Decentralized Parameter and Random Field Estimation with Wireless Sensor Networks

PhD dissertation by

Javier Matamoros Morcillo

PhD Advisor

Dr. Carles Antón-Haro
Centre Tecnològic de Telecomunicacions de Catalunya

Universitat Politècnica de Catalunya

Para Estela,

Abstract

In recent years, research on Wireless Sensor Networks (WSN) has attracted considerable attention. This is in part motivated by the large number of applications in which WSNs are called to play a pivotal role, such as parameter estimation (namely, moisture, temperature), event detection (leakage of pollutants, earthquakes, fires), or localization and tracking (for e.g. border control, inventory tracking), to name a few.

This PhD dissertation is focused on the design of *decentralized* estimation schemes for wireless sensor networks. In this context, sensors observe a given phenomenon of interest (e.g. temperature). Consequently, sensor observations are conveyed over the wireless medium to a Fusion Center (FC) for further processing. The ultimate goal of the WSN is the *estimation* or *reconstruction* of the phenomenon with minimum distortion. The problem is addressed from a signal processing and information-theoretical perspective. However, the interplay with some selected functionalities at the link layer of the OSI protocol stack (e.g. scheduling protocols) or network topologies (flat/hierarchical) are also taken into consideration where appropriate.

First, this dissertation addresses the power allocation problem in amplify-and-forward wireless sensor networks for the estimation of a *spatially-homogeneous* parameter. This study is mainly devoted to the analysis of a class of Opportunistic Power Allocation (OPA) strategies which operate with low complexity and stringent signalling requirements. Several problems of interest in WSNs are considered: *i*) the minimization of distortion, *ii*) the minimization of transmit power and, *iii*) the enhancement of network lifetime. Finally, hierarchical network topologies are introduced for those situations where sensor-to-FC channel links suffer from severe path losses. In this context, the analysis is aimed to identify the power allocation strategy that provides the best performance trade-off between the estimation accuracy and the signaling requirements.

Second, sensor nodes are allowed to transmit their observations *digitally*. In this setting, two encoding strategies are analyzed: Quantize-and-Estimate (Q&E) encoding and Compress-and-Estimate (C&E) encoding, which operate with and without side information at the decoder, respectively. This PhD dissertation addresses a number of issues of interest: *i*) the impact of different channel models (Gaussian, Rayleigh-fading channels with/without transmit CSI) on the accuracy of the estimates, *ii*) the optimal number of sensors to be deployed and, *iii*) the

impact of *realistic* contention-based multiple-access protocols on the estimation distortion.

Finally, this PhD dissertation focuses on the estimation of *spatial* random fields. In this scenario, the spatial variability of the parameter of interest is taken into account, rather than assuming the estimation of a *single* (i.e. spatially-homogeneous) parameter. Two different scenarios are considered, namely, delay-*constrained* networks and delay-*tolerant* networks. In addition, the case where sensors cannot acquire instantaneous transmit CSI (CSIT) is addressed. In this context, the outage events experienced in the sensors-to-FC links result in a random sampling effect which is investigated.

Resumen

Las redes de sensores inalámbricas están compuestas por un gran número de dispositivos de bajo coste y bajo consumo energético llamados sensores. Estos sensores incluyen funcionalidades como: sensado, técnicas básicas de procesado de señal y un transceptor RF. Las aplicaciones más comunes de las redes de sensores inalámbricas son: monitorización medioambiental, detección de eventos, monitorización de objetos, seguridad doméstica, aplicaciones médicas y militares, entre otras.

Principalmente, el objetivo de esta tesis doctoral es el diseño de esquemas de estimación descentralizados para redes de sensores inalámbricas. En todos los escenarios considerados en esta tesis, los sensores observan y muestrean un fenómeno de interés (e.g. temperatura, presión, humedad...). Posteriormente, las muestras almacenadas en los sensores son transmitidas a través de un canal inalámbrico hacia un centro de fusión para su procesamiento. El principal objetivo de la red de sensores es la estimación o reconstrucción del fenómeno de interés con la mínima distorsión. El problema se plantea desde un punto de vista de procesado de señal y teoría de la información. Sin embargo, también se considera la interacción con algunas funcionalidades de la capa de enlace de la pila de protocolos OSI (e.g. protocolos de scheduling) además de diferentes topologías de red (plana y jerárquica).

En primer lugar, esta tesis se centra en el problema de asignación de potencia en redes de sensores. En particular, los sensores amplifican y retransmiten sus observaciones hacia el centro de fusión (i.e. comunicaciones analógicas). En este contexto, se proponen y analizan varias técnicas de asignación de potencia oportunistas, cuyas características son su baja complejidad y requisitos de señalización. Se consideran varios problemas específicos de una red de sensores: i) la minimización de la distorsión, ii) la minimización de la potencia transmitida y, iii) el aumento del tiempo de vida de la red. Finalmente, se introducen topologías de red jerárquicas con el objetivo de paliar las pérdidas por propagación comunes en escenarios donde los sensores están situados a una gran distancia del centro de fusión. En este escenario, el objetivo es identificar la estrategia de asignación de potencia más apropiada, teniendo en cuenta la calidad de estimación y los requisitos de señalización de ésta.

En segundo lugar, se considera el caso en que los sensores codifican sus observaciones usando un determinado número de bits (i.e. comunicaciones digitales). En este escenario,

se analizan dos estrategias de codificación: Quantize-and-Estimate (Q&E) y Compress-and-Estimate (C&E). A diferencia de Q&E, la estrategia C&E permite incorporar la información disponible en el receptor en la codificación de las observaciones, obteniendo así una menor distorsión en el centro de fusión. En esta tesis se tratan varios problemas de interés: i) el impacto de diferentes modelos de canal (canales Gausianos y canales con desvanecimientos Rayleigh con/sin información instantánea del canal) en la calidad de las estimaciones, ii) el número óptimo de sensores que se debe desplegar para minimizar la distorsión y; iii) el impacto de protocolos de contención de acceso al medio en la distorsión.

Por último, esta tesis se centra en la estimación de campos espaciales. En este contexto, se adopta un modelo de correlación que, a diferencia de los estudios anteriores, tiene en cuenta la variabilidad del parámetro en el espacio. En este contexto, el estudio se centra en dos tipos de aplicaciones: redes de sensores con restricciones de retardo en la estimación y redes de sensores con una cierta tolerancia en el retardo de la estimación. Finalmente, se analiza el caso más realista, en el que los sensores no disponen de información instantánea del canal y por lo tanto no pueden transmitir sus datos de manera fiable. Por consiguiente, el objetivo es el análisis del impacto de este fenómeno en el muestreo del campo y de esta forma en la distorsión del campo reconstruido.

Acknowledgements

En primer lugar, quisiera agradecer al CTTC por brindarme la posibilidad de realizar esta tesis doctoral. Durante estos 4 años he vivido experiencias que nunca olvidaré.

Querría ofrecer mi más sincero agradecimiento a Carles Antón, director de esta tesis. Carles, has estado siempre dispuesto a discutir cualquier aspecto técnico y has conseguido proporcionarme la motivación y apoyo necesarios para la realización de esta tesis doctoral. Gracias.

Gracias a mis colegas doctorandos (algunos ya doctores): Aitor, Bego, Miguel Ángel, Musbah, Giuseppe, Jaime, Ana María, Fran, Jalonso, Alessandro, Dani, Iñaki, Javi Arribas, Angelos, Pol. Gracias por las innumerables tertulias y discusiones (casi nunca técnicas ;–)), lo cual ayuda a crear un maravilloso entorno de trabajo donde me he sentido como en casa. Luego están mis amigos y compañeros, David y Jesusín, gracias a los dos por vuestro apoyo durante estos últimos 4 años, tanto en los buenos como en los no tan buenos momentos.

Gracias también a todos los miembros del club *tupperware* (Ana Collado, Selva, Fermín, Toni, Miquel, Luís, . . .), con los que he mantenido infinidades de charlas durante la comida. Querría resaltar a mi compañero *culé* Ricardo, con el que he discutido temas futboleros. No querría dejarme a nadie por el camino así que gracias a TODOS mis compañeros del CTTC.

Agradezco también a mis viejos y grandes amigos Marta, Marcos, Víctor y Patri por el apoyo y porque la amistad es de lo más bonito que existe. Gracias Marcos y Víctor por esas timbas de *Póker* y por esas partidas al *PES*, con esto todo ha sido más fácil.

Gracias a mi madre, que siempre está ahí y que siempre tiene una palabra de ánimo y cariño. Gracias María, la mejor hermana del mundo, y gracias a mi padre, que aunque no pudo ver esto me transmitió unos valores e ideales que de alguna forma me empujaron a ello.

Por último, Estela mil gracias, has estado siempre ahí apoyándome y dándome ánimos, me has sufrido durante estos años y nunca podré estarte los suficientemente agradecido. Simplemente, te quiero.

Javier Matamoros Morcillo

Contents

List of acronyms

A&F	Amplify-and-Forward
ADC	Analog-to-Digital Converter
AWGN	Additive White Gaussian Noise
BLUE	Best Linear Unbiased Estimator
BS	Base Station
C&E	Compress-and-Estimate
CDF	Cumulative Density Function
CEDC	Compress-and-Estimate for Delay-Constrained applications
CEDT	Compress-and-Estimate for Delay-Tolerant applications
CEO	Chief Executive Officer
CH	Cluster-Head
CRLB	Cramer-Rao Lower Bound
CSI	Channel State Information
CSIT	Transmit Channel State Information
DC	Delay-Constrained
DPH	Discrete PHase-type
DT	Delay-Tolerant
FC	Fusion Center
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
GMOU	Gaussian Markov Ornstein-Ulenhbek
LT	Lifetime
LMMSE	Linear Mean Squared Error
MAC	Multiple-Access Channel
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MUD	Multi-User Diversity
MVU	Minimum Variance Unbiased estimator

OPA	Oportunistic Power Allocation
OPA-D	Oportunistic Power Allocation for the minimization of Distortion
OPA-DR	Robust Opportunistic Power Allocation for the minimization of Distortion
OPA-LT	Opportunistic Power Allocation for the minimization of Distortion
OPA-P	Oportunistic Power Allocation for the minimization of transmit Power
OSI	Open System Interconnection
pdf	Probability Density Function
PFS	Proportional Fair Scheduling
PHY	PHYsical layer
pmf	Probability Mass Function
Q&E	Quantize-and-Estimate
QEDC	Quantize-and-Estimate for Delay-Constrained applications
QEDT	Quantize-and-Estimate for Delay-Tolerant applications
REI	Residual Energy Information
SENMA	Sensor Networks with Mobile Access
SMUD	Selective Multi-User Diversity
SNR	Signal-to-Noise Ratio
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
UPA	Uniform Power Allocation
WF	Waterfilling
WF-D	Waterfilling for the minimization of distortion
WF-P	Waterfilling for the minimization of transmit power
WSN	Wireless Sensor Network

Notation

$\mathbf{X}$	Matrix
$\mathsf{diag}\left[\mathbf{X}\right]$	Vector constructed with the elements in the diagonal of matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$
$\mathbf{x}$	Column vector
^{T}	Transpose operator
$\left\|\cdot\right\|$	Euclidean norm
$\mathbf{I}_n$	Identity matrix of dimension $n \times n$
$\sim$	Distributed according to / scale as
$\triangleq$	Definition
$\lim$	Limit
$\log\left(\cdot\right)$	base-2 logarithm
$\ln\left(\cdot\right)$	Natural logarithm
$\arg$	Argument
$\max, \min$	Maximum and minimum
$\mathsf{W}_0\left(\cdot\right)$	First real branch of the Lambert W function
$\mathsf{W}_{-1}\left(\cdot\right)$	Negative real branch of the Lambert W function
$\left[x\right]^+$	Positive part of x, i.e. $\max\left\{0, x\right\}$
$\Pr\left(\cdot\right)$	Probability
$\mathbb{E}\left[\cdot\right]$	Mathematical expectation
$\mathrm{H}\left(\cdot\right)$	Entropy
$\mathrm{I}\left(\cdot;\cdot\right)$	Mutual information

Chapter 1

Introduction

1.1 Motivation

Wireless Sensor Networks (WSN) consist of a potentially large number of energy-constrained sensing devices (the *sensors*) which are capable of conveying data over wireless links to a Fusion Center (FC) where it is further processed. A non-exhaustive list of applications for WSNs encompass environmental monitoring (e.g. determination of the concentration of pollutants, temperature, pressure), event detection (leakage of substances, earthquakes, fire), localization and tracking of assets, healthcare (remote patient monitoring), or military applications (surveillance, border control) to name a few (see e.g. [1–4] for more examples). Due to such a broad range of applications, in the coming years wireless sensor networks are called to play a pivotal role in our daily lives. This has driven substantial advances in e.g. energy harvesting techniques, microelectronics, decentralized signal processing, wireless communications or networking.

Wireless *sensor* networks differ from other *data* networks in many aspects. To start with, sensors are typically equipped with batteries which, unlike in e.g. mobile phones, are often difficult or impossible to replace. Consequently, the development of energy-efficient signal processing techniques and communication protocols capable of enhancing network lifetime becomes a priority. Besides, both the computational and data transmission capabilities (rate,

range, etc.) of sensor nodes are, in general, rather limited. Moreover, in many applications sensors have to be deployed in remote and/or large areas. All of the above, clearly advocates for the adoption of *decentralized* signal processing techniques and networking protocols. This, on the one hand, minimizes the need for nodes to coordinate with a central authority (which translates into energy savings due to reduced signalling) and, on the other, allows for a more sophisticated processing of data by leveraging on the computational capabilities of the individual sensor nodes.

Another distinctive feature of WSNs is that their design is *application dependent* whereas wireless *data* networks (such as Wi-Fi, 3G or WiMax) are typically conceived as generic-purpose. In terms of performance metrics, for instance, a WSN aimed to detect fires in a forest should be optimized to attain the best trade-off between the probabilities of detection and false alarm (latency consideration could also be taken into account). On the contrary, in environmental monitoring applications one is more concerned about estimating the parameter of interest (e.g. temperature) with the highest possible accuracy. Ultimately, those differences in terms of purpose and performance metrics translate into a number of differences concerning architectural, terminal or communication protocol designs.

Other particularities of WSNs stem from the fact that the number of sensors in such networks is potentially large. Consequently, the unitary cost of those devices needs to be kept low which often translates into an error-prone behavior. This has a number of implications. To start with, network designs should be robust to such imperfections and possible malfunctioning. For example, routing schemes should bear in mind that sporadically a specific node might be unable to forward data. Besides, one should ensure that the designed algorithms and protocols scale well with an increasing number of sensor nodes. For instance, detection rules with a computational complexity which is e.g. exponential in the number of nodes might not be suitable for large sensor networks.

Finally, an important aspect which impacts on the design of WSNs is the fact that sensor measurements are often correlated (e.g. temperature measured by sensors which are close to each other). This assumption, which seldom holds in wireless *data* networks, triggers a number of interesting design trade-offs. For instance, one could think of successive encoding schemes capable of removing redundancy in the transmitted data and, by doing so, achieve substantial energy savings. However, the resulting encoding scheme becomes more sensitive to channel outages and dropped frames which, ultimately, might entail the re-transmission of the whole set of measurements. Clearly, this would be barely desirable in terms of latency and/or energy consumption.

This PhD dissertation is focused on the design of decentralized estimation schemes for wireless sensor networks. The problem is mostly tackled from a signal processing and information-theoretical perspective. Still, the interplay with some selected functionalities at the link layer of the OSI protocol stack (e.g. scheduling protocols) or network topologies (flat/hierarchical)

will be taken into consideration where appropriate. Ultimately, this PhD dissertation attempts to find answers to questions such as what is the power allocation scheme that exhibits the best trade-off in terms of performance, complexity and signalling requirements? What is the price to be paid, in terms of estimation accuracy, in order to enhance network lifetime? What is the optimal number of sensor nodes needed to be deployed as a function of the spatial variability of the parameter to be estimated? Is it worth using successive encoding schemes in practical wireless sensor networks? What is the impact of contention-based (vs. contention-free) multiple-access schemes on the attainable distortion?

In subsequent sections, we outline the contents and organization of this PhD Dissertation. Next, we provide a list of journal and conference publications that have resulted from the realization of this work.

1.2 Outline

This PhD dissertation is focused on the design and analysis of estimation techniques for WSNs. Chapter 3 addresses the case of amplify-and-forward WSNs in which the sensor observations are scaled by an amplifying factor for their transmission to the FC (i.e. *analog* transmissions). On the contrary, in Chapters 4 and 5 sensors encode/quantize data into a number of bits (i.e. *digital* transmissions) before sending them to the FC. Regarding the estimation problem, Chapters 3 and 4 are devoted to the estimation of a spatially-homogeneous parameter, whereas Chapter 5 addresses the more realistic case of the estimation of spatial random fields. In all cases, the main performance metric is the distortion in the estimates.

This PhD dissertation is organized as follows:

Chapter 2 reviews a number of concepts which will be used in this PhD Dissertation. First, we present an overview of hardware and network topologies issues in wireless sensor networks. Next, some basic concepts on estimation theory, information theory and opportunistic communications are outlined, respectively.

Chapter 3 addresses the problem of estimating a spatially-homogeneous *parameter* with amplify-and-forward WSNs. In this setting, the main contribution is the design of an Opportunistic Power Allocation (OPA) scheme with low signalling requirements. The scheme is particularized for several problems of interest: *i*) the minimization of distortion, *ii*) the minimization of transmit power and, *iii*) the enhancement of network lifetime. Besides, a hierarchical network topology is also proposed for scenarios in which sensors are placed at large distances from the FC. In this setting, sensors are grouped into clusters and a cluster-head (CH) is in charge of consolidating and sending a local estimate to the FC. For this network topology, an exhaustive comparison of different power allocation strategies is conducted. In particular, all

the strategies are compared in terms of estimation distortion and Channel State Information (CSI) requirements.

In Chapter 4, the problem at hand continues to be the estimation of a spatially-homogeneous parameter but, unlike in the previous chapter, sensor nodes transmit their observations *digitally*. In particular, two encoding strategies are analyzed: *i*) Quantize-and-Estimate (Q&E) and *ii*) Compress-and-Estimate (C&E). First, Chapter 4 addresses a scenario with *orthogonal* sensor-to-FC channels (i.e. TDMA/FDMA). In this setting and by constraining *bandwidth* and *power* irrespectively of the network size, the distortion behavior is analyzed for different channel models (Gaussian, Rayleigh-fading, with/without Transmit CSI (CSIT)). For scenarios in which sensors operate without instantaneous CSIT, a *constant* and *common* encoding rate must be used. Such encoding rate must be carefully designed since, ultimately, it determines the outage probability experienced at the sensors-to-FC channels, *and* the resolution at which sensor observations are encoded (and both phenomena have an impact on the accuracy in the estimates). For the C&E encoding strategy, the impact of the encoding order on the distortion, which arises from the successive encoding/decoding structure of the strategy, is also investigated. Next, Chapter 4 addresses a (more realistic) case in which sensors seize the channel via contention-based multiple-access mechanisms (e.g. ALOHA). Furthermore, a hierarchical topology is adopted and the performance of reservation-based protocols and contention-based multiple-access protocols at different levels of the hierarchy is analyzed.

Chapter 5 goes one step beyond and focuses on the estimation of spatial random *fields*. In this setting, the correlation between the sensor observations is determined by the distance between sensors. Two different scenarios are considered, namely, delay-*constrained* networks and delay-*tolerant* networks. Besides, two different encoding strategies are adopted: *i*) quantize-and-estimate encoding and, *ii*) compress-and-estimate where each sensor exploits (at the encoder) the correlation between adjacent observations. Finally, the case in which sensors cannot acquire instantaneous transmit CSI (CSIT) is addressed. In this context, as in Chapter 4, a *constant* and *common* encoding rate is adopted at the sensor nodes which along network size are optimized.

Chapter 6 concludes this PhD dissertation with a summary and a discussion of the main results of this work. Some suggestions for future work are also outlined.

1.3 Contribution

Chapter 3

The main contributions of Chapter 3 have been published in 1 journal paper and 4 conference papers while another 1 letter is under review.

- J. Matamoros, C. Antón-Haro, Opportunistic Power Allocation and Sensor Selection Schemes for Wireless Sensor Networks, IEEE Transactions on Wireless Communications, vol. 9, no. 2, pp. 534–539, Feb. 2010.
- J. Matamoros, C. Antón-Haro, Scaling Law of an Opportunistic Power Allocation Scheme for Amplify-and-Forward Wireless Sensor Networks, submitted to the IEEE Communications Letters.
- J. Matamoros, C. Antón-Haro, Opportunistic Power Allocation in Wireless Sensor Networks with Imperfect Channel State Information, in Proceedings of the ICT-Mobile Summit 2008, 10-12 June 2008, Stockholm (Sweden).
- J. Matamoros, C. Antón-Haro, Opportunistic Power Allocation Schemes for the Maximization of Network Lifetime in Wireless Sensors Networks, in Proceedings of IEEE Int'l Conference on Audio, Speech and Signal Processing (ICASSP 2008), Apr. 2008, Las Vegas, Nevada (USA).
- J. Matamoros, C. Antón-Haro, Hierarchical Organizations of Sensors for Decentralized Parameter Estimation, In Proc. IEEE Int'l Symposium on Signal Processing and Information Technology (ISSPIT), Dec. 2007, El Cairo (Egypt).
- J. Matamoros, C. Antón-Haro, Opportunistic Power Allocation Schemes for Wireless Sensor Networks, IEEE Int'l Symposium on Signal Processing and Information Technology (ISSPIT), Dec. 2007, El Cairo (Egypt).

Chapter 4

Contributions of Chapter 4 have been published in part in 5 conference papers, and 1 journal paper is under review.

- J. Matamoros, C. Antón-Haro, Optimal Network Size and Encoding Rate for Wireless Sensor Network-based Decentralized Estimation under Power and Bandwidth Constraints, submitted to IEEE Transactions on Wireless Communications.
- J. Matamoros, C. Antón-Haro, Hierarchical Wireless Sensor Networks with Contention-based Multiple-Access Schemes - A PHY/MAC Cross-layer Design, in Proceedings of Second International Workshop on Cross Layer Design, (IWCLD 2009), 11-12 June 2009, Palma de Mallorca (Spain).
- J. Matamoros, C. Antón-Haro, Optimized Constant-Rate Encoding for Decentralized Parameter Estimation with Wireless Sensor Networks, in Proceedings of IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC 2009), 21-24 June 2009, Perugia (Italy).

- J. Matamoros, C. Antón-Haro, To Sort or not to Sort: Optimal Sensor Scheduling for Successive Compress-and-Estimate Encoding, in Proceedings of IEEE International Conference in Communications (ICC 2009), 14-18 June 2009, Dresden (Germany).
- J. Matamoros, C. Antón-Haro, Bandwidth Contraints in Wireless Sensor Networks for Rayleigh Fading Channels, in Proceedings of NEWCOM++, ACoRN Joint Workshop, 30-1 April 2009, Barcelona (Spain).
- J. Matamoros, C. Antón-Haro, Bandwidth Constraints in Wireless Sensor-based Decentralized Estimation Schemes for Gaussian Channels, in Proceedings of IEEE Global Conference on Communications (GLOBECOM 2008), 30 November-4 December 2008, Louisiana, New Orleans (USA).

Chapter 5

Finally, contributions of Chapter 5 have been published in part in 2 conference papers and, 1 journal paper and 2 conference papers are under review.

- J. Matamoros, C. Antón-Haro, Random Field Estimation with Delay-constrained and Delay-tolerant Wireless Sensor Networks, submitted to EURASIP Journal on Wireless Communications and Networking.
- J. Matamoros, C. Antón-Haro, Optimal network size and encoding rate for random field estimation with wireless sensor networks, in Proceeding of The 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2009). 13-16 December 2009, Aruba (Dutch Antilles).
- J. Matamoros and C. Antón-Haro, Delay-tolerant vs. delay-constrained estimation of spatial random fields, in Proceedings Future Network & Mobile Summit 2010, Florence, Italy, 16-18 June 2010.
- J. Matamoros and C. Antón-Haro, Quantize-and-estimate encoding schemes for random field estimation with delay-constrained and delay-tolerant wireless sensor networks, submitted to IEEE PIMRC 2010.
- J. Matamoros and C. Antón-Haro, Random field estimation with delay-tolerant wireless sensor networks: Quantize-and-estimate vs. compress-and-estimate encoding, submitted to IEEE Globecom 2010.

Other contributions not presented in this dissertation

During the first year of the PhD studies, two conference papers were published. In these works, a novel random access technique is proposed for decentralized parameter estimation which requires only local CSI at the sensor nodes.

- J. Matamoros, C. Antón-Haro, Opportunistic Random Access for Distributed Parameter Estimation in Wireless Sensor Networks, European Signal Processing Conference (EUSIPCO'07), Poznan (Poland), Sept. 2007, pp. 2439-2443.

- J. Matamoros, C. Antón-Haro, Distributed Scheduling in Wireless Sensor Networks under Heterogeneity and Imperfect Channel State Information, IEEE Statistical Signal Processing Workshop (SSP), August 2007.

Chapter 2

Background

In this chapter we review a number of concepts and mathematical tools which will be used in this PhD dissertation. First, in Section 2.1 we provide an overview of a number of hardware and networking issues in WSNs. Then, in Section 2.2, we introduce several basic concepts in estimation theory and some recent results on decentralized estimation for wireless sensor networks. Next, we establish the link between estimation theory and information theory in Section 2.3. Finally, in Section 2.4, we introduce the concept of multi-user diversity and opportunistic communications and their exploitation in a context of wireless sensor networks.

2.1 Wireless sensor networks: hardware and network issues

Nowadays, Crossbow Motes[1] are perhaps the most popular sensor nodes due to their versatility. These commercial sensor nodes include all basic operations: sensing, simple digital signal processing and an IEEE 802.15.4 RF transceiver. The key features of IEEE 802.15.4 technology [5] are low cost, low complexity, low energy consumption and low data rates. A sensor node is mainly composed of a microprocessor, data storage components, Analog-to-Digital Converters (ADCs), sensors, an RF transceiver and a battery (see Fig. 2.1). Research on hardware is mainly aimed to build and design small electronic components with a low energy

[1] For further details see http://www.xbow.com

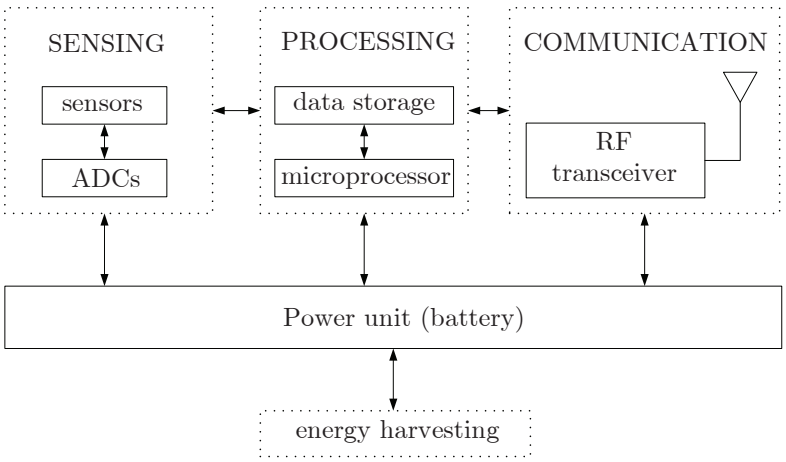


Figure 2.1: Block diagram of a sensor node.

consumption requirements. This fact, along with energy harvesting methods [6] (e.g. wind energy, solar cells), will enable the design of more energy-efficient wireless sensor nodes.

As far as network topologies are concerned, *single-hop* transmission, in which a number of *static* sensor nodes transmit their observations *directly* to the FC, is, by far, the most popular one [7–12]. Still, *multi-hop* transmission is also used in WSNs. In a multi-hop network, the information from the source node hops over a set of intermediate nodes in order to reach the destination (see Figure 2.2). In this context, the authors in [13] propose a multi-hop network where a source transmits to the destination with the help of a set of tier of sensors acting as relays. The authors derive expressions for the ergodic capacity and the outage probability for different fading distributions. In [14] instead, the sensors are grouped into clusters where they cooperate to form MIMO channels between clusters. The authors derive the optimal power allocation (and time sharing) within intra-cluster and inter-cluster communications, in order to minimize the end-to-end outage probability.

Hierarchical network topologies have also been addressed in a number of works (see [15] and references therein). In this setting, sensors are organized into clusters, each of which is under the supervision of a Cluster-Head (CH). Each CH is in charge of consolidating cluster data and conveying such information to the FC (see Fig. 2.3). The nodes which act as the cluster-heads can be determined in advance (e.g. more powerful ones) or, alternatively, can be selected depending on the *current* network conditions (e.g. the one with the strongest channel gain to the FC or, the one with the higher residual energy).

In [16], the authors consider two different scenarios, *i*) multihop transmissions and, *ii*) SENMA (Sensor Networks with Mobile Access), originally introduced in [17] (see Fig. 2.4). This paper studies the scaling behavior of the energy consumption by taking into account the transmission energy and the listening energy. Besides, the case of multiple FCs (both for cooperative and non-cooperative scenarios) is investigated as well.

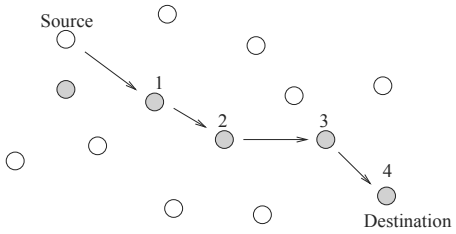


Figure 2.2: Multi-hop network.

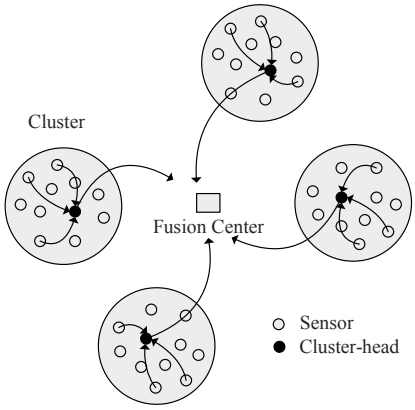


Figure 2.3: Hierarchical network.

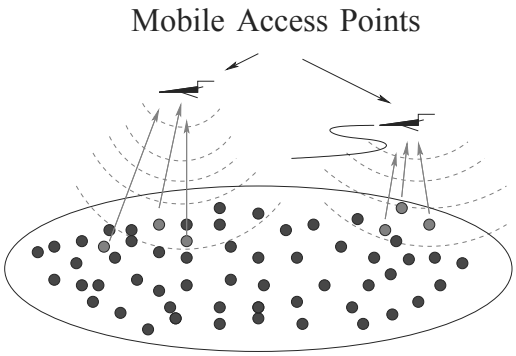


Figure 2.4: Sensor Networks with Mobile Access.

2.2 Estimation theory

This PhD dissertation addresses the problem of the *estimation* of unknown parameters or random field with WSNs. Hence, we start by reviewing some basic concepts of estimation theory. First, we present the well-known problem of *centralized* parameter estimation. Next, we focus on the problem of *decentralized* parameter estimation, where the observations containing information of the parameter of interest are *geographically* distributed as it is the case in sensor networks.

2.2.1 Centralized parameter estimation

The estimation of an unknown parameter is a classical problem [18]. This problem can be approached from two different perspectives: the *classical* estimation and the *Bayesian* estimation.

Classical estimation

In classical estimation, the parameter of interest, denoted in the sequel as θ, is assumed to be deterministic but unknown. The deterministic assumption follows from the fact that no prior statistical information of θ is available. Standing on these basis, the goal is to obtain an estimate of θ, that is $\hat{\theta}$, with minimum distortion. Usually, the distortion D in the estimate $\hat{\theta}$, is defined as the Mean Squared Error (MSE), that is

$$D = \mathbb{E}\left[\left(\theta - \hat{\theta}\right)^2\right] = \int \left(\theta - \hat{\theta}\right)^2 p\left(\mathbf{x};\theta\right) d\mathbf{x}, \tag{2.1}$$

where $\mathbf{x}$ stands for the vector of observations, and $p\left(\mathbf{x};\theta\right)$ is the pdf of the observation vector parameterized by the unknown parameter θ. The interest typically lies in unbiased estimators, i.e. estimator for which $\mathbb{E}[\hat{\theta}] = \theta$. In this context, the so-called Cramer-Rao-Lower-Bound (CRLB), whose definition can be found in [18], constitutes the absolute benchmark. The estimator with the minimum variance (i.e. the one which minimizes (2.1)) for all θ is the so-called Minimum Variance Unbiased (MVU) estimator and, further, if it attains the CRLB it is said to be efficient. Unfortunately, the MVU estimator is in general difficult to find or may not even exist. In this case, one can resort to the Maximum Likehood (ML) estimator, defined as

$$\hat{\theta}_{\mathrm{ML}} \triangleq \arg\max_{\theta} p\left(\mathbf{x};\theta\right), \tag{2.2}$$

which can be shown to be unbiased *and* efficient for an asymptotically large sample size.

An interesting case where the MVU estimator can always be determined is for linear data models, namely

$$\mathbf{x} = \mathbf{h}\theta + \mathbf{n}, \tag{2.3}$$

where $\mathbf{h} = [h_1, \ldots, h_N]^T$ is a known vector and $\mathbf{n} \sim \mathcal{N}\left(0, \mathbf{I}_N\sigma_n^2\right)$ denotes the additive white Gaussian noise (AWGN). By imposing unbiasedness, i.e. $\mathbb{E}_{\mathbf{y}}\left[\hat{\theta}(\mathbf{y})\right] = \theta$, and according to the distortion criteria of (2.1), the minimum variance unbiased estimator reads

$$\hat{\theta} = \left(\sum_{k=1}^{N} \frac{h_k^2}{\sigma_n^2}\right)^{-1} \left(\sum_{k=1}^{N} \frac{h_k x_k}{\sigma_n^2}\right), \tag{2.4}$$

with distortion given by

$$D = \left(\sum_{k=1}^{N} \frac{h_k^2}{\sigma_n^2}\right)^{-1}. \tag{2.5}$$

In this case, this estimator is efficient since it attains the CRLB. Besides, it turns out to be linear in the data and, for linearized data models, it is often referred to as the Best Linear Unbiased Estimator (BLUE).

Bayesian estimation

Unlike in the classical estimation theory, in Bayesian estimation the unknown parameter is assumed to be *random* with a known prior pdf. Consequently, one can exploit this fact by incorporating this prior information into the design of the estimator. In this case, the distortion metric is typically the *bayesian* mean squared error, namely

$$D = \mathbb{E}\left[\left(\theta - \hat{\theta}\right)^2\right] = \int\int \left(\theta - \hat{\theta}\right)^2 p\left(\mathbf{x}, \theta\right) d\mathbf{x} d\theta, \tag{2.6}$$

where the error is averaged over the joint pdf $p\left(\mathbf{x}, \theta\right)$. It is straightforward to show that the optimal estimator in terms of MSE is given by the posterior mean, that is

$$\hat{\theta}(\mathbf{x}) = \mathbb{E}[\theta|\mathbf{x}], \tag{2.7}$$

with distortion given by

$$D = \mathrm{Var}[\theta|\mathbf{x}]. \tag{2.8}$$

As an example, we consider the following linear model

$$\mathbf{x} = \mathbf{h}\theta + \mathbf{n}, \tag{2.9}$$

where $\mathbf{h} = [h_1, \ldots, h_N]^T$ is a known vector and $\mathbf{n} \sim \mathcal{N}\left(0, \mathbf{I}_N \sigma_n^2\right)$ denotes AWGN. Here, we assume that the prior pdf of the parameter θ is available and given by $\theta \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$. Consequently, from (2.7) we have that

$$\hat{\theta} = \mathbb{E}[\theta|\mathbf{x}] = \left(\frac{1}{\sigma_\theta^2} + \sum_{k=1}^{N} \frac{h_k^2}{\sigma_n^2}\right)^{-1} \left(\sum_{k=1}^{N} \frac{h_k x_k}{\sigma_n^2}\right), \tag{2.10}$$

with distortion given by

$$D = \mathrm{Var}[\theta|\mathbf{x}] = \left(\frac{1}{\sigma_\theta^2} + \sum_{k=1}^{N} \frac{h_k^2}{\sigma_n^2}\right)^{-1}. \tag{2.11}$$

Again, the optimal estimator turns out to be linear in the data and corresponds to the Linear Mean Squared Error (LMMSE) estimator [18, Chapter 12]. By comparing (2.11) with (2.5), we see that the prior information about the parameter of interest helps decrease the distortion in the estimates.

2.2.2 Decentralized parameter estimation

In wireless sensor networks, sensor observations are geographically distributed and, hence, the aforementioned estimators have to be designed to operate in a decentralized manner. Furthermore, the sensor network topology has to be taken into account for the design of the decentralized estimation technique. In the sequel, we define *infrastructure-based* networks as a network *with* a FC gathering and processing the information, and *infrastructureless* network as a network *without* any central device or coordinator.

Infrastructure-based networks

For correlated Gaussian sources, the *analog* re-transmission of the observations is known to scale optimally in terms of distortion [12]. Motivated by this result, under orthogonal channels and different local observation qualities, in [11] the optimal power allocation is derived for two different situations: *i*) the minimization of distortion subject to a sum-power constraint, and *ii*) the minimization of transmit power subject to a maximum distortion target. Alternatively, the authors in [9] address the problem of decentralized estimation, where each sensor is only allowed to send *binary* observations to the FC. Interestingly, this paper introduces a class of ML estimators that attain a variance close to the CRLB with merely 1 bit per observation. Besides, by relaxing the bandwidth constraint, the best possible estimator under binary observations is constructed. Similar analysis are conducted in [10] under unknown noise pdf leading to the so-called *universal* (pdf-unaware) estimators. Universal estimators based on *quantized* sensor data have been introduced in [8, 19]. In particular, the work of [19] suggests that the optimal decentralized estimation scheme with 1-bit per observation should allocate 1/2 of the sensors to estimate the first bit of the unknown parameter, 1/4 of the sensors to estimate the second bit, and so on. In addition, [20] proposes a simple probabilistic quantization scheme in order to obtain an unbiased binary message. By doing so, one can simply use a suboptimal and of low complexity estimator, such estimator is the Best Linear Unbiased Estimator (BLUE).

Several models have been proposed in the literature (see [21] and references therein) to characterize the spatial correlation associated to random *fields*. In this context, the GMOU (Gaussian Markov Ornstein-Ulenhbek) model [22] is commonly used in the literature (e.g. see [23–25]) and lends itself to a mathematical tractability. For a general Gaussian correlation model, the interested reader is referred to [26]. The authors in [26] propose a bayesian framework for adaptive quantization at the sensor nodes, which requires a feedback channel.

Infrastructureless networks

Infrastructureless approaches for distributed estimation have been considered in e.g. [27–31]. In [27] each sensor has a first-order dynamical system initialized with the local measurements and, only communication between nearby nodes is allowed, exchanging their local states. In this paper, the authors prove that each node converges to the globally optimal ML estimator under some stability conditions. In this class of estimators, some latency in the estimation turns up, since the process to achieve consensus is iterative in nature. Hence, the energy consumption, which is proportional to the total number of iterations, increases, as well. One could think of decreasing the transmit power to reduce the energy consumption in each iteration but, in that case, the connectivity would decrease as more iterations would be needed to achieve consensus [28]. Besides, this paper studies the impact of the network topology on the energy consumption and it concludes that a random deployment is preferable to a regular grid of

sensors.

Some algorithms have also been proposed in the context of target tracking with a WSN [32–34]. In [34], for instance, the authors consider a linear dynamical system and propose a bandwidth-constrained distributed Kalman filter. More precisely, each sensor is only allowed to broadcast the sign of the innovation (1-bit) but, surprisingly, its performance is shown to be close to that of the traditional (i.e. *analog*) Kalman filter.

2.3 Information theory

In this section, we attempt to establish a link between the estimation theory and the information theory. This will be needed in Chapters 4 and 5 where a number of information theoretical approaches will be adopted to encode observations at the sensor nodes. After a reminder of some definitions, in Section 2.3.2, we introduce the concept of *lossless* compression for discrete random variables. Next, in Section 2.3.3 we outline the principles of *lossy* compression. Finally, the source-channel separation theorem is discussed in Section 2.3.4.

2.3.1 Reminder of definitions

Let X, Y be two discrete memoryless sources with *joint* pmf $p_{X,Y}(x)$ and marginal pmf's $p_X(x)$ and $p_Y(y)$, respectively. We introduce the following definitions:

Entropy The entropy of a random variable is defined as follows:

$$H(X) \triangleq -\sum_x p_X(x) \log p_X(x).$$

Joint entropy Likewise, the joint entropy of two random variables X and Y is given by

$$H(X,Y) \triangleq -\sum_{x,y} p_{X,Y}(x,y) \log p_{X,Y}(x,y).$$

Conditional entropy The conditional entropy of X given Y reads

$$H(X|Y) \triangleq -\sum_{x,y} p_{X,Y}(x,y) \log p_{X|Y}(x|y).$$

Mutual information The mutual information of X and Y is defined as follows:

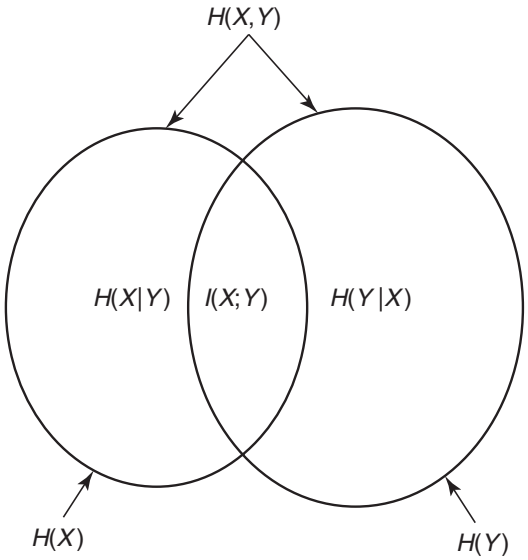


Figure 2.5: Graphical interpretation [35, Chap. 2].

$$\begin{aligned} I(X;Y) &\triangleq \sum_{x,y} p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \\ &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= I(Y;X). \end{aligned} \tag{2.12}$$

In Fig. 2.5, the circles corresponding to $H(X)$ and $H(Y)$ denote the information of X and Y. Likewise, the joint entropy $H(X,Y)$ is the union of the information of X and Y. Therefore, the conditional entropy $H(X|Y)$ denotes the quantity of information of X independent of Y. Finally, the mutual information $I(X;Y)$ is the intersection of the information of X and Y.

2.3.2 Lossless compression

In a lossless compression setting, the source observed at the encoder can be compressed to a finite number of bits and still be almost perfectly reconstructed. Let X be a memoryless discrete source with a pmf $p_X(x)$. For *lossless* compression of X, the average number of bits per sample must satisfy:

$$R_X \geq H(X). \tag{2.13}$$

This compression rate can only be achieved by encoding large blocks of samples. To show that, consider a length-n vector of independent realizations of X, i.e. $\mathbf{x} = x^{(1)}, \ldots, x^{(n)}$ with probability $\Pr(\mathbf{x}) = \prod_{i=1}^{n} p_X(x^{(i)})$. For large n, the total number of *typical* sequences is approximately $2^{nH(X)}$ and all typical sequences are equiprobable [35, Chapter 3]. Consequently, the encoding-decoding process could be as follows:

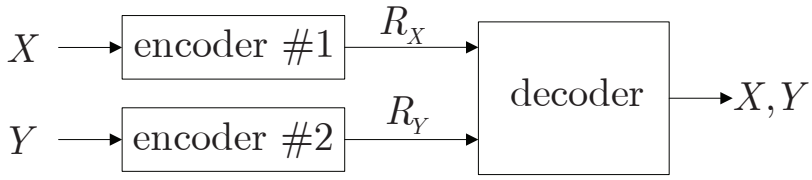


Figure 2.6: Separate encoding of X and Y.

1. **At the encoder**: Randomly generate a codebook $\mathcal{C}_X$ containing all typical sequences, i.e. $2^{nH(X)}$ codewords, and reveal it to the decoder. Each codeword has an associated index denoted by $s \in \left[1, \ldots, 2^{nH(X)}\right]$. Since $\mathbf{x}$ is a typical codeword with high probability, it will be represented with probability close to 1 in $\mathcal{C}_X$. Select the corresponding index s corresponding to codeword $\mathbf{x}$ and send it to the decoder.

2. **At the decoder**: Receive index s. Select the codeword corresponding to the index s in $\mathcal{C}_X$ and obtain $\mathbf{x}$.

Correlated random variables

Typically, sensor observations are correlated. By properly encoding such observations so that redundant information is removed before transmission, substantial energy savings can be achieved. To illustrate that, in this section we review the optimal encoding strategy for two correlated sources.

Let X, Y be two discrete memoryless sources with *joint* pmf $p_{X,Y}(x)$ and marginal pmf's $p_X(x)$ and $p_Y(y)$, respectively. According to the previous result, a rate of $R_{XY} \geq H(X,Y)$ bits per sample suffices to encode a large length-n sequence $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$. On the contrary, if X and Y are observations available at *separate* encoders (sensors), as depicted in Fig. 2.6, by choosing $R_X \geq H(X)$ and $R_Y \geq H(Y)$ we can reconstruct X and Y perfectly at the decoder. However, in the seminal paper of Slepian and Wolf [36], it is shown that $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$ can be perfectly reconstructed at the decoder, if and only if the corresponding rates satisfy the following conditions:

$$R_X \geq H(X|Y) \tag{2.14}$$

$$R_Y \geq H(Y|X) \tag{2.15}$$

$$R_X + R_Y \geq H(X,Y). \tag{2.16}$$

This rate region is depicted in Fig. 2.7. In other words, one can adopt an encoding strategy with a sum rate identical to that of the centralized case, where both sources X and Y are available at the (joint) encoder. For instance, if encoder #1 encodes data at a rate of $R_X \geq H(X)$

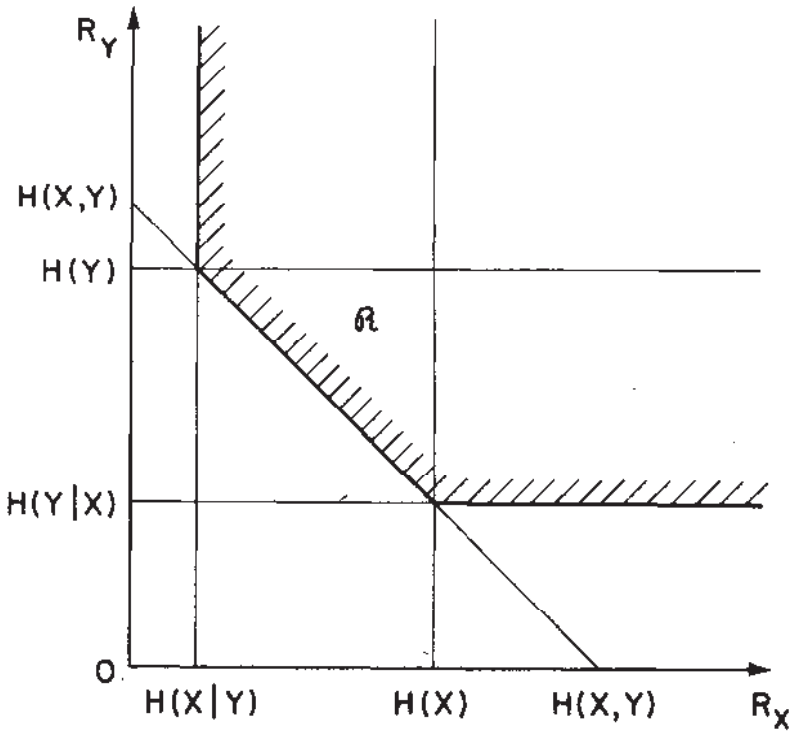


Figure 2.7: Achievable rate region [36].

then encoder #2, can assume that X will be available at the decoder and, thus, encode its observations at a rate $R_Y \geq H(Y|X)$. This corresponds to one of the corner points of the rate region shown in Fig. 2.7. Finally, we outline the corresponding encoding-decoding strategy which allows the system to operate at one of the corner points of the achievable rate region:

1. **At encoder #1**: Randomly generate a codebook $\mathcal{C}_X$ containing all typical sequences, i.e. $2^{nH(X)}$ codewords, and reveal it to the decoder. Each codeword has an associated index denoted as $s_1 \in \left[1, \ldots, 2^{nH(X)}\right]$. Then, look for the codeword which is jointly typical with the length-n source vector $\mathbf{x}$. Since, $\mathbf{x}$ is a typical codeword, it will be represented with probability 1 in $\mathcal{C}_X$. Select the corresponding index s_1 and send it to the decoder.

2. **At encoder #2**: Randomly generate a codebook $\mathcal{C}_Y$ containing all typical sequences, i.e. $2^{nH(Y)}$ codewords, and reveal it to the decoder. Randomly partition the codebook into 2^{nR_Y} *bins* and reveal the partition to the decoder. Next, send the index of the bin $s_2 \in \left[1, \ldots, 2^{nR_Y}\right]$ to which the codeword belongs

3. **At the decoder**: First, receive index s_1 and extract $\mathbf{x}$. To decode $\mathbf{y}$, the decoder looks for the codeword $\mathbf{y}$ which is jointly typical with $\mathbf{x}$ in the bin pointed by index s_2. To prevent from ambiguity, the number of codewords in each bin must be less than $2^{nI(X;Y)}$, which yields $R_Y \geq H(Y|X)$.

It is worth noting that the remaining points of the rate region of Fig. 2.7 can be achieved

through time-sharing.

2.3.3 Lossy compression

In some applications, allowing some distortion in the reconstruction can be acceptable. For instance, in the context of WSNs, one could think of decreasing the amount of transmitted data (and, thus, the energy consumption that it entails) at the price of increasing distortion in the resulting estimate. Besides, for continuous (i.e. *analog*) sources, an infinite number of bits would be needed to achieve zero distortion in the estimates, which is not realistic. For this reason, in subsequent sections, we review some basic results on rate-distortion trade-offs in lossy data compression.

Rate-distortion function

Let $\mathbf{x} = x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ be the set of observations and $\hat{\mathbf{x}} = \hat{x}^{(1)}, \ldots, \hat{x}^{(n)}$ their estimates at the decoder. Then, for a given distortion metric $d(\cdot,\cdot)$ the distortion for large n is given by

$$D = d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n}\sum_{i=1}^{n} d\left(x^{(i)}, \hat{x}^{(i)}\right) \tag{2.17}$$

$$= \mathbb{E}_{X,\hat{X}}\left[d\left(X, \hat{X}\right)\right] \tag{2.18}$$

which follows from the law of large numbers. From [35, Chapter 13], the rate-distortion function can be defined as:

$$R(D) \triangleq \min_{f_{\hat{X}|X}(\hat{x}|x):\mathbb{E}_{X,\hat{X}}\left[d\left(X-\hat{X}\right)\right]\leq D} I\left(X;\hat{X}\right)$$

where the minimization is over all conditional distributions $f_{\hat{X}|X}(\hat{x}|x)$ for which the distortion constraint is satisfied.

Gaussian source: For a zero-mean Gaussian source $X \sim \mathcal{N}(0, \sigma_x^2)$, we have that [35, Chapter 13] (see also Fig. 2.8)

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma_x^2}{D} & 0 \leq D \leq \sigma_x^2 \\ 0 & D > \sigma_x^2 \end{cases}.$$

The encoding-decoding process would be as follows:

1. **At the encoder**: Randomly generate a Gaussian codebook $\mathcal{C}$ containing $2^{nR(D)}$ codewords, and reveal it to the decoder. Each codeword has associated an index denoted as $s \in \left[1, \ldots, 2^{nR(D)}\right]$. Then, look for a codeword $\hat{\mathbf{x}}$ which is *distortion typical* [2] with the length-n source vector $\mathbf{x}$. Select the corresponding index s and send it to the decoder.

[2] The definition for distortion typical can be found in [35, Chapter 13]

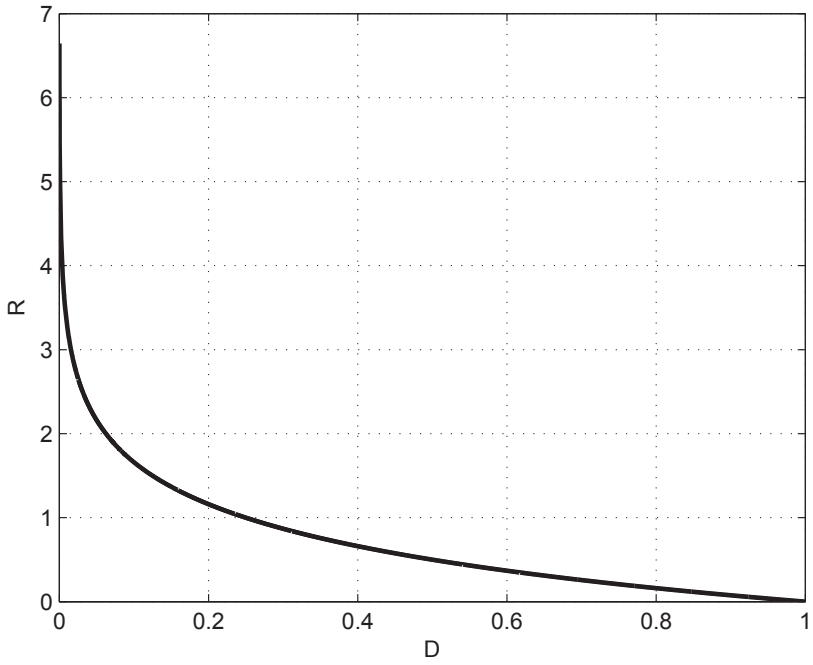


Figure 2.8: Rate-distortion function for a Gaussian source ($\sigma_x^2 = 1$).

2. **At the decoder**: Receive index s. Select the codeword corresponding to the index s in $\mathcal{C}$ and obtain $\hat{\mathbf{x}}$.

Rate-distortion function with side information at the decoder

In this section, we ask ourselves about the impact of having as side information at the decoder some random variable Y which is correlated with X. To that aim, let X, Y be two continuous memoryless sources with joint probability density function $f_{X,Y}(x, y)$ and marginal pdf's denoted by $f_X(x)$ and $f_Y(y)$, respectively. From [37], the rate-distortion function with side information Y at the decoder reads

$$R_Y(D) \triangleq \min_{f_{W|X}(w|x), g:\mathbb{E}_{X,W,Y}[d(x,g(y,w))] \leq D} \left(I\left(X;W\right) - I\left(Y;W\right)\right)$$

where W stands for an auxiliary random variable denoting the encoded version of X. In the next paragraphs, we outline the encoding-decoding strategy where we assume both the probability density function $f_{W|X}$ and the reconstruction function g to be known.

1. **At the encoder**: From $f_{W|X}(w|x)$, compute $f(w) = \int f_X(x) f_{W|X}(w|x) dx$. Then randomly generate a codebook $\mathcal{C}$ containing 2^{nR_1} codewords $\mathbf{w}(s) \sim \prod_{i=1}^{n} f_W(w^{(i)})$ indexed by $s \in 1, \ldots, 2^{nR_1}$ with $R_1 = I\left(X;W\right)$. Randomly partition the codebook into

2^{nR} *bins*. Next, look for the codeword $\mathbf{w}$ which is jointly typical with the source vector $\mathbf{x}$ and send the index of the bin where the codeword belongs to.

2. **At the decoder**: First, receive the index of the bin where the codeword $\mathbf{w}$ belongs to. From this, select the codeword which is jointly typical with the side information given by $\mathbf{y}$. To prevent from ambiguity and ensure that the only jointly typical codeword with $\mathbf{y}$ is the intended transmitted $\mathbf{w}$, the number of codewords in each bin must be less than $2^{nI(W;Y)}$, which leads to $R \geq I(X;W) - I(Y;W)$. Finally, compute the per sample estimate, i.e. $\hat{x}^{(1)}, \ldots, \hat{x}^{(n)} = g(w^{(1)}, y^{(1)}), \ldots, g(w^{(n)}, y^{(n)})$, with average distortion D.

It is worth noting that this problem is similar to that of lossless compression with correlated sources. Unfortunately, the extension of the setting of Fig. 2.6 for a lossy compression scenario continues to be an open problem, and only some problems of interest have been fully characterized (e.g. the quadratic Gaussian CEO problem [38]).

2.3.4 Source-channel coding separation principle

In a sensor network, sensor nodes not only have to compress the collected samples but also they have to transmit them over a noisy channel to the FC. From [35, Chapter 8], in point-to-point communications, source channel separation is optimal. More precisely, a discrete source can be perfectly reconstructed at the decoder if the following inequality is satisfied

$$nH(X) \leq mC, \tag{2.19}$$

where, in the above expression, C denotes the capacity (in bits per channel use) of a memoryless channel characterized by $f(y|z)$ (see Fig. 2.9), and $\frac{m}{n}$ denotes the ratio of channel uses per source sample. The encoding/decoding process is as follows:

- **At the encoder:** First, the n samples of the source X are encoded and represented by an index s which, as commented in Section 2.3.2, $s \in 1, \ldots, 2^{nH(x)}$. The index is used as an input for the channel coding stage. The channel codebook consists of at most 2^{mC} codewords. A one-to-one mapping of each source codeword into a channel codeword exists if $nH(X) \leq mC$. Finally, the channel codeword corresponding to index s is transmitted to the decoder.

- **At the decoder:** The decoder receives $\mathbf{z}$ (see Fig. 2.9) and, since the encoder is transmitting at the maximum rate which can be reliably supported by the channel, i.e. C, the transmitted codeword $\mathbf{y}$ (see Fig. 2.9) is decoded without errors. Next, the channel decoder propagates the index of the transmitted codeword to the source decoding stage. Finally, the decoder looks for the source codeword associated to index s and obtains $\mathbf{x}$.

