusion units of the system. In order to decide what state of the phenomenon, H_0 or H_1, is present the considered sensor network undertakes the following steps. Firstly, a specific set of $N-1$ devices performs observations of the environment and transmits its measurements to a fusion sensor, $S_{1,j}$, over error free parallel access channels. Secondly, in order to make a binary local decision, u_j, regarding the presence or absence of the event of interest, the aforementioned fusion sensor uses the local decision rule $\gamma_j(u_{j-1}, \boldsymbol{y}_j)$ with $u_0 = \emptyset$. Using this rule the j-th fusion unit fuses the local decision u_{j-1} taken by its preceding fusion sensor $S_{1,j-1}$, and the vector formed by its own measurement as well as the $N-1$ received measurements, $\boldsymbol{y}_j = [y_{1,j}, y_{2,j}, \ldots, y_{N,j}]^T \in \mathcal{Y}_{j,k}$ where $\mathcal{Y}_{j,k}$ equals to the set of all possible values that the random variable $\boldsymbol{Y}_j$ can take on under hypothesis H_k with $k \in \{0, 1\}$. Finally, the resulting local decision is transmitted to the successor of the considered fusion stage, i.e. $S_{1,j+1}$. Thus, the final decision of the distributed detection system matches the one made by the last fusion node in the tandem architecture, in other words, $S_{1,L}$ in Figure 5.1.

For the previous sensor network we suppose that the sensor observations may be dependent under both hypothesis. This can be a result of the presence of correlation sources in signals such as the noise or the one associated

with the event of interest in the detection problem. At the same time, this implies that the local decisions performed at each fusion stage, U_j, and the sensor observations, $Y_{i,j}$, may be dependent under hypothesis H_k with $i = \{1, 2, \ldots, N\}$, $j = \{1, 2, \ldots, L\}$ and $k \in \{0, 1\}$. Nevertheless, we have to remark that we consider a correlation scenario where, under both hypothesis and according to the following definition, the number of non-maximally dependent sensor observations is strictly increasing with the number of fusion stages, L.

Definition 1. *The observations, $\boldsymbol{Y}_j$ and $\boldsymbol{Y}_{j'}$, received by two different fusion stages with indices j and j' are said to be maximally dependent under hypothesis H_k only if they have the same cardinality,*

$$\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{Y}_j|\boldsymbol{Y}_{j'},H_k}\right) = 0$$

and

$$\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{Y}_{j'}|\boldsymbol{Y}_j,H_k}\right) = 0$$

where $tr(\cdot)$ denotes the trace operator and $\boldsymbol{C}_{\boldsymbol{Y}_j|\boldsymbol{Y}_{j'},H_k}$ is equal to covariance matrix of $\boldsymbol{Y}_j$ given $\boldsymbol{Y}_{j'}$ and hypothesis H_k with $k \in \{0, 1\}$, $j \neq j'$ and $j, j' \in \{1, 2, \ldots, L\}$.

Intuitively, the previous definition states that two sensor observations $\boldsymbol{Y}_j$ and $\boldsymbol{Y}_{j'}$, are maximally dependent only if there is no uncertainty in $\boldsymbol{Y}_j$ once $\boldsymbol{Y}_{j'}$ is known and vice-versa. Consequently, if we take into account that many phenomena present micro-scale variations caused by the so-called nugget effect (see [Møller, 2003]), we can easily realize that the major part of the detection scenarios belong to the setting considered in this problem statement.

Regarding the detection process of the network the probability of error associated with the j-th fusion stage is

$$P_e(j) = P(H_0) \cdot P_{_{FA}}(j) + (1 - P(H_0)) \cdot P_M(j) \tag{5.1}$$

with $j \in \{1, 2, \ldots, L\}$, $P(H_0) \in (0, 1)$ equal to the prior probability of hypothesis H_0,

$$P_M(j) = 1 - P_D(j) = P(U_j = 0|H_1) \tag{5.2}$$

denoting the corresponding probability of misdetection ($P_D(j)$ being the probability of detection) and

$$P_{_{FA}}(j) = P(U_j = 1|H_0) \tag{5.3}$$

equal to the probability of false alarm at the j-th fusion node. Under the assumption that each sensor device knows its fusion node and that each fusion node knows its successor, i.e. the fusion node receiving its local decision, we suppose that the distributed detection system solves the following binary hypothesis testing problem under a specific formulation

$$H_k\text{:} \quad \boldsymbol{Y} \sim f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k) \tag{5.4}$$

where, for $k \in \{0, 1\}$,

- $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_{L-1}^T, \boldsymbol{Y}_L^T]^T$ is the set of observations performed by the devices that form the network.

- $f_{\boldsymbol{Y}|H}(\boldsymbol{y}|H_k)$ is the joint p.d.f. of the sensor observations under hypothesis H_k.

Therefore, given the definitions provided in (5.2) and (5.3), under the Neyman-Pearson formulation this means that the tandem network applies the set of local fusion rules,

$$\Gamma^{\text{NP}} = \{\gamma_j^{\text{NP}}(u_{j-1}, \boldsymbol{y}_j)\}_{j=1}^L, \tag{5.5}$$

that maximizes the probability of detection of the final decision of the system, $P_D(L)$, when its corresponding probability of false alarm is lower than or equal to a constant $\alpha_L \in (0, 1)$, i.e.

$$P_{_{FA}}(L) = P(U_L = 1|H_0) \leq \alpha_L \in (0, 1). \tag{5.6}$$

On the contrary, if the tandem network solves (5.4) under the Bayesian set up, the applied set of local fusion rules

$$\Gamma^{\text{B}} = \{\gamma_j^{\text{B}}(u_{j-1}, \boldsymbol{y}_j)\}_{j=1}^L, \tag{5.7}$$

minimizes the probability of error of the system, $P_e(L)$, defined in (5.1) for $j = L$.

5.3 System analysis

The optimal strategies that, under dependent sensor observations, have to be applied by a tandem network performing a Bayesian and a Neyman-Pearson test are derived in [Xiang and Wang, 2006] and [Yan and Blum, 2001] respectively. However, as far as the authors are concerned, little is known about necessary and sufficient conditions that, under both formulations, make the corresponding probability of error go to zero as the number of fusion stages approaches infinity. The only results addressing this issue are obtained in [Papastavrou and Athans, 1992] under the Bayesian formulation. In that work the authors restricted to a scenario where the sensor observations are i.i.d., $N_j = 1$ for all $j \in \{1, 2, \ldots, L\}$, and all the fusion stages are identical in terms of their receiving operating curve (ROC). Motivated by this last fact, without the conditional independence assumption and under both the Bayesian set up and the Neyman-Pearson formulation we derive necessary and sufficient conditions for asymptotic perfect detection.

5.3.1 The Neyman-Pearson case

Under the Neyman-Pearson formulation the tandem network applies the strategy defined in (5.5). Consequently, if we want to determine when asymptotic perfect detection happens we have to derive conditions ensuring that, for any arbitrary upper bound on the overall probability of false alarm, the corresponding probability of misdetection goes to zero as the number of fusion stages approaches infinity. Specifically, we do so by firstly defining the following quantities

$$\Lambda\left(\boldsymbol{Y}_j\right) = \ln\left(\frac{f_{\boldsymbol{Y}_j|H}(\boldsymbol{y}_j|H_1)}{f_{\boldsymbol{Y}_j|H}(\boldsymbol{y}_j|H_0)}\right), \tag{5.8}$$

$$\Lambda\left(\boldsymbol{Y}_j|\, u_{j-1}\right) \begin{cases} \Lambda\left(\boldsymbol{Y}_1\right) & \text{if } j = 1, \\ \ln\left(\frac{f_{\boldsymbol{Y}_j|U_{j-1},H}(\boldsymbol{y}_j|u_{j-1},H_1)}{f_{\boldsymbol{Y}_j|U_{j-1},H}(\boldsymbol{y}_j|u_{j-1},H_0)}\right) & \text{if } j \neq 1, \end{cases} \tag{5.9}$$

$$t_j(u_{j-1}) = \begin{cases} \tau_1 & \text{if } j = 1, \\ \tau_j - \ln\left(\frac{P_{U_{j-1}|H}(u_{j-1}|H_1)}{P_{U_{j-1}|H}(u_{j-1}|H_0)}\right) & \text{if } j \neq 1 \end{cases} \tag{5.10}$$

and

$$F_j^{(u_{j-1})}(s|H_k) = \begin{cases} P(\Lambda(\boldsymbol{Y}_1) < s|\, H_k) & \text{if } j = 1, \\ P(\Lambda(\boldsymbol{Y}_j|\, u_{j-1}) < s|\, H_k) & \text{if } j \neq 1, \end{cases} \tag{5.11}$$

with $j \in \{1, 2, \ldots, L\}$, $u_0 = \emptyset$, τ_j denoting a bounded constant dependent on the fusion stage and $f_{\boldsymbol{Y}_j|U_{j-1},H}(\boldsymbol{y}_j|u_{j-1}, H_k)$ equal to the conditional p.d.f. of the sensor observations $\boldsymbol{Y}_j$ given the binary local decision U_{j-1} under H_k.

Since u_0 is equal to the empty set, note that the quantities defined in (5.9)-(5.11) do not depend on the value of u_{j-1} when $j = 1$. In fact, realize that $t_1 = \tau_1$ and that the probability density functions $f_{\boldsymbol{Y}_1|H}(\boldsymbol{y}_1|H_k)$, with $k \in \{0, 1\}$, are the only ones involved in (5.9) and (5.11). However, bearing in mind this substantial difference in the definition, for the sake of simplicity in the notation and from now on every quantity conditioned on the preceding local decision will be equally denoted for every fusion stage with index $j \in \{1, 2, \ldots, L\}$. This way, we are now at the position to prove the subsequent lemma.

Lemma 1. *Consider an optimal tandem team that, as it is described in Section 5.2, is formed by L fusion stages and performs a Neyman-Pearson hypothesis test on dependent observations. Then, for any arbitrary constraint* $P_{FA}(L) \leq \alpha_L \in (0, 1)$*, the limit of the optimal probability of misdetection as L goes to infinity does not converge to zero as long as the support of* $f_{\boldsymbol{Y}_L|U_{L-1},H}(\boldsymbol{y}_L|u_{L-1}, H_k)$ *consists of two events, unambiguously indexed by* $I_L \in \{0, 1\}$*, and distributed as shown in Table 5.1 with*

$$A^{(u_{L-1})}(L) \in \begin{cases} (0.5, 1) & \textit{if } u_{L-1} = 0, \\ (0.5, 1] & \textit{if } u_{L-1} = 1, \end{cases} \tag{5.12}$$

and

$$B^{(u_{L-1})}(L) \in \begin{cases} (0.5, 1] & \textit{if } u_{L-1} = 0, \\ (0.5, 1) & \textit{if } u_{L-1} = 1. \end{cases} \tag{5.13}$$

Proof. From classical detection theory detailed in [Poor, 1994, section II.D] as well as the results obtained in [Viswanathan et al., 1988] and [Yan and Blum, 2001, Section V] we can easily check that, in a tandem network performing a Neyman-Pearson test, the last fusion stage applies the

H \ $\boldsymbol{Y}_L$	$I_L = 0$	$I_L = 1$
H_0	$A^{(u_{L-1})}(L)$	$1 - A^{(u_{L-1})}(L)$
H_1	$1 - B^{(u_{L-1})}(L)$	$B^{(u_{L-1})}(L)$

Table 5.1: *Special probability distribution of the N_L observations taken by the L-th fusion unit, $\boldsymbol{Y}_L$, when the preceding fusion node decides $u_{L-1} \in \{0, 1\}$.*

subsequent LLRT

$$u_L = \gamma_L\left(u_{L-1}, \boldsymbol{y}_L\right) = \begin{cases} 1 & \text{if } \ln\left(\frac{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_1)}{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_0)}\right) > \tau_L, \\ \gamma & \text{if } \ln\left(\frac{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_1)}{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_0)}\right) = \tau_L, \\ 0 & \text{otherwise,} \end{cases} \tag{5.14}$$

where, for $k \in \{0, 1\}$,

- $f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_k)$ denotes the joint p.d.f. of the sensor observations $\boldsymbol{Y}_L$ and the binary local decision U_{L-1} under H_k.

