UNIVERSITY OF MACEDONIA

DEPARTMENT OF APPLIED INFORMATICS

Efficient Compressed Sensing for Wireless Communication Systems

Doctoral Thesis

of

Theofanis Xifilidis

Thessaloniki, 2022

Doctoral Thesis Committee:

Konstantinos E. Psannis, Associate Professor, University of Macedonia (supervisor)

Manos Roumeliotis, Professor, University of Macedonia

Konstantinos Margaritis, Professor, University of Macedonia

EFFICIENT COMPRESSED SENSING FOR WIRELESS COMMUNICATION SYSTEMS

Theofanis Xifilidis

Thesis submitted

in partial fulfillment of the requirements for the Doctorate Degree at

University of Macedonia

Examination Committee:

Konstantinos E. Psannis, Associate Professor, University of Macedonia (supervisor)

Manos Roumeliotis, Professor, University of Macedonia

Konstantinos Margaritis, Professor, University of Macedonia

Eleftherios Angelis, Professor, Aristotle University of Thessaloniki

Panagiotis Sarigiannidis, Associate Professor, University of Western Macedonia

Petros Nikopolitidis, Associate Professor, Aristotle University of Thessaloniki

Sofia Petridou, Assistant Professor, University of Macedonia

Author: Theofanis Xifilidis

Thessaloniki, 2022

In memory of my father

Acknowledgements

First of all, I would like to express my gratitude to my supervisor Professor Kostas Psannis for his valuable guidance and support throughout the preparation of this thesis and his endless patience, dedication and greatly contributive feedback throughout the research and writing of this thesis.

I would also like to thank my advisory committee, Professor Manos Roumeliotis and Professor Konstantinos Margaritis for their continuous support and valuable collaboration for competing my thesis.

Most importantly, I am full of gratitude to my beloved mother and sister, for their unconditional love, endless support and guidance throughout the completion of this thesis.

Abstract

The currently deployed wireless communication systems have been inextricably linked to the Big Data deluge conveying an explosive growth of transmitted and processed heterogeneous data that demand a low complexity and resource efficient information extraction. The above necessity constitutes the main motivation for a thorough understanding of advantages and limitations that govern the communication system performance analysis and design.

As with other scientific areas, the performance evaluation of modern communication systems implies the careful formulation of an optimization problem with relevant objective and constraints that best model the practical requirements of the specific problem at hand. To that end, a priori information for the specific problem is proven to contribute to the acquiring of an optimal solution. CS comprises a set of optimization tools that effectively convert a problem of high intractability i.e. combinatorial nature to a reduced dimensionality problem with low computational and implementation complexity. Thus, CS has already emigrated to wireless communication system performance evaluation due to the essential fact that the convenient property of structure as additional information is also verified in the wireless communication field. Efficient algorithms that jointly consider physical limitations of the wireless channel, in terms of the distortion it induces to the transmitted signal, are already present and rapidly evolving.

This thesis constitutes the integration of CS and information theory in statistical modeling of the wireless channel from the essential statistical independence assumption point of view. The random distorting mechanisms of the wireless channel are accounted for thus additive noise and diverse fading. Moreover, the concept of resource efficiency adapts to this thesis in terms of information accessibility in a large scale deployed network. The above statistical assumption provides useful performance insights which extent in terms of comparison to the more practical correlation case with respect to additive noise and fading distributions.

Finally, the merit of the results derived in this thesis is verified by stating feasible application of these results in terms of CS theory and principles to the current Fifth Generation communication systems. Through this relation, limitations of the current communication system generation can further boost performance and provide guidelines in terms of practical application specific considerations serving as motive to designing improved future communication systems.

Keywords

Compressed Sensing, Wireless Communications, Information theory, Performance, Optimization, Complexity, Sparsity, Compressibility, Fading, Noise, Fifth Generation Networks, Statistics, Independence, Correlation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

4G	Fourth Generation Networks
5G	Fifth Generation Networks
6G	Sixth Generation Networks
ADC	Analog-To-Digital Converter
AOA	Angle-of-Arrival
AOD	Angle-of-Departure
BPDN	Basis Pursuit DeNoising
BS	Base Station
CDI	Channel Distribution Information
CDMA	Code Division Multiple Access
CLT	Central Limit Theorem
CR	Cognitive Radio
CS	Compressed Sensing
CSI	Channel State Information
D2D	Device To Device
DCS	Distributed Compressed Sensing
ED	Energy Detection
FDMA	Frequency Division Multiple Access
IoT	Internet-of-Things
LMS	Learning Management System
LOS	Line-of-Sight
LRT	Likelihood Ratio Test
MAP	Maximum A Posteriori
MEC	Mobile Edge Caching
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
mMIMO	Massive Multiple Input Multiple Output
MRC	Maximal Ratio Combining
NLOS	Non-Line-of-Sight
NOMA	Non Orthogonal Multiple Access
NP	Non Polynomial
NSP	Null Space Property
OFDM	Orthogonal Frequency Division Multiplexing
OMP	Orthogonal Matching Pursuit
PU	Primary User
QoS	Quality of Service
RIP	Restricted Isometry Property
SIMO	Single Input Multiple Output

SINR	Signal to Interference Noise Ratio
SIR	Signal to Interference Ratio
SN	Social Network
SNR	Signal to Noise Ratio
SU	Secondary User
TDMA	Time Division Multiple Access
UWB	Ultra Wide Band
WSN	Wireless Sensor Network

1. INTRODUCTION

1.1 Wireless Communication Systems Fundamentals

Ever since the pioneering works of Maxwell, Hertz and Tesla which established the wireless transmission of information and Marconi devising the first system of radio communication the area of wireless communications continues to rapidly flourish and develop emerging technologies that offer an increasingly extensive capability of information transmission around the globe with new technologies, standards and services resulting in ubiquitous communication infrastructures and implemented communication systems.

Wireless communications [1],[2] involve the transfer of information, a term which permits many diverse interpretations depending on what we are transmitting and the effect of the classification of this term in the overall performance and case-dependent goals which are to be achieved in a communication scenario. It is clear that information transmission relies on the well-established principle that the receiver is not aware of the information content the transmitter intends to send, at least not in a manner that would render the whole communication scheme unnecessary. This concept verifies the fundamental property of wireless information transmission being not deterministic but random. It is clear thus that statistics and stochastic process theory constitute the main cornerstone of the mere feasibility and moreover performance evaluation along with understanding tradeoffs and limitations in the respective theory so as to further boost performance and develop next generation advanced systems.

In every communication scenario the source generates the information. The latter term has taken a plethora of meanings and interpretations [3],[4] that extend to totally abstract aspects as can be found in the information theory literature, which will be subsequently discussed. However, the effective subset of these interpretations that are applicable in wireless communication systems terminology and metrics have greatly aided to the mathematical formalism and communication system performance evaluation and implementation. Thus, qualitative and quantitative explanation of the term information is a necessity in the wireless communication field of research. The concept of amount is not as crucial as what kind of information is being transmitted. Although the concept of information transmitter briefly detailed as electric current and its transformation to electromagnetic wave propagated through the wireless channel remains common as the wireless communication systems have evolved, the query as to what is being transmitted is the main fact that drives the communication systems evolution. Hence, the transition from text to image and multimedia transmission is one of the key elements that differentiate the communication systems and its capabilities through the evolution from past digital generation systems to the current 5G and future 6G communication systems which seem to shortly become a reality as a consequence of the 5G system saturation in the context of the extent of advantages and solutions provided as opposed to what future needs will dictate. The next ingredient of a communication to be discussed is the system receiver skipping the introduction to the channel and its impact on communication efficiency as this is the core element on which the analysis of this thesis is based and will be accounted thereafter.

The wireless receiver operates on the fundamental principle of not only inverting the electromagnetic wave into a compatible for the receiver circuitry form but also inverting all processes conducted in the transmitter prior to transmission. Indeed, this property of invertibility is exactly what verifies the optimal performance of decoding and acquiring the transmitted signal as though no impairments were present or in a more accurate manner with an acceptable distortion that does not cancel the successful reception of the specific information sequence. As a complementary remark, the limitations imposed by the channel characterized as additive noise and fading are also present in the processing in transmitter and receiver circuitry. This reflects the fact that although processing aims at providing information integrity and compatible form for minimizing distortion, the concept of channel impairments is practically inevitable in any process conducted or any communication system implemented.

The analysis now proceeds with the introduction to the wireless channel [5], based on whose optimization the current thesis relies on. Three are the main sources of impairments induced by the channel: additive noise, fading and interference. Additive noise is modeled as a Gaussian random variable as a means to state that the signal passing through the medium is perturbed by a random distorting process that is not deterministic in any way and can only be statistically described as a consequence of realizations dictated by the respective distribution. System performance analysis also refers to non-Gaussian noise forms such as impulsive noise which are more complex and require more efficient analysis and statistical modeling. Analysis of communication system performance in this thesis restricts to Gaussian additive noise consideration. Proceeding to fading which can also be Gaussian modeled but also following other distribution [6],[7] it is defined as random fluctuations in power due to the multipath effect i.e. waves reaching the receiver from different directions, the latter fitting in the context of multi-antenna systems [8],[9], where a main challenge is the phase correction of the incident waves thus mitigating the destructive contribution of incident waves in approximately opposite phase. The concept of independence is crucial in this context as it ensures that deep fading of an independence branch is not likely to be accompanied by deep fade in other incident waves at the receiver. Finally, interference is the result of multiple users transmitting in the same frequencies or in same time interval thus rendering received signal decoding difficult and sometimes infeasible if interference cancellation techniques are not employed. Orthogonality is a main strategy of transmitting signals with minimum interference, which is applied in the problem of resource allocation as in users transmitting in a cell or nearby cells. It must be however noted that non-orthogonality has also attracted attention for accommodating more users in a given cellular network scheme where alternative techniques are utilized and orthogonality condition is relaxed in pursuit of fairness. The scheme explained also extends to CR systems where the goal that must be achieved is opportunistic user transmission even in the same interval or frequency bin but with acceptable interference. This constitutes an important optimization problem still under dense investigation. The goal of the communication system is to handle redundancy stemming from analog signals to digital signals. The reduction of redundancy results in diversity. This enhances system capacity as most evident in MIMO systems where diversity is attained in the spatial dimension. However, redundancy can be exploited to ensure reliable information transmission, a concept realized in many ways along with

complexity reduction and design optimization. Redundancy is observed in the analog signal while discrete signals [10] are produced in a manner that in general leads to inevitable information loss. This property with numerous relations to many scientific branches dictates the necessity to devise methods that minimize this information loss, as digital signals are the basis of current communication systems.

Diversity is a key property that ensures capacity increase [11],[12] by means of different techniques that are solely based on the so called degrees of freedom of the communication system from which the most essential are given i.e. time, frequency, space and code. MIMO wireless systems also utilize the concept of angle as a consequence of the use of multiple antenna at transmitter and/or receiver and Direction-of-Departure and Direction-of-Arrival derivations. The design and implementation of each communication system as well as its metrics and requirements are not identical but entirely depend on the application at hand as well as the goals to be achieved translated as system performance metrics. The following paragraphs detail the degrees of freedom along with the concept of diversity interpreted to each of the latter.

As intuitively deduced, time involves information exchange in successive time intervals a scheme known as TDMA [13]. In this scheme, each scheduled user transmits information at preassigned intervals in a manner of introducing guard zones so as to avoid interference. As a degree of freedom, time can be translated in many ways. Scheduling along with synchronization is a fundamental issue that defines multi-user communications and ensures reliable communication and successful decoding at the receiver side. Although asynchronous communication is also extensively studied, synchronization remains a key role for successful detection and decoding at the receiver. Time has been the primary metric of communication system performance and translated to information transmission and reception achieving optimal performance by suitable scaling of transmission and reception time intervals and scheduling information exchange in different time intervals for each user thus utilizing time diversity.

Frequency diversity is another metric based on the latter degree of freedom that relies on the frequency content of the transmitted sequence. Hence diversity is achieved by assigning different frequency bins to each user transmission, a scheme known as FDMA [14]. The concept of allocated bandwidth has been the major resource that is to be optimally allocated in each user always in accordance to the requirements and tradeoffs of each application-dependent communication scenario. The current state-of-the-art networks are characterized by the following limitations: spectrum scarcity and spectrum underutilization. The former concerns the limited bandwidth resulting from the numerous networks to which portions of available bandwidth have been assigned leading to small unoccupied bandwidth along with the constraining fact that only certain frequency bands can be allocated to users. The use of higher unoccupied frequency bands has emerged as a viable solution but introduces several drawbacks as a consequence of using higher frequencies for transmission such as more attenuation and required dimensions of antennas. Spectrum underutilization reflects the fact that the instantaneously used spectrum from its legitimate users is a considerably small fraction of the allocated bandwidth. This observation constitutes the reason spectrum sensing is the major performance functionality that has encompassed CS as a means of exploiting this observation. Moreover, the differentiation between narrowband and

wideband scenarios is a fundamental bottleneck in terms of performance and spectrum sensing complexity. The latter two characterizations are closely related particularly regarding the method of dividing the respective spectrum to smaller spectrum slices related to OFDM [15] method. CR networks are primarily concerned with the need for efficient spectrum sensing employing a diverse range of sensing methods and opportunistic spectrum use either assigning resources for each use or allowing secondary users to use the spectrum zone without causing harmful interference to primary users. Hence, frequency diversity is achieved in this manner.

The concept of code has come to address system performance as an approach alternative to increasing signal-to-noise-ratio or allocating more bandwidth. As opposed to assigning different time intervals or frequency bins for optimal communication mitigating interference, CDMA [16] assigns a different coding sequence, i.e. signature to each user that renders the user distinct from the others. The straightforward advantage that code sequences serve to achieve the required diversity in a manner that allows users to transmit at same time intervals or utilize the same frequency bands. This achieved diversity can also be thought as a signaling of increased dimensionality which can be directly related to performance optimization, increased complexity and means of increasing transmission rate always considering the error probability of a specific communication scenario. Furthermore, coding schemes consist of inserting redundant bits to information sequences so as to combat channel impairments i.e. additive noise, fading and interference. Thus, the goal is lower error probability and robustness to varying channel conditions.

Finally, the spatial degree of freedom relies upon achieving such diversity by employing multiple antennas in transmitter and/or receiver side. The so called spatial multiplexing scheme is based on many incident waves impinging on multiple receive antennas by means of scatterers in the environment thus creating reflected signals that, as stated above, contribute either constructively thus aiding in correct detection and information decoding or destructively degrading receiver performance or increasing complexity by requiring phase correction so as to exploit the multiple signal of delayed replicas arriving at receiver side. Moreover, the achieved spatial diversity is also reflected in increased channel capacity in a linear manner with respect to the number of multiple antennas at transmitter or receiver. Stated in other words, capacity increases without increasing SNR or related bandwidth. Hence, capacity optimization comes at no additional cost in terms of bandwidth which means no extra noise as well as no SNR increase which is critical in SNR limited communication applications. The optimization of channel capacity has been extended to accommodate fading channels with the related channel models. These multiple-input-multiple-output channels are characterized as statistically independent i.e. realizing the achievable diversity, an effect obtained by sufficient spacing between multiple antennas that are deployed.

1.2 Wireless communication systems evolution: from 1G to 5G and beyond

The first reported cellular telephony known as 1G was introduced in 1980s and involved voice service delivery in an analog format. Amplitude modulation or narrowband frequency modulation were used in two frequency bands of width equal to 25MHz

namely 824 to 849MHz and 869 to 894MHz. 1G systems suffered from poor voice quality, small coverage and no encryption.

The main differentiating point of 2G networks was the use of digitally modulating techniques. In a digital context, pure waveform reconstruction at receiver was not the case rather than identifying a waveform matching one of available candidate waveforms. 2G systems appeared in the 1990s and offered increased capacity and low rate data application support. Global System for Mobile Communications (GSM) was the representative 2G paradigm. Apart from voice, text message transmission was also feasible. Rates from 80 to 120kbps were supported in 2G systems.

It was the 3G wireless systems that reaped the benefits of CDMA. Such systems were available in the 2000s and supported a variety of services including telephony, high speed data, video, paging and messaging. 3GPartnership project was available in 1998 using CDMA and thus providing high data rates namely up to low Mbps for downlink and uplink.

4G systems mainly addressed the issues of high data rates together with mobility. Thus, maximum downlink data rates of 1Gbps for low mobility and 100 Mbps for high mobility solutions were achieved. OFDM technology and LTE Advanced were considered as key 4G wireless system technologies. The spectral efficiencies brought about by OFDM i.e. the amount of information that can be contained in certain bandwidth for a target QoS were the main benefits from this generation. Multiple antenna techniques leveraged spatial diversity to further boost transmission rates with a cost in hardware implementation.

5G appeared at the Big Data era i.e. the enormous growth of diverse types of data to be exchanged at a global scale such as video and multimedia streaming in a mobility aware and energy efficient manner. High data amounts along with low latency and high reliability transmission requirements are the main advantages of the current generation systems namely 5G, which was first deployed in south Korea in 2019. Data rates on the order of Gbps for both uplink and downlink are enabled and mobility can also be extensively supported. Cloud computing services and IoT network deployment, being an order of magnitude larger than cellular networks in terms of scalability, are already a reality with billions of devices connected worldwide along with the Big Data deluge. mmWave communication technologies, mMIMO, beamforming and Artificial Intelligence tools are some of the key technologies in the 5G and beyond context.

1.3 Wireless communications and Information Theory

Information Theory [17],[18] involves the modeling and thorough understanding of performance limitations of wireless communication systems founded by Shannon's pioneering work: a mathematical theory of communications [19] which studied the impact of noise on signal transmission through the wireless channel and the necessity of exploiting the signal structure in a statistical sense of the information to be transmitted and reconstructed at receiver side. Though Shannon's work concentrated on additive noise channels thus not accounting for fading together with the state-of-the-art technology of communication systems at the time of this publication, his findings have provided profound understanding of communication systems performance limits

that are still consistent after the so many decades of progress made in communication system theory and practical design.

Information theory is a revolutionary field applicable not only in wireless communication theory and design but also in diverse fields where uncertainty, inherent to the concept of randomness, governs many areas such as signal and image processing, language processing as well as biology and genetics. In wireless communications it addresses two fundamental queries. The first one is to what extent can we compress an information source. It is proven that there exists a critical determined in each case level, namely entropy below which compression is not possible. Given a distinct meaning in the term information generating source in the wireless communication regime, statistical tools quantifying entropy are applied for evaluating performance limits in a universal sense. The second query refers to what extent can we maximize transmission rate answered by the channel capacity limit [20],[21]. Thus, as long as transmission rate falls below channel capacity, communication with arbitrarily low error can be achieved. Otherwise, transmission error is bounded away from zero. In a clarifying comment, as long as entropy is below channel capacity asymptotic to be exact error free transmission is achievable.

1.4 Information Theory and Compressed Sensing

Concluding, on the basis of information theory, the DCS approach [22] is beneficial in terms of signal, image and data processing where the convenient concept of structure and sparsity in certain domains has been verified. Furthermore, in sensor network scenarios, distributed source coding exploits correlation and hence, many compression algorithms have already been devised. Hence, coding techniques falling under the umbrella of information theory successfully integrate CS in a way to optimize wireless communication system theory. Channel coding also relates to compressibility thus effectively enhancing data sequence by additional bits for combating channel impairments. DCS also relate to computational asymmetry thus joint decoder is far more complex and resource demanding while the simplified encoder only confines its role to computing incoherent projections as dictated by CS principles. It must also be mentioned that the concept of adaptivity is dominated by non-adaptive CS based reconstruction methods, hence the above contrast could potentially lead to a refinement of CS optimization in the future.

1.5 Thesis Motivation and Scope

Motivation

In the era of explosive data growth namely the Big Data Deluge and the unrolling of 5G communication systems, intense research activity has focused on handling theses large data quantities in a resource efficient manner. Hence, methods of signal and data processing of reduced computational and implementation complexity while still obtaining the actual information content already exhibit significant progress with near optimal processing efficiency and low complexity. Another parameter to be considered are the limitations of currently deployed communication systems which are the main motivation for evolution of next generation communication systems. In the above context, CS has arisen as a set of powerful of optimization tools successfully addressing

the issue of useful information extraction. Together with universal approaches offered by CS, efforts to devising optimization methods in an adaptive sense tailored to a specific wireless communication performance scenario are also under extensive investigation. Statistical modeling integrated with entropy and channel coding have already been considered as effective schemes that not only enhance CS optimality but also offer reduced complexity methods for confronting the optimization problem at hand. The latter goal constituted the main motivation of this thesis.

Scope

The scope of this thesis is focused on the assessment of wireless communication system performance by considering channel statistical models under the prism of statistical independence assumption employing CS and information theory entropy calculations in a combined manner. The results of this thesis can provide useful insights compared to the practical assumption of correlation and assist in the thorough understanding of communication systems and effectively promote next generation wireless system analysis and design.

1.6 Thesis Contribution

This thesis proposes a performance evaluation of the wireless channel assisted by entropy calculations for deriving the average code length for describing the channel. Our channel modeling approach includes practical constraints specifically fading conditions and additive noise and is based on the assumption of statistical independence. Fading channel performance is assessed by capacity, average code length and error probability are derived for diverse fading conditions under the assumption of independence. An energy detection scheme is proposed based on convolutional statistics for the fading channels considered including additive Gaussian noise. Moreover, correlation is being assumed in a WSN scenario using the symmetric Gaussian distribution in a fully randomized model deriving reconstruction error and energy estimation error with CS application by joint correlation and sparsity considerations. The content caching methodology is also accounted for performing a probabilistic comparison by means of three distinct distributions. File segmentation scenario in sparse and dense cases was proven to exhibit convergence. A brief notion of privacy issues in SNs along with application of CS to medicine and healthcare data as well as smart education by proposition of interactive classroom is provided. The results in this thesis are directly related to performance issues in 5G networks supporting their practical merit. Finally, conclusions were drawn along with feasible extensions of this thesis for future research.

1.7 Thesis outline

After this comprehensive introduction, the thesis is structured as follows.

Chapter 2 provides a detailed notion on CS theory and its potential to reduce computational and implementation complexity by exploiting structure defined by sparsity. The mathematical preliminaries commonly used throughout this thesis are provided. Chapter 3 addresses channel distribution knowledge by combining entropy and CS based distribution reconstruction. Moreover, fading channel performance is

evaluated by CS compressibility consideration along with WSN performance based on the statistical independence and correlation assumptions. Chapter 4 investigates content caching by probabilistic scheme and file segmentation consideration. Security issues in SNs, healthcare data management and proposition of smart interactive classroom are also derived. Chapter 5 derives application of CS and the results from the previous chapters to the current 5G wireless systems challenges and performance. Chapter 6 draws conclusions and highlights future research directions.

1.8 Publications

The scientific findings of this thesis have been published in international peer-reviewed journals and international conference proceedings. A complete list of these publications is given below:

Publications in Journals:

[J1] T. Xifilidis and K. E. Psannis, " Caching hit probability and Compressive Sensing perspective for mobile cellular networks," Elsevier Simulation Modelling Practice and Theory, vol. 87, pp. 92-98, Sept. 2018.

[J2] T. Xifilidis, K.E. Psannis, “Wireless fading channels performance based on Taylor expansion and compressed sensing: A comparative approach”, in Wiley International Journal of Communication Systems, vol.34, issue 8, e4794, March 2021.

[J3] T. Xifilidis and K.E. Psannis, “Correlation-based wireless sensor networks performance: the compressed sensing paradigm,” Springer Cluster Computing, vol. 25, issue 2, pp. 965-981, Nov. 2021.

Publications in Conferences:

[C1] C. Stergiou, K. E. Psannis, T. Xifilidis, A. P. Plageras and B. B. Gupta, "Security and privacy of big data for social networking services in cloud," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2018, pp. 438-443.

[C2] I. Kakalou, D. Papadopoulou, T. Xifilidis, K. E. Psannis, K. Siakavara and Y. Ishibashi, "A survey on spectrum sensing algorithms for cognitive radio networks," 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), 2018, pp. 1-4.

[C3] C. Stergiou, A. P. Plageras, K. E. Psannis, T. Xifilidis, G. Kokkonis, S. Kontogiannis, K. Tsarava and A. Sapountzi, "Proposed High Level Architecture of a Smart Interconnected Interactive Classroom," 2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM), 2018, pp. 1-6.

[C4] T. Xifilidis and K. E. Psannis, "Practicing Medicine: A Compressed Sensing Approach," in Proceedings of New Technologies in Health: Medical, Legal and Ethical Issues Conference, 21-22 November 2019, Thessaloniki, Greece.

[C5] T. Xifilidis, K. E. Psannis, G. Minopoulos, G. Kokkonis and Y. Ishibashi, "Convolution Based Energy Detection Scheme for Cognitive Radio Systems," 2019 2nd World Symposium on Communication Engineering (WSCE), 2019, pp. 58-62.

[C6] T. Xifilidis and K. E. Psannis, "Fading Channel Coding Based on Entropy and Compressive Sensing," 2020 3rd World Symposium on Communication Engineering (WSCE), 2020, pp. 44-48.

2. COMPRESSED SENSING THEORY AND MATHEMATICAL PREREQUISITES

2.1 Compressed Sensing Theory and Principles

CS [23],[24] is a set of optimization tools aiming at drastically reducing the computational and implementation complexity in a way that guarantees optimal signal detection/estimation or reconstruction in a variety of signal processing oriented scientific research areas that are entangled with the information technology that strictly demands efficient information technology and resource efficient knowledge extraction from large or massive amounts of data.

CS is fundamentally based on the concept of sparsity [25] hence a dataset with the bulk of elements equal to zero except for a small fraction of the data being nonzero, hence this sparse set. CS achieves this complexity reduction by claiming the reduction of sampling rate given a signal while not degrading its reconstruction. The applicability of such an approach is traced in the conventional computationally expensive methodology of first sampling acquisition and, proceeding further, discarding a large portion of the samples prior to their processing stage. This is where CS enters the stage of simultaneously sampling and compression. This translates to sampling at a sub-Nyquist rate [26] as opposed to Shannon Nyquist sampling theorem [27], [28] the latter stating that the minimum rate for achieving an accurate signal reconstruction is twice the highest frequency present in the signal to be recovered. This is what formulates the term of CS. It must also be stressed that CS mathematical tools employed were not novel but existed many years before their inclusion in CS optimization theory and principle.

CS principle transforms an initially intractable problem, the latter defined as a computationally exhaustive problem, into another form that admits a more convenient solution even if such an approach results in a sub-optimal solution. The mathematical tools essentially involve random matrix algebra and l_p norm algebra. The initial problem formulation assumes p=0 thus counting the number of nonzero elements in a mathematical structure such as a column matrix. It must also be noted that mathematically the l_p norm for p equal to zero is not strictly a norm as it does not satisfy the triangle inequality. Nevertheless, it is mentioned as such and it provides the fundamental CS problem formulation. However, this form requires a combinatorial expensive solution and the straightforward consideration entails the transformation of this problem to a tractable one. The mathematical formula of the l_p norm quantity for arbitrary p value is given below:

$$\|x\|_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}} \qquad (1)$$

Having derived the meaning of the l_p norm for zero valued p, we also mention that for $p \rightarrow \infty$ the expression derives the greatest value of the elements in the aforementioned column matrix structure. The general problem statement of CS theory is the inverse of the problem that derives the output of a system with a sparse vector as input, the matrix

transforming the sparse vector known as sensing or measurement matrix and the perturbation quantity termed additive noise in the wireless communication literature. Hence, it must be noted that the problem in CS theory is underdetermined. This states that the measurement vector dimensionality is smaller than the ambient dimensionality of the sparse vector [29] to be recovered. Either assuming the noiseless regime thus zero valued additive noise vector or the consideration of existing noise, the problem is at first glance computationally intractable or in terms of combinatorial theory shown to be NP-Hard. Hence, a transformation is required for this problem to be converted to a solvable problem with the additional requirement to produce an optimal solution. This accounts for the case where the vector to be recovered is not sparse in its initial form. In this case, a so called sparsifying matrix is multiplied with the measurement matrix and vector product in order to transfer the problem in another domain where the resulting vector fulfills the sparsity requirement. Representative sparsifying matrices are Discrete Fourier Transform, Discrete Cosine Transform and Discrete Wavelet transforms. As an extension of the above procedure to achieve a «universal» transformation [30], the dimensions of time, frequency, space and code should always be considered not separately but also jointly in order to reap benefits of a general CS problem transformation procedure. Bearing in mind that information signals perceived in the above sparse signal context are random rather than deterministic in nature, this not only does not penalize the value of CS theory but further establishes the random matrix optimization framework of CS [31],[32] and renders the statement of signal recovery with overwhelming probability valid and conceptually meaningful. This is also justified by the fact that signal processing methods for random signals are simply extensions of their deterministic counterparts. As already stated, the CS problem is by definition underdetermined. This dictates that additional constraints must be imposed to narrow down possible problem solutions and effectively reach the optimal one. This constraints necessity is equivalent to assuming the sparse property as an optimization problem constraint.

Throughout the evolution of CS theory and principles, numerous solution methodologies have been proposed that are mainly divided in two subcategories. The first involves the replacement of parameter p value from zero to one, also termed as l_1 minimization [33] or linear programming. This method has been proven to be accurate but is characterized as computationally expensive. The antipode of the above methodologies are iterative algorithms known as greedy algorithms [34],[35] which are much faster but suffer from inferior solution accuracy. Thus, the former solution relates to linear optimization and also complies to the property of sparsifying matrix and measurement matrix incoherence so as to approach an optimal solution. Thus, considering the CS problem in an algorithmic sense, there exists a quantity to be minimized accompanied by a quantity acting as constraint in order to narrow down feasible solutions. A straightforward claim is whether this formulation can be considered not separately but in a combined manner where each quantity is multiplied by a relevant weight component. Such expressions jointly utilize norms with different p parameter values for each weighted term and are termed as Lagrange multipliers. Together with the property that different p values lead to differentiated geometrical models, the randomness related statistical distributions pose insightful modeling approaches and also provide feasible asymptotic evaluations by making use of CLT in

a Gaussian distributed assumption. Moreover, Banach or Hilbert spaces intimately relate to CS theory in terms of defining spaces equipped with l_p norms of inner product, the latter being significant in CS theory as low coherence involves a necessary calculation step in order to assess the optimality of results reached by CS methodology. Alternatively, the greedy algorithms select one element in each iteration from the vector space in which the signal resides. It must also be noted that CS theory has been characterized by the property of non-adaptivity hence the algorithms designed are universal in a sense that they apply to diverse sparsity scenarios. The intuitive query of whether this characteristic is due to the limited evolution of CS theory hence adaptive methodologies will be devised with improved performance characteristics constitutes the most crucial challenge to be investigated.

As already stated, CS utilizes random measurement matrices i.e. matrices whose elements are randomly valued variables modeled by specific statistical distributions. A valuable property of such matrices is enclosed in the following statement which is closely related to CS solution methodology: if according to sparsity concept a subset of the columns of a random matrix are randomly selected these columns will be linearly independent. Also, if the RIP i.e. the multiplication of measurement matrix with the sparse vector does not significantly affect the sparse vector then solution obtained is optimal. However, RIP requires extensive computations in order to verify whether it holds in a specific setting. Another issue to be accounted for is the case where problem parameters share common properties which permit joint processing [36], a case proven to be beneficial for the CS framework. In the latter context, the grouping or clustering schemes could significantly reduce computational burden and problem dimensionality thus direct problem solvability exactly in the CS methodology orientation aiding useful information extraction. Joint sparsity of two random variables could also provide substantial aid to selecting the transformation sparsifying matrix as well as ensuring the incoherence property required by CS theory. Hence, combinations of random matrices and sufficient number of elements from each base in order to acquire a complete representation of the signal preserving its information content poses a challenging necessity that is also described by information theory emigrating to wireless communication scientific research. Furthermore, the representation of a signal by a set of not necessarily orthogonal bases must be conducted in a manner that the contribution of each base in the resulting representation must be clearly depicted besides the completeness of the expression. Hence, in a joint scenario, CS transformation matrices must be selected in the above manner including all related problem parameters. Introducing the concept of approximating sparse vector by the concept of compressibility [37], which dictates preserving subset of largest in magnitude elements given a certain threshold, the issue of optimal representation is of major concern in signal reconstruction at the receiver side of a communication system. As optimization problems in wireless communication literature are characterized by increased dimensionality, the joint or independent consideration of the parameters involved must be carefully conducted. As such, low dimensionality projection and information content retrieval at the expense of a tolerable information loss is of utmost importance in such optimization problems.

Returning to the initial problem formulation of recovering a signal of large dimensionality from acquired measurements whose cardinality is less than the signal ambient dimension via l_1 norm, linearity property is verified along with reduced complexity. However, there are crucial requirements that must be fulfilled in order for this problem transformation to provide the optimal tailored to the specific scenario solution. This is the very RIP verification test described above with its accompanying need for computational excessive burden. Another mathematical test is termed as NSP. In the following paragraphs, the two latter tests will be briefly analyzed.

NSP [38] is a test that quantifies the similarity, via inner product operation, of two distinct columns of a CS sensing/measurement matrix. Given that the columns of such matrices must be in a geometrical sense different, the value of this tests must produce smallest possible values. In turn, even if NSP portrays how «optimal» our solution is, it still exhibits low resilience to noise leading to measurement distortion as in a more practical case.

RIP [39] can be effectively applied in noisy scenarios and confirms that the effect of sparse vector transformation via measurement matrix induces a bounded effect on the norm of the sparse vector to be recovered alone. The constant involved must be contained in the [0,1] range and the smaller its value the more bounded the effect of the vector transformation is. It must also be noted that the l_2 norm is used in the related mathematical expression of the RIP. For clarity, the equation modeling RIP test is given below:

$$(1-\delta)\|x\|_2 \leq \|A^* x\|_2 \leq (1+\delta)\|x\|_2 \qquad (2)$$

For a concise reference to the two categories of methodologies for solving CS optimization problems, greedy algorithmic approaches demonstrate many versions each with different convergence guarantees and speed. One or multiple elements may be selected in each iteration along with residual formed from the previous iterations. One very specific class of algorithms are learning algorithms which as they are executed learn the basis consistent with the problem. Moreover, the properties characterizing each algorithm in general also apply in greedy algorithms such as complexity or equivalently the determination of the calculation imposing the dominant computational burden and algorithm termination criterion for example maximum number of iterations. These are shaped depending on the problem and the degrees of freedom to be exploited for optimal resource efficient solution derivations. It is also imperative to note that the objective functions and corresponding constraints may also appear multiplied by weighting quantities which could alter the very convergence or solvability of the optimization problem and also demonstrate practical tradeoffs in their algorithmic context.

CS sparsifying matrices may be known or in a more practical sense learned [40],[41] during execution of algorithms. Hence, it is usually the case that sparsifying matrix is mostly characterized to belong to a countable set of candidate basis whereas the measurement matrix applied to the problem once the sparsity domain has been specified is drawn from random distributions ensemble. The crucial point that CS requires to emerge as a tool for low complexity underdetermined problem solving is the low

similarity of the two above matrices i.e. sparsifying and measurement matrix. Thus, the design for the latter could not only assume statistical independence but also measurement history to be exploited or correlation that better fit into scenarios for wireless communication signal reconstruction receiver side in a closed loop feedback assisted setting. To illustrate the latter, constraints of the optimization problem not only include past measurement history but also dictate the proper exploitation of rate of feedback information so as not to become outdated i.e. not constructive for signal detection and reconstruction.

Summarizing the main ingredients of CS are briefly mentioned in the following:

a) CS consists of a family of optimization tools aiming at alleviating computational and implementation complexity in a manner of obtaining an near optimal solution accounting for the sparsity constraint or equivalently in non-convex setting the compressibility constraint both of the latter implying the beneficial existence of structure within the data so as to avoid the conventional resource costly data sampling and discarding the latter being replaced by simultaneous data sampling and compression drastically reducing the required computational overhead. Key problem parameters are identified and l_p norm values for CS problem formulation balance problem solution characteristics such as solution uniqueness and tractability.
b) Relative to wireless communications but also in the areas of signal processing and information theory, CS exploits this convenient structure or patterns to ease computations but also provide performance benchmarks for example in practical communication scenarios which may promote thorough understanding of channel characteristics, resource allocation and impact of practical constraints on the algorithm employed. In a clarifying comment, CS may not only be applied in current 5G networks but also contribute to revisiting methods addressing problems of previous communication systems generations and also provide insights as to how to optimally interpret and confront future network challenges and shortcomings in a way to promote 6G and beyond communication systems.
c) The claimed alleviation of computational complexities in CS theory surpassing traditional sampling theorems and data processing methods especially addressing exponentially increasing complexity combinatorial problems by narrowing down feasible solutions is aimed at preserving information content and orienting sampling to the dimensionality of the actual information content rather than ambient dimensions. Moreover, the tradeoff between randomness and structure is what bears all the advantages and shortcomings of CS applicability to realistic problem. Hence randomness promotes efficient mathematical modeling with statistical tools while structure is what explicitly defines implementation feasibility in wireless communication system design. The above two aspects coexist in every scenario and define the extent to which optimization tools actually accomplish the desired goals. In a similar context, the latter determine the objective functions and constraints included in the design of effective algorithms that model practical problems.

Finally, we provide a brief notion on integration of CS in information theory. CS embodiment in information theory already counts many decades as the aspect of sparsity was already present before the establishment of CS in data compression and coding [42],[43]. As coding addresses the problem of adding redundancy to the transmitted data sequence to combat channel impairments and minimize distortion in a diverse tradeoffs context, CS promises tremendous benefits in effective data compression while quantifying information content via entropy and uncertainty metrics. As measurements are essentially bits, the lower bounds for required measurements and for successful reconstruction [44] are derived. Relative to acceptable signal recovery, quantifying distortion at target SNR also provides performance limitations from an information theoretic perspective.

2.2 Mathematical Prerequisites

In this section the commonly used mathematical formulae throughout this thesis are provided along with brief comments on their interpretations.

Firstly, we derive the statistical distributions describing each fading channel along with the additive noise distribution. Throughout the thesis, second order statistics are assumed thus mean and variances are derived. Thus, although Gaussian symmetric distribution is fully described by such assumption, the skewed fading distributions are partially modeled, in a manner sufficient for the analyses conducted. The Rayleigh fading distribution is given by the expression below followed by the mean and variance formulae:

$$f_x(x) = \frac{x}{sigma^2} e^{-\frac{x^2}{2sigma^2}} \tag{3}$$

$$\mu = sigma\sqrt{\frac{\pi}{2}} \tag{4}$$

$$\sigma^2 = sigma^2\left(\frac{4-\pi}{2}\right) \tag{5}$$

Thus, the distribution is entirely defined by value selection of parameter sigma which also determines the mean and variance quantities. Proceeding to the Rician fading distribution, the distribution expression along with mean and variances are provided below:

$$f_x(x) = \frac{x}{sigma^2} e^{-\frac{\left(x^2+v^2\right)}{2sigma^2}} \mathrm{I}_0\left(\frac{xv}{sigma^2}\right) \tag{6}$$

$$\mu = sigma\sqrt{\frac{\pi}{2}} * L_{1/2}\left(-\frac{v^2}{2sigma^2}\right) \tag{7}$$

$$\sigma^2 = 2sigma^2 + v^2 - \frac{\pi sigma^2}{2} * L^2_{1/2}\left(-\frac{v^2}{2sigma^2}\right) \tag{8}$$

Hence, Rician fading distribution is defined by two parameters namely parameter v and sigma. Both are required to adequately describe the distribution. The distribution expression contains Bessel function of first kind and zero order, while mean and variance expression involve Laguerre polynomial which can be quantified by the hypergeometric function of first kind.

Proceeding to the Nakagami-m fading distribution, the distribution and mean and variance expressions are given below:

$$f_x(x) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} e^{-\frac{mx^2}{\Omega}} \tag{9}$$

$$mean = \frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{\frac{1}{2}} \tag{10}$$

$$\sigma^2 = \Omega\left(1 - \frac{1}{m}\left(\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\right)^2\right) \tag{11}$$

As is apparent, Nakagami-m fading distribution is also described by two parameters namely m the shape parameter and Ω the spread parameter. It can be observed that Γ(m) is the gamma function appearing in all three expressions. Moreover, the case m=0.5 results in an expression that resembles the Gaussian function whereas m=1 results in an expression close to the Rayleigh fading distribution.

The Gaussian distributed being the most fundamental distribution in various disciplines and also in wireless communication is a symmetric distribution and as such completely described by second order statistics i.e. mean and variance. In the context of this thesis, Gaussian distribution is used to model additive noise as a practical inherent property of the wireless channel as well as CLT distribution. The distribution expression is given below. Note that, unlike the fading distributions, the Gaussian distribution expression

includes mean and variance instead of parameters of fading distributions that separately define distribution along with mean and variance.

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{12}$$

Proceeding further, the statistical moments given a specific distribution of nth order are given by the integral below:

$$E[x^n] = \int_{-\infty}^{+\infty} x^n * f_x(x) dx \tag{13}$$

Resulting from the above formula the variance can be computed as the difference between the second order statistical moment i.e. n=2 in the above formula and the square of the first order statistical moment i.e. derived by n=1 in the formula above:

$$\sigma^2 = \mathrm{E}[x^2] - (\mathrm{E}[x])^2 \tag{14}$$

We now proceed to information theoretic formulae, Shannon entropy, relative entropy and channel capacity. The first quantifies the uncertainty within an outcome of a random variable and is a strictly positive quantity by minus sign insertion as given below:

$$H(f_x) = -\sum_x f_x \log_2(f_x) \tag{15}$$

The relative entropy involves two distributions, the accurate distribution and the approximating one in a practical scenario of partial distribution knowledge. It is thus a measure of dissimilarity between the two distributions namely the accurate f_x and the approximating p_x. In other words, it quantifies the distance between the former and latter distribution. The respective formula is given below:

$$D(f_x \parallel p_x) = \sum_x f_x \log_2\left(\frac{f_x}{p_x}\right) \tag{16}$$

The sum of the Shannon entropy and relative entropy involving the distributions modeling the wireless fading channel expresses the average number of bits required to describe the channel. The first term, Shannon entropy, represents the bits due to the true distribution while the second term, relative entropy, express the redundant bits as a consequence of approaching the true distribution with the approximating one.

Finally, the Shannon capacity is given by the formula below including the squared channel gain to account for fading channel cases:

$$C = BW * \log_2\left(1 + \frac{|h|^2 P}{N_0 BW}\right) \qquad (17)$$

In this formula, the bandwidth of the channel appears both as a linear term and as a quantity inside the logarithm. The squared channel gain is multiplied by the transmit power and divided by the product of the bandwidth and the noise power spectral density.

Moving on to CS theory, the Gaussian case effectively quantifies the number of measurements much less than the initial dimensionality of the information signal that ensure signal reconstruction in this subcase of CS regime. The related formula is given below where N is the initial number of elements in the signal vector and k is the number of nonzero elements in the former vector. Hence, k divided by N is termed as the sparsity ratio:

$$M = k * \log_2\left(\frac{N}{k}\right) \qquad (18)$$

We conclude the mathematical preliminary section with the initial CS problem formulation using the l_0 norm as well as the convex relaxed version of this problem in both noise free and noisy cases with minimization of the l_1 norm. These problem formulations are termed Basis Pursuit (BP) and Basis Pursuit Denoising (BPDN), respectively. It must be stressed that although this problem forms are not addressed throughout this thesis, they are noted in order to give prominence to the derivation of the distribution of the sum of independent random variables via the convolution of the distributions of the variables that comprise the sum contrary to the investigation of a correlation based analysis of the expression deriving the channel output as the sum of the input weighted by the fading distribution channel gain plus the additive noise inserted as an inherent property of the channel. This expression is given below:

$$y = h * s + n \qquad (19)$$

The initial CS problem formulation via l_0 norm expression is given in the following:

$$\arg\min \|x\|_0 \quad \text{s.t.} \quad Y = A * x \qquad (20)$$

The l_1 minimization problem expression is given below for the noiseless and additive noisy cases i.e. BP and BPDN respectively:

$$\arg\min \|x\|_1 \quad \text{s.t.} \quad Y = A * x \tag{21}$$

$$\arg\min \|x\|_1 \quad \text{s.t.} \quad \|Y - A * x\|_2 \leq e \tag{22}$$

BPDN has the modified constraint of bounded noise via the l_2 norm expression while still preserving the function to be minimized as in the noise free case. Towards a clarifying comment, the statistical convolution derives the distribution of y solely based on the statistical independence assumption thus bypassing the consideration of the above Eq.(19). A concluding comment is that there is no notion of a specific sparsifying matrix. Thus, in the analysis of this thesis the sparsifying matrix is equal in a matrix multiplication sense, to the identity matrix. In other words, the signal is already sparse and the application of the measurement matrix suffices.

3. COMPRESSED SENSING IN WIRELESS COMMUNICATION SYSTEMS

CS promotes improved performance in sampling and signal acquisition that is solely based on certain structure properties, specifically sparsity or compressibility, relying on random matrix theory to validate its potential for optimal signal recovery applicable also in wireless communication [45],[46],[47] performance assessment. The mere requirements of CS as a principle is the existence of a sparsity domain reached by means of a non-identity matrix multiplication in a general case when such transformation is needed by a measurement matrix which is randomized following a certain statistical distribution. In this context, Gaussian is the most favorable distribution directly qualifying for measurement sparsification with a known formula of adequate number of measurements for successful reconstruction.

Relative to the core contributions in this chapter, the channel estimation issue [48],[49] translated in our case as distribution modeling of fading channels can be effectively promoted by sparsity or equivalently compressibility as in the subsequent analysis. Under this constraint, system performance can be stated in an accurate context in terms of many practical assumptions such as dominant cluster multipath model even in a NLOS scenario implying dominant scatter component and multiple antennas implementation i.e. trading off hardware complexity with spatial diversity exploitation.

Proceeding further, CR based spectrum sensing can successfully incorporate CS as sparsity relates to underutilized spectrum in other words small portion of the licensed bandwidth instantaneously utilized. Finally, in densely deployed networks, correlation in time and space can be efficiently exploited along with sparsity scenario to reduce overhead and promote energy efficiency and reconstruction quality either in a centralized or decentralized case. Concluding this short reference, CS is already proven to guarantee performance gains by designing effective algorithms with diverse convergence, speed and reconstruction accuracy criteria. However, the challenges of unknown or rapidly varying sparsity or model mismatch in terms of learned sparsifying dictionary also translate to practical constraints in the wireless system performance formulating interesting problems at a post CS applicability verification stage where many of these problems admit further investigation.

3.1 Chapter outline

This chapter investigates the application of CS to fading channels combined with fading channel coding which quantifies the average number of information bits required to describe the wireless channel. Moreover, statistical CS is employed the latter characterization stemming from applying a method of minimizing l1 norm, which corresponds to the convex relaxation of the initially intractable l_0 norm CS recovery problem. This method effectively tackles the much more excessive method of conventional l_1 norm minimization but also provides the optimal solution by means of first and second order moment statistics. The fading channel distributions considered are Rayleigh fading, Rician Fading and Nakagami fading distributions as given in Eqs. (3),(6) and (9) respectively. The common distribution modeling additive noise is the Gaussian distribution. The performance metrics are channel capacity, equivalent channel distribution variance, average code description length based on Shannon

entropy and relative entropy as well as symbol error probability. The distribution inference problem is either formulated as an inverse problem by using channel coding and CS based distribution reconstruction results in the first section and by CS compressibility rule and Taylor approximation in the second section. Moreover, the analysis reviews the spectrum sensing methodologies in CR systems with an algorithmic flavor along with main consideration of sub-Nyquist CS techniques. A convolutional probabilistic ED based scheme including the CS, CLT [50] and Rayleigh, Rician and Nakagami-m fading cases is proposed. Finally, a probabilistic scheme assessing WSN network performance in terms of reconstruction error and energy estimation is adopted. Gaussian statistics are the main distribution for all performance derivations of this section. The results are directly applicable to promote network effective topology and routing.

