Ph.D. Dissertation

Communications in Wireless Sensor Networks: Compression, Energy Efficiency and Secrecy

Joan Enric Barceló Lladó

Thesis Advisors: Antoni Morell Pérez and Gonzalo Seco Granados
Department of Telecommunications and Systems Engineering,
Escola d'Enginyeria,
Universitat Autònoma de Barcelona (UAB).

July, 2012.

Communications in Wireless Sensor Networks: Compression, Energy Efficiency and Secrecy

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Universitat Autònoma de Barcelona.

Doctoral study: Telecomunicació i enginyeria de sistemes.

Author: Joan Enric Barceló Lladó

Thesis advisor: Antoni Morell Pérez

Thesis advisor: Gonzalo Seco Granados

Dept. de telecomunicació i enginyeria de sistemes.
Escola d'Enginyeria.
Universitat Autònoma de Barcelona.

Bellaterra, July 2012

Abstract

Wireless Sensor Networks (WSNs) have emerged as one of the most promising wireless communication systems in the last decade. They can be used in a wide variety of applications such as environmental monitoring, natural disaster prediction, healthcare, transportation, indoor positioning, and military tasks. The cost and the complexity of the nodes within a WSN are typically low, which results in constraints such as energy limitation, low computational speed, and reduced communication bandwidth. With the advances in wireless communications and the growing demand of new and more complex applications, WSNs must be optimized in order to overcome their intrinsic limitations in terms of complexity and power.

In this dissertation, and according to these constraints, we propose a set of techniques that provide to a WSN the following interesting features:

1. Distributed operation without the need of signaling among sensing nodes.
2. Energy-efficient communications.
3. Low complexity at the sensing nodes.
4. Low resource (i.e., bandwidth, time, etc.) utilization.
5. Low distortion level at the receiver.

6. Secret communications at the physical layer.

First, we study the zero-delay downsampling transmission. This technique allows the system to reduce the number of transmissions and hence decrease the total energy spent. In particular, we study the performance of deterministic, probabilistic and conditional downsampling encoding-decoding pairs for the case of the autoregressive signal model. We obtain closed form expressions for the quadratic error of the deterministic and probabilistic encoder-decoders, while accurate approximations are derived for the quadratic error of the conditional downsampling schemes.

Second, we analyze data compression applied to large WSNs. For the realistic case where the correlation parameters are not known a priori, we obtain two enhanced correlation estimators: *i*) one for the linear Wiener filter vector and *ii*) one for the achieved mean square error. Both estimators are employed in the two key steps of the distributed source coding algorithm. These estimators notably improve the performance of the algorithm in comparison to the application of classical sample estimators, specially when the dimension of the observation vector is comparable in magnitude to the number of samples used in the training phase.

Then, we propose a distributed and energy-efficient communication scheme named *Amplify-and-Forward Compressed Sensing*. This scheme is based on compressed sensing and exploits the correlation present in the signal in order to reduce both the resource utilization and the energy consumption. More specifically, the system is designed according to a cost function that controls the trade-off between the quadratic error in the reconstruction and the energy consumption of the network. In order to aid the system design, a simple model that accurately approximates the performance of the proposed scheme in terms of the quadratic error has been derived. Furthermore, we contribute to the compressed sensing theory with a tighter relationship between the minimum number of measurements that are required for a given network dimension and the sparsity level of the

transmitted signal.

Finally, the proposed Amplify-and-Forward Compressed Sensing scheme is also studied in terms of secrecy and wiretap distortion at the physical layer. It is shown that the proposed scheme is perfectly secret in the presence of one or even a small group of eavesdroppers whereas for a larger eavesdropping set, it is still possible to notably deteriorate its espionage capabilities thanks to a proposed technique specifically designed to introduce extra uncertainty only in the channel estimation of the eavesdroppers.

Resum

Les xarxes de sensors sense fils (WSNs) han esdevingut un dels sistemes de comunicació amb més projecció d'aquesta dècada. Abasten una àmplia varietat d'aplicacions tals com la monitorització del medi ambient, la predicció de desastres naturals, en medicina, en transport, posicionament en interiors, i tasques militars. Els nodes que composen la xarxa, són típicament de baix cost, cosa que atorga una sèrie de limitacions en termes d'energia, velocitat de càlcul i d'ample de banda. Amb els avenços de les comunicacions sense fils i la creixent demanda de noves i més complexes aplicacions, les WSNs s'han d'optimitzar per tal de minimitzar aquestes limitacions.

Aquesta tesi proposa un conjunt de tècniques que proporcionen a una WSN les següents característiques:

1. Implementació distribuïda sense necessitat de senyalització entre nodes sensors.
2. Comunicacions energèticament eficients.
3. Poca complexitat als nodes sensors.
4. Empra pocs recursos (temps, ample de banda, etc.).
5. Presenta un error quadràtic mig baix en reconstrucció al receptor.
6. Comunicacions secretes a capa física.

Primer, s'estudia la transmissió seqüencial de mostreig reduït. Aquesta tècnica permet la disminució del nombre de transmissions i, per tant, reduir la despesa energètica associada a la comunicació a la xarxa. En particular, s'estudia el rendiment dels codificadors determinístics, probabilístics i condicionals de mostreig reduït per senyals autoregressius. S'obtenen expressions tancades de l'error quadràtic mig pel cas de mostreig reduït determinístic i probabilístic, mentre que pel cas condicional es deriven aproximacions ajustades.

A continuació, s'analitza la compressió de la informació per WSNs grans. Pel cas on els paràmetres de correlació del senyal són desconeguts a priori, es proposen dos estimadors millorats: *i*) un per la predicció emprant el filtre de Wiener i *ii*) un per l'error quadràtic mig obtingut. Ambdós estimadors s'empren pels dos passos claus de l'algorisme de codificació distribuida de canal. Aquests estimadors milloren notablement el rendiment de l'algorisme en comparació amb els estimadors de mostres clàssics, especialment quan la dimensió del vector d'observacions és comparable en magnitud amb el nombre de mostres usades a la fase d'entrenament de l'algorisme.

Posteriorment, es proposa un esquema de comunicació distribuïda i energèticament eficient anomenat *Amplify-and-Forward Compressed Sensing*. Aquest esquema es basa en la tècnica de sensat comprimit i aprofita la correlació existent al senyal rebut per tal de reduir tant el nombre de recursos emprats com les despeses energètiques del sistema. Específicament, el sistema es dissenya seguint una funció de cost que controla el compromís existent entre error quadràtic i consum energètic de la xarxa. Per aconseguir aquest disseny, es deriva un model simple que aproxima el rendiment de l'esquema proposat en termes d'error quadràtic mig. A més, es contribueix a la teoria de sensat comprimit amb una nova i més ajustada relació entre el mínim nombre de mesures necessàries donades unes determinades propietats del senyal.

Finalment, s'estudia l'esquema proposat Amplify-and-Forward Com-

pressed Sensing des d'un punt de vista de secretisme a capa física. Es demostra que aquest esquema assoleix secretisme perfecte sota la presència d'un o d'un grup reduït d'espies, mentre que per un nombre més gran, és possible deteriorar notablement les seves capacitats d'espionatge gràcies a una tècnica proposta especialment dissenyada per introduir un extra d'incertesa solament a l'estimació dels espies.

Resumen

Las redes de sensores inalámbricas (WSNs) se han convertido en uno de los sistemas de comunicación con mayor proyección de la década. Abarcan una gran variedad de aplicaciones tales como la monitorización del medio ambiente, la predicción de desastres naturales, la medicina, el transporte, posicionamiento en interiores, y tareas militares. Los nodos que componen una WSN son comúnmente de bajo coste, lo que otorga una serie de limitaciones en términos tales como de energía, velocidad de cálculo y de ancho de banda. Con los recientes avances en el campo de las comunicaciones inalámbricas y el continuo crecimiento de la demanda de nuevas y más complejas aplicaciones, las WSNs deben ser optimizadas para minimizar estas limitaciones.

Esta tesis propone un conjunto de técnicas que proporcionan a una WSN las siguientes características:

1. Implementació distribuida sin necesidad de señalización entre los nodos sensores.

2. Comunicaciones energéticamente eficientes.

3. Se requiere poca complejidad en los nodos sensores.

4. Utiliza pocos recursos (tiempo, ancho de banda, etc.)

5. Presenta un bajo error cuadrático medio en reconstrucción en el receptor.

6. Comunicaciones secretas en capa física.

Primero, se estudia la transmisión secuencial de muestreo reducido. Esta técnica permite disminuir el número de transmisiones y, por tanto, reducir el gasto energético asociado a la comunicación en la red. En particular, se estudia el rendimiento de los codificadores determinísticos, probabilísticos y condicionales de muestreo reducido para señales autoregresivas. Se obtienen expresiones cerradas del error cuadrático medio para el caso de muestreo reducido determinístico y probabilístico, mientras que para el caso condicional se derivan aproximaciones ajustadas.

A continuación, se analiza la compresión de información para WSNs grandes. Para el caso práctico donde los parámetros de correlación de la señal no se conocen a priori, se proponen dos estimadores mejorados: *i*) uno para la predicción usando el filtro de Wiener y *ii*) otro para el error quadrático medio obtenido. Ambos estimadores se usan en los dos pasos clave del algoritmo de codificación distribuida de canal. Estos estimadores mejoran notablemente el rendimiento del algoritmo en comparación a los estimadores de muestras clásicos, especialmente cuando la dimensión del vector de observaciones es comparable al número de muestras usadas en la fase de entrenamiento del algoritmo.

Posteriormente, se propone un esquema de comunicación distribuido y energéticamente eficiente llamado *Amplify-and-Forward Compressed Sensing*. Este esquema está basado en sensado comprimido y aprovecha la correlación existente en la señal recibida para reducir tanto el número de recursos utilizados como el gasto de energía. Especificamente, el sistema se diseña según una función de coste que controla el compromiso existente entre el error cuadrático y consumo energético de la red. Para conseguir este diseño, se deriva un modelo simple que aproxima el rendimiento del esquema propuesto en términos de error cuadrático medio. Además, se contribuye

a la teoría de sensado comprimido con una nueva y mas ajustada relación entre el mínimo número de medidas necesarias dadas unas determinadas propiedades de la señal.

Finalmente, se estudia el esquema propuesto Amplify-and-Forward Compressed Sensing desde un punto de vista de secretismo en capa física. Se demuestra que el esquema logra secretismo perfecto bajo la presencia de uno o un grupo reducido de espías mientras que para un número más grande, es posible deteriorar notablemente sus capacidades de espionaje gracias a una técnica propuesta especialmente diseñada para introducir una extra de incertidumbre solamente a la estimación de los espías.

Notation

In general, boldface upper-case letters denote matrices ($\mathbf{A}$), boldface lower-case letters denote column vectors ($\mathbf{a}$), and italics denote scalars (a).

$\mathbf{A}^T, \mathbf{A}^*, \mathbf{A}^H$	Transpose, complex conjugate, and conjugate transpose (Hermitian) of matrix $\mathbf{A}$, respectively.
$\mathbf{A}^{-1}$	Inverse of a full-rank matrix $\mathbf{A}$.
$\mathbf{a}^T, \mathbf{a}^*, \mathbf{a}^H$	Transpose, complex conjugate, and conjugate transpose (Hermitian) of vector $\mathbf{a}$, respectively.
$[\mathbf{A}]_{i,j}$	The (ith, jth) element of matrix $\mathbf{A}$.
$[\mathbf{a}]_i$	The ith element of vector $\mathbf{a}$.
$[\mathbf{A}]_i$	The ith row of matrix $\mathbf{A}$.
$[\mathbf{A}]_{:,i}$	The ith column of matrix $\mathbf{A}$.

$\|a\|$	Absolute value of the scalar a.
$\|\mathbf{a}\|_{l^0}$	l^0- (pseudo)norm of vector $\mathbf{a}$. It counts the number of non-zero elements in a vector $\mathbf{a}$.
$\|\mathbf{a}\|_{l^1}$	l^1- norm of vector $\mathbf{a}$, $\|\mathbf{a}\|_{l^1} = \sum_i \|[\mathbf{a}]_i\|$.
$\|\mathbf{a}\|_2$ or $\|\mathbf{a}\|$	Euclidean norm of vector $\mathbf{a}$, $\|\mathbf{a}\|_2 = (\mathbf{a}^H\mathbf{a})^{1/2}$.
$\|\mathbf{A}\|$ or $\|\mathbf{A}\|_F$	Frobenius norm of a matrix $\mathbf{A}$, $\|\mathbf{A}\| = \sqrt{\mathrm{Tr}(\mathbf{A}^H\mathbf{A})}$.
$\mathbf{I}_M$	Identity matrix of dimension M.
$\mathbf{0}$	All-zero matrix with convenient dimensions.
$\mathbf{e}_i$	Canonical vector with all the elements being zero except the ith position, which is equal to one.
$\mathbf{A}^{1/2}$	Positive definite Hermitian square root of A, $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.
$\mathrm{Tr}(\mathbf{A})$	Trace of $\mathbf{A}$.
$\hat{\mathbf{A}}$, $\hat{\mathbf{a}}$, $\hat{a}$	Estimation of matrix $\mathbf{A}$, vector $\mathbf{a}$, and scalar a, respectively.
$\mathbf{a}_K$	K-sparse approximation of vector $\mathbf{a}$, $\|\mathbf{a}\|_{l^0} = K$.
$\mathbb{E}[\cdot]$	Statistical expectation.
$\mathbb{E}[\cdot\|a < b]$	Conditional expectation given the condition $a < b$.
$\mathrm{var}(\cdot)$	Statistical variance.
$\mathrm{var}(\cdot\|a < b)$	Conditional variance given the condition $a < b$.
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$	Gaussian vector distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$.
$\sigma^2_{\mathbf{x}}$	Variance of $\mathbf{x}$.
$\mathrm{erf}(x)$	Error function of variable x, $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.
$\mathrm{erfc}(x)$	Complementary error function of variable x, $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$.

$a!$	Factorial of a, $a! = \prod_{i=1}^{a} i$.
$(a)^+$	Maximum between the real value a and 0, $\max(0, a)$.
$\binom{S}{E}$	Binomial coefficient, $\binom{S}{E} = \frac{S!}{(S-E)!E!}$
$\Pi(x)$	Rectangular function, $\Pi(x) = 0$ if $\|x\| > 0.5$, $\Pi(x) = 1$ if $\|x\| < 0.5$, and $\Pi(x) = 0.5$ if $\|x\| = 0.5$.
$\delta(x)$	Dirac delta function.
$(\cdot)^\star$	Optimal value.
$\lceil \cdot \rceil$	Ceil function
$\mathcal{I}(a \leq b)$	Indicator function of the condition $a \leq b$.
$\log(\cdot)$	Natural logarithm.
$\log_b(\cdot)$	Logarithm in base b.
$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$,	The set of natural, integer, real and complex numbers, respectively.
$\mathbb{R}^M$, $\mathbb{C}^M$	The set of M-dimensional vectors with real and complex entries, respectively.
$\mathbb{R}^{M\times N}$, $\mathbb{C}^{M\times N}$	The set of $M \times N$ matrices with real and complex entries, respectively.
(a, b), $[a, b]$	Open interval ($a < x < b$) and closed interval ($a \leq x \leq b$), respectively.
$\cup$, $\cap$	Intersection and union of sets.
$\varnothing$	Empty set.
$\setminus$	Set substraction.
$\asymp$	Almost surely convergence.
$:=$	Defined as.
$\sim$	Distributed according to.
$\approx$ or $\eqsim$	Approximate equivalence.

Acronyms

a.s.	almost surely
A/D	Analog-to-Digital
AF	Amplify-and-Forward
AF-CS	Amplify-and-Forward Compressed Sensing
AR	Auto-Regressive
AR-1	Auto-Regressive process of order 1
AWGN	Additive White Gaussian Noise
bps	Bits per Second
CA	Classical Approach
CDE	Conditional Downsampling Encoder
cdf	Cumulative Density Function
CDMA	Code Division Multiple Access
CS	Compressed Sensing
CSD	Compressed Sensing Decoder
CSI	Channel State Information
CSIR	Channel State Information at the Receiver
CWS	Compressed Wireless Sensing
DCT	Discrete Cosine Transform
DDE	Deterministic Downsampling Encoder

DL	Diagonal Loading
DPCM	Differential Pulse Code Modulation
DSC	Distributed Source Coding
ECSI	Eavesdropper Channel State Information
edf	Empirical Distribution Function
FC	Fusion Center
FDMA	Frequency Division Multiple Access
GCWS	Generalized Compressed Wireless Sensing
GSA	Generalized Statistical Analysis
i.i.d.	Independent and Identically Distributed
LASSO	Least Absolute Shrinkage and Selection Operator
LWF	Linear Wiener Filter
MAC	Multiple Access Channel
MC	Markov Chain
MIMO	Multiple-Input Multiple-Output
MLE	Maximum Likelihood Estimatior
MMSE	Minimum Mean Square Error
MSE	Minimum Square Error
NP	Nondeterministic Polynomial time
OFDM	Orthogonal Frequency Division Multiplexing
PCA	Principal Component Analysis
PCM	Pulse Code Modulation
PD	Predictive Decoder
PDE	Probabilistic Downsampling Encoder
pdf	Probability Density Function

PER	Packet Error Rate
PFA	Probability of False Alarm
PHY-Layer	Physical Layer
PoD	Probability of Detection
RIP	Restricted Isometric Property
RMT	Random Matrix Theory
s.t.	subject to
SCM	Sample Covariance Matrix
SD	Step Decoder
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access
WSN	Wireless Sensor Network

Index

Chapter 1

Introduction

1.1 Motivation

Wireless Sensor Networks (WSNs) have become one of the hottest research topics in the last ten years. They have been intensively and extensively studied from different disciplines such as information theory, signal processing, and communications among many others since WSNs collect a wide variety of challenging constraints and peculiarities. Hence, WSNs have been used as the ideal testbed scenario in order to apply and study the performance of different techniques under such features. In this dissertation, we mainly focus on the following ones.

1.1.1 Energy-limited network

A WSN is typically composed of many tiny, low-powered and inexpensive wireless sensing nodes to monitor a certain physical measurement, such as temperature, humidity, pressure, light, pollution, etc, and transmit their measurement wirelessly to a central entity that is usually called fusion center. In most cases, the sensing nodes are powered by small batteries that cannot be recharged due to practical reasons. Hence, they are strongly energy-limited and the lifetime of the WSN is strongly constrained to the

lifetime of the battery of the sensing nodes. Therefore, the usage of energy-efficient techniques is in many cases mandatory.

Following with this motivation, one can find many energy-efficient strategies in order to mitigate the energy costs and hence increase the lifetime of the WSN. Without the aim of being exhaustive, we point out some examples:

- Energy-aware routing for cooperative WSNs and ad-hoc networks [Toh01] [You04]. These techniques seek the optimum path that minimizes the total spent energy in multi-hop WSNs.

- Signal processing techniques for minimum-power distributed transmission schemes [Mud09] [Zar11]. Using distributed beamforming techniques, the nodes can decrease the transmitted power at the same time that they increase the total throughput of the network.

- Data-aware techniques to reduce energy by using efficient information processing [Pra02]. By means of signal processing techniques, the network exploits the inherent structure and properties of the measured signal in order to compress the data and therefore reduce the associated energy costs.

Our study falls in the third category and may be complementary to the other approaches.

1.1.2 Distributed network

The sensing nodes of the WSN are spatially distributed over a given area of coverage, where each one of the sensing nodes acts as an autonomous entity. This derives to a distributed configuration at the transmitted side. On the other hand, the receiver is typically one entity that gathers (and processes) the information coming from the sensing nodes.

Although a large number of works deal with cooperative communications, where the signaling among sensing nodes is totally allowed, we restrict as much as we can the communication among sensing nodes mainly because of energy constraints. It is clear that the more signaling load, the more number of transmissions and hence the more power power.

In order to be consequent with the energy constraints of WSNs exposed before, we are motivated to seek fully distributed architectures at the transmitter side (i.e., the set of sensing nodes), where the communication among them is minimized.

1.1.3 Space-time correlated sources

As we have already mentioned above, WSNs are typically composed of a large number of sensing nodes within a delimited area of coverage. Therefore, it is common to assume that the sensing nodes are spatially close enough that their measurements present some similarities with each other. The degree of similarity usually depends on the distance between the nodes.

Furthermore, the measured signal is usually slow-varying in time (e.g., the temperature of a field is expected to change slowly). In addition to that, the sampling frequency of the sensing nodes is high enough that the sample taken in one time instant retain some information from the past measurements and also gives information about future samples.

Accordingly, we assume that the measured signal presents correlations in both the spatial and the temporal domain. Therefore, the final aim of this dissertation is to exploit these characteristics of the signal in order to design and assess distributed and energy-efficient techniques for the WSN scenario.

1.2 Organization of Dissertation

1.2.1 Outline

The present dissertation is organized in six chapters. The outline is as follows.

- *Chapter 1* presents the motivation and the context of this research work. It also presents the outline of the dissertation where the relations and dependences among de chapters are specified.

- *Chapter 2* introduces a particular type of encoding-decoding strategies, the zero-delay downsampling transmissions. These strategies are studied for the particular case of autoregressive signal models, which conveniently describe physical measurements as the ones present in WSNs. In particular, we study the performance for both deterministic and statistic downsampling encoder-decoders. We obtain closed form expressions for the quadratic error of the deterministic and probabilistic encoders, while error approximations are derived for the conditional encoders.

- *Chapter 3* proposes two enhanced correlation estimators of the well-known Linear Wiener Filter (LWF) and its derived Mean Square Error (MSE) for the two key steps of a practical distributed source coding scheme. It is shown in Chapter 3 that the proposed enhanced estimators notably improve the classic sample estimations when the number of samples used in the training phase is comparable in magnitude to the dimension of the observation vector. On the other hand, when the number of samples is much higher than the observation dimension, our proposed enhanced estimators perform as the classical estimators.

- *Chapter 4* considers compressed sensing as a convenient signal processing tool in order to design energy-efficient communication schemes

in a distributed way. In this chapter, we propose a novel compressed sensing transmission scheme named *Amplify-and-Forward Compressed Sensing* (AF-CS) that significantly reduces the resource utilization and the energy consumption. We also propose a simple model that accurately approximates the reconstruction error introduced by the proposed scheme. Furthermore, we contribute to the compressed sensing literature with a new and tighter relation among the minimum number of measurements for a given dimension and sparsity level of the original vector. The analytical model and this new relation allows us to dimension the WSN (i.e., number of active sensing nodes and number of relays) based on a cost function that controls the trade-off between reconstruction error and energy consumption. We also show by simulation that the AF-CS outperforms other techniques in the literature in terms of distortion and energy-efficiency, providing at the same time, significant reduction in the number of channel uses.

- *Chapter 5* addresses the physical layer secrecy topic in wireless sensor networks. In particular, the AF-CS scheme is proposed as a secret scheme against passive eavesdropping. The secrecy performance of our proposed technique is studied in terms of perfect secrecy and wiretap distortion. Chapter 5 also proposes a channel estimation technique based on random pilots that allows the system to introduce extra uncertainty only in the channel estimation of the eavesdroppers. It is shown how this technique can dramatically increase the secrecy of AF-CS even for the case of several coordinated eavesdroppers listening at the same time.

- Finally, *Chapter 6* concludes the dissertation with a summary of the main contributions and listing the remaining tasks for future work.

1.2.2 Relations and dependences among chapters

Albeit each chapter has been written in the most self-contained way possible for clarity and readability, there exist strong relations and dependences among them. The existing relations and dependences are illustrated in Fig. 1.1 and they are properly referenced during the reading of the dissertation. With some abuse of language, we will say that a Chapter A depends on a Chapter B when the reading of Chapter B is *needed* in order to properly understand Chapter A. Moreover, we will say that a Chapter C is related to a Chapter D when some of the results in Chapter D can be used (but they are not necessary) in Chapter C.

The relations and dependences of Chapters 1, 6 with Chapters 2-5 are obvious since Chapters 1, 6 introduce and conclude Chapters 2-5. The rest are linked as follows.

- *Chapter 2* uses the LWF estimation in one of the proposed encoders and one of the proposed decoders. However, it assumes perfect knowledge of the correlation matrix. In a real transmission environment, this quantity needs to be estimated. Therefore, Chapter 2 is *related* to Chapter 3 since this last one proposes enhanced correlation estimators that can be used in Chapter 2 in order to estimate the linear Wiener filter and its related MSE with fewer samples than for the conventional sample estimators.

- *Chapter 3* follows the $AR-1$ signal model, which is detailed in Chapter 2. Chapter 3 also applies the convenient matrix notation of the $AR-1$ process introduced in Chapter 2. Hence, Chapter 3 is *related* to Chapter 2.

- *Chapter 4* presents the AF-CS. This scheme is based on the Conditional Downsampling Encoder (CDE) and the Predictive Decoder (PD), that are studied in detail in Chapter 2. Thus, *Chapter 4 dependends* on Chapter 2. Furthermore, both CDE and PD use the LWF

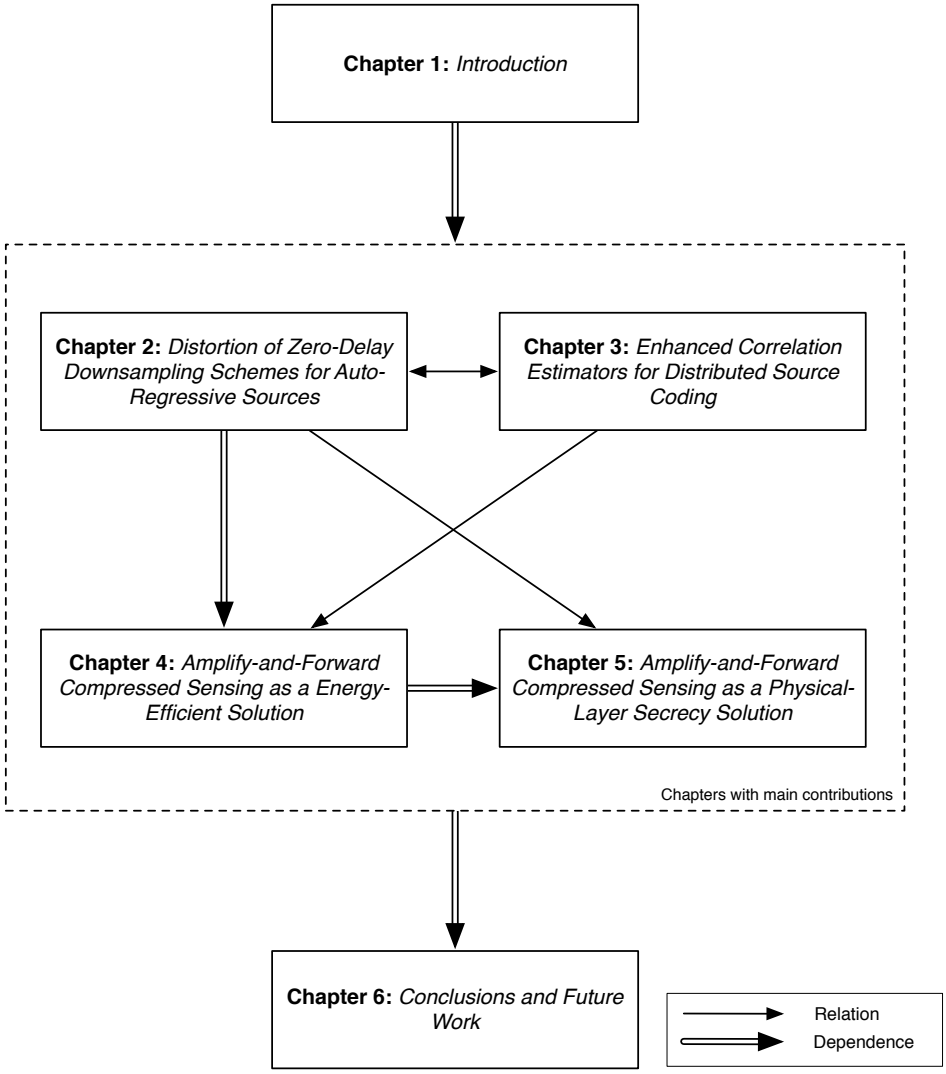

Figure 1.1: Relations and dependences among chapters.

as a predictor, and hence the estimators proposed in Chapter 3 can be used. In addition, Chapter 4 presents an analytical model that describes the performance of AF-CS in terms of quadratic distortion. Hence, the MSE estimator proposed in Chapter 3 can be used if the correlation parameters are estimated (instead of the true ones). We conclude that Chapter 4 is related to Chapter 3 as well.

- Finally, *Chapter 5* explores the physical-layer secrecy properties of the AF-CS introduced and detailed in Chapter 4. Hence, *Chapter 5* strongly *depends* on Chapter 4. Furthermore, the signal model follows the one in Chapter 2, and hence the reading of Chapter 2 is recommended (but not needed) before reading Chapter 5.

1.3 Research contributions

The related research contributions are concentrated in Chapters 2-5 and listed as follows:

Chapter 2 The results regarding to downsampling encoding can be found in the following Journal paper reference

> [BL12c] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Distortion of Zero-Delay Downsampling Schemes for Auto-Regressive Sources with Incomplete Observation Vectors,"* EURASIP Journal on Advances in Signal Processing (submitted June 2012).

Chapter 3 The derivation of the proposed enhanced correlation estimations has been introduced in the following conference paper and studied in detail in the following journal paper

> [BL10] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Distributed Source Coding for Large Wireless Sensor Networks,"* in Proc. 44th Asilomar Conf. on Signals, Systems and Computers, Nov 2010, Monterey, CA.

[BL12e] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Enhanced Correlation Estimators for Distributed Source Coding in Large Wireless Sensor Networks,"* IEEE Sensors Journal (accepted for publication).

Chapter 4 The AF-CS scheme has been introduced in the following conference paper and studied in detail in the following journal paper

[BL11] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Optimization of the Amplify-and-Forward Transmission in a Wireless Sensor Network Using Compressed Sensing,"* in Proc. 19th European Conference on Signal Processing (EUSIPCO), Aug 31 2011, Barcelona, Spain.

[BL12b] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Amplify-and-Forward Compressed Sensing as an Energy-efficient Solution for Wireless Sensor Networks,"* ACM Transactions on Sensor Networks, (submitted March 2012).

Chapter 5 The PHY-Layer secrecy properties of the AF-CS scheme have been introduced in the following conference paper and detailed in the following journal paper

[BL12a] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Amplify-and-Forward Compressed Sensing as a PHY-Layer Secrecy Solution for Wireless Sensor Networks,"* in Proc. 7th Sensor Array and Multichannel Signal Processing (SAM 2012), 2012, New Jersey, US.

[BL12d] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, *"Amplify-and-Forward Compressed Sensing as a Physical Layer Secrecy Solution for Wireless Sensor Networks,"* IEEE Transactions on Information Forensics and Security, (submitted July 2012).

Chapter 2

Distortion of Zero-Delay Downsampling for Auto-Regressive Sources

2.1 Summary

In this chapter, we assess the performance of various zero-delay encoding-decoding strategies. The proposed analytical study addresses the performance degradation at the receiver side when some of the samples are missing or have been removed at a given ratio. Concretely, a Gaussian autoregressive signal model is considered. Hence, the analyzed encoder-decoder pairs exploit their knowledge about the signal structure in different ways. In particular, we study the performance in terms of quadratic distortion of three downsampling encoders, which are the deterministic downsampling encoder (DDE), the probabilistic downsampling encoder (PDE), and the conditional downsampling encoder (CDE), combined with two decoders: the step decoder (SD) and the predictive decoder (PD). We derive closed form expressions of the quadratic distortion for the pairs DDE-SD, DDE-PD, PDE-SD and PDE-PD. For the case of CDE-SD and CDE-PD we

derive closed approximations for their quadratic distortion. Alse, we propose two strategies in order to design the threshold of the condition of the CDE that decides whether a sample is transmitted or not at a given time instant. Numerical simulation validates our analytical results, and the accuracy of the approximations regarding CDE-SD and CDE-PD. Moreover, we compare the obtained quadratic distortion and extract the conclusions of the capabilities of the studied encoding-decoding schemes.

2.2 Introduction

2.2.1 Motivation and previous work

With the emergence of increasingly heterogeneous communication technologies, the encoding-decoding schemes have been diversified in order to satisfy the constraints for a given system. This makes it impossible to obtain a single optimal encoding-decoding solution valid for all transmission schemes. Instead, one of the most important steps in the design of a transmission scheme is the selection of the encoder-decoder pair that meets the system requirements.

Many transmission schemes use non-causal transmissions such as block-coding. In these cases, the source collects a number of contiguous samples in order to compress them by removing part of (or all) the redundancy among them. Within this group of encoders-decoders a large amount of different techniques can be found. Albeit these transmissions are very appropriate for high-rate transmissions and/or delay-tolerant communications, these non-causal transmission schemes may not be applicable because block transmissions are not always allowed due to delay constraints and/or low symbol rates of the source.

For delay sensitive applications such as real-time monitoring applications, where the reconstruction of the signal must take place at the same time instant as the corresponding input sample, causal source codes are

more convenient. Hence, a source code is said to be causal if the nth decoded sample depends on the output signal only through its first n components, or in other worlds, depends on the past and present outputs but not on future ones. Quantizers, delta modulators, differential pulse code modulators, and adaptive versions of these are all causal in the above sense. The basic properties of causal source codes have been introduced in 1982 in [Neu82], and related works have been expanded so far. The work in [Wei05] extends the general results of [Neu82] for the case where *side information*, i.e. extra information that is correlated with the source, is available at the encoder and the decoder.

In addition, a causal source code is called zero-delay or sequential code if both the encoder and the decoder are causal (note that for the causal source code definition, the assumption of causality is only at the decoder) [Der12], [Vis00]. Some applications of zero-delay schemes can be found in the speech coding or the encoding of a certain physical phenomenon that is monitored in real time, such as in wireless sensor networks.

In the literature, there are several zero-delay coding systems. One of the most common is the well-known Differential Pulse Code Modulation (DPCM). In a nutshell, the current sample to be coded is predicted from previously coded samples. This prediction is used as a reference and it is compared with the current sample. Hence, the output of the encoder is the prediction error. The inverse operation takes place at the decoder side. According to [Zam08], DPCM was first introduced in a U.S. patent by C. Cutler in 1952. Since then, many results have been appeared. In particular, the autoregressive (AR) model has received a special attention for the study of the zero-delay coding schemes. Some of the early works on AR models date back to the 60s. The works in [O'N71a] and [Pro67] analyze the quadratic rate distortion of DPCM (the reader can find an extended description of the rate distortion in Chapter 13 of [MC91]). The work in [O'N71b] extends this results assuming a Gaussian distribution

of the predicted error. Other works proposed algorithms for nonuniform quantizers optimized in order to minimize the distortion rate [Far85].

Later works, as the one in [Gul01], try to particularize the results obtained by the DPCM also for the case of low bit rates. In such cases the system performance becomes worse. Then, the classic DPCM encoder is modified in order to achieve better performance in terms of rate distortion for a low bit rate regime.

Recent works on this field have tried to unify the theoretical limits of the DPCM (and other zero-delay schemes) for AR models with other information theory concepts. The authors in [Zam08] provide analytical results for the existing duality between the rate distortion of an AR process with the capacity of the inter-symbol interference channels.

By contrast, other works as [Der12] adjust upper and lower limits of the rate distortion for generic zero-delay schemes from a information theoretical point of view using the mutual information as a measure of the rate.

2.2.2 Our contribution

Our proposed work also follows the same sequential transmission approach presented above. In particular, we study downsampling encoding-decoding schemes in which the samples of an input signal are either blocked or transmitted following a given criterion. We analyze the performance loss of different encoding-decoding pairs when the number of samples is reduced by a factor γ. We study the following three downsampling criteria: *i*) a Deterministic Downsampling Encoder (DDE), a Probabilistic Downsampling Encoder (PDE), and a Conditional Downsampling Encoder (CDE).

A DDE works as a decimator, i.e., it reduces the number of samples following a deterministic pattern. Hence, the DDE selects only one in γ^{-1} samples, where γ^{-1} is typically a natural number.

A PDE slightly differs from a common decimator since it reduces the number of samples following a probabilistic pattern, i.e., one sample will

be transmitted with probability γ and otherwise blocked with probability $1-\gamma$. This method eliminates the restriction of γ^{-1} to be a natural number. However, we analytically show that a DDE outperforms a PDE in terms of quadratic distortion.

A CDE uses side information in order to sequentially elaborate the decimator pattern. Basically, it predicts the current sample using linear estimation and uses this prediction as a reference. Then, the transmission is blocked if the prediction error does not exceed a given threshold, and transmitted otherwise. It is clear that a key step of the CDE design is to determine the threshold that ensures a sample rate reduction of a factor γ. Therefore, two different threshold designs are proposed in this chapter.

Clearly, this last encoder presents some similarities with the DPCM in the sense that both schemes use (linear) prediction as a reference in order to encode the input signal. However, they present important differences as well, which can be summarized as:

- A DPCM produces an outcome sample for each input sample. In other words, it does not change the sample rate. On the contrary, the CDE (and also the DDE and the PDE) reduces the sample rate. This behavior is very convenient in some energy-constrained scenarios, such as wireless sensor networks, since the total number of transmissions is reduced by a factor γ, increasing the energy efficiency of the network.

- While a DPCM works at symbol level, the CDE does at sample level. Thus, the downsampling encoder-decoder schemes studied in this paper are not exclusive to the DPCM or other zero-delay coding techniques. Actually, they can be used on top of them when the signal is transmitted.

On the other hand, the decoder recovers the original sampling rate by upsampling the signal. We study two possible decoders: *i*) a Step Decoder (SD), and a Predictive Decoder (PD). A SD reconstructs the missing sam-

ples by replicating the last decoded samples. This do not require any side information knowledge. On the contrary, the PD reconstructs the missing samples by linear prediction (as in the CDE case). We analytically show the improvements in terms of quadratic distortion when the samples have been predicted rather than simply replicated.

We give analytical expressions for the quadratic distortion of the following downsampling encoding-decoding pairs: DDE-SD, DDE-PD, PDE-SD, and PDE-SD. Furthermore, we also provide accurate approximations for the quadratic distortion of CDE-SD and CDE-PD. Numerical simulations support our proposed analytical expressions.

In order to analytically address this problem, we focus in this chapter on an AR model. This probabilistic model is particularly interesting for both its analytical simplicity and its applicability in modeling real physical sources [Has80].

2.2.3 Organization of the chapter

The rest of the chapter is organized as follows. In Section 5.3 we introduce the problem statement and the scenario considered throughout the chapter. Section 2.4 presents the encoding-decoding schemes under study. The analytical expressions of the downsampling distortion for each encoder-decoder pair are detailed in Section 2.5. Two different design strategies for the case of conditional encoding and their performance are presented in Section 2.6. Simulation results are shown in Section 5.6. Conclusions and suggestions for future research are drawn in Section 5.7.

2.3 System Model and Assumptions

The transmission model under consideration is illustrated in Fig. 5.3.

Let $x(n)$ be a real and time-discrete auto-regressive (AR) model of order

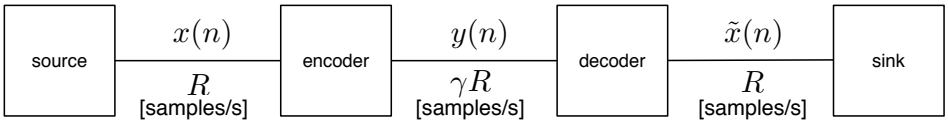

Figure 2.1: Block diagram of a generic encoder-decoder transmission scheme. The source is generating the desired signal $x(n)$ at a rate R samples/s. The signal $y(n)$ is the encoded version of $x(n)$ and it is transmitted at a rate γR, where $\gamma \leq 1$. The signal $\tilde{x}(n)$ is the reconstruction of the desired signal $x(n)$ after the decoder.

N and variance σ_x^2, sampled at a rate $\mathcal{R}$. It is defined as

$$x(n) = \mathbf{w}^T \mathbf{x}(n) + z(n), \text{ for } n = 1, 2, \ldots \tag{2.1}$$

The vector $\mathbf{x}(n) \in \mathbb{R}^N$, stacks the previous N samples of the stochastic process, i.e., $\mathbf{x}(n) = [x(n-1)\, x(n-2) \ldots x(n-N)]^T$, the *auto-regression coefficients* are denoted by the vector $\mathbf{w} \in \mathbb{R}^N$ and assumed constant during the transmission. The vector $z(n)$ is a Gaussian process with zero mean and variance σ_z^2. Hence, note that $\mathbf{x}(n)$ has also zero mean and it is uncorrelated with $z(n)$ but not with $z(n-i)$ for $i = 1, 2, ...$

We assume $x(n)$ to be a continuous-valued process. Although a continuous random measurement requires infinite precision in order to be digitally sent with zero error [MC91], we assume that the quantization error to be negligible in comparison with the quadratic distortion of the downsampling encoder-decoder pair. This assumption is supported by the rate-distortion theory, since we can select a symbol rate such as the quantization error would be as small as we want.

Furthermore, we require that the signal $x(n)$ is transmitted in a zero-delay manner from the source to the destination. Throughout this chapter, we understand for zero-delay transmission when for each sample at time n the receiver will have a reconstruction of the signal $x(n)$. Furthermore, for

time instant n we are not interested in $x(n-1)$ anymore, so delay tolerant strategies (such as block encoding schemes) are not feasible. Following this constraint, we will look for encoders that allows us to reduce the sample rate sample-by-sample in real time.

Hence, we consider a non-linear encoder with a coding rate γ. In our particular case, the encoder selects which samples from $x(n)$ are going to be transmitted with a rate of γ and the rest will be discarded. The selected samples are represented by $y(n)$, therefore, note that $y(n)$ is only defined for those time slots in which the encoder decides to transmit. In the following Section, we will revise different downsampling encoding criteria and we study their performance.

Moreover, we consider non-linear decoders in order to recover an approximation of $x(n)$, i.e. $\tilde{x}(n)$, from $y(n)$. Roughly speaking, the decoder will construct $\tilde{x}(n)$ copying the samples of $y(n)$ when the transmission exists and predicting the rest otherwise.

Definition 2.1 *For a given pair of encoder-decoder, the sink will receive $\tilde{x}(n)$ with a given* downsampling distortion*. It defines the quadratic distortion introduced by the given downsampling encoder-decoder pair e-d, as*

$$\mathcal{D}(e,d) = \mathbb{E}[(x(n) - \tilde{x}(n))^2]. \tag{2.2}$$

In the next sections we present the different encoder-decoder pairs and we evaluate their downsampling distortion.

2.4 Zero-Delay Transmission Schemes and Their Downsampling Distortion

2.4.1 Available information about the desired signal $x(n)$

Let $\mathbf{R} \in \mathbb{R}^{N\times N}$ be the covariance matrix of the observation vector $\mathbf{x}(n)$, defined as

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & \cdots & r_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N-1} & r_{N-2} & \cdots & r_0 \end{bmatrix}, \tag{2.3}$$

where $r_i = \mathbb{E}[x(n)x(n-i)]$. This matrix is assumed to be known since it can be efficiently estimated after a training phase of M samples as

$$\mathbf{R} = \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{M} \mathbf{x}(n)\mathbf{x}^T(n). \tag{2.4}$$

There are many ways to estimate the coefficients of an AR process. We select the Linear Wiener Filter (LWF) solution since it is optimal in terms of the Minimum Square Error (MSE). The LWF solution, $\mathbf{w}$ satisfies the Wiener-Hopf equations [Hay01]:

$$\mathbf{r} = \mathbf{R}\mathbf{w} \quad \Rightarrow \quad \mathbf{w} = \mathbf{R}^{-1}\mathbf{r}, \tag{2.5}$$

where $\mathbf{r} = [r_1\, r_2\, \ldots\, r_N]^T \in \mathbb{R}^N$ is the cross-correlation vector defined as $\mathbf{r} = \mathbb{E}[x(n)\mathbf{x}(n)]$.

Hence, one can predict $x(n)$ using the LWF with a given observation vector $\tilde{\mathbf{x}}(n) = [\tilde{x}(n-1)\ \tilde{x}(n-2)\ \ldots\ \tilde{x}(n-N)]$ as

$$\hat{x}(n) = \mathbf{w}^T\tilde{\mathbf{x}}(n) = (\mathbf{R}^{-1}\mathbf{r})^T\tilde{\mathbf{x}}(n). \tag{2.6}$$

where $\tilde{x}(n)$ is the decoded sample at time instant n.

For the case that the observation vector $\tilde{\mathbf{x}}(n)$ is $\mathbf{x}(n)$, the MSE in the prediction of an AR process for an arbitrary $\mathbf{w}$ is:

$$\begin{aligned} \text{MSE}(\mathbf{w}) &= \mathbb{E}[(x(n)-\hat{x}(n))^2] = \mathbb{E}[x(n)^2] - \mathbb{E}[x(n)\hat{x}(n)] + \mathbb{E}[\hat{x}(n)^2], \\ &= \sigma_x^2 - 2\mathbf{w}^T\mathbf{r} + \mathbf{w}^T\mathbf{R}\mathbf{w}. \end{aligned} \tag{2.7}$$

Lemma 2.1 *The MSE lower-bound MSE*$(\mathbf{w}) = \sigma_z^2$ *can be achieved by the LWF solution.*

Proof The MSE for the LWF solution, i.e. $\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}$, is

$$\text{MSE}(\mathbf{R}^{-1}\mathbf{r}) = \sigma_x^2 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r}. \tag{2.8}$$

Following the initial model in (2.1), we represent σ_x^2 in terms of $\mathbf{R}$, $\mathbf{r}$ and $\mathbf{w}$, and we get

$$\sigma_x^2 = \mathbb{E}[x(n)^2] = \mathbb{E}[(\mathbf{w}^T\mathbf{x}(n) + z(n))^2]. \tag{2.9}$$

Since $z(n)$ is uncorrelated with $\mathbf{x}(n)$, we obtain

$$\begin{aligned} \sigma_x^2 &= \mathbf{w}^T\mathbb{E}[\mathbf{x}(n)^T\mathbf{x}(n)]\mathbf{w} + \mathbb{E}[z(n)^2] \\ &= \mathbf{w}^T\mathbf{R}\mathbf{w} + \sigma_z^2. \end{aligned} \tag{2.10}$$

Substituting the result in (2.10) in (2.8), and taking into account that using the LWF solution in (2.5) the term $\mathbf{w}^T\mathbf{R}\mathbf{w} = \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r}$, we get the well-known result

$$\text{MSE}(\mathbf{R}^{-1}\mathbf{r}) = \mathbf{w}^T\mathbf{R}\mathbf{w} + \sigma_z^2 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} = \sigma_z^2 \tag{2.11}$$

2.4.2 The Auto-Regressive model of order 1, $AR-1$

In signal processing it is usual to assume time correlated signals in order to model real sources. One of the correlation models commonly used is $[\mathbf{R}]_{n,n-i} = r_i = \rho^i$, where $\rho \in [0,1]$ is the correlation factor. One example

is [Ram10] where this model is used to represent the spatial correlation among sensors in a Wireless Sensor Network monitoring light, temperature and humidity, and compares the obtained results for a correlation factor of $\rho = 0.95$ with the ones obtained from a real environment.

This correlation model turns out to be an auto-regressive model of order 1 ($AR{-}1$). It is easy to observe that for $[\mathbf{R}]_{n,n-i} = r_i = \rho^i$ the autoregressive coefficients obtained from the LWF solution in (2.5) are of the form $\mathbf{w} = [\rho\ 0\ 0\ \dots\ ]^T$. Hence, the AR model can be written as an $AR-1$ as

$$x(n) = \mathbf{w}^T\mathbf{x}(n) + z(n) = \rho\, x(n-1) + z(n). \tag{2.12}$$

Without loss of generality, we assume that $\sigma_x^2 = 1$. Therefore, the MSE obtained using the LWF solution is:

$$\begin{aligned} \mathrm{MSE}(\mathbf{w}) &= \mathbb{E}[|x(n) - \rho x(n-1)|^2] \\ &= 1 - 2\rho\mathbb{E}[x(n)x(n-1)] + \rho^2\mathbb{E}[x(n-1)x(n-1)] \\ &= 1 - 2\rho^2 + \rho^2 = 1 - \rho^2. \end{aligned} \tag{2.13}$$

It means that we can predict our signal of interest $x(n)$ with an error bound of $1 - \rho^2$ for the case we have perfect knowledge of the correlation parameters $\mathbf{R}$ and $\mathbf{r}$ and $x(n)$ models an $AR-1$ process. This function is illustrated in Fig. 2.2.

2.4.3 Different encoding alternatives

We select three downsampling encoders among many other possibilities. They have been chosen for their simplicity and because many other strategies can be derived from them.