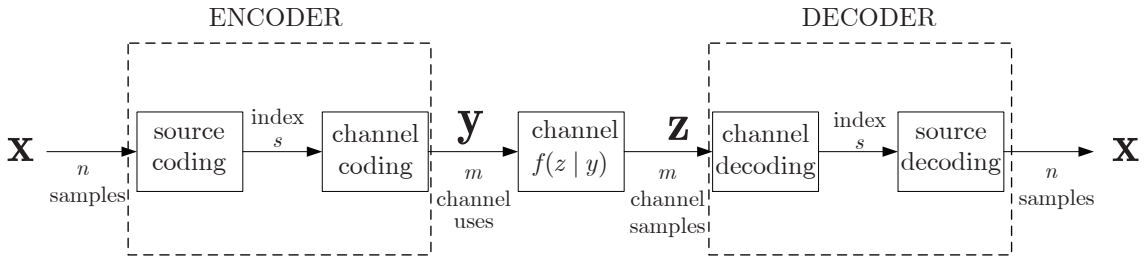


Figure 2.9: Separate source and channel coding.

Clearly, the fact that the source and channel coding can be treated (and *optimally* solved) as independent problems leads to a high degree of modularity in the implementation of communication systems.

Unfortunately, this optimality does not hold for multi-terminal settings such the Chief Executive Officer (CEO) problem of [39]. In the quadratic Gaussian CEO problem, N sensors/terminals observe a common source of interest x embedded into (independent) Gaussian noise n_i ; $i = 1, \ldots, N$. Sensors encode their observations for transmission over a multiple-access channel. The destination, or fusion center, receives the data and produces an estimate of x, that is, $\hat{x}$. For this setting, the separation of source and channel coding was shown to be suboptimal for asymptotically large WSNs [12]. To that aim, the authors proved that for Amplify-and-Forward (A&F) strategies, where sensors transmit scaled versions of their observations, the distortion decreases in the number of sensor nodes as in the centralized case, that is

$$D_{\mathsf{A\&F}} \sim \frac{1}{N},$$

whereas in a system where source and channel coding is carried out separately,

$$D_{\mathsf{sep}} \sim \frac{1}{\log N}.$$

Still, such optimality can only be achieved if *all* the A&F sensors can be fully synchronized (which is difficult to achieve in practical scenarios).

2.4 Multi-user diversity and opportunistic communications

One intrinsic characteristic of wireless channels is the fluctuation of the channel strength due to constructive and destructive interference. This fluctuation, known as *fading*, can be combated by creating a number of independent paths between the transmitter and the receiver through time, space or frequency diversity. Besides, in multi-terminal networks one can also exploit the so-called *Multi-User* Diversity (MUD).

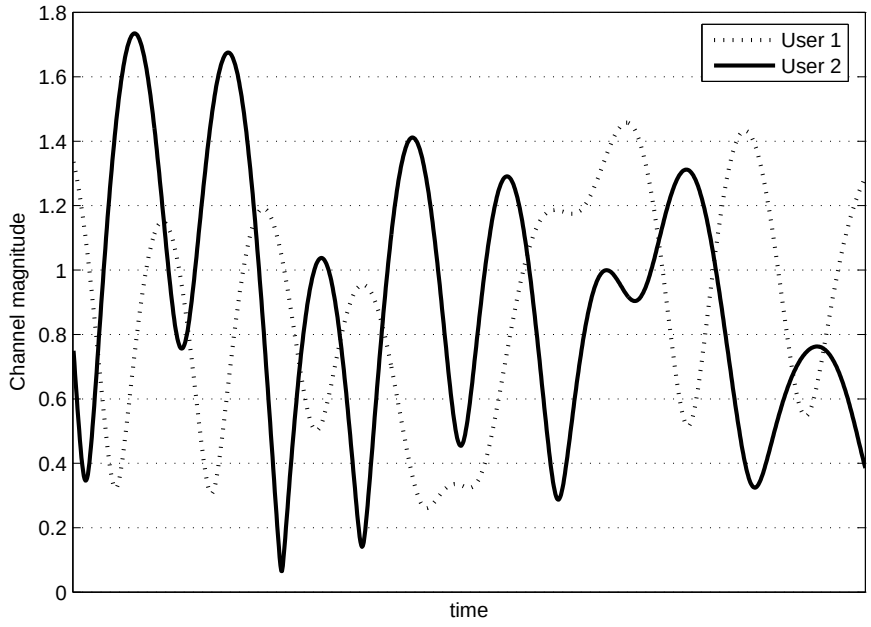


Figure 2.10: Channel fluctuations for two different users.

2.4.1 Opportunistic communications in wireless data networks

Multi-user diversity is the result of having a large population of users with independent fading conditions. In their seminal work, Knopp and Humblet [40] established the roots of opportunistic communications. Their work showed that in the uplink of single-antenna multi-user networks, the sum-rate under a sum-power constraint can be maximized by granting access to the user experiencing the most favorable channel conditions (see also [41]). Similar results were derived for the parallel broadcast (i.e. downlink) channel in [42]. In Figure 2.10, we depict the channel magnitude for two different users in the uplink. In this example, diversity appears in two dimensions: time and users. Here, one can exploit multi-user diversity by selecting at each time instant the user experiencing the most favorable channel condition to the Base Station (BS). Clearly, by increasing the number of terminals (N), the probability of having a user with a stronger channel gain increases too.

With independent and identical fading conditions, opportunistic approaches exhibit long-term fairness since, on *average*, each user is scheduled the same number of times. Conversely, if the fading coefficients are *non-identically* distributed these strategies become unfair. In the WSN context, this could entail, for instance, that sensors closer to the FC would die earlier, which is not desirable. To avoid that, one can resort to Proportional Fair Scheduling (PFS) strategies where the metric for the user selection is the *accumulated* throughput in a sliding observation window which ensures short-term fairness [43, 44]. It is worth noting that all these strategies assume that channels are fast-fading. For slow-fading scenarios, one can induce

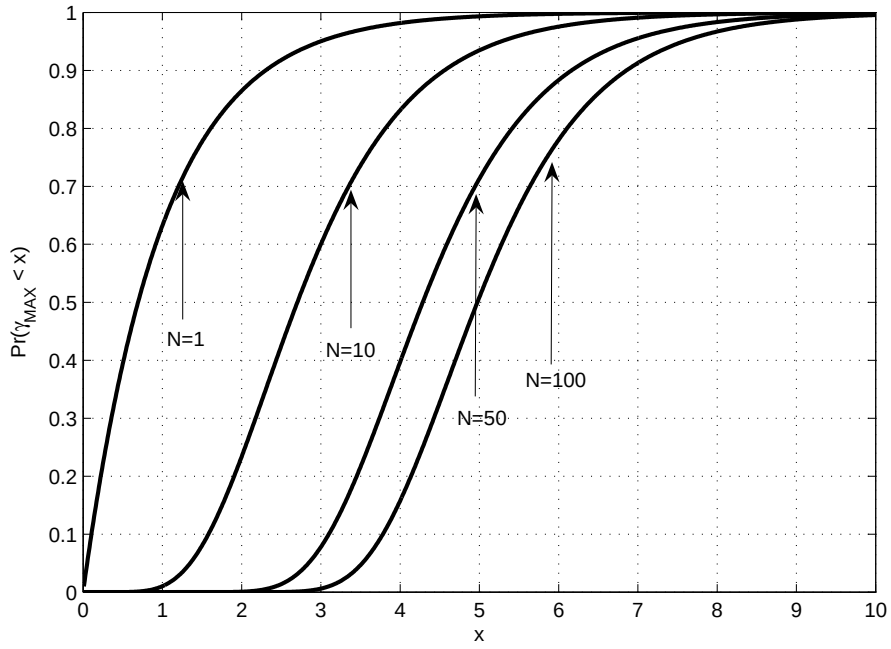


Figure 2.11: CDF of the strongest channel gain for different number of users for Rayleigh-fading channels.

pseudo-random fading by adopting the approach of [43].

The major drawback of all works cited above is the need for *global* and *perfect* CSI at the Base Station (BS). For this reason, [45,46] analyze the impact of delayed and noisy CSI estimates on multi-user diversity. To alleviate the need for global CSI, the authors in [47] proposed a simple thresholding strategy, by which only those users with channel gains above a given threshold report them to the BS. In the literature, this strategy is known as Selective Multi-User Diversity (SMUD). By doing so, the load in the feedback channel decreases at the expense of a small loss in terms of sum-rate. This follows from the fact that there exists an outage scheduling probability for which no user reports its CSI to the BS. In this situation, the BS randomly schedules one of the users.

However, in the previous algorithm analog feedback is still required. The case of quantized feedback is considered in [48], where merely 1 bit of feedback suffices to capture the optimal growth in capacity for an increasing number of users. That is, for large N the capacity scales as $C \sim \log\log N$. Similar results are obtained in [49] for multi-user MIMO settings.

An opportunistic variation of the well-known ALOHA protocol [50,51] is introduced in [52] by which the scheduling decision made by the *terminals* are on the basis of *local* CSI only. Clearly, this scheduling protocol suffers from packet collisions but, still, it is shown to be asymptotically optimal and to achieve the same capacity growth rate as a *centralized* scheduler. More precisely, the ratio of throughputs for the opportunistic ALOHA and the *centralized* schedulers is shown

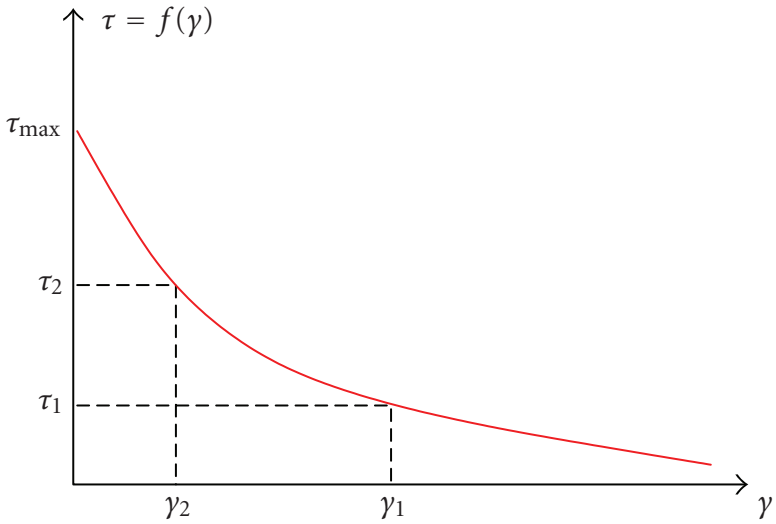


Figure 2.12: Opportunistic carrier sensing of [54].

to be $1/e$ for a large N. The reader is referred to [53] for the the case that the receiver can handle multiple packet reception. .

2.4.2 Opportunistic schemes in wireless sensor networks

Although, the aforementioned strategies were derived in the context of wireless *data* networks, opportunistic schemes are also suitable for wireless *sensor* networks. For instance, in a WSN with a large population of sensors and a fixed communication rate, one can schedule each time instant the sensor for which the transmission would result in the lowest energy consumption or, alternatively, the one with the larger residual energy.

In [54, 55], the authors proposed an opportunistic backoff strategy where sensors choose their backoff periods by mapping their corresponding channel strength onto a common backoff function. The backoff function is aimed at minimizing the energy consumption and, hence, it prioritizes the sensors with the most favorable channel conditions by assigning them the shorter backoff times. For instance, for two sensor nodes with channel gains γ_1 and γ_2 with $\gamma_1 > \gamma_2$, sensors select τ_1 and τ_2 as their respective backoff times according to Fig. 2.12. Therefore, the sensor node with the strongest channel gain γ_1 is the one actually scheduled in a *distributed* fashion to transmit its information, since $\tau_1 < \tau_2$ and the second sensor will not transmit.

Opportunistic communications can also be useful for the enhancement of network lifetime [7, 56, 57]. The definition of the Network Lifetime [58] is application dependent but, for simplicity and mathematical tractability, is typically considered as the time elapsed until one sensor runs out of energy. The work in [7] considers the sensor scheduling problem with different levels

of information, namely, CSI, Residual Energy Information (REI) and both. The conclusion is that one should simultaneously use, REI and CSI to maximize the network lifetime. The idea behind that is to schedule sensors experiencing the most favorable channel conditions when the network is young and sensors with higher residual energies when the network grows older [59].

Chapter 3

Opportunistic Power Allocation Schemes for Wireless Sensor Networks

In this chapter, the focus of our study is the analysis, in terms of complexity and CSI requirements, of different power allocation strategies for decentralized parameter estimation via WSNs. First, we propose and analyze a class of Opportunistic Power Allocation (OPA) schemes. In all cases, only sensors experiencing favorable conditions (e.g. with channel gains above a threshold) participate in the estimation process by adjusting their transmit power on the basis of local Channel State Information (CSI) and, in some cases, Residual Energy Information (REI). Interestingly, the signaling and CSI requirements associated with the OPA schemes are substantially lower than those of the optimal (i.e. waterfilling-like) approaches, which demand global CSI information in analog form and, still, their performance is virtually identical. Next, for situations in which sensors are situated at a large distances from the FC, we adopt a hierarchical topology where sensors are grouped into clusters. In each cluster, a cluster-head is in charge of processing and sending a cluster estimate to the FC. For this network topology, we carry out an exhaustive performance assessment of different power allocation schemes. Throughout the chapter, the proposed strategies are compared in terms of distortion and CSI requirements.

3.1 Introduction

The source-channel coding separation theorem by which source and channel coding can be regarded as decoupled problems and thus be solved independently [35, Ch. 8], turns out to provide suboptimal solutions in the case of *Multiple Access* channels (MAC) with correlated sources [12]. Conversely, an amplify-and-forward (A&F) strategy is known to scale optimally in terms of estimation distortion, when the number of users grows without bound. However, such asymptotic optimality is achieved if distributed synchronization of the sensor signals can be orchestrated at the physical layer in order to achieve beamforming gains. In the more realistic case of orthogonal sensors-to-FC channels, the authors in [11] derived the optimal power allocation for two different problems of interest: *i*) the minimization of distortion subject to a sum-power constraint, and *ii*) the minimization of transmit power subject to a maximum distortion target. In both cases, the optimal power allocation is given by a kind of water-filling solution (referred to in the sequel as WF-D and WF-P) in which sensors with poor channel gains or noisy observations should remain inactive to save power. This finding builds a bridge between opportunistic communications (originally addressed in a wireless *data* network context for the multiple-access [40] and broadcast [42] channels, respectively) and the problem of decentralized parameter estimation with wireless *sensor* networks.

The main drawbacks of [11, 40, 42] are *i*) the need for *global* (namely, the terminal-to-BS channel gains for *all* the terminals in the network) and *instantaneous* CSI at the Base Station or Fusion Center; and *ii*) the computational complexity that water-filling solutions entail. Concerning CSI requirements, they can be alleviated by resorting to thresholding rules, e.g. [47], by which only terminals with channel gains above a predefined threshold are allowed to feed back information to the BS. By doing so, the signaling load decreases at the expense of a very moderate performance loss [47]. Going one step beyond, [48] proved that, for an asymptotically high number of terminals, just one bit of feedback (instead of *analog*) per terminal suffices to capture the optimal capacity growth-rate of capacity. As for the high computational burden that water-filling solutions entail, it is addressed in [60] by assuming that power is evenly allocated over a subset of terminals. This results in a simplified water-filling scheme from which the subset of active users can be easily determined.

Notwithstanding, not only energy efficiency but also network lifetime is of interest in WSNs. The definition of the network lifetime (LT), namely, the amount of time for which the network is operational, is clearly application-dependent. However, for simplicity and mathematical tractability, one typically defines network LT as the time elapsed until one sensor runs out of energy. In recent works [7], the authors show how a sensible use at the scheduler of Residual Energy Information (REI) in combination with CSI information is key to extend network LT.

3.1.1 Contribution

In this chapter, we propose and analyze a class of Opportunistic Power Allocation (OPA) schemes suitable for decentralized parameter estimation with WSNs. We adopt the amplify-and-forward technique proposed in [61] [11] and convey sensor observations to the FC through a set of orthogonal channels. Inspired by [47] [49], all the OPA schemes proposed here have one feature in common: only sensors experiencing certain local conditions (i.e. channel gain and/or residual energy above a threshold) are allowed to participate in the estimation process. This strategy is aimed at retaining as much performance as possible of the corresponding *optimal* power allocation scheme while keeping network signalling and energy consumption under control. More precisely, the proposed opportunistic schemes merely require *i*) the sensor-to-FC channel gains of the *subset* of active nodes plus some *statistical* CSI[1] at the FC (in [11, 60] the channel gains of *all* sensor nodes are needed), *ii*) *one* bit of feedback per sensor (instead of analog signaling as in [47] or [11]); and *iii*) *local* CSI and, possibly REI, at each sensor node. In particular, we derive opportunistic power allocation schemes for the following optimization problems:

1. Minimization of *distortion* (OPA-D)
2. Minimization of transmit *power* (OPA-P)
3. Enhancement of network *lifetime* (OPA-LT)

We also address the case in which the local channel state information available in the sensor nodes is subject to impairments (e.g. noisy or delayed CSI estimates). For brevity, we focus on deriving an improved version of the OPA-D scheme, referred to in the sequel as OPA-DR, which is *robust* to such imperfect CSI estimates. However, the extensions to the OPA-P and OPA-LT schemes are relatively straightforward, as well. For all the above-mentioned cases, we obtain closed-form expressions of the global reporting threshold (only numerical methods are used in [60] to compute the optimal cut-off point which, in turn, determines the subset of active nodes), and we derive the associated power allocation rule on the basis of *local* CSI only.

Next, we adopt a *hierarchical* topology which is suitable for scenarios with severe path loss in the sensor-to-FC channels. Here, sensors are grouped into clusters where a cluster-head acts as a local fusion center and consolidates the data gathered in the cluster. The cluster-heads are coordinated by the Fusion Center where the final estimation is obtained. Unlike in previous works [62], our goal is to *estimate* a parameter and, to that aim, we explicitly consider the impact of the network topology on the attainable accuracy. By doing so, and unlike [11], we can take advantage of the intra-cluster channel gains. We also show that balancing the available power between the sensors and the cluster-heads is of paramount importance and, in

[1] In some cases, REI information is also needed

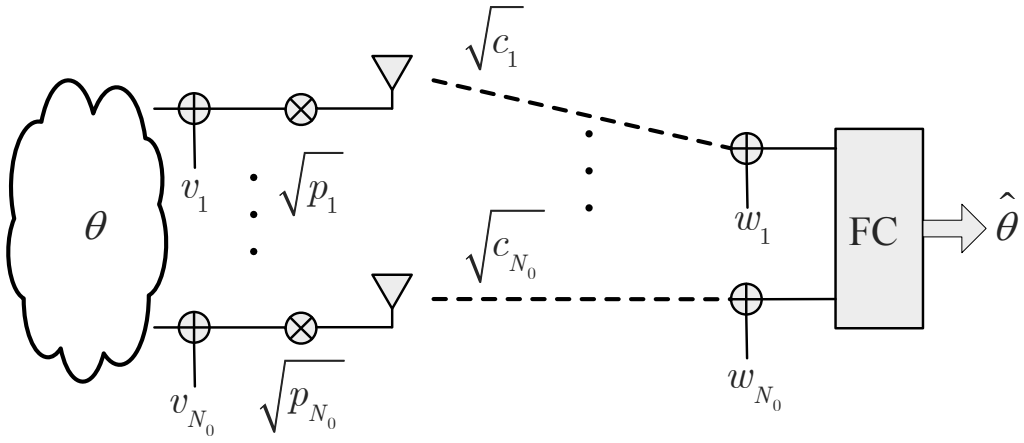


Figure 3.1: System model.

particular, we derive the optimal fraction of power dedicated to each subset for the Uniform Power Allocation (UPA) case. Last, we discuss some hybrid solutions which combine UPA and WF power allocation schemes at the sensor and cluster-head levels.

The contents of this chapter have been partly published in references [63–68].

The chapter is organized as follows. First, in Section 3.2, we present the signal model. For completeness, we review the optimal power allocation strategies in Section 3.3. Next, in Section 3.4, we introduce the proposed opportunistic power allocation strategy and the associated communication protocol in a general framework. In Section 3.5, we particularize the algorithm to the problem of the minimization of distortion and derive the corresponding reporting threshold and power allocation rule. In Sections 3.6 and 3.7, we focus our attention on the transmit power and network lifetime enhancement problems, respectively. Next, in Section 3.8 we present some additional results for a hierarchical network topology. Finally, we close the chapter by summarizing the main findings in Section 3.9.

3.2 Signal model

Consider a WSN composed of one Fusion Center (FC) and a large population of N_o energy-constrained sensors which have been deployed to estimate an unknown scalar, slowly-varying and spatially-homogeneous parameter θ. The observation at sensor i can be expressed as

$$x_i = \theta + v_i \quad ; \quad i = 1, \ldots, N_o. \tag{3.1}$$

where v_i denotes AWGN noise of variance σ_v^2 (i.e. $v_i \sim \mathcal{CN}(0, \sigma_v^2)$). We adopt an amplify-and-forward re-transmission strategy and, consequently, the observation at each sensor is scaled by a factor $\sqrt{p_i}$ before transmission. Hence, the received signal at the FC (see Fig. 3.1) can be

modeled as[2]:

$$y_i = \sqrt{p_i}\sqrt{c_i}\,(\theta + v_i) + w_i = \sqrt{p_i c_i}\theta + \sqrt{p_i c_i}v_i + w_i \quad ; \quad i = 1, \ldots, N_o, \tag{3.2}$$

where w_i stands for the i.i.d. AWGN noise (i.e. $w \sim \mathcal{CN}\,(0, \sigma_w^2)$) and c_i denotes the channel *power* gain. For non-frequency selective block Rayleigh-fading channels, c_i turns out to be an exponentially-distributed random variable of mean μ_c, that is,

$$f_c(x) = \frac{1}{\mu_c} e^{-\frac{x}{\mu_c}}, \tag{3.3}$$

which is assumed to be independent and identically distributed (i.i.d) over sensors. In each time-slot, only a subset of $N \leq N_o$ *active* sensors transmit their observations to the FC over a set of orthogonal channels (e.g. FDMA). Consequently, the $N \times 1$ received signal vector $\mathbf{y}$ reads

$$\mathbf{y} = \mathbf{h}\theta + \mathbf{z}, \tag{3.4}$$

with $\mathbf{h} = \left[\sqrt{p_1 c_1}, \ldots, \sqrt{p_N c_N}\right]^T$ and with $\mathbf{z}$ standing for AWGN with (diagonal) covariance matrix $\mathbf{C}$ given by $\mathrm{diag}\,[\mathbf{C}] = [p_1 c_1 \sigma_v^2 + \sigma_w^2, \ldots, p_N c_N \sigma_v^2 + \sigma_w^2]^T$. In an attempt to make our estimator simple and universal (i.e. independent of any particular distribution of the noise), we adopt the Best Linear Unbiased Estimator (BLUE) [18, Ch. 6]. The estimate at the FC is thus given by

$$\hat{\theta} = \left(\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}\right)^{-1} \mathbf{h}^T \mathbf{C}^{-1} \mathbf{y}. \tag{3.5}$$

This estimator is known to be efficient (and, of course, unbiased) for the linear signal model described above and, hence, we can adopt the variance as a distortion measure D:

$$D = \mathsf{Var}(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = \left(\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}\right)^{-1}. \tag{3.6}$$

Since matrix $\mathbf{C}$ is diagonal, the above equation can be written as

$$D = \mathsf{Var}(\hat{\theta}) = \left(\sum_{i=1}^{N} \frac{p_i c_i}{p_i c_i \sigma_v^2 + \sigma_w^2}\right)^{-1}, \tag{3.7}$$

from which it becomes apparent that the actual distortion depends on the power allocation strategy *and* the number of active sensors N.

3.3 Optimal power allocation strategies

In this section, we review the *optimal* power allocation strategy derived by Cui *et al.* in [11]. More precisely, the authors addressed two problems of interest, namely, *i*) the minimization of distortion for a given sum-power constraint and, *ii*) the minimization of the transmit power for a given distortion target.

[2]Implicitly, we also assume pair-wise synchronization between each sensor node and the FC.

3.3.1 Minimization of distortion

The power allocation rule that minimizes the distortion for a given sum-power constraint is given by the solution to the following problem:

$$\begin{aligned} \min_{p_1,\ldots,p_{N_o}} \quad & \left(\sum_{i=1}^{N_o} \frac{p_i c_i}{p_i c_i \sigma_v^2 + \sigma_w^2} \right)^{-1} \\ \text{s.t.} \quad & (\sigma_v^2 + U^2) \sum_{i=1}^{N_o} p_i \leq P'_{\mathrm{T}} \end{aligned} \tag{3.8}$$

where P'_{T} stands for the total transmit power and $\{-U \ldots U\}$ denotes the dynamic range of the sensors[3]. From [11], it is given by the following waterfilling-like (WF-D) solution:

$$p_i^* = \frac{\sigma_w^2}{\sigma_v^2 c_i} \left[\frac{\sqrt{c_i}}{\sqrt{\lambda_0}\sigma_w} - 1 \right]^+ \quad ; \quad i = 1, \ldots, N_o. \tag{3.9}$$

In this last expression, the operator $[x]^+$ is defined as $[x]^+ = \max\{x, 0\}$ and λ_0 denotes the optimal water-level which is computed at the FC from $c_i \, ; i = 1 \ldots N_o$ in order to meet the sum-power constraint. Clearly, only sensors with strong channels to the FC will be allocated positive power ($p_i > 0$) and, thus, will become part of the subset of N active nodes. However, the price to be paid for the optimality of such solution is two-fold: *i*) the need for *global* CSI at the FC (the whole set of channel gains); and *ii*) the need for the FC to inform the sensor nodes, on a *frame-by-frame* basis, about the optimal water-level. This unavoidably entails an extensive signalling between the FC and the sensor nodes and, ultimately, an increased energy consumption (which is barely desirable in WSNs).

When no CSI is available at the FC or in the absence of signalling channels between the FC and the sensors, one can alternatively resort to a Uniform Power Allocation (UPA) rule. In this case, *all* the sensors remain active (regardless of their channel conditions) and evenly allocate transmit power according to

$$p_i = \frac{P_{\mathrm{T}}}{N_o}; \quad i = 1, \ldots, N_o. \tag{3.10}$$

Reasonably, a substantial performance gap can be expected between the WF (optimal) and UPA strategies in many scenarios.

[3]For the ease of notation, in the sequel we re-define $P_{\mathrm{T}} = P'_{\mathrm{T}}/(\sigma_v^2 + U^2)$.

3.3.2 Minimization of transmit power

From [11], the power allocation rule that minimizes the total transmit power under a prescribed distortion target D_{T}, i.e.

$$\begin{aligned} \min_{p_1,\ldots,p_{N_o}} \quad & \sum_{i=1}^{N_o} p_i \\ \text{s.t.} \quad & D \leq D_{\mathrm{T}} \end{aligned} \tag{3.11}$$

is given by the following waterfilling-like (WF-P) solution:

$$p_i^* = \frac{\sigma_w^2}{\sigma_v^2 c_i}\left[\frac{\sqrt{c_i\lambda_0}}{\sigma_w} - 1\right]^+ \quad ; \quad i = 1, \ldots, N_o. \tag{3.12}$$

Again, λ_0 denotes the optimal water-level which is computed at the FC from $c_i\,; i = 1 \ldots N_o$ in order to meet the sum-power constraint. Clearly, only sensors experiencing high gains in the sensors-to-FC channels will be allocated non-zero power ($p_i > 0$) and, thus, will become part of the subset of N active nodes. As in WF-D, the drawbacks are again the need to obtain *global* CSI at the FC and the need for the FC to report, on a *frame-by-frame* basis, about the optimal water-level.

3.4 Opportunistic power allocation: general framework

In an attempt to keep signalling as low as possible while retaining part of the optimality of the water-filling solution, we propose a novel Opportunistic Power Allocation (OPA) strategy. Before particularizing OPA to a number of problems of interest (minimization of distortion, or transmit power, or enhancement of network lifetime), we briefly describe the corresponding communication protocol in a general framework, and discuss the associated CSI requirements.

3.4.1 Communication protocol

The Opportunistic Power Allocation (OPA) schemes operate according to the following communication protocol:

1. **Initialization**: Compute and broadcast the reporting threshold γ_{th}. This threshold ultimately depends on the design criterion: minimization of the transmit power, maximization of the overall distortion, or the enhancement of network lifetime.

2. **Identification of the subset of active sensors**: Each sensor node notifies the FC whether it will actually participate in the estimation process or not (see Fig. 3.2). Only sensors

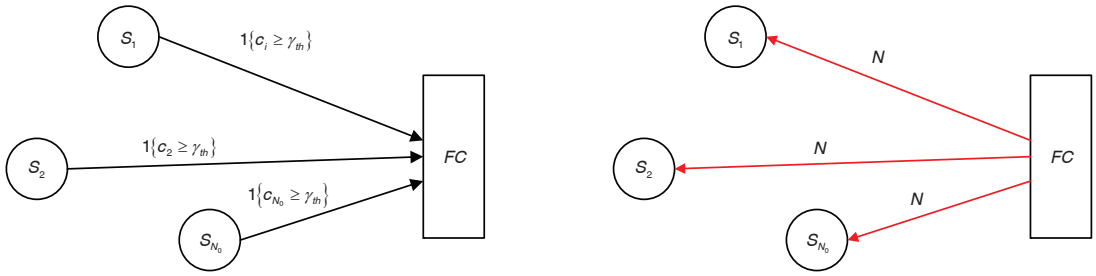


Figure 3.2: Identification of the subset of active sensors: sensors notify the FC about their intention to participate (left) and the FC informs about the number of active sensors (right).

above the threshold will participate. The *number* of active sensors in each timeslot s, $N = N[s]$, is then broadcasted by the FC (see Fig. 3.2).

3. **Power Allocation and Transmission**: The N active sensor nodes adjust their transmit power accordingly and send their observations to the FC[4].

4. **Go to Step 2**

3.4.2 CSI requirements

Prior to further formalizing the algorithms, we will briefly summarize the signalling and CSI requirements associated with this protocol.

- **At the Fusion Center**: As will be shown in subsequent sections, only *statistical* CSI (and, in some cases, REI) is needed in order to compute the closed-form expressions of the reporting threshold in Step #1. The channel gains of the *subset of active nodes* are also necessary to estimate the underlying parameter θ according to (3.5), whereas in [11, 60] all the channel gains must be known to the FC. As illustrated in Section 3.5.3, the average number of active nodes is on the order of 10-20% of the whole population. Consequently, the savings in terms of signalling and energy consumption are potentially very high.

- **At the sensor nodes**: Each sensor must be aware of its *own* channel gain[5] (i.e. *local* Channel State Information) and, possibly, REI in order to *i*) determine whether it belongs to the subset of active nodes (Step #2); and *ii*) adjust its transmit power accordingly (Step #3). Besides, the number of active sensors in each timeslot must also be broadcasted by the FC.

[4]The task of scheduling active sensors on orthogonal channels is delegated to the MAC layer and, therefore, is out of the scope of this work.

[5]To that extent, a training sequence could be sent by the FC at the beginning of each timeslot. However, most of the energy consumption here is restricted to the transmitter (the FC) rather than the receiver (the energy-constrained sensor node).

Finally, one signalling bit is needed for each sensor to indicate to the FC whether it belongs or not to the subset of active nodes in the current time-slot (Step #2).

Interestingly, in a waterfilling-like solution the computational complexity at the FC is a consequence of the sorting algorithm. The computational complexity of the best sorting algorithm is $\mathcal{O}\left(N_o \log(N_o)\right)$, whereas for *all* OPA schemes, the only operation carried out at the FC is a sum in order to obtain the total of number active sensor nodes (see step #2 in Section 3.4.1).

3.5 OPA for the minimization of distortion (OPA-D)

Here, we attempt to find a threshold $\gamma_{\rm th}$ that minimizes the expected distortion (w.r.t. the channel realizations and the number of active sensors) subject to a sum-power constraint:

$$\begin{aligned} \gamma_{\rm th}^{*} \quad &= \quad \arg\min_{\gamma_{\rm th}} \left\{ \mathbb{E}_{\{c_i\}_{i=1}^{N},N;\gamma_{\rm th}} \left[\left(\sum_{i=1}^{N} \frac{p_i c_i}{p_i c_i \sigma_v^2 + \sigma_w^2} \right)^{-1} \right] \right\} \\ &\text{s.t.} \qquad \sum_{i=1}^{N_0} p_i \leq P_{\rm T}. \end{aligned} \tag{3.13}$$

We propose to uniformly allocate the available transmit power among the set of *active* sensors only, namely

$$p_i = \begin{cases} \frac{P_{\rm T}}{N} & \text{if } c_i > \gamma_{\rm th} \quad ; \quad i = 1, \ldots, N_o \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

since, in this way, we avoid wasting resources in sensors experiencing non-favorable channel conditions (e.g. as occurs in UPA schemes, where *all* sensors transmit with identical power levels). From Figure 3.3, the idea behind the OPA-D scheme is to mimic the optimal sensor *selection* of the waterfilling-like solution but, differently from (3.8), the transmit power for each sensor node (after selection) is selected regardless its channel gain. Accordingly, the OPA-D strategy retains:

1. The simple power allocation of the uniform power allocation.
2. Some of the optimality of the WF-D solution by only activating those sensors experiencing favorable channel conditions.

In these conditions, the optimal threshold $\gamma_{\rm th}^{*}$ can be found by solving the following optimization problem:

$$\gamma_{\rm th}^{*} = \arg\min_{\gamma_{\rm th}} \left\{ \mathbb{E}_{N;\gamma_{\rm th}} \left[\mathbb{E}_{\{c_i\}_{i=1}^{N}|N;\gamma_{\rm th}} \left[\left(\sum_{i=1}^{N} \frac{\frac{P_{\rm T}}{N} c_i}{\frac{P_{\rm T}}{N} c_i \sigma_v^2 + \sigma_w^2} \right)^{-1} \right] \right] \right\}. \tag{3.15}$$

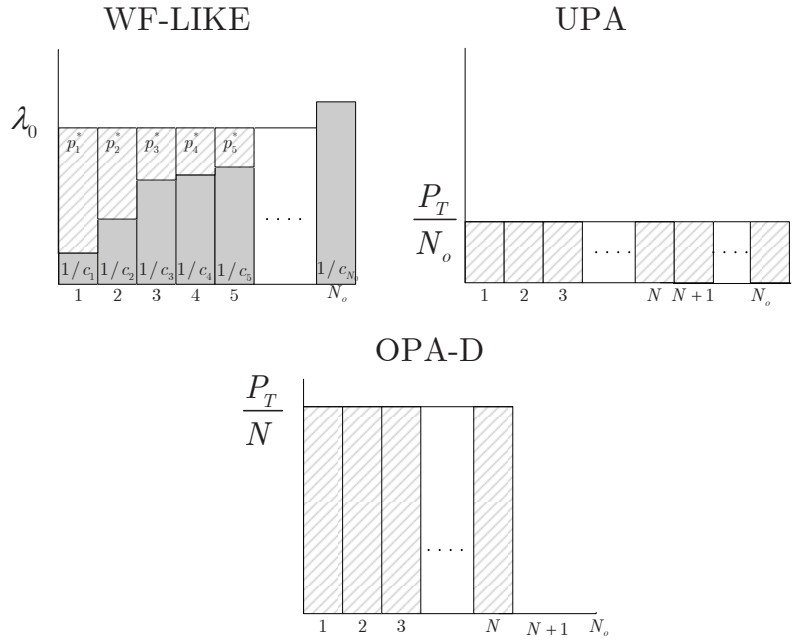


Figure 3.3: Graphical interpretation of the OPA-D strategy.

Unfortunately, this last expression is barely tractable. Instead, we find a lower bound of the argument in (3.15) which entails the use of the joint pdf of the random variables $\{c_i\}_{i=1}^N|N;\gamma_{\text{th}}$ (or $\{c_i\}_{i=1}^N;\gamma_{\text{th}}$ in short); and the pmf of $N;\gamma_{\text{th}}$ that will be derived next. Since $\{c_i\}_{i=1}^N;\gamma_{\text{th}}$ are i.i.d. random variables, it suffices to find the pdf of the marginal truncated random variable $c_i;\gamma_{\text{th}}$. One can easily prove that:

$$f_{c_i;\gamma_{\text{th}}}(x) = \frac{f_{c_i}(x)}{1-F_{c_i}(\gamma_{\text{th}})} = \frac{e^{\frac{\gamma_{\text{th}}}{\mu_c}}}{\mu_c} e^{-\frac{x}{\mu_c}}, \quad x \in [\gamma_{\text{th}}, \infty), \tag{3.16}$$

where $F_{c_i}(\cdot)$ denotes the CDF function [6] of the r.v. c_i. Besides, for each truncated r.v. we have that $\mathbb{E}_{c_i;\gamma_{\text{th}}}[x] = \int_{\gamma_{\text{th}}}^{\infty} x f_{c_i;\gamma_{\text{th}}}(x) = \mu_c + \gamma_{\text{th}}$. Concerning $N;\gamma_{\text{th}}$, it clearly follows a binomial distribution:

$$\Pr\{N = n;\gamma_{\text{th}}\} = \binom{N_o}{n} p^n (1-p)^{N_o-n}. \tag{3.17}$$

with individual probability of activation given by $p = 1 - F_{c_i}(\gamma_{\text{th}}) = e^{-\frac{\gamma_{\text{th}}}{\mu_c}}$. Bearing all the above in mind, expression (3.15) can be lower-bounded as follows,

$$\begin{aligned} \mathbb{E}_{N;\gamma_{\text{th}}} & \left[\mathbb{E}_{\{c_i\}_{i=1}^N|N;\gamma_{\text{th}}} \left[\left(\sum_{i=1}^{N} \frac{\frac{P_{\text{T}}}{N} c_i}{\frac{P_{\text{T}}}{N} c_i \sigma_v^2 + \sigma_w^2} \right)^{-1} \right] \right] \\ \geq & \quad \mathbb{E}_{N;\gamma_{\text{th}}} \left[\left(\mathbb{E}_{\{c_i\}_{i=1}^N;\gamma_{\text{th}}} \left[\sum_{i=1}^{N} \frac{\frac{P_{\text{T}}}{N} c_i}{\frac{P_{\text{T}}}{N} c_i \sigma_v^2 + \sigma_w^2} \right] \right)^{-1} \right] \end{aligned}$$

[6] To recall, c_i is an exponentially-distributed r.v. with mean μ_c

$$\geq \mathbb{E}_{N;\gamma_{\mathrm{th}}}\left[\left(\frac{P_{\mathrm{T}}\left(\mu_c+\gamma_{\mathrm{th}}\right)}{\frac{P_{\mathrm{T}}}{N}\left(\mu_c+\gamma_{\mathrm{th}}\right)\sigma_v^2+\sigma_w^2}\right)^{-1}\right]$$

$$\geq \left(\frac{P_{\mathrm{T}}\left(\mu_c+\gamma_{\mathrm{th}}\right)}{\frac{P_{\mathrm{T}}}{N_o e^{-\frac{\gamma_{\mathrm{th}}}{\mu_c}}}\left(\mu_c+\gamma_{\mathrm{th}}\right)\sigma_v^2+\sigma_w^2}\right)^{-1}. \tag{3.18}$$

The first inequality holds because $\mathbb{E}\{1/g(x)\} \geq 1/\mathbb{E}\{g(x)\}$ as long as $g(x)$ is a positive and concave function [69, Ch. 3]. The two remaining inequalities follow from the fact that the arguments in the expectation terms are convex in c_i and N, respectively, and thus, the Jensen inequality applies. Finally, since (3.18) is convex in γ_{th}, its optimal value, $\check{\gamma}_{\mathrm{th}}^*$, can be found by setting its first derivative to zero, which leads to the following expression:

$$\frac{1}{2\mu_c}(\gamma_{\mathrm{th}}+\mu_c)e^{\frac{\gamma_{\mathrm{th}}+\mu_c}{2\mu_c}} = \frac{1}{2}\sqrt{\frac{N_o\sigma_w^2\mu_c e}{P_{\mathrm{T}}\sigma_v^2\mu_c}}. \tag{3.19}$$

By defining $x = \frac{\gamma_{\mathrm{th}}+\mu_c}{2\mu_c}$, we have

$$x = \mathsf{W}_0\left(\frac{1}{2}\sqrt{\frac{N_o\sigma_w^2\mu_c e}{P_{\mathrm{T}}\sigma_v^2\mu_c}}\right). \tag{3.20}$$

and, finally,

$$\check{\gamma}_{\mathrm{th}}^* = \left[2\mu_c\mathsf{W}_0\left(\frac{1}{2}\sqrt{\frac{N_o\sigma_w^2 e}{P_{\mathrm{T}}\sigma_v^2\mu_c}}\right)-\mu_c\right]^+ \tag{3.21}$$

where $\mathsf{W}_0(x)$ stands for the positive real branch of the Lambert function which is defined as $x = \mathsf{W}_0(x)\, e^{\mathsf{W}_0(x)}$.

Figure 3.4 shows the actual distortion value (computed numerically) and the convex lower bound given by equation (3.18) as a function of γ_{th}. Clearly, the bound is tight, in particular, for large networks when the Jensen inequalities above become even tighter. Consequently, we will incur in marginal performance loss resulting from the use of the approximate threshold $\check{\gamma}_{\mathrm{th}}^*$ instead of the actual one[7].

In order to give some insight into the behavior of the (approximate) threshold $\check{\gamma}_{\mathrm{th}}^*$, we depict in Fig. 3.5 the corresponding individual probability of activation, i.e., $p = e^{-\frac{\check{\gamma}_{\mathrm{th}}^*}{\mu_c}}$ as a function of the transmit power, P_{T}. First, one can observe that for an increasing transmit power, the probability of activation grows, as well. In other words, since power is not a scarce resource anymore, a higher number of sensors are allowed to participate in the estimation process (even if their contribution might be somewhat marginal due to less favorable channel conditions). Second, the growth rate of the individual probability of activation clearly depends on the quality of the sensor observations. For observations with poor quality (e.g. $\sigma_v^2 = 0.1$), the system tends

[7] To insist, the important aspect here is that the approximate threshold is very accurate; the fact that it was obtained from a lower bound is incidental.

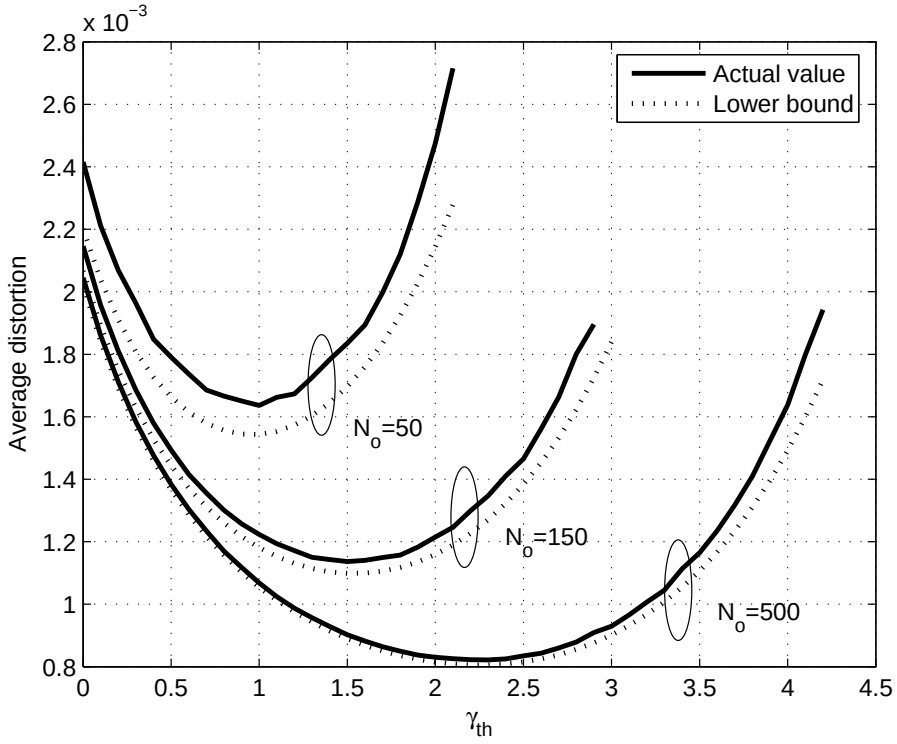


Figure 3.4: Actual distortion and lower bound as a function of $\gamma_{\rm th}$. $(N_o = 500, P_{\rm T} = 50, \sigma_v^2 = 0.01, \sigma_w^2 = 0.1, \mu_c = 1)$.

to activate more sensors in order to average out the observation noise. Conversely, in scenarios with higher observation qualities (e.g. $\sigma_v^2 = 0.0001$), to select the sensors with strongest channel gains is more beneficial.

3.5.1 Asymptotic analysis of the distortion rate

In this section, we analyze the rate at which the distortion decreases when the number of sensors grows without bound. To that aim, we resort to the derivation of asymptotic lower and upper bounds for the distortion attainable by the OPA-D and WF-D strategies, respectively.

OPA-D: asymptotic upper bound

According to the previous section, the threshold $\check{\gamma}_{\rm th}^*$ stands for the minimum channel gain for a sensor to be active and, thus, the channel gains of all *active* sensor nodes can be lower-bounded by $\check{\gamma}_{\rm th}^*$. By doing so, the distortion for a particular realization of N can be upper-bounded as follows:

$$D_{\mathsf{OPA-D}} \leq D_{\mathrm{OPA-D,UB}}^{N_o} = \left(\frac{P_T \check{\gamma}_{\rm th}^*}{\frac{P_T}{N} \check{\gamma}_{\rm th}^* \sigma_v^2 + \sigma_w^2} \right)^{-1}. \tag{3.22}$$

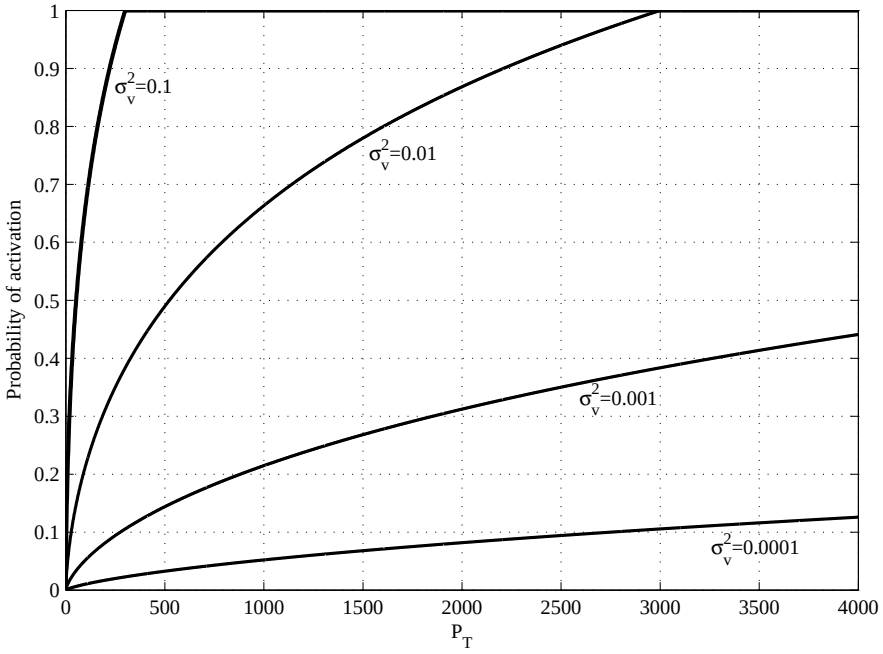


Figure 3.5: Probability of activation vs. total transmit power for different values of the observation noise variance σ_v^2 ($N_o = 300$, $\sigma_w^2 = 0.1$).

On the other hand, in Appendix 3.A.5 we prove that

$$\lim_{N_o \to \infty} \frac{\frac{P_T \check{\gamma}_{\text{th}}^*}{\sigma_w^2}}{\frac{P_T \check{\gamma}_{\text{th}}^*}{\frac{P_T}{N} \check{\gamma}_{\text{th}}^* \sigma_v^2 + \sigma_w^2}} \stackrel{\text{P}}{=} 1, \tag{3.23}$$

where $\stackrel{\text{P}}{=}$ denotes convergence in probability. This result states that distortion for the OPA-D scheme decreases *at least* with a rate given by

$$D_{\text{OPA-D,UB}}^{\infty} \sim \frac{\sigma_w^2}{P_T \check{\gamma}_{\text{th}}^*} \sim \frac{\sigma_w^2}{P_T \mathsf{W}_0\left(N_o\right)}. \tag{3.24}$$

As expected, in the OPA-D strategy adding sensors to the network pays off. Conversely, in the case of Uniform Power Allocation (UPA) over all sensors, increasing the network size is known to be not worthwhile, since distortion converges to a constant value [11].

WF-D: asymptotic lower bound

The fact that the optimal power allocation (WF-D) is computed by means of a waterfilling-like algorithm, makes the asymptotic analysis of the distortion rate extremely involved. Alternatively, we derive an absolute lower bound for any power allocation strategy. To that aim, note that distortion in (3.7) can be lower-bounded by considering noiseless sensor observations,

namely $\sigma_v^2 = 0$. By doing so, we have

$$D \geq D_{\mathsf{LB}} = \left(\sum_{i=1}^{N_o} \frac{p_i}{\sigma_w^2} c_i \right)^{-1}. \tag{3.25}$$

It is straightforward to show that the optimal power allocation which minimizes this lower bound (subject to a sum-power constraint $\sum_{i=1}^{N_o} p_i = P_T$) is to allocate all the available power P_T to the sensor with the highest channel gain, that is, $p_i = P_T$ if $i = \arg\max_i c_i$. Therefore, the distortion for the optimal power allocation of [11] for any $\sigma_v^2 > 0$ can be lower-bounded by

$$D_{\mathsf{WF-D}} \geq D_{\mathsf{WF-D,LB}} = \left(\frac{P_T}{\sigma_w^2} \max_i c_i \right)^{-1}. \tag{3.26}$$

On the other, for a large number of sensor nodes one can prove that

$$\lim_{N_o \to \infty} \frac{\max_i c_i}{\mathbb{E}\left[\max_i c_i\right]} \overset{\mathrm{P}}{=} 1, \tag{3.27}$$

which follows from Tchebychev's inequality. Besides, from [70] we have that

$$\mathbb{E}\left[\max_i c_i\right] = \sum_{i=1}^{N_o} i^{-1}$$

and, further,

$$\sum_{i=1}^{N_o} i^{-1} \sim \log(N_o).$$

Hence, from (3.25) (3.26) and (3.27) one finally concludes that the distortion for any power allocation strategy and a large network size decays *at most* at a rate given by

$$D^{\infty}_{\mathsf{WF-D,LB}} \sim \frac{\sigma_w^2}{P_T \log(N_o)}. \tag{3.28}$$

OPA-D: asymptotic distortion rate

For an arbitrary number of sensors, the *distortion* attained by OPA-D with the *optimal* threshold γ^*_{th} necessarily lies between those of OPA-D with an approximate threshold $\check{\gamma}^*_{\mathrm{th}}$ and WF-D. This also holds true for networks with an asymptotically large number of sensors. In this circumstances, expressions (3.24) and (3.28) revealed that the *rate* at which distortion decreases in OPA-D can be upper- and lower-bounded by $D^{\infty}_{\mathrm{OPA-D,UB}}$ and $D^{\infty}_{\mathsf{WF-D,LB}}$, respectively. From [71], it is straightforward to show that

$$\frac{D^{\infty}_{\mathsf{WF-D,LB}}}{D^{\infty}_{\mathsf{OPA-D,UB}}} = \lim_{N_o \to \infty} \frac{W_0(N_o)}{\log(N_0)} = 1, \tag{3.29}$$

namely, the rate at which the distortion for, on the one hand, the OPA-D scheme with the approximate threshold and, on the other hand, WF-D schemes decrease is *identical*. Consequently, the distortion associated to OPA-D (with the *optimal* threshold γ_{th}^*) also decreases at the same rate that WF-D does when the number of sensors grows without bound. In other words, there is no penalty (in terms of distortion rates) associated to the use of OPA-D instead of WF-D.

3.5.2 Imperfect channel state information: OPA-DR scheme

In realistic scenarios, only imperfect (e.g. noisy or delayed) CSI estimates are available at the sensors. Under this assumption, we derive next the corresponding reporting threshold.

To start with, let h_i and $\hat{h}_i$ denote the actual channel response and its estimate, and c_i and $\hat{c}_i$ denote their respective squared magnitudes. We can model the channel estimate as [72, Ch. 8]:

$$\hat{h}_i = h_i + e_i \quad ; \quad i = 1, \ldots, N_o \tag{3.30}$$

where e_i is the estimation error which is i.i.d. over the sensors and independent of h_i. Furthermore, e_i is modeled as a complex circular Gaussian random variable of variance σ_e^2. With these assumptions, h_i and $\hat{h}_i$ turn out to be related through a Gaussian model and, hence, the conditional random variable $h_i|\hat{h}_i$ follows a Gaussian distribution, that is,

$$h_i|\hat{h}_i \sim \mathcal{CN}(\eta_i \hat{h}_i, \sigma_i^2), \tag{3.31}$$

with

$$\eta_i = \frac{\mu_c}{\mu_c + \sigma_e^2} \quad \sigma_i^2 = \frac{\mu_c \sigma_e^2}{\mu_c + \sigma_e^2}. \tag{3.32}$$

Hereinafter, we attempt to minimize of distortion with such imperfect channel estimates. The expected distortion w.r.t. the *actual* channel realizations (which determine the distortion in the estimate), the *estimates* of the channel gains (on the basis of which sensors decide whether they belong to the active subset) and the number of active sensors reads

$$\mathbb{E}_{N;\gamma_{\text{th}}} \left[\mathbb{E}_{\{c_i\}_{i=1}^N, \{\hat{c}_i\}_{i=1}^N | N;\gamma_{\text{th}}} \left[\left(\sum_{i=1}^{N} \frac{\frac{P_{\text{T}}}{N} c_i}{\frac{P_{\text{T}}}{N} c_i \sigma_v^2 + \sigma_w^2} \right)^{-1} \right] \right] \tag{3.33}$$

$$\geq \mathbb{E}_{N;\gamma_{\text{th}}} \left[\left(\mathbb{E}_{\{c_i\}_{i=1}^N, \{\hat{c}_i\}_{i=1}^N | N;\gamma_{\text{th}}} \left[\sum_{i=1}^{N} \frac{\frac{P_{\text{T}}}{N} c_i}{\frac{P_{\text{T}}}{N} c_i \sigma_v^2 + \sigma_w^2} \right] \right)^{-1} \right]$$

$$= \mathbb{E}_{N;\gamma_{\text{th}}} \left[\left(\sum_{i=1}^{N} \mathbb{E}_{c_i, \hat{c}_i | N;\gamma_{\text{th}}} \left[\frac{\frac{P_{\text{T}}}{N} c_i}{\frac{P_{\text{T}}}{N} c_i \sigma_v^2 + \sigma_w^2} \right] \right)^{-1} \right]$$

$$= \mathrm{E}_{N;\gamma_{\text{th}}} \left[\left(\sum_{i=1}^{N} \mathrm{E}_{\hat{c}_i | N;\gamma_{\text{th}}} \left[\mathrm{E}_{c_i|\hat{c}_i} \left[\frac{\frac{P_{\text{T}}}{N} c_i}{\frac{P_{\text{T}}}{N} c_i \sigma_v^2 + \sigma_w^2} \right] \right] \right)^{-1} \right]. \tag{3.34}$$

Again, the first inequality holds because $\mathbb{E}\left[1/g(x)\right] \geq 1/\mathbb{E}\left[g(x)\right]$ as long as $g(x)$ is a positive and concave function [69, Ch. 3] w.r.t. c_i (notice that the argument in the expectation term does not depend on $\hat{c}_i$). The last equality holds because the random variable c_i is independent of the selection process given $\hat{c}_i$. From [73, Ch. 2], we have that

$$\mathbb{E}_{\hat{c}_i;\gamma_{\rm th}}\left[\mathbb{E}_{c|\hat{c}_i}\left[c\right]\right] = \mu_c + \left(\frac{\mu_c}{\mu_c+\sigma_e^2}\right)^2 \gamma_{\rm th}. \tag{3.35}$$

From the above expressions and by repeatedly applying Jensen's inequalities, a lower bound of (3.34) the average distortion (3.33) reads

$$\left(\frac{\mu_c + \left(\frac{\mu_c}{\mu_c+\sigma_e^2}\right)^2 \gamma_{\rm th}}{\frac{P_{\rm T}}{Me^{-\frac{\gamma_{\rm th}}{\mu_c}}}\left(\mu_c + \left(\frac{\mu_c}{\mu_c+\sigma_e^2}\right)^2 \gamma_{\rm th}\right)\sigma_v^2 + \sigma_w^2}\right)^{-1}. \tag{3.36}$$

where the effect of the uncertainty in the channel estimates becomes apparent. To be more precise, the larger the uncertainty, i.e. $\sigma_e^2 \to \infty$, the larger the lower bound of (3.36).