3.2. Fading channel coding and distribution estimation with Compressed Sensing

3.2.1 Introduction

The results of this section are related to our publication [C6]-([51]) utilizing CS for assessing samples required for statistical density reconstruction utilizing Rayleigh and Rician fading distributions. The required number of samples for each probability density function considered for its reconstruction are derived by a search for minimal l_1 norm optimization problem that bypasses the inherent complexity of the traditional l_1 minimization method but is also applicable due to distribution convexity. The latter is verified by the asymmetric bell curved distributions which do not exhibit any other extremum but decay to zero as the abscissa increases towards infinity. The number of required bits is equal to the sum of Shannon entropy and relative entropy. Shannon entropy expresses the number of bits used to describe the channel according to the distribution that it follows. Relative entropy is considered when the accurate distribution is approximated by a different one. Hence, it quantifies the dissimilarity or distance between the accurate and approximating distribution in terms of additional number of bits required to describe the wireless channel. The structure of the proposed method is the following: the pairs of CLT related distribution and Rayleigh/Rician fading distributions are used to derive the maximum number of samples for which finite values of entropy were observed are fed into the CS based l_1 optimization problem for the required number of samples of probability density reconstruction. Thus, for each distribution pair the required number of samples is derived which may refer to the same distribution in different pairs but differs in terms of values as each case accepts a different entropy as the problem input. The CLT case models the approximation of a Gaussian distribution of the fading distribution that is defined by a sum of 10 independent fading variables either of the fading noiseless case or as the sum of 10 variables following a distribution equal to the convolution of the fading distribution and Gaussian additive noise distribution. The results on required average number of bits are formed by pairs of CLT related distribution and all four distribution Rayleigh and Rician both noiseless and noisy cases included as well four additional cases excluding the CLT case and forming the respective pairs of the latter distributions. The results are fed into the CS based required number of samples, a problem that is stated by a Lagrangian optimization expression that derives the minimum l_1 norm and computes

the required number of samples based on the two initial statistical moment problem formulated by deriving the extremum of an equivalent function. The inverse problem of distribution identification is thus addressed deciding in favor of either NLOS or LOS component existence.

The above constitutes a requirement for the random channel to be modeled with finite uncertainty and also directly relates to the maximum entropy problem, which has proven to be greatly contributive for channel modeling from an information theoretic point of view. The remarkable merit of such an approach is that it provides consistent models based on the concept of uncertainty as our prior maximized information. It is also highly relevant to Bayesian inference which in this section is confined to the distinction between Rayleigh and Rician fading, as stated.

3.2.2. Past related work

Relative to past related literature, [52] and [53] investigate the statistical properties of MIMO fading channels providing a restricted version of the ML method for deriving mean and covariance matrices in a computationally efficient way. The main mathematically translated differentiation of the above approach is naturally the absence of nonzero quantities of covariance since the assumption of fading and noise independence is what the analysis of this section is based on. Furthermore, the discrete problem of block fading channel is addressed along with continuous fading assumption. Although the notion of entropy value derivation comes along in a discrete form but the fading distributions support the continuous assumption, the essential assumption for the analysis of this section is the channel coherence assumption. This emigrates to the fact that the distribution knowledge remains valid, leaving as a future extension of this work the issue of coherence in terms of channel gain values but also the noncoherent scenario where the problem of outdated channel estimation measurements emerges. Debbah and Muller in [54] and Biglieri et al. in [55] address the maximum entropy problem as a way of modeling the MIMO channel with the only constraint that of finite energy. It is quite interesting to relate these works with the analysis of the current section which is based on entropy as the information content. The latter is straightforward applied to the fading case with the additional restriction of Rayleigh and Rician fading distributions which is the element that renders this analysis of high merit as the NLOS and LOS are assumed. This conveys that the channel is accurately modeled as the distinction is based on two opposite scenarios. Hence, there is no derivation as to how severe fading the distribution implies but to whether a LOS component exists, which is different to the complementary Rayleigh fading assumption. These papers also highlight that the Gaussian assumption valid for modeling the channel is a consequence of considering the two first statistical moments i.e. second order statistics, commenting therefore that if additional moments are available, channel matrix could be modeled otherwise. The crucial conclusion reached in this paper, however, is that incorporating more constraints for channel modeling optimization problem by means of maximum entropy could lead to a suboptimal channel description. Thus, the main direction of this thesis being the integration of CS with the entropy problem the challenge of whether CS combined with entropy could guarantee optimality for channel modeling is stated as an extension of the latter work. In [56] and [57], the relative entropy is calculated via the integral of the quantity of the difference of the linear mean squared error from the nonlinear minimum

mean squared error induced by the penalty of adopting linearity. The latter statement could offer an intriguing extension of the derivation of CS number of samples for distribution reconstruction in terms of the reconstruction error i.e. CS reconstruction error of the respective distribution. Moreover, these works are closely related to our analysis as they consider non-Gaussian distributions, the fading distributions in our case and Gaussian contamination, which translates to the standard Gaussian noise distribution considered that formulates the noisy cases of the analysis. It is important to note that the relative entropy and mean square error related derivations refer to the output that results in the sum of the convolved channel and transmitted signal quantities with the Gaussian additive noise addition. Hence, the results are based on the above standard expression whereas our analysis is based on the convolved channel and noise distributions as stated in the mathematical preliminaries section always in the context of independence. Moreover, the finite second order moment is consistent in our analysis compared to [56] whereas it is stated that, being essential for deriving the mean squared error formula, a one-to-one mapping is necessary as a transformation for the finite second moment assumption to hold. Aligned with the analysis of this thesis, the noise is considered independent between two different values without any assumption of memory or stationarity as well as independent to the channel realizations that are either modeled as noiseless or noisy fading distributions. The above stem from the strict notion of independence. It is also noted that the quantification or required number of bits derived by the sum of Shannon entropy is also related to how data compression deviates above the minimum entropy related value and as a consequence could be expressed as the difference on mean and minimum mean square error. Proceeding further to [58] by Kontoyannis and Verdu, the case of an optimal compressor is investigated by the fact that it is transparent to the achievable limits it must satisfy as well as the prefix condition provided the decoder does not need to know where and when a compressed file starts and ends. In accordance to the analysis of this thesis, memory is not exploited in the optimization compression problem, however, the proposed Huffman coding scheme is in agreement with the assumption of this section as the channel statistics are known to derive the average code length required for each distribution case but is in contrast to the inverse problem in which with the aid of CS based required number of samples for reconstruction, the channel distribution is inferred, or more accurately decided upon by means of the results of the previous section. Relative to investigating the optimal distribution of the symbol length, the query of whether in our case this length is known at the decompressor we argue that in order for the inverse problem to admit a correct solution, each distribution case, fading noiseless or noisy, must be related to the symbol length permitting its identification. In light of a theorem of this work, the extension of an assumption regarding a small alphabet of a random variable packed together with the exponential increase of alphabet for fixed-to-variable case to the capability of CS-based sparse alphabet in the context of the Rayleigh/Rician fading assumption is considered challenging. Concerning the asymptotic normality of optimal coding length, the authors bring forth the argument of excluding the prefix constraint in order to tackle the large deviation behavior. Finally, the determination of optimal code length statistics is another route to take in our work in conjunction with the Central Limit Theorem as a comparison distribution in our analysis that relates to the asymptotic Gaussian approximation of the latter work.

Another work [59] closely related to our analysis but assuming a finite state memory scenario investigates the normalized empirical expectation code length as opposed to the probabilistic-wise entropy rate, the latter matching our proposed approach with the difference that this work deals with source coding as opposed in the consideration of channel coding in our case. Apart from the relative entropy which is the only individual sequence analogue in our case, the authors in [59] seek a deterministic analogue of the Renyi entropy, which although intimately related to the standard Shannon entropy used in our case, is investigated separately. Furthermore, the relation of the commented length fluctuations of source blocks are generally not ruled out in our case since the derivations concern the average channel code length. The influence of certain correlated side information content on the encoded string length can be thought as additional constraint to the compression optimization problem. Therefore, the connection to our analysis deriving average code description length for channel coding could be the relative entropy as a measure of redundant bits resulting from partial distribution knowledge.

3.2.3 CS based distribution reconstruction

The approach adopted relies on the fundamental equation that computes the variance of a distribution by subtracting the square of the first order moment, i.e. the distribution mean from the second order moment as in Eq.(14). Proceeding with the result derivation the second order moment is equal to the sum of squares divided by the number of samples N i.e. the squared l2 norm divided by N. The squared first order moment is equal to the squared l1 norm divided by the square of the number of samples i.e. N^2. From the above and solving for l1 the following inequality is derived:

$$l_1 \geq \sqrt{N l_2^{\,2} - N^2 \sigma^2} \tag{23}$$

The minimum number of the l1 norm is thus obvious from the above equation. To proceed with the derivation of the minimum the partial derivatives are computed with respect to variables: N, the number of samples and l2 the equivalent norm. The derivation produces the following equations:

$$\frac{\partial l_1}{\partial N} = \frac{\left(l_2^{\,2} - 2N\sigma^2\right)}{2\sqrt{\mathrm{N} l_2^{\,2} - N^2\sigma^2}} \tag{24}$$

$$\frac{\partial l_1}{\partial l_2} = \frac{N l_2}{\sqrt{\mathrm{N} l_2^{\,2} - N^2\sigma^2}} \tag{25}$$

According to fundamental theorem of calculus, the derivatives need to be set to zero. This results in the following solution:

$$l_2^{\,2} = 2N\sigma^2 \tag{26}$$

from the first derivative. The result also complies with the positivity of the quantity inside the root expression that must be verified. The second derivative does not

contribute to the system solution. Hence, the minimum value of l1 norm based on this condition is equal to the following quantity:

$$l_1 = N\sigma \tag{27}$$

Hence, the primary quantity evaluated is the l_2 norm, which by CS theory has non-unique solutions as depicted in the equivalent spherical geometry. The l_1 norm is thus uniquely quantified in the second step as also dictated by CS principle. Based on the above, the following optimization problem is formulated with the assigned weights as Lagrange multipliers. The complete statement of the optimization problem is given below:

$$L = l_1 + \lambda_2\left(\int_{-\infty}^{N} x f_x(x)dx - \mu\right) + \lambda_3\left(\int_{-\infty}^{N} x^2 f_x dx - \mu^2 - \sigma^2\right) \tag{28}$$

After some algebraic manipulations the following equations are formed:

$$-\lambda_2 - 2\mu\lambda_3 = 0 \tag{29}$$

$$\mathrm{N} - \lambda_2 + 2\sigma\lambda_3 = 0 \tag{30}$$

$$\mathrm{N} = -2(\mu + \sigma)\lambda_3 \tag{31}$$

Based on the number of finite entropy and parameter values for μ and σ, multiplier λ_3 is evaluated, being positive according to Eq. (31) and from Eq. (29) parameter λ_2 is evaluated. Finally, the l1 norm quantity is evaluated which enables the derivation of the number of required samples for CS-based distribution reconstruction. It should also be noted that the integrals expressions are considered with the entropy number of samples which verifies the nonzero value of the weighted expressions in the Lagrange optimization function.

The above optimization problem results in the data of number of samples for which finite entropy was observed and two values representing the CS number of samples for density reconstruction in each of the respective pairs formed by the assumption of the accurate and approximating distribution for the relative entropy expression.

3.2.4 Algorithm formulation

The related algorithmic formulation of the above stated problem is provided below.

Algorithm

1. **Input**: Rayleigh-Rician fading noiseless and noisy, mean and variance of distributions, CLT Gaussian distribution, Shannon/relative entropies, channel code length, l_1 minimization inequality.
2. Form pairs of distributions for Shannon/relative entropies
3. Calculate Shannon Entropies for assumed distributions and relative entropies for all distribution pairs.

4. Calculate channel code length based on derived Shannon and relative entropies as the sum of bits with respect to each term.
5. Calculate number of samples for each distribution reconstruction based on CS theory and the above pairs and entropies related numbers of samples.
6. Based on channel code lengths and CS based number of samples for distribution reconstruction identify whether the distribution is Rayleigh or Rician.
7. **Output:** Channel code lengths for each distribution pair, CS-based number of samples, identification of distribution as Rayleigh or Rician.

3.2.5 Simulation results

Relative to second-order statistics, the following Table 3.1 lists the means and variances of each distribution forming the pairs considered for the relative entropies derivations calculated from the related contribution parameters. The ten times multiple mean of the equivalent Gaussian distribution is the generic property of the mean of a sum or random variables while the 10 times multiple variance is a consequence of the fact that the sum comprises of independent variables, hence the parameters of the approximating Gaussian distribution.

TABLE 3.1: Distributions Statistical Parameters

Distributions	Mean μ	Variance σ^2
Rayleigh noiseless	1.25	0.43
Rayleigh noisy	1.25	1.43
Rician noiseless	1.54	0.61
Rician noisy	1.54	1.61
Gaussian noise	0	1
CLT Gaussian	10μ	$10\sigma^2$

Along with number of bits required evaluated by the Shannon and relative entropy in Eq.(15) and Eq. (16), respectively, the sum referred to as channel code length, the values of number of samples for finite entropy were observed and CS based required number of samples for each distribution reconstruction given in Eq. (27) for all pairs considered, the results are summarized in the following Table 3.2:

TABLE 3.2: Maximum samples for finite entropy, CS number of samples for distribution reconstruction and average code lengths

Distribution pairs	N_{max} samples for finite entropy	N_1 samples for reconstruction of first distribution in pair	N_2 samples for reconstruction of second distribution in pair	Average code lengths
CLT/Rayleigh noiseless	14	27	27	4
CLT/Rayleigh noisy	13	15	16	2

Distribution pairs	N_{max} samples for finite entropy	N_1 samples for reconstruction of first distribution in pair	N_2 samples for reconstruction of second distribution in pair	Average code lengths
CLT/Rician noiseless	17	38	45	26
CLT/Rician noisy	16	22	23	3
Rayleigh noiseless/Rayleigh noisy	5	1	1	4
Rician noiseless/Rician noisy	2	1	1	2
Rayleigh noiseless/Rician noiseless	5	1	2	6
Rayleigh noisy/Rician noisy	6	1	1	2

The first observation requiring an accurate interpretation are the significantly less bits required for CS reconstruction for the Rayleigh and Rician noisy cases paired with the equivalent CLT distribution compared to the noiseless cases. Though contrary to the fact that considering noise ought to increase number of required bits, the interpretation is straightforward. Hence, this stems from the independence assumption between fading and noise distributions, which is impractical as correlation always exists. Moreover, the general observation of Rician fading paired with CLT distribution requiring additional bits from the rest of the cases is verified not only in terms of CS based samples for reconstruction but also for the value up to which finite entropy was observed. In the same context, the approximation of Rayleigh fading by Rician fading requires same number of samples for Rayleigh reconstruction but one extra sample for the Rician distribution reconstruction. The opposite i.e one extra sample holds for the number of maximum finite entropy based values. These results are directly justified from the modified Bessel function term in the mathematical expression of the Rician fading distribution.

Relative to the decimal values of Shannon and Relative entropies and with respect to CLT cases, noiseless Rayleigh conveys larger Shannon entropy and less relative entropy while for the additive noise case the opposite with larger relative entropy was observed. This translates that the uncertainty of the CLT related approximation has more uncertainty when related to the Rayleigh approximation than when related to Rician fading, which can be attributed to the contrast between NLOS and LOS component. Noisy cases have less uncertainty than noiseless cases. Moreover, approximation of Rayleigh noiseless by Rayleigh noisy has more uncertainty than approximation of Rician noiseless by Rician noisy. Hence, the effect of NLOS conditions resulting in greater uncertainty as opposed to Rician LOS is apparent in this case where the Rayleigh and Rician are paired instead of the CLT based cases.

Specifically, in CS number of samples the number of samples for the accurate and approximating distributions seem to be in the same magnitude for all cases except the Rician fading case with the CLT as the accurate distribution. The results express the greater number of samples for reconstructing the Rician distribution compared to its CLT case. Another observation is the reduced number of samples for the Rayleigh and Rician cases compared to the cases where CLT was considered. As a complementary justification for the difference in number of samples in the aforementioned divided groups of results, the input to the CS derivations are the different number of bits related to finite entropy values. Hence, even though the same distributions are considered the results are different and dependent on the approximating distributions by means of the relative entropy calculations.

Finally, the channel code length derivations result in useful observations. The average code lengths for the case quantifying the distance of the CLT related distribution and Rician noiseless distribution requires the greatest average code length. The observation of noisy case in CLT pairs requiring fewer bits is also confirmed with the case of CLT and Rician noisy as opposed to Rician noiseless pair requiring far fewer bits for channel coding. The case of approximating Rayleigh noiseless distribution by Rayleigh noisy requires two less bits than the approximation of the same with Rician noiseless. This actually states that the noisy Rayleigh case is a lower complexity approximation compared to Rician noiseless fading case the latter characterization being valid for any of the above code length comparisons.

As a concluding remark, the results of the simulations section are indicative of the distribution they result from as they are redundant in the sense that number of values for finite entropy, CS based required number of samples for distribution reconstruction and average code lengths support a certain decision or they are sufficiently diverse referring to the case conveying ambiguity with respect to a specific result, the other results provide distinction and enable a certain derivation. This latter observation supports the decision of Rayleigh or Rician including noiseless or noisy cases, which is the inverse problem addressed in the next subsection.

3.2.6 Channel distribution knowledge inverse problem

This section investigates the inverse problem i.e. channel distribution identification which subsequently results in channel knowledge. In the context of variable length channel coding and non-varying or slowly varying channel distribution the results exhibit structure enabling solution to the inverse problem.

Firstly, the significantly greater number of bits for finite entropy along with the CS number of samples with the CLT consideration regarding Rician fading can be made accurately by such an observation. Rician noiseless and noisy fading can also be identified from the Rayleigh cases with greater bits and samples values. The identification of Rician noiseless instead of Rayleigh noisy, given that the approximating distribution is Rician noisy can be made by the different in number of entropy based bits. The decision over noiseless fading compared to noisy fading can be made by the noticeably smaller values not only for the number of bits up to which finite entropy values were observed but also for the CS based number of samples for density

function reconstruction for all fading distribution cases. The identification of Rayleigh noiseless approximated by its noisy counterpart from Rician noiseless case approximated by its noisy counterpart is also feasible from the entropy bits as well as derived code length. Finally, the mean and variances can be the initial values that can be solved for the specific parameters of each distribution thus leading to their complete mathematical formulation. From the initial point of view that only Rayleigh and Rician fading distributions are assumed, the channel estimation related application of the above analysis is furthermore beneficial. Considering that Rayleigh fading introduces severe distortion and multiple scatterers and Rician involves a dominant LOS component, the wireless channel properties are more accurate. Moreover, the Gaussian statistics of in-phase and quadrature components having zero means for Rayleigh case and nonzero for Rician case could be accurately modeled with the additional knowledge of their variance. Finally, the exact distribution modeling the channel could be achieved. Starting from the known variance and with the formulas deriving mean and variance the required parameter could be achieved. The following describe the latter calculations.

In the Rayleigh fading case, the knowledge of the variance or mean lead directly to the calculation of sigma parameter. For the Rician fading case, however, the formulas deriving the mean and variance are interrelated, thus pointing out the requirement of numerical solution of the equations for calculating parameters sigma and v. Specifically, by inserting approximation of mean and variance values and Taylor expansion of Laguerre polynomial the iterative solution of the resulting polynomials could be leveraged to determine the desired parameters.

It is apparent that the approach relates the number of bits for channel coding as a measure of uncertainty while the CS based optimization relies on the reconstruction accuracy of the models of randomness i.e. the statistical distributions used in the former coding derivations. At a MIMO scenario of considerable complexity, the identification of channel fading distribution and the verification of whether additive noise must be included in the channel model is a promising application of such a combined coding and channel model reconstruction in 5G massive multi-antenna environments. Together with dominant cluster paths assumed as sparse, the CS optimization can further boost detection performance and reduce decoding complexity by means of available channel state information at transmitter and receiver. From a decision perspective, the choice of more than two distributions for identification leading to an optimization problem of multiple possible outcomes could be addressed by exploiting similar observations on the basis of structure of the problem. Methodologies based on CS for optimal solution compensating for ambiguities as in the current analysis where CS complements results from entropy calculations to reach accurate decision could be introduced. Optimal performance in terms of tradeoffs and reduced complexity may also be derived. These remarks set the stage for considering Rayleigh, Rician and Nakagami-m fading cases the latter being an approximation of the former two by proper parameter selection. This constitutes the background of the next section.

3.2.7 Conclusions and future work

This section has investigated a channel distribution estimation and CS based reconstruction in two separate stages. The first relied on entropy based derivations for Rayleigh and Rician fading distributions both with and without additive noise consideration relating to maximum entropy problem. The second stage used the aforementioned results as input to the CS distribution reconstruction property thus combining both stages in order to address the inverse channel distribution estimation problem for the above two distributions.

As future work, the results based on the derived measurements from a statistical independence estimation assumption can be extended to the correlated cases including additional fading distributions i.e a diverse fading environment together with CS optimization principles.

3.3 Wireless fading channel performance by means of Compressive Sensing and Taylor approximation

3.3.1 Introduction

The results in this section are related to our publication [J2]-([60]). As already stated, performance of wireless communication systems is mainly based on the concept of randomness through which the wireless channel is modeled. The notions of Shannon entropy and relative entropy quantify the number of bits to achieve channel coding i.e. efficiently describe the random channel. Thus, the statistical distributions used to model the wireless channel are characterized by their variance, common to all distributions, the latter being a measure of spread of random values with respect to the mean of the distribution. The above concept of uncertainty is also quantified by the Shannon entropy assuming the respective distribution is used, while relative entropy is utilized as a measure of dissimilarity when an approximating distribution with respect to the accurate is used. The resulting redundant bits are required for channel description.

In this section, the Rayleigh, Rician and Nakagami-m fading channels are considered in terms of performance metrics: capacity [61], variance estimation, required number of bits to describe the channel and symbol error probability. The aforementioned exact distributions are considered along with the inferred distribution of the same kind resulting from keeping only the largest in magnitude channel gains as dictated by the CS principle. This problem can be thought as a more generalized problem of identifying the exact distribution by a set of data samples i.e. the largest samples in the CS case. A more accurate remark on the problem formulated in this section is that the optimization problem has the specific two constraints that a certain given fraction of channel gains is preserved and that the optimal distribution is of the same kind as the initial one which was the initialization of this optimization problem. Moreover, the consideration of additive noise is also efficiently encapsulated in the probability density function that serves as the initial one. The mathematical manipulation is, as stated before, the same: the equivalent distribution of the case including additive Gaussian noise $N(0,1)$ i.e. modeled by standard Gaussian distribution, is the convolution of the fading distribution and additive Gaussian distribution, as dictated by independent fading channel and noise

distributions. Hence, the relation to the concept of structure involved in CS reconstruction method in terms of an initially intractable problem or a problem with infinite solution is evident by the verified results that the aforementioned constraints lead to the unique most optimal result. Furthermore, the Taylor approximation of the exact distribution expressions is considered. Such an approach by modeling fading distributions in an approximating manner not only effectively captures randomness as opposed to modeling results in term of produced realizations but also justifies validity of results in a manner of the respective distribution approximation. In order to achieve a representative approximation capturing curvature of the three distributions a second order Taylor polynomial approximation is applied. Complying to the aforementioned validity of result the method of formulating the Taylor polynomial with randomized coefficients is left as related future work extension. Hence, the exact, CS inferred and Taylor polynomial approximated expressions are used as channel gain generating numbers.

3.3.2 Past related work

A representative work [62] by Pramanik et.al referring to multi antenna systems at a mMIMO scale, a technology under the 5G communications evolution, proposes a channel estimation scheme in a CS framework promoted by the validated, for this case, concept of sparsity. In general, the consideration of increased overhead in a mMIMO system in terms of increased training for CSI acquisition, so as not to be outdated, in order to improve performance essentially leads to smaller fraction of information symbols transmitted which can translate to the findings of this section regarding average code description as a low rate code length. The zero additional bits required concerning the CS effect as will be derived in the simulation section stresses the necessity of designing high rate codes, which in the context of [62] renders sparsity of multipath as a promising way of increasing transmission rate with a lower training overhead, the channel estimation CS approach having witnessed significant progress already. The goal of channel estimation through LDPC coding is achieved by modified StOMP CS algorithm which falls under the category of CS greedy algorithm while our approach is l_1 norm minimization oriented. Since the analysis of this section is based on independence statistics in the derivation of the noisy fading channel matrix, the advantage of convergence speed of StOMP against precision of convex optimization-oriented method is left as an alternative route of solving the problem of fading channel estimation and optimized performance. Similar to our symbol error derivations for Rayleigh, Rician and Nakagami-m fading channels the improvement in BER is provided at an LDPC decoder with side information available. This a priori information serving as constraints for the optimization problem can either be translated as preserved number of channel gains due to CS compressibility rule or information about the extent of approximation achieved by the Taylor polynomial not only because of the requirement for confinement to positive valued integral but also depending on the polynomial degree considered. The formulation of channel estimation problem exploiting sparsity in matrix notation is conducted by means of FDD whereas the analysis of this thesis concerns temporal and spatial domains, in the wireless sensor network performance optimization section, leaving the case of frequency as future indicative research. The proposed channel estimation algorithm in the aforementioned

work exhibited improved performance over conventional methods in low SNR regime. Although the required SNR for the channel coding in our case is not mentioned, CS provides reduced complexity and symbol error which could lead to reduced transmission power and fewer retransmission thus improving SNR. A work not only related to the effect of precoding as a means to achieve interference mitigation by Taylor expansion but also addressing convergence and complexity is conducted in [63]. While our analysis is solely based on independence, the authors in this paper aim at selecting most relevant correlation terms optimizing MMSE. The essence of this work using Taylor expansion is to circumvent the issue of complexity arising as a result of matrix inversion related divisions. It is also argued that matrix multiplications favor hardware design efficiency. The diagonally dominant matrix structure preserving only largest magnitude can be directly connected to the CS compressibility rule employed in this section by a predefined threshold, with the difference in the already stated notion of independence. Offline computation method is the reason for negligible complexity regarding optimized coefficients derivation or small number of updating the latter values in the first data frame. Munkhammar et.al in [64] address the problem of distribution reconstruction, namely Weibull, and highlight the proposed polynomial approximation for use in distribution convolution which intimately relates to the statistical prerequisite of independence in our analysis as well as the derivation of channel distribution accounting for fading and additive noise. The concept of moment based distribution approximation closely relates to the previous section where the Lagrangian optimization expression includes the weighted first two moments, whereas this section compares the derived variance values based on CS and the Taylor polynomial approximation of the fading distributions. The steps followed for distribution estimation resemble the method adopted for this section in the following manners: second order statistics are available, the distribution problem specific interval is determined as the interval in which the Taylor approximation of the distributions investigated in this section remains positive. This stems from the degree of the Taylor polynomial, i.e. second degree as considered which in turn determines the achievable approximation. Moreover, the authors comment on the check of quality of convergence which in our case can be encompassed in the following statement: the goodness of fit is estimated indirectly by the capacity of the Taylor approximation, the calculated variances, the average channel code length based on Shannon and relative entropies as well as symbol error probabilities. Hence, the penalty resulting from the Taylor approximation with a fixed assumption of degree and expansion point constitutes a measure of goodness of fit. The suggestion of an approximating polynomial constructed as a product of a normal distribution and an Hermite-based polynomial could provide an extension of our work for fading distribution along with Gaussian additive noise consideration. Another crucial remark that could be related to the CS compressibility rule in a case of multiple extrema of a distribution i.e. in a bimodal or in general multimodal distribution mentioned in [64], the degree of, Taylor in our case, polynomial, along with a suitable expansion point, could be reduced up to a maximum value so as to restrict the approximating property in a certain interval, isolating to the maximum extent, the desired extremum. In this context, Taylor approximation could provide a better fit compared to CS inference. The CS related inferred distribution derived in this section could also be applied to this case which states an interesting

research extension. A nonlinear equalization scheme with the aim of mitigating nonlinear distortion and interference, which have a severe impact on 5G mmWave communications, is proposed in [65], by a local-linearization observation model. The authors address the nonlinear distortion and frequency selectivity, the latter assumption differing to our case in this section where the channel coherence is a prerequisite for deriving channel gain realizations from the distributions under consideration. As opposed to the authors' statement that nonlinearity of output symbols induces memory which can be thought of as correlation via dependence, our analysis is stated as quadratic approximation of fading distributions, i.e. in such a manner that permits admitting strict independence between fading channel realizations as well as between fading and additive Gaussian noise. Proceeding further, the extent to which marginal a priori distribution is known or high dimensional intractability of a posterior distribution estimation (via particle filtering), is tackled in our analysis as we assume distributions are known i.e. the equivalent distribution resulting from independence based convolution. This means that a CDI is supported with a SIMO model as stated in the system model section. It is the piecewise continuity of the related curves that allows the linearized i.e. first order Taylor polynomial to provide an informative approximation. Also, the problem of recursive channel statistics update is further simplified by the Gaussian assumption providing a solvable iterative problem of determining mean and covariance matrix. Hence, it is an interesting question left for future research how the memory property will impact Taylor polynomial approximation merit, with arbitrary degree and how the memory in symbols will be fully exploited. In an interference mitigation framework, [66] focuses on the low complexity of guard zone based interference management. Along with uniformly distributed users in a geographical area of cell, the derivation of average throughput is addressed by Taylor expansion and inner radius of guard zone optimized. Successive interference cancellation is quantified in terms of decoding threshold, variance and user-base station as well as D2D-to base station distances and probability expressions are provided. The convexity property of related probability expressions confirm the feasibility of user transmission optimization. The analysis in this section includes diverse fading as compared to Rayleigh scenario in the above work but also exploits equivalent distribution convexity so as to apply CS compressibility rule. The idea of considering a randomized distance excluding the circular ring area, where transmission is prohibited so as not to lead to low probability, is combined with the proposed algorithm for inner radius optimization in order to minimize performance degradation. The idea of applying Taylor expansion is in accordance with the analysis of this section, thus, in the absence of a closed form expression. The convexity of average throughput is also present in this paper and utilized by adjusting inner cell radius. The degree of Taylor expansion considers the values of second degree and also fourth degree while we consider only second degree as the minimal degree of producing curved approximation. An interesting approach of using Taylor expansions of Doppler frequencies, AOAs, AODs, multipath delays and powers is adopted in [67] with the nonstationary assumption introduced together with varied velocity receiver mobility and scatterers as opposed to receivers moving with constant velocities and fixed scatterrers. The Doppler frequency arising as the result of vehicle and scatterers movement is equivalently expressed by a second degree polynomial as is our case investigated whereas another distinction is the fact that the

expressions derived are confined to a SISO model while our analysis assumes SIMO model. The mean of AOAs and AODs are modeled by a first degree Taylor polynomial as well as multipath delays and powers, the latter two approached by first order polynomial and exponential power model. The main differentiation of the above work and our case is that: effect and modeling of Doppler frequency is not considered in the case of movement in our work nor are multipath delays and their relation to bounce models in a scattering environment. Another differentiation is the temporal correlation between fading samples assumed whereas in our case independence is not relaxed by introducing correlation. An information theoretic approach for channel modeling which very closely resembles the approach adopted in this thesis is [68]. Here, the authors justify the use of relative entropy, which we supported by the term of distribution mismatch, as a partial knowledge of distribution case, the latter being very practical in multiple antenna systems. Specifically, the authors translate this, due to partial knowledge, approximation by a class of distributions that result in a maximum value of relative entropy the latter up to a defined threshold, in a Rayleigh channel statistical case. The main motivation of recent active research is the channel variations being tracked at receiver. This paper not only admits symbol coherence time as a restriction, as in the analysis of this section, but also supports such statement by the fact that transmit antennas number increasing above coherence time will not provide capacity gains. Addressing channel capacity optimization, channel uncertainty emerges as a modeling tool and the related maximin problem is stated. The Lagrangian multiplier resembles, as a mathematical tool, the analysis over Rayleigh vs. Rician distribution inference in the previous section which was formulated by CS principle and did not include relative entropy as a weighted optimization term. This section generalizes both analyses, that of previous section and that of [68] in the sense that encompasses diverse fading: Rayleigh, Rician and Nakagami-m fading distributions, assessing performance of all fading scenarios not only by capacity but also entropy based channel coding and symbol error probability. Hence, the extension of the current section towards a bounded relative entropy optimization problem is a challenging one that could provide performance in multi-antenna systems in generalized fading conditions always from an information theoretic derivation of classes of optimal distributions, an approach that could refine the solution produced in this paper that the solution occurs at the relative entropy boundary along with a satisfactory channel coding interpretation. Contrary to our standard Gaussian additive noise assumption included as a statistical prerequisite in the entire thesis, a ML noise variance estimator for mMIMO channel estimation [69] addresses the more practical problem where noise variance is not known and is either same or different in the receiver antennas. As noise variance directly impacts channel estimation, the authors propose a method that is non-iterative as is the case in this section that tackles the high complexity of ML based estimator as well as the inversion of covariance matrix in noise estimation for conventional MIMO systems, the latter aim being similar to that in the previous section where l_1 minimization formulated by the equivalent CS algorithms was replaced by a moment based optimization problem along with a Lagrangian optimization expression. Another essential metric investigated in the above work are the AOAs in the mMIMO setting whereas our analysis refers to a SIMO model focusing on independent branches at receiver with no notion of the channel characteristics except for the fading distribution modeled by along with additive noise

accounted for. Another point of mathematical similarity between this paper and our case is the operator max in Eq.(32), which forces to zero any result that indicates negative valued variance and the operator that forces to zero a relative entropy of negative value which by definition would quantify the redundant bits of average code description length. Referring to the same variance estimation and due to the fact that our analysis is confined to conventional MIMO systems, one could argue whether the variance comparisons conducted could migrate to the mMIMO setting. This would certainly increase computations and problem dimensionality even if same variance values could apply to different antenna clusters. Hence, the issues of complexity and computational burden demand careful investigation in such a query. Our case matches more closely the equal variance case and further the unit variance assumption also considered for this scenario in [69]. Additionally, the observation of the symbol error rate decreasing with increasing antenna number offers an interesting extension of the current analysis in this section as it considers not only diverse fading conditions along with additive noise cases but also the CS based inferences and Taylor approximations as performance comparisons benchmarks, all this to be applied in the 5G mMIMO case. A quite representative paper thoroughly investigating compressibility conditions for distributions is [70] where incompressibility is based on Laplace distribution while compressibility considers Pareto distribution and Gaussian distributed «encoder» as implicitly stated in the paper. By performing comparison to CS related assumptions, the Laplace distribution that is accounted for as a poor candidate for compressibility does not affect analysis in this section as the fading distributions are all defined in the positive axis. Moreover, the reconstruction error relative to unbounded second moment that can be made arbitrarily small is not adopted in our case but could serve as a performance benchmark quantifying the error which in our case is not implicitly conveyed but indirectly by derivation of the finite variances for each distribution case. Another valuable comment relates the aforementioned unbounded second moment with sufficient compressibility providing the asymptotic infinite dimensional optimality of l_1 based decoder under the Gaussian encoder constraint. This could provide the means of either extending our fading distribution cases to the mMIMO setting along with asymptotic performance benchmarks ensuring the property of convexity. As opposed to CS inference of the respective distributions, the issue of the extent to which under-sampling guarantees compressibility using sparse method constitutes an interesting point of extension of this section in terms of conditions of choosing CS methods as opposed to least squares. In overall, the main derivation of [70] being the value ranges of under-sampling ratio based on conditions for compressibility regarding second and fourth moments, the fading distributions and additive noise forming the independence based distributions by convolution operation, could be further investigated. Specifically, the CS inference could provide a comparison framework if used in the context of sparse estimators and the extent of their optimality. The Laplace distribution widely examined in this paper could also be set for comparison for all fading cases in this section. In a relevant context of not whether fourth moment is bounded or not but considering a large valued fourth moment of multipath amplitude distribution, the authors in [71] derive the conditions for CS channel estimation based on OMP performs almost well as BPDN, a property known as a characterization of greedy algorithms offering lower computational complexity. The promising contribution of this work is

to extend CS estimation algorithmic performance in a diverse set of amplitude models in a non-Gaussian sense. The justification of non-Gaussian multipath component model concerns the property of differences between delays and the utilization of the CLT related modeling, a scenario differing from the aforementioned paper, where the amplitude is separately modeled. The multipath clustering assumption induces memory which is not a prerequisite for our analysis in this section. The delay discretization error in the context of CS OFDM channel estimation and its impact on multipath clusters could be equivalently modeled as a Taylor polynomial with the respective non-negative delay value as the expansion point, which could translate to a transition from MIMO to mMIMO scale along with the optimized antenna number and spacing so as to exploit the correlated case. Another remark supporting the paper contribution is the l_1 minimization based algorithm BPDN which requires weaker condition where similar magnitude of coefficients is assumed as opposed to rapid magnitude decay. This fits into our analysis relevant to the methodology replacing the noisy l_1 minimization algorithm (BPDN) with one based on independence and statistical convolution, as stated in the mathematical preliminaries section. Regarding the use of specific values of l_p norm parameter value p in CS literature so as to balance complexity and solution accuracy along with observation that the compressibility index introduced is based on non-Gaussianity of multipath components can be expressed as l_p norms with values of p equal to two and four, the derivations in [71] offer interesting applicability to fading distributions not only by independence dictated but also introducing memory and correlation constraints for mean-square-error analysis. Thus, with a fourth moment magnitude condition, multipath components are more compressible which means sparsity can be stricter while achieving same performance, the latter in a statistical sense which very closely matches our case in terms of mathematical formalism. Among other findings of this paper is that CS estimators outperform non-sparse estimators, which is the conclusion reached in our case that are fully justified by two interpretations: the extent to which the number of measurements chosen lead to representative results along with CS optimized superior performance as a result of compressibility. Chandra and Bose in [72] derive analytic expressions for coherent Rician fading channels for symbol error probability which follows our main course of deriving the related curves for all fading distributions considered generated according to CS inferred case and Taylor approximation case. The assumption on which the analysis is based is the short symbol duration such that a negligible coherent loss, which also abides by our analysis, so as no frequency selectivity is accounted for. Moreover, since mathematical manipulations involve only approximations of distributions, the most accurate model that closely relates to our case is a partial distribution CDI model, while the CSI at receiver side requires more than what is considered as the optimization problem input and constraints. The observation of higher SNR requirement in the case of increasing constellation size also complies to our results where higher SNR leads to smaller symbol error probability for all distributions evaluated. In overall, the essential element is the construction of closed-form expressions for symbol error probability by means of Gauss hypergeometric function in coherent Rician fading together with consideration of Rician factor K=0 leading to Rayleigh and thus NLOS component distribution. Hence, the Taylor approximation of such expressions could be an interesting extension, as the impact of polynomial quantities in performance, along with CS compressibility

rule applied. Proceeding further, another work [73] in the same context of coherent but Nakagami-m fading along with the equivalence to Rayleigh fading by proper parameter selection relates to our consideration of the latter fading distribution. The main contribution of this paper being built upon gamma function and Q-function approximation, the latter is worth noting serving as a polynomial approximation with improvement if the relevant degree is increased. Combined with the argument that despite the apparent complexity of the error expression derived by the authors, it constitutes a simple paradigm compared to existing literature formula for calculations on Nakagami-m fading channels. The similarity of the mathematical derivations of this paper that matches our analysis is the derivation of the frequency domain expression of the moment generating function that is based on a product relevant to convolution in the original domain which is dictated by the sum of the equivalent SNRs. This paves the way for a frequency domain expression of our case for further derivations by CS principle of compressibility and Taylor approximation adopting a polynomial flavor to the problem of fading channel performance. As the main conclusion reached for [73] the accuracy of the approximation is verified and thus can also be used in the derivations of independence based scenarios as well as incorporating the challenge of correlation as a far more practical assumption where Q-function approximations are also computationally attractive due to its widespread use in closed form formulae.

3.3.3 System Model

Before proceeding to the results based on simulations, the system model is provided: we consider a SIMO scenario i.e. a transmitter equipped with single antenna and a multiple antenna receiver thus independent branches complying with receiver diversity and also incoherence principle of CS optimization criterion. As already stated, the investigation in this section does not account for practical correlation and the related antenna spacing problem to achieve a directional beam width to a certain user. Moreover, user mobility causing Doppler frequency calculations as well as interference management and mitigation techniques, the latter caused by neighboring transmissions are not considered. In summation, time coherence is assumed or stated more accurately symbol duration is short enough that negligible coherence loss occurs.

3.3.4 Simulation Results

Distribution curves and capacity derivations

The simulations were conducted using Matlab software. The fading distributions considered are Rayleigh fading, Rician fading and Nakagami-m fading each of the distributions considering additive noise as an additional case apart from the noiseless case. The additive noise follows a standard Gaussian distribution i.e. modeled as $N(0,1)$. All fading cases assume independence with respect to noise thus allowing the derivation of the noisy fading distribution as a convolution of the fading density function and the standard Gaussian noise density. Based on the aforementioned distributions, there are two approximations taken into account: the CS inferred distribution and the Taylor approximation distribution with respect to the exact one. All three distributions serve as channel gains generating functions. The first approximation is generated according to CS compressibility rule. Hence, a fraction of the largest valued channel gains is

preserved while the rest are discarded. Based on these preserved channel gains the distribution of the same kind based on these coefficients is derived i.e. an equivalent inference problem with the CS compressibility constraint quantified. Regarding the second approximation, the Taylor polynomial is derived with a second degree assumption, i.e. the minimum degree required for capturing curvature, which is also the property of the distribution curves. The two above approximation methods are then utilized to assess performance i.e. capacity, variance calculated for each case, average code description length based on Shannon and relative entropy and symbol error probability evaluation for all distributions considered.

The structure of this section is as follows: first the distributions based on these approximations are derived for each fading case along with the capacity derivations by means of Eq.(17). The average values are derived for each case. The next stage concerns variance calculations based on first two order moments given by Eq.(13) accounting for the respective approximations and performing comparison with means and variances for Rayleigh distribution Eq.(4) and Eq.(5), Rician Distribution from Eq.(7) and Eq.(8) and Nakagami-m distribution from Eq.(10) and Eq.(11). The average code description length is derived and the results are fully justified for each fading and approximation cases. Finally, the last section of the simulation results evaluates symbol error probabilities for all the above explained cases considered.

Rayleigh Noiseless Fading

The first section depicts the three considered curves produced as above for Rayleigh noiseless fading case. The Figure 3.1 is given below:

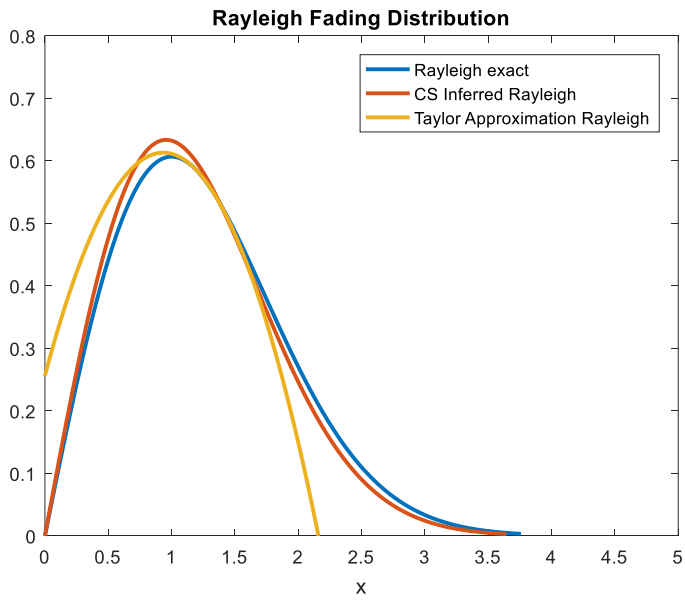

FIGURE 3.1: Rayleigh noiseless fading distributions: exact, CS inferred and Taylor approximation

As observed, all three curves are closely matched, the more accurate curve being the CS inferred one. The Taylor polynomial based curve can be substantially improved by increasing the polynomial degree though the second degree is considered in this case.

The capacity given three Rayleigh fading distribution curves is given in Figure 3.2 below:

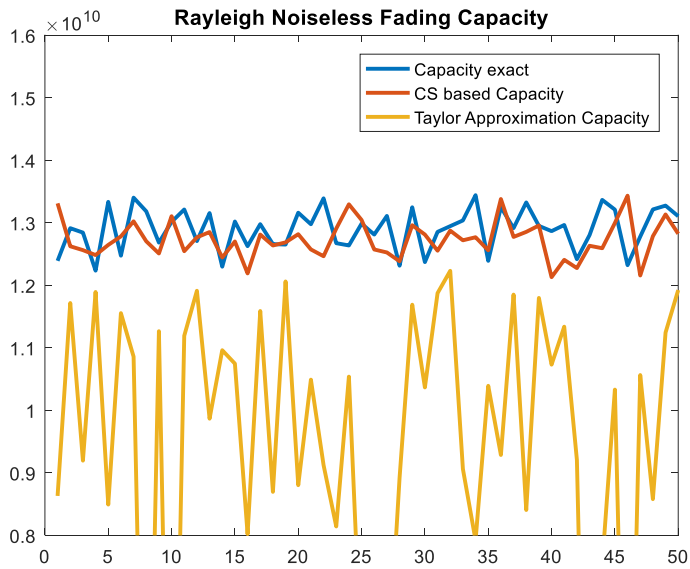

FIGURE 3.2: Channel Capacity for noiseless Rayleigh Fading

The shape of the resulting curves is a direct consequence of channel gains randomness. In order to reach meaningful results, the average value of the capacity is calculated and compared for the three above cases. Indicatively, the exact distribution resulted in 12.8Gbps while CS inferred case equal to 12.7Gbps and Taylor approximated curve resulted in 9.82Gbps. Hence, a 0.78% capacity penalty was observed for CS case and 23.3% penalty for the Taylor approximation case. Hence, CS induces a near optimal result while Taylor approximation induces an observable penalty, while Taylor approximation produces more variation around its average value.

Rayleigh Noisy Fading

The inclusion of additive noise is accompanied by the independence noise assumption, hence the exact distribution is equal to the convolution of fading distribution and Gaussian noise of zero mean and unit variance. This result is then used to produce CS inferred and Taylor approximation in the same manner as above. Thus, the three distributions for this section are plotted below in Figure 3.3:

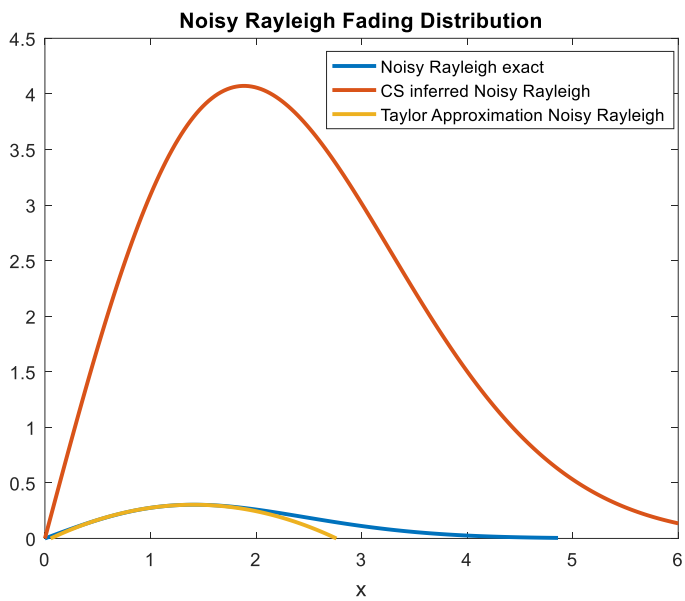

FIGURE 3.3: Rayleigh Noisy Fading Distributions: exact, CS inferred and Taylor approximations

The most crucial observation concerns the CS inferred case where due to keeping only the channel gains of larger magnitude results in the distribution curve of larger area and above the exact distribution. As apparent in the Rayleigh noisy case, the Taylor approximation provides the best fit to the exact distribution extracted from the fading and additive noise convolution operation. The resulting capacity curves are plotted below in Figure 3.4.