In order to describe the selected encoders, we first need to introduce the following definition:

Definition 2.2 *The* transmission support function *of an encoder e, named $g_e(n)$, is an indicator function which takes the value one when the transmission exists and zero otherwise.*

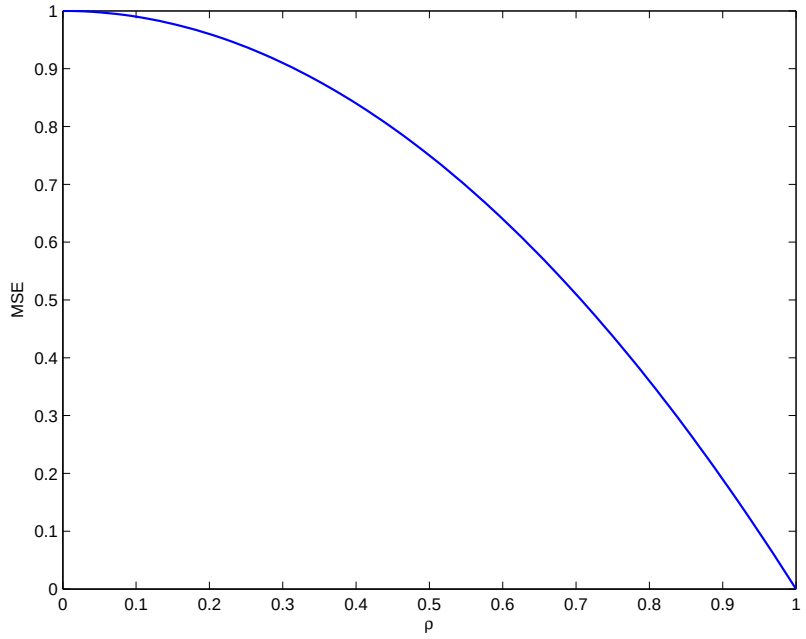

Figure 2.2: MSE lower-bound in the prediction of a $AR-1$ process as a function of the AR coefficient.

2.4.3.1 Deterministic Downsampling Encoder (DDE)

This encoder is the simplest one and acts as a decimator. Its transmission support function is:

$$g_{\mathrm{DDE}}(n) = \begin{cases} 1 & \text{when} \quad n \bmod \gamma^{-1} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

Note that for uniform downsampling, the DDE is only defined for compression rates γ of the form $\gamma^{-1} \in \mathbb{N}$.

2.4.3.2 Probabilistic Downsampling Encoder (PDE)

This encoder solves the limitation of DDE that γ^{-1} is a natural number. Basically, the symbol $x(n)$ will be transmitted following a given probabilistic pattern. Thus, the transmission support function is:

$$g_{\text{PDE}}(n) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases} \tag{2.15}$$

It is straightforward to see that in order to guarantee a compression rate of γ, the value of the transmission probability p should be $p = \gamma$.

2.4.3.3 Conditional Downsampling Encoder (CDE)

Previous encoders do not assume any memory or prior information of the signal of interest $x(n)$. On the contrary, the CDE uses the available information in order to decide whether the signal should be transmitted or not. In particular, we analyze the cases where the available information is either the last decoded sample $\tilde{x}(n-1)$ or a linear prediction using the LWF solution in (2.5) with a given observation vector $\tilde{\mathbf{x}}(n)$. The available information is compared with the signal of interest $x(n)$. If the absolute value of the difference is higher than a given threshold Δ, the encoder will transmit the signal. Otherwise, if the difference is below Δ, the transmission is blocked. Mathematically, for the first case,

$$g_{\text{CDE}}(n) = \begin{cases} 1 & \text{if } |x(n) - \tilde{x}(n-1)| > \Delta \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

For the LWF prediction, the CDE is

$$g_{\text{CDE}}(n) = \begin{cases} 1 & \text{if } |x(n) - \hat{x}(n)| > \Delta \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

Although this scheme is quite simple, it has two main complications: *i*) the LWF predictor assumes the knowledge of the correlation parameters $\mathbf{R}$

and $\mathbf{r}$ or at least good estimates of them, and *ii*) the threshold Δ should be designed in such a way that it ensure a coding rate of γ. The first problem adds some complexity to the system but can be efficiently solved using existing correlation estimators [BL12e] (as it is detailed in Chapter 3). The second one is addressed later in Section 2.5.

2.4.4 Different decoding alternatives

As for the encoding strategies, we select two decoders from a bunch of possible solutions. The first one is probably the simplest and it does not require any knowledge of the correlation parameters, while the second uses this knowledge in order to reconstruct the received signal.

2.4.4.1 Step Decoder (SD)

It is the simplest decoder. It just copies the value of $y(n)$ into $\tilde{x}(n)$ when $g_e(n) = 1$ or maintains the last decoded value $\tilde{x}(n-1)$ if $g_e(n) = 0$. The decoder function is described as

$$d_{\mathrm{SD}}(n) = \begin{cases} \tilde{x}(n) = y(n) & \text{if } g_e(n) = 1 \\ \tilde{x}(n) = \tilde{x}(n-1) & \text{otherwise.} \end{cases} \tag{2.18}$$

This approach is very typical when the source is sensing a given time-correlated phenomena. Since it is assumed slow changing, the magnitude is maintained until we receive an update.

2.4.4.2 Predictive Decoder (PD)

If we take advantage of the time correlation properties of $x(n)$, we can obtain lower downsampling distortion than for the SD case. The behavior is similar than the previous decoder SD, but in this case, when $g_e(n) = 0$, the PD predicts $x(n)$ using LWF instead of replicating $\tilde{x}(n)$. Mathematically,

$$d_{\mathrm{PD}}(n) = \begin{cases} \tilde{x}(n) = y(n) & \text{if } g_e(n) = 1 \\ \tilde{x}(n) = \hat{x}(n) & \text{otherwise.} \end{cases} \tag{2.19}$$

2.5 Downsampling Distortion of the Encoder-Decoder Pairs

2.5.1 Signal prediction using incomplete observation vectors

Let the observation vector $\tilde{\mathbf{x}}(n) \in \mathbb{R}^N$ is $\tilde{\mathbf{x}}(n) = [\tilde{x}(n-1)\, \tilde{x}(n-1) \cdots\ \tilde{x}(n-N)]^T$ be an incomplete version of $\mathbf{x}(n)$. The vector $\tilde{\mathbf{x}}(n)$ is constructed using the N last decoded samples. This is because the decoder does not necessarily know all the values of $\mathbf{x}(n)$ and only knows the decoded ones. Hence, some values of $\tilde{\mathbf{x}}(n)$ are replicas of $\mathbf{x}(n)$ and the rest are predicted values $\hat{x}(n)$.

Definition 2.3 *Let the vector $\tilde{\mathbf{x}}_t$ be an instance of $\tilde{\mathbf{x}}(n)$ where the last true sample was received at time $n - t$. Mathematically,*

$$[\tilde{\mathbf{x}}_t(n)]_j = \begin{cases} \hat{x}(n-j) & \text{if } j < t \\ x(n-j) & \text{if } j = t \end{cases} \tag{2.20}$$

Theorem 2.1 *If $\tilde{\mathbf{x}}_t(n)$ is used as the observation vector of the LWF, the MSE is degraded as*

$$MSE_t = 1 - \rho^{2t} \tag{2.21}$$

Proof It is proved by induction. First let us assume the case where the vector $\tilde{\mathbf{x}}_2(n)$ is of the form $\tilde{\mathbf{x}}_2(n) = [\hat{x}(n-1)\ x(n-2) \dots x(n-N)]^T$, that is, all the positions in the vector correspond to true measurements except

the first one. In this case,

$$\begin{aligned}\mathbb{E}[|x(n)-\mathbf{w}^H\tilde{\mathbf{x}}_2(n)|^2] &= \mathbb{E}[|x(n)-\rho\hat{x}(n-1)|^2]\\ &= \mathbb{E}[|x(n)-\rho\mathbf{w}^H\tilde{\mathbf{x}}_1(n-1)|^2]\\ &= \mathbb{E}[|x(n)-\rho^2x(n-2)|^2]\\ &= 1-2\rho^2\mathbb{E}[x(n)x(n-2)]+\rho^4\mathbb{E}[x(n-2)x(n-2)]\\ &= 1-\rho^4. \end{aligned} \tag{2.22}$$

For the case where $\tilde{\mathbf{x}}_3(n)$ is of the form $\tilde{\mathbf{x}}_3(n) = [\hat{x}(n-1)\ \hat{x}(n-2)\ x(n-3)\dots x(n-N)]^T$, the MSE is degraded as

$$\begin{aligned}\mathbb{E}[|x(n)-\mathbf{w}^H\tilde{\mathbf{x}}_3(n)|^2] &= \mathbb{E}[|x(n)-\rho\hat{x}(n-1)|^2]\\ &= \mathbb{E}[|x(n)-\rho\mathbf{w}^H\tilde{\mathbf{x}}_2(n-1)|^2]\\ &= \mathbb{E}[|x(n)-\rho^2\hat{x}(n-2)|^2]\\ &= \mathbb{E}[|x(n)-\rho^2\mathbf{w}^H\tilde{\mathbf{x}}_1(n-2)|^2]\\ &= \mathbb{E}[|x(n)-\rho^3x(n-3)|^2]\\ &= 1-2\rho^3\mathbb{E}[x(n)x(n-3)]+\rho^6\mathbb{E}[x(n-3)x(n-3)]\\ &= 1-2\rho^6+\rho^6 = 1-\rho^6. \end{aligned} \tag{2.23}$$

It is straightforward to conclude that, for the general case where $\tilde{\mathbf{x}}_t(n)$ is of the form $\tilde{\mathbf{x}}_t(n) = [\hat{x}(n-1)\dots\hat{x}(n-t+1)\,x(n-t)\dots x(n-N)]^T$, the MSE is degraded as

$$\mathbb{E}[|x(n)-\mathbf{w}^H\tilde{\mathbf{x}}_t(n)|^2] = 1-\rho^{2t}. \tag{2.24}$$

Corollary 2.1 *For a given ρ, the MSE is only a function of the position of the last true measurement in the observation vector for an $AR-1$ process. Furthermore, it is not dependent on the dimension N of $\tilde{\mathbf{x}}_t(n)$.*

Proof The proof of the first statement is straightforward and it is enough to verify that the MSE obtained by $\tilde{\mathbf{x}}_t(n)$ and $\tilde{\mathbf{x}}'_t(n)$, where

$$\tilde{\mathbf{x}}'_t(n) = [\hat{x}(n-1)\dots x(n-t)\dots\hat{x}(n-N)]^T, \tag{2.25}$$

is the same. Then, let us consider for example $t = 2$,

$$\begin{aligned}\mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_2'(n)|^2] &= \mathbb{E}[|x(n) - \rho\hat{x}(n-1)|^2] \\ &= \mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_2(n)|^2] = 1 - \rho^4. \quad (2.26)\end{aligned}$$

Moreover, for observation vectors that only contain estimated measures (i.e., $t > N$), the MSE also follows (2.21). One can see that if $t = N + 1$, then the MSE is:

$$\begin{aligned}\mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_{N+1}(n)|^2] &= \mathbb{E}[|x(n) - \rho^T \hat{x}(n-N)|^2] \\ &= \mathbb{E}[|x(n) - \rho^N \mathbf{w}^H \tilde{\mathbf{x}}_1(n-N)|^2] \\ &= \mathbb{E}[|x(n) - \rho^{N+1} x(n-N-1)|^2] \\ &= 1 - \rho^{2(N+1)}. \quad (2.27)\end{aligned}$$

Hence, the probability that the last true sample of the vector $\tilde{\mathbf{x}}(n)$ is in the position t depends directly on the downsampling criteria used at the encoder. Therefore, in order to compute the downsampling distortion for a given encoder-decoder pair, we need to compute the probability of occurrence of the event t, or what is the same, the probability that the observation vector $\tilde{\mathbf{x}}$ is actually $\tilde{\mathbf{x}}_t$. Next, we illustrate this problem using a Markov Chain (MC) model.

2.5.2 The Markov Chain solution for the incomplete observation vector case

Let a MC model a discrete-time process where a random variable $E(n)$ is changing in time. The MCs have the property that to be in a state t, i.e. $E(n) = t$, only depends on the previous state, i.e., $E(n-1)$. This property is very interesting in order to model $AR-1$ processes. Moreover, a MC is said to be homogeneous when the probability of transition between the states of $E(n)$ is invariant in time, i.e.,

$$p_{i,j} = P(E(n) = j | E(n-1) = i) \in [0, 1]. \quad (2.28)$$

where $i, j = 0, 1, \ldots, T-1$. For a given i, the $p_{i,j}$ follows a given distribution probability, and hence

$$\sum_{j=0}^{T-1} p_{i,j} = 1, \tag{2.29}$$

since in any step, $E(n-1) = i$ can change to any $E(n) = j$ with a given probability $p_{i,j} \in [0, 1]$ and they are mutually exclusive. Let us introduce the following two definitions.

Definition 2.4 *Let the matrix $\mathbf{T} \in \mathbb{R}^{T \times T}$ denote the transition matrix of an homogeneous MC process of T states where*

$$\mathbf{T} = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,T-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{T-1,0} & p_{T-1,2} & \cdots & p_{T-1,T-1} \end{bmatrix}, \tag{2.30}$$

and each row represents a probability distribution as in (2.29), so $[\mathbf{T}^T]_i \mathbf{1} = 1$.

Definition 2.5 *Let the vector $\mathbf{p} \in \mathbb{R}^T$ denote the stationary probability vector of an homogeneous MC process of T states any vector that holds the stationary condition*

$$\mathbf{p}^T = \mathbf{p}^T \mathbf{T}, \text{ and } \mathbf{p}^T \mathbf{1} = 1 \tag{2.31}$$

where $\mathbf{p} = [P_0\, P_1\, \ldots\, P_{T-1}]^T$ contains the probabilities to be in each state $t = 0, 1, \ldots, T$ in the stationary regime of the MC process.

2.5.3 The downsampling distortion of the encoder-decoder pairs

In this section we analytically evaluate the performance of the proposed encoder-decoder pairs in terms of the downsampling distortion.

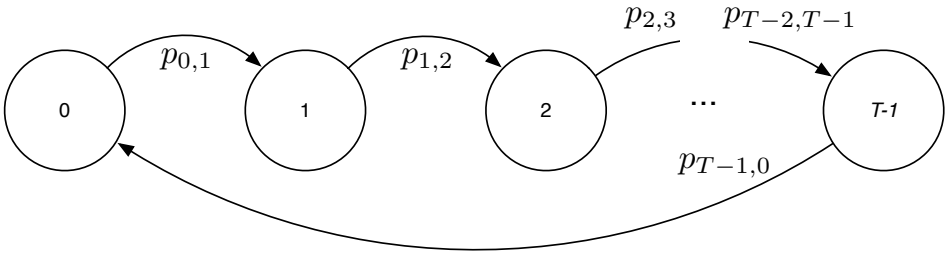

Figure 2.3: Finite Markov Chain of T states that models the encoder DDE.

2.5.3.1 The pair DDE-SD

The DDE can be modeled following the finite Markov Chain of T states in Fig. 2.3. The state $E(n) = 0$ means that in time n the transmission exists. Similarly, the state $E(n) = t$ means that the sample $n-t$ was the last to be transmitted. The transition matrix that describes the process of the DDE is:

$$\mathbf{T}_{\text{DDE}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{2.32}$$

where the first diagonal above the main diagonal and the position $(1, T)$ are loaded with ones while the rest are zeros.

Lemma 2.2 *The DDE-SD pair introduces at the state t the following error:*

$$MSE_t^{DDE\text{-}SD} = 2\left(1 - \rho^t\right). \tag{2.33}$$

Proof The index t denotes the time spacing between the last available sample with the current one. Thus, we can compute the $\text{MSE}_t^{\text{DDE-SD}}$ as

$$\begin{aligned}\text{MSE}_t^{\text{DDE-SD}} &= \mathbb{E}[(x(n)-x(n-t))^2] \\ &= \mathbb{E}[(x(n))^2] - 2\mathbb{E}[(x(n)x(n-t))] + \mathbb{E}[(x(n-t))^2] \\ &= 1 - 2\rho^t + 1 = 2\left(1-\rho^t\right). \end{aligned} \tag{2.34}$$

Theorem 2.2 *The downsampling distortion for the DDE-SD pair is:*

$$\mathcal{D}(DDE,SD) = 2 - 2\gamma\frac{\rho^{1/\gamma}-1}{\rho-1}. \tag{2.35}$$

Proof The downsampling distortion will be the sum of the MSE contributions for each state. Applying the definition of stationary probability vector in Definition 2.5 we extract that $P_i = P_j$ for all $i, j = 0, 1, \ldots, T$. Since we impose a coding rate of γ, the probability of transmission, i.e. P_0, is $P_0 = 1/T = \gamma$. The stationary probability vector is $\mathbf{p} = \gamma\mathbf{1}$. Hence, the downsampling distortion can be computed as

$$\begin{aligned}\mathcal{D}(\text{DDE,SD}) &= \sum_{t=0}^{T-1} P_t \text{MSE}_t^{\text{DDE-SD}} = \frac{2}{T}\sum_{t=0}^{T-1}(1-\rho^t) = 2 - \frac{2}{T}\sum_{t=0}^{T-1}\rho^t \\ &= 2 - \frac{2}{T}\frac{\rho^T-1}{\rho-1} = 2 - 2\gamma\frac{\rho^{1/\gamma}-1}{\rho-1}. \end{aligned} \tag{2.36}$$

2.5.3.2 The pair DDE-PD

The MC in Fig. 2.3 also models the behavior of the DDE-PD pair. However, the knowledge of the correlation parameters are available at the PD and hence it can predict the non-transmitted samples using the LWF. Therefore the MSE associated to the state t obeys Theorem 2.1.

Theorem 2.3 *The downsampling distortion for the DDE-PD pair is:*

$$\mathcal{D}(DDE,PD) = 1 - \gamma\frac{\rho^{2/\gamma}-1}{\rho^2-1}. \tag{2.37}$$

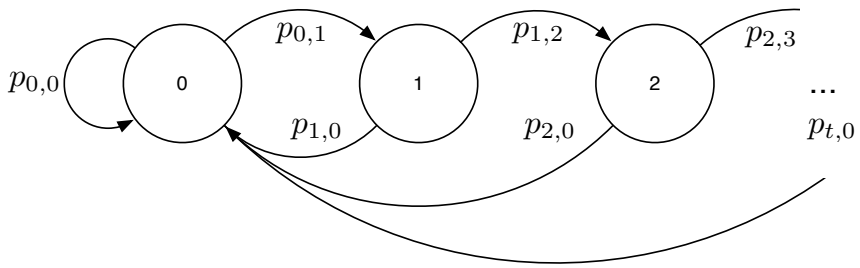

Figure 2.4: Infinite Markov Chain that models the encoder PDE.

Proof Following Theorem 2.1, the $\text{MSE}_t^{\text{DDE-PD}} = 1-\rho^{2t}$. Hence, the downsampling distortion can be computed as

$$\begin{aligned}\mathcal{D}(\text{DDE,PD}) &= \sum_{t=0}^{T-1} P_t \text{MSE}_t^{\text{DDE-PD}} = \frac{1}{T}\sum_{t=0}^{T-1}(1-\rho^{2t}) = 1 - \frac{1}{T}\sum_{t=0}^{T-1}\rho^{2t} \\ &= 1 - \frac{1}{T}\frac{\rho^{2T}-1}{\rho^2-1} = 1 - \gamma\frac{\rho^{2/\gamma}-1}{\rho^2-1}. \end{aligned} \tag{2.38}$$

2.5.3.3 The pair PDE-SD

The PDE can be modeled following the infinite MC in Fig. 2.4. As defined before, the state $E(n) = 0$ means that the transmission exists in time n. Similarly, the state $E(n) = t$ for $t \neq 0$, means that the sample $n - t$ was the last to be transmitted. The transition matrix (with dimension $T \to \infty$)

that describes the process of the PDE is:

$$\mathbf{T}_{\text{PDE}} = \begin{bmatrix} p_{0,0} & p_{0,1} & 0 & \cdots \\ p_{1,0} & 0 & p_{1,2} & \\ \vdots & \vdots & & \ddots \end{bmatrix}. \tag{2.39}$$

From the stationary condition in (2.31) we can obtain the following relations.

$$P_t = p_{t-1,t} P_{t-1}, \quad \text{thus} \quad P_t = P_0 \prod_{i=1}^{t} p_{i-1,i}. \tag{2.40}$$

where by definition $\sum_{i=1}^{\infty} P_i = 1 - P_0$. Moreover, and after some algebraic manipulations

$$\frac{1 - P_0}{P_0} = \sum_{t=1}^{\infty} \left(\prod_{j=1}^{t} p_{j-1,j} \right). \tag{2.41}$$

It is easy to observe that there are infinite solutions for the transition probabilities $p_{i,j}$. For simplicity, we assume that all $p_{t-1,t}$ are equal, i.e. the *uniform probability* case. It gives us two main advantages:

1. It is the easiest solution to be implemented in practice. The source decides either to transmit or not regardless of what is the current state t.

2. It reduces the problem to a closed from solution.

Lemma 2.3 *The uniform solutions of the non-zero transition probabilities and for the stationary probability vector are:*

$$p_{0,0} = \gamma, \tag{2.42}$$

$$p_{t-1,t} = 1 - \gamma, \ \textit{for } t = 1, 2, \ldots \tag{2.43}$$

$$P_t = \gamma (1 - \gamma)^t. \tag{2.44}$$

Proof Let us first impose that $P_0 = \gamma$. Hence, for the uniform probability case $p_{t-1,t} = p$, and using (2.41)

$$\begin{aligned} \frac{1-\gamma}{\gamma} &= \sum_{t=1}^{\infty} p^t, \\ \frac{1-\gamma}{\gamma} - 1 &= \frac{1}{1-p}, \\ p &= 1-\gamma. \end{aligned} \tag{2.45}$$

So, if $p_{0,1} = 1-\gamma$, we obtain that $p_{0,0} = \gamma$. In order to compute the probability of each state, and considering (2.40), we get

$$P_t = \gamma p^t = \gamma(1-\gamma)^t. \tag{2.46}$$

Theorem 2.4 *The downsampling distortion for the PDE-SD pair is:*

$$\mathcal{D}(PDE,SD) = 2\left(1 - \frac{\gamma}{1-\rho(1-\gamma)}\right). \tag{2.47}$$

Proof Using the MSE_t of the decoder SD in Lemma 2.2, we obtain

$$\begin{aligned} \mathcal{D}(\mathrm{PDE,SD}) &= \sum_{t=0}^{\infty} P_t \mathrm{MSE}_t^{\text{PDE-SD}} \\ &= \sum_{t=0}^{\infty} \gamma(1-\gamma)^t 2(1-\rho^t) = 2\gamma \sum_{t=0}^{\infty} \left((1-\gamma)^t - \rho^t(1-\gamma)^t\right) \\ &= 2\gamma\left(\frac{1}{1-(1-\gamma)} - \frac{1}{1-\rho(1-\gamma)}\right) \\ &= 2\left(1 - \frac{\gamma}{1-\rho(1-\gamma)}\right). \end{aligned} \tag{2.48}$$

2.5.3.4 The pair PDE-PD

The MC in Fig. 2.4 also models the behavior of the PDE-PD pair. The MSE associated to the state t obeys Theorem 2.1.

Theorem 2.5 *The downsampling distortion for the PDE-PD pair is:*

$$\mathcal{D}(PDE,PD) = 1 - \frac{\gamma}{1-\rho^2(1-\gamma)}. \tag{2.49}$$

Proof The downsampling distortion can be computed as

$$\begin{aligned}
\mathcal{D}(\text{PDE,PD}) &= \sum_{t=0}^{\infty} P_t \text{MSE}_t^{\text{PDE-PD}} \\
&= \sum_{t=0}^{\infty} \gamma(1-\gamma)^t(1-\rho^{2t}) = \gamma \sum_{t=0}^{\infty} \left((1-\gamma)^t - \rho^{2t}(1-\gamma)^t\right) \\
&= \gamma\left(\frac{1}{1-(1-\gamma)} - \frac{1}{1-\rho^2(1-\gamma)}\right) \\
&= 1 - \frac{\gamma}{1-\rho^2(1-\gamma)}.
\end{aligned} \tag{2.50}$$

2.5.3.5 The pairs CDE-SD and CDE-PD

The CDE can be also modeled using the infinite MC of Fig. 2.4. Hence, the transmission matrix $\mathbf{T}_{\text{CDE}}$ has the same structure than $\mathbf{T}_{\text{PDE}}$ in (2.39) and the expressions (2.40) and (2.41) are valid as well. However, the rest is different. We address their performance and design in the next section.

2.6 Design and Performance of CDE-SD and CDE-PD

Following the scheme in (2.17), our aim is to design the threshold value Δ in order to guarantee that the source only transmits a fraction γ of the total samples. For the general case, we may have different values of Δ according to each state t of the MC. Therefore, we define the threshold Δ_t as the threshold value applied to the state t.

The condition in (2.17) modifies the *pdf* of the error.

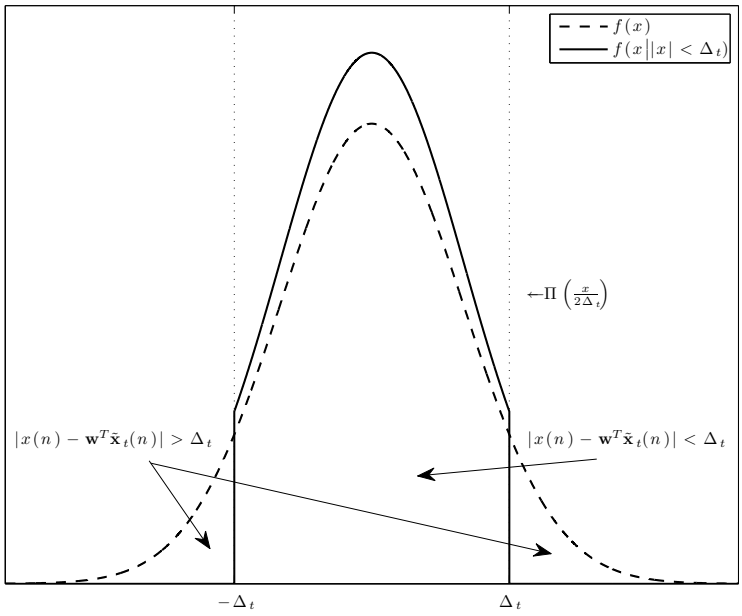

Figure 2.5: Qualitative representation of the conditional *pdf* $f(x\big||x|<\Delta_t)$ due to that the measurements that $|x_s(n)-\mathbf{w}^T\tilde{\mathbf{x}}_t(n)|<\Delta_t$ are not introducing error since they are not estimated. The parameter Δ_t should be chosen in order to guarantee that a fraction γ of the total measurements are transmitted.

Definition 2.6 *Let the conditional pdf* $f(x\big||x|<\Delta_t)$ *be the pdf of* x *conditioned to* $|x|<\Delta_t$*. Mathematically,*

$$f(x\big||x|<\Delta_t)=\beta(\Delta_t)^{-1}f(x)\ \Pi\left(\frac{x}{2\Delta_t}\right), \tag{2.51}$$

where $f(x)$ *is the original* pdf *of* x *and* $\beta(\Delta_t)\in(0,1)$ *is:*

$$\beta(\Delta_t)=\int_{\Delta_t}^{\Delta_t} f(x)dx. \tag{2.52}$$

Moreover, the rectangular function $\Pi(x)$ *is defined as follows:* $\Pi(x)=0$ *if* $|x|>0.5$, $\Pi(x)=1$ *if* $|x|<0.5$*, and* $\Pi(x)=0.5$ *if* $|x|=0.5$*. This definition is summarized in Fig. 2.5.*

Lemma 2.4 *Let* $x\sim\mathcal{N}(0,\sigma^2)$*. Then, the variance of the conditional pdf* $f(x\big||x|<\Delta_t)$ *is:*

$$var(x\big||x|<\Delta_t)=\frac{2}{\sqrt{2\pi\sigma^2}}\left(-\Delta_t\sigma^2\ e^{\frac{-\Delta_t^2}{2\sigma^2}}+\frac{1}{2}\sqrt{2\pi\sigma^6}\ erf\left(\frac{\Delta_t}{\sqrt{2\sigma^2}}\right)\right). \tag{2.53}$$

Proof Let x' define the random variable

$$x'\sim\{x_1\big||x|<\Delta_t\} \tag{2.54}$$

where $x_1\sim\mathcal{N}(0,\sigma^2)$. Hence,

$$var(x')=var(x\big||x|<\Delta_t)=\int_{-\infty}^{\infty}x^2f(x\big||x|<\Delta_t)dx. \tag{2.55}$$

Using the relation:

$$f(A|B)=\frac{f(A,B)}{P(B)}. \tag{2.56}$$

we obtain

$$var(x')=\int_{-\infty}^{\infty}x^2\frac{f(x,|x|<\Delta_t)}{P\{|x|<\Delta_t\}}dx. \tag{2.57}$$

The term $P\{|x| < \Delta_t\}$ in the denominator is:

$$P\{|x| < \Delta_t\} = \int_{-\Delta_t}^{\Delta_t} f(x)dx = \beta(\Delta_t). \tag{2.58}$$

So,

$$var(x') = \beta^{-1}(\Delta_t) \int_{-\infty}^{\infty} x^2 f(x, |x| < \Delta_t)dx. \tag{2.59}$$

Applying the same relation than in (2.56), we obtain

$$var(x') = \beta^{-1}(\Delta_t) \int_{-\infty}^{\infty} x^2 f(x) P\{|x| < \Delta_t \big| x\}dx, \tag{2.60}$$

where the term $P\{|x| < \Delta_t \big| x\}$ is

$$P\{|x| < \Delta_t \big| x\} = \Pi\left(\frac{x}{2\Delta_t}\right). \tag{2.61}$$

Thus,

$$\begin{aligned} var(x') &= \beta^{-1}(\Delta_t) \int_{-\Delta_t}^{\Delta_t} x^2 f(x)dx \\ &= \frac{2}{\beta(\Delta_t)\sqrt{2\pi\sigma^2}} \left(-\Delta_t \sigma^2 \; e^{\frac{-\Delta_t^2}{2\sigma^2}} + \frac{1}{2}\sqrt{2\pi\sigma^6}\; \mathrm{erf}\left(\frac{\Delta_t}{\sqrt{2\sigma^2}}\right)\right), \end{aligned} \tag{2.62}$$

that comes from the relation

$$\int_0^{\epsilon} x^2 e^{-\alpha x^2} dx = -\frac{\epsilon}{2\alpha} e^{-\alpha\epsilon^2} + \frac{1}{4}\sqrt{\frac{\pi}{\alpha^3}} \mathrm{erf}\left(\epsilon\sqrt{\alpha}\right). \tag{2.63}$$

Definition 2.7 *We define the conditional function* $h(\sigma^2|\Delta_t) : \mathbb{R} \to \mathbb{R}$ *as*

$$h(\sigma^2|\Delta_t) = var(x \big| |x| < \Delta_t) \tag{2.64}$$

2.6.1 Approximations for the downsampling distortion of CDE-PD and CDE-SD

For simplicity, we have changed the previous order and we assess first the pair CDE-PD.

2.6.1.1 The pair CDE-PD

As we mentioned previously, some of the results for the CDE encoder are the same than for the PDE encoder. However, the knowledge of some prior information about the signal notably reduces the MSE at the decoder. This is because only the samples with lower MSE are predicted, i.e. the ones that satisfy $|x(n) - \mathbf{w}^T\tilde{\mathbf{x}}_t(n)| < \Delta_t$, since they introduce less noise power at the decoder.

Lemma 2.5 *Let the $MSE_t^{CDE\text{-}PD}$ define as the mean square error when the observation vector is $\tilde{\mathbf{x}}_t(n)$. Then, the $\underline{MSE}_t^{CDE\text{-}PD}$ is an approximation of $MSE_t^{CDE\text{-}PD}$ (i.e., the error introduced by the CDE-PD pair at the state t) and it is defined as*

$$\underline{MSE}_t^{CDE\text{-}PD} = h\left(1-\rho^2+\rho^2\underline{MSE}_{t-1}^{CDE\text{-}PD}|\Delta_t\right) \simeq MSE_t^{CDE\text{-}PD}. \quad (2.65)$$

Proof For $t = 1$, the error $\mathrm{MSE}_1^{\text{CDE-PD}}$ follows the conditional variance[1] such that

$$\begin{aligned}
&\mathrm{MSE}_1^{\text{CDE-PD}} = \mathbb{E}\left[(x(n)-\mathbf{w}^T\tilde{\mathbf{x}}_1(n))^2\big||x(n)-\mathbf{w}^T\tilde{\mathbf{x}}_1(n)|<\Delta_1\right]\\
&= \mathbb{E}\left[(\rho x(n-1)+z(n)-\rho x(n-1))^2\big||\rho x(n-1)+z(n)-\rho x(n-1)|<\Delta_1\right]\\
&= \mathbb{E}\left[z(n)^2\big||z(n)|<\Delta_1\right] = \int_{-\infty}^{\infty} z(n)^2 f'(z(n)\big||z(n)|<\Delta_1)dz(n), \quad (2.66)
\end{aligned}$$

[1]The *conditional variance* of a continuous random variable X given the condition $Y = y$ is defined as $\mathrm{var}(X|Y=y) = \mathbb{E}[X^2|Y=y] = \int_{-\infty}^{\infty} x^2 f(X|Y=y)dx$, where $f(X|Y=y)$ is the conditional *pdf* of X given $Y = y$.

Using Definition 2.7 and since $z(n) \sim \mathcal{N}(0, \sigma_z^2)$ where $\sigma_z^2 = 1 - \rho^2$, the $\mathrm{MSE}_1^{\text{CDE-SD}}$ is

$$\mathrm{MSE}_1^{\text{CDE-PD}} = h(1 - \rho^2 | \Delta_1). \tag{2.67}$$

For $t = 2$ the available knowledge is twofold; i) we know that $|x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_2(n)| < \Delta_2$, and ii) we also know that in $t = 1$ the error was $|z(n-1)| < \Delta_1$. Therefore, the $\mathrm{MSE}_2^{\text{CDE-PD}}$ can be written as

$$\begin{aligned}
\mathrm{MSE}_2^{\text{CDE-PD}} &= \mathbb{E}\left[(x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_2(n))^2 \middle| \right. \\
&\quad \left. |x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_2(n)| < \Delta_2, |z(n-1)| < \Delta_1 \right], \\
&= \mathbb{E}\left[(\rho x(n-1) + z(n) - \rho \mathbf{w}^T \tilde{\mathbf{x}}_1(n-1))^2 \middle| \right. \\
&\quad \left. |x(n) - \rho \mathbf{w}^T \tilde{\mathbf{x}}_1(n-1)| < \Delta_2, |z(n-1)| < \Delta_1 \right], \\
&= \mathbb{E}\left[(\rho z(n-1) + z(n))^2 \middle| \right. \\
&\quad \left. |\rho z(n-1) + z(n)| < \Delta_2, |z(n-1)| < \Delta_1 \right].
\end{aligned} \tag{2.68}$$

The expectation in (2.68) can be computed as

$$\begin{aligned}
\mathrm{MSE}_2^{\text{CDE-PD}} &= \iint_{-\infty}^{\infty} (z(n) + \rho z(n-1))^2 f(z(n) + \rho z(n-1) \big| \\
&\quad |z(n-1)| < \Delta_1, |\rho z(n-1) + z(n)| < \Delta_2) dz(n) dz(n-1).
\end{aligned} \tag{2.69}$$

This expression is actually the computation of the variance of a bivariant truncated normal distribution. The solution of a singly truncated bivariate distribution can be found in [Ros61]. For higher orders, i.e. $t > 2$, the solution refers to the calculation of the variance of a truncated multivariate normal distributions [Man09]. Although a solution already exists in the literature, it turns out to be quite complex. Moreover, its complexity increases in t. For that reason, we are considering the following approximation:

$$\begin{aligned}
&\{\rho z(n-1) + z(n) \big| |z(n-1)| < \Delta_1\} \sim \\
&\sim \mathcal{N}(0, \mathbb{E}\left[(\rho z(n-1) + z(n))^2 \big| |z(n-1)| < \Delta_1\right]).
\end{aligned} \tag{2.70}$$

but in the general case, it does not necessarily follow a Gaussian distribution. The variance $\mathbb{E}\left[(\rho z(n-1)+z(n))^2 \middle| |z(n-1)|<\Delta_1\right]$ can also be expressed as

$$\begin{aligned}\mathbb{E}\left[(\rho z(n-1)+z(n))^2 \middle| |z(n-1)|<\Delta_1\right] &= \\ &= \mathbb{E}\left[z(n)\right] + \rho^2 \mathbb{E}\left[z(n-1) \middle| |z(n-1)|<\Delta_1\right], \\ &= 1-\rho^2+\rho^2 \mathrm{MSE}_1^{\text{CDE-PD}}, \end{aligned} \tag{2.71}$$

so, the MSE introduced at $t=2$ is approximated by

$$\mathrm{MSE}_2^{\text{CDE-PD}} \simeq h\left(1-\rho^2+\rho^2 \mathrm{MSE}_1^{\text{CDE-PD}}|\Delta_2\right). \tag{2.72}$$

It is easy to conclude that for the general case t, the $\underline{\mathrm{MSE}}_t^{\text{CDE-PD}}$ is:

$$\mathrm{MSE}_t^{\text{CDE-PD}} \simeq \underline{\mathrm{MSE}}_t^{\text{CDE-PD}} = h\left(1-\rho^2+\rho^2 \underline{\mathrm{MSE}}_{t-1}^{\text{CDE-PD}}|\Delta_t\right). \tag{2.73}$$

and hence the $\mathcal{D}$(CDE,PD) is approximated by;

$$\mathcal{D}(\text{CDE,PD}) \simeq \sum_{t=0}^{\infty} P_t \underline{\mathrm{MSE}}_t^{\text{CDE-PD}}. \tag{2.74}$$

However, this is still an open problem. It is because the values of P_t are not determined yet. We study this issue afterwards in Section 2.6.2.

2.6.1.2 The pair CDE-SD

If $\hat{x}(n)$ is constructed from a linear prediction using the LWF, the MSE in prediction is directly the power of the noise $z(n)$ as we mentioned in (2.11). However, using other strategies, the error will increase. In particular, the pair CDE-SD constructs $\hat{x}(n)$ as the last transmitted sample, i.e., $\hat{x}(n) = x(n-t)$. This prediction scheme not only introduces error due to $z(n)$ but also due to $x(n)$.

Lemma 2.6 *The $\underline{MSE}_t^{CDE\text{-}SD}$ is an approximation of $MSE_t^{CDE\text{-}SD}$ (i.e., the error introduced by the CDE-SD pair at the state t) and it is defined as*

$$\underline{MSE}_t^{CDE\text{-}SD} = h\left(1-\rho^2 + \underline{MSE}_{t-1}^{CDE\text{-}SD}|\Delta_t\right) \leq MSE_t^{CDE\text{-}SD}. \tag{2.75}$$

Proof Similarly to the CDE-SD, for $t=1$ the error $\mathrm{MSE}_1^{\text{CDE-SD}}$ follows the conditional variance such that

$$\begin{aligned}
\mathrm{MSE}_1^{\text{CDE-SD}} &= \mathbb{E}\left[(x(n)-x(n-1))^2\middle||x(n)-x(n-1)|<\Delta_1\right] \\
&= \mathbb{E}\left[(z(n)-(1-\rho)x(n-1))^2\middle|\right. \\
&\quad \left.|z(n)-(1-\rho)x(n-1)|<\Delta_1\right] \\
&= \mathbb{E}\left[z'(n)^2\middle||z'(n)|<\Delta_1\right] \\
&= \int_{-\infty}^{\infty} z'(n)^2 f(z'(n)\big||z'(n)|<\Delta_1)dz'(n),
\end{aligned} \tag{2.76}$$

where $z'(n)=z(n)-(1-\rho)x(n-1)$ contains both the error contribution due to $z(n)$ and $x(n)$ with variance $\sigma_z'^2$ equal to

$$\begin{aligned}
\sigma_z'^2 &= \mathbb{E}\left[(z(n)-(1-\rho)x(n-1))^2\right] \\
&= \mathbb{E}\left[z(n)\right] + (1-\rho)^2\mathbb{E}\left[x(n-1)\right] = 2(1-\rho).
\end{aligned} \tag{2.77}$$

Therefore, the $\mathrm{MSE}_1^{\text{CDE-SD}}$ is

$$\mathrm{MSE}_1^{\text{CDE-SD}} = h(2(1-\rho)|\Delta_1). \tag{2.78}$$

For $t=2$ the available information is twofold; *i*) we know that $|x(n)-x(n-2)|<\Delta_2$, and *ii*) we also know that in $t=1$ the error was $|z'(n-1)|<\Delta_1$. Therefore, the $\mathrm{MSE}_2^{\text{CDE-SD}}$ can be written as

$$\begin{aligned}
\mathrm{MSE}_2^{\text{CDE-SD}} &= \mathbb{E}\left[(x(n)-x(n-2))^2\middle|\right. \\
&\quad \left.|x(n)-x(n-2)|<\Delta_2, |z'(n-1)|<\Delta_1\right] \\
&= \mathbb{E}\left[(\rho z(n-1)+z(n)-(1-\rho^2)x(n-2))^2\middle|\right. \\
&\quad |\rho z(n-1)+z(n)-(1-\rho^2)x(n-2)| \\
&\quad \left.<\Delta_2, |z'(n-1)|<\Delta_1\right].
\end{aligned} \tag{2.79}$$

To solve the $\mathrm{MSE}_t^{\text{CDE-SD}}$ in a recursive way may be harder than for the CDE-PD case. It is because we cannot apply directly the conditional function since the expectation in (2.79) is not of the form $h(\sigma_x^2|\Delta) = \mathbb{E}[x^2\big||x| < \Delta]$. Hence, in order to simplify, we propose a lower-bound for (2.79) such as

$$\begin{aligned}\mathrm{MSE}_2^{\text{CDE-SD}} \quad \geq \quad & \mathbb{E}\left[(z'(n-1)+z(n))^2\big|\right. \\ & \left.|z'(n-1)+z(n)|<\Delta_2, |z'(n-1)|<\Delta_1\right]. \quad (2.80)\end{aligned}$$

One can easily check that it is in fact an lower-bound since

$$\begin{aligned}\mathbb{E}\left[(z(n)-(1-\rho)x(n-1))^2\right] \quad &\leq \quad \mathbb{E}\left[(\rho z(n)-(1-\rho^2)x(n-1))^2\right] \\ (1-\rho^2) \quad &\leq \quad 2(1-\rho). \quad (2.81)\end{aligned}$$

Our proposed lower-bound is very close to the real value for high values of ρ. Using the same approximation as in the CDE-PD case, and after some simple algebra, we can find a lower-bound of (2.79) as

$$\underline{\mathrm{MSE}}_2^{\text{CDE-SD}} = h\left(1-\rho^2+\mathrm{MSE}_1^{\text{CDE-SD}}|\Delta_2\right) \leq \mathrm{MSE}_2^{\text{CDE-SD}}. \quad (2.82)$$

It is easy to conclude that for the general case t, the $\underline{\mathrm{MSE}}_t^{\text{CDE-SD}}$ is:

$$\underline{\mathrm{MSE}}_t^{\text{CDE-SD}} = h\left(1-\rho^2+\underline{\mathrm{MSE}}_{t-1}^{\text{CDE-SD}}|\Delta_t\right) \leq \mathrm{MSE}_t^{\text{CDE-SD}}. \quad (2.83)$$

and hence the $\mathcal{D}$(CDE,SD) is lower-bounded by;

$$\mathcal{D}(\text{CDE,SD}) \quad \geq \quad \sum_{t=0}^{\infty} P_t \underline{\mathrm{MSE}}_t^{\text{CDE-SD}}. \quad (2.84)$$

As for the case of the CDE-PD pair, this is still an open problem and it is studied afterwards in Section 2.6.2.

2.6.2 Design of the CDE-SD and the CDE-PD

From the design point of view, our aim is to obtain a set of Δ_t's that assure a coding rate at the CDE of γ. However, there are infinite solutions. That is why we propose two possible approaches to face with the design of Δ_t.

- Fixed Δ_t, i.e., $\Delta_t = \Delta$ for all t.
- Variable Δ_t in order to maintain constant transition probabilities, i.e., $p_{t-1,t} = p$ for all t.

2.6.2.1 Fixed Δ_t design

This is probably the simplest approach in order to design the CDE since the encoder do not have to change the value of Δ_t according to the current state since $\Delta_t = \Delta$ for all t.

First, we want to make explicit the existing relation between Δ and $p_{t-1,t}$, as

$$p_{t-1,t}(\Delta) = \int_{-\Delta}^{\Delta} f_t(x)dx. \tag{2.85}$$

where $f_t(x)$ is the pdf of the error at state t.

As we have been doing, we assume that $x(n) - \hat{x}(n)$ follows a Gaussian distribution with zero mean and variance $\mathrm{MSE}_t^{\mathrm{CDE}}(\Delta)$, where

$$\mathrm{MSE}_t^{\mathrm{CDE}}(\Delta) = \begin{cases} 1 - \rho^2 + \mathrm{MSE}_t^{\text{CDE-SD}}(\Delta) & \text{if CUE-SD} \\ 1 - \rho^2 + \rho^2 \mathrm{MSE}_t^{\text{CDE-PD}}(\Delta) & \text{if CUE-PD.} \end{cases} \tag{2.86}$$

Thus[2],

$$p_{t-1,t}(\Delta) = 1 - 2\int_{\Delta}^{\infty} f_t(x)dx = \mathrm{erf}\left(\frac{\Delta}{\sqrt{2\mathrm{MSE}_{t-1}^{\mathrm{CDE}}(\Delta)}}\right), \tag{2.87}$$

[2]It comes from the definition of the cumulative density function of a Gaussian variable such that $\int_{-\infty}^{a} f(x)dx = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{a}{\sqrt{2\sigma_a^2}}\right)\right)$

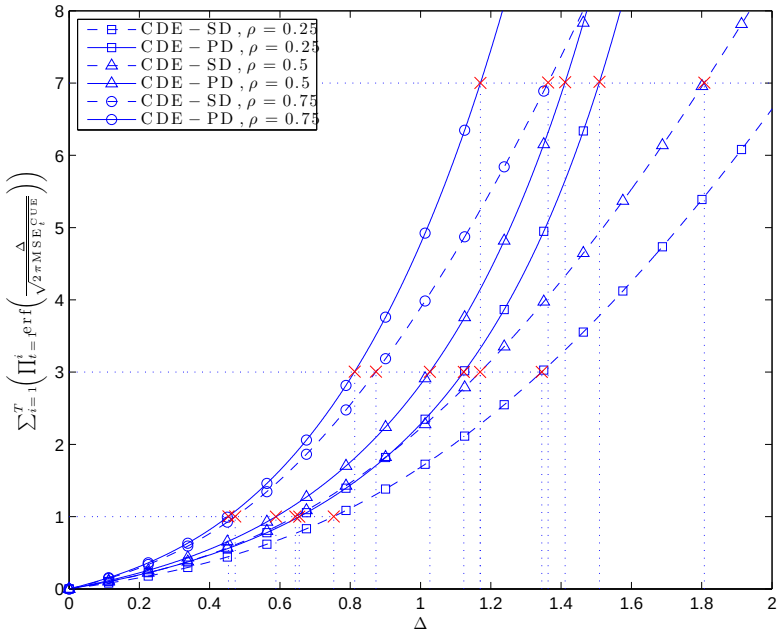

Figure 2.6: Numerical solution of Δ for the fixed Δ_t design. The values of γ are $\gamma = \{0.125, 0.25, 0.5\}$, the values of ρ are $\rho = \{0.25, 0.5, 0.75\}$ and the value of T is 100. The (red) $\times$ are the Δ solutions for a given ρ and γ.

where $\mathrm{erf}(x)$ is the error function of x. Solving for Δ we obtain

$$\Delta = \sqrt{2\mathrm{MSE}_{t-1}^{\mathrm{CDE}}(\Delta)}\mathrm{erf}^{-1}\left(p_{t-1,t}\right), \quad \text{for } t = 0, 1, \ldots \tag{2.88}$$

Using the result in (2.41), we can numerically approximate Δ that assures $P_0 = \gamma$ as the unique solution of

$$\sum_{i=1}^{T}\left(\prod_{t=1}^{i}\mathrm{erf}\left(\frac{\Delta}{\sqrt{2\mathrm{MSE}_{t}^{\mathrm{CDE}}(\Delta)}}\right)\right) = \frac{1-\gamma}{\gamma}, \quad \text{for } T \to \infty. \tag{2.89}$$

The solution of Δ for different values of γ and ρ can be graphically seen in Fig. 2.6.

2.6.2.2 Variable Δ_t design

This approach allows for a slightly easier computation of the values of Δ_t. The main difference with the previous design scheme is that we can use the result in Lemma 2.3, such that, $p_{t-1,t} = 1-\gamma$ and $p_{0,0} = \gamma$. Hence, Δ_t is directly

$$\Delta_t = \sqrt{2\mathrm{MSE}_{t-1}^{\mathrm{CDE}}(\Delta_{t-1})}\mathrm{erf}^{-1}\left(1-\gamma\right). \tag{2.90}$$

To graphically validate our design framework, we have proposed the following experiment.