Finally, after some algebra, the approximate threshold $\check{\gamma}_{\rm th}^*$ with *imperfect* CSI can be expressed in closed-form as follows:

$$\check{\gamma}_{\rm th}^* = \left[\left(\mu_c+\sigma_e^2\right)\left(2\mathsf{W}_0\left(\frac{1}{2\mu_c}\sqrt{\frac{N_o\sigma_w^2\left(\mu_c+\sigma_e^2\right)e^{\frac{\mu_c+\sigma_e^2}{\mu_c}}}{P_{\rm T}\sigma_v^2}}\right) - \frac{\mu_c+\sigma_e^2}{\mu_c}\right)\right]^+. \tag{3.37}$$

In the sequel, the opportunistic power allocation scheme which operates with such reporting threshold will be referred to as Robust OPA-D (or OPA-DR). As expected, with perfect CSI (i.e. $\sigma_e^2 \to 0$) the above threshold converges to that of OPA-D which is given by equation (3.21). Conversely, in scenarios with very poor CSI qualities ($\sigma_e^2 \to \infty$) the system mimics the behavior of a UPA scheme, namely $\check{\gamma}_{\rm th}^* \to 0$ (see proof in Appendix 3.A.2). Indeed, when no reliable selection of sensors can be carried out because of very poor CSI on sensor-to-FC channel conditions, the best thing to do is to let all the sensors participate in the estimation process.

3.5.3 Simulations and numerical results

In Figure 3.6, we depict the average distortion attained by the OPA-D scheme as a function of the network size (N_o) for a given sum-power constraint. First of all, one observes that the proposed opportunistic power allocation scheme performs remarkably better than its uniform power allocation counterpart: in OPA-D curves the overall distortion is $150-280\%$ lower than in UPA. As expected, saving the available power for those sensors which experience better channel conditions definitely pays-off. More importantly, the performance of OPA-D is virtually identical to that of the WF-D (i.e. optimal) power allocation scheme. To insist, the WF-D

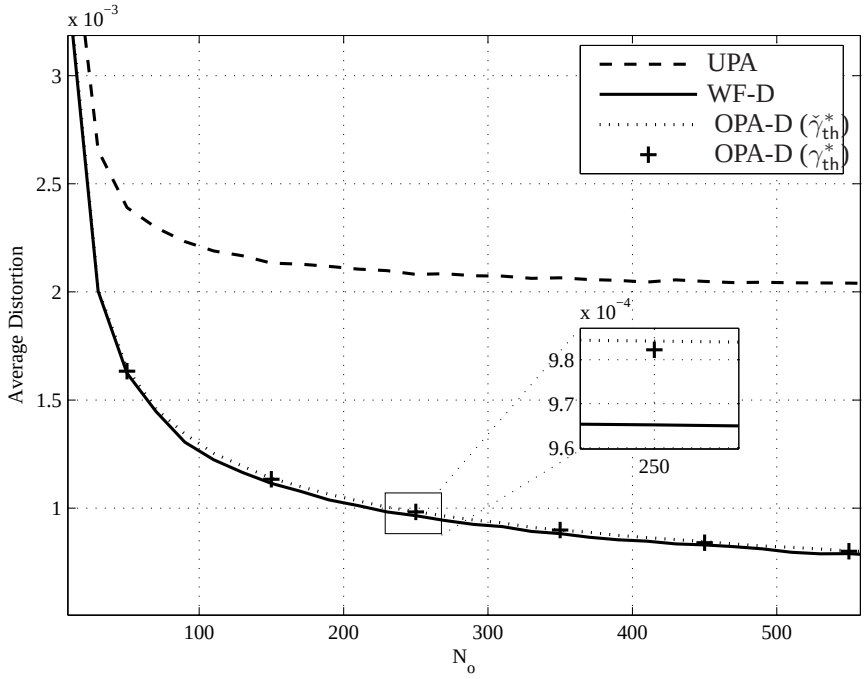


Figure 3.6: Average distortion vs. network size ($P_{\rm T} = 50$, $\sigma_v^2 = 0.01$, $\sigma_w^2 = 0.1$). The performance of OPA-D was evaluated with the approximate threshold $\check{\gamma}^*_{\rm th}$ in (3.21), whereas markers on that curve ($+$) show results with the true optimal threshold $\gamma^*_{\rm th}$ that was computed numerically.

scheme requires full and instantaneous CSI from *all* the sensors in the network, whereas in OPA-D this is only needed for the subset of active nodes, along with some *statistical* CSI. Besides, OPA-D effectively exploits multi-user diversity (as we proved in Section 3.5.1) whereas UPA quickly saturates, as already pointed out in [11]. Finally, the performance loss resulting from the use of the approximate optimal threshold $\check{\gamma}^*_{\rm th}$ computed with the closed-form expression (3.21) instead of the actual one (which can only be computed numerically) is negligible for the whole range of values of N_o considered. That is, the inequalities that we resorted to in the derivation of the lower bound are tight for $N_o = 50, \ldots, 650$.

The gain of OPA with respect to UPA is better illustrated in Figure 3.7. For a low transmit power constraint, OPA-D and WF-D schemes exhibit a substantial gain with respect to UPA. Conversely, this gain decreases for an increasing transmit power $P_{\rm T}$. In that case, both WF-D and OPA-D tend to activate the whole set of sensors.

In Figure 3.8, we depict the average number of active sensors for the OPA-D and WF-D schemes. Interestingly, the number of active sensors is much lower for the OPA-D scheme. However, the gain that the WF-D strategy attains with an increased number of active sensors was shown to be marginal. Consequently, it is preferable to uniformly allocate power to a smaller subset of sensors with high channel gains (OPA-D case) rather than spread resources thinner and allocate some power to sensors with low channel gains (WF-D) that, ultimately,

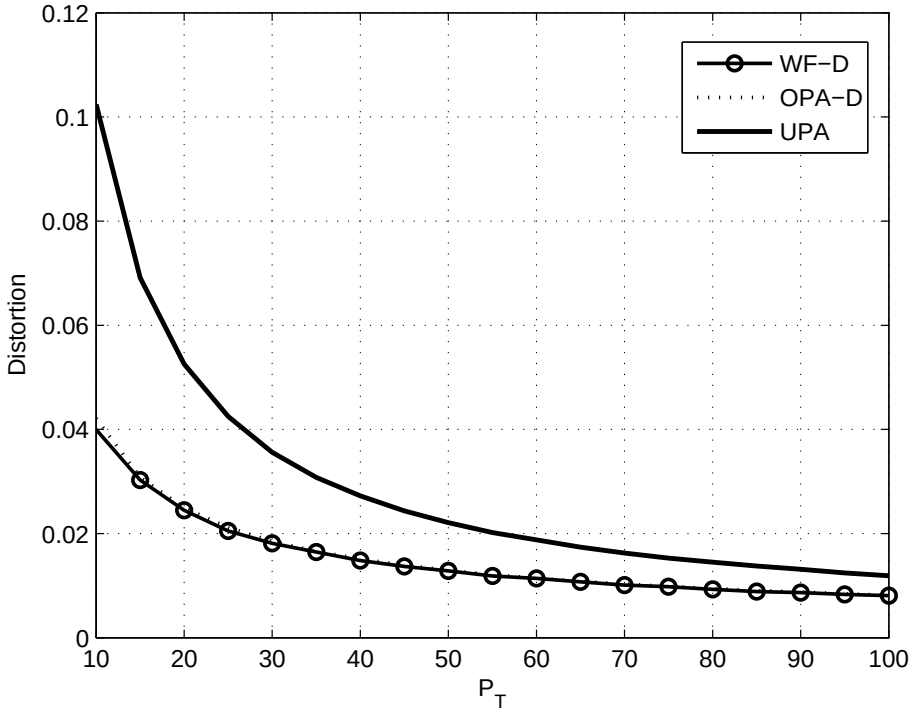


Figure 3.7: Distortion vs. sum-power constraint P_T ($\sigma_w^2 = 1, \sigma_v^2 = 0.1, N_o = 100$)

would have a very limited contribution to the reduction of the overall distortion in the estimate. Such reduced number of active sensors translates into a more relaxed requirement in terms of *i*) the number of orthogonal FC-sensor channels needed; and *ii*) the number of channel gains to be estimated at the FC. In Figure 3.8, we also observe that both strategies tend to activate more sensors as the transmit power P_{T} increases. As previously discussed, for very high values of P_{T} the optimal solution is to uniformly allocate the power among the sensors (i.e. same as in UPA).

In Figure 3.9, we plot the average distortion attained by the OPA-DR scheme as a function of the population size, and for different levels of CSI uncertainty $\Delta_e = 10\log(\mu_e/\sigma_e^2)$ for a given network size ($N_o = 500$ sensors). Interestingly, for all the OPA-DR curves, the rate at which the distortion decreases mostly mimics that of the OPA-D (with perfect CSI) and WF schemes. Hence, OPA-DR is capable of exploiting multi-user diversity in the same way as such schemes do even for high values of Δ_e (e.g. $\Delta_e = 0dB$). Complementarily, in Fig. 3.10 we depict the average distortion vs. the amount of CSI uncertainty Δ_e. For $\Delta_e = 15\mathrm{dB}$ the performance is virtually identical to the case of perfect CSI, and, more importantly, with $\Delta_e = 0\mathrm{dB}$ it is still significantly better than that of UPA. Indeed, the OPA-DR curve only approaches the UPA bound (this meaning that no actual sensor selection is carried out) when the channel estimates are of extremely poor quality ($\Delta_e = -15\mathrm{dB}$).

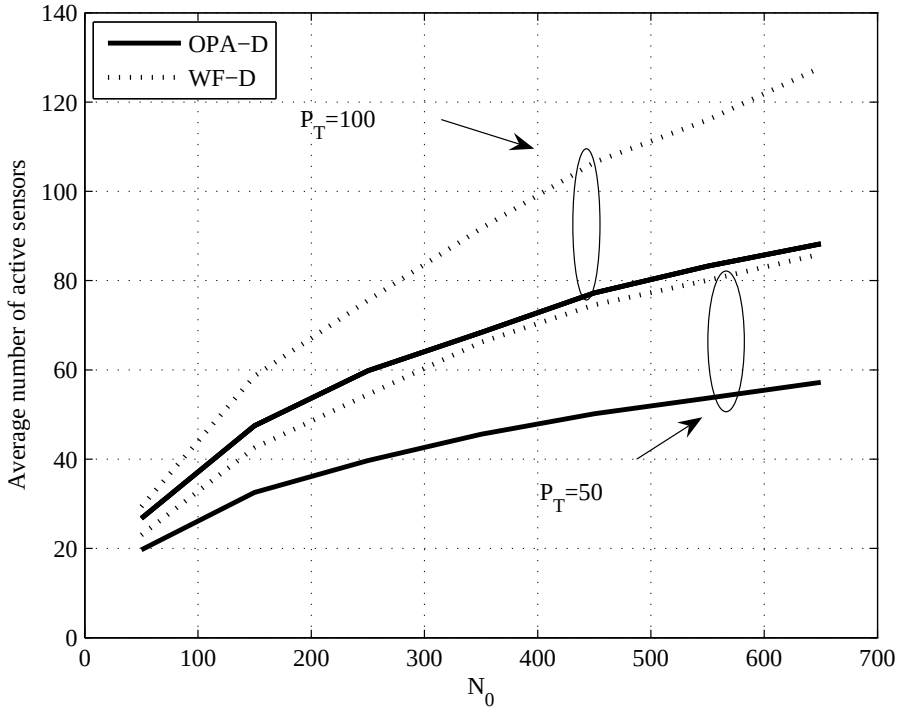


Figure 3.8: Average number of active sensors vs. network size for the minimization of distortion.

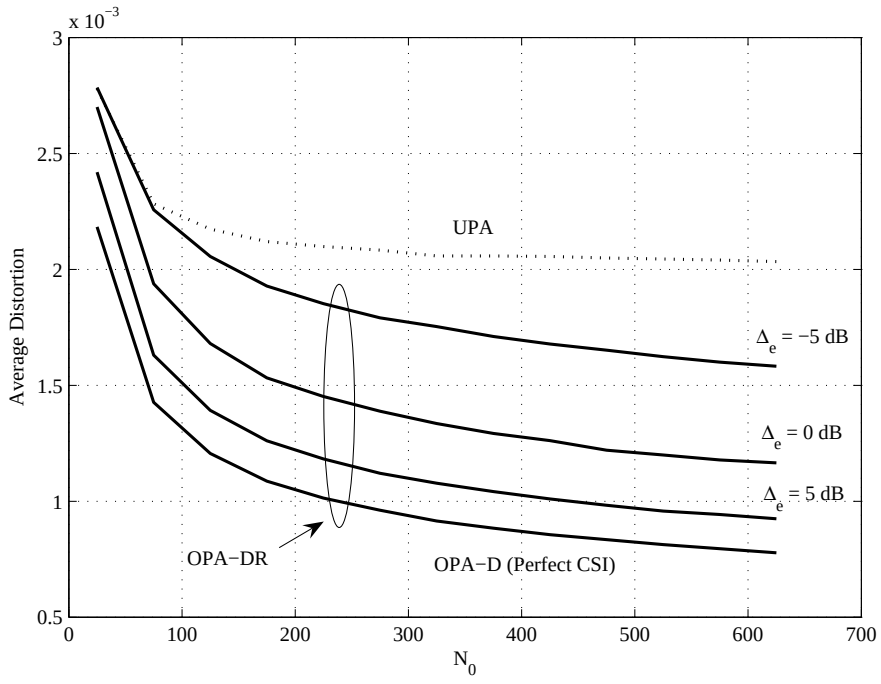


Figure 3.9: Average distortion of the robust OPA scheme vs. network size for different values of CSI uncertainty Δ_e($P_{\mathrm{T}} = 50$, $\sigma_v^2 = 0.01$, $\sigma_w^2 = 0.1$, approximate threshold).

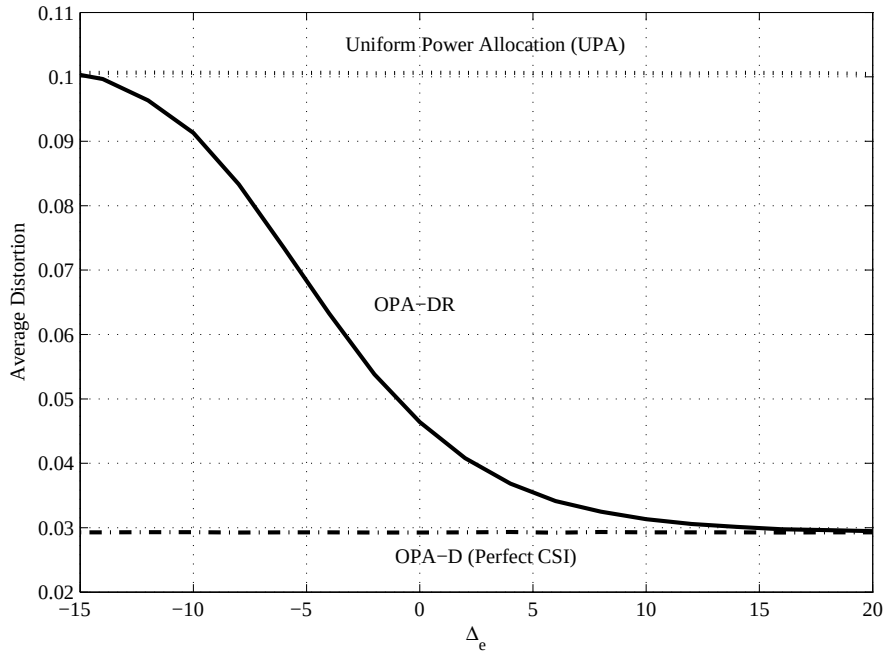


Figure 3.10: Average distortion vs. CSI uncertainty. The solid curve depicts the performance exhibited by the OPA-DR scheme ($\sigma_v^2 = 0.01, \sigma_w^2 = 0.1, N_0 = 500, P = 1$, approximate threshold).

3.6 OPA for the minimization of transmit power (OPA-P)

Energy efficiency is of paramount importance in wireless sensor networks. Hence, we change our design criterion and now we attempt to find a reporting threshold which minimizes the total transmit *power* subject to a given distortion constraint:

$$\gamma_{\rm th}^* = \arg\min_{\gamma_{\rm th}} \mathbb{E}_{\{c_i\}_{i=1}^N, N; \gamma_{\rm th}} \left[\sum_{i=1}^{N} p_i \right] \tag{3.38}$$

$$\text{s.t.} \quad D = D_{\rm T}, \tag{3.39}$$

where D and $D_{\rm T}$ stand for the actual and target distortion, respectively. From (3.7) the overall distortion D can be readily expressed in terms of the individual contributions D_i of each active sensor node, namely

$$D = \left(\sum_{i=1}^{N} \frac{1}{D_i} \right)^{-1}. \tag{3.40}$$

Note that D_i stands for the distortion if only sensor i transmits its observation. Likewise, for a given N, (3.40) stands for the distortion when the subset of N *active* sensors transmit their observations. Since only *local* CSI can be assumed to be available at the sensor nodes, we further impose their individual contributions to the overall distortion to be identical. To guarantee that the constraint in (3.39) is met, we let $D_i = ND_{\rm T}$ and force each sensor to adjust

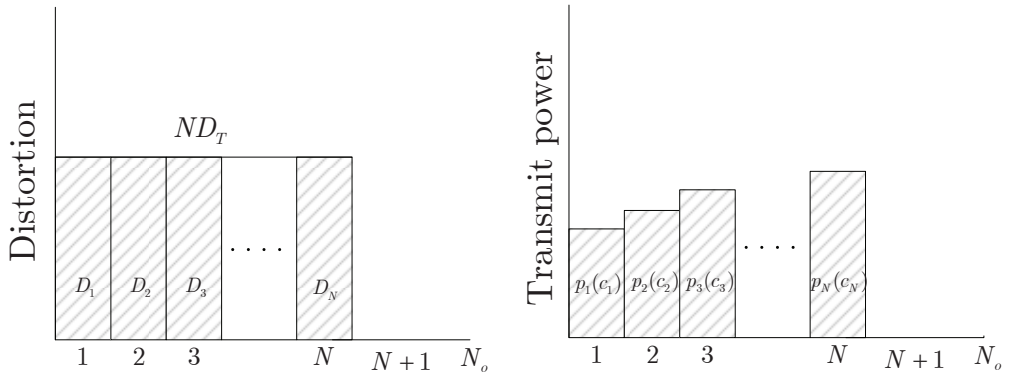


Figure 3.11: Graphical interpretation of the OPA-P strategy.

locally its transmit power accordingly (see Figure 3.11). From (3.7), we have that necessarily

$$p_i = \begin{cases} \frac{\frac{1}{ND_{\mathrm{T}}}\sigma_w^2}{c_i\left(1-\frac{1}{ND_{\mathrm{T}}}\sigma_v^2\right)} & c_i > \gamma_{\mathsf{th}} \\ 0 & \text{otherwise.} \end{cases} \tag{3.41}$$

for $i = 1, \ldots, N_o$. Finally, the optimization problem can now be re-written as

$$\gamma_{\mathrm{th}}^* = \arg\min_{\gamma_{\mathrm{th}}} \mathbb{E}_{\{c_i\}_{i=1}^N, N; \gamma_{\mathrm{th}}} \left[\sum_{i=1}^{N} \frac{\frac{1}{ND_{\mathrm{T}}}\sigma_w^2}{c_i\left(1-\frac{1}{ND_{\mathrm{T}}}\sigma_v^2\right)} \right]. \tag{3.42}$$

Again, the expression above is barely tractable. For that reason, we derive a lower bound by repeatedly applying Jensen's inequality:

$$\mathbb{E}_{N;\gamma_{\mathrm{th}}} \left[\mathbb{E}_{\{c_i\}_{i=1}^N;\gamma_{\mathrm{th}}} \left[\sum_{i=1}^{N} \frac{\frac{1}{ND_{\mathrm{T}}}\sigma_w^2}{c_i\left(1-\frac{1}{ND_{\mathrm{T}}}\sigma_v^2\right)} \right] \right] \geq \mathbb{E}_{N;\gamma_{\mathrm{th}}} \left[\frac{\frac{1}{D_{\mathrm{T}}}\sigma_w^2}{(\mu_c+\gamma_{\mathrm{th}})\left(1-\frac{1}{ND_{\mathrm{T}}}\sigma_v^2\right)} \right] \tag{3.43}$$

$$\geq \frac{\frac{1}{D_{\mathrm{T}}}\sigma_w^2}{(\mu_c+\gamma_{\mathrm{th}})\left(1-\frac{1}{D_{\mathrm{T}}N_o e^{-\gamma_{\mathrm{th}}/\mu_c}}\sigma_v^2\right)} \tag{3.44}$$

The argument of the first expression is clearly convex in c_i. As for (3.43), the argument is convex in N as long as $N \geq \lceil \sigma_v^2/D_{\mathrm{T}} \rceil$. This means, in turn, that the target distortion D_{T} can be actually met since otherwise the transmit power p_i would take negative values (see equation (3.41)))[8]. As it is shown in Appendix 3.A.3, for large N_o, the probability of the event $\{N \geq \lceil \sigma_v^2/D_{\mathrm{T}} \rceil\}$ can be made arbitrarily close to 1 and, thus, the bound we are deriving is almost surely valid.

Finally, we have to prove that the lower bound in (3.44) is convex in γ_{th}. Note that the denominator in (3.44) is concave and positive for $\gamma_{\mathrm{th}} \in [0, \mu_c \log(D_{\mathrm{T}} N_o/\sigma_v^2))$. Since $f(x) = 1/x$ is convex and non-increasing in $x \in \mathbb{R}^+$ by composition [69, Ch. 3] we conclude that (3.44)

[8] Actually, $N_{\mathsf{cen}} = \lceil \sigma_v^2/D_{\mathrm{T}} \rceil$ can be interpreted as the minimum number of observations needed in a *centralized* scenario to attain a prescribed distortion level D_T with noisy observations of variance σ_v^2.

is convex in $\gamma_{\rm th} \in [0, \mu_c \log(D_{\rm T} N_o/\sigma_v^2))$ which, as detailed in Appendix 3.A.1, is the only domain of $\gamma_{\rm th}$ in practice. Setting its derivative to zero yields

$$\check{\gamma}_{\rm th}^* = \left[\mu_c \mathsf{W}_0\left(\frac{D_{\rm T} N_o e^2}{\sigma_v^2}\right) - 2\mu_c\right]^+. \tag{3.45}$$

As in previous designs, the tighter the inequalities, the closer the approximate threshold $\check{\gamma}_{\rm th}^*$ will be to its optimal value $\gamma_{\rm th}^*$.

3.7 OPA for the enhancement of network lifetime (OPA-LT)

As far as this section is concerned, we define network lifetime (LT) as the time elapsed until the first sensor runs out of energy [58]. When this occurs, the remaining $N-1$ active sensors scheduled in a timeslot are not capable of attaining the prescribed distortion level. Such *estimation outage* occurs because power was allocated under the assumption of having N active sensors (see Eq. 3.41) whereas only $N-1$ conveyed their observations to the FC.

Clearly, any sensor scheduling scheme aimed at increasing network LT should take into account not only the channel propagation conditions (as done in the previous sections) but also the information on the residual energy in the nodes (REI). In the spirit of [7], we let sensor i participate in the estimation process if and only if the product of its residual energy in time-slot s, $\varepsilon_i[s]$, and the channel gain is above a threshold, namely, $\varepsilon_i[s]c_i > \gamma_{\rm th}[s]$. In other words, sensors experiencing favorable channel conditions *and* sufficient residual energy are scheduled with probability

$$\Pr\left(\varepsilon_i[s]c_i > \gamma_{\rm th}[s]\right) = e^{-\frac{\gamma_{\rm th}[s]}{\mu_c \varepsilon_i[s]}}. \tag{3.46}$$

This selection strategy is known to enhance the network lifetime while, as we will see later on, it keeps the transmit power reasonably low [7]. However, it introduces individual thresholds for *each* sensor (instead of a single reporting threshold, as in OPA-P and OPA-D) which have to be re-computed during network lifetime and not only in the initialization phase. Note also that the energy vector[9] $\boldsymbol{\varepsilon}[s] = [\varepsilon_1[s], \ldots, \varepsilon_{N_o}[s]]$ is a non-stationary stochastic process the individual entries of which are locally updated as follows,

$$\varepsilon_i[s+1] = \varepsilon_i[s] - p_i[s]T_s \quad \text{with } \varepsilon_i[0] = \varepsilon_o, \tag{3.47}$$

where $p_i[s]$ denotes the transmit power in slot s, T_s is the duration of the timeslot and ε_o stands for the initial energy. As for the power allocation rule, we again force each active sensor to evenly contribute to the overall distortion, that is, each sensor adjusts locally its transmit power according to (3.41).

[9] We assume that the energy budget is dominated by energy consumption during wireless transmission

In this context, the optimal threshold $\gamma_{\text{th}}^{*}[s]$ is the one which minimizes the total transmit power under this REI-based selection rule[10], namely

$$\gamma_{\text{th}}^{*}[s] = \arg\min_{\gamma_{\text{th}}[s]} \left\{ \mathbb{E}_{\{c_i\}_{i=1}^{N}, N; \gamma_{\text{th}}[s], \boldsymbol{\varepsilon}[s]} \left[\sum_{i=1}^{N} \frac{\frac{1}{ND_{\text{T}}}\sigma_w^2}{c_i \left(1 - \frac{1}{ND_{\text{T}}}\sigma_v^2\right)} \right] \right\}. \tag{3.48}$$

This problem is barely tractable and, again, we must resort to a lower bound. First, though, we need to introduce three inequalities that will be useful for the derivation of the bound. Without loss of generality, let $\boldsymbol{\varepsilon}[s]$ be an *ordered* vector, namely $\varepsilon_1[s] > \varepsilon_2[s] > \ldots > \varepsilon_{N_o}[s]$. By resorting to Jensen's inequality, the average number of active sensors (which, on the basis of equation (3.46), can be computed as the summation of the individual activation probabilities of N_o different binomial random variables) can be lower-bounded as follows:

$$\mathbb{E}_{N; \gamma_{\text{th}[s]}, \boldsymbol{\varepsilon}[s]}\left[N\right] = \sum_{i=1}^{N_o} e^{-\frac{\gamma_{\text{th}}[s]}{\varepsilon_i[s]\mu_c}} \geq N_o e^{-\frac{\gamma_{\text{th}}[s]}{\mu_c N_o}\sum_{i=1}^{N_o}\frac{1}{\varepsilon_i[s]}}. \tag{3.49}$$

Besides, for an ordered vector of energies and for some $N_o' < N_o$ the average number of active sensors can also be upper-bounded (see proof in Appendix 3.A.4) by:

$$\mathbb{E}_{N; \gamma_{\text{th}[s]}, \boldsymbol{\varepsilon}[s]}\left[N\right] = \sum_{i=1}^{N_o} e^{-\frac{\gamma_{\text{th}}[s]}{\varepsilon_i[s]\mu_c}} \leq N_o e^{-\frac{\gamma_{\text{th}}[s]}{\mu_c N_o'}\sum_{i=1}^{N_0'}\frac{1}{\varepsilon_i[s]}} \tag{3.50}$$

for $0 \leq \gamma_{\text{th}}[s] \leq \gamma^{'}$, with γ' being defined in equation (3.56) ahead. The interest in letting $N_o' > 1$ (for $N_o' = 1$ the inequality is trivial for any $\gamma_{\text{th}}[s]$) lies in the fact that the higher N_o', the tighter the resulting upper bound. Still, for $N_o' > 1$ the bound is only valid for part of the function domain and, hence, one should first identify $\gamma^{'}$ and then let N_o' take the highest value possible for which the inequality holds. We will go back to this issue later in this section.

From equation (3.49), it is straightforward to obtain the last inequality that we need:

$$\mathbb{E}_{c; \gamma_{\text{th}}[s], \boldsymbol{\varepsilon}[s]}\left[c\right] \leq \mu_c + \frac{\gamma_{\text{th}}[s]}{H(\boldsymbol{\varepsilon}[s])_{1:N_o}}. \tag{3.51}$$

with $H(\boldsymbol{\varepsilon}[s])_{1:N_o} = N_o \left(\sum_{i=1}^{N_o} \varepsilon_i[s]^{-1}\right)^{-1}$ standing for the harmonic mean of the first N_o elements of vector $\boldsymbol{\varepsilon}[s]$.

Now, by repeatedly applying Jensen's inequality along with these inequalities (as displayed in

[10] As discussed in the previous paragraphs, the optimal threshold $\gamma_{\text{th}}^{*}[s]$, which depends on the vector of residual energies, has to be re-computed on a timeslot-by-timeslot basis.

the equations below), we can finally obtain the lower bound of the score function (3.48):

$$\mathbb{E}_{N;\gamma_{\mathrm{th}}[s],\boldsymbol{\varepsilon}[s]} \left[\mathbb{E}_{\{c_i\}_{i=1}^N;\gamma_{\mathrm{th}}[s],\boldsymbol{\varepsilon}[s]} \left[\sum_{i=1}^{N} \frac{\frac{1}{ND_{\mathrm{T}}}\sigma_w^2}{c_i\left(1-\frac{1}{ND_{\mathrm{T}}}\sigma_v^2\right)} \right] \right] \tag{3.52}$$

$$\overset{(3.51)}{\geq} \mathbb{E}_{N;\gamma_{\mathrm{th}}[s],\boldsymbol{\varepsilon}[s]} \left[\frac{\frac{1}{D_{\mathrm{T}}}\sigma_w^2}{\left(\mu_c + \frac{\gamma_{\mathrm{th}}[s]}{H(\boldsymbol{\varepsilon}[s])_{1:N_o}}\right)\left(1-\frac{1}{ND_{\mathrm{T}}}\sigma_v^2\right)} \right] \tag{3.53}$$

$$\geq \frac{\frac{1}{D_{\mathrm{T}}}\sigma_w^2}{\left(\mu_c + \frac{\gamma_{\mathrm{th}}[s]}{H(\boldsymbol{\varepsilon}[s])_{1:N_o}}\right)\left(1-\frac{1}{\sum_{i=1}^{N_o} e^{-\frac{\gamma_{\mathrm{th}}[s]}{\mu_c \varepsilon[s]}} D_{\mathrm{T}}}\sigma_v^2\right)} \tag{3.54}$$

$$\overset{(3.50)}{\geq} \frac{\frac{1}{D_{\mathrm{T}}}\sigma_w^2}{\left(\mu_c + \frac{\gamma_{\mathrm{th}}[s]}{H(\boldsymbol{\varepsilon}[s])_{1:N_o}}\right)\left(1-\frac{1}{D_{\mathrm{T}}N_o e^{-\frac{\gamma_{\mathrm{th}}[s]}{\mu_c H(\boldsymbol{\varepsilon}[s])_{1:N_o'}}}}\sigma_v^2\right)}. \tag{3.55}$$

The argument in the first expression is clearly convex in c_i. As for (3.54), the argument is convex in N as long as $N \geq \lceil \sigma_v^2/D_{\mathrm{T}} \rceil$, as discussed in the previous section. The highest value of $\gamma_{\mathrm{th}}[s]$ for which (3.55) is still a convex function occurs when the second term in parenthesis in the denominator, which is a decreasing function in $\gamma_{\mathrm{th}}[s]$, tends to zero (for negative values, the bound is not a convex function anymore). Hence, we have:

$$\gamma' = \mu_c H(\boldsymbol{\varepsilon}[s])_{1:N_o'} \ln\left(\frac{N_o D_t}{\sigma_v^2}\right) \tag{3.56}$$

and, from this value, the FC can compute the highest value of N_o' for which inequality (3.50) holds true. Finally, by setting its derivative respect to $\gamma_{\mathrm{th}}[s]$ to zero, we obtain the threshold $\gamma_{\mathrm{th}}^*[s]$ which minimizes the bound, that is,

$$\gamma_{\mathrm{th}}^*[s] = \mu_c H(\boldsymbol{\varepsilon}[s])_{1:N_o'} \left[\mathsf{W}_0\left(\frac{D_{\mathrm{T}} N_o e^{\frac{H(\boldsymbol{\varepsilon}[s])_{1:N_o'} + H(\boldsymbol{\varepsilon}[s])_{1:N_o}}{H(\boldsymbol{\varepsilon}[s])_{1:N_o'}}}}{\sigma_v^2} \right) - \frac{H(\boldsymbol{\varepsilon}[s])_{1:N_o} + H(\boldsymbol{\varepsilon}[s])_{1:N_o'}}{H(\boldsymbol{\varepsilon}[s])_{1:N_o'}} \right]^+. \tag{3.57}$$

which can be shown to lie within $[0, \gamma')$ (the analysis is similar to that in Appendix 3.A.1). From the equation above, one notices that the threshold $\gamma_{\mathrm{th}}^*[s]$ depends on the residual energy vector $\boldsymbol{\varepsilon}[s]$ and thus, the FC needs REI for its computation. However, there is no need for sensors to send updates of their REI. Instead, $\boldsymbol{\varepsilon}[s]$ can be locally updated at the FC as in (3.47), since both the individual sensors that are scheduled to send data and their channel gains c_i are known to it.

Finally, in the case of identical residual energies, $\varepsilon_i[s] = \varepsilon[s]$; $i = 1, \ldots, N_o$ equation (3.50) holds with equality for up to $N_o' = N_o$. Thus, we have $H(\boldsymbol{\varepsilon}[s])_{1:N_o'} = H(\boldsymbol{\varepsilon}[s])_{1:N_o} = \varepsilon[s]$ and, by replacing (3.57) into (3.46) we realize that actual sensor selection rule is identical to that of the OPA-P case which simply disregards REI information.

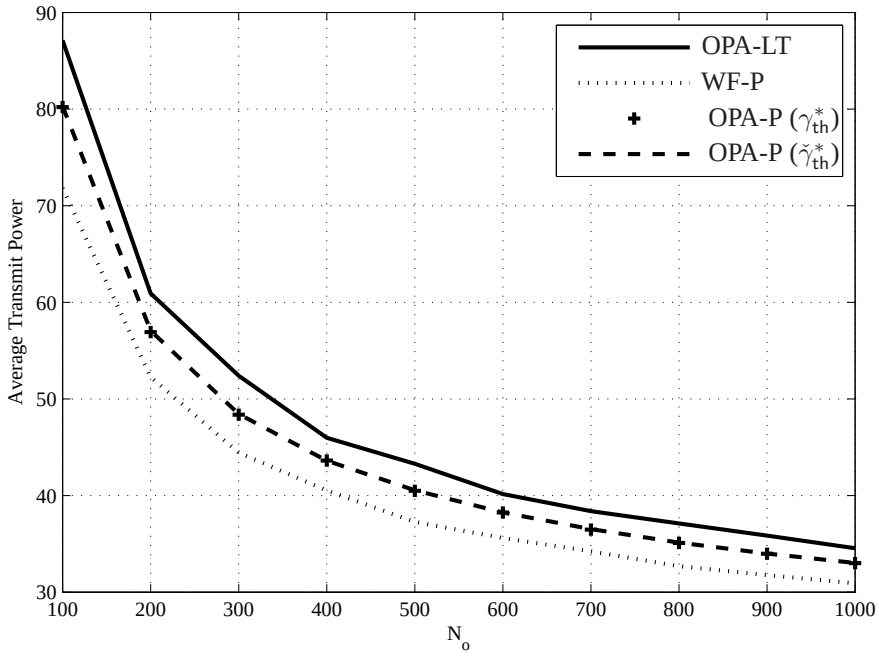


Figure 3.12: Average transmit power vs. network size ($D_{\rm T} = 0.001$, $\sigma_v^2 = 0.01$, $\sigma_w^2 = 0.1$). The performance of OPA-P (dashed curve) was evaluated with the approximate threshold $\check{\gamma}^*_{\rm th}$ in (3.45), whereas markers on that curve ($+$) show results with the true optimal threshold $\gamma^*_{\rm th}$ that was computed numerically.

3.7.1 Simulations and numerical results

In Fig. 3.12, we compare the average transmit power as a function of the network size for a given distortion target. First, we observe that the performance of OPA-P is close to that of the WF-P (i.e. optimal) power allocation scheme. Note, however, that such a marginal gain of WF-P entails a much larger amount of FC-sensor signalling and exchange of information. Besides, the increase in the transmit power associated with the use of OPA-LT can also be regarded as very moderate ($8 - 10\%$); this is despite of the fact that the sensor(s) experiencing the best channel conditions might not be scheduled in some situations, for instance, when some other sensor is running out of batteries. It is worth noting that, in the OPA-LT case, it is not possible (within a reasonable time frame) to numerically compute the true optimal thresholds and, as in the OPA-P case, to check the performance loss w.r.t. the approximate ones derived with the bound. Still, such curve would necessarily lie in between those of OPA-LT (upper bound, given by the approximate threshold) and OPA-P (lower bound, given by a threshold which actually disregards REI) which, as commented above, are very close to each other.

For completeness, Figure 3.13 depicts the average number of active sensors for the OPA-P and WF-P schemes. Again, the number of active sensors is substantially lower for the OPA scheme. Besides, one can notice that, the higher the observation noise, the higher the number of active

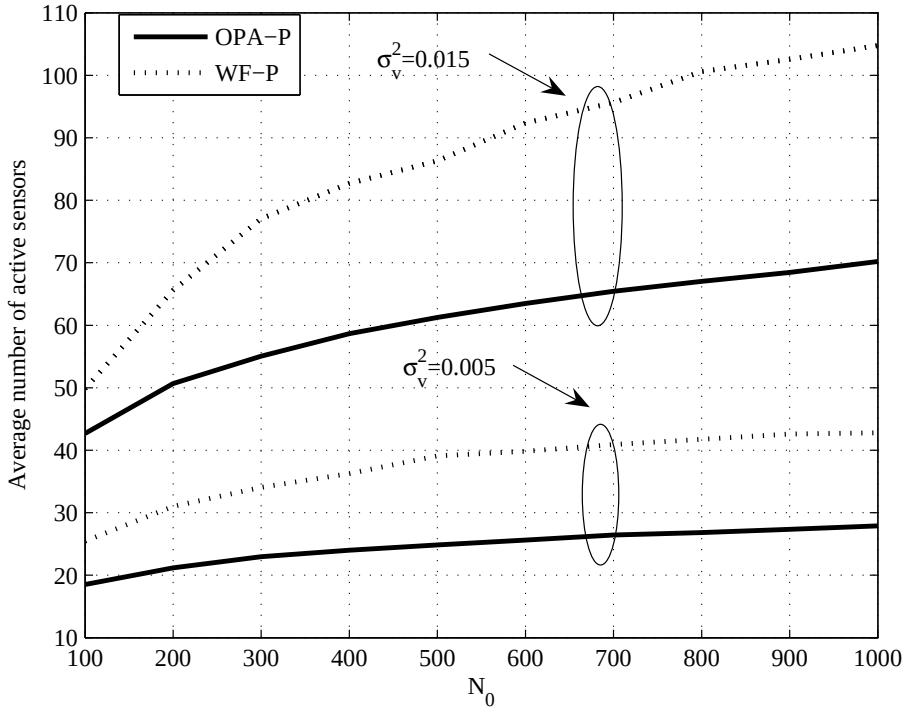


Figure 3.13: Average number of active sensors vs. network size for transmit power minimization.

sensors for both strategies. Clearly, as long as σ_v^2 increases more sensors must be activated in order to meet the pre-defined distortion target.

In Fig. 3.14, we depict the average network lifetime vs. the network size for a given distortion target. First, one realizes that WF-P and OPA-P yield comparable network lifetimes. More importantly, OPA-LT almost doubles the LT obtained with the other two solutions thanks to a sensible use of REI information. However, as long as the scheduling rule and the reporting threshold do not minimize the energy consumption at each time-slot anymore, the average transmit power of OPA-LT, is slightly higher now (see Fig. 3.12).

If one incorporates REI into the scheduling process, the sensors with higher residual energies are more prone to participate in the estimation task. Roughly speaking, by properly combining REI with CSI into the scheduling process, one has a means to enforce energy to be uniformly spent over sensors time-slot after time-slot. This extent is illustrated in Figure 3.15, where we plot the energy dispersion defined as $\chi_\varepsilon = \frac{\sigma_{\boldsymbol{\varepsilon}[s]}}{\mu_{\boldsymbol{\varepsilon}[s]}}$ and with

$$\mu_{\boldsymbol{\varepsilon}[s]} = \frac{1}{N_o} \sum_{i=1}^{N_o} \varepsilon_i[s]$$

and

$$\sigma_{\boldsymbol{\varepsilon}[s]} = \sqrt{\frac{1}{N_o} \sum_{i=1}^{N_o} \left(\varepsilon_i[s] - \mu_{\boldsymbol{\varepsilon}[s]}\right)^2}$$

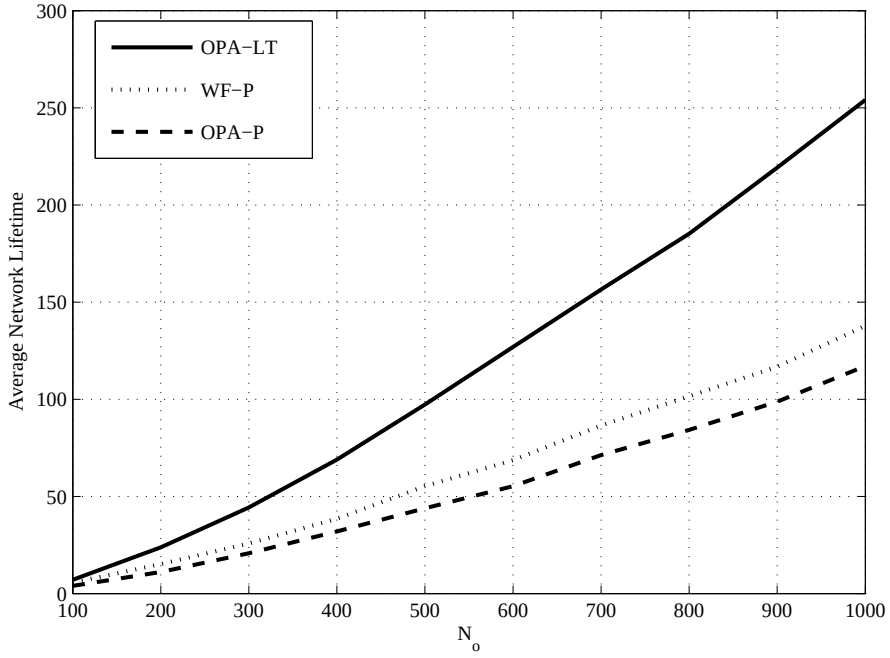


Figure 3.14: Average network lifetime vs. network size ($D_{\mathrm{T}} = 0.001$, $\sigma_v^2 = 0.01$, $\sigma_w^2 = 0.1, \varepsilon_0 = 10$).

denoting the mean and the standard deviation of the vector of residual energies. In Fig. 3.15, one clearly observes that both strategies, OPA-P and OPA-LT, yield similar energy dispersion values in *young* networks. The explanation for this behavior is quite straightforward: during the first iterations all sensors have approximately the same residual energies, i.e. the energy dispersion is already low, and, hence, the scheduler for both solutions relies mostly on the CSI. However, as time elapses the OPA-LT scheme effectively exploits the REI information and appropriately balances the residual energy in the network, this resulting into lower values of χ_ε. More formally, such balancing is carried out through *i*) the introduction of the harmonic mean of the energy vector into the threshold given by (3.57); and *ii*) the fact that the r.v. which is checked against such threshold encompasses the product of CSI *and* local REI.

3.8 Power allocation strategies for hierarchical sensor networks

So far, we have considered a flat network topology where all the sensors transmit their observations to a *single* coordinator, i.e the FC. Notwithstanding, in situations where there exists a strong path loss between the sensors and the FC, e.g. due to a large distance between the FC and the sensor nodes, flat networks might not be appropriate. For this reason, in this section we re-visit the problem of decentralized parameter estimation to analyze what power allocation

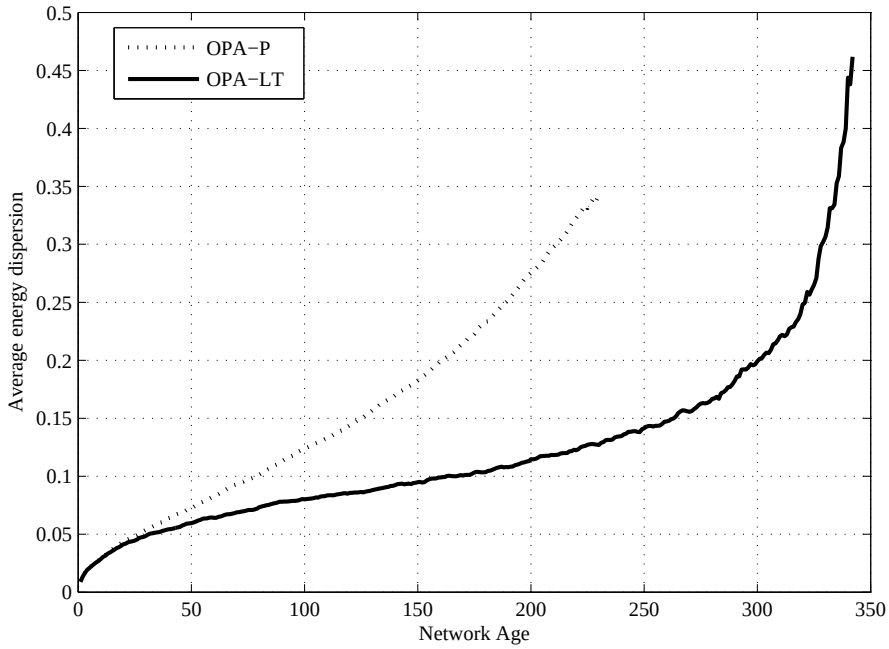


Figure 3.15: Energy dispersion vs. network age ($D_{\mathrm{T}} = 0.001, \varepsilon_0 = 30, \sigma_v^2 = 0.01, \sigma_w^2 = 0.1, N_o = 500$).

strategies are more suitable for hierarchical sensor networks.

Hierarchical topologies for Wireless Sensor Networks have been addressed in a number of works (see [62] and references therein), where the purpose of clustering is either to minimize the number of hops to the FC or to consolidate the amount of data sent. In our context, a hierarchical structure is mostly introduced in order to reduce the complexity of the system, in terms of CSI, and in some cases to increase the accuracy of the estimates.

3.8.1 Network Model

Again, our goal is to estimate a scalar, slowly-varying and spatially-homogenous parameter θ. To this aim, we adopt a hierarchical structure which is composed of the following network elements:

- **Sensors,** which are energy-constrained devices mainly aimed at sampling the unknown parameter θ. The N_0 sensors nodes in the WSN are grouped into N_c clusters of size N (namely, $N_0 = NN_c$).
- **Cluster-heads**: The purpose of the cluster-head is two-fold: to coordinate the sensors in the cluster in order to obtain a local estimate of θ and, also, to transmit such an estimate

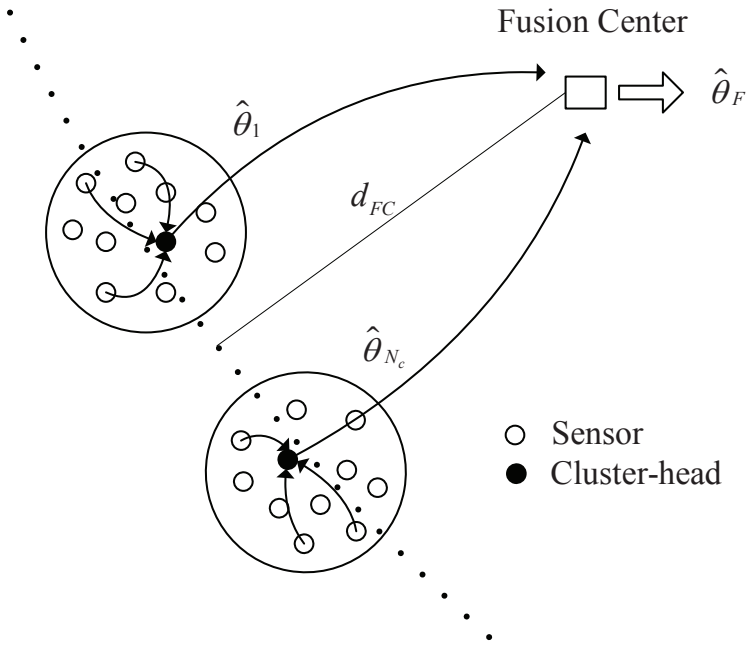


Figure 3.16: Hierarchical organizations of sensors

to the FC. As detailed in Section 3.8.3, the sensors within each cluster take turns in becoming cluster-heads.

- **Fusion Center**: Its main task is to coordinate the N_c cluster-heads and, also, to provide the final estimate of the parameter θ to the user.

Hence our hierarchical WSN is organized in two layers. The first (i.e. lower) layer is composed of the N_c clusters and their corresponding sensor nodes. The second (i.e. upper) layer encompasses the N_c cluster-heads and the fusion center. Again, we consider orthogonal transmissions by which each sensor in the first layer uses an orthogonal channel to convey its observation to the cluster-head, this resulting into a maximum of $N - 1$ orthogonal channels per cluster. The cluster-head could just send again the entire vector of observations to the FC but, clearly, this would result into a waste of resources in layer 2. Instead, we adopt a more scalable *estimate*-and-*forward* strategy by which each cluster-head re-transmits its local estimate. As a result, the maximum number of orthogonal channels required in layer 2 is restricted to N_c, regardless of the network size ($N_0 >> N_c$).

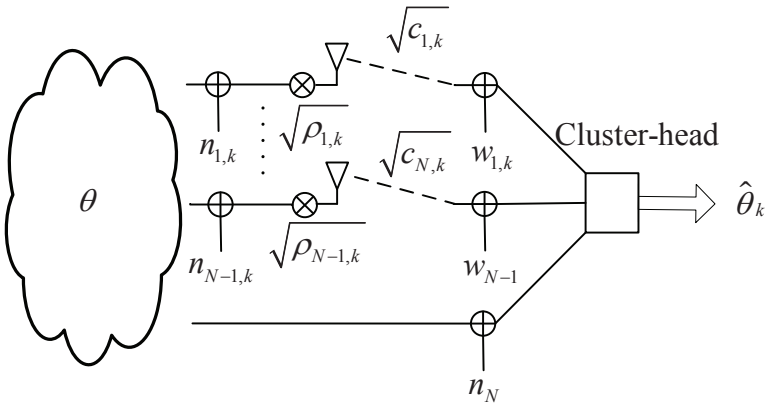


Figure 3.17: System model.

3.8.2 Distortion analysis

Layer 1

The observation at sensor i in the j-*th* cluster can be expressed as

$$x_{i,j} = \theta + v_{i,j}. \tag{3.58}$$

where the random variable $v_{i,j}$ denotes AWGN noise of variance σ_v^2 (i.e. $n_i \sim \mathcal{CN}(0, \sigma_v^2)$). Again, in each sensor the observation is scaled by a factor $\sqrt{\rho_{i,j}}$ before being transmitted to the cluster-head (i.e. *amplify*-and-*forward*). In the sequel, we assume non-frequency selective Rayleigh block-fading and, further, pair-wise synchronization between *i*) each sensor node and the cluster-head and, *ii*) between each cluster-head and the FC. Hence, the signal received at the j-*th* cluster-head (see Fig. 3.17) can be written as:

$$y_{i,j} = \sqrt{\rho_{i,j}}\sqrt{c_{i,j}}\,(\theta + v_{i,j}) + w_{i,j} \tag{3.59}$$

where $w_{i,j}$ stands for i.i.d. AWGN (i.e. $w \sim \mathcal{CN}(0, \sigma_w^2)$) and $c_{i,j}$ denotes the channel power gain which is modeled as an exponentially-distributed random variable with mean μ_c. Furthermore, we assume that such channel gains are i.i.d across sensors and, there is no path-loss within the clusters (i.e. $\mu_c = 1$). In each time-slot, $N' \leq N$ sensors transmit their observations to the cluster-head over a set of orthogonal channels (e.g. FDMA) and, thus, the $(N'+1) \times 1$ received signal vector $\mathbf{y}_j$ reads

$$\mathbf{y}_j = \mathbf{h}_j\theta + \mathbf{z}_j, \tag{3.60}$$

with $\mathbf{h}_j = \left[\sqrt{\rho_{1,j}c_{1,j}}, \ldots, \sqrt{\rho_{N',j}c_{N',j}}, 1\right]^T$ and $\mathbf{z}_j$ standing for AWGN with (diagonal) covariance matrix $\mathbf{C}_j$ given by $\mathrm{diag}\,[\mathbf{C}_j] = [\rho_{1,j}c_{1,j}\sigma_v^2 + \sigma_w^2, \ldots, \rho_{N',j}c_{N',j}\sigma_v^2 + \sigma_w^2, \sigma_v^2]^T$. The last element in $\mathbf{h}_j$ and $\mathrm{diag}\,[\mathbf{C}_j]$ accounts for the effect of the local observation at the cluster-head which is also capable of taking measurements. The BLUE [18] estimate at each cluster-head can be computed as

$$\hat{\theta}_j = \left(\mathbf{h}_j^T\mathbf{C}_j^{-1}\mathbf{h}_j\right)^{-1}\mathbf{h}_j^T\mathbf{C}_j^{-1}\mathbf{y}_j. \tag{3.61}$$

with variance given by

$$\mathsf{Var}(\hat{\theta}_j) = \mathbb{E}\left[\left(\hat{\theta}_j - \theta\right)^2\right] = \left(\mathbf{h}_j^T \mathbf{C}_j^{-1} \mathbf{h}_j\right)^{-1}. \tag{3.62}$$

Since matrix $\mathbf{C}_j$ is diagonal, the equation above can be written in compact form as

$$\sigma_j^2 = \mathsf{Var}(\hat{\theta}_j) = \left(\sum_{i=1}^{N'} \frac{\rho_{i,j} c_{i,j}}{\rho_{i,j} c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2}\right)^{-1}. \tag{3.63}$$

The BLUE estimator is unbiased and, thus, the resulting estimate at the j*-th* cluster-head can be modeled as,

$$\hat{\theta}_j = \theta + e_j, \tag{3.64}$$

where e_j denotes AWGN noise with variance σ_j^2.

Layer 2

Each cluster-head re-transmits its local estimate scaled by a factor $\sqrt{\psi_k}$ over one of the orthogonal channels. Hence, the signal received at the FC from the j*-th* cluster-head reads

$$r_j = \sqrt{\psi_k g_j^*}\,(\theta + e_j) + w_j, \tag{3.65}$$

where w_j stands for i.i.d. AWGN (i.e. $w \sim \mathcal{CN}(0, \sigma_w^2)$) and g_j^* denotes the channel power gain from the cluster-head to the FC. Again, we assume that the channel gains are i.i.d across cluster-heads but, unlike in the intra-cluster case, we introduce a path-loss model. Hence, g_j^* is selected from a set of N i.i.d exponentially-distributed random variables with mean $\mu_g = d_{FC}^{-\delta}$, with δ standing for the path-loss coefficient, and d_{FC} denoting the distance from the clusters to the FC. It is worth noting that g_j^* will actually depend on the cluster-head selection method (see Section 3.8.3). In each time-slot, $N_c' \leq N_c$ cluster-heads re-transmit their observations to the FC over a set of orthogonal channels and, finally, the $N_c' \times 1$ received signal vector $\mathbf{r}$ reads

$$\mathbf{r} = \mathbf{b}\theta + \boldsymbol{\nu}, \tag{3.66}$$

where $\mathbf{b}$ is a $N_c' \times 1$ column vector defined as $\mathbf{b} = \left[\sqrt{\psi_1 g_1}, \ldots, \sqrt{\psi_{N_c'} g_{N_c'}}\right]^T$ and, $\boldsymbol{\nu}$ is AWGN with (diagonal) covariance matrix given by $\mathrm{diag}\left[\mathbf{C}_\nu\right] = \left[\psi_1 g_1 \sigma_1^2 + \sigma_w^2, \ldots, \psi_{N_c'} g_{N_c'} \sigma_{N_c'}^2 + \sigma_w^2\right]^T$. The variance of the BLUE estimator at the FC, which we will take as a distortion measure D_F, is given by:

$$D_F = \mathsf{Var}(\hat{\theta}_F) = \left(\sum_{j=1}^{N_c'} \frac{\psi_j g_j^*}{\psi_j g_j^* \sigma_j^2 + \sigma_w^2}\right)^{-1}, \tag{3.67}$$

with

$$\hat{\theta}_F = \left(\mathbf{b}^T \mathbf{C}_\nu^{-1} \mathbf{b}\right)^{-1} \mathbf{b}^T \mathbf{C}_\nu^{-1} \mathbf{r}. \tag{3.68}$$

From (3.63) and (3.67), one concludes that the overall distortion is a function of the power allocation in *both* layers. Therefore, for a given sum-power constraint there exists a trade-off between the fraction of power allocated to every cluster in layer 1 (and their associated estimation variances), and the power allocated to layer 2 (which also impacts on the overall distortion). In Section 3.8.4, we derive some strategies aimed at carefully balancing the power allocation between layers in such hierarchical organizations.

3.8.3 Selection of the cluster-head

Let $g_{i,j}$ denote the sensor-to-FC channel gain of the i-*th* sensor in cluster j. At each time instant, the sensor in each cluster experiencing the most favorable channel conditions becomes the cluster-head, that is[11]

$$g_j^* = \max_i \{g_{i,j}\} \qquad 1 \leq j \leq N_c. \tag{3.69}$$

Note that g_j^* is the first order statistic of an exponential parent distribution drawn from a population of size N. Hence, its pdf is given by

$$f_{g_k^*}(x) = NF(x)^{N-1} f(x), \tag{3.70}$$

where $F(x)$ and $f(x)$ stand for the CDF and pdf of $g_{i,j}$. This cluster-head selection method has two advantages: *i*) the sensor experiencing the most favorable channel conditions is the one which actually conveys the information to the FC, this resulting in a lower final distortion; and, *ii*) the selection method is fair, since each sensor has the same probability of becoming a cluster-head (i.e. the energy is uniformly spent over sensors).

[11]This can be accomplished in a decentralized way (i.e. without participation of the FC) by resorting to the distributed back-off strategy proposed in [74]. Besides, we assume a TDD (Time Division Duplex) duplexing scheme, thus, the uplink channel gains can be derived from the downlink estimates obtained with the pilot symbols broadcasted by the FC.