The results obtained from the figure are the average capacities: 9.22Gbps for exact curve, 9.44Gbps for CS inferred case thus 2.3% larger capacity than the exact distribution case and 9.48Gbps for Taylor approximation hence 2.7% larger capacity. The above results being the approximations producing higher average capacities are both interpreted by the additive noise consideration along with the extent to which the number of samples obtained are sufficient for reaching accurate results. This however does not degrade the merit of the observation that CS results in performance improvement in this case of application of inferring a distribution in a manner of being of the same kind. Considering the above and relative to the CS case, an optimal capacity has been reached. For the Taylor approximation, the justification for higher capacity is the combined effect of number of extra samples needed and the convolution of the fading distribution with the noise distribution leading to the random channel gains that derive higher capacity. This is a consequence of better fit of Taylor expansion given the resulting noisy fading expression.

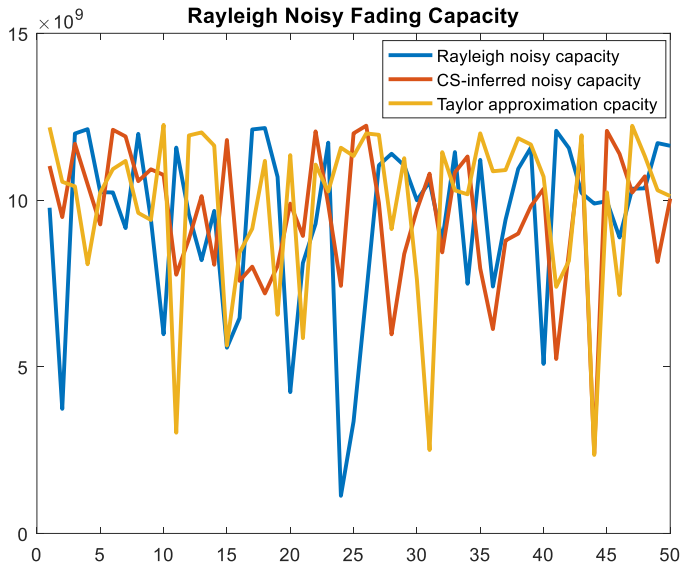

FIGURE 3.4: Channel Capacity for Noisy Rayleigh Fading

Having completed the Rayleigh fading cases, the following subsection is based on Rician fading case both noiseless and noisy cases along with the two approximation methods outlined above.

<u>Rician Noiseless Fading</u>

The Rician noiseless fading case is simulated and the results are shown below in Figure 3.5 and the related capacity curves are shown in Figure 3.6.

As observed from the above figure, the exact distribution results in average capacity of 13.5Gbps, the CS inferred distribution in capacity of 14.2Gbps, hence, 5% larger capacity and the Taylor approximation resulted in 10Gbps capacity thus in a 26% penalty compared to the exact distribution case.

As observed, the variation of the Taylor approximated curve about its average value is much larger compared to the exact and CS inferred case, which show the same smaller variation. This observation was consistent with the Rayleigh noiseless case but did not adapt to the case including additive noise. Moreover, compared to Rayleigh noiseless case where CS compressibility and inference results in small penalty, the inference for Rician noiseless case verifies the first observation that translates to CS optimality, i.e. larger capacity interpreted by CS compressibility along with sufficiency of number of samples for deriving a representative result. The section proceeds with the Rician noisy fading case.

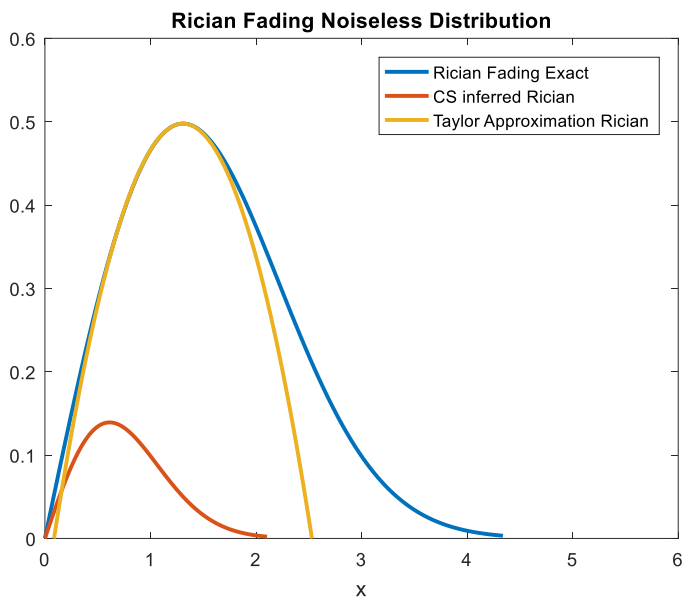

FIGURE 3.5: Rician Noiseless Fading Distributions: exact, CS inferred and Taylor. approximation

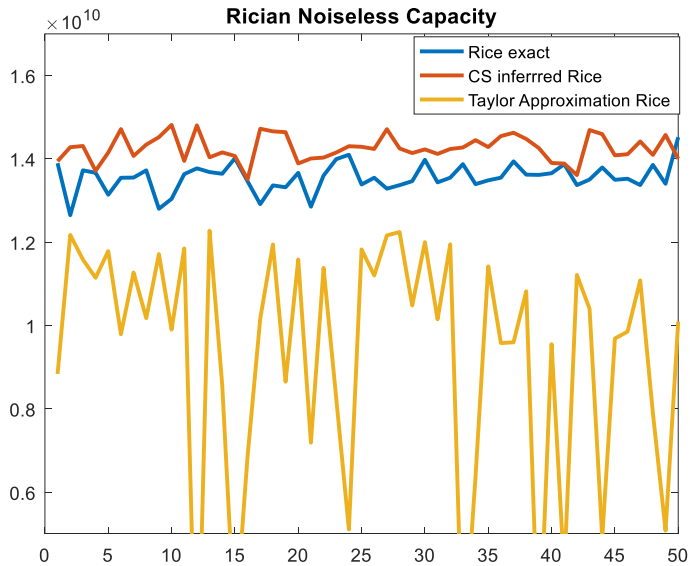

FIGURE 3.6: Channel Capacity for Noiseless Rician fading

<u>Rician Noisy Fading</u>

The Rician fading with additive noise included results in the following distribution curves and capacity derivations in Figures 3.7 and 3.8 respectively.

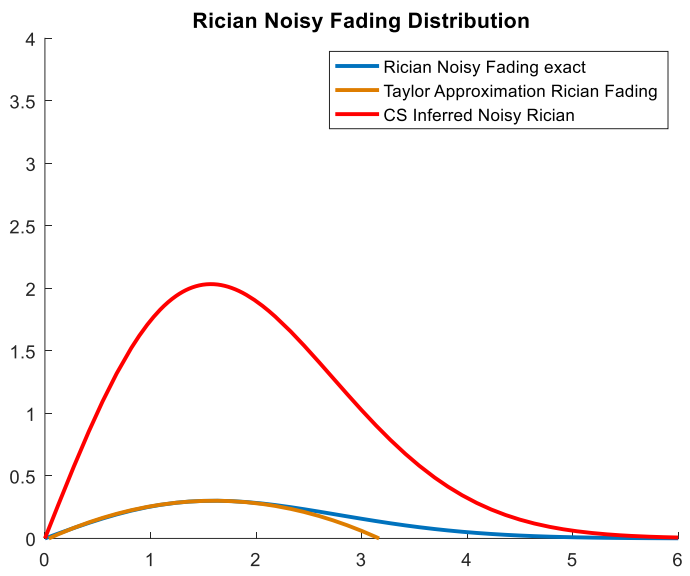

FIGURE 3.7: Rician noisy fading distributions: exact, CS inferred and Taylor approximation

The same observations for the noisy Rayleigh fading case hold also for the noisy Rician fading concerning the CS inferred distribution case. Moreover, Taylor results in better fit to the exact noisy fading distribution. For the noisy case, the CS inferred distribution follows the same pattern, i.e. larger space below the curve and higher peak of the distribution curve. As will be observed and interpreted, this pattern will be the core finding of this section regarding CS inferred case average code length derivation.

As is also apparent, Taylor approximation provides an accurate and close curve fitting which appears to be slightly superior compared to the Rician noiseless case examined above. The average capacity for the exact distribution was found equal to 9.78Gbps, for the CS inferred case equal to 8.89Gbps i.e. 9.1% penalty compared to the exact one and for the Taylor approximated almost 9.78Gbps showing an optimal value match to the exact value. This is a very interesting result that refers to the additive noise where at first glance, can be mathematically attributed to the closer approximation for this curve. This translates to the fact that for the specific degree and expansion point the match resulting is closer, which is also related to our statistical assumption of independence. Finally, the variations of the capacity curves are also similar, as in noisy Rayleigh fading case.

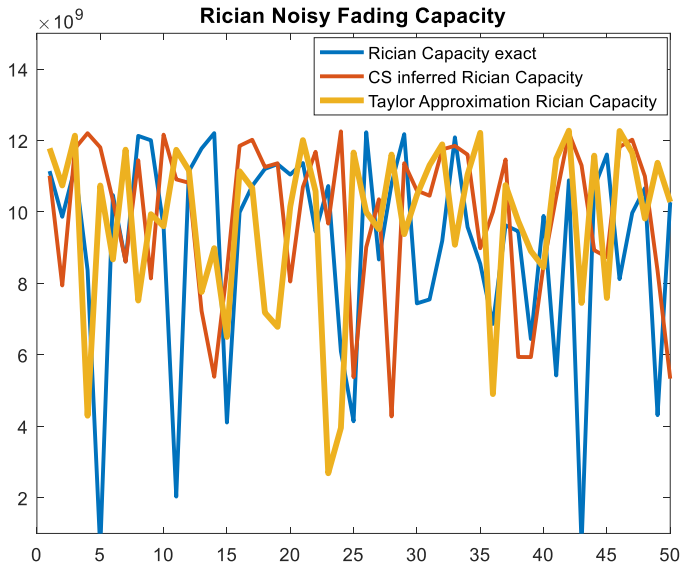

FIGURE 3.8: Channel Capacity for noisy Rician fading

<u>Nakagami-m Noiseless Fading</u>

Proceeding in the same manner, the three respective curves, exact, CS inferred and Taylor approximation for the Nakagami-m fading case without additive noise are depicted below in Figure 3.9. The merit of the analysis in this section relies on the fact that the Nakagami-m fading distribution can well approximate the Rayleigh and Rician fading distributions, as will be stated in the mathematics interpretation section.

An important observation regarding this case is the CS inferred distribution curve indicating the trend observed in the previous noisy cases with the exception that this curve is much narrower hence a much smaller variance as will be shown in the variance estimation section. Hence, the CS based inferred curve being above the exact in the Nakagami-m fading noiseless regime was only observed in the previous noisy cases. Another observation is the close curve fit for the Taylor approximation of the exact curve. Regarding capacity evaluations the results for this case are given in Figure 3.10 and enhanced with the following comments.

The exact average capacity was calculated to be equal to 11.9Gbps while the CS inferred case resulted in 12Gbps, i.e. 0.83% greater capacity. The Taylor approximation derived an average capacity of 9.58Gbps thus a 19.4% average capacity penalty. Finally, the CS inferred distribution curve appears to be slightly displaced with respect to the peak value of the exact distribution and also be characterized by symmetry as opposed to the other cases investigated.

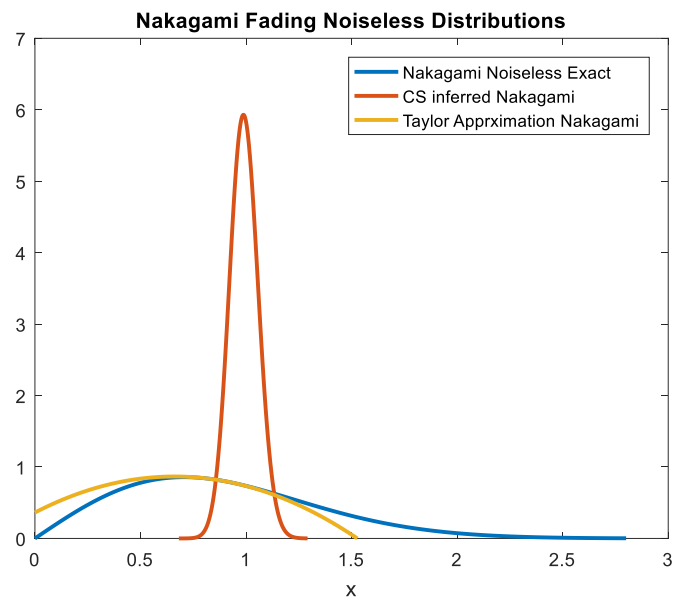

FIGURE 3.9: Nakagami-m Fading Noiseless Distributions: exact, CS inferred and Taylor approximation

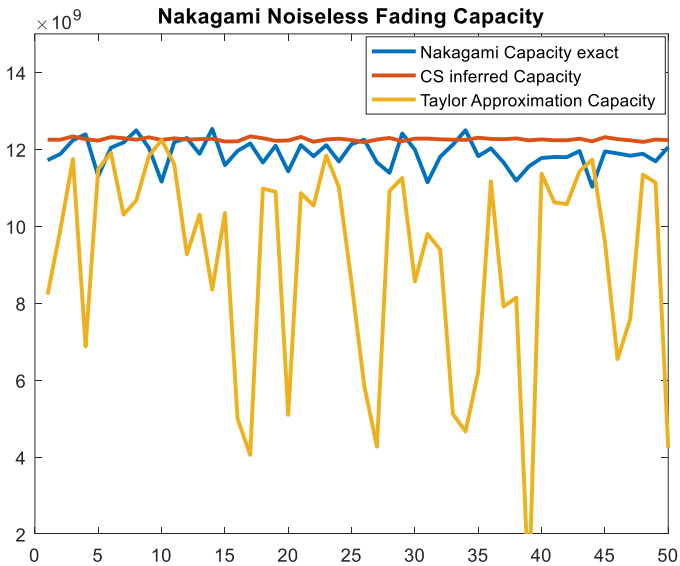

FIGURE 3.10: Channel Capacity for Noiseless Nakagami-m fading

<u>Nakagami-m Noisy Fading</u>

For the Nakagami-m fading case with additive noise consideration, Figure 3.11 depicts the similar trend in the distribution curves with the exception that the CS inferred in this case is narrower hence a smaller variance as in the noiseless case.

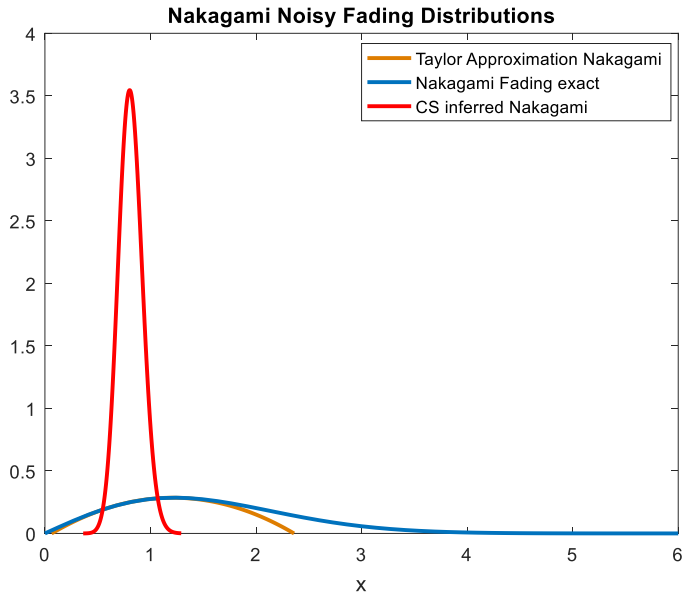

FIGURE 3.11: Nakagami-m Noisy Fading Distributions: exact, CS inferred and Taylor approximation

The capacity derivations are depicted in Figure 3.12. For this noisy case, the exact average capacity value was found equal to 8.7 Gbps, for the CS inferred case the average capacity equal to 9.5Gbps i.e. a 8.4% larger capacity and for the Taylor approximation case the average capacity was equal to 9.6Gbps, thus 9.3% larger average capacity. The slightly larger average capacity values are justified. For the CS inferred case a combination of requiring more samples for reaching representative values along with optimal CS results are the justification for the result. The reason for the greater average capacity derived from the Taylor approximation fit is that the channel gains produced are larger valued which stems from the Taylor curve fit.

It is worth noting from the simulations that the CS inferred average capacity, a trend also observed in previous cases, is above the average of the capacity generated from the exact distribution. This is justified by the independence assumption as well as the additive noise distribution. What remains in terms of future research is the investigation of correlation followed by CS inference as well as optimized noise variance which may not be known in advance. Summing up, the asymptotic CLT case for all derived distributions similar to the previous section could provide performance benchmark in the large sample regime and is left as future extension of the above cases.

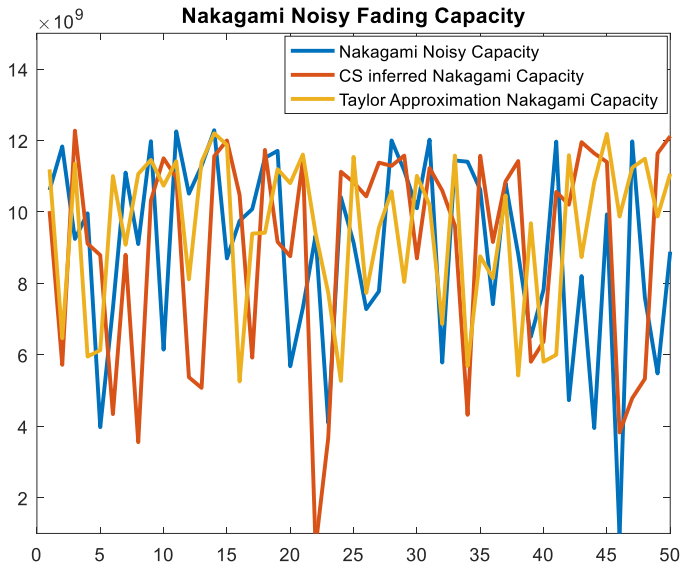

FIGURE 3.12: Channel Capacity for Noisy Nakagami-m fading

Variance estimation

This section is devoted to calculating variances of the distributions produced by means of the above: exact, CS inferred and Taylor approximation for each fading case considering both noiseless and noisy cases. The results of the extensive simulations conducted are summarized in the Table 3.3 below:

TABLE 3.3: Statistical Variances for Noiseless and Additive Noise cases

Fading Distributions/No noise-Noise included	Noiseless case	Additive noise included
Formula calculation/Rayleigh fading exact	0.429/0.429	1.429/1.258
Rayleigh CS inferred	0.3935	0.1206
Rayleigh Taylor approximation	0.2	0.7
Formula calculation/ Rician fading exact	0.61/0.61	1.61/1.481
Rician CS inferred	0.1	0.2
Rician Taylor approximation	0.468	0.9
Formula calculation/ Nakagami-m fading exact	0.214/0.214	1.214/0.946
Nakagami-m CS inferred	0.005	0.012
Nakagami-m Taylor approximation	0.1459	0.476

Before accounting for the results for each fading distribution some remarks must be provided for overall results interpretation. The first comparison is with respect to the variance calculated from the specific formula for each distribution and compared to the variance derived from the moment integrals based on Eqs. (13),(14).

Proceeding further, the latter equation is used to evaluate variance with the CS inferred distribution in the integral expression. The third variance is calculated based on the integrals expression but using the Taylor approximation of the respective distribution.

Based on the above, the initial important observation is the perfect match between the variance calculated from the formula and that of the integral expression in the noiseless cases. The next result relates to the increase in variance values for the noisy cases compared to the noiseless ones. The only exception is the CS inferred Rayleigh distribution. The final remarkable result concerns the variance value mismatch between the value obtained from the formulas and the integral expressions for the noisy cases. Thus, the formula based calculation is equal to the fading distribution variance plus one, resulting from the independence assumption and the covariance term cancellation. However, the result from the integral expression is of smaller value due to the fact that correlation is included in the integral calculations. It must also be noted that the variance calculation results are in perfect agreement with the distribution curves depicted in the previous section.

<u>Rayleigh fading case</u>

In the noiseless case, the formula and integral based derived values are in perfect match as already justified. For the CS inferred the slight difference in variance compared to the exact is verified. The Taylor approximation based result indicated smaller variance.

In the noisy case, the difference between formula and integral based expression values are evident and justified above, with independence calculation as opposed to correlation based moment integrals. The CS inferred integral based expression results in smaller variance thus narrower curve, while the Taylor approximation is characterized by a value closer to the exact variance.

<u>Rician fading case</u>

In the noiseless case, the match between formula based and integral based values is also verified. The CS inferred case indicates a lower value while the Taylor approximation has a smaller difference from the exact value.

In the noisy cases, the mismatch explained above occurs in this case as well. The CS inferred variance is smaller than the Taylor approximation case which more closely approaches the exact variance value.

<u>Nakagami-m fading case</u>

The formula based and integrals based value match is verified in this section as well. The narrow curve depicted in the Figure 3.9 is verified by small variance value. Though concerning a noiseless case the narrow curve of the CS inferred case above the exact is observed in this case. The Taylor approximation based integral expression result produces a variance value closer to the exact.

In the noisy case, the mismatch between formula based and integral expression based is observed here. The CS inferred case produces a slightly larger value compared to the CS inferred similar noiseless case. Taylor approximation results in larger variance value.

Required number of bits for fading channels

This section derives the number of bits required to describe the fading channels considering all the above cases. Thus it quantifies the uncertainty of the channel, in terms of Shannon entropy with respect to the wireless channel distribution. Additionally, the relative entropy quantifies the redundant bits required in the case of approximating the accurate distribution with a different one.

Regarding the derivation of the results of this section, the Shannon entropy was calculated based on the exact distribution expressions and the relative entropy was calculated utilizing the CS inferred and Taylor approximating case as the approximating distributions as input to the relative entropy cases. Before proceeding to the required number of bits for each of the cases considered, the equations of Shannon and relative entropy must be carefully examined.

First of all, the Shannon entropy is by definition a positive entropy resulting from the minus sign since the values inside the logarithm are probability values hence smaller than unity which leads to a negative logarithm. Thus the bits quantifying uncertainty being positive by definition are expressed via the Shannon entropy. For the relative entropy, the coding theory assumes that it is a positive quantity. From a mathematical point of view, this means that the logarithm is positive i.e. the quantity inside the logarithm is greater than unity. This is actually a consequence of Information Inequality that dictates the fact that code length difference must be positive as a means of certifying the quantification that relative entropy realizes.

However, some cases considering the aforementioned approximations derive negative values, which are not supported by coding theory. From the results of the cases where the approximated curve is above the exact one in CS inferred cases, it is apparent that the formulation of the relative entropy in these cases results in a fraction less than unity since the accurate distribution is in the numerator and the approximating distribution in the denominator. Thus, a negative result is obtained. The interpretation of such results complies with the operator below:

$$\text{Relative entropy} = \max\left(0, \sum_{x} f_x \log_2\left(\frac{f_x}{p_x}\right)\right) \qquad (32)$$

Hence, the interpretation of the negative valued cases that constitutes one of the core findings of this thesis is that zero additional bits are required and the required number of bits are equal to the Shannon entropy. The extension of such interpretation to the context of reducing required number of bits due to negative valued relative entropy is beyond thesis scope and left as a future research direction. The above justified observations are evident in the noisy cases. Hence, the CS based resulting distribution is characterized by greater uncertainty compared to the exact when additive Gaussian noise is considered. Towards an efficient interpretation, noise increases uncertainty in

CS distribution case but translates to zero complexity as no extra bits are required for channel coding. The justified results are also partly due to the contribution of the assumption of independence in the formulation of the noisy exact distributions. Accordingly, the independence assumption along with additive noise consideration leads to an incoherence that is defined as the effect or reduced uncertainty of the CS inferred distribution compared to the true distribution which is the very reason of zero additional complexity. Finally, the relative entropy is finite which is fulfilled by definition since the case that distribution value of zero for distribution f_x holds for p_x as well. Having noted the above interpretation, the results for each fading case are analyzed below in Table 3.4.

TABLE 3.4: Shannon, relative entropy values and average required numbers of bits

Fading distributions/ entropies	Shannon entropy exact distribution	Relative entropy exact vs. CS based inferred/ exact vs. Taylor approximation	Average number of bits required
Rayleigh noiseless fading	23.5	3.6/-1.7	28/24
Rayleigh noisy fading	17.6	-16.4/ 5.4	18/24
Rician noiseless Fading	24.4	82.8/17.8	108/43
Rician noisy fading	22.6	-29.3/11.4	23/35
Nakagami-m noiseless fading	17.8	-70/ -3.2	18/18
Nakagami-m noisy fading	26	-53/0.86	26/27

Rayleigh fading case

Related to the noiseless case, 24 bits are required based on Shannon entropy and 4 additional bits are required for the CS inferred case verifying the close match between the two distribution curves. For the Taylor approximation negative valued relative entropy was found the reason being the Taylor curve being slightly above the exact curve. Hence, 28 bits are required for the CS inferred case as an approximation and 24 bits for the Taylor curve.

The noisy case investigation resulted in 18 bits for Shannon entropy. For CS inferred case the relative entropy was negative as the equivalent distribution curve being above the exact, whereas for the Taylor approximation 6 extra bits are required. Hence, for CS inferred case 18 bits are required and for the Taylor approximation 24 bits are required.

Rician fading case

Regarding the noiseless case, Shannon entropy was equal to 25 which for the CS inferred case 83 extra bits are needed. This indicated the significantly smaller

uncertainty of the CS inferred case its curve being well below the exact curve. For the Taylor approximation curve, 18 extra bits are required. The reason for this is the observable gap at the right side of the distributions. In overall, CS inferred case requires 108 total bits while the Taylor approximation case requires 43 bits.

In the noisy Rician fading, 23 bits are due to Shannon entropy. The CS inferred case results in negative valued relative entropy while the Taylor approximation requires 12 additional bits. Hence the CS inferred case requires a total of 23 bits whereas the Taylor approximation requires 35 bits in total.

Nakagami-m fading case

The Shannon entropy results in 18 bits for the noiseless case. For the CS inferred approximation the significantly negative value of relative entropy is verified by the narrow curve being above the exact. Moreover, the Taylor approximation was less negative as indicated by a part of the Taylor approximation curve being slightly above the exact distribution. Hence, both cases require 18 bits.

For the noisy case, Shannon entropy derives a value of 26 bits required whereas the same trend observed for the CS inferred case resulting in relative entropy as in the noiseless case. For the Taylor approximation case, 1 additional bit is required. Hence for the CS inferred case 26 bits are required and for the Taylor approximation case 27 bits are required in overall.

Wireless fading performance assessment via symbol error probability

This section derives symbol error probability with respect to SNR for all fading cases investigated. The exact distribution case, CS inferred case and Taylor approximation case are considered.

Rayleigh fading case

For the Rayleigh fading case, in noiseless assumption the exact distribution based error curve is similar in error magnitude with the CS inferred one, hence, initial observation for CS inferred case is no performance degradation. For Taylor approximation, higher error probability was observed.

For the noisy case, CS inferred curve indicates a smaller symbol error probability. This is a combined effect of how representative the curve is with respect to samples taken along with improved performance for the CS inferred case. Moreover, the Taylor approximation based curve indicates a smaller error not only compared to the exact noisy curve but also compared to the noiseless Taylor approximation curve. Such smaller penalty can be attributed to the closer match of the average capacity with respect to the exact case when additive noise is considered compared to the noiseless case. The figure for this case is given below in Figure 3.13:

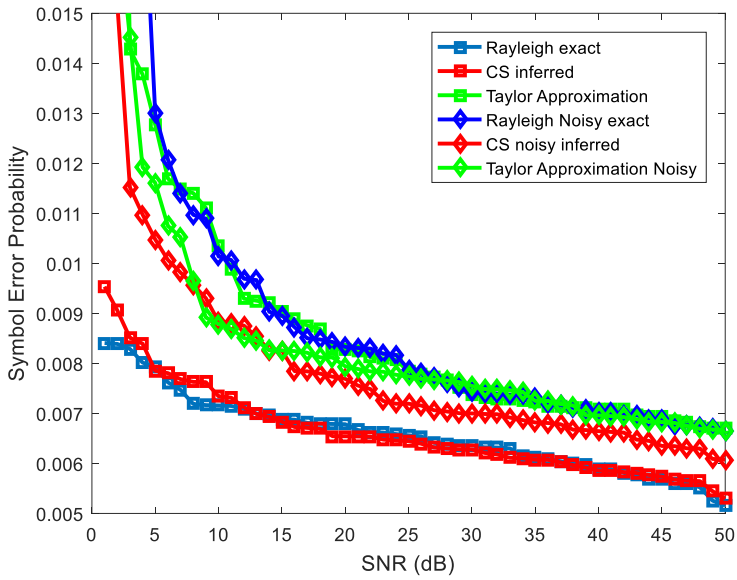

FIGURE 3.13: Symbol Error Probability for Rayleigh fading noiseless –noisy cases

Towards a thorough interpretation of the above figure curves, it is observed that in the noiseless case the symbol error probabilities of exact and CS inferred distributions are almost identical thus producing similar error values. The Taylor approximating case depicts a performance penalty and it is quite indicative that the curve closely approaches the additive noisy curves, specifically the exact noisy distribution curve. Also indicative is the observation that the noisy cases of CS inferred distribution and Taylor approximation distribution produce symbol error probability curves that are below the Taylor approximating curve for Rayleigh noiseless case. This conveys the significant performance penalty of the Taylor approximation of the respective case.

Regarding inclusion of additive noise, the merit of applying CS based inference is evident enhancing the concept of optimality given the fact that noise contaminates the channel. Thus, performance improvement results in this more practical and also distortion case. Moreover, the Taylor approximation for this noisy case depicts similar performance but with another justification i.e. a polynomial based approximation of the exact noisy distribution. Hence, from different angles, the CS applying compressibility rule and Taylor approximation approaching a curve more closely given the polynomial degree and specific expansion point result in improved performance. This again traces back to our assumption of independence as well as noise variance parameter considered.

<u>Rician fading case</u>

The symbol error probability curves for Rician fading are shown in Figure 3.14:

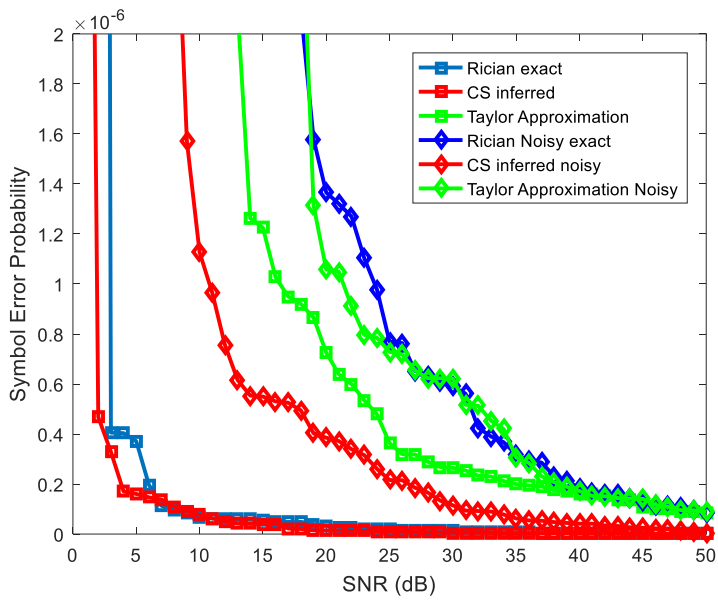

FIGURE 3.14: Symbol Error Probability for Rician fading noiseless-noisy cases

As observed in the Rayleigh fading case, the close match of exact and CS inferred case in Rician noiseless fading verifies optimal performance. The Taylor approximation case results in significant performance penalty, which is higher than CS inferred curve for the noisy Rician fading case but lower than symbol error probability curves produced for the exact noisy case as well as the Taylor approximation of the additive noise case included.

Regarding the additive noise consideration, the exact and Taylor approximating cases are almost similar in terms of symbol error probability, while the CS inferred distribution results in lower error probability curve. This once again highlights the advantage of applying CS compressibility rule not only resulting in no performance degradation but also inducing improvement relative to the exact fading case.

<u>Nakagami-m fading case</u>

The symbol error probability curves for this fading case are depicted below in Figure 3.15. In contrast to the previous cases, CS inferred curve for the noiseless setting shows a significant performance degradation which directly relates to the narrow curve being above the exact one. Taylor approximation case conveys a smaller error penalty compared to the CS latter case.

For the noisy case, the CS inferred case results show a performance degradation as well. The Taylor approximation case results in smaller error and more closely matches the exact distribution error probability curve.

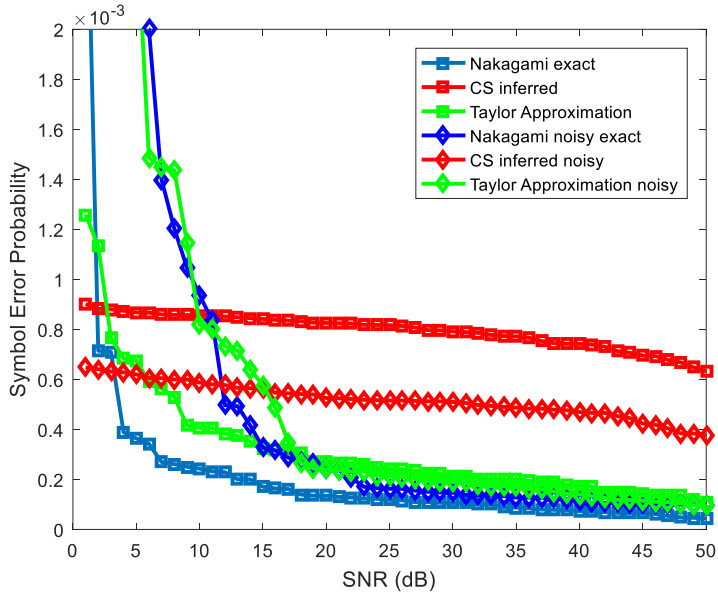

FIGURE 3. 15: Symbol Error Probability for Nakagami fading noiseless-noisy cases

Another observation worth highlighting is that the CS inferred error curve for the noisy case is below the noiseless curve. This leads to the fact that because of independence assumption the remarkable CS inference for noise contaminated fading results in lower symbol error probability. As observed from the symbol error derivations, the Nakagami-m fading case resulted in smaller error than Rayleigh fading case and higher error compared to the Rician fading case. Thus, Nakagami-m fading is a compromise between NLOS Rayleigh fading and Rician LOS fading. The mathematical explanation of the above is that Nakagami-m fading can model both Rayleigh and Rician fading with proper parameter selection. Finally, additive noise consideration results in larger errors for all cases.

3.3.5 Mathematical interpretations

The concept of convexity is verified by the shape of the fading distributions as the peak value is the unique extremum and the curve decays to zero as the abscissa approaches zero. Relative to the shape of the fading distributions, Rayleigh distribution curve includes an exponential decaying factor similar to the Gaussian distribution plus a linear term the above resulting in a non-symmetric convex curve. The Rician distribution consists of the aforementioned mathematical components plus the modified zero order Bessel function term, the latter having the monotonically decreasing property. The above contribute to the non-symmetric Rician distribution curve. Similar observations hold for the Nakagami-m fading case.

As already stated, the Taylor polynomial is assumed to be of second degree, the minimum degree of smooth convexity and the minimum required for capturing curvature. The extent to which the exact fading distributions are represented by their Taylor is directly dependent on the Taylor polynomial degree. The main goal being the accurate representation of the respective fading distribution the straightforward analytic calculation of complex integrals deriving entropy and statistical moments, is weighted together with the truncated version of the approximating curve which is defined by the

extent of the neighborhood of the polynomial expansion point. Hence, in order to improve approximation, the aforementioned degree must be increased the tradeoff being the resulting increase in complexity. Additionally, the increase in complexity grows much more rapidly compared to the increased approximation of the fading distribution achieved by the Taylor polynomial of higher order. Hence, although second order is the minimum degree for complying with curvature of fading distribution, the increase in polynomial order is expected to be confined to few extra degrees with respect to the second degree assumed in our case, thus settling to a slowly decreasing error.

While the Taylor approximation is more accurate than the CS inferred cases as observed in the noisy regimes, the CS comes along with no performance degradation in the CS inferred case while the CS is characterized by the same kind of distribution but with different parameters. A crucial remark is that the CS inferred distribution is of the same kind as the exact one but considering only larger values of channel gains. Thus, the resulting distribution is based on a subset of the initial channel gains. As a straightforward conclusion is that as more elements are added, the inferred distribution approaches the exact one in a tractable manner. Hence, in a topological sense the inclusion of successively more elements for CS inferring represents an intersection of the subset and original set which becomes a set equality relation. Moreover, an accurate justification of the produced distribution curves in all noisy cases as well as the Nakagami-m noiseless case is to what extent the number of channel gains are sufficient to accurately «produce» the actual i.e. exact distribution from the selected subset of channel gains.

An alternative method defining the maximum value of relative entropy and selecting an approximating distribution belonging to the uncertainty set of distribution that result in relative entropy equal or smaller to the maximum value is an interesting extension to be investigated. Another future research issue concerning application of CLT as in the previous section where it served as a distribution to be approximated is the CS inference and Taylor approximation of the respective asymptotic distribution. CS inference could therefore adequately approximate the distribution of each case by the same kind constraint. Contrary to the manner of approximation above, the Taylor expansion relates to the distribution in a polynomial-wise manner i.e. a polynomial curve with increasing precision as the polynomial degree increases. The quality of the curve fit would be justified by the properties of the curve and the fading conditions modeled by this curve as well as the optimal Taylor polynomial degree and the expansion point chosen. The latter could point to a separate investigation if the convergence focused on a certain interval of the respective distribution. All the above essentially boil down to the fact that fading channel performance constitutes a complex optimization problem with a function or metric to be optimized by enforcing specific constraints that correspond to the practical case it refers to. Optimality is not always guaranteed and if it does, as in the sections analyzed above, translates to certain tradeoffs which result in gains from a point of view to losses from another aspect. The vital conclusion from this section formulates the fact that entropy and consequently coding along with CS principles provide promising results as performance benchmarks even in the case of impractical and unrealistic assumption of statistical independence.

3.3.6 Applications to wireless communication systems

Regarding performance evaluation of wireless systems, the metrics considered in the previous sections are sufficient to completely assess system performance. The capacity results determine the upper bound of achievable rate through the channel. The symbol error probability quantifies the fraction of transmitted bits anticipated to be erroneously received by the wireless receiver. Furthermore, the variance of the fading distributions represents the spread of values with respect to the mean value which is also a measure of uncertainty. The number of bits also convey the uncertainty of the channel distribution and provide the number of redundant bits resulting from the approximation of the accurate distribution with another one.

The benefits of CS theory for the fading channels performance are twofold. The first is with regard to symbol error probability which results to no performance degradation and even improved i.e. lower error probability curve. The second is due to the fact that relative entropy values were computed to be negative. This was translated to zero extra bits in the context of the mismatch between the accurate distribution and the approximating one, which states that zero complexity was introduced in such a CS inferred case.

The Taylor polynomial advantage, as already stated, was the transformation of a mathematical quantity into a polynomial thus providing the calculation of complex integrals which would otherwise require the use of numerical methods of high computational complexity. The inherent limitation of approximation of the optimal result is thus traded off for mathematical manipulation. Examples of cases of system analysis employing Taylor polynomial approximation are derivations of error probability, modulation schemes and error rate evaluation, interference management in cellular networks, channel estimation or equalization methods and resource allocation in terms of time, frequency, code or power.

Indicative examples from a mathematical point of view are clustered MIMO channel modeling of multipaths where Taylor polynomial with respect to estimated delay value is employed. Another example is hardware design specifically amplifier characteristic the latter being of nonlinear nature in the case of excessive power received. Thus Taylor approximation of second degree or higher can be used to approximate the related curve. The linear polynomial can also be considered in case of amplifier moving away from saturation. Moreover, Taylor polynomial can be used to follow certain threshold crossings such as zero crossings provided that the Taylor polynomial order is sufficiently high that can accurately approach the zeros of the nonlinear curve to be approximated. As a final case of applicability, signaling waveforms of given shape and finite duration can be modeled by Taylor polynomial in the expression of the received symbol. It should also be reminded that approximation quality depends on the function i.e. curve to be approximated as well as the polynomial degree and expansion point. As signaling waveforms may be characterized by various shapes, the issue of complexity that constitutes the tradeoff for approximation achieved must be carefully balanced to ensure low complexity and properties of the curve that must remain valid such as zero

crossings, as stated above or close match in a predefined interval that preserves the validity of the optimization problem.

3.3.7 Conclusions and future work

This section realizes the benefits of applying CS theory and Taylor polynomial approximations in the context of wireless fading channel performance assessment. Both noiseless and noisy regimes of Rayleigh, Rician and Nakagami-m fading channels have been investigated. The metrics considered are capacity, distribution variances, number of bits for coding the channel and symbol error probability characterization.

As future work, the detailed statement of the CS optimization problem either by linear programming or by the use of greedy algorithms integrated with the approach of this thesis is a promising alternative. An extension of such an approach to nonconvex problem investigation has the potential of effectively modeling and analyzing 5G communication systems performance.

The convenient integration and differentiation techniques applied to polynomials could enable the derivation of near optimal solutions of complex mathematical expressions encountered often in the evaluation of wireless communication metrics. Finally, summarizing the benefits of applying CS in this paper the results of optimal performance and no additional complexity can be integrated with the optimization problems that enable computational and implementation complexity alleviation, in near-optimal decoding and receiver hardware implementation.

3.4 Cognitive Radio and Compressed Sensing optimization

3.4.1 Introduction

The results of this section are related to publication [C2]-([74]) which was co-authored with Kakalou et al. Spectrum utilization has emerged as one of most crucial resource to be allocated in all users exchanging information. Spectrum utilization limitations are twofold. The first issue is the excessive spectrum usage resulting into spectrum scarcity. In order for increasing number of users to face the spectrum scarcity challenge, high frequency bands are attempted to be exploited, the main limitation of which is the greater attenuation in wireless systems. The second issue is spectrum underutilization which leverages the observation that the instantaneous spectrum usage i.e. occupancy is a small fraction of the entire allocated spectrum. Hence, the spectrum is not fully exploited.

To address the spectrum sensing and allocation, CR [75] was verified as a promising technology for dynamic and effective spectrum usage, introducing the concept of primary and secondary users. The former are the legitimate users that have been allocated a spectrum zone while the latter opportunistically access the same spectrum either when the primary user is not using it or under the constraint that they do not cause harmful interference. Thus, CR is fundamentally based on opportunistic spectrum usage and employs various schemes to achieve optimal performance. Such schemes are transmission control to compensate for varying channel conditions and cooperative schemes, the latter possessing the potential to provide tremendous performance benefits

and optimal resource allocation. The CR technology has presented many spectrum sensing techniques [76]. Among those, the Energy Detection scheme [77] is the simplest to implement requiring no a priori information about primary user shape and parameters. Thus, it accumulates energy level and reaches a decision with the consideration of a predefined threshold. However, the Energy Detection method is proven to be vulnerable to noise uncertainty which includes noise variance variation, a practical assumption in numerous communication scenarios. Energy Detection also addresses the challenge of leveraging non-flat signal characteristics [78] in a computationally inexpensive way and handle practical SNR levels.

Compressive spectrum sensing [79] is a relatively recently emerging set of optimization tools which when applied to CR enables the reduction of samples along with accurate estimation. Hence, as dictated by CS principle it aims at alleviating complexity.

3.4.2 Cognitive Radio: Cooperative vs. Non-cooperative Schemes

In general, cooperation among communicating users [80] can provide benefits that overcome limitations in non-cooperative schemes. Towards mentioning representative issues where cooperation emerges as a promising strategy are: varying fading channel conditions, shadowing effects and hidden node problem, the latter commonly encountered in cellular networks. Fading channel conditions admit diverse statistical modeling approaches which reflect the properties of the environment that lies between transmitting and receiver locations. Coding strategies are extensively employed to mitigate distortion and achieve accurate decoding at receiver side with low complexity and reconstruction error. Relative to fading, shadowing occurs when transmitter and receiver are obstructed by large objects and hence, communication is realized via scattering and reflection causing the known multipath effect. Employing transmit power control, frequency planning along with CDMA technique are mainly leveraged to reduce shadowing caused impairments in a MIMO environment [81] thus also taking advantage of spatial diversity. The hidden node problem is another issue confronted by cooperation in wireless networks. Thus, in order to avoid limited communication with an intended node a dense node deployment scenario is selected as an effective solution always with the cost of additional infrastructure and transmission scheduling and coordination.

In our context, cooperation takes place between the opportunistic secondary users in an attempt to effectively use available spectrum under tolerable interference level with the primary user. The spectrum sensing task is initially performed at user level and the multiple decisions are sent to fusion center where a rule oriented decision is made with the aid of predefined threshold. The latter functionality can be characterized as centralized with the issues of complexity and computational overhead arising, decentralized where distributed algorithms are utilized and decision is made at user level and a hybrid scheme combining the above definitions in order to mitigate interference effects and achieve a synchronized and low complexity sensing framework. Cluster based cooperation is usually exploited in an attempt to create groups of users whose communication is separately coordinated. Hence, a cooperative scheme reduces overhead, provides efficient spectrum management [82] and compensates for cases of increased scalability and low complexity in a weighted

contribution of each cluster to the decision making procedure. Relay-assisted decision making [83] handles communication of users with diverse channel conditions and improves spectral efficiency. Routing discovery in a dynamic topology network poses a great challenge in a varying channel conditions scenario along with tolerable definition serving as the constraint to the fusion center based sensing optimization problem.