Experiment 2.1 *We have simulated the CUE-SD and the CUE-PD for* $\gamma = [1/8\ 1/4\ 1/2]$ *and for* $\rho \in [0, 1]$. *The signal has been generated following the AR-1 process of* 5000 *samples (for each value of* ρ*). We have computed the probability of transmission* P_0 *obtained using our threshold design framework.*

From Experiment 2.1, we have plotted the probability of transmission P_0 as a function of ρ and for each value of γ. We have used the variable Δ_t

design. In Fig. 2.7, we have compared the obtained results with the target coding rate γ and we have observed that for the case of CUE-PD, the fitting is very accurate. For the case of CUE-SD, is slightly worse. It is due to the approximation in (2.81). We said that this approximation improves when $\rho \to 1$. This behavior can be observed in Fig. 2.7.

2.7 Performance Evaluation

In this section, we evaluate and compare the performance of the different encoder-decoder pairs as a function of the downsampling distortion. Moreover, we introduce an experimental evaluation in order to confirm the validity of our theoretical results. For that, we have generated a signal $x(n)$ as a sequence of 5000 samples using the $AR-1$ model in (2.12) and for different values of the autoregressive parameter $\rho \in [0, 1]$ with resolution 0.01. The results are computed for $\gamma = [1/8,\ 1/4,\ 1/2]$.

2.7.1 The pair DDE-SD and the pair DDE-PD

We analyze the downsampling distortion for the DDE-SD and the DDE-PD pair. We compare the theoretical results with the experimental results.

So, Fig. 2.8 confirms the validity of our theoretical model for the downsampling distortion.

Also we compare the difference in performance according to the decoder used. The PD takes into account the signal correlation information in the decoding process and hence, the total performance is increased notably for low values of ρ. On the contrary, if $\rho \to 1$, both decoders perform similarly since $x(n) - \rho^t x(n-t) \approx x(n) - x(n-t)$.

In Fig 2.8, we can also graphically evaluate the impact of γ. In our scenario, the signal $x(n)$ is transmitted by the DDE one in $\{8,\ 4,\ 2\}$ times following a uniform pattern. It is easy to see that the larger the γ the lower

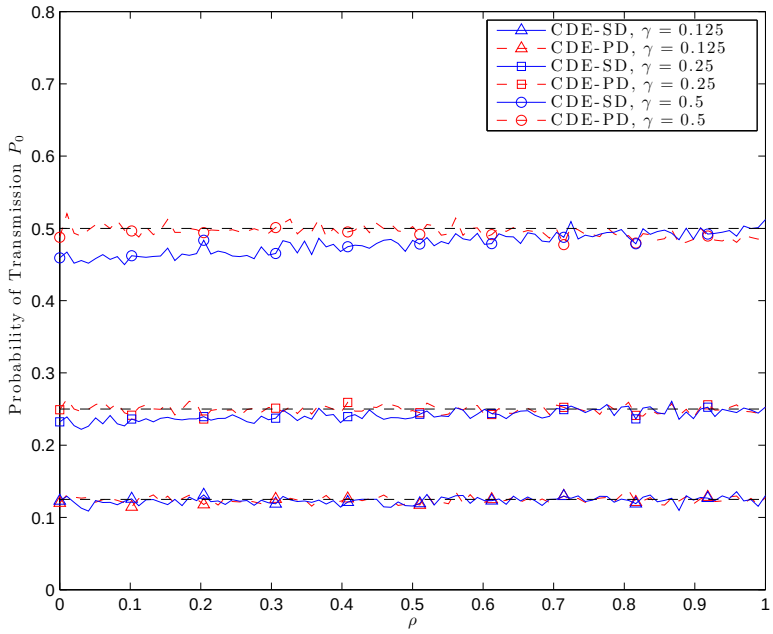

Figure 2.7: Experimental results from Experiment 2.1. The empirical probability of transmission is compared with the target coding rate γ for the CUE-SD and CUE-PD schemes. We have used the variable threshold design.

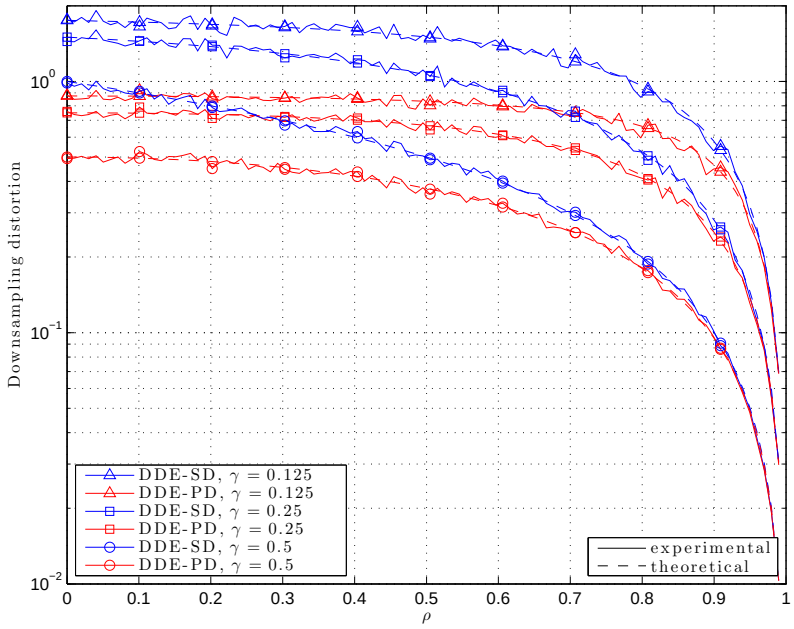

Figure 2.8: Experimental and theoretical downsampling distortion of the pairs DDE-SD and DDE-PD as a function of ρ. The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$.

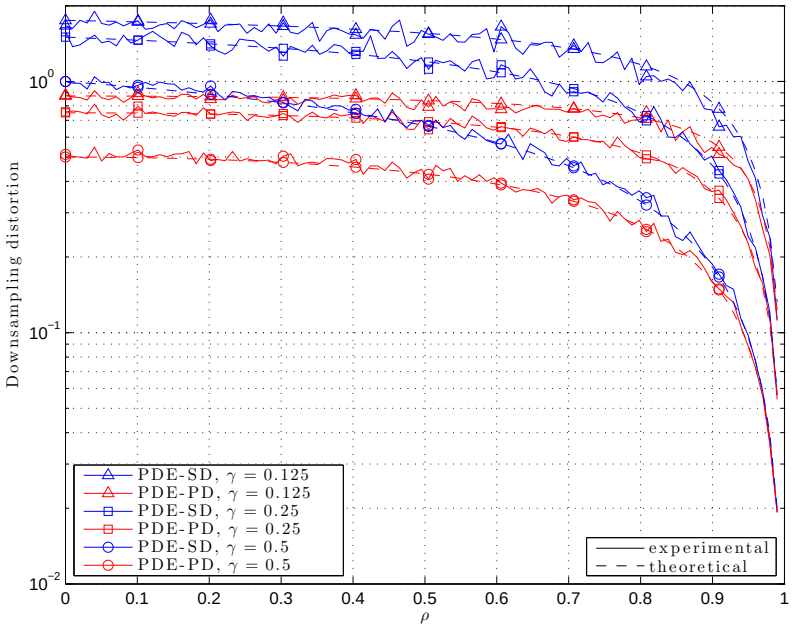

Figure 2.9: Experimental and theoretical downsampling distortion of the pairs PDE-SD and PDE-PD as a function of ρ. The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$.

the distortion. However, there exists a trade of between the downsampling distortion and the compression rate.

2.7.2 The pair PDE-SD and the pair PDE-PD

The downsampling distortion for the PDE-SD and the PDE-PD is plotted in Fig. 2.9. However, the conclusions that one can extract from these results are basically the same than for the pairs DDE-SD and DDE-PD. In order to be concise, we compare the downsampling distortion performance of the different pairs later in Section 2.7.4.

2.7.3 The pair CDE-SD and the pair CDE-PD

The performance of the previous encoder-decoder pairs can be notably improved by using conditional transmission at the encoder site. In particular, we study and compare the downsampling distortion of the two design approaches, i.e., the fixed Δ_t design and the variable Δ_t design (with uniform transition probabilities), depicted in Fig. 2.10 and Fig. 2.11, respectively. As in the previous pairs, we compare both the experimental results with the theoretical results. However, in that case our theoretical results are limited to an approximation rather than the real system performance. Even so, we can observe that the approximations are very accurate for all the different simulations. For the case of CUE-PD, the approximation is so close to the system performance that the difference cannot be observed because it is masked by the small amount of noise due to the simulation. For the case of CUE-SD, the difference is slightly bigger because the approximation in (2.81).

Another conclusion is that the downsampling distortion is notably higher for the fixed design. It is because their transition probabilities $p_{t-1,t}$ are increasing in t, and it facilitates to achieve higher states t in the MC with higher probability (i.e., higher MSE_t's). On the contrary, the variable design concentrates the states in lower t values.

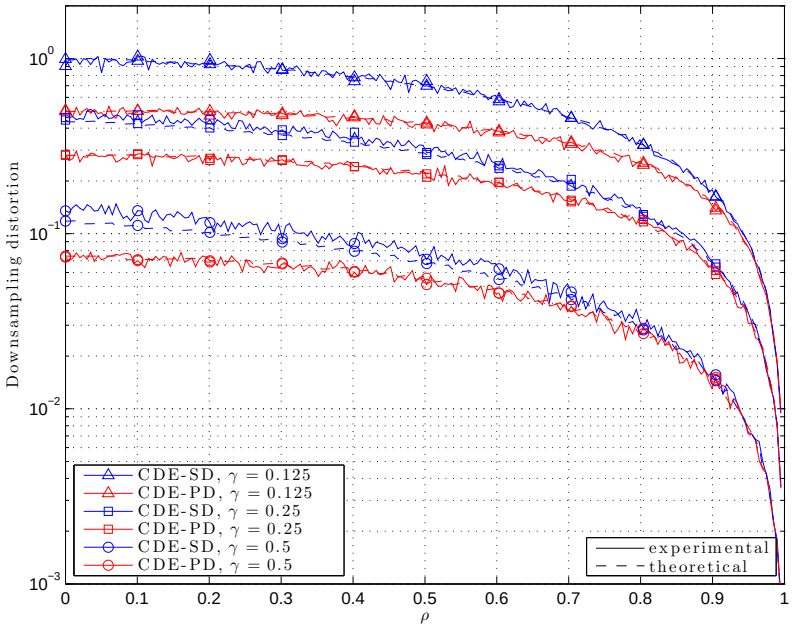

Figure 2.10: Experimental and theoretical approximation of the downsampling distortion of the pairs CDE-SD and CDE-PD following a fixed Δ_t design. The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$.

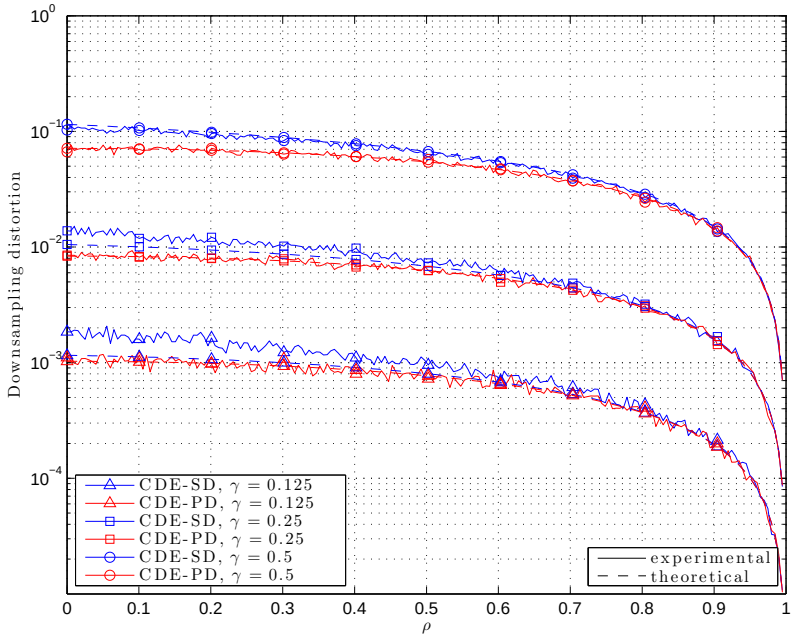

Figure 2.11: Experimental and theoretical approximation of the downsampling distortion of the pairs CDE-SD and CDE-PD following a variable Δ_t design. The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$.

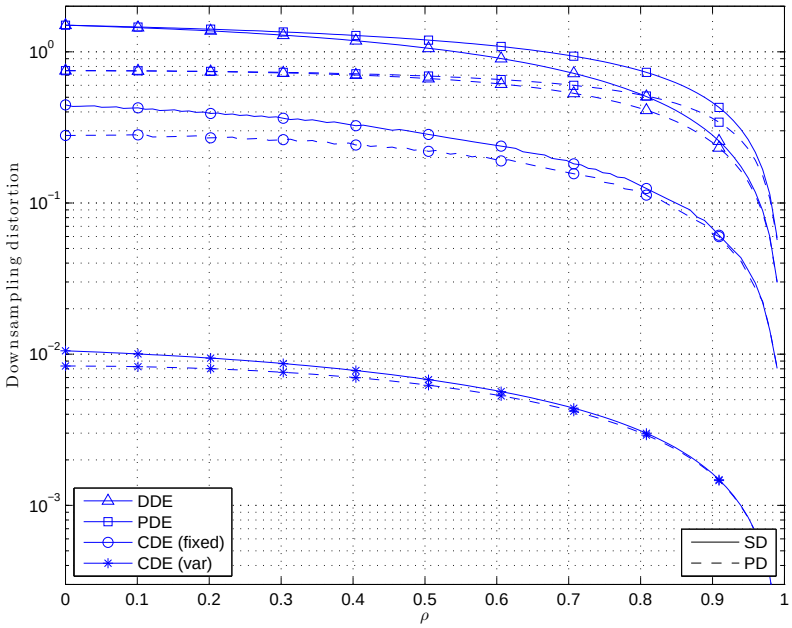

Figure 2.12: Comparison of the downsampling distortion of the different encoding-decoding pairs as a function of ρ. The coding rate is $\gamma = 0.25$.

From a practical point of view, the CDE is simpler if it follows a fixed design since the encoder only needs to know the value of Δ and also it does not need to track the current state t. However, from a computational point of view, the variable approach is simpler since it can be computed analytically, instead of numerically.

2.7.4 Comparison of the downsampling distortion

Finally, we compare the performance of the different encoder-decoder pairs. Although Fig. 2.12 does not provide any extra information, it allows us to better compare the performance of the different schemes. For the sake of simplicity, we only compare the theoretical results for the case of $\gamma = 0.25$.

One can observe that the performance of the DDE and PDE encoders are similar. However, the deterministic encoder works slightly better since it only uses the lowest γ^{-1} states of the finite MC, while PDE uses higher states that are related to higher errors. However, the main disadvantage of the DDE encoder in front of the PDE is its lack of flexibility, since the uniform solution is only valid for natural values of γ^{-1}. Furthermore, the PDE with uniform transition probabilities do not need to track the current state t of the process and hence it is simpler.

The big hop in performance is observed for the CDE. This encoder eliminates the transmissions of the samples with most redundant information. Thus, only the most "unpredictable" samples are transmitted.

2.8 Conclusions

In this chapter, we have evaluated the performance of different encoding-decoding strategies in order to reduce the number of transmitted samples. In particular, we define the downsampling distortion function in order to evaluate the performance of the combination of three downsampling encoders, which are the deterministic downsampling encoder (DDE), the probabilistic

downsampling encoder (PDE), and the conditional downsampling encoder (CDE), with two decoders: the step decoder (SD) and the predictive decoder (PD).

We have obtained closed form expressions for the pairs DDE-SD, DDE-PD, PDE-SD and PDE-PD and accurate approximations for CDE-SD and CDE-PD. Moreover, we have proposed two strategies in order to design the threshold of the condition in the CDE, i.e., the fixed threshold design and the variable threshold design.

The simulation results validate our theoretical results. Furthermore, we have compared the performance of the different pairs and we have showed the impact of taking into account the signal model in the encoding-decoding process. Hence, the pair CDE-PD (with variable threshold design) outperforms by far the rest of the studied strategies. However, finding the optimum threshold values in terms of minimizing the downsampling distortion remains as an open problem.

Chapter 3

Enhanced Correlation Estimators for Distributed Source Coding

3.1 Summary

In this chapter, we propose two estimators based on correlation parameters for the two key steps of a practical *distributed source coding* scheme, namely: *i*) the computation of the side-information at the receiver side, and *ii*) the estimation of the required number of bits to compress the readings in order to guarantee a certain symbol error probability. We show that using the proposed enhanced estimators, the distributed source coding algorithm performs better in terms of both the compression rate and the symbol error rate. In particular, this improvement is specially significative when the number of snapshots used in the training phase is only slightly larger than the dimension of the observation vector. On the contrary, when the number of snapshots is much higher than the observation dimension, our proposed estimators perform similarly to the classical estimators.

3.2 Introduction

Sample estimators are widely used in statistical signal processing and it is well-known that their performance is highly conditioned to the number of considered samples [Hay01]. In particular, sample correlation estimators perform the best when the number of samples is sufficiently large in comparison with the dimension of the observation vector. However, when both magnitudes are similar, the performance may be severely degraded and other techniques should be addressed.

In this chapter, we propose two enhanced correlation estimators derived from Generalized Statistical Analysis (GSA) (introduced by V. L. Girko in [Gir90] and extended in [Gir98]). This discipline comes from Random Matrix Theory (RMT) [Meh91, Tul04] and provides consistent estimators when both the number of snapshots of the training phase N and the observation dimension M are arbitrarily large and comparable in magnitude.

Following this approach, the main motivation is to include such derived estimators for Distributed Source Coding (DSC) applied to a large Wireless Sensor Network (WSN) framework, which typically are formed by a large number of high space-time correlated sources (e.g. fire control in forests monitoring the temperature or humidity levels, or tracking the location of the products in large stores), where DSC may be used to remove the inherent redundancy in such a correlated readings [Pra02] , [Pra03] and hence send compressed messages with the subsequent energy savings.

3.2.1 Previous results of DSC in WSNs

Surprisingly, existing results from information theory (precisely, from the work of Slepian and Wolf [Sle73]) show that this compression can be executed in a fully blind manner, i.e., only with the knowledge of the local data. It means that sensors compress the data without the knowledge of the signals of the other sensors, and interestingly, without any loss of per-

formance in comparison with the centralized approach. Theoretically, the DSC achieves the maximum sum rate, however, practical algorithms still perform far from the theoretical limits [Pra03].

However, practical (and suboptimal) solutions can be found in the literature. For a star-topology WSN, the authors in [Old08b] propose a DSC scheme divided in two phases: the training phase and the compression phase. During the training phase, the correlation parameters are estimated. Hence, the duration of this phase depends on the network configuration and the requirements of the application. In particular, they consider a network composed only of two source nodes and one sink. For higher number of sources, the number of snapshots used in the correlation estimation notably increases. The authors extend their results in [Old08a] to a cluster-based WSN, where each cluster manages a total of four nodes and acts separately to the rest of the clusters. However, in both [Old08b] and [Old08a], the estimation of the correlation parameters is not detailed.

For a relay WSN scenario, the authors in [Tan07] also present a two-phase DSC algorithm. As in [Old08b] and [Old08a], they assume that the training phase is large enough to achieve the desired accuracy in the correlation estimation for an arbitrary number of source nodes. Even so, the sources are managed into smaller groups or clusters. Following this grouping approach, the DSC algorithm cannot fully exploit all the spatial correlations within the network, since only the correlations among the sensors of a cluster are used. Therefore, a lot of useful information is missed.

For a multi-hop WSNs, the scheme proposed in [Wan08] exploits the "redundancy free" nature of the DSC to optimize jointly the DSC and the routing paths in order to increase the lifetime of the network. More recent results in [Sax10] extend the DSC algorithm for a multistage scheme. The authors particularize the results for two sources and two layers.

Many other interesting DSC algorithms are also being actively studied, with new alternatives that continuously improve many aspects of DSC. In

this chapter, we extend the work in [Cho03], where the authors propose a simple DSC algorithm in order to compress the signal from multiple space-time correlated sources. Although their analysis is for an arbitrarily large number of nodes, they particularize for the case of two nodes.

3.2.2 Our contribution

In the literature of DSC, there exists a lack of study regarding the performance drop in the correlation estimation due to a large number of correlated sources.

In order to overcome this limitation, we address the case of a DSC algorithm applied to a large WSN scenario, where the observation dimension M is typically large (since it depends on the number of sensors that composes the network) and classical sample estimators may fail unless a very long training phase is considered (becoming in most cases unpractical). On the contrary, our proposed estimators improve the trade-off between the training phase duration N and the accuracy of the estimation of the correlation parameters. The main contributions of this chapter are summarized as follows:

i) We analyze the performance of the DSC algorithm for large WSNs, and in particular, we study the correlation estimation problem in such a scenario.

ii) We propose two enhanced estimators to mitigate the performance drop of DSC algorithm when the number of sources is arbitrarily large and conventional estimators are used.

iii) We numerically compare the performance of conventional sample estimators with our enhanced estimators. Our enhanced DSC algorithm turns out to decrease largely the training phase duration and allows us to reduce the number of transmitted bits in comparison with the conventional approach.

Although our DSC algorithm is based on the Algorithm in [Cho03], it presents many differences. The most significative ones are:

1. We incorporate our enhanced estimators that improve the DSC performance in a large WSN scenario, which is the core of the Section III.

2. We use different assumptions of the prediction error than the ones in [Cho03]. In particular, we study Gaussian signals (the ones most present in the nature), and therefore our expressions are derived accordingly to their distributions. On the contrary, in [Cho03] they do not assume any statistical structure of their signals.

3. The authors in [Cho03] perform the correlation tracking and the signal compression simultaneously at the same phase since they use an adaptive LMS approach. Instead, we use a two phase algorithm, i.e. sensing phase and compression phase, as the ones in [Old08b], [Old08a] and [Tan07]. A detailed description can be found next in Algorithm 3.1: sensing node and in Algorithm 3.2: fusion center.

3.2.3 Organization of the chapter

The rest of the chapter is organized as follows: In Section 3.3 we describe the system model. The DSC algorithm is presented in Section 3.4. The derivation of the enhanced estimators is detailed in Section 3.5. Simulation results are given in Section 3.6, and conclusions are drawn in Section 3.7.

3.3 System Model and Assumptions

We assume a large and dense WSN scenario that measures a certain physical phenomena such as temperature or humidity. For *large* WSN we mean that the number of sensing nodes may be arbitrarily large, i.e. of hundreds or thousands of nodes, and for *dense* WSN we mean that the sensing nodes

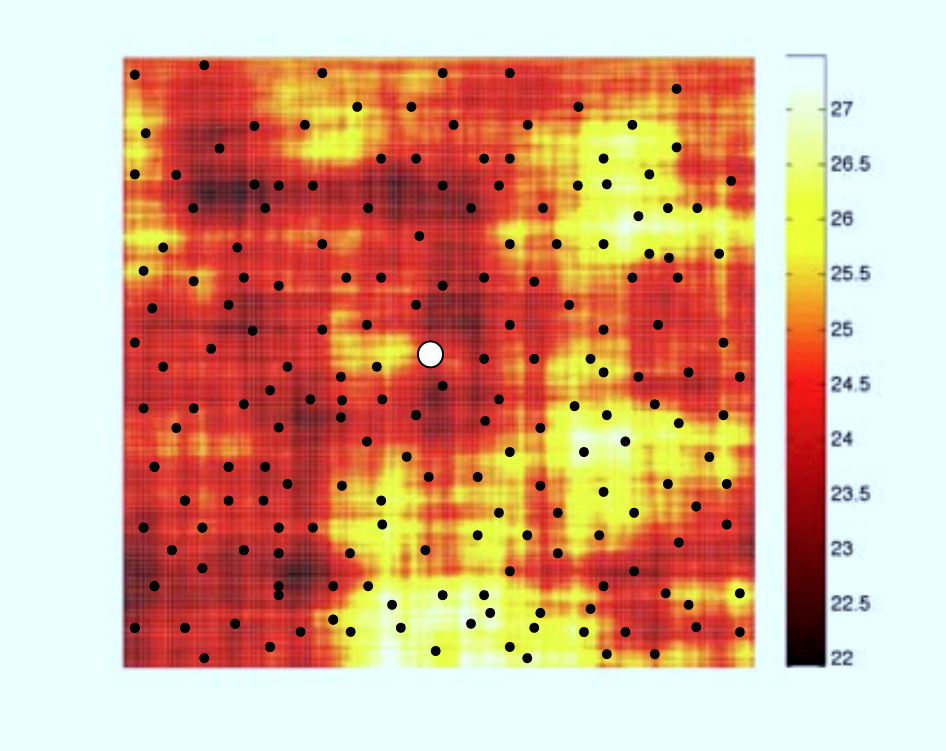

Figure 3.1: Illustrative example of a correlation dominated large WSN field, composed of a set $\mathcal{S}$ of S sensing nodes (black dots) measuring a certain scalar magnitude and transmitting their readings to one fusion center (white dot).

are close enough to present spatial correlations in their measured data. This scenario is graphically summarized in Fig. 3.1. The final interest of this chapter is to study the impact when the fusion center receives the information from a large number of sensing nodes. Thus, although we assume a WSN configured in star topology for simplicity, our proposed algorithm is also compatible with multi-hop techniques or with other network configurations.

Therefore, the network is composed of two types of nodes: *i*) a set $\mathcal{S}$ of S sensing nodes that transmit the measurements when they are requested, and *ii*) one fusion center that manages the sensing nodes, and gathers and processes their measured data. We assume that the limitations in terms of computing power and energy consumption are in the sensing nodes, instead we assume no constraints for the fusion center.

We consider that the signals are space-time correlated and modeled as an S-dimensional stochastic process, namely,

$$\mathbf{X} = [\mathbf{x}(1)\,\mathbf{x}(2)\,\ldots\,\mathbf{x}(N)\,] = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(N) \\ x_2(1) & x_2(2) & \cdots & x_2(N) \\ \vdots & \vdots & & \vdots \\ x_S(1) & x_S(2) & \cdots & x_S(N) \end{bmatrix}, \tag{3.1}$$

where $x_s(n)$ denotes the measurement of the sth sensor at the sample time n and N denotes the number of time samples in the observation window.

The main assumptions throughout this chapter are collected as follows:

3.3.1 Assumptions on the signal model

Let $x_s(n)$ be a real and time-discrete auto-regressive model of order 1 ($AR-1$), which is commonly assumed in the signal processing literature in order to model real sources [Has80]. It is defined as:

$$x_s(n) = \rho_t x_s(n-1) + z_t(n), \quad \text{for } n = 1, 2, \ldots \tag{3.2}$$

The auto-regression time coefficient is denoted by $\rho_t \in [0,1]$ and assumed to be constant during the transmission. The random process $z_t(n)$ is a sequence of Gaussian distributed and independent random variables with zero mean and variance σ_t^2.

In the same way, $x_s(n)$ can be also modeled as a space $AR-1$ following,

$$x_s(n) = \rho_x x_{s-1}(n) + z_s(n), \quad \text{for } s = 1, 2, \dots S. \tag{3.3}$$

The auto-regression space coefficient is denoted by $\rho_x \in [0,1]$ and it is also assumed to be constant during the transmission. The random process $z_s(n)$ is a sequence of Gaussian distributed and independent random variables with zero mean and variance σ_s^2.

Hence, following with the results in Chapter 2, the time-covariance matrix of the time sequence of length T, i.e., $[x_s(n)\ x_s(n-1) \cdots x_s(n-T+1)]^T$ follows

$$[\mathbf{R}_t]_{n,n-i} = \rho_t^i. \tag{3.4}$$

and the space-covariance matrix of the spatial vector $[x_1(n)\ x_2(n) \cdots x_S(n)]^T$ follows

$$[\mathbf{R}_s]_{s,s-i} = \rho_x^i. \tag{3.5}$$

Without loss of generality, we assume that $\sigma_x^2 = 1$. Therefore, following the Lemma 2.1, the variance of the noise $z_t(n)$ is $\sigma_t^2 = 1 - \rho_t^2$, while $\sigma_s^2 = 1 - \rho_x^2$ denotes the variance of $z_s(n)$.

3.3.2 Assumptions on the channel and the system model

We assume noiseless communication from $\mathcal{S}$ to the fusion center. Although any real application measurement will be corrupted by at least a small amount of noise, we consider noise-free communication paths in order to better evaluate the system performance, since the communication noise

will only affects by incrementing the total Symbol Error Rate (SER) at the fusion center.

We assume Time Division Multiple Access (TDMA) with successive decoding. First, sensor 1 transmits its reading using only the knowledge of its own past samples. After, sensor 2 codifies its readings according to its past samples and the reading of sensor 1, and so on. Without loss of generality, we focus on the study of the sth sensor where already S' sensors have been decoded. Hence, sensor s uses the information from the S' previously decoded sensors and its past readings to codify $x_s(n)$.

3.4 Distributed Source Coding Algorithm

3.4.1 DSC background

In order to explain the idea behind DSC, the concept of entropy is needed. If X and Y denote two discrete random variables, the entropy of a discrete random variable $\mathcal{H}(X)$ can be seen as the minimum number of bits required to encode X without any loss of information. Similarly, the joint entropy of two discrete random variables $\mathcal{H}(X,Y)$ can be seen as the minimum number of bits needed to encode X and Y jointly. If X contains any information about Y (i.e., they are somehow correlated), then $\mathcal{H}(X,Y) < \mathcal{H}(X)+\mathcal{H}(Y)$. One can first encode Y to $\mathcal{H}(Y)$ and then encode X to $\mathcal{H}(X|Y)$, which is the entropy of X if Y is known. By definition, $\mathcal{H}(X,Y) = \mathcal{H}(Y)+\mathcal{H}(X|Y)$. In Fig. 3.2(a), one can see an example scheme of how this can be performed.

The main novelty of DSC (introduced first in [Sle73] by Slepian and Wolf in 1973) is that the coding rate (i.e., $\mathcal{H}(X|Y)$ bits/sample) can be guaranteed without loss of information even when the encoder does not have full access to the random variable Y. Hence, the knowledge of Y is only assumed available at the decoder, as in Fig. 3.2(b). Three years later, Wyner and Ziv extend these results to the case of lossy encoding of continuous-valued gaussian variables [Wyn75].

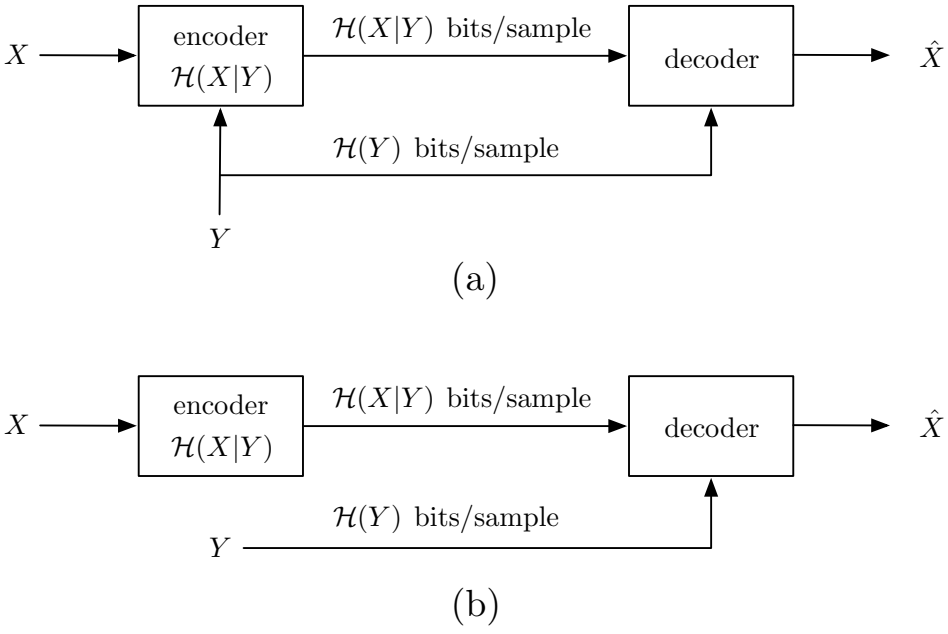

Figure 3.2: (a) shows the block diagram of encoding the random variable X to $\mathcal{H}(X|Y)$ bits/sample where Y is known at both the encoding and the decoding blocks. Otherwise, in (b), X is encoded to $\mathcal{H}(X|Y)$ bits/sample, where Y is only known at the decoding block.

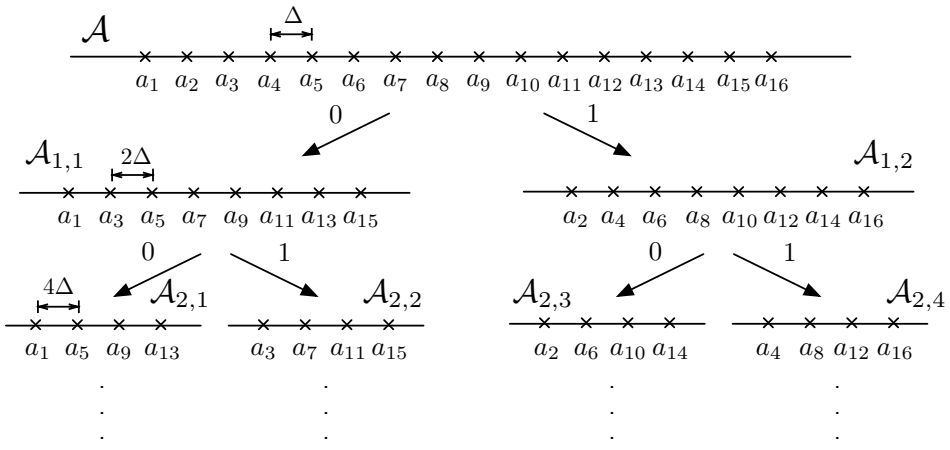

Figure 3.3: Graphical representation of the sub-codebooks using a tree-based scheme. In this example, the alphabet $\mathcal{A}$ contains 16 symbols, and a 4th level of sub-codebooks is shown.

3.4.2 Practical DSC algorithm

Since there are no practical techniques to achieve the theoretical limits of [Sle73] and [Wyn75], suboptimal algorithms are used instead. In this chapter, we follow the approach in [Cho03], where the authors propose the construction of a codebook based on the decomposition of a given finite alphabet $\mathcal{A}$ in several sub-codebooks. Fig. 3.3 gives a graphical intuition on how the codebook can be decomposed in several sublevels.

In general, the DSC algorithm is divided in two phases that involve both the sensing nodes and the fusion center (see Algorith 3.1 and Algorithm 3.2 respectively):

1. The training phase of length N, where the sensing node maps its l-bit reading $x_s(n)$ according to the alphabet $\mathcal{A} = \{a_i\}_{i=1,2...,2^l}$, with a quantization step of $|a_{i+1} - a_i| = \Delta$, and sends an uncompressed version of its data coded in l-bits. After collecting the N snapshots of the training phase, the fusion center estimates the correlation parameters for each sensor.

2. The coding phase, where a given *side-information* $y(n)$ is available at the fusion center and the sensing node can encode its reading using only $b(n) \leq l$ bits. Hence, the sensor transmits only the index B of a sub-codebook $\mathcal{A}_B \subseteq \mathcal{A}$ (B is codified in $b(n)$ bits) that contains the mapped reading $x_s(n)$. Thus, the fusion center receives the sub-codebook identifier B, and selects the symbol in $\mathcal{A}_B$ closer to the side-information $y(n)$,

$$x_s(n) = \arg \min_{a_i \in \mathcal{A}_B} |y(n) - a_i|. \tag{3.6}$$

Let us concentrate on the following two steps of the coding phase.

Step A. Compute the side-information $y(n)$

First, let us define the observation vector $\mathbf{x}(n) \in \mathbb{R}^M$ with covariance matrix $\mathbf{R} \in \mathbb{R}^{M\times M}$ as the information available at the fusion center and $\mathbf{r}_x$ is the cross-correlation vector, $\mathbf{r}_x = \mathbb{E}[\mathbf{x}(n)x_s(n)]$. The vector $\mathbf{x}(n)$ collects: *i*) the K past readings of the sensor, and *ii*) the readings of the set $\mathcal{S}'$ of already-decoded sensors in time slot n (where $\mathcal{S}' \subset \mathcal{S}$ with cardinality S'), hence $M = K + S'$. Note also that the covariance matrix $\mathbf{R}$ will be constructed from the corresponding entries of $\mathbf{R}_s$ and $\mathbf{R}_t$. Then, the side-information $y(n)$ is a linear prediction of $x_s(n)$ and it is computed as a linear combination of the entries of $\mathbf{x}(n)$, i.e.,

$$y(n) = \mathbf{w}^H \mathbf{x}(n), \tag{3.7}$$

following the Linear Wiener Filter (LWF) solution. The LWF solution, $\mathbf{w}^\star = \mathbf{R}^{-1}\mathbf{r}_x$, is known to be optimal in the Mean Square Error (MSE) sense [Hay01]. Mathematically,

$$\mathrm{MSE}(\mathbf{w}) = \sigma^2_{x_s} - 2\,\mathrm{Re}[\mathbf{w}^H\mathbf{r}_x] + \mathbf{w}^H\mathbf{R}\mathbf{w}. \tag{3.8}$$

$$\begin{aligned} \frac{\partial \mathrm{MSE}(\mathbf{w})}{\partial \mathbf{w}^H} &= -\mathbf{r}_x + \mathbf{w}^H\mathbf{R} = 0; \\ \mathbf{w}^\star &= \mathbf{R}^{-1}\mathbf{r}_x, \end{aligned} \tag{3.9}$$

and then, the MSE achieved is minimum and is given by

$$\mathrm{MSE}(\mathbf{w}^\star) = \sigma^2_{x_s} - \mathbf{r}_x^H\mathbf{R}^{-1}\mathbf{r}_x. \tag{3.10}$$

However, to compute $\mathbf{w}^\star$ the perfect knowledge of $\mathbf{R}^{-1}$ and $\mathbf{r}_x$ is necessary but not available. Classical methods replace $\mathbf{R}^{-1}$ and $\mathbf{r}_x$ directly by their sample estimators denoted by $\hat{\mathbf{R}}^{-1}$ and $\hat{\mathbf{r}}_x$, respectively. Although when $N \gg M$ this classical approach provides good results, better estimators can be used instead when N has the same order of magnitude as M, but still $N > M$.

Algorithm 3.1 Sensing node

1. Training phase

Get l-bit reading from A/D converter

Transmit l-bit symbol.

2. coding phase

Get l-bit reading from A/D converter

Encode and transmit $b(n)$-bit codeword.

Step B. Compute the number of bits in transmission $b(n)$

In order to determine the number of bits $b(n)$ to encode $x_s(n)$ without decoding error, one must guarantee that $|x_s(n)-y(n)| < 2^{b(n)-1}\Delta$. However, since the reading $x_s(n)$ is not yet available at the fusion center, we compute the number of bits to encode $x_s(n)$ in order to guarantee a given Symbol Error Rate threshold, $\mathrm{SER_t}$.

Assuming $x_s(n) - y(n) \sim \mathcal{N}(0, \mathrm{MSE}(\mathbf{w}))$, the SER can be expressed as

$$\mathrm{SER} \quad = \quad \mathrm{erfc}\left(\frac{2^{b(n)-1}\Delta}{\sqrt{2\mathrm{MSE}(\mathbf{w})}}\right). \tag{3.11}$$

We have focused on the particular case of Gaussian prediction errors. For a general case, other approaches can be used, as e.g., the Chebychev's inequality in [Cho03].

Solving for $b(n)$ in (3.11) for a given $\mathrm{SER_t}$, we get

$$b(n) \quad \geq \quad \left\lceil \log_2\left(\frac{\sqrt{2\mathrm{MSE}(\mathbf{w})}}{\Delta}\mathrm{erfc}^{-1}(\mathrm{SER_t})\right) + 1 \right\rceil. \tag{3.12}$$

It should be iteratively repeated for every sensing node as it is represented in Algorithm 3.2.

In large WSNs, the number of already-decoded sensors S' (and hence M) is typically large. Therefore, maintaining a training phase such that $N \gg M$

Algorithm 3.2 Fusion center

1. Training phase
for $n = 1$ to N **do**
  for $s = 1$ to S **do**
    Request sth sensor for a l-bit reading (i.e., uncoded).
  end for
end for
Estimate the correlation parameters for each sensor, i.e., $\mathbf{R}$ and $\mathbf{r}_x$ and compute $\hat{\mathbf{w}}$ and $\widehat{\mathrm{MSE}}(\hat{\mathbf{w}})$ as in (3.35) and (3.45), respectively.
2. Coding phase
for $n > N$ to *end* **do**
  for $s = 1$ to S **do**
    Step A. Compute side-information as $y(n) = \hat{\mathbf{w}}^H \mathbf{x}(n)$.
    Step B. Compute $b(n)$ following (3.12).
    Request sth sensor for a $b(n)$-bit reading (i.e., encoded).
    Decode $x_s(n)$ using (3.6).
  end for
end for.

may become inefficient in most cases. On the other hand, eq. (3.12) requires an accurate estimation of MSE($\mathbf{w}$) in order to obtain the smallest $b(n)$ possible, while $\mathrm{SER_t}$ is guaranteed. Thus, our aim is to look for enhanced estimators for both $\mathbf{w}$ and MSE($\mathbf{w}$) that improve the *classical* estimators when N and M are large and comparable in magnitude.

3.5 Enhanced Correlation Estimators

First, let us consider a collection of N random observations of a certain M-dimensional stochastic process, denoted by $\mathbf{X}_N = [\mathbf{x}(1)\,\mathbf{x}(2)\,\ldots\,\mathbf{x}(N)]$. We assume, without loss of generality, that these observations have zero

mean $\mathbb{E}[\mathbf{x}(n)] = 0$, and $\mathbb{E}[\|\mathbf{x}(n)|^2] = 1$, and covariance matrix $\mathbf{R}$.

The Sample Covariance Matrix (SCM), here denoted by $\hat{\mathbf{R}}$, is constructed from the observations as in [Rub09],

$$\begin{aligned}\hat{\mathbf{R}} &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}(n)^H \\ &= \frac{1}{N}\mathbf{X}_N\mathbf{X}_N^H = \frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}^H\boldsymbol{\Xi}\mathbf{R}^{1/2},\end{aligned} \tag{3.13}$$

where $\boldsymbol{\Xi}$ defines a $N \times M$ random matrix with i.i.d. complex entries, zero mean and unit variance. Moreover, let $\hat{\mathbf{r}}_x$ be the sample cross-correlation vector between the observation vector $\mathbf{x}(n)$ and the desired response $x_s(n)$, defined as

$$\hat{\mathbf{r}}_x = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)x_s(n). \tag{3.14}$$

The classical estimator $\hat{\mathbf{w}}_{\text{class}}$ for the solution of the LWF (3.9) is given by

$$\hat{\mathbf{w}}_{\text{class}} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}_x. \tag{3.15}$$

3.5.1 Enhanced estimator for the Linear Wiener Filter

It is well-known that the classical LWF estimator (3.15) is a N-consistent estimator of the LWF solution, i.e., $|\hat{\mathbf{w}}_{\text{class}} - \mathbf{w}| \to 0$, as $N \to \infty$.

In practice, $\hat{\mathbf{w}}_{\text{class}}$ provides good estimates when the training phase N is sufficiently large compared to the observation dimension M. However, when $M \to \infty$, while $M/N \to c \in (0, 1)$, it does not necessary provide N, M-consistency (indeed, [Mes08] shows that (3.15) is not N, M-consistent), and better estimators can be derived. Mathematically,

$$|\hat{\mathbf{w}}_{\text{class}} - \mathbf{w}| \nrightarrow 0, \text{ as } N, M \to \infty;\ M/N \to c. \tag{3.16}$$

In practice, it may occur when the training phase is short and comparable in magnitude with the dimension of the observation vector.

In the literature of consistent estimation, structures of the type of (3.9) are usually addressed assuming that the vector $\mathbf{r}_x$ is a non-random deterministic vector [Mes06]. Thus, from the best of the author's knowledge, the estimation of (3.9) where both $\mathbf{R}$ and $\mathbf{r}_x$ are random and statistically *dependent* is still an open problem. However we have checked using numerical simulations that the results for the random case addressed here behaves similarly to what is expected for the case where $\mathbf{r}_x$ is deterministic. The deterministic case is already solved in the RMT literature, e.g., [Gir98]. Considering this, we can improve the classical estimator in (3.15) and propose an enhanced estimator for the LWF.

We focus on the estimation of scalar functionals of the inverse of $\hat{\mathbf{R}}$, i.e., $\varphi(\hat{\mathbf{R}}^{-1}) : \mathbb{R}^{M\times M} \rightarrow \mathbb{R}$ of the type:

$$\varphi(\hat{\mathbf{R}}^{-1}) = \mathbf{a}^H \mathbf{R}^{-1} \mathbf{b}. \tag{3.17}$$

First let us to introduce the following definitions that can be found in [Tul04] and [Mes05]. Let $\mathbf{A}$ denote a generic positive semidefinite $M \times M$ matrix.

Definition 3.1 *Let the function $F_{\mathbf{A}} : \mathbb{R} \rightarrow [0, 1]$ be the empirical spectral distribution of the eigenvalues of $\mathbf{A}$, here denoted as λ_m:*

$$F_{\mathbf{A}}(x) = \frac{1}{M} \sum_{m=1}^{M} \mathcal{I}(\lambda_m \leq x), \tag{3.18}$$

whose Stieltjes transform *is defined by (for both the continuous and the finite size cases)*

$$s_{\mathbf{A}}(z) = \int \frac{1}{\lambda - z} dF_{\mathbf{A}}(\lambda) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\lambda_m - z}. \tag{3.19}$$

Definition 3.2 *Let the function* $H_{\mathbf{A}} : \mathbb{R} \rightarrow [0,1]$ *be an instance of the empirical distribution of the eigenvalues and eigenvectors of* $\mathbf{A}$*, denoted as* $\boldsymbol{\nu}_m$*:*

$$H_{\mathbf{A}}(x) = \sum_{m=1}^{M} \mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b}\, \mathcal{I}(\lambda_m \leq x), \tag{3.20}$$

whose Stieltjes transform is defined by (again for both the continuous and the finite size cases)

$$\begin{aligned} m_{\mathbf{A}}(z) &= \int \frac{1}{\lambda - z} dH_{\mathbf{A}}(\lambda) = \sum_{m=1}^{M} \frac{\mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b}}{\lambda_m - z} \\ &= \mathbf{a}^H \left(\mathbf{A} - z\mathbf{I}_M\right)^{-1} \mathbf{b}, \quad z \in \mathbb{C}. \end{aligned} \tag{3.21}$$

Furthermore, we make use of the *Marčenco-Pastur Theorem* [Mar67, Theorem 1], for matrices of the form $\boldsymbol{\Phi} = \boldsymbol{\Upsilon} + \frac{1}{N}\boldsymbol{\Xi}\mathbf{R}\boldsymbol{\Xi}^H$, where:

- $\boldsymbol{\Upsilon}$ is an arbitrary Hermitian $N \times N$ matrix.
- $\boldsymbol{\Xi}$ is an $N \times M$ matrix such that its entries are iid complex random variables with zero mean and variance 1, i.e. $[\boldsymbol{\Xi}]_{i,j} \in \mathbb{C}, \mathbb{E}[\boldsymbol{\Xi}_{i,j}] = 0$ and $\mathbb{E}[\|\boldsymbol{\Xi}_{i,j}\|^2] = 1$.
- $\mathbf{R}$ is the true covariance matrix, and the empirical distribution function of its eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ converges almost surely in distribution to a nonrandom cumulative distribution function $F_{\mathbf{R}}(\lambda)$ as $N \rightarrow \infty$.