3.8.4 Hierarchical power allocation strategies

Accordingly, the optimization problem can be posed as follows,

$$\min_{\psi_1,\ldots,\psi_{N_c},\rho_{1,1},\ldots,\rho_{N-1,N_c}} \left(\sum_{j=1}^{N_c} \frac{\psi_i g_j^*}{\psi_i g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1} \tag{3.71}$$

$$\text{s.t}: \quad \sum_{j=1}^{N_c} \left(\psi_j + \sum_{i=1}^{N-1} \rho_{i,j} \right) \leq P_{\text{T}} \tag{3.72}$$

$$\text{For} \quad 1 \leq j \leq N_c \quad \sigma_j^2 = \left(\sum_{i=1}^{N-1} \frac{\rho_{i,j} c_{i,j}}{\rho_{i,j} c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1}. \tag{3.73}$$

The above problem is barely tractable since the optimization variables, $\psi_1, \ldots, \psi_{N_c}$, $\rho_{1,1}, \ldots, \rho_{N-1,N_c}$, are coupled through the sum-power constraint (3.72). Furthermore, a solution to (3.71) would entail coordination and CSI exchange between layers which can be costly and/or impractical. Consequently, we introduce an additional parameter $\alpha \in [0, 1)$ which determines the percentage of transmit power allocated to each layer. By doing so, we can decouple the sum-power constraint leading to the following simplified problem:

$$\min_{\psi_1,\ldots,\psi_{N_c},\rho_{1,1},\ldots,\rho_{N-1,N_c}} \left(\sum_{j=1}^{N_c} \frac{\psi_i g_j^*}{\psi_i g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1} \tag{3.74}$$

$$\text{s.t}: \quad \sum_{j=1}^{N_c} \psi_j \leq (1-\alpha) P_{\text{T}} \tag{3.75}$$

$$\text{For} \quad 1 \leq j \leq N_c \quad \sum_{i=1}^{N-1} \rho_{i,j} \leq \alpha \frac{P_{\text{T}}}{N_c} \tag{3.76}$$

$$\sigma_j^2 = \left(\sum_{i=1}^{N-1} \frac{\rho_i c_{i,j}}{\rho_i c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1}. \tag{3.77}$$

Note that in the expression above we have introduced an *individual* power constraint for *each* cluster in layer 1. This reflects a situation where each cluster allocates power independently from the remaining ones. Furthermore, the N_c individual constraints in (3.76) are identical since so are clusters and, in addition, the number of sensors in each cluster is high (i.e. clusters are *statistically* identical). From all the above, we can decompose the minimization problem in

the following way [69]:

$$\min_{\psi_1,\ldots,\psi_{N_c}} \min_{\rho_{1,1},\ldots,\rho_{N-1,1}} \cdots \min_{\rho_{1,N_c},\ldots,p_{N-1,N_c}} \left(\sum_{j=1}^{N_c} \frac{\psi_i g_j^*}{\psi_i g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1}$$

s.t :

$$\sum_{j=1}^{N_c} \psi_j \leq (1-\alpha) P_{\mathrm{T}} \tag{3.78}$$

For $1 \leq j \leq N_c$

$$\sum_{i=1}^{N-1} \rho_{i,j} \leq \alpha \frac{P_{\mathrm{T}}}{N_c} \tag{3.79}$$

$$\sigma_j^2 = \left(\sum_{i=1}^{N-1} \frac{\rho_{i,j} c_{i,j}}{\rho_{i,j} c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1}. \tag{3.80}$$

Since σ_j^2 exclusively depends on $\rho_{1,j}, \ldots, \rho_{N-1,j}$, then it is straightforward to show that the optimization problem can be decomposed into $N_c + 1$ parallel problems:

$$\min_{\rho_{1,j},\ldots,\rho_{N-1,j}} \left(\sum_{i=1}^{N-1} \frac{\rho_{i,j} c_{i,j}}{\rho_{i,j} c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1}$$

s.t :

$$\sum_{i=1}^{N-1} \rho_{i,j} \leq \alpha \frac{P_t}{N_c} \tag{3.81}$$

for $1 \leq j \leq N_c$ and, also,

$$\min_{\psi_1,\ldots,\psi_{N_c}} \left(\sum_{j=1}^{N_c} \frac{\psi_j g_j^*}{\psi_j g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1}$$

s.t :

$$\sum_{j=1}^{N_c} \psi_j \leq (1-\alpha) P_T. \tag{3.82}$$

As commented above, α plays an important role in the optimization problem. In our analysis, we will determine its optimum value, α^*, on the basis of partial (i.e. statistical) CSI only, namely,

$$\min_{\alpha} \left\{ \mathbb{E}_{\left\{\{c_{i,j}\}_{i=1}^{N-1}\right\}_{j=1}^{N_c}, \{g_j^*\}_{j=1}^{N_c}} \left[\left(\sum_{j=1}^{N_c} \frac{\psi_j g_j^*}{\psi_j g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1} \right] \right\}$$

s.t :

$$\sigma_j^2 = \left(\sum_{i=1}^{N-1} \frac{\rho_{i,j} c_{i,j}}{\rho_{i,j} c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1}; \quad 1 \leq j \leq N_c$$

$$\alpha \in [0, 1).$$

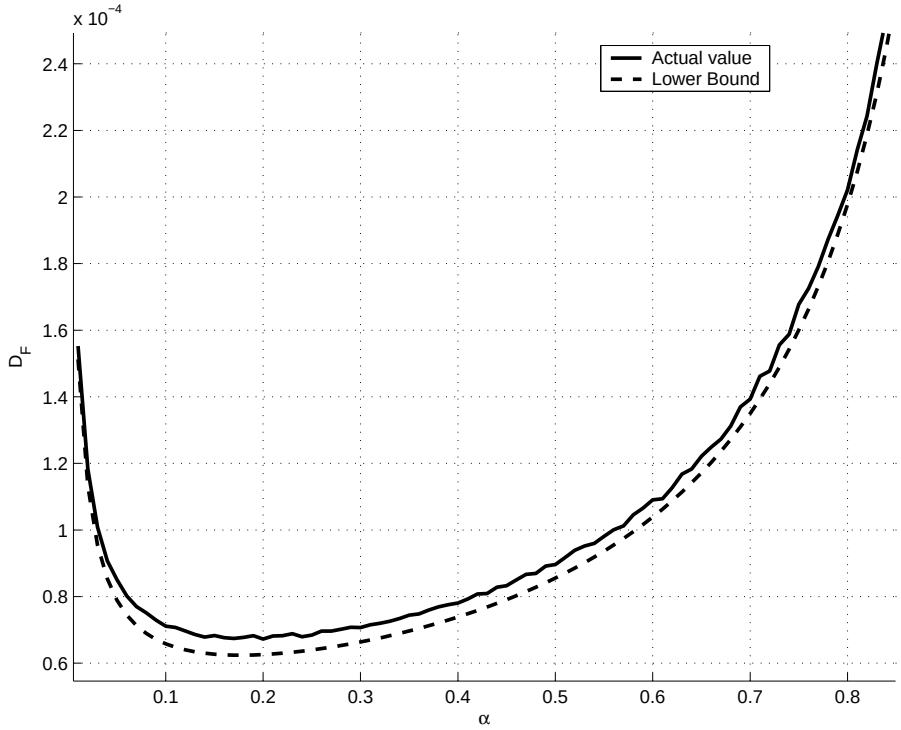


Figure 3.18: Actual distortion and lower bound for the UPA case as a function of α ($N = 40, N_c = 4, P_{\mathrm{T}} = 500, \sigma_v^2 = 0.001, \sigma_w^2 = 0.001$).

From the expression above, one concludes that the optimal power split will depend on a number of system parameters (such as N_c, N, etc), the power allocation rule (i.e. uniform, waterfilling) through its dependency on $\rho_{i,j}$ and ψ_j and, also, on statistical CSI. As an example, the optimal power split α^* will attempt to compensate for the path-loss effects between the cluster-heads and the FC by allocating more power to layer 2 (i.e. by forcing α^* to take smaller values).

In the following subsection, we compute the optimal power split between layers for the uniform power allocation (UPA) case and, next, we discuss hybrid solutions featuring optimal (i.e. waterfilling) power allocation scheme in at least one out of the two layers.

Uniform power allocation in both layers

In the absence of CSI at the cluster-heads *and* the FC, the best thing one can do is to uniformly allocate the transmit power. Hence, each sensor transmits with power $\alpha\frac{P_{\mathrm{T}}}{(N-1)N_c}$, and each cluster-head with $(1-\alpha)\frac{P_{\mathrm{T}}}{N_c}$. The optimal fraction of power α^* is the one which minimizes the following expression,

$$\min_{\alpha} \left\{ \mathbb{E}\left[\left(\sum_{j=1}^{N_c} \frac{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} g_j^*}{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1} \right] \right\} \tag{3.83}$$

$$\text{s.t}:$$
$$\sigma_j^2 = \left(\sum_{i=1}^{N-1} \frac{\alpha \frac{P_{\mathrm{T}}}{(N-1)N_c} c_{i,j}}{\alpha \frac{P_{\mathrm{T}}}{(N-1)N_c} c_{i,j} \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1} ; 1 \leq j \leq N_c$$
$$\alpha \in (0,1).$$

Unfortunately, the resulting optimization problem is barely tractable and, in general, does not have a closed-form solution. Instead, we will compute a lower bound for the cost function in (3.83):

$$\mathbb{E}_{\left\{\{c_{i,j}\}_{i=1}^{N-1}\right\}_{j=1}^{N_c}, \{g_j^*\}_{j=1}^{N_c}} \left[\left(\sum_{j=1}^{N_c} \frac{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} g_j^*}{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} g_j^* \sigma_j^2 + \sigma_w^2} \right)^{-1} \right]$$
$$\geq \mathbb{E}_{\left\{\{c_{i,j}\}_{i=1}^{N-1}\right\}_{j=1}^{N_c}} \left[\left(\sum_{j=1}^{N_c} \frac{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} \mu_{g^*}}{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} \mu_{g^*} \sigma_j^2 + \sigma_w^2} \right)^{-1} \right]$$
$$\geq \left(\frac{(1-\alpha) P_{\mathrm{T}} \mu_{g^*}}{(1-\alpha)\frac{P_{\mathrm{T}}}{N_c} \mu_{g^*} \overline{D} + \sigma_w^2} \right)^{-1} \tag{3.84}$$

with,

$$\overline{D} = \left(\frac{\alpha \frac{P_{\mathrm{T}}}{N_c} \mu_c}{\alpha \frac{P_{\mathrm{T}}}{(N-1)N_c} \mu_c \sigma_v^2 + \sigma_w^2} + \frac{1}{\sigma_v^2} \right)^{-1} \tag{3.85}$$

and where the expectation of the first order statistic of an exponential parent distribution, μ_{g^*}, can be efficiently computed as [70, Chapter 3]

$$\mu_{g^*} = \mathbb{E}\{g_j^*\} = \sum_{k=1}^{N} \frac{\mu_g}{k}. \tag{3.86}$$

The first inequality follows from the fact that $\mathbb{E}[g(x)^{-1}] \geq 1/g(\mathbb{E}[x])$ provided that $g(x)$ is a positive and concave function in x. The second inequality is due to the fact that the argument in the expectation is convex in the sequence of random variables $c_{i,k}$. Finally, one can prove that (3.84) is convex in α and, hence, by setting its first derivative to zero we can obtain its optimal value as follows:

$$\alpha^* = \left[\frac{(N-1)\left(-\sqrt{\mu_{g^*}\mu_c^3} P_{\mathrm{T}} \sigma_v^2 - N_c \sigma_w^2 \sqrt{\mu_c} (N-1) \sqrt{\mu_{g^*}} + \mu_c \left(N \mu_{g^*} P_{\mathrm{T}} \sigma_v^2 + N N_c \sigma_w^2 - \mu_{g^*} \right) \right)}{P_{\mathrm{T}} \mu_c \sigma_v^2 \left(N^2 \left(\mu_{g^*} - \mu_c \right) + \mu_{g^*} \left(1 - 2N \right) \right)} \right]^+ \tag{3.87}$$

where $[x]^+ = \max\{x, 0\}$.

In Figure 3.18, we depict the lower bound in (3.84) versus its actual value as a function of the power split. Clearly, the bound is tight for the whole range of α values and, as a result, only marginal performance loss can be expected when approximating the true value α^* by the one obtained with the lower bound.

Hybrid WF-UPA solutions

If we can assume that full CSI is locally available either at the cluster-heads or at the FC (or in both), then we can compute the optimal power allocations by solving equations (3.81) and (3.82), respectively. According to Section 3.3.1 such optimal solutions are given by:

$$\rho_{i,j}^{*} = \frac{\sigma_w^2}{\sigma_v^2 c_{i,j}} \left[\sqrt{\frac{c_{i,j}}{\lambda_j^{*} \sigma_w^2}} - 1 \right]^{+} \tag{3.88}$$

$$\psi_j^{*} = \frac{\sigma_w^2}{\sigma_j^2 g_j} \left[\sqrt{\frac{g_j}{\beta^{*} \sigma_w^2}} - 1 \right]^{+}, \tag{3.89}$$

where β^* and λ_j^* ($j = 1 \dots N_c$) stand for the optimal water-levels which must be computed numerically as in [11]. For this very same reason, the optimum α^* does not admit a closed-form expression anymore and, thus, we must resort to numerical methods.

3.8.5 Simulations and numerical results

As far as computer simulations are concerned, we consider a network with $N_c = 4$ clusters and $N = 40$ sensors in each (i.e. $N_o = 160$ sensors in total). For the wireless links between the cluster-heads and the FC, we assume a path-loss coefficient $\delta = 2$. In Fig. 4, we depict the overall distortion attained by the different combinations of *i*) hierarchical and flat (i.e. non-hierarchical) network models, and *ii*) power allocation schemes used in each layer. The cases with a flat network model are used for benchmarking purposes only.

To start with, one can clearly observe the huge gap between the UPA/UPA (i.e. uniform power allocation in both layers) and UPA (i.e. flat network structure and UPA scheme) curves for the whole range of distances to the FC. The introduction of a network hierarchy and the computation of the optimal power split turn out to be very useful in ensuring that the transmit power is efficiently spent in obtaining accurate estimates in layer 1 clusters, rather than in forcing every sensor to overcome the severe path-loss in the wireless links to the FC (to recall, in the hierarchical case this task is conducted by the cluster heads, only). Besides, the performance exhibited by the UPA/UPA scheme is comparable to (or, in some cases, even slightly better than) that of a flat network scheme with WF power allocation which, additionally, would require full CSI at the FC. Indeed, some additional gain can be obtained by using WF in the second layer (i.e. UPA/UPA vs. UPA/WF) curves. In the light of the increased CSI requirements at the FC, though, such gain can be regarded as marginal, in particular, for $d_{FC} < 120$ m, namely, low-to-mid values. However, as we increase the cluster-head-to-FC distance and, consequently, decrease the SNR in layer 2, the use of WF in the second layer becomes more and more necessary (asymptotically, a single cluster-head would send data to the FC only).

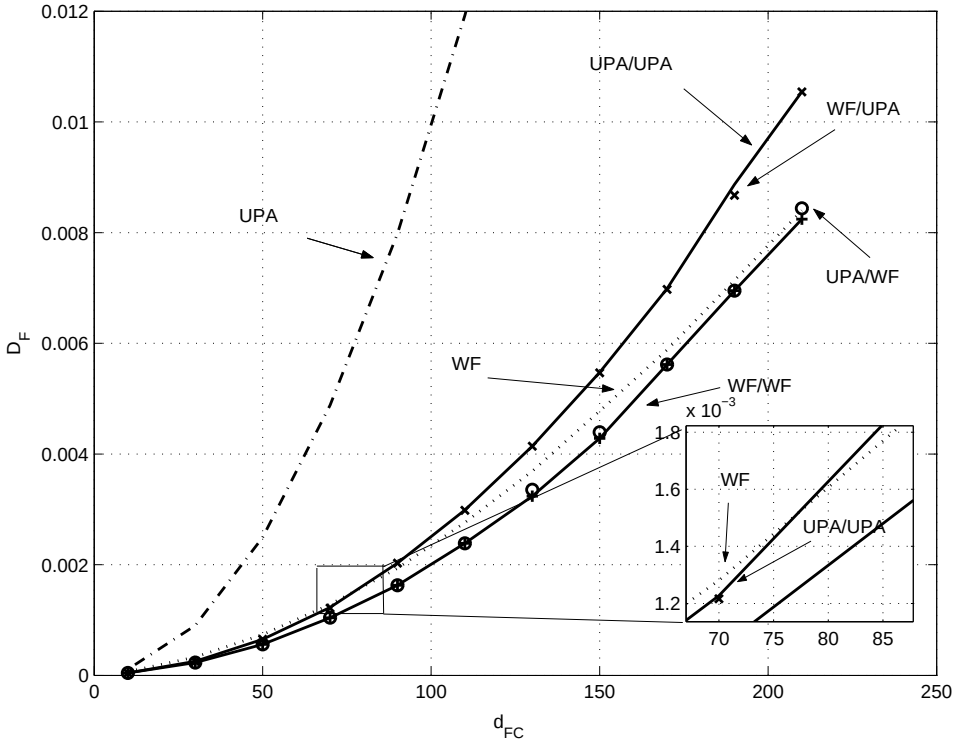


Figure 3.19: Overall distortion attained by the different schemes. The curves labeled with X/Y correspond to the cases with hierarchical structures, with X and Y denoting the power allocation scheme (i.e. UPA or WF) in the first and second layers, respectively. The curves labeled with only one power allocation scheme correspond to the benchmark cases with flat network structure. ($N = 40$, $N_c = 4$, $P_{\mathrm{T}} = 500$, $\sigma_v^2 = 10^{-3}$, $\sigma_w^2 = 10^{-3}$).

Next, one can clearly observe that, irrespectively of the power allocation strategy adopted in layer 2, little or no improvement results from the use of WF in layer 1. This is due to the fact that the average SNR in layer 1 is potentially high, both because of the low noise level ($\sigma_w^2 = 10^{-3}$) and the absence of path-loss effects.

Finally, it is worth noting that the UPA/WF (or WF/WF) and WF curves converge for an increasing d_{FC}. Certainly, when path-loss effects becomes more severe, the optimal power split for the hierarchical schemes reserves virtually no power for layer 1 ($\alpha^* \to 0$). In these circumstances, cluster-heads merely rely on their own parameter estimate (to recall, cluster-head takes measurements, as well) which is then conveyed to the FC. As long as the WF scheme starts activating N_c or less sensors, both the hierarchical and flat solutions are formally identical. The same effect can be observed when the measurement noise (σ_v^2) tends to zero. When this occurs, the additional power required to further decrease the estimation noise in the cluster-heads (see equation (3.73)) is too high and, thus, power is better spent in layer 2 (i.e. $\alpha^* \to 0$ again).

3.9 Chapter summary and conclusions

In this chapter, we have addressed the problem of power allocation in a context of amplify-and-forward WSNs for decentralized parameter estimation. First, we have focused on a *flat* topology and we have proposed a class of opportunistic power allocation schemes. In such OPA schemes, only sensors experiencing favorable conditions (i.e. above a threshold) participate in the estimation process by adjusting their transmit power on the basis of local CSI (and, in some cases, REI) information only. We have addressed a number of classical problems of interest such as the minimization of estimation distortion, the minimization of transmit power or the enhancement of network lifetime. To that extent, we have particularized the general OPA framework to each problem of interest (OPA-D, OPA-P and OPA-LT, respectively) which entails the derivation of an approximate but tight closed-form expression of the threshold along with the corresponding power allocation rule. Furthermore, we have also addressed the case with imperfect CSI to derive an improved version of the OPA-D scheme (OPA-DR) which is *robust* to such imperfections. Computer simulation results reveal that with OPA-D the overall distortion is $150 - 280\%$ lower than that of UPA and virtually identical to that of the optimal waterfilling (WF-D) scheme. More significantly, we have proved that the rate at which distortion decreases for the OPA-D and WF-D is identical when the number of sensor nodes grows without bound. To stress, that the signaling and CSI requirements for all OPA schemes are far more relaxed than those of WF approaches. We have also observed that the robust version performs close to systems operating with perfect CSI even with moderate values of CSI uncertainty ($\Delta_e = 0\text{dB}$). From the comparison of OPA-LT with OPA-P, we have concluded that OPA-LT leads to a two-fold extension of the network lifetime (due to a more balanced energy consumption over sensors) at the expense of a slight increase in the transmit power ($8 - 10\%$). Finally, we have found that, when compared to WF approaches, the whole family of OPA schemes tend to activate a lower number of sensors. This is beneficial in terms of the number of orthogonal channels needed and hence, the delay.

Finally, we have adopted a hierarchical network structure in which each cluster is governed by a cluster-head which is the sensor experiencing the most favorable channel conditions to the FC. Cluster-heads gather the measurements in their respective clusters and, then, forward the cluster estimate to the FC. We have shown that by balancing the power between both layers in the hierarchy, the minimization problem can be decomposed into smaller sub-problems and, furthermore, we have derived a closed-form expression for the optimal power split between layers for the UPA case. Computer simulation results reveal that a hierarchical network with UPA schemes in both layers constitutes the best trade-off in terms of the performance (namely, estimation accuracy) vs. CSI requirements for almost the whole range of cluster-head-to-FC distance values. For the scenario under consideration, we have also shown that using WF in the second layer pays-off whereas it renders no or little gain in layer 1.

3.A Appendix

3.A.1 Derivation of $\check{\gamma}_{\mathrm{th}}^{*}$'s domain

We want to identify in what circumstance $\check{\gamma}_{\mathrm{th}}^{*} \in [0, \mu_c \log(D_{\mathrm{T}} N_o/\sigma_v^2))$, with $\check{\gamma}_{\mathrm{th}}^{*}$ given by (3.45). Since, by definition, $\check{\gamma}_{\mathrm{th}}^{*}$ takes only positive values, only the upper value of the range has to be checked, namely

$$\mu_c \mathsf{W}_0\left(\frac{D_{\mathrm{T}} N_o e^2}{\sigma_v^2}\right) - 2\mu_c < \mu_c \log\left(\frac{D_{\mathrm{T}} N_o}{\sigma_v^2}\right). \tag{3.90}$$

Bearing in mind that $2 = \log(e^2)$, we can re-write the inequality above as:

$$\mathsf{W}_0\left(\frac{D_{\mathrm{T}} N_o e^2}{\sigma_v^2}\right) < \log\left(\frac{D_{\mathrm{T}} N_o e^2}{\sigma_v^2}\right). \tag{3.91}$$

which holds true if and only if the argument in the W_0 and log functions is greater than e or, equivalently, $N_o > \frac{\sigma_v^2}{D_{\mathrm{T}} e}$.However, in the process of deriving the lower bound we already imposed $N \geq \lceil \sigma_v^2/D_{\mathrm{T}} \rceil$. Since $N_o \geq N$, the inequality above automatically holds true and, thus, the interval $[0, \mu_c \log(D_{\mathrm{T}} N_o/\sigma_v^2))$ is the only possible domain of γ_{th}^{*}.

3.A.2 Proof of the convergence of OPA-DR to UPA for poor quality CSI

From equation (3.37), we have to show that the threshold $\check{\gamma}_{\mathrm{th}}^{*} \to 0$ as $\sigma_e^2 \to \infty$. Consider first the asymptotic expansion of the Lambert function $\mathsf{W}_0(z)$ as given in [71]:

$$\mathsf{W}_0(z) = \log z - \log\log z + \nu. \tag{3.92}$$

Hereafter, we neglect the contribution of the ν term since it vanishes for $z \to \infty$ (which holds true for $\sigma_e^2 \to \infty$). If we define $z = c\sqrt{xe^x}$ with c being a constant and positive value, we only have to prove (see equation (3.37)) that

$$\mathsf{W}_0\left(c\sqrt{xe^x}\right) - \frac{x}{2} \leq 0 \tag{3.93}$$

when x grows without bound. By using the asymptotic expansion presented above, we can derive the following inequality:

$$\begin{aligned}
\mathsf{W}_0\left(c\sqrt{xe^x}\right) - \frac{x}{2} &= \log\left(c\sqrt{xe^x}\right) - \log\log\left(c\sqrt{xe^x}\right) - \log\left(\sqrt{e^x}\right) \\
&= \log\left(c\sqrt{x}\right) - \log\log\left(c\sqrt{xe^x}\right) \\
&\leq \log\left(c\sqrt{x}\right) - \log\left(\log c + \frac{1}{2}x\right).
\end{aligned} \tag{3.94}$$

Clearly, the right hand of the last expression takes negative values for sufficiently high x, which concludes the proof.

3.A.3 Proof of $\Pr\{N > \lceil \sigma_v^2/D_{\mathrm{T}} \rceil\} \to 1$ for large N_o

Let $\alpha = N_{\mathsf{cen}}/N_o$ be the ratio between the number of observations needed in a centralized scenario $N_{\mathsf{cen}} = \lceil \sigma_v^2/D_{\mathrm{T}} \rceil$ and the total number of sensors in the network N_o. We want to show that $\Pr\{N \leq N_{\mathsf{cen}}\} \underset{N_o, N_{\mathsf{cen}} \to \infty}{\longrightarrow} 0$. That is, we decrease the distortion target as the population grows. Hence, by resorting to the Chernoff's bound for a binomial random variable, the probability of the event of interest can be now upper bounded as

$$\Pr\{N \leq N_{\mathsf{cen}}\} \leq e^{-\frac{1}{2}N_o(p-\alpha)^2} \tag{3.95}$$

where p is the probability of activation and can be easily computed by using the optimal threshold of (3.45). The bound of (3.95) is only valid if $p \geq \alpha$ and, hence, we only need to satisfy

$$p = \alpha \mathsf{W}_0\left(\frac{e^2}{\alpha}\right) > \alpha$$

which holds true for $\alpha < e$. Therefore, it is now straightforward to show that

$$\Pr\{N \leq N_{\mathsf{cen}}\} \leq e^{-\frac{1}{2}N(p-\alpha)^2} \underset{N_o, N_{\mathsf{cen}} \to \infty}{\longrightarrow} 0. \tag{3.96}$$

3.A.4 Proof of inequality (3.50)

Here, we want to prove that, for a given cut-off value γ', the following inequality holds true

$$\mathbb{E}_{N;\gamma_{\mathrm{th}}[s]|\boldsymbol{\varepsilon}[s]}[N] = \sum_{i=1}^{N_o} e^{-\frac{\gamma_{\mathrm{th}}[s]}{\varepsilon_i[s]\mu_c}} \leq N_o e^{-\frac{\gamma_{\mathrm{th}}[s]}{\mu_c N_o'}\sum_{i=1}^{N_0'}\frac{1}{\varepsilon_i[s]}}. \tag{3.97}$$

for some range of values of N_o' and for all $\gamma_{\mathrm{th}}[s] \in [0, \gamma')$. Let f and g be functions defined as

$$f(\gamma_{\mathsf{th}}[s]) = N_o e^{-\frac{\gamma_{\mathrm{th}}[s]}{\mu_c N_o'}\sum_{i=1}^{N_0'}\frac{1}{\varepsilon_i[s]}} \quad \text{and} \quad g(\gamma_{\mathsf{th}}[s]) = \sum_{i=1}^{N_o} e^{-\frac{\gamma_{\mathrm{th}}[s]}{\varepsilon_i[s]\mu_c}}. \tag{3.98}$$

Clearly, f and g are continuous, convex and non-increasing functions in $\gamma_{\mathsf{th}}[s]$. Besides, we have that $f(0) = g(0)$ and, also, that their respective derivatives fulfill $f'(0) > g'(0)$. Hence, $f(\gamma)$ does act as an upper bound of $g(\gamma)$ for some positive range of values of γ around 0 (which, ultimately, depends on μ_c, N_o' and $\varepsilon_i[s]$).

Next, it is straightforward to verify that the above inequality always holds true for $N_o' = 1$. Besides, from Jensen's inequality, one concludes that the above expression is strictly false for $N_o' = N_o$. Thus, for each γ' there exists some N_o' which turns to be the highest value of in $[1, N_o)$ for which (3.97) holds true. Such value must be found with numerical methods.

3.A.5 Convergence of the distortion for the OPA scheme for large N_o

Recall the ratio of (3.23) as g_N. Therefore, we have to prove that

$$\lim_{N_0 \to \infty} \Pr\left(|g_N - 1| \geq \epsilon\right) = 0 \tag{3.99}$$

with $\epsilon > 0$. Hence, by substituting the definition of g_N in the argument of the limit we obtain

$$\Pr\left(|g_N - 1| \geq \epsilon\right) = \Pr\left(N \leq \frac{P_{\mathrm{T}} \check{\gamma}_{\mathsf{th}} \sigma_v^2}{\sigma_w^2 \epsilon}\right) \tag{3.100}$$

Clearly, N is a binomial r.v. with mean given by $\mathbb{E}[N] = N_o e^{-\check{\gamma}^*_{\mathsf{th}}} = \frac{e W_0^2\left(\beta\sqrt{N_o}\right)}{\beta^2}$, where β is a constant not depending on N or N_o. The probability of (3.100) can be upper-bounded by using Chernoff's bound as follows

$$\Pr\left(N \leq \frac{P_{\mathrm{T}} \check{\gamma}_{\mathsf{th}} \sigma_v^2}{\sigma_w^2 \epsilon}\right) \leq e^{-\frac{\left(\mathbb{E}[N] - \frac{P_{\mathrm{T}} \check{\gamma}_{\mathsf{th}} \sigma_v^2}{\sigma_w^2 \epsilon}\right)^2}{2\mathbb{E}[N]}}, \tag{3.101}$$

which holds true[12] for $\mathbb{E}[N] > \frac{P_{\mathrm{T}} \check{\gamma}_{\mathsf{th}} \sigma_v^2}{\sigma_w^2 \epsilon}$. After some algebra, one can easily prove that

$$\lim_{N_0 \to \infty} e^{-\frac{\left(\mathbb{E}[N] - \frac{P_{\mathrm{T}} \check{\gamma}_{\mathsf{th}} \sigma_v^2}{\sigma_w^2 \epsilon}\right)^2}{2\mathbb{E}[N]}} = 0 \tag{3.102}$$

which concludes the proof of (3.23).

[12] For large N_o one can easily prove that $\mathbb{E}[N] > \frac{P_{\mathrm{T}} \check{\gamma}_{\mathsf{th}} \sigma_v^2}{\sigma_w^2 \epsilon}$

Chapter 4

Encoding Schemes in Bandwidth-constrained Wireless Sensor Networks

In this chapter, we continue to focus on the problem of decentralized *parameter* estimation via WSNs. However, we go one step beyond the amplify-and-forward (i.e. *analog*) retransmission strategies and assume that sensors are capable of encoding their observations for *digital* transmission. In particular, we consider two different encoding strategies, namely, Quantize-and-Estimate (Q&E) and Compress-and-Estimate (C&E) and assume first that sensor observations are conveyed to the Fusion Center (FC) over a number of *orthogonal* Gaussian or Rayleigh-fading channels. We constrain both power *and* bandwidth to be constant irrespectively of the network size and find approximate closed-form expressions of the optimal number of sensors for a number of cases of interest. Besides, we derive the optimal encoding rate for the Q&E scheme when, in the absence of Transmit Channel State Information (CSIT), sensors must encode their observations at a *common* and *constant* rate. For the (successive) C&E strategy, we also determine the encoding order that minimizes the resulting distortion in the FC estimates. We complement the analysis by deriving an expression of the asymptotic distortion when the number of sensors grows without bound, and the rate at which distortion decreases in the high-SNR regime. Finally, we introduce contention-based control multiple-access protocols in the

system model. In the context of hierarchical networks we analyze the impact of the packet collisions that contention-based multiple-access protocols entail on the distortion in the estimates.

4.1 Introduction

As discussed in the previous chapter, in the context of Multiple-Access channels (MAC) an amplify-and-forward strategy is known to scale optimally in terms of estimation distortion. However for the case of *orthogonal* channels considered here, solutions based on separate source and channel coding outperform A&F-based ones. In this context, in [75] the source-channel separation is shown to be optimal for the quadratic Gaussian CEO problem.

As far as source coding is concerned, the work in [76] constitutes a generalization to sensor trees of Wyner and Ziv's pioneering study [37]. More precisely, the authors compare two different coding schemes, namely Quantize-and-Estimate (Q&E) and Compress-and-Estimate (C&E). The former is a particularization of Wyner-Ziv's problem to the case where no side information is available at the decoder; whereas the latter is a *successive* Wyner-Ziv-based coding strategy capable of exploiting the correlation among sensor observations. Yet sub-optimal, the performance of this successive encoding scheme is not far from that of other more complex joint (over sensors) coding strategies. Besides, in [77] it is shown to achieve any point of the rate-distortion region of the quadratic Gaussian CEO problem. In both cases, it is assumed that sensors experience *Gaussian* (or, more generally, *deterministic*) channels, this meaning that the transmissions rates are known and, consequently, all sensors convey their observations *reliably*. Going one step beyond, the authors in [78] derive the rate-distortion region for the case of erasure channels where a *fixed* and *known* number of observations k are reliably received. Nevertheless, when the erasure pattern is unknown and/or random, joint coding techniques or successive encoding techniques such as [76], exhibit poor performance. This problem is addressed in [23], where a number of tradeoffs between reliability and efficiency are examined for a *lossless* compression setting.

In scenarios with non-reciprocal (e.g. Frequency Division Duplexing (FDD)) fading channels, it is often assumed that only *statistical* CSI is available at the transmitter. Consequently, the encoding rate at the sensor nodes cannot be dynamically adjusted to match instantaneous channel conditions. In this context, the point-to-point communication of Gaussian sources over fading channels has been recently investigated in [79] and [80] with the ultimate purpose of minimizing the expected distortion at the FC. These works stand on the seminal work of [81], in which the source is encoded in multiple layers (each one representing a different channel state) by exploiting the so-called successive refinement property [82]. Accordingly, the receiver adaptively decodes the information according to the realization of the channel state.

In addition, in [75] and [76], it is considered that each sensor-to-FC communication occurs in

a reserved orthogonal channel, i.e. sensors operate in a TDMA/FDMA fashion coordinated by the FC. Nonetheless, in realistic WSN scenarios, where the number of nodes is potentially very high, the centralized coordination that reservation-based multiple access schemes entail is barely desirable. Alternatively, the design and performance analysis of *contention-based* mechanisms (e.g. ALOHA [50], Carrier-Sense Multiple-Access with Collision Avoidance CSMA/CA) in wireless *data* networks has been studied for years. For instance, in [52] the authors designed a channel-aware version of the ALOHA protocol capable of exploiting multi-user diversity in a distributed manner. In the same vein, an opportunistic variation of the CSMA/CA scheme has been recently proposed in [83].

4.1.1 Contribution

In this chapter, we extend and carry out an in-depth analysis of the Q&E and C&E encoding strategies presented in [76]. First, we focus on a scenario with *orthogonal* Gaussian channels where, as commented above, the source-channel coding separation theorem holds. We go one step beyond [76] and further constrain not only power but also *bandwidth* to be constant irrespectively of the number of sensors in the network. With these assumptions, we ask ourselves whether increasing the population of sensors is always worth doing (in terms of estimation accuracy) or if, alternatively, an optimal operating point exists. To answer this question, we derive an approximate closed-form expression of the optimal number of sensors in the network. Next, we move onto the case of orthogonal *Rayleigh-fading* channels. For benchmarking purposes, we initially assume that sensors are capable of acquiring instantaneous transmit CSI (CSIT). Being C&E a successive encoding/decoding scheme, we analytically determine the optimal encoding order (over sensors) which minimizes the resulting distortion. We complement the analysis by deriving the asymptotic distortion attainable by the Q&E and C&E strategies for both Gaussian and Rayleigh-fading channels.

Next, we realistically constrain the sensor nodes in the *network* (and, by doing so, we go beyond the *point-to-point* analysis of [79] and [80]) to operate without *instantaneous* transmit CSI. Accordingly, the observations must be encoded at a *constant* and *common* (i.e. identical for all sensors) rate that we determine on the basis of *statistical* CSI only. Such encoding rate has an impact on *i*) the number of quantization bits allocated to the encoders; and *ii*) the actual number of observations reliably received at the FC due to the outage probability experienced in the sensor-to-FC fading channels. This is in stark contrast with [78] and [23] where the quantization process and the transmission problems are regarded as decoupled. In other words, we analytically find the *optimal* trade-off in terms of quantization bits (more finely-quantized observations are beneficial in terms of the overall distortion) vs. the number of observations actually received at the FC (the higher the number observations, the better the smoothing of the observation noise). We solve the problem for two cases of interest, namely, sensors with high

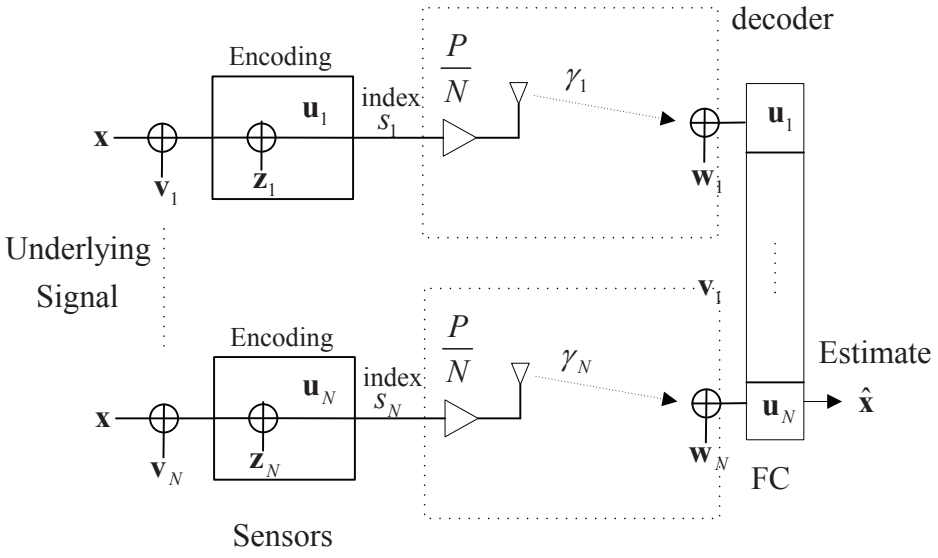


Figure 4.1: System model.

and low observation noise. Finally, we determine the asymptotic rate at which the distortion decreases in the high-SNR regime of a large sensor network.

Finally, we go one step beyond [75] [76] [78] and explicitly take into account realistic multiple-access schemes in a context of *hierarchical* WSNs. More precisely, we analyze the impact of a *contention-based* mechanism (ALOHA), and the packet collisions that it entails. Unlike [52] and [83], our goal is to minimize the distortion in the estimates rather than maximize network throughput. Since the number of received observations now becomes a random variable (due to packet collisions) and for the sake of scalability, we depart from successive/joint coding schemes and, instead, adopt in the sensor nodes and CHs the Q&E encoding strategy of [76].

The contents of this chapter have been partly published in references [84–89].

The chapter is organized as follows. First, in Section 4.2, we present the signal model. In Sections 4.3 and 4.4 we present the distortion analysis for the Q&E and the C&E encoding strategies. Next, Section 4.5 is devoted to the analysis of the aforementioned strategies for Gaussian channels. Subsequently, in Section 4.6 we address the case of Rayleigh-fading channels with CSIT at the sensor nodes. Section 4.7 addresses the case where sensors do not have instantaneous CSIT. Next, the comparison of contention-based and Reservation-based multiple access schemes is addressed in Section 4.8. Finally, we close the chapter by summarizing the main findings in Section 4.9.

4.2 Signal model

Consider a WSN composed of one Fusion Center (FC) and N energy-constrained sensor nodes. The sensors observe a *common* source of interest which can be modeled as a length-n (with n sufficiently large) vector $\mathbf{x} = [x^{(1)}, \ldots, x^{(n)}]^T$ of independent and identically-distributed (i.i.d) complex circular Gaussian random variables, namely, $p\left(x^{(1)}, \ldots, x^{(n)}\right) = \prod_{i=1}^{n} p\left(x^{(i)}\right)$, of variance σ_x^2. Our goal is to estimate $\mathbf{x}$ from the sensor observations with the highest possible accuracy. According to Fig. 5.12, the observation at sensor k can be expressed as

$$\mathbf{y}_k = \mathbf{x} + \mathbf{v}_k \tag{4.1}$$

where $\mathbf{v}_k$ denotes memoryless AWGN noise (measurement noise) of variance σ_v^2 and i.i.d. over sensors. In order to convey the measurements to the Fusion Center, each sensor encodes a block of n consecutive observations $\mathbf{y}_k = [y_k^{(1)}, \ldots, y_k^{(n)}]^T$, where n denotes the temporal variable, into a (length-n) codeword $\mathbf{u}_k(s) \in \mathcal{C}$. Next, the corresponding indexes[1] $s_k \in \{1, \ldots, 2^{nR_k}\}; k = 1 \ldots N$ are sent to the FC over the set of N *orthogonal* channels. At this point, we introduce W_s as the *temporal* sampling rate of the source and, hence, for a reliable transmission to occur, the encoding rate R_k must satisfy:

$$W_s R_k \leq w_k \log_2\left(1 + \frac{p_k \gamma_k}{w_k \mathsf{N_o}}\right) \quad [\mathrm{b/s}] \tag{4.2}$$

where w_k and p_k stand for the transmit power and bandwidth allocated to the k-th sensor/channel, and $\mathsf{N_o}$ denotes the noise spectral density. For the ease of notation and without loss of generality, in the sequel we consider $W_s = 1$. The channel (squared) gains, $\gamma_k; k = 1, \ldots, N$, are subsequently modeled as Gaussian channels (namely, $\gamma_k = 1\, \forall k$), or as independent exponentially-distributed unit-mean random variables (i.e. Rayleigh-fading channels). Besides, we further impose a total bandwidth constraint, W, and a sum-power constraint, P. For simplicity, we uniformly allocate power *and* bandwidth to the set of N sensors, this yielding $p_k = P/N$ and $w_k = W/N$. Hence, the transmission rate at the output of the k-th encoder reads

$$R_k \leq \frac{W}{N} \log_2\left(1 + \mathsf{SNR} \cdot \gamma_k\right) \quad [\mathrm{b/s}] \tag{4.3}$$

with $\mathsf{SNR} \triangleq \frac{P}{W \mathsf{N_o}}$. Finally, the FC decodes the received signals and produces an estimate $\hat{\mathbf{x}}$ of the source. Due to the measurement noise and the resource constraints, some distortion results which, in the sequel, will be characterized by the following metric:

$$D = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[|\hat{x}_i - x_i|^2\right]. \tag{4.4}$$

[1] As it will become apparent later, the codebook $\mathcal{C}$ consists of, at most, 2^{nR_k} codewords.

4.3 Quantize-and-Estimate (Q&E): distortion analysis

In the Q&E scheme, each observation is encoded regardless of any side information that could be made available by the FC. From [35], the following inequality holds for the rate at the output of the k-th encoder (quantizer):

$$R_k \geq \mathrm{I}\left(y_k; u_k\right) \quad [\mathrm{b/sample}] \tag{4.5}$$

with $\mathrm{I}\left(\cdot;\cdot\right)$ standing for the mutual information. The encoding process is modeled through the auxiliary variable $\mathbf{u}_k = \mathbf{y}_k + \mathbf{z}_k$, with $\mathbf{z}_k \sim \mathcal{N}(0, \sigma_{z_k}^2 \mathbf{I})$ and statistically independent of $\mathbf{y}_k$. From this, one concludes that $u_k \longleftrightarrow y_k \longleftrightarrow x$ form a Markov chain and, hence, the mutual information reads:

$$\begin{aligned} \mathrm{I}\left(y_k; u_k\right) &= \mathrm{H}(u_k) - \mathrm{H}(u_k|y_k) \\ &= \log\left(1 + \frac{\sigma_x^2 + \sigma_v^2}{\sigma_{z_k}^2}\right) \quad [\mathrm{b/sample}] \end{aligned} \tag{4.6}$$

From (4.3), (4.5) and (4.61), we have that

$$\frac{W}{N}\log\left(1 + \mathsf{SNR}\gamma_k\right) \geq R_k \geq \log\left(1 + \frac{\sigma_x^2 + \sigma_v^2}{\sigma_{z_k}^2}\right). \tag{4.7}$$

In the sequel, we further impose each sensor to encode its observation at the maximum rate that can be reliably supported by the channel. By taking the above expressions with equality, the variance of the *quantization* noise yields:

$$\sigma_{z_k}^2 = \frac{\sigma_v^2 + \sigma_x^2}{\left(1 + \mathsf{SNR}\gamma_k\right)^{\frac{W}{N}} - 1}. \tag{4.8}$$

With the codewords sent by the set of N sensors, the FC produces an MMSE estimate of $\mathbf{x}$ with total distortion given by [18, Ch.10]:

$$D_{N,\mathsf{Q\&E}} = \sigma_{x|u_1,\ldots,u_N}^2 = \left(\frac{1}{\sigma_x^2} + \sum_{k=1}^{N}\frac{1}{\sigma_v^2 + \sigma_{z_k}^2}\right)^{-1}. \tag{4.9}$$

Finally, by replacing (4.8) into the expression above, the resulting distortion can be re-written as

$$D_{N,\mathsf{Q\&E}} = \left(\frac{1}{\sigma_x^2} + \sum_{k=1}^{N}\frac{\left(1 + \mathsf{SNR}\gamma_k\right)^{\frac{W}{N}} - 1}{\sigma_v^2\left(1 + \mathsf{SNR}\gamma_k\right)^{\frac{W}{N}} + \sigma_x^2}\right)^{-1}. \tag{4.10}$$

4.4 Compress-and-Estimate (C&E): distortion analysis

Here, the encoders do exploit the statistical knowledge of the decoder's data as side information. More precisely, we adopt the successive encoding scheme of [76] by which, in encoding

data, the k-th sensor is aware of the distortion level attained with the previous $k-1$ transmissions. By doing so, the encoding process can be adjusted in such a way that most of the redundant information is removed before transmission. In this sense, we refer to this second approach as *Compress-and-Estimate* (C&E) coding.

Let π be a given ordering of the N sensors in the network. For an arbitrary sensor k, its encoding rate R_k in the presence of side information (resulting from the $u_{\pi(1)},\ldots,u_{\pi(k-1)}$ codewords transmitted by the previous first $k-1$ sensors) verifies [76]:

$$R_{\pi(k)} \geq \mathrm{I}\left(y_{\pi(k)}; u_{\pi(k)}|u_{\pi(1)},\ldots,u_{\pi(k-1)}\right) \quad \text{[b/sample]} \tag{4.11}$$

with $\mathbf{u}_{\pi(k)} = \mathbf{x} + \mathbf{v}_{\pi(k)} + \mathbf{z}_{\pi(k)}$ and $\mathbf{z}_{\pi(k)} \sim \mathcal{N}(0, \sigma^2_{z_{\pi(k)}}\mathbf{I})$). The above expression can be re-written as follows:

$$\begin{aligned} R_{\pi(k)} &\geq \mathrm{I}\left(y_{\pi(k)}; u_{\pi(k)}|u_{\pi(1)},\ldots,u_{\pi(k-1)}\right) \\ &= \mathrm{H}\left(u_{\pi(k)}|u_{\pi(1)},\ldots,u_{\pi(k-1)}\right) - \mathrm{H}\left(u_{\pi(k)}|y_{\pi(k)}\right) \\ &= \log\left(1 + \frac{\sigma^2_{x|u_{\pi(1)},\ldots,u_{\pi(k-1)}} + \sigma^2_v}{\sigma^2_{z_{\pi(k)}}}\right) \quad \text{[b/sample]} \end{aligned} \tag{4.12}$$

where the first equality results from the fact that $u_{\pi(k)} \longleftrightarrow y_{\pi(k)} \longleftrightarrow u_{\pi(1)},\ldots,u_{\pi(k-1)}$ necessarily form a Markov chain [76] and, thus, $u_{\pi(k)}$ is conditionally independent of $u_{\pi(1)},\ldots,u_{\pi(k-1)}$ given $y_{\pi(k)}$. Besides, the conditional variance $\sigma^2_{x|u_{\pi(1)},\ldots,u_{\pi(k-1)}}$ is, by definition, the distortion at the output of the MMSE estimator at the FC upon reception of $k-1$ measurements, namely, $D^{(\pi)}_{k-1,\mathrm{C\&E}}$ (with $D^{(\pi)}_0 = \sigma^2_x$). By imposing again each sensor to encode its observation at the maximum rate that can be reliably supported by the channel, the variance of the encoding noise yields:

$$\sigma^2_{z_{\pi(k)}} = \frac{\sigma^2_v + D^{(\pi)}_{k-1,\mathrm{C\&E}}}{\left(1+\mathsf{SNR}\gamma_k\right)^{\frac{W}{N}} - 1}. \tag{4.13}$$

Since the encoding processes themselves are statistically independent, the distortion after receiving N observations reads:

$$\begin{aligned} D^{(\pi)}_{N,\mathrm{C\&E}} &= \left(\frac{1}{\sigma^2_x} + \sum_{k=1}^{N}\frac{1}{\sigma^2_v + \sigma^2_{z_{\pi(k)}}}\right)^{-1} \\ &= \left(\frac{1}{\sigma^2_x} + \sum_{k=1}^{N}\frac{\left(1+\mathsf{SNR}\gamma_{\pi(k)}\right)^{\frac{W}{N}} - 1}{\sigma^2_v\left(1+\mathsf{SNR}\gamma_{\pi(k)}\right)^{\frac{W}{N}} + D^{(\pi)}_{k-1,\mathrm{C\&E}}}\right)^{-1}. \end{aligned} \tag{4.14}$$

Alternatively, at each step of the decoding structure the distortion can be computed in the following recursive form:

$$D^{(\pi)}_{k,\mathrm{C\&E}} = \left(\frac{1}{D^{(\pi)}_{k-1,\mathrm{C\&E}}} + \frac{\left(1+\mathsf{SNR}\gamma_{\pi(k)}\right)^{\frac{W}{N}} - 1}{\sigma^2_v\left(1+\mathsf{SNR}\gamma_{\pi(k)}\right)^{\frac{W}{N}} + D^{(\pi)}_{k-1,\mathrm{C\&E}}}\right)^{-1} \quad ; \quad k = 1,\ldots,N. \tag{4.15}$$

It is worth noting that the additional computational complexity associated to the C&E scheme is restricted to the successive decoder needed at the FC. Conversely, the complexity of the encoders in the sensor nodes is comparable in both cases.

4.5 Gaussian channels

In Gaussian channels, all sensors experience identical channel conditions ($\gamma_k = 1 \ \forall k$ in the above expressions). Bearing this in mind, we derive some optimal operating points and/or asymptotic distortion limits for the Q&E and C&E schemes.

4.5.1 Quantize-and-Estimate: optimal network size and asymptotic distortion

From (4.10), the distortion attained by the Q&E scheme is given by,

$$\frac{1}{D_{N,\mathsf{Q\&E}}} = \frac{1}{\sigma_x^2} + \frac{N\left(\left(1+\mathsf{SNR}\right)^{\frac{W}{N}}-1\right)}{\sigma_v^2\left(1+\mathsf{SNR}\right)^{\frac{W}{N}}+\sigma_x^2} \tag{4.16}$$

First, we want to show that, for a given bandwidth W, there exists an optimal network size which minimizes the overall distortion. To show that, we relax $N \in \mathbb{R}^+$ and prove in Appendix 4.A.1 that (4.16) is a *quasiconvex* function in N and, therefore, there exists a *single* optimal operating point N^*. The intuition behind this fact is the following: for an increasing number of sensors, the FC is capable of better smoothing the observation noise and, thus, the distortion decreases (i.e. a more accurate estimate results). However, the available bandwidth has to be shared among a higher number of sensors and, hence, the measurements undergo a rougher quantization before transmission. As soon as this second effect dominates, the distortion increases again.

Unfortunately, a closed-form expression of the optimal number of sensors, N^*, cannot be obtained for the general case. Instead, we consider the following *approximation* for the second summation term in (4.16)

$$\frac{N\left(\left(1+\mathsf{SNR}\right)^{\frac{W}{N}}-1\right)}{\sigma_v^2\left(1+\mathsf{SNR}\right)^{\frac{W}{N}}+\sigma_x^2} \approx \frac{N\left(1+\mathsf{SNR}\right)^{\frac{W}{N}}}{\sigma_v^2\left(1+\mathsf{SNR}\right)^{\frac{W}{N}}+\sigma_x^2} \tag{4.17}$$

which is valid for $\frac{W}{N} \gg 1$. On the one hand, by setting the first derivative of (4.17) to zero, the following two possible solutions yield:

$$N^* \approx \left\{ \frac{W\ln\left(1+\mathsf{SNR}\right)}{1-\mathsf{W}_{-1}\left(-\frac{\sigma_v^2 e}{\sigma_x^2}\right)}, \frac{W\ln\left(1+\mathsf{SNR}\right)}{1-\mathsf{W}_{0}\left(-\frac{\sigma_v^2 e}{\sigma_x^2}\right)} \right\} \tag{4.18}$$

with $\mathsf{W}_0(\cdot)$ and $\mathsf{W}_{-1}(\cdot)$ standing for the two real branches of the Lambert function [71], where $\mathrm{dom}\{\mathsf{W}_{-1}(x)\} = (-1/e, 0)$ and $\mathrm{dom}\{\mathsf{W}_0(x)\} = (-1/e, \infty)$. On the other, the approximation (4.17) can be shown to be concave for

$$N \leq N_{\mathsf{th}} = \frac{W \ln(1+\mathsf{SNR})}{\ln\left(\frac{\sigma_x^2}{\sigma_v^2}\right)} \tag{4.19}$$

and convex otherwise. Now, notice that $\mathsf{W}_{-1}(-x) \leq \ln(-x) \leq \mathsf{W}_0(-x)$ for $x \in (0, 1/e)$ and, hence, from (4.19), the (approximate) solution belonging to the concave domain of (4.17), that is, for $N^* \leq N_{\mathsf{th}}$, can only be given by

$$N^* \approx \frac{W \ln(1+\mathsf{SNR})}{1 - \mathsf{W}_{-1}\left(-\frac{\sigma_v^2 e}{\sigma_x^2}\right)}. \tag{4.20}$$

From this last expression and the aforementioned domain of the $\mathsf{W}_{-1}(x)$ function, the approximate solution of (4.20) is feasible (that is, $N \in \mathbb{R}^+$) if and only if $\sigma_v^2 \leq \sigma_x^2/e^2$. For this range of values, the solution of (4.20) gives a very accurate approximation of the actual value of N^*, as shown in Fig. 4.2. Besides, one also observes that increasing the overall SNR leads to a higher N^*: the higher the SNR the higher the number of observations that can be accommodated (which results into an improved estimation accuracy). Conversely, for each curve, if the correlation $\rho = \mathrm{Cov}(y_k, y_l)/\sigma_{y_k}\sigma_{y_l} = \sigma_x^2/(\sigma_x^2 + \sigma_v^2)$ between observations increases, i.e. σ_v^2 decreases, then the optimal number of sensors decreases. In other words, one should refrain from conveying many observations to the FC because of their correlation and because the bandwidth and power *per observation* would be smaller.

Next, we compute the asymptotic distortion when the number of sensors grows without bound, that is,

$$D_{\infty,\mathsf{Q\&E}} = \left(\frac{1}{\sigma_x^2} + \frac{W \ln(1+\mathsf{SNR})}{\sigma_x^2 + \sigma_v^2}.\right)^{-1}. \tag{4.21}$$

Interestingly, despite that power and bandwidth are spread thinner and thinner, the asymptotic distortion converges to a finite value $D_{\infty,\mathsf{Q\&E}} < \sigma_x^2$. In other words, performance is never worse than that of a wild guess on the parameter x.