- τ_L is the smallest number such that $\varsigma \leq \alpha_L \in (0, 1)$ with

$$\varsigma = P\left(\ln\left(\frac{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_1)}{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_0)}\right) > \tau_L \middle| H_0\right). \tag{5.15}$$

- $\gamma \in \{0, 1\}$ is a randomization function that takes on the value 1 with the subsequent probability

$$\nu = \begin{cases} 1 & \text{if } \vartheta = 0, \\ \frac{\alpha_L - \varsigma}{\vartheta} & \text{otherwise} \end{cases} \tag{5.16}$$

with

$$\vartheta = P\left(\ln\left(\frac{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_1)}{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_0)}\right) = \tau_L \middle| H_0\right). \tag{5.17}$$

If we now use the chain rule of probability detailed in [Papoulis and Pillai, 2002]

$$\ln\left(\frac{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_1)}{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_0)}\right) = \Lambda\left(\boldsymbol{Y}_L | \, u_{L-1}\right) + \ln\left(\frac{P_{U_{L-1}|H}(u_{L-1}|H_1)}{P_{U_{L-1}|H}(u_{L-1}|H_0)}\right) \tag{5.18}$$

notice that (5.14) is equivalent to

$$u_L = \gamma_L\left(u_{L-1}, \boldsymbol{y}_L\right) = \begin{cases} 1 & \text{if } \Lambda\left(\boldsymbol{Y}_L | \, u_{L-1}\right) > t_L(u_{L-1}), \\ \gamma & \text{if } \Lambda\left(\boldsymbol{Y}_L | \, u_{L-1}\right) = t_L(u_{L-1}), \\ 0 & \text{otherwise,} \end{cases} \tag{5.19}$$

with $t_L(u_{L-1})$ defined in (5.10) when τ_L is equal to the minimum integer ensuring that $\varsigma \leq \alpha_L \in (0,1)$. Consequently, under the set of conditional observation models described by Table 5.1 together with (5.12) and (5.13), taking into account the definitions given in (5.9)-(5.11), the conditional ROC curve of the last fusion stage given the preceding local decision, $\mathrm{ROC}_L(u_{L-1})$, matches the plot shown in Figure 5.2 with

$$\begin{aligned} P_D^{(u_{L-1})}(L|t_L(u_{L-1})) &= P(U_L = 1|u_{L-1}, t_L(u_{L-1}), H_1) \\ &= 1 - F_L^{(u_{L-1})}\left(t_L(u_{L-1})|H_1\right) \end{aligned} \tag{5.20}$$

and

$$\begin{aligned} P_{FA}^{(u_{L-1})}(L|t_L(u_{L-1})) &= P(U_L = 1|u_{L-1}, t_L(u_{L-1}), H_0) \\ &= 1 - F_L^{(u_{L-1})}\left(t_L(u_{L-1})|H_0\right) \end{aligned} \tag{5.21}$$

for all $u_{L-1} \in \{0,1\}$. In addition, taking into account that the two possible values of $\boldsymbol{y}_L$ can be unambiguously indexed by I_L, we also have that the last fusion stage of the considered tandem network can only apply six different fusion rules, $\gamma_L\left(u_{L-1}, I_L\right)$, satisfying the monotonicity property given in [Varshney, 1997]. Next, using the expressions provided in Table 5.2 we can easily check that, even in the most optimistic scenario where the penultimate fusion stage has zero probability of error, none of them but one, $\gamma_L\left(u_{L-1}, I_L\right) = u_{L-1}$, achieves zero probability of misdetection for any arbitrary constraint on the corresponding probability of false alarm. To be more precise, evaluating the expressions given in Table 5.2 at $P_D(L-1) = 1$ and $P_{FA}(L-1) = \alpha_L \in (0,1)$, none of them but $\gamma_L\left(u_{L-1}, I_L\right) = u_{L-1}$ achieves $P_D(L) = 1 - P_M(L) = 1$ at the same time as $P_{FA}(L) \leq \alpha_L \in (0,1)$.

We now know that, under the Neyman-Pearson formulation and under the observation model considered in this lemma, an optimal tandem network performing a Neyman-Pearson test can only achieve asymptotic perfect detection in two cases. On the one hand, one of these two settings would be a scenario where the first fusion stage achieves $P_M(1) = 0$ for $P_{FA}(1) \leq \alpha_L \in (0,1)$ and where $\gamma_j\left(u_{j-1}, I_j\right) = u_{j-1}$ for all $j \in \{2, \ldots, L\}$.

Nevertheless, we can realize that the deployment of a tandem network would not be practical in that scenario. On the other hand, the second setting occurs when $\gamma_L(u_{L-1}, I_L) = u_{L-1}$ and when the penultimate fusion stage has zero probability of misdetection with $P_{_{FA}}(L-1) \leq \alpha_L \in (0,1)$. However, as long as the fusion stages transmit binary messages to their successors this scenario cannot happen because, even in the most optimistic scenario, the other five fusion rules cannot achieve zero probability of misdetection for any arbitrary constraint on the corresponding probability of false alarm. This means that, if the last fusion stage presents one of the conditional observation models expressed in Table 5.1 together with (5.12) and (5.13), the optimal probability of misdetection of the infinite tandem team does not converge to zero for any arbitrary constraint on the overall probability of false alarm. This way, we conclude the proof. □

In Lemma 1 we have provided a set of conditional observation models for which there is no asymptotic perfect detection in a tandem network performing a Neyman-Pearson hypothesis test. Consequently, we have proved the following corollary.

Corollary 3. *Under the Neyman-Pearson formulation consider the tandem network described in Section 5.2. Then, for any arbitrary constraint on the overall probability of false alarm, the corresponding probability of misdetection does not have to converge to zero as the number of fusion stages, L, approaches infinity.*

At this point, we have shown that the result of the previous corollary happens when the last fusion stage of the tandem network has one of the conditional observation models belonging to the set described in Table 5.1 and parametrized by (5.12) and (5.13). However, under the Neyman-Pearson set up and dependent observations we still do not know necessary and sufficient conditions ensuring asymptotic perfect detection. Due to this last fact, in the subsequent theorem we initially derive necessary conditions that have to be satisfied in a scenario where the optimal probability of misdetection of an infinite tandem team goes to zero for an arbitrary upper bound on the corresponding probability of false alarm.

Theorem 3. *Under the Neyman-Pearson formulation consider the tandem network described in Section 5.2. For $L \geq 2$ define $m_{0,L}^{(u_{L-1})}$ and $m_{1,L}^{(u_{L-1})}$*

as the initial and final slopes of the conditional ROC curve $\mathrm{ROC}_L(u_{L-1})$ *respectively. Then, for any* $P_{_{FA}}(L) \leq \alpha_L \in (0,1)$*, in order to achieve*

$$\lim_{L\to\infty} P_M(L) = 0$$

$m_{0,L}^{(0)} = \infty$ *or* $m_{1,L}^{(1)} = 0$ *are necessary conditions.*

Proof. To prove the theorem assume that

$$m_{0,L}^{(0)} < \infty, \tag{5.22}$$

and that

$$m_{1,L}^{(1)} > 0. \tag{5.23}$$

In addition, taking into account that, in an optimal tandem network performing Neyman-Pearson test, the last fusion stage applies the LLRT given in (5.19), we also have that the conditional ROC curve, $\mathrm{ROC}_L(u_{L-1})$, associated with the last fusion stage is concave downward for all $u_{L-1} \in \{0,1\}$. Under these assumptions, we can easily note that the detection performance of the last fusion stage is worse than or equal to the detection performance of a tandem network where the conditional ROC curves of the last fusion stage belong to the set characterized by Figure 5.2 together with (5.12) and (5.13). At the same time, by Lemma 1 we already know that, for this kind of tandem networks, the limit of the optimal probability of misdetection as the number of fusion stages goes to infinity could stay bounded away from zero under some upper bounds on the corresponding probability of false alarm. Therefore, if an infinite tandem network with dependent observations consists of fusion units whose conditional ROC curves satisfy (5.22) and (5.23), it will never achieve zero probability of misdetection for any arbitrary constraint on the overall probability of false alarm. This concludes the proof of the theorem. □

Using the necessary conditions provided in the previous theorem it can be checked if a tandem network cannot achieve asymptotic perfect learning under the Neyman-Pearson formulation. Note that, for either discrete or continuous sensor observations, if (5.22) and (5.23) are satisfied, in an infinite tandem team the overall probability of misdetection cannot converge to zero for any arbitrary constraint on the overall probability of false alarm. On the contrary, if either (5.22) or (5.23) is violated, asymptotic perfect detection

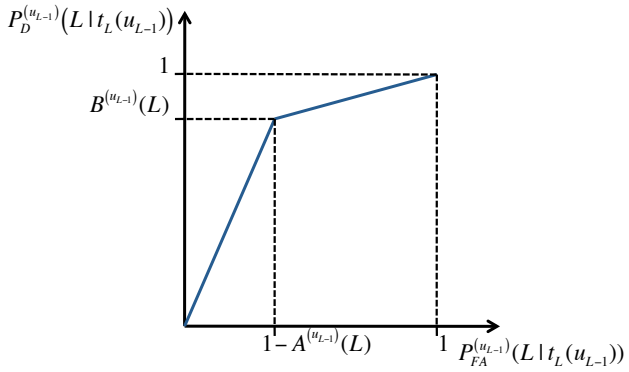


Figure 5.2: *Conditional ROC curve, $ROC_L(u_{L-1})$, associated with the last fusion stage of an optimal tandem network performing a Neyman-Pearson test and whose conditional observation model is given by Table 5.1 with $u_{L-1} \in \{0,1\}$.*

cannot be guaranteed for an optimal tandem network performing a Neyman-Pearson test. Indeed, in order to ensure this, we need to check the sufficient conditions provided in the following theorem.

Theorem 4. *Under the Neyman-Pearson formulation consider the tandem network described in Section 5.2. Additionally, assume that, under H_0 and H_1, the p.d.f. of $\Lambda(\boldsymbol{Y}_j|u_{j-1})$ contains no point masses for all $j \in \{1,2,\ldots,L\}$ and $u_{j-1} \in \{0,1\}$. For $j \in \{1,2,\ldots,L\}$ and $u_0 = \emptyset$ also define $m_{0,j}^{(u_{j-1})}$ and $m_{1,j}^{(u_{j-1})}$ as the initial and final slopes of the conditional ROC curve $\mathrm{ROC}_j(u_{j-1})$ respectively. Then, for any $P_{FA}(L) \leq \alpha_L \in (0,1)$, in order to achieve*

$$\lim_{L\to\infty} P_M(L) = 0$$

$m_{0,j}^{(0)} = \infty$ or $m_{1,j}^{(1)} = 0$ for all $j \in \{1,2,\ldots,L\}$ are sufficient conditions.

Proof. Basically, assuming that the sufficient conditions hold the proof of this theorem results in proposing fusion rules so that, although they are not necessarily optimal, the probability of misdetection of the resulting tandem network asymptotically converges to zero under the Neyman-Pearson formulation. Equivalently, we have to show that given any $\delta \in (0,1)$ and any constraint $P_{FA}(L) \leq \alpha_L \in (0,1)$, if the proposed decision rules are employed and the sufficient conditions occur, the probability of misdetection of the tandem team can be made less than δ, as L goes to infinity. This proof is

$\gamma_L\left(u_{L-1}, I_L\right)$	$P_{_{FA}}(L)$	$P_D(L)$
0	0	0
1	1	1
u_{L-1}	$P_{_{FA}}(L-1)$	$P_D(L-1)$
I_L	$\left(1-P_{_{FA}}(L-1)\right)\left(1-A^{(0)}(L)\right)$ $+P_{_{FA}}(L-1)\left(1-A^{(1)}(L)\right)$	$\left(1-P_D(L-1)\right)B^{(0)}(L)$ $+P_D(L-1)B^{(1)}(L)$
$\mathrm{OR}(u_{L-1}, I_L)$	$\left(1-P_{_{FA}}(L-1)\right)\left(1-A^{(0)}(L)\right)$ $+P_{_{FA}}(L-1)$	$\left(1-P_D(L-1)\right)B^{(0)}(L)$ $+P_D(L-1)$
$\mathrm{AND}(u_{L-1}, I_L)$	$P_{_{FA}}(L-1)\left(1-A^{(1)}(L)\right)$	$P_D(L-1)B^{(1)}(L)$

Table 5.2: *Probability of false alarm, $P_{_{FA}}(L)$, and probability of detection, $P_D(L)$ associated with each one of the local fusion rules that can be applied by the last fusion stage belonging to an optimal tandem team performing a Neyman-Pearson test and whose conditional observation model is described in Table 5.1.*

---

detailed in Appendix 5.A for $m_{0,j}^{(0)} = \infty$ for all $j \in \{1, 2, \ldots, L\}$. For the sake of brevity, due to the fact that it requires an analogous development, we will omit the proof for $m_{1,j}^{(1)} = 0$ for all $j \in \{1, 2, \ldots, L\}$. □