Then, the Steltjes transform of $F_{\boldsymbol{\Phi}}$ can be written as

$$s_{\boldsymbol{\Phi}}(z) = s_{\boldsymbol{\Upsilon}}\left(z - c\int \frac{\lambda dF_{\mathbf{R}}(\lambda)}{1 + \lambda s_{\boldsymbol{\Phi}}(z)}\right). \tag{3.22}$$

Furthermore, we assume

$$\boldsymbol{\Upsilon} = \mathbf{0}_N, \text{and } \boldsymbol{\Phi} = \frac{1}{N}\boldsymbol{\Xi}\mathbf{R}\boldsymbol{\Xi}^H = \frac{1}{N}\boldsymbol{\Xi}\mathbf{R}^{1/2}\mathbf{R}^{1/2}\boldsymbol{\Xi}^H. \tag{3.23}$$

Hence, the Stieltjes transform of $\boldsymbol{\Upsilon}$ is given by

$$s_{\boldsymbol{\Upsilon}}(z) = \frac{1}{0-z} = -z^{-1}. \tag{3.24}$$

Using (3.22), we get the equation

$$s_{\boldsymbol{\Phi}}(z) = -\left(z - c\int \frac{\lambda dF_{\mathbf{R}}(\lambda)}{1+\lambda s_{\boldsymbol{\Phi}}(z)}\right)^{-1}. \tag{3.25}$$

Typically, to make $\boldsymbol{\Phi}$ define an arbitrary SCM $\hat{\mathbf{R}}$, it should have dimension $M \times M$ rather than $N \times N$. So, we introduce the $M \times M$ SCM as $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}^H\boldsymbol{\Xi}\mathbf{R}^{1/2}$. Matrix $\hat{\mathbf{R}}$ has the same structure than in (3.13).

Note that the non-zero eigenvalues of $\boldsymbol{\Phi}$ and $\hat{\mathbf{R}}$ are the same, however $\boldsymbol{\Phi}$ has $N-M$ zero eigenvalues extra. So, we can relate the eigenvalue distributions (and hence their Stieltjes transforms [Bai07]) for both $\boldsymbol{\Phi}$ and $\hat{\mathbf{R}}$ as follows,

$$\begin{aligned}
\frac{dF_{\Phi}}{d\lambda} &= \frac{M}{N}\frac{dF_{\hat{\mathbf{R}}}}{d\lambda} + \frac{(N-M)}{N}\delta(\lambda), \\
F_{\boldsymbol{\Phi}} &= \frac{M}{N}F_{\hat{\mathbf{R}}} + \frac{(N-M)}{N}u(\lambda), \\
s_{\boldsymbol{\Phi}}(z) &= cs_{\hat{\mathbf{R}}}(z) - \frac{(1-c)}{z}.
\end{aligned} \tag{3.26}$$

Substituting (3.26) in (3.25), and after some algebraic manipulations, we obtain (for both the continuous and the finite size approach)

$$s_{\hat{\mathbf{R}}}(z) = \int \frac{dF_{\mathbf{R}}(\lambda)}{\left(1-c-czs_{\hat{\mathbf{R}}}(z)\right)\lambda - z}, \tag{3.27}$$

$$= \frac{1}{M}\sum_{m=1}^{M} \frac{1}{\left(1-c-czs_{\hat{\mathbf{R}}}(z)\right)\lambda_m - z}. \tag{3.28}$$

Now, let the function $H_{\mathbf{R}}(x)$ be an instance of the empirical distribution of the eigenvalues (denoted as λ_m) and eigenvectors of $\mathbf{R}$ (denoted as $\boldsymbol{\nu}_m$)

as in [Mes05]:

$$H_{\mathbf{R}}(x) = \sum_{m=1}^{M} \mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b}\, \mathcal{I}(\lambda_m \leq x), \tag{3.29}$$

whose Stieltjes transform is defined by (for both the continuous and the finite size cases)

$$\begin{aligned} m_{\mathbf{R}}(z) &= \int \frac{1}{\lambda - z} dH_{\mathbf{R}}(\lambda), \\ &= \sum_{m=1}^{M} \frac{\mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b}}{\lambda_m - z} \\ &= \mathbf{a}^H \left(\mathbf{R} - z\mathbf{I}_M\right)^{-1} \mathbf{b}, \quad z \in \mathbb{C}. \end{aligned} \tag{3.30}$$

where vectors $\mathbf{a}$ and $\mathbf{b}$ are two generic and deterministic vectors.

Following some general assumptions, the asymptotic behavior of $s_{\mathbf{A}}(z)$ and $m_{\mathbf{A}}(z)$ is the same [Mes08, Th. 1], and hence one can apply the results above for $m_{\hat{\mathbf{R}}}$, and evaluate it for the case of $z = 0$. Then

$$\begin{aligned} m_{\hat{\mathbf{R}}}(z) \asymp \int \frac{dH_{\mathbf{R}}(\lambda)}{w(z)\lambda - z} &= \frac{1}{M} \sum_{m=1}^{M} \frac{\mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b}}{w(z)\lambda_m - z} \\ &= \mathbf{a}^H \left(w(z)\mathbf{R} - z\mathbf{I}\right)^{-1} \mathbf{b}, \end{aligned} \tag{3.31}$$

where $w(z) = 1 - c - czs_{\hat{\mathbf{R}}}(z)$, where $s_{\hat{\mathbf{R}}}(z)$ is defined in [Mes08] as the unique solution to the following equation in the set $\{s_{\hat{\mathbf{R}}}(z) \in \mathbb{C} : -(1-c)/z + cs_{\hat{\mathbf{R}}}(z) \in \mathbb{C}^+\}$:

$$s_{\hat{\mathbf{R}}}(z) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{w(z)\lambda_m - z}. \tag{3.32}$$

Evaluating $m_{\hat{\mathbf{R}}}(z)$ for the case of $z = 0$ one can easily observe that

$$(1-c)\, \mathbf{a}^H \hat{\mathbf{R}}^{-1} \mathbf{b} \asymp \mathbf{a}^H \mathbf{R}^{-1} \mathbf{b}. \tag{3.33}$$

In our case, the vector $\mathbf{a}$ is selected as an all-zero vector with a one at the ith position (usually represented as $\mathbf{e}_i$) and $\mathbf{b} = \hat{\mathbf{r}}_x$, then

$$[\mathbf{w}^\star]_i \asymp (1-c)\mathbf{e}_i^H \hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}_x. \tag{3.34}$$

Hence, an enhanced estimator of the LWF solution is given by

$$\hat{\mathbf{w}} = (1-c)\hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}_x. \tag{3.35}$$

The estimator in (3.35) can be seen as a scaled version of the classical LWF estimator as:

$$\hat{\mathbf{w}} = \alpha^\star \hat{\mathbf{w}}_{\text{class}}, \tag{3.36}$$

where α is a scaling factor and $\alpha^\star$ is its optimal value in terms of MSE and computed as

$$\alpha^\star = \underset{\alpha}{\text{argmin}}\{\text{MSE}(\alpha\hat{\mathbf{w}}_{\text{class}})\}. \tag{3.37}$$

We test by simulation that the minimum MSE is obtained when α is actually $\alpha^\star = (1-c)$ (see Fig. 3.4 of the Numerical Results section).

The intuition behind the estimator in (3.35) can be seen as follows: The parameter $\alpha \in (0,1)$ represents the confidence in the classical estimator. If $\hat{\mathbf{w}}_{\text{class}}$ has been estimated with a large number of samples in comparison with M, the degree of confidence will be high and $\hat{\mathbf{w}} \simeq \hat{\mathbf{w}}_{\text{class}}$ for $i = 1, \ldots, M$. Otherwise, when $N > M$ but comparable in magnitude, $\hat{\mathbf{w}}_{\text{class}}$ is not expected to be the best weighting vector. In order to mitigate the performance reduction due to the missadjustment in $\hat{\mathbf{w}}_{\text{class}}$, the vector is attenuated.

3.5.2 Enhanced estimator for the Mean Square Error

A traditional approach to estimate the MSE is by simply replacing the true correlations by their sample estimators. From (3.8), one can derive an estimator of the MSE given $\hat{\mathbf{w}}_{\text{class}}$ as

$$\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}}_{\text{class}}) = \hat{\sigma}_{x_s}^2 - \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}_x. \tag{3.38}$$

When $\hat{\mathbf{w}}$ is given, the theoretical expression of the MSE is:

$$\mathrm{MSE}(\hat{\mathbf{w}}) = \sigma_{x_s}^2 - 2\,\mathrm{Re}[\hat{\mathbf{w}}^H \mathbf{r}_x] + \hat{\mathbf{w}}^H \mathbf{R} \hat{\mathbf{w}}. \tag{3.39}$$

Using the classical approach, one can estimate $\mathrm{MSE}(\hat{\mathbf{w}})$ as:

$$\begin{aligned} \widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}}) &= \hat{\sigma}_{x_s}^2 - 2(1-c)\hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x \\ &+ (1-c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x, \end{aligned} \tag{3.40}$$

where $\hat{\sigma}_{x_s}^2$ is the sample estimator of the signal variance $\sigma_{x_s}^2$, defined as:

$$\hat{\sigma}_{x_s}^2 = \frac{1}{N} \sum_{n=1}^{N} x_s(n)^2. \tag{3.41}$$

The estimator $\widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}})$ is proved to be N-consistent (one can directly check the case when $c \to 0$), but indeed it is not consistent when the observation dimension M increases without bound and at the same rate as N.

In order to overcome this problem, we proposed an enhanced estimator of (3.39). The first two terms of $\mathrm{MSE}(\hat{\mathbf{w}})$ are directly estimated by their sample estimators, i.e., $\hat{\sigma}_{x_s}^2 - 2\,\mathrm{Re}[\hat{\mathbf{w}}^H \hat{\mathbf{r}}_x]$, since they do not involve unknown matrices in the estimation. Hence, the critical part resides in the estimation of the last term $(1-c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$ which is a function of the true covariance matrix. Hence we define the function

$$\beta(z) = \mathbf{h}_1^H (\hat{\mathbf{R}} - z\mathbf{I})^{-1} \mathbf{R} \mathbf{h}_2, \tag{3.42}$$

where $\mathbf{h}_1 = \hat{\mathbf{r}}_x$, and $\mathbf{h}_2 = \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$. Using the result in (3.31) we can rewrite $\beta(z)$ as

$$\beta(z) \asymp \mathbf{h}_1^H (w(z)\mathbf{R} - z\mathbf{I})^{-1} \mathbf{R} \mathbf{h}_2, \tag{3.43}$$

and evaluating $\beta(z)$ for $z = 0$, one can estimate $\beta(0)$ as

$$\hat{\beta} = (1-c)^{-1} \mathbf{h}_1^H \mathbf{h}_2. \tag{3.44}$$

Once we have the enhanced estimator $\hat{\beta} = \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$ for the term $\beta(0) = \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$, we substitute each term of (21) for its estimate. Hence, an enhanced estimator of MSE($\hat{\mathbf{w}}$) is given by:

$$\widehat{\mathrm{MSE}}(\hat{\mathbf{w}}) = \hat{\sigma}_{x_s}^2 - (1-c)\hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \tag{3.45}$$

Note that the approach taken in this chapter is slightly different to the MSE estimator in [Rub09], where the authors give an N, M-consistent estimator for the optimal MMSE. Otherwise, in this chapter we are interested in estimating the practical MSE obtained by using a certain weighting vector (in our case $\hat{\mathbf{w}}$ in (3.35)), which not necessarily provides the MMSE lower bound.

3.6 Numerical Results

The simulated scenario is composed of 200 sensing nodes and one fusion center configured in star topology. Their measurements are assumed to be space-time correlated following the correlation model $[\mathbf{R}_s]_{i,i+k} = [\mathbf{R}_t]_{i,i+k} = \rho^{|k|}$, where $\rho = 0.9$. Although any real application measurement will be corrupted by at least a small amount of noise, we have considered noise-free communication paths in order to better evaluate the system performance.

In particular, we study the behavior of the following figures as a function of c and how they affect to the SER performance:

1. MSE($\mathbf{w}$): the MSE for a given filter $\mathbf{w}$ in (3.8).
2. $\widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}}_{\mathrm{class}})$: the classical MSE estimator in (3.38).
3. $\widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}})$: the classical MSE estimator in (3.40).
4. $\widehat{\mathrm{MSE}}(\hat{\mathbf{w}})$: the proposed MSE estimator in (3.45).

Table 3.1: simulation parameters

Parameter	**Value**
Number of *fusion centers*:	$F = 1$
Already-decoded *sensing nodes*:	$S' = 200$
Number of past samples:	$K = 200$, hence $M = 400$
Length of the training phase:	$N = 1000$
Aspect ratio (M/N):	$c = 0.4$
Correlation model, $[\mathbf{R}_s]_{i,i+k} = [\mathbf{R}_t]_{i,i+k} = \rho^{\lvert k \rvert}$:	$\rho = 0.9$
SER threshold:	$\mathrm{SER_t} = 10^{-2}$
A/D converter depth:	$l = 12$ bits

The 2-4) have been computed using the corresponding mathematical expressions. On the other hand, the MSE($\mathbf{w}$) has been computed experimentally. We have developed a WSN simulation environment in Matlab and we have implemented in it our proposed DSC algorithm.

Table 5.1 summarizes the parameters that configure the basic setup of the simulation environment.

3.6.1 Performance of the proposed LWF estimator, $\hat{\mathbf{w}}$

In this subsection, we evaluate the MSE performance obtained by simulation of the proposed LWF estimator involved in the side-information $y(n)$.

Following the approach exposed in Subsection 3.5.1, Fig. 3.4 draws the simulation results for the MSE obtained with the LWF as a function of the parameter α and for different configurations of M and N. From this simulation experiment, we can compute the optimal α as in (3.37), which is represented in Fig. 3.4 as solid line with markers +. Moreover, we show

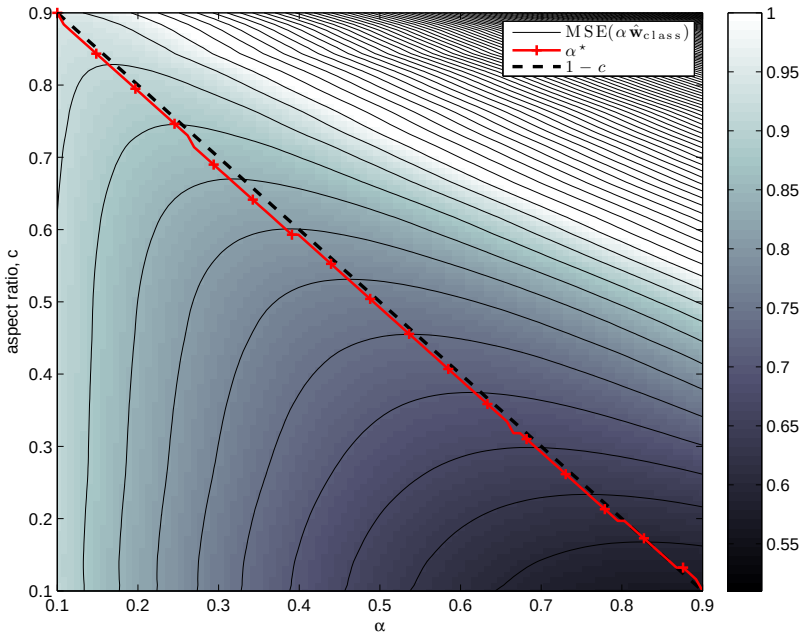

Figure 3.4: MSE performance of the LWF estimator for different values of α is represented by the colormap and the contour lines. The optimal α obtained as in (3.37) (solid line with marker +) is also compared with the expression $1 - c$ (dashed line). The configuration is $M = 200$ and $c = [0.1,\ 0.9]$. This figure has been averaged over 100 realizations.

that $\alpha^\star$ actually fits with the theoretical limit $\alpha^\star = 1 - c$ (dashed line), as predicted in Section 3.5.

Fig. 3.5 compares the performance in terms of MSE of our proposed estimator with some of the most popular estimation techniques, i.e., the classical sample estimator (3.15), the sample estimator with Diagonal Loading (DL), and three instances of the Principal Component Analysis (PCA) method.

However, there is still a gap between the MSE obtained with the proposed method and the one obtained assuming full correlation knowledge of $\mathbf{R}$ and $\mathbf{r}_x$.

3.6.1.1 Classical sample estimator

The behavior is clear; for low values of c^{-1}, the proposed LWF estimator outperforms the classical method. On the other hand, when we let c^{-1} increase, both estimators perform similarly.

3.6.1.2 DL estimator

Namely,

$$\hat{\mathbf{w}}_{\mathrm{DL}} = (\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1} \hat{\mathbf{r}}_x. \tag{3.46}$$

Although for low values of c^{-1} DL presents lower MSE, our proposed method shows two important advantages; *i*) DL is not consistent when $c^{-1} \to \infty$, and *ii*) the optimum loading factor $\gamma^\star$ that minimizes the MSE may vary according to the scenario, and in the literature there is not a clear expression to obtain $\gamma^\star$ analytically but only iteratively or by simulation. We use $\gamma = 0.8$, which gives the minimum MSE for $c = 0.4$.

3.6.1.3 PCA estimator

Keeping only the $M' < M$ largest eigenvalues of $\hat{\mathbf{R}}$ (because the smallest are more difficult to be estimated and hence they may introduce higher

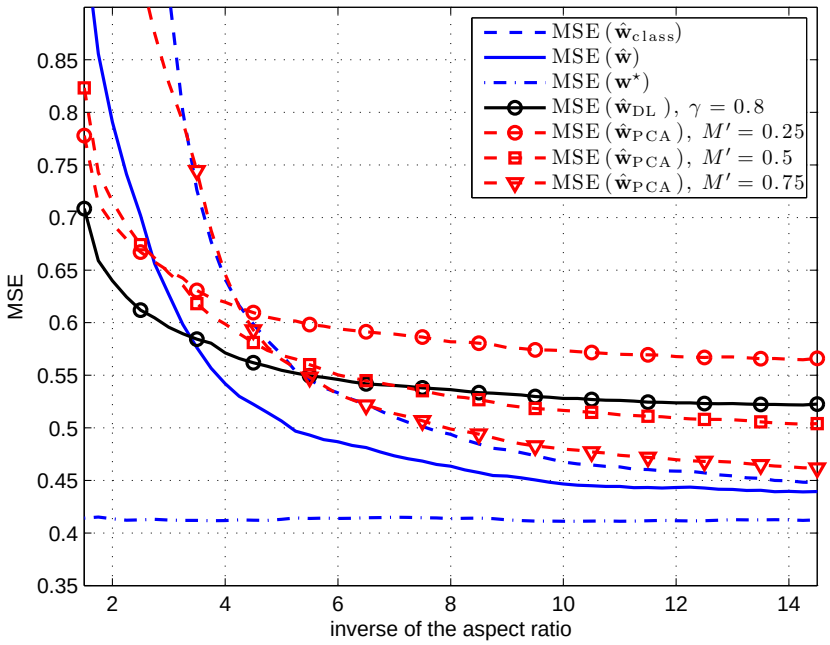

Figure 3.5: MSE performance of the classical, DL, PCA, and the proposed estimators as a function of the inverse of the aspect ratio c^{-1}.

errors), the MSE can be improved [Jol02]. Therefore, $\hat{\mathbf{R}}_{\text{PCA}}$ is a lower rank projection onto the subspace generated by the M' larger eigenvalues of $\hat{\mathbf{R}}$. Thus

$$\hat{\mathbf{w}}_{\text{PCA}} = \hat{\mathbf{R}}_{\text{PCA}}^{-1}\hat{\mathbf{r}}_x. \tag{3.47}$$

PCA presents the same limitations as DL but with the difference that this trade-off is balanced changing M'. In addition, when $M, N \to \infty$, the eigendecomposition may become hard to handle by practical small sensors.

In order to analyze the impact of the proposed estimator $\hat{\mathbf{w}}$ on the system performance, we compare in Fig. 3.10 the experimental SER for each of the LWF estimators as a function of the compression level $b(n)/l$.

One can observe that for large values of $b(n)/l$, e.g., $b(n)/l = 0.75$, the SER obtained is eight times smaller for $\hat{\mathbf{w}}$ than for $\hat{\mathbf{w}}_{\text{class}}$. Even so, one may make the following argument: If we want to achieve a certain SER_{t} (e.g. $\text{SER}_{\text{t}} = 10^{-2}$), we can compress up to 0.75 using the classical estimator, and 0.72 using the proposed. At first glance, it seems that the gain is quite moderate. However, we show next in Example 3.1 that it has an important impact on the total system performance when both the proposed estimators are combined.

3.6.2 Performance of the proposed MSE estimator, $\widehat{\text{MSE}}(\hat{\mathbf{w}})$

The MSE is involved in the computation of $b(n)$ in (3.12). Hence, a good estimation of the MSE is required in order to not overestimate (getting a too conservative result) or underestimate (inducing potential errors) the parameter $b(n)$, and thus maintain the system requirements, such as the SER_{t}.

Fig. 3.7 plots the MSE curve obtained experimentally (solid line). It is compared with our proposed MSE estimator (3.45) and the classical approach of (3.40). It is easy to see that our proposed estimator fits considerably better with the experimental results, while the classical estimator is clearly underestimating, especially for low values of c^{-1}. In fact, the

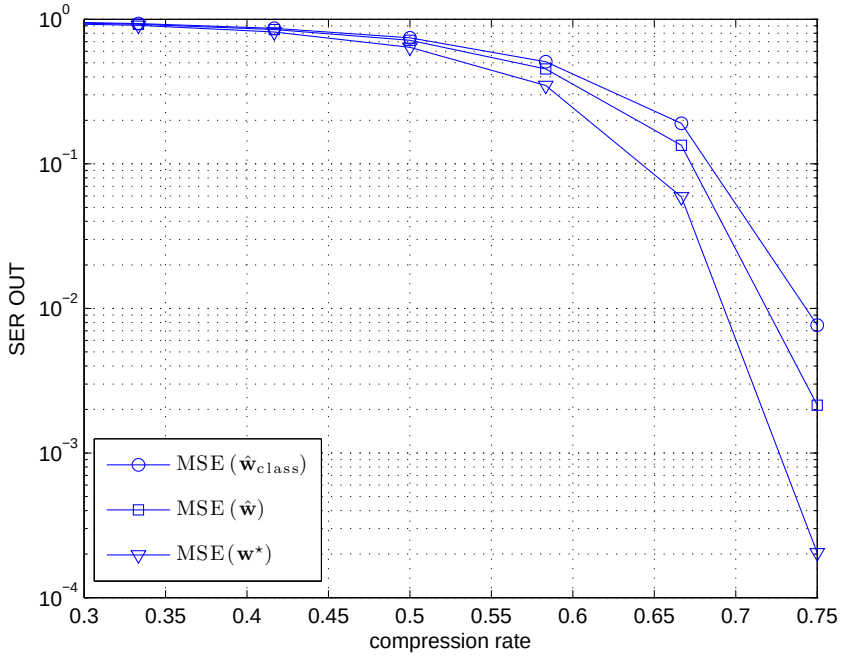

Figure 3.6: Performance of the experimental SER using the classical $\hat{\mathbf{w}}_{\mathrm{class}}$ and the proposed $\hat{\mathbf{w}}$ estimators as a function of the compression rate for $c = 0.4$.

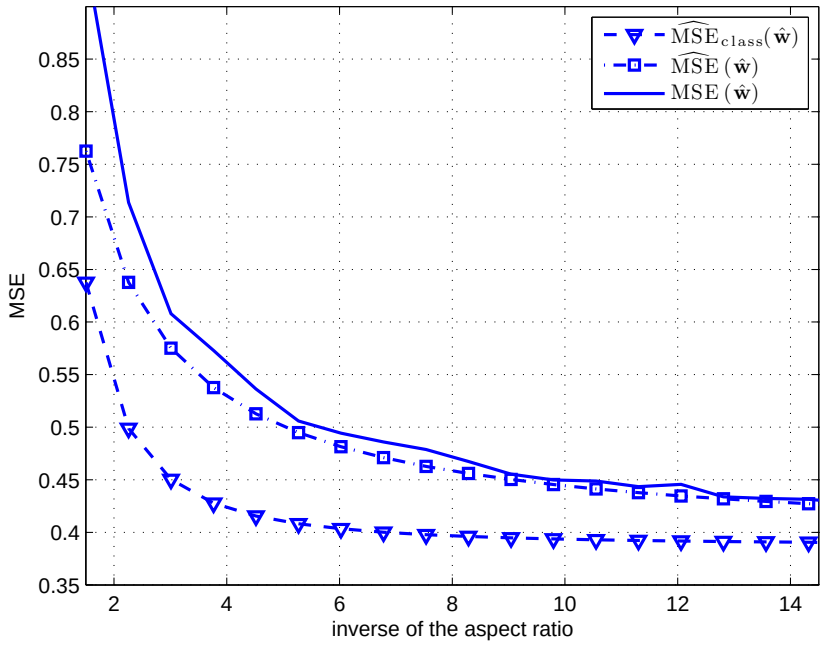

Figure 3.7: Performance of the classical $\widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}})$ and the proposed $\widehat{\mathrm{MSE}}(\hat{\mathbf{w}})$ estimators compared to the experimental reference $\mathrm{MSE}(\hat{\mathbf{w}})$ (solid line) as a function of the inverse of the aspect ratio c^{-1}.

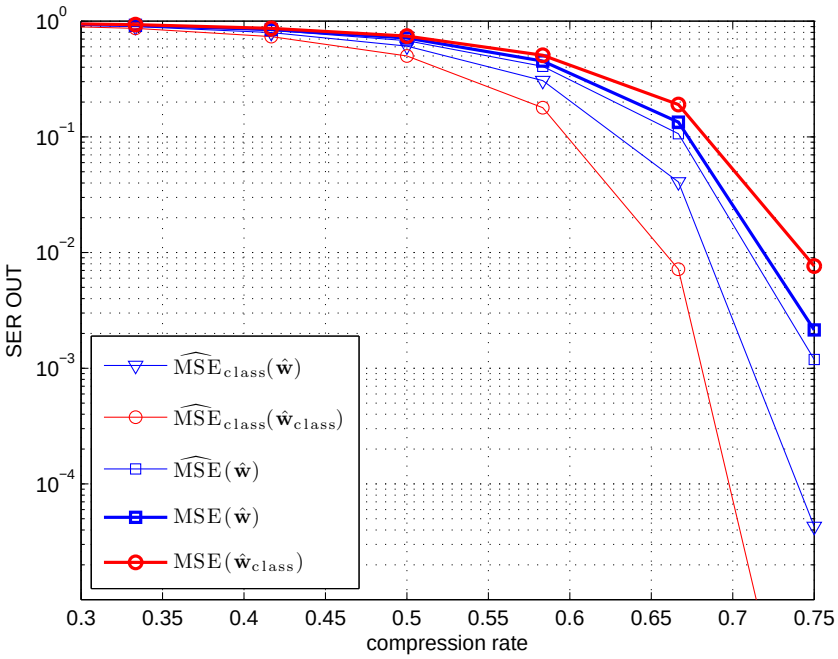

Figure 3.8: Performance of the predicted SER using different MSE estimators. It is compared to the experimental SER for $c = 0.4$.

classical approach is underestimating the MSE. Following (3.12), the DSC algorithm will be stingy with the number of bits used, and almost certainly, the SER requirements will not be achieved.

From a user point of view, the experimental SER curves of Fig. 3.10 are not available *a priori*, so the user should use a predicted version of the SER instead to determine which is the maximum compression rate that one can apply in order to guarantee a given SER_t.

In Fig. 3.8 we compare the experimental SER with the following:

1. Predicted SER when the classical MSE estimator is used and $\hat{\mathbf{w}}$ is given, i.e., $\widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}})$.

2. Predicted SER when the proposed MSE estimator is used and $\hat{\mathbf{w}}$ is given, i.e., $\widehat{\mathrm{MSE}}(\hat{\mathbf{w}})$.

3. Predicted SER when the classical MSE estimator is used and $\hat{\mathbf{w}}_{\mathrm{class}}$ is given, i.e., $\widehat{\mathrm{MSE}}_{\mathrm{class}}(\hat{\mathbf{w}}_{\mathrm{class}})$.

They are calculated using the formula (3.11) replacing the $\mathrm{MSE}(\mathbf{w})$ of the denominator by their respective estimators.

In Fig. 3.8 we observe that the proposed estimator curve fits the best with the experimental SER (which is also shown in Fig. 3.10). However, all the estimators are indeed underestimating. The consequences of this fact are illustrated in the following example.

Example 3.1 *Let us take as a system requirement* $SER_t = 10^{-2}$*. Observing the predicted SER curves in Fig. 3.8, we may decide to compress our messages with a ratio of* 0.66 *if we are using* $\widehat{MSE}_{class}(\hat{\mathbf{w}}_{class})$*,* 0.68 *if* $\widehat{MSE}_{class}(\hat{\mathbf{w}})$*, and 0.71 if* $\widehat{MSE}(\hat{\mathbf{w}})$*. Now, we map these three points to their respective experimental curves, i.e.,* $\widehat{MSE}_{class}(\hat{\mathbf{w}}_{class})$ *to* $MSE(\hat{\mathbf{w}}_{class})$*, and* $\widehat{MSE}_{class}(\hat{\mathbf{w}})$ *and* $\widehat{MSE}(\hat{\mathbf{w}})$ *to* $MSE(\hat{\mathbf{w}})$*. The real output SER of the system would be* $2 \cdot 10^{-1}$ *and about* 10^{-1} *for the first and second option respectively,*

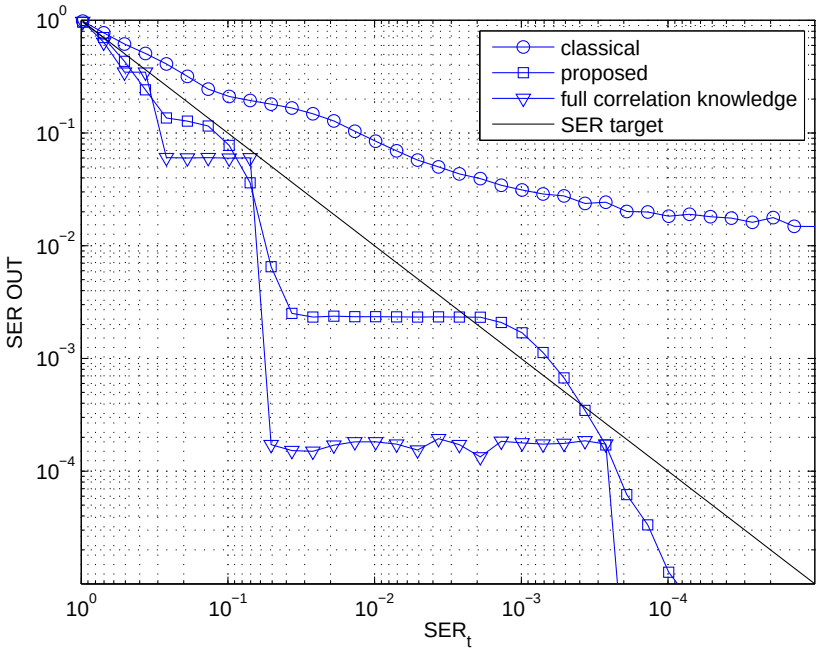

Figure 3.9: Comparison of the SER as a function of the $\text{SER}_\text{t} = 10^{-3}$ for both the classical and proposed methods for the case of $c = 0.4$. This figure has been averaged over 100 realizations.

which is one order of magnitude larger than the expected SER. On the other hand, we get $1.2{\cdot}10^{-2}$ *(instead of* 10^{-2}*, so it is still slightly underestimated), obtaining a more accurate solution.*

Hence, using the proposed estimators the gain in front of the classical methods is twofold; on the one hand we can obtain higher compression (thus higher energy savings) since $\mathrm{MSE}(\hat{\mathbf{w}}) < \mathrm{MSE}(\hat{\mathbf{w}}_{\mathrm{class}})$, and on the other hand the proposed estimators adjust substantially better to the system requirements than the classical estimators do. Moreover, the more stringent the $\mathrm{SER_t}$ the higher the gain.

3.6.3 Symbol Error Rate as a function of $\mathrm{SER_t}$

The final purpose of DSC is the reduction of the transmitted bits in order to reduce power consumption. However, system requirements must be taken into account in the design phase. Thus we analyze the performance in terms of the SER fidelity. In other words, we compare the SER obtained with the proposed and classical techniques as a function of the $\mathrm{SER_t}$.

Graphically, Fig. 3.9 shows the SER performance for the case $c = 0.4$. Ideally, the SER performance curve should be below but as close as possible to the SER threshold (the solid line in the figure). Our proposed method is actually below the threshold, except for a small area around $\mathrm{SER_t} = 10^{-3}$. The curve is staircase-shaped due to the ceiling function of (3.12).

On the contrary, the classical approach does not fit the system requirements. One possible solution to counteract this effect is to increase the training phase N (i.e., decrease c). This conclusion is supported by the following result.

3.6.4 Symbol Error Rate as a function of c^{-1}

We set the SER threshold of 10^{-2}. Fig. 3.10 shows that the tendency of the classical DSC approach is to fulfill the SER requirements only for high

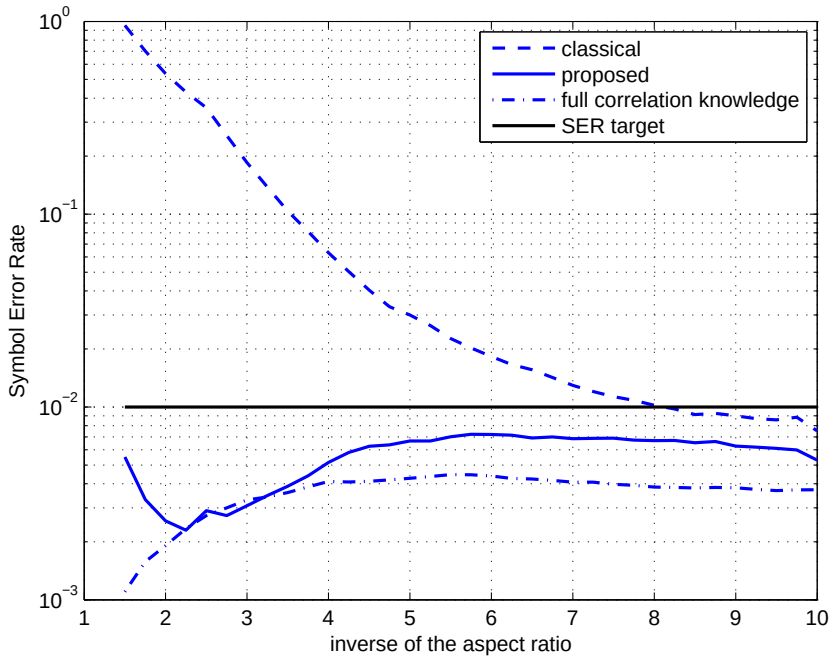

Figure 3.10: Comparison of the SER performance of classical and the proposed methods imposing a $\mathrm{SER_t} = 10^{-2}$. It has been averaged over 100 realizations.

values of c^{-1} due to its N-consistency property. However for low values of N (i.e., when the ratio c increases), it cannot fit with the system requirements. Therefore, the system experiments higher SER than expected.

One can observe that the classical method requires a training phase at least eight times longer than the observation dimension M to fulfill the $\mathrm{SER}_{\mathrm{t}}$. On the other hand, using the proposed estimators, one can guarantee the requirements even for values of c close to one.

3.7 Conclusions

This chapter has proposed two enhanced correlation estimators for the Linear Wiener Filter and the Mean Square Error to operate when the number of snapshots N and the dimension of the observation vector M are large and comparable in magnitude, or equivalently for short training phases. This scenario is very suitable for large WSNs due to its large number of sensors. Concretely, the enhanced estimators have been designed to carry out the two key steps in a Distributed Source Coding algorithm, i.e., the computation of the side-information $y(n)$ based on the existing space-time correlations, and the computation of the minimum number of bits to encode the readings in order to guarantee a certain Symbol Error Rate.

Numerical results show that our proposed estimators perform far better for values of the aspect ratio M/N close to one. Furthermore, they perform as the corresponding sample estimators when $M/N \to 0$ (i.e., for very long training phases). In practice, it allows us to reduce the number of transmitted bits (and hence reduce the energy consumption) at the same time that the system requirements (in terms of maximum Symbol Error Rate) are guaranteed. On the contrary, it does not happen when the conventional estimators are used. Therefore, it allows us to decrease largely the training phase in Distributed Source Coding schemes.

Chapter 4

Amplify-and-Forward Compressed Sensing as an Energy-Efficient Solution

4.1 Summary

Sensor measurements typically show space-time correlations and compressed sensing techniques can exploit this feature in order to improve the existing trade-off among reconstruction error, energy consumption and resource utilization. However, due to the distributed structure of wireless sensor networks, most of the compressed sensing algorithms are hardly applicable, or at least they are very costly in terms of signaling due to the centralized nature of compressed sensing approaches. In this chapter, we propose a novel distributed compressed sensing transmission scheme which is referred to as amplify-and-forward compressed sensing (AF-CS). The underlying idea is twofold: First, we take advantage of the time correlation in order to produce sparse versions of the signal vector, which collects the transmitted signals of the sensors. Second, we take advantage of the multiple access channel in order to perform random measurements of the signal

vector. We also propose a simple model that accurately approximates the distortion introduced by the proposed scheme. It allows us to dimension the network (i.e., number of active nodes and relays) based on a cost function that controls the trade-off between reconstruction error and energy consumption. Simulation results show that our proposed algorithm outperforms other techniques in terms of distortion and number of transmissions, providing at the same time, energy savings and significant reduction in the number of channel uses.

4.2 Introduction

Wireless Sensor Networks (WSNs) design is currently one of the most challenging topics in the wireless communications field. In particular, WSNs are severely energy-constrained because they consist of many small, cheap and power limited nodes, whose batteries cannot be recharged in most cases. Hence, the application of energy-efficient algorithms turns out to be crucial.

4.2.1 Is the compressed sensing a good candidate to build energy-efficient WSNs?

Usually, WSNs are designed to perform one specific task such as the detection of some chemical agents; the measurement of temperature, humidity or light; location, estimation and positioning. Hence, each node senses some specific physical magnitude from the environment. When the measurements $x(n)$, $x(n-1)$, $x(n-2)$, and so on, are sampled at a sufficiently high rate, the elements of the signal of interest are assumed to be time-correlated, i.e.,

$$\mathbb{E}[x(n)x(n-t)] \neq 0, \tag{4.1}$$

where t is the sampling difference between samples $x(n)$ and $x(n-t)$.

Furthermore, as the WSNs are conformed of many different sensing nodes, the measures at a given sample time n are not totally independent,

and therefore spatial correlation among nodes is also assumed. In the same way,

$$\mathbb{E}[x_i(n)x_j(n)] \neq 0, \quad i,j \in \mathcal{S}, \tag{4.2}$$

where $x_i(n)$ and $x_j(n)$ are the measurements of the ith and the jth sensing nodes respectively, and $\mathcal{S}$ describes the set of total sensing nodes in the WSN.

Under space-time correlated assumptions, the resulting signal turns out to be a smooth and slowly varying signal in the time and space domains. Hence, the level of "uncertainty" is quite low in such signals. This is the reason why Compressed Sensing (CS) appears as a good candidate to take advantage of this fact.

In a nutshell, CS allows to recover a given signal $\mathbf{x} \in \mathbb{R}^S$ (under some assumptions) from a small number of measurements. This is possible when $\mathbf{x}$ can be accurately (or exactly) represented by a linear combination of K vectors taken form a desirable basis $\mathbf{\Psi}$ [Don06b]. If so, we say that the vector $\mathbf{x}$ is compressible or that it has a K-sparse representation in $\mathbf{\Psi}$. Roughly speaking, the CS theory says that it is possible to recover the signal $\mathbf{x}$ with a number of measurements proportional (up to a logarithmic factor [Can06a]) to the information contained in the signal $\mathbf{x}$, i.e. K, instead of to the number of samples S.

As we can easily see, the CS framework is very suitable for WSN with correlated sources. In theory we are able to recover a given signal $\mathbf{x}$ with a number of measurements much smaller than the original number of samples. This approach reduces notably the number of measurements, hence it turns out to be an energy-efficient solution scheme for WSNs. In other words, the question formulated at the title of the section has a positive answer.

4.2.2 Are the current CS schemes good energy-efficient solutions for WSNs?

Recent signal processing results exploit Compressed Sensing (CS) techniques as a powerful solution to compress the information based on the fact that the signal has an approximate sparse representation in a given transformed linear basis. However, the application of CS to *distributed* systems is not straightforward and this is why CS has not been widely extended to WSNs yet.

Before answering the question of the title, we review some of the different CS-based techniques that have been recently appeared in WSN literature. Authors in [Men09] propose a *detection* technique that uses CS in order to significantly reduce the number of active sensors or, what is the same, to reduce the sampling rate. In [Nik11], a *localization* mechanism exploits the sparse nature of the position of the nodes within a grid to apply CS techniques. Although that paper presents the results for a wireless local area network, they could also be extended to a WSN scenario.

The CS literature related to *distributed communications* is quite heterogeneous, nevertheless we point out some examples of different applications. Authors of [Luo09, Gup05] deal with data compression in tree-based networks. The compression is carried out by the data gathering nodes, but all the sensors need to be active. In [Cho09], projection methods for multihop networks are proposed. The message is distributed from the source to the sensors following a given route (projection), where each node adds its measurement. In such techniques all sensors have to be listening, and for the generic case, they need to transmit once per measurement. It results in over-expensive star-WSNs in terms of energy consumption. In order to mitigate this effect, the authors of [Cho09] present some heuristics that modifies the CS scheme.

Probably one of the most relevant work (and also one of the most referenced) in the field of CS applied to WSNs is [Baj06], where the authors first

proposed what they called Compressive Wireless Sensing (CWS). Although we do not go into the details of the work in [Baj06], we briefly explain the main idea and extract some straightforward conclusions[1].

Let $\mathbf{x}(n) \in \mathbb{R}^S$ be the measured vector at the sample time n,

$$\mathbf{x}(n) = [x_1(n)\, x_2(n)\, \ldots\, x_S(n)]^T, \tag{4.3}$$

where $x_s(n)$ is the measured signal of the sth sensor at the sample time n. No assumptions are made for $\mathbf{x}(n)$ other than it is compressible in a certain (and previously fixed) orthonormal basis $\mathbf{\Psi} \in \mathbb{R}^{S\times S}$. Up to some scaling factors needed to meet the power constraints of the sensors and the corresponding additive noise of the wireless channel, the signal model is

$$\boldsymbol{\omega}(n) = \mathbf{\Psi}\mathbf{x}(n), \tag{4.4}$$

where the vector $\boldsymbol{\omega}(n) \in \mathbb{R}^S$ is the (pseudo) sparse representation of $\mathbf{x}(n)$ in the basis $\mathbf{\Psi}$. The paper [Baj06] includes an extra (and hard) assumption that is the knowledge of the ordering of the elements in $\boldsymbol{\omega}(n)$, or in other words, the authors in [Baj06] assume that the elements of $\boldsymbol{\omega}(n)$ are sorted in a descending absolute value, i.e.,

$$|[\boldsymbol{\omega}(n)_i]| \geq |[\boldsymbol{\omega}(n)]_j|, \quad i < j. \tag{4.5}$$

Furthermore, it defines $\hat{\mathbf{x}}(n)$ as the best K-term approximation of $x(n)$ in terms of $\mathbf{\Psi}$ as:

$$\hat{\mathbf{x}}(n) = \mathbf{\Psi}_K^T \boldsymbol{\omega}_K(n), \tag{4.6}$$

where $\mathbf{\Psi}_K \in \mathbb{R}^{S\times S}$ is a replica of the first K rows of $\mathbf{\Psi}$ and the rest are zeros, and $\boldsymbol{\omega}_K(n) \in \mathbb{R}^S$ is a replica of the first K entries of $\boldsymbol{\omega}(n)$ and the rest are zeros. In order to simulate these computations from a WSN perspective, they propose an iterative algorithm where each sensor sends in

[1] let us make use of our notation for the sake of homogeneity of the document.

the kth round the measurement $[\mathbf{\Psi}]_{k,s}[\mathbf{x}]_s(n)$ to the fusion center following an uncoded coherent transmission scheme as in [Che06] and [Gas03]. Under such transmission schemes, the fusion center receives (again up to some constants and noise)

$$[\boldsymbol{\omega}(n)]_k = \sum_{s=1}^{S} [\mathbf{\Psi}]_{k,s} x_s(n). \tag{4.7}$$

This transmission should be repeated K times in order to get an approximation of $\hat{\mathbf{x}}(n)$ and the system uses $K \ll S$ channel uses for a total of $K \times S$ transmissions. Similarly, we can compare this algorithm with the classical approach where each sensor uses one channel use to send its measurement to the fusion center. Similarly to (4.8), and ignoring the channel loses and the noise, the signal at the fusion center can be modeled as:

$$\boldsymbol{\omega}(n) = \mathbf{I}_S \mathbf{x}(n), \tag{4.8}$$

where it is clear that "only" S transmissions are used. So, in theory, the transmission scheme proposed in [Baj06] will be energy efficient if the matrix $\mathbf{\Psi}$ holds the following condition,

$$\|\mathbf{\Psi}_K\|^2 < \|\mathbf{I}_S\|^2 = S, \tag{4.9}$$

where $\|\mathbf{A}\|$ indicates the Frobenius norm of a matrix $\mathbf{A}$ of dimension $A \times B$ and it is defined as:

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{A} \sum_{j=1}^{B} |[\mathbf{A}]_{i,j}|^2} \tag{4.10}$$

Following the same approach than in [Baj06], but relaxing the assumption in (4.5), we reformulate the problem in a more general form (using a general CS framework such as the one in [Hau08]). This is crucial given the fact that the assumption of the knowledge of the ordering of $\boldsymbol{\omega}(n)$ directly

violates one of the most important premises in CS, that is: it is not required the knowledge of the location of the K principal components of $\boldsymbol{\omega}(n)$ and one can still recover the original vector with a number of measurements proportional to K (up to a logarithmic factor) [Can06a].

This modification is referred as Generalized Compressed Wireless Sensing (GCWS) and it is detailed as follows

$$\mathbf{y}(n) = \boldsymbol{\Phi}\boldsymbol{\omega}(n) = \boldsymbol{\Phi}\boldsymbol{\Psi}\mathbf{x}(n), \tag{4.11}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{R\times S}$ is the sensing matrix (it will be properly defined next) that models R generic and nonadaptive measurements of $\boldsymbol{\omega}(n)$. In fact, the reconstruction will be exact if $\boldsymbol{\omega}(n)$ is sparse and the number of measurements is proportional to its sparseness. Now we reformulate the energy-efficient condition in (4.9) according to this more generic scheme as,

$$\|\boldsymbol{\Phi}\boldsymbol{\Psi}\|^2 < \|\mathbf{I}_S\|^2 = S. \tag{4.12}$$

Albeit this is easily met for some pairs $\boldsymbol{\Phi}\boldsymbol{\Psi}$, the energy-efficient condition in (4.9) and (4.12) is far to be realistic. This is because it is proved that the sensors spend most of their power while they are active and, otherwise, the energy cost is negligible while they remain silent (in sleep mode) [Rug07]. We can graphically see this behavior in Fig. 4.1 with the data borrowed from [Rug07]. It shows us that minimizing the total amount of transmitted power at the expense of the number of transmissions turns out to be not the best option in terms of energy consumption. Instead, it seems that we can obtain better results with schemes focused on directly reducing the number of transmissions (in consequence, they also reduce the total transmission power and meets the condition in (4.9) is still satisfied).

Hence, the answer of the question at the heading of this section is that better and more realistic energy-efficient algorithms for WSNs can be formulated.

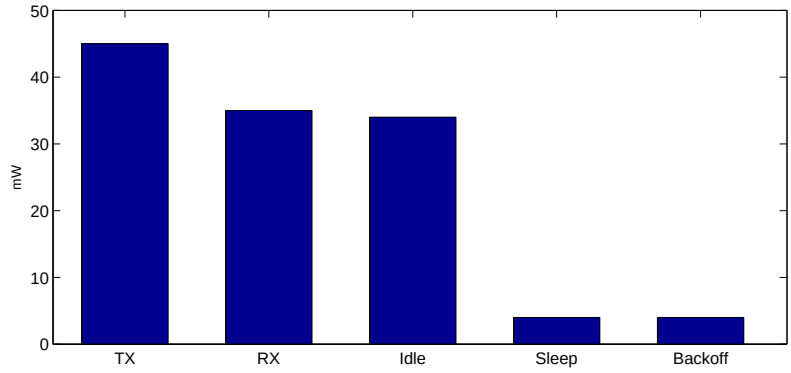

Figure 4.1: Sensor node power requirements.
This figure is not obtained by any experiment carried out by the authors, instead it is borrowed from [Rug07].

4.2.3 Our contribution

The main contribution of this chapter is to provide an energy-efficient compressed sensing approach for WSNs. The resulting technique is called *Amplify-and-Forward Compressed Sensing* (AF-CS) that relax some of the strong assumptions or limitations introduced by the other works in the literature, as it is explained next. The AF-CS was first introduced in [BL11] and has been detailed in [BL12b].

We do not assume previous knowledge of either the position or the ordering of the principal components of the transformed vector. This knowledge would only be available with the presence of an "oracle". Furthermore, we do not assume pure sparse input vectors. On the contrary, our scheme may face with more general signals, which are not exactly sparse (also called pseudosparse). Indeed, our proposed design is able to obtain a pure sparse version of the signal (using lossy transformations) in a fully distributed

way, which (up to the best of the author's knowledge) is novel in the literature about CS applied to distributed communications. This approach is notably more realistic in the sense that signals corresponding to physical phenomena are never exactly sparse even in a transformed domain.