Underlay cognitive radio model [84] allows primary user and secondary opportunistic users to communicate simultaneously as long as interference is retained under an application specific threshold. In the overlay case [85], the secondary user listens to the primary user transmission and aims at using spectrum for its own transmission while optimizing primary user communication by means of coding techniques. Finally, interweave systems [86] detect available slots in time or frequency or code dimensions to ensure effective spectrum access. Among quality metrics that must be accounted for are detection probability, false alarm probability and misdetection probability, metrics that are of generic nature in the area of wireless communication mathematical framework. Detection probability refers to the convergence of decision about PU presence and the actual PU presence. The latter must be maximized in order to ascertain efficient balanced spectrum usage for both PU and SU in the context of non-harmful interference. False alarm probability is defined as the differentiation of decision of PU presence when the actual PU is not present. False alarm probability and detection probability are characterized by a tradeoff which must be given special attention for CR system performance. Finally, misdetection probability is defined as the decision in favor of PU absence in cases where the actual PU is present. All three above metrics must be carefully optimized so as to avoid spectrum sensing degradation. Below we provide the hypothesis test that includes PU decision of PU presence as H1 case and PU absence in noise only H0 case:

$$\text{H1: } y(n)=x(n)+w(n) \tag{33}$$

$$\text{H0: } y(n)= w(n) \tag{34}$$

In the above test formulation x(n) is the PU signal and w(n) is the additive noise term. Energy detection based on predefined threshold will be used in our derivation of decision optimization for fading channel conditions considered in the next section. Proceeding further with the cooperative strategies for decision making, hard decision scheme is founded upon related outcome represented as 1 for PU presence and 0 for PU absence combined with soft decision at fusion center level by rules such as AND rule, OR rule, Majority rule and K out of N rule. Thus, global decision based on number of decisions in favor of presence as well as absence with respect to a finite number of trials is based on the following formulae:

$$Q_D = \sum_{i=k}^{N} \left(1 - P_D\right)^{N-i} P_D{}^{i} \tag{35}$$

$$Q_F = \sum_{i=k}^{N} \left(1 - P_F\right)^{N-i} P_F{}^{i} \tag{36}$$

Also worth stressing is the fact that global decision performed by the fusion center is dependent on a plethora of factors such as network topology dynamics, energy constraints along with antenna selection as well as scheduling and fading channel conditions as optimization problem constraints. The complexity issue is another element that must be investigated.

3.4.3 Spectrum Sensing Techniques Classification

In this section, the CR based spectrum sensing techniques are categorized providing brief comments for each technique. The common aim is to maximize detection probability and spectrum hole detection.

Energy detection method conducts decision about PU presence via a predetermined threshold. Prior PU signal characteristics knowledge is not needed and complexity is kept low. However, performance deteriorates at low SNR regime and under network interference. Regarding further implications concerning ED method, noise variance must be provided where in case of variation [87] poses a challenge for the respective method. False alarm probability is also leveraged to form the ED scheme via knowledge of noise variance. In a sub-band division scenario the complexity must be kept low at the same time with mitigating spectrum leakage. A comment related to this aforementioned sub-band division is the concept of independence as a fundamental assumption. In this case, the derivation of convolution of distributions as the definition of the distribution of the sum of independent random variables can be translated to spectral domain by substituting convolution with multiplication. In this case, overlapping and sub-band consideration is a challenging approach.

Cyclostationarity is another sensing technique exploiting the homonymous property to perform sensing task. Thus, spectral correlation aids the differentiation of signal from noise. Estimation of spectral autocorrelation function enables blind identification of licensed PUs and remains consistent under low SNR models [88].

Matched filtering based detection is an optimal SNR maximizing approach with the strict constraint of signal characteristics knowledge at receiver side. The optimized performance in the presence of additive noise, as is the case of the next section, comes along with rapid sensing at the cost of synchronization ability [89] in order to achieve this SNR maximization. This, however, formulates an impractical assumption rendering this technique unpopular with the additional shortcoming of increased complexity.

Waveform based sensing [90] is based on exploitation of signal patterns about preambles or transmitted pilots and conducts PU detection based on cross-correlation metric of received signal and transmitted signal. Correlation aids in the above sensing process requiring synchronization and a fixed predetermined threshold. Hence, regarding this technique, correlation is what quantifies decision metric and, as being a practical assumption, does not incorporate the independent random variable model.

Covariance sensing method focuses its effectiveness on the fact that signal covariance [91] differs from additive noise covariance. In the inclusion of fading channel modeling, this method is a suitable sensing method and is usually requires bandpass

filters with Slepian sequences having the convenient property of bulk of energy in a specific frequency band. Spectral estimate reduced variance comes along with small spectral leakage.

Eigen value sensing [92] requires no noise variance knowledge but suffers from high computational complexity. By taking advantage of eigenvalue correlation, the technique achieves improved detection probability in a channel and noise blind manner while reducing noise uncertainty effect.

Wavelet packet transform sensing technique [93] relies on the property that the signal to be detected usually occupies low frequencies while additive noise is characterized by high frequency content. Thus frequency band decomposition leads to accurate detection. Additionally, this method favors wideband sensing by tracing spectral structure irregularities and dynamically adapts to such structures. One possible computational bottleneck arises when high rates are desired. This is where CS sub-Nyquist technique discussed in the subsequent part emerges in order to balance between structure and randomness always retaining low complexity and simple implementation.

Spectral estimation [94] derives properties in the frequency domain by leveraging the autocorrelation function and specifically its Fourier transform. The technique then searches for patterns such as spikes which lead to the decision of existing periodicities. CS could also aid PU signal detection in cases of partial spectral support knowledge.

Finally, multi-antenna technique [95] introduces the benefit of received SNR increase. Interference mitigation along with increased reliability and sensing at low SNR levels are the key advantages that this method provides. The basic technique is MRC. The logically straightforward limitation from adopting this approach is increased complexity and implementation. It is also indicative that CS could also aid in compensating for these limitations as they both are the exact benefits that CS theory guarantees i.e. reduced complexity and more efficient sensing.

3.4.4 Narrowband vs. Wideband Sensing

Depending on the range of frequency content as wideband sensing [96], the literature refers to task of sensing a frequency content that significantly exceeds the channel coherence bandwidth. To confront such a computationally and resource demanding task, the frequency content is divided in terms of utilized bandwidth in narrow spectra in order for the hypothesis testing to be applied and conduct decision making about PU presence. Wideband sensing requires high rate ADCs [97] along with a series of parallel filters which drastically increases implementation complexity. Following conventional signal processing, Nyquist criterion dictates that sampling rate must be at least twice the highest frequency present in the mathematical expression of the signal. Deducing from the above, CS appears as a very attractive methodology and optimization tools set to alleviate computational and implementation complexity. Further notion for CS theory and principles will not be detailed at this point as is extensively formulated in the mathematical background section. Hence, omitting problem transformation to a form for tractable solution and fundamental sparsity and compressibility properties upon which CS optimization is founded, we only provide the wideband sensing related benefit of applying CS as sensing a given frequency band with fewer samples obtained

or equivalently a larger frequency band for a given number of samples. Limitations of CS for spectrum sensing include noise uncertainty and recovery uncertainty, RIP proof [98] (a significantly computationally expensive task of verification) as well as unknown varying sparsity order and high memory storage requirements. As in other applications, spectrum sensing via employing CS mainly depends on the balance of randomness as the direction taken for optimal CS based signal reconstruction and structure existing which is what favors feasible implementation in such scenarios. Relative to implementation, it is the efficient manner to avoid costly or impractical ADCs for wideband sensing that verifies the merit of CS optimization. As in the entire thesis adopted, additive noise is modeled as Gaussian and the case of unknown variance is also addressed in CS literature as a more realistic assumption. A last comment concerns CS application for sensing in CR systems as twofold: centralized or distributed [99]. The latter emigrates to WSN design dealt with by a fully randomized Gaussian scheme in section 3.6. Thus, a comment sufficing is the feasibility of applying CS in a hybrid scheme which is the crucial benefit resulting from the model assumed in the next section.

3.4.5 Spectrum sensing in related dimensions

Spectrum sensing techniques are not limited in one dimension such as time but can be extended to other dimensions. In this subsection, we briefly outline the manner in which sensing is conducted in each dimension in the following table:

TABLE 3.5: Spectrum Sensing Dimensionalities

Dimension	What is sensed?	Comments
Frequency	Opportunity in frequency domain	Availability in the frequency spectrum. Not all bands used simultaneously.
Time	Opportunity of a band in time	Availability of sub-band in time.
Geographical Space	Location/Distance of PUs	Occupied or available spectrum in the same geographical area.
Code	Spreading code, time hopping or frequency hopping. Synchronization estimation can be avoided with long random code usage.	Use of orthogonal codes to mitigate interference. Opportunity in code domain i.e. detecting spectrum usage and also multipath parameters.
Angle	Directions of PUs beam and locations of PUs	Spectrum opportunity in angle dimension. PU transmitting in a certain direction while SU transmitting in different direction causing no interference

A concluding remark that must be made accounts for the fact that sensing can be performed simultaneously in more than one dimension [100] and can leverage code domain to effectively overcome limitations such as channel distortion confronted by retransmission or expensive feedback as well as computational complexity for timely and accurate decoding.

3.4.6 Algorithmic derivations for Spectrum sensing

In this subsection, several algorithms and their performance are analyzed from existing literature approaches. The derivations as well as comments on problem assumptions and constraints are thoroughly interpreted in terms of the investigation of the next section in this chapter. Also noted that this subsection reviews past related work in a more generic context while the subsection reviewing past literature in the next subsection presents our analysis in a more concentrated on the specific Energy Detection literature.

A triple algorithm proposition from a game theoretic perspective is provided in [101] along with incomplete network information distributed learning algorithm classified as evolutionary algorithm by means of robust game theory. The issue of excessive overhead by requiring channel statistics to be known, as is the case with our proposition in next section is mentioned to be tackled in the algorithm proposed. Moreover, the scheme is non-cooperative and leverages imperfect CSI. A technical remark included in this paper is that tolerable interference comes along with spectrum underutilization, hence CS application being based on such observation could optimize interference even further. Incorporating fading channel models an energy detection based effort with statistical methodology as in the next section could integrate sensing scheme simplicity and mitigate interference. Regarding spectrum access, the optimization problem posed is of probabilistic nature similar to the mathematics of the next section for both accurate and inaccurate spectrum state assumptions. An interesting point of view adopted relates to the statement that sensing strategies diversity enables collision avoidance and interference management. As stated by the authors in the above work, the evolutionary algorithm may induce overhead and hence an alternative version applicable in a distributed sense is proposed with the desired property of no excessive information exchange. Verifying the relation of the above statement to the results in the next section, complexity may indeed arise as an issue as a consequence of independence and convolution operation. Investigation of how a Nash equilibrium strategy could relate to the above along with the practical constraint of inaccurate spectrum state is left as future research. A finalizing remark for this paper concerns the parameter m as number of SUs which possess a value for optimal performance. This is in agreement with choosing optimal number of samples for largest percentage of deciding in favor of PU presence. High throughput and low collision could be investigated in the future under the mathematical framework adopted in the next section. Another work [102] proposes an artificial immune algorithm in a centralized, cooperative and interweave spectrum access scheme as an energy detection paradigm for enhancing capacity and mitigating interference. The first observation matching the forthcoming analysis in the next section is the fading realizations and additive noise independence in prerequisite of our case and the above paper. The optimization problem either in constrained or unconstrained form in the immune algorithm context can also be applied in the derivations of the next

section by accounting for fading channel models and dynamic threshold adaptation. Finally, threshold and sensing time optimization could be defined by the CS related principle as a comparison benchmark along with extension of convexity consideration which, as CS theory dictates, could be solved by linear programming or greedy algorithms. The integration of the latter with clone selection algorithm could provide performance insights for the energy detection sensing schemes. An excellent work of Bkassiny and Jayaweera [103] proposes an active channel detector in a wideband sensing regime with non-Gaussian noise statistical assumptions. Proceeding one step further, the case of partial noise model knowledge is considered and a robust detection scheme with reference to a nominal distribution is proposed based on quadratic complexity nonlinear regression. The convolutional scheme of the next scheme can be made to address the spectral activity after dividing the wideband range in to smaller sub-bands. However, the independence assumption can lead to excessive computations given its integration with the wideband sensing regime. Nevertheless, the merit of non-Gaussian assumption translates to the analysis of the next section with the difference that fading distributions refer to a multiplicative effect as opposed to additive noise. Moreover, the concept of contaminated distribution could potentially build further upon the non-Gaussian noise models and also question its robustness in terms of CS constraints. Hence, the analysis of the next section can be extended for contaminating all fading distributions considered. Specifically, l_1 norm selection replacing l_2 norm relative to energy can also be integrated with CS principle for the case of heavy-tailed distributions. A quite representative work [104] proposing an energy detection based scheme for hard decision interweave spectrum access addresses the tradeoff between energy conservation and sensing performance by means of evolutionary game and coalition formation algorithm. A very clarifying comment concerns the issue of reducing participating users in a cooperative scheme when sensing performance is acceptable. However, insufficient number of users may degrade this performance. This carefully balanced tradeoff could provide an interesting extension of the next section where the number of i.i.d. samples could correspond to a case specific user selection of the cooperative scheme. This user selection is also essential in order to alleviate excessive overhead. Furthermore, the optimization of detection and false alarm probabilities in this user selection context could provide dynamic decision threshold which in our case could relate reliability with diverse fading condition consideration as is the case in the next section. Also worth noting is that our analysis can be extended by including memory of fixed or varying order not only in terms of deciding the proper action to take from a selfish point of view, as formulated in [104], but also to integrate the diverse fading conditions of our analysis for optimizing the selfish action to be taken in a scenario that favors non symmetric distributions. Varying channel statistics also pose a modeling challenge. The latter selfishness attribute poses an entire novel optimization problem by differentiating solvability in the cooperation context. Another point of similarity is the hard decision making in both the aforementioned paper and our analysis. Regarding majority vote rule, our analysis could benefit from setting a percentage above which sensing is acceptable instead of merely comparing the percentages of each case. One step further, the beneficial requirement of not being predictable so that not to be exploited could well fit to the CS case in terms of incoherence so as to interpret the derive percentage in our case. A concluding remark

on the aforementioned work is the entropy increase which causes secondary users to be hesitant of sensing the respective channel, an applicable consideration of the fading cases in our analysis. An attempt to improve sensing performance by differential evolution is proposed in [105] in a cognitive sensor network thus introducing the scalability assumption. The probabilistic flavor of the stages of differential evolution could also be applied for performance of energy based spectrum sensing in our analysis. The related iteration process could either define our analysis in the independence concept in terms of optimality or excessive computational complexity reached thereby terminating the process. What stresses the merit of this paper is the ability of differential evolution approach to allow the rapid withdrawal of SU opportunistically using a frequency band so as to effectively reduce interference. This statement does not match a property of our analysis the latter only implying a complexity issue. This resulting complexity could be alleviated by certain a priori knowledge in the sensing optimization problem, the latter made feasible by this rapid SU termination of using a given frequency band, an approach that is based on scheduling. In overall, the concept of cognitive sensor network being dense sensor deployment could build upon our approach in the next section by simultaneously considering traffic congestion, intelligent differential evolution spectrum sensing along with diverse fading conditions and CS sparsity rule. An excellent paper by Huang et al. [106] proposes a cluster-based cooperative interweave spectrum access scheme via non-parametric Bayesian rule. The authors in this work provide a thorough statement about advantages and shortcomings of centralized sensing methods. Hence, in our work the distributed class of sensing strategies can be utilized at the first stage along with cooperation to alleviate complexity. The above cooperation implies correlation of sensing data in space dimension and is a promising direction for addressing excessive overhead. It is also intuitive that intra-cluster spectrum information implies temporal correlation which is to be exploited as well for improving sensing performance. Furthermore, the cooperative proposition made in the latter paper also confronts orthogonality which could ensure improved performance along with resources reuse. Sensing data diversity in heterogeneous CR networks or equivalently in a network where SUs are different distances apart from PUs and thus experience diverse channel conditions. This statement is directly related to our case and it is important to trace the merit of cooperation in this scheme integrated with CS sparsity and incoherence. Sensing data heterogeneity could also fit to our analysis of diverse fading conditions that could be incorporated in a single varying fading network. Regarding fading channel model assumed, the authors in the latter work assume the envelope related Rayleigh fading as both components are Gaussian zero mean distributed while our analysis assumes diverse fading scenarios. The bridging point for the two above could be the Nakagami-m fading distribution which approaches the Rayleigh fading by proper parameter selection and could be used to assess sensing performance with the independence principle preserved. Additionally, the Dirichlet process related distribution derivations could relate our analysis by incorporating prior knowledge in terms of adaptively setting detection and false alarm probability values and fading distribution cases. As already stated, complexity is a major issue arising from the analysis in the next section. Hence, the reduction of computations could originate from either introducing correlation as opposed to independence or the CS principle exploiting structure, the

latter resulting from mathematical relevance of considered non-symmetric fading distributions. The proposed method also exhibits robustness to noise and has the potential to leverage dynamic channel availability by not considering number of data groups. A last essential tradeoff arises in [106] as that of cluster number versus sensing performance. This could lead to improvement and relate to our analysis if an area of interest in the coverage of the CR network is preset, which could combat fading impairment by considering the same diverse fading conditions. Another remarkable work [107] introduces a cooperative centralized spectrum sensing scheme by Bayesian rule based on deep sensing. The latter is what renders this approach beneficial compared to existing methods along with the reference of mobility and time varying fading having a detrimental effect on sensing performance. Translating to our analysis in next section, derived optimal number of samples could have different outcomes based on mobility as this concept differentiates performance optimization problem including mobility as a constraint to be accounted for and also model fading by non-static distributions. These sample derivations could also enhance a dynamic model for fading time variation along with claiming independence principle with respect to time which fits into our convolutional operation calculations assuming statistical independence as a prerequisite. Furthermore, while our analysis considers fixed threshold for all cases and invariant false alarm probability, the authors in the above work comment further on the optimality of determining dynamic threshold from which false alarm probability is determined by gamma function calculations. The above contrast could provide a benchmark for sensing performance assessment in diverse fading conditions in a dynamic state space sense. CS consideration as a special case in our analysis could be extended via the latter paper which implies the varying sparsity order along with hidden states estimation. In overall, the combined PU existence state and fading state could be embodied to the variety of fading conditions in accordance to false alarm probability optimization and complexity related measurement history. It is also worthwhile noticing that the authors in this paper contemplate the transitions of PU states by setting a counter that counts transitions, an approach which could extend our analysis not by merely calculating percentages but also transitions in terms of fading cases. In direct relation to our analysis is the statement in [107] that a large sample size, as in the derivations of our case, may increase complexity as well as sensing time and transmission efficiency. The latter issue is closely related to fading conditions and could thus improve dynamic spectrum utilization. Proceeding one step further in the analysis, the correlation, as opposed to independence in our case, could improve performance and be integrated with fading channel dynamics via the deep sensing scheme. The contrast of a dynamically adjusted false alarm probability, as opposed to a fixed scenario where the respective probability is set to the target value could provide sensing comparison in the already stated fading cases. As a concluding remark on this paper, our case could be beneficially extended to mobile scenario and flexible resource allocation in a manner including correlation effects and compensating for emerging issues on 5G cognitive systems sensing performance. A centralized ED based method of spectrum sensing based on modified majority rule and cooperative hard decision is proposed in [108] relying on the practical multiple/adaptive transmit power levels assumption which is proven to be optimal in performance loss compared to cases of considering lowest, greatest and average power levels in existing algorithms. The

reviewing comment of multiple antennas at SU i.e. exploiting spatial diversity/degree of freedom is an intriguing consideration while the contrast of the former to an equivalent cooperative scheme is left as future work based on the derivations of our analysis. Moreover, the PU transmit power knowledge promotes spectrum sharing and enhances the «intelligence» of the cognitive system. The merit of this work is traced in prior probability assignment and the adoption of multiple hypotheses concept. It is also very interesting to query whether the multiple power level and the multiple SUs as assumptions could lead to a resource allocation scheme of increased combinatorial complexity as well as the ascending order sorting could have an impact on the multiple hypothesis problem. Together with CS methodology and non-Gaussian signal consideration, a research extension is formulated. Another rather intuitive issue arising from the analysis in [108] is the impact of false alarm probability on the unnecessary complexity order in terms of the approach that first decides over PU presence and, as a second stage, determining the transmit power level in use. Integrated with dynamic sensing threshold and fading conditions, an indicative means of improving performance is the sensing history and the correlation effect contrary to independence assumption of our case. Accordingly, decision region formed as a result could provide valuable insights. The notion of seldom use of power level corresponding to a decision region could also admit an interpretation of an entropic concept which would attribute large uncertainty of these decision regions and also jointly consider noise variance magnitude. Regarding cooperative decision making, the number of SUs for a certain decision could be combined with the optimal number of samples corresponding to the maximum percentage in favor of PU presence. The particular case of equal majority votes with respect to different states could potentially be addressed by CS sparsity rule if existing structure is exploited. Finally, the coupling of PU state and power level determination in a multiple power level case implies that the error probability in either stage needs to be carefully balanced and also compensate the fading conditions and number of samples in favor of PU presence as investigated in the next section. Referring to the work of Salman et al. [109], a relay-assisted cooperative soft decision ED based scheme is introduced for number of users and received signal information joint detection. A strategy reviewed in this paper is that of SUs with high detection probability aiding SUs with low detection probability. This could be related to the goal of achieving a target value and be also integrated with the local level different fading conditions and interference levels. The algorithmic derivation of the respective paper also focuses on the minimization of mean square error and correlation errors which can very easily relate to our search for optimal number of samples achieving the goal of highest percentage in favor of PU presence. The combined effect of genetic algorithm and CS sparsity/compressibility rule on the basis of multiple extrema and the ability of the latter algorithm not being trapped in local minima constitutes a challenging topic for further research. It must also be noted that the relay-based calculations come in a weighted manner in accordance to the two aforementioned error derivations. Finalizing with our comments for [109], the increase of number of sensors in each ULA induces a complexity issue which is not investigated in combination with SNR levels and resulting errors existing in this paper. Regarding to the relation with our analysis in next section, a comment would suffice stating that fading conditions in relay scenarios are already maturely studied in literature leaving the notion of joint PU presence and signal

related parameters estimation as a case to be investigated. The paper of Hwang and Lee [110] addresses cooperative spectrum sensing performance proposing imperfect feedback based approach via differential evolution algorithm in a centralized hard decision scheme. This imperfect feedback assumption is investigated under Rayleigh fading, thus can be applied to our fading diversity and additive noise based analysis with the channel impairments affecting decision information feedback channel and assessing the impact of such imperfection on decision making and overall sensing performance. In the same context, sensor selection and mobility assumptions could provide performance benefits as well. The probability of detection is sought to be maximized while false alarm probability is assumed to be upper bounded i.e. a Neuman-Pearson criterion. A very interesting remark in the above work concerns the probability derivation of PU presence to be expressed as a function of probability of PU absence. This could provide the means for accurate sensing performance as well as complexity reduction. In our case, this could translate separately for each fading assumption and also preserve independence in the related optimal number of samples for each case. Another important point to be highlighted is the outperforming proposing scheme as opposed to MRC, based on which a more detailed consideration of a MIMO environment with multiple antennas at transmitting and/or receiving side with the intuitive issue of implementation complexity accounted for. Furthermore, the merit of the proposed quantization combining is verified by same detection probability with less information requirement. However, the reduced sensitivity to channel state estimation error requires investigation of whether it relates to a performance tradeoff. Another centralized ED based cooperative soft decision underlay spectrum access method is presented in [111] by investigating whether interference is harmful or not by leveraging Bayesian Active Learning rule. The objective of this approach is to learn interference channel gains by allowing reverse link feedback. In the analysis of the next section, fading conditions could be included in interference modeling and provide a challenging model for joint consideration of the above particularly if the non-Gaussian statistics are adopted. A similarity point between the latter paper and our analysis is that optimal number of samples derivations are also dependent on the distributions parameters. Hence, this property combined with SINR conditions for interference learning provide a challenging multiple parameter optimization problem where fading and additive noise can be jointly considered. However, a requirement of high data rate may result in an outdated feedback or detrimental channel distortion along with increased feedback history required to combat uncertainty. Hence, the modulation and coding selection scheme mentioned by the authors in [111] may require re-consideration. Following the methodology in this paper, the main assumption is the grouping and spreading of multiple users across available bandwidth, an approach that in terms of spreading concept is already applicable in communications literature with respect to the different dimensions of time, frequency, space and code. Bridging the above with our case in next section, the optimal number of samples derived could change with respect to multiple user constraint in the optimization problem along with bandwidth allocation for each group. Towards a low complexity and reduced overhead approach, CS sparsity principle could be efficiently applied by exploiting for instance structure in the latter constrained optimization problem and achieve balance with the probabilistic harmful or tolerable interference in user subsets. The algorithmic prerequisite assumption in this

paper is that of initial maximum uncertainty regarding interference channel gains and raises the logical question of the number of probing attempts required to derive the desired gains from a diverse fading condition point of view. Another very indicative comment concerns the requirement of diverse direction for deriving interference gains which is fulfilled by incorporating randomness, an aspect directly related to CS random measurement practice for overcoming deterministic limitations. This can further interpret the derivations of our analysis in the next section enhanced by the hyperplane geometry analysis conducted in [111]. Furthermore, regarding this uniform hypersphere point picking, potential structure as a measure of dimensionality could be effectively leveraged by CS to reduce complexity without compromising derived algorithm results optimality. Solution uniqueness could also be considered with convex relaxation methods as in CS theory and principles. Elaborating more on CS, the context of solving MAP estimation problem by alternative to convex relaxation methods could also be extended for a convex programming and greedy algorithm performance comparison to the other methods. Towards the investigation of a centralized sensing approach, the authors in the aforementioned paper provide a statement of the exponential increase of computations as a consequence of network size increase. This statement admits a double interpretation: the extent to which centralized scenario is adopted as opposed to a distributed one for relaxation of the localized computational burden and the feasibility of applying complexity alleviation-oriented algorithms that are applicable in both architectures with corresponding sensing performance tradeoffs. The above contrast constitutes a promising future research direction. Concluding our remarks, the observation of faster estimation causing less interference could lead to the proposition of predicting interference in a diverse fading environment by merely observing speed by learning statistical parameters. Following the same context, instead of deriving optimal number of samples as in our case in the next section, a worst case analysis relative to fading and interference as well as network dimensionality and computational burden could be jointly optimized. A centralized cooperative soft decision interweave spectrum access scheme is proposed by Wu et al. in [112] by accounting for average error probability minimization and throughput maximization by means of low complexity algorithm and optimized parameters investigation. The first statement worth stressing in this paper is the already fundamental hard vs. soft decision tradeoff. The latter provides accuracy at the expense of increased overhead and consumed energy. Moreover, quantization levels and embedding localized decisions in most significant bit of quantized bits promotes the feasibility of an efficient coding scheme for effective information transmission by simultaneously achieving reduced complexity decoding and sensing performance. Regarding system model prerequisites, the event of parameter changes directly requires updating, an approach which could translate to our analysis as optimal and stable updating of fading distribution parameters to accurately learn the channel in a real time setting combined with error probability and throughput optimization. Hence, the multi-bit regime could result in overhead that could potentially be balanced by efficient variable-length coding. A natural question arises as to what extent the sensing performance can be optimized under the low complexity requirement by considering the raw measurement statistics. Clearly, the quantized soft decision fusion scheme results in a global decision but the statistical impact in a non-Gaussian setting remains a challenge especially by considering the CLT

case as in our analysis and a varying sparsity CS rule as a practical assumption. Inclusion of fading conditions and their impact on the linear search algorithm implemented in [112] is also a feasible future consideration. Therefore, this extra parameter could require a more detailed investigation in terms of optimal results reached. As a final point worth highlighting is the further query of the effect of sampling frequency with CS based sub-Nyquist sampling strategy along with number of cognitive sensors and throughput maximization in this sparsity regime.

We now proceed with spectrum sensing literature review with CS methodology applied. The exceptional work of Cohen and Eldar [113] proposes a sub-Nyquist sensing scheme which exploits cyclostationarity structure for sensing efficiency and noise robustness. The authors state in a thorough manner that cyclostationarity is a compromise between low SNR sensitive energy detection scheme and matched filter approach requiring perfect knowledge or received signal. This statement clearly points towards a future research direction concerning our analysis with a more concise CS sub-Nyquist sampling concept consideration in the context of the latter two other methods. Another important issue raised is the method for conducting sampling in order to acquire correlation measurements. Hence, in a CS sub-Nyquist regime compressibility and existing pattern/structure could provide a viable implementable solution to this problem. Cyclic spectrum recovery can be particularly investigated in a bounded as well as non-Gaussian noise assumption, a problem to which CS sub-Nyquist sampling could offer attractive ways of overcoming cyclic spectrum recovery limitations. Towards a bridging between our next section analysis and the indicative results reached in this paper, the multiple frequencies resulting from the fundamental notion of Fourier series i.e. higher order harmonics can be integrated with the noise robustness of fading environments along with combining optimal number of samples with sub-Nyquist rate in the context of the simplification of the whole process requiring only estimation, user presence in a frequency band and not signal reconstruction. The latter harmonics implying transmission correlations pose an interesting problem in inserting small correlation properties to compare with our independence based analysis. The authors also provide an insightful geometrical interpretation of the correlation induced observations i.e. self-correlation and cross-correlation between frequency bands. This analysis admits an integration to our case in the same manner mentioned above. The reference of multicoset sampling introduces the very crucial issue of irregular sampling as a sample selective version of regular sampling and its impact on sub-Nyquist sampling exploiting cyclostationarity. Hence, samples may be characterized by an optimal number and locations. Extensive analysis of the autocorrelation matrices stemming from the geometrical self and cross correlations preceded is conducted by determining the location of nonzero entries with respect to frequency shift. A straightforward claim is the application of CS compressibility rule to the related observations and results with the additional claim that nonzero location randomness can also be consistent as a proposed approach by emigrating from random or irregular sampling to optimization random transformation matrix. Relevant to the resulting sensing performance the authors in [111] provide a notion of a feasible increase in false alarm probability in power spectrum recovery problem in the ED comparison benchmark that they use which holds for low SNR regime as opposed in cyclostationary detection. Concluding with this work, the impact of noise deteriorating performance is

also verified in the proposed cyclostationary detection which enhances the understanding of the noise robustness property relative to performance of spectrum recovery. We continue our literature review with a power spectrum sensing approach [114], which as stated above in [113], is a special case of cyclic spectrum recovery by the same authors, namely Cohen and Eldar. Thus, minimal required sampling rates are derived in the noiseless regime, an approach that could admit the extension of considering additive Gaussian noise and specified parameter impact on performance. With respect to the noisy case as an extension, including the non-Gaussian non-symmetric property and studying the impact on performance and how these assumptions translate as a mere rate adjustment, correlated measurement history or consideration of existing structure in a different manner. A crucial remark on the proposed approach in [114] involves the blind power spectrum estimation as well as the non-blind scenario where the reconstruction is based on occupied frequency bands. This could further aid the sensing accuracy of the respective problem. The independent case as adopted in our case may not be valid for considering this approach as opposed to the practical correlation assumption. An issue of concern is the derived transformation matrix from which the Nyquist samples are filtered to derive sub-Nyquist samples. In such an approach, complexity may be an issue to overcome especially by incorporating noise with a minimal cost in excess samples and structure exploitation in both time and frequency domains. Towards achieving sub-Nyquist sampling efficiency in the same context the number of samples in our analysis could also be included in a sampling pattern consideration thus alleviating computations and simplifying hardware in a fading environment along with considering a penalty of minimal sampling rate in an equivalent blind scenario. With the aforementioned paper and our analysis both being conducted in the time domain, we convolve distributions for calculating the ratio above or below a threshold by the independence assumption while the authors in the latter work derive autocorrelation matrix by convolving samples a similarity that could provide the means for extending our analysis. Another contrast is the dimensionality reduction requirement stemming from the multiple potentially occupied bands considered simultaneously as opposed to our analysis involving a single hypothesis test problem. To that end, CS sparsity rule could greatly aid in achieving the aforementioned goal. As in the above work, a parametric evaluation of sub-Nyquist sampling emigrated to our case could feasibly consider the parameter value selection of the fading distributions jointly with the varying parameter case in the paper for sensing performance. We now finalize with a detailed and concise survey [115] by Cichoń outlining energy efficiency in cooperative spectrum sensing context. The first remark in this paper, rather straightforward, is that channel information (initially at receiver side) and measurement history i.e. system memory combined with cooperative principle result in improved performance. Again the detection and false alarm probabilities tradeoff is highlighted in terms of cooperation, noting that either of the above can be modeled by a specific target value while optimizing the other in a joint consideration. Hence, our case considering fixed false alarm probability in a non-cooperative scenario can be obviously extended. In a relative future work scenario, the entropy based evaluation of such probability metrics poses a great challenge to be addressed. Moreover, in the context of a priori knowledge, the definition of a blind sensing problem could be defined in a variety of ways by incorporating full or partial

knowledge of many parameters. Concerning interference, the energy efficiency bottleneck could be integrated in an approach that not only validates whether interference falls below a certain threshold but also in a dynamic case of continuously improving sensing performance as interference level continuous reducing its value. The soft decision quantifies better the decision quality compared to hard decision and may also utilize mathematical tools such as convolutional statistics in the analysis of the next section at the expense of increased complexity and computational overhead. However, the quantization schemes applied to soft decisions are also points of concern. Cluster based approaches could be modeled as discretized constraints to the optimization problem on the basis of cluster head selection and the choice of performing local decisions or forwarding data to the fusion center the latter forming the related global cooperative decision. As for distributed schemes, the notion of node neighborhood must be considered, in terms of relaying absence, along with confronting impact of probability metrics mentioned above. The strategy of dynamic user selection with which to share the global decision reached could also enhance security of the CR network reflecting whether a centralized, cluster-based or distributed topology is assumed. Relevant to time domain stages of sensing commented on this survey and the user selection for decision circulation, the latency issue naturally arises as well as its impact on sensing accuracy and false alarm probability. In the same sense, the complexity of user detection in time could be potentially reduced by a probabilistic analysis that predicts frequency band vacancy in a certain time interval. As an extension of sensing and spectrum access time which must be jointly balanced the fading conditions assumed pose certain limitations in the first place to the intervals being solutions of the above joint optimization problem and also require revisiting of the derived throughput results. To this end, cooperation further improves sensing and throughput. User selection is also reinterpreted by means of a subset conducting spectrum sensing functionality and another subset for spectrum access the latter distinction relying on cooperation as is understood. It is also straightforward that a varying CS based sparsity order that may characterize spectrum utilization along with channel coding technique may alter transmitted sequence and promote energy savings. Stemming from network perspective, cross-layer sensing optimization poses a state-of-the-art challenge in current CR systems. Rather than the simple formulation of the total consumed power in the survey, the derivation of consumed power may be excessively more complex and also involve nonlinearities given the type of constraints included in the optimization problem. The independence assumption can be substituted for correlation of sensor readings that may permit selection of sensor subset as well as excluding users that experience severe fading limitations from their environment. The logical query from the above approach is to what extent the sensing performance degradation stemming from user exclusion can be minimized. Concerning bitwise energy consumption, efficient coding techniques can significantly lower required energy requirements and be combined with less sensing time while achieving the same sensing performance. The node/user selection strategy is also reported in this survey [115] and made clear that it can be applied under a variety of assumptions i.e. the disjoint grouping or censoring approach and the dimensions of space, time, frequency or code. Hence, the formulation of sensor selection boils down to a well posed constrained optimization problem in a probabilistic and application specific sense.

Clearly, SNR criterion in a diverse fading environment may be ineffective due to rapid variations and thus may compromise performance. High SNR node selection also admits CS compressibility rule application, as a sparsity approximation problem, given fading and additive noise, for low complexity optimal performance. An SNR-based alternative approach favors uncorrelated or weakly correlated sensing decisions closely related to our approach. This could translate to optimal number of samples based on uncorrelated node selection scheme along with the challenging delay minimization requirement. Furthermore, the methodology of voting is such that severe fading or shadowing effect could directly degrade performance. This indicates a clear direction of future investigation of our analysis with the relevant channel condition partial knowledge as prerequisite and a priori knowledge of PU presence information considered as a constraint in the sensing algorithm applied fitting to a practical scenario. In the same context, a point of concern would be jointly optimizing detection and false alarm probabilities. Cooperative selfishness integrated with relaying is another promising direction of research as a means to combat channel impairments and boost energy efficiency. The final assessment of this survey is the consideration of the optimization problem in a multidimensional setting effectively formed by the metrics involved, their quantification and, as a complementary notion, the dimensions of time, frequency, space and code. Concluding, we suggest the dimensionality structure-induced reduction offered by CS principle to further assist to lowering complexity, improving sensing accuracy and achieving higher energy efficiency. What must be stressed in applying CS theory is the careful identification of degrees of freedom of this problem.

3.4.7 Conclusions and future work

This section has surveyed spectrum sensing techniques providing analysis and derivations of the respective advantages and limitation arising from their practical implementation in CR systems. The methodology of confronting wideband sensing cases by dividing spectrum to narrowband regions has been extended to the CS spectrum sensing methodology allowing fewer samples for spectrum sensing efficiency outperforming Shannon-Nyquist sampling theorem and achieving low complexity and resource management gains.

The extensions of this section to improved sensing methods in the context of future generation systems exhibits CS spectrum sensing applicability by exploiting structure i.e. patterns in the spectrum of PU and SU context.

3.5 Convolution Energy Detection based Scheme for Cognitive Radio

3.5.1 Introduction

The results of this section are related to our publication [C5]-([116]), a probabilistic convolution based detection scheme is adopted the decision being based on a threshold the value of which is assumed by the Gaussian statistical case. The numerator of the fraction is the sample generating distribution indicating primary user presence to the noise distribution thus primary user absence. The percentage of number of fraction values above threshold is evaluated for the following cases: the CLT theorem based

number of samples, the CS theory based number of samples and the asymmetrical, compared to the Gaussian case, fading channel distributions, namely Rayleigh, Rician and Nakagami-m fading distributions expressed by Eqs.(3),(6),(9). An algorithm is formulated and simulation results are interpreted.

3.5.2 Mathematical Preliminaries

The threshold on which the decision is based is given by the following equation:

$$\lambda = \sigma_w^{\ 2}\left(Q^{-1}\left(P_{fa}\right)\sqrt{2N} + N\right) \qquad (37)$$

where N is the number of samples, σ_w^2 is the noise variance and P_{fa} is the probability assumed to take the value of 0.01, a value that conveys accurate estimation. The CS based formula for the calculation of preserved number of samples is given by Eq.(18).

The sparsity k is related to the number of samples N by the sparsity ratio. The assumed value for this ratio stems from the percentage of utilized spectrum in the 5-6GHz zone that is 10% of the allocated spectrum. This results in sparsity ratio equal to 0.1 using Eq.(18).

3.5.3 Past related work

Before we proceed to the formulation of our convolutional approach we present indicative past literature that specifically fits into the context of the analysis of this section i.e. ED sensing methodology.

Compressive spectrum sensing approach for Rayleigh channel ED based scheme is proposed in [117] by means of compressed measurements. Thus, this work aims at addressing fading environments the lognormal fading being already investigated in past literature. Contrary to our work assuming integer fading parameters, the non-integer regime poses a great challenge that could be confronted in diverse fading with the feasibility of soft decision accurate sensing scheme. According to the CS based scheme applied, the Fourier series representation stemming from the DCT sensing matrix could provide the means of comparing results from time and frequency domain in a sub-optimal sense as a result of truncation of respective series representation. It is also straightforward to extend this analysis to other fading cases initially from the Nakagami-m properly selecting m value. The latter extension should also incorporate detection and false alarm probabilities in closed-form considered to be beneficial in terms of tractability. Finally, the low compression ratio assumption must also be included as a means of lowering complexity. A paper by Singh and Mitra [118] remarkably illustrates performance degradation induced by multipath fading as well as introducing CLT related results as an increased number of samples in a practical scenario. The latter work also boldly stresses the maximum detection probability related number of samples which fits into our analysis considering Rayleigh, Rician and Nakagami-m fading. The notion of error-floor for accurately setting the decision threshold is related to our standard Gaussian distribution model for the additive noise considered which due to its probabilistic nature can be linked to the approximation we can only calculate regarding noise floor. Given that our analysis can be extended to the frequency domain, the number of samples derivation in the aforementioned work can

be related to time-bandwidth product which in turn relates to uncertainty principle concept. The latter can be investigated from the diverse fading and additive noise conditions assumed. Arriving at the mathematical framework of this paper, the similarity is apparent and enhanced in our case by convolution operation along with likewise fixed false alarm probability. Another remark could include parametric evaluation such as false alarm and detection probabilities, fading parameters reflecting severity as well as CLT by quantifying approximation of normal distribution which also compensates for noise variance and user signal distribution if assumed random. Furthermore, Kishore et al. in [119] address energy detection performance including Rayleigh and Rician fading. The differentiation with respect to our work is the Nakagami-m fading distribution which, although separately parameterized and evaluated in our analysis, can be mathematically linked to the two former fading distribution by means of proper parameter selection. Hence, this intermediate fading case also constitutes an interesting direction. As to the prolonging of sensing time thus difference in energy quantities, the threshold must be optimally reset. To that end, cooperation and relaying could assist in sensing refinement always at the expense of implementation and computational complexity. Required SNR derivations for Rayleigh and Rician fading verify the LOS component prerequisite in Rician fading translating to lower SNR required with respect to Rayleigh. Compared to our analysis, there are two issues emerging: the additive noise consideration in the fading scenarios is assumed in our case but not in the latter paper while the second issue is the very interpretation of the findings of this paper that sets the stage for CS based spectrum sensing in a manner of refining sensing and reducing complexity beginning with the non-adaptive approach and leaving the adaptive sensing methodology as a feasible future investigation. The aforementioned remark on the interpretation of the Nakagami-m distribution verifies the merit of another work [120] which proposes collaborative sensing as a means to improve energy detection under Rayleigh and Nakagami-m fading conditions. The authors in this paper also perfectly distinguish advantages and limitations of spectrum sensing methods in cooperative and non-cooperative manner. As comments briefly summarizing the tables in this paper, ED is beneficially blind but is ambiguous in low SNR cases and cannot be applied in spread spectrum signaling. The statement that false alarm probability is not a function of fading thus promotes our fixed value assumption in our analysis. The challenge of considering fading is also evident by the closed form expressions of detection probability emigrating from Marcum q function in AWGN to complicated exponential and hypergeometric functions emerging in Rayleigh and Nakagami-m fading cases. In the collaborative scenario, it is apparent that the resulting simplified expressions are solely based on independence which we assume in our analysis as well. Worth mentioning reviewing this paper is the performance degradation with increased time and bandwidth product. As already stated, this product metric could feasibly be combined with CS for detection probability maximization. Another observation of the latter paper is the outperforming of the collaborative scenario with respect to the AWGN channel, which indicates an interesting performance advantage given the optimal compared to fading, Gaussian case. However, it is straightforward that complexity arises as a tradeoff for performance. Concluding with this paper, the above comments point towards an attractive future research. ED method for fading distributions considered above i.e.

Rayleigh and Nakagami-m is also derived in [121] by Yadav and Agrawal. The contribution of calculated detection threshold in a dynamic manner is traced to the derivation of detection and false alarm probabilities whereas in our case the inverse calculation is utilized i.e. fixed false alarm probabilities and threshold derivation. The latter estimation in a dynamic manner constitutes an alternative approach for future research under the fading environments considered. Moreover, another distinction should be noted: the authors in the above work considered AWGN channel thus unit fading channel gain whereas our analysis includes noisy fading environments as a more practical approach. In order to clarify the analysis of the latter paper, it is interesting to note that the probabilistic derivations regarding the Rayleigh fading assumption are based on parameter selection of a Nakagami-m fading distribution as a special case. In our case, the Nakagami-m fading distribution is not characterized by the parameter value that reduces to Rayleigh fading. This notation permits us to identify and propose addressing the problem of detection performance as a matter of scaling between the latter two distributions integrated with CS optimization. The simulation results in [121] verify the optimality of performance of Nakagami-m with increasing value of factor m compared to Rayleigh fading. The last issue requiring attention is the inclusion of other fading distributions which extent beyond our analysis as well in the additional context of introducing cooperation among SUs. As opposed to static threshold selection as in our analysis, the work [122] by Arjoune et al. proposes an advantageous dynamic threshold selection, always in the context of ED based methodology by conducting measurements of power of noise contained in the received signal. Convolution of received samples with pilots and averaging can be related to the convolution operation for deriving the equivalent distribution of the sum of channel gains considered. Thus the mathematical framework of the latter paper can be applied on top of convoluting to determine the distribution along with comparison of ED with matched filtering sensing schemes. Besides, our derivation of optimal number of samples for maximum percentage in favor of PU presence may take different and relatively low values thus in general our approach compensates for sensing time and complexity. The optimization perspective of CS principles could further impact sensing threshold in a sparsity scenario initially in a non-adaptive sense. CS could further aid the double threshold scheme commented in this paper by reducing the «middle-range» ambiguity leading to a consistent decision even in these cases. However, the authors in [122] elaborate on the covariance matrix focusing on the distinction of signal and noise eigenvalues with the assumption of independent noise similar to our case where we also assume independent fading realizations. This provides the basis for deriving the total covariance matrix. Moreover, the statistical covariance matrix concept can be further enhanced by the CLT theorem which in our analysis is merely adopted as a comparison case. The calculation of noise variance via an equivalent minimization problem involves a theoretical and empirical distribution comparison which could be embodied to our analysis for estimating the different of optimal number of samples in each case, the latter providing the feasibility of closed form expression. Regarding simulations, the SNR values for optimized performance should be further investigated apart from the observation of detection probability improvement to a more advanced consideration of noisy fading environment where decision threshold is dynamically adjusted. As the context of CS promises reduced computational and implementation complexity, we

turn our attention to a representative work [123] that employs CS for fewer samples requirement and quantifies performance in the Gaussian noise reduced samples regime. The first issue differentiation of this work to our analysis is that authors consider stochastic signal while our user signal is deterministic. In addition, the fixed alarm probability assumption and detection probability optimization comply to our analysis. Another detail that must be stressed regarding this paper that closely relates to our work is the consideration of maximum false alarm probability that accompanies the likelihood ratio test. An indicative comment concerning the CS based modeling of this problem ideal for future research is the l_∞ norm operator as equivalent for maximum value of a variable. Moreover, the modification of sensing threshold as a necessity emerging from reduced measurement dimensionality applying CS is an interesting extension of this work with the additional challenge of diverse fading and noise channel impairments. As a concluding remark, the effect of compression ratio and SNR decrease on detection and false alarm probabilities conveys performance deterioration which is nothing else but the statement of CS based performance vs. complexity tradeoff with respect to the moderate SNR regime. Proceeding to the work of Dikmese et al. in [78] briefly commented in the previous section, we review it from the aspect of this section. Hence, it efficiently proposes a wideband methodology for spectrum analysis. Fast Fourier Transform along with deviating from the boxcar model are utilized for PU spectral characteristics estimation for low complexity sensing. Reference to multi antennas scenarios along with cooperative schemes done in a complete manner would require an extensive amount of analysis and performance that is beyond the scope of this thesis and is therefore left as future work. However, a brief comment would suffice: spatial diversity is exploited at the expense of increased hardware design complexity. Moving on to the core of the analysis of this paper, the authors account for the practical case of frequency selectivity regarding the channel, according to moderate SNR value range, along with PU reappearance during SU transmission duration. Concerning paper contribution the sliding window concept is employed and the spectrum band division in order to derive flat sub-bands is employed. Towards a bridging remark between this paper and our analysis, it is clear that we do not focus on a frequency domain approach hence the essential component of this paper i.e. robust multicarrier techniques formulate an interesting deviation of our analysis for future consideration. To this specific end, the Parseval's theorem connects time and frequency domain by the identical test statistics derivation property. Moreover, a crucial question that arises is to what extent sensing performance is optimized by frequency tuning i.e. dissimilarity of band location of PU spectrum with respect to the sensing band. Simultaneous sensing achieved at this paper is what highlights the merit of their approach. A CS based probabilistic approach could be designed such that partial match of the PU band and sensing band is optimized as a constrained optimization approach while accounting false alarm and detection probabilities. Contrary to the derivation of [78] regarding PU presence decision probability increasing with interval length, our scheme produces results that characterize each fading distribution with an optimal number of samples that are related to highest percentage of decision in favor of PU presence. Hence, increased probability translates to maximum percentage of PU presence decision. Moreover, sub-band resulting from wideband division clearly promotes the optimized weighting process and accurate Energy detection performance.