It can be straightforwardly checked that the sufficient conditions derived in Theorem 4 are valid when, for all $u_{j-1} \in \{0, 1\}$ and all $j \in \{1, 2, \ldots, L\}$, $\Lambda\left(\boldsymbol{Y}_j | u_{j-1}\right)$ is a continuous scalar random variable whose p.d.f. does not contain any point masses of probability under both hypothesis. Since the selection of a specific strategy Γ, optimal or suboptimal, involves a partition of the observation space we have that, for all $j \in \{2, \ldots, L\}$ and $u_{j-1} \in \{0, 1\}$,

$$\Lambda\left(\boldsymbol{Y}_j | \, u_{j-1}\right) = \begin{cases} \Lambda\left(\boldsymbol{Y}_1\right) & \text{if } j = 1, \\ \\ \Lambda^*\left(\boldsymbol{Y}_j\right) + C_j^*(u_{j-1}) & \text{if } j \in \{2, 3, \ldots, L\} \end{cases} \tag{5.24}$$

where

$$C_j^*(u_{j-1}) = \begin{cases} \frac{1-P_{_{FA}}(j-1)}{1-P_D(j-1)} & \text{if } u_{j-1} = 0, \\ \\ \frac{P_{_{FA}}(j-1)}{P_D(j-1)} & \text{if } u_{j-1} = 1 \end{cases} \tag{5.25}$$

and

$$\Lambda^*\left(\boldsymbol{Y}_j\right) = \log\left(\frac{\int_{\boldsymbol{Y}^{j-1}\in\mathcal{V}_{u_{j-1}}^{j-1}(\Gamma)} f_{\boldsymbol{Y}^j|H}(\boldsymbol{y}^j|H_1)\,\mathrm{d}\boldsymbol{y}^{j-1}}{\int_{\boldsymbol{Y}^{j-1}\in\mathcal{V}_{u_{j-1}}^{j-1}(\Gamma)} f_{\boldsymbol{Y}^j|H}(\boldsymbol{y}^j|H_0)\,\mathrm{d}\boldsymbol{y}^{j-1}}\right) \tag{5.26}$$

with

$$\boldsymbol{Y}^j = [\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_j^T]^T, \tag{5.27}$$

$$\mathcal{Y}_k^{j-1} = \mathcal{Y}_{1,k} \times \mathcal{Y}_{2,k} \times \ldots \times \mathcal{Y}_{j-1,k} \tag{5.28}$$

and $\mathcal{V}_{\hat{k}}^{j-1}(\Gamma)$ denoting the partition of the observation space $\mathcal{Y}^{j-1} = \mathcal{Y}_0^{j-1} \cup \mathcal{Y}_1^{j-1}$ that yields $u_{j-1} = \hat{k} \in \{0,1\}$ under the strategy Γ, i.e.

$$\mathcal{V}_{\hat{k}}^{j-1}(\Gamma) = \left\{ \boldsymbol{Y}^{j-1} \middle|\, u_{j-1}(\Gamma) = \hat{k} \right\}. \tag{5.29}$$

For the sake of clarity but only for this time, the dependency of u_{j-1} w.r.t. Γ has been made explicit. By noting that $C_j^*(u_{j-1})$ is equal to a constant once u_{j-1} equals zero or one, the expression provided in (5.24) means that the aforementioned assumption is guaranteed when the partition $\mathcal{V}_{u_{j-1}}^{j-1}(\Gamma)$ and the observation model make $\Lambda\left(\boldsymbol{Y}_1\right)$ and $\Lambda^*\left(\boldsymbol{Y}_j\right)$ not contain any point masses under H_0 and H_1. In many dependence scenarios this partition can be intricate. Nevertheless, we can check that, independently of the partition, as it is illustrated in one example of Section 5.4, the probability density functions of $\Lambda\left(\boldsymbol{Y}_1\right)$ and $\Lambda^*\left(\boldsymbol{Y}_j\right)$ do not contain point masses in many practical scenarios where the sensor observations $\{\boldsymbol{Y}_j\}_{j=1}^L$ are continuous under both hypothesis. Thus, under these scenarios and under Neyman-Pearson formulation Theorems 3 and 4 provide some convenient tests that determine whether the probability of misdetection always goes to zero or not in an infinite tandem team. It is clear that these tests are based on a set of conditions given in terms of features associated with the conditional ROC curves given the preceding local decision. However, taking full advantage of the properties associated with continuous likelihood ratio tests these conditions can also be expressed in the following equivalent form.

Corollary 4. *Consider the tandem network of Theorem 4 where, for all* $j \in \{1,2,\ldots,L\}$ *and* $u_{j-1} \in \{0,1\}$*, the p.d.f. of* $\Lambda\left(\boldsymbol{Y}_j|u_{j-1}\right)$ *does not contains any point masses under* H_0 *and* H_1*. Also define*

$$\Lambda_j^{\sup}(0) = \sup_{Y_j\in\mathcal{Y}_j^{(0)}} \left\{\Lambda\left(\boldsymbol{Y}_j|\,0\right)\right\} \tag{5.30}$$

and

$$\Lambda_j^{\inf}(1) = \inf_{Y_j \in \mathcal{Y}_j^{(1)}} \{\Lambda(\boldsymbol{Y}_j|\, 1)\} \tag{5.31}$$

where $\mathcal{Y}_j^{(u_{j-1})}$ *equals the set of possible values that* $\boldsymbol{Y}_j$ *can take on when it is distributed according to*

$$\begin{aligned} f_{\boldsymbol{Y}_j|U_{j-1}}(\boldsymbol{y}_j|u_{j-1}) =& f_{\boldsymbol{Y}_j|U_{j-1},H}(\boldsymbol{y}_j|u_{j-1}, H_0) \cdot P(H_0) \\ &+ f_{\boldsymbol{Y}_j|U_{j-1},H}(\boldsymbol{y}_j|u_{j-1}, H_1) \cdot (1 - P(H_0)) \end{aligned} \tag{5.32}$$

with $P(H_0) \in (0,1)$ *equal to prior probability of* H_0, $u_0 = \emptyset$, *and* $u_{j-1} \in \{0,1\}$ *if* $j \in \{1,2,\ldots,L\}$. *Then, for any* $P_{_{FA}}(L) \leq \alpha_L \in (0,1)$, *in order to achieve*

$$\lim_{L\to\infty} P_M(L) = 0$$

$\Lambda_L^{\sup}(0) = \infty$ *or* $\Lambda_L^{\inf}(1) = -\infty$ *are necessary conditions and*

$$\Lambda_-^{\sup}(0) = \min_{j\in\{1,\ldots,L\}} \left\{\Lambda_j^{\sup}(0)\right\} = \infty \tag{5.33}$$

or

$$\Lambda_+^{\inf}(1) = \max_{j\in\{1,\ldots,L\}} \left\{\Lambda_j^{\inf}(1)\right\} = -\infty \tag{5.34}$$

are sufficient conditions.

Proof. In order to prove the corollary we firstly have to recall that the suboptimal strategy employed in the proof of Theorem 4 is based on LLRT of $\Lambda(\boldsymbol{y}_j|u_{j-1})$ when $\boldsymbol{Y}_j = \boldsymbol{y}_j$ and $u_{j-1} = \hat{k} \in \{0,1\}$. Additionally, we have to take into account that, under dependent sensor observations and under the Neyman Pearson formulation, the optimal fusion rule applied by the last fusion stage of a serial configuration matches the LLRT given in (5.19) (see the properties given in [Poor, 1994, section II.D] as well as the results obtained in [Viswanathan et al., 1988] and [Yan and Blum, 2001, Section V]). Since the p.d.f. of $\Lambda(\boldsymbol{y}_j|u_{j-1})$ does not contain any point masses for all $j \in \{1,2,\ldots,L\}$ and all $u_{j-1} \in \{0,1\}$, the proof is completed by showing that the conditions given in Theorems 3 and 4 occur if and only if the conditions provided in the Corollary hold. In particular, consider the following two expressions

$$P_D^{(u_{j-1})}(j|s) = 1 - F_j^{(u_{j-1})}(s|H_1) \tag{5.35}$$

and

$$P_{FA}^{(u_{j-1})}(j|s) = 1 - F_j^{(u_{j-1})}(s|H_0) \tag{5.36}$$

with $F_j^{(u_{j-1})}(s|H_k)$ defined in (5.11) and $s \in (-\infty, \infty)$ denoting the threshold used in a LLRT based on $\Lambda(\boldsymbol{Y}_j|\, u_{j-1})$ for $k \in \{0,1\}$, $u_0 = \emptyset$ and $u_{j-1} \in \{0,1\}$ when $j \in \{2,\ldots,L\}$. Subsequently, if we use the well-known property

$$\frac{\mathrm{d}\, P_D^{(u_{j-1})}(j|s)}{\mathrm{d}\, P_{FA}^{(u_{j-1})}(j|s)} = e^s \tag{5.37}$$

with the corresponding evaluation of u_{j-1}, we can easily check that, for any $j \in \{1,2,\ldots,L\}$, $m_{0,j}^{(0)} = \infty$ if and only if $\Lambda_j^{\sup}(0)$ is not upper bounded by a finite constant. Analogously, from the same property we can notice that, for any $j \in \{1,2,\ldots,L\}$, $m_{1,j}^{(1)} = 0$ if and only if $\Lambda_j^{\inf}(1)$ is not lower bounded by a finite constant. At this point, in order to conclude the proof we only need to realize that $\Lambda_-^{\sup}(0)$ and $\Lambda_+^{\inf}(1)$ are not upper or lower bounded as long as the involved terms, $\Lambda_j^{\sup}(0)$ and $\Lambda_j^{\inf}(1)$, are not upper or lower bounded for all $j \in \{1,2,\ldots,L\}$ respectively. □

According to Corollary 4 and concluding the subsection we can easily check a pair of aspects. First, we can realize that an optimal tandem network performing a Neyman-Pearson test cannot achieve perfect detection as long as there exists some finite constant, $B > 0$, such that

$$\Lambda_L^{\sup}(0) < B \tag{5.38}$$

and

$$\Lambda_L^{\inf}(1) > -B \tag{5.39}$$

for a specific observation model where, for all $j \in \{1,2,\ldots,L\}$ and $u_{j-1} \in \{0,1\}$, the conditional p.d.f. of $\Lambda(\boldsymbol{Y}_j|u_{j-1})$ does not contains any point masses given H_0 and H_1. To put it differently, if (5.38) and (5.39) are satisfied under those kind of observation models, the distributed detection system with tandem architecture will not discriminate H_1 against H_0 with zero probability of misdetection and any arbitrary upper bound on the overall probability of false alarm. Furthermore, from the previous corollary we know that, under the same kind of observation models, the addition of extra fusion stages can make the system achieve any desired level of performance as long as $\Lambda_-^{\sup}(0)$ is not upper bounded or $\Lambda_+^{\inf}(1)$ is not lower bounded. It is of great value to remark that, under the setting studied in [Papastavrou and Athans, 1992],

these conditions result in the necessary and sufficient conditions derived in that work. This shows that the results derived along this section are consistent with previous related works.

5.3.2 The Bayesian case

In this case, in order to solve (5.4) we consider a scenario where the tandem network applies the strategy (5.7) minimizing the probability of error of the system. Consequently, if we want to determine when asymptotic perfect detection happens, this time we have to derive conditions under which the last fusion node converges in probability to deciding the right hypothesis as the number of fusion stages becomes large. Although the formulation is different from the one of the preceding subsection, we can straightforwardly check that the necessary and sufficient conditions are the same. In particular, we also have that the corresponding proofs are closely related to the ones used in Theorems 3 and 4. Due to this last fact and for the sake of brevity they have been omitted. Instead we will only comment the key aspects as well as the minor modifications that have to be taken into account when the conditions are derived under the Bayesian set up.

First of all, note that Lemma 1 can be straightforwardly extended to the Bayesian set up. Similar to the proof carried out under the Neyman-Pearson formulation, considering one of the conditional observation models expressed in Table 5.1 together with (5.12) and (5.13), the key is that, even in the most optimistic scenario where $P_e(L-1)=0$, none of the fusion rules but one, $\gamma_L\left(u_{L-1}, I_L\right)=u_{L-1}$, achieves perfect detection at the last fusion stage. However, in order to check that, under any of the conditional observation models expressed in Table 5.1 together with (5.12) and (5.13), none of the fusion rules achieves $P_e(L)=0$, in this case we have to evaluate the expressions given in Table 5.2 at $P_M(L-1)=P_{FA}(L-1)=0$. At the same time this procedure automatically extends Corollary 3 under the Bayesian set up.

Next, from the results of [Xiang and Wang, 2006] recall that, under the Bayesian set up, the optimal fusion rule associated with the last fusion node is given by the subsequent likelihood ratio test

$$u_L=\gamma_L\left(u_{L-1}, \boldsymbol{y}_L\right)=\begin{cases} 1 & \text{if } \ln\left(\frac{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_1)}{f_{\boldsymbol{Y}_L, U_{L-1}|H}(\boldsymbol{y}_L, u_{L-1}|H_0)}\right) \geq \frac{C_F}{C_D}, \\ 0 & \text{otherwise,} \end{cases} \tag{5.40}$$

where

$$C_F = P(H_0)\ (C_{10} - C_{00}) > 0 \tag{5.41}$$

and

$$C_D = (1 - P(H_0))\ (C_{01} - C_{11}) > 0 \tag{5.42}$$

with C_{ik}, for $i, k \in \{0, 1\}$, denoting the cost of global decision being H_i when H_k is present. Therefore, as it happened under the Neyman-Pearson formulation, if we use (5.18) together with (5.9)-(5.11), we have that the conditional ROC curve of the last fusion stage given the preceding local decision, $\mathrm{ROC}_L(u_{L-1})$, matches the plot shown in Figure 5.2 with $P_D^{(u_{L-1})}(L|t_L(u_{L-1}))$ and $P_{FA}^{(u_{L-1})}(L|t_L(u_{L-1}))$ defined in (5.20) and (5.21) respectively. At this point and after having extended the results provided in Lemma 1, we can follow the proof of Theorem 3 in order to extend its results under the Bayesian set up. In this way, we are able to show that the necessary conditions for asymptotic detection derived in Theorem 3 still hold under the Bayesian set up.