As we will see in detail next, the sparse version of the input signal can be obtained in a distributed way by using downsampling encoding. A downsampling encoder with ratio $\gamma = K/S$ that ensures an average of K active sensors at a time. In particular, we assume throughout the explanation of AF-CS that the Conditional Downsampling Encoder (CDE) is used at the sensing node side (introduced and detailed in Chapter 2). We have selected it because of its low Minimum Square Error (MSE) at the reconstruction . The price to pay for such a good performance is the need to know the time correlation coefficients among the samples of the same sensor or at least a good estimations of them (the correlation estimation issue is also extensively revised in Chapter 3). It will be clear that taking advantage of the time correlation instead of the spatial correlation keeps the distributed character of the proposed technique. This is due to that the time information only depends on the sensor itself and not on the measurements from other sensors.

Furthermore, the proposed scheme reduces the required number of transmissions. This technique is based on the spatial diversity that is introduced using several relays. Hence, R sensors act as relays performing random measurements $\mathbf{y}(n)$ over the set of $\mathcal{S}$ sensors. These random measurements are in fact achieved by the proper random nature of the Multiple Access Channel (MAC) taking advantage of its randomness. On the other hand, the scheme in [Baj06] tries to compensate it with power control techniques. These coeffitients are in fact represented by the sensing matrix $\mathbf{\Phi}$ previously introduced in (4.11). Nevertheless, it implies perfect channel state information at the receiver (CSIR), which the reader will know that it is a common assumption in the wireless communication and signal

processing literature.

One of the main problems (if not the biggest) is the selection of the minimum number of measurements R required to obtain an accurate recovery of the signal of interest. At the time of writing this chapter, the literature agrees that a signal with K non-zero entries can be faithfully recovered from about $K \log S$ random measurements, i.e.,

$$R \leq C_0 K \log S, \tag{4.13}$$

where the only information we know about C_0 is that it is a positive (and not very big) constant.

The reconstruction of the signal $\boldsymbol{\omega}_K(n)$ is carried out by the fusion center. The most extended CS decoder (for the noiseless case) is the l^1-norm minimization program $\mathcal{P}1$,

$$\begin{aligned} \mathcal{P}1: \quad & \underset{\hat{\boldsymbol{\omega}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\hat{\boldsymbol{\omega}}_K(n)\|_{l^1} \\ & \text{subject to} \quad \mathbf{y}(n) = \boldsymbol{\Phi}\hat{\boldsymbol{\omega}}_K(n), \end{aligned} \tag{4.14}$$

which is a convex relaxation of the original NP-hard problem $\mathcal{P}_0$ (with the l^0-norm) as in e.g., [Don06b], [Can06a]:

$$\begin{aligned} \mathcal{P}0: \quad & \underset{\hat{\boldsymbol{\omega}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\hat{\boldsymbol{\omega}}_K(n)\|_{l^0} \\ & \text{subject to} \quad \mathbf{y}(n) = \boldsymbol{\Phi}\hat{\boldsymbol{\omega}}_K(n)), \end{aligned} \tag{4.15}$$

We also take into account the noisy case where the measurements of the relays are contaminated with some additive noise vector $\mathbf{w}(n) \in \mathbb{R}^R$ at the measurements, such as $\mathbf{w}(n) \sim \mathcal{N}(0, \sigma_w^2)$, and we study its impact in the reconstruction process. For this purpose, we consider the minimization program which solves the l^1-norm [Can08a],

$$\begin{aligned} \mathcal{P}2: \quad & \underset{\hat{\boldsymbol{\omega}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\hat{\boldsymbol{\omega}}_K(n)\|_{l^1} \\ & \text{subject to} \quad \|\mathbf{y}(n) - \boldsymbol{\Phi}\hat{\boldsymbol{\omega}}_K(n)\|_2 < \varepsilon, \end{aligned} \tag{4.16}$$

for ε as an upper bound on the magnitude of the noise, i.e, $\varepsilon \geq \|w(n)\|_2$. We can also find other approaches in the literature as the unconstrained LASSO problem for the noisy case in [Can11], [Par11].

In addition to the distributed CS scheme, we also propose a framework in order to model the distortion. This model allows us to predict the introduced error as a function of the system parameters (i.e. K, R, and S). This is extremely important since it allows us to also propose an analytical design scheme in order to control the existing trade-off between energy consumption and reconstruction accuracy.

4.2.4 Organization of the chapter

The rest of the chapter is organized as follows. In Section 5.3 we present the problem statement and the assumptions considered throughout the chapter. Section 4.4 explains the proposed CS algorithm. The expressions of the distortion analysis are detailed in Section 4.5. The cost function and the subsequent optimization problem are presented in Section 4.6. Simulation results are shown in Section 5.6, and conclusions are drawn in Section 5.7.

4.3 System Model and Assumptions

Let us consider a WSN configured in star-topology that monitors a given physical scalar magnitude (e.g., temperature or humidity). In particular, let us assume the scheme in Fig. 4.2, where:

- A set $\mathcal{S}$ of S sensing nodes is connected (wirelessly) to one fusion center.

- A subset $\mathcal{K}(n) \subseteq \mathcal{S}$ (of average cardinality K) of active sensors are transmitting at sample time n. The remaining sensors in $\mathcal{Q}(n) = \mathcal{S} \setminus \mathcal{K}(n)$ (of average cardinality Q) stay silent (sleep mode).

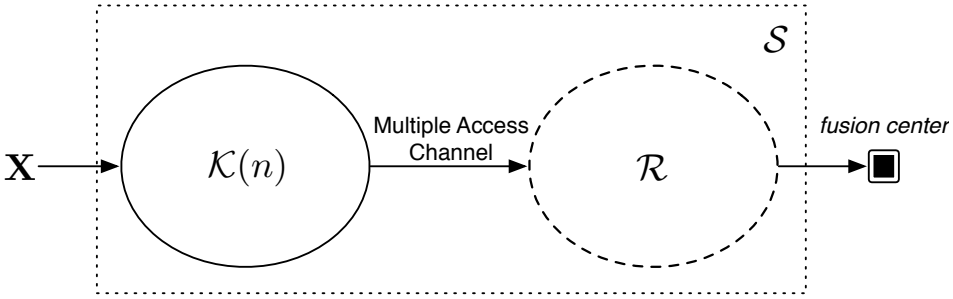

Figure 4.2: Multiple access channel scenario composed by K active sensing nodes, R relay nodes and one fusion center.

- A subset $\mathcal{R}(n) \subseteq \mathcal{S}$ (of cardinality R) acts as Amplify-and-Forward (AF) relay sensors at sample time n.

Note that the subsets $\mathcal{K}(n)$ and $\mathcal{R}(n)$ are not necessarily disjoint (i.e., they do not need to satisfy $\mathcal{K}(n) \cap \mathcal{R}(n) = \varnothing$).

In CS nomenclature, K also corresponds to the number of non-zero elements of the transmitted vector $\mathbf{x}(n) \in \mathbb{R}^S$, and R is the number of projections used in the reconstruction process, i.e., the number of rows of the sensing matrix, $\mathbf{\Phi} \in \mathbb{R}^{R \times S}$ (defined next in 4.20), where typically $K < R < S$.

We consider that the signals are space-time correlated and modeled as an S-dimensional stochastic process, namely,

$$\mathbf{X} = [\mathbf{x}(1)\,\mathbf{x}(2)\,\ldots\,\mathbf{x}(N)\,] = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(N) \\ x_2(1) & x_2(2) & \cdots & x_2(N) \\ \vdots & \vdots & & \vdots \\ x_S(1) & x_S(2) & \cdots & x_S(N) \end{bmatrix}, \quad (4.17)$$

where $x_s(n)$ denotes the measurement of the sth sensor at the sample time n and N denotes the number of time samples in the observation window.

The main assumptions throughout this chapter are collected as follows:

4.3.1 Assumptions on the signal model

Let $x(n)$ be a real and time-discrete auto-regressive model of order 1 ($AR-1$), which is commonly assumed in the signal processing literature in order to model real sources [Has80]. It is defined as:

$$x(n) = \rho x(n-1) + z(n), \quad \text{for } n = 1, 2, \dots \tag{4.18}$$

The auto-regression coefficient is denoted by $\rho \in [0, 1]$ and assumed to be constant during the transmission. The random process $z(n)$ is a sequence of Gaussian distributed and independent random variables with zero mean and variance σ_z^2.

As we showed in Chapter 2, the covariance matrix of an $AR-1$ model follows

$$[\mathbf{R}]_{n,n-i} = \rho^i. \tag{4.19}$$

That is why it is also referred to as the correlation factor.

Without loss of generality, we assume that $\sigma_x^2 = 1$. Therefore, following the Lemma 2.1, the variance of the noise is $\sigma_z^2 = 1 - \rho^2$.

We also assume that $x(n)$ is a continuous-valued process or, in other words, that the quantization error is assumed to be zero. Although a continuous random measurement requires infinite precision in order to be digitally sent with zero error, we assume the quantization error to be negligible in comparison with the distortion introduced by our proposed scheme. This assumption is supported by the rate-distortion theory [MC91], since we can select a symbol rate such that the quantization error is as small as desired. However, we want to stress out that this assumption only responds to pedagogical reasons, and that our proposed scheme works in practice with quantized signals as well.

4.3.2 Assumptions on the channel and the system model

We assume perfect channel state information (CSI) at the fusion center for all the links that go from a sensing node to a relay node. This assumption is quite common in the communications and signal processing literature. It is traditionally addressed with the transmission of a training sequence composed by pilot symbols with known amplitude that allows us to estimate the channel at the receiver side. The same can be achieved using blind methods that exploit the knowledge of the structure of the transmitted signals. In particular, we assume that the channel matrix, here relabeled as sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{R \times S}$ follows the Gaussian measurement ensemble, where:

$$[\mathbf{\Phi}]_{i,j} \sim \mathcal{N}(0, R^{-1}). \tag{4.20}$$

The variance of the sensing matrix R^{-1} is a convention in the literature in order to maintain the relation

$$\mathbb{E}\left[\|\mathbf{\Phi}\mathbf{x}\|^2\right] = \mathbb{E}\left[\|\mathbf{x}\|^2\right] \tag{4.21}$$

for an arbitrary vector $\mathbf{x}$. This assumption is just for convenience and it does not affect in the generality of the model since the channel gain can be adjusted at the receiver if needed.

Moreover, we do not assume anything regarding the links from $\mathcal{R}$ to the fusion center further than these links are controlled by a certain orthogonal policy that requires R channel uses for each sample time n in order to transmit the data from the relays to the fusion center.

4.4 Amplify and Forward Compressed Sensing

This section details our proposed CS scheme called AF-CS. It is divided in three phases, *i*) the sensing phase, *ii*) the projection phase, and *iii*) the signal reconstruction phase (see *i*) and *ii*) graphically represented in Fig. 4.3).

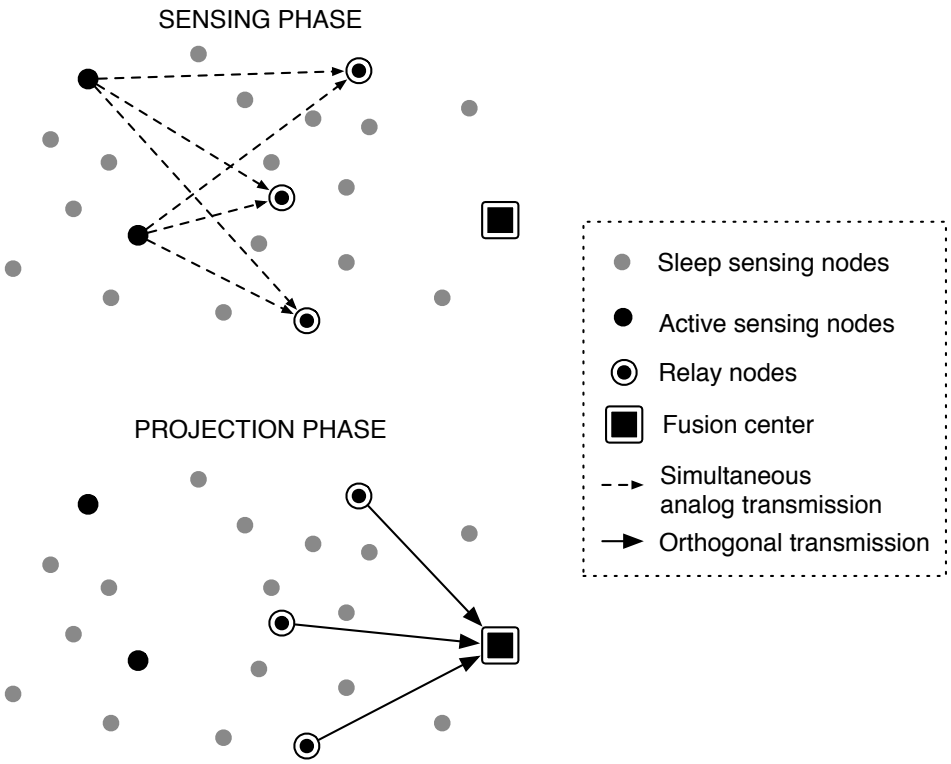

Figure 4.3: WSN scenario composed by K active sensing nodes, Q quiet sensing nodes, R relay sensors, and one fusion center. The signal reconstruction phase is carried out by the fusion center after the projection phase.

4.4.1 Sensing phase

Compressed sensing exploits sparsity to acquire high-dimensional signals represented in a low-dimensional subspace. A priori, physical phenomena are not necessarily sparse, but expected to be correlated in time and space. Thus, correlated signals may have a (pseudo)sparse representation if they are expressed in a proper basis. Wavelets are in general considered as a good candidate to construct the sparsity basis matrix, $\mathbf{\Psi} \in \mathbb{C}^{S\times S}$ [Hau08]. The main difficulty in projecting a signal onto a given wavelet basis is that this is a centralized problem. In order to carry out this task, some central entity is needed to gather all the measurements (i.e., collect $\mathbf{x}(n)$) and compute all the wavelet coefficients $\mathbf{\Psi x}(n)$ (or at least the K most important). This is not a problem in centralized scenarios as image processing, but to do so in a distributed fashion is not straightforward and strong assumptions appear in the literature (e.g., [Baj06]) as we have detailed previously.

In a WSN scenario, the computation of the transformed coefficients requires all the S sensing nodes to transmit their readings towards this central entity in order to get a sparse version of $\mathbf{x}(n)$ afterwards. One can easily see that this approach is signaling intensive and highly energy consuming.

In order to overcome this problem, one possible solution is to artificially create a sparse representation of $\mathbf{x}(n)$ with only K loaded entries and to set the rest to zero. One can easily do that by substituting the linear transformation $\mathbf{\Psi x}(n)$ by a non-linear downsampling encoder for each sensor. In principle, this procedure is not efficient at all since we are drastically removing a huge amount of information about the vector $\mathbf{x}(n)$, in particular $\frac{S-K}{S}$ of the total entries (remember that $K \ll S$). In order to counteract this effect, we propose to use a Conditional Downsampling Encoder (CDE), which benefits from the time correlation properties of the signal. As we previously showed in Chapter 2, the distortion introduced for high-correlated signals (i.e., $\rho \to 1$) is quite low, even for high downsampling rates.

Shortly, the CDE uses the time correlation of the signal in order to

Algorithm 4.1 Sensing nodes

for $n = 1$ to end **do**
  during the sensing phase
  for each $s \in \mathcal{S}$ **do**
    get the sth measurement $x_s(n)$.
    compute linear prediction $\hat{x}_s(n)$ as in (4.22).
    compute $x_s(n) - \hat{x}_s(n)$.
    if $|x_s(n) - \hat{x}_s(n)| > \Delta$ **then**
      active sensor mode (i.e., belongs to $\mathcal{K}(n)$)
      broadcast $[\mathbf{x}_K(n)]_s = x_s(n)$.
      store $x_s(n)$.
    else
      stay in sleep mode (i.e., belongs to $\mathcal{Q}(n)$)
      store $\hat{x}_s(n)$.
    end if
  end for
end for.

decide whether the current sample should be transmitted or not. In particular, the CDE uses a linear prediction of $x_s(n)$ denoted as $\hat{x}_s(n)$. The value $\hat{x}_s(n)$ is compared to the signal of interest $x_s(n)$. If the absolute value of the difference is higher than a given threshold Δ, the encoder transmits the sample. Otherwise, if the difference is below Δ, the transmission is blocked. The value of Δ can be chosen to ensure that K active sensors are active in mean, as it is detailed in Section 4.5.

Then, the goal is to obtain a predictor $\hat{x}_s(n)$ of each element of $\mathbf{x}(n)$ based on the N past readings of each sensor. This allows us to obtain a sparse version of $\mathbf{x}(n)$, named $\mathbf{x}_K(n)$, containing (on average) only the K most relevant readings with low distortion.

The predictor $\hat{x}_s(n)$ can be computed as a linear combination of the N

previously decoded readings at the sth sensor. We use the Linear Wiener Filter (LWF) to predict $x_s(n)$ because it is optimal in terms of the MSE. Mathematically,

$$\hat{x}_s(n) = \mathbf{w}^H \tilde{\mathbf{x}}_s(n), \tag{4.22}$$

where $\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}$ is the N-dimensional LWF solution, $\mathbf{r}$ is known and denotes the $N \times 1$ cross-correlation vector between the past stored samples and the desired measurement $x_s(n)$, and the observation vector $\tilde{\mathbf{x}}_s(n) \in \mathbb{R}^N$ collects the N past decoded values of sth sensor by the fusion center. It is defined as:

$$[\tilde{\mathbf{x}}_s(n)]_j = \begin{cases} x(n-j) & \text{if } s \in \mathcal{K}(n-j) \\ \hat{x}(n-j) & \text{otherwise} \end{cases} \tag{4.23}$$

Note that both the sensing nodes and the fusion center should have the same version of the observation vector. That is why the observation vector is not constructed simply as $[\tilde{\mathbf{x}}_s(n)]_j = x(n-j)$. Thus if the sth sensor transmits, $x_s(n)$ is stored in the observation vector at the sensing node. Otherwise, if the sth sensor is silent, it stores $\hat{x}(n)$ (see Algorithm 4.1).

Following this approach, only the subset of sensors $\mathcal{K}(n)$ transmit their readings while the rest $\mathcal{Q}(n)$ remain sleep. Algorithm 4.1 reviews the action executed by the sensing nodes.

4.4.2 Projection phase

Some existing CS-based techniques solve the problem of making projections by assuming a tree-based WSN and computing them in the gathering nodes [Luo09]. Others circulate the message from the source through the network and each node adds its contribution [Cho09]. Then, the message returns to the sink with all the contributions forming one projection.

On the contrary, in [Baj06] each sensor sends its contribution $[\mathbf{\Psi}]_{k,s} x_s(n)$ for $k = 1, \ldots, K$, under the unfeasible assumption that the

Algorithm 4.2 Relay nodes

for $n = 1$ to end **do**
  for each $r \in \mathcal{R}$ **do**
    while during the sensing phase **do**
      collect readings from $\mathcal{K}(n)$.
    end while
    compute projection $y_r(n) = [\mathbf{\Phi}]_r \mathbf{x}_K(n) + w(n)$.
    transmit $y_r(n)$ to the fusion center.
  end for
end for.

transformed vector is sorted in decreasing order of the absolute value of its entries as in (4.5). The fusion center receives each contribution as

$$y_k(n) = [\mathbf{\Psi}]_k \mathbf{x}(n) + w(n), \quad \text{for } k = 1, \ldots, K, \tag{4.24}$$

where $w(n)$ models the additive white Gaussian noise (AWGN) with zero mean and variance σ_w^2 present in the wireless channel. Each sensor repeats this algorithm K times per sample.

Furthermore, if we assume the GCWS scheme that we have introduced in Section 4.2, the fusion center receives the sum of all contributions as

$$y_r(n) = [\mathbf{\Phi\Psi}]_r \mathbf{x}(n) + w(n), \quad \text{for } r = 1, \ldots, R. \tag{4.25}$$

Obviously, the approaches in [Cho09] and [Baj06] may become expensive in terms of the number of transmissions since, in the worst case, $K \cdot S$ transmissions for each field measurement in [Baj06] and $R \cdot S$ for the GCWS are required. To cope with this problem, the authors in [Cho09] search (heuristically) K-sparse projection vectors to compensate the high energy costs (at the expenses of reconstruction accuracy).

On the contrary, our proposed algorithm does not assume either a previous knowledge of the principal components of the signal or the sparse

projection vectors and still its cost is $(K+R) \ll (R \cdot S)$ transmissions, or even $(K+R) \ll (K \cdot S)$.

First, let $y_r(n)$ define the received signal at the rth relay node as:

$$\begin{aligned} y_r(n) &= \sum_{s \in \mathcal{K}} [\mathbf{\Phi}]_{r,s} x_s(n) + w(n), \\ &= [\mathbf{\Phi}]_r \mathbf{x}_K(n) + w(n), \quad \text{for } r = 1, \ldots, R. \end{aligned} \tag{4.26}$$

where $[\mathbf{\Phi}]_{r,s}$ models the flat-fading channel from the active sensor s towards the relay r. In the CS literature, random Gaussian matrices with i.i.d. entries have been extensively used as sensing matrices like $\mathbf{\Phi}$ due to its simplicity and incoherence properties with respect to any fixed basis $\mathbf{\Psi}$ [Can08b].

The reader may notice that the matrix $\mathbf{\Phi}$ drawn from the Gaussian ensemble may model the channel only when $\mathcal{K}(n) \cap \mathcal{R} = \varnothing$. In other works, when any of the relay sensors does not act as active sensor simultaneously. If we relax this assumption, i.e. we allow $\mathcal{K}(n) \cap \mathcal{R} \neq \varnothing$, the sensing matrix $\mathbf{\Phi}$ is not a purely Gaussian random matrix anymore. It is because the element $[\mathbf{\Phi}]_{k,r}$ (which refers to the same physical node that acts as the kth active sensor in the sensing phase and as the rth relay in the projection phase) is $[\mathbf{\Phi}]_{k,r} = 1$ or even $[\mathbf{\Phi}]_{k,r} = 0$ if this sensor adds its contribution after receiving $y_r(n)$ or not. This fact may degrade the incoherence property of Gaussian matrices and may be a problem because the coherence between basis is very important in the accuracy of the reconstruction process [Can08b], [Can11]. For that reason, it needs to be preserved.

In order to overcome this problem, the active node k acting also as relay r can add its contribution as $[\mathbf{\Phi}]_{k,r} x_k(n)$ where $[\mathbf{\Phi}]_{k,r}$ is pseudo randomly generated as $\mathcal{N}(0, R^{-1})$ as in [Baj06]. In principle it may incur in extra signaling. However, each sensor can locally draw the elements of $[\mathbf{\Phi}]_{k,r}$ in an efficient manner by using the seed of a pseudo-random generator and its network identifier. Similarly, the fusion center only needs to know the seed in order to reconstruct $[\mathbf{\Phi}]_{k,r}$. Therefore the fusion center does not need to

Algorithm 4.3 fusion center

for $n = 1$ to end **do**
  while during the sensing phase **do**
    compute available information $\tilde{\mathbf{x}}_s(n)$ for each sensor.
  end while
  while during the projection phase **do**
    collect $y_1(n) \dots y_R(n)$ projection from relays.
  end while
  solve $\mathcal{P}'1$ or $\mathcal{P}'2$ in (4.28) and (5.15) respectively to recover $\hat{\mathbf{x}}_K(n)$.
  obtain $\hat{\mathbf{x}}(n)$ replacing the zeros by their estimations $\hat{x}_s(n)$.
end for.

signal extra data.

From a medium access control point of view, there is no need to make orthogonal transmissions during the sensing phase, so this phase has a cost of one channel use as we can see in (4.26). Differently, during the projection phase, it is assumed that the relay nodes transmit through R orthogonal channels to send a coded version of the projected values $y_r(n)$. This phase has a total cost of R channel uses (this process is summarized in Algorithm 4.2).

4.4.3 Signal reconstruction phase

The fusion center gathers all the received measurements in the R dimensional projection vector, denoted as $\mathbf{y}(n) = [y_1(n)\, y_2(n) \dots y_R(n)]^T$. Hence, the goal of the reconstruction phase is to recover an approximation of $\mathbf{x}_K(n)$, i.e. $\hat{\mathbf{x}}_K(n)$, given $\mathbf{y}(n)$ and $\mathbf{\Phi}$. The most prevalent decoder for the noiseless case in the CS literature is the l^1-norm minimization program which is a convex relaxation of the original NP-hard problem (with the

l^0-norm) as in e.g., [Don06b], [Can06a]:

$$\mathcal{P}1: \quad \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\boldsymbol{\Psi}\hat{\mathbf{x}}(n)\|_{l^1} \qquad \text{subject to} \quad \mathbf{y}(n) = \boldsymbol{\Phi}\boldsymbol{\Psi}\hat{\mathbf{x}}(n). \tag{4.27}$$

Note that this problem is the same than in (5.18) with the notation of $\boldsymbol{\Psi}\hat{\mathbf{x}}(n) = \hat{\boldsymbol{\omega}}_K(n)$.

Unfortunately we cannot directly apply $\mathcal{P}1$ as the decoder of the proposed AF-CS scheme. This is because we have substituted the linear transformation $\boldsymbol{\Psi}\mathbf{x}(n)$ by the non-linear encoder CDE in the sensing phase. Hence, we propose the following AF-CS decoder, which has two main building blocks: *i*) a CS decoder (CSD) block, called $\mathcal{P}'1$, that recovers $\hat{\mathbf{x}}_K(n)$, and *ii*) a Prediction Decoder (PD) with $\hat{x}(n)$ as its output.

Thus, the CS decoder $\mathcal{P}'1$ for the noiseless case is defined as:

$$\mathcal{P}'1: \quad \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\hat{\mathbf{x}}(n)\|_{l^1} \qquad \text{subject to} \quad \mathbf{y}(n) = \boldsymbol{\Phi}\hat{\mathbf{x}}(n). \tag{4.28}$$

For the noisy case, we use a modification of the minimization problem previously formulated in (4.16) as:

$$\mathcal{P}'2: \quad \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\hat{\mathbf{x}}(n)\|_{l^1} \qquad \text{subject to} \quad \|\mathbf{y}(n) - \boldsymbol{\Phi}\hat{\mathbf{x}}(n)\|_2 < \varepsilon. \tag{4.29}$$

Afterwards, the algorithm should replace the $S-K$ zeros by their correspondent predicted entries (see Algorithm 4.3). Therefore, we propose to use an instance of the PD with the following non-linear decoding function,

$$d_{\text{PD}}(n) = \begin{cases} [\hat{\mathbf{x}}(n)]_s = [\hat{\mathbf{x}}_K(n)]_s & \text{if } [\hat{\mathbf{x}}_K(n)]_s \neq 0 \\ [\hat{\mathbf{x}}(n)]_s = \hat{x}_s(n) & \text{otherwise} \end{cases} \tag{4.30}$$

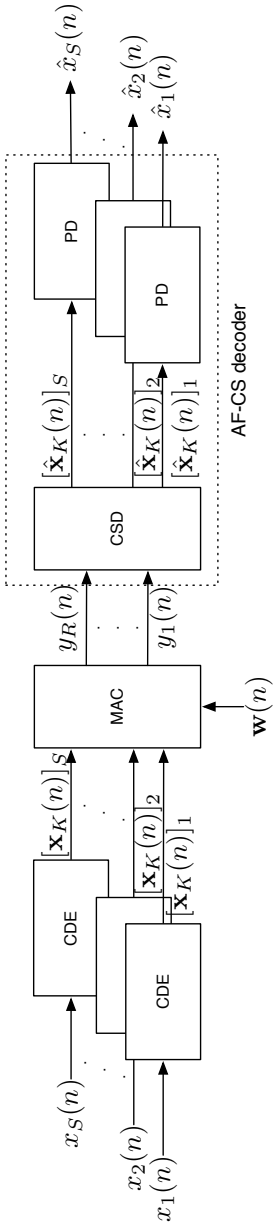

Figure 4.4: Block diagram of the AF-CS transmission scheme.

Furthermore, we assume that both the sensing nodes and the fusion center have perfect knowledge of the time correlation parameters $\mathbf{R}$ and $\mathbf{r}$. In practice, one can use different methods to reduce the number of snapshots needed to obtain good correlation estimators, as we show in Chapter 3.

The proposed technique has three main sources of distortion; *i*) the l^1-norm minimization problem in the reconstruction of the active node's measurements, *ii*) the error in the prediction of the LWF to reconstruct the silent sensors, and *iii*) the noise of the wireless channel.

We showed in [BL11] that the distortion introduced by the proposed algorithm is very sensitive to the values of K, R, and S and hence we need an error model in order to configure a priori the network. In the following section we propose a framework to model the error introduced by the first two sources pointed out above.

4.5 Distortion analysis of the AF-CS decoder

In this section, we asses the error introduced by the AF-CS decoder, $\mathcal{D}_{\text{AF-CS}}$. As we mentioned previously, it can be seen as a combination of two partial decoders: *i*) a CS decoder, named CSD, and *ii*) a predictive decoder named PD. In the following we analyze the distortion introduced by both steps separately.

4.5.1 Distortion due to the CSD step

In this subsection we focus our analysis on the distortion introduced by the CS decoder, denoted as $\mathcal{D}_{\text{CSD}}$:

$$\mathcal{D}_{\text{CSD}} = \mathbb{E}\left[\|\mathbf{x}_K(n) - \hat{\mathbf{x}}_K(n)\|_2^2\right]. \tag{4.31}$$

In particular, we study the design conditions for K, R, and S that

guarantee an upper-bound of the distortion as in [Can11], i.e.,

$$\mathcal{D}_{\mathrm{CSD}} < \left(\frac{K}{R}\sigma_w^2 \log S\right). \tag{4.32}$$

The reader will notice that this bound is applicable to both the noiseless and the noisy problems $\mathcal{P}'1$ and $\mathcal{P}'2$ respectively. For the noiseless case, we directly get $\mathcal{D}_{\mathrm{CSD}} = 0$, so it means that we look for the design conditions for K, R, and S in such a way that we can expect perfect recovery of the sparse sensor $\mathbf{x}_K(n)$.

What we know from the current state-of-the-art of the CS theory [Can11] is not much more than if the number of random measurements R are on the order of $K \log S$ (with $R \ll S$), it is possible to recover $\mathbf{x}_K(n)$ with an error bounded for (4.32). This condition was first introduced for the Fourier basis case in [Can06a] and for the Gaussian ensemble in [Can06b]. From this result, other variations to $R > C_0 K \log S$ have appeared for the Gaussian ensemble case, e.g., $R > C_0 K \log\left(\frac{S}{K}\right)$ in [Che98], [Can06a], $R > C_0 K \log\left(\frac{S}{R}\right)$ in [Don06a], $R > C_0 K \log^2(S)$ in [Can11]. However, little more than $C_0 > 0$ is known. Because of this limitation, we can also find some practical results as R should be of the order of $3K$ to $5K$ [Can08a], [Can08b], and maybe others that we have unintentionally omitted. One can conclude that the bounds in the literature are quite heterogeneous. In fact, it is still unknown whether or not the perfect recovery (for the noiseless case) can be guaranteed when the number of measurements R is on the order of $K \log S$ [Can11], and to solve it is still an open problem in the CS literature (at the time of writing the current chapter).

Now, the most widely used tool for addressing such problems is the so-called *Restricted Isometry Property* (RIP), which was first introduced by Candès and Tao in [Can05].

If a given sensing matrix $\mathbf{\Phi}$ verifies the RIP, it means that it behaves like a nearly orthogonal system but only for sparse linear combinations. It is shown in [Can08b] and [Can06a] that this condition allows for the exact

reconstruction of sparse linear combination of these vectors.

Definition 1 [Can05]: *A matrix* $\mathbf{\Phi}$ *satisfies the RIP of order* K *with restricted isometry constant* $\delta_K \in (0,1)$ *if*

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1+\delta_K)\|\mathbf{x}\|_2^2, \tag{4.33}$$

where $\mathbf{\Phi}_K \in \mathbb{R}^{R\times K}$ *is formed by retaining any set of* K *or less columns from* $\mathbf{\Phi}$, $\mathbf{x}$ *is any* K*-sparse vector of the appropriate size, and* δ_K *is the smallest number (and not too close to one) that holds the RIP condition for each integer* $K = 1, 2, ...$

In order to see the connection between CS and RIP, we can observe the following example.

Example 4.1 *[Can11] Suppose unique reconstruction of a* K*-sparse vector failed. Then, there would exist at least two* K*-sparse vectors,* $\mathbf{x}$ *and* $\mathbf{x}'$*, that obey* $\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi x}'$*. Thus we have*

$$\mathbf{\Phi c} = \mathbf{0}, \tag{4.34}$$

where $\mathbf{c} = \mathbf{x} - \mathbf{x}'$ *is a (at most)* $2K$*-sparse vector lying in the null space of* $\mathbf{\Phi}$*. This cannot be possible if RIP holds or* $\delta_{2K} < 1$ *[Can06a], so we get a contradiction. In other words if we apply norms of both sides to (4.34) and apply the RIP condition, we have,*

$$(1-\delta_{2K})\|\mathbf{c}\|_2^2 \leq \|\mathbf{\Phi c}\|_2^2 \leq (1+\delta_{2K})\|\mathbf{c}\|_2^2, \tag{4.35}$$

and (5.19) only can be held if $\delta_{2K} \geq 1$*, so we reach the same contradiction.*

We can also relate the RIP with $\mathcal{P}0$ and $\mathcal{P}1$ (or even $\mathcal{P}'1$). In theory [Can08b], $\mathcal{P}0$ and $\mathcal{P}1$ are formally equivalent in the following sense:

i) if $\delta_{2K} < 1$ the $\mathcal{P}0$ has a unique K-sparse solution.

ii) if $\delta_{2K} < \sqrt{2} - 1$ the solution of $\mathcal{P}1$ is identical to the one of $\mathcal{P}0$. In other works, the solution is unique as well.

Following the results of Example 4.1 for δ_{2K}, the condition in equation (4.33) is equivalent to require all the eigenvalues of $\mathbf{\Sigma} = \mathbf{\Phi}_K^T \mathbf{\Phi}_K, \mathbf{\Sigma} \in \mathbb{R}^{K \times K}$ to be inside the interval $[1 - \delta_{2K}, 1 + \delta_{2K}]$ as it is said in [Bar07], where $\mathbf{\Phi}_K \in \mathbb{R}^{R \times K}$ is formed by retaining any set of K or less columns from $\mathbf{\Phi}$, i.e.,

$$1 - \delta_{2K} < \lambda_{\min} < \lambda_{\max} < 1 + \delta_{2K}, \tag{4.36}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the limiting support of the eigenvalues of $\mathbf{\Sigma}$. Thus, using the asymptotic results from the work of Marčenko and Pastur [Mar67], we characterize the matrix $\mathbf{\Sigma}$ as a Wishart matrix and thus the asymptotic density function of its eigenvalues, $f_{\mathbf{\Sigma}}(\lambda)$, follows the well-known Marčenko-Pastur distribution:

$$f_{\mathbf{\Sigma}}(\lambda) = \left(1 - \frac{1}{\alpha^*}\right)^+ \delta(\lambda) + \frac{\sqrt{(\lambda - \lambda_{\min})^+ (\lambda_{\max} - \lambda)^+}}{2\pi\alpha^*\lambda}, \tag{4.37}$$

where $\lambda_{\min} = (1 - \sqrt{\alpha}^*)^2$ and $\lambda_{\max} = (1 + \sqrt{\alpha}^*)^2$ are the support region boundaries of $f_{\mathbf{\Sigma}}(\lambda)$ and $\alpha^* = \lim_{K,R \to \infty} K/R$.

Let us first consider the case of $\mathcal{P}0$ for simplicity. To keep $\delta_{2K} < 1$, one needs to consider only the $\lambda_{\max}$ since the $\lambda_{\min}$ is always positive because $\mathbf{\Sigma}$ is positive semidefinite by definition. Then $\lambda_{\max}$ must obey:

$$\lambda_{\max} < 1 + \delta_{2K} < 2 \Longrightarrow (1 + \sqrt{\alpha})^2 < 2. \tag{4.38}$$

Then, the condition for the ratio α, denoted as $\mathcal{C}_{\mathcal{P}0}$, is given by

$$\mathcal{C}_{\mathcal{P}0}: \quad \alpha \ < \ (\sqrt{2} - 1)^2 = 0.1716. \tag{4.39}$$

In Fig. 4.5 we plot $f_{\mathbf{\Sigma}}(\lambda)$ for different values of α^*. Surprisingly these results are excellent approximations even for quite small systems [Tul04]. In Fig. 4.5, we also compare with the normalized histogram of the average of 100 empirical distribution functions (*edf*) for a quite small Wishart ensemble

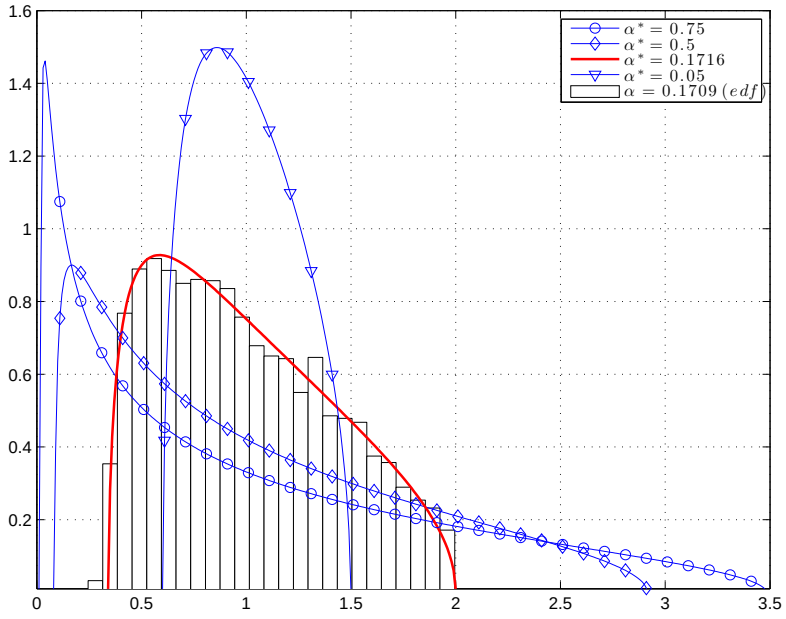

Figure 4.5: The Marčenko-Pastur density function of the eigenvalues of a Wishart matrix for different values of α. We also compare with the normalized histogram of the average of 100 *edf*s for a Wishart matrix of size $K = 20$, $R = 117$ and $\alpha = 0.1709$ which is the best approximation for $\alpha^\star = 0.1716$.

of size $K = 20$ and $\alpha = 0.1709$ in order to graphically show the fitness accuracy.

Similarly to (4.39), we can also find the matrix ratio α that makes $\delta_{2K} < \sqrt{2}-1$ hold as a function of the maximum and minimum eigenvalues of $\mathbf{\Sigma}$ as

$$1 - (\sqrt{2} - 1) < \lambda_{\min} < \lambda_{\max} < 1 + (\sqrt{2} - 1), \tag{4.40}$$

One can solve the system of inequalities as:

$$1 - (\sqrt{2} - 1) < (1 - \sqrt{\alpha})^2 \Longrightarrow \alpha < 0.0551 \tag{4.41}$$

$$1 + (\sqrt{2} - 1) < (1 + \sqrt{\alpha})^2 \Longrightarrow \alpha < 0.0358 \tag{4.42}$$

Then, the condition for the ratio α of the problem $\mathcal{P}1$, $\mathcal{C}_{\mathcal{P}1}$, is related to the condition for $\lambda_{\max}$ because it is the most restrictive.

$$\mathcal{C}_{\mathcal{P}1}: \quad \alpha < 0.0358. \tag{4.43}$$

In order to graphically observe the performance of these two limits, we show the result of the following experiment.

Experiment 4.1 *We have simulated for $S = 200$, $R = [0, S/2]$, and $K = 10$, solving the $\mathcal{P}'1$ problem for the noiseless case. We have used* `CVX`*, a package for specifying and solving convex programs [Gra11, Gra08]. In our experiment, $\mathbf{x}_K(n)$ have been generated as a S dimensional all-zero vector except for K loaded entries with independent and Gaussian values with zero mean and unit variance, and located randomly along the dimension of $\mathbf{x}_K(n)$. The $R \times S$ matrix $\mathbf{\Phi}$ have been generated following the Gaussian ensemble with entries distributed according to $\mathcal{N}(0, R^{-1})$. We have averaged the problem 1000 times for each value of R using in each iteration different realizations of $\mathbf{x}_K(n)$ and $\mathbf{\Phi}$.*

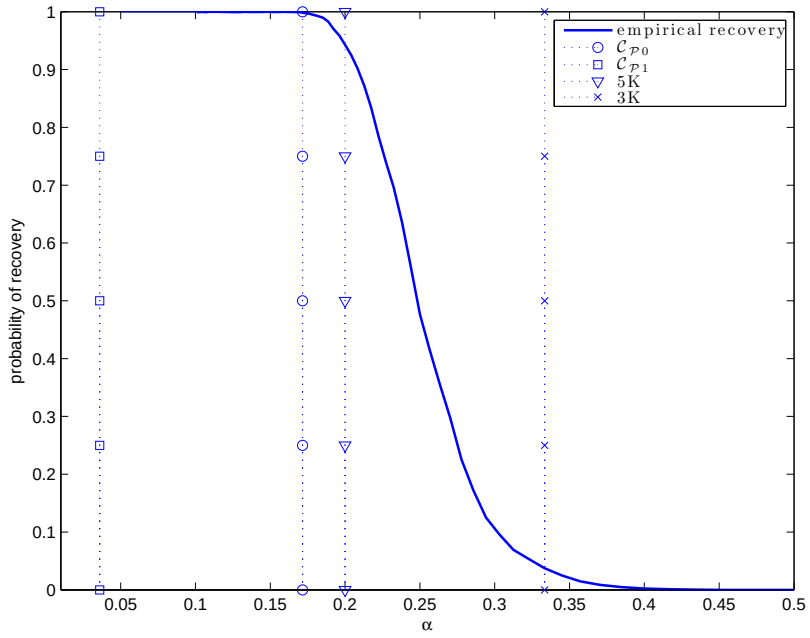

Figure 4.6: Empirical results from Experiment 4.1. The empirical probability of recovery is compared with the recovery bounds $\mathcal{C}_{\mathcal{P}0}$, $\mathcal{C}_{\mathcal{P}1}$, $3K$, and $5K$. The vertical lines mean that the left-hand side of them is for perfect recovery while the right hand side is where the recovery is not guaranteed with high probability.

From Experiment 4.1, we have plotted the empirical probability of recovery in Fig. 4.6. We have also compared the experimental results with the $\mathcal{C}_{\mathcal{P}0}$, $\mathcal{C}_{\mathcal{P}0}$, and the practical conditions $3K - 5K$ from [Can08a].

Although the condition $\mathcal{C}_{\mathcal{P}0}$ is derived for $\mathcal{P}0$, it probably brings the best accuracy with the experimental results. The case $3K - 5K$ is shown to be more aggressive and thus the signal is perfectly reconstructed with higher probability for the values closed to $5K$ case and with lower probability for the values closed to $3K$. On the other hand, the condition $\mathcal{C}_{\mathcal{P}1}$ turns out to be too conservative.

It is clear that the RIP is intimately related to the maximum and minimum eigenvalues of $\mathbf{\Sigma}$ throughout (4.36). Therefore, in order to relate the probability of recovery with the eigenvalues of $\mathbf{\Phi}$ (as the RIP suggests) we have proposed the following experiment.

Experiment 4.2 *The setup is the same as in Experiment* 4.1*, with the difference that we have only simulated for the values* $\alpha = [1/4,\ 1/5,\ 1/6,\ 1/7]$*. In this case, we represent the probability of recovery as a function of the eigenvalues of* $\mathbf{\Sigma}$ *instead of* α*. To obtain accurate results, we have averaged the problem* $3 \cdot 10^4$ *times for each value of* R*.*

Although the curves in Fig. 4.7 are quite noisy in the extremes due to the low probability of occurrence of those values of $\lambda_{\{\min,\max\}}$, we can extract two important conclusions about the relation between the eigenvalues of $\mathbf{\Sigma}$ and the probability of recovery. Those are,

i) We can observe that only the minimum eigenvalue has a significative impact in the recovery capabilities of $\mathcal{P}_1$. Instead, we can extract from the results that there is not a restriction regarding to the maximum eigenvalue. It contradicts the second inequality of the RIP statement in (4.36).

ii) Matrices with the same $\lambda_{\{\min,\max\}}$ but with different ratios α perform

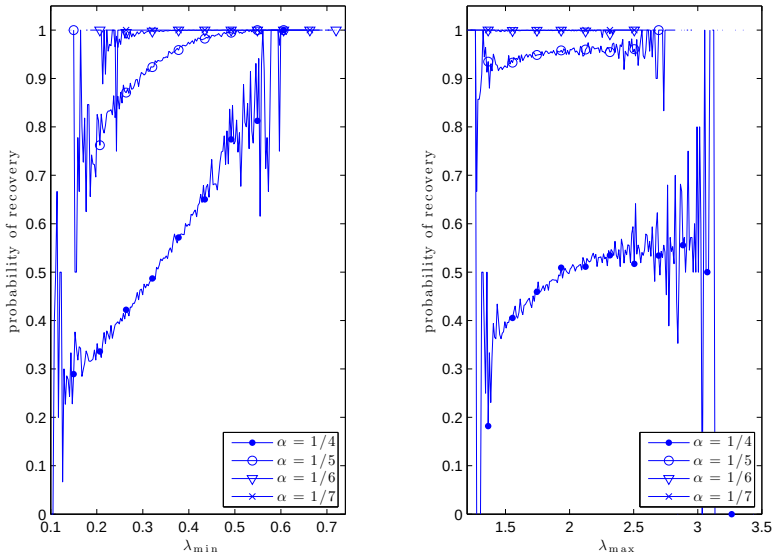

Figure 4.7: Empirical results for Experiment 4.2. We plotted the probability of recovery for different configurations of α as a function of the minimum eigenvalue $\lambda_{\min}$ (left) and maximum eigenvalue $\lambda_{\max}$ (right) of the matrix $\mathbf{\Sigma}$.

differently. Therefore, one can conclude from this observation, that it is not sufficient to focus on the eigenvalue support region of $\mathbf{\Sigma}$.

From Experiment 4.1 and Experiment 4.2, we can conclude that there is still a lot of research to do in the field of modeling the performance of CS as a function of the matrix $\mathbf{\Phi}$ and the number of measurements R. Actually, a recent work in [Can11], have proposed a new approach to face with this problem using RIPless theory.

Because these results are still at an early stage, we use empirical results in order to obtain an accurate model that describes the performance precisely. To do so, we introduce the following experiment.

Experiment 4.3 *The setup is the same as in Experiment* 4.1 *and Experiment* 4.2*, but with the following input parameters:* $S = 200$*,* $K = [0, S]$*, and* $R = [0, S]$*. We averaged 5 times for each combination of* K *and* R*.*

In Fig. 4.8 we represent the results of the Experiment 4.3. We plot the probability of recovery as a function of both K and R. The grey zone indicates that the probability of recovery is zero. The white zone means that perfect recovery occurs with high probability. We can see that the division of the recovery and non-recovery zones is well defined.

In the same plot, we also compare the empirical performance of $\mathcal{P}1$ with the different relations in the literature that we introduced at the top of the section. Although we set all constants $C_0 = 1$ for simplicity, one can see that even tuning C_0 a posteriori using the knowledge of the experimental results, none of the models accurately adjusts to the result of the experiment for the whole range of values.