The derived weights can be fitted in terms of the impact of respective sub-band to the sensing outcome, whereas our case provides the diverse fading consideration with contaminating Gaussian noise. In the same context, subset of sub-bands to be selected can lead to lower complexity while at suitable values of false alarm, detection probabilities and SNR regime. Hence, our analysis can comply to this subset selection in a cooperative scenario as intriguing future work. The frequency selectivity effect is also an issue to be addressed on the basis of the fact that only PU presence statistics are affected, an observation that validates applicability to our analysis. Also, as the authors provide remark on the ambiguity of whether non-white spectrum is due to transmitted signal spectrum or channel spectrum, this issue demands both separate and joint investigation to quantify performance with necessary measures accounted for. Relative to this paper, the results also raised an important issue to be taken into account, namely, the synchronization coherence filter design with optimum weights as the proposition to further validate performance. A subject of future investigation are the filter design characteristics of stopband attenuation and implementation complexity.

3.5.4 Proposed energy detection scheme

Energy detection scheme is based on the «binary» hypothesis stated below:

$$y[n] = w[n] \qquad \text{: PU Absent} \qquad (38)$$

$$y[n] = h[n] \otimes s[n] + w[n] \qquad \text{:PU Present} \qquad (39)$$

where y[n] is the discrete received signal, h[n] is the channel gain, s[n] is the user signal which is assumed deterministic and of unit amplitude in each time instant and w[n] is the zero mean unit variance noise following a Gaussian distribution N(0,1). Although the method of estimating moments for accumulated power are a well-established method, the comparison of the ratio of the equivalent probability density function in favor of primary user presence to the probability density function resulting in only noise is adopted here. Thus, the test statistic for LRT is given below:

$$\frac{p_n(n \mid H1)}{p_n(n \mid H0)} > \lambda \qquad (40)$$

where λ is the defined by Eq.(37) as the threshold and case larger ratio than λ decides in favor of primary user presence and smaller ratio in favor of primary user absence, hence, only noise. The numerator is equivalent to the n times of convolution of the fading channel by itself and a single convolution with the Gaussian probability density function modeling the noise in the communication channel. On the other hand, the denominator is the noise probability density function considered above. Another clarifying comment that must be made is that our proposed approach does not consider the squared amplitude channel gains as to derive the energies calculated in a common ED sensing problem. It considers channel gain related distributions in a sense that the channel gain magnitude is directly indicative of the energy of the sample. To validate such an approach, the sensing time interval is identical for all cases investigated.

The two comparison benchmarks consist of the CLT based statistics assuming Gaussian distributions as the following:

$$N\left(\sigma_w^{\ 2}, \frac{\sigma_w^{\ 4}}{N}\right) \quad \text{:PU Absence} \tag{41}$$

$$N\left(\sigma_x^{\ 2} + \sigma_w^{\ 2}, \frac{\left(\sigma_x^{\ 2} + \sigma_w^{\ 2}\right)^2}{N}\right) \quad \text{:PU Presence} \tag{42}$$

And Gaussian related formula Eq.(12) . In (41),(42) N is the number of samples, σ_x^2 is the user signal variance set to 3 and σ_w^2 is the equivalent noise variance set to one. Regarding the aforementioned formulae, the Gaussian statistics are assumed. However, they are applied to fading cases as an optimistic set of measures due to the fact that fading has a more detrimental effect on performance compared to additive noise. The analysis of this paper relies on the fading distributions namely Rayleigh, Rician and Nakagami-m fading with their expressions given in Eqs. (3),(6) and (9) respectively.

3.5.5 Algorithm formulation

The algorithm below analyzes in detail the steps of calculations for all fading cases and the comparison based on two CLT and CS based performance benchmarks.

Algorithm

1. **Input** LRT Test Statistics: CLT case (i=1), Rayleigh (i=2), Rice (i=3), Nakagami (i=4), number of samples N.
   Iterations:
2. for distribution **i**←*1 to 4*
3. Randomly select a value of samples i.e. between 50 and 100. Insert different values smaller or larger than selected value.
4. Calculate percentage of values above threshold forming the ratio with respect to input number of samples for the CLT case and with respect to the total number dictated by the convolution for the inserted values for the fading cases, respectively.
5. Also, calculate the CS-based number of samples using Eq. (18).
6. **While** no maximum has been achieved i.e. a sample number where smaller percentages are estimated for lesser or larger values of the input
   then expand to more distant to the initial input values.
   endwhile
7. **Select** the number of samples corresponding to the maximum reached and compare the resulting percentage to the CLT case percentage and CS-based percentage, respectively.
   end for
8. **Output:** optimal number of samples for which maximum percentage of values of LRT above dynamic threshold for CLT case, Rayleigh fading, Rician fading and Nakagami-m fading.

3.5.6 Simulation results

All simulations were conducted using Matlab software. By inserting different number of samples, the number of samples resulting in maximum percentage of decision values in favor of primary user presence is calculated and compared to the CLT and CS related percentages. Regarding simulations, the number of samples of the noise component in denominator is not determined by our sample number input to each respective case but by the order of magnitude larger value resulting from the convolution operation in the numerator or the likelihood ratio. As a remark concerning the visualization of the LRT ratio, the results from simulations implied a steep slope thus ratio values were either significantly low or high with respect to the threshold. This is the reason for omitting this depiction.

CLT case

As evident from the simulations, the number of samples in favor of primary user presence does not carry useful information. Instead, the percentage of number of samples above threshold to the total number of samples resulting from the convolution operation is what conveys meaningful results.

In the CLT case, inserting N=20, 50, 100, 150 and 180 samples the maximum percentage of values above threshold was 99% and was observed for M=100 samples. At greater numbers of samples, the percentage gradually decreased.

Rayleigh fading Channel

After initiating with 4, 5 and 10 samples and ending with N=40, 50 and 100 the maximum percentage of samples above threshold equal to 68% was found by inserting N=5. The CS based case indicated a percentage of 33.3%. For other values of N regarding Rayleigh fading, N=30 resulted in 9.4%, N=50 in 3.84% and N=100 resulted in 1.02%.

Rician fading Channel

Inserting N=17, 20, 25 and 30 as input parameter, a highest percentage of 1.54% was found for N=18, while for N=25 the percentage decreased to the value of 0.64%. The small percentage of Rician fading is a result of the Bessel term and the 18-times convolution of Rician distribution in the formulation of LRT test statistics.

Nakagami-m fading Channel

In this fading case, the maximum percentage of 96% was found by inserting N=5. For N=10 the percentage falls to 62% and for N=50 to 4.64%. This result can be also interpreted mathematically by the choice of parameter m equal to 0.5.

3.5.7 Overall results interpetations

The important result of the above simulations is the optimal performance of the CLT case which however is an asymptotic case regardless of the initial distribution considered as CLT theory dictates. Hence, an insightful remark is that of contemplating fading cases as a «transient» phase with the CLT characterized by the steady state

property. Another remark is the asymmetry of the fading distributions which by theory dictates the necessity of calculating the third and fourth order moments to adequately define the distribution curve contrary to the Gaussian distribution being symmetric and described by mean and variance. This second order statistics reflects the entire thesis and serves as an approximation for both analysis and simulation results obtained. The CS based resulting percentage of 33.3% is proven to be an intermediate result accounting of CS property of sub-optimal approximation. As for future work, the conditions and optimization problem formulation for improving this percentage will be considered in terms of sparsity and fading related compressibility rule.

As the convolution operation has a smoothening effect on the original distribution and also has the property of approaching a Gaussian distribution, the smaller percentages for the fading cases are interpreted in the context of deviating from the asymptotic Gaussian symmetric curve. Hence, the values of N resulting in maximum percentages were shifted towards lower values, as the asymmetry of the fading cases justify.

Another remark is that the simulation results based on convolution rely on the assumption of independence between fading channel realizations and between channel and noise distribution as well. Moreover, the trend observed in each case investigated of an optimal number of samples with greatest percentage of decision in favor of PU presence is a direct consequence of sufficient sample number reached for lesser values and correlation arising from further increasing this variable value beyond the optimal one derived. The practical case being correlation leads to different performance. As final remark, the maximum percentages derived for all cases considered directly impact accurate detection of user presence or absence.

3.5.8 Conclusions and future work

This section investigates energy detection scheme for Rayleigh, Rician and Nakagami-m fading conditions. The results are based on convolution statistics with a predefined threshold and the CLT and CS based cases as comparison benchmarks. Related algorithm and results interpretations are accounted for. As future work, the moment estimation method as ED scheme is considered integrated with CS problem formulation regarding convexity or non-convexity are promising directions of research. Extension of the proposed scheme to other CR spectrum sensing schemes and the state-of-the-art wideband sensing challenge are worth investigating.

3.6 Correlation based WSN performance: The CS paradigm

3.6.1 Introduction

The results in this section are related to our publication [J3]-([124]). WSNs consist of densely deployed sensors collecting environmental readings such as temperature, humidity, light e.t.c or performing the operation of target tracking which is more dynamic in nature and requires synchronization constrained functionality. WSN are autonomously deployed and their main challenges are energy efficiency, decentralized operation and node mobility as a means of addressing network connectivity and localized fault tolerance.

The main limitation of WSN operation is limited energy for performing sensing, computation and communication tasks. Thus this issue has spurred research for developing strategies to minimize energy consumption or maximize energy savings. Together with the small size and formidable cost of sensors, this bottleneck is a state-of-the-art challenge for WSN design. Dynamic topology changes or node energy depletion or failure could lead to operation disruption. Moreover, decentralized operation is a critical challenge compared to centralized operation where heavy computations are shifted to the base station also referred to as sink node. Network heterogeneity and asymmetry leading to unevenly dividing computation and communication tasks through the network are other major issues.

Temporal, spatial and spatiotemporal correlation have proven to be efficient properties verified for WSN operation and data processing that boost network performance and improve data reconstruction quality. The above concepts of correlation have already been integrated in WSN design and offered effective information exchange and processing. Another set of optimization tools that has been combined with correlation in WSNs is CS.

CS theory aims at providing the benefits of drastically reducing the computation and implementation complexity, issues being most important for optimal WSN design, deployment and operation. CS theory main concept is sparsity or compressibility. The former being the fundamental notion, obtains only the small fraction of nonzero elements of a signal vector and discards the bulk of zero valued elements. On the other hand, compressibility is an alternative approach that keeps the largest in magnitude elements and discards the rest. Temporal, spatial and most importantly spatiotemporal correlation along with CS have already provided low complexity data processing and efficient information extraction.

The contribution of this section is a probabilistic scheme assessing WSN network performance in terms of reconstruction error and energy estimation is conducted. Gaussian statistics are the main distribution for all performance derivations. Concerning mathematical scheme formulation, the information vector x comprises of mean values of average values of multiple sensor readings following a Gaussian distribution with zero mean and variance equal to the inverse of the number of readings forming the average values. The values obtained by this manner approach the i.i.d. CLT scenario asymptotically described by Gaussian distribution. The transformation matrix Φ which is multiplied by vector x, consists of elements of Gaussian zero mean and same variance distribution forming the sum and consequently the average values and mean values as elements of the matrix. The vector Y consists also of Gaussian elements of same variance but nonzero mean. The mean is derived from the fundamental equation relating first moment i.e. mean, second moment and variance of the Gaussian distribution. The second moment expressing power is assumed equal to one for the temporal correlation case and equal to the inverse of the number of mean values for spatial correlation case which specifically corresponds to a random sensor selection scheme in a given neighborhood. The sum of such elements representing readings lead to average values and consequently to the mean values realizations produced. For the spatiotemporal correlation the two cases above are fairly accounted for. Based on the error between vector Y and product of matrix Φ and vector x, the reconstruction error

is derived. The cases considered are noiseless and noisy cases respectively. In each case, the dense independence based case, the dense correlation based case and CS compressibility based case considering low and high compressibility ratios. Taking the energy of each Gaussian element and the related sums the above cases are also considered to derive mean values with the difference that results express energy estimation error. Benefits reaped from the above analysis are outlined in terms of network topology and routing along with applicability in communication systems, particularly IoT deployment networks. Conclusions and feasible research extensions are provided at the end of the section.

3.6.2 Past related work

The early work of Akyildiz et.al [125] is a survey that collectively mentioned the performance issues of a WSN referring to all functionality layers of the OSI model. The issues arising by emigrating from traditional ad hoc networks to orders of magnitude denser deployment of wireless sensors are analyzed and interrelated. For example, the power and routing awareness are crucial issues also affecting topology. Additionally, the balanced power allocation is determined by the sensor neighborhood or more general resources of the network. A straightforward extension of the analysis in this paper is traced to the concept of mobility and whether it is static or fully randomized, the latter characterization implying a varying neighborhood. The main concept of data redundancy could on first thought result in low necessity of mobile nodes covering the area of interest but also contribute to diversity considering that various data types are available to a subset of sensors forming a portion of the entire network. The consideration of spatial, temporal or joint spatiotemporal correlation could all define the means of exploiting sensor mobility in the network to the fullest extent. Relevant to the above, synchronization and switching off mechanisms are such delicate issues that could even question the validity of the results in our analysis of this section. Nevertheless, data redundancy as well as correlation in a monitoring network could smoothen the requirements of the network and point to the right direction for achieving energy conservation. A point stressed in [125] is the issue of energy consumption as well as limited memory of sensor nodes in terms of the scalability constitutes an interesting extension of the decentralized property that can be supported by our analysis based in temporally and spatially correlated data. Therefore, the protocols addressing information flow in a sensor network mentioned comprehensively in this early work can benefit from the randomized perspective as in the analysis of this section and provide optimal results by imposing inherent constraints that describe these networks by definition as well as additional ones that incorporate optimal statistical distributions for all resources to be allocated in an energy-efficient and effective data routing manner. Following the aforementioned work but being a much more complete survey, [126] highlights the main design issues for a WSN. Based on the categorization of sensor nodes, a straightforward issue arising is that of randomized mobility which in our analysis can be imposed as to whether, due to redundancy, mobile sensors are needed for further performance improvement. Effective routing is also another challenge relative to mobility and synchronization. Another issue discussed in this survey is the sleep and awake scheduling mechanisms adopted in an application specific WSN deployment. Idle sensors still consume energy and scheduling can even cause problems

apart from energy consumption. Relative to the analysis of this section, environmental monitoring is assumed thus energy conservation is the main bottleneck and latency is not of primary concern thus less stringent. Wireless standards are extensively analyzed in order to compensate for energy conservation as well as infrastructure, scalability and interoperability. As in multi-resolution storage of sensed data, the temporal and spatial redundancy directly relates to the analysis of this section expanded by the joint spatiotemporal correlation proven to provide even more optimal results of data reconstruction and energy estimation. Another challenging aspect is failure management where there are methods that aim at determining the cause of failure and the extent to which it significantly impacts network functionality. Given the correlation based analysis in our case, a capability of bypassing a failure event is more straightforward than actually detecting the failure in a localized manner. Another common point is traced at the relation of reconstruction error being bounded with the synchronization quality achieved which in turn affects energy conservation, a concept related to the energy estimation error analysis in our case. The extension of this derivation is the setting of a predefined error threshold in an application-dependent manner as well as algorithm development that have the property of low complexity fast convergence. Moreover, our main assumption being environmental monitoring sensor network, effective coverage in this case relaxes the stringent degree of overlapping coverage regions of a subset of sensors, which can be significantly relaxed as opposed to other applications such as real-time target tracking. The respective survey [126] underlines this fact by the statement of optimized minimum number of sensors guaranteeing coverage. Regarding compression and data aggregation, this section calculates average values which statistically interpreted are less varying than individual sensing values. That is based on the property of correlated, as in our case, readings and absence of abnormal values. Hence, aggregation based on average value can be combined with compression which translates to the sparsity ratios considered. Another similarity of this section adopting a fully randomized model regarding sensing values, transformation matrix as well as transformed vector is the issue of reliability mentioned in the survey the probabilistic version of which assumes the delivery guaranteed to a randomly selected subset of sensors near the coverage area. This also complies to the random sensing matrix as a CS optimization preference for probabilistically optimal results. A point of interest regarding the case of retransmissions needed in a packet loss scenario is whether a «uniform» retransmission of packets takes place or a specific packet which reflects additional knowledge of the lost packet identity. Relevant to spatial correlation in the sensed data in our case which implies the formation of a neighborhood of sensors, the survey provides a comment on correlation filtering for selective transmission from a certain node. This could contribute to enhanced security, improved scheduling as well as better resource allocation i.e. energy efficiency. Integration of wireless performance and low complexity with the CS principles with the resulting zero additional complexity induced by inference in a cross-layer optimization perspective is another future research direction. Finally, the most crucial bridging commentary of the survey to our analysis is the statement that fundamentally the derivations for improved performance rely on an optimization problem formulation with its objective function and its application-dependent constraints. Hence, this fits into the general approaches dictating that this approach takes many forms but obeys the

mathematical principle of the optimization problem and the design of an effective algorithm, as is the CS case. All these approaches in a cross-layer optimization manner make use of the interaction between layers towards achieving energy efficiency, effective routing under dynamic network topology as well as low complexity information extraction. Proceeding with another survey that deals entirely with energy efficiency as the primary prerequisite for extension of network lifetime, [127] provides a taxonomy of energy preserving methodologies from different points of view. The first distinction of energy consuming tasks states the following: sensing relates to negligible energy consumption while communication dominates energy expenditure compared to processing of measurements task. Hence, our analysis concerning sensing values and energies via a probabilistic Gaussian transformation representation can be related as a stochastic methodology, even by a non-Gaussian non-symmetric model, to the remaining sensor networks tasks. As in the analysis of this section, temporal and spatially correlated sensor readings can provide insightful directions for duty-cycling optimized scheduling of sleep and awake states further aided by data redundancies. Our analysis of spatial correlation involves a uniform power allocation over a node neighborhood which supports mobility as stated in [127]. In the same redundancy context, the subset of sensors to enable information flow can be chosen as in an energy-efficient, localization aware or secure wise manner. Redundant samples combined with sensing subsystem energy mitigation is a feasible extension of the analysis in this section. As the randomized mobility mentioned above, this survey states the problem of whether the mobility is controlled or not i.e. how randomness and a related pattern balance in order to optimize network connectivity and coverage. This randomization concept can be also extended in order to capture randomness in a modeling of delay and duty cycling by random distribution in manner covering network holes and optimize sleep schedule in an energy efficient and effective routing way. Collision in a randomized scheme can also be mitigated by data redundancy which results as a consequence of temporal and spatial correlation in the analysis of this section. It is the interrelation of low duty cycling with synchronization and allowance for flexible awake and sleep times that can provide additional energy efficiency and reduction of traffic and packet collisions. In an environmental monitoring data redundant network as in our case, a relaxation of the synchronization requirement can be assumed. One step ahead, the decentralized achievability is also based on this synchronization relaxation as well as energy efficiency. The effect of correlation in our analysis along with a randomized sensor wakeup scheme as described in [127] could well compensate for topology changes throughout operation which in our case could translate to node failure due to energy depletion, the above supported by a dense deployment scenario. Regarding the tradeoff in an environmental monitoring sensor network of energy efficiency versus latency, the former being more of a prerequisite, the question imposed by the survey is addressed in such a manner that in an energy limited scenario for this type of networks redundancy can further promote energy savings and, as stated above, functionality with no or relaxed coordinator nodes. An interesting similarity of our analysis which builds upon Gaussian modeled distributed readings, transformation matrix and equivalent, is the probabilistic data prediction concept which underlines the need of model validity by means of sampled data. Hence, the extension in the non-Gaussian case with the investigation of reconstruction error and energy estimation error as in this section brings

forth the extent of validity of the chosen distribution in both centralized scheme compensating for computation burden, as well as decentralized scheme based on correlated data assumption of our case. As opposed to randomization, a priori information supporting model validity can be thought as involved structure, which can be embodied to the performance optimization at hand. Additionally, the issues of outliers or sharp variations in data are not primarily considered due to correlation assumption. Adaptive sampling describes our case very accurately. Returning to the crucial concept of mobility either referring to sink node or sensor nodes, a certain energy saving observation can cannot be guaranteed in every network deployment. Instead, the balance between mobility pattern and the means of achieving mobility have to be taken into account as thoroughly examined in the above survey. A feasible comment in our case is that mobility can fulfill its energy consumption reduction goal as it is supported by the spatiotemporal correlation. The latter can be integrated with CS optimized performance to provide low reconstruction error and robustness of the sensor network. An excellent paper [128] that so closely relates to our analysis including the present section as well as channel estimation section estimates temporal correlation to adjust sleep scheduling by means of optimizing sensing of the observation area state leading to energy efficient performance improvement and promoting the mitigation of network functionality dependence on sink calculations. A particular similarity that must be stressed is the so called method of filtering in the paper i.e. the computation of the distribution by means of the observations to our CS based inference based on preserving largest in magnitude elements. In fact, this relation could constitute a CS integrated future research extension. Concerning the main concept of [128], temporal correlation decreases entropy whereas the entropy «floor» reached could lead to a characterization of the specified observation distribution. Hence, the effect on our analysis could be two-fold: the effect of separate temporal ad spatial correlation on entropy decrease as well as the effect of the joint spatiotemporal assumption and the entropy decrease in a non-Gaussian case relative to noise as well as composite fading cases. Another difference of this work is the linear increment of correlation and exponentially decrement while in our analysis we assume linear increment of correlation that is combined with CS based sparsity ratio to produce the representative results of reconstruction error and energy estimation error. The entropies, conditional and relative entropy, can also be used to derive consumed energy by comparing the distributions of transmitted and actual data. This could be the basis of a probabilistic extension in terms of Gaussian statistics by employing the method of inference as in section 3.3 based on CS compressibility rule or in a generic probabilistic analysis aiming at energy efficiency quantification. The errors derived in our case can be also derived by entropy related method boldly stressed in [128]. A concluding remark relative to this work is the dynamic determination of sleep duration in a network of sensed values of the same type. Hence, as in our case, the heterogeneous network case either resulting from a segmentation of a large network or an integration of networks covering an area but measuring different environmental quantities remains a challenge based on correlation-induced data redundancy. The property of redundancy stemming from spatial correlation, justified as the consequence of dense node deployment, is addressed in [129] where a cluster network is assumed where intra-cluster region is divided in spatially correlated regions, the same manner in which the network

neighborhood of a given sensing node is characterized by a spatial correlation value in our analysis. The cooperation between nodes is also leveraged in this paper towards the achievement of energy efficiency in the respective sensor subset. However, the authors comment on the required number of active sensors as an application specific consideration relative to data discrepancies, while our analysis in this section focuses entirely on reconstruction error transparent to the latter property highlighting that our scheme supports correlation inducing redundancy which is exactly the feasibility of supporting energy efficiency and gradual independence of a centralized sink node. Furthermore, the spatial correlation in the above work is considered in an adaptive manner not accounted for in our analysis where simply incremental increase of correlation is assumed for assessing sensor network performance. A natural extension of our work is the randomness inherent in initial network deployment as well as choice of subset of nodes providing sufficient energy for operation as well as low distortion in a way exploited in [129]. Relative to cluster head to sink communication compromise which is briefly commented, it is also a feasible extension of our case which can be thought of as a less probable, in probabilistic-wise statement, scenario due to flexibility in selecting optimal route from node to cluster head and sink induced by data redundancy. Another probabilistic statement, which adheres to Gaussian statistics in our analysis, but could be extended to arbitrary but realistic in modeling sensor coverage and percentages of active nodes with respect to total number of deployed sensors is the modeling of active node density based on event proximity and adaptive spatial correlation as a constraint to the optimization problem resulting from performance assessment. A final comment regarding the discussed paper is the derivation of results of minimizing distortion which can be considered in an entropy wise approach and average coding length initially in the Gaussian case, which admits a tractable entropy calculation, as well as to fading distributions thus taking fading channel in node communications into account together with correlation and energy efficiency. The latter two quantities pose a challenging optimization problem which can pave the way for effective exploitation of sensor capabilities and performance limitations. Maivizhi and Yogesh propose a data aggregation based technique in [130] to achieve redundancy elimination of redundant data resulting from correlation as already stated in detail above. The important assumptions of their work are the following: static topology and computation-efficient base station, hence, deviating from a decentralized feasibility perspective. The concept that differentiates our analysis considering correlation values from this work is that in [130] the match between sensed data given a specific threshold is considered in a zero-one derivation depending on the similarity of sensed measurements. This takes place at the source level aggregation. At aggregator level, the correlation is characterized as high by means of a determined threshold whereas our analysis makes no use of such a parameter in order to eliminate redundant data. Another basic assumption of this work is the one dimensional type of sensed data i.e. temperature. The paper derives a redundant data elimination method to estimate data accuracy in a different manner than our approach computing reconstruction error by Gaussian statistics. Also, [130] supports reduction in energy consumption by means of data aggregation while energy estimation error in our case is concentrated to energy of sensor readings. A concluding remark concerning the latter paper, the authors assume a statics network, while our work proceeds one step ahead as

it leaves a margin for supporting a decentralized scheme as well as reconstruction error and energy estimation error, from which mobility as well as effective routing from leaf nodes to cluster head and sink could be realized. A remarkable work [131] studies the coverage and energy efficiency tradeoff in terms of applying optimal scheduling to balance coverage when application dependent requirements for network lifetime extensions arise that must be fulfilled. The CS principle adopted in our analysis could also be applied to this paper due to the NP-hardness of the scheduling optimization problem and the utilization of a greedy algorithm. Indeed, the transformation method used in CS theory is also identical in the manner in which this polynomial property is verified in the paper. The insightfulness of this paper is the consideration of the sub-area of an area coverage of a sensor that is covered by at least two active sensors simply coordinated by the optimal on-schedule. This on-schedule is to be solved for in a thorough statement of an optimization problem including life cycle duration, battery life duration and derived area coverage. Moreover, the problem is re-formulated to consider maximum network lifetime compensating for the redundant area coverage. Concerning the notion of number of neighboring nodes related to required computational effort also implies that our case could be extended from the mere assignment of spatial correlation value in a given neighborhood to minimizing coverage redundancy as carried out in [131] along with the additional intuitive query of what the shape of the sensor neighborhood and coverage area is. The similarity of method of solution of greedy algorithm in this paper and in CS theory is verified with the latter suggesting that the algorithm could be modified by selecting more than one sensors in each iteration. According to the authors, the distributed sub-optimal equivalent problem adopts a randomized schedule. This is similar to the pair-wise correlation value in an incrementally deterministic increasing value regime along with the Gaussian distribution assumption in deriving reconstruction and energy estimation errors in the analysis of this section. What should also be stressed is the convergence to local optimum in the distributed version which, in terms of CS optimization, is nothing else but compressibility approximation relative to the sparsity assumption. Moreover, the centralized network property is the initial stage from which to emigrate to distributed property, the latter claimed to be achievable in terms of reconstruction accuracy in our analysis based on the correlation-induced redundancy. A statement worth highlighting is that randomly deployed sensors provide performance gain as the node scalability increases and the detection probability also rises justified by logical deduction. A concluding remark about [131] is that the merit of the results obtained is more useful in real time sensor network applications such as target tracking as opposed to our case dealing with environmental monitoring networks. Signal reconstruction based on sampling optimization in a sparse setting is addressed in [132]. Here, the authors mention that reconstruction is achieved via a low-dimensional sparsity promoting representation, i.e. CS methodology. Specifically, their work employs Deep Learning as the method of exploiting nonlinear mappings for efficient signal reconstruction compensating for the linearity based CS methodology. Deep Learning is also preferable to the widely used Principal Component Analysis. The primary point of interest expressed in this paper is the generative model used to derive model distributions which can aid reconstruction of sample data, bearing similarity to CS inference conducted in the previous section by setting a predefined threshold as dictated by CS compressibility

rule as well as modeling Gaussian statistics with specific threshold to encapsulate CS compressibility performance in the analysis of this section. Moreover, we employ linear Gaussian statistics based encoding of the sensor readings which is captured in the resulting vector. The statistical nature of our approach complies to the approach of location optimization. The intuitive statement that spatial correlation is closer to being exploited the authors in [132] confirm the inclusion temporal correlation as well which is also consistent to our analysis. The encoding of the sensor readings accounting for the relevant mismatch is also identical in the operators in this paper by means of the weighting matrix similar to transformation matrix Φ, and the bias vector which relates to the additive noise in our case modeled as standard Gaussian distribution. What should be stressed concerning results of this work are the requirement for low mean square error and low variance of the spatial prediction. This translates to our analysis as minimum reconstruction error and representative mean values derived in our analysis that do not alter significantly in a temporal and spatial sense. The last point of convergence of this paper to our analysis is the joint consideration of temporal and spatial domain: although the authors devote two separate stages for the above domain considerations our analysis in this section considers a joint consideration as a complementary case of considering temporal and spatial cases separately. We assume the value of these correlations as equal and investigate performance with a fair contribution of the above two domains in the derived performance. Another work taking advantage of spatio-temporal correlation [133] adjusts sampling rate according to low or high correlation data in order to mitigate data redundancy and excessive energy consumption. Along with the latter strategy, the non-sampled data is predicted by means of a reconstruction algorithm executed at the sink node. The important concept to note concerning this work is that the required computational task is performed at cluster head and sink which deviates from supporting a decentralized scenario as boldly commented in our analysis in this section based on the correlation induced redundancy similar to the prerequisite assumption in [133]. A crucial remark in this work differentiates the compression and aggregation operation for reducing energy consumption as opposed to adaptive sampling. This also sheds light in our analysis where mean values derivations determines a certain aggregation method and CS based compressibility can be considered as non-adaptive method. Therefore, adaptivity in a CS sense remains a challenging hypothesis towards achieving improved performance gains as opposed to non-adaptive CS. Furthermore, the issue of selecting a sensor node as «representative» either by the correlation value criterion or the residual node energy may provide useful results. Nonetheless, our analysis considers temporal and spatial correlation and supports the claim that subset nodes selection on the basis of redundant sensed data provides energy efficiency as data reduction drastically reduces transmission and computations. The authors in the above work consider a complete characterization of energy consumed for all sensor node processes conducted whereas our probabilistic approach considers sensed value energies to derive the resulting energy estimation error after deriving reconstruction error. Additionally, the percentage of required sampling rate modification according to correlation value is performed at cluster level in a manner which exploits correlation of neighboring nodes. A limitation of this work is the case of correlated sensors not compensating for each other, which could benefit from a probabilistic analysis such as our analysis in this section or by

considering other criteria such as residual node energies. Besides the limitations pointed out, the authors examine how their proposed method could handle variations in the sensed data relative to increase of sampling interval between measurements. Hence, this could translate to the following question in our analysis: how robust is the computation of means of average values taking simultaneously into account the rarity of abnormal readings in the scenario assumed. This question of robustness is intended for future research. The above could contemplate non-Gaussian statistics as extension of our case investigated. Following this robustness of the mean values generated in our case, we further support a decentralized design whereas [133] proposes a method that is heavily dependent on heavy calculations executed at sink node, where energy scarcity is a much more relaxed issue from a computational expensive point of view. To introduce the impact of CS on wireless sensor network performance improvement, we present a detailed review [134] which collectively analyzes the different applications of CS in the latter type of networks. Distributed CS is highlighted to achieve energy conservation by means of temporal and spatial correlation, with a priori sparsity knowledge being an impractical assumption. The issue of sensor deployment is also underlined as crucial be it addressed in a probabilistic manner or a pattern-wise (regular/irregular) manner. The problem of sparsifying domain knowledge in CS optimization techniques is crucial and addressed by learning optimal domain in a plethora of cases. Relative to the above formulation, our case only quantifies the measurement matrix and assumes the sparsifying matrix to be the identity matrix i.e. no transformation is required. A brief notion concerning the generalization of such an assumption is that non-Gaussian cases could dictate a transformation to acquire compressed samples for a monitoring WSN. In our analysis, sensing overhead has been reduced by leveraging CS compressibility in the unique case of Gaussian statistics an assumption enabling the quantification of preserved larger magnitude samples in a non-adaptive, as defined in literature, manner. Hence, the extent to which this approach provides benefits is subtly stated in [134]. An extension of our case to arbitrary non-Gaussian cases implies an adaptive sparsity promoting approach which could provide optimized dynamic topology and effective mobility based routing along with the issue of randomly deployed sensor locations provided as a concluding remark in the aforementioned review. Proceeding to the inclusion of CS sparsity effect on WSN performance, we first outline [135] as indicative of the latter consideration along with temporal correlation of environmental monitoring data. Benefits are introduced by the well-known property of existing structure in data integrated with sparsity, the latter magnified by the temporal correlation assumption. Sparsity is refined by proper measurement matrix design. Thus, integration of memory with CS sparsity offers the improved results of this paper. A similar to intuition of applying CS is the transmission of the time difference instead of sensor reading, which in our analysis could provide extra noise immunity given that we only consider correlation of sensed values in time and space. In fact, the additional noise immunity due to sensor level computation could translate as case optimality of our analysis as we support the decentralized property and as a consequence of slowly varying mean values in both time and space. The vectorization operation used to produce a single measurement vector provides increased complexity in real-time calculations, which in our case could be reduced as CS compressibility is applied to sensor reading value level. Moreover, the authors elaborate

on weight assignment inversely proportional to previous signal values with an additional factor included to avoid instabilities due to overwhelming weight values obtained. The notion of weight in our analysis is traced in the «weighted» contribution of spatial and temporal correlation assumptions when this joint correlation model is considered, which is verified by reconstruction and energy estimation error results. Simulations conducted in [135] consider different noise levels whereas the noise assumption which forms the core concept of our analysis is independence and identical Gaussian distribution with zero mean and unit variance. A concluding remark of this case is the issue of spatio-temporal correlation left as future work, an issue thoroughly investigated in our analysis comparing results of separate to joint correlation consideration. Hu and Yang in [136] present a well-defined work exploiting spatial correlation along with distributed CS principle for energy efficient observation reconstruction. In general, distributed CS is especially suited for handling network scalability and computational asymmetry scenarios. This asymmetry is only limited to initial information processing in our analysis since the need for sink based calculations gradually diminishes due to correlation and a decentralized network functionality is supported. However, the latter work clearly accounts for a centralized cluster-based network architecture with data collected at cluster head level and reconstruction done at sink level. Determining in-network radius based on error threshold is one step further towards contemplating for correlation. A similar concept to this one is the spreading of available power in the spatial correlation case in a sensor neighborhood, the area size of each is not quantified but assumed to define the subset of sensors that are characterized by a spatial correlation value. Regarding this paper, the decoding is solely conducted on sink level thus omitting the potential of an autonomous coordinator-independent network where computations are locally performed at sensor node level. Another indicative remark is low complexity observation which does not encumber the decoding process done at sink with data forwarded by cluster-head nodes. Moreover, contrary to the effect of zero approaching error as number of measurement increases, we clearly, in our analysis define the dense case, both independent and correlated, along with CS compressibility case, deriving our results. Another spatial correlation based event detection scheme [137] carried out in fusion center performs detection based on uncompressed data covariance matrix by means of compressed data, thus without accounting for raw data observations. Stated once again, distance between sensor or neighborhood dimensionality are not quantified and our analysis only involves correlation values. The merit of this work is the fact that noise threshold is not necessary along with the statement for independent noise similarity to our analysis. Regarding the mathematical formulation of [137], the hypothesis testing framework is adopted which makes a decision by means of diagonal structure in the signal absence scenario and off-diagonal matrix structure when signal is present, a concept not followed in our analysis where fading values are always present and the cases of no noise or noise contaminated sensor readings are assumed always under the essential assumption of statistical independence. The covariance matrix appears to have this banded property by merely exploiting the statement that as distance increases, correlation rapidly decays, the latter considering the power exponential model. On the contrary, our analysis is based on the following statement: the subset of sensors in a neighborhood all possess the same correlation, the value of which is incrementally varied and characterizes the same

sensor subset. To finalize with examining [137], the robustness of the proposed covariance based event detector given noise power uncertainty is what justifies its contribution with the additional comparison result that, in the absence of uncertainty, energy detector is preferable. This further verifies performance under a practical constraint. As opposed to this work, our analysis only uses a standard Gaussian statistical model for the additive noise and studies its effect with following assumptions: a) Gaussian distribution with specific assigned parameters and b) obeying the independence principle. These provide our results with an incremental correlation model. Zhang et al integrate spatial and temporal correlation in [138] to achieve improved reconstruction accuracy with the constraints of the optimization problem being utilization of Kronecker CS and cluster network topology. Moreover, the reconstruction takes place at sink level thus assuming a centralized approach. The random walk approach for this problem provides some issues that induce high overhead and thus dissipate energy. Hence, given the comments in this paper and our fully randomized Gaussian statistical approach the query that arises is whether a randomized approach could compensate for link reliability and a compact topology that ensures the feasibility of considering significant spatial correlation values along with temporal correlation of the sensed data at sensor node level observations. Another extension of our work could follow the investigations in this paper i.e. a randomized scheme for optimally selecting the sensor nodes with readings that could enhance reconstruction accuracy. Clearly, this could require non-Gaussian statistics assumptions and an optimal distribution modeling approach under a Bayesian inference optimization problem. Sleep scheduling separately at each cluster could also be accordingly coordinated for this reconstruction accuracy achieve and its corresponding tradeoff with network overall energy efficiency. Relating the CS methodology in the aforementioned paper, RIP condition is required to be fulfilled as well as the l_1 norm based convex relaxation for sink level decoding. A comment that suffices concerning this verification is the Gaussian statistical assumption which verifies the CS sparsity ratio of node level randomly sensed data. The already provided comment in this section is the issue of modeling the problem of optimal sensor deployment and how it could incorporate posterior connectivity and reconstruction accuracy requirements. Hence, the claim that randomness, as in CS optimization random matrix theory, could compensate for deficiencies such as covering network holes, as opposed to the issue of negligible correlation as nodes are far apart, leads to the question of whether this case is all about the optimal distribution in an application-specific manner. The non-Gaussian approach is also a strong candidate in this problem as well and the uniform deployment could prove to be outperformed in terms of energy efficiency based constraints. Still, our Gaussian random model admits an extension as to what extent this distribution could provide optimal performance in a clustered sink coordinated topology as well as a decentralized optimization scenario. The authors in [138] employ vectorization operation for joint consideration of spatial and temporal correlations. As our analysis relies upon a fair combination of the latter two correlation models through appropriate parameterization of the vector that represents the transformed matrix of randomly Gaussian sensor readings via the Gaussian transformation matrix, the above approach could include Kronecker CS for derived results comparison. This could accompany the DFT and DCT sparsity models for temporal and spatial correlation, respectively. As a

decentralized extension remains, as already stated, a challenge the question arises from the above paper as to how the role of cluster head and its corresponding calculations could be shared among nodes and how the node subset selectivity still applies. The results presented in the latter work verify that packet loss does not impact problem formulation. Relative to our analysis, the randomized model along with correlation induced redundancy could account for packet loss impact absorption and, one step further, to reduced effect of packet loss in a gradually autonomous network. Completing our comments on this paper, the energy efficiency needs further attention which could differ from our similar claim in the analysis of this section due to the fact that the former rely upon a clustered sink-coordinated network. A representative work bearing similarities with the last one examined was conducted by Hooshmand et al in [139] and utilizes spatio-temporal correlation structure based on Kronecker CS. The compression scheme exploits signal structure in an energy-efficient manner comparable to existing schemes. A remark worth noting is that correlation and consequently redundancy permits a certain tolerable compression loss which, along with energy efficiency, indicates an interesting research extension integrated with our Gaussian statistical assumption identical to the signal statistical model in the latter paper. Sensor selection problem for optimal reconstruction strategy is also commented in this paper whereas our analysis only provides a remark that correlation induced sensor subset for information forwarding is feasible on the basis of slowly varying mean values of sensor readings as contrary to sensor reading actual values. Besides reconstruction accuracy, the greatest benefit achieved, as stated in [139], can be realized by admitting different interpretations as the energy efficiency requirement or latency (clearly not a crucial issue in environmental monitoring networks) or even security as we briefly comment on the WSN applicability section. All the above can be combined with CS compressibility principle to provide those performance gains. The comparison benchmark of covariance structure can also be considered along with independence assumption as considered in our analysis. The main concept relating this work to our analysis is that temporal and spatial correlation are firstly separately investigated in terms of temporal sparsifying transforms, Discrete Cosine Transform and spatial correlation approach by Distributed Source Coding. Then, spatio-temporal compression algorithms are analyzed. On the contrary, our analysis depicts that Gaussian random values are already sparse and we merely leverage the equivalent transformation matrix. Regarding spatial correlation, this paper uses quantization which considering our case with already slow varying mean values, quantization could result in even greater savings with the possible loss induced left as future investigation. Clearly, the authors claim negligible quantization error and the issue of complexity of their method that estimates sensed data based on error with respect to side information leaves the problem of overhead as a point of concern which could be addressed by employing advanced coding schemes. All sensor selection algorithms presented in the aforementioned work based on a scheduling scheme could be contrasted against uniform deployment. Hence, a Gaussian model could be assumed due to its simplicity and symmetry and could compensate for requirements tailored for environmental monitoring networks i.e. focusing on connectivity and robustness against additive noise or, more related to sensor selection, node failures. A straightforward extension of our analysis given the self-tuning remark with error input is the correlated noise consideration integrated in a

CS optimization problem with the l_2 norm quantified error tolerance based on predefined threshold. Nevertheless, our statistical independence based convolutional approach could benefit from this self-tuned statement by selecting an arbitrarily small correlation value determined by parameters as the sensed data properties, the dynamic topology and the required reliability and robustness of the sensor network. It is also worth noting that our Gaussian statistics based CS sparsity ratio can also be related to the randomized scheme of applying a small subset of sensors for sampling and compare to deterministic sensor selection schemes. The above contrast could emigrate to how differently derived mean values in our analysis vary by employing different sensor selection schemes. Moreover, another problem worth investigating is how the neighborhood is shaped along with random subset selection and whether this selection scheme implies that spatial correlation value is low and how temporal correlation contributes to mean value representative values. It is clear that proposed scheme in [139] is heavily dependent on sink for reconstruction and no comment is provided otherwise. It is also interesting to query whether our Gaussian randomized reconstruction and energy estimation error derivations could be modified in the spirit of the latter paper where critical correlation value ranges are considered and additionally combined with a noise tolerance threshold instead of its mere statistical description. Concluding comparisons of our work over the latter paper, the statement of a transient phase of a windowed version induces the case of observing the extent of a similar case of transience in our analysis from the moment the sink begins conducting the calculations till the instant when calculations have reached a stage where representative values are known at sensor level and consequently the steady state is realized where the desired decentralized network operation is then feasible. A complementary investigation would consider mean values strategy as in our case along with raw data as well as other operators resulting in different quantities of data exchanged within the network. Proceeding further, the authors in [140] leverage spatiotemporal correlation in an environmental monitoring network with heterogeneous data types with the aid of CS methodology. Therefore, it is stated that correlation among nodes comes along with sensor selection schemes to reduce data rates confronting energy consumption. Hence, correlation between diverse data types integrated with CS principle encapsulates the merit of this work. An indicative CS application is the exploitation of sparse side information as opposed to the actual signal as stated by the authors. An interesting fact concerns the inclusion of weights in sums formed by sensors, the sums being forwarded to sink. Hence, the question regarding our work arises whether the sensor data forming the average values which in turn provide the mean values could include weights to optimize reconstruction error, noise robustness and energy efficiency, on the basis of reducing communication and transmissions. The copula function tool also implies a probabilistic extension of our work as we consider independent realizations along with correlation cases. Hence, copula function can be exploited to bridge distributions resulting from independence with those from arbitrary correlation values. In fact, spatio-temporal correlation induced dependencies constitute a challenge for shaping this copula function element in a sense of deriving the mean values in our analysis. Following the same concept, the Markovian assumption injecting memory property in the distribution derivation problem could also enable the extension to combined correlation and dependence on past sensor readings. The above

could render the investigation of resulting complexity versus slowly varying property of the mean values derived. The tradeoff of complexity as opposed to energy efficiency could also be a challenging aspect. Contrary to the latter paper, the variance of each of the sensor readings in the set is kept constant and thus unambiguously contributes to the sum of measurements and consequently to the average values derived. Moreover, correlation matrix fitting via likelihood estimation is not applied in our analysis whereas we are limited to investigating performance relative to incrementally varying temporal as well as spatial correlation. Extension of our Gaussian randomized scheme to other distributions requires balance between computational complexity and modeling capacity as stressed by the authors in [140]. Also worth noting is the separate energy consumption evaluation of the processing and transmission phases which, in our case, are clearly reduced due to slow variations, caused by rare abnormal readings of the mean values derived.

3.6.3 System model

The optimal system model regarding the proposed probabilistic scheme would be a set of concentric circles each defining a circular ring as a node neighborhood where spatial correlation holds as a property. Thus, instead of a square area of deployment or an area with the sink at the center of concentric circles, the tree network model is considered as the suitable topology serving as a compromise between centralized and in-network scenarios. Modeled as a graph G(V,E) with vertices as nodes and edges as bidirectional links for information exchange, data flows are feasible between any set of nodes. A clustered network setting could also apply while for temporal and spatial correlation properties a sampling rate adequate for temporal correlation and adequately dense sensor deployment are assumed to be fulfilled. Either quantifying the distance between node neighborhoods or adopting d-hops distance away from sink could be well suited for the correlated scheme proposed. For the completeness of the above statement, the sensor nodes are assumed to be dense in the defined neighborhood of the network thus spatial correlation is constant for the respective region. As for temporal correlation, the sampling frequency could be assumed such that correlation values are supported in both time and space dimensions.

3.6.4 Mathematical preliminaries

The fundamental statistical tool used throughout the forthcoming analysis is the Gaussian distribution Eq.(12), which is widely applied in wireless communications. It is characterized by symmetry and is fully described by the mean and variance. The sum of such variables is also Gaussian with a mean equal to the sum of the means and variance equal to the sum of variances stemming from the independence assumption by means of which the covariance terms are cancelled. Furthermore, for the correlated case the resulting sum of variables with equal pairwise correlation has the following variance:

$$\mathrm{var}(S) = N * \sigma^2 \left[1 + (\mathrm{N} - 1)\rho\right] \qquad (43)$$

where σ^2 is the variance of each of the variables and ρ is equal to the common correlation. It must be noted that the independent as assumed additive noise is also Gaussian distributed for both reconstruction error and energy estimation error cases.

For the above cases, the expression of the problem is stated below:

$$Y=\Phi*x+e \tag{44}$$

where x is the vector of mean values from average values of sensor readings. Matrix Φ is the transformation matrix also Gaussian distributed with zero mean and variance equal to the inverse of number of variables comprising the sum. The average values are computed and the elements of the matrix represent the mean values of average values in the same way. The vector Y is Gaussian distributed as well with equal variance for each element forming the sum and nonzero mean stemming from the equation:

$$mean = \sqrt{\mathrm{E}\left[x^2\right] - \sigma^2} \tag{45}$$

where for temporal correlation the second moment assumes a value of one and for spatial correlation equal to the inverse of mean values considered. Finally, the energy estimation error utilized the (non-central) chi square distribution with k degrees of freedom equal to the number of terms in the sum and non-centrality parameter for the vector Y nonzero mean inclusion. However, analysis also resorted to an approximation of the chi square distribution by a Gaussian distribution as given in [141]:

$$N\left(k+s, \sqrt{2k+4s}\right) \tag{46}$$

where both mean and variance are functions of degrees of freedom k and non-centrality parameter s. The reason behind this is that the CS cases quantify the number of elements preserved thus effectively tackling such derivation for the chi square distribution case.