4.5.2 Compress-and-Estimate: discussion

The distortion associated to the C&E strategy, i.e. $D_{N,\mathrm{C\&E}}$, is known to be a monotonically-decreasing function in N, except for $\sigma_v^2 \to 0$ (i.e. $\rho = 1$, fully correlated observations). In this case, the particularization of (4.15) for $\sigma_v^2 = 0$ yields

$$D_{N,\mathrm{C\&E}} = \sigma_x^2 (1+\mathsf{SNR})^{-W} \tag{4.22}$$

which, clearly, is not a function of N. In this particular case, the distortion attained by C&E equals that of Q&E since, for $\sigma_v^2 \to 0$, the optimal network size for the Q&E strategy can

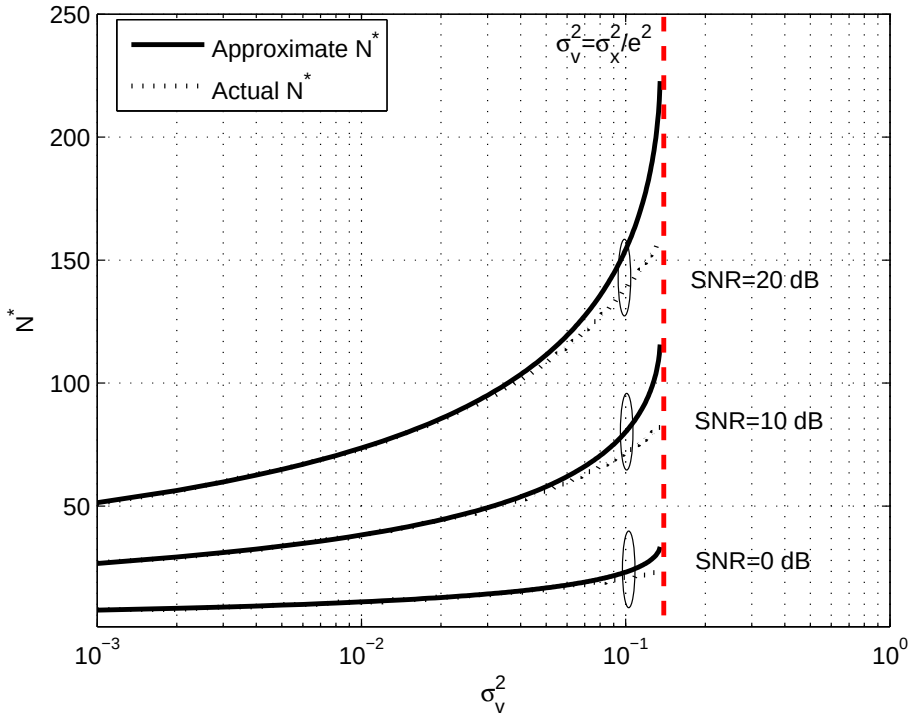


Figure 4.2: Optimal number of sensors vs. observation noise variance σ_v^2 ($W = 100$, $\sigma_x^2 = 1$).

be shown to be $N^* = 1$. Likewise, for large values of σ_v^2, i.e. uncorrelated observations, the distortion for the C&E strategy is identical to that of Q&E as given by (4.16) particularized for $N^* \to \infty$. For an arbitrary value of σ_v^2 and when the number of sensors increases without bound, the asymptotic distortion $D_{\infty,\mathrm{C\&E}}$ of (4.14) for $\gamma_k = 1\ \forall k$ is given by the (numerical) solution to the following equation [76]:

$$W \ln\left(1 + \mathsf{SNR}\right) = \frac{\sigma_v^2}{\sigma_x^2}\left(\frac{\sigma_x^2}{D_{\infty,\mathrm{C\&E}}} - 1\right) + \log\frac{\sigma_x^2}{D_{\infty,\mathrm{C\&E}}} \tag{4.23}$$

4.5.3 Simulations and numerical results

In Fig. 4.3, we depict the distortion associated with the Q&E scheme as a function of the network size (Gaussian channels). When the observation noise is low ($\sigma_v^2 = 0.001$), the distortion function is sharp and, hence, optimizing on the number of sensors pays off. Conversely, by increasing σ_v^2 the curves become flatter and, consequently, there exists some flexibility in the number of sensors (performance degrades gracefully in the vicinity of N^*). For scenarios with very noisy observations ($\sigma_v^2 = 0.5$), distortion turns out to be a monotonically decreasing convex function in N: increasing the number of sensors is worth doing since it allows for a better smoothing of the observation noise. Besides, the larger the overall SNR the higher the optimal number of sensors since, with additional transmit power, a higher number of sensor observations can be accommodated.

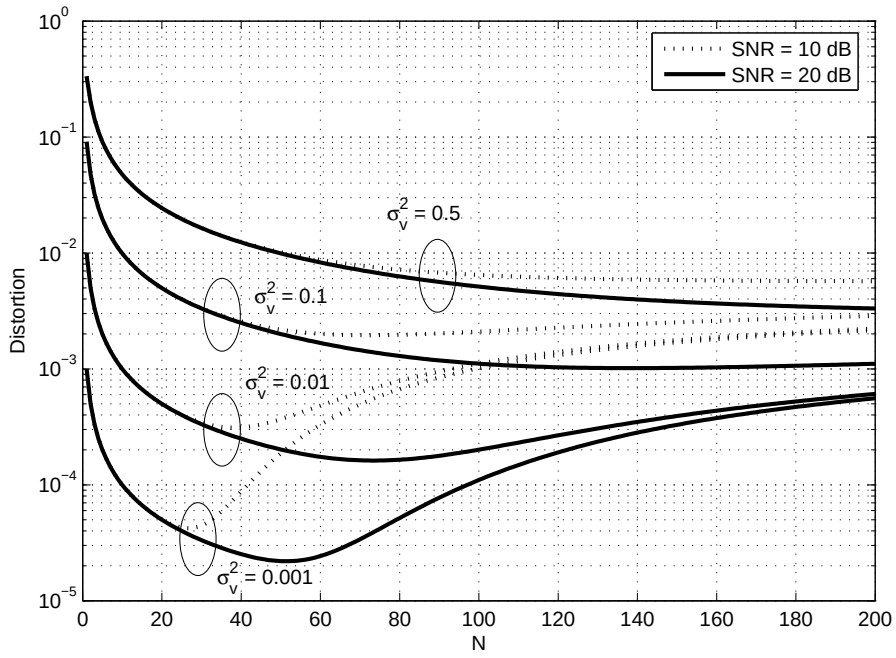


Figure 4.3: Distortion for the Q&E strategy (Gaussian channels) vs. network size N ($W = 100$, $\sigma_x^2 = 1$).

In Figure 4.4, we depict the distortion attained by the Q&E encoding strategy evaluated at the true optimal point N^* (namely, $D_{\text{Q\&E},N^*}$) and the distortion attained by a large sensor network (that is, $D_{\text{Q\&E},\infty}$). For scenarios with low observation noise (small σ_v^2), carefully designing the network size pays off. On the contrary, as σ_v^2 increases, one can simply deploy a high number of sensors (in order to average out the observation noise) without incurring in a substantial performance loss with respect to the asymptotic case.

4.6 Rayleigh-fading channels with transmit CSI

For Rayleigh-fading channels, each sensor in the network experiences different channel conditions. As a result, the distortion in the estimates at the FC depends on the specific set of γ_k values. This has diverse implications for the two strategies considered here. In Q&E encoding, on the one hand, *local* channel state information is needed at the sensor nodes in order to *locally* adjust the encoding rate. On the other hand, *global* CSI is needed by the C&E strategy since the encoding rate at the sensor nodes depends not only on their current local channel gains but also on other sensor-to-FC channel gains. This will be further elaborated in the next subsections.

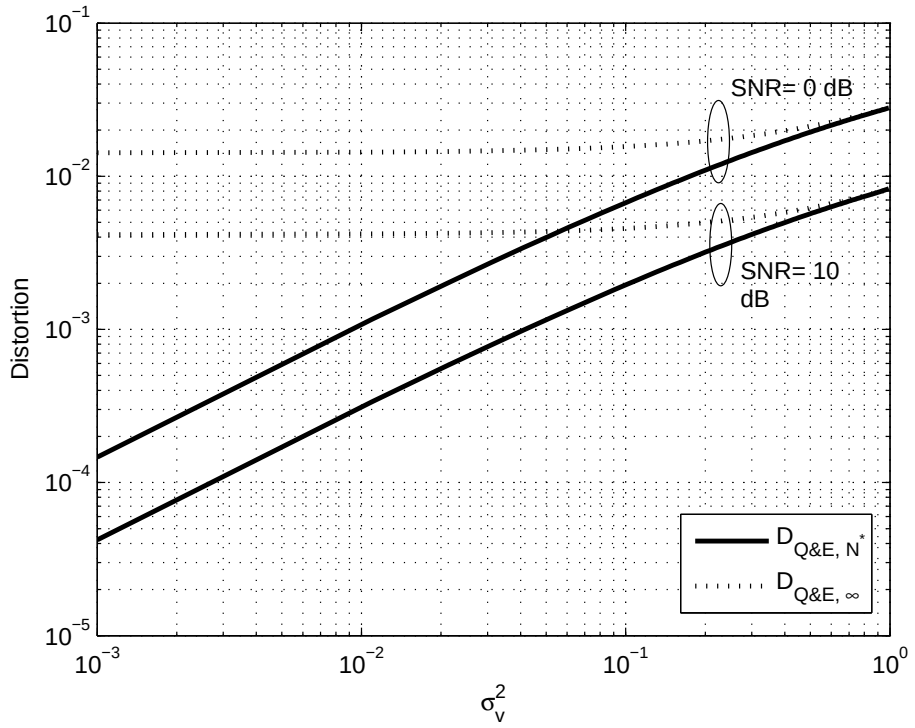


Figure 4.4: Distortion for the Q&E strategy vs. observation noise variance σ_v^2 ($W = 100$, $\sigma_x^2 = 1$).

4.6.1 Quantize-and-Estimate: asymptotic distortion

In this section, we provide some asymptotic results on the attainable distortion when the number of sensor nodes grows without bound. To that end, we focus on the second term in (4.14) and then we define:

$$\sum_{k=1}^{N} \frac{(1+\mathsf{SNR}\gamma_k)^{\frac{W}{N}}-1}{\sigma_v^2\,(1+\mathsf{SNR}\gamma_k)^{\frac{W}{N}}+\sigma_x^2} \triangleq \sum_{k=1}^{N} \frac{g(\gamma_k,N)}{f(\gamma_k,N)}. \tag{4.24}$$

On the one hand, in Appendix 4.A.2 we prove that for $N \to \infty$

$$\sum_{k=1}^{N} \frac{g(\gamma_k,N)}{f(\gamma_k,N)} - \sum_{k=1}^{N} \frac{g(\gamma_k,N)}{\sigma_x^2+\sigma_v^2} \xrightarrow{\mathrm{P}} 0 \tag{4.25}$$

where $\xrightarrow{\mathrm{P}}$ denotes convergence in probability. Consequently, the left-hand side in (4.24) converges (in probability) to the more manageable expression $\frac{1}{\sigma_x^2+\sigma_v^2}\sum_{k=1}^{N} g(\gamma_k,N)$ in the case of an asymptotically-large WSN. On the other hand, from Appendix 4.A.3 we know that

$$\sum_{i=1}^{N} g(\gamma_i,N) \quad \xrightarrow{\mathrm{P}} \quad We^{\frac{1}{\mathsf{SNR}}}\Gamma\left(0,\frac{1}{\mathsf{SNR}}\right) \tag{4.26}$$

with $\Gamma(a,x)=\int_x^\infty e^{-t}t^{a-1}dt$. From all the above, the asymptotic distortion when the number of sensors grows without bound yields

$$D_{\infty,\mathsf{Q\&E}} \xrightarrow{\mathrm{P}} \left(\frac{1}{\sigma_x^2}+\frac{We^{\frac{1}{\mathsf{SNR}}}\Gamma\left(0,\frac{1}{\mathsf{SNR}}\right)}{\sigma_v^2+\sigma_x^2}\right)^{-1}. \tag{4.27}$$

By comparing the expression above these lines with that of Gaussian channels in (4.21) and noticing that $e^{\frac{1}{x}}\Gamma\left(0,\frac{1}{x}\right)<\ln(1+x)\,\forall\, x\geq 0$, one concludes that fading has a *negative* impact in the (asymptotic) accuracy of the estimates.

4.6.2 Compress-and-Estimate: optimal sensor ordering and asymptotic distortion

Being C&E a successive encoding strategy (as it follows from (4.14)) and due to the fact that sensors experience different sensor-to-FC channel gains, one easily concludes that the ordering π has an impact on the attainable distortion. For that reason, we are interested in determining the optimal and worst-case ordering of sensors and, on that basis, analyze the corresponding performance gap.

Optimal ordering

Consider two sensors with channel gains γ_i ; $i=1,2$. Without loss of generality, assume that $\gamma_1\geq\gamma_2$ and, define the two possible orderings (decreasing/increasing) as $\pi_{\mathrm{D}}=(1,2)$ and $\pi_{\mathrm{I}}=(2,1)$. In this section, for the ease of notation and without loss of generality, we let $\mathsf{SNR}=1$ and $\frac{W}{N}=1$ and, further, we define $\beta_i^{(\pi)}\triangleq\frac{1}{D_{i,\mathsf{C\&E}}^{(\pi)}}$. Hence, equation (4.15) can be conveniently re-written as follows,

$$\beta_2^{(\pi)}=\beta_1^{(\pi)}+\frac{\beta_1^{(\pi)}\gamma_{\pi(2)}}{\beta_1^{(\pi)}\left(\gamma_{\pi(2)}\sigma_v^2+\sigma_v^2\right)+1} \tag{4.28}$$

where the term $\beta_0^{(\pi)}=1/\sigma_x^2$, which is needed to be compute $\beta_1^{(\pi)}$, does not depend on the particular ordering ($\pi_{\mathrm{I}}/\pi_{\mathrm{D}}$). The goal is to determine which ordering attains the lowest distortion or, equivalently, the highest value of β_2, that is, whether

$$\begin{aligned}\beta_2^{(\pi_{\mathrm{D}})}-\beta_2^{(\pi_{\mathrm{I}})} &= \beta_1^{(\pi_{\mathrm{D}})}-\beta_1^{(\pi_{\mathrm{I}})}\\ &+ \frac{\beta_1^{(\pi_{\mathrm{D}})}\gamma_2}{\beta_1^{(\pi_{\mathrm{D}})}\left(\gamma_2\sigma_v^2+\sigma_v^2\right)+1}-\frac{\beta_1^{(\pi_{\mathrm{I}})}\gamma_1}{\beta_1^{(\pi_{\mathrm{I}})}\left(\gamma_1\sigma_v^2+\sigma_v^2\right)+1}\gtrless 0.\end{aligned} \tag{4.29}$$

After some tedious manipulations, we have that

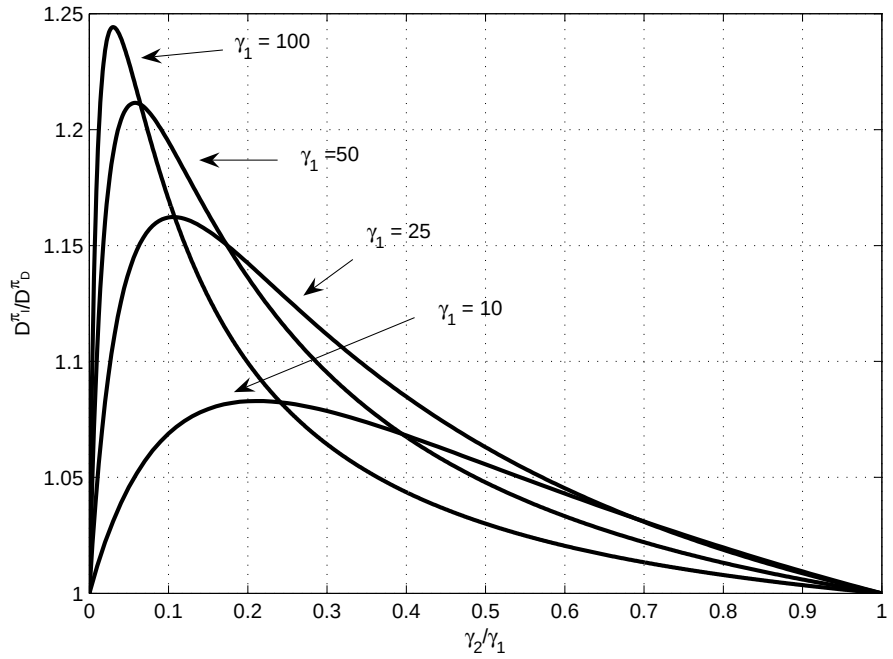


Figure 4.5: Distortion associated to the optimal and worst-case scheduling order. Two-sensor case ($\rho = 0.9$, $P = 1$).

$$\beta_2^{(\pi_{\mathrm{D}})} - \beta_2^{(\pi_{\mathrm{I}})} = \frac{\gamma_1\gamma_2\left(\gamma_2+1\right)\left(\gamma_1+1\right)\left(\gamma_1-\gamma_2\right)\left(\beta_0\sigma_v^2+1\right)\sigma_v^4\beta_0^3}{\prod_{i=1,j\neq i}^{2}\left(1+\beta_0^2\sigma_v^4(\gamma_2+1)(\gamma_1+1)+\beta_0\sigma_v^2(\gamma_i+1)(\gamma_j+2)\right)}. \tag{4.30}$$

Under the initial assumption of $\gamma_1 \geq \gamma_2$, expression (4.30) is always greater than (or equal to) zero, this meaning that the observation of the sensor experiencing the most favorable channel condition should be encoded first (i.e. in *decreasing* order of channel gains, π_{D}) in order to increase the side information at the FC at the next step. The equality in (4.29) holds in two degraded cases only: *i*) for $\gamma_1 = \gamma_2$, i.e. identical channel gains and, *ii*) $\gamma_i = 0$, for, at least, one sensor. In Fig. 4.5, we plot the ratio $D^{(\pi_{\mathrm{I}})}/D^{(\pi_{\mathrm{D}})}$ for several values of γ_1 and $\gamma_2 \in [0, \gamma_1]$. As expected, for $\frac{\gamma_2}{\gamma_1} \to \{0, 1\}$ both orderings exhibit a comparable performance (degraded cases). However, there exists a range of values for which π_{D} clearly outperforms π_{I}. Besides, the higher γ_1 (i.e. high-SNR scenario), the more important using the optimal ordering can be (see maxima on the curves).

Now, consider n sensors with channel gains γ_i ; $i = 1, \ldots, n$. Without loss of generality, assume that $\gamma_1 \geq \gamma_2 \geq, \ldots, \gamma_n$. By induction, the optimal ordering for *any* group of n sensors is given by $\pi^* = (1, 2, \ldots, n)$, this meaning that the sensors must be sorted and their data encoded in a *decreasing* order of channel gains.

Next, we focus on the case with $n + 1$ sensors. First, we divide the $(n + 1)!$ possible orderings into $n + 1$ groups, where each group is denoted by $\mathcal{G}_i$. Group $\mathcal{G}_i$ is composed of all the $n!$

orderings having i as its first element, namely,

$$\pi_{\mathcal{G}_i,k} = (i, \pi'_{\mathcal{G}_i,k}); \quad k = 1, \dots n! \quad \text{and} \quad i = 1, \dots n+1 \tag{4.31}$$

where $\pi'_{\mathcal{G}_i,k}$ denotes the k-*th* permutation of the remaining n elements. With this arrangement, we can start discarding some orderings.

Remark 1 *For all the orderings within a group, the distortion after the first iteration of the decoder ($D_1^{(\pi_{\mathcal{G}_i,k})}$) is, by definition, identical. Therefore, it only remains to determine the best ordering (i.e. the one that attains the lowest distortion after $n+1$ transmissions) for the last n elements. By induction, such ordering is given when sensors are sorted in a decreasing order of channel gains.*

As a result of Remark 1, for each group we only need to retain the best ordering, which is given by:

$$\pi^*_{\mathcal{G}_i} = \begin{cases} (1, 2, \dots, n, n+1) & ; i = 1 \\ (i, 2, \dots, i-1, i+1, \dots, n+1) & ; i = 2 \dots n \\ (n+1, 1, 2, \dots, n) & ; i = n+1 \end{cases} .$$

Remark 2 *According to* (4.15), $D_{n+1}^{(\pi)}$ *(namely, the distortion after the ($n+1$)-*th *iteration) is a monotonically increasing function in $D_n^{(\pi)}$.*

It is worth noting that the first n orderings, ($\pi^*_{\mathcal{G}_i}; i = 1 \dots n$), have their last element (i.e. $n+1$) in common. Hence, from *Remark 2*, the ordering that minimizes the distortion after the n-th iteration ($D_n^{(\pi^*_{\mathcal{G}_i})}$), does minimize the distortion after the ($n+1$)-*th* iteration ($D_{n+1}^{(\pi^*_{\mathcal{G}_i})}$), as well. From all the above, one concludes that the best ordering out of the first n groups turns out to be:

$$\pi^*_{\mathcal{G}_1} = (1, 2, \dots, n+1) . \tag{4.32}$$

Finally, it only remains to determine which ordering out of $\pi^*_{\mathcal{G}_1}$ and $\pi^*_{\mathcal{G}_{n+1}}$ attains the lowest distortion. However, one of the previously-discarded orderings within the first group, namely, $\pi_{\mathcal{G}_1,k_0} = (1, 2, \dots, n+1, n)$ is known to have a smaller distortion than $\pi^*_{\mathcal{G}_{n+1}}$. This follows again from *Remark 2*, from the fact that, *i*) by induction, $D_n^{(\pi^*_{\mathcal{G}_{n+1}})} \geq D_n^{(\pi_{\mathcal{G}_i,k_0})}$; and *ii*) again, $\pi^*_{\mathcal{G}_i}$ and $\pi_{\mathcal{G}_i,k}$ have their last element (n) in common. As a result, $D_{n+1}^{(\pi^*_{\mathcal{G}_{n+1}})} \geq D_{n+1}^{(\pi_{\mathcal{G}_i,k})} \geq D_{n+1}^{(\pi^*_{\mathcal{G}_1})}$ and, hence, $\pi^*_{\mathcal{G}_{n+1}}$ can be discarded. From all the above, the optimal ordering with $n+1$ sensors turns out to be:

$$\pi^* = \pi^*_{\mathcal{G}_1} = (1, 2, \dots, n+1) . \tag{4.33}$$

Interestingly, this ordering minimizes the distortion at *each* iteration of the decoder in the receiver. Unfortunately, this does not hold true for the general case, as it occurs when sensors experience identical channel gains and *different* observation noise variances (see [85]).

Worst-case ordering

In order to gain some insight on the performance of a particular ordering (e.g. a pre-defined ordering), we attempt to identify the sorting that leads to the highest distortion, that is, the worst-case ordering π^w. This, along with the optimal ordering, can be used as upper and lower performance bounds, respectively.

By induction and on the basis of the results of previous section, the worst-case ordering for any group of n sensors is given by $\pi^w = (n, n-1, \ldots, 1)$. For $n+1$ sensors, the $(n+1)!$ possible orderings can be divided into $n+1$ groups, as done before. By resorting to Remark 1, we can discard again all elements but one from each group, which leads to the following subset of worst-case orderings:

$$\pi^w_{\mathcal{G}_i} = \begin{cases} (1, n+1, n, \ldots, 2) & ; i = 1 \\ (i, n+1, n, \ldots, i+1, i-1, \ldots, 1) & ; i = 2 \ldots n \\ (n+1, n, \ldots, 1) & ; i = n+1. \end{cases}$$

Now, from *Remark 2*, the worst-case ordering out of the $i = 2 \ldots n+1$ groups turns out to be

$$\pi^w_{\mathcal{G}_{n+1}} = (n+1, n, \ldots, 1)\,. \tag{4.34}$$

Finally, it only remains to determine which ordering out of $\pi^w_{\mathcal{G}_{n+1}}$ and $\pi^w_{\mathcal{G}_1}$ attains the highest distortion level. However, one of the previously-discarded orderings within the $(n+1)$-*th* group $\pi_{\mathcal{G}_{n+1},k_0} = (n+1, n, \ldots, 3, 1, 2)$ is known to have a higher distortion than $\pi^w_{\mathcal{G}_1}$. Therefore, $\pi^w_{\mathcal{G}_1}$ can be discarded and, consequently, the worst-case ordering with $n+1$ sensors yields

$$\pi^w = \pi^w_{\mathcal{G}_{n+1}} = (n+1, n, \ldots, 1)\,. \tag{4.35}$$

In conclusion, the performance of an arbitrary ordering π can be upper- and lower-bounded by those of the optimal and worst-case orderings, that is,

$$D^{(\pi^*)}_{n+1} \leq D^{(\pi)}_{n+1} \leq D^{(\pi^w)}_{n+1}, \tag{4.36}$$

as defined in expressions (4.33) and (4.35), respectively.

Asymptotic distortion expressions

To close this section, we derive a lower bound of the asymptotic distortion for $N \to \infty$. By setting $D^{(\pi)}_{k-1,\mathsf{C\&E}} = 0\ \forall k$ in (4.14), the following lower bound results

$$\begin{aligned} D^{\pi}_{\infty,\mathrm{C\&E}} \quad &\geq \quad \left(\frac{1}{\sigma_x^2} + \sum_{k=1}^{N} \frac{\left(1 + \mathsf{SNR}\gamma_{\pi(k)}\right)^{\frac{W}{N}} - 1}{\sigma_v^2 \left(1 + \mathsf{SNR}\gamma_{\pi(k)}\right)^{\frac{W}{N}}} \right)^{-1} \\ &\xrightarrow{\mathrm{P}} \quad \left(\frac{1}{\sigma_x^2} + \frac{We^{\frac{1}{\mathsf{SNR}}}\Gamma\left(0, \frac{1}{\mathsf{SNR}}\right)}{\sigma_v^2} \right)^{-1} \end{aligned} \tag{4.37}$$

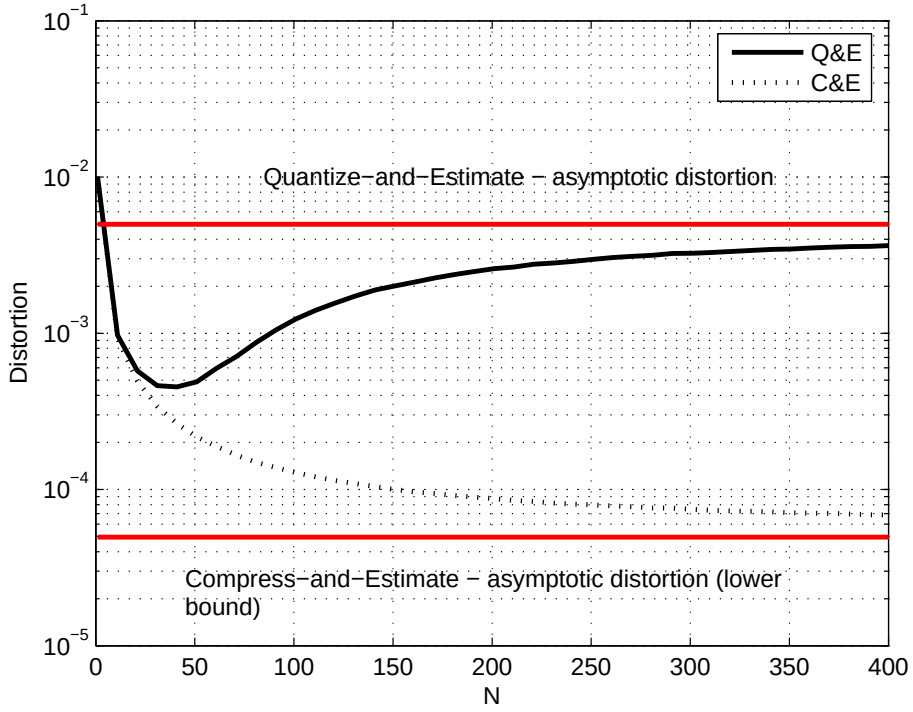


Figure 4.6: Distortion for Rayleigh-fading channels vs. network size ($\sigma_x^2 = 1, W = 100, \mathsf{SNR} = 10\text{dB}, \sigma_v^2 = 0.01$).

which, interestingly, does not depend on the specific ordering π. From (4.10) and (4.14), it follows that $D^{(\pi^*)}_{N,\text{C\&E}} \leq D_{N,\text{Q\&E}}$ for *all* N. Besides, in the high-SNR regime both *i*) the lower bound of $D^{\pi}_{\infty,\text{C\&E}}$ given by (4.37); and *ii*) $D_{\infty,\text{Q\&E}}$ behave as $e^{\frac{1}{\mathsf{SNR}}}\Gamma\left(0, \frac{1}{\mathsf{SNR}}\right) \sim \frac{1}{\ln \mathsf{SNR}}$. From this, one concludes that, necessarily, the actual distortion $D^{(\pi^*)}_{\infty,\text{C\&E}}$ decreases as $\frac{1}{\ln \mathsf{SNR}}$, as well.

4.6.3 Simulations and numerical results

In Figure 4.6, we illustrate the behavior of the Q&E and C&E encoding schemes with CSIT for a varying number of sensors (Rayleigh-fading channels). For a small number of sensors ($N < 40$), the performance of both encoding schemes is virtually identical. Although Q&E cannot avoid sending redundant information, the bandwidth per sensor in this region is still high, the observations can be accurately encoded and the noise-averaging effect (which is identical for both strategies) dominates. As N grows, the messages undergo a rougher encoding (quantization) process. This can be partly compensated by the C&E scheme which, by successively encoding data, is able to remove the correlation in the observations. Consequently, and unlike in Q&E encoding, distortion continues to decrease. Besides, one also observes that the distortion attained by the Q&E and C&E schemes converge to the asymptotic limits derived in the text and, interestingly, the asymptotic bound (4.37) is shown to be tight.

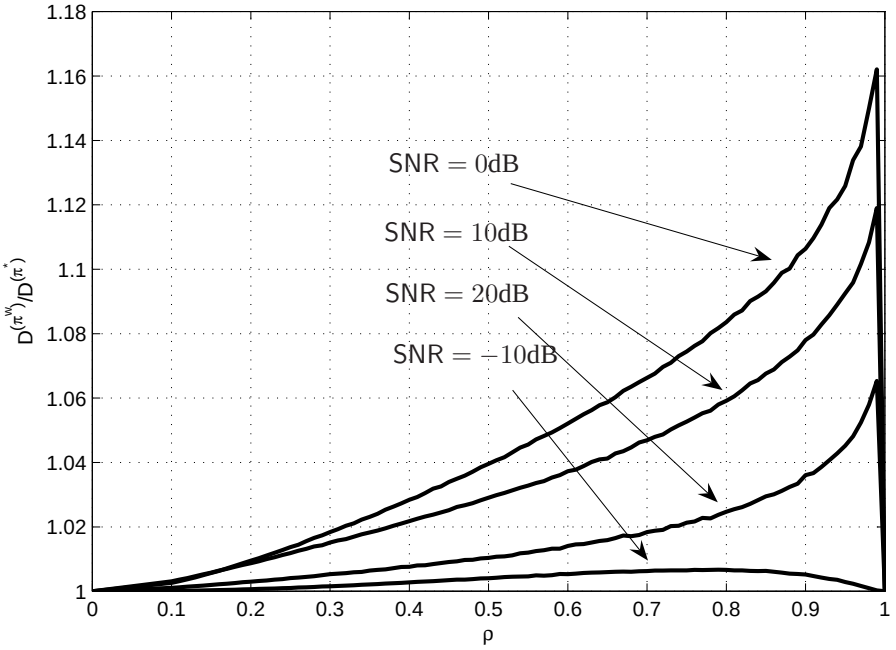


Figure 4.7: Ratio of distortions associated to the optimal and worst-case scheduling order. Different channel gains ($N = 25$ sensors, and $\frac{W}{N} = 1$).

By plotting the ratio between the optimal and worst-case distortions ($D^{(\pi^w)}/D^{(\pi^*)}$), we evaluate the impact of the encoding order on the C&E scheme (Fig. 4.7). For a low correlation coefficient ρ (i.e. noisy observations), the impact of the ordering vanishes: the quality of the observations is very poor and the impact of sorting the observations differently is marginal. For $\rho \to 1$ (namely, $\sigma_v^2 \to 0$), the observations are identical and, likewise, it does not really matter in which order they are encoded. For intermediate values of ρ, the ordering does play a role, although the higher the SNR (i.e. the available transmit power), the lesser the impact of the ordering. Still, for very low transmit power ($\mathsf{SNR} = -10$ dB) the achievable rates over the sensor-to-FC channels are small, the (roughly encoded) observations do not provide significant gains when used as side information at the FC and, again, all orderings attain similar distortion levels.

4.7 Rayleigh-fading channels without transmit CSI

In the absence of instantaneous transmit CSI, neither can the encoding rate be dynamically adjusted to channel conditions, nor is suitable the C&E scheme[2]. Hence, we focus on the Q&E scheme and propose a modification by which each sensor observation is encoded at a *common*

[2]Being C&E a successive encoding scheme, as soon as one sensor is in outage and the observation cannot be successfully decoded at the FC (see further details ahead in the text), the distortion in the estimate increases dramatically.

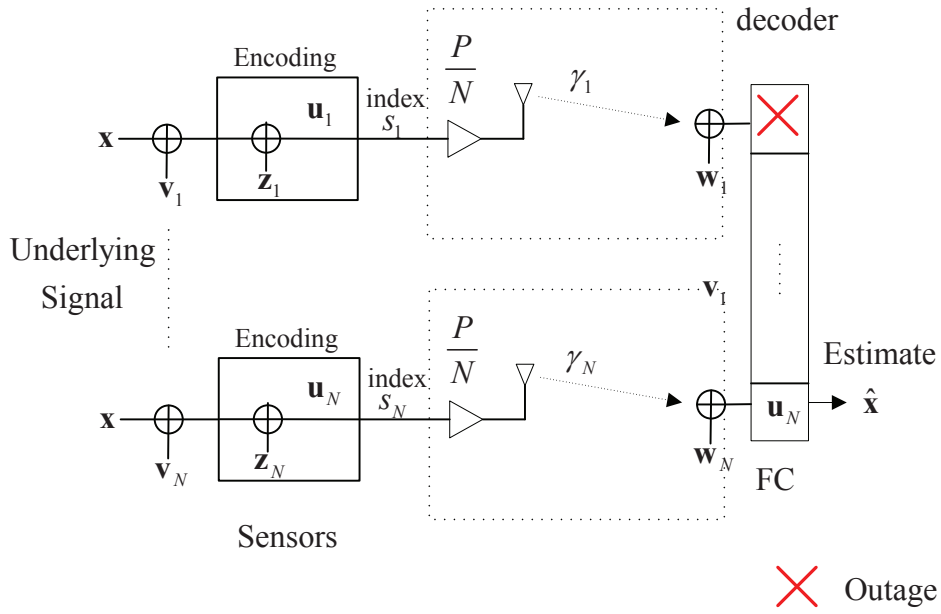


Figure 4.8: System model for no transmit CSI at the sensor nodes.

and *constant* rate given by

$$R = R_k = \log_2\left(1 + \frac{\sigma_x^2 + \sigma_v^2}{\sigma_z^2}\right) \quad ;\forall k. \tag{4.38}$$

Clearly, the goal now is to find the optimal value of R, as a function of *statistical* CSI only, which minimizes the distortion attained at the FC. To that end, we first compute the outage probability, that is, the probability that the encoding rate R exceeds the instantaneous channel rate, namely

$$\begin{aligned} p_{\mathsf{out}} &= \Pr\left(\log_2\left(1 + \frac{\sigma_x^2 + \sigma_v^2}{\sigma_z^2}\right) > \frac{W}{N}\log_2(1 + \mathsf{SNR}\cdot\gamma_k)\right) \\ &= F_\gamma\left(\frac{\left(1 + \frac{\sigma_x^2+\sigma_v^2}{\sigma_z^2}\right)^{\frac{N}{W}} - 1}{\mathsf{SNR}}\right) \end{aligned} \tag{4.39}$$

where $F_\gamma(x)$ stands for the CDF of γ_k; $k = 1, \ldots, N$. According to (4.39), the variance of the quantization noise can be expressed as a function of p_{out}, as

$$\sigma_z^2(p_{\mathsf{out}}) = \frac{\sigma_x^2 + \sigma_v^2}{\left(1 + \mathsf{SNR}\cdot F_\gamma^{-1}\left(p_{\mathsf{out}}\right)\right)^{\frac{W}{N}} - 1}, \tag{4.40}$$

with $F_\gamma^{-1}(\cdot)$ standing for the inverse of the CDF of γ. From the $N_{\mathsf{d}} \leq N$ codewords reliably decoded at the FC, an MMSE estimate of $\mathbf{x}$ with an overall distortion given by [18, Ch.10]

$$D_{N_{\mathsf{d}},\mathsf{Q\&E}} = \sigma^2_{x|u_1,\ldots,u_{N_{\mathsf{d}}}} = \left(\frac{1}{\sigma_x^2} + \frac{N_{\mathsf{d}}}{\sigma_v^2 + \sigma_z^2(p_{\mathsf{out}})}\right)^{-1} \tag{4.41}$$

can be obtained. Clearly, the expression above depends on N_{d} which is a random variable. Due to the fact that sensors experience independent and identical fading conditions, it follows a binomial distribution with probability of activation $p = 1 - p_{\mathsf{out}}$. For Rayleigh-fading channels, $F_{\gamma}^{-1}(\cdot)$ turns out to be the inverse of the CDF of an exponentially-distributed random variable, namely

$$F_{\gamma}^{-1}(p_{\mathsf{out}}) = \ln\left(\frac{1}{1-p_{\mathsf{out}}}\right). \tag{4.42}$$

By replacing (4.40) and (4.42) into (4.41) the expected distortion yields

$$D_{\mathsf{Q\&E}} = \mathbb{E}_{N_{\mathsf{d}}}\left[\left(\frac{1}{\sigma_x^2} + \frac{N_{\mathsf{d}}\left(\left(1-\mathsf{SNR}\cdot\ln(1-p_{\mathsf{out}})\right)^{\frac{W}{N}}-1\right)}{\sigma_v^2\left(1-\mathsf{SNR}\cdot\ln(1-p_{\mathsf{out}})\right)^{\frac{W}{N}}+\sigma_x^2}\right)^{-1}\right].$$

Unfortunately, a closed form expression of the expected distortion is extremely difficult to obtain and, hence, it cannot be used as a score function to optimize N. However, the argument in the expectation function is convex in N_{d} and, thus, we can resort to the following lower bound:

$$D_{\mathsf{Q\&E}} \geq D_{\mathsf{Q\&E}}^{\mathsf{LB}} = \left(\frac{1}{\sigma_x^2} + \frac{N(1-p_{\mathsf{out}})\left(\left(1-\mathsf{SNR}\cdot\ln(1-p_{\mathsf{out}})\right)^{\frac{W}{N}}-1\right)}{\sigma_v^2\left(1-\mathsf{SNR}\cdot\ln(1-p_{\mathsf{out}})\right)^{\frac{W}{N}}+\sigma_x^2}\right)^{-1} \tag{4.43}$$

which follows from Jensen's inequality. This bound can be shown to be tight as N grows without bound (see Appendix 4.A.4) and, interestingly, it also performs reasonably well for practical values of N as shown in Fig. 4.9. In this plot, we also observe that this bound is a convex function in p_{out} and, hence, there exists a single optimal operating point p_{out}^*. Therefore, the subsequent analysis focuses on the bound given by (4.43) which, due to its tightness, allows us to find an accurate approximation of the optimal encoding rate, R^*, along with the optimal number of sensor nodes of the network, N^*. From (4.38) and (4.40), R is, in turn, a function of N and p_{out} and, therefore, we equivalently minimize with respect to these two variables. This optimization problem can be re-written as follows [69]:

$$\min_{N,p_{\mathsf{out}}} D^{\mathsf{LB}}(N,p_{\mathsf{out}}) = \min_{p_{\mathsf{out}}}\left(\min_{N} D^{\mathsf{LB}}(N,p_{\mathsf{out}})\right) \tag{4.44}$$

$$\triangleq \min_{p_{\mathsf{out}}} \tilde{D}^{\mathsf{LB}}(p_{\mathsf{out}}) \tag{4.45}$$

with $0 \leq p_{\mathsf{out}} \leq 1$. In general, this problem is non-convex and, thus, difficult to solve. However it can be efficiently solved in two cases of interest, namely, scenarios with high and low observations noises. Complementarily, we analyze the rate at which the distortion decreases in the high-SNR regime.

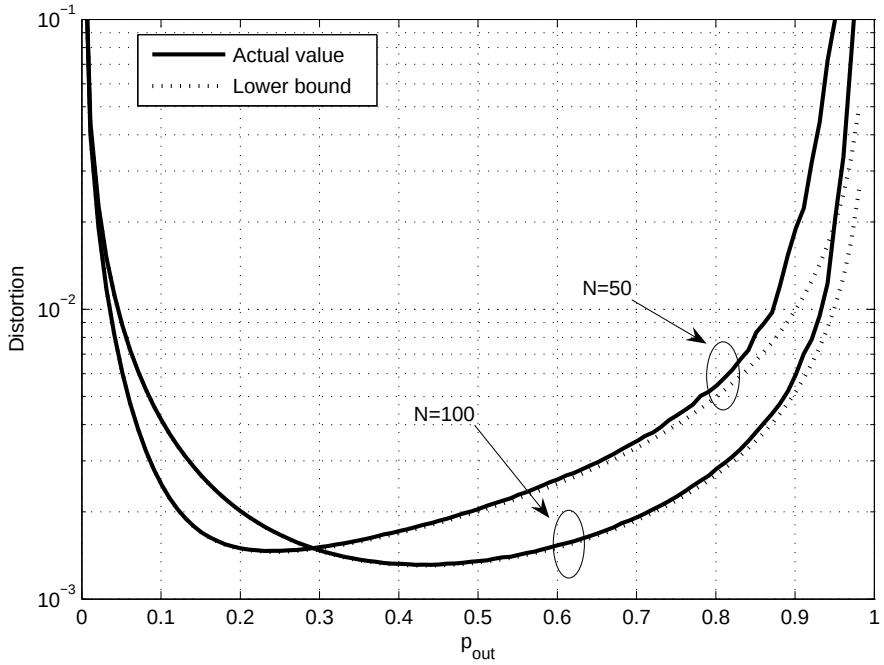


Figure 4.9: Distortion vs. outage probability p_{out} for different network sizes ($W = 200, \sigma_x^2 = 1, \sigma_v^2 = 0.05, \mathsf{SNR} = 10\text{dB}$).

4.7.1 Scenario 1: high observation noise

Here, we consider that $\sigma_v^2 \gg \sigma_x^2$, and, consequently, the *inner* minimization problem in (4.44) can be approximated as follows

$$\min_N D^{\mathsf{LB}}(N, p_{\mathsf{out}}) \equiv \max_N \left(\frac{N(1-p_{\mathsf{out}}) \left((1 - \mathsf{SNR} \cdot \ln(1-p_{\mathsf{out}}))^{\frac{W}{N}} - 1 \right)}{\sigma_v^2 (1 - \mathsf{SNR} \cdot \ln(1-p_{\mathsf{out}}))^{\frac{W}{N}} + \sigma_x^2} \right) \quad (4.46)$$

$$\approx \max_N \left(\frac{N(1-p_{\mathsf{out}}) \left((1 - \mathsf{SNR} \cdot \ln(1-p_{\mathsf{out}}))^{\frac{W}{N}} - 1 \right)}{\sigma_v^2 (1 - \mathsf{SNR} \cdot \ln(1-p_{\mathsf{out}}))^{\frac{W}{N}}} \right) \quad (4.47)$$

The resulting cost function turns out to be concave and monotonically increasing in N and, hence, the optimal number of sensors is $N^* \to \infty$ (that is, many sensors are needed in order to smooth the observation noise). Bearing this in mind, the *outer* minimization problem yields:

$$\min_{p_{\mathsf{out}}} \tilde{D}^{\mathsf{LB}}(p_{\mathsf{out}}) = \min_{p_{\mathsf{out}}} \left(\frac{1}{\sigma_x^2} + \frac{W(1-p_{\mathsf{out}}) \ln(1 - \mathsf{SNR} \cdot \ln(1-p_{\mathsf{out}}))}{\sigma_v^2} \right)^{-1}, \quad (4.48)$$

which can be easily shown to be a convex problem. Setting its first derivative to zero yields

$$p_{\mathsf{out}}^* = 1 - e^{\frac{1}{\mathsf{SNR}} - \frac{1}{\mathsf{W}_0(\mathsf{SNR})}}. \quad (4.49)$$

As far as distortion is concerned, increasing the average $\mathsf{SNR} = \frac{P}{W\mathsf{N}_o}$ in the sensor-to-FC channels has two beneficial effects. On the one hand, p_{out}^* decreases and, hence, a higher

percentage of observations can be successfully decoded at the FC. On the other, observations can be more finely quantized, as it follows from (4.40).

4.7.2 Scenario 2: low observation noise

Here, we consider that $\sigma_v^2 \ll \sigma_x^2$. Following the same steps as in the case of Gaussian channels, (Section 4.5), the *inner* minimization problem in (4.44) yields:

$$N^*(p_{\mathsf{out}}) \approx -\frac{W \ln\left(1 - \mathsf{SNR} \ln\left(1 - \mathsf{p}_{\mathsf{out}}\right)\right)}{\mathsf{W}_{-1}\left(-\frac{\sigma_v^2 e}{\sigma_x^2}\right) - 1}. \tag{4.50}$$

From (4.50) the *outer* optimization problem now reads,

$$\min_{p_{\mathsf{out}}} \tilde{D}^{\mathsf{LB}}\left(p_{\mathsf{out}}\right) \approx \max_{p_{\mathsf{out}}} C\left(1 - p_{\mathsf{out}}\right) \ln\left(1 - \mathsf{SNR} \ln\left(1 - p_{\mathsf{out}}\right)\right) \tag{4.51}$$

with C standing for a positive constant. Surprisingly, the resulting optimization problem is equivalent to that of (4.48) and, hence, the optimal $\{N^*, p_{\mathsf{out}}^*\}$ pair is given by

$$N^* \approx \frac{W \ln\left(\frac{\mathsf{SNR}}{\mathsf{W}_0(\mathsf{SNR})}\right)}{\mathsf{W}_{-1}\left(1 - \frac{\sigma_v^2 e}{\sigma_x^2}\right)}, \tag{4.52}$$

$$p_{\mathsf{out}}^* \approx 1 - e^{\frac{1}{\mathsf{SNR}} - \frac{1}{\mathsf{W}_0(\mathsf{SNR})}}. \tag{4.53}$$

Finally, by subsequently substituting N^* and p_{out}^* into (4.40) and (4.38), we have that

$$R^* \approx \mathsf{W}_{-1}\left(-\frac{\sigma_v^2 e}{\sigma_x^2}\right) - 1. \tag{4.54}$$

Interestingly, as long as $\sigma_v^2 \ll \sigma_x^2$ the encoding rate exclusively depends on the variance of the observation noise. On the contrary, the optimal number of sensor nodes is a function of the available bandwidth and power (through $\mathsf{SNR} = \frac{P}{W\mathsf{N}_o}$), as well.

4.7.3 Asymptotic law in the high-SNR regime

From Appendix 4.A.4, the average distortion $D_{\mathsf{Q\&E}}$ converges in probability to its lower bound $D_{\mathsf{Q\&E}}^{\mathsf{LB}}$ for large N. This holds true for any p_{out} and, in particular, for $p_{\mathsf{out}}^* = 1 - e^{\frac{1}{\mathsf{SNR}} - \frac{1}{\mathsf{W}_0(\mathsf{SNR})}}$ which can be shown to be the optimal outage probability as $N \to \infty$. Hence, from this last result and (4.43), we have

$$\lim_{N\to\infty} \frac{D_{\mathsf{Q\&E}}(p_{\mathsf{out}}^*)}{D_{\mathsf{Q\&E}}^{\mathsf{LB}}(p_{\mathsf{out}}^*)} = \lim_{N\to\infty} \frac{D_{N_d,\mathsf{Q\&E}}(p_{\mathsf{out}}^*)}{\left(\frac{1}{\sigma_x^2} + \frac{We^{\frac{1}{\mathsf{SNR}} - \frac{1}{\mathsf{W}_0(\mathsf{SNR})}} \ln\left(\frac{\mathsf{SNR}}{\mathsf{W}_0(\mathsf{SNR})}\right)}{\sigma_v^2 + \sigma_x^2}\right)^{-1}} \xrightarrow{P} 1. \tag{4.55}$$

Besides, it is straightforward to prove that the denominator in this last equation satisfies

$$\lim_{\mathsf{SNR}\to\infty} \frac{\frac{W\ln(\mathsf{SNR})}{\sigma_v^2+\sigma_x^2}}{\frac{1}{\sigma_x^2}+\frac{We^{\frac{1}{\mathsf{SNR}}-\frac{1}{\mathsf{W_0(SNR)}}}\ln\left(\frac{\mathsf{SNR}}{\mathsf{W_0(SNR)}}\right)}{\sigma_v^2+\sigma_x^2}} = 1. \tag{4.56}$$

Finally, from the numerators in (4.55) and (4.56), we conclude that

$$D_{\mathsf{Q\&E}} \sim \frac{\sigma_x^2+\sigma_v^2}{W\ln(\mathsf{SNR})}. \tag{4.57}$$

In other words, for large sensor networks the (optimal) distortion in the high-SNR regime decreases as $1/\ln(\mathsf{SNR})$. Interestingly, this law is identical to that of (4.21) although the latter was derived in the (more favorable) case of Q&E encoding *with* CSIT at the sensor nodes, where *all* the observations can be successfully decoded by the FC.

4.7.4 Simulations and numerical results

Figure 4.10 illustrates the impact of CSIT on the behavior of the Q&E scheme (Rayleigh-fading channels). As expected, the lack of CSIT translates into an increased distortion (Fig. 4.10a). Notwithstanding, the gap between both strategies is relatively small and *constant* for the whole range of (sensor-to-FC) SNR values. Complementarily, Figure 4.10b depicts the optimal number of sensor nodes with and without CSIT. First, we observe that the approximation for low observation noise scenarios of (4.50) is very accurate for the whole range of SNRs. Second, the optimal number of sensor nodes without CSIT is smaller than with CSIT (constant and variable encoding rates, respectively). The intuition behind this is the following: for a constant encoding rate, some outage probability will unavoidably occur and, in order to partly compensate for that, one should increase the transmit power *per sensor* by allocating the available transmit power to a *reduced* number of sensor nodes.

Next, Fig. 4.11 depicts the asymptotic distortion (for large networks) as a function of the SNR. As discussed in the last paragraphs of sections 4.6 and 4.7, the rate at which the distortion decreases in the high-SNR regime is identical for the Q&E and C&E with CSIT, and Q&E without CSIT encoding strategies ($1/ln(\mathsf{SNR})$). In the high-SNR regime, the impact of CSIT or the encoding scheme restricts to a scale factor in the attainable distortion.

4.8 Contention-based vs. reservation-based multiple-access schemes

In previous sections, we have considered orthogonal channels to the FC and, implicitly, the adoption of a reservation-based multiple-access scheme (e.g. FDMA/TDMA). In realistic sce-

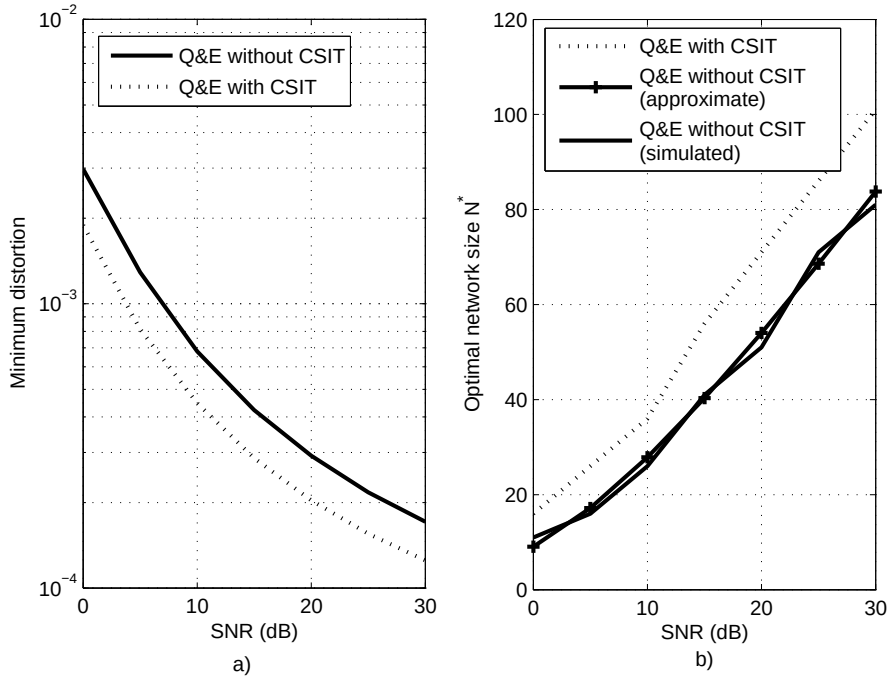


Figure 4.10: a) Minimum distortion for Rayleigh-fading channels vs. SNR and b) optimal network size vs. SNR ($W = 100, \sigma_v^2 = 0.01, \sigma_x^2 = 1$).

narios, where the number of sensor nodes is potentially very high, the centralized coordination that reservation-based multiple access schemes entail is barely desirable. Instead, letting sensors seize the transmission medium via contention-based mechanisms (e.g. CSMA/CA or ALOHA protocols), is far more attractive.

Therefore, to close this chapter we extend the previous analysis to encompass the effect of contention-based multiple-access schemes. We focus our attention on the (more challenging) scenario of *hierarchical* wireless sensor networks.

4.8.1 Signal and network model

Consider a hierarchical WSN where sensor nodes are grouped into N_c clusters[3] with N sensors each (see Fig. 4.12a). Each cluster is coordinated by a Cluster-Head (CH) and, in turn, the set of CHs is coordinated by the Fusion Center (FC). Within each cluster, the (common) source of interest is modeled as a length-n (with n sufficiently large) vector $\mathbf{x}_i = [x_i^{(1)}, \ldots, x_i^{(n)}]^T$ of independent and identically-distributed zero mean Gaussian random variables of variance σ_x^2, with joint pdf $p\left(x_i^{(1)}, \ldots, x_i^{(n)}\right) = \prod_{j=1}^{n} p\left(x_i^{(j)}\right)$. The *noisy* observation at the k-th sensor in

[3]We assume a quasi-static network topology and, hence, we neglect the impact of the signalling associated with the clustering protocol.

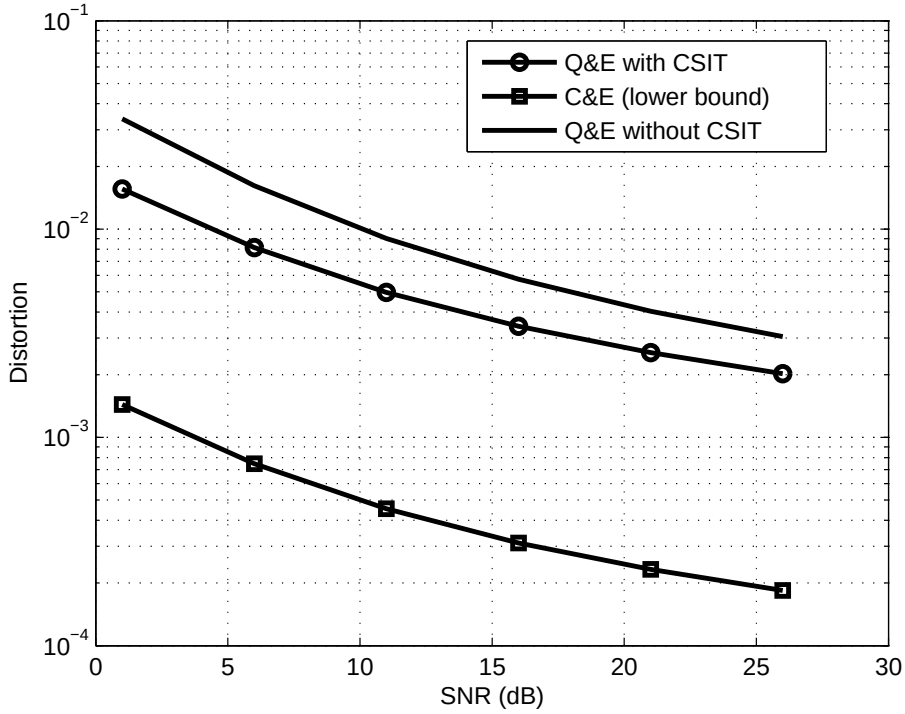


Figure 4.11: Asymptotic distortions vs. SNR ($W = 100$, $\sigma_x^2 = 1$, $\sigma_v^2 = 0.1$).

cluster i reads

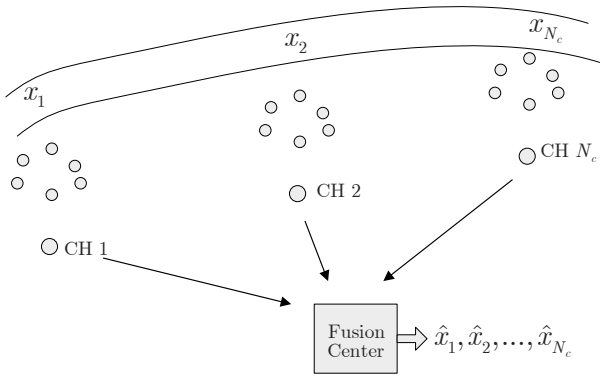
$$\mathbf{y}_{k,i} = \mathbf{x}_i + \mathbf{v}_{k,i} \tag{4.58}$$

where $\mathbf{v}_k$ denotes memoryless AWGN noise (measurement noise) of variance σ_v^2 and i.i.d. over sensors and clusters. Next, each sensor encodes its observation $\mathbf{y}_{k,i} = [y_{k,i}^{(1)}, \ldots, y_{k,i}^{(n)}]^T$ into a length-n codeword $\mathbf{u}_{k,i}(s) \in \mathcal{C}$ and, finally, the corresponding index $s_{k,i}$ is transmitted. As shown in Fig.4.12b, data communication is organized in two phases within each timeslot (i.e. half-duplex operation). In the first phase, of duration $T_1 = \alpha T_s$, sensors convey their encoded data to the corresponding cluster-head. In the second phase, of duration $T_2 = (1-\alpha)T_s$, cluster-heads send the processed data to the Fusion Center (FC). In all cases, we assume (for mathematical tractability) that communications take place over Gaussian channels. From all the above, the effective rates in bits per sample for each phase/layer in the hierarchy are given by

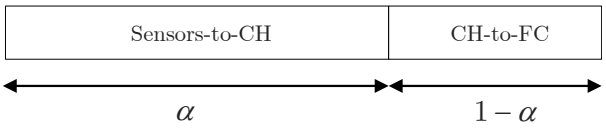
$$R_1 = \quad \alpha R_1' \quad = \alpha W \log_2 (1 + \mathsf{SNR}_1)\,, \tag{4.59}$$

$$R_2 = (1-\alpha)\, R_2' \; = (1-\alpha)\, W \log_2 (1 + \mathsf{SNR}_2) \tag{4.60}$$

where SNR_1 and SNR_2 stand for the average signal-to-noise ratio in each layer. Besides, we assume that the sources of interest which are monitored by the clusters are statistically independent (e.g. a random field being sampled by *distant* clusters of sensors). Consequently, the quality of the k-th element in the vector of FC estimates $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{N_c}$ (see Fig.4.12b), *exclusively* depends on the data being received from the k-th cluster head. For that reason, in the



(a) Network model.



(b) Communication phases.

Figure 4.12: Hierarchical wireless sensor network.

following we will focus our analysis on an arbitrary cluster in the network and, accordingly, the cluster index i will be dropped.