Moreover, note from the results that two boundary points must be necessarily included. The first one is the trivial point $(K, R) = (0, 0)$, and all the proposed models already contain this point. The second point that should be present in the model is $(K, R) = (S, S)$. The proof is straightforward; $\mathbf{\Phi}$ is a square full rank matrix (with almost sure probability [Fen07]),

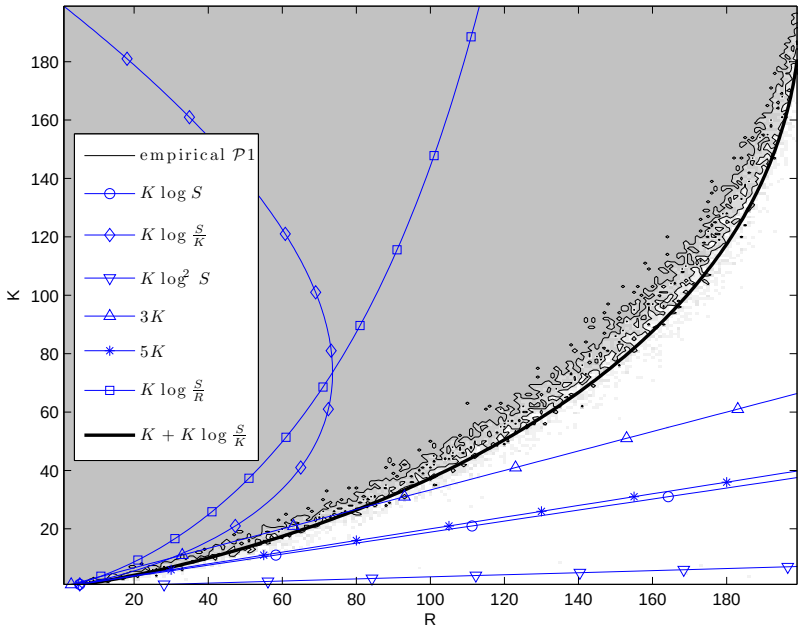

Figure 4.8: Empirical results for Experiment 4.3. The empirical probability of recovery map for each value of $K, R \in 1, ..., S$ is compared with different models in the literature. The white area means perfect recovery with high probability while the grey area means that the probability of recovery is zero. The light grey tones mean the intermediate values of probability of reconstruction. The area below the curves indicates which are the values (K, R) that the model relates to high probability of recovery. We omitted the parameter C_0 present in some of the models for simplicity.

and hence one may solve $\mathcal{P}1$ (or even $\mathcal{P}'1$) with zero error by solving $\hat{\mathbf{x}}_K(n) = \mathbf{\Phi}^{-1}\mathbf{y}(n)$ for any value of K (even for $K = S$). However, none of the proposed methods contain this second point. This is probably because they are only valid under the assumption $K \ll S$.

In order to overcome this problem, we propose an empirical model called $\mathcal{C}_{\text{CS}}$ that accurately fits for all the points, namely

$$\mathcal{C}_{\text{CS}}: \quad R > K + K \log\left(\frac{S}{K}\right) \tag{4.44}$$

and it is plotted as a double-width solid line in Fig. 4.8.

This can be seen as a modification of the condition $R > K \log\left(\frac{S}{K}\right)$ in [Can06a], that takes into account the entire set of values for K and R. Additionally, some multiplicative constants may be added in order to improve the accuracy in certain scenarios as in the other models in the literature.

The mathematical validation of this model is still missing, however it has been numerically tested in several other experiments.

4.5.2 Distortion due to the PD step

In this section we focus on the distortion introduced by the Predictive Decoder (PD) step, denoted as $\mathcal{D}_{\text{PD}}$:

$$\mathcal{D}_{\text{PD}} = \mathbb{E}\left[\|\mathbf{x}(n) - \hat{\mathbf{x}}(n)\|_2^2\right], \tag{4.45}$$

assuming that no distortion is propagated from the previous CSD step, or in other words, that we have achieved perfect recovery of the vector $\mathbf{x}_K(n)$, i.e., $\mathbf{x}_K(n) = \hat{\mathbf{x}}_K(n)$.

In the following, we focus, without loss of generality, on the performance of the sth sensor. The transmission block diagram for the sth sensor is simplified as Fig. 4.9.

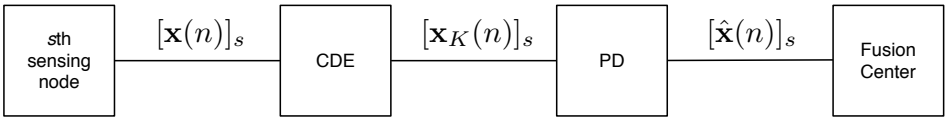

Figure 4.9: Simplified transmission block diagram for a given sensor s under the assumption that there is no distortion due to the CS decoder.

Following this simplified approach, the resulting system turns out to be similar than the one analyzed in the Chapter 2 for the particular case of the CDE-PD decoder pair[2].

Let us assume the $AR-1$ process as the signal model of sensor s,

$$x_s(n) = \rho x_s(n-1) + z(n), \tag{4.46}$$

where $z(n)$ models the uncertainty in the model as $z(n) \sim \mathcal{N}(0, \sigma_z^2)$.

The CDE-PD performs Linear Wiener Filter (LWF) prediction at both the encoder and the decoder. The LWF predictor is optimal in the sense that its MSE is σ_z^2. This is proved in Lemma 2.1, and it also relates the correlation factor with the LWF prediction error as $\sigma_z^2 = 1 - \rho^2$. However, this error only can be achieved if the observation vector contains the true value of the last transmitted sample, i.e., $x_s(n-1)$.

As we have previously detailed in the sensing phase, the observation vector at the CDE (sensing node) should be identical to the one at the PD (fusion center). Hence, it follows the structure introduced in (4.23). Using this "incomplete" version of the observation vector, the MSE is degraded as the sensor s does not transmit. This degradation is approximated by (Lemma 2.5):

$$\mathrm{MSE}_t \simeq h\left(1 - \rho^2 + \rho^2 \mathrm{MSE}_{t-1} | \Delta_t\right), \tag{4.47}$$

[2]for a more self-contained explanation of Chapter 4, we particularize some results already introduced in Chapter 2 with the corresponding change of notation.

where $\text{MSE}_0 = 0$, t indicates the time index of the last transmitted sample, Δ_t is the threshold value applied in the CDE at time t, and the conditional function $h(\sigma^2|\Delta_t)$ is introduced in Definition 2.7 as:

$$h(\sigma^2|\Delta_t) = \frac{2}{\sqrt{2\pi\sigma^2}}\left(-\Delta_t\sigma^2\ e^{\frac{-\Delta_t^2}{2\sigma^2}} + \frac{1}{2}\sqrt{2\pi\sigma^6}\ \text{erf}\left(\frac{\Delta_t}{\sqrt{2\sigma^2}}\right)\right). \quad (4.48)$$

Furthermore, following the results in Chapter 2, the value of Δ_t can be related to the compression rate $\gamma = K/S$ as:

$$\Delta_t = \sqrt{2\pi\text{MSE}_t}\text{erf}^{-1}(1-\gamma) \quad (4.49)$$

It is easy to conclude that the distortion introduced by the pair CDE-PD can be numerically approximated by

$$\mathcal{D}_{\text{PD}} \ \simeq \ \sum_{t=0}^{\infty} P_t \text{MSE}_t. \quad (4.50)$$

where $P_t = \gamma(1-\gamma)^t$ (see Lemma 2.3).

4.6 Design of the Parameters K and R: Error versus Energy Trade-off

The parameters K and R have a direct influence on the distortion level of AF-CS decoder. The higher K, the lower the distortion of the PD. At the same time, a sufficiently large number of R ensures zero distortion at the CSD.

On the other hand, K and R also influence in terms of energy consumption as we have seen previously.

One of the possible approaches to design the network is to add a new degree of freedom. We define the utility function $\nu(\beta)$ as:

$$\nu(\beta) = \beta\mathcal{D}_{\text{AF-CS}} + (1-\beta)\mathcal{E}, \quad (4.51)$$

Table 4.1: Simulation Parameters

Parameter	Value
Number of *fusion nodes*:	$F = 1$
Number of *sensing nodes*:	$S = 200$
Number of *active sensors*:	$K = [0, 200]$
Number of *relay nodes*:	$R = [0, 200]$
Cost factor:	$\beta = 0.5$
Correlation model, $[\mathbf{R}]_{n,n+k} = \rho^{\lvert k \rvert}$:	$\rho = 0.99$

where $\mathcal{E}$ is the energy consumption modeled as the number of transmissions normalized by the total number of sensors, i.e., $\mathcal{E} = \frac{K+R}{S}$, since the most of the energy consumption occurs when the sensor is in active mode [Rug07]. Therefore, the *cost factor* β controls the trade-off between energy consumption and error. Hence, for a given β, the design problem corresponds to the solution of $K^\star$ and $R^\star$ that minimizes $\mathcal{P}3$ as:

$$\begin{aligned} \mathcal{P}3: \quad & \underset{K,R}{\text{minimize}} \quad && \nu(\beta), \\ & \text{subject to} && \mathcal{C}_{\text{CS}}. \end{aligned} \tag{4.52}$$

As discussed before, the distortion of the AF-CS decoder can be modeled as $\mathcal{D}_{\text{AF-CS}} = \mathcal{D}_{\text{PD}}$ if $\mathcal{C}_{\text{CS}}$ holds. Thus, this problem can be solved very fast since $\mathcal{D}_{\text{PD}}$ is only a function of K and one can solve $\mathcal{P}3$ for K and then compute the lowest value of R that satisfies $\mathcal{C}_{\text{CS}}$.

4.7 Numerical Results

In this section, we provide simulation results to show the performance of our proposed energy-efficient scheme. Table 5.1 summarizes the parameters that we consider in our simulations.

Table 4.2: Results of the Utility Function $\nu(\beta)$

	K	R
$\min\{\tilde{\nu}(\beta)\}$	$K_{\text{sim}} = 13$	$R_{\text{sim}} = 38$
$\min\{\nu(\beta)\}$ s.t. $\mathcal{C}_{\text{CS}}$	$K^{\star} = 11$	$R^{\star} = 43$
$\min\{\nu(\beta)\}$ s.t. $\mathcal{C}'_{\text{CS}}$	$K'^{\star} = 9$	$R'^{\star} = 48$

4.7.1 Results about the proposed CS design rules.

In order to study the accuracy of our analytical design model, we compare the theoretical to the real (simulated) results. Fig. 4.10 shows the map formed by the contour lines of the empirical cost function $\tilde{\nu}(\beta)$, defined as

$$\tilde{\nu}(\beta) = \beta \mathcal{D}^{\text{sim}}_{\text{AF-CS}} + (1-\beta)\mathcal{E}. \tag{4.53}$$

It replaces the $\mathcal{D}_{\text{AF-CS}}$ calculated using the analytical model in Section 4.5 by the empirical $\mathcal{D}^{\text{sim}}_{\text{AF-CS}}$, defined next in (4.54). In this simulation, we set $\beta = 0.5$ in order to prioritize equally the energy consumption and the reconstruction error. Clearly, we can observe a cold region for low values of K and R (i.e., the contour lines around the marker $\times$). This is a direct consequence of the CS principia; the signal can be accurately recovered from a small amount of the total data. On the contrary, for high values of K and R, the proposed CS scheme performs inefficiently due to either a wrong CS recovery or higher energy consumption, or due to both cases.

We have also plotted in Fig. 4.10 the feasible regions and the optimal results for the minimization problem $\mathcal{P}3$ for two different constraints for the number of measurements R, i.e.: *i*) the proposed $\mathcal{C}_{\text{CS}}: R > K + K\log\left(\frac{S}{K}\right)$, and *ii*) the classical $\mathcal{C}'_{\text{CS}}: R > K\log S$.

In Fig. 4.11 we evaluate the existing gap between the design parameters obtained in (4.52) using both $\mathcal{C}_{\text{CS}}$ and $\mathcal{C}'_{\text{CS}}$, i.e., $K^{\star}$ and $R^{\star}$, and $K'^{\star}$ and $R'^{\star}$, respectively, and the ones obtained a posteriori minimizing $\tilde{\nu}(\beta)$ by simulation, i.e., K_{sim} and R_{sim}.

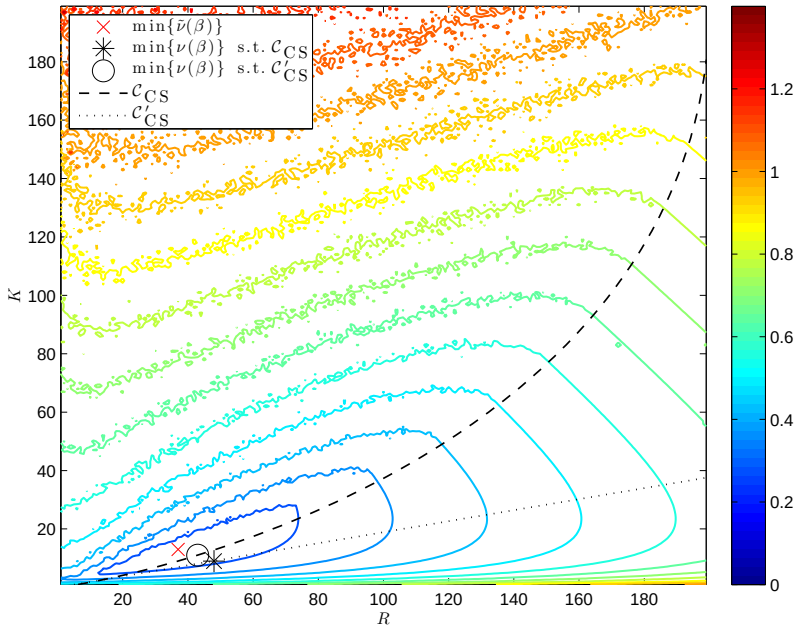

Figure 4.10: Contour map of $\tilde{\nu}(\beta)$ for $\beta = 0.5$. Cold colors represent low values and high values are hot-colored (in color printed version). For black and white version, contours near the cross marker are showing the minimum values. This figure has been averaged over 100 realizations. The areas under the curves $\mathcal{C}_{\mathrm{CS}}$ and $\mathcal{C}'_{\mathrm{CS}}$ correspond to the feasible set of $\mathcal{P}3$ subject to $\mathcal{C}_{\mathrm{CS}}$ and $\mathcal{C}'_{\mathrm{CS}}$ respectively.

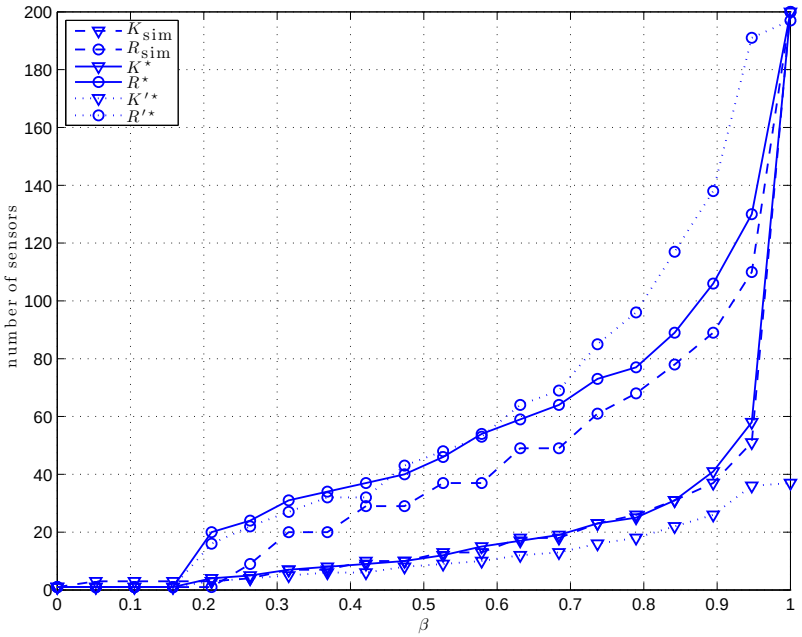

Figure 4.11: Comparison between the optimal K and R parameters obtained analytically (solid lines) with the optimal simulated ones using $\mathcal{C}_{\text{CS}}$ (dotted lines) and $\mathcal{C}'_{\text{CS}}$ (dashed lines).

Although $\mathcal{C}'_{\mathrm{CS}}$ guarantees an exact reconstruction of $\mathbf{x}_K(n)$, the results in Fig. 4.11 show that this constraint is too conservative. On the other hand, we can observe that our proposed model using $\mathcal{C}_{\mathrm{CS}}$ performs much closer to the real performance of the system. Indeed, it slightly upper-bounds the actual results, giving a slightly conservative solution for the network design parameters. However, the gap between $R^\star$ and R_{sim} is 9.8 sensors on average (it means only a relative error of 4.9% over the total number of sensors). The gap between $K^\star$ and K_{sim} is even smaller, 1.4 on average (0.7%). On the other hand, using $\mathcal{C}'_{\mathrm{CS}}$, the error increases up to 17.5 in mean (8.75%) for $R'^\star$ and 12.35 (6%) for $K'^\star$.

For low values of β, our approach prioritizes the energy savings decreasing the number of active sensors. Although it provides less accuracy, this situation may be interesting for some cases in WSN monitoring, for instance: temperature or humidity monitoring for long periods of time where changes are not expected, or for periods of less interest, e.g., during the nights. On the contrary, other situations would require higher accuracy levels in the measurements at the cost of being more energy expensive. Using the proposed approach, one can tune the parameter β in order to accommodate all these situations. In the following subsections, we only consider the results of $\mathcal{P}2$ using $\mathcal{C}'_{\mathrm{CS}}$.

4.7.2 Comparison with standard CS schemes available in the WSN literature

We compare the performance of AF-CS scheme with three reference systems in the WSN literature:

- Classical Approach (CA). The group of $\mathcal{S}$ sensing nodes transmit their measurements directly to the fusion node each time slot.
- Compressive Wireless Sensing (CWS). Following the approach in [Baj06], the group of $\mathcal{S}$ sensing nodes simultaneously transmit their

readings multiplied previously by the corresponding element of $\mathbf{\Psi}$ to the fusion center. This process should be repeated K times.

- Generalized Compressed Wireless Sensing (GCWS). It is based on CWS but relaxing the assumption about the prior knowledge of the transform coefficients in (4.5). Instead of sending the larger K transformed coefficients, it produces R random measurements as in (4.11) and (4.24) for the noiseless and the noisy case, respectively.

For a fair comparison, we add a space correlation (i.e., a correlation among the measures of the sensors) equal to the applied time correlation in Table 5.1, since algorithms of the type of the CWS scheme use this property to compress the readings. In addition, the matrix $\mathbf{\Psi}$ is constructed following a DCT basis and $\mathbf{\Phi}$ is a gaussian random matrix (with i.i.d. entries). In principle, this type of transform is suitable for temperature-like readings, as it is showed in [Cho09].

We compare the following figures of merit:

1. *Empirical* distortion, $\mathcal{D}^{\text{sim}}_{\text{AF-CS}}$, defined as:

$$\mathcal{D}^{\text{sim}}_{\text{AF-CS}} = \mathbb{E}[|\mathbf{x}(n) - \hat{\mathbf{x}}(n)|^2], \tag{4.54}$$

   averaged over 100 realizations and for different instances of $\mathbf{\Phi}$.

2. The *relative energy consumption* $\mathcal{E}$ measures the energy consumption in comparison to a standard star-topology WSN (in terms of the number of transmissions), so:

   - $\mathcal{E} = 1$ for the case of CA. All S sensors transmit their readings to the fusion center.
   - $\mathcal{E} = K$ for CWS. All S sensors transmit using the same resource to the fusion center to perform one projection. This process should be repeated K times.

- $\mathcal{E} = R$ for GCWS. All S sensors transmit using the same resource to the fusion center to perform one projection. This process should be repeated R times.
- $\mathcal{E} = \left(\frac{R+K}{S}\right)$ for AF-CS. The subset of sensors $\mathcal{K}$ broadcast their messages towards the R relays, involving K transmissions. In addition, the relay nodes retransmit the computed R projections to the fusion center.

3. The number of channel uses measures the number of the required resources for each scheme, thus;
   - CA permforms S channel uses assuming a given orthogonal MAC policy.
   - CWS requires K channel uses. This is only possible under the assumption that all the nodes knows a priori the correct order of the transformed coefficients.
   - GCWS requires R channel uses.
   - AF-CS uses one channel use in the sensing phase and R in the relay phase, hence a total of $R + 1$ channel uses.
4. The empirical cost function $\tilde{\nu}(\beta)$ is an indicator of the trade-off between energy savings and accuracy. We have set $\beta = 0.5$ for a better viewing of such a trade-off.

Fig. 4.12 summarizes the results for the configuration $K^\star = 11$, $R^\star = 43$ for $\beta = 0.5$ obtained in Fig. 4.10 for the evaluated topologies.

We consider the link among relays and the fusion center as error-free because this link can include error correction protocols making the communication more robust. However, this is not possible in the link between active nodes and relays.

First, in terms of MSE, it is obvious that the best performance is obtained with the CA, since there is no reconstruction error. Moreover, our

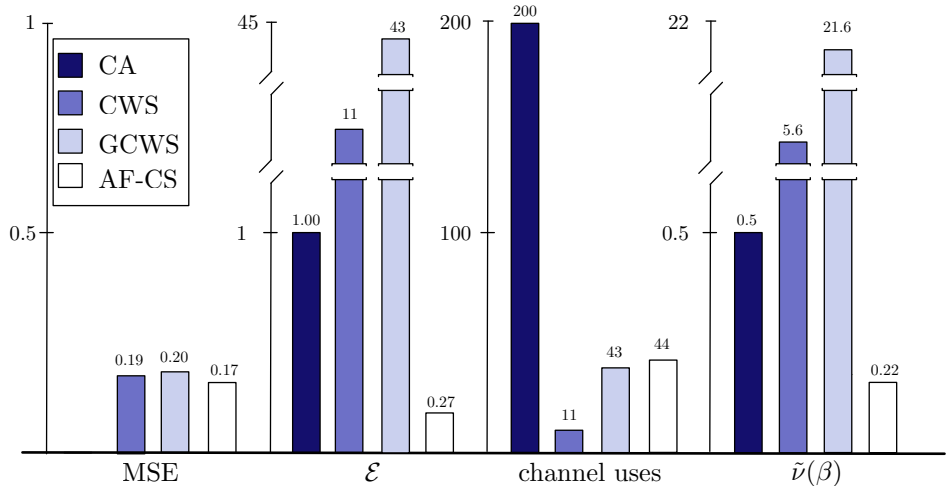

Figure 4.12: Graphic bar comparison of the different figures of merit among the studied CS schemes. Lower values are better.

proposed method performs slightly better than CWS (0.17 against 0.19, or even 0.20). Although the DCT is very suitable for smooth signals (the signal is already quite smooth since $\rho = 0.99$), for smoother signals the performance of CWS improves.

The major difference among the schemes can be observed in terms of the energy consumption. Our method performs the best compared to the other evaluated techniques. One can see that CWS performs even worse than the CA, and that the GCWS is extremely inefficient. The same conclusions can be extracted from $\tilde{\nu}(\beta)$.

Regarding the number of channel uses, CWS performs the best with 11 channel uses. However, our proposed method performs similar to GCWS with 44 in front of 43. Besides, a big improvement in comparison to CA can be appraciated with 200 channel uses.

4.7.3 Robustness against additive white Gaussian noise

Most of the wireless systems are contaminated by some amount of noise in the communication link. In this section, we evaluate by simulation the performance reduction due to the AWGN comparing our method with three different instances of the CWS approach:

- *Classic* CWS: It is the case studied in the previous section where we keep the K most important DCT coefficients of $\boldsymbol{\omega}_K = \boldsymbol{\Psi}_K \mathbf{x}(n)$, while the rest are set to zero.

- *Random* CWS: The unfeasible assumption that the sensors know the ordering of the K-largest DCT coefficients is relaxed, and thus, K coefficients are randomly selected.

- *Low* CWS: The fusion center selects the first K DCT coefficients, because typically smooth signals concentrate the most information in low frequencies.

- GCWS: It is the case proposed in the previous sections where R linear combinations of the transformed vector are selected, that is $\boldsymbol{\Phi}\boldsymbol{\Psi}\mathbf{x}(n)$.

Simulation results in Fig. 4.13 show the performance of the methods analyzed as a function of the SNR. For the AF-CS, the SNR is defined as the ratio between the power of the useful signal $\boldsymbol{\Phi}\mathbf{x}_K(n)$ and the power of the noise $\mathbf{w}(n)$:

$$\mathrm{SNR}_{\text{AF-CS}} = 10 \log \left(\frac{\|\boldsymbol{\Phi}\mathbf{x}_K(n)\|^2}{\|\mathbf{w}(n)\|^2} \right). \tag{4.55}$$

For the CWS, the SNR has de form,

$$\mathrm{SNR}_{\text{CWS}}^{\{\text{classic, random, low}\}} = 10 \log \left(\frac{\|\boldsymbol{\omega}_K(n)\|^2}{\|\mathbf{w}(n)\|^2} \right), \tag{4.56}$$

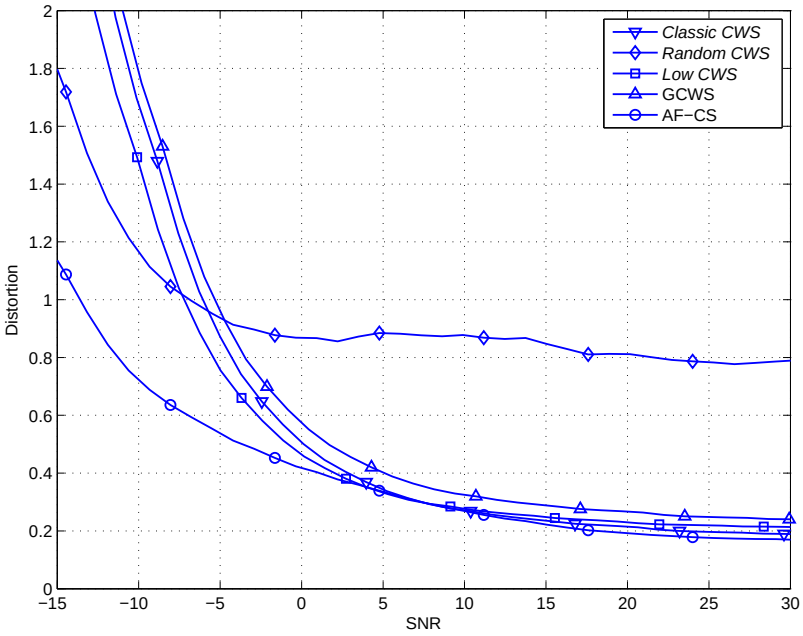

Figure 4.13: Comparison between the MSE as a function of the SNR for different instances of CWS, the GCWS, and AF-CS by solving $\mathcal{P}2$ and $\mathcal{P}'2$, respectively. The simulation setup is for $K = 11$ and $R = 43$ (results in Table 4.2). The figure has been averaged over 100 realizations.

where

$$\mathrm{SNR}_{\mathrm{CWS}}^{\mathrm{random}} \leq \mathrm{SNR}_{\mathrm{CWS}}^{\mathrm{low}} \leq \mathrm{SNR}_{\mathrm{CWS}}^{\mathrm{classic}}, \tag{4.57}$$

holds for the same configuration setup. For the case of GCWS, the SNR is approximated by,

$$\mathrm{SNR}_{\mathrm{GCWS}} = 10\log\left(\frac{\|\boldsymbol{\omega}_K(n)\|^2}{\|\mathbf{w}(n)\|^2}\right) \simeq 10\log\left(\frac{R\|\mathbf{x}(n)\|^2}{S\|\mathbf{w}(n)\|^2}\right). \tag{4.58}$$

The physical intuition behind (4.57) and (4.58) is that the Classical CWS selects the K larger coefficients. They concentrate the most power of the transformed vector $\boldsymbol{\omega}(n)$. Hence for a given configuration, the SNR is higher than the other cases. Similarly, the Low CWS selects high power coefficients with higher probability than the Random CWS case. On the other hand, GCWS measures R linear combinations among all the coefficients of $\boldsymbol{\omega}(n)$, thus the power is a fraction R/S of the total power of $\mathbf{x}(n)$.

First we study the impact of the selection of the DCT coefficients. It is easy to conclude from the simulation results that a random selection of the coefficients is not a good choice for the simulated scenario. Even for high SNR levels, the distortion is not lower than 0.8. Otherwise, the *low* CWS achieves a distortion performance quite close to the actual Classical CWS performance. So, it can be a good alternative in order to relax the assumption of the K largest coefficients.

Furthermore, we evaluate and compare the robustness against the noise contamination. We observe that the Classical CWS, the Low CWS, and the GCWS have similar performance while AF-CS outperforms clearly the rest of the schemes in terms of the SNR. That is why for a given configuration, the AF-CS spends much less energy per time slot. However, we can see that the AF-CS scheme is more robust than the CWS-based schemes since the slope of its curve decays more slowly.

4.8 Conclusions

This chapter has introduced a distributed solution for a compressed sensing implementation in a Wireless Sensor Network scenario, which is referred to as Amplify-and-Forward Compressed Sensing (AF-CS). In particular, it presents an energy-efficient design for a star-topology sensor network based on an Amplify-and-Forward configuration. First, the sensing nodes exploit inner time correlation in order to reduce the number of transmissions, only keeping as active nodes the K sensors with the most unpredictable readings. In such a way, we distributedly transform the vector of interest $\mathbf{x}(n)$ in a K-sparse approximation $\mathbf{x}_K(n)$. Second, the relay nodes receive random projections formed from the linear combination of the K readings of the active sensors, each one multiplied by its channel gain. In order to do so, the active nodes transmit time-synchronized. Thus, the number of channel uses may be drastically reduced. The relay nodes retransmit the received random projections to the fusion center. Afterwards, the fusion center reconstructs the original vector using an l^1-norm minimization (widely used in the compressed sensing framework).

Furthermore, we have described an analytical procedure in order to characterize the distortion caused by both the linear prediction and the compressed sensing reconstruction process. The obtained error model allows us to dimension a priori the sensor network, i.e., the number of active sensors K and the minimum number of relay nodes R that are needed in order to guarantee the established performance requirements, which take into account the energy consumption and the signal distortion. To do so, we propose a design criteria based on a cost function that controls the existing trade-off between the energy consumption and the signal distortion.

The simulation results (we show the case when the number of transmissions and the distortion are equally weighted) indicate that our proposed scheme drastically reduces the number of transmissions and the number of channel uses compared to a classical transmission scheme. the AF-CS

scheme also outperforms other distributed compressed-sensing-based techniques for wireless sensor networks not only in terms of energy-efficiency but also in terms of robustness agains noise.

In summary, the main contributions of our proposed scheme in front of other CS techniques in the literature are listed as:

- AF-CS presents a practical and distributed scheme that does not require some unrealistic assumptions of other schemes in the literature.
- AF-CS focuses on the minimization of the number of transmissions and not only the transmitted power. This approach is proved to be way more realistic and more suitable for energy-constrained scenarios as WSNs.
- AF-CS reduces drastically the number of channel uses in front of classical schemes and at the same time maintain similar number of channel uses than other distributed CS techniques.

Chapter 5

Amplify-and-Forward Compressed Sensing as a Physical-Layer Secrecy Solution

5.1 Summary

Physical layer secrecy is an emerging security concept that achieves secure data transmission in presence of eavesdropping nodes at the physical layer. Results in compressed sensing show that this technique can be applied to wireless sensor networks in order to reduce the power consumption and the amount of channel uses. In this chapter we extend the results on the amplify-and-forward compressed sensing scheme (AF-CS) with the study of the physical layer secrecy performance for the case when malicious eavesdropping nodes are listening. In particular, we investigate the robustness of the AF-CS scheme in presence of a group of coordinated eavesdropping nodes under the assumption that they have a corrupted channel state information. In order to fulfill this assumption, we propose a channel esti-

mation technique based on random pilots. This technique introduces extra uncertainty only in the channel estimation of the eavesdroppers. Our simulation results evaluate the physical layer robustness as a function of the total number of coordinated eavesdroppers and the level of channel estimation distortion of the eavesdroppers. We demonstrate that this scheme achieves perfect secrecy in presence of a small number of eavesdroppers (the meaning of small will be defined later on). We also show that a very high number of eavesdropping nodes are required to perfectly recover the signal in comparison to other distributed compressed sensing schemes in the literature.

5.2 Introduction

Wireless communications have been extended to virtually all the possible scenarios and applications. Unfortunately, security risks are inherent in any wireless transmission. This is mainly because the communication channel is open to any intruder. Therefore, unauthorized entities may exploit this scenario in order to obtain confidential information, to corrupt the transmitted data, to degrade the network performance, etc.

For our convenience, we differentiate these attacks in two groups [Sri08]:

- *Physical layer attacks.* They benefit from the wireless connection nature. Mainly, there exist two kind of attacks at physical layer: *i*) the eavesdropping that refers to the existence of one/many unauthorized receiver/s trying to extract information from the signal present in the wireless channel, and *ii*) the jamming [Kas04, Sha05], that refers to the existence of a malicious transmitter (or a group of them) that intentionally degrades the intended communication by introducing additional interference. Although both approaches are extremely interesting, we focus our study on the robustness of the AF-CS scheme against eavesdropping attacks.

- *Upper layer attacks*. They are mostly related to the application layer and hence they can be performed in both wired and wireless systems. The analysis of these attacks is beyond the scope of this thesis.

5.2.1 Physical-Layer secrecy background

Physical-Layer Security proposes different mechanisms in order to protect the wireless communication against mainly malicious jammers and/or unauthorized eavesdroppers. We address the latter case, and throughout this document, this case is referred to as Physical-Layer (PHY-layer) secrecy. Therefore, the basic aim of the PHY-layer secrecy is to allow reliable transmission of confidential messages over a wireless link in presence of eavesdroppers.

This issue has been traditionally addressed using spread spectrum techniques such as Code Division Multiple Access (CDMA). Thanks to the pseudorandom codes that can be seen as secret keys, the intended signal is converted in a noise-like signal for any receiver that does not possess the code. However, in general CDMA has the limitation of short keywords (about 24 bits [Sri08]) and a persistent eavesdropper could get the key with some little effort. That is why upper-layer cryptographic techniques have been used so far. These techniques assume large keys that make the message almost impossible to decipher. However, cryptographic mechanisms have two main problems in the wireless scenario: first, the distribution of the key over a public medium, and second, the high computational complexity of the cryptographic mechanisms which goes beyond the hardware and power limitations of some devices such as in Wireless Sensor Networks (WSNs).

For these reasons and according to the proliferation of wireless communications, the interest on secrecy mechanisms at physical layer has dramatically grown in the last decade. However, this concept is not new in the literature and comes from 1949, when Shannon postulated the foundations

of the modern cryptography in his seminal work [Sha49]. The proposed scheme assumes that a transmitter sends encrypted information using a non-reusable key over a noiseless channel and with the presence of an eavesdropper that has access of the transmit coding scheme and the transmitted signal. In that paper, the author postulates the conditions for the code to ensure perfect secrecy, term that is properly defined next. This work opened a whole branch of key-based secrecy research.

Later in the seventies, Wyner opened a new branch of research about keyless secrecy techniques in [Wyn75]. In particular, the author assumes that the eavesdropper has the additional effect of the non-ideal wiretap channel [Car77], which can be seen as a degraded version of the main channel. Under this assumption, it obtains the maximum rate over the main channel while the amount of information is negligible in the wiretap channel. For the subsequent works, the interested reader may find an excellent review about PHY-layer secrecy in [Muk10].

In order to quantify the level of secrecy of a proposed method, we can define the following figures of merit.

- *Mutual Information.* This metric measures the mutual dependence of two continuous random variables, the original message X and the decoded message Y. It is mathematically formulated as

$$\begin{aligned} I(X;Y) &= \mathcal{H}(X) - \mathcal{H}(X|Y) \\ &= \int_Y \int_X f(x,y) \log\left(\frac{f(x,y)}{f(x)f(y)}\right) dxdy. \end{aligned} \tag{5.1}$$

  where $\mathcal{H}(X)$ and $\mathcal{H}(X|Y)$ denote the entropy of X and the conditional entropy of X when Y is known. For the discrete random variables case, the double integral is replaced by summations as

$$I(X;Y) = \sum_{y\in Y} \sum_{x\in X} f(x,y) \log\left(\frac{f(x,y)}{f(x)f(y)}\right). \tag{5.2}$$

For the case where X and Y to be independent, the mutual information is $I(X;Y) = 0$, which means *perfect secrecy*.

- *Equivocation Rate* [Mar11]. It is a measure of the amount of information that the eavesdropper can get from the message and is defined as

$$R_e = \frac{\mathcal{H}(X|Y)}{N}, \tag{5.3}$$

where N is the codeword length. When $R_e = \mathcal{H}(X)/N$, then the mutual information $I(X;Y) = 0$, which means *perfect secrecy*.

- *Wiretap Distortion*. It measures the normalized squared error of the eavesdropper decoded signal with respect to the intended one, i.e.,

$$\mathcal{D}_e = \mathbb{E}\left[\frac{\|X - Y\|^2}{\|X\|^2}\right]. \tag{5.4}$$

- Other common metrics in the literature are *Secrecy Rate* [Ogg08], and *Secrecy Capacity* [Sha07].

One of the common ways to protect the intended message against eavesdropping is to use opportunistic transmissions when the wiretap channel experiments a fading [Gop08]. Doing so, one can obtain reliable transmissions even when the eavesdropper has a better average SNR than the legitimate receiver [Li07]. This approach can be extended to a general MIMO scenario, where the transmitter adds a precoding matrix which is orthogonal to the wiretap channel matrix. [Zha10].

These techniques require a perfect knowledge of the Eavesdropper Channel State Information (ECSI). On the other hand, relatively fewer studies consider the case of a complete absence of the wiretap channel knowledge by the intended pair transmitter-receiver. However, other works require only the knowledge of the statistics of the wiretap channel [Neg05, Goe08]. They

use an artificial noise injected to the signal in order to degrade the quality of the wiretap channel. Moreover, the authors of [Muk10] show the poor performance of waterfilling techniques when no information is available regarding the eavesdropper channel.

Many other studies regarding different network configurations are also being actively studied, with new alternatives and different secrecy performances. We do not review these other exciting activities, but focus our attention on the secrecy schemes that we can build using Compressed Sensing (CS) strategies.

5.2.2 Is the compressed sensing a good candidate to build PHY-Layer secret strategies?

Compressed Sensing is a novel tool that allows us to sample the signals below the Nyquist rate [Don06b], and it is specially powerful in scenarios where the signals are sparse or compressible in a certain basis domain, as in image signal processing or detection (the interested reader is encouraged to visit the Rice's CS database at `dsp.rice.edu/cs`). However, only a very few works relates CS with secrecy. In fact, we review the most relevant contributions up to date, that are collected in the following four conference papers: [Rac08], [May10] for the key-based secrecy case, and [Agr11], [Ree11] for the keyless secrecy case.

The works in [Rac08], and [May10] consider the scenario in Fig 5.1, with one source, one receiver, and one eavesdropper. The product of the transform matrix and the sensing matrix, i.e., $\mathbf{\Phi\Psi}$ can be seen as an encryption key, which is assumed to be unknown by the eavesdropper[1].

Thus, the eavesdropper receives an exact copy of the transformed measurements, i.e.,

$$\mathbf{y}(n) = \mathbf{\Phi\Psi x}(n). \tag{5.5}$$

[1] All the CS parameters are properly defined in the next section.

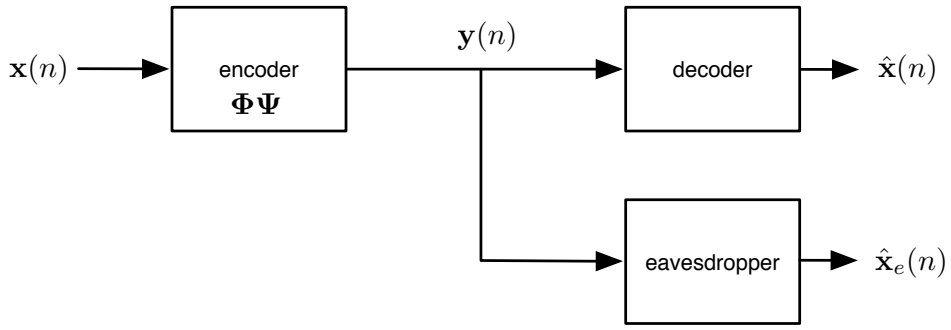

Figure 5.1: Block diagram of the key-based PHY-Layer secrecy scenario.

Therefore, the paper studies how difficult it is for the eavesdropper to recover $\mathbf{x}(n)$ from the measurements $\mathbf{y}(n)$ without the knowledge of the key $\mathbf{\Phi\Psi}$. Actually, it has been proved (see Lemma 1 in [Rac08]) that compressed sensing encryption does not achieve perfect secrecy, i.e., $I(\mathbf{x}(n);\mathbf{y}(n)) > 0$. The mutual information $I(\mathbf{x}(n);\mathbf{y}(n)) = \mathbf{0}$ would be zero if and only if $\mathbf{x}(n)$ and $\mathbf{y}(n)$ are independent. However, since $\mathbf{\Phi\Psi}$ is lineal, $\mathbf{x}(n) = \mathbf{0}$ implies that $\mathbf{y}(n) = \mathbf{0}$, and hence $P(\mathbf{y}(n) = \mathbf{0}|\mathbf{x}(n) = \mathbf{0}) \neq P(\mathbf{y}(n) = \mathbf{0})$, meaning statistical dependence.

Furthermore, the authors in [Rac08] introduce the concept of *Computational Secrecy*. It is applied to the cases when the encrypted data contains complete information about the message but extracting this information will be equivalent to solve a computational problem (NP-hard discovery). According to that definition, they proved that the eavesdropper cannot decode the message using a different (wrong) key $\mathbf{\Phi'\Psi'}$ in the reconstruction with probability one, for the case when $\mathbf{\Phi\Psi}$ and $\mathbf{\Phi'\Psi'}$ are independent.

The work in [May10] extends the results in [Rac08] and considers the perfect secrecy problem using CS. They show that under certain conditions, perfect secrecy via CS is achievable. The conditions are:

1. The signal $\mathbf{x}(n)$ has a uniform distribution over a given alphabet.

2. The key $\mathbf{\Phi\Psi}$ is $R \times S$, where R is the number of measurements, S is the dimension of the signal $\mathbf{x}(n)$, and K is the number of non-zero values (sparsity level) of $\mathbf{x}(n)$. Thus, the condition $R > 2K$ is imposed.

3. The matrix $\mathbf{\Phi}$ holds *Restricted Isometry Property* (RIP) [Can06a] (already introduced in eq. 4.33).

4. The number of source messages goes to infinity.

On the other hand, the works in [Agr11] and [Ree11] consider the scenario in Fig. 5.2. This scenario is different than the one in Fig. 5.1 because *i*) the CS encoding matrix $\mathbf{\Phi\Psi}$ is also known by the eavesdropper, and *ii*) they consider the effect of the intended channel and the wiretap channel. The signal model is:

$$\mathbf{y}(n) = \mathbf{H\Phi\Psi x}(n) + \mathbf{w}(n) \tag{5.6}$$

for the legitimate user, and

$$\tilde{\mathbf{y}}(n) = \tilde{\mathbf{H}}\mathbf{\Phi\Psi x}(n) + \tilde{\mathbf{w}}(n) \tag{5.7}$$

for the eavesdropper, where $\mathbf{H} \in \mathbb{R}^{S\times S}$ and $\tilde{\mathbf{H}} \in \mathbb{R}^{S\times S}$ are the intended and the wiretap (flat-fading) channel matrices respectively, and $\mathbf{w}(n)$ and $\tilde{\mathbf{w}}(n)$ model the Additive White Gaussian Noise (AWGN) of the wireless channels, both distributed as $\mathcal{N}(0, \sigma^2_{\mathbf{w}})$

Differently from [Rac08], the authors in [Agr11] do not focus on neither perfect secrecy nor computational secrecy. They introduce the *Wolfowitz* secrecy, which states that the eavesdropper is unable to decode the message with high probability, or equivalently, that the eavesdropper's probability of recovery can be made arbitrarily small. So, it lies between perfect secrecy (from the information theory perspective) and computational secrecy.

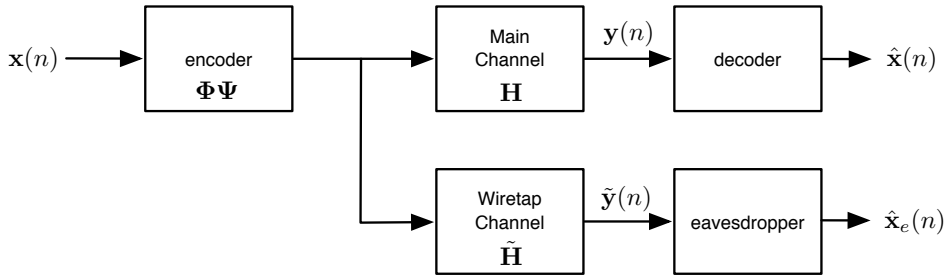

Figure 5.2: Block diagram of the keyless PHY-Layer secrecy scenario.

The authors demonstrated that it is possible to ensure Wolfowitz secrecy if the average singular value of $\tilde{\mathbf{H}}$ is less than a given constant multiple of the minimum singular value of $\mathbf{H}$.

In the same line, the authors in [Ree11] study secrecy capacity of the wiretap channel in Fig. 5.2 when CS-like matrices are using to encode the message.

5.2.3 Are the current CS schemes a good PHY-Layer secrecy solution for WSNs?

Following the references addressed above, CS can be seen as a good candidate in order to provide PHY-layer secrecy agains eavesdropping in addition to the other CS benefits.

However, the scenario that they consider is a point-to-point communication that involves only a single transmitter that compresses the signal, one receiver and one eavesdropper. Hence, this scenario follows a centralized approach that is not directly applicable to our scheme due to the distributed nature of the WSN environment. These limitations have been already addressed in Section 4.2 and can be summarized as follows:

- If the signal $\boldsymbol{\omega}(n) = \boldsymbol{\Psi}\mathbf{x}(n)$ is not purely sparse, the CS encoding cannot be directly applied in a distributed fashion, since the K-largest coefficients of $\boldsymbol{\omega}(n)$ have to be selected first.

- High energy consumption and channel uses per measurement are required.

In order to overcome these problems, a Compressed Wireless Sensing (CWS)-like approaches can be proposed, as the one in [Baj06] or the Generalized CWS (GCWS) introduced in Section 4.2. Although they are designed to face with energy-efficient communications, one can expect that they could provide some PHY-layer secrecy as well.

In order to do so, two main issues have to be taken into account:

- Find a way to send the CS matrix (key) securely over the wireless channel.

- Design a robust CS matrix that cannot be easily discovered from the measurements. The authors in [And12] propose an algorithm that allows to discover the CS matrix from only a few measurements when the matrix has some structure, e.g., Fourier matrix, Discrete Cosine Transform (DCT) matrix, Toeplitz matrix, etc.

In conclusion, the current PHY-Layer secure CS schemes are not suitable for distributed scenarios. Moreover, the current distributed CS techniques are not designed to provide PHY-Layer secrecy. Hence, new CS schemes are needed in order to implement secret systems in WSNs.

5.2.4 Our contribution

As a solution of the problems above, we propose the AF-CS as a distributed and secret CS scheme. The AF-CS was first introduced in [BL11] and detailed in [BL12b]. As we have already shown in Chapter 4, AF-CS is able

to reduce the energy consumption using, at the same time, a very limited number of channel uses and following the distributed approach of WSNs.

In this chapter, we address the secrecy level of AF-CS scheme in presence of not only one but a group of coordinated and passive eavesdroppers.

In order to provide the so-called PHY-layer secrecy, the system takes advantage of the linear combinations that are produced on the air thanks to the Multiple Access Channel (MAC). This idea comes from the Network Coding theory, where the messages are not treated as indivisible, but instead, algebraic manipulations are allowed.

In fact, the MAC matrix is used as the CS matrix and it can be also seen as the encryption key of the PHY-layer secrecy scheme following the key-based approach in [Rac08] and [May10]. On the other hand, we also consider the case that the eavesdroppers suffer from the effect of the wiretap channel as in [Agr11], [Ree11], where we assume that the eavesdroppers only have access to a degraded estimation of the wiretap channel matrix.

Actually, our approach cannot be classified as either a key-based or a keyless approach. Instead, it follows a more general scheme that can be seen as a combination of both approaches. The subsequent advantages and contributions with respect to each of the approaches are listed next.

- *Key-based PHY-secrecy.* One of the challenges in key-based schemes is to securely exchange the key matrices. In AF-CS, there is no exchange of CS matrix since the sensing nodes do not need the matrix to encode the message. Instead, the encoding is performed *on the air*. It reduces the computational complexity derived from the encryption process. Hence, it has a great impact on the design of the WSNs because the sensing nodes are very hardware limited.

- *Keyless PHY-secrecy.* Although we consider that both the intended and the wiretap communications are perturbed by the wireless channel as in the wiretap scenario, we do not assume any knowledge of the wiretap channel by the intended nodes. As it is said in [Muk10], very

few works in the literature consider the case of no ECSI at either the transmitter or the intended receiver.

In order to relax the assumption that the eavesdroppers only have access to a degraded estimation of the wiretap channel matrix, we also propose a secure channel estimation technique based on random pilots that allows the system to control the distortion introduced to the channel estimate of the eavesdroppers and hence to guarantee a desired secrecy level. This strategy is similar than the artificial noise injection proposed in [Neg05, Goe08], but with the main difference that the noise is not used to encode or mask the signal but only the pilots in the channel estimation. Doing so, the energy consumption is reduced in comparison to the artificial noise injection method.

Moreover, we compare our proposed AF-CS with CWS-like schemes and we find out that AF-CS dramatically increases the protection agains eavesdropping at physical layer. We also analyze the secrecy performance in front of two different approaches: *i*) a conventional star-topology WSN monitoring a physical scalar magnitude, e.g., temperature or humidity, and *ii*) a detection scenario where only few sensors are active at a time, like in a wildfire detection application.