Looking at the proof of Theorem 4 we can verify that the sufficient conditions for asymptotic perfect detection under the Neyman-Pearson formulation also hold under the Bayesian set up. In order to show this, we have to follow the same steps as in the proof of Theorem 4. In particular, assuming that the sufficient conditions occur, we will use the suboptimal fusion rules proposed in the proof of Theorem 4 in order to show that, given any arbitrary $\delta \in (0, 1)$, the probability of error at the last fusion node of the resulting tandem network is lower than an arbitrary $\delta \in (0, 1)$, i.e. $P_e(L) < \delta$. Consequently, in comparison with the proof followed in Theorem 4 under the Neyman-Pearson formulation, if we take into account (5.1), the only difference is that the parameter $\alpha_L \in (0, 1)$ has to be replaced by δ in (5.87).

Finally, we focus on the extension of Corollary 4 under the Bayesian set up. In this case note that the corresponding proof consists in showing that the conditions given in Theorem 3 and 4 occur if and only if the conditions provided in the Corollary hold. Toward this goal, recall that, under the Neyman-Pearson formulation, we employed properties of the ROC curves associated with LLRTs. Due to the fact that those properties are independent of the considered formulation, Bayes or Neyman-Pearson, it is clear that, once we have extended Theorems 3 and 4 under the Bayesian set up, we are at the position to automatically extend Corollary 4 under the formulation considered in this subsection.

5.4 Illustrative examples

In Section 5.3 we have provided necessary and sufficient conditions that allow us to determine if a tandem network can achieve asymptotic perfect detection under some specific observation model. Next, we are going to analyze several practical scenarios where the application of the aforementioned conditions is shown.

Initially we consider a hypothesis testing problem involving the distributed detection of a constant signal in correlated Gaussian noise. In this problem, each fusion node performs one observation of the environment according to the following observation model under each one of the two possible hypothesis

$$\begin{array}{ll} H_0: & y_j = v_j \\ H_1: & y_j = s_j + v_j \end{array} \tag{5.43}$$

with $j \in \{1, 2, \ldots, L\}$. For the previous observation model, the Gaussian noise process $\{v_j\}_{j=1}^{L}$ has zero mean and covariance function

$$\rho(i, j) = E\{v_i v_j\} = \sigma^2 \rho^{|i-j|} \tag{5.44}$$

with $i, j \in \{1, 2, \ldots, L\}$ and $0 \leq \rho < 1$. Additionally, the samples $\{s_j\}_{j=1}^{L}$ are drawn from an uniformly bounded deterministic signal, i.e., there exists a constant C_a such that $|s_j| \leq C_a$ for all $j \in \{1, 2, \ldots, L\}$. Hence, if $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the joint p.d.f. of the sensor observations under each one of the two hypothesis is given by

$$\begin{array}{ll} H_0: & \boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_v) \\ H_1: & \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{s}, \boldsymbol{\Sigma}_v) \end{array} \tag{5.45}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_L]^T$ and

$$\boldsymbol{\Sigma_v} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{L-1} \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho^{L-1} & \cdots & \cdots & 1 \end{bmatrix}. \tag{5.46}$$

Precisely, from those joint probability density functions we will be able to prove the subsequent corollary.

Corollary 5. *For both the Bayesian set up and the Neyman-Pearson formulation, an optimal tandem network can achieve asymptotic perfect detection when solving the hypothesis testing problem provided in* (5.45) *and* (5.46)*.*

Proof. As it can be expected from the theoretical results, for both the Bayesian set up and the Neyman-Pearson formulation, we have to check the same conditions in order to proof the corollary. Focusing on one of those conditions, we firstly have to show that, given any strategy Γ, for all $u_{j-1} \in \{0,1\}$ and all $j \in \{1,2,\ldots,L\}$ the conditional p.d.f. of $\Lambda\left(\boldsymbol{Y}_j|u_{j-1}\right)$ does not contain any point masses under both hypothesis, H_0 and H_1. For $j=1$ this can be easily seen since

$$\Lambda\left(Y_1|u_0\right) = \Lambda\left(Y_1\right) = Y_1\frac{s_1}{\sigma^2} + C_b \tag{5.47}$$

with

$$C_b = -\frac{s_1^2}{2\,\sigma^2}. \tag{5.48}$$

Given the marginal p.d.f. of Y_1 under H_0 and H_1, from (5.47) and (5.48) we can straightforwardly note that, for $j=1$, the p.d.f. of $\Lambda\left(Y_j|u_0\right)$ does not contain any point masses under both hypothesis. On the contrary, when we want to proof this for $j>1$, we have to undertake a more elaborated analysis.

Under any arbitrary but fixed strategy Γ as well as the observation model described in (5.45) and (5.46), we initially have that, for $u_{j-1}, \hat{k} \in \{0,1\}$ and $j \in \{2,3,\ldots,L\}$,

$$\begin{aligned}\Lambda\left(Y_j|u_{j-1}=\hat{k}\right) &= \Lambda\left(Y_j\right) - \log\left(P(U_{j-1}=\hat{k}|H_1)\right) \\ &\quad + \log\left(P(U_{j-1}=\hat{k}|y_j,H_1)\right) \\ &= Y_j\frac{s_j}{\sigma^2} + C_c + \log\left(P(U_{j-1}=\hat{k}|y_j,H_1)\right)\end{aligned} \tag{5.49}$$

where

$$C_c = C_b - \log\left(P(U_{j-1}=\hat{k}|H_1)\right) \tag{5.50}$$

and

$$P(U_{j-1}=\hat{k}|y_j,H_1)$$
$$=\begin{cases}\int_{\mathcal{Y}_{1,1}} P(U_1=\hat{k}|y_1)\, f_{Y_1|Y_2,H}(y_1|y_2,H_1)\mathrm{d}y_1 & \text{if } j=2\\ \int_{\mathcal{Y}_{j-1,1}} f_{Y_{j-1}|Y_j,H}(y_{j-1}|y_j,H_1) & \\ \quad\times\Big[P(U_{j-1}=\hat{k}|y_{j-1},u_{j-2}=1)P(U_{j-2}=1|\boldsymbol{y}_{j-1}^{j}) & \\ \quad+P(U_{j-1}=\hat{k}|y_{j-1},u_{j-2}=0)P(U_{j-2}=0|\boldsymbol{y}_{j-1}^{j})\Big]\,\mathrm{d}y_{j-1} & \text{if } j>2\end{cases} \tag{5.51}$$

with $\boldsymbol{y}_{j-1}^{j}=[y_{j-1},y_j]$. In the hypothesis testing problem expressed in (5.45) and (5.46) if we assume that, for $j\in\{3,4,\ldots,L\}$,

$$P(U_{j-1}=1|y_{j-1},u_{j-2}=1)\geq P(U_{j-1}=1|y_{j-1},u_{j-2}=0), \tag{5.52}$$

and that, for some $y_{j-1}\in\mathcal{Y}_{j-1,1}=(-\infty,\infty)$,

$$P(U_{j-1}=1|y_{j-1},u_{j-2}=1)<1 \tag{5.53}$$

and

$$P(U_{j-1}=1|y_{j-1},u_{j-2}=0)>0, \tag{5.54}$$

with $j\in\{2,3,\ldots,L\}$ and $u_0=\emptyset$, from (5.51) and after some algebraic manipulations we have that, for all $y_j\in\mathcal{Y}_{j,1}=(-\infty,\infty)$ and all $j\in\{2,3,\ldots,L\}$,

$$P(U_{j-1}=\hat{k}|y_j,H_1)\in(0,1) \tag{5.55}$$

with $\hat{k}=1$. Realize that the assumptions given in (5.52)-(5.54) are not restrictive under the considered observation model. On the one hand, as it is stated in [Ekchian and Tenney, 1982], the first of them given in (5.52) not only disambiguates two possible symmetric solutions when designing the optimal fusion rules of an optimal tandem network under a specific set up. It is also the natural way of modelling the effect of u_{j-2} on its successor, i.e. the $(j-1)-th$ fusion stage. On the other hand, if the assumptions provided in (5.53) and (5.54) were not satisfied, independently of y_{j-1} the $(j-1)$-th fusion stage would make the same binary decision as its predecessor. This kind of strategy would only make sense if

$$P_e(j-2)=0 \tag{5.56}$$

under Bayesian set up, or if, under the Neyman-Pearson formulation,

$$P_M(j-2) = 0 \tag{5.57}$$

for $P_{_{FA}}(j-2) \leq \alpha_L \in (0,1)$. Nonetheless, considering the observation model described in (5.45) and (5.46) this situation cannot happen under both formulations. Realize that, for the aforementioned observation model, a centralized system, whose detection performance is better than the one achieved by a tandem network, cannot achieve zero probability of error or zero probability of misdetection under the Bayesian set up or under the Neyman-Pearson formulation respectively.

Next, by noting that C_c is equal to a constant when u_{j-1} is equal to a specific value $\hat{k} \in \{0,1\}$, (5.49) and (5.55) entail that, under both hypothesis, the p.d.f. of $\Lambda(\boldsymbol{Y}_j|u_{j-1})$ has the same support as $\Lambda(\boldsymbol{Y}_j)$ with $j \in \{2,3,\ldots,L\}$. At the same time this last fact implies that, under both hypothesis and any $j \in \{2,3,\ldots,L\}$, the p.d.f. of $\Lambda(\boldsymbol{Y}_j|u_{j-1})$ can only contain point masses if and only if

$$P(U_{j-1} = \hat{k}|y_j, H_1) \propto e^{\left(-Y_j \frac{s_j}{\sigma^2}\right)} \tag{5.58}$$

for several values of $y_j \in (a,b)$ with $-\infty < a < b < \infty$. However, using (5.51) and taking into that, for all $j \in \{2,3,\ldots,L\}$,

$$f_{Y_{j-1}|Y_j,H}(y_{j-1}|y_j, H_1) = \mathcal{N}\left(s_{j-1} + \rho\,(y_j - s_j), \sigma^2\,(1-\rho^2)\right), \tag{5.59}$$

we can check that, given any strategy Γ,

$$P(U_{j-1} = \hat{k}|y_j, H_1) = \sum_i C_i Q(f_i(y_j)) \tag{5.60}$$

where $\{f_i(y_j)\}_{i=1}$ equals a sequence of an affine functions of y_j, $C_i \in \mathbb{Z}$ and

$$Q(x_0) = \frac{1}{\sqrt{(2\,\pi)}} \int_{x_0}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x \tag{5.61}$$

with $\mathbb{Z}$ denoting the set of integers. Due to the fact that the linear combination given in the r.h.s. of (5.60), cannot be proportional to the r.h.s. of (5.59), by contradiction we show that, for any $j \in \{2,3,\ldots,L\}$ and any $u_{j-1} \in \{0,1\}$, the p.d.f. of $\Lambda(\boldsymbol{Y}_j|u_{j-1})$ does not contain any point masses under H_0 and H_1.

At this point, we only need to check if the sufficient conditions of Corollary 4 are verified under the observation model characterized in (5.45) and (5.46). In particular, we are going to show that $\Lambda_{+}^{\inf}(1)$ is not lower bounded. To do so, from (5.47) and (5.48) we initially check that, under the observation model described by (5.45) and (5.46), the quantity $\Lambda_{1}^{\inf}(1)$ defined in (5.31) is not lower bounded. Consequently, recalling (5.34) we conclude the proof as long as we show that $\Lambda_{j}^{\inf}(1)$ is not lower bounded for all $j \in \{2, 3, \ldots, L\}$.