4.8.2 Reservation-based multiple access

Throughout this section, we assume that a reservation-based multiple-access scheme (e.g. TDMA or FDMA) is in place. Hence, the allocation of the orthogonal channels to data packets is either static or, alternatively, it is organized by a centralized scheduler. Consequently, no packet collisions occur. This multiple-access scheme will be used as a benchmark for *contention-based* ones, to be presented later.

With these assumptions, the available rate *per sensor* in Layer 1 turns out to be $\frac{R_1}{N}$ and, hence, the codebook $\mathcal{C}$ consists of, at most, $2^{n\frac{R_1}{N}}$ codewords $\mathbf{u}_k(s)$ with $s \in \{1, 2, \ldots, 2^{n\frac{R_1}{N}}\}$. We adopt the Q&E strategy[4] where the encoding process is modeled through the auxiliary variable $\mathbf{u}_k = \mathbf{y}_k + \mathbf{z}_k$ with $\mathbf{z}_k \sim \mathcal{N}(0, \sigma_z^2\mathbf{I})$ and statistically independent of $\mathbf{y}_k$. Consequently, $u_k \longleftrightarrow y_k \longleftrightarrow x$ form a Markov chain with $\mathbf{u}_k = \mathbf{x} + \mathbf{v}_k + \mathbf{z}_k$. Bearing this in mind, the encoding rate must satisfy:

$$\begin{aligned} \frac{R_1}{N} &\geq \mathrm{I}(y_k; u_k) = \mathrm{H}(u_k) - \mathrm{H}(u_k|y_k) \\ &= \log_2\left(1 + \frac{\sigma_x^2 + \sigma_v^2}{\sigma_z^2}\right) \end{aligned} \tag{4.61}$$

[4]We adopt the Q&E strategy since the encoding process at the sensors is carried out independently. This will be particularly important when, in the sequel, we consider the packet losses at the MAC layer.

and, hence, the variance of the quantization noise for the lowest possible encoding rate reads

$$\sigma_z^2 = \frac{\sigma_x^2 + \sigma_v^2}{2^{\frac{R_1}{N}} - 1}. \tag{4.62}$$

The distortion of the MMSE estimate of $\mathbf{x}$ at the CH is given by [18]:

$$\begin{aligned} D_{\mathsf{CH},N} &= \left(\frac{1}{\sigma_x^2} + \frac{N}{\sigma_z^2 + \sigma_v^2}\right)^{-1} \\ &= \left(\frac{1}{\sigma_x^2} + \frac{N\left(2^{\frac{R_1}{N}} - 1\right)}{\sigma_v^2 2^{\frac{R_1}{N}} + \sigma_x^2}\right)^{-1}. \end{aligned} \tag{4.63}$$

Such CH estimate can now be modeled as

$$\mathbf{y}_{\mathsf{CH}} = \mathbf{x} + \mathbf{v}_{\mathsf{CH}} \tag{4.64}$$

where $\mathbf{v}_{\mathsf{CH}} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2_{\mathbf{v}_{\mathsf{CH}}}\mathbf{I}\right)$ stands for the *equivalent* observation noise at the CH with variance given by[5]

$$\sigma^2_{v_{\mathsf{CH}}} = \left(\frac{1}{D_{\mathsf{CH},N}} - \frac{1}{\sigma_x^2}\right)^{-1} = \frac{\sigma_v^2 2^{\frac{R_1}{N}} + \sigma_x^2}{N\left(2^{\frac{R_1}{N}} - 1\right)}. \tag{4.65}$$

Next, the CH encodes $\mathbf{y}_{\mathsf{CH}}$ into the auxiliary random variable $\mathbf{u}_{\mathsf{CH}}$ with rate $\frac{R_2}{N_c}$ (to recall, N_c orthogonal channels are available in Layer 2). Again, the codeword $\mathbf{u}_{\mathsf{CH}}$ can be modeled as,

$$\mathbf{u}_{\mathsf{CH}} = \mathbf{y}_{\mathsf{CH}} + \mathbf{z}_{\mathsf{CH}} \tag{4.66}$$

where $\mathbf{z}_{\mathsf{CH}} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2_{\mathbf{z}_{\mathsf{CH}}}\mathbf{I}\right)$ denotes the quantization noise at the CH, with variance given by

$$\sigma^2_{z_{\mathsf{CH}}} = \frac{\sigma^2_{v_{\mathsf{CH}}} + \sigma_x^2}{2^{\frac{R_2}{N_c}} - 1}, \tag{4.67}$$

which is computed similarly to (4.62). From (4.65) and (4.67), the distortion of the estimate of x at the FC can be finally expressed as follows:

$$D_{\mathsf{FC},N} = \left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma^2_{v_{\mathsf{CH}}} + \sigma^2_{z_{\mathsf{CH}}}}\right)^{-1} \tag{4.68}$$

$$= \left(\frac{1}{\sigma_x^2} + \frac{2^{\frac{R_2}{N_c}} - 1}{\sigma^2_{v_{\mathsf{CH}}} 2^{\frac{R_2}{N_c}} + \sigma_x^2}\right)^{-1}. \tag{4.69}$$

4.8.3 Contention-based multiple access

In this section, we assume that a *contention-based* multiple-access scheme is adopted in *both* layers of the hierarchical network. For mathematical tractability, we focus our analysis on the

[5] This follows from equation (4.63) for $N = 1$.

results achieved with the ALOHA protocol[6], which relieves sensor nodes/CHs from sensing the medium before transmitting data. Besides, we further assume that no packet collisions result from simultaneous transmissions in *different* clusters (i.e. distant clusters).

A quick overview of the ALOHA protocol

In the classical ALOHA protocol [50] , the distribution of the (initial) transmission time of a packet follows a uniform distribution in $(0, T)$, where T stands for the duration of the corresponding timeslot[7]. For a fully-loaded system, a packet duration of T_p seconds and by neglecting the border effects, the probability that two packets collide can be computed as:

$$p_{\mathsf{col}} = 1 - \left(1 - \frac{2T_p}{T}\right)^{N-1}. \tag{4.70}$$

Now, we re-define $T = NT_p$ where N is the number of terminals (sensors or CHs). From (4.70), the probability of collision yields

$$p_{\mathsf{col}} = 1 - \left(1 - \frac{2}{N}\right)^{N-1}. \tag{4.71}$$

Next, we are interested in characterizing the pmf of the random variable N_s, namely, the number of successful packet transmissions in a given timeslot (with $0 \leq N_s \leq N$). Clearly, we have that

$$\Pr(N_s = n) = \binom{N}{n} p_n q_{N-n} \tag{4.72}$$

where p_n stands for the probability that one particular subset of n sensors (or CHs) successfully transmit their data, and q_{N-n} accounts for the probability that the packets from the remaining $N - n$ sensors (or CHs) collide. Unfortunately, this probability (and pmf) turns out to be extremely complex to characterize. Instead, in Appendix 4.A.5 we show that one can approximate the pmf of N_s for *large* N by that of a *binomial* random variable N_b, that is,

$$\Pr(N_s = n) \approx \Pr(N_b = n) = \binom{N}{n} (1 - p_{\mathsf{col}})^n \, p_{\mathsf{col}}^{N-n}. \tag{4.73}$$

In Figure 4.13, we plot the actual CDF of N_s and its binomial counterpart. Clearly, for $N = 100$, the binomial approximation is quite accurate. For low and moderate values of N (i.e. $N = 20, 50$), the approximation continues to be acceptable.

[6]Clearly, by using more sophisticated MAC protocols like CSMA/CA more realistic results would follow. However, for an initial analysis like this, the ALOHA protocol constitutes a fairly simple and attractive alternative.

[7]Here, T plays the same role as T_1 and T_2 in Section 4.8.1.

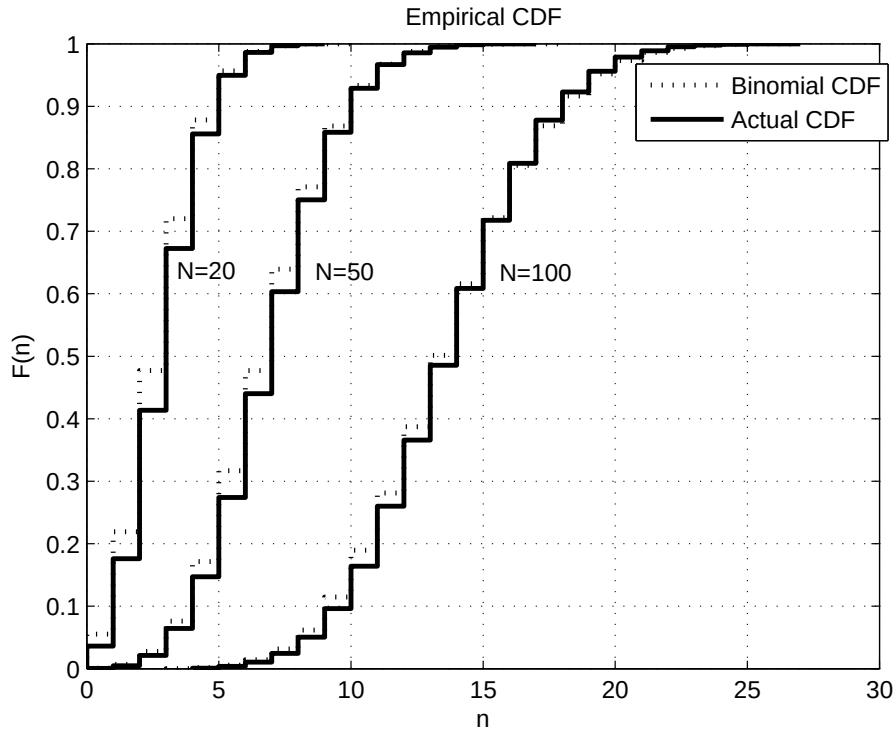


Figure 4.13: Cumulative density functions: actual vs. binomial approximation.

Distortion analysis

On the basis of the Bayes theorem, the average distortion attained at the FC can be expressed as:

$$\begin{aligned} \overline{D}_{\mathsf{FC}} &= p_{\mathsf{col},2}\mathbb{E}\left[D_{\mathsf{FC}}|\text{col.}\right] + (1-p_{\mathsf{col},2})\,\mathbb{E}\left[D_{\mathsf{FC}}|\text{no col.}\right] \\ &= p_{\mathsf{col},2}\sigma_x^2 + (1-p_{\mathsf{col},2})\,\mathbb{E}\left[D_{\mathsf{FC},N_s}\right] \end{aligned} \tag{4.74}$$

with $p_{\mathsf{col},2}$ standing for the probability of collision in Layer 2 (which follows from replacing N with N_c in (4.71)). In the case of a packet collision (first term in the summation), the FC simply outputs the statistical mean of x, this resulting into a conditional distortion of $\mathbb{E}\left[D_{\mathsf{FC}}|\text{col.}\right] = \sigma_x^2$. On the contrary, if the packet is successfully received by the FC (second term in the summation), the distortion depends on the actual number of packets successfully received in Layer 1 (N_s), with expected value given by $\mathbb{E}\left[D_{\mathsf{FC}}|\text{no col.}\right] = \mathbb{E}_{N_s}[D_{\mathsf{FC},N_s}]$, namely,

$$\mathbb{E}_{N_s}[D_{\mathsf{FC},N_s}] = \mathbb{E}_{N_s}\left[\left(\frac{1}{\sigma_x^2} + \frac{2^{\frac{R_2}{N_c}}-1}{\sigma_{v_{\mathsf{CH}}}^2(N_s)2^{\frac{R_2}{N_c}} + \sigma_x^2}\right)^{-1}\right]. \tag{4.75}$$

In the expression above, variable $\sigma_{v_{\mathsf{CH}}}^2(N_s)$ stands for the variance of the equivalent observation noise at the cluster-head observation, that is

$$\sigma_{v_{\mathsf{CH}}}^2(N_s) = \left(\frac{1}{D_{\mathsf{CH},N_s}} - \frac{1}{\sigma_x^2}\right)^{-1} = \frac{\sigma_v^2 2^{\frac{R_1}{N}} + \sigma_x^2}{N_s\left(2^{\frac{R_1}{N}}-1\right)}. \tag{4.76}$$

Unfortunately, a closed-form expression of the expected distortion given by (4.75) is extremely difficult to obtain. Instead and by realizing that the argument in the expectation term of (4.75) is a convex function in N_s, one can resort to Jensen's inequality and derive the following lower bound:

$$\mathbb{E}_{N_s}[D_{\mathsf{FC},N_s}] \geq \left(\frac{1}{\sigma_x^2} + \frac{2^{\frac{R_2}{N_c}} - 1}{\sigma_{v_{\mathsf{CH}}}^2 \left(\overline{N}\right) 2^{\frac{R_2}{N_c}} + \sigma_x^2} \right)^{-1} \tag{4.77}$$

where we have defined $\overline{N} = \mathbb{E}[N_s]$. According to Section 4.8.3, we can now replace $\overline{N} \approx (1 - p_{\mathsf{col},1})\, N$, namely, the mean of the binomial pmf approximation of (4.73), with $p_{\mathsf{col},1}$ given by (4.71). Interestingly, this bound can be shown to be tight for $N \to \infty$. Finally, by replacing (4.77) into (4.74) a (tight) lower bound for the overall distortion follows. As a remark, it is worth noting that by particularizing (4.74) for $p_{\mathsf{col},1} = 0$ and (4.77) for $p_{\mathsf{col},2} = 0$, we obtain the distortion associated to the reservation-based protocol presented in the previous section.

4.8.4 Resource allocation problem

Here, we attempt to minimize the expected distortion at the FC with respect to $\alpha \in [0, 1]$, which determines the time devoted in each timeslot to sensor-to-CH and CH-to-FC communications ($T_1 = \alpha T_s$ and $T_2 = (1 - \alpha) T_s$, respectively). To that extent, we realize that the only term in (4.74) involved in the minimization w.r.t. α turns out to be (4.75). Therefore, by *i*) recalling from (4.59) and (4.60) that $R_1 = \alpha R_1'$ and $R_2 = (1 - \alpha)\, R_2'$; and *ii*) resorting to the lower bound of (4.77), the minimization problem now reads

$$\min_{\alpha \in [0,1]} \frac{2^{(1-\alpha)\frac{R_2'}{N_c}} \sigma_{v_{\mathsf{CH}}}^2 \left(\overline{N}\right) + \sigma_x^2}{2^{(1-\alpha)\frac{R_2'}{N_c}} - 1}. \tag{4.78}$$

In the sequel, we assume that $2^{(1-\alpha)\frac{R_2'}{N_c}} \gg 1$ or, in other words, that each CH-to-FC link in Layer 2 is capable of conveying large amounts of information[8]. Bearing this in mind, the minimization problem above these lines can be approximated as follows:

$$\min_{\alpha \in [0,1]} \sigma_{v_{\mathsf{CH}}}^2 \left(\overline{N}\right) + \frac{\sigma_x^2}{2^{(1-\alpha)\frac{R_2'}{N_c}}}. \tag{4.79}$$

In subsequent sections, we compute the optimal α[9] for two cases of interest in Layer 1, namely, *i*) high data rate per sensor and *ii*) low data rate per sensor.

[8] The underlying assumption here is that the number of cluster-heads is relatively low.

[9] Strictly speaking, the minimization of (4.79) yields a *quasi*-optimal value of α since (4.79) turns out to be an approximation to the actual distortion.

High data rate per sensor in Layer 1

First, we address the case where $2^{\alpha\frac{R_1'}{N}} \gg 1$ which holds when the cluster size N is small compared to the available channel rate R_1'. In these conditions, the argument in the minimization problem (4.79), re-defined as $f(\alpha)$, simplifies to:

$$f(\alpha) \approx \frac{1}{N\left(1-p_{\mathsf{col},1}\right)}\left(\sigma_v^2+\frac{\sigma_x^2}{2^{\alpha\frac{R_1'}{N}}}\right)+\frac{\sigma_x^2}{2^{(1-\alpha)\frac{R_2'}{N_c}}}. \tag{4.80}$$

This problem is now convex in α and, hence, a closed-form solution α^* can be found by just setting the first derivative of (4.80) to zero, namely

$$\alpha^*=\left(\frac{R_1'}{N}+\frac{R_2'}{N_c}\right)^{-1}\left(\frac{R_2'}{N_c}+\log_2\left(\frac{R_1'N_c}{R_2'N^2\left(1-p_{\mathsf{col},1}\right)}\right)\right). \tag{4.81}$$

From this expression, one concludes that the system tends to allocate more resources (time) to the layer with the lowest channel rate. If $R_1' \to \infty$ (and R_2' does not) then $\alpha^* \to 0$ which prioritizes CH-to-FC transmissions. Conversely, if $R_2' \to \infty$ then $\alpha^* \to 1$, this meaning that sensor-to-CH transmissions become a priority. Besides, the optimal α is clearly an increasing function in the probability of collision in Layer 1 (the higher the probability of collision, the longer the time devoted to Layer 1 to partly compensate for this effect).

Low data rate per sensor in Layer 1

Here, we address the realistic case where the number of sensors in Layer 1 is high and, hence, $2^{\alpha\frac{R_1'}{N}} \to 1$. To start with, we compute

$$\lim_{N\to\infty}\sigma_{v_{\mathsf{CH}}}^2\left(\overline{N}\right)=\frac{\sigma_v^2+\sigma_x^2}{\alpha R_1\left(1-p_{\mathsf{col},1}\right)\ln(2)}. \tag{4.82}$$

and, next, we substitute this into minimization problem (4.79) which is now convex. As in the previous section, a closed-form expression of the optimal operating point can be easily found and it reads

$$\alpha^*=\frac{2N_c}{\ln(2)R_2'}W_0\left(\frac{1}{2}\sqrt{\frac{R_2'\left(\sigma_x^2+\sigma_v^2\right)}{N_cR_1'\left(1-p_{\mathsf{col},1}\right)\sigma_x^2}}e^{\frac{R_2'\ln(2)}{2N_c}}\right) \tag{4.83}$$

with $W_0(\cdot)$ standing for the Lambert function [71]. Similar conclusions to those of the high data rate per sensor case can be drawn from this last expression. However, the optimal α depends now on the quality of the observations at the sensor nodes, as well. For noisy observations (namely, high values of σ_v^2), it is necessary to increase α^* in order not to introduce excessive quantization noise in Layer 1.

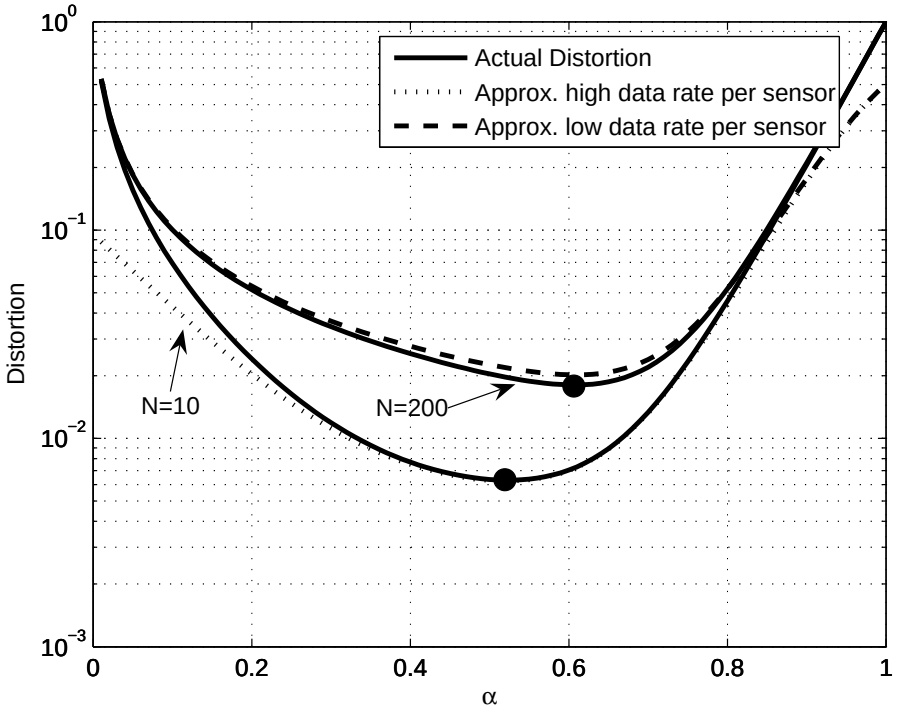


Figure 4.14: Distortion for reservation-based mechanisms in Layer 1 and Layer 2 ($N_c = 3$, $\mathsf{SNR}_1 = 20\text{dB}$, $\mathsf{SNR}_2 = 10\text{dB}$, $W' = 20$, $\sigma_v^2 = 0.05$, $\sigma_x^2 = 1$). Markers on the curves denote the optimal operating points given by (4.81) and (4.83).

4.8.5 Simulations and numerical results

Figure 4.14 illustrates the accuracy of the approximations of the optimization problem given by (4.79) in both high *and* low data rate per sensor scenarios. In particular, we focus on the case where a reservation-based multiple access mechanism is adopted in both layers (i.e. $p_{\mathsf{col},1} = p_{\mathsf{col},2} = 0$). In scenarios with high data rate per sensor ($N = 10$), the approximate distortion given by (4.80) is quite tight and, hence, the optimal value of α can be accurately computed with (4.81). On the contrary, in scenarios with low data rate per sensor ($N = 200$) the approximation (4.80) turns out to be loose. Hence, one has to resort to (4.83) to determine the optimal operating point α^*.

Next, in Fig. 4.15, we show the impact of reservation-based and contention-based mechanisms on the overall performance. Clearly, adopting reservation-based schemes in both layers (curve labeled with 'TDMA Layer 1, TDMA Layer 2') yields the lowest possible distortion for the whole range of α. As expected, the introduction of contention-based mechanisms (and the packet collisions that they entail) results into an increased distortion level. Contention-based mechanisms are particulary harmful in Layer 2 since a packet collision in a CH-to-FC link prevents *all* the data collected by that specific CH from being used to estimate the parameter. The impact of contention-based mechanisms in Layer 1 is moderate: when a packet is dropped, the (noisy) observations sent by other sensors are still helpful for the estimation of the parameter

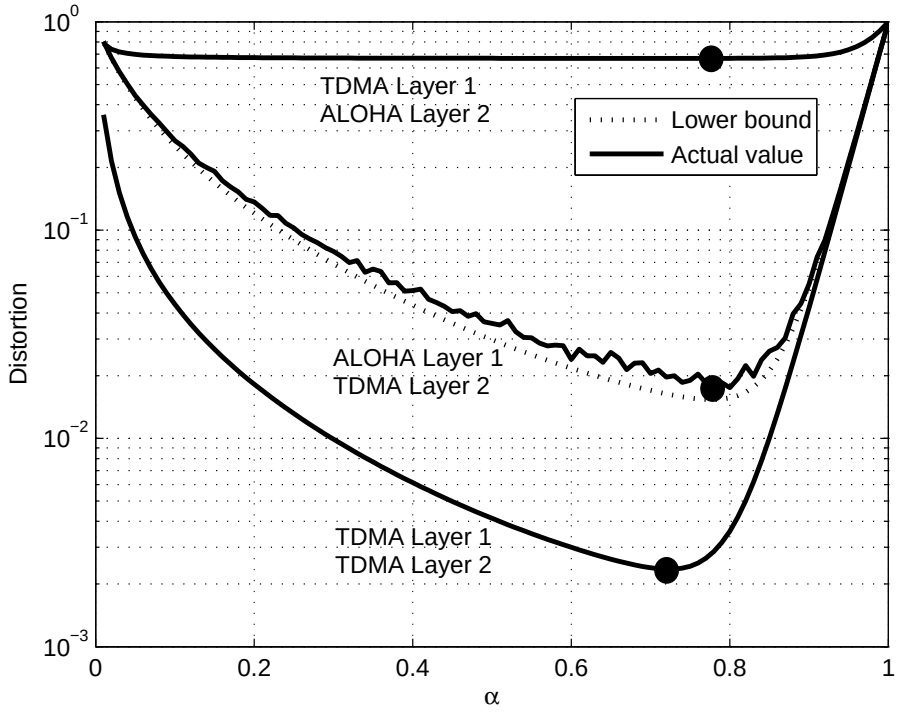


Figure 4.15: Impact of reservation-based and contention-based mechanisms on distortion ($N_c = 3$, $N = 40$, $\mathsf{SNR}_1 = 20\text{dB}$, $\mathsf{SNR}_2 = 10\text{dB}$, $W' = 40$, $\sigma_v^2 = 0.05$, $\sigma_x^2 = 1$). Markers on the curves denote the optimal operating points given by (4.81) and (4.83).

of interest. Besides, we observe that the lower bound that was found by substituting (4.77) into (4.74) is tight (dotted curve). This validates the optimal resource allocations given by (4.81) and (4.83). Finally, the presence of collisions in Layer 1 leads to an increased value of α^*. This effect is captured by the closed-form solutions given by (4.81) and (4.83), as commented in Section 4.8.4.

Finally, Fig. 4.16 depicts the expected distortion at the FC as a function of the signal to noise ratio experienced in Layer 1, SNR_1 (for the optimal α, high number of sensors per cluster case). Interestingly, the *rate* at which the distortion decreases with reservation-based and contention-based is *identical* in both cases. In other words, when the number of sensors per cluster is high, only a constant penalty in terms of distortion can be expected.

4.9 Chapter summary and conclusions

In this chapter, we have first conducted an in-depth analysis of the Quantize-and-Estimate (Q&E) and Compress-and-Estimate (C&E) encoding strategies in (orthogonal) Gaussian and Rayleigh-fading channels under power *and* bandwidth constraints. For the Q&E scheme, we have proved that there exists an optimal number of sensor nodes which minimizes the overall

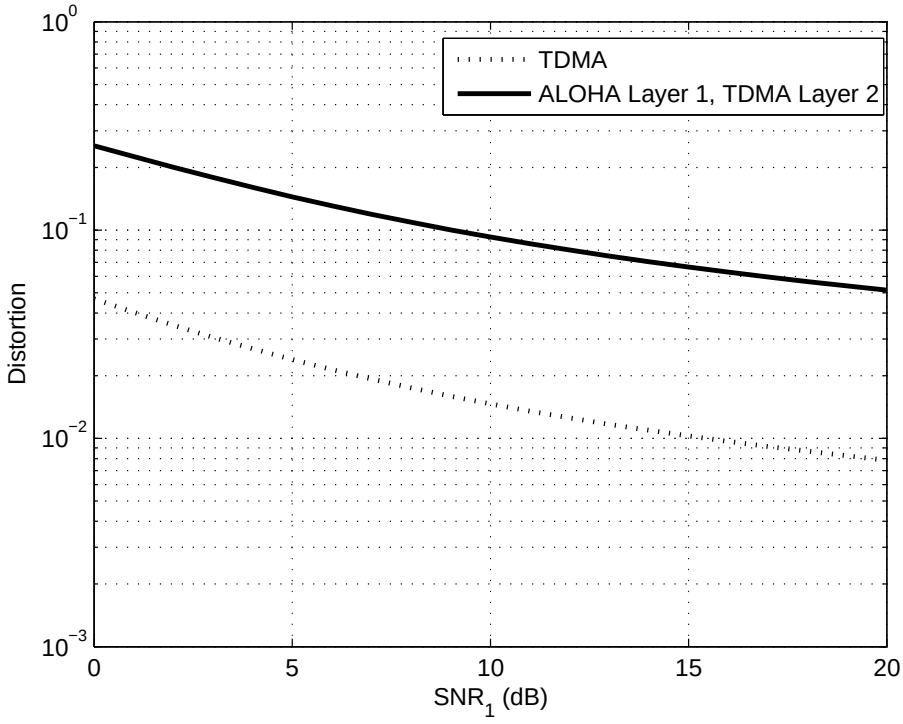


Figure 4.16: Impact of SNR_1 on distortion for a high number of sensors per cluster case ($N_c = 3$, $\mathsf{SNR}_2 = 10\text{dB}$, $W' = 20$, $\sigma_v^2 = 0.05$, $\sigma_x^2 = 1$).

distortion in the estimates. Conversely, in C&E encoding, increasing the number of sensors always pays off. For the Q&E scheme, we have derived an approximate closed-form expression of its optimal operating point (Gaussian channels and some cases of interest in Rayleigh-fading channels without CSIT) and concluded that optimizing on the number of sensors is particularly useful when the observation noise is low. For the C&E scheme, we have analytically shown that encoding the observations in a decreasing order of (sensor-to-FC) channel gains minimizes the resulting distortion. Computer simulation results reveal that ordering is particularly important in scenarios with moderate observation noise or transmit power. We have also derived, in a context of Rayleigh-fading channels, closed-form expressions of the distortion attained by the Q&E and C&E (lower bound) schemes for an asymptotically-high number of sensors. From this, we conclude that, as expected, distortion is lower in the C&E case. Besides, in the absence of CSIT, we have found the optimal value of the *common* and *constant* encoding rate of the Q&E scheme. In other words, we have identified the optimal trade-off in terms of quantization bits vs. the number of observations actually received at the FC (due to outage effects). We have approximately solved the problem for two cases of interest, namely, sensors with high and low observation noise and found out that, interestingly, the lack of CSIT translates into a moderate increase of distortion for the whole range of SNR values.

Second, and unlike the previous analysis where each sensor-to-FC communication occurs in a *reserved* orthogonal channel (e.g. TDMA/FDMA), we have addressed a more realistic sce-

nario, where sensors seize the channel via contention-based multiple-access protocols. We have adopted a hierarchical topology where sensors are grouped into clusters and each cluster is governed by a cluster-head, which is in charge of consolidating the cluster estimate and send it to the FC. First, we have derived a *closed-form* expression of the distortion attained at the FC with a reservation-based protocol (e.g. TDMA) which has been used as a benchmark. Next, we have extended the analysis to encompass the effect of packet collisions stemming from the use of contention-based schemes. Specifically, we have found an approximate (yet tight) expression of the distortion associated to the ALOHA protocol. On that basis, we have identified the optimal time split, α^*, for sensor-to-CH (Layer 1) and CH-to-FC (Layer 2) communications. Furthermore, we have derived (approximate) closed-form expressions of α^* for two cases of interest, namely, high data rate and low data rate per sensor. Simulation results reveal that the adoption of contention-based mechanisms is particulary harmful in Layer 2 whereas their impact in Layer 1 is moderate. Besides, we have found (both analytically and numerically) that the presence of packet collisions in Layer 1 leads to an increased value of α^*. Finally, we have also observed that the *rate* at which the distortion decreases with the SNR_1 is *identical* for reservation-based and contention-based schemes.

4.A Appendix

4.A.1 Quasiconvexity of the distortion function for Q&E encoding and Gaussian channels

We want to prove that the distortion given by (4.16) is a quasiconvex function in N (we relax $N \in \mathbb{R}$). Mathematically, the distortion $D_{N,\mathsf{Q\&E}}$ is a quasiconvex function if its domain and all its sublevel sets

$$\mathcal{S}_\alpha = \left\{ N \in \mathbb{R}^+ \middle| D_{N,\mathsf{Q\&E}} \leq \alpha \right\} \tag{4.84}$$

for $\alpha \in \mathbb{R}$ are convex (i.e. continuous) [69, Chapter 3]

The problem is equivalent to prove that the second term in (4.16) is a *quasiconcave* function or, mathematically, that its domain and all its *superlevel* sets (see definition below) are convex (i.e. continuous). To that aim, we re-write the superlevel sets of (4.84) as follows:

$$\mathcal{S}_\alpha = \left\{ N \in \mathbb{R}^+ \middle| D_{N,\mathsf{Q\&E}} \leq \alpha \right\} \tag{4.85}$$

$$= \left\{ N \in \mathbb{R}^+, \frac{N\left((1+\mathsf{SNR})^{\frac{W}{N}} - 1\right)}{\sigma_v^2 (1+\mathsf{SNR})^{\frac{W}{N}} + \sigma_x^2} \geq \beta \right\} = \mathcal{S}_\beta, \tag{4.86}$$

with $\beta = \frac{1}{\alpha} - \frac{1}{\sigma_x^2} \in \mathbb{R}$. After some manipulations, the above sets can be re-written as:

$$\mathcal{S}_\beta = \left\{ N \in \mathbb{R}^+, f(N) \geq \sigma_x^2 + \sigma_v^2 \right\} \tag{4.87}$$

with

$$f(N) = \left(\frac{N}{\beta} - \sigma_v^2 \right) \left((1+\mathsf{SNR})^{\frac{W}{N}} - 1 \right). \tag{4.88}$$

Hence, the problem is equivalent to prove that $f(N)$ is also a quasiconcave function in N. On the one hand, we have that $f(N)$ asymptotically converges to

$$\lim_{N\to\infty} f(N) = \frac{W}{\beta} \ln(1+\mathsf{SNR}). \tag{4.89}$$

On the other, from the second derivative of $f(N)$ w.r.t. N it easily follows that, for $\beta < \frac{W\log(1+\mathsf{SNR})}{2\sigma_v^2}$

$$f(N) \rightarrow \begin{cases} \text{concave if} & N < \frac{W\log(1+\mathsf{SNR})\sigma_v^2\beta}{W\log(1+\mathsf{SNR}) - 2\sigma_v^2\beta} \\ \text{convex if} & N > \frac{W\log(1+\mathsf{SNR})\sigma_v^2\beta}{W\log(1+\mathsf{SNR}) - 2\sigma_v^2\beta} \end{cases}. \tag{4.90}$$

whereas for $\beta > \frac{W\log(1+\mathsf{SNR})}{2\sigma_v^2}$, $f(N)$ is concave for all $N > 0$. According to this analysis along with the asymptotic value computed in (4.89), $f(N)$ is necessarily a quasiconcave function. Besides, it has (at most) one change of sign in its first derivative. From all this, one concludes that S_α are convex sets and, hence, distortion is a quasiconvex function in N with a single optimal value N^*.

4.A.2 Convergence in probability of $\sum_{i=1}^{N} \frac{g(\gamma_i,N)}{f(\gamma_i,N)}$ for large N

We want to prove that

$$\sum_{k=1}^{N} \frac{g(\gamma_k, N)}{f(\gamma_k, N)} \xrightarrow{\text{P}} \frac{1}{\sigma_x^2 + \sigma_v^2} \sum_{k=1}^{N} g(\gamma_k, N) \tag{4.91}$$

or, alternatively, that

$$\sum_{k=1}^{N} \frac{g(\gamma_k, N)}{f(\gamma_k, N)} \left(\frac{f(\gamma_k, N) - (\sigma_x^2 + \sigma_v^2)}{\sigma_x^2 + \sigma_v^2} \right) \xrightarrow{\text{P}} 0 \tag{4.92}$$

for $N \to \infty$. Besides, from their respective definitions in (4.24), we know that $f(\gamma_i, N)$ and $g(\gamma_i, N)$ are related through

$$g(\gamma_i, N) = \frac{f(\gamma_i, N) - (\sigma_x^2 + \sigma_v^2)}{\sigma_v^2}. \tag{4.93}$$

Replacing (4.93) into (4.92) yields

$$\sum_{k=1}^{N} \frac{\left(f(\gamma_k, N) - (\sigma_x^2 + \sigma_v^2)\right)^2}{\sigma_v^2 f(\gamma_k, N) (\sigma_x^2 + \sigma_v^2)} \tag{4.94}$$

$$\leq \sum_{k=1}^{N} \frac{\left(f(\gamma_k, N) - (\sigma_x^2 + \sigma_v^2)\right)^2}{\sigma_v^2 (\sigma_x^2 + \sigma_v^2)^2} \tag{4.95}$$

$$\leq N \frac{\left(f(\gamma_{(1:N)}, N) - (\sigma_x^2 + \sigma_v^2)\right)^2}{\sigma_v^2 (\sigma_x^2 + \sigma_v^2)^2} \tag{4.96}$$

Inequality of (4.95) follows from the fact that $f(\gamma_i, N) \geq \sigma_x^2 + \sigma_v^2$. Inequality (4.96), where $\gamma_{(1:N)} = \max_{i=1..N}\{\gamma_i\}$ denotes the first order statistic of a set of N random variables, is a straightforward upper bound on the summation term. From this last expression, we want to show that

$$\lim_{N\to\infty} \Pr\left\{ N \left(f\left(\gamma_{(1:N)}\right) - \left(\sigma_x^2 + \sigma_v^2\right)\right)^2 \leq \epsilon \right\} = 1. \tag{4.97}$$

For the sake of clarity and without loss of generality, in the sequel we particularize the expressions for the $\mathsf{SNR} = 1$ and $W = 1$ case and, hence, this last expression can be re-written as:

$$\lim_{N\to\infty} \Pr\left\{ \gamma_{(1:N)} \leq \left(\frac{\sqrt{\epsilon}}{\sigma_v^2 \sqrt{N}} + 1 \right)^N - 1 \right\}. \tag{4.98}$$

By using the CDF of the first order statistic $\gamma_{(1:N)}$, which is defined as $F_{\gamma_{(1:N)}}(x) = F_{\gamma}^{N}(x) = (1 - e^{-x})^N$, one finally obtains

$$= \lim_{N\to\infty} \left(1 - \exp\left(- \left(\frac{\sqrt{\epsilon}}{\sigma_v^2 \sqrt{N}} + 1 \right)^N + 1 \right) \right)^N = 1,$$

which concludes the proof.

4.A.3 Convergence in probability of $\sum_{i=1}^{N} g(\gamma_i, N)$ for large N

By resorting to the power series $a^x = \sum_{k=0}^{\infty} \frac{x^k \ln^k(a)}{k!}$ [90, 1.211.2], we can factorize the summation term as follows:

$$\begin{aligned}\sum_{i=1}^{N} g(\gamma_i, N) &= \frac{1}{N}\sum_{i=1}^{N} W\ln\left(1+\mathsf{SNR}\gamma_i\right) \\ &+ \sum_{i=1}^{N}\sum_{k=2}^{\infty}\frac{W^k \ln^k\left(1+\mathsf{SNR}\gamma_i\right)}{k!N^k}.\end{aligned} \tag{4.99}$$

The second term in (4.99) vanishes as $N \to \infty$. As for the first term, by the weak law of large numbers, we have

$$\frac{1}{N}\sum_{i=1}^{N} W\ln\left(1+\mathsf{SNR}\gamma_i\right) \xrightarrow{\mathrm{P}} \mathbb{E}_\gamma\left[W\ln\left(1+\mathsf{SNR}\gamma_i\right)\right] \tag{4.100}$$

where the expectation term can be easily computed as

$$\mathbb{E}_\gamma\left[W\ln\left(1+\gamma_i\mathsf{SNR}\right)\right] = We^{\frac{1}{\mathsf{SNR}}}\Gamma\left(0, \frac{1}{\mathsf{SNR}}\right) \tag{4.101}$$

with $\Gamma(a,x)$ standing for the incomplete Gamma function [90, 8.350.2]. In conclusion, we have that

$$D_{\infty,\mathsf{Q\&E}} \xrightarrow{\mathrm{P}} \left(\frac{1}{\sigma_x^2} + \frac{We^{\frac{1}{\mathsf{SNR}}}\Gamma\left(0,\frac{1}{\mathsf{SNR}}\right)}{\sigma_v^2+\sigma_x^2}\right)^{-1}, \tag{4.102}$$

which concludes the proof.

4.A.4 Proof of the tightness of bound (4.43)

In this section, we prove that the bound derived in (4.43) is asymptotically tight for large N or, in other words, that the probability $\Pr\left\{\left|\frac{\frac{1}{\sigma_x^2}+X_N}{\frac{1}{\sigma_x^2}+\mu_N} - 1\right| \geq \delta\right\}$ can be made arbitrarily small for any $\delta > 0$, where X_N is a random variable with an arbitrary distribution of mean $\mu_N > 0$ and variance σ_N^2. For any $\delta > 0$, we have

$$\begin{aligned}\Pr\left\{\left|\frac{\frac{1}{\sigma_x^2}+X_N}{\frac{1}{\sigma_x^2}+\mu_N} - 1\right| \geq \delta\right\} &= \Pr\left\{\left|\frac{X_N-\mu_N}{\frac{1}{\sigma_x^2}+\mu_N}\right| \geq \delta\right\} \\ &\leq \frac{1}{\delta^2}\frac{\sigma_N^2}{\left(\mu_N+\frac{1}{\sigma_x^2}\right)^2} \qquad (4.103) \\ &\leq \frac{\sigma_x^2}{2\delta^2}\frac{\sigma_N^2}{\mu_N}, \qquad (4.104)\end{aligned}$$

where inequality (4.103) follows from Tchebychev's bound. Since, in our case,

$$X_N = \frac{N_{\mathsf{d}} \left(\left(1 + \mathsf{SNR} \cdot F_{\gamma}^{-1}(p_{\mathsf{out}})\right)^{\frac{W}{N}} - 1 \right)}{\sigma_v^2 \left(1 + \mathsf{SNR} \cdot F_{\gamma}^{-1}(p_{\mathsf{out}})\right)^{\frac{W}{N}} + \sigma_x^2}$$

turns out to be a binomial random variable, it is straightforward to compute σ_N^2 and μ_N to realize that the ratio $\frac{\sigma_N^2}{\mu_N} \to 0$ as N grows without bound. Therefore, from (4.104) we have that

$$\lim_{N\to\infty} \frac{\frac{1}{\sigma_x^2} + X_N}{\frac{1}{\sigma_x^2} + \mu_N} \stackrel{\mathrm{P}}{=} 1 \tag{4.105}$$

where $\stackrel{\mathrm{P}}{=}$ denotes convergence in probability. Since the point-wise limit $\lim_{N\to\infty} \frac{1}{\sigma_x^2} + \mu_N = \frac{1}{\sigma_x^2} + \mu_\infty$, i.e. converges to a constant value, we have that

$$\lim_{N\to\infty} \frac{1}{\sigma_x^2} + X_N \stackrel{\mathrm{P}}{=} \frac{1}{\sigma_x^2} + \mu_\infty. \tag{4.106}$$

This fact means that the bound derived in (4.43) is asymptotically tight in N, which concludes the proof.

4.A.5 Binomial approximation of N_s

By neglecting the border effects, the value of p_n for the random variable N_s reads

$$p_n = \prod_{i=0}^{n-1} \left(1 - \frac{2}{N-i}\right)^{N-i-1}. \tag{4.107}$$

By considering that for large N the probability that n sensors packets are received without collisions, when n is close to N, is negligible then, for a fixed and relatively small n and large N we have that

$$\begin{aligned} p_n &= \lim_{N\to\infty} \prod_{i=0}^{n} \left(1 - \frac{2}{N-i}\right)^{N-i-1} = \\ &= \lim_{N\to\infty} (1 - p_{\mathsf{col}})^n = e^{-2n}. \end{aligned} \tag{4.108}$$

Now, by substituting (4.108) into (4.72) and due to the fact that the sum of probabilities of the approximate pmf must be 1, one concludes that $q_{N-n} \approx \lim_{N\to\infty} p_{\mathsf{col}}^{N-n} = (1 - e^{-2})^{N-n}$.

Chapter 5

Estimation of Random Fields with Wireless Sensor Networks

In this chapter, we study the problem of random field estimation with wireless sensor networks. We consider two encoding strategies, namely Compress-and-Estimate (C&E) and Quantize-and-Estimate (Q&E), which operate with and without side information at the decoder, respectively. We focus our attention on two scenarios of interest: *delay-constrained* networks, in which the observations collected in a particular timeslot must be immediately encoded and conveyed to the Fusion Center (FC); and *delay-tolerant* (DT) networks, where the time horizon is enlarged to a number of consecutive timeslots. For both scenarios and encoding strategies, we extensively analyze the distortion in the reconstructed random field. In DT scenarios, we find closed-form expressions of the optimal number of samples to be encoded in each timeslot (Q&E and C&E cases). Besides, we identify buffer stability conditions and a number of interesting distortion vs. buffer occupancy trade-offs. Latency issues in the reconstruction of the random field are addressed as well. Finally, we address the case in which the system operates without instantaneous transmit CSI at the sensor nodes (for a delay-constrained scenario). As in the previous chapter, we consider that the sensors adopt a *common* and *constant* encoding rate. The constant encoding rate along with the network size are optimized in order to minimize the attainable distortion in the reconstruction of the spatial random field.

5.1 Introduction

In many cases, the physical phenomena observed by sensor networks (e.g. environmental parameters, crop conditions) can be modelled as a spatial random field. The set of observations captured by different sensor nodes are, thus, correlated in space. Therefore, the goal now is the reconstruction of the spatial random field at *all* the spatial points (see e.g. [21, 25, 91, 92]).

In a context of random field *estimation* with WSNs, the pioneering work of [93] introduced the so-called "bit-conservation principle". The authors prove that, for spatially *bandlimited* processes, the bit budget per Nyquist-period can be arbitrarily re-allocated along the quantization precision and/or the space (by adding more sensor nodes) axes, while retaining the same decay profile of the reconstruction error. In [94] and, again, for bandlimited processes with *arbitrary* statistical distributions, the authors propose a mathematical framework to study the impact of the random sampling effect (arising from the adoption of contention-based multiple-access schemes) on the resulting estimation accuracy. For *Gaussian* observations, [26] presents a feedback-assisted Bayesian framework for adaptive quantization at the sensor nodes.

From a different perspective but still in the context of random field estimation, [25] proposes a novel MAC protocol which minimizes the number of attempts to transmit correlated data. By doing so, not only energy but also bandwidth is preserved. Besides, in [24] the authors investigate the impact of *random* sampling, as opposed to deterministic sampling (i.e. equally-spaced sensors) which is difficult to achieve in practice, in the reconstruction of the field. The main conclusion is that, whereas deterministic sampling pays off in the high-SNR regime, both schemes exhibit comparable performances in the low-SNR regime.

In scenarios with non-reciprocal (e.g. FDD systems) fading channels, it is often assumed that only statistical CSI is available at the transmitter. Consequently, the encoding rate at the sensor nodes cannot be dynamically adjusted to match instantaneous channel conditions. In this context, the estimation of a spatially homogeneous parameter without instantaneous CSI has been considered in the previous Chapter (see also [79, 80]). Unlike previous works, for spatial random fields the outage events experienced in the sensor-to-FC links modifies the sampling pattern and, hence, needs to be investigated.

5.1.1 Contribution

In this chapter, we go one step beyond Chapters 3 and 4 and address the problem of (non-necessarily bandlimited) random field estimation via wireless sensor networks. To that aim, we adopt the Q&E and C&E encoding schemes of [76] and analyze their performance in two scenarios of interest: *delay-constrained* (DC) and *delay-tolerant* (DT) sensor networks. In DC scenarios, the observations collected in a particular timeslot must be immediately encoded

and conveyed to the FC. In DT networks, on the contrary, the time horizon is enlarged to L consecutive timeslots. Clearly, this entails the use of local buffers but, in exchange, the distortion in the reconstructed random field is lower. To capitalize on this, we derive closed-form expressions of the distortion attainable in DT scenarios (unlike in [24,25,94], we explicitly take into account quantization effects) and, from this, we determine the optimal number of samples to be encoded in each of the L timeslots as a function of the channel conditions of that particular timeslot. Along with that, we identify under which circumstances the buffers are stable (i.e. buffer occupancy does not grow without bound) and, besides, we study a number of distortion vs. buffer occupancy trade-offs. Complementarily, we analyze the latency in the reconstruction of n consecutive realizations (i.e. those collected in one timeslot) of the random field.

Finally and unlike in previous works, we address the case where sensors operate in the absence of transmit CSI (for delay-constrained applications). Consequently, we propose a *constant-rate* encoding strategy which unavoidably entails some outage probability in Rayleigh-fading scenarios. This effect, along with the spatial sampling process and the power and bandwidth constraints that we impose, results into some distortion that we attempt to minimize by carefully selecting the optimal number of sensor nodes to be deployed and the corresponding encoding rate.

The contents of this chapter have been partly published in [95–99].

The chapter is organized as follows. First, in Section 5.2, we present the signal model, the communication model and the distortion analysis respectively. Next, Section 5.3 focuses on the strategies for delay-constrained WSNs. In Sections 5.4 and 5.5, we study the compress-and-estimate and quantize-and-estimate strategies for delay-tolerant WSNs. Subsequently, Section 5.6 addresses the latency analysis for the delay-constrained strategies. Next, in Section 5.7, in the context of delay-constrained applications, we consider the case where sensors operate in the absence of instantaneous transmit CSI. Finally, we close the chapter by summarizing the main findings in Section 5.8.

5.2 Signal model and distortion analysis

Let $Y(s)$ be a one-dimensional random field defined in the range $s \in [0, d]$, with s denoting the spatial variable. As in [23–25], we adopt a stationary homogeneous Gaussian Markov Ornstein-Uhlenbeck (GMOU) model [100] to characterize the dynamics and spatial correlation of $Y(s)$. GMOU random fields obey the following linear stochastic differential equation:

$$dY(s) = \theta Y(s)\, ds + \sigma W(s) \tag{5.1}$$

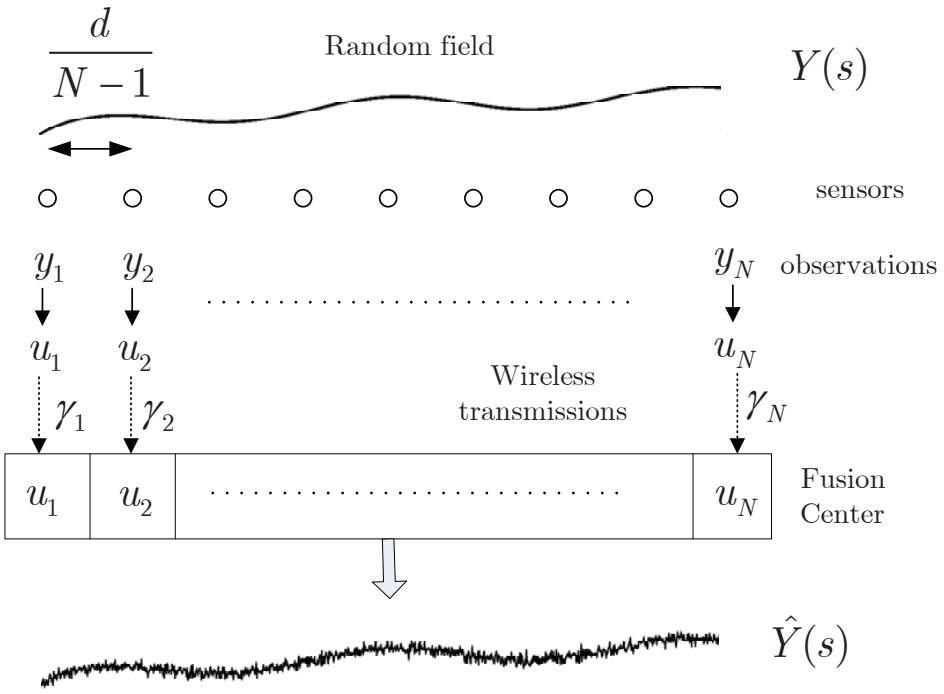


Figure 5.1: System model.

where, by definition, $Y(s) \sim \mathcal{N}\left(0, \sigma_y^2\right)$ with $\sigma_y^2 = \frac{\sigma}{2\theta}$, $W(s)$ denotes Brownian Motion with unit variance parameter, and θ, σ are constants reflecting the (spatial) variability of the field and its *noisy* behaviour, respectively. According to this model, the autocorrelation function is given by $R_Y(s_1, s_2) = \sigma_y^2 e^{-\theta|s_2 - s_1|}$ and, hence, the process is not (spatially) bandlimited.

The random field is uniformly sampled by N sensor nodes, with inter-sensor distance given by $d/(N-1) \simeq d/N$ (see Fig. 5.1). The spatial samples can thus be readily expressed as follows [22]:

$$y_k = Y\left(k\frac{d}{N}\right) = e^{-\theta\frac{d}{2N}} y_{k-1} + n_k \quad ; k = 1, \ldots, N \tag{5.2}$$

where $n_k \sim \mathcal{N}\left(0, \sigma_y^2\left(1 - e^{-\theta\frac{d}{N}}\right)\right)$.

5.2.1 Communication Model

As shown in Fig. 5.2, each time slot is composed of two distinctive phases: *i*) the *sensing* phase and, *ii*) the *transmission* phase. In the former, each sensor collects and stores in a local buffer a large block of n independent and consecutive observations $\mathbf{y}_k = [y_k^{(1)}, \ldots, y_k^{(n)}]^T$. Next, in the transmission phase, the length-n vector of observations, $\mathbf{y}_k$, is block-encoded into a length-n codeword $\mathbf{u}_k(v_k) \in \mathcal{C}$ at a rate of R_k bits per sample. The encoding (quantization) process is modeled through the auxiliary random variable $\mathbf{u}_k = \mathbf{y}_k + \mathbf{z}_k$ with $\mathbf{z}_k \sim \mathcal{N}(0, \sigma_{z_k}^2 \mathbf{I})$ and

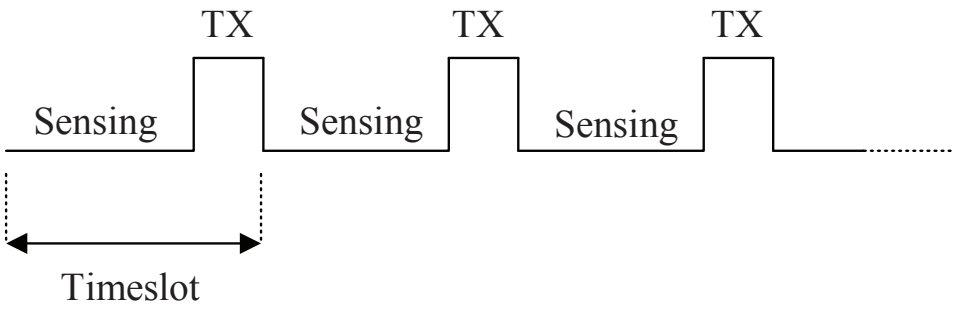


Figure 5.2: Sensing and transmission phases.

statistically independent of $\mathbf{y}_k$. The corresponding index[1] $v_k \in \{1, \ldots, 2^{nR_k}\}; k = 1 \ldots N$ is then conveyed[2] to the FC, in a total of $\frac{m}{N}$ channel uses, over one of the N *orthogonal* channels available. For a reliable transmission to occur, the encoding rate R_k must satisfy:

$$nR_k \leq \frac{m}{N} \log_2 (1 + \mathsf{SNR}\gamma_k) \quad [\text{b/s}] \tag{5.3}$$

where SNR stands for the average signal-to-noise ratio experienced in the sensor-to-FC channels. Besides, $\gamma_1, \ldots, \gamma_N$ denote the channel (squared) gains that, in the sequel, we model as independent and exponentially-distributed unit-mean random variables (Rayleigh-fading channels). We further assume that the channel gains are independent over time slots (block fading assumption).

From the *set* of decoded codewords, the FC reconstructs the random field $Y(s)$ for *all* $s \in [0, d]$. As a result of the spatial sampling process and the channel bandwidth constraint, the reconstructed field $\hat{Y}(s)$ is subject to some distortion which will be characterized by the following metric:

$$D(s) = \mathbb{E}\left[\left|\hat{Y}(s) - Y(s)\right|^2\right] \quad ; \forall s \in [0, d]. \tag{5.4}$$

5.2.2 Distortion analysis: a general framework

For the distortion metric given by (5.4), the optimal estimator turns out to be the posterior mean given all the codewords[3] $\mathbf{u}_r = [u_1, \ldots, u_N]^T$, that is, the MMSE estimator [18, Ch. 10]:

$$\hat{Y}(s) = \mathbb{E}\left[Y(s)|\mathbf{u}_r\right] \quad ; \forall s \in [0, d]\,. \tag{5.5}$$

For mathematical tractability, however, only the *two closest* decoded codewords, namely u_{k-1} and u_k, will be used to reconstruct $Y(s)$ for *all* the corresponding intermediate spatial points

[1] As it will become apparent later, the codebook $\mathcal{C}$ consists of, at most, 2^{nR_k} codewords.

[2] In the case of random binning, instead of sending the index of the codeword, the sensor sends the index of the bin where the codeword $\mathbf{u}_k$ is contained. In this case, one can re-define nR_k as the number of bins and, hence, the actual bits per sample needed to send $\mathbf{u}_k$. It is worth noting that random binning is assumed in the CEDC and CEDT strategies ahead. For further details, the reader is referred to Section 2.3.3.

[3] Without loss of generality, we focus on the per-sample distortion.

(see Fig. 5.1), i.e.

$$\hat{Y}(s) = \mathbb{E}\left[Y(s)|u_{k-1}, u_k\right]; \quad \forall s \in \left[(k-1)\frac{d}{N}, k\frac{d}{N}\right], \quad k = 2, \ldots, N. \tag{5.6}$$

For the ease of notation and without loss of generality, in the sequel we will assume $k = 1$ and, hence, the interval between observations becomes $s \in \left[0, \frac{d}{N}\right]$. The distortion associated to the estimator (5.6) reads [18, Ch. 10]

$$D_k(s) = \sigma^2_{Y(s)|u_{k-1},u_k} = \sigma^2_{Y(s)|u_{k-1}} - \frac{\mathrm{Cov}^2\left(Y(s), u_k|u_{k-1}\right)}{\sigma^2_{u_k|u_{k-1}}} \tag{5.7}$$

where

$$\sigma^2_{Y(s)|u_{k-1}} = \left(\frac{1}{\sigma_y^2} + \frac{e^{-\theta s}}{\left(1 - e^{-\theta s}\right)\sigma_y^2 + \sigma^2_{z_{k-1}}}\right)^{-1}. \tag{5.8}$$

After some algebra, we obtain

$$\begin{aligned} \mathrm{Cov}\left(Y(s), u_k|u_{k-1}\right) &= \mathbb{E}\left[\left(Y(s) - \mathbb{E}\left[Y(s)\middle|u_{k-1}\right]\middle|u_{k-1}\right)\left(u_k - \mathbb{E}\left[u_k\middle|u_{k-1}\right]\middle|u_{k-1}\right)\right] & (5.9)\\ &= \sqrt{e^{-\theta\left(\frac{d}{N}-s\right)}}\sigma^2_{Y(s)|u_{k-1}}, & (5.10)\end{aligned}$$

and

$$\sigma^2_{u_k|u_{k-1}} = e^{-\theta\left(\frac{d}{N}-s\right)}\sigma^2_{Y(s)|u_{k-1}} + \left(1 - e^{-\theta\left(\frac{d}{N}-s\right)}\right)\sigma_y^2 + \sigma^2_{z_k}.$$

It is worth noting that the variance of the quantization noise $\sigma^2_{z_{k-1}}$ and $\sigma^2_{z_k}$ are determined by the encoding strategy in use at the sensor nodes.