5.2.5 Organization of the chapter

The rest of the chapter is organized as follows. In Section 5.3 we present the problem statement and the assumptions considered throughout the chapter. Section 5.4 details the secrecy properties of the AF-CS. The random pilot technique is introduced in Section 5.5. Simulation results are given in Section 5.6, and conclusions are drawn in Section 5.7.

5.3 System Model and Assumptions

We consider a WSN configured in star-topology that monitors a given physical scalar magnitude (e.g., temperature, humidity) or detects a physical event (e.g., wildfire). In particular we assume the scheme in Fig. 5.3, that is:

- A set $\mathcal{S}$ of S sensing nodes connected (wirelessly) to one fusion center, that acts as the intended receiver. Their measurements at discrete time n are represented by $\mathbf{x}(n)$.

- A subset $\mathcal{K}(n) \subseteq \mathcal{S}$ (of cardinality K) of active sensors that are transmitting at a given time n. The transmitted vector is $\mathbf{x}_K(n)$ where only K positions are different to zero. The remaining sensors in $\mathcal{Q}(n) = \mathcal{S} \setminus \mathcal{K}(n)$ (of cardinality Q) remain silent.

- A subset $\mathcal{R} \subseteq \mathcal{S}$ (of cardinality R) acts as relay nodes in Amplify-and-Forward (AF) mode.

- A set $\mathcal{E}$ (of cardinality E) of malicious and passive eavesdropping nodes.

According to the CS nomenclature, K also corresponds to the number of non-zero elements of the transmitted vector $\mathbf{x}(n) \in \mathbb{R}^S$, and R is the number of measurements used in the reconstruction process at the fusion center, i.e., the number of rows of the sensing matrix, $\mathbf{\Phi} \in \mathbb{R}^{R \times S}$, where typically $K < R < S$. On the contrary, the eavesdroppers use E measurements to try to decode the signal, i.e., the number of rows of the sensing matrix, $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{E \times S}$, used by the eavesdroppers.

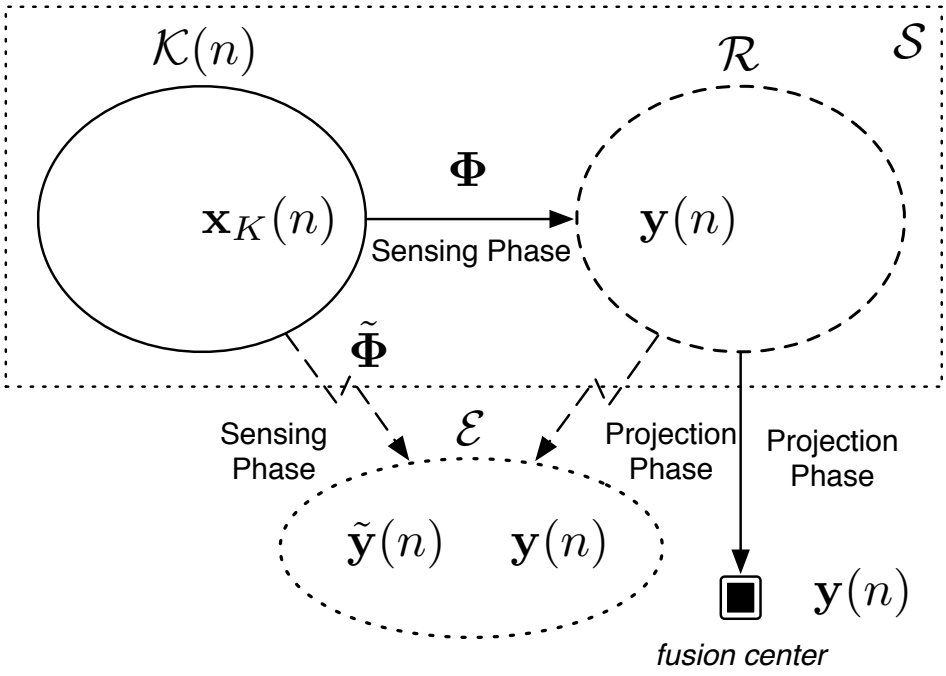

Figure 5.3: Multiple active channel scenario composed by K active sensing nodes, R relay nodes, E eavesdropping nodes, and one fusion center.

We consider that the signals are space-time correlated and modeled as an S-dimensional stochastic process, namely,

$$\mathbf{X} = [\mathbf{x}(1)\,\mathbf{x}(2)\,\ldots\,\mathbf{x}(N)\,] = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(N) \\ x_2(1) & x_2(2) & \cdots & x_2(N) \\ \vdots & \vdots & & \vdots \\ x_S(1) & x_S(2) & \cdots & x_S(N) \end{bmatrix}, \qquad (5.8)$$

where $x_s(n)$ denotes the measurement of the sth sensor at the sample time n and N denotes the number of time samples in the observation window.

The main assumptions throughout this chapter are collected as follows:

5.3.1 Assumptions on the signal model

The assumption on the signal model are assumed to be the same as in Chapter 4, and they are summarized as follows.

Let $x(n)$ be a real and time-discrete auto-regressive model of order 1 ($AR-1$) defined as:

$$x(n) = \rho x(n-1) + z(n), \quad \text{for } n = 1, 2, \ldots \tag{5.9}$$

The auto-regression coefficient is denoted by $\rho \in [0, 1]$ and it is assumed to be constant during the transmission. The random process $z(n)$ is a sequence of Gaussian distributed and independent random variables with zero mean and variance σ_z^2.

The covariance matrix of an $AR-1$ model follows

$$[\mathbf{R}]_{n,n-i} = \rho^i. \tag{5.10}$$

Without loss of generality, we assume that $\sigma_x^2 = 1$ and that the noise variance $\sigma_z^2 = 1 - \rho^2$. We also assume that $x(n)$ is a continuous-valued process or, in other words, that the quantization error is assumed to be zero.

5.3.2 Assumptions on the channel and the system model

The assumptions regarding the channel and the system model are also similar to the ones in Chapter 4 with the main difference that we take here into account also the presence of a set of eavesdroppers.

We assume perfect Channel State Information (CSI) at the fusion center for all the links that go from any sensing node to a relay node in $\mathcal{R}$. In particular, we assume that the channel matrix, also referred to as the sensing

matrix $\mathbf{\Phi} \in \mathbb{R}^{R \times S}$ follows the Gaussian measurement ensemble, where:

$$[\mathbf{\Phi}]_{i,j} \sim \mathcal{N}(0, R^{-1}). \tag{5.11}$$

The variance of the sensing matrix R^{-1} is a convention in the literature in order to maintain the relation

$$\mathbb{E}\left[\|\mathbf{\Phi}\mathbf{x}\|\right] = \mathbb{E}\left[\|\mathbf{x}\|\right] \tag{5.12}$$

for an arbitrary vector $\mathbf{x}$. This assumption is just for convenience and it does not affect in the generality of the model since the channel gain can be adjusted at the receiver if needed.

On the other side, we assume partial knowledge of the CSI at the eavesdroppers for all the links that go from any sensing node to a relay node in $\mathcal{R}$. The wiretap channel matrix, also referred to as wiretap sensing matrix $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{E \times S}$ follows the Gaussian measurement ensemble, where:

$$[\tilde{\mathbf{\Phi}}]_{i,j} \sim \mathcal{N}(0, E^{-1}). \tag{5.13}$$

However, there is no mutual channel knowledge in the sense that the eavesdroppers do not have access to $\mathbf{\Phi}$, and that the fusion center does not need to know $\tilde{\mathbf{\Phi}}$. Therefore, the typical assumption of perfect or partial ECSI at the intended receiver is relaxed.

Moreover, we do not assume anything regarding the links from $\mathcal{R}$ to the fusion center other than these links are controlled by a certain orthogonal policy that requires R channel uses for each sample time n in order to transmit the data from the relays to the fusion center.

5.4 Eavesdropping the Amplify-and-Forward Compressed Sensing Scheme

In this chapter, we consider the AF-CS algorithm already presented in Chapter 4, which is summarized in the following three phases:

1. *Sensing phase*. It proposes a distributed method in order to select the K most relevant readings of the transmitted vector $\mathbf{x}(n) \in \mathbb{R}^S$ based on the inner time correlation. These readings are collected in a K-sparse vector, $\mathbf{x}_K(n) \in \mathbb{R}^K$ and broadcasted time-synchronized using uncoded transmissions to the relay nodes.

2. *Projection phase*. Each relay has received linear combinations of $\mathbf{x}_K(n)$ thanks to the MAC, modeled by the sensing matrix, $\mathbf{\Phi}$. Then, it relays them in AF mode to the fusion center using a given orthogonal transmission (e.g., time multiplexing).

3. *Reconstruction phase*. The fusion center collects the projections from all the relays in the vector $\mathbf{y}(n)$ and solves the l^1-norm minimization program $\mathcal{P}1$ [Don06b],

$$\begin{aligned} \mathcal{P}1: \quad & \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} && \|\hat{\mathbf{x}}_K(n)\|_{l^1} \\ & \text{subject to} && \mathbf{y}(n) = \mathbf{\Phi}\hat{\mathbf{x}}_K(n) \end{aligned} \tag{5.14}$$

   in order to obtain an accurate reconstruction of $\mathbf{x}_K(n)$, named $\hat{\mathbf{x}}_K(n)$, for the noiseless case. For the noisy case, the fusion center solves

$$\begin{aligned} \mathcal{P}2: \quad & \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} && \|\hat{\mathbf{x}}(n)\|_{l^1} \\ & \text{subject to} && \|\mathbf{y}(n) - \mathbf{\Phi}\hat{\mathbf{x}}(n)\|_2 < \varepsilon, \end{aligned} \tag{5.15}$$

   for ε as an upper bound on the magnitude of the noise. Afterwards, the fusion center completes the remaining Q entries of the vector $\mathbf{x}(n)$ using a linear prediction in order to get the full $\hat{\mathbf{x}}(n)$.

Next, we assess the PHY-layer secrecy performance of the sensing and projection phases. Note that we do not study the eavesdropping in the reconstruction phase because wireless transmissions are involved only in the first two phases.

5.4.1 Eavesdropping during the sensing phase

During this phase, all the sensors in $\mathcal{K}(n)$ broadcast their readings, and hence the relay sensors receive linear combinations due to the nature of the MAC, namely,

$$\mathbf{y}(n) = \mathbf{\Phi}\mathbf{x}_K(n) + \mathbf{w}(n), \tag{5.16}$$

where the vector $\mathbf{y}(n) \in \mathbb{R}^R$ stacks all the received signals of the nodes in $\mathcal{R}$, the sensing matrix $\mathbf{\Phi}$ models the channel between $\mathcal{S}$ and $\mathcal{R}$ as a random matrix with i.i.d. Gaussian entries with zero mean and variance $\sigma^2_{\mathbf{\Phi}} = R^{-1}$. Finally, $\mathbf{w}(n)$ denotes white Gaussian noise with zero mean and variance $\sigma^2_{\mathbf{w}}$.

Similarly to (5.16), the received signal at the eavesdroppers is:

$$\tilde{\mathbf{y}}(n) = \tilde{\mathbf{\Phi}}\mathbf{x}_K(n) + \tilde{\mathbf{w}}(n). \tag{5.17}$$

where $\tilde{\mathbf{y}}(n) \in \mathbb{R}^E$ stacks the signals received by the nodes in $\mathcal{E}$, and $\tilde{\mathbf{\Phi}}$ models the channel between $\mathcal{S}$ and $\mathcal{E}$ as a random matrix with i.i.d. Gaussian entries with zero mean and variance $\sigma^2_{\tilde{\mathbf{\Phi}}} = E^{-1}$ and $\tilde{\mathbf{w}}(n)$ denotes white Gaussian noise with zero mean and variance $\sigma^2_{\mathbf{w}}$.

For the sake of simplicity and without loss of generality, we focus on the noiseless problem. Moreover, we will address two different cases: *i*) PHY-Layer secrecy with perfect CSI at the eavesdroppers, and *ii*) PHY-Layer secrecy with imperfect CSI.

5.4.1.1 Eavesdroppers with perfect CSI

Here, we assume that the eavesdroppers have perfect knowledge of the wiretap sensing matrix $\tilde{\mathbf{\Phi}}$. Then, the eavesdroppers would have to jointly

solve the following problem:

$$\mathcal{P}1_{\mathrm{e}}: \quad \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad \|\hat{\mathbf{x}}_K(n)\|_{l^1} \qquad \text{subject to} \quad \tilde{\mathbf{y}}(n) = \tilde{\mathbf{\Phi}}\hat{\mathbf{x}}_K(n). \tag{5.18}$$

This problem is the classical CS decoder, CSD. The most common way to address its performance is by means of the Restricted Isometric Property (RIP), which is detailed in Definition 4.33. We reproduce it here using the new nomenclature for the eavesdropper set as:

Definition 5.1 *[Can05]: A matrix $\tilde{\mathbf{\Phi}}$ satisfies the RIP of order K with restricted isometry constant $\delta_K \in (0,1)$ if*

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \leq \|\tilde{\mathbf{\Phi}}\mathbf{x}\|_2^2 \leq (1+\delta_K)\|\mathbf{x}\|_2^2, \tag{5.19}$$

where $\tilde{\mathbf{\Phi}}_K \in \mathbb{R}^{E\times K}$ is formed by retaining any set of K or less columns from $\tilde{\mathbf{\Phi}}$, $\mathbf{x}$ is any K-sparse vector of the appropriate size, and δ_K is the smallest number (and not too close to one) that holds the RIP condition for each integer $K = 1, 2, ...$

Most of the CS literature agree that if the elements of the matrix $\tilde{\mathbf{\Phi}}$ are selected from an i.i.d. Gaussian measurement ensemble (as in 5.13), then $\tilde{\mathbf{\Phi}}$ will satisfy the RIP with overwhelming probability for $E \geq C_0 K \log S$ [Can08a] or even $E \geq C_0 K \log(S/K)$ [Can08b, Don06b], where C_0 is some positive constant. In addition to this already-existing results in the literature, we have incorporated a new relation, which is

$$\mathcal{C}_{\mathrm{CS}}: \quad E < K + K\log\left(\frac{S}{K}\right) \tag{5.20}$$

that fits better with the experimental results as we show in Chapter 4. Therefore, we will use $\mathcal{C}_{\mathrm{CS}}$ as the CS condition in order to determine if $\mathbf{x}_K(n)$ can be recovered from $\tilde{\mathbf{y}}(n)$ with high probability or not.

Although it is very difficult to predict what happens when $E \sim K + K\log(S/K)$ [Can11], we will differentiate three possible cases: *i*) low values of E, i.e., $E \leq K$, *ii*) moderate values of E, i.e., $K < E \leq K + K\log(S/K)$, and *iii*) high values of E, i.e., $E > K + \log(S/K)$, or what is the same, that E satisfies $\mathcal{C}_{\text{CS}}$.

For low values of E. We address the PHY-layer secrecy of the sensing phase with $E \leq K$ throughout the following lemmas and theorems.

Lemma 5.1 *Let $\mathcal{X}$ denote the solution set of the eavesdroppers that is composed of all the possible recovered vectors at the eavesdropper's decoder. Hence, the cardinality of $\mathcal{X}$ is at least*

$$M = \binom{S}{E} = \frac{S!}{(S-E)!E!}, \tag{5.21}$$

which means that the solution is not unique.

Proof For simplicity, we use the original (but computationally intractable) decoder $\mathcal{P}0_e$ instead of its convex relaxation $\mathcal{P}1_e$ [Don06b], which is defined as,

$$\begin{aligned} \mathcal{P}0_{\text{e}}: \quad & \underset{\hat{\mathbf{x}}_K(n)\in\mathbb{R}^S}{\text{minimize}} \quad && \|\hat{\mathbf{x}}_K(n)\|_0 \\ & \text{subject to} && \tilde{\mathbf{y}}(n) = \tilde{\mathbf{\Phi}}\hat{\mathbf{x}}_K(n), \end{aligned} \tag{5.22}$$

where $\|\cdot\|_0$ denotes the l_0 (pseudo)norm, defined as the sparsity of the signal.

In order to show that the solution is not unique, it is enough to prove that there exists at least one E-sparse vector $\mathbf{x}_E$ with loaded entries according to any subset of indices $\Omega_E \subset \mathcal{S}$ of cardinality E. Therefore, let the matrix $\tilde{\mathbf{\Phi}}_{\Omega_E}$ denote a $E \times E$ measurement matrix obtained by selecting the E columns of $\tilde{\mathbf{\Phi}}$ corresponding to the indices Ω_E and let the E-dimensional vector $\mathbf{x}_{\Omega_E}$ collect the E loaded entries of $\mathbf{x}_E$. It is verified that

$$\mathbf{y}(n) = \tilde{\mathbf{\Phi}}_{\Omega_E}\mathbf{x}_{\Omega_E} \tag{5.23}$$

for any Ω_E. Since the matrix $\tilde{\mathbf{\Phi}}_{\Omega_E}$ is full rank with overwhelming probability, a vector $\mathbf{x}_{\Omega_E} = \tilde{\mathbf{\Phi}}_{\Omega_E}^{-1}\tilde{\mathbf{y}}(n)$ exists for any Ω_E.

Since a number of

$$M = \binom{S}{E} = \frac{S!}{(S-E)!E!}, \tag{5.24}$$

different index sets of E elements over the set $\mathcal{S}$ can be generated, the proof of Lemma 5.1 is concluded.

Theorem 5.1 *For a given* $\mathbf{y}(n)$ *and* $\mathbf{\Phi}$*, let* $\mathcal{X}$ *denote the set of* E*-sparse vectors* $\hat{\mathbf{x}}$ *that are solution of* $\mathcal{P}0_e$ *in (5.22) and with cardinality* M*. Let* $\mathcal{X}'$ *denote the set of* K*-sparse vectors* $\mathbf{x}'$ *with cardinality* M' *that are solution of*

$$\mathbf{y}(n) = \mathbf{\Phi}\mathbf{x}'. \tag{5.25}$$

Therefore, if $E < K$*, the mutual information between any recovered vector* $\hat{\mathbf{x}}$ *and the original vector* $\mathbf{x}_K(n)$ *is zero, which means perfect secrecy.*

Proof We focus on $\mathcal{P}0_e$ because, if perfect secrecy holds for $\mathcal{P}0_e$, automatically it does for $\mathcal{P}1_e$ as well. In order to compute the mutual information between $\hat{\mathbf{x}}$ and $\mathbf{x}'$ we have

$$\begin{aligned} I(\hat{\mathbf{x}};\mathbf{x}') &= \mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}') \\ &= \mathcal{H}(\hat{\mathbf{x}}) - \sum_{\mathbf{x}\in\mathcal{X}'} \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x})P(\mathbf{x}'=\mathbf{x}) \\ &= \mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\hat{\mathbf{x}}) \sum_{\mathbf{x}\in\mathcal{X}'} P(\mathbf{x}'=\mathbf{x}) = 0. \end{aligned} \tag{5.26}$$

Note that the conditioned entropy $\mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x})$ is actually $\mathcal{H}(\hat{\mathbf{x}})$ since $\mathbf{x}$ is not a solution of $\mathcal{P}0_e$ and hence it does not modify $\mathcal{X}$.

The concept of Theorem 5.1 is graphically represented in Fig. 5.4, where the domains of $\mathbf{x}_K(n)$ and $\hat{\mathbf{x}}_K(n)$ are disjointed. It is also shown that $\mathbf{x}_K(n)$ cannot be recovered in any case from $\hat{\mathbf{x}}$.

Theorem 5.2 *For the case $E = K$, the mutual information between a recovered vector $\hat{\mathbf{x}}$ and the original vector $\mathbf{x}_K(n)$ is also zero, which means perfect secrecy.*

Proof This case slightly differs from the case where $E < K$ since $\mathcal{X} = \mathcal{X}'$ and the original vector $\mathbf{x}(n)$ is a possible solution of $\mathcal{P}0_e$. Even knowing that $\mathbf{x}_K(n)$ is inside the solution set, the decoder is not able to discriminate the correct solution among the others since all possible K-sparse (or E-sparse) $\hat{\mathbf{x}}$ minimize $\mathcal{P}0_e$. This fact is similar to the realization of the roll of a dice, where you know that the solution is one of the six possibilities, but all solutions are equally possible. Knowing that, $\mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}) = \mathcal{H}(\hat{\mathbf{x}})$. Hence, the mutual information equals zero.

Although for both cases $E < K$ and $E = K$ the system achieves perfect secrecy, some differences exist in the reconstruction probability.

Lemma 5.2 *If $E < K$, the eavesdroppers will recover the signal $\mathbf{x}_K(n)$ with zero probability, which is defined as*

$$P(\hat{\mathbf{x}}_K(n) = \mathbf{x}_K(n)) = 0. \tag{5.27}$$

Proof The rank of $\tilde{\mathbf{\Phi}}$ is $\text{rank}(\mathbf{\Phi}) = E$ with overwhelming probability [Fen07], and thus the sparsity on the reconstruction of $\mathbf{x}_K(n)$ cannot be higher than an E-sparse signal (instead of K-sparse) [Rac08].

Lemma 5.3 *If $E = K$, the eavesdroppers will recover the original vector $\mathbf{x}_K(n)$ with non-zero probability. However, it is asymptotically zero in S.*

Proof Since for the case $E = K$, the vector $\mathbf{x}_K(n)$ is in the solution set $\mathcal{X}$ (with cardinality M), there is a probability of selecting it among all possible solutions. Without loss of generality, we assume that the output of the $\mathcal{P}0_e$ is uniformly distributed among $\mathcal{X}$. Hence,

$$P(\mathbf{x}_K(n) = \mathbf{x}) = \frac{1}{M}, \tag{5.28}$$

where M can be computed in the same way that in (5.21). Thus,

$$P(\mathbf{x}_K(n) = \mathbf{x}) = \frac{(S-K)!K!}{S!}. \tag{5.29}$$

In the asymptotic regime we have

$$\lim_{S\to\infty} \frac{K!(S-K)!}{S!} = 0. \tag{5.30}$$

This asymptotic result make sense in real scenarios since $S \gg K$ is typically assumed in CS schemes. In Fig 5.5, the probability of recovery is plotted for different values of K. This probability decreases quickly as S increases. Actually, even for small ratios of S/K the probability of recovery is almost negligible.

For moderate values of E, perfect secrecy cannot be guaranteed for the case $E > K$ with perfect CSI.

Theorem 5.3 *For a given* $\mathbf{y}(n)$ *and* $\mathbf{\Phi}$*, let* $\mathcal{X}$ *denote the set of vectors* $\mathbf{x}'$ *that are solution of*

$$\mathbf{y}(n) = \mathbf{\Phi}\mathbf{x}'. \tag{5.31}$$

Therefore, if $E > K$*, the mutual information between the recovered vector* $\hat{\mathbf{x}}$ *and the original vector* $\mathbf{x}_K(n)$ *is not zero, which means that perfect secrecy is not achieved.*

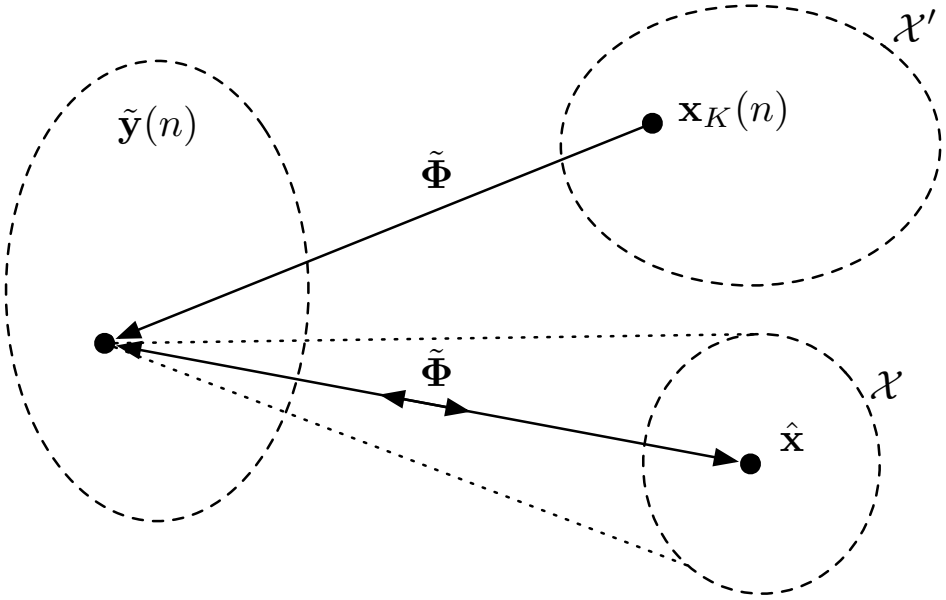

Figure 5.4: Graphical representation of the PHY-layer secrecy for the $E < K$ case. Each circle represents the set of the possible values of each vector. The matrix $\tilde{\mathbf{\Phi}}$ maps any $\mathbf{x}_K(n)$ onto a point in the corresponding set of $\tilde{\mathbf{y}}(n)$. Then, the decoder maps back $\tilde{\mathbf{y}}(n)$ onto a set of the possible values of $\hat{\mathbf{x}}_K(n)$, which is a disjoint set with the one that corresponds to $\mathbf{x}_K(n)$. Finally, the decoder only have the information to project the solution from the set corresponding to $\hat{\mathbf{x}}_K(n)$ to the one of $\tilde{\mathbf{y}}(n)$ and viceversa. Since they are disjointed, the recovery of $\mathbf{x}_K(n)$ is impossible.

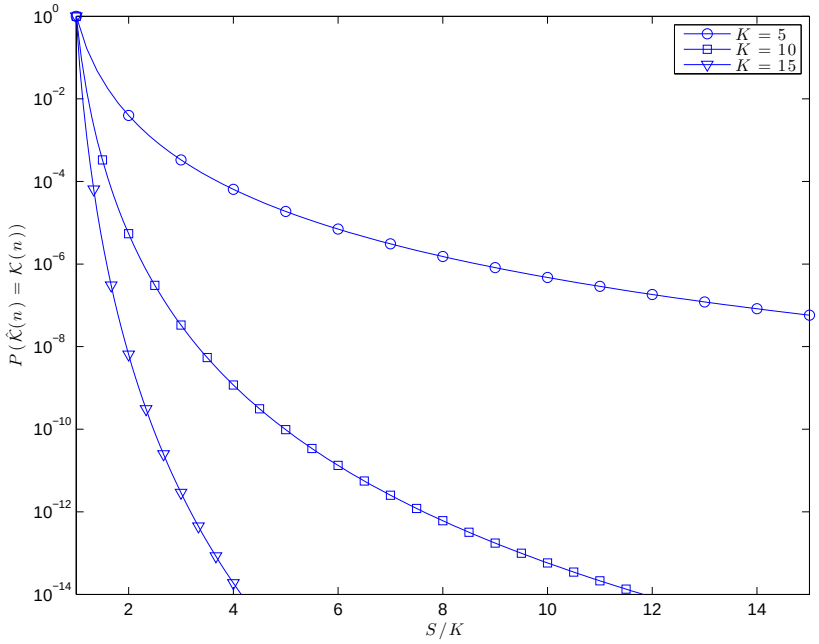

Figure 5.5: Probability of recovery for the case $E = K$ as a function of S.

Proof The proof follows the same approach than Theorem 5.1. Thus,

$$\begin{aligned} I(\hat{\mathbf{x}};\mathbf{x}') &= \mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}') \\ &= \mathcal{H}(\hat{\mathbf{x}}) - \sum_{\mathbf{x}\in\mathcal{X}} \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x})P(\mathbf{x}'=\mathbf{x}) \\ &= \mathcal{H}(\hat{\mathbf{x}}) - \sum_{\mathbf{x}\in\mathcal{X},\mathbf{x}\neq\mathbf{x}_K(n)} \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x})P(\mathbf{x}'=\mathbf{x}) \\ &\quad -\mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x}_K(n))P(\mathbf{x}'=\mathbf{x}_K(n)) \\ &\overset{(i)}{=} \mathcal{H}(\hat{\mathbf{x}}) - \sum_{\mathbf{x}\in\mathcal{X},\mathbf{x}\neq\mathbf{x}_K(n)} \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x})P(\mathbf{x}'=\mathbf{x}) \\ &\overset{(ii)}{\geq} \mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\hat{\mathbf{x}}) \sum_{\mathbf{x}\in\mathcal{X},\mathbf{x}\neq\mathbf{x}_K(n)} P(\mathbf{x}'=\mathbf{x}) > 0. \end{aligned} \tag{5.32}$$

where:

- (i) is because $\mathbf{x}_K(n)$ is the unique K-sparse solution of $\mathcal{P}0_e$. Then, $\mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x}_K(n))=0$ because if the decoder knows $\mathbf{x}_K(n)$, the solution set is reduced to a set of cardinality one.
- (ii) is because $\mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}'=\mathbf{x}) \leq \mathcal{H}(\hat{\mathbf{x}})$, and $\sum_{\mathbf{x}\in\mathcal{X},\mathbf{x}\neq\mathbf{x}_K(n)} P(\mathbf{x}'=\mathbf{x}) < 1$ since $\mathbf{x}_K(n)$ is in the set $\mathcal{X}$.

However, the eavesdroppers are only able to successfully decode the signal with low probability. This probability increases with E. Even for the cases that the vector $\mathbf{x}_K(n)$ is successfully decoded, it only contains a small amount of the sensor readings since the condition $S \gg K$ holds. The remaining Q measurements have to be estimated by a LWF predictor at the eavesdropper's side. Therefore, the observation vector used in the LWF for the eavesdroppers will contain, with high probability, erroneous samples that come from the bad decoding of previous samples and degenerates the estimation of the remaining Q samples of $\mathbf{x}(n)$.

For high values of E, if the condition $\mathcal{C}_{\text{CS}}$ is satisfied, the eavesdropper set $\mathcal{E}$ can decode the signal with high probability. Moreover, the eavesdroppers obtain an accurate version of the observation vector for the LWF prediction. In this situation, the eavesdroppers obtain a similar performance as the fusion center.

5.4.1.2 Eavesdroppers with corrupted CSI

Here, we assume that the eavesdroppers have imperfect knowledge of the wiretap sensing matrix $\tilde{\mathbf{\Phi}}$. Let $\hat{\mathbf{\Phi}}$ denote corrupted wiretap sensing matrix, which is modeled as

$$\hat{\mathbf{\Phi}} = \tilde{\mathbf{\Phi}} + \mathbf{\Sigma}, \tag{5.33}$$

where $\mathbf{\Sigma} \in \mathbb{R}^{E \times S}$ is a random matrix with i.i.d. Gaussian entries with zero mean and variance $\sigma^2_{\mathbf{\Sigma}}$ that models the errors in the channel estimation. This perturbation in the sensing matrix results in a *multiplicative noise*, which is more difficult to analyze than the additive noise (as in Chapter 4) since it is correlated with the signal of interest [Her10].

Then, the coordinated eavesdroppers would have to jointly solve the following problem:

$$\begin{aligned} \mathcal{P}'1_{\text{e}}: \quad & \underset{\hat{\mathbf{x}}_K(n) \in \mathbb{R}^S}{\text{minimize}} \quad && \|\hat{\mathbf{x}}_K(n)\|_{l^1} \\ & \text{subject to} && \|\hat{\mathbf{\Phi}}\hat{\mathbf{x}}_K(n)\|^2 < \varepsilon. \end{aligned} \tag{5.34}$$

for some $\varepsilon > 0$.

Some recent results in the literature study similar problems. The work in [Yan11] analyzes the effect of a structured perturbation in the sensing matrix. In particular, the model under study is

$$\hat{\mathbf{\Phi}} = \mathbf{\Phi} + \mathbf{B}\mathbf{\Delta} \tag{5.35}$$

where $\mathbf{B} \in \mathbb{R}^{R \times S}$ is known a priori and $\mathbf{\Delta}$ is a diagonal matrix of uniformly distributed and bounded entries. The authors in [Her10] deals with more general perturbations in the sensing matrix and follows the model in (5.33). Their studies are focused on small perturbations, understanding for small perturbations when $\|\mathbf{\Sigma}\|/\|\tilde{\mathbf{\Phi}}\| < 1$. Under such condition, they show that an upperbound for the error at the receiver grows linearly with the perturbation level.

For low values of E, the same results as in the perfect CSI case also hold, which means perfect secrecy.

For moderate and high values of E, the condition C_{CS} is not valid anymore and we cannot use it as an orientative bound to decide if the signal can be decoded with low or high probability. Now, it depends also on the power of the perturbation introduced in the estimation process. In Section 5.5, we introduce a new technique in order to control the amount of distortion in the estimation of the wiretap channel matrix by the eavesdroppers. Here, a convenient metric to be studied in the results section is how the channel perturbation affects to the relative wiretap distortion at the eavesdroppers.

5.4.2 Eavesdropping during the projection phase

This phase is very robust against malicious and passive eavesdropping. Here, the derived results are not dependent on the number of eavesdroppers, since each eavesdropper in $\mathcal{E}$ has full access to the signal sent by the relays, i.e., $\mathbf{y}(n)$, to the fusion center (assuming that $\mathbf{y}(n)$ is not encrypted) as it is represented in Fig. 5.3.

This problem is similar to the key-based PHY-layer secrecy works in [Rac08] and [May10]. However, the main difference is that the sensing nodes do not encrypt the signal with any key. Instead, they send uncoded signals and the MAC implicitly performs random linear combinations modeled according to the matrices $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}}$.

Actually, this key-less coding mechanism is not new and comes from

the well-known discipline of Network Coding [Yeu05], where the signals from different sources are not handled individually and algebraic operations among them are allowed instead. So, sending linear combinations of the signals offers a natural way of protection [Fra07].

For perfect secrecy, it is enough to proof that if the eavesdroppers try to reconstruct the signal with a wrong sensing matrix $\mathbf{\Phi}'$ (understanding 'wrong' as independent to $\mathbf{\Phi}$), the eavesdroppers will recover a R-sparse vector, instead of the K-sparse original one.

Lemma 5.4 *Let $\mathbf{\Phi}$ and $\mathbf{\Phi}'$ be two $R \times S$ independent matrices following the Gaussian measurement ensemble. For a K-sparse vector $\mathbf{x}_K(n)$, let $\mathbf{y}(n) = \mathbf{\Phi}\mathbf{x}_K(n)$. Then, all $\hat{\mathbf{x}}_K(n)$ that satisfy $\mathbf{y}(n) = \mathbf{\Phi}\hat{\mathbf{x}}_K(n)$ are E-sparse with probability one.*

Proof The proof is the same than the one of Theorem 1 in [Rac08].

Remark 5.1 *The main difference with [Rac08] is that they obtained computational secrecy since they assume a finite set of key matrices. Hence, an eavesdropper with unlimited computational complexity may try among all the possibilities until the recovery was K-sparse. On the contrary, we can ensure perfect secrecy because there are infinite number of i.i.d. possible matrices. Hence, the eavesdroppers have zero probability to guess the correct one if no further information is provided.*

5.5 Channel estimation based on random pilots

In this section, we evaluate the assumption about the imperfect CSI at the eavesdroppers. The corrupted estimation of the wiretap sensing matrix decreases the eavesdropping capabilities during the sensing phase. Thus, we propose a novel technique to support this assumption.

Although one may think that even for the case of perfect channel estimation, the required E may become unpractical for relatively high values of

K, one can design the system in order to protect even more the information against passive eavesdropping.

We propose a novel technique that allows the sensing nodes to corrupt the channel estimation of the eavesdroppers without decreasing their own estimation. In particular we propose a training phase where each sensing node sends N random pilots with amplitude $A + s(n)$, where A is a known and constant value and $s(n)$ is a (pseudo)random sequence distributed as $s(n) \sim N(0, \sigma_s^2)$, which has been previously agreed. The pilot signal from sth sensing node at the rth relay is:

$$p_{r,s}(n) = (A + s(n))[\mathbf{\Phi}]_{r,s} + [\mathbf{w}(n)]_r. \tag{5.36}$$

On the other hand, the eth eavesdropper will receive the pilot signal from the sth sensing node as

$$p_{e,s}(n) = (A + s(n))[\tilde{\mathbf{\Phi}}]_{e,s} + [\tilde{\mathbf{w}}(n)]_e. \tag{5.37}$$

We assume the general scheme that the eavesdroppers do not have complete information of the pilot amplitudes. Instead, they only know partial information of the pilot sequence. The part that they know is A, while it is assumed that the eavesdroppers do not have access to $s(n)$.

5.5.1 Performance of the random pilots technique

The performance of the random pilots technique is summarized in the following two lemmas.

Lemma 5.5 *Let $[\hat{\mathbf{\Phi}}]_{r,s}$ be the estimate of the channel coefficient between sth sensing node and rth relay. Then, the fusion center can achieve $\mathbb{E}[\left|[\mathbf{\Phi}]_{r,s} - [\hat{\mathbf{\Phi}}]_{r,s}\right|^2 < \varepsilon$ for an arbitrary small $\varepsilon > 0$.*

Proof For notation within this lemma, let the N dimensional vector $\mathbf{p}$ represent the collected N pilots samples $p_{r,s}(n)$, ϕ represent $[\mathbf{\Phi}]_{r,s}$, $\hat{\phi}$ represent

$[\hat{\mathbf{\Phi}}]_{r,s}$ and $w(n)$ is $[\mathbf{w}(n)]_r$. To actually find the Maximum Likelihood Estimation (MLE) of ϕ, we first write the *pdf* of $\mathbf{p}$ as a function of ϕ as [Kay93]

$$f(\mathbf{p};\phi) = \frac{1}{(2\pi\sigma_{\mathbf{w}}^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma_{\mathbf{w}}^2}\sum_{n=1}^{N}([\mathbf{p}]_n - (A+s(n))\phi)^2\right]. \quad (5.38)$$

The log-likelihood function of ϕ becomes as

$$\ln f(\mathbf{p};\phi) = -\frac{N}{2}(2\pi\sigma_{\mathbf{w}}^2) - \frac{1}{2\sigma_{\mathbf{w}}^2}\sum_{n=1}^{N}([\mathbf{p}]_n - (A+s(n))\phi)^2. \quad (5.39)$$

After some simple algebra and taking its derivative produces

$$\frac{\partial \ln f(\mathbf{p};\phi)}{\partial\phi} = \frac{1}{\sigma_{\mathbf{w}}^2}\sum_{n=1}^{N}[\mathbf{p}]_n(A+s(n)) - \frac{\phi}{\sigma_{\mathbf{w}}^2}\sum_{n=1}^{N}(A+s(n))^2, \quad (5.40)$$

and setting it equal to zero and solving for $\hat{\phi}$ we obtain the MLE

$$\hat{\phi} = \frac{\sum_{n=1}^{N}[\mathbf{p}]_n(A+s(n))}{\sum_{n=1}^{N}(A+s(n))^2}. \quad (5.41)$$

The MSE of the MLE can be computed as

$$\mathbb{E}\left[(\phi-\hat{\phi})^2\right] = \mathbb{E}\left[\left(\phi - \frac{\sum_{n=1}^{N}[\mathbf{p}]_n(A+s(n))}{\sum_{n=1}^{N}(A+s(n))^2}\right)^2\right] \quad (5.42)$$

If $[\mathbf{p}]_n$ is replaced by $(A+s(n))\phi + w(n)$, we obtain

$$\mathbb{E}\left[(\phi-\hat{\phi})^2\right] = \mathbb{E}\left[\left(\frac{\sum_{n=1}^{N}w(n)(A+s(n))}{\sum_{n=1}^{N}(A+s(n))^2}\right)^2\right], \quad (5.43)$$

Since the term $\sum_{n=1}^{N}(A+s(n))^2$ is known by the intended receivers, we can replace it by its equivalent mean power $N(A^2+\sigma_s^2)$. Hence

$$\mathbb{E}\left[(\phi-\hat{\phi})^2\right] = \frac{1}{N^2(A^2+\sigma_s^2)^2}\mathbb{E}\left[\left(\sum_{n=1}^{N}w(n)(A+s(n))\right)^2\right], \quad (5.44)$$

and since $w(n)$ is independent of $A + s(n)$ we can write

$$\begin{aligned}\mathbb{E}\left[(\phi - \hat{\phi})^2\right] &= \frac{1}{N^2(A^2+\sigma_s^2)^2}\mathbb{E}\left[\sum_{n=1}^{N} w(n)^2(A+s(n))^2\right], \\ &= \frac{\sigma_{\mathbf{w}}^2}{N(A^2+\sigma_s^2)}. \end{aligned} \tag{5.45}$$

So the system may decrease the estimation error as much as desired by increasing the power of the pilots $A^2 + \sigma_s^2$ and/or the number of pilots N (with the subsequent energy and signaling costs). Therefore, we assume perfect channel state information at the relays.

Similarly, the eavesdroppers can estimate the channel coefficient $[\tilde{\mathbf{\Phi}}]_{e,s}$ with the difference that we assume $s(n)$ only known by the intended nodes but not by the eavesdropping nodes.

Lemma 5.6 *Let $[\hat{\tilde{\mathbf{\Phi}}}]_{\mathbf{e,s}}$ be the estimate of the channel coefficient between sth sensing node and eth eavesdropper. Then, the eavesdropper can achieve a mean square error $\mathbb{E}[\left|[\mathbf{\Phi}]_{r,s} - [\hat{\mathbf{\Phi}}]_{r,s}\right|^2 > \varepsilon$ where $\varepsilon > 0$ is a nonzero error floor. Furthermore, ε is a function of σ_s^2.*

Proof For notation within this lemma, let the N dimensional vector $\mathbf{p}$ represent the collected N pilots samples $p_{e,s}(n)$, ϕ represent $[\hat{\mathbf{\Phi}}]_{e,s}$, $\hat{\phi}$ represent $[\hat{\tilde{\mathbf{\Phi}}}]_{e,s}$ and $w(n)$ is $[\mathbf{w}(n)]_e$.

Since $s(n)$ is unknown and treated as multiplicative noise by the eavesdroppers, the result signal model at the eth eavesdropper is:

$$[\mathbf{p}]_n = (A + s(n))\phi + w(n) = A\phi + s(n)\phi + w(n), \tag{5.46}$$

where the term $A\phi$ can be seen as the desired signal and $s(n)\phi + w(n)$ as the noise term with variance $\sigma_t^2 = \phi^2\sigma_s^2 + \sigma_w^2$, or for the general case (and with some abuse of notation) for any entry of the wiretap matrix $\tilde{\mathbf{\Phi}}$, $\sigma_t^2 = \sigma_{\mathbf{\Phi}}^2\sigma_s^2 + \sigma_w^2$.

To actually find the Maximum Likelihood Estimation (MLE) of ϕ, we first write the *pdf* of $\mathbf{p}$ as a function of ϕ as

$$f(\mathbf{p};\phi) = \frac{1}{(2\pi\sigma_t^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma_t^2}\sum_{n=1}^{N}([\mathbf{p}]_n - A\phi)^2\right]. \tag{5.47}$$

The log-likelihood function of ϕ becomes as

$$\begin{aligned} \ln f(\mathbf{p};\phi) &= -\frac{N}{2}(2\pi\sigma_t^2) - \frac{1}{2\sigma_t^2}\sum_{n=1}^{N}([\mathbf{p}]_n - A\phi)^2, \\ &= -\frac{N}{2}(2\pi\sigma_t^2) - \frac{1}{2\sigma_t^2}\left(\sum_{n=1}^{N}[\mathbf{p}]_n^2 - 2A\phi\sum_{n=1}^{N}[\mathbf{p}]_n + NA^2\phi^2\right) \end{aligned} \tag{5.48}$$

Taking its derivative we get

$$\frac{\partial \ln f(\mathbf{p};\phi)}{\partial\phi} = -\frac{1}{2\sigma_t^2}\left(-2A\sum_{n=1}^{N}[\mathbf{p}]_n + 2NA^2\phi\right), \tag{5.49}$$

and setting it equal to zero and solving for $\hat{\phi}$ we obtain the MLE

$$\hat{\phi} = \frac{\sum_{n=1}^{N}[\mathbf{p}]_n}{NA}. \tag{5.50}$$

The MSE of the MLE can be computed as

$$\mathbb{E}\left[(\phi - \hat{\phi})^2\right] = \mathbb{E}\left[\left(\phi - \frac{\sum_{n=1}^{N}[\mathbf{p}]_n}{NA}\right)^2\right] \tag{5.51}$$

Replacing $[\mathbf{p}]_n$ by $(A+s(n))\phi + w(n)$, we obtain

$$\begin{aligned}
\mathbb{E}\left[(\phi-\hat{\phi})^2\right] &= \mathbb{E}\left[\left(-\frac{\sum_{n=1}^{N}(A+s(n))\phi+w(n)}{NA}\right)^2\right], \\
&= \mathbb{E}\left[\left(\frac{\sum_{n=1}^{N}s(n)\phi+w(n)}{NA}\right)^2\right], \\
&= \frac{1}{N^2A^2}\mathbb{E}\left[\left(\sum_{n=1}^{N}s(n)\phi+w(n)\right)^2\right] \\
&= \frac{N\sigma_t^2}{N^2A^2} = \frac{\sigma_s^2\sigma_{\mathbf{\Phi}}^2+\sigma_{\mathbf{w}}^2}{NA^2}. \qquad (5.52)
\end{aligned}$$

Clearly, the introduction of the pseudorandom sequence $s(n)$ achieves a double improvement. On the one hand, it reduces the channel estimation error at the intended nodes, which is an expected consequence since the system is spending more power in pilots. On the other hand, it introduces additional error in the estimation of the eavesdroppers.

We evaluate the performance of the proposed technique in the following two experiments.

Experiment 5.1 *We have simulated the channel estimation for* $N = [1, 150]$*,* $A = 5$*,* $\sigma_{\mathbf{w}}^2 = 0.1$*, with target values* $[\mathbf{\Phi}]_{r,s} = [\tilde{\mathbf{\Phi}}]_{e,s} = 1$*. In our experiment, we have generated* $s(n)$ *as a Gaussian sequence of zero mean and variance* $\sigma_s^2 = 10$ *(for a ratio* $\sigma_s^2/\sigma_{\mathbf{\Phi}}^2 = 10$*).*

From Experiment 5.1, we have plotted the results of channel estimation for both the MLE of the intended receiver in (5.41) and the MLE of the eavesdropper in (5.50). We can graphically see the difference in terms of performance of both estimators even for high values of N (in favor of (5.41)). As the intended receiver achieves good accuracy for values of N close to 20 (or even lower), the eavesdroppers require much higher values of N to achieve similar accuracies.

Experiment 5.2 *The setup is the same as in Experiment 5.1 with the difference that in this case, we set* $N = 50$ *and we evaluate the performance in terms of quadratic error as a function of the ratio* $\sigma_s^2/\sigma_{\mathbf{\Phi}}^2$. *We also compare the results with the analytical expressions of (5.42) and (5.51).*

The results of Experiment 5.2 have been plotted in Fig. 5.7. On the one hand we observe that the fitting of the analytical MSE estimator (dashed line) with the experimental squared error (solid line) is sharp.

Furthermore, we can also observe that if we increase the power of the sequence $s(n)$, we obtain the announced double improvement, that is, increasing the MSE at the eavesdroppers while the MSE at the relays is decreased.

5.5.2 Secrecy of the random pilots sequence

In this subsection we clarify the assumption that the sequence $s(n)$ is only known by the intended nodes and not by the eavesdroppers.

In principle, this assumption may seem to be hard in the sense that if we assume that the sequence is used in a secure scenario, it has to be previously sent using any other technique. In other words, it looks like a circular argument where in order to send information securely, it is assumed some prior information (in this case $s(n)$) that has been already sent securely.

However, the sensing nodes and the fusion center can use "off air" techniques in order to agree with a given sequence $s(n)$. For example, each sensor can locally generate a sequence $s(n)$ (which is different for each sensing node) in an efficient manner by using the same seed of a pseudorandom generator. This seed may be a function of the network identifier (or any other parameter related to the management of the network or to the hardware of the device) that is unknown by any alien node.

Hence, since there is no transmission over the communication channel, i.e. the airwave, the pseudorandom sequence is protected against eavesdropping.

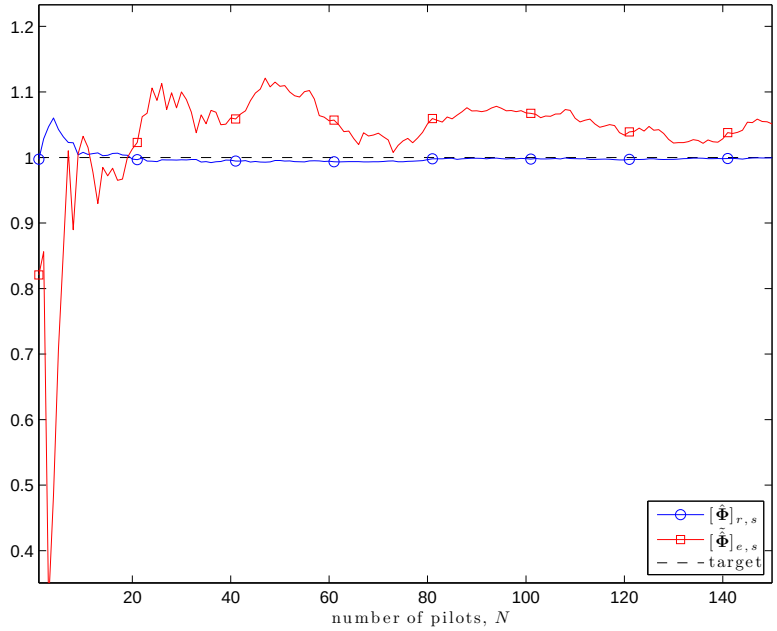

Figure 5.6: Channel estimation for the intended nodes (with marker ◯) and the eavesdropping nodes (with marker □) as a function of the number of pilots N. The actual value of the channel is equal in both cases and is plotted in dashed line.