When showing the aforementioned conditions, from (5.49) we firstly have to realize that, under the observation model expressed in (5.45) and (5.46),

$$\Lambda\left(Y_j|\,1\right) \leq -\ln\left(P_D(j-1)\right) + \Lambda\left(Y_j\right) \tag{5.62}$$

with $j \in \{2, 3, \ldots, L\}$. Subsequently, note that (5.62) yields

$$\Lambda_{j}^{\inf}(1) \leq -\ln\left(P_D(j-1)\right) + \inf_{Y_j \in \mathcal{Y}_j^{(1)}} \left\{\Lambda\left(Y_j\right)\right\} = -\ln\left(P_D(j-1)\right) + \Lambda_{j}^{\inf} \tag{5.63}$$

with $\Lambda_{j}^{\inf}(1)$ defined in (5.31). At the same time, taking into account that the support of $f_{Y_j|U_{j-1},H}(y_j|1, H_0)$, $\mathcal{Y}_{j,0}^{(1)}$, equals $\mathbb{R}$ because

$$f_{Y_j|U_{j-1},H}(y_j|u_{j-1}, H_0) = f_{Y_j|H}(y_j|H_0) = \mathcal{N}(0, \sigma^2) \tag{5.64}$$

is verified when $u_{j-1} \in \{0, 1\}$, from the evaluation of (5.32) at $u_{j-1} = 1$ we also have that

$$\mathcal{Y}_{j}^{(1)} = \mathcal{Y}_{j,0}^{(1)} \cup \mathcal{Y}_{j,1}^{(1)} = \mathbb{R} \tag{5.65}$$

for all $j \in \{2, \ldots, L\}$ and independently of the support of $f_{Y_j|U_{j-1},H}(y_j|1, H_1)$, $\mathcal{Y}_{j,1}^{(1)}$. Hence, due to the fact that $P_D(j-1) > 0$ when (5.55) holds for all $j \in \{2, \ldots, L\}$, and that

$$\Lambda_{j}^{\inf} = \inf_{Y_j \in \mathcal{Y}_j^{(1)}} \left\{\Lambda\left(Y_j\right)\right\} \tag{5.66}$$

is not lower bounded when (5.65) occurs, (5.63) implies that $\Lambda_{j}^{\inf}(1)$ is not lower bounded. Equivalently, if we recall the proof where we have already shown that $\Lambda_{1}^{\inf}(1)$ is not lower bounded, the previous statement means that, under the observation model described in (5.45) and (5.46), the quantity $\Lambda_{+}^{\inf}(1)$, defined in (5.34), is not lower bounded. In this way the proof is concluded.

□

In the previous example we have provided a detection scenario where the sufficient conditions of Corollary 4 occur. Next, we analyze a scenario where the necessary conditions to achieve asymptotic perfection do not hold. Specifically, in the next illustrative example we consider a tandem network solving this hypothesis testing problem under the Bayesian set up or under the Neyman-Pearson formulation

$$H_k: \ \{y_j\}_{j=1}^{L} \text{are samples drawn from a first-order regular Markov chain, } C_k, \text{ with } M\text{-ary state space and transition probability matrix } \mathbf{\Pi}_k. \tag{5.67}$$

where, for $k \in \{0,1\}$,

$$\mathbf{\Pi}_k(m,n) = P(y_j = n | y_{j-1} = m, H_k) \in (0,1) \tag{5.68}$$

for all $m,n \in \{1,2,\ldots,M\}$, $k \in \{0,1\}$ and $j \in \{2,3,\ldots,L\}$. Under this observation model using the inequalities

$$P_{Y_j|U_{j-1},H}(n|u_{j-1}H_k) \geq \min_{m=1,2,\ldots,M} \left\{ P_{Y_j|Y_{j-1},H}(n|m,H_k) \right\} \tag{5.69}$$

and

$$P_{Y_j|U_{j-1},H}(n|u_{j-1}H_k) \leq \max_{m=1,2,\ldots,M} \left\{ P_{Y_j|Y_{j-1},H}(n|m,H_k) \right\} \tag{5.70}$$

we can easily see that

$$P_{Y_j|U_{j-1},H}(n|u_{j-1}H_k) \in (0,1) \tag{5.71}$$

for all $n \in \{1,2,\ldots,M\}$, $u_{j-1} \in \{0,1\}$ and $j \in \{2,3,\ldots,L\}$. Recalling the optimal fusion rule applied by the last fusion stage of a tandem network performing a Bayesian or a Neyman-Pearson test (see [Xiang and Wang, 2006] and [Yan and Blum, 2001]), (5.71) implies that, for both formulations, $m_{0,L}^{(0)} < \infty$ and $m_{1,L}^{(1)} > 0$. If we now realize that these inequalities cannot happen when the necessary conditions for asymptotic perfect detection hold, we automatically prove the following corollary.

Corollary 6. *Under both the Bayesian set up and the Neyman-Pearson formulation, an optimal tandem network can never achieve asymptotic perfect detection when discriminating two first-order Markov chains defined on a M-ary state space and whose transition probability matrices only have strictly positive elements.*

The last illustrative example consists in the particularization of the previous scenario for $M = 2$. With this simplified setting we are able to show that, under both the Bayesian set up and the Neyman-Pearson formulation, a tandem network could only achieve any desired level of performance if the Markov chains present under both hypothesis are different and if at least one of them is absorbing. This way, as it is stated in the subsequent corollary, we generalize the result of Corollary 6 to all possible kinds of regular first-order Markov chains defined on a binary state space.

Corollary 7. *Under both the Bayesian set up and the Neyman-Pearson formulation, an optimal tandem network can never achieve asymptotic perfect detection when discriminating two first-order regular Markov chains defined on a binary state space.*

Proof. See Appendix 5.B. □

Finally, we will complement the analysis of this last example by considering a scenario where the Markov chain present under one of the two hypothesis is not regular. In particular, we will assume that the Markov chain present under H_1, C_1, is regular and has a transition probability matrix characterized by

$$\mathbf{\Pi}_1(m, n) \in (0, 1) \tag{5.72}$$

for all $m, n \in \{0, 1\}$. Additionally, we will consider that the Markov chain associated with H_0, C_0 is absorbing with

$$\mathbf{\Pi}_0(0, 1) = 0 \tag{5.73}$$

and

$$\mathbf{\Pi}_0(1, 0) \in (0, 1]. \tag{5.74}$$

Assuming that the Markov chains, C_0 and C_1, are in their stationary regime when they are observed by the tandem network, as a consequence of (5.73), (5.74) and the subsequent expression

$$P(Y_j = 1|H_0) = 1 - P(Y_j = 0|H_0) = \frac{\mathbf{\Pi}_0(0, 1)}{\mathbf{\Pi}_0(1, 0) + \mathbf{\Pi}_0(0, 1)} = 0 \tag{5.75}$$

with $j \in \{1, 2, \ldots, L\}$, we have that, for all $j \in \{1, 2, \ldots, L\}$,

$$P_{Y_j|U_{j-1},H}(1|0, H_0) = (1 - \mathbf{\Pi}_0(1, 0))\, P_{Y_{j-1}|U_{j-1},H}(1|0, H_0) = 0. \tag{5.76}$$

Next, if we take into account that the last decision rule is given by (5.40) and (5.19) under the Bayesian set up and the Neyman-Pearson formulation respectively, from (5.9) and (5.18) we can see that (5.76), evaluated at $j = L$, ensures $m_{0,L}^{(0)} = \infty$ under the observation model resulting from the particularization of (5.67) for $M = 2$. It should be noted that this result is consistent if we take into account that, in a tandem network whose observation model matches the one expressed in (5.72)-(5.74), there exists a straightforward strategy achieving perfect detection under both the Bayesian set up and the Neyman-Pearson formulation.

5.5 Summary

In this chapter, we addressed the analysis of detection performance of a distributed detection system with serial architecture. In order to solve a distributed detection problem where the sensor observations might be dependent under each one of the two possible hypothesis, each fusion stage applies a binary local decision rule. For the previously described scenario and under the Neyman-Pearson formulation, we initially derived necessary and sufficient conditions for asymptotic perfect learning. In other words, we were able to obtain necessary and sufficient conditions ensuring that, for a specific constraint on the probability of false alarm of the system, the corresponding probability of misdetection go to zero as the number of fusion stages increases. Later, under the same scenario these conditions were also extended to optimal tandem networks performing a Bayesian hypothesis test on conditionally dependent observations. Finally, we provided some illustrative examples where the application of the previous necessary and sufficient conditions is shown.

5.A Proof of Theorem 4

Firstly choose some large value η^* and consider a tandem network whose fusion stages apply the following decision rules for some integer L^*. At the first fusion node

$$u_1 = \gamma_1(\boldsymbol{y}_1) = \begin{cases} 0 & \text{if } \Lambda(\boldsymbol{y}_1) \leq \log(\eta^*), \\ 1 & \text{if } \Lambda(\boldsymbol{y}_1) > \log(\eta^*) \end{cases} \tag{5.77}$$

with

$$\Lambda(\boldsymbol{y}_1) = \ln\left(\frac{f_{\boldsymbol{Y}_1|H}(\boldsymbol{y}_1|H_1)}{f_{\boldsymbol{Y}_1|H}(\boldsymbol{y}_1|H_0)}\right). \tag{5.78}$$

For the fusion units whose index $j \in \{2, 3, \ldots, L^*\}$

$$u_j = \gamma_j(\boldsymbol{y}_j, u_{j-1}) = \begin{cases} 0 & \text{if } \Lambda(\boldsymbol{y}_j|0) \leq \log(\eta^*) \text{ and } u_{j-1} = 0, \\ 1 & \text{otherwise,} \end{cases} \tag{5.79}$$

with $\Lambda(\boldsymbol{y}_j|0)$ equal to the evaluation of $\Lambda(\boldsymbol{Y}_j|u_{j-1})$ at $\boldsymbol{Y}_j = \boldsymbol{y}_j$ and $u_{j-1} = 0$. Finally, for the fusion stages with index $j \in \{L^* + 1, L^* + 2, \ldots\}$

$$u_j = \gamma_j(\boldsymbol{y}_j, u_{j-1}) = u_{j-1}. \tag{5.80}$$

According to the previous strategy, only the first L^* fusion stages influence the final decision of the team. In particular, a decision $u_j = 0$, with $j \in \{1, 2, \ldots, L^*\}$, is propagated through the following $L - j$ fusion stages of the team until some fusion node j^*, with $j^* \in \{j+1, j+2, \ldots, L^*\}$, performs an observation $\boldsymbol{y}_{j^*}$ which makes it decide $u_{j^*} = 1$. In that case, the team decision decision will be $u_L = 1$; otherwise the team decision will be $u_L = 0$. To put it differently, under the proposed strategy, the final decision is only H_0 when $u_1 = u_2 = \ldots = u_{L^*} = 0$. Consequently, using the described strategy the tandem network has the following probability of false alarm

$$P_{_{FA}}(L) = 1 - P(U_L = 0|H_0) = 1 - \prod_{j=1}^{L^*}(1 - P_{_{FA}}^{(0)}(j|\eta^*)) \tag{5.81}$$

and the subsequent probability of detection

$$P_D(L) = 1 - P(U_L = 0|H_1) = 1 - \prod_{j=1}^{L^*}(1 - P_D^{(0)}(j|\eta^*)) \tag{5.82}$$

where, for all $j \in \{1, \ldots, L^*\}$,

$$P_{_{FA}}^{(0)}(j|\eta^*) = \begin{cases} P(U_1 = 1|\eta^*, H_0) & \text{if } j = 1 \\ P(U_j = 1|U_{j-1} = 0, \eta^*, H_0) & \text{if } j \neq 1 \end{cases} \tag{5.83}$$

and

$$P_D^{(0)}(j|\eta^*) = \begin{cases} P(U_1 = 1|\eta^*, H_1) & \text{if } j = 1 \\ P(U_j = 1|U_{j-1} = 0, \eta^*, H_1) & \text{if } j \neq 1 \end{cases} \tag{5.84}$$

denote the operating point for the j-th fusion stage.

From basic detection theory provided in [Van Trees, 2001, Section 2.2.2], we can verify that, as long as $m_{0,j}^{(0)} = \infty$ and the p.d.f. of $\Lambda\left(\boldsymbol{Y}_j|u_{j-1}\right)$ contains no point masses for all $j \in \{1, 2, \ldots, L\}$, the operating points

$$\left\{\left(P_{FA}^{(0)}(j|\eta^*), P_D^{(0)}(j|\eta^*)\right)\right\}_{j=1}^{L^*} \tag{5.85}$$

can always be obtained regardless of how large the value η^* is. At the same time we can note that the probabilities defining the operating points are very close to zero since η^* is chosen sufficiently large. Actually, in the considered strategy the probability $P_{FA}^{(0)}(j|\eta^*)$ or $P_D^{(0)}(j|\eta^*)$, with $j \in \{2, 3, \ldots, L^*\}$, is only equal to zero when the observations $\boldsymbol{Y}_{j-1}$ and $\boldsymbol{Y}_j$ are maximally dependent under H_0 or H_1 according to Definition 1 respectively. In that case, as it can be easily checked from (5.81) and (5.82), the corresponding fusion stage does not have effect on the detection performance of the network. Consequently, ignoring that kind of fusion stages and assuming that the number of fusion nodes performing non-maximally dependent observations under both hypothesis is strictly increasing w.r.t. L, without loss of generality we can consider an equivalent network where $P_{FA}^{(0)}(j|\eta^*)$, $P_D^{(0)}(j|\eta^*) > 0$.