3.6.5 Algorithm formulation

For the two error estimation cases investigated i.e. reconstruction error and energy estimation error the related algorithms are given below:

Algorithm 1 Reconstruction error estimation

1. **Input**: Temporal correlation casetemp, spatial correlation casespat, spatiotemporal correlation casespattemp. Number of readings in each sensor (temporal/spatial cases), number of sensing periods and power consumption for temporal case/number of sensors and power consumption for spatial case/equal number of sensing periods and sensors for spatiotemporal case. Additive Gaussian noise distribution as zero mean unit variance $N(0,1)$.
   For casetemp noiseless case
2. Calculate mean and variance of each sensor reading.
   For the independence case
3. Calculate the encoded measurement vector Y, the transformation matrix Φ and vector x.

4. Estimate reconstruction error of means of average values of sensor readings (vector x), transformation matrix Φ and vector Y. Evaluate according to error magnitude and sign.
   For correlation case
5. Modify variance of sum according to correlation value and repeat step 3 and 4.
   For Compressed Sensing case
6. Determine correlation and sparsity ratio values, and repeat steps 3 and 4 for low and high sparsity ratio values.
   For casetemp noisy case
7. Increase sensor reading variance by one and repeat steps 3-6.
   For casespat noiseless case
8. Modify mean value of distribution of elements based on which vector Y is formed and repeat steps 3-6.
   For casespat noisy case
9. Increase sensor reading variance by one and repeat steps 3-6.
   For casespattemp noiseless case
10. Modify mean value of distribution of elements based on which vector Y is formed and repeat steps 3-6.
    For casespattemp noisy case
11. Increase sensor reading variance by one and repeat steps 3-6.
    End
12. **Output:** Reconstruction errors for temporal, spatial and spatiotemporal correlation cases.

As indicated by the above algorithm all three correlation cases are included with the separate assumptions of noiseless and additive Gaussian noise in each case.

Algorithm 2 Energy error estimation

---

1. **Input:** Temporal correlation casetemp, spatial correlation casespat, spatiotemporal correlation casespattemp. Number of readings in each sensor (temporal/spatial/spatiotemporal cases), number of sensing periods and power consumption for temporal case/number of sensors and power consumption for spatial case/equal number of sensing periods and sensors for spatiotemporal case. Additive Gaussian noise distribution as $N(0,1)$. Gaussian distribution approximation of non-central chi-square distribution.

For casetemp, noiseless case

2. Calculate mean and variance of each sensor reading.
3. Calculate degrees of freedom and non-centrality parameter of the (non-central) chi-square distribution modeling the sum of energies of the random variables.

4. Based on Gaussian approximation, calculate the encoded vector Y, transformation matrix Φ and vector x of mean values of average values of energies.
5. Estimate energy error and evaluate according to error magnitude and sign.

For correlation case

6. Modify non-centrality parameter according to correlation value and repeat step 3-5.

For Compressed Sensing case

7. Determine correlation and sparsity ratio values, modify non-centrality parameter and repeat steps 3-5 for low and high sparsity ratio values.

For casetemp noisy case

8. Increase sensor reading variance by one, modify non-centrality parameter and repeat steps 2-7.

For casespat noiseless case

9. Modify non-centrality parameter and repeat steps 2-7.

For casespat noisy case

10. Increase sensor reading variance by one, modify non-centrality parameter and repeat steps 2-7.

For casespattemp noiseless case

11. Modify non-centrality parameter and repeat steps 2-7.

For casespattemp noisy case

12. Increase sensor reading variance by one, modify non-centrality parameter and repeat steps 2-7.

End

13. **Output:** Energy estimation errors for temporal, spatial and spatiotemporal correlation cases.

3.6.6 Simulation results

All simulations were conducted using Matlab software. The simulations are divided in two subsections: the reconstruction error estimation and the energy error estimation. Both error estimations are based on Eq.(44) the mismatch between elements of Y and the transformed via matrix Φ, elements of vector x.

Reconstruction error

Concerning vector Y, the mean value of the elements that form the sum and consequently the mean values of the vector Y is computed according to Eq.(45) with unit power assumption and variance equal to 1/40 i.e. the inverse of the number of variables comprising the sum. The dimension of Y is defined by 30x1, i.e. 30 mean values considered as realizations in the simulation. The matrix Φ is thus of dimensions 30x30 with equal number of mean values that resulted from average values from the aforementioned sum. The elements comprising the sum are also zero mean Gaussian distributed with the above variance. Finally, for vector x where the variables forming

the sum are equivalent to sensor readings, the zero mean and same variance are considered as well. For the correlated cases, the correlation was assumed to take the values 0.1 to 0.9 with an increment of 0.1. In the CS cases, simulations were conducted considering a sparsity ratio from 0.1 to 0.9 with an increment of 0.1 whereas the correlation values were assumed likewise. The respective values were pairwise considered and the pair with smallest estimation error was chosen for simulations.

Temporal correlation noiseless case

The simulations for this case include the independence case plotted by blue curve, the correlation case plotted by the red curve and the CS cases of low and high sparsity ratios with green and magenta curves. The same trend for the above cases investigated is followed also in the next subsections. CS based cases preserved only a subset of the average values in each realization. Moreover, the value obtained for each case is the fifteenth realization which is indicative due to curve symmetry with respect to this value.

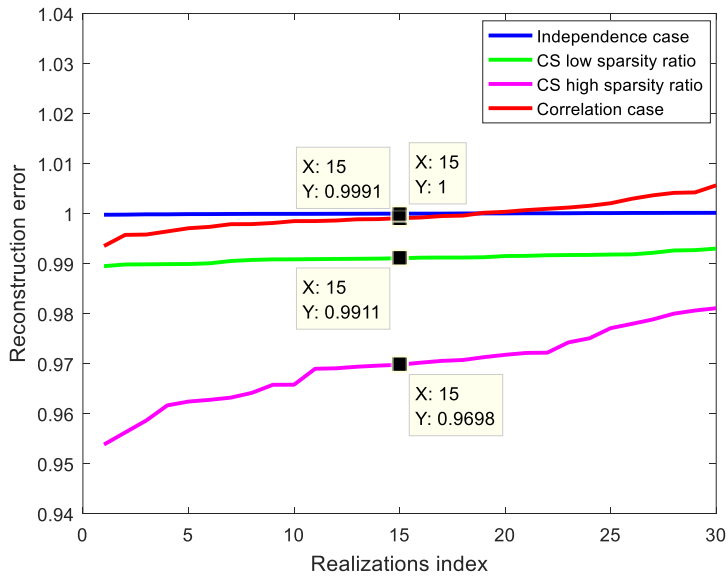

FIGURE 3.16: Reconstruction error for temporal noiseless correlation case

The findings of this section plotted in Figure 3.16 indicate that independence case performs the worst while the optimal correlation value of 0.9 resulted in negligible error compared to independence applicable in cases where high precision errors are considered up to fourth decimal digits. The first observation of low CS error is made in this section. For sparsity ratio and correlation values equal to 0.1 the lower error was verified by the green curve. For the high sparsity ratio of 0.9 and correlation equal to 0.8 the error was found even lower. Hence, for low sparsity ratio complying with incoherence CS principle the effect was verified. On the other hand, the large sparsity ratio i.e. small fraction of values discarded and large correlation resulted in lower error, meaning that correlation has a more beneficial impact as the dense case is approached. Thus, in an either dense or CS based cases, correlation induces lower error as a consistent observation.

Temporal correlation noisy case

This subsection deals with temporal correlation along with additive noise consideration contaminating the sensor readings relative to vector x and variables based on which the elements of matrix Φ and vector Y are formed. This noise is Gaussian distributed zero mean and unit variance. Hence, the key concept differentiating the noisy case from the previous one is that the small decimal variance value of readings/variables is increased to a large extent as the unit noise variance is added to this decimal value. Consequently, the variance of the sum of the Gaussian variables is greatly magnified thus resulting in a much wider Gaussian bell curve. The related Figure 3.17 is given below:

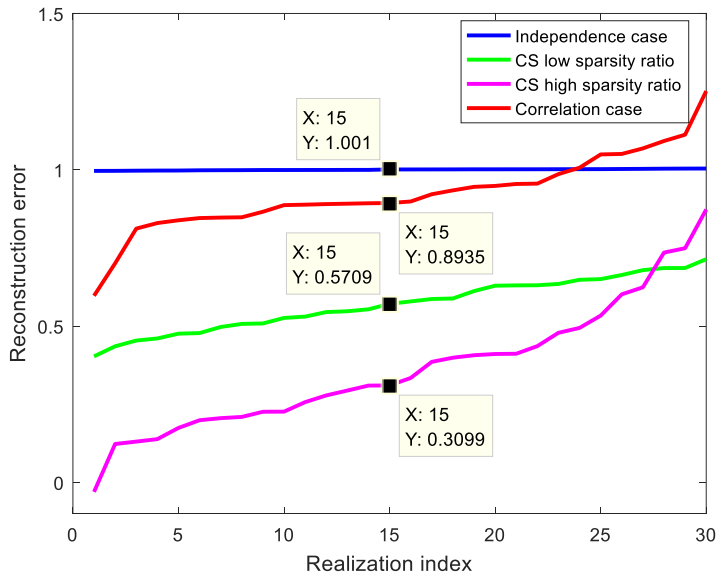

FIGURE 3.17: Reconstruction error for temporal noisy correlation case

The CS green curve assumes a sparsity ratio of 0.1 and correlation value equal to 0.3 while the high sparsity ratio equal to 0.9 and correlation at the value of 0.8. The independence performs the worst in this section as well, whereas the other values especially in the CS based cases depict lower error. This is contrary to the effect that noise deteriorates performance. However, this result is justified as follows.

The distributions of elements of vector x, matrix Φ and vector Y have a much larger variance due to noise and their Gaussian bell curves overlap. Thus, variables assume values in a much wider range and more distant values are assumed with higher probability. In some simulations, results indicated even negative errors which are also justified by the aforementioned curves overlap which is a result of independent additive noise and the much larger variance compared to the variables variance and the errors shift towards lower values. Hence, in a practical scenario, adjusting the correlation in a specified value range improves the performance and lower error is obtained as opposed to the ideal case of independence. A final remark for CS cases is that results follow the same trend, i.e. lower errors were observed in comparison to the other cases in this subsection and the previous noiseless section as well.

Spatial correlation noiseless case

Compared to the temporal correlation case, the spatial correlation assumes a smaller nonzero mean of the random variables that result in average values and consequently the mean values that represent the related matrix and vectors. This smaller mean results from Eq.(45) by assuming the second moment equal to 1/30 which indicates an even spreading of power among the neighboring nodes set to communicate by a random sensor selection scheme. The denominator of this second moment was chosen as the minimum number of independent random variables that can be well approximated by a Gaussian distribution in terms of the CLT. This can be considered as a spreading in the power domain, i.e. a uniform power allocation among all such neighboring nodes. The variance concerning the variables forming the sum remain the same. Figure 3.18 for the spatial noiseless case is given below:

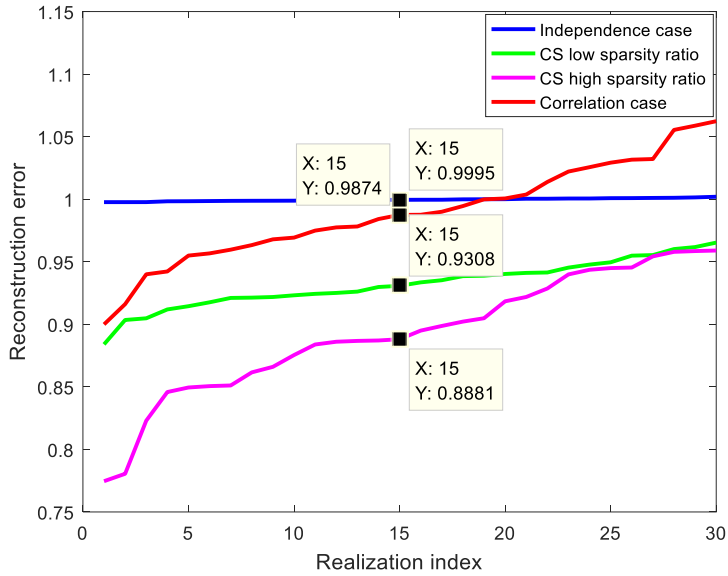

FIGURE 3.18: Reconstruction error for spatial noiseless correlation case

The main observation from the figure in this section are the lower errors for all assumptions compared to the temporal noiseless case investigated in the respective subsection. The CS low sparsity ratio case depicted by the green curve assumed sparsity ratio equal to 0.1 and correlation value to 0.3. For the high sparsity ratio case, the sparsity ratio was equal to 0.9 and correlation was set to 0.8. For the CS low reconstruction errors the same trends of the previous sections also hold. Stated once again, independence dense case performs the worst while dense correlation of value 0.9 results in slightly lower reconstruction error.

Spatial correlation noisy case

As in temporal noisy case, the variance of each of the variables in the sum is increased by one due to *N(0,1)* additive noise included in the model. The result of this section are depicted below in Figure 3.19:

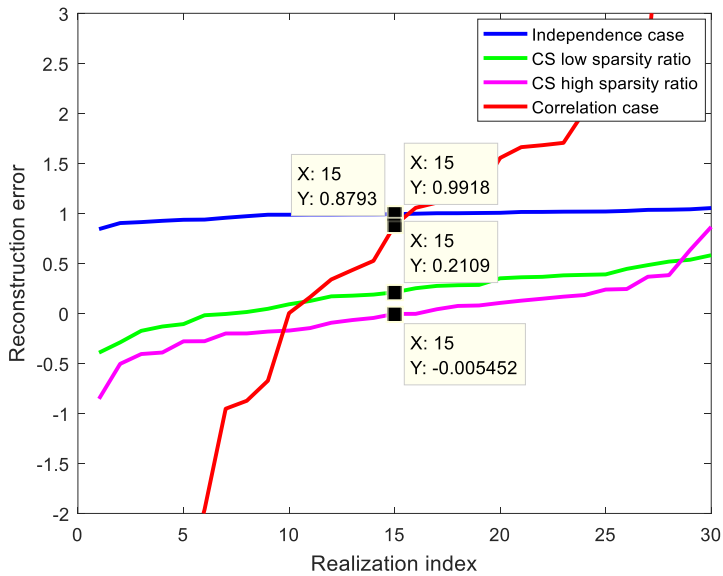

FIGURE 3.19: Reconstruction error for spatial noisy correlation case

The observations for this section is that the related figure indicates lower errors compared to spatial noiseless case and even lower errors than the temporal noisy case as well. The CS cases also convey lower errors with respect to the low sparsity case with sparsity ratio set to 0.1 and correlation set to 0.3 and the high sparsity ratio equal to 0.9 and correlation set to 0.9. The second observation was the increased number of negative errors in this subsection justified by the distributions overlapping being a more contributive result due to the smaller distance of distribution of the elements of the sum relative to vector Y from zero mean accounting for the distributions of vector x and matrix Φ. A final remark concerning the CS cases is the significantly lower error for low sparsity ratio case and the practically zero error i.e. no mismatch for the high sparsity and correlation valued case shown by magenta curve.

<u>Spatiotemporal correlation noiseless case</u>

The spatiotemporal correlation is a joint correlation scheme integrating both temporal and spatial correlation assumptions. Relative to the simulations of this section the joint consideration of equal number of temporal sensing periods and spatial number of neighboring nodes is made. The variances of the variables in the sums of independent, correlated and CS based case remain the same whereas the mean value of elements of the sum in vector Y is jointly quantified i.e. by 20 times the mean value based on temporal correlation and 20 times the mean value based on spatial correlation. Another important assumption is that the correlation value is constant between variables in each group and between variables in the same group. This is the reason why variance of sum remains the same. Figure 3.20 below presents the results for this case:

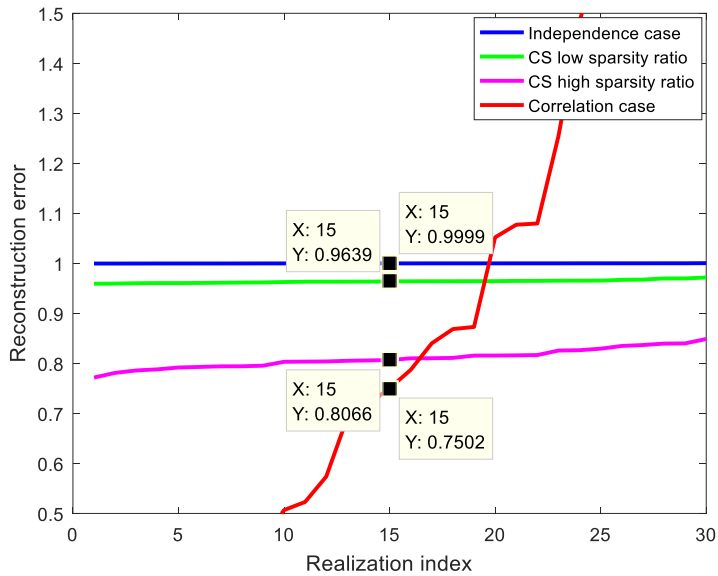

FIGURE 3.20: Reconstruction error for spatiotemporal noiseless correlation case

It is obvious that the independence case has the same error magnitude. The dense correlation case depicts a significantly lower error compared to other noiseless cases verifying the improved performance effect of joint spatiotemporal correlation case. The CS low sparsity ratio exhibits lower error than the temporal noiseless case and higher error than the spatial noiseless case assuming sparsity ratio equal to 0.2 and correlation equal to 0.4. This can be interpreted as the joint contribution of temporal and spatial correlation cases leading to this result. Concerning CS high sparsity ratio equal to 0.9 and correlation equal to 0.8 exhibited smaller error than the low sparsity ratio case as in previous sections.

<u>Spatiotemporal correlation noisy case</u>

The assumptions of the previous noiseless case hold for this case as well along with the increase of the variance of the variables in the sums considered by one due to additive noise consideration similar to the above noisy cases. Figure 3.21 depicting the results for this cases is given below:

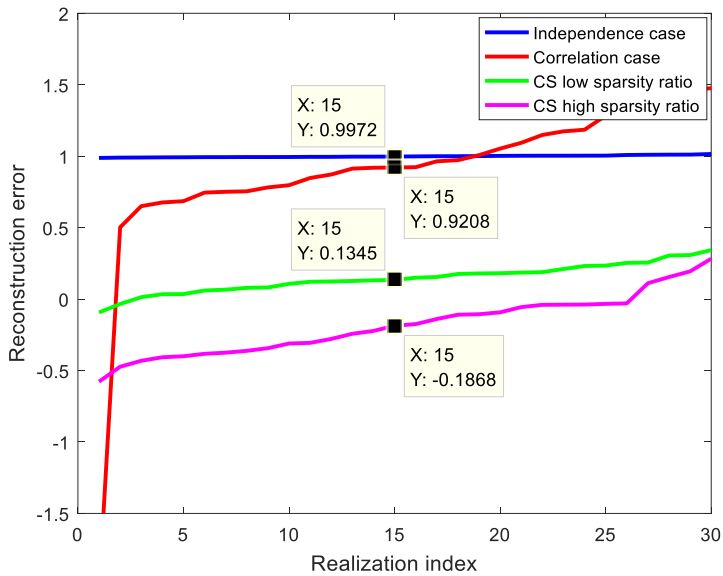

FIGURE 3.21: Reconstruction error for spatiotemporal noisy correlation case

The crucial observation for this last case is the fact that applying CS principle results in the drastically lowered error compared to all cases. Explaining the findings in detail, the independence case exhibits a decrease in reconstruction error. The red curve representing the dense correlation case also conveys a lower error value. The CS low sparsity ratio case with sparsity ratio equal to 0.2 and correlation equal to 0.5 exhibits the lowest error particularly in comparison with noisy cases in which case the comparison is fair. About CS high sparsity ratio case, the negative sign is indicative assuming sparsity ratio and correlation values both equal to 0.9. Interpreting the latter results in terms of absolute value of error, this section conveys the lowest error along with the spatial noisy case. Mathematically interpreted, the values of vector Y underestimate the resulting values of the product of matrix Φ and vector x. Related to assumption of fair temporal and spatial mean values, the distance of the mean value of elements of the sum for vector Y is between the mean value considered for spatial and that of temporal case.

<u>Energy estimation error</u>

This section presents simulation results regarding energy of the variables forming the sums in the previous section. Thus, the average energy values were computed and their final mean values as in the previous section. Towards this investigation, the non-central chi square distribution is utilized with degrees of freedom equal to the number of variables in the sum and non-centrality parameter fully defining the distribution.

However, all simulations conducted consider a Gaussian approximation of the latter chi-square distribution the reason being that CS theory relative to Gaussian statistics accurately quantifies the number of samples i.e. average values in this case adequate to ensure reconstruction. Contrary to the chi-square, the Gaussian approximation effectively tackling this issue. As a final remark, all temporal, spatial and spatiotemporal noiseless and noisy cases are investigated in terms of energy estimation error as well.

Temporal noiseless energy estimation error case

This first section considers temporal correlation case without additive noise considered. Correlation regards the variables the sum of energies of which is calculated and consequently the mean values as elements of the vectors Y and x and matrix Φ. Independence, dense correlation and CS low and high sparsity ratios are all included in the simulations conducted. Figure 3.22 for this subsection is given below:

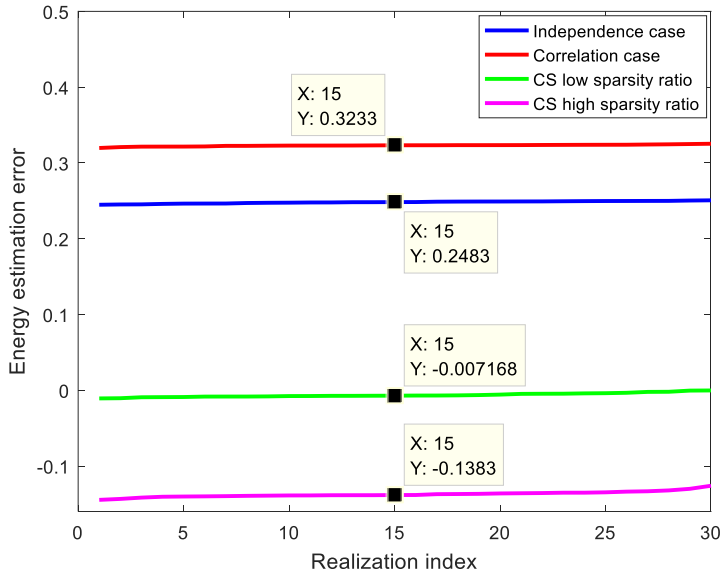

FIGURE 3.22: Energy estimation error for temporal noiseless correlation case

The first observation contrary to the reconstruction error is that the independence case performs considerably better than the dense correlation case in terms of energy estimation error. Concerning CS cases, the low sparsity ratio with value of 0.4 and correlation value of 0.7 performs better than the high sparsity ratio with sparsity ratio equal to 0.9 and correlation equal to 0.2. From these results, it is apparent that former CS case exhibits a higher sparsity ratio and higher correlation as well, while the latter considers a lower correlation value. This must be accounted for together with the fact that in reconstruction error the CS high sparsity ratio case performs better than the low sparsity ratio while in this energy estimation error case the opposite is observed. Hence, sparsity ratio increase in the former case dictates that more samples must be preserved to reach the optimal energy estimation error. Additionally, the CS based negative signed errors although of small magnitude convey that another property of the model considered is that it can be characterized as lossy, i.e. the elements of vector Y are of smaller magnitude compared to elements of the product of matrix Φ and vector x. Hence, according to this representation of transformed vector x by vector Y energy is lost.

Temporal noisy energy estimation error case

In this subsection, variance of the elements the energies of which are considered are increased by one as in all previous noisy cases. This modifies the non-centrality parameter of the chi-square distribution and consequently the parameters of the approximating Gaussian distribution. Figure 3.23 for this noisy case is given below:

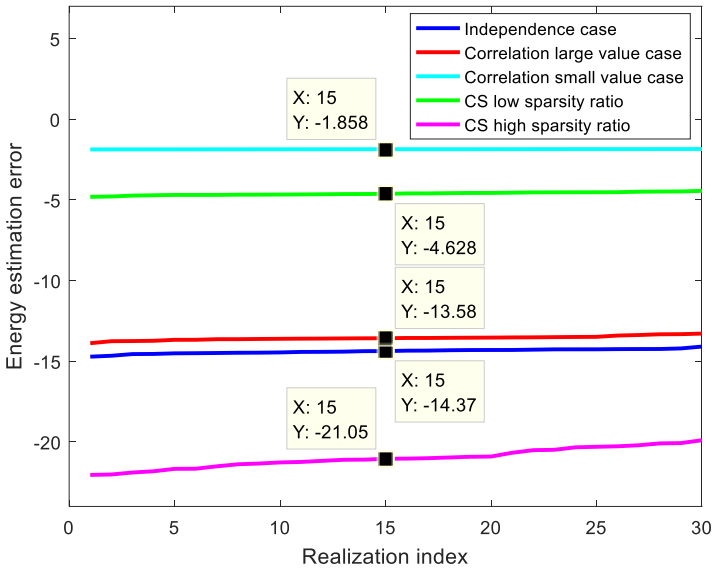

FIGURE 3.23: Energy estimation error for temporal noisy correlation case

The first observation for this section is the negative sign of the errors for dense and CS based cases investigated. Furthermore, the decrease of correlation leads to smaller error as depicted by the cyan colored curve. Hence, in a practical application, the correlation can be decreased so as to achieve a more accurate energy estimation as opposed to the independence case. The low sparsity ratio assumed a value of 0.4 while high sparsity ratio value was set to 0.9. The common correlation value of 0.1 is justified as follows: as the correlation increased the error was shown to take more negative values represented by continuously steeper curves. Thus with error equal to one for zero correlation the optimal correlation for zero error approaches lies between zero and 0.1. Relative to value assumed for simulation the smallest value of 0.1 was considered, as the minimum value assumed given the specified range. Finally, the property of lossy representation commented in the previous noiseless subsection was found to increase according to the negative values of the figure for this subsection.

<u>Spatial noiseless energy estimation error case</u>

As in the spatial cases already investigated, the mean value of elements whose energies form the sum is modified in the same way by setting the second moment equal to 1/30 by the random sensor selection scheme justification that was derived above. The related Figure 3.24 for this case is given below:

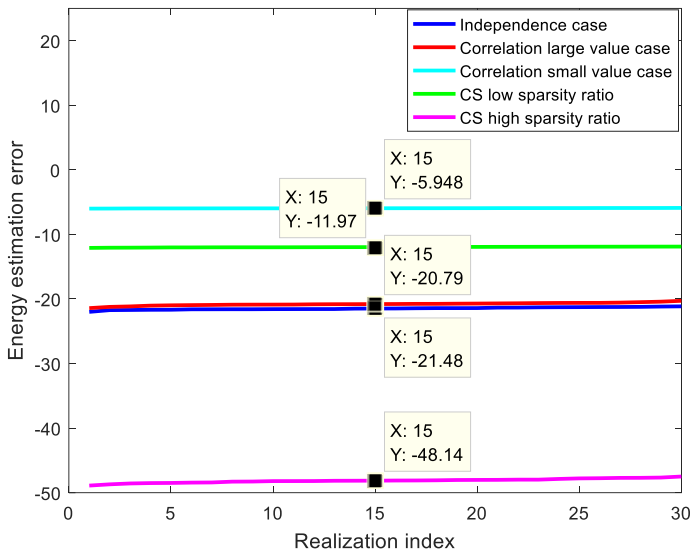

FIGURE 3.24: Energy estimation error for spatial noiseless correlation case

A similar observation is the negative valued energy estimation errors for dense and CS based cases. The independence and dense correlated cases indicate negligible performance difference. Similar to previous subsection small correlation i.e. equal to 0.1 results in lower error. CS low sparsity ratio with a ratio value equal to 0.4 performed slightly better than the high sparsity ratio case with a ratio of 0.9. Another interesting observation is that dense correlation case with minimum correlation, as depicted by cyan curve, performs better than CS low sparsity ratio case shown by green curve.

<u>Spatial noisy energy estimation error case</u>

Increasing variance of the elements whose energies form the sum by one is the main assumption as in all noisy cases. Figure 3.25 for the spatial noisy case is given below:

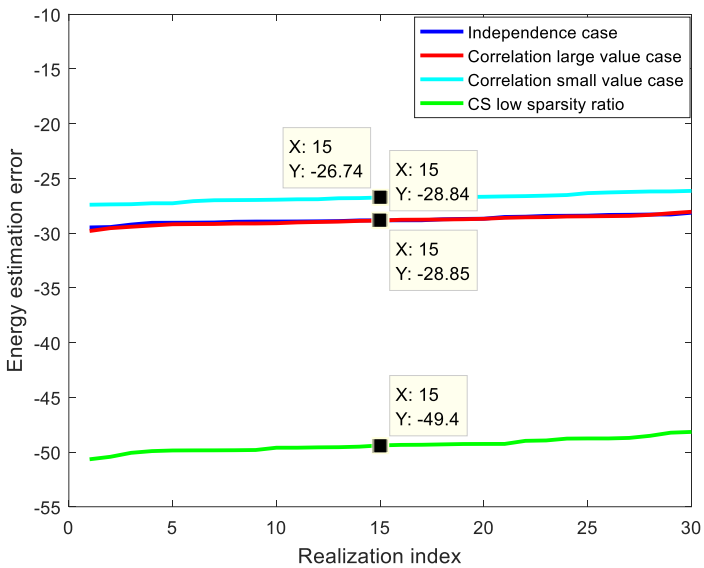

FIGURE 3.25: Energy estimation error for spatial noisy correlation case

The independence and dense correlation cases perform the same, while the small dense correlation case depicted with a value of 0.1 conveys improved performance. However,

the most important finding of this subsection is that CS low sparsity error case, with sparsity ratio equal to 0.4 and correlation equal to 0.1, performs much worse. The CS high sparsity ratio indicated even larger negative error and is not depicted due to figure scaling. Additionally, this section derives the more pronounced effect of lossy representation evident from the negative signed energy estimation error values. Hence, the model parameters must be accordingly modified to compensate for the quantification issue.

Spatiotemporal noiseless energy estimation error case

This section fairly considers the temporal and spatial correlation cases into a joint perspective with same assumptions as in reconstruction error spatiotemporal cases. The findings of the simulations for this case are shown below in Figure 3.26:

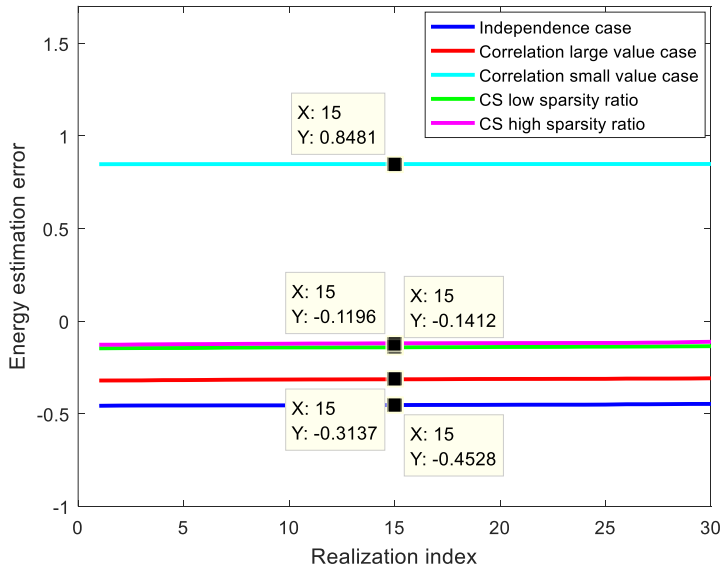

FIGURE 3.26: Energy estimation error for spatiotemporal noiseless correlation case

It is apparent that both CS based cases perform best indicating a reasonably small error i.e. achieving an almost perfect match. The worst performance was observed for dense correlation case with a minimum value of 0.1 concerning correlation, as shown by the cyan curve. Independence case is slightly outperformed by the dense correlation case showing a smaller estimation error. The crucial observation in this case is that it exhibits larger errors compared to the previous temporal noiseless case and smaller compared to the spatial noiseless case. Hence, this «weighted» contribution of the two separate cases is verified with the temporal noiseless contributing the most as with slightly smaller errors compared to this spatiotemporal noiseless case.

Spatiotemporal noisy energy estimation error case

This section similarly assumes a fair consideration of temporal and spatial correlation and an increase of variance of elements whose energies form the sum by one as in all previous noisy cases. The results for spatiotemporal noisy case are provided in Figure 3.27:

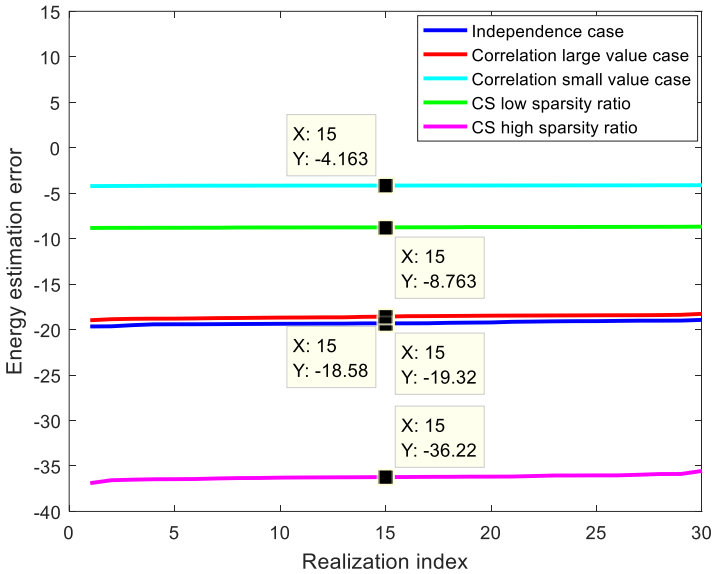

FIGURE 3.27: Energy estimation error for spatiotemporal noisy correlation case

The first observation regarding this case is similarly the inefficiency of the representation of the transformed, via matrix Φ, vector x by vector Y due to largely negative errors. The aforementioned «weighted» contribution of temporal noisy case and the spatial noisy case to this spatiotemporal case is also verified. Furthermore, independence case shows similar value in error with the dense correlated value with correlation equal to 0.9, while the dense correlation case with minimum correlation of value 0.1 exhibits the lowest error in this case. Finally, the CS case with low sparsity ratio equal to 0.4 and correlation equal to 0.1 is only outperformed by the dense low correlation case. CS high sparsity ratio case performs the worst. An interpretation of this observation is its high inefficiency given the model and simulation assumptions.

3.6.7 Overall results interpretation

Concerning the independence case in reconstruction error derivation the worst performance was observed. Hence, in a practical scenario introducing the maximum possible correlation will directly improve the performance thus lowering the reconstruction error. Moreover, all related cases both noiseless and including additive noise, achieved low reconstruction error with small values of sparsity ratio either 0.1 or 0.2. This is in accordance with achieving low errors while discarding the most of the data samples. The straightforward impact of such a finding is the significant reduction in computational complexity. Another important remark is the lower error achieved in the high sparsity ratio case. This indicates that approaching the dense case with the highest correlation value greatly improves the performance as well. The errors in this section are also characterized by a significantly wide range of values.

Regarding the noisy cases together with independence principle of elements comprising the sums involved in the derived average values and consequently mean values and the Gaussian N(0,1) assumption resulted in distribution overlapping. The quantification of such observation depends to a large extent on the simulation assumptions. This translates to the distance of distribution of vector Y and that of the product of matrix Φ

and vector x. The assumed noise variance is order of magnitude larger than the decimal value of the elements mentioned above. Thus, as the distributions overlap the effect of negative valued errors arises with the meaning that values of vector Y underestimate the values of the product of matrix Φ and vector x. The above translates that with the distribution curves overlapping the probability that the error remains relatively in a short value range is with low probability. As distribution variance further increases it is likely that the errors derived are characterized by value ranges each with an asymptotically specific probability i.e. a determined number of occurences. Relative to energy estimation error cases, the temporal noiseless case results show that independence outperforms the dense correlation case, as opposed to the previous section of reconstruction errors. In the rest of the cases, the results indicated either similar performance or slightly lower error for the dense correlation case. The negative signed errors are a consequence of lesser values of vector Y compared to values of the vector resulting from the multiplication of matrix Φ and vector x. Although the simulation results for the noisy cases assume independent noise with respect to variables forming the sum in Eq.(19), a practical scenario would include noise correlation which would improve performance. Another remark concerning the CS low sparsity ratio cases is the shift observed to higher sparsity ratio values for achieving the lowest error. Considering the minimum correlation value of 0.1, reducing correlation in a practical scenario can achieve an arbitrarily low estimation error.

3.6.8 Effects in topology and routing of Wireless Sensor Network design

The initial comment regarding the optimal network topology is a set of concentric circles each of which defines a neighborhood with a subset of nodes. Considering the feasibility of adopting standard topologies such as square grid where the node deployment would result in sub-optimal area coverage the tree network structure is assumed, as already stated. This is based on the balance between area coverage and circular symmetry. In order to ensure network operation, the data similarity could aid in optimal dynamic topology scheme. Such redundancy could contribute to fault tolerance and be further compensated by subset of nodes mobility. The randomized model assumed in the case of this section could fill in the gap of covering network holes and ensuring coverage and connectivity. Together with redundancy topology adjustment due to network failure or energy depletion could promote decentralized operation of the network. The proposed probabilistic scheme leading to quantification of reconstruction error and energy estimation error could further integrate correlation and CS compressibility rule to minimize energy consumption given the constraints of delay and bandwidth.

Aiming at optimizing routing, the spatiotemporal correlation aspect could provide optimal error values and support decentralized operation which is achieved as follows. In the first stage and in order to achieve representative values the heavy computations should be shifted to the sink node, on the basis of not receiving abnormal readings. Given the resulting redundancy, the network nodes gradually gain knowledge of the representative values thus not requiring sink nodes. Hence, the proposed scheme supports localized computations and achieve decentralized operation. However, all the above are confined to the environmental monitoring case. This clearly implies that abnormal readings are rarely measured, temporal correlation complies to the slow

variation of current readings from the past values at sensor level and spatial correlation assumes slow variation in the readings in a given sensor neighborhood. In a more generalized setting with reduced scalability i.e. in cellular networks or MIMO cooperative relay networks the initial WSN based sensor selection scheme translates to the relay selection scheme for massive connectivity. The achievable decentralized property is directly equivalent to reduced feedback requirement which not only reduces overhead but also ensures reliable communication due to the emerging redundancy. Despite the lossy characteristic of the model in the energy estimation error cases, the efficiency of the scheme is verified.

3.6.9 Applications to Wireless Sensor Networks

Wireless sensor network design in the framework of communication systems is characterized by common limitations. Regarding design issues, a fundamental problem is the achievement of optimal results on terms of specific metrics given the definition and formulation of application specific constraints.

Despite the rapid evolution of WSNs, energy efficiency still remains the issue of primary concern. Since energy harvesting techniques were not proven adequate to compensate for node energy limitation, efficient duty cycling techniques and optimal scheduling of node sleep intervals have been developed to coordinate active and inactive sensors during dynamic network operation.

Scalability is the next important issue. Data redundancy can provide the concept of effective scalability based on sensor subset selection. To further exploit correlation and redundancy introduced bidirectional information exchange between pair of nodes is also supported which will directly result in the decentralized property mentioned above.

In the context of dense deployment, interference mitigation is a major issue. Directional antennas and optimal sensor selection scheme can help alleviate this issue combined with distributed transform coding in a post correlation learning stage. The important extension of correlation introduced by the proposed scheme is the balance that must be achieved between sufficiently large correlation in terms of reconstruction error case and small value in terms of energy estimation error case. Due to capacity, effort to reduce redundancy and increase of data diversity in case of heterogeneous network or a network portioned in several segments is of paramount importance. Another issue is privacy and data security. Exploiting redundancy relative to the proposed scheme, subset of sensors less prone to security breach and overhearing could be selected. The effect of handoff management in cellular networks is of major concern especially when applied in an area not covered by node deployment. Given that the proposed scheme supports node failure management, the proposed randomized scheme can improve handoff management and ensure connectivity. Relative to the QoS requirements the decentralized property can contribute to reduced complexity and achieve optimal performance. The bandwidth allocation and issue of latency especially in scenarios such as healthcare monitoring, disaster event detection and target tracking are also of major concern for future wireless sensor networks.

As a final remark, IoT networks are also characterized by the need for energy efficiency improvement. Additionally, heavy traffic load also emerges as an issue which is

addressed by smart node deployment with much larger scalability. The key performance issues for IoT networks [142] is the high heterogeneity, location awareness, effective routing and high data processing complexity. Considering the different types of data involved, our proposed probabilistic scheme can be applied in a densely deployed IoT network. Finally, cost is the critical issue determining the design of a WSN and is the contribution of many aspects such as low power and autonomous sensing nodes and the feasibility of deployment and constraints considered by a specific application.

3.6.10 Conclusions and future work

A fully randomized probabilistic scheme is proposed for evaluating reconstruction error and energy estimation error accounting for temporal, spatial and joint spatiotemporal correlation metric. In the former case, sum of readings or generally random variables forming sum leading to average values and consequently mean values as elements of the related vectors and matrix. The latter case assumes a likewise scenario with the difference that the sum of energies of the related variables are considered. Gaussian statistics are applied in all cases considered. The dense independent and dense correlated cases are investigated along with CS based compressibility rule and low/high sparsity ratios values. The results are indicative of WSN performance with application to WSN communication systems.

The proposed scheme could be integrated with CS initial l_0 norm problem formulation which could be solved by the known convex relaxation or by means of a greedy algorithm. The results of error magnitude could be generalized to performance optimization based on correlation which can be assumed as varying throughout time or space i.e. different in each node. Channel fading conditions along with non-Gaussian or correlated noise also pose an attractive case for investigation. Topology and routing are also promising directions for introducing probabilistic scheme and distributed CS principle. To conclude, cross-layer optimization could also be addressed by translating the proposed scheme to the benefits of each respective layer and formulating performance as a CS based optimization problem with diverse constraints that not only cover environmental monitoring but also extent to real time requirements such as target tracking, disaster event monitoring or advance healthcare monitoring networks.

4. WIRELESS CONTENT CACHING AND CS APPLICATION TO SOCIAL NETWORKS, MEDICINE AND SMART EDUCATION

Dynamic web content, in terms of volume and type diversity, poses significant challenges for storing, exchanging and processing of data in current communications networks more than ever given the Big Data deluge characterizing the rapidly evolving communications network generations [143], particularly mobile network data, which are already becoming a reality. Referring to the issue of latency, which many networks face, content caching in terms of the spatial dimension will store frequently accessed files for quick acquisition by a user that specifically makes a file request. The latter acquisition may provide the solution for network fault tolerance and efficient operation, even if a reconfiguration is needed.

It is clear that the concept of frequent access along with file popularity generally admit different interpretations. Hence, a comment would suffice that dynamic content delivery relies on the basis of available content copies in a way of not becoming obsolete. To that end, the concept of a constrained optimization problem is completely valid and the issues of resource allocation and complexity also emerge. A relevant problem to the latter is to what extent a cached file needs to be updated and in which portion of the network this requirement is to be fulfilled. Hence, the architecture and dynamic topology and routing of the network are inextricably related to content caching. Caching is also a means to combat information redundancy in the context of content or information centric networking under the name of edge caching. Indeed, frequently accessed files are to account for duplicate file downloads and caching confronts this very issue. A major tradeoff is that of bandwidth vs. storage cost in a dense mobile network. It is also evident from a plethora of networks that frequently accessed portion of files is a small fragment of the total files stored hence the rule of CS based sparsity is a first attempt to relate CS to the analysis of this file. As a clarifying comment on the caching strategy efficiency, many requests for popular files may be locally satisfied [144] thus alleviating mobile traffic and overhead. Still, this statement is also adept to introduce shortcomings in terms of file popularity and diversity along with the localized memory storage size of each user in the network. As for the latter storage, one strategy is the equal spreading of files in users cache that may efficiently define a localized content caching operation but may hinder file diversity. Surely, there is a tradeoff in terms of a cached file in multiple locations as opposed to file diversity that could serve numerous requests in a timely and resource efficient manner. To that end, content update information is another issue that may require increased information exchange that may not be limited to a localized scale. Additionally, referring to the centralized scenario as opposed to a distributed one, the latter is first applied to verify whether it could cover specific file request otherwise the central data base is accessed. This directly impacts latency and complexity issues along with energy consumption bottleneck commented in previous chapter. This section provides a networking performance analysis in terms of the hit probability. This concerns cellular networks that are one step less densely deployed compared to WSNs of the previous section. Thus, the issue of heavy data traffic management is a key performance metric in such networks. Caching hit probability is a technique aiming at alleviating backbone network congestion and latency reduction by storing frequently accessed files i.e. files of high

popularity as stated in literature as opposed to the rest stored in base station entity. The hit probability is defined as the probability of a file requested by a user together with the successful delivery of the file to the user.

4.1 Chapter outline

In this chapter, caching hit probability is investigated in the context of communication networks considering uniform, Zipf-like and normal probability together with comparisons and interpretations. Similar to CS theory, the cases of few and many files cached is analyzed by simulations conducted whereby considering file segmentation convergence is verified. A brief notion on privacy issues for SNs is provided along with application of CS in practice of medicine and smart education proposing an interactive classroom paradigm.

4.2 Probabilistic content caching and CS

4.2.1 Introduction

The results in this section are related to our publication [J1]-([145]). Categorization of wireless networks to cellular and ad-hoc networks [146] also relate to the challenge of designing WSNs that operate in a decentralized autonomous manner rather than a centralized operation. Thus, cellular networks required base station entity for information exchange coordination while ad-hoc networks are without infrastructure. The benefits of ad-hoc networks are autonomous operation, self-healing and self-configuration and flexibility as well. However, their constraints are requirement of optimal node-level performance, limited throughput due to system loading and extensive latency. Hence, centralized coordination provides benefits such as traffic management along with latency mitigation.

Cellular networks coverage area is divided in hexagonal cells. Each cell consists of its own base station with transmitter, receiver and control unit and an antenna with its frequency set. The use of the same frequency must not take place in adjacent cells so as to reduce interference. Moreover, CDMA scheme is widely employed as a means of transmitting through the same channel by assigning a specific code sequence to each transmitter. Thus, frequency diversity, multipath resistance and privacy are ensured as well as no performance degradation when many users communicating are involved. The drawbacks are self-jamming, near-far problems and QoS degradation. Cellular networks achieve high capacity, less local interference along with reduced transmission power especially when employing networks of very small size such as cellular networks. The disadvantages are need for infrastructure and the necessity for handovers as mobile users move from one cell to another.

Caching as investigated in this paper is used in offline as well as online and cooperative schemes in cellular as well as ad-hoc networks. However, the case investigated concerns cellular networks.

4.2.2 Why mobility-aware caching

Information exchange in cellular networks is confronted with latency and connectivity issues as reliable communication must be guaranteed for optimal operation. Backbone

network congestion must also be alleviated to avoid QoS degradation. This is exactly where caching emerges as a key solution. Caching enables storage of popular in terms of user requests files in an information-centric manner, hence prioritizing information as opposed to host requesting the data. Caching reduces the need for base station intervention except only in cases where files are not stored in users' cache. This results in reduced overhead and latency, improved overall performance, reduction of heavy traffic load as well as reliability and bandwidth saving. Static as well as randomized schemes have been developed the latter fitting into the approach of this thesis.