5.3 Delay-constrained WSNs

In delay-constrained applications, the n samples collected in the sensing phase of a given timeslot must be necessarily encoded and transmitted to the FC in the subsequent transmission phase. The goal of this section is to particularize the analysis of Section 5.2.2 and compute the average distortion for the cases of Delay-Constrained Quantize-and-Estimate (QEDC) and Compress-and-Estimate (CEDC) encoding strategies.

5.3.1 Quantize-and-Estimate: average distortion

In this approach, each sensor encodes its observation regardless of any side information that could be made available by the FC. From [35], the following inequality should hold for the rate at the output of the k-th encoder (quantizer):

$$R_k \geq \mathrm{I}\left(y_k; u_k\right) \quad [\mathrm{b/sample}] \tag{5.11}$$

with $\mathrm{I}(\cdot;\cdot)$ standing for the mutual information. As discussed before, the encoding (quantization) process is modeled (see e.g. [76, 78] for further details) through the auxiliary variable $\mathbf{u}_k = \mathbf{y}_k + \mathbf{z}_k$ with $\mathbf{z}_k \sim \mathcal{N}(0, \sigma_{z_k}^2 \mathbf{I})$ and statistically independent of $\mathbf{y}_k$. From this, the minimum rate per sample can be expressed as follows:

$$\mathrm{I}(y_k; u_k) \;=\; \mathrm{H}(u_k) - \mathrm{H}(u_k|y_k) = \log\left(1 + \frac{\sigma_y^2}{\sigma_{z_k}^2}\right) \quad [\text{b/sample}]\,. \tag{5.12}$$

From (5.3), (5.11) and (5.12) we have that, necessarily

$$\frac{m}{N}\log_2\left(1 + \mathsf{SNR}\cdot\gamma_k\right) \geq n\log_2\left(1 + \frac{\sigma_y^2}{\sigma_{z_k}^2}\right). \tag{5.13}$$

By letting equality hold in (5.13), the minimum variance of the *quantization* noise yields

$$\sigma_{z_k}^2 = \frac{\sigma_y^2}{\left(1 + \mathsf{SNR}\gamma_k\right)^{\frac{W}{N}} - 1} \quad ; \quad k = 1, \ldots, N \tag{5.14}$$

with $W = \frac{m}{n}$ standing for the channel uses-to-samples ratio. By substituting (5.14) into (5.7), the distortion in an arbitrary spatial point s in the k-*th* segment reads

$$D_k^{\mathsf{QEDC}}(s) \;=\; \left(\frac{1}{\sigma_{Y(s)|u_{k-1}}^2} + \frac{e^{-\theta\left(\frac{d}{N}-s\right)}\left(\left(1+\mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N}} - 1\right)}{\left(\left(1+\mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N}} - 1\right)\left(1 - e^{-\theta\left(\frac{d}{N}-s\right)}\right)\sigma_y^2 + \sigma_y^2}\right)^{-1} \tag{5.15}$$

with

$$\sigma_{Y(s)|u_{k-1}}^2 = \left(\frac{1}{\sigma_y^2} + \frac{e^{-\theta s}\left(\left(1+\mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N}} - 1\right)}{\left(\left(1+\mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N}} - 1\right)\left(1 - e^{-\theta s}\right)\sigma_y^2 + \sigma_y^2}\right)^{-1}. \tag{5.16}$$

The average distortion (over the spatial variable s) in the $k-th$ network segment can be computed as

$$\overline{D}_k^{\mathsf{QEDC}} = \frac{N}{d}\int_0^{\frac{d}{N}} D_k^{\mathsf{QEDC}}(s)ds, \tag{5.17}$$

and, from this, the average distortion (over channel realizations) follows:

$$\overline{D}^{\mathsf{QEDC}} = \mathbb{E}_{\gamma_1,\ldots,\gamma_N}\left[\frac{1}{N-1}\sum_{k=1}^{N-1}\overline{D}_{k+1}^{\mathsf{QEDC}}\right]. \tag{5.18}$$

5.3.2 Compress-and-Estimate: average distortion

In this approach, we allow each sensor (encoder) to use the side information provided by its neighbors. For simplicity, we let each sensor to encode its current observation u_k based only

on the adjacent sensor (encoded) observation[4] u_{k-1}. Accordingly, we have that the minimum rate per sample can be expressed as follows:

$$\begin{aligned} R_k \geq \mathrm{I}\left(y_k; u_k | u_{k-1}\right) &= \mathrm{H}(u_k|u_{k-1}) - \mathrm{H}(u_k|y_k, u_{k-1}) \\ &= \mathrm{H}\left(y_k + z_k|u_{k-1}\right) - \mathrm{H}\left(y_k + z_k|y_k\right) \\ &= \log_2\left(1 + \frac{\sigma^2_{y_k|u_{k-1}}}{\sigma^2_{z_k}}\right) \quad [\mathrm{b/sample}]. \end{aligned} \tag{5.19}$$

where the second equality is due to the fact that $u_k \leftrightarrow y_k \leftrightarrow u_{k-1}$ form a Markov chain. Bearing this in mind, for a reliable transmission we must satisfy:

$$\frac{m}{N}\log_2\left(1 + \mathsf{SNR}\cdot\gamma_k\right) \geq n\log_2\left(1 + \frac{\sigma^2_{y_k|u_{k-1}}}{\sigma^2_{z_k}}\right). \tag{5.20}$$

By taking equality in (5.20), we can compute the minimum variance of the *quantization* noise $\sigma^2_{z_k}$ as

$$\sigma^2_{z_k} = \frac{\sigma^2_{y_k|u_{k-1}}}{\left(1 + \mathsf{SNR}\gamma_k\right)^{\frac{W}{N}} - 1} \quad ; \quad k = 1, \ldots, N, \tag{5.21}$$

where $\sigma^2_{y_k|u_{k-1}}$ can be easily computed as follows:

$$\sigma^2_{y_k|u_{k-1}} = e^{-\theta\left(\frac{d}{N} - s\right)}\sigma^2_{Y(s)|u_k} + \left(1 - e^{-\theta\left(\frac{d}{N} - s\right)}\right)\sigma^2_y. \tag{5.22}$$

From (5.7), the distortion at an arbitrary spatial point s reads:

$$D_k^{\mathsf{CEDC}}(s) = \frac{\sigma^2_y\sigma^2_{Y(s)|u_{k-1}}\left(e^{\theta\left(\frac{d}{N} - s\right)} - 1\right)}{\sigma^2_y\left(e^{\theta\left(\frac{d}{N} - s\right)} - 1\right) + \sigma^2_{Y(s)|u_{k-1}}} + \frac{\sigma^4_{Y(s)|u_{k-1}}\left(1 + \mathsf{SNR}\gamma_k\right)^{-\frac{W}{N}}}{\sigma^2_y\left(e^{\theta\left(\frac{d}{N} - s\right)} - 1\right) + \sigma^2_{Y(s)|u_{k-1}}}. \tag{5.23}$$

with

$$\sigma^2_{Y(s)|u_{k-1}} = \left(\frac{1}{\sigma^2_y} + \frac{e^{-\theta s}\left(\left(1 + \mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N}} - 1\right)}{\left(\left(1 + \mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N}} - 1\right)\left(1 - e^{-\theta s}\right)\sigma^2_y + \sigma^2_{y_{k-1}|u_{k-2}}}\right)^{-1}. \tag{5.24}$$

The average distortion for each network segment can be computed as follows:

$$\overline{D}_k^{\mathsf{CEDC}} = \frac{N}{d}\int_0^{\frac{d}{N}} D_k^{\mathsf{CEDC}}(s) \tag{5.25}$$

and, finally, the average distortion (over the channel realizations and network segments) yields:

$$\overline{D}^{\mathsf{CEDC}} = \mathbb{E}_{\gamma_1,\ldots,\gamma_N}\left[\frac{1}{N-1}\sum_{k=1}^{N-1}\overline{D}_{k+1}^{\mathsf{CEDC}}\right]. \tag{5.26}$$

[4] Alternatively, we could use *all* the sensor observations but due to the (spatial) Markov property of the random field model, this would not decrease significantly the encoding rate.

5.4 Delay-tolerant WSNs with Quantize-and-Estimate encoding

Here, we only impose a *long-term* delay constraint: the Ln samples collected in L consecutive timeslots must be conveyed to the FC in such L timeslots. In other words, sensors have now the flexibility to encode and transmit a variable number of samples in each time slot. This provide additional degrees of freedom to adjust the (per-sample) encoding rate to the actual channel conditions and, by doing so, attain a lower distortion.

Let $n_k(i) = \alpha_k(i)n$ be the number of samples encoded in m/N channel uses by sensor k in time-slot i. As in the previous section, we need

$$\frac{m}{N}\log_2\left(1+\mathsf{SNR}\cdot\gamma_k(i)\right) \geq \alpha_k(i)n\log_2\left(1+\frac{\sigma_y^2}{\sigma_{z_k}^2}\right) \quad ; \quad k=1,\ldots,N. \tag{5.27}$$

By replacing $\sigma_{z_k}^2$ from (5.27) into (5.7), the distortion per timeslot yields

$$D_{k,\alpha_k(i)}^{\mathsf{QEDT}}(s) = \left(\frac{1}{\sigma_{Y(s)|u_{k-1}}^2}+\frac{e^{-\theta\left(\frac{d}{N}-s\right)}\left(\left(1+\mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N\alpha_k}}-1\right)}{\left(\left(1+\mathsf{SNR}\gamma_k\left(i\right)\right)^{\frac{W}{N\alpha_k}}-1\right)\left(1-e^{-\theta\left(\frac{d}{N}-s\right)}\right)\sigma_y^2+\sigma_y^2}\right)^{-1} \tag{5.28}$$

The ultimate goal is to minimize the *average* distortion over L timeslots at an arbitrary spatial point s (the average distortion over the entire random field will be computed in Section 5.4.1 ahead). Hence, the optimization problem can be posed as follows[5]:

$$\min_{\alpha_k(1),\ldots,\alpha_k(L)} \quad \frac{1}{L}\sum_{i=1}^{L}\alpha_k(i)D_{k,\alpha_k(i)}^{\mathsf{QEDT}}(s) \tag{5.29}$$

$$\text{s.t.} \qquad \sum_{i=1}^{L}\alpha_k(i)n = Ln \tag{5.30}$$

where the constraint in (5.30) is introduced to ensure the stability of the system. Unfortunately, a closed form solution cannot be obtained for the general case. Alternatively, we consider a suboptimal encoding strategy: sensor k will assume that the FC does not exploit u_{k-1} (the codeword sent by the adjacent sensor) but only u_k in order to reconstruct the random field $Y(s)$ in $s \in \left[(k-1)\frac{d}{N}, k\frac{d}{N}\right]$.[6] The new cost function can be readily expressed as follows:

$$\check{D}_{k,\alpha_k(i)}^{\mathsf{QEDT}}(s) = \sigma_{Y(s)|u_k}^2 = \sigma_y^2\left(1-e^{-\theta s}\right)+\sigma_y^2 e^{-\theta s}\left(1+\mathsf{SNR}\gamma_k(i)\right)^{-\frac{W}{N\alpha_k(i)}}.$$

[5]Implicitly, we are assuming that the (k-1)-*th* sensor encodes at a constant rate over timeslots. This will be verified later on in this section.

[6]Still, the FC continues to use both u_k and u_{k-1} to reconstruct the random field. Yet suboptimal, this solution still outperforms those obtained in delay-constrained scenarios (see computer simulations section).

Clearly, only the second term in the summation of the cost function $\check{D}^{\mathsf{QEDT}}_{k,\alpha_k(i)}(s)$ is relevant to the optimization problem, which can be re-written as

$$\begin{aligned} \min_{\alpha_k(1),\ldots,\alpha_k(L)} \quad & \frac{1}{L}\sum_{i=1}^{L}\alpha_k(i)\left(1+\mathsf{SNR}\gamma_k(i)\right)^{-\frac{W}{N\alpha_k(i)}} \\ \text{s.t.} \quad & \frac{1}{L}\sum_{i=1}^{L}\alpha_k(i)=1. \end{aligned} \tag{5.31}$$

It is straightforward to show that this is a convex problem. Hence, one can construct the Lagrangian function as follows:

$$\begin{aligned} \mathcal{L}\left(\lambda,\alpha_k(1),\ldots,\alpha_k(L)\right) &= \frac{1}{L}\sum_{i=1}^{L}\alpha_k(i)\left(1+\mathsf{SNR}\gamma_k(i)\right)^{-\frac{W}{N\alpha_k(i)}} \\ &+ \lambda\left(\frac{1}{L}\sum_{i=1}^{L}\alpha_k(i)-1\right) \end{aligned} \tag{5.32}$$

where λ is the Lagrange multiplier. Therefore, by setting the first derivative of (5.32) w.r.t. $\alpha_k(i)$ to zero we obtain

$$\alpha_k^*(i)=\frac{W}{N}\frac{\ln\left(1+\mathsf{SNR}\gamma_k(i)\right)}{1-\mathsf{W}_{-1}\left(\frac{\lambda}{e}\right)} \tag{5.33}$$

with $\mathsf{W}_{-1}(\cdot)$ denoting the negative real branch of the Lambert function [71]. Apparently, the future channel gains $(\gamma_k(i+1),\ldots,\gamma_k(L))$ would also be needed in order to compute λ^*. However, as $L\to\infty$ this non-casuality requirement vanishes: by the law of large numbers, we have that

$$\lim_{L\to\infty}\frac{1}{L}\sum_{i=1}^{L}\alpha_k^*(i) = \frac{W}{N}\frac{\mathbb{E}_\gamma\left[\ln\left(1+\mathsf{SNR}\gamma\right)\right]}{1-\mathsf{W}_{-1}\left(\frac{\lambda}{e}\right)} \tag{5.34}$$

where γ a exponential distributed random variable. Hence, λ^* can be readily obtained by replacing into the constraint of (5.31), namely

$$\lambda^*=-\sigma_y^2\left(\frac{W}{N}\overline{R}\ln(2)+1\right)e^{-\frac{W}{N}\overline{R}\ln(2)} \tag{5.35}$$

where we have defined

$$\overline{R}\triangleq\mathbb{E}_\gamma\left[\log_2\left(1+\mathsf{SNR}\gamma\right)\right]. \tag{5.36}$$

Finally, replacing λ^* into (5.33) yields

$$\alpha_k^*(i)=\frac{\log_2\left(1+\mathsf{SNR}\gamma_k(i)\right)}{\overline{R}} \quad ; \quad \begin{matrix} i=1,\ldots,L \\ k=1,\ldots,N \end{matrix}, \tag{5.37}$$

and, by substituting $\alpha_k^*(i)$ into (5.43), the quantization noise for the k-*th* sensor node reads:

$$\sigma_z^2=\sigma_{z_k}^2=\frac{\sigma_y^2}{2^{\frac{W}{N}\overline{R}}-1} \quad ; \quad \begin{matrix} i=1,\ldots,L \\ k=1,\ldots,N \end{matrix}. \tag{5.38}$$

5.4.1 Average distortion in the reconstructed random field

By inserting $\alpha_k^*(i)$ into the *original* cost function of (5.28), the distortion for an arbitrary point in the k-*th* network segment reads:

$$D_{k,\alpha_k(i)}^{\mathsf{QEDT}}(s) \;=\; D_k^{\mathsf{QEDT}}(s) = \left(\frac{1}{\sigma^2_{Y(s)|u_{k-1}}} + \frac{e^{-\theta\left(\frac{d}{N}-s\right)}\left(2^{\frac{m}{n}\overline{R}}-1\right)}{\left(2^{\frac{m}{n}\overline{R}}-1\right)\left(1-e^{-\theta\left(\frac{d}{N}-s\right)}\right)\sigma_y^2+\sigma_y^2} \right)^{-1} \quad (5.39)$$

Interestingly, distortion is not a function of the channel gain experienced by the k-*th* sensor in timeslot i (i.e. distortion does not depend on $\alpha_k^*(i)$). As a result and unlike in QEDC encoding, the distortion experienced in every timeslot $i = 1, \ldots, L$ is identical. This can be useful in applications where a constant distortion level is needed.

After some tedious manipulations, the average distortion in the *whole* reconstructed random field can be expressed as:

$$\overline{D}^{\mathsf{QEDT}} \;=\; \frac{1}{N-1}\sum_{k=1}^{N-1}\frac{N}{d}\int_0^{\frac{d}{N}} D_{k+1}^{\mathsf{QEDT}}(s)ds \quad (5.40)$$

$$\;=\; \frac{\left(\left(\sigma_y^2+\sigma_z^2\right)^2 e^{\frac{\theta d}{N}}+\sigma_y^4\right)\frac{\theta d}{N} - 2\sigma_y^4\left(\sigma_y^2+\sigma_z^2\right)\left(e^{\frac{\theta d}{N}}-1\right)}{\left(\left(\sigma_y^2+\sigma_z^2\right)^2 e^{\frac{\theta d}{N}}-\sigma_y^4\right)\frac{\theta d}{N}}. \quad (5.41)$$

5.4.2 Buffer stability considerations

In order to derive a closed-form solution of the optimal number of samples to be encoded in each time slot ($\alpha_k^*(i)$), in (5.34) we let the number of timeslots L grow to infinity. Clearly, this might lead to a situation were buffer occupancy grows without bound, that is, to buffer unstability. To avoid that, we will encode and transmit a (slightly) higher number of samples per timeslot, namely

$$\alpha_k'(i)n = \frac{\log_2\left(1+\mathsf{SNR}\gamma_k(i)\right)}{\overline{R}-\delta}n > \alpha_k^*(i)n \quad (5.42)$$

with $0 < \delta < \overline{R}$. By doing so, one can prove (see Appendix 5.A.1) that buffers are stable. Unsurprisingly, this come at the expense of an increased distortion in the estimates (see computer simulation results in Section 5.4.3).

5.4.3 Simulations and numerical results

Figure 5.3 depicts the (per-timeslot) distortion in the reconstructed random field for both the QEDC and QEDT encoding strategies and different SNR values. For the QEDC strategy, we show the average value along with the $\pm\sigma$ confidence interval (to recall that, unlike in the

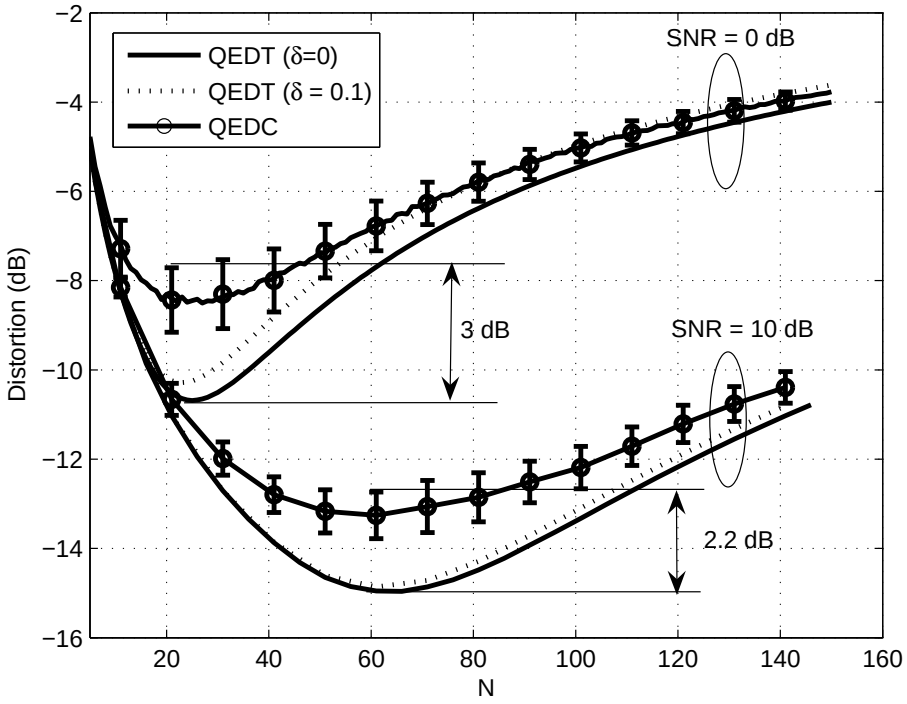


Figure 5.3: Average distortion vs. network size N ($W = 150$, $\theta d = 10$).

QEDT case, the distortion in QEDC encoding varies from timeslot to timeslot). Several conclusions can be drawn. First, for each curve there exists an optimal operating point, that is, a network size for which distortion can be minimized. The intuition behind this fact is that, despite that spatial variations of the random field are better captured by a denser grid of sensors, for a total bandwidth constraint the available rate per sensor progressively diminishes, this resulting in a more rougher quantization of the observations. Thus, the optimal trade-off between these two effects needs to be identified. Second, the distortion associated to delay-tolerant strategies is, as expected, lower than for the delay-constrained ones. Moreover, the lower the average SNR in the sensor-to-FC channels (namely, sensors with lower transmit power), the higher the gain (up to 3 dB for SNR=0 dB). Third, guaranteing buffer stability in the QEDT scheme only results into a marginal penalty in distortion, as shown in the curves labeled with $\delta = 0$ and $\delta = 0.1$. Complementarily, in Fig. 5.4, we depict buffer occupancy for several values of δ. For $\delta = 0$, the system is clearly unstable. Conversely, by letting δ take positive values, e.g. for $\delta = 0.1$ as in Fig.5.3, the average buffer occupancy can be kept under control (with a relatively small average buffer occupancy of $3n$ samples, in this case). Clearly, increasing δ has a two-fold effect: the average buffer occupancy diminishes but, simultaneously, the resulting distortion increases.

Finally, the rate at which the distortion decreases for the QEDC and QEDT schemes (evaluated at their respective optimal operating points) for an increasing SNR is shown in Figure 5.5.

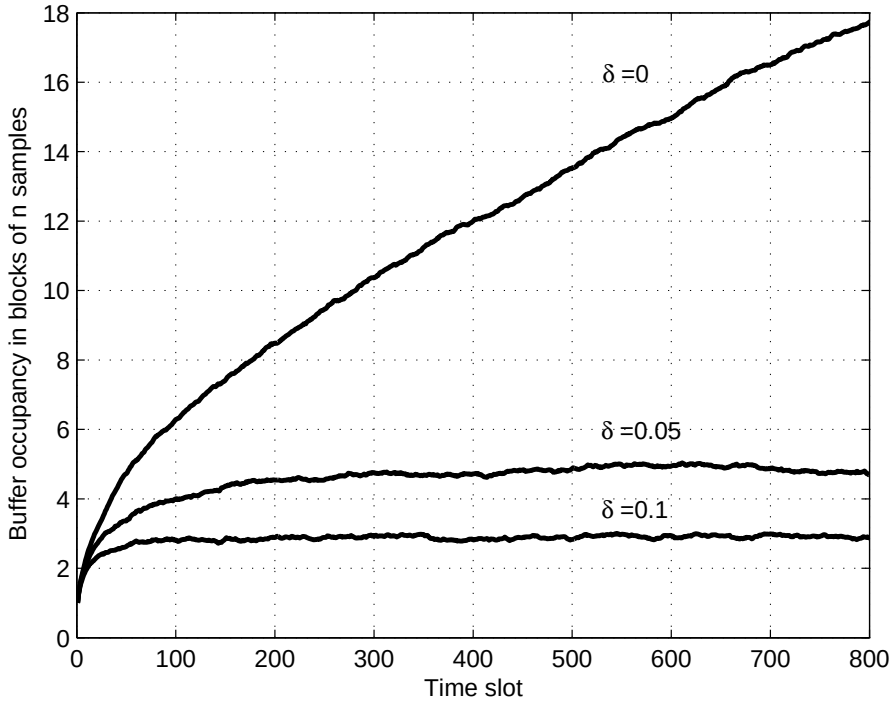


Figure 5.4: Average buffer size vs. timeslot ($\mathsf{SNR} = 0$ dB).

For intermediate distortion values, the gap is approximately 4 dB. That is, for a prescribed distortion level, the energy consumption in delay-constrained networks is 2.5 times higher.

5.5 Delay-tolerant WSNs with Compress-and-Estimate encoding

As in the previous section, let $n_k(i) = \alpha_k(i)n$ be the number of samples encoded in m/N channel uses (i.e. one timeslot). For reliable decoding at the FC, the rate at the output of the C&E encoder must satisfy:

$$\frac{m}{N}\log_2\left(1+\mathsf{SNR}\cdot\gamma_k(i)\right) \geq \alpha_k(i)n\log_2\left(1+\frac{\sigma^2_{y_k|u_{k-1}}}{\sigma^2_{z_k}}\right). \tag{5.43}$$

To stress that expression (5.43) differs from (5.27) in that the C&E encoder assumes that the FC will use u_{k-1} to *decode* u_k and, hence, $\sigma^2_{y_k}$ has been replaced by $\sigma^2_{y_k|u_{k-1}}$. Therefore, from (5.7) and the definition of $\sigma^2_{y_k|u_{k-1}}$ in (5.22), we have that for the current block of $\alpha_k(i)n$ samples the distortion reads

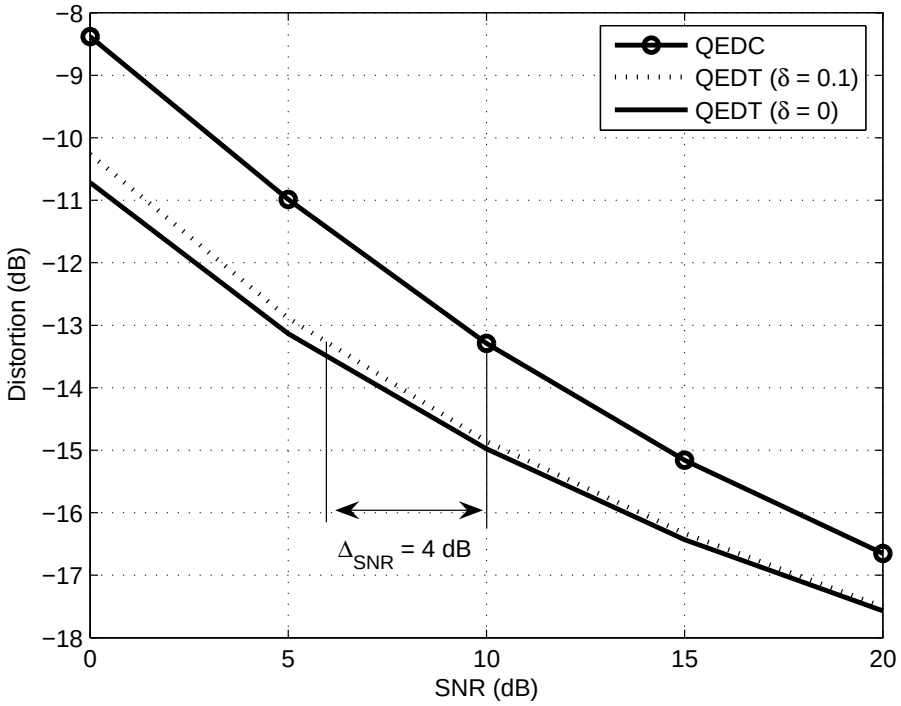


Figure 5.5: Average distortion vs. SNR ($W = 150$, $\theta d = 10$).

$$D^{\mathsf{CEDT}}_{k,\alpha_k(i)}(s) = \frac{\sigma_y^2 \sigma^2_{Y(s)|u_{k-1}} \left(e^{\theta\left(\frac{d}{N}-s\right)}-1\right)}{\sigma_y^2\left(e^{\theta\left(\frac{d}{N}-s\right)}-1\right)+\sigma^2_{Y(s)|u_{k-1}}} + \frac{\sigma^4_{Y(s)|u_{k-1}}\left(1+\mathsf{SNR}\gamma_k)(i)\right)^{-\frac{m}{\alpha_k(i)n}}}{\sigma_y^2\left(e^{\theta\left(\frac{d}{N}-s\right)}-1\right)+\sigma^2_{Y(s)|u_{k-1}}}. \quad (5.44)$$

and by averaging over L timeslots, the following problem results:

$$\min_{\alpha_k(1),\ldots,\alpha_k(L)} \quad \frac{1}{L}\sum_{i=1}^{L} \alpha_k(i) D^{\mathsf{CEDT}}_{k,\alpha_k(i)}(s) \quad (5.45)$$

$$\text{s.t.} \qquad \sum_{i=1}^{L} \alpha_k(i) n = Ln. \quad (5.46)$$

Solving this problem leads to a closed-form solution that is identical to that of the QEDT case, namely,

$$\alpha_k^*(i) = \frac{\log_2\left(1+\mathsf{SNR}\gamma_k(i)\right)}{\overline{R}}. \quad (5.47)$$

Finally, by replacing $\alpha_k^*(i)$ into (5.43) yields

$$\sigma^2_{z_k} = \frac{\sigma^2_{y_k|u_{k-1}}}{2^{\frac{W}{N}\overline{R}}-1} \quad ; \quad \begin{matrix} i = 1,\ldots,L \\ k = 1,\ldots,N \end{matrix}. \quad (5.48)$$

As in the QEDT case, this last expression reveals that all sensors encode their observations at a constant rate. This was implicitly assumed in the score function (5.46). To remark, the stability analysis of Section 5.4.2 also applies here.

5.5.1 Average distortion in the reconstructed random field

By inserting $\alpha_k^*(i)$ into the original cost function of (5.46), the distortion for an arbitrary point in the k-*th* segment reads

$$D_{k,\alpha_k(i)}^{\mathsf{CEDT}}(s) = \frac{\sigma_y^2 \sigma_{Y(s)|u_{k-1}}^2 \left(e^{\theta\left(\frac{d}{N}-s\right)}-1\right)}{\sigma_y^2\left(e^{\theta\left(\frac{d}{N}-s\right)}-1\right)+\sigma_{Y(s)|u_{k-1}}^2} + \frac{\sigma_{Y(s)|u_{k-1}}^4 2^{-\frac{W}{N}\overline{R}}}{\sigma_y^2\left(e^{\theta\left(\frac{d}{N}-s\right)}-1\right)+\sigma_{Y(s)|u_{k-1}}^2} \quad (5.49)$$

As in the QEDT case, distortion is not a function of the channel gain experienced by the k-*th* sensor in timeslot i. Hence, the distortion experienced in every timeslot $i = 1, \ldots, L$ is identical. Therefore, the average distortion for each network segment can be computed as follows:

$$\overline{D}_k^{\mathsf{CEDT}} = \frac{N}{d}\int_0^{\frac{d}{N}} D_k^{\mathsf{CEDT}}(s) \quad (5.50)$$

$$= \frac{\left(\left(\sigma_y^2+\sigma_{z_{k-1}}^2\right)\left(\sigma_y^2+\sigma_{z_k}^2\right)e^{\frac{\theta d}{N}}+\sigma_y^4\right)\frac{\theta d}{N} - 2\sigma_y^4\left(2\sigma_y^2+\sigma_{z_{k-1}}^2\sigma_{z_k}^2\right)\left(e^{\frac{\theta d}{N}}-1\right)}{\left(\left(\sigma_y^2+\sigma_{z_{k-1}}^2\right)\left(\sigma_y^2+\sigma_{z_k}^2\right)e^{\frac{\theta d}{N}}-\sigma_y^4\right)\frac{\theta d}{N}} \quad (5.51)$$

Finally, the average distortion in the *whole* reconstructed random field can be expressed as:

$$\overline{D}^{\mathsf{CEDT}} = \frac{1}{N-1}\sum_{k=1}^{N-1}\overline{D}_{k+1}^{\mathsf{CEDT}}. \quad (5.52)$$

Interestingly, the average distortion has a simple closed-form expression, this being in a stark contrast with the CEDC strategy where, in general, a closed-form expression for the average distortion (over different channel realizations) cannot be found.

5.5.2 Simulations and numerical results

Figure 5.6 illustrates the average distortion in the reconstructed random field for the CEDC and CEDT encoding strategies. As in quantize-and-estimate encoding, there exists an optimal number of sensors nodes. Finding such N^* is particularly useful for random fields with low SNR per sensor, since the curve is sharper in this case. The gap between the minimum distortion attainable by the CEDC and CEDT schemes (which results from an adequate exploitation of channel fluctuation in the delay-tolerant approach) is approximately 2-3 dB. Concerning buffer occupancy-distortion trade-offs, the same comments as in the quantize-and-estimate case apply.

Finally, in Fig. 5.7, we compare the distortion attained by QEDT/CEDT encoding strategies for random fields with low and high spatial variabilities ($\theta d = 1$, $\theta d = 10$, respectively). Due to the fact that CEDT is capable of exploiting spatial correlation, it always outperforms QEDT. Moreover, the higher the spatial correlation ($\theta d = 1$), the larger the gap between the curves.

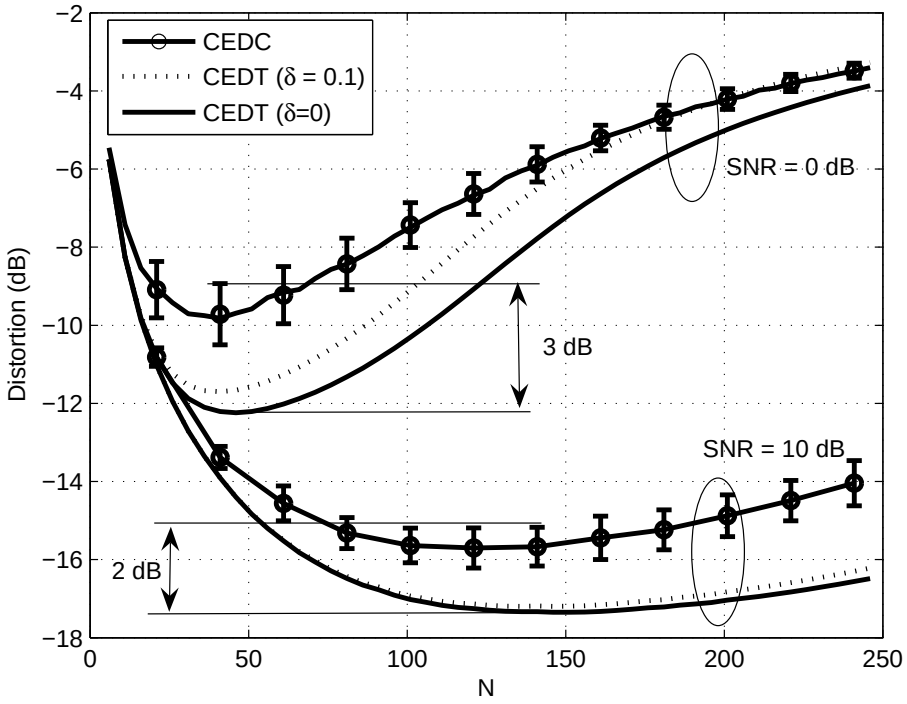


Figure 5.6: Average distortion vs. network size N ($W = 150$, $\theta d = 10$).

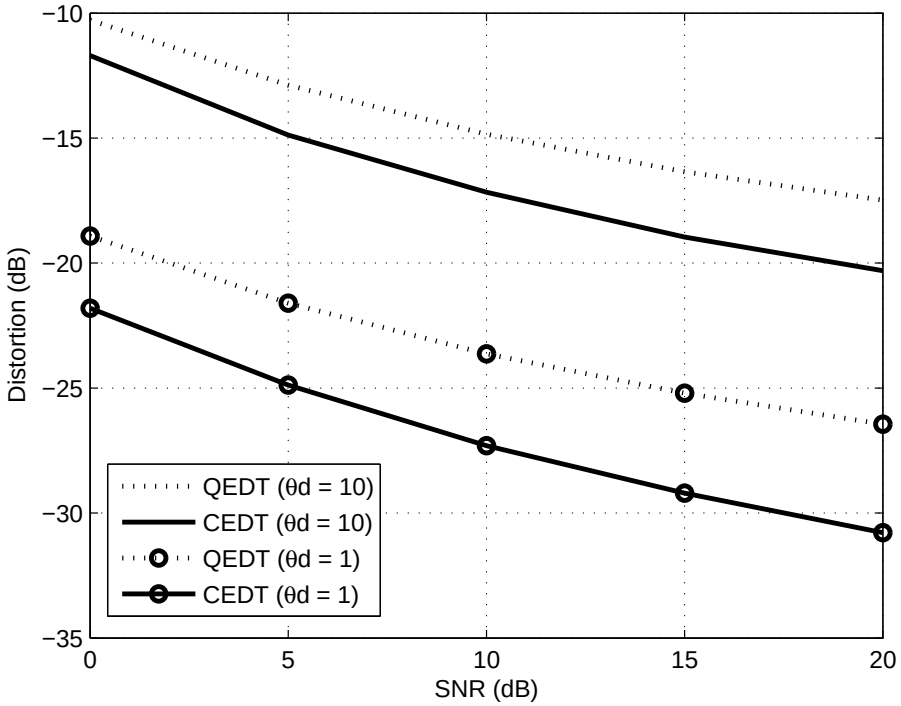


Figure 5.7: Distortion vs. SNR ($W = 150$).

5.6 Latency analysis

As discussed in the previous sections, in delay-tolerant networks the number of samples encoded in each timeslot is not constant. Unavoidably, this introduces some delay in the reconstruction of the random field for each block of n consecutive samples (whereas in the case of DC scenarios, the n consecutive realizations of the random field can be immediately reconstructed).

In the sequel, we analytically assess the latency at the FC for reconstructing the entire random field. To that end, first we propose a model, which accounts for the latency in receiving n consecutive samples from *one* particular sensor node. Next, we derive the latency of the QEDT and CEDT encoding strategies, respectively.

5.6.1 Latency analysis for a single sensor node

Let $n_k^*(i) = \lfloor \alpha_k^*(i) n \rfloor$ the number of samples encoded in $\frac{m}{N}$ channel uses in timeslot i. The probability that $l = 0, \ldots, n-1$ samples are encoded in an arbitrary timeslot i can be expressed as

$$p_l = \Pr\left(n_k^*(i) = l\right) \tag{5.53}$$

$$= \Pr\left(\frac{l}{n} \leq \alpha_k^*(i) < \frac{l+1}{n}\right) \quad ; \quad l = 0, \ldots, n-1. \tag{5.54}$$

Besides, we define

$$p_n = \Pr\left(n_k^*(i) \geq n\right) \tag{5.55}$$

$$= \Pr\left(\alpha_k^*(i) \geq 1\right). \tag{5.56}$$

On that basis, we model our system as an absorbing Markov chain [101, Chapter 8] with n *transient* states ($\mathcal{S}_1, \ldots, \mathcal{S}_{n-1}$) and one *absorbing* state ($\mathcal{S}_n$) defined as follows:

$$\mathcal{S}_l = \begin{cases} l \text{ samples have been transmitted in previous timeslots} & ; \quad l = 0, \ldots, n-1 \\ n \text{ or more samples have been transmitted in previous timeslots} & ; \quad l = n \end{cases}. \tag{5.57}$$

The transition matrix $\mathbf{P}$ of an absorbing Markov chain has the following canonical form [101, Chapter 8]:

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{r} \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{5.58}$$

where $\mathbf{Q}$ denotes the $(n+1) \times (n+1)$ transient matrix, $\mathbf{r}$ is a $(n+1) \times 1$ non-zero vector (otherwise the absorbing state could never be reached from the transient states). The entries of the matrix $\mathbf{Q}$ can be computed as follows:

$$q_{l,j} = \begin{cases} 0 & j < l \\ p_{j-l} & \text{otherwise} \end{cases}, \tag{5.59}$$

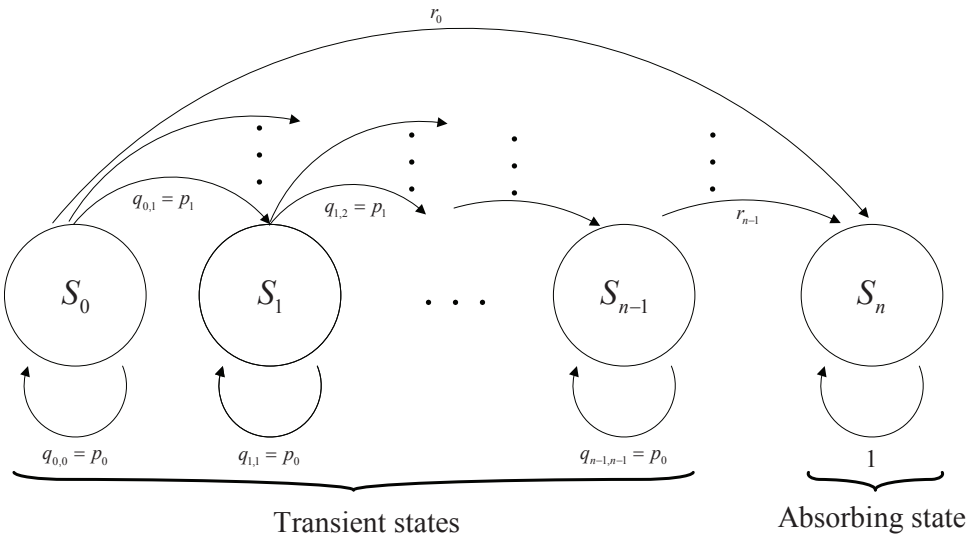


Figure 5.8: An absorbing Markov chain.

and, the entries of the $(n+1) \times 1$ $\mathbf{r}$ vector, which denote the probability of absorbtion from each transient states, are given by

$$r_l = 1 - \sum_{j=0}^{n-1} q_{l,j} \quad ; \quad l = 0, \ldots, n-1. \tag{5.60}$$

Our goal is to characterize the time elapsed until the absorbing state is reached or, in other words, the time needed to transmit n consecutive samples of the local observation of the random field at sensor k (i.e. the latency). For an absorbing Markov chain defined as in (5.58), the random variable τ, standing for the time to absorbtion, obeys the so-called Discrete PHase-type (DPH) distribution. From [102], the probability mass and cumulative distribution functions can be expressed as:

$$f_\tau(t) = \Pr(\tau = t) = \boldsymbol{\pi}^T \mathbf{Q}^{t-1} \mathbf{r} \quad ; \quad t = 1, \ldots, \infty \tag{5.61}$$

$$F_\tau(t) = \Pr(\tau \leq t) = 1 - \boldsymbol{\pi}^T \mathbf{Q}^t \mathbf{1} \quad ; \quad t = 1, \ldots, \infty \tag{5.62}$$

where in the expressions above the $(n+1) \times 1$ vector $\boldsymbol{\pi}$ denotes the initial conditions. Since we assume that in the beginning no samples have been transmitted, it yields

$$\boldsymbol{\pi}^T = [1, 0, \ldots, 0]^T.$$

The average time to absorbtion reads:

$$\mathbb{E}[\tau] = \sum_{t=1}^{\infty} t f_\tau(t). \tag{5.63}$$

Alternatively, from [101, Chapter 8], one can compute

$$\mathbf{u} = (\mathbf{I} - \mathbf{Q_1})^{-1} \mathbf{1} \tag{5.64}$$

with the first element of $\mathbf{u}$ denoting the average time to absorbtion, i.e.

$$\mathbb{E}[\tau] = \mathbf{u}(1).$$

Finally, the only missing point is to find a closed-form expression for the set of probabilities $\{p_0, p_1, \ldots, p_n\}$ defined in (5.54) and (5.56). From (5.37), we have that

$$\alpha_k^*(i) = \frac{\log_2(1 + \mathsf{SNR}\gamma_k(i))}{\overline{R}}. \tag{5.65}$$

with $\overline{R} = \mathbb{E}_\gamma[\log_2(1 + \gamma\mathsf{SNR})]$ and, hence,

$$\begin{aligned} p_l &= \Pr\left(\frac{l}{n} \leq \alpha_k^*(i) < \frac{l+1}{n}\right) && (5.66)\\ &= \Pr\left(\frac{l}{n}\overline{R} \leq \log_2(1 + \mathsf{SNR}\gamma_k(i)) < \frac{l+1}{n}\overline{R}\right) && (5.67)\\ &= F_\gamma\left(\frac{2^{\frac{l+1}{n}\overline{R}} - 1}{\mathsf{SNR}}\right) - F_\gamma\left(\frac{2^{\frac{l}{n}\overline{R}} - 1}{\mathsf{SNR}}\right) && (5.68) \end{aligned}$$

for $l = 0, \ldots, n-1$ and $p_n = 1 - F_\gamma\left(\frac{2^{\overline{R}}-1}{\mathsf{SNR}}\right)$. For Rayleigh-fading channels, the CDF of the channel gain is given by $F_\gamma(x) = 1 - e^{-x}$.

5.6.2 Latency analysis for QEDT encoding

At this point, the interest lies in characterizing the time elapsed until N sensors in the network encode and transmit their corresponding n first samples of the random field. That is, we attempt to characterize the latency in reconstructing the first n consecutive realizations of the *whole* random field. Let Ψ be a random variable which models such latency, namely

$$\Psi = \max_{k=1,\ldots,N} \tau_k, \tag{5.69}$$

where τ_k stands for the latency associated to the individual sensor k as defined in the previous section. Since, on the one hand, sensors are assumed to experience i.i.d fading channels to the FC and, on the other, the FC decodes the samples received from each sensor independently, $\tau_1, \ldots, \tau_N$ turn out to be i.i.d. DPH random variables with marginal pmf's and CDFs given by (5.61) and (5.62), respectively. From all the above, the CDF of the latency associated to QEDT encoding reads

$$\begin{aligned} F_\Psi(t) &= \Pr(\Psi \leq t) = \Pr\left(\max_k \tau_k \leq t\right) && (5.70)\\ &= \Pr(\tau_1 \leq t, \tau_2 \leq t, \ldots, \tau_N \leq t) && (5.71)\\ &= F_\tau^N(t) = \left(1 - \boldsymbol{\pi}^T\mathbf{Q}^t\mathbf{1}\right)^N \quad t = 1, \ldots, \infty && (5.72) \end{aligned}$$

From (5.72), it follows that the probability mass function reads

$$\begin{aligned} f_{\Psi}(t) &= \Pr(\Psi = t) && (5.73)\\ &= F_{\Psi}(t) - F_{\Psi}(t-1) && (5.74)\\ &= \left(1 - \boldsymbol{\pi}^T \mathbf{Q}^t \mathbf{1}\right)^N - \left(1 - \boldsymbol{\pi}^T \mathbf{Q}^{t-1} \mathbf{1}\right)^N \quad t = 1, \ldots, \infty. \end{aligned}$$

Finally, the average latency yields:

$$\mathbb{E}[\Psi] = \sum_{t=1}^{\infty} t f_{\Psi}(t). \tag{5.75}$$

It is worth noting that the average latency of the system will increase with the number of sensor nodes N, since for a large network size the higher the time to absorbtion for the slowest sensor node.

5.6.3 Latency analysis for CEDT encoding

Due to the successive encoding of the observations that C&E strategies entail, the latency analysis here is far more involved and, in general, does not allow for the derivation of closed-form expressions. To circumvent that, we will resort to an approximate (yet accurate) approximation.

In order for the FC to successfully decode the codeword received from sensor k, the codeword sent by the adjacent sensor $k-1$ must have been decoded first. Clearly, this means that the codeword received from sensor N will be the last one to be decoded. Due to the fact that sensors experience i.i.d. fading conditions (and, thus, the number of observations received from different sensors are not time-aligned), when the first n samples sent by sensor N are ready to be decoded, a total of $n + c_o n$ (instead of only n) samples from sensor $N-1$ have already been decoded on average. Accordingly, a total of $n + (N-1)\, c_o n$ samples from sensor #1 have already been decoded too (see Fig. 5.9). Hence, the first n realizations of the *entire* random field can be reconstructed if, equivalently, $n + (N-1)\, c_o n$ samples sent by the first sensor have already been decoded by the FC. Since the encoding process for the first sensor is identical in C&E and Q&E encoding, in order to compute the latency for the reconstruction of the *random field*, it suffices to compute the time to absorbtion for an *individual* sensor (sensor #1). The only change with respect to the model given in (5.58) is that the Markov chain has now a total of $n + (N-1)\, c_o n$ states (instead of n) and, hence, the size and elements of matrix $\mathbf{Q}$ and vectors π and $\mathbf{r}$ in (5.61) and (5.62) must be adjusted accordingly.

As for parameter c_o (which exclusively depends on the pdf of the sensor-to-FC channel gains), it can only be determined empirically (see next Section).

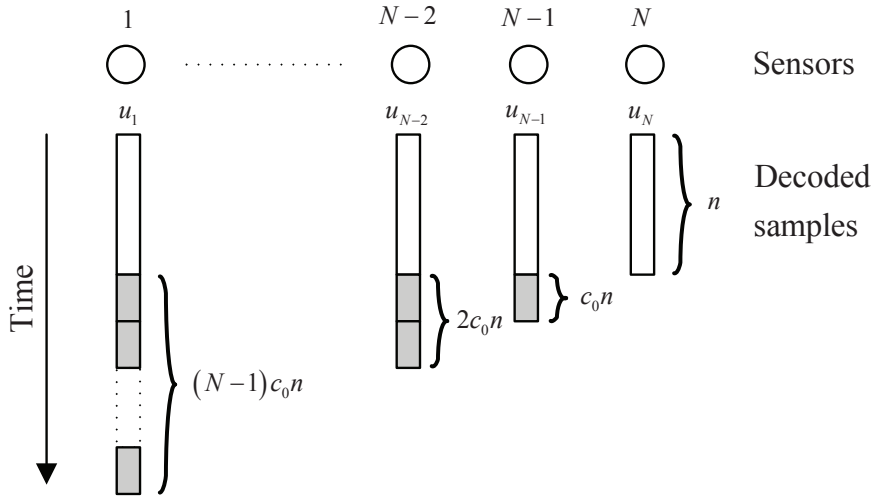


Figure 5.9: Approximate CEDT decoding for the analysis of the latency.

5.6.4 Simulations and numerical results

In Figures 5.10 and 5.11 we depict the average latency for the QEDT and CEDT strategies, respectively. Interestingly, there exists a trade-off in terms of the attainable distortion vs. latency. Whereas in CEDT encoding latency exhibits a *linear* increase in the number of sensors, in QEDT encoding latency grows *logarithmically* (i.e. more slowly). However, CEDT schemes attain lower distortions than QEDT ones. Besides, in Fig. 5.10 it is also worth noting the perfect match between the simulations and the numerical results and, unsurprisingly, that the higher the average SNR, the lower the latency. Also, Figure 5.11 reveals that by using an appropriate value of c_o (i.e. $c_o = 0.6$), the latency associated to the approximate model described in Section 5.6.3 matches the actual one.

5.7 Random field estimation without transmit CSI

In the absence of (instantaneous) transmit CSI, neither can the encoding rate be dynamically adjusted to channel conditions, nor is the successive C&E encoding strategy applicable[7]. Hence, we focus on the Q&E scheme and propose a modification by which each sensor observation is *independently* encoded at a *common* and *constant* rate given by

$$R = R_k = \log_2\left(1 + \frac{\sigma_y^2}{\sigma_z^2}\right) \quad ; \forall k. \tag{5.76}$$

The outage probability, namely, the probability that the encoding rate R exceeds the instantaneous channel rate (and, hence, the codeword cannot be successfully decoded at the FC), can

[7] See discussion in Section 4.6.

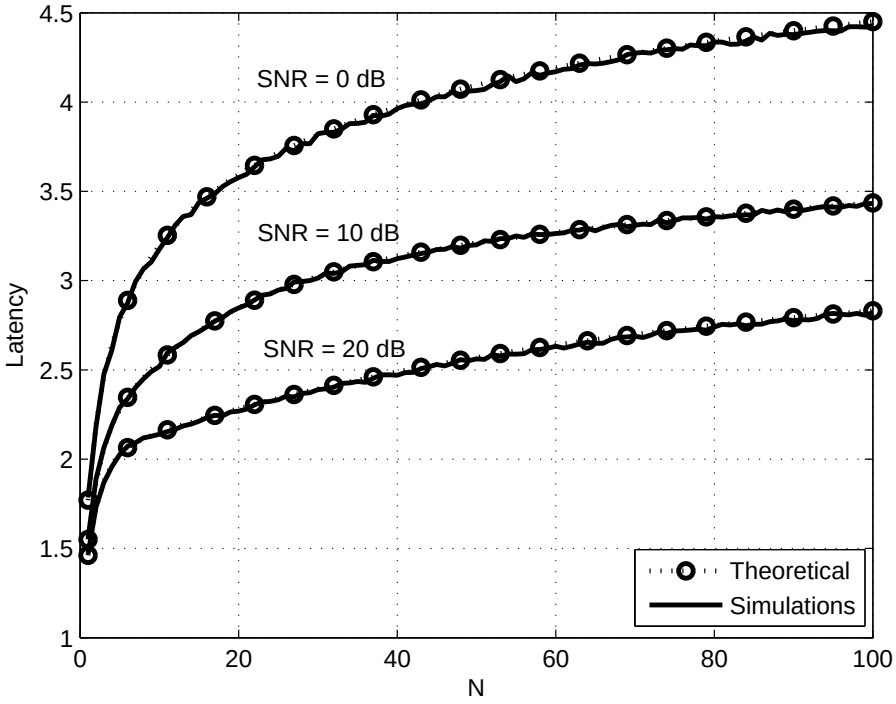


Figure 5.10: Average latency vs. network size N.

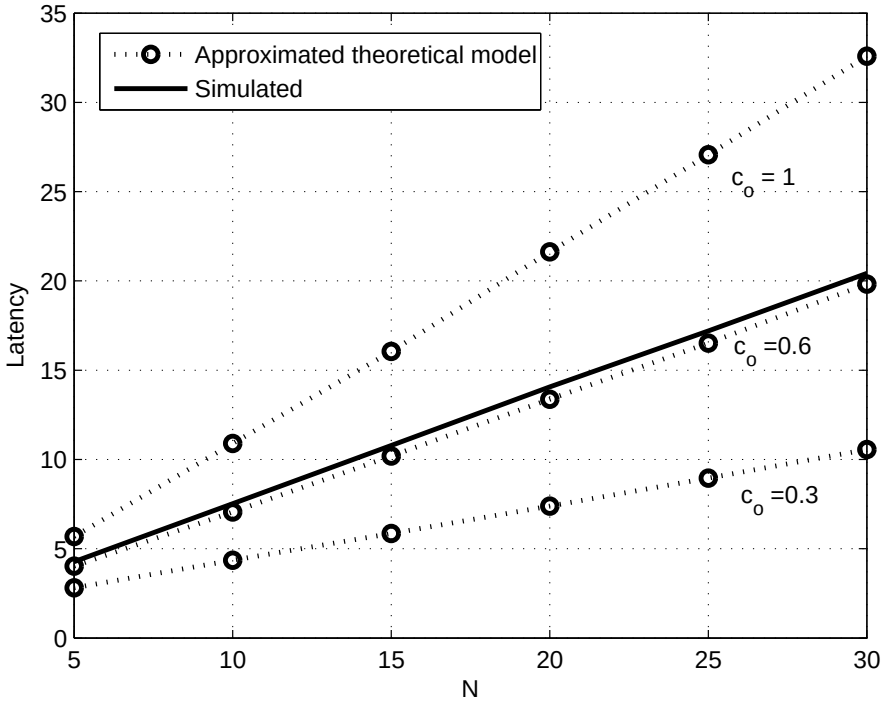


Figure 5.11: Average latency vs. network size (SNR $= 10$ dB).

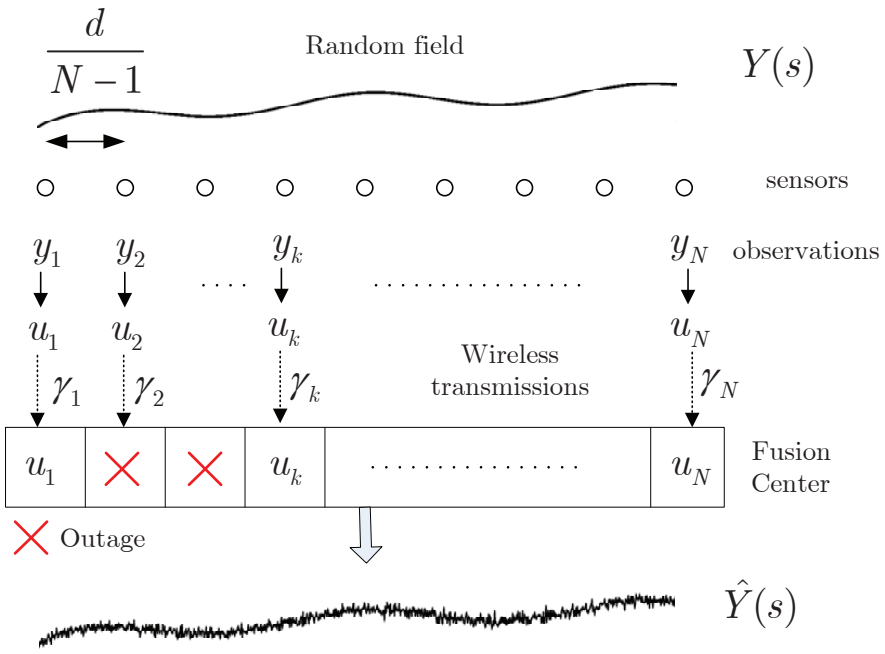


Figure 5.12: System model.

be expressed as

$$\begin{aligned} p_{\mathsf{out}} &= \Pr\left(\log_2\left(1+\frac{\sigma_y^2}{\sigma_z^2}\right) > \frac{W}{N}\log_2(1+\mathsf{SNR}\cdot\gamma_k)\right) \\ &= F_\gamma\left(\frac{1}{\mathsf{SNR}}\left(1+\frac{\sigma_y^2}{\sigma_z^2}\right)^{\frac{N}{W}} - \frac{1}{\mathsf{SNR}}\right) \end{aligned} \tag{5.77}$$

where $F_\gamma(x)$ stands for the CDF of the channel gains γ_k. From (5.77), the variance of the quantization noise reads

$$\sigma_z^2(N, p_{\mathsf{out}}) = \frac{\sigma_y^2}{\left(1+\mathsf{SNR}\cdot F_\gamma^{-1}\left(p_{\mathsf{out}}\right)\right)^{\frac{W}{N}} - 1}, \tag{5.78}$$

where $F_\gamma^{-1}\left(p_{\mathsf{out}}\right) = \log\left(\frac{1}{1-p_{\mathsf{out}}}\right)$. Interestingly, the actual (spatial) sampling pattern is defined by two system parameters: *i*) the number of sensors (N) and, *ii*) the outage probability experienced in the sensor-to-FC links (p_{out}). The former determines the inter-sensor distance, whereas the latter introduces a random sampling effect which modifies the spatial sampling pattern (see Fig. 5.12).