Next, from (5.81) and (5.82) check that the integer L^* has to satisfy an upper and a lower bound in order to achieve

$$P_M(L) < \delta \in (0, 1) \tag{5.86}$$

for any upper bound

$$P_{FA}(L) \leq \alpha_L \in (0, 1). \tag{5.87}$$

On the one hand, if we take into account that for all $j \in \{1, 2, \ldots, L^*\}$

$$\frac{P_D^{(0)}(j|\eta^*)}{P_{FA}^{(0)}(j|\eta^*)} > \eta^* \tag{5.88}$$

holds because conditional ROC curves, $\mathrm{ROC}_j(0)$, are concave downward when $\Lambda\left(\boldsymbol{Y}_j|u_{j-1}\right)$ does not contain any point masses of probability under either hypothesis, from (5.83) we can check that (5.87) occurs as long as

$$\sum_{j=1}^{L^*} \log\left(1 - \frac{P_D^{(0)}(j|\eta^*)}{\eta^*}\right) \geq \log(1 - \alpha_L). \tag{5.89}$$

On the other hand, knowing that $P_M(L) = 1 - P_D(L)$, (5.84) and (5.86) yield

$$\sum_{j=1}^{L^*} \log\left(1 - P_D^{(0)}(j|\eta^*)\right) \leq \log(\delta). \tag{5.90}$$

Thus, for any of the constraints provided in (5.87) the probability of misdetection of the system, $P_M(L)$, is less than an arbitrary $\delta \in (0,1)$ as long as an integer L^* satisfying (5.89) and (5.90) indeed exists.

Finally, we conclude the proof by showing the existence of an integer L^* satisfying both (5.89) and (5.90). Toward this goal, recall that

$$\log(1+x) \approx x \tag{5.91}$$

for $x \approx 0$, and that $P_D^{(0)}(j_{\min}|\eta^*)$ is very closed to zero and greater than zero if η^* is chosen sufficiently large in the equivalent tandem network where there are no fusion stages performing maximally dependent observations. Therefore, considering that $\alpha_L > \delta$ without loss of generality, the inequalities provided in (5.89) and (5.90) hold as long as

$$\frac{-\log(1-\delta)\,\eta^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} \geq 1 \tag{5.92}$$

and

$$\frac{-\log(\delta)}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} \leq 1. \tag{5.93}$$

If we multiply by L^* both sides of the two previous inequalities, we can easily check that they are satisfied when the subsequent expression occurs

$$\left\lfloor \frac{-\log(\delta)\,L^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} \right\rfloor \geq \; L^* \geq \left\lceil \frac{-\log(1-\delta)\,\eta^*\,L^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} \right\rceil \tag{5.94}$$

with $\lfloor x \rfloor$ equal to the integer part of x and $\lceil x \rceil$ denoting the smallest integer greater than or equal to x. Now, check that if

$$\eta^* \geq \frac{\log(\delta)}{\log(1-\delta)} - \frac{2}{\log(1-\delta)} \tag{5.95}$$

we have that

$$\begin{aligned}\frac{-\log(1-\delta)\,\eta^*\,L^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} &> \frac{-\log(\delta)\,L^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} + \frac{2\,L^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} \\ &\geq \frac{-\log(\delta)\,L^*}{\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*)} + 2\end{aligned} \tag{5.96}$$

where the last inequality follows since $\sum_{j=1}^{L^*} P_D^{(0)}(j|\eta^*) \leq L^*$. Consequently, from (5.96) it is straightforward to show that choosing η^* sufficiently large, which can be done because the $m_{0,j}^{(0)}$ is assumed to be infinite for all $j \in \{1, 2, \ldots, L^*\}$, an integer satisfying both (5.89) and (5.90) can always be obtained. At the same time, knowing that the detection performance of the proposed strategy is worse than or equal to the detection performance of the optimal strategy, we show that, for any of the constraints given in (5.89), (5.90) also occurs in an optimal tandem network performing a Neyman-Pearson test on dependent observations. In other words, for any arbitrary constraint on the probability of false alarm of an infinite tandem team, the optimal probability of misdetection always goes to zero as long as $m_{0,j}^{(0)} = \infty$ and the p.d.f. of $\Lambda\left(\boldsymbol{Y}_j|u_{j-1}\right)$ contains no point masses under either hypothes and for all $j \in \{1, 2, \ldots, L^*\}$. This concludes the proof.

5.B Proof of Corollary 7

Taking into account the statement of Corollary 6 we need to show that $m_{0,L}^{(0)} < \infty$ holds in two specific settings. On the one hand, in one of these settings

$$\boldsymbol{\Pi}_0(0,1) = 1, \tag{5.97}$$

$$\boldsymbol{\Pi}_0(1,0) \in (0,1) \tag{5.98}$$

and the Markov chain present under H_1, C_1, is not absorbing, i.e.

$$\boldsymbol{\Pi}_1(0,1), \boldsymbol{\Pi}_1(1,0) > 0. \tag{5.99}$$

On the other hand, the second scenario is described by (5.99) as well as the subsequent constraints

$$\boldsymbol{\Pi}_0(1,0) = 1, \tag{5.100}$$

$$\boldsymbol{\Pi}_0(0,1) \in (0,1]. \tag{5.101}$$

In order to prove that the necessary conditions derived in Section 5.3 do not hold when the Markov chains, C_0 and C_1, are not absorbing we should also analyze that $m_{1,L}^{(1)} > 0$ is verified under two additional sets of constraints. However, this proof has been omitted since it is analogous to the one required when showing that $m_{0,L}^{(0)} < \infty$ holds under the two sets considered in this appendix.

To show that $m_{0,L}^{(0)} < \infty$ holds in the setting characterized by (5.97)-(5.99) we can employ the subsequent expression

$$\begin{aligned} & P_{Y_j|U_{j-1},H}(1|u_{j-1},H_k) \\ & \quad = \sum_{y_{j-1}=0}^{1} P_{Y_j|Y_{j-1},U_{j-1},H}(1|y_{j-1},u_{j-1},H_k) P_{Y_{j-1}|U_{j-1},H}(y_{j-1}|u_{j-1},H_k) \\ & \quad = \mathbf{\Pi}_k(0,1) + \rho_k \, P_{Y_{j-1}|U_{j-1},H}(1|u_{j-1},H_k) \\ & \quad = 1 - \mathbf{\Pi}_k(1,0) - \rho_k \, P_{Y_{j-1}|U_{j-1},H}(0|u_{j-1},H_k), \end{aligned} \tag{5.102}$$

with $k \in \{0,1\}$ and

$$\rho_k = 1 - \mathbf{\Pi}_k(1,0) - \mathbf{\Pi}_k(0,1). \tag{5.103}$$

From (5.97), (5.98) and these last two expressions we can easily realize that $P_{Y_j|U_{j-1},H}(1|u_{j-1},H_0) = 0$ if

$$P_{Y_{j-1}|U_{j-1},H}(1|u_{j-1},H_k) = \frac{1}{\mathbf{\Pi}_k(1,0)} > 1 \tag{5.104}$$

with $u_{j-1} \in \{0,1\}$ and $j \in \{2,\ldots,L\}$. However, this is not possible since $P_{Y_{j-1}|U_{j-1},H}(y_{j-1}|u_{j-1},H_k) \leq 1$ by definition. Therefore, if we take into account that the last decision rule is given by (5.40) and (5.19) under the Bayesian set up and the Neyman-Pearson formulation respectively, from (5.9) and (5.18) we can see that $m_{0,L}^{(0)} < \infty$ does not hold as long as (5.97)-(5.99) occur.

Next, if we recall again the form of the optimal fusion rule applied by the last fusion stage of a tandem network performing a Bayes test or a Neyman-Pearson test, from (5.102) we can check that, for both formulations and under the set of constraints given in (5.99)-(5.101), $m_{0,L}^{(0)} = \infty$ holds only if

$$P_{Y_{L-1}|U_{L-1},H}(1|0,H_0) = \mathbf{\Pi}_0(0,1) \, P_{Y_{L-1}|U_{L-1},H}(0|0,H_0) = 0. \tag{5.105}$$

Due to the fact that, under the considered observation model, $\mathbf{\Pi}_0(0,1) > 0$, the previous expression results into

$$P_{Y_{L-1}|U_{L-1},H}(0|0,H_0) = 0. \tag{5.106}$$

Subsequently, we are going to analyze the implications associated with this last constraint when the Markov chains present under both hypothesis are in their stationary regime, i.e.

$$P(Y_j = 0|H_k) = 1 - P(Y_j = 0|H_k) = \frac{\mathbf{\Pi}_k(1,0)}{\mathbf{\Pi}_k(1,0) + \mathbf{\Pi}_k(0,1)} \tag{5.107}$$

for all $j \in \{1,\ldots,L\}$ and $k \in \{0,1\}$. In particular, if we initially consider that $P_{_{FA}}(L-1) < 1$, using (5.107) we can realize that, under the set of constraints given in (5.99)-(5.101),

$$\begin{aligned} P_{Y_{L-1}|U_{L-1},H}(0|0,H_0) &= \frac{P_{U_{L-1}|Y_{L-1},H}(0|0,H_0)}{1 - P_{FA}(L-1)} P(Y_{L-1} = 0|H_0) \\ &= \frac{P_{U_{L-1}|Y_{L-1},H}(0|0,H_0)}{(1 - P_{FA}(L-1))\,(1 + \mathbf{\Pi}_0(0,1))} \end{aligned} \tag{5.108}$$

This last expression together with (5.106) implies that

$$P_{U_{L-1}|Y_{L-1},H}(0|0,H_0) = 0. \tag{5.109}$$

At this point, if we recall the monotonicity property associated with fusion rules whose inputs are binary, we can verify that (5.109) yields $P_{_{FA}}(L-1) = 1$. Nonetheless, note that this is contradictory with the initial assumption $P_{_{FA}}(L-1) < 1$. Consequently, under the considered observation model and under the constraints given in (5.99)-(5.101), (5.105) cannot happen if $P_{_{FA}}(L-1) > 0$. On the contrary, if we assume that $P_{_{FA}}(L-1) = 1$, from

$$\begin{aligned} P_D(L) =& P_D(L-1) \cdot \left(1 - F_L^{(1)}\left(t_L(1)|H_1\right)\right) \\ &+ (1 - P_D(L-1)) \cdot \left(F_L^{(0)}\left(t_L(0)|H_1\right)\right) \end{aligned} \tag{5.110}$$

and

$$\begin{aligned} P_{_{FA}}(L) =& P_{_{FA}}(L-1) \cdot \left(1 - F_L^{(1)}\left(t_L(1)|H_0\right)\right) \\ &+ (1 - P_{_{FA}}(L-1)) \cdot \left(1 - F_L^{(0)}\left(t_L(0)|H_0\right)\right) \end{aligned} \tag{5.111}$$

we can straightforwardly check that, even in the most optimistic situation where $P_D(L-1)=1$, when (5.99) holds the last fusion stage cannot perform binary fusion rules achieving any desired level of performance under the Bayesian set up or under the Neyman-Pearson formulation. In this way, we show that $m_{0,L}^{(0)} < \infty$ as long as the constraints (5.99)-(5.101) occur.

Finally, we have to remark that all the constraints expressed in (5.97), (5.98), (5.100) and (5.101) parametrize the set of all the possible non-absorbing Markov chains, C_0, present under H_0. At this point, if we take into account that (5.99) represents all the possible non-absorbing Markov chains, C_1, present under H_1, this means that $m_{0,L}^{(0)} < \infty$ and $m_{1,L}^{(1)} > 0$ hold as long as none of the Markov chains, C_0 and C_1, is absorbing. To put it differently, we have that, under both the Bayesian set up and the Neyman-Pearson formulation, a tandem network can only achieve any desired level of performance if at least one of the binary Markov chains present under one of the two hypothesis is absorbing.

Chapter 6

Conclusions and future research lines

In this dissertation we faced the design and analysis of distributed detection systems with conditionally dependent sensor observations and binary quantization rules at the sensors. In particular, under the Neyman-Pearson formulation, we characterized the overall detection performance as a function of different physical and design parameters associated with parallel and tandem networks where the local processing scheme provides a binary summary of the sensor observations.

Initially, we undertook the aforementioned characterization for parallel distributed detection systems formed by a large number of binary local detectors as well as a data fusion center. In this kind of system, based on the observations of each device, the binary local decisions are taken with no kind of cooperation and they are transmitted to the fusion center over an error free bank of parallel access channels. Additionally, we also considered that the devices are randomly deployed along a straight line and that the corresponding sensor spacings are drawn independently from a common p.d.f., $f_D(d)$. At the same time, we assumed that, under both hypothesis H_0 and H_1, the dependence among the local decisions is modelled by means of a 1-D MRF with nearest-neighbour dependency and binary state space. Under this scenario, we firstly derived a closed-form error exponent for the Neyman-Pearson test performed at the fusion center when this one only knows the p.d.f. of the sensor spacings. After studying the analytical properties and evaluating the derived error exponent for a specific physical model associated with the conditional probabilities of the assumed MRF,

we were able to study the dependency of the overall detection performance w.r.t. different parameters of networks deployed with a different p.d.f. of the sensor spacings, $f_D(d)$. Regarding the discrimination of two different 1-D Markov random fields defined on a binary state space and where the involved random variables are not maximally dependent, the result of the previous work yielded the subsequent conclusions for the considered network:

- When the binary local decisions are independent under H_0 and H_1, the derived error exponent is always consistent with the Neyman-Pearson error exponent stated by Stein's Lemma.