Energy efficiency is a fundamental issue that characterizes cellular networks as well as WSNs as already stated. Addressing the limitation of energy budget, user mobility is applied with the common assumption of arbitrary movement inside the considered cell. User velocity is proven to be a crucial parameter for data exchange. Users with high velocity must cache most popular files to address high contact rate while users with low velocity must also cache most popular files to meet their own requirements. Finally, medium velocity users must store least popular files to avoid duplicate caching.

In the context of content caching, file segmentation [147] which are stored in different users' cache has a straightforward impact on performance. However, number of segments required for successful file recovery is dependent on file size, which is not further investigated. The derivation of specific patterns regarding user mobility along with randomized schemes addresses many fixed network topology issues. The users assumed in the investigation of the three respective distributions though are free to arbitrarily change position in certain areas of specific radius, the main condition being the number of users in each area becomes less as the transition of inner to outer radius areas takes place. The mobility model for this analysis is either the random walk supporting arbitrary movement inside pico-cell and temporal dependence model accounting for correlated parameters such as user velocity.

4.2.3 Past related work

We now review past literature that directly relates to our analysis and enhances its merit by performing thorough comparisons. We review general works that compare to our analysis in a generic manner that can be integrated with caching strategies. Proceeding further, works specifically tailored to caching schemes propositions are then reviewed with detailed comparison to our approach also conducted.

We first review a well-structured survey [148] which highlights the main issues addressed in the context of communication systems which closely relate to caching strategies. The first issue is network users scalability and high traffic volume and diversity, issues that impact on the necessity of content caching in a manner of minimal replication of cached content and serving of file requests in a distributed manner so as to avoid centralized approaches with increased signaling and information updating overhead. The mMIMO technology could provide an effective solution for diversified file content popularity serving a selected subset of users that have made the particular content request which in our case could translate to performance comparison of hit probability in a massive setting exploiting correlation and setting the effective number of cached files considering the three distributions as in our analysis. Dense small cell

deployment is another direction that 5G systems take in order to limit power and balance frequency reuse in a real-time scenario while interference issue emerges as a limitation. In turn, caching hit probability may decrease as a consequence of misidentification of user related to file content request if interference is not accurately modeled in a probabilistic manner. In a cell level perspective, related handoffs could even complicate the situation. As for bandwidth usage, caching strategies also demand more use which in the 5G network could be satisfied with higher unused spectrum bands with the shortcoming of increase attenuation as a result of higher frequencies use. This could also fundamentally degrade hit probability and also question the efficiency of the Zipf-like distribution as a means to model caching hit probability. Besides high data rates, latency minimization optimizes energy efficiency of the whole network which is still a target metric to be achieved in caching file networks by eliminating duplication and redundancy. As stated by the authors in [148], spectral efficiency connected to bandwidth allocation is a major issue. This issue may be addressed by reducing file exchange and localizing the bandwidth resource problem. NOMA scheme may address the increase in user accommodation which in turn gives prominence to content caching for low latency fetching from users' memory storage. For each cached file, code signature uniqueness can be the key methodology for frequency reuse and secure caching in an ultra-dense infrastructure. Another review paper [149] accurately reflects the 5G goals to be achieved i.e. even higher data rates above those reached by 4G and energy efficiency which is anticipated to play a significant role as a bottleneck for future networks. Concerning these two metrics, a certain conclusion that can be reached is that their optimization is dependent on many parameters. With respect to caching, diversified popularity is indicative of data diversity that are involved in storage for frequent rapid accessibility which, in a probabilistic context, depends on statistical modeling and communication channel conditions in the portion of the network area where caching is conducted. Mobile caching is also a major scheme which may serve as a means for increased connectivity and efficient file availability but may also negatively impact achieved data rate. Finally, the content caching strategies could adapt to indoor or outdoor networks in an entirely different way and one promising extension of our probabilistic analysis may be realized by selecting either of these scenarios with different distributions considered and different optimality results reached. Proceeding to past literature with direct caching scheme proposition, a work [150] adopts content centric networking to introduce payment for caching of a user's files. An online caching algorithm is proposed that assumes caching and retrieval costs otherwise blind to file popularity. Emitting from the related comment in this paper, our analysis could further provide an optimal distinction between content cached as in our case and location awareness the latter achieving a distributed caching scheme at the expense of tolerating the drawbacks resulting from centralized coordinator absence in the network. Edge caching contrary to core network caching could also benefit from our probabilistic analysis along with justification of file segmentation approach as a function of network area of interest. File popularity also impacts energy efficiency as a straightforward deduction due to reduced energy by storing frequently accessed data and jointly optimizing routing as well. An interesting approach to the content caching problem which ignores popularity and decision is made on content request could benefit from CS principle when applied to such networks. Specifically, a decision could be made as

to simultaneously evaluate popularity prior to request as an a priori information for effective low complexity decision along with prediction model for request given a certain file. A conditional probabilistic setting could account for request history so as to optimize future access to a certain file. Indeed, a well-posed problem could be that of applying CS sparsity as a means of few cached popular files in a user's memory. An equivalent problem could formulate the problem of not caching a file if its popularity is below a predefined or dynamic traffic aware threshold or as a terminating criterion of the algorithm the case when cache is full of content stored. Furthermore, the so called hit probability that encapsulates the file request together with successful file delivery is tantamount to caching and retrieval costs assumed in the latter paper. These costs surely admit a modeling approach with an additional cost-aware constraint to the caching optimization problem. The proposed online scheme uses a minimum cost path whereas our analysis also assumes models such as the random walk model or the temporal dependence model accounting for user velocity hence a mobility assumption. Decentralized topology is also assumed that combines with popularity nonawareness and low overhead. Although our analysis of transmitted power adheres to increased user density with network cell radius decrease, the above assumption relaxes the necessity of this rule and fits into a more dynamic topology setting. Also worth noting is the case of tracing a cached file at multiple locations which can be thought of as a measure of redundancy. This could be confronted by narrowing down the possible sources from which the file can be retrieved by additional criteria such as minimized cost and low latency. Statistical modeling can be refined combined with CS application as a means for rendering tractable a complex optimization problem. One step further form assuming caching problem independence, inter content correlation could contribute to reduce overhead but also be dependent on popularity. This statement could also affect the parameterized Zipf's distribution which by definition determines how many files are significantly popular by selecting parameter value as in our analysis. Although our discussion given [150] is ignorant of popularity the algorithmic derivation as a consequence of determining cost by answering whether a cached file is not cached at another location could assist in calculating its popularity. This could be well integrated with the three distributions considered in our case. The flooding strategy affects the observation of a file not cached at a given user but also implies overhead. Moreover, caching time intervals have diverse impacts on latency and need for information updating which can be considered as another dimension of the optimization problem. To that end, content eviction is not to be accounted for alone but in the context of the subset of users from which the cached content is evicted. Concluding, randomized caching costs, caching time intervals and cached file popularities are promising directions of extending distribution types for the modeling of the latter. We now proceed to the work of Shiral [151] which serves as the backbone of comparing cellular and ad-hoc networks in the content caching context. With reference to cellular networks, the asymmetric, in terms of computations, property favors a centralized scenario and paves the way for investigating the content caching at optimal locations, including mobility concept as in our analysis. The main properties of a sensor network is order of magnitude larger user deployment resulting in traffic volume and heterogeneity and energy efficiency. Further notion for WSNs is provided in detail in section 3.6. Worth mentioning is the multi-hop pattern relative to the delivery of a

cached file after a request to the desired user/destination. This consideration could alter the parameter of the Zipf-like distribution and deviated from the uniform distribution we adopted as a comparison benchmark. As stated above, file redundancy could confine caching to diverse files while degrading data rates. Regarding ad-hoc networks, their property of base station absence renders them vulnerable to issues such as dynamic packet request which in the language of content caching may lead to varying hit probability and popularity value of each cached file. Hence, a localized file exchange in each portion of a bigger network could more accurately describe functionality of an ad-hoc network. The next paper by Tehrani et al. [152] deals with the hybrid architecture of devices communicating via base station and devices communicating autonomously with each other, a scheme that dictates meticulous routing throughout the network, interference management and, relative to our analysis, content caching modeling in a static framework but also integrating D2D communications with mobility aware caching. To that end, the latter consideration must coordinate resource allocation in a cross-layer perspective. In the previous section, cooperation in the CR context was boldly stressed as an efficient means of refining and improving spectrum sensing performance. However, cooperation can also contribute to file duplication reduction and low latency with popularity aware caching by proper setting of information routing and MAC scheduling. Another methodology to be vastly researched is relying for effective content caching along with the hardware and coding constraint to be compensated for. Additionally, the analysis in this section deriving transmission power can also be incorporated with spatial correlated information exchange and specifically investigating the channel effects such as shadowing which may cause deviation from the model of lower transmission power in a decreasing radius network cell. As for interference management, the most well-posed proposition is the latter two-tier hybrid scheme where base station may also contribute to coordination but localized coordination may also promote interference mitigation. In the aforementioned context of caching due to payment benefit, relaying may also involve payment as an incentive, hence, file popularity can derive the payment for promoting storage at a user level. A probabilistic approach as our Bernoulli-wise certain probability that a file is cached and the complementary probability concerning a file not cached can extent to modifying payment according to the value of aforementioned probability as a function of other parameters such as maximum latency or harmful interference. An interesting perspective regarding interference is that spectrum may be allocated by means of payment to more than one user promoting a cooperative scenario with the critical constraint to the problem that of sufficient user distance so as not to be impaired by the level of related interference. To that end, game theory is an intriguing methodology that could refine our probabilistic analysis in search for a stable solution. Finalizing with the contrast of this work to our analysis, the localized «learning» of two devices in proximity could provide the means of autonomous operation and caching of files with strictly high popularity. Following the same context of D2D mobility aware caching and user velocity issue, the authors in [153] introduce a low complexity dynamic programming algorithm and a time efficient greedy algorithm as an efficient approach for content caching. Moreover, cached file density with respect to popularity seem to be interrelated, a property which could translate to transmission power related user density from a different point of view. The idea of ultra-dense network can be applied

to a certain extent from an interference point of view and content caching may also demonstrate limitations that become dominant when exceeding deployment density beyond a certain scale. This contrast may also be reflected by caching statistical modeling. The rateless fountain code use complies to our file segmentation assumption and the gap of the total and equivalent probabilities can be utilized to assess file recovery quality on the basis of sufficient segments collected. Similarly, linear complexity proposed algorithm can be modified by CS principle application with reference to the case of sparse and dense scenarios. Hence, this NP-hard problem can be alternatively solved with reported low computational burden. A well-posed statement which reflects the randomized availability of a certain segment to be recovered by a user that makes the request is that concerning the maximum delay constraint after the expiry of which segment is recovered from BS. This can be also statistically modeled jointly with Poisson contact times along with exploiting memoryless Poisson property. A very insightful feasibility of applying CS principle in the caching placement algorithm is offered by the dynamic programming algorithmic concept. In a comprehensive statement, CS may be recursively applied but also in a final refining level to derive a solution with constraints corresponding to realistic conditions satisfied. We also stress the similarity of the content placement strategy to a CS greedy algorithm so as to reduce complexity. The latter can also be related to the concept of decreasing gain as the algorithm element subset grows which may for instance be realized by an accurate real-time distribution modeling the cached file segments along with revisiting algorithm termination conditions. It is our opinion that random caching serving as a comparison benchmark in [153] should be further optimized jointly with mobility mechanism as opposed to the proposed greedy algorithm, an approach which could facilitate CS application with respect to varyingly parametrized Zipf distribution. The random waypoint model is also applicable in our analysis as it fulfills the requirement of user movement speed and location in a probabilistic manner. Among the simulation results in this paper, it is worth pointing out the result stating that high enough velocity implies assuming that each user has access to all users' cache. This is quite indicative as a mobility pattern and can clearly be combined with our analysis in order to probabilistically encapsulate user velocity in specific value intervals and to thus learn the optimal distribution based on file popularity and user velocity. Concluding, as the authors emphasize, user density also referred to in our analysis can be jointly optimized in the dynamic programming content in order to avoid duplicate files cached and serve the needs of the entire large sized network. Even further, another interesting paper by Liu et al. [154] abiding by mobility awareness and mobile edge computing, introduces a coded probabilistic caching scheme for compensating for caching efficiency and throughput maximization all the above in small scale cellular networks. The merit of this work is the MEC-enabled tradeoff investigation of mobility, caching and channel selection diversity. Hence, this provides a unified perspective of integrating our analysis, comparatively probabilistic as well, by accounting for fading channel conditions along with Zipf distribution parameterization and different statistical modeling consideration. In order to accurately state system model in the latter paper, interference and frequency reuse are excluded issues from the problem assumptions. This paper is very insightful as the particular point of modeling the remaining data to be collected also assumes the uniform

distribution as opposed to the optimal one. This is identical to our considerations where evaluation of the popularity model includes uniform, Zipf-like and normal distributions. Motivated by the analysis of [154], a joint probabilistic problem formulation could model both file popularity and user density and throughput in a given area from which the content request can be satisfied while minimizing the segment recovery from the small BS. Additionally, the increase in caching hit probability demonstrates an increase in throughput, a result which requires the mobility intensity diversity, as stated by the authors in the respective paper. As concluding remarks, the popularity skewness and mobility intensity must be carefully balanced to avoid sensitivity induced deviations from acceptable optimized parameters values. Hence, skewness must be increased to a certain extent otherwise the observed gain will be outweighed. The most crucial point of merit concerning the analysis in this paper, is the exploitation of mobility and distributed storage to address backhaul link issues. The above could relate to our analysis by Zipf parameter optimization and inclusion of channel conditions as a means to further achieve optimal performance due to this additional knowledge inserted in the optimization problem. CS optimization theory could provide the means for low complexity derivations in the same context of the heuristic algorithms used to reach optimality. Another work [155] derives the benefit of applying mobility concept to compensate for the limitation of retrieving file portions from distant devices as a consequence of interference and channel conditions. The gains achieved are quantified in terms of coverage probability. The positive effect of mobility is serving as a solution to the issue of not caching the entire file in a certain device and thus, given this practical limitation, improves coverage probability. An indicative assumption of this paper is the independence of the localized area from which the file portion is retrieved with respect to the location of this area in the network. This renders performance evaluation applicable to a subset of users irrespective of their location. The context of the scheme applied is based on ignoring thermal noise and considering interference power instead. Moreover, the authors set the prerequisite of independence as a means of using the product rule to derive the total interference distribution calculating in turn the coverage probability. It is straightforward to adopt these calculations in comparing caching hit probability by conditioning on not only Rayleigh but diverse fading distributions. These channel conditions could contribute to asymptotic performance by varying the relative coverage areas with respect to cached files in the optimization problem. As also stated by the authors, multiple cached files divided in portions cached in a distributed manner is a challenging extension which also relates to our probabilistic analysis, the latter being possible treated as an increased dimensionality problem. A representative paper by Wen et al. [156] addresses caching hit probability optimization with respect to content placement probability employing random caching strategy all the above in the large scale heterogeneous network model. In this large scale context, MIMO wireless technology is another paradigm to boost throughput and improve performance in terms of content placement, coded caching and hit probability related reliability with the fundamental tradeoff referring to implementation complexity. It is straightforward to relate the optimality of probabilistic placement compared to its deterministic equivalent to the CS random matrix perspective as a means to overcome deterministic limitations. Hence, CS can extend our analysis by utilizing sparsity rule for cached content with investigation of parameter of Zipf distribution reflecting file popularity diversity.

Considering the effect of fading and additive noise that commonly characterize channel conditions, the assumption of the authors in [156] about interference being the dominant bottleneck is still valid and the integration of all the above poses a challenge in terms of impact of each channel parameter on hit probability along with accurate statistical modeling. The authors employ a quite insightful approach of storage capacities modeling the latter as varying between different tiers which is all about verifying the probabilistic content placement concept for extending our hit probability derivations. The consideration of both link reliability and content availability admit the rule that content may be available but reliability not guaranteed and vice versa. Given a communication scenario with increased path loss exponent introduces the limitation that transmission power increase does not have the dominant effect on association probability which simply translates to the fact that coded caching approaches and exploitation of other degrees of freedom improves content placement and deployment density. Relative to the joint contribution of physical layer parameters and content popularity, a reasonable question is whether additional channel conditions can be efficiently included in the path loss exponent value and additionally whether identical probabilistic model for popularity and channel conditions can have a major overall effect on reliable low latency content caching. The placement probabilities in a more practical case are contained in the [0,1] interval and hence as dictated by moderate popularity the spreading of cached content in BSs is valid. This could provide interesting results if combined with uniform SIR thresholds in the network area. It is also imperative to note the context similarity of non-convexity of the multi-tier content placement problem with CS based non-convex problems and therefore suggest CS compressibility application for such a problem since the derivation of associated thresholds are thoroughly considered. Finally, coded caching and cooperation can result in an apparent performance improvement via probabilistic analysis such as in the next subsection. Proceeding one step further in the context of one-hop D2D caching strategy, Zhang et al. in [157] propose scheduling and power allocation problem by solving for scheduling satisfying SIR and transmit power constraints along with the second problem of power allocation under the rate maximization. Zipf-like distribution is accordingly utilized and the non-convexity statement traced once again also strongly implies the CS principle consideration with the similarity of file segments considered in our analysis. Clearly, the approach adopted assumes single file request, hence the multiple file extension with the file segmentation context may lead to computation intractability but may point towards near optimal approaches in a similar to our analysis randomized context. Clearly, for achieving near optimal performance the one-hop distance assumption not only implies spatial correlation and also neighboring file availability, both of which must be investigated as to the heterogeneous network assumption and throughput maximization. The contribution of this paper also bypasses the communication reliability issue between any two users as it merely considers cached file availability. Thus, integrating channel status is a straightforward claim which may well result in non-closed form expressions necessitating numerical methodology for solution. The authors also provide detailed comments on the use of the Zipf-like distribution, based on which our uniform assumption and non-skewed symmetric normal distribution indeed provide statistical insights on the popularity defined problem. Probabilistic analysis given D2D communication could also provide

modeling of request file diversity i.e. no multiple request of file from a common helper along with mobility-aware coded caching so as to allow frequency band reuse in a relaxed reduced interference manner. Contrary to the latter past literature reviewed in this section, the differentiation of [157] is that additive noise plus interference is formed in the SINR calculation thus a more complete characterization of the link reliability model. An interesting assumption in this paper is that SINR may be satisfied in terms of solving the respective scheduling problem in a manner of links sharing the same user subset but not working simultaneously. Moreover, the optimization problem posed in this paper accurately reflects the principle of optimal solution feasibility in terms of SINR constraints and link removal when needed. It is our understanding that this iterative process can be revisited by applying CS greedy approach and investigate the accuracy versus speed algorithmic tradeoff in a mitigated interference attempt. In the latter algorithm serving as an extension the increment as to increase SINR, the significance of the increment value remains to be seen as a measure of complexity and power allocation efficiency. The optimal distance between users is carefully balanced in the reported results and deviates from the «hot» server referred to in our analysis thus content is more diversified. Finalizing with this paper review, the strategy of selecting better channel conditions i.e. rich become richer could demonstrate interesting results if compared to a more uniform perspective of allocated resources in the same scenario and also combine with our hit probability comparisons. Therefore, a clear limitation could be complexity but also issues of file duplication in sparse as opposed to dense deployment areas and optimized latency involved, reflecting effective file popularity variations. Cooperative scenarios and multiple antenna spatial diversity exploiting scenarios constitute promising approaches as well which could further refine optimization results in a decentralized manner at the well-established expense of complex implementation and hardware constraints. The final two papers reviewed are inherent to the mathematical background concerning Zipf law as a benchmark to assess addressing the use of the respective quantity in the simulations of our next section. Contrary to independence based assumption as the main pylon of this thesis, Zipf convergence dictates consideration of past file accesses to further refine the caching hit ratio. Furthermore, cache everywhere strategy may be hindered by weak channel conditions or intolerable interference level. It is also the large file size that renders our file segmentation approach valid, the sparse or dense scenario of which in multiple cached files regime remains as a step further towards Zipf law file popularity. An interesting implication of Zipf's law is the derived number of access times given a certain object which is obtained provided the number of accesses is large enough. Hence, the intuitive query is whether our analysis assuming normal distribution for file hit probability could be modified to leverage CLT and directly compare to the normal approximation and the detailed interpretations of our comparison. Our analysis could well incorporate the confidence interval relevant to the reliability of the caching model assumed. To that end, alternative a priori knowledge could be leveraged besides past access history in a multiple file size and requests setting. Completing our remarks with this paper, 5G oriented bandwidth and storage demanding multimedia streaming can potentially redefine Zipf-like distribution enhanced with dynamic file popularity based on different measurement history. Concluding this section, Breslau et al. in [158] provide a detailed mathematical analysis focusing on the modeling efficiency of web

caching by Zipf-like distribution which is a deviation from Zipf' law. The authors in this paper also demonstrated weak relation of access frequency to the file size, a statement which could further determine the above relation via an uncertainty based approach where frequent request translates to reduced uncertainty and rare request frequency relates to greater uncertainty. Hence, the concept of independence or equivalently weak correlation can be extended to a wider sense, for instance, independent deep fading which can further elaborate on integrating diverse fading with caching hit probability and network deployment density. Relative to Zipf's law applicability, user selection strategy could pose deviations to Zipf's law modeling along with dynamic topology based randomized caching. Hence, an overall probabilistic optimization problem considering Zipf's law could boil down to optimal coefficient derivation for cached files along with optimal distribution for other parameters of the problem. The accurate remark relative to weak correlation between file change rate and access frequency is what confines possible future research to small correlation values with the extension of asymptotic results in large file volume regime. Hence, the crucial question that arises is to what extent the Zipf-like distribution based independence assumption constitutes an oversimplifying one in terms of file popularity and availability investigating the «hot» and diverse «cold» server cases. The accompanying algorithm referred to as cache replacement algorithm addresses a challenging problem from a probabilistic point of view. The comparison in terms of hit ratio performance could very well extend our probabilistic analysis in a manner that given the cached files in a certain cache could be evicted from the cache if a certain length of the cache finite size does not contain the respective file. However, this has potential implications on the designed algorithm. A related property could be that diversity of cached files implies that a pattern exists such that a cached file repeatedly appears as requested in a maximum memory file length. Hence, file request and cache sizes have a dominant role in cached files contained in a dynamically varying cache. Clearly, the latter property is also computationally demanding, an effect to be magnified in a large scale network regime. Closing our remarks for [158], the asymptotic case investigated in this work could well invoke the CLT, hence, Gaussian distribution serving as a comparison benchmark in our analysis could well be utilized asymptotically by leveraging the crucial independence statistical prerequisite followed by interpreting popularity and file requests both changing as a practical assumption.

4.2.4 Caching hit probability distributions and pico-cell network analysis

Every file stored in users' file is characterized by a certain hit probability derived by the assume statistical distribution. The current analysis considers popularity based strategy i.e. most popular files cached or proactive strategy exploiting user relationships. The simulations conducted considered a varying number of cached files along with three corresponding distributions: the uniform distribution, normal distribution and Zipf-like distribution. The uniform distribution relates to equal hit probability for all files, hence as number of files cached increases the hit probability decreases as mathematically deduced. For normal distribution instead of using Q-function, the approximation formula below is considered:

$$p(x) = \frac{2m^m}{\Gamma(m)} x^{2m-1} e^{-mx^2} \tag{47}$$

Where m is a parameter set to 0.5 for accurate approximation. The third distribution is the Zipf-like with parameter γ_r expressed by the following formula:

$$p(x) = \frac{x^{-\gamma_r}}{\sum_{\zeta=1}^{N} \zeta^{-\gamma_r}} \tag{48}$$

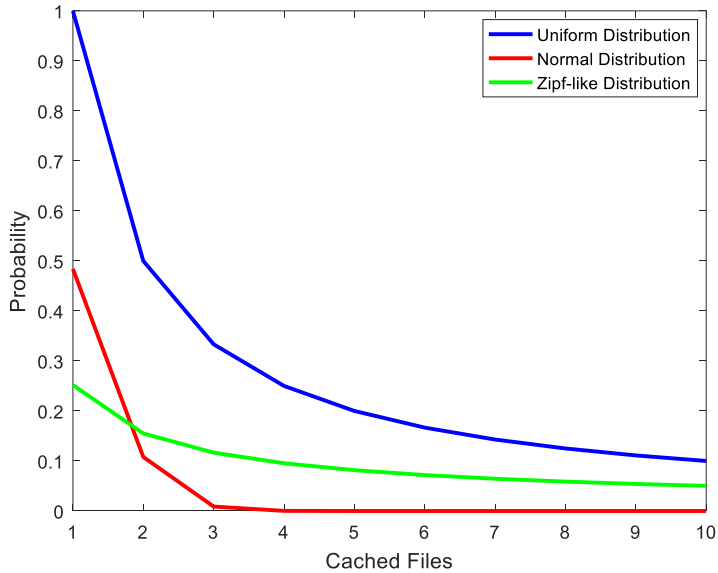

FIGURE 4. 1: Probability vs. Cached Files for Uniform, Normal and Zipf-Like Distributions

where N is the number of files and γ_r is distribution parameter reflecting file popularity. The file popularity distribution considered i.e. Zipf-like differentiates itself from the known Zipf's law which results by assuming values of γ_r=0.986 which is sufficiently close to one. As opposed to this characterization for values of 0.64 to 0.83 the referred to distribution is derived. High γ_r denote few files popularity meaning that the server is «hot» and the balance of distance of users and file popularity is achieved by caching most popular files. On the other hand, lower γ_r values means more «diversified» popularity and files with different levels of popularity are cached. As a remark relating uniform to Zipf-like distribution, the value of γ_r=0 derives uniform from Zipf-like distribution.

Concerning assumptions of simulations conducted we set number of files N=10 and parameter γ_r=0.7 as a value chosen balancing between aforementioned marginal values. The uniform distribution case is proven to be optimal for all N that integrates the concept of equal file popularity. It must also be noted that the trend of this curve does not reach a certain floor value as may be claimed at first size but gradually approaches the abscissa with an increasing rate. The Zipf-like distribution is outperformed by

uniform distribution but is superior in terms of higher caching hit probability compared to the normal distribution approximation. The mere deviation from this observation is the interval of up to two cached files where normal distribution outperforms its Zipf-like distribution counterpart. Beyond the two cached file number, Zipf-like distribution demonstrates a greater hit probability. Moreover, a performance gap is observed pairwise for all three curves. The performance gap is shown to decrease as N increases. In overall, the distribution performing the worst is the normal distribution observably. There are two related reasons for such performance: the approximation formula used and the second reason is thoroughly explained by the next comments comparing normal and Zipf-like distributions.

Zipf-like distribution assumes sorting of files by decreasing popularity, which does not hold for the normal distribution fully described by the mean and variance parameters. The variance indicates how concentrated the values are around the mean while mean represents the symmetry of the distribution. Hence, variance indicates how many files are most popular while the mean serves as an index of which the most popular file is.

Another manner of interpreting the results in Figure 4.1 above is the information theoretic entropy/uncertainty concept. Hence, the uniform distribution outperforming both Zipf-like and normal distributions is the one with the highest uncertainty a property emanating from the flat curve of this distribution. Moreover, the entropy quantification of the normal distribution is strictly a function of its variance hence location blind in terms of mean value. This also justifies our zero mean assumption. To that end, one could argue that the normal distribution may take negative values, which is contrary to probability positive values. However, this can be justified by the fact that the expression leveraged to our analysis is a Nakagami-m based approximation. Hence, the positive valued requirement is covered by the respective nature of Nakagami-m distribution. The second reason for utilizing the aforementioned approximation is that Nakagami-m distribution, as stated in previous section, represents a compromise between the Rayleigh and Rician fading related distribution, by means of proper parameter selection. This is one step further to investigate skewed distributions along with their impact on caching hit probability. Finally, the Zipf-like distribution is observably characterized by less uncertainty than uniform distribution. The conclusion is thus straightforward: optimal hit probability is achieved by large uncertainty. Also noted is the steep slope of the normal approximation case descending to zero beyond the three cached files number.

<u>Pico cell network analysis</u>

We now consider a wireless cellular link region, which is divided in an inner area of radius R_o=50m, a value between R_o=50m and R_1=100m, a value between R_1=100m and R_2=150m and a value between R_2=150m and R_3=200m. The received power is given below:

$$P_r = \frac{P_t h_k}{\|X\|_a} \tag{49}$$

Where P_t is transmission power, h_k=e, where e is Euler's number as the Rayleigh fading channel gain and ||X|| is equivalent to distance inside the cell. The minimum received power is calculated for a receiver of 100MHz bandwidth, a noise figure of 1.5dB at temperature T=290K and SNR of 10dB thus equal to -82.5dBm or equivalently 5.6pW. With loss exponent equal to a=4 and rearranging the above formula the transmission power with respect to the radius values is derived based on the following formula:

$$P_t = \frac{P_r R^a}{h_k} \tag{50}$$

The derived values are contained in the following Table 4.1:

TABLE 4.1: Pico cell radii and transmission powers

Area Zone of Pico Cell	Radius (meters)	Transmission Power
0	50	$12.8*10^{-6}$ W
1	100	$0.2*10^{-3}$ W
2	150	$1.04*10^{-3}$ W
3	200	$3.3*10^{-3}$ W

The result concerning the above calculations is that denser user deployment must be realized in cells of small radius as larger received power is available thus enabling greater energy efficiency. Furthermore, larger transmission power is needed as the radius increases as intuitively deduced. The case of smaller received power in small radius as opposed in the statement above due to channel impairments as shadowing is not accounted for in current analysis. Complying with the denser user deployment in small radius areas, the caching hit probability increases along with less energy consumed.

4.2.5 Extension to compressed sensing sparsity concept

This section deals with a probabilistic analysis relative to caching hit probability of a specific file with probability p and not cached in user memory with probability 1-p, thus a Bernoulli distribution scenario. The formula of total probability is given below:

$$P_{TOT} = p^n (1-p)^{N-n} \tag{51}$$

where it is assumed that N=10. The CS sparsity principle is represented with the case n<<N with n as cached files and N-n not cached, while the dense case is expressed by n≈N. For the former case $p^n \rightarrow 1$ thus $P_{TOT} \rightarrow (1-p)^N$ while for the latter case $(1-p)^{N-n} \rightarrow 1$ and $P_{TOT} \rightarrow p^N$. For the value n=1 the Figure 4.2 is given below:

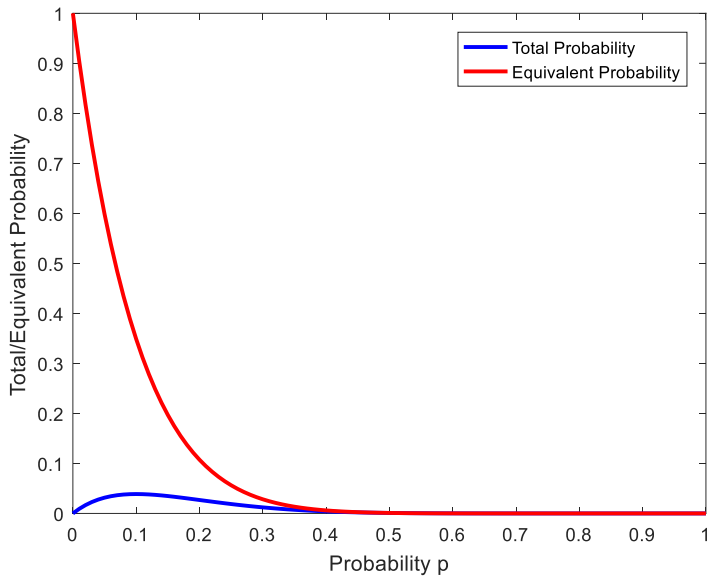

FIGURE 4.2: Total and equivalent total Probability for n=1

In order to better realize convergence of the total and equivalent probability cases, we introduce file segmentation strategy, i.e. decimal values for number of cached files. The next Figure 4.3 given below assumes n=0.2:

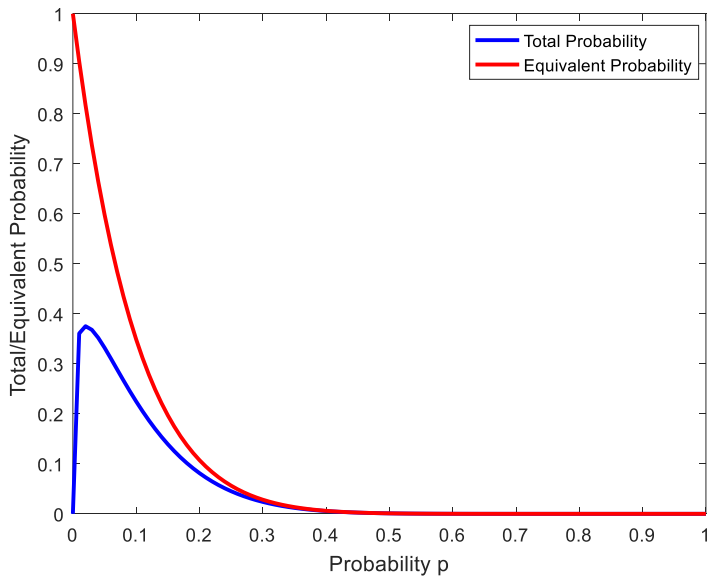

FIGURE 4.3: Total and equivalent total Probability for n=0.2

As apparent from the figure the convergence for decimal value representing segmented files is greater and «sparse» case is analyzed completely.

The final concept is the convergence observed in the dense case as n approaches N i.e. N-n→0. Figure 4.4 for integer value n=9 is provided below:

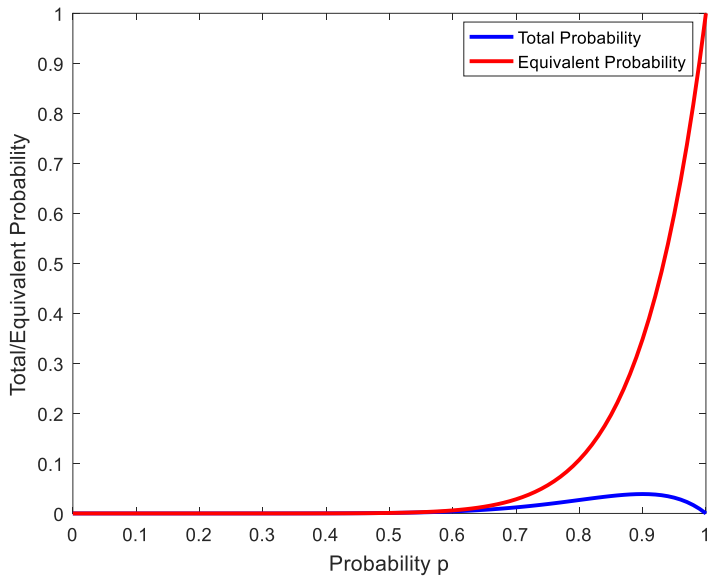

FIGURE 4.4: Total and equivalent total Probability for n=9

The figure verifies the convergence for the dense case assumed above. Following the same context i.e. decimal values representing cached file segments the value n=9.8 is considered in Figure 4.5:

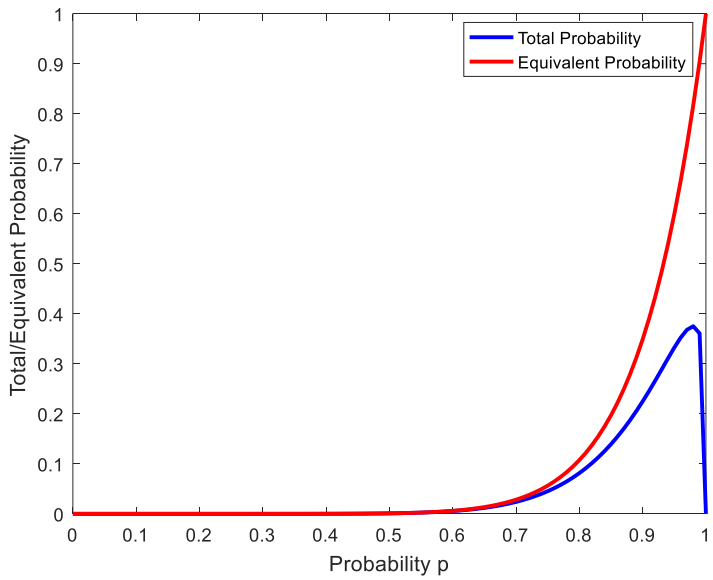

FIGURE 4.5: Total and equivalent total Probability for n=9.8

A final remark is that CS approximately approaches the optimal solution in the sparse and dense case as well. File segmentation is proven to compensate for the sub-optimality of integer values. Thus considering the above as an optimization problem together with the fact that number of segmented parts, in which a file may be divided is dependent on the size of the file and the specific application, the validity of optimal result along with theoretic CS derivations can provide either a solution or a constraint in the formulation of the problem of using the concept of sparsity or compressibility together with caching files according to popularity by means of Zipf-like distribution.

The analysis conducted in this section relative to other approximating distributions as normal uniform can contribute to different total probability expressions and provide insights on caching files in a cellular network framework. As a clarifying remark, the probabilistic analyses in previous section considering the uniform, Zipf-like and normal distribution can be effectively combined with both CS related sparse and dense segmented files left for further investigation. Moreover, the last section can be accordingly modified to quantify entropy in a higher dimensionality regime which also, in terms of the latter segmentation, could fit into the CLT case as segmented file parts grows in number. The balance of Zipf-like distribution coefficient values reflecting file request and cache sizes, could provide an interesting approach integrated with diverse fading conditions and additive noise apart from interference levels describing the channel.

4.2.6 Conclusions and future work

This section investigates a cache-enabled cellular network with caching hit probability considered for probability of certain file stored in users' cache. The need for caching scheme is highlighted while simulations account for three distinct distributions: uniform, normal and Zipf-like distributions. The results of hit probability are technically interpreted by comparing hit probabilities for each case. Evaluating transmission power and the relation to the strategy of dense user deployment in a cell of small radius, hit probability is quantified as greater in dense cellular network areas. Furthermore, assuming Bernoulli distribution for the case of file cached with probability p and consequently not cached with probability 1-p the convergence of exact total probability and equivalent probability curves is verified. The latter more pronounced effect is evident in the case of adopting the segmented file assumption i.e. decimal values instead of integers. The results can be translated by CS sparse and dense cases reaching a sub-optimal result.

Directions of future research include cellular networks of larger radius together with applying coded caching and user mobility in the respective cell. Caching information capacity combined with optimal data exchange based on performance metrics is another promising direction. Cooperative caching is also applicable with the aim of designing robust efficient algorithms and reduction of delay along with adequate QoS at the expense of increased computational complexity. User mobility and resulting patterns can also be leveraged.

Finally, CS principles and exploitation of time variant distributions is also a future issue.

The domains of time, frequency, code or space provide potential extension of the derived results translated in the context of benefits or drawbacks for each of the above domains along with joint dimensionality approach in a cross-layer optimization constrained equivalent problem.

4.3 Compressed Sensing Integration with Cloud assisted Social Networking

4.3.1 Introduction

The results of this section are related to publication [C1]-([159]), which was co-authored with Stergiou et al. Current wireless communication systems face the critical challenge of exchanging and processing of large datasets that are characterized by excessive volume, heterogeneity as well as different speed relative to their generation, processing and access. Towards an accurate remark, Big Data refers to the explosive growth of data in our era of information technologies that affects the majority of technologies and scientific areas. Hence, big data analytics provide the very means for analysis, processing and information extraction from these big data quantities. Information extraction addresses this challenging problem by exploiting pattern or structure within the data in a resource efficient manner. The realistic attribute of Big Data is unstructured and semi-structured property as opposed to the more convenient structured property, which is exactly the driving motivation for advanced analytics tools for processing and analysis of these datasets. The latter may contribute to processing and analysis simplification but nevertheless requires careful modeling so as to minimize complexity and processing overhead and achieve analysis in an optimal, robust, secure and flexible framework. It must also be stressed that, as is imperative in many Big Data analytics scenarios, the problem is formulated as an optimization problem, whereas optimization is carried out in multiple dimensionalities such as time space e.t.c. as in communication performance optimization problems. In such optimization problems a crucial issue involves the insertion of a pre-analysis phase so as to assess certain attributes of the data set that will greatly aid processing efficiency. Hence, although this strategy requires a portion of the total time available for data analysis, its inclusion may play a dominant role in the resulting information extraction and induced complexity. Moreover, problem transformation and learning of data attributes and structure are major stages prior to analysis. The former is quite intuitive as it may provide the means of modifying raw data to a form that eases processing and better understanding of dataset results at the expense of a transformation induced shortcoming. Clearly, a dataset of large volume may possess a certain structure that is difficult to realize due to large dimensionality implying computational complexity. Learning as a phase of data processing may be resource consuming but may greatly aid the overall process by more accurately predicting future outcomes and optimization trends. In light of CS principle, dictating prior discarding of data for efficient and low complexity estimation and reconstruction, data analytics also apply this rule which is exactly what abides to information extraction from large volume data and also improve failure robustness. Probabilistic modeling constitutes a part of an overall data analysis phase referred to a data model building. This is the next most important phase following data structure understanding as it includes testing and validation in a practical case but also provides useful insights in an asymptotic regime as a feasible consideration of the latter phase. An example with respect to modeling is the testing of whether a hypothesis is verified in a qualitative or quantitative and more complex setting. At this point, it is necessary to highlight the merit of statistical modeling. The stages of model building and evaluation assist predictability, robustness, interaction and dependences among independent variables of interest along with quantifying the extent to which the model

provides further performance of data processing and exchange. Data volume also relates to obtaining sufficient adequate measurements and data samples to claim obtaining a comprehensive model as to approximating an asymptotic model predefined. Based on dataset attributes, clustering is a method of grouping data judging from properties they possess, a technique that may pose a computationally expensive combinatorial optimization problem, which calls upon complexity reducing structure commented above. The clustering based attributes can be equivalently modeled as constraints to the optimization problem, the inclusion of which could alter the results achieved. Thus, the CS optimization principles applicability becomes increasingly apparent not only on the large dataset volume but also due to the practical and convenient property of possessing structure in diverse Big Data quantities such as cloud computing, wireless communications, healthcare and smart devices environments. Specifically, another issue of wrongly relying on certain observations stems from the relationship of those either as properties of the dataset or a result from a specific recursion of the algorithm employed. Based on probability, the above results could be found to be characterized by independence or correlation concepts.

To further elaborate on mathematical tools leveraged for extracting knowledge from large datasets, regression is an inherent method when it comes to probabilistic modeling thus identifying input variables with a dominant role on the algorithm employed along with sensitivity analysis with respect to certain parameters. To that end, linear regression as deduced from its definition define a linear input-output relationship. Although applicability of this method is a result of sub optimal relaxation of certain problems that are highly intractable, linear modeling is very well-posed from a probabilistic point of view. Hence, expected values, uncertainty measure and abrupt variations as well as computed error components can all be traced by such methods. Additionally, CS theory promises a logarithmic relation to many problems and can thus be effectively integrated with linear regression schemes. Logistic regression is another form of addressing large datasets and serves as the basis of classification methods. These methods label each observation by a specific label in a predefined dataset and formulates an effective method of dynamically categorizing data. Categorization may be decided based on diverse attributes of the data and utilizes greedy algorithms that iteratively determine categories of available data in the same algorithmic manner employed in the CS counterpart of convex relaxation methods. Correlation may be the criterion applied to lead the algorithm to deriving a specific optimal solution. Regarding time series tools, the problem is solely dependent on data sample history with the memory order defined in an application-specific manner aiding future forecasting by the already commented structure exploitation. As a finalizing notion on big data analytics, we mention in-database analytics that rid of the necessity to transfer data to analytic tool for processing. Moreover, accurate data analysis may often require a transformation prior or posterior to data processing as to effectively translate in the needs and constraints each problem contains.

The Big Data deluge including all properties such as volume, heterogeneity and velocity have spurred the generation of many information technologies related computing methodologies, the most widespread known under the name of Cloud Computing.

Hence, data analytics briefly mentioned above are inseparably related to Cloud computing.

Cloud computing [160] refers to a local network of servers for storage, easy access and processing of data. In current communication systems, the necessity of cloud computing stems from the multi-technology integration along with the shaping of challenges and limitations that motivate the design and evolution of future wireless networks. Cloud computing provides the means for building and managing networks in a secure privacy preserving framework. Moreover, mobile cloud computing is a step further in that the transition from data center to mobile users along with inheriting the advantages and limitations that mobile technology and distributed access induce.

This section proposes a Cloud computing assisted system for Social networking from a security perspective in the Big Data regime for efficient information exchange. As an extension of this work verifying its merit, a novel database is provided for assessing user interactions statistics with Social networking. Furthermore, CS principle applicability is attempted to be combined with the proposition of this section to shed light on proposed secure framework.

Social networking is a social web-based structure that enables multiuser interactions in a way though that compromises security and data privacy of sensitive data to be accessed by unauthorized third parties. Hence, a secure environment as a consequence of CS based optimization, cryptography and encryption along with transmission and communication cost mitigation can be realized. Without a doubt, the feature of Social networking enabling coherent pattern definition from localized interventions. This renders pattern finding feasible in a decentralized setting promoting autonomous operation with optimal performance as the computations required at a base station to achieve desirable performance is relaxed by this convenient property mentioned above. This fits into structure exploitation at the core of CS optimization. It could also be thought as redundancy resulting from minimum interactions. Hence, the less the interactions required, the less computational burden falls upon the social network to improve performance in a low complexity manner. Another bridging point between cloud computing and CS is the fact that in a cloud environment time and space are transparent dimensionalities. This could either serve as a means to reach performance penalties imposed by CS in time and space but also promote the non-adaptive nature of optimizing algorithms. As commented above, is all about the aforementioned data availability by using mobile devices. Hence, the concept of random movement as a means to compensate for deterministic fixed routes limitations e.g. covering network holes. This enables redefining resource allocation for Big Data handling in SN of low cost infrastructure. However, this raises security issues due to the very nature of Big Data and requires resilient data analytics to combat security vulnerabilities along with the requirement for efficient management of finite memory storage and processing power.

4.3.2 CS method for Social Networking security issues

Relative to SNs, users have the ability to dispense and send information to any set of users they know and thus develop a friendly web-based community. The effects of Big

Data and sharing of user related private information have resulted in security issues compromising sensitive data. Thus security breaches may be interpreted as unauthorized access to a site content or private information access which are generally not distinct since realization of one implies the other.

Given SNs, the Big Data processing and sharing clearly verifies not just large volume but all three attributes defining the Big Data deluge: volume, velocity, variety i.e. diversity. As the transition to petabytes or higher processed has already become a reality, CS structure exploitation has arisen to be more imperative than ever such that integrated with statistical analysis leads to resource efficient information extraction from these datasets but may also be leveraged to enhance resilience to security breaches. Regarding Big Data attributes impact on Big Data characteristics, only a comment will suffice: Big Data volume is the most related attribute to the characteristics such as data hiding, user identification and public data. Moreover, public data category is more influenced by diverse analysis tools and services. The processing, storage and transfer of data are quantified by the so called encryption rate, the value of which introduces an additional limitation for real time processing of social networking data.

4.3.3 CS applicability on proposed system

The main framework to support system proposition is essentially based on the beneficial integration of encryption algorithms that separately reach high encryption rates and together prove to be accurate and efficient. Similarly, encryption rate increases as data transmission evolves which indicates the necessity of implementing the encryption process prior to transmission and during transmission stage. This is the point of convergence with respect to CS based encryption by reducing transmission time intervals and applies sparsity rule. Computational security and encryption efficiency in a scenario of adversary presence in an overwhelming probabilistic context as a consequence of random measurement matrix that CS optimization leverages.