As far as the distortion analysis is concerned, it is similar to that of Section 5.2.2. For each network segment, the random field is estimated based on the two closest *successfully-decoded* codewords, namely u_k and u_{k+N_h}, that is,

$$\hat{Y}(s) = \mathbb{E}\left[Y(s)|u_k, u_{k+N_h}\right]; \quad \forall s \in \left[k\frac{d}{N}, (k+N_h)\frac{d}{N}\right] \tag{5.79}$$

where N_h is a geometrically-distributed random variable with probability of success given by $p = 1 - p_{\mathsf{out}}$. Accordingly, the distortion associated to the estimator of (5.79) is given by

$$D_{N_h}(s) = \sigma^2_{Y(s)|u_k,u_{k+N_h}} = \sigma^2_{Y(s)|u_k} - \frac{\mathrm{Cov}^2\left(Y(s)|u_k, u_{k+N_h}\right)}{\sigma^2_{u_{k+N_h}|u_k}} \tag{5.80}$$

where

$$\sigma^2_{Y(s)|u_k} = \left(\frac{1}{\sigma_y^2} + \frac{e^{-\theta s}}{\left(1-e^{-\theta s}\right)\sigma_y^2 + \sigma_z^2}\right)^{-1}, \tag{5.81}$$

$$\mathrm{Cov}^2\left(Y(s), u_{k+N_h}|u_k\right) = e^{-\theta\left(N_h\frac{d}{N}-s\right)}\sigma^4_{Y(s)|u_k}, \tag{5.82}$$

$$\begin{aligned}\sigma^2_{u_{k+N_h}|u_k} &= e^{-\theta\left(N_h\frac{d}{N}-s\right)}\sigma^2_{Y(s)|u_k} \\ &+ \left(1-e^{-\theta\left(N_h\frac{d}{N}-s\right)}\right)\sigma_y^2 + \sigma_z^2.\end{aligned} \tag{5.83}$$

5.7.1 Optimization problem: optimal network design and encoding rate

Here, we are interested in a min-max design by which we attempt to minimize the maximum value of distortion in each segment. For each network segment, namely $s \in \left[0, N_h\frac{d}{N}\right]$, the highest distortion is attained at the middle point (see [24]), i.e. $s = \frac{1}{2}N_h\frac{d}{N}$, and reads

$$D_{N_h}(s) \leq \max_{s\in\left[0,N_h\frac{d}{N}\right]} D_{N_h}(s) = D^{\mathsf{max}}_{N_h} = \sigma_y^2\frac{1-e^{-\theta N_h\frac{d}{2N}}+\sigma_z^2}{1+e^{-\theta N_h\frac{d}{2N}}+\sigma_z^2}.$$

For networks with a sufficiently-high number of sensors and by neglecting border effects, this yields an average value of

$$D^{\mathsf{max}} = D^{\mathsf{max}}(N, p_{\mathsf{out}}) = \sigma_y^2\mathbb{E}_{N_h}\left[\frac{1-e^{-\theta N_h\frac{d}{2N}}+\sigma_z^2}{1+e^{-\theta N_h\frac{d}{2N}}+\sigma_z^2}\right] \tag{5.84}$$

$$\leq D^{\mathsf{ub}}(N, p_{\mathsf{out}}) = \sigma_y^2\frac{1-e^{-\theta\frac{d}{2N(1-p_{\mathsf{out}})}}+\sigma_z^2}{1+e^{-\theta\frac{d}{2N(1-p_{\mathsf{out}})}}+\sigma_z^2}, \tag{5.85}$$

where again the upper bound follows from the application of Jensen's inequality.

Our goal is, thus, to find the optimal number of sensor nodes N in the network and the corresponding encoding rate R which minimizes the cost function (5.85). From (5.76) and (5.78), R is, in turn, a function of N and p_{out} and, therefore, we equivalently minimize with respect to these two variables. This optimization problem can be solved in two steps [69]:

$$\min_{N,p_{\mathsf{out}}} D^{\mathsf{ub}}(N, p_{\mathsf{out}}) = \min_{p_{\mathsf{out}}}\left(\min_N D^{\mathsf{ub}}(N, p_{\mathsf{out}})\right) \tag{5.86}$$

$$= \min_{p_{\mathsf{out}}} \tilde{D}^{\mathsf{ub}}(p_{\mathsf{out}}). \tag{5.87}$$

Hence, in the next subsections, we first solve for the optimal $N(p_{\mathsf{out}})$ in the inner minimization problem of (5.86), and then we solve (5.87) for the optimal p_{out}.

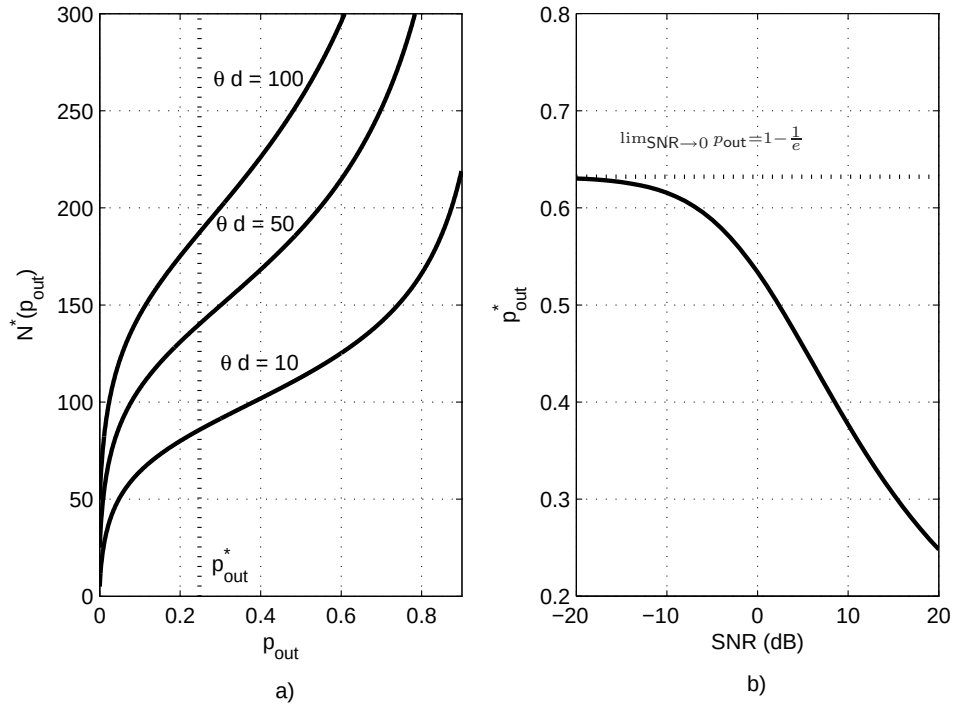


Figure 5.13: Optimal network size vs. outage probability for $W = 100$ and $\mathsf{SNR} = 20\mathrm{dB}$ (left) and optimal outage probability vs. SNR (right).

5.7.2 Optimal network size for arbitrary outage probability

According to (5.86), we have

$$N^*(p_{\mathsf{out}}) = \arg\min_N D^{\mathsf{ub}}(N, p_{\mathsf{out}}). \tag{5.88}$$

By relaxing $N \in \mathbb{R}^+$, the objective function becomes convex in N and a closed-form expression of N can be readily obtained from its first derivative:

$$N^*(p_{\mathsf{out}}) = \frac{WR(p_{\mathsf{out}})}{(1-p_{\mathsf{out}})\log_2\left(1+\frac{2WR(p_{\mathsf{out}})}{\theta d}\right)}, \tag{5.89}$$

with $R(p_{\mathsf{out}})$ given by

$$R(p_{\mathsf{out}}) = (1-p_{\mathsf{out}})\log_2(1-\mathsf{SNR}\ln(1-p_{\mathsf{out}})). \tag{5.90}$$

As shown in Fig. 5.13a, $N^*(p_{\mathsf{out}})$ is a monotonically increasing function in p_{out}. For high encoding rates, we have that $p_{\mathsf{out}} \to 1$ and, hence, the number of sensors must be increased to partly compensate for such lost codewords. Conversely, if sensors encode at low data rates then $p_{\mathsf{out}} \to 0$, a more uniform sampling of the random field results and then a minimum of $N^* \to \frac{1}{2}\theta d$ sensors suffices.

5.7.3 Optimal outage probability, network size and encoding rate

Now, replacing the optimal number of sensor nodes of (5.89) into (5.86) yields:

$$\tilde{D}_{\mathsf{ub}}\left(p_{\mathsf{out}}\right)=\sigma_y^2\frac{1-\left(1+CR\left(p_{\mathsf{out}}\right)\right)^{-\frac{1}{CR(p_{\mathsf{out}})}}+\frac{1}{CR(p_{\mathsf{out}})}}{1+\left(1+CR\left(p_{\mathsf{out}}\right)\right)^{-\frac{1}{CR(p_{\mathsf{out}})}}+\frac{1}{CR(p_{\mathsf{out}})}} \tag{5.91}$$

with $C=\frac{2\ln(2)}{\theta d}$ and $R\left(p_{\mathsf{out}}\right)$ given by (5.90). On the other, the objective function of (5.91) can be shown to be monotonically decreasing in $R\left(p_{\mathsf{out}}\right)$ and, hence, the original optimization problem of (5.87) is equivalent to

$$\min_{p_{\mathsf{out}}}\tilde{D}_{\mathsf{ub}}\equiv\max_{p_{\mathsf{out}}}R\left(p_{\mathsf{out}}\right). \tag{5.92}$$

The cost function $R\left(p_{\mathsf{out}}\right)$ is concave in p_{out} and the optimal value of p_{out} can be readily shown to be

$$p_{\mathsf{out}}^{*}=1-e^{\frac{1}{\mathsf{SNR}}-\frac{1}{\mathsf{W}_0(\mathsf{SNR})}} \tag{5.93}$$

where $\mathsf{W}_0\left(\cdot\right)$ stands for the positive real branch of the Lambert function [71]. As depicted in Fig. 5.13b, the optimal outage probability is a monotonically decreasing function in the SNR and, from (5.93), it converges to a constant value in the low SNR regime, namely

$$\lim_{\mathsf{SNR}\to 0}p_{\mathsf{out}}=\frac{1}{e}. \tag{5.94}$$

From (5.89) and (5.93), the optimal network size is given by

$$N^{*}=\frac{W\log_2\left(\frac{\mathsf{SNR}}{\mathsf{W}_0(\mathsf{SNR})}\right)}{\log_2\left(1+\frac{2W}{\theta d}e^{\frac{1}{\mathsf{SNR}}-\frac{1}{\mathsf{W}_0(\mathsf{SNR})}}\ln\left(\frac{\mathsf{SNR}}{\mathsf{W}_0(\mathsf{SNR})}\right)\right)} \tag{5.95}$$

and, finally, from (5.76) and (5.78) the optimal encoding rate results:

$$R^{*}=\log_2\left(1+\frac{2W}{\theta d}e^{\frac{1}{\mathsf{SNR}}-\frac{1}{\mathsf{W}_0(\mathsf{SNR})}}\ln\left(\frac{\mathsf{SNR}}{\mathsf{W}_0\left(\mathsf{SNR}\right)}\right)\right). \tag{5.96}$$

5.7.4 Simulations and numerical results

In Fig. 5.14, we plot the optimal number of nodes N^{*} and the optimal encoding rate R^{*} given in (5.95) and (5.96), respectively. First, we can observe that both parameters are increasing functions in the overall $\mathsf{SNR}=\frac{P}{W\mathsf{N}_o}$: the higher the available transmit power, the higher the number of nodes that can be accommodated and the higher the achievable rates. Second, for random fields with low spatial variability, i.e. $\theta d=10$, the optimal encoding rate is higher and the network size lower, this meaning that it is better to sparsely sense the field with high resolution observations. On the contrary, for random fields with high spatial variability, this

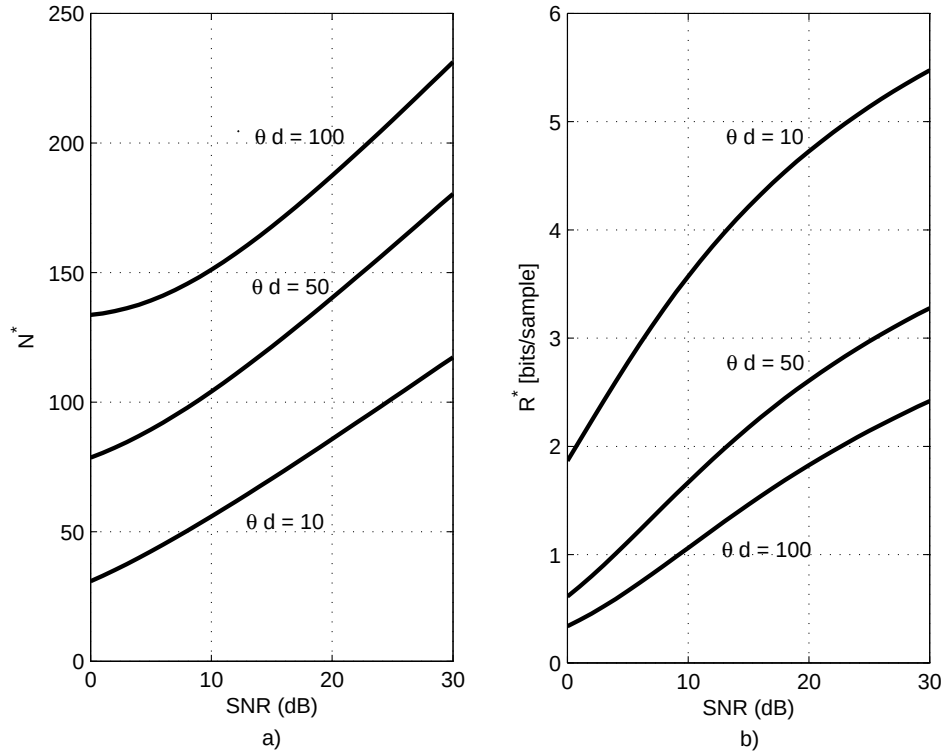


Figure 5.14: Optimal network size (left) and encoding rate (right) vs. SNR. (W=100).

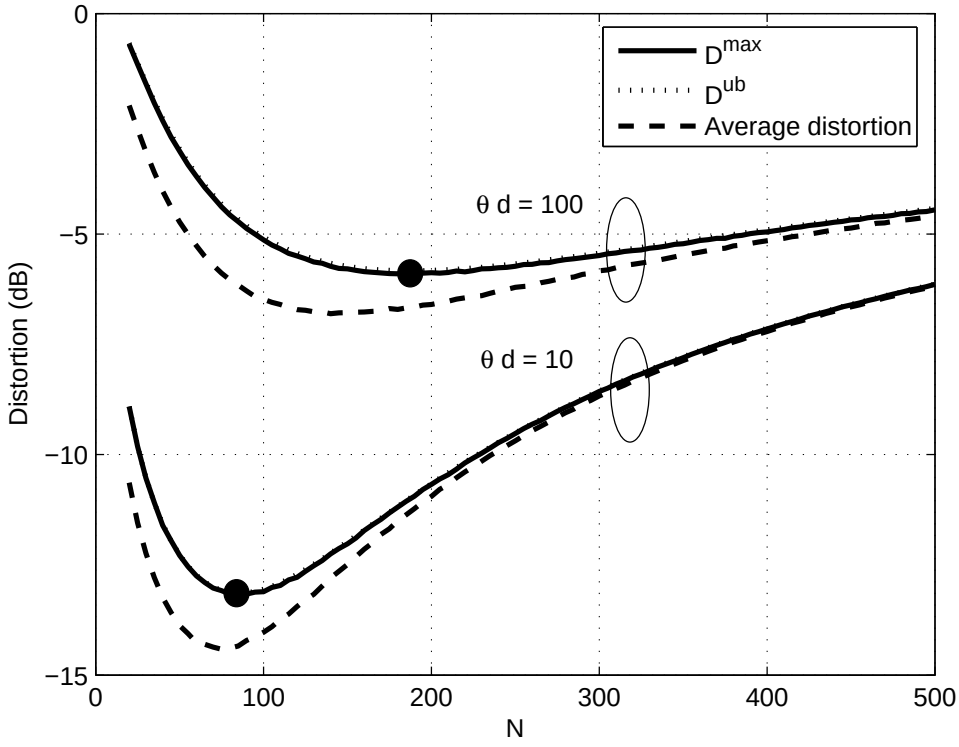


Figure 5.15: Distortion vs. network size (W=100, SNR $= 20dB$).

effect must be captured by increasing the number of sensor nodes and, due to the bandwidth and power constraints, decrease the number of bits per sensor accordingly.

Next, in Fig. 5.15, we depict the *min-max* distortion evaluated at $R^*(N)$ (computed numerically) along with the corresponding upper bound. For completeness, we also show the *average* distortion in the reconstruction of the random field which, as expected, constitutes a lower bound of the min-max distortion. In all cases, an optimal operating point N^* exists, which is illustrated through the markers on the curves. Finding such N^* (along with R^*) reveals particularly useful for random fields with low spatial variability ($\theta d = 10$) since the curve is sharper in this case. More importantly, the upper bound of the distortion given by (5.85) is tight for the whole range of N. Consequently, the solutions of (5.95) and (5.96) turn out to be accurate approximations of the true optimal pair $N^* - R^*$ that minimizes the resulting distortion. Finally, one can also observe that, indeed, the higher the variability of the random field, the higher the distortion for all N.

5.8 Chapter summary and conclusions

In this chapter, we have extensively analyzed the problem of random field estimation with wireless sensor networks. In order to characterize the dynamics and spatial correlation of the random field, we have adopted a stationary homogeneous Gaussian Markov Ornstein-Uhlenbeck model. We have considered two scenarios of interest: delay-constrained (DC) and delay-tolerant (DT) networks. For each scenario, we have analyzed two encoding schemes, namely, quantize-and-estimate (Q&E) and compress-and-estimate (C&E). In all cases (QEDC, QEDT, CEDC and CEDT), we have carried out an extensive analysis of the average distortion experienced in the reconstructed random field. Moreover, for the QEDT and CEDT strategies we have derived closed-form expressions of *i*) the average distortion in the estimates, and *ii*) the optimal number of samples of the random field to be encoded in each timeslot (under some simplifying assumptions). Interestingly, the resulting per-timeslot distortion in DT scenarios is deterministic and constant whereas, in DC scenarios, it ultimately depends on the fading conditions experienced in each timeslot. Next, we have focused on the latency associated to the QEDT and CEDT strategies. We have modeled our system as an absorbing Markov chain and, on that basis, we have fully characterized the pdf, CDF, and the average latency for the QEDT case. For CEDT encoding, we have identified an approximate system model suitable for the computation of the average latency. Simulation results reveal that, under a total bandwidth constraint, there exists an optimal number of sensors for which the distortion in the reconstructed random field can be minimized (QEDC, QEDT, CEDC and CEDT cases). This constitutes the best trade-off in terms of, on the one hand, the ability to capture the spatial variations of the random field and, on the other, the per-sensor channel bandwidth available to encode observations. Besides, the distortion associated to delay-tolerant strategies is, as expected, lower than

for delay-constrained ones: some 2-3 dB for both the QE and CE encoding schemes. Moreover, buffer occupancy can be kept at very moderate levels (3 timeslots) with a marginal penalty in terms of distortion (less than 0.3 dB). We also observe that CE schemes effectively exploit the spatial correlation and, by doing so, attain a lower distortion than their QE counterparts (DC and DT scenarios). As far as latency is concerned, we have empirically shown that CEDT exhibits a *linear* increase in the number of sensors whereas in QEDT encoding latency grows *logarithmically* (i.e. more slowly). However, CEDT schemes attain a lower distortion than QEDT ones. Besides, for the QEDT case, there is a perfect match between simulations and the theoretical model and, for the CEDT case, latency can be accurately represented by adequately parameterizing the aforementioned approximate system model.

Finally, we have addressed scenarios where sensor nodes operate without transmit CSI. We have proposed a constant-rate encoding strategy which unavoidably entails some outage probability in Rayleigh-fading scenarios. This effect, along with the spatial sampling process and the power and bandwidth constraints that we impose, results in some distortion that we attempt to minimize by carefully selecting the optimal number of sensor nodes to be deployed and the corresponding encoding rate. On the basis of a (tight) upper bound on the maximum distortion in each network segment, we have derived closed-form expressions of the corresponding optimal values. Computer simulations reveal that random fields with low spatial variability should be sparsely sampled with high resolution observations. Contrarily, in scenarios with high spatial variability, the random field is better represented (i.e. the distortion is lower) if the number of sensors is increased despite that the encoding rate must be necessarily lower due to the bandwidth and power constraints. Finding an optimal operating pair N^*-R^* reveals particularly useful for random fields with low spatial variability due to the fact that the distortion curve is sharper in this case.

5.A Appendix

5.A.1 Stability analysis

We want to prove that buffers are stable (i.e. their occupancy is bounded) for large L. Let $b_k(i)$ denote the number of samples in the buffer of the k-*th* sensor in time slot i, with initial conditions given by $b_k(0) = L_0 n$. After L timeslots, the increase in the number of samples stored in the buffer can be expressed as

$$b_k(L) - b_k(0) = Ln - \sum_{i=1}^{L} \alpha'_k(i) n \tag{5.97}$$

where Ln accounts for the number of samples generated in those L timeslots, and $\sum_{i=1}^{L} \alpha'_k(i) n$ with

$$\alpha'_k(i) = \frac{\log_2\left(1 + \mathsf{SNR}\gamma_k(i)\right)}{\overline{R} - \delta} > \alpha^*_k(i) \tag{5.98}$$

stands for the actual number of samples encoded and transmitted by the k-*th* sensor node. The probability of experiencing an increase greater than ϵn in the number of samples stored reads

$$\begin{aligned} \Pr\left(b_k\left(L\right) - b_k\left(0\right) \geq \epsilon n\right) &= \Pr\left(Ln - \sum_{i=1}^{L} \alpha'_k(i) n \geq \epsilon n\right) & (5.99)\\ &= \Pr\left(\sum_{i=1}^{L} \alpha'_k(i) \leq L - \epsilon\right). & (5.100) \end{aligned}$$

for any $\epsilon > 0$. Replacing (5.98) into this last expression yields:

$$\begin{aligned} &\Pr\left(b_k\left(L\right) - b_k\left(0\right) \geq \epsilon n\right) & (5.101)\\ &= \Pr\left(\sum_{i=1}^{L} \frac{\log_2\left(1 + \mathsf{SNR}\gamma_k(i)\right)}{\overline{R} - \delta} \leq L - \epsilon\right) & (5.102)\\ &= \Pr\left(\sum_{i=1}^{L} \log_2\left(1 + \mathsf{SNR}\gamma_k(i)\right) - L\overline{R} \leq \left(\epsilon - L\right)\delta - \epsilon\overline{R}\right) & (5.103)\\ &= \Pr\left(\sum_{i=1}^{L} \frac{\log_2\left(1 + \mathsf{SNR}\gamma_k(i)\right) - \overline{R}}{\sqrt{L\mathrm{Var}\left(R\right)}} \leq \frac{\left(\epsilon - L\right)\delta - \epsilon\overline{R}}{\sqrt{L\mathrm{Var}\left(R\right)}}\right) & (5.104) \end{aligned}$$

where we have defined

$$\mathrm{Var}\left(R\right) \triangleq \mathbb{E}_\gamma\left[\left(\log_2\left(1 + \mathsf{SNR}\gamma(i)\right) - \overline{R}\right)^2\right].$$

For large L, we can resort to the central limit theorem by which

$$Z = \sum_{i=1}^{L} \frac{\log_2\left(1 + \mathsf{SNR}\gamma_k(i)\right) - \overline{R}}{\sqrt{L\mathrm{Var}\left(R\right)}} \sim \mathcal{N}\left(0, 1\right). \tag{5.105}$$

Hence, as long as δ takes strictly positive values ($\delta > 0$), we have that

$$\lim_{L\to\infty} \Pr\left(b_k\left(L\right) - b_k\left(0\right) \geq \epsilon\, n\right) \;=\; \lim_{L\to\infty} \Pr\left(Z \leq \frac{\left(\epsilon - L\right)\delta - \epsilon\overline{R}}{\sqrt{L\mathrm{Var}\left(R\right)}}\right) = 0. \quad (5.106)$$

This result states that, as long as we encode a slightly higher number of samples per timeslot (which depends on parameter δ) the probability that the increase in buffer occupancy exceeds ϵn samples (for a *finite* value of ϵ) can be made arbitrary small for large L. That is, buffers are stable. Conversely, $\delta = 0$ yields

$$\lim_{L\to\infty} \Pr\left(b_k\left(L\right) - b_k\left(0\right) \geq \epsilon n\right) \stackrel{\delta=0}{=} \frac{1}{2}, \quad (5.107)$$

this meaning that, even for arbitrarily large values of ϵ, the probability that buffer occupancy increases beyond $\epsilon\, n$ is unavoidably $1/2$ (i.e. unstable buffers).

In addition to this main result, the probability for buffers to drain after L timeslots can be expressed as

$$\begin{aligned} p_{\mathsf{drain}} &= \Pr\left(b_k(L) = 0\right) && (5.108)\\ &= \Pr\left(\sum_{i=1}^{L} \alpha_k'(i) n \geq \left(L + L_0\right) n\right) && (5.109)\\ &= \Pr\left(\sum_{i=1}^{L} \frac{\log_2\left(1 + \mathsf{SNR}\gamma_k(i)\right)}{\overline{R} - \delta} \geq L + L_0\right). && (5.110) \end{aligned}$$

By resorting again to the central limit theorem, we have that for any positive value of δ

$$\lim_{L\to\infty} p_{\mathsf{drain}} = \lim_{L\to\infty} \Pr\left(Z \geq \frac{L_0\overline{R} - \left(L + L_0\right)\delta}{\sqrt{L\mathrm{Var}\left(R\right)}}\right) = 1 \quad (5.111)$$

and, thus, buffers will drain with probability one after a sufficiently large number of timeslots.

5.A.2 Decoding structure

In this appendix, we provide a decoding structure that makes possible the decoding of the data when sensors adopt the CEDT strategy. Without loss of generality, consider two sensor nodes $k \in \{1, 2\}$, where sensor $k = 1$ will be referred as the *master* sensor node and, sensor node $k = 2$, as the *slave* sensor node, respectively.

Encoding and decoding master sensor's data

At the sensor node: The sensor constructs a codebook consisting of $2^{n_1(i)R_1(i)}$ codewords and

sends the index of the codeword[8] $\mathbf{u}_1(i)$, that is $v_1(i) \in 1, \ldots, 2^{n_1(i)R_1(i)}$.

At the FC: Since $n_1(i)R_1(i) \leq mR_c(\gamma_1(i))$ where

$$R_c(\gamma) = \log_2\left(1 + \mathsf{SNR}\gamma\right),$$

the index $v_1(i)$ is successfully decoded and so $\mathbf{u}_{i,1}$, which contains $n_1(i)$ samples.

Next, $\mathbf{u}_1(i)$ is stored into a local buffer for a future use. Therefore, after L' timeslots we have $\mathbf{u}_{L'} = \left[\mathbf{u}_1^T(i), \ldots, \mathbf{u}_1^T(i+L')\right]^T$ containing a total of $n_{L'} = \sum_{l=0}^{L'-1} n_1(i+l)$ quantized samples.

Encoding and decoding slave sensor's data

At the sensor node: In each timeslot, the sensor construct a codebook consisting of $2^{n_2(i)R_2'(i)}$ codewords. Then, the codebook is partitioned into $2^{n_2(i)R_2(i)}$ bins as in [35, Theorem 14.9.1]. The sensor sends the index of the bin $v_2(i) \in 1, \ldots 2^{n_2(i)R_2(i)}$ where the codeword $\mathbf{u}_2(i)$ belongs to.

At the FC: Since $n_2(i)R_2(i) \leq mR_c(\gamma_2(i))$ the index $v_2(i)$ is successfully decoded. Therefore, the FC has located the bin where the true codeword $\mathbf{u}_2(i)$ is. For a reliable decoding, the FC has to be available to distinguish the true codeword $\mathbf{u}_2(i)$ within bin $v_2(i)$. To that extent, we first assume the set of samples $\mathbf{u}_{L'}$ at the decoding buffer of the master sensor is of length $n_{L'} \geq n_2(i)$. In this case, the minimum number of bins, i.e. $2^{n_2(i)R_2(i)}$ or, equivalently, the minimum rate per sample $n_2(i)R_2(i)$ must satisfy:

$$\begin{aligned} n_2(i)R_2(i) &\geq \mathrm{I}\left(\mathbf{y}_2(i); \mathbf{u}_2(i)|\mathbf{u}_{L'}\right) && (5.112)\\ &= \sum_{n=1}^{n_2(i)} \mathrm{I}\left(y_2(i); u_2(i)|u_{L'}(i)\right) && (5.113)\end{aligned}$$

where the second equality is due to the memoryless property of the source y and, due to causality, the set of samples contained in $\mathbf{u}_{L'}$ which are *temporally correlated* with $\mathbf{u}_2(i)$ are the first $n_2(i)$ samples. Note that if the master sensor encodes its blocks of samples according to (5.37) then, it is straightforward to show that the encoding rate per sample is the same in each block. Namely, we have that $\mathbf{u}_{L'} = \mathbf{y_1} + \mathbf{z_1}$ where $\mathbf{z_1} \sim \mathcal{CN}\left(0, \sigma_{z_1}^2\mathbf{I}\right)$ where $\sigma_{z_1}^2$ is defined as in (5.48). Consequently, from (5.113) we can drop the sample index and obtain

$$\begin{aligned} n_2(i)R_2(i) &\geq \sum_{n=1}^{n_2(i)} \mathrm{I}\left(y_2(n); u_2(n)|u_{L'}(n)\right) && (5.114)\\ &= n_2(i)\mathrm{I}\left(y_2; u_2|u_1\right) && (5.115)\\ &= n_2(i)\log_2\left(1 + \frac{\sigma_{y_2|u_1}^2}{\sigma_{z_2}^2}\right), && (5.116)\end{aligned}$$

[8] To recall, i stands for the timeslot index

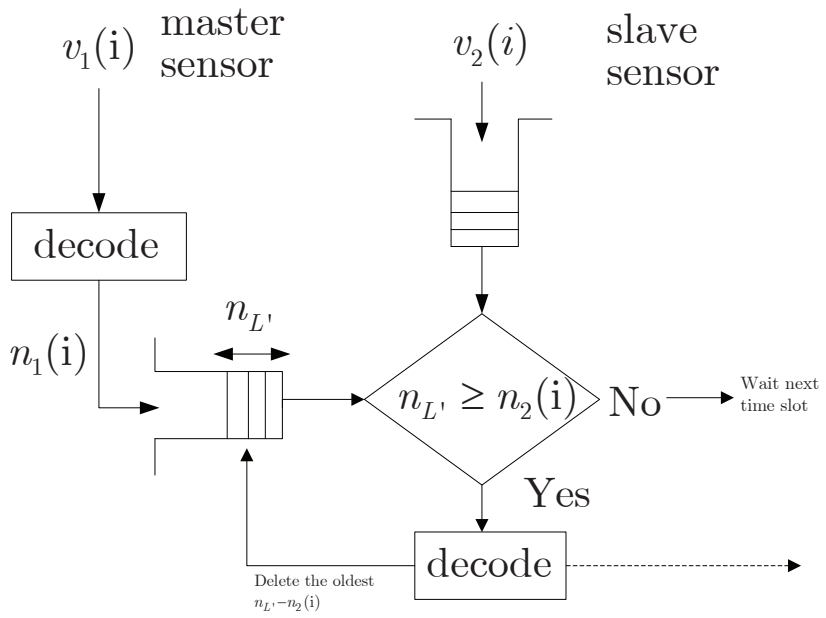


Figure 5.16: Decoding structure for the CEDT encoding strategy.

where $\sigma^2_{y_2|u_1}$ does *not* depend on the timeslot index i and $\sigma^2_{z_2}$ is a function only of *local* sensor parameters. In particular, the value of $\sigma^2_{z_2}$ is set *locally* according to (5.48) and, hence, we can decode $\mathbf{u}_2(i)$ reliably. Next, the block of decoded samples are deleted from the buffers at the FC.

Contrarily, if the set of samples $\mathbf{u}_{L'}$ at the decoding buffer of the master sensor is of length $n_{L'} \leq n_2(i)$, the FC will store the index $v_2(i)$ and wait until $n_{L'} \geq n_2(i)$. The decoding structure is summarized in Figure 5.16.

Generalization to an arbitrary number of sensor nodes

The strategy above can be easily generalized to an arbitrary number of sensor nodes N. In particular, a slave sensor will act as a master sensor for the next sensor node in the encoding chain. Interestingly, this strategy can be implemented at the sensor nodes by using *local* information only, whereas in the CEDC strategy the encoding rate used at the master node must be known in advance at the slave sensor node for a reliable decoding at the FC.

Chapter 6

Conclusions and Future Work

In this PhD dissertation, we have focused on the design of decentralized estimation schemes for wireless sensor networks. Essentially, we have addressed the problem from a signal processing and information theoretical perspective. Still, we have also considered the impact of some selected functionalities at the link layer of the OSI protocol stack (e.g. scheduling protocols) or network topologies (flat/hierarchical). First, in Chapter 3, we have addressed the power allocation problem in amplify-and-forward WSNs, where sensor observations are scaled by an amplifying factor and transmitted to the FC (i.e. *analog* transmissions). Conversely, in Chapters 4 and 5, we have focused on the case where sensors encode their observations into a number of bits (i.e. *digital* transmissions) before sending them to the FC. Regarding the estimation problem, in Chapters 3 and 4 we have addressed the problem of the estimation of a spatially-homogeneous parameter, whereas, in Chapter 5, we have focused on the more realistic case of the estimation of spatial random fields. In the sequel, we summarize the main findings of each chapter of this dissertation.

6.1 Conclusions

Chapter 3

In this chapter, we have addressed the problem of power allocation in a context of amplify-and-forward WSNs for decentralized parameter estimation. First, we have focused on a flat topology and we have proposed a class of opportunistic power allocation schemes. Such OPA schemes have one feature in common: only sensors experiencing certain local conditions (i.e. channel gain and/or residual energy above a threshold) are allowed to participate in the estimation process. We have addressed a number of classical problems of interest such as the minimization of estimation distortion (OPA-D), the minimization of transmit power (OPA-P) or the enhancement of network lifetime (OPA-LT). In all cases, we have derived a closed-form expression of an approximate but tight closed-form expression of the threshold along with the corresponding power allocation rule. Furthermore, we have also addressed the case with imperfect CSI to derive an improved version of the OPA-D scheme (OPA-DR) which is robust to such imperfections.

Interestingly, computer simulation results have revealed that the performance of the OPA-D and OPA-P schemes are virtually identical to that of the optimal schemes (WF-D and WF-P, respectively). More significantly, we have proved that the rate at which distortion decreases for the OPA-D and WF-D is identical when the number of sensor nodes increases without bound. We have also observed that the robust version (OPA-DR) performs close to systems operating with perfect CSI even with moderate values of CSI uncertainty. From the comparison of OPA-LT with OPA-P, we have concluded that a sensible use of REI results in a two-fold extension of network lifetime.

Finally, we have adopted a hierarchical network structure which is suitable for scenarios with severe path loss in the sensor-to-FC channels. In this scenario, we have addressed the power allocation problem for the minimization of the attainable distortion at the FC. We have shown that the minimization problem can be decomposed into smaller sub-problems and, furthermore, we have derived a closed-form expression for the optimal power split between layers for the UPA case. Computer simulation results have shown that a hierarchical network with UPA schemes in both layers constitutes the best trade-off in terms of performance (namely, estimation accuracy) vs. CSI requirements.

Chapter 4

In Chapter 4, we have focused again on the problem of decentralized parameter estimation with WSNs. Unlike in previous chapter, we have considered that sensors are capable of encoding their observations for digital transmission.

First, we have conducted an in-depth analysis of the Quantize-and-Estimate (Q&E) and Compress-

and-Estimate (C&E) encoding strategies in (orthogonal) Gaussian and Rayleigh-fading channels under power *and* bandwidth constraints. For the Q&E scheme, we have shown that there exists an optimal network size which minimizes the overall distortion in the estimates. Additionally, we have derived an approximate closed-form expression of its optimal operating point (Gaussian channels and some cases of interest in Rayleigh-fading channels without CSIT). For the C&E scheme, we have analytically shown that encoding the observations in a *decreasing* order of (sensor-to-FC) channel gains minimizes the resulting distortion. Computer simulation results reveal that ordering is particularly important in scenarios with moderate observation noise or transmit power. We have also derived, in a context of Rayleigh-fading channels, closed-form expressions of the distortion attained by the Q&E and C&E (lower bound) schemes when the number of sensor nodes increases without bound. Next, we have constrained sensor nodes to operate without instantaneous transmit CSI. In this context, we have proposed to adopt a constant and common rate for all sensors. Next, we have approximately computed the optimal encoding rate for two cases of interest, namely, sensors with high and low observation noise and found out that, interestingly, the lack of CSIT translates into a moderate increase of distortion for the whole range of SNR values.

Finally, we have explicitly taken into account realistic multiple-access schemes in a context of *hierarchical* WSNs. More precisely, we have analyzed the impact of a contention-based mechanism (ALOHA), and the packet collisions that it entails. First, we have derived a *closed-form* expression of the distortion attained at the FC with a reservation-based protocol (e.g. TDMA) which has been used as a benchmark. Next, we have extended the analysis to encompass the effect of packet collisions stemming from the use of contention-based schemes. Specifically, we have found an approximate (yet tight) expression of the distortion associated to the ALOHA protocol. On that basis, we have identified the optimal time split, for sensor-to-CH (Layer 1) and CH-to-FC (Layer 2) communications. Furthermore, we have derived (approximate) closed-form expressions of the optimal time split for two cases of interest, namely, high data rate and low data rate per sensor. By means of computer simulations, we have shown that the adoption of contention-based mechanisms is particulary harmful in Layer 2 whereas their impact in Layer 1 is moderate.

Chapter 5

In Chapter 5, we have extensively analyzed the problem of random field estimation with wireless sensor networks. In order to characterize the dynamics and spatial correlation of the random field, we have adopted a stationary homogeneous Gaussian Markov Ornstein-Uhlenbeck model.

First, we have considered two scenarios of interest: delay-constrained (DC) and delay-tolerant (DT) networks. For each scenario, we have analyzed two encoding schemes, namely, Quantize-and-Estimate (Q&E) and Compress-and-Estimate (C&E). In all cases (QEDC, QEDT, CEDC and CEDT), we have carried out an extensive analysis of the average distortion experienced in

the reconstructed random field. For delay-tolerant cases, we have derived closed-form expressions of *i*) the average distortion in the estimates, and *ii*) the optimal number of samples of the random field to be encoded in each timeslot. Interestingly, the resulting per-timeslot distortion in DT scenarios is deterministic and constant whereas, in DC scenarios, it ultimately depends on the fading conditions experienced in each timeslot. Simulation results reveal that, under a total bandwidth constraint, there exists an optimal number of sensors for which the distortion in the reconstructed random field can be minimized (QEDC, QEDT, CEDC and CEDT cases). This constitutes the best trade-off in terms of, on the one hand, the ability to capture the spatial variations of the random field and, on the other, the per-sensor channel bandwidth available to encode observations. Besides, the distortion associated to delay-tolerant strategies is, as expected, lower than for delay-constrained ones. Moreover, buffer occupancy can be kept at very moderate levels with a marginal penalty in terms of distortion. We have also observed that CE schemes effectively exploit the spatial correlation and, by doing so, attain a lower distortion than their QE counterparts (DC and DT scenarios). As far as latency is concerned, we have shown that CEDT exhibits a *linear* increase in the number of sensors whereas in QEDT encoding latency grows *logarithmically* (i.e. more slowly). However, CEDT schemes attain a lower distortion than QEDT ones.

Finally, we have closed the chapter by addressing the case where sensors operate without CSIT. We have proposed a constant-rate encoding strategy which unavoidably entails some outage probability in Rayleigh fading scenarios. On the basis of a (tight) upper bound of the maximum distortion in each network segment, we have derived closed-form expressions of the corresponding optimal values. Results have revealed that the optimal operating point is particularly useful for random fields with low spatial variability due to the fact that the distortion curve is sharper in this case.

6.2 Future work

In this section, we discuss a number of research areas and related topics for further work in the field of decentralized parameter estimation via wireless sensor networks.

- **Opportunistic power allocation schemes with different sensor observation qualities**. The OPA framework could be extended by considering different observations qualities at the sensor nodes (namely, different observation noises). In this case, the sensor selection algorithm should consider not only the uplink channel quality of each sensor node but also the quality of the sensor observations.

- **Joint source-channel coding with non-synchronized sensors**. Amplify-and-forward schemes are known to be asymptotically optimal where sensor transmissions are fully

synchronized. However, distributed synchronization of the sensor signals can be difficult to achieve in practice. An interesting extension would the consideration of synchronization errors and their impact on the rate at which distortion decreases.

- **Encoding strategies and signal processing techniques for hierarchical network topologies**. An important open issue is how data is processed at the cluster-heads. Various options are possible: *i*) consolidation of the cluster information into a cluster estimate and re-transmission to the FC, *ii*) re-transmission of all the sensors measurements to the FC or, *iii*) hybrid strategies. In all cases, a number of interesting trade-offs (reliability, accuracy, etc) turn up. Besides, it would be interesting the analysis of different encoding strategies (e.g. amplify-and-forward, quantize-and-estimate and compress-and-estimate) on such hierarchical WSN topology.

- **Successive refinement techniques with random access mechanisms**. By adopting successive refinement techniques along with superposition coding techniques, sensors can adaptively encode their observation by considering some *levels* of side information at the FC. Motivated by this fact, it would be interesting the design and analysis of such successive refinement techniques in the context of contention-based multiple access channels, where sensors are not aware of the current side information available at the FC.

- **Opportunistic random field estimation**. Future work on random field estimation also encompasses the extension of the analysis carried out in Chapter 5 by addressing the case where only those sensors experiencing favorable channel conditions actually sample the field.

- **Uniform/nonuniform sampling of spatial random fields**. Here, the focus would be on assessing the impact of *random* (vs. deterministic) deployment of the sensor nodes in the reconstruction of the random field.

Bibliography

[1] I. F. Akyildiz, W. Su, Y. Sankarsubramanian, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, Mar. 2002.

[2] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed diffusion for wireless sensor networking," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 2–16, Feb. 2002.

[3] W. Tsujita, S. Kaneko, T. Ueda, H. Ishida, and T. Moriizumi, "Sensor-based air-pollution measurement system for environmental monitoring network," *Proceedings IEEE Int. Conf. Solid-State Sensors, Actuators and Microsystems*, vol. 1, pp. 544–547, Jun. 2003.

[4] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, *Wireless Sensor and Actuator Networks: Technologies, analysis and design*, Elsevier, 2008.

[5] Institute of Electrical and Electronic Engineers, "IEEE std. 802.15.4-2003, IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements - part 15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low rate wireless personal area networks (LR-WPANs)," *IEEE Press*, October 2003.

[6] Energy harvesting forum, "Energy harvesting electronic solutions for wireless sensor networks & control systems," .

[7] Yunxia Chen, Qing Zhao, V. Krishnamurthy, and D. Djonin, "Transmission scheduling for optimizing sensor network lifetime: A stochastic shortest path approach," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2294–2309, May 2007.

[8] Zhi-Quan Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[9] A. Ribeiro and G.B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, March 2006.

[10] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks part II: Unknown probability density function," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2784–2796, Jul. 2006.

[11] S. Cui, J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683 – 4695, Sept. 2007.

[12] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science*, April 2003, pp. 162–177.

[13] M. Dohler, A. Gkelias, and A. H. Aghvami, "Capacity of distributed PHY-layer sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 622–639, March 2006.

[14] A. del Coso, U. Spagnolini, and C. Ibars, "Cooperative distributed MIMO channels in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 402–414, Feb. 2007.

[15] S. Bandyopadhyay and E. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)*, 2003.

[16] G. Mergen, Q. Zhao, and L. Tong, "Sensor networks with mobile access: energy and capacity considerations," *IEEE Transaction on Wireless Communications*, vol. 54, no. 11, pp. 2033–2044, Nov. 2006.

[17] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proceedings IEEE MILCOM'03*, Monterey (CA), USA, Nov. 2003, pp. 688–693.

[18] Steven M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice Hall Signal Processing Series, 1993.

[19] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[20] Jin-Jun Xiao, S. Cui, Zhi-Quan Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 431–422, Feb. 2006.

[21] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Elsevier Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004.

[22] Steven Finch, "Ornstein-Uhlenbeck Process," May 2004.

[23] D. Marco and D.L. Neuhoff, "Reliability vs. efficiency in distributed source coding for field-gathering sensor networks," in *IEEE IPSN 2004*, Berkeley, California, USA, April 2004, pp. 161–168.

[24] Min Dong, Lang Tong, and B.M. Sadler, "Impact of data retrieval pattern on homogeneous signal field reconstruction in dense sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4352–4364, Nov. 2006.

[25] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 316–329, 2006.

[26] A. Dogandžić and Kun Qiu, "Decentralized random-field estimation for sensor networks using quantized spatially correlated data and fusion-center feedback," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 6069–6085, Dec. 2008.

[27] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3456–3470, July 2007.

[28] G. Scutari, S. Barbarossa, and L. Pescosolido, "Distributed decision through self synchronizing sensor networks in the presence of propagation delays and asymetric channels," *IEEE Transactions on Signal Processing*, (to appear).

[29] G. Scutari, S. Barbarossa, and A. Swami, "Achieving consensus in self-organizing wireless sensor networks: the impact of network topology on energy consumption," *IEEE Transactions on Signal Processing*, (to appear).

[30] G. Scutari, S. Barbarossa, and A. Swami, "Achieving consensus in self-organizing wireless sensor networks: The impact of network topology on energy consumption," in *Proceedings IEEE ICASSP 2007*, Honolulu, Hawaii, USA, Apr. 2007, vol. 2, pp. 841–844.

[31] S. Barbarossa and G. Scutari, "Distributed decision through self-synchronization," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 26–35, 2007.

[32] J. Aslam, Z. Butler, F. Constantin, V. Crespi, G. Cybenko, and D. Rus, "Tracking a moving object with a binary sensor network," in *in Proc. 1st Int. Conf. Embedded Networked Sensor Systems*, 2007, pp. 150–161.

[33] P.M. Djuric, M. Vemula, and M.F. Bugallo, "Tracking with particle filtering in tertiary wireless sensor networks," in *IEEE ICASSP 2005*, Philadelphia, USA, March 2005, vol. 4, pp. 757–760.

[34] A. Ribeiro, G.B. Giannakis, and S.I. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782–4795, Dec. 2006.

[35] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley Series in Telecommunications, 1993.

[36] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul 1973.

[37] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[38] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proceedings IEEE ISIT 2004*, June 2004.

[39] T. Berger, Zhen Zhang, and H Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 887–902, May 1996.

[40] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *IEEE ICC 1995*, June 1995, vol. 1, pp. 331–335.

[41] D.N.C. Tse and S.V. Hanly, "Multiaccess fading channels I. polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov 1998.

[42] D. N. C. Tse, "Optimal power allocation over parallel gaussian broadcast channels," in *Proceedings of ISSIT 1997*, Ulm, Germany, June 1997.

[43] D. N. C. Tse, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[44] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, Apr. 1998.

[45] J. L. Vicario and C. Antón-Haro, "A unified approach to the analytical assesment of multi-user diversity with imperfect channel state information," *European Transactions on Telecommunications*, vol. 18, pp. 573–582, Oct. 2007.

[46] J. L. Vicario and C. Antón-Haro, "Analytical assessment of multi-user vs. spatial diversity trade-offs with delayed channel state information," *IEEE Communications Letters*, vol. 10, no. 8, pp. 588–590, 2006.

[47] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?," in *IEEE Int'l Conference on Communications*, June 2004, pp. 234–238.

[48] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4363–4372, Nov. 2007.

[49] S. Sanayei and A. Nosratinia, "Opportunistic beamforming with limited feedback," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2765 – 2771, Aug. 2007.

[50] N. Abramson, "The aloha system - another alternative for computer communication," in *Fall Joint Computer Conference*, 1970.

[51] D. Bertsekas and R. Gallager, *Data networks*, Prentice Hall, second edition, 1992.

[52] X. Qin and R.A. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.

[53] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.

[54] Q. Zhao and L. Tong, "Opportunistic carrier sensing for energy-efficient information retrieval in sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 2, pp. 231–241, April 2005.

[55] Qing Zhao and Lang Tong, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Processing Magazine*, vol. 2006, no. 4, pp. 84–93, July 2006.

[56] R. Madan and S. Lall, "Distributed algorithms for maximum lifetime routing in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2185–2193, Aug. 2006.

[57] R. Madan, S. Cui, S. Lal, and A. Goldsmith, "Cross-layer design for lifetime maximization in interference-limited wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3142–3152, November 2006.

[58] Yunxia Chen and Qing Zhao, "On the lifetime of wireless sensor networks," *IEEE Communications Letters*, vol. 9, no. 11, pp. 976–978, Nov. 2005.

[59] Y. Chen and Q. Zhao, "Maximizing the lifetime of sensor network using local information on channel state and residual energy," in *Conference on Information Sciences and Systems*, March 2005.

[60] W. Yu and M. Cioffi, "Constant-power waterfilling: performance bound and low-complexity implementation," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 23–28, Jan. 2006.

[61] S. Cui, J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks: analog vs. digital," in *Proceedings ICASSP 2005*, Philadelphia, USA, March 2005, pp. 745–748.

[62] Seema Bandyopadhyay and E.J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, March-3 April 2003, vol. 3, pp. 1713–1723.

[63] J. Matamoros and C. Antón-Haro, "Opportunistic power allocation and sensor selection schemes for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 534–539, Feb. 2010.

[64] J. Matamoros and C. Antón-Haro, "Scaling law of an opportunistic power allocation scheme for amplify-and-forward wireless sensor networks," *submitted to IEEE Communications Letters*.

[65] J. Matamoros and C. Antón-Haro, "Opportunistic power allocation in wireless sensor networks with imperfect channel state information," in *Proceedings ICT-Mobile Summit 2008*, Stockholm, Sweden, June 2008.

[66] J. Matamoros and C. Antón-Haro, "Opportunistic power allocation schemes for the maximization of network lifetime in wireless sensor networks," in *Proceedings IEEE ICASSP 2008*, Las Vegas, USA, April 2008, pp. 2273–2276.

[67] J. Matamoros and C. Antón-Haro, "Opportunistic power allocation schemes for wireless sensor networks," in *Proceedings IEEE ISSPIT*, El Cairo, Egypt, December 2007, pp. 220–224.

[68] J. Matamoros and C. Antón-Haro, "Hierarchical organizations of sensors for decentralized parameter estimation," in *Proceedings IEEE ISSPIT*, El Cairo, Egypt, December 2007, pp. 513–518.

[69] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 1993.

[70] H. A. David and H. N. Nagaraja, *Order statistics*, Wiley Series in Probability and Statistics, third edition, 2003.

[71] R.M. Corless, G.H. Gonnet, D.E.G. Hare, and D.J. Jeffrey, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.

[72] A. B. Gershman and N. D. Sidiropoulos, *Space-time processing for MIMO communications*, J. Wiley & Sonds, Ltd, 2005.

[73] J. G. Proakis, *Digital communications*, McGraw-Hill International Editions, fourth edition, 2001.

[74] Qing Zhao and Lang Tong, "Opportunistic carrier sensing for energy-efficient information retrieval in sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 2, pp. 231–241, 2005.

[75] Jin-Jun Xiao and Zhi-Quan Luo, "Multiterminal source channel communication over an orthogonal multiple-access channel," *IEEE Trans. on Information Theory*, vol. 53, no. 9, pp. 3255–3264, Sep. 2007.

[76] S.C. Draper and G.W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 966–976, Aug. 2004.

[77] J. Chen and T. Berger, "Successive Wyner-Ziv coding scheme and its application to the quadratic Gaussian CEO problem," *Submitted to IEEE Transactions on Information Theory*.

[78] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 765–775, Apr. 2005.

[79] C. Tian, A. Steiner, S. Shamai, and S.N. Diggavi, "Successive refinement via broadcast: optimizing expected distortion of a gaussian source over a gaussian fading channel," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2903–2918, July 2008.

[80] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Optimal power distribution and minimum expected distortion in Gaussian layered broadcast coding with successive refinement," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 5074–5086, November 2009.

[81] S. Shamai, "A broadcast strategy for the gaussian slowly fading channel," in *IEEE International Symposium on Information Theory 1997*, Ulm, Germany, Jun-4 Jul 1997.

[82] W.H.R. Equitz and T.M. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 269–275, Mar 1991.

[83] Chan-Soo Hwang, Kibeom Seong, and J. Cioffi, "Throughput maximization by utilizing multi-user diversity in slow-fading random access channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2526–2535, July 2008.

[84] J. Matamoros and C. Antón-Haro, "Optimal network size and encoding rate for wireless sensor network-based decentralized estimation under power and bandwidth constraints," *submitted to IEEE Transactions on Wireless Communications*.

[85] J. Matamoros and C. Antón-Haro, "To sort or not to sort: optimal sensor scheduling for successive compress-and-estimate encoding," in *Proceedings of IEEE International Conference in Communications (ICC 2009)*, Dresden, Germany, June 2009.

[86] J. Matamoros and C. Antón-Haro, "Hierarchical wireless sensor networks with contention-based multiple-access schemes - a PHY/MAC cross-layer design," in *Proceedings of second international workshop on cross layer design (IWCLD 2009)*, Palma de Mallorca, Spain, June 2009.

[87] J. Matamoros and C. Antón-Haro, "Optimized constant-rate encoding for decentralized parameter estimation with wireless sensor networks," in *Proceedings of IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC 2009)*, Perugia, Italy, June 2009, pp. 677–681.

[88] J. Matamoros and C. Antón-Haro, "Bandwidth constraints in wireless sensor-based decentralized estimation schemes for Gaussian channels," in *Proceedings of IEEE Global Conference on Communications (GLOBECOM 2008)*, New Orleans, USA, December 2008, pp. 1–5.

[89] J. Matamoros and C. Antón-Haro, "Bandwidth contraints in wireless sensor networks for Rayleigh fading channels," in *Proceedings of NEWCOM++ ACoRN Joint Workshop*, Barcelona, Spain, April 2009.

[90] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Academic Press, sixth edition, 2000.

[91] S. Pattem, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," in *International Symposium on Information Processing Sensor Networks*, April 2004, pp. 28–35.

[92] Berger J.O., De Oliveira V., and Sanso B., "Objective Bayesian analysis of spatially correlated data," *Journal of the American Statistical Association*, vol. 96, pp. 1361–1374, December 2001.

[93] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A "bit-conservation principle"," *Lecture Notes in Computer Science*, vol. 2634/2003, pp. 17–31, Jan. 2003.

[94] D. Dardari, A. Conti, C. Buratti, and R. Verdone, "Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 7, pp. 790–802, July 2007.

[95] J. Matamoros and C. Antón-Haro, "Random field estimation with delay-constrained and delay-tolerant wireless sensor networks," *submitted to EURASIP Journal on Wireless Communications and Networking*.

[96] J. Matamoros and C. Antón-Haro, "Optimal network size and encoding rate for random field estimation with wireless sensor networks," in *Proceedings of The 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2009)*, December 2009.

[97] J. Matamoros and C. Antón-Haro, "Delay-tolerant vs. delay-constrained estimation of spatial random fields," in *Proceedings Future Network & Mobile Summit 2010*, Florence, Italy, 16-18 June 2010.

[98] J. Matamoros and C. Antón-Haro, "Quantize-and-estimate encoding schemes for random field estimation with delay-constrained and delay-tolerant wireless sensor networks," *submitted to IEEE PIMRC 2010*.

[99] J. Matamoros and C. Antón-Haro, "Random field estimation with delay-tolerant wireless sensor networks: Quantize-and-estimate vs. compress-and-estimate encoding," *submitted to IEEE Globecom 2010*.

[100] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Springer-Verlag, 1988.

[101] C. D. Mayer, *Matrix analysis and applied linear algebra*, SIAM: Society for Industrial and Applied Mathematics, 2001.

[102] M. F. Neuts, *Matrix-geometric solutions in stochastic models: an algorithmic approach, Chapter 2: probability distributions of phase type*, Dover Publications Inc., 1981.