- Independently of $f_D(d)$ and for any arbitrary constraint $\alpha \in (0, 1)$ on the probability of false alarm of the system, the corresponding probability of misdetection can be arbitrarily low as the number of observations increases.

- Independently of the sensor spacing and the physical model followed by the conditional probabilities of the MRF, the maximum discrimination between the null and the alternative hypothesis is not always achieved when the sensor observations are conditionally independent under both hypothesis. In particular, the dependence among the local decisions might be more informative than the conditional independence. Actually, when a correlation structure is only present under one of the two hypothesis, there will always be a range of mean correlation strengths under which the detection performance of the network is better than the one associated with conditionally independent sensor observations. On the contrary, the existence of the aforementioned range of values of the mean correlation strength cannot be ensured in a scenario where, under both hypothesis, the local decisions are dependent and are distributed according to a 1-D MRF. We showed that this occurs since different trends of detection performance can appear with increasing dependency if we assume different sets of physical and design parameters of the network and if the local decisions are dependent under H_0 and H_1.

- Under exponentially spaced sensors with failures, given a specific mean correlation strength, the error exponent, K, for the Neyman-Pearson fusion of the Markov local decisions remains unaltered although the failure rate, q, of the devices varies. Consequently, there might be

detection scenarios under which the error exponent, and therefore, the asymptotic detection performance of the system, is completely determined from the mean correlation strength between neighbour local decisions. However, this does not hold for all the probability density functions of the sensor spacings as well as all the scenarios of dependence among the binary local decisions. Nevertheless, for the cases where this is verified, the error exponent for the fusion rule can be the same for two different sets of physical and design parameters of the network that result in the same mean correlation strength among neighbour nodes. Consequently, under this kind of scenarios, important savings in the consumption of network resources can be made. Toward this goal, among the sets of network parameters that achieve a required level of asymptotic detection performance with a specific mean correlation strength, it suffices to choose the one minimizing the corresponding network resources.

Later, we characterized the performance associated with a 2-D version of the previous distributed detection system. However, this time the devices are located on a rectangular lattice so that sensors belonging to a specific row or column are equally spaced. Additionally, for each hypothesis H_0 and H_1, the correlation structure of the local decisions is modelled with a 2-D causal field where the rows and columns are outcomes of the same first-order binary Markov chain. Similarly to the 1-D setting, we derived, analyzed and evaluated a closed-form error exponent for the Neyman-Pearson fusion of the local decisions. As a result of those tasks, the characterization of the detection performance associated with the previously described 1-D network, and therefore, the previous results, are generalized to a 2-D scenario where the devices are deterministically deployed,

Finally, we addressed the analysis of the detection performance of a serial detection system formed by L fusion nodes. In order to solve a hypothesis testing problem where the sensor observations might be dependent under each one of the two possible hypothesis, the tandem network applies a strategy based on binary decision rules at each one of the fusion stages. Specifically, the binary local decision rule associated with each fusion stage only depends on the result of fusing its own measurement, the measurements received from the sensor nodes that directly communicate with it over error free parallel access channels, and the local decision taken by the preceding fusion node. Under the Neyman-Pearson strategy we were able to obtain

necessary and sufficient conditions ensuring that, for a specific constraint on the probability of false alarm of the system, the corresponding probability of misdetection goes to zero as the number of fusion stages increases. Later, under the same scenario these conditions were also extended to optimal tandem networks performing a Bayesian hypothesis test on conditionally dependent observation. If $\Lambda(\boldsymbol{Y}_j|\, u_{j-1})$ denotes the likelihood ratio of the sensor observations performed by the j-th fusion stage when the preceding fusion node decides $u_{j-1} \in \{0,1\}$ with $j \in \{1,2,\ldots,L\}$ and $u_0 = \emptyset$, for the considered tandem network, under both the Bayesian and the Neyman-Pearson set ups, the analysis of the previous conditions as well as some illustrative examples yield the following conclusions:

- The considered tandem network cannot achieve asymptotic perfect detection if $\Lambda(\boldsymbol{Y}_L|\, 0)$ is upper bounded and $\Lambda(\boldsymbol{Y}_L|\, 1)$ is lower bounded.

- The addition of extra fusion stages can make the considered tandem network achieve any desired level of performance if, for all $j \in \{1,2,\ldots,L\}$, either $\Lambda(\boldsymbol{Y}_j|\, 0)$ is not upper bounded or either $\Lambda(\boldsymbol{Y}_j|\, 1)$ is not lower bounded.

- The considered tandem network can achieve any desired level of performance when detecting a deterministic signal in correlated Gaussian noise.

- Asymptotic perfect detection cannot be achieved when the considered tandem network wants to discriminate between two regular first-order Markov chains satisfying some conditions. Specifically, if the Markov chains associated with each hypothesis are defined on a M-ary state space, the previous result holds when the corresponding transition probability matrices only have strictly positive elements. In the case that they were defined on binary state space, this result can be applied to a detection problem where the Markov chains are regular under H_0 and H_1.

Next, we give a description of the possible future research lines for the problems addressed in this thesis dissertation. Regarding the characterization of the detection performance associated with two-stage parallel distributed systems, an interesting future research direction would be

the derivation of the Neyman-Pearson error exponent when the devices are randomly located over a 2-D region. Furthermore, during this dissertation we have assumed that the local decisions are noiselessly transmitted to the fusion center. However, for both the 1-D and the 2-D deployments, an appealing extension would be the characterization of the error exponent when the data fusion center receives noisy observations of the binary quantizations undertaken at the sensors. Moreover, it would be of great value to extend all the previous research lines when the data fusion center performs a Bayes hypothesis test.

As it happens for parallel networks with dependent observations and binary quantization rules at the sensors, there is a big number of possible future research lines when a tandem architecture is considered. Concluding this section we will only mention a couple of them. For instance, for the tandem network considered in the last part of this dissertation, an interesting research line would be the study of conditions ensuring that Neyman-Pearson tests at the fusion nodes maximize the overall probability of detection s.t. a fixed constraint on the corresponding probability of false alarm. Some results associated with this problem have already been published in [Plata-Chaves et al., 2011b]. Additionally, in order to provide more design guidelines for this kind of networks, an appealing extension of [Tay et al., 2008b] would be the derivation of upper bounds on the best rate of decay that, for an arbitrary upper bound on the overall probability of false alarm, the corresponding probability of misdetection can achieve when the aforementioned tandem network implements a myopic strategy under the Neyman-Pearson formulation. The results of that task, which have being gathered in [Plata-Chaves et al., 2012a], allow us to show that the rate of decay for the overall probability of misdetection is always subexponential although the correlation structure of the sensor observations yields an exponential rate of decay under a parallel architecture. At the same time, besides obtaining some design guidelines, from those bounds we will also be able to quantify the performance of the network in terms of the best subexponential rate of decay that the overall probability of misdetection can achieve. Likewise, it would be of great interest to extend the previously described future research lines as well as the results presented along this thesis to tree architectures and/or a scenario where, for $M > 2$, the fusion nodes perform M-ary quantizations of the available sensor observations.

Appendix A

Acronyms and abbrevations

- 1-D: one-dimensional.
- 2-D: two-dimensional.
- AWGN: additive white Gaussian noise.
- CGF: cumulant generating function.
- GLRT: generalized likelihood ratio test.
- GMRF: Gauss-Markov random field.
- i.i.d.: independent and identically distributed.
- l.h.s.: left-hand side.
- LLRT: log-likelihood ratio test.
- MGF: moment generating function.
- MRF: Markov random field.
- PBPO: person-by-person optimization.
- p.d.f.: probability density function.
- p.m.f.: probability mass function.
- r.h.s.: right-hand side.

- ROC: receiving operating curve.
- SPRT: sequential probability ratio test.
- s.t.: subject to.
- UMP: uniformly most powerful.
- w.r.t.: with respect to.

List of Figures

List of Tables

Bibliography

[Aldosari and Moura, 2007] Aldosari, S. A. and Moura, J. M. F. (2007). Detection in sensor networks: The saddlepoint approximation. *IEEE Transactions on Signal Processing*, 55(1):327–340.

[Anandkumar et al., 2009] Anandkumar, A., Tong, L., and Swami, A. (2009). Detection of Gauss–Markov random fields with nearest-neighbor dependency. *IEEE Transactions on Information Theory*, 55(2):816–827.

[Appadwedula et al., 2005] Appadwedula, S., Veeravalli, V. V., and Jones, D. L. (2005). Energy-efficient detection in sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):693–702.

[Artés-Rodríguez and Lázaro, 2007] Artés-Rodríguez, A. and Lázaro, M. (2007). Optimal sensor selection in heterogeneous sensor networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, pages 1017–1020.

[Artés-Rodríguez et al., 2005] Artés-Rodríguez, A., Lázaro, M., and Sánchez-Fernández, M. (2005). Decentralized detection in sensor networks with censored transmissions. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2005. ICASSP 2005*, pages 817–820.

[Artés-Rodríguez et al., 2004] Artés-Rodríguez, A., Torres-López, R., and Tong, L. (2004). Distributed detection by multiple tests in sensor networks using range information. In *12th European Signal Conference, 2010. EUSIPCO 2004*, pages 797–800.

[Artés-Rodríguez, 2004] Artés-Rodríguez, A. (2004). Decentralized detection in sensor networks using range information. In *IEEE International*

Conference on Acoustics, Speech and Signal Processing, 2004. ICASSP 2004, pages 265–268.

[Blahut, 1987] Blahut, R. E. (1987). *Principles and practice of information theory*. Addison-Wesley, Inc. Boston, MA, USA.

[Blum, 1996] Blum, R. S. (1996). Necessary conditions for optimum distributed sensor detectors under the Neyman–Pearson criterion. *IEEE Transactions on Information Theory*, 42(3):990–994.

[Blum et al., 1997] Blum, R. S., Kassam, S. A., and Poor, H. V. (1997). Distributed detection with multiple sensors: Part II – Advanced topics. *Proceedings of the IEEE*, 85(1):64–79.

[Booth et al., 1994] Booth, J. G., Hall, P., and Wood, A. T. A. (1994). On the validity of edgeworth and saddlepoint approximations. *Journal of Multivariate Analysis*, 51(1):121–138.

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

[Chamberland and Veeravalli, 2004] Chamberland, J. F. and Veeravalli, V. V. (2004). Asymptotic results for decentralized detection in power constrained wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6):1007–1015.

[Chamberland and Veeravalli, 2006] Chamberland, J. F. and Veeravalli, V. V. (2006). How dense should a sensor network be for decentralized detection with correlated observations. *IEEE Transactions on Information Theory*, 52(11):5099–5106.

[Champagnat et al., 1998] Champagnat, F., Idier, J., and Goussard, Y. (1998). Stationary Markov random fields on a finite rectangular lattice. *IEEE Transactions on Information Theory*, 44(7):2901–2916.

[Chen et al., 2011] Chen, H., Chen, B., and Varshney, P. K. (2011). A new framework for distributed detection with conditionally dependent observations. *IEEE Transactions on Signal Processing*, 60(3):1409–1419.

[Cherikh, 1989] Cherikh, M. (1989). *Optimal decision and detection in the decentralized case*. PhD thesis, Case Western Reserve University.

[Cover, 1969] Cover, T. M. (1969). Hypothesis testing with finite statistics. *The Annals of Mathematical Statistics*, 40(3):828–835.

[Cover and Thomas, 2006] Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory*. Wiley-Interscience, New York.

[Dembo and Zeitouni, 1998] Dembo, A. and Zeitouni, O. (1998). *Large deviations techniques and applications*. Springer.

[Drakopoulos and Lee, 1991] Drakopoulos, E. and Lee, C. C. (1991). Optimum multisensor fusion of correlated local decisions. *IEEE Transactions on Aerospace and Electronic Systems*, 27(4):593–606.

[Ekchian and Tenney, 1982] Ekchian, L. K. and Tenney, R. R. (1982). Detection networks. In *IEEE 21st Conference on Decision and Control, 1982*, pages 686–691.

[Grenander and Szegö, 1984] Grenander, U. and Szegö, G. (1984). *Toeplitz forms and their applications*. American Mathematical Society.

[He and Blum, 2011] He, Q. and Blum, R. S. (2011). New hypothesis testing-based methods for fault detection for smart grid systems. In *45th Annual Conference on Information Sciences and Systems, 2011. CISS 2011*, pages 1–6.

[Hellman and Cover, 1970] Hellman, M. E. and Cover, T. M. (1970). Learning with finite memory. *The Annals of Mathematical Statistics*, 41(3):765–782.