The system proposition relies on the use of cloud computing environment that renders the authentication enabled by providers to be transparent to network scalability i.e. it will be available to all users in a SN. This property further promotes CS non-adaptivity in the sense that user number variation does not impact security robustness provided by integrating cloud environment. Hence, the algorithmic integration with the additional assumption of multimedia content transmission leverages segmentation to smaller packets for minimizing packet loss. Another trend observed is the absence of stuck overflow via the latter integration thus ensuring smooth data transmission. To this extent, CS could further provide security and take advantage of this beneficial integration by leveraging sparsity on the effective interactions of each user with the Social network. Another issue of concern is whether the non-negligible impact on the number of users possesses a certain limit hence if IoT scenario influences CS encryption and resilience to attacks.

4.3.4 Past related work

In this subsection, we proceed to a brief review of related work. We only highlight major principles related to security issues along with challenges for future considerations.

In the first paper by Gao et al. [161], security issues in SNs are derived and their interrelations and defense mechanisms are analyzed. We will not delve into further detailed categorization of security threat types but instead confine ourselves to security breaches classification. The issue of security breach from a service provider raises the question on how to take advantage of centralized architecture by simultaneously mitigating security compromise. Thus, encryption along with user selection can protect sensitive data and even benefit from a localized information sharing, a strategy which can further be aided by CS based encryption. Thus, instead of pursuing a decentralized scenario the concept of provider induced security breach implies a central coordinator scenario for enhancing data integrity. The complementary case is user induced security breach, where providers prevent unauthorized access. In such cases, localized interaction may prove to be prone to security breach. Hence, impersonating a friendly user, data may be exposed to various attackers. Concerning security breach from third parties, the key mechanism for these breaches is the lack of monitoring the functionalities employed by third parties for information accessing that would be preventive with respect to security breaches. Hence, through a dynamic request for information accessing, the specification of strict set of data that the user allows access for constitutes a main defense mechanism. Moreover, as CS sparsification is all about exploiting structure as a balance to randomness based mathematical formalism, structure can prove to be the cause for a security breach if for instance a link pattern is observed by third parties with reference to user interaction inside the social network, which may be simple to derive requiring small overhead and being applied to similar networks in terms of topology. This could be effectively confronted by CS encryption. From an attack launching perspective, the attacker may require moderate or extensive knowledge given the network, which also reflect the threat level posed to compromised users and the complexity of security enhancing schemes. Relative to CS assisted security ciphering framework in the distributed setting of cloud edge environment, [162] introduces data security double layer scheme in a resource efficient manner tailored to such scenarios such as densely deployed IoT networks. The familiar concept of comparing measurement matrix of encoded and reconstructed signal for verifying whether data integrity is compromised is also commented thus assigning the computational burden on the decoder. Clearly, the dimensionality involved is a measure of the feasibility of a successful security breach. Another aspect of CS based security is the key updating method which relates to CS incoherence principle as a dimensionality reduction scheme. The latter scheme combined with revisiting computational asymmetry in an edge cloud environment constitutes a promising research direction. Regarding sorting out noise sources in a security preserving communication scenario, we only stress the integration of channel impairment induced performance degradation and security enhancing schemes. An additional strategy regards the determination of location of data that have been exposed to attackers. Hence, an attempt to protect crucial data part, the attacker could be directed to access

the part that do not constitute the important locations rather than locations that do not compromise security and do not degrade decoder size reconstruction. Another very interesting issue concerns correlation between sparse measurements and signal to be reconstructed in the sense that exposure of the former may lead to determination of the latter. It is this matter where CS dictating incoherence between sparsifying domain and measurement matrix could aid in reducing correlation while still ensuring optimal reconstruction in a reduced complexity context. Either by considering signal reconstruction failure or inserting false data in the sequence, CS based encryption can provide enhanced robustness to attacks. An interesting extension is that besides robustness, additional information about the attacker can be obtained as a complementary measure of sensitive data protection. Proceeding further with a sensor network and the challenge of distributed information exchange, security is verified along with low complexity achieved by user-level encryption in a cloud assisted scenario. An issue of data integrity relates to the random mapping of original data and its inverse mapping for data reconstruction is all about CS random matrix theory and optimal reconstruction with overwhelming probability context. Moreover, sensitive data being a small fraction of transmitted ones could also contribute to security enhancement. On the basis of CS based methodology of solving a sparsity based problem, the extension of l_1 minimization commented to preserve security leaves a promising direction of future research. Furthermore, the issue of data retrieval conducted at the cloud must also be investigated along with feasibility of security compromise at the cloud-user communication phase. Also, a security breach in a SN could be discouraged if an excessive number of permutations are needed based on raw data transformation i.e. attacker faces a large complexity.

4.3.5 Conclusions and future work

SN is a networking paradigm that as many other entities is dependent on exchange and processing of large sets of data. As Big Data has already emigrated to currently deployed communication systems, the aspects of security and privacy preservation of users information constitute an inherent challenge and necessity in the SN context. To that end, CS provides the encryption capability when integrated to the specific scenario at hand. CS related optimal transformation matrices in a random manner are the main target of research in this area. In a high scalability network, technologies such as Cloud Computing for easy acquiring of frequently accessed data also impacts data security.

The combination of CS lightweight encryption prior and during transmission stage and multiuser interactions monitoring in web-based communities can address the data privacy issue. Hence, Big Data analytics combined with CS for processing and information extraction either in terms of the information content or the data security parameters pose a significant research challenge.

4.4 CS Integration with Medicine Practice

The evolution of Big Data has already resulted in the utilization of different technologies that are assigned to address the issue of communication, storage and processing of this large volume and heterogeneous amount of data which have reached the order of zettabytes and continue to grow. This motivation has inevitably emigrated

to healthcare data management which have formed various optimization problems similar to the general context but with application specific additional constraints stemming from the requirements posed from medical science. Among the straightforward constraints are low latency in a real time measurement gathering and processing environment and data reconstruction accuracy so as to timely alert involved clinicians for a certain situation calling upon medical treatment and to accuracy to assess disease presence and severity.

4.4.1 Introduction

The results of this section are related to publication [C4]-([163]). The science of medicine relies just as in the wireless communication optimization methodology relies upon certain problems that by definition must be well-posed in order to be solvable by existing methods. However, CS theory has demonstrated that ill-posed problems can be optimally solved exploiting structure as the problem constraint. This property has emigrated to practice of medical science.

Medicine deals with diagnosing and effectively curing diseases assisted by numerous information technologies that promote healthcare data management and efficient processing for useful information extraction. Hence, a priori knowledge is incorporated to the clinical procedure in order to conduct early diagnosis and prevention of health deteriorating. There is also a posteriori knowledge that follows diagnosis and relates to curing diseases. It is thus understandable that accurate and timely diagnosis is imperative that also adopts to technologies for fast data communication from the patient to clinician in order to effectively treat the respective disease.

4.4.2 Biomedical Big Data and Information technologies

Efficient information gathering, storage, processing and communication are required to refine disease understanding and classification. Hence, the Big Data properties are all satisfied when it comes to healthcare data. Similar to heterogeneous data types such as text, images and video, healthcare data regime implies different diseases, different disease severities and different patients classified via diverse criteria.

Computers and information technologies have introduced crucial capabilities to address this Big Data deluge and the shift to resource expensive real time applications. Hence, in a clinical environment processing for clinical interpretation of data has been enabled through the use of entities such as wireless networks and cloud environment for fast on demand access of patient data without compromising their confidentiality. Centralized or decentralized approaches have been deployed with compensating for scheduling, low latency and energy efficiency of the network functionality. Technological tools such as Artificial Intelligence or the massively deployed IoT paradigm involving numerous devices connected around the globe intelligently handling biomedical data and processing them in a timely manner. As integration of technologies expands, clinical databases share information employing encryption to preserve privacy and avoid information leakage to third parties.

4.4.3 Past Related Work

We review two highly relevant works focusing on IoT-Healthcare CS application along with security oriented CS based considerations.

The first paper reviewed by Siddique [164] et al. investigates the effect of redundancy of data in healthcare wearable devices in order to assume a tolerable error thus enabling energy savings as a consequence of this relaxation. Hence, healthcare data consideration in this paper fits into the WSN energy consumption bottleneck extensively analyzed in section 3.6. Cloud assisted environment is also valid in this context along with data compression techniques. The most important issue pertained to CS application is the quantification of required measurements to achieve optimal diagnosis related results. The fundamental inherent information loss caused by compression algorithms is also a concern in healthcare data derived in this paper by jointly considering diagnosis accuracy. The additive Gaussian margin i.e. error tolerance is also concisely derived in this section which straightforward implies denoising algorithm tailored to the above tolerance. As a finalizing remark concerning this paper, the latency and energy efficiency tradeoff, as mentioned in this subsection, must be further elaborated as it is a major requirement for IoT biomedical wearable technology, in order to assist in timely treatment of a certain reported disorder.

In the past related work in [165], Yuan et al. base their work on CS fast and secure acquisition, indexing and processing of healthcare multimedia data amounts as a consequence of their exponential increase in a cloud-assisted framework. Sampling, compression and recovery are analyzed and secure encrypted design along with memory efficiency are achieved. A well-posed statement in this work is that large multimedia data volume results in increased security breach potential in practical network scenarios. Contrary to the previous paper review, low latency is commented to be related with data retrieval quality. Security of compressed samples is enabled by prior to cloud transmission while reduced volume does not come at the expense of value for the clinician accessing those healthcare data. However, this preservation of content is what carries the cost of a privacy attack in cloud server. Bandwidth issues are also emerging in real-time healthcare wearable devices as rapid informing of the clinician is a prerequisite in such scenarios. This issue becomes more imperative in a large global database with dynamic information sharing. Concluding with this paper, the accuracy latency tradeoff is also very well investigated and stated to depend on the specific healthcare data application. Hence, accuracy may be required from a clinician to further ascertain interpretation and accurate diagnosis. However, in the case of alerting the clinician of an abnormal reading sensed by the wearable device requires very low latency which comes at the expense of accuracy loss. Finally, CS based optimization by employing randomness may encumber the use of security keys for future for data sample grouping. Thus, the latter poses a challenging future investigation.

4.4.4 Compressed Sensing assisted Healthcare Data Management

Medicine practice requires excessive amount of data to interpret and evaluate in order to perform an accurate diagnosis and decide in favor of a disease based on clinical symptoms. Thus, this appears to be in contrast to CS practice that involves selective

consideration of data and most importantly data compression to offer low complexity and assist in time efficient treatment and tolerable information loss thus not affecting diagnosis accuracy. It is therefore clear that CS based healthcare data processing must provide a careful selection of the related degrees of freedom in a problem so as to ensure tractability and solution optimality.

In order for the above to be achieved, a parallelization consists of clinical diagnosis combined with laboratory test results that must be effectively integrated based on the clinician's expertise. Common symptoms that relate to diverse diseases must be carefully examined. At this point, CS can be leveraged as to aid to determining correct diseases identification based on constraints that appear together with healthcare data structure and thus filtering the range of practical results. One possible claim in healthcare data processing could suggest that structure or pattern emerging in data could require excessive amounts of clinical examinations so as to reach a safe diagnosis. However, the potential for low cost and low complexity data processing has already been exhibited in the healthcare data deluge and thus resource efficient information extraction is indeed feasible. Moreover, it must be stressed that all clinical examinations and diagnoses inevitably involve a certain amount of uncertainty thus implying an information theoretic approach that can be adopted. The latter also validates an information theoretic analysis either for statistical inference or providing healthcare data oriented performance benchmarks. No detailed notion for the above will be further provided except for the entropy wise interpretation that frequent symptoms arising with respect to a class of diseases convey small amount of uncertainty while rare symptoms are characterized by more information content. Hence, the greater the uncertainty the more resources needed for clinical evaluation.

It is also a fact that medicine like other sciences, utilize statistical methods for conducting diagnoses and aid the decision making just as in other information technology disciplines. CS leveraging advance statistical methods can boost data management efficiency and extend present practices in medicine practice. Another concept is that of an insertion of perturbation in the test results increasing uncertainty. The broader concept of such an additive quantity referred to as additive noise in wireless communication literature constitutes an error inherent in the clinical results conducted by a medical device. Hence, the model of inserting noise also complies to medical devices and instruments from a mathematical point of view as well as a quantitative perspective characterized by randomness necessitating statistical methods. This is nothing else but applying random optimization CS based techniques. The resulting imprecision of results must be effectively removed always attempting to preserve computational complexity tolerably low. Towards an accurate relation, randomly injected perturbation is required to evaluate so as to minimize the potential effect of leading the clinician to incorrect diagnosis as a result of induced deviation in the healthcare data processing and significant information context extraction. Concluding, given the globally interconnected healthcare databases, cloud computing enabling on demand information access always considering the need for privacy preservation along with web content caching [165] for fast access of frequently requested data are both promising technologies that enable a whole new perspective for promoting medical science ability to handle severe diseases in both a diagnosis and

treatment aspect. CR is also a key technology to enable accommodating excessive users in a network, both licensed and opportunistic, while not hindering communication quality and information exchange.

4.4.5 Legal and Moral Issues Arising in Information Technologies Era

As already derived in previous subsection, the global biomedical data or medical images are being shared and processed collaboratively, thus raising security and confidentiality concerns. As stated above, CS encryption could provide a lightweight and efficient data protection technique. Nevertheless, information leakage is still a concern and CS could potentially improve data integrity by its corresponding sparsity concept. This could be applied to integrating compression and encryption while selectively protecting fractions of the data sequences transmitted. At a digital identity verification scenario, the attacker could be discouraged in his attempt to decrypt and access data if he encountered a varying excessive complexity i.e. required a combinatorial exhaustive search, which in addition could provide adequate time for the legitimate users to identify the attempt and prevent data from being revealed to the attacker. Relative to social networking security breach scenario commented in the previous subsection, healthcare data also constitutes similar context for all threat types mentioned in the former.

Concerning healthcare data management, there are critical moral issues to be addressed caused by the very widespread exchange and sharing of data. The issue of whether access of private medical data to a non-clinician set of users raises a most important issue given that medical data confidentiality must be preserved. Moreover, the application of a medical treatment raises the issue of necessity versus obligation. Another issue reflecting the latter is the imperative requirement that medical treatment should be accessible to all patients regardless of social distinctions and beliefs. In addition to the latter, each patient must have the freedom and ability to choose and benefit from a certain treatment in the sense of best possible clinical result given a certain disease. It is thus imperative that CS low complexity integration by means of information technologies should be oriented to society welfare and collective benefit given the practice of medical science.

4.4.6 Conclusions and future work

It is clear the CS principle could emigrate to healthcare data management efficiency promising computational burden alleviation and validate accurate and timely diagnosis capability for clinicians by sharing data through a global database infrastructure. Combination with cloud computing technology and the massive deployment of the IoT paradigm, data assessment and effective processing introduces revolutionary healthcare data handling. Together with advances in electronics and thus medical devices capabilities along with progress in medical science in terms of diseases understanding and treatment generation, information technologies will lead to a new data science era that will motivate progress in these technologies even further.

4.5 CS in Smart Classroom Architecture

Smart education has emerged as one of the most essential part of smart deployment scenarios that has emigrated from conventional methods to complex deployed network assisted infrastructure as a means to provide students, educators and administrators with powerful tools to promote interactive acquisition of knowledge and better understanding of the course materials and subjects that the educator attempts to the students irrespective of individual properties of the students that could potentially handicap the education process and the main concept of each delivered course.

A diverse number of scientific areas have significantly contributed in improving smart education processes such as communication technologies, sensor networks, cloud computing and advanced information processing. These technologies have opened new frontiers in smart education via massively deployed sensor networks and enabled a much better educational level since educator can continuously evaluate whether each student has thoroughly understood the course being taught but most importantly the emotional and intellectual state of the student which can serve as a valuable feedback to the teacher for explaining the concepts taught in a manner that best suits each individual student.

It is also understandable that Big Data deluge ideally fits into the diversified data involved in daily teaching while social networking is also a paradigm technology that indicates the applicability of large databases and information sharing for smart education in a global scale. Hence, CS low complexity useful information extraction could not only simplify the education process in the sense of educational quality but also greatly aid the educator in assessing specific parameters for each student ruling out many parameters and keeping the desired ones to be investigated in an advanced pedagogical practice environment.

4.5.1 Introduction

The results in this section are related to publication [C3]-([166]), which was co-authored with Stergiou et al. LMS already provide tools for an interactive teaching and testing methodology but has been confined to a two dimensional environment. Forum, Wiki, games and aggregated games are part of the education process in an asynchronous functionality context while real-time and mobile audio visual course conducting services constitute the synchronous functionality. Thus, it is the hybrid asynchronous and synchronous model integration that is the core contribution concept of this section along with state-of-the-art sensor and haptic equipment implementation for sensing and touch information in an interconnected interactive smart classroom. IoT, Big Data and cloud services thus interact in the above context and jointly progress in a rapid manner applicable to many information sharing network functionalities.

As already stated, Cloud computing enable remote use of hardware and software. Integrated with IoT and even further to Internet-of-Everything in 6G and beyond paves the way for a highly interactive classroom architecture with sharing of diverse entities that forms a new generation of smart education methodologies. The latter entities are both of large volume and variety thus complying to Big Data characterization in both structured and unstructured form that enable CS optimization method application.

Regarding communication technologies, sensor networks have already been analyzed in Chapter 3 and thus further notion will not be given with the exception of the comment that this section is oriented to indoor applications for interactive streaming hence reliable data gathering is imperative while signal attenuation is a more relaxed requirement contrary to outdoor long range applications. At this point, 5G networks becoming a reality along with low power sensor networking are to be fully exploited in the smart classroom architecture. Moreover, virtual reality interrelated with haptic sense can redefine real-time experiments conducting and distance learning combined with thorough understanding from the student side in courses such as Physics or Chemistry where experimentation is a vital part of theory interpreting governing laws and their physical context. Hence, complex problems will be significantly simplified by taking advantage of CS sparsity rule enabling clarified interpretation even when parameters are filtered and confined to a certain subset as a consequence of structure involved in the education process of teaching physical sciences. Moreover, the CS principle can be considered as an advantageous set of tools for the very performance of wireless networks and cloud computing tools that support the educational process. Long-term benefits and effective training of young scientists will therefore be achieved.

4.5.2 Past related work

Relative to paper proposition, the concept of dividing a complex problem [167] into simpler ones is also adept to CS principle where the effect of clustering applied in the educational process simplifies solving method approach and lowers complexity of the overall computational load. Hence, this approach not only applies to smart classroom operability but also constitutes a remarkable method of approaching courses that involve mathematical thinking and logical deductions. The collaboration of students in a haptic sense enabled course not only further encourages high performance from each student, by awarding outstanding performance but also validates a thorough understanding of the concepts taught. Aligned with clustering of students and courses being taught in the smart classroom along with the three dimensional enhancement, Valsamidis et al. [168] utilizes these methods to boost educational process efficiency. The reported shortcoming by the authors relative to statistics useful only for platform administration, combination with CS could render statistical tools most capable of assessing large educational data efficient processing in real-time teaching environment. To that end, data mining is a mechanism for exploiting patterns identical to CS principle. A bridging remark of this paper stating data visualization to our proposition is the projection of the real world to the virtual world where the issue of dimensionality arises. Hence, given the problem of integrated technologies the latter must be accompanied by data visualization methods for ensuring educational process quality and efficient student interaction. A final comment regards the statistical interpretation of the virtual courses which can be combined with an information theoretic perspective given the methods that are applicable to solving the majority of problems of low uncertainty while the methods applicable to a small portion of the course problems are characterized by greater uncertainty. This consideration could promote understanding of the issues unsolved that each problem in the course addresses. Moreover, uncertainty quantification of clustered student mental states could more closely characterize the effect of each smart education process has on the participating students. Gounopoulos

et al. [169] proposes an evaluation framework for web-based courses in the transparent to time and geographical location Learning Management system. As it is the student profile knowledge that best shapes the needs and characteristics of the corresponding educators given the students participating to the virtual course, the questionnaire aims at acquiring this knowledge. A measure of richness of the course is the ratio quantifying whether the set of visited web pages is diverse or only consists of a small subset of web pages. Relative to our proposition, the number of students attending a specific course and the number of web pages visited constitute correlated parameters which could promote the merit of our integrated technology based classroom architecture, all the above to a globally interconnected context. As a concluding remark, the above parameters could quantify the complexity required for CS based useful information extraction. The last paper reviewed [170] introduces the concept of blended learning which elaborates on the transition from conventional teaching to e-learning environment. The latter results from integrating digital educational content with online support services. It must also be highlighted that it is the hybridization of conventional learning and e-learning that ensures the advantages of each process are included in the smart classroom as in our derived proposition. Finally, statistical evaluation integrated with CS as already stated above could provide a clear picture of the extent to which educational content is adequately absorbed by the students provided that involved virtual teaching parameter dependencies are not only known to the educator but also fully analyzed in a manner that will allow information technologies to adapt their functionalities for virtual interaction and efficient educational content delivery to the participating students.

4.5.3 CS and the High level architecture proposition

The proposition of this section is termed Interconnected Interactive Classroom. Haptic assisted user sense transfer and distant interaction. The Haptic sense interactive interface workstations along with sensor actuator devices serving as human machine interfaces comprises the integrated architecture. The proposition however also requires real-time protocols at a separate wise application level or at the LMS data transfer or sensor interoperability system. In detail, the proposed architecture consists of a cloud computing server where user and sensors data streams are received stored in a device set via local area networks. This is the point where CS encryption should be optimized as to whether content is known at the cloud along with data compression in a secure manner so as not to compromise data privacy irrespective of whether the compressed samples of reduced size can be exploited in the same way as raw data. Hence, privacy preservation will ensure that content will not be subject to unauthorized access. Proceeding further with proposed implementation, haptic set of devices with application and data transfer protocols together with virtual and augmented reality headsets will allow human machine interoperability with the created virtual world. To this end, permitting feedback will not only greatly improve human machine interface operation and the overall smart interconnected classroom but also effectively utilize memory of the whole system and also energy efficient prediction the latter coordinated by scheduling the feedback functionality in terms of parameter estimation quality. Moreover, the smart classroom instructor will have access to virtual context of the course thus enabling real-time evaluation of sensor data. Additionally, a haptic

equipment control station will provide user interaction monitoring and data usage, which will coordinate the smooth operation of the virtual world environment and the supporting local network. This proposed architecture will allow the transition from a conventional two dimensional LMS to a system of three dimensions with extended capabilities such as virtual exercises, virtual self-evaluation, virtual questionnaires and forum wiki. The capabilities of the student will also be enriched and consisting of dynamic document download through tablet or computer, a realistic three dimensional sense and improved interface with diverse selections available given the educational material. Participating in a modern virtual educational process, advance supervisory tools will be available along with augmented reality enhanced learning procedure. Finally, the property of choosing physical presence in this interconnected classroom or wireless connection enables to expand the impact that the benefits of this classroom demonstrates in an international level.

4.5.4 Virtual Services integration with Smart Interactive Classroom proposition

The categorization of services that are contained in the proposed smart interactive classroom are briefly analyzed in this subsection with possible CS applicability suggestions.

The virtual classroom service is the first component that enables three dimensional immersion through artificial imaging. This is a crucial point where CS can provide the advantage of low complexity in image processing and visualization while not degrading the merit of the educational process. Appropriate three dimensional modeling and haptic devices enable this student interface interaction with the proper protocol support for text or messaging and streaming services. The second service is not confined to augmented reality related service but also includes bio-readings and bioactivity monitoring, the latter as logically enabling valuable feedback to the educator for taking the proper actions to optimize student perception as well as mental and psychological state. It is also a fact that energy efficiency is still a bottleneck despite the state-of-the-art use of sensors which could be addressed by energy harvesting techniques. CS pattern exploitation and artificial intelligence algorithms can be part of a cognitive service for the proposed classroom architecture with two fold benefits real time evaluation of student response to virtual world and the interactive course material possibly including experiments and tests and feedback to the educator. The third service component reflects the real indoor position of a student contrary to his virtual class position in terms of determining the former for the projection to the latter. The fourth component relates to the touch interaction haptic service utilizing technologically advanced gloves with pressure sensors and infrared transceivers. Apart from providing low latency, streaming quality and feedback control, their use can be realized to visually impaired students. Moving on to the fifth component, three dimensional modeling of the proposed smart classroom and advance toolboxes use, the student are given the advantage of using three dimensional printers for implementing real world to the ideal world. The sixth component concerns virtual reality audiovisual recording. The on-demand content, directly implying cloud technology adaptation and three dimensional user actions to further refine the educational process. A matter of overhead given the recording service can be balanced by cloud computing and CS dimensionality reduction. The final component referred to as virtual course assessment service accurately evaluates student

mental state and performance by intelligent algorithms and clustering techniques for sensor and haptic data, the latter admitting a balanced optimization problem formulation for minimum resource expenditure in exchange with a tolerable estimation error of student performance as this component is responsible for various tasks such as evaluation reports through diverse devices, laptops, tablets, mobile phones e.t.c. It must also be stressed that information sharing between students, while in virtual course, could tremendously improve smart education interactive classroom quality as a measure of cooperation. Thus, thorough understanding will be promoted along with feedback for the educator in order to assess collective performance evaluation and also instrument modifications to the smart education process tailored to separate groups of students.

4.5.5 Smart Interactive Classroom implementation

The tools that will contribute to smart classroom result reliability are internal and external quality control. The former is conducted from the research team and collaborating institutions for project realization, while the latter performed from those accessing the virtual environment for filling in components for project development.

Concerning implementation, it is the aforementioned integration of wireless technologies that contribute to the WiFi Local Area Network consisting of an adapter, a router with access points deployed in the rooms facilitating the smart classroom environments along with indoor antennas in a relay assisted environment for amplifying received signal power. As a concluding remark, hardware implementation must essentially be integrated with software that define the runtime environment and interface development of the virtual class. Thus, gathering of data resulting from a certain indoor network infrastructure along with compression and low latency and complexity processing are the issues defined by software use in the proposed smart classroom.

4.5.6 Conclusions and future work

This section proposed smart interactive interconnected classroom and the beneficial application to virtual environment teaching procedure with the use of diverse technologies and powerful computational tools that holistically improve the educational process in real-time scenarios and implementation. Extension of this proposition to a pure laboratory class and augmented reality environment in different languages are challenging issues for future research. This specialized laboratory smart education scenario transparent to nationalities in terms of course content will surely benefit from the smart classroom paradigm in primary and secondary education. It seems also apparent that the societal impact of this advanced smart education class will be more than merely an advanced educational scheme but also a motivation for the evolution of smart classroom realizations with the use of the most recent communication and information sharing and processing technologies.

5. APPLICATION TO FIFTH GENERATION COMMUNICATIONS SYSTEMS

This chapter is devoted to investigating application of the core concepts considered as assumption prerequisites in this thesis and all technical results from the previous chapters that resulted due to the consideration of the former in the 5G wireless communications systems.

5G wireless systems [171],[172] constitute the current widely deployed but also rapidly evolving communication systems generation which, as each newly adopted communication technology, promises to successfully address bottlenecks and limitations of past communication system generations and also further boost performance, the latter admitting numerous interpretations and metrics considered according to the requirements of the communication scenario. To this end, 5G systems have evolved in parallel with the information technology and the data explosion deluge fundamentally termed as Big Data with all its corresponding properties successfully emigrating to wireless communication performance evaluation as well. Thus, 5G networks have already reached a mature scientific research stage and are becoming a reality in their deployment.

As the 4G LTE networks are gradually surpassed, 5G paradigm promises high data rates specifically on the order of Gbps, energy efficiency, low latency and increased capacity as well as improved QoS in the context of demanding multimedia video streaming thus characterized by resource expensive requirements. It is the latter issue together with mobility that conveys the need for network performance under such practical constraints. This comes along with the Big Data deluge that poses even greater challenge of tremendously increased data traffic management. Cooperative communication schemes along with multiple antenna implementation, initially at receiver side and also extended to transmitter side with closed loop consideration and implementation bottleneck further promise improved performance whereas the optimization of CS along with information theory and statistical independence could potentially provide asymptotic performance as opposed to essential correlation in a practical sense. Correlation must be set according to maximum value dictated for reconstruction error minimization and minimum value dictate for energy estimation error. As by definition, the above schemes imply correlation based performance optimization and thus comparison is not only feasible but indicative of independence based derivations. Furthermore, communication in the range of 3 to 300 GHz termed mmWave communications are encompassed in the 5G network regime to increase capacity while exploiting this bandwidth range. However, the very small portion of the above bandwidth is being utilized which by definition implies the CS applicability as in the case of CR spectrum sensing. The anywhere anytime 5G requirement also blends in with cloud computing paradigm and caching at mobile users, the latter being equivalent in the context of low latency and on demand availability in real time information exchange. Referring to the dense cellular network deployment, increased interference versus reduced energy and high connectivity should be carefully balanced, in a caching randomized sense and efficient statistical modeling. Correlation may thus reduce throughput but improve CS based useful information extraction in terms of

induced redundancy. Dense deployment also involves directional smart antennas for mitigating interference and optimize achieved throughput. Further notion on related beamforming technique will be provided below where, relative to statistical independence, interference could be minimized by independent radiation patterns with CS applicability in terms of incoherence between sparsifying transformation and measurement matrices. Moreover, channel modeling and estimation emerge as issues more imperative than ever in the 5G system modeling. Hence, quantification of channel information in receiver and, as an extension to transmitters, admits the distinction between channel state information and channel distribution information. The practical cases investigated in 5G systems usually account for the latter, hence, channel partial knowledge which is perfectly related to our derivations employing relative entropy as a means for assessing redundant bits required when modeling an accurate distribution by an approximating one. The results of Chapter 3 promise to provide insightful properties of the CS inference based fading channels from an information theoretic channel coding perspective and its inherent additional complexity. In a multi-antenna setting, the knowledge of such channels is intractable hence CS could point to an effective quantity for the latter. As in this thesis fading channel distributions are considered along with the additive noise constraint, the LOS path leads to improved connectivity whereas NLOS conditions require various mechanisms for combating severe distortion as well as equalizers that further add complexity and may introduce latency. Hence, applying CS questions the feasibility of a reduced complexity such approach along with the optimal effect on the channel distortion, leaving the latter as future challenge to be addressed. Relative to multipath fading in a MIMO equivalent channel model the strategy of beamforming dictates steering the antenna array beam to a specific direction so as to increase throughput and reduce interference. However, in a multipath environment the received signal components either add constructively or destructively hence a demanding implementation complexity approach would require phase shifters to correct the multipath phases in order to provide beneficial information for signal decoding. To that end, as opposed to independence assumption, correlation is desirable and CS may provide the means for reducing complexity while still ensuring accurate signal decoding. Clearly, the mMIMO schemes ensure constructive addition of wavefronts at the cost of hardware complexity. The correlation concept is more adaptable as opposed to orthogonality, the latter representing an ideal case which does not practically hold. Moreover, this case is relaxed for boosting excessive user accommodation and spectral efficiency in the context of NOMA. Moving on to spectral efficiency methods, full duplex communication promises to boost capacity and feedback quality as well as 5G network security. Relative to our analysis in Chapter 3, the claimed user selection scheme for security resilience formulates an interesting problem of whether full duplex compromises data integrity or aids security ensuring communication overhead in an environment of legitimate users and sparse number of attackers.

Regarding mm-wave communications [173] that involve climbing upwards in frequency bands from 30GHz to 300 GHz introduces both merits as well as drawbacks. The promise for increase of capacity anticipated to be expected is a major driving force for exploiting unused spectrum. The concept of CS incoherence serving as a measure of diversity is an indicative step towards this direction which however will require

advanced signal processing techniques, a context that could relate to CS with dynamic sparsity assumption and investigation for coding schemes tailored to these frequency ranges. Hence, it is channel modeling, potentially in a CS framework, that must be progressed as attenuation increases by climbing up in frequency. The sparse dominant signal components in such scenarios ought to be exploited. To combat this path loss, highly directional antennas for steerable beamforming also constitutes an imperative issue to advance along with correlation assumption given the multipath components. Thus, the independence based results of this thesis can be further incorporated in a challenging UWB analytical channel model but could also be extended to a physical model accounting for electromagnetic properties of this channel. Entropy based consideration and CS exploitation of degrees of freedom pose an interesting channel distribution knowledge future problem. Nonetheless, CS nonadaptivity could also provide a compact characterization of the entire mmWave band by a channel model instead of producing several, as successively higher bands are attempted to be exploited. Clustering could thus be considered in a temporal as well as spatial domain which constitute two of the many degrees of freedom such as frequency, code, power or angle.

Focusing on the mMIMO technology [174], [175] the use of compact antenna sets comprising of a large number is a step further with the increased hardware complexity problem. However, it is an essential prerequisite in order to achieve beamforming gains by verifying that the radiation pattern beam is directive and narrow enough so as to increase capacity of the network and suppress interference. Integration of the above to precoding, particularly in the challenging NLOS channel assumption, is an interesting problem also viewable from an information theoretic point of view. In this context, antenna spacing being a function of wavelength defined in the mmWave band regime also depends on correlation which verifies the practicality of the latter as opposed to independence related diversity for achieving capacity. mMIMO promises to further boost spectral efficiency which in a CS context could be contemplated with the low complexity information extraction already benefited with optimal exploitation of spectrum. To this end, correlation could hinder capacity increase. Nonetheless, it should be considered as a practical assumption in the system design problem. At this point, it is worth noting that asymptotic system performance evaluation on the basis of number of antennas tending to infinity, channel orthogonality emerges as a consequence. Hence, the independence based analysis of this thesis could indeed provide valuable insights to this asymptotic wise consideration along with applying the mathematical tool of CLT and performing comparison to practical case of finite antennas number as well as correlation of mMIMO channel statistical modeling. Moreover, reducing complexity in such an asymptotic case could effectively emigrate to a practical massive antenna deployment scenario. Such tradeoffs remain to be investigated by employing CS as well as effective entropy based analysis quantifying uncertainty of the wireless channel. In a beamforming scenario, sparsity of RF chains as opposed to number of antennas constitutes a well posed CS performance optimization problem. In this context, interference could more easily be cancelled and energy efficiency can be achieved. As an indication of correlation being a practical assumption, the de-correlating process for promoting spatial multiplexing is an essential step and could directly relate to a fully scaled problem including all channel impairments, namely,

additive noise, fading and interference followed by the optimal derivation of correlation value. Another issue of mMIMO technology is antenna interleaving which as number of antenna elements grows could evolve in a computationally demanding combinatorial problem to be solved. CS could thus be utilized together with a compromise of beamforming gain achieved. Another point of similarity is the computational exhaustive CSI in the mMIMO setting. Hence, a reduced complexity approach for the latter could lead to the partial distribution knowledge problem adopted in this thesis along with variable length coding quantified by Shannon and relative entropies. Channel reciprocity is another convenient property that could improve channel estimation. From the above, it can be easily deduced that CS optimization is indeed applicable in the mMIMO regime for reducing complexity emerging as a consequence of massive antenna deployment, potentially assuming optimization in a case specific adaptive sense. It is also noted that in frequency selective channel model, the occurrence of a deep fade is highly unlikely when statistical independence is assumed. This implies the application of CS and uncertainty quantification in the latter fading model. Additionally, in high mobility scenarios, the correlation of mobility parameters and channel dynamics points towards performance improvement of the 5G networks. Also, due to high data rates achieved, the measurements for estimating the channel in a scheme relying on measurement history could become outdated and thus of limited use. Moreover, an estimation error could likely propagate in future computations thus severely degrading performance. Furthermore, it is the signaling overhead and critical resource allocation in the mMIMO setting that poses the major complexity problem, which could relate to our statistical fading channel analysis since partial distribution knowledge fits into the scheduling problem as a practical constraint. Concerning mMIMO detection, sparsity based algorithms are already applicable, hence, the notion of compressibility as a measure of deviation from strict sparsity can be combined for improved detection performance providing robustness to fading, an observation admitting further extensions as to the statistical channel model considered. Performance could also be enhanced by claiming energy efficiency achieved by increasing antennas thus reducing transmit power in a «spreading» sense approaching a uniform allocation model. The latter could be compared to a biased power allocation scheme accompanied by a reduced CS based complexity adaptive scheme. Similar to the thesis derivations, a combined entropy based channel coding scheme by initially narrowing down channel multipaths arrived at receiver could also compensate for partial distribution knowledge along with indicating presence of interference. In such complex scenarios, CS compressibility could render error rate performance improvement feasible and also address the algorithmic expensive computations such as matrix inversion as the channel matrix dimensionality grows. Proceeding to a relative mathematical formulation, if such matrices are composed as a sum of diagonal and off-diagonal matrix and the inversion applies on the diagonal case, significant computation savings are achieved. Remaining at a mathematical fading channel model perspective, the relation of entropy as opposed to mathematical expression constitutes an interesting approach extending beyond statistical independence to a correlation-aware scenario. It is also worth mentioning that the correlation practical value is encapsulated in the fact that in a mMIMO setting, transmitted and received signals are correlated, a property to be accounted for. This mutual dependence could aid reconstruction quality as well as low

complexity resulting from the CS theory. A very attractive approach of sphere decoding which by definition limits constellation points to those within a specified radius is the consideration of the criterion that defines this radius. This could translate to a CS constrained optimization problem providing refined results.

Proceeding to dynamic spectrum sharing in a 5G CR network [176], [177], the dimensions of time and space jointly define the principle of frequency reuse feasibility and provide potential for dynamic spectrum, as a valuable resource in CR networks without degrading performance. What defines the complexity of spectrum sensing is the requirement to adjust software and hardware in this dynamic problem. Hence, in this dynamic scenario of exploitation of higher frequency bands, spectrum sensing techniques must be accordingly tailored to the specific applications. A challenging issue regards the spectrum sensing accuracy with the requirement for continuous sensing, which requires effective algorithmic design. Algorithmic complexity arises in the context of making use of past sensing history deemed necessary in this dynamic setting. In this higher frequency regime, CS may introduce the problem of UWB sparsity which could demand increased interactions to estimate in a prior to sensing phase. Moreover, an interference-aware bandwidth allocation scheme could prove to be of combinatorial nature thus admitting, under CS theory defined conditions, a computationally efficient method for addressing such a problem. CS could also be integrated with channel coding and spectrum access strategies, namely overlay and underlay to optimize spectrum utilization and decrease sensing time delay. Information theoretic channel modeling could also aid in optimizing SUs spectrum access accounting for tolerable interference from the PU activity side. The strategy of NOMA, for which detailed remarks will be given in the subsequent paragraph, is a strategy that unequally allocates power to a large number of users along with interference cancellation, a valuable result in the 5G CR context. The spectrum sensing efficiency is based on spectrum inference which essentially involves spectrum prediction based on past spectrum occupancy statistics exploiting correlation properties. Thus, spectrum inference poses a tremendous challenge for applying the probabilistic approach in this thesis accounting for additive noise and fading channel distribution as practical constraints along with exploitation of past decisions as well as occupancy related statistics. Compressive spectrum sensing addressing the wideband channel sensing issue can be effectively combined in order to reduce complexity and loosen the stringent requirements on ADCs functionality as an impact on hardware complexity. The time, space and frequency domains are the straightforward ones for dynamic spectrum inference and access. However, code domain also conveys promising results by assigning code signatures related to sensing of a certain set of frequency bands. With reference to time domain and the zero/one problem formulation for expressing channel availability or occupancy respectively, inference can incorporate both stationary and non-stationary models given the variation of the respective probabilities. Along with dynamic probabilities of detection and false alarm along with fast varying channel conditions, the latter setting constitutes a feasible extension of our work that adopts a target false alarm probability scenario. Given the spatial spectrum prediction model along with our analysis including CRs and WSNs spectrum occupancy variations encompass all dimensions namely, time, frequency and space and it is in this sense that correlation must be considered. It is thus imperative from an algorithmic design sense, that the channel parameter, sensing of spectrum as

well as interference level are evaluated. To that end, a comparison must be made between CS linear projections of reduced dimensionality and linear prediction employed for inference algorithms. However, the claim that nonlinearity, as more flexible, could result in improved adaptive CS optimization algorithms along with convexity being applicable can characterize efficient spectrum inference in a digital signal processing context. Weighting in the characterization of the significance of past observations for future ones can provide the means for optimal spectrum sensing and inference and also point towards the optimal algorithm for the problem at hand. Thus, even in the case of known distribution the assignment of probabilities for each decision, busy or idle, can prove to be computationally expensive. In such cases, additional a priori knowledge or entropy related analysis can provide the key tools for alleviating complexity. Moreover, Bayesian inference is a well performing methodology based on a priori distribution to predict a posteriori distribution. This can be combined in the CS inference context to provide insights about performance evaluation by incorporating temporal as well as spatial correlations. Entropy can be successfully integrated with our convolution based probabilistic scheme to improve spectrum sensing in the context of predictability. In this randomized setting, mobility arises as a useful strategy compensating for localized poor channel conditions as well as requirements for ensuring connectivity and continuous transmission always in a pattern-wise context. This could further aid dynamic topology configurations and sensor failure management in distributed sensor network setting. As an additional comment, a joint domain inference approach i.e. including the essential code as well as angle domains i.e. in multiple antenna setting could further boost prediction quality and spectrum sensing efficiency.

Moving on to NOMA [178],[179], utilization of a resource block accommodating more users is achieved by relaxing orthogonality. Providing the means of achieving spectral efficiency and increased throughput, NOMA brings forth the question of non-adaptive allocation of resources in a diverse set of fading channel conditions and also the contrast of this approach with CS compressibility criterion which more closely resembles the strategy of choosing users with good channel conditions. Furthermore, fairness could serve as a practical constraint to the above optimization problem for optimal allocation of available resources. In other words, NOMA ensures massive network deployment which, combined with spatio-temporal correlation consideration, further enhances user accommodation. Given the straightforward claim that NOMA pursuing fairness could degrade performance, the benefits of such strategy is spectral efficiency which admits further extension in terms of an information theoretic limit accounting for achieved rate. A probabilistic analysis similar to the one in this thesis based on convolutional statistics could adopt NOMA in order to orient optimization in a QoS perspective, coined by the term CR-NOMA. An essential comment is the application of hybrid NOMA where a closely related to CS combinatorial problem approach is the grouping of users in order to assign same resource to each group. Performance issues with correlations arise in this context particularly relative to decoding complexity. Furthermore, the differentiation of user channel conditions providing the performance gain could well be related to CS principle of incoherence which is also a prerequisite for CS performance optimization. Moreover, the derivation of average code length for describing the channel, as adopted in the analysis of this thesis, combined with capacity achieving

coding is a fruitful direction to be considered. Moreover, user ordering employing entropy criterion is a feasible extension of the thesis analysis. An important issue is that of user ordering according to channel feedback. There exist many parameters to be considered for this problem such as the issue of outdated data due to rapid channel variations, the complexity of acquiring useful data as well as correlation property which could relax the need for high overhead. CS compressibility is a potentially useful tool for the latter problem formed by assuming different number of antennas at the BS and at users' side. Multiple users with diverse requirements also resemble a cooperative scheme with NOMA applicability. In this context, full duplex communication can provide the superior performance of NOMA compared to half duplex relaying. User ordering according to QoS requirements, instead of channel conditions, is another direction to be investigated by means of an optimal probabilistic model accounting for correlation as a more practical case compared to independence. To that end, channel conditions, specifically attenuation as a consequence of using higher mmWave frequency bands, is another parameter to be considered. Channel distribution knowledge and approximation using a different distribution thus employing relative entropy could lead to assessment of performance limitations and interference mitigation resulting from practical assumptions of NOMA scheme. Additionally, in the CS approach, learning the optimal sparsifying domain along with a randomized measurement matrix stemming from statistical channel model is an attractive alternative. From the above, it is apparent that the benefits of applying NOMA principles integrated with other 5G technologies further boost performance capabilities and compensate for issues such as energy efficiency together with the well justified user scalability. From a coding perspective, the strategy of treating other users' signals as noise in the decoding process is already mathematically supported in terms of achieved capacity in the NOMA fairness context. Towards achieving interference mitigation, processing power as a requirement in 5G networks could benefit from the statistical assumptions of independence versus correlation combined with NOMA as a means of achieving massive connectivity. CS based sparse number of users compared to magnitude of antennas deployed is also a viable point of investigation which could provide optimized bounds of performance. NOMA oriented cooperative communications involving resource sharing i.e. power but also computational burden could be further elaborated in the query of whether a decentralized scenario such as in a WSN could be supported in this asymmetric context. In such an attempt, the promise for more user accommodation by NOMA could be beneficially integrated. A measure of the merit of rate achieving performance by employing NOMA is determined by the assumption of low SNR which also extends to CR-ED scheme efficiency in such regime. Integration of fading channel model with beamforming based beam misalignment cases and the impact of the latter on NOMA performance degradation could be combined with uncertainty quantification approach and channel coding. Hence, misalignment can be compensated for by NOMA for serving maximum number of users by each beam. Relative to NOMA cellular network cases, the typical assumption states that far users are poorly served as opposed to nearby users. However, the shadowing effect which could render the above statement inaccurate requires further attention in terms of the manner on which NOMA enables spectral efficiency and extensive user accommodation along with the impact on interference cancellation

decoding task. Stated differently, the imperfect channel knowledge could result in decoding order ambiguity, an issue that could be alleviated by additional a priori information. Relaying is also based on this principle. Additionally, the pursuit of fairness by NOMA complying to the uniform allocation strategy as the analysis of WSN in this thesis, further stresses the need of evaluating CR scheme in terms of opportunistic spectrum use in a manner of not inducing harmful interference for PU. Energy harvesting constitutes a step further towards NOMA oriented selection of users with good channel conditions in order to assist weak powered users. Having already stated the merit of randomized mobility, the integration with NOMA exploiting a certain pattern, could enable the massive connectivity requirement of such networks. Regarding NOMA and cloud/edge computing integration the asymmetric computation property along with easy access of frequently requested information could provide practical performance enhancements and assist in low latency and minimized energy allocated in the NOMA sense.

Proceeding to the energy efficiency requirement for 5G networks [180], the fundamental statement encapsulates the fact that energy efficiency is characterized by certain tradeoffs that relate to spectral efficiency discussed above in the NOMA context, deployment efficiency, delay and bandwidth. In plain words, energy efficiency, as a desired goal to be achieved, is an outcome of many parameters relative to wireless system performance. Hence, the consideration of CLT which in theory applies to Gaussian and non-Gaussian distribution models, as in our analysis can potentially encompass various practical channel models and convey performance benchmarks by simultaneously exploiting statistical correlation. The OFDM scheme based on sub-band division, promotes energy efficiency by optimally utilizing narrowband channels for transmission thus promoting energy efficiency, and resilience to frequency selective fading and interference. In a practical sense, energy efficiency and achievable rate can be a vital set of conflicting objectives. To that end, previously discussed NOMA pursuit of fairness can provide a viable performance balance. Moreover, delay constraints can translate to transmission rate requirements. A target value capacity can also provide a solution to the above issue. Furthermore, heterogeneous network design, diverse fading conditions and asymmetric computations assisted by correlation parameter and uniform energy consumption can redefine performance bounds in a most practical manner answering 5G network requirements and result in effective algorithm design for the above. Diversity in terms of power consumed by hardware circuitry is another parameter affecting energy efficiency, rendering CS optimization tools applicable in UWB communications thus simplifying implementation complexity. It must also be noted that properly selected coding scheme is energy efficient as well