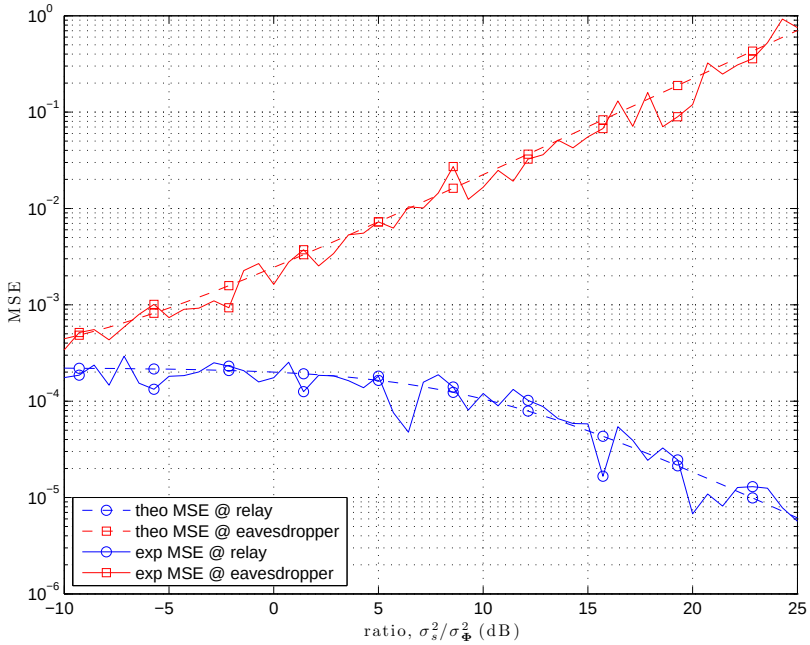

Figure 5.7: Comparison of the experimental results of the MSE for both MLEs in (5.41) and (5.50). They are also compared with the analytical expressions in (5.42) and (5.51) respectively. The values for the experimental results have been averaged 20 times.

This technique is not new and it has been used for different purposes in the literature, e.g., in [Baj06] it is used to generate the same sensing matrix in a distributed way for the transmitters and the receiver, and in [Rac08], it is used to create a random key which is know by the transmitter and the receiver but not by any intruder.

5.6 Numerical Results

In this section, we first summarize the theoretical results obtained in Section 5.4. Then, we evaluate the PHY-layer secrecy performance of the AF-CS throughout simulation.

Table 5.1 summarizes the parameters that we consider in our simulations.

Table 5.1: Simulation Parameters

Parameter	**Value**
Number of *fusion nodes*:	$F = 1$
Number of *sensing nodes*:	$S = 200$
Number of *active sensors*:	$K = 10$
Number of *relay nodes*:	$R = 60$
Number of *eavesdropping nodes*:	$E = [0, 110]$
Compressed Sensing Condition:	$\mathcal{C}_{\text{CS}} : E > 40$

We also define the following figures of merit.

- *Channel estimation distortion*, $\mathcal{D}$. This metric measures the ratio between the power of the estimation degradation and the variance of the channel coefficients, namely,

$$\mathcal{D} = 10 \log \left(\frac{\sigma^2_{\mathbf{\Sigma}}}{\sigma^2_{\mathbf{\Phi}}} \right) = 10 \log \left(\frac{\sigma^2_{\mathbf{\Phi}} \sigma^2_s + \sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{\Phi}} N A^2} \right). \tag{5.53}$$

- *Packet Error Rate* (PER). It measures the reconstruction failure rate of $\mathbf{x}_K(n)$, i.e., PER $= P(\hat{\mathbf{x}}_K(n) \neq \mathbf{x}_K(n))$. For practical reasons, we consider that two vectors are different when $\|\mathbf{x}_K(n) - \hat{\mathbf{x}}_K(n)\|^2 / \|\mathbf{x}_K(n)\|^2 > 0.01$.

- *Probability of detection*, PoD $= P(\mathcal{K}(n) \subseteq \hat{\mathcal{K}}(n))$, where $\hat{\mathcal{K}}(n)$ is the estimation of the active sensors set of $\mathbf{x}_K(n)$. In other words, it measures the probability that $\mathcal{E}$ succeeds in detecting the active nodes.

- *Probability of False Alarm*, PFA $= P(\mathcal{Q}(n) \nsubseteq \hat{\mathcal{Q}}(n))$, where $\hat{\mathcal{Q}}(n)$ is the estimation of the silent sensors subset. It measures the probability that $\mathcal{E}$ fails in detecting the silent sensors.

5.6.1 Summary of the Theoretical Results

Table 5.2: Summary of Theoretical Results

		perfect CSI at $\mathcal{E}$	**corrupted CSI at $\mathcal{E}$**
Sensing Phase	$E \leq K$	Perfect secrecy	Perfect secrecy
	$K < E \leq \mathcal{C}_{\text{CS}}$	Low PoR	Low PoR
	$E > \mathcal{C}_{\text{CS}}$	High PoR	Low PoR
Projection Phase		Perfect secrecy	

Table 5.2 summarizes the PHY-secrecy performance for each of the possible cases. We have divided the analysis of the sensing phase in four cases depending on the number of eavesdroppers.

For low values of E (i.e., $E \leq K$), perfect secrecy can be guaranteed even in the case that the eavesdroppers have perfect channel estimation and zero probability of recovery. For the particular case $E = K$, the probability of recovery is not zero but asymptotically zero in S.

For moderate values of E (i.e., $K < E \leq \mathcal{C}_{\text{CS}}$), the eavesdroppers cannot recover $\mathbf{x}_K(n)$ with high probability getting high wiretap distortion. However, perfect secrecy cannot be guaranteed for that configuration.

Only for high values of E (i.e., $E > \mathcal{C}_{\text{CS}}$) and with perfect CSI, the eavesdroppers can decode the signal $\mathbf{x}_K(n)$ with high probability. However, thanks to the introduction of the random pilots technique, the intended nodes can corrupt the CSI of the eavesdroppers and make the wiretap distortion grow linearly with the introduced noise power.

The projection phase achieves perfect secrecy in any case.

5.6.2 Probability of detection as a function of the number of eavesdroppers

First, we consider the simple detection scenario case. In such a scenario, a malicious eavesdropping set may be interested in detecting only the supporting set of $\mathbf{x}_K(n)$, or in other words, which subgroup $\mathcal{K}(n) \subseteq \mathcal{S}$ is in active mode in each timeslot. Hence, in this case the actual value of the message is not essential.

The numerical simulation has been run in Matlab as follows. For each realization, a new wiretap sensing matrix $\tilde{\mathbf{\Phi}}$ of dimension $E \times S$ has been randomly generated following a Gaussian measurement ensemble $\mathcal{N}(0, E^{-1})$. Next, for each channel distortion value $\mathcal{D}$, a perturbation matrix $\mathbf{\Sigma}$ has been generated with entries according to $\mathcal{N}(0, \sigma^2_{\mathbf{\Sigma}})$. The K non-zero entries of a random vector $\mathbf{x}_K(n)$ of sparsity K are distributed as $\mathcal{N}(0, \sigma^2_{\mathbf{x}})$ and uniformly located across the S possible positions. Finally, the decoder $\mathcal{P}'1_e$ in (5.34) has been implemented using `CVX`, a package for specifying and solving convex programs [Gra11, Gra08].

In Fig. 5.8 we plot the Probability of Detection (PoD) in terms of the number of eavesdroppers E, or what is the same, the probability that the estimated support set, $\hat{\mathcal{K}}(n)$ contains actually all $\mathcal{K}(n)$ nodes.

For perfect CSI at the eavesdroppers, the simulation supports that for

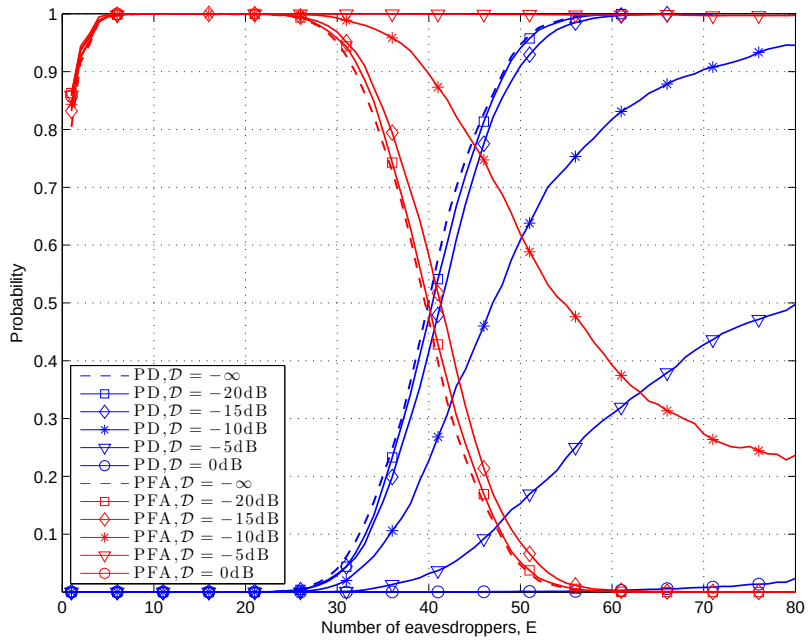

Figure 5.8: Probability of detection and probability of false alarm as a function of the number of eavesdroppers and for different values of channel estimation distortion. This figure has been averaged over 1000 realizations.

small $E < K$, the recovery is infeasible, getting a PoD of 0. Moreover, for moderate values of E, i.e., $K < E < \mathcal{C}_{\text{CS}}$ the support set $\mathcal{K}(n)$ is recovered with low probability. On the other hand, for values of E similar or greater than $\mathcal{C}_{\text{CS}}$, the eavesdropping set can recover $\mathcal{K}(n)$ with high probability following the $\mathcal{C}_{\text{CS}}$ condition. According to Fig. 5.8, we observe that the bound $\mathcal{C}_{\text{CS}}$ (i.e., $E = 40$) divides the low and high PoD for values smaller and bigger than 0.5, respectively.

For corrupted CSI at the eavesdroppers, the simulation shows how the PoD is degraded. Even for small values of $\mathcal{D}$, e.g. $\mathcal{D} = -10$dB, the PoD degenerates drastically and PoD close to 1 can only be achieved for very large values of E ($E > 80$ nodes). For values of $\mathcal{D} = 0$dB (which means that the introduced perturbation is of the order of channel variance), the supporting set $\mathcal{K}(n)$ can be recovered with negligible probability.

The probability of detection is not enough to quantify the detection performance. This is because one can set the vector $\hat{\mathbf{x}}_K(n)$ with all the entries loaded and therefore $\mathcal{K}(n) \subseteq \hat{\mathcal{K}}(n)$ with probability one. That is why the Probability of False Alarm (PFA) is also plotted in Fig. 5.8. Here, we observe that the reconstruction phase is even less accurate since the reconstructed signal is not purely K-sparse, and instead, many undesired spikes appear in other positions outside the support set $\mathcal{K}(n)$.

5.6.3 Probability of detection compared to CWS-like techniques

Following the same detection scenario case, we compare the AF-CS with CWS-like methods [Baj06]. Both methods are CS-based distributed schemes. Although CWS has not been designed from a PHY-layer secure perspective, we assess its secrecy performance since its approach is one of the most extended CS schemes in WSN literature (as it is detailed in Chapter 4).

Simulation results show that a single eavesdropper with a channel dis-

tortion of less that -15dB suffices in decoding the transmitted signal with high probability. Furthermore, it can be seen in Fig. 5.9 that for the 110 eavesdroppers configuration, the AF-CS achieves the same performance as the CWS with only 2 eavesdroppers. The reason of such a big difference is the natural protection that gives the spacial diversity introduced by the relays. Therefore, in CWS a single eavesdropper can capture the entire signal that it is sent by the K sensing nodes (or R for the case of Generalized CWS, also detailed in Chapter 4).

Furthermore, Fig. 5.9 also shows the PoD results as a function of the channel estimation distortion $\mathcal{D}$. For very small distortion values (i.e, $\mathcal{D} < -20$dB) the performance drop is negligible. However, it degrades fast for values of $\mathcal{D} > -15$dB. For the case of $\mathcal{D} = 0$dB, the probability of detection is negligible for any configuration.

5.6.4 Packet error rate as a function of the fumber of eavesdroppers

In this subsection, our approach is slightly different than in the previous cases because now we do focus on the actual values of the vector $\mathbf{x}_K(n)$. Hence, we take as a figure of merit the Packet Error Rate (PER). From a theoretical point of view, the PER is defined as the success ratio that the eavesdroppers recover exactly $\mathbf{x}_K(n)$ from $\tilde{\mathbf{y}}(n)$. However, the case where $\mathbf{x}_K(n) = \hat{\mathbf{x}}_K(n)$ can only be achieved for the noiseless case. From a practical point of view, we set a threshold in order to determine wether the vector $\hat{\mathbf{x}}_K(n)$ is an acceptable reconstruction of $\mathbf{x}_K(n)$ or not, that is if the relative wiretap distortion is

$$\mathcal{D}_e = \mathbb{E}\left[\frac{\|\mathbf{x}_K(n) - \hat{\mathbf{x}}_K(n)\|^2}{\|\mathbf{x}_K(n)\|^2}\right] < 0.1, \tag{5.54}$$

we consider that the eavesdroppers succeed in the recovery.

Although it is a different approach, we observe in Fig. 5.10 a similar behavior than in the PoD analysis, where for the ideal case $\mathcal{D} = -\infty$dB.

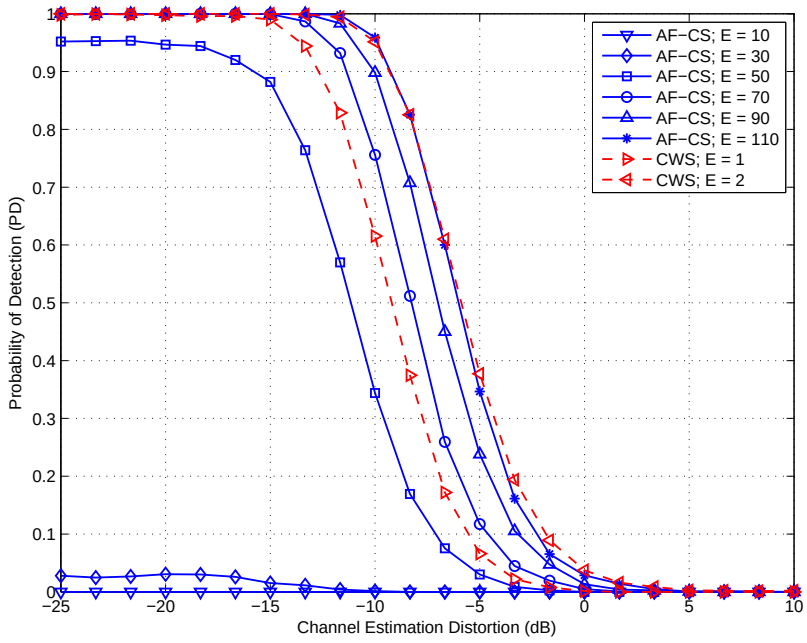

Figure 5.9: Probability of recovery as a function of the channel estimation distortion for different number of coordinated eavesdroppers for $K = 10$ and $S = 200$. Solid lines represent the performance of AF-CS while dashed lines denote CWS. This figure has been averaged over 1000 realizations.

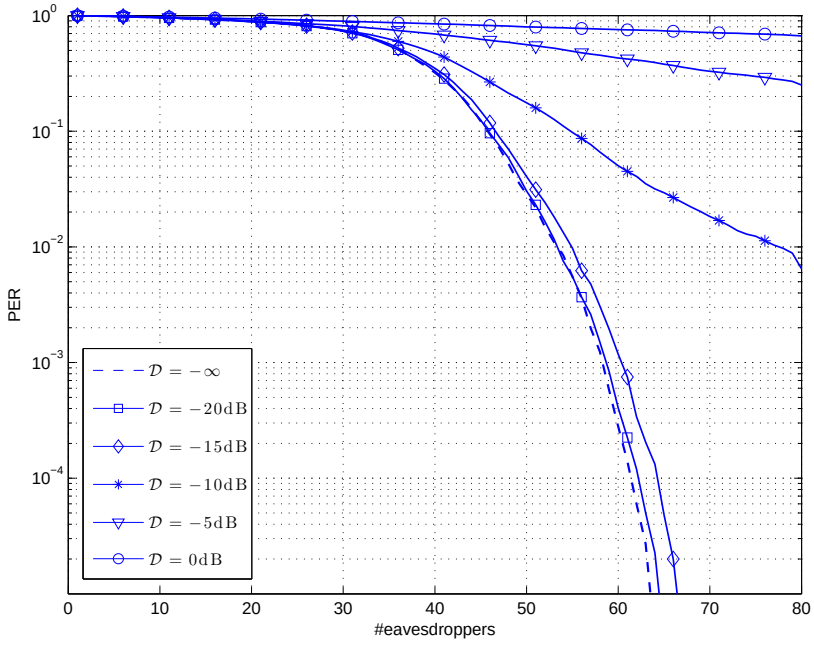

Figure 5.10: Packet Error Rate in the eavesdroppers reconstruction as a function of the number of nodes and for different values of channel estimation distortion. This figure has been averaged over 1000 realizations.

In other works, the eavesdroppers achieve good PER for values of E bigger than $\mathcal{C}_{\text{CS}}$, whereas for low values of E the PER is zero.

5.6.5 Relative wiretap distortion as a function of the estimation channel distortion

Here we study the performance of the relative wiretap distortion as a function of the channel estimation distortion.

This study actually extends the one in [Her10] and confirm some of their results. Mainly, we show (as in [Her10]) that the distortion at the receiver grows linearly with the power of the channel estimation distortion for values $\mathcal{D} < 0$dB. This is true (up to some error floor) not only for the case $\mathcal{D} < 0$dB but also when $\mathcal{D} > 0$dB, as we can graphically see in Fig. 5.11.

We also show that the relative wiretap distortion decreases for lower values of $\mathcal{D}$ up to some error floor, which depends on the number of eavesdroppers. It means that even for the ideal case of $\mathcal{D} = -\infty$dB, the relative wiretap distortion cannot be decreased further than the error floor.

However, probably the most relevant result of this subsection is the following: for values of $\mathcal{D} = 0$dB all the configurations achieve a similar relative wiretap distortion of 1. It means that the distortion of the reconstruction phase performed at the eavesdroppers is equal to the actual variance of the signal. That is to say, the eavesdroppers do not know anything about the signal $\mathbf{x}_K(n)$ as it can be appreciated in the following example.

Example 5.1 *A given decoder that does not receive $y(n)$ and does not have any further information about $\mathbf{x}_K(n)$ than their entries are zero mean can guess a decoder vector with a relative wiretap distortion of 1 by setting $\hat{\mathbf{x}}_K(n) = \mathbf{0}$. One can easily check it as*

$$\mathcal{D}_e = \mathbb{E}\left[\frac{\|\mathbf{x}_K(n) - \mathbf{0}\|^2}{\|\mathbf{x}_K(n)\|^2}\right] = 1, \tag{5.55}$$

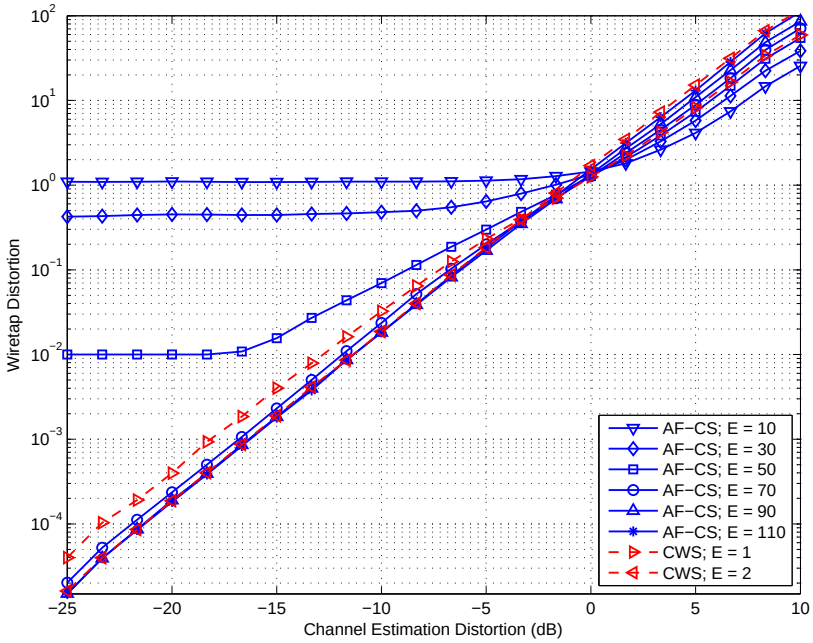

Figure 5.11: Wiretap distortion as a function of the channel estimation distortion for different number of coordinated eavesdroppers for $K = 10$ and $S = 200$. Solid lines represent the performance of AF-CS while dashed lines denote CWS. This figure has been averaged over 1000 realizations.

Hence, an important conclusion is that if the intended nodes set $\mathcal{D} = 0$dB, the eavesdroppers will achieve a relative wiretap distortion of 1 independently of E.

5.7 Conclusions

In this chapter, we have evaluated the Amplify-and-Forward Compressed Sensing (AF-CS) as a physical layer secrecy solution for Wireless Sensor Networks (WSNs). In particular, we have studied the robustness for each of the different phases of the given scheme against a passive eavesdropper agent composed by several malicious and coordinated nodes.

We have analytically demonstrated that AF-CS achieves perfect secrecy for the cases when the number of eavesdroppers E is less than or equal to the sparsity level of the signal K.

For larger number of eavesdroppers, we have proposed a secure training phase based on random pilots that contaminates their channel estimation. In fact, the relative wiretap distortion at the eavesdroppers grows linearly with the power of the introduced perturbation.

The simulation results for both the communication and the detection scenarios support our claim that the scheme under study is perfectly secret at physical layer when the number of eavesdropping nodes is less than the sparsity level of the signal. On the other hand, and assuming the ideal case of perfect channel estimation at the eavesdropper's side, high decoding rates (higher than 0.5) are only achievable when the number of eavesdropping nodes is high enough to hold the restricted isometric property condition.

Moreover, we show that the required number of eavesdroppers increases fast as a function of their channel estimation degradation and therefore the system can adapt the level of introduced distortion in order to control packet error rate or probability of detection of the eavesdroppers. Actually, we have observed that for channel perturbations similar to the channel

variance, the eavesdroppers obtain the same relative wiretap distortion that almost without any knowledge about the signal. However, the price to pay is that the more distortion we add at the eavesdroppers, the higher the energy cost at the sensing nodes.

Furthermore, AF-CS drastically outperforms other distributed compressed sensing solutions for WSNs in terms of physical layer secrecy.

Chapter 6

Conclusions and Future Work

This dissertation has studied the energy limitation of the Wireless Sensor Networks (WSNs) from a communications perspective. In particular, it has focused on the realistic case where correlations in the temporal and spatial domain are present in the input signal. Particular attention has been placed on the analysis and design of distributed schemaes that exploit this space-time correlation in order to obtain energy-efficient communication solutions. Based on existing results in signal processing as Compressed Sensing, a simple framework called *Amplify-and-Forward Compressed Sensing* (AF-CS) has been developed throughout this dissertation in order to design efficient communication schemes with several desired features for WSNs such as energy-efficient, resource-limited, low complex, high reconstruction accuracy, and protection agains eavesdropping. Simulation results have been provided in order to support the theoretical results and quantify the performance of the proposed schemes. As a result, the proposed schemes have been shown to improve other results in the literature in many of the analyzed metrics such as symbol/packet error rate, relative energy consumption, channel uses, mean square error, probability of detection, and perfect secrecy.

6.1 Conclusions

The motivation and the organization of the present dissertation has been given in Chapter 1. Chapter 1 has also emphasized that Chapters 2-5 contain the main research contributions. Therefore the main conclusions of this dissertation are grouped as follows.

Chapter 2 has evaluated the performance of different encoding-decoding strategies in order to reduce the number of transmitted samples. Concretely, the six three encoders and two decoders have been proposed, those are the Deterministic Downsampling Encoder (DDE), the Probabilistic Downsampling Encoder (PDE), the Conditional Downsampling Encoder (CDE), the Step Decoder (SD), and the Predictive Decoder (PD). Each of the six possible encoding-decoding combinations have been analytically studied and closed forms for their distortion have been obtained (approximations for the CDE-SD and CDE-PD). Simulation results have been used in order to validate the theoretical expressions. Moreover, we have concluded that the pair CDE-PD drastically outperforms the rest of strategies.

Chapter 3 has proposed two enhanced estimations based on the correlation parameters for both the Linear Wiener Filter (LWF) and its derived Mean Square Error as the main contribution of this chapter. In particular, they have been incorporated in two key steps of a practical Distributed Source Coding (DSC) scheme. It is shown that the DSC performs better in terms of compression rate and symbol error rate when the proposed estimators replace the classical sample estimators. This improvement is specially significative for the cases when the number of snapshots used in the training phase and the dimension of the observation vector are similar.

Chapter 4 has introduced the novel AF-CS as a distributed solution for energy-efficient WSNs. Since it is based on CS, some of the most important results on this topic are reviewed and some new contributions have been proposed, such as a revised relation among the number of required measurements as a function of the sparsity and the dimension of the input

signal. According to those results, an analytical model is proposed in order to design and characterize the AF-CS scheme. Furthermore, the simulation results have shown that AF-CS drastically reduces the number of transmissions and the number of channel uses compared to other transmission schemes present in the literature.

Chapter 5 has extended the results of Chapter 4 and has evaluated the AF-CS as a physical-layer secrecy solution for WSNs. In particular, the presence of a eavesdropper set of nodes has been considered. The secrecy level for each of the cases according to the number of eavesdroppers have been separately studied. It has been demonstrated that AF-CS achieves perfect secrecy for the cases when the number of eavesdroppers is less than the sparsity of the signal. For larger number of eavesdroppers, a random pilot technique has been proposed as a training phase in order to contaminate the channel estimation of the eavesdroppers and therefore decrease their performance. Simulation results have supported the theoretical results and they have validated the random pilot technique as a good candidate to increase the protection of AF-CS against a large number of eavesdropping nodes.

6.2 Future Work

Some lines of research regarding Chapters 2-5 remain as future work.

Regarding the downsampling encoders proposed in Chapter 2, the main open problems are

- To obtain exact analytical expressions for the downsampling distortion of the CDE-SD and CDE-PD pairs instead of approximations.
- To obtain the optimal set of threshold values for the condition in CDE that minimize the downsampling distortion of CDE-SD and CDE-PD pairs.

Regarding the enhanced correlation estimators in Chapter 3, the remaining issues are

- To obtain analytical results for both estimators for the case that the correlation matrix and the cross-correlation vector are dependent each other.
- To explore the benefits of the enhanced correlation estimators over other transmission techniques rather than DSC such as zero-delay downsampling transmissions and therefore apply them for a more realistic AF-CS scheme.

With respect to the CS theory in Chapter 4, the main open problems is

- Providing an analytical expression for the proposed relation among the number of required measurements as a function of the sparsity and the dimension of the input signal instead of the empirical one.

Regarding to the AF-CS in Chapter 4, some extensions can be proposed, such as

- Extending the results for the case of decode-and-forward transmission schemes.
- Finding new AF-CS-based schemes for partial or no channel state information.
- Exploring the possibilities of adding new degrees of freedom such as Multiple-Input Multiple-Output (MIMO) or Orthogonal Frequency Division Multiplexing (OFDM) at the transmitters and/or the receivers.

With respect to the AF-CS as a physical-layer solution in Chapter 5, the following issue remains as future work

- Obtaining a more accurate solution in terms of perfect secrecy for the cases when the number of eavesdroppers is higher than the sparsity level of the input signal.

Bibliography

[Agr11] S. Agrawal, S. Vishwanath, "Secrecy Using Compressive Sensing", *Information Theory Workshop (ITW), 2011 IEEE*, pags. 563 –567, oct. 2011.

[And12] H. S. Anderson, "On Discovering the Compressive Sensing Matrix From Few Signal/Measurement Pairs", Tech. rep., Sandia National Laboratories, Available at: http://wifs11.org/Documents/Poster011.pdf, 2012.

[Bai07] Z. D. Bai, B. Q. Miao, G. M. Pan, "On Asymptotics Behavior of Eigenvectors of Large Sample Covariance Matrices", *Ann. Probab.*, Vol. 35, nº 4, pags. 457–483, 2007.

[Baj06] W. Bajwa, J. Haupt, A. Sayeed, R. Nowak, "Compressive Wireless Sensing", *Proc. Information Processing in Sensor Networks, 2006. IPSN 2006. The Fifth International Conference on*, pags. 134 –142, New York, NY, USA, Apr. 2006.

[Bar07] R. G. Baraniuk, M. A. Davenport, R. A. DeVore, M. B. Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices", *Constructive Approximation*, 2007.

[BL10] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Distributed Source Coding in Large Wireless Sensor Networks", *Proc. Signals, Systems and Computers, 2010 44th Asilomar Conference on*, Nov. 2010.

[BL11] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Optimization of the Amplify-and-Forward in a Wireless Sensor Networks Using Compressed Sensing", *Proc. 19th European Signal Processing Conference (EUSIPCO '11)*, Barcelona, Spain., Aug. 2011.

[BL12a] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Amplify-and-Forward Compressed Sensing as a PHY-Layer Secrecy Solution for Wireless Sensor Networks", *Proc. 7th Sensor Array and Multichannel Signal Processing (SAM 2012)*, Hoboken, NJ, USA., Jun. 2012.

[BL12b] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Amplify-and-Forward Compressed Sensing as an Energy-Efficient Solution for Wireless Sensor Networks", *ACM Transactions on Sensor Networks*, Submitted for publication at March 2012.

[BL12c] J. E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Distortion of Zero-Delay Downsampling Schemes for Auto-Regressive Sources with Incomplete Observation Vectors", *EURASIP Journal on Advances in Signal Processing*, Submitted for publication at June 2012.

[BL12d] J.E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Amplify-and-forward compressed sensing as a physical layer secrecy solution for wireless sensor networks", *IEEE Trans. Info. Forensics and Security*, submitted for publication at July 2012 2012.

[BL12e] J.E. Barcelo-Llado, A. Morell, G. Seco-Granados, "Enhanced Correlation Estimators for Distributed Source Coding in Large Wireless Sensor Networks", *IEEE Sensors Journal*, 2012.

[Can05] E.J. Candès, T. Tao, "Decoding by Linear Programming", *Information Theory, IEEE Transactions on*, Vol. 51, nº 12, pags. 4203–4215, Dec. 2005.

[Can06a] E.J. Candès, J. Romberg, T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", *Information Theory, IEEE Transactions on*, Vol. 52, nº 2, pags. 489–509, Feb. 2006.

[Can06b] E.J. Candès, T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?", *Information Theory, IEEE Transactions on*, Vol. 52, nº 12, pags. 5406 –5425, Dec. 2006.

[Can08a] E Candès, "The Restricted Isometry Property and its Implications for Compressed Sensing", *Comptes Rendus Mathematique*, Vol. 346, nº 9-10, pags. 589–592, 2008.

[Can08b] E.J. Candès, M.B. Wakin, "An Introduction To Compressive Sampling", *Signal Processing Magazine, IEEE*, Vol. 25, nº 2, pags. 21–30, March 2008.

[Can11] E.J. Candès, Y. Plan, "A Probabilistic and RIPless Theory of Compressed Sensing", *Information Theory, IEEE Transactions on*, Vol. 57, nº 11, pags. 7235 –7254, Nov. 2011.

[Car77] A. Carleial, M. Hellman, "A Note on Wyner's Wiretap Channel (Corresp.)", *Information Theory, IEEE Transactions on*, Vol. 23, nº 3, pags. 387 – 390, May 1977.

[Che98] S. S. Chen, D. L. Donoho, Michael, A. Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, Vol. 20, pags. 33–61, 1998.

[Che06] B. Chen, L. Tong, P. K. Varshney, "Channel-Aware Distributed Detection in Wireless Sensor Networks", *Signal Processing Magazine, IEEE*, Vol. 23, nº 4, pags. 16 – 26, Jul. 2006.

[Cho03] J. Chou, D. Petrovic, Kannan Ramachandran, "A Distributed and Adaptive Signal Processing Approach to Reducing Energy Consumption in Sensor Networks", *Proc. IEEE INFOCOM 2003.*, Vol. 2, pags. 1054 – 1062, Mar. 2003.

[Cho09] C. Chou, R. Rana, W. Hu, "Energy Efficient Information Collection in Wireless Sensor Networks Using Adaptive Compressive Sensing", *Local Computer Networks, 2009. IEEE 34th Conference on*, pags. 443–450, Zurich, Switzeland, Oct. 2009.

[Der12] M. Derpich, "Improved Upper Bounds to the Causal Quadratic Rate-Distortion Function for Gaussian Stationary", *Information Theory, IEEE Transactions on*, Vol. 58, nº 99, pags. 3131 – 3152, May 2012.

[Don06a] David L. Donoho, "Thresholds for the Recovery of Sparse Solutions via l1 Minimization", *Proc. ell-1 minimization, Conf. on Information Sciences and Systems*, Princeton, NJ, USA, Mar. 2006.

[Don06b] D.L. Donoho, "Compressed Sensing", *Information Theory, IEEE Transactions on*, Vol. 52, nº 4, pags. 1289–1306, Apr. 2006.

[Far85] N. Farvardin, J. Modestino, "Rate-Distortion Performance of DPCM Schemes for Autoregressive Sources", *Information The-*

ory, IEEE Transactions on, Vol. 31, nº 3, pags. 402 – 418, May 1985.

[Fen07] X. Feng, Z. Zhang, "The Rank of a Random Matrix", *Applied Mathematics and Computation*, Vol. 185, pags. 689–694, 2007.

[Fra07] C. Fragouli, E. Soljanin, *Network Coding Fundamentals*, now Publishers, 2007.

[Gas03] M. Gastpar, M. Vetterli, "Source-Channel Communication in Sensor Networks", *Proc. of the 2nd international conference on Information processing in sensor networks*, IPSN'03, pags. 162–177, Palo Alto, CA, USA, 2003.

[Gir90] V. L. Girko, "$G_{25}-$Estimators of Principal Components", *Theory. Probab. Mathematical Statistics*, Vol. 40, pags. 1–10, 1990.

[Gir98] V. L. Girko, *An Introduction to Statistical Analysis of Random Arrays*, VSP, The Netherlands, 1998.

[Goe08] S. Goel, R. Negi, "Guaranteeing Secrecy using Artificial Noise", *Wireless Communications, IEEE Transactions on*, Vol. 7, nº 6, pags. 2180 –2189, Jun. 2008.

[Gop08] P.K. Gopala, L. Lai, H. El Gamal, "On the Secrecy Capacity of Fading Channels", *Information Theory, IEEE Transactions on*, Vol. 54, nº 10, pags. 4687 –4698, Oct. 2008.

[Gra08] M. Grant, S. Boyd, "Graph Implementations for Nonsmooth Convex Programs", V. Blondel, S. Boyd, H. Kimura (eds.), *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pags. 95–110, Springer-Verlag Limited, 2008.

[Gra11] M. Grant, S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 1.21", Apr. 2011.

[Gul01] O.G. Guleryuz, M.T. Orchard, "On the DPCM Compression of Gaussian Autoregressive Sequences", *Information Theory, IEEE Transactions on*, Vol. 47, nº 3, pags. 945 –956, Mar. 2001.

[Gup05] H. Gupta, V. Navda, S. R. Das, V. Chowdhary, "Efficient Gathering of Correlated Data in Sensor Networks", *MobiHoc '05: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, pags. 402–413, ACM, New York, NY, USA, 2005.

[Has80] T. Hashimoto, S. Arimoto, "On the Rate-Distortion Function for the Nonstationary Gaussian Autoregressive Process", *Information Theory, IEEE Transactions on*, Vol. 26, nº 4, pags. 478 – 480, jul 1980.

[Hau08] J. Haupt, W.U. Bajwa, M. Rabbat, R. Nowak, "Compressed Sensing for Networked Data", *Signal Processing Magazine, IEEE*, Vol. 25, nº 2, pags. 92–101, Mar. 2008.

[Hay01] S. Haykin, *Adaptive Filter Theory (4th Edition)*, Prentice Hall, Sep. 2001.

[Her10] M.A. Herman, T. Strohmer, "General Deviants: An Analysis of Perturbations in Compressed Sensing", *Selected Topics in Signal Processing, IEEE Journal of*, Vol. 4, nº 2, pags. 342 –349, Apr. 2010.

[Jol02] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, USA, 2nd ed., 2002.

[Kas04] A. Kashyap, T. Basar, R. Srikant, "Correlated Jamming on MIMO Gaussian Fading Channels", *Communications, 2004 IEEE International Conference on*, Vol. 1, pags. 458 – 462 Vol.1, Jun. 2004.

[Kay93] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*, Prentice Hall, Inc., 1993.

[Li07] Z. Li, R. Yates, W. Trappe, "Secret Communication with a Fading Eavesdropper Channel", *Proc. IEEE ISIT*, Nice, France, Jul. 2007.

[Luo09] C. Luo, F. Wu, J. Sun, C. W. Chen, "Compressive Data Gathering for Large-Scale Wireless Sensor Networks", *Proc. ACM Mobicom'09*, pags. 145–156, Beijing, China, Sep. 2009.

[Man09] B. G. Manjunath, S. Wilhelm, "Moments Calculation for the Double Truncated Multivariate Normal Density", *Available at http://dx.doi.org/10.2139/ssrn.1472153*, Sep. 2009.

[Mar67] V. A. Marčenco, L. A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices", *Math USSR Sbornik*, Vol. 1, pags. 457–483, 1967.

[Mar11] N. Marina, H. Yagi, H.V. Poor, "Improved Rate-Equivocation Regions for Secure Cooperative Communication", *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pags. 2871 –2875, Aug. 2011.

[May10] M. R. Mayiami, B. Seyfe, H. G. Bafghi, "Perfect Secrecy Using Compressed Sensing", *CoRR*, Vol. abs/1011.3985, 2010.

[MC91] J. A. Thomson. M. Cover, *Elements on Information Theory*, John Wiley & Sons, Inc., 1991.

[Meh91] M. L. Mehta, *Random Matrices*, Academic Press, Boston, 2nd ed., 1991.

[Men09] J. Meng, H. Li, Z. Han, "Sparse Event Detection in Wireless Sensor Networks Using Compressive Sensing", *Proc. Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pags. 181 –185, Mar. 2009.

[Mes05] X. Mestre, M. Lagunas, *Diagonal Loading for Finite Sample Size Beamforming: an Asymptotic Approach*, nº Chapter in: *Robust Adaptive Beamforming* Ed. J. Li and P. Stoica,, John Wiley & Sons, New York, NY, USA, 2005.

[Mes06] X. Mestre, M.A. Lagunas, "Finite Sample Size Effect on Minimum Variance Beamformers: Optimum Diagonal Loading Factor for Large Arrays", *Signal Processing, IEEE Transactions on*, Vol. 54, nº 1, pags. 69 – 82, Jan. 2006.

[Mes08] X. Mestre, "On the Asymptotic Behavior of the Sample Estimates of Eigenvalues and Eigenvectors of Covariance Matrices", *Signal Processing, IEEE Transactions on*, Vol. 56, nº 11, pags. 5353 –5368, Nov. 2008.

[Mud09] R. Mudumbai, D.R. Brown, U. Madhow, H.V. Poor, "Distributed Transmit Beamforming: Challenges and Recent Progress", *Communications Magazine, IEEE*, Vol. 47, nº 2, pags. 102 –110, Feb. 2009.

[Muk10] A. Mukherjee, S. A. Fakoorian, Jing Huang, A. L. Swindlehurst, "Principles of Physical Layer Security in Multiuser Wireless Networks: A Survey", *Wireless Communications, IEEE Transactions on*, Vol. abs/1011.3754, 2010.

[Neg05] R. Negi, S. Goel, "Secret Communication Using Artificial Noise", *Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd*, Vol. 3, pags. 1906 – 1910, sept., 2005.

[Neu82] D. Neuhoff, R. Gilbert, "Causal Source Codes", *Information Theory, IEEE Transactions on*, Vol. 28, nº 5, pags. 701 – 713, sep 1982.

[Nik11] S. Nikitaki, P. Tsakalides, "Localization in Wireless Networks Based on Jointly Compressed Sensing", *Proc. 19th European Signal Processing Conference (EUSIPCO '11)*, Barcelona, Spain., Aug. 2011.

[Ogg08] F. Oggier, B. Hassibi, "The Secrecy Capacity of the MIMO Wiretap Channel", *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pags. 524 –528, Jul. 2008.

[Old08a] F. Oldewurtel, J. Ansari, P. Mahonen, "Cross-Layer Design for Distributed Source Coding in Wireless Sensor Networks", *Proc. Sensor Technologies and Applications, 2008. SENSORCOMM '08. Second International Conference on*, pags. 435 –443, Aug. 2008.

[Old08b] F. Oldewurtel, M. Foks, P. Mahonen, "On a Practical Distributed Source Coding Scheme for Wireless Sensor Networks", *Proc. Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pags. 228 –232, May 2008.

[O'N71a] J. B. O'Neal, "Delta-Modulation Quantizing Noise-Analytic and Computer Simulation Results for Gaussian and Television Input Signals", *Bell Syst. Tech. J.*, Vol. 45, pags. 117–141, Jan. 1971.

[O'N71b] J. B. O'Neal, "Signal to Quantization Noise Ratio for Differential PCM", *IEEE Trans. Commun. Technol.*, Vol. vol COM-19, pags. 568–569, Aug. 1971.

[Par11] J.L. Paredes, G.R. Arce, "Compressive Sensing Signal Reconstruction by Weighted Median Regression Estimates", *Signal Processing, IEEE Transactions on*, Vol. 59, nº 6, pags. 2585 – 2601, Jun. 2011.

[Pra02] S.S. Pradhan, J. Kusuma, K. Ramchandran, "Distributed Compression in a Dense Microsensor Network", *Signal Processing Magazine, IEEE*, Vol. 19, nº 2, pags. 51 –60, Mar. 2002.

[Pra03] S.S. Pradhan, K. Ramchandran, "Distributed Source Coding Using Syndromes (DISCUS): Design and Construction", *Information Theory, IEEE Transactions on*, Vol. 49, nº 3, pags. 626 – 643, Mar. 2003.

[Pro67] E. N. Protonotarios, "Slope Overload Noise in Differential Pulse Code Modulation Systems", *Bell Syst. Tech. J.*, Vol. 46, pags. 2119–2161, 1967.

[Rac08] Y. Rachlin, D. Baron, "The Secrecy of Compressed Sensing Measurements", *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pags. 813 –817, sept. 2008.

[Ram10] S. Ramaswamy, K. Viswanatha, A. Saxena, K. Rose, "Towards Large Scale Distributed Coding", *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pags. 1326–1329, 14-19 2010.

[Ree11] G. Reeves, N. Goela, N. Milosavljevic, M. Gastpar, "A Compressed Sensing Wire-tap Channel", *Information Theory Workshop (ITW), 2011 IEEE*, pags. 548 –552, Oct. 2011.

[Ros61] S. Rosenbaum, "Moments of a truncated bivariate normal distribution", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 23, nº 2, pags. pp. 405–408, 1961.

[Rub09] F. Rubio, X. Mestre, "Consistent Reduced-Rank LMMSE Estimation With a Limited Number of Samples per Observation Dimension", *Signal Processing, IEEE Transactions on*, Vol. 57, nº 8, pags. 2889 –2902, Aug. 2009.

[Rug07] R. Rugin, A. Conti, G. Mazzini, "Experimental Investigation of the Energy Consumption for Wireless Sensor Network with Centralized Data Collection Scheme", *Proc. Software, Telecommunications and Computer Networks, 2007. SoftCOM 2007. 15th International Conference on*, pags. 1 –5, Sep. 2007.

[Sax10] A. Saxena, K. Rose, "On Scalable Distributed Coding of Correlated Sources", *Signal Processing, IEEE Transactions on*, Vol. 58, nº 5, pags. 2875 –2883, May 2010.

[Sha49] C. E. Shannon, "Communication Theory of Secrecy Systems", *Bell Sys. Tech. Journ.*, Vol. 28, pags. 656–715, 1949.

[Sha05] S. Shafiee, S. Ulukus, "Capacity of Multiple Access Channels with Correlated Jamming", *Military Communications Conference, 2005. MILCOM 2005. IEEE*, pags. 218 –224 Vol. 1, Oct. 2005.

[Sha07] S. Shafiee, S. Ulukus, "Achievable Rates in Gaussian MISO Channels with Secrecy Constraints", *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pags. 2466 –2470, Jun. 2007.

[Sle73] D. Slepian, J. Wolf, "Noiseless Coding of Correlated Information Sources", *Information Theory, IEEE Transactions on*, Vol. 19, nº 4, pags. 471 – 480, Jul. 1973.

[Sri08] M. Srivatsa, "Who is Listening? Security in Wireless Networks", *Proc. Signal Processing, Communications and Networking, 2008. ICSCN '08. International Conference on*, pags. 167 –172, Jan. 2008.

[Tan07] Z. Tang, I.A. Glover, A.N. Evans, J. He, "Energy Efficient Transmission Protocol for Distributed Source Coding in Sensor Networks", *Proc. Communications, 2007. ICC '07. IEEE International Conference on*, pags. 3870 –3875, Jun. 2007.

[Toh01] C. K. Toh, "Maximum Battery Life Routing to Support Ubiquitous Mobile Computing in Wireless Ad Hoc Networks", *Communications Magazine, IEEE*, Vol. 39, nº 6, pags. 138 –147, Jun. 2001.

[Tul04] A. M. Tulino, S. Verdú, *Random Matrix Theory and Wireless Communications*, Foundations and Trends in Communications and Information Theory 1 (1), Jun. 2004.

[Vis00] H. Viswanathan, T. Berger, "Sequential coding of correlated sources", *Information Theory, IEEE Transactions on*, Vol. 46, nº 1, pags. 236 –246, jan 2000.

[Wan08] H. Wang, D. Peng, W. Wang, H. Sharif, H. Chen, "Cross-Layer Routing Optimization in Multirate Wireless Sensor Networks for Distributed Source Coding Based Applications", *Wireless Communications, IEEE Transactions on*, Vol. 7, nº 10, pags. 3999 –4009, Oct. 2008.

[Wei05] T. Weissman, N. Merhav, "On Causal Source Codes with Side Information", *Information Theory, IEEE Transactions on*, Vol. 51, nº 11, pags. 4003 – 4013, Nov. 2005.

[Wyn75] A. D. Wyner, "The Wire-tap Channel", *Bell Sys. Tech. Journ.*, Vol. 54, nº 1355-1387, 1975.

[Yan11] Z. Yang, C. Zhang, L. Xie, "Robustly Stable Signal Recovery in Compressed Sensing with Structured Matrix Perturbation", *CoRR*, Vol. abs/1112.0071, 2011.

[Yeu05] R. W. Yeung, S.-Y. R. Li, N. Cai, Z. Zhang, *Network Coding Theory*, now Publishers, 2005.

[You04] O. Younis, S. Fahmy, "HEED: a Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks", *Mobile Computing, IEEE Transactions on*, Vol. 3, nº 4, pags. 366 – 379, Oct. 2004.

[Zam08] R. Zamir, Y. Kochman, U. Erez, "Achieving the Gaussian Rate Distortion Function by Prediction", *Information Theory, IEEE Transactions on*, Vol. 54, nº 7, pags. 3354 –3364, Jul. 2008.

[Zar11] K. Zarifi, S. Zaidi, S. Affes, A. Ghrayeb, "A Distributed Amplify-and-Forward Beamforming Technique in Wireless Sensor Networks", *Signal Processing, IEEE Transactions on*, Vol. 59, nº 8, pags. 3657 –3674, Aug. 2011.

[Zha10] L. Zhang, R. Zhang, Y. Liang, Y. Xin, S. Cui, "On the Relationship Between the Multi-Antenna Secrecy Communications and Cognitive Radio Communications", *Communications, IEEE Transactions on*, Vol. 58, nº 6, pags. 1877 –1886, Jun. 2010.