[Helstrom, 1994] Helstrom, C. W. (1994). *Elements of signal detection and estimation*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

[Hoballah and Varshney, 1989] Hoballah, I. Y. and Varshney, P. K. (1989). Distributed Bayesian signal detection. *IEEE Transactions on Information Theory*, 35(5):995–1000.

[Huber, 1965] Huber, P. (1965). A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, 36(6):1753–1758.

[Justesen, 2005] Justesen, J. (2005). Fields from Markov chains. *IEEE Transactions on Information Theory*, 51(12):4358–4362.

[Kailath et al., 2000] Kailath, T., Sayed, A. H., and Hassibi, B. (2000). *Linear estimation*. Prentice-Hall.

[Kam et al., 1992] Kam, M., Zhu, Q., and Gray, W. S. (1992). Optimal data fusion of correlated local decisions in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, 28(3):916–920.

[Kassam and Thomas, 1988] Kassam, S. A. and Thomas, J. B. (1988). *Signal detection in non–Gaussian noise*, volume 5. Springer-Verlag New York.

[Kemeny and Snell, 1976] Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov chains*. Springer.

[Koplowitz, 1975] Koplowitz, J. (1975). Necessary and sufficient memory size for m-hypothesis testing. *IEEE Transactions on Information Theory*, 21(1):44–46.

[Lauritzen, 1996] Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press.

[Lázaro et al., 2005] Lázaro, M., Artés-Rodríguez, A., and Sánchez-Fernández, M. (2005). Decentralized detection in binary dense sensor networks: To transmit or not to transmit. In *Proceedings of IEEE International Workshop on Wireless Ad-Hoc Networks, 2005. IWWAN 2005*.

[Lázaro et al., 2009] Lázaro, M., Sánchez-Fernández, M., and Artés-Rodríguez, A. (2009). Optimal sensor selection in binary heterogeneous sensor networks. *IEEE Transactions on Signal Processing*, 57(4):1577–1587.

[Levy, 2008] Levy, B. (2008). *Principles of signal detection and parameter estimation*. Springer Verlag.

[Li and Dai, 2007] Li, W. and Dai, H. (2007). Distributed detection in wireless sensor networks using a multiple access channel. *IEEE Transactions on Signal Processing*, 55(3):822–833.

[Marano et al., 2009] Marano, S., Matta, V., and Tong, L. (2009). Distributed detection in the presence of byzantine attacks. *IEEE Transactions on Signal Processing*, 57(1):16–29.

[Misra and Tong, 2008] Misra, S. and Tong, L. (2008). Error exponents for the detection of Gauss–Markov signals using randomly spaced sensors. *IEEE Transactions on Signal Processing*, 56(8):3385–3396.

[Møller, 2003] Møller, J. (2003). *Spatial statistics and computational methods*. Springer, New York.

[Papastavrou and Athans, 1992] Papastavrou, J. D. and Athans, M. (1992). Distributed detection by a large team of sensors in tandem. *IEEE Transactions on Aerospace and Electronic Systems*, 28(3):639–653.

[Papoulis and Pillai, 2002] Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables and stochastic processes*. McGraw-Hill Education, New York.

[Pickard, 1977] Pickard, D. K. (1977). A curious binary lattice process. *Journal of Applied Probability*, pages 717–731.

[Plata-Chaves, 2009] Plata-Chaves, J. (2009). Neyman–Pearson decentralized detection with dependent observations. Master's thesis, Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Spain. Supervised by Dr. Marcelino Lázaro.

[Plata-Chaves and Lázaro, 2011] Plata-Chaves, J. and Lázaro, M. (2011). Closed-Form error exponent for the Neyman–Pearson fusion of dependent local decisions in a one-dimensional sensor network. *IEEE Transactions on Signal Processing*, 59(3):1239–1254.

[Plata-Chaves et al., 2011a] Plata-Chaves, J., Lázaro, M., and Artés-Rodríguez, A. (2011a). Neyman-Pearson detection in sensor networks with dependent observations. In *Workshop on Topics in Information Theory and Communications, WTITC 2011*.

[Plata-Chaves et al., 2011b] Plata-Chaves, J., Lázaro, M., and Artés-Rodríguez, A. (2011b). Optimal Neyman–Pearson fusion in two-dimensional sensor networks with serial architecture and dependent

observations. In *14th International Conference on Information Fusion, 2011. Fusion 2011*, pages 1–6.

[Plata-Chaves et al., 2012a] Plata-Chaves, J., Lázaro, M., and Artés-Rodríguez, A. (2012a). Asymptotic performance of the Neyman–Pearson test in tandem Networks with dependent observations. *In preparation*.

[Plata-Chaves et al., 2012b] Plata-Chaves, J., Lázaro, M., and Artés-Rodríguez, A. (2012b). Distributed detection in large tandem networks with dependent observations. *Submitted to IEEE Transactions on Signal Processing*.

[Plata-Chaves and Lázaro, 2009] Plata-Chaves, J. and Lázaro, M. (2009). Closed-form error exponent for the Neyman–Pearson fusion of Markov local decisions. In *IEEE/SP 15th Workshop on Statistical Signal Processing, 2009. SSP 2009*, pages 533–536.

[Plata-Chaves and Lázaro, 2010] Plata-Chaves, J. and Lázaro, M. (2010). Closed-form error exponent for the Neyman–Pearson fusion of two-dimensional Markov local decisions. In *18th European Signal Conference, 2010. EUSIPCO 2010*, pages 164–168.

[Plata-Chaves et al., 2008] Plata-Chaves, J., Lázaro, M., and Artés-Rodríguez, A. (2008). Decentralized detection in a dense wireless sensor network with correlated observations. In *Workshop on Information Theory for Sensor Networks, 2008. WITS 2008*, pages 3–10.

[Poor, 1994] Poor, H. V. (1994). *An introduction to signal detection and estimation*. Springer, New York.

[Reibman and Nolte, 1987] Reibman, A. R. and Nolte, L. W. (1987). Design and performance comparison of distributed detection networks. *IEEE Transactions on Aerospace and Electronic Systems*, 23(6):789–797.

[Spitzer, 1971] Spitzer, F. (1971). Markov random fields and Gibbs ensembles. *American Mathematical Monthly*, pages 142–154.

[Srinivasan, 1986a] Srinivasan, R. (1986a). Distributed radar detection theory. *IEE Proceedings F of Communications, Radar and Signal Processing*, 133(1):55–60.

[Srinivasan, 1986b] Srinivasan, R. (1986b). A theory of distributed detection. *Signal Processing*, 11(4):319–327.

[Sung et al., 2008a] Sung, Y., Poor, H. V., and Yu, H. (2008a). Large deviations analysis for the detection of 2D hidden Gauss–Markov random fields using sensor networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pages 3893–3896.

[Sung et al., 2009] Sung, Y., Poor, H. V., and Yu, H. (2009). How much information can one get from a wireless ad hoc sensor network over a correlated random field? *IEEE Transactions on Information Theory*, 55(6):2827–2847.

[Sung et al., 2006] Sung, Y., Tong, L., and Poor, H. V. (2006). Neyman–Pearson detection of Gauss–Markov signals in noise: closed-form error exponent and properties. *IEEE Transactions on Information Theory*, 52(4):1354–1365.

[Sung et al., 2008b] Sung, Y., Zhang, X., Tong, L., and Poor, H. V. (2008b). Sensor configuration and activation for field detection in large sensor arrays. *IEEE Transactions on Signal Processing*, 56(2):447–463.

[Tang et al., 1991] Tang, Z. B., Pattipati, K. R., and Kleinman, D. L. (1991). Optimization of detection networks: Part I – Tandem structures. *IEEE Transactions on Systems, Man and Cybernetics*, 21(5):1044–1059.

[Tartakovsky, 1998] Tartakovsky, A. G. (1998). Asymptotic optimality of certain multihypothesis sequential tests: Non-iid case. *Statistical Inference for Stochastic Processes*, 1(3):265–295.

[Tartakovsky and Veeravalli, 2005] Tartakovsky, A. G. and Veeravalli, V. V. (2005). General asymptotic Bayesian theory of quickest change detection. *Theory of Probability and its Applications*, 49(3):458–497.

[Tay, 2011] Tay, W. P. (2011). The value of feedback in decentralized detection. *Submitted to IEEE Transactions on Information Theory*, [Online]. Available: http://aps.arxiv.org/abs/1108.6121.

[Tay and Tsitsiklis, 2011] Tay, W. P. and Tsitsiklis, J. N. (2011). Error exponents for decentralized detection in feedback architectures. In *IEEE*

International Conference on Acoustics, Speech and Signal Processing, 2011. ICASSP 2011, pages 2496–2499.

[Tay et al., 2007] Tay, W. P., Tsitsiklis, J. N., and Win, M. Z. (2007). Asymptotic performance of a censoring sensor network. *IEEE Transactions on Information Theory*, 53(11):4191–4209.

[Tay et al., 2008a] Tay, W. P., Tsitsiklis, J. N., and Win, M. Z. (2008a). Data fusion trees for detection: Does architecture matter? *IEEE Transactions on Information Theory*, 54(9):4155–4168.

[Tay et al., 2008b] Tay, W. P., Tsitsiklis, J. N., and Win, M. Z. (2008b). On the subexponential decay of detection error probabilities in long tandems. *IEEE Transactions on Information Theory*, 54(10):4767–4771.

[Tay et al., 2009] Tay, W. P., Tsitsiklis, J. N., and Win, M. Z. (2009). Bayesian detection in bounded height tree networks. *IEEE Transactions on Signal Processing*, 57(10):4042–4051.

[Tenney and Sandell, 1981] Tenney, R. R. and Sandell, N. R. (1981). Detection with distributed sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 49(4):501–510.

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

[Thomopoulos et al., 1989] Thomopoulos, S. C. A., Viswanathan, R., and Bougoulias, D. K. (1989). Optimal distributed decision fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 25(5):761–765.

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

[Tsitsiklis, 1993] Tsitsiklis, J. N. (1993). Decentralized detection. *Advances in Statistical Signal Processing*, 2:297–344.

[Vajda, 1989] Vajda, I. (1989). *Theory of statistical inference and information*. Kluwer Academic, Dordrecht, The Netherlands.

[Van Trees, 2001] Van Trees, H. L. (2001). *Detection, estimation, and modulation theory, Vol. I.* Wiley-Interscience.

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

[Varshney, 1997] Varshney, P. K. (1997). *Distributed detection and data fusion*. Springer, New York.

[Villard and Bianchi, 2011] Villard, J. and Bianchi, P. (2011). High-rate vector quantization for the Neyman–Pearson detection of correlated processes. *IEEE Transactions on Information Theory*, 57(8):5387–5409.

[Villard et al., 2010] Villard, J., Bianchi, P., Moulines, E., and Piantanida, P. (2010). High-rate quantization for the Neyman–Pearson detection of hidden Markov processes. In *Information Theory Workshop, 2010. ITW 2010*, pages 1–5.

[Viswanathan et al., 1988] Viswanathan, R., Thomopoulos, S. C. A., and Tumuluri, R. (1988). Optimal serial distributed decision fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 24(4):366–376.

[Wainwright and Jordan, 2008] Wainwright, M. J. and Jordan, M. I. (2008). *Graphical models, exponential families, and variational inference*. Now Publishers Inc.

[Wald, 1947] Wald, A. (1947). *Sequential analysis*. Wiley Online Library.

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

[Willett and Warren, 1992] Willett, P. and Warren, D. (1992). The suboptimality of randomized tests in distributed and quantized detection systems. *IEEE Transactions on Information Theory*, 38(2):355–361.

[Xiang and Wang, 2006] Xiang, M. and Wang, L. (2006). Distributed signal detection with serial structures. In *9th International Conference on Information Fusion, 2006. ICIF 2006*, pages 1–6.

[Yan and Blum, 2000] Yan, Q. and Blum, R. S. (2000). On some unresolved issues in finding optimum distributed detection schemes. *IEEE Transactions on Signal Processing*, 48(12):3280–3288.

[Yan and Blum, 2001] Yan, Q. and Blum, R. S. (2001). Distributed signal detection under the Neyman–Pearson criterion. *IEEE Transactions on Information Theory*, 47(4):1368–1377.

[Zeitouni et al., 1992] Zeitouni, O., Ziv, J., and Merhav, N. (1992). When is the generalized likelihood ratio test optimal? *IEEE Transactions on Information Theory*, 38(5):1597–1602.

[Zhang et al., 2012] Zhang, Z., Pezeshki, A., Moran, W., Howard, S. D., and Chong, E. K. P. (2012). Error probability bounds for balanced binary relay trees. *To appear in IEEE Transactions on Information Theory*.

[Zoumpoulis et al., 2010] Zoumpoulis, S. I., Tsitsiklis, J. N., and Kreidl, O. P. (2010). On decentralized detection with partial information sharing among sensors. *IEEE Transactions on Signal Processing*, 59(4):1759–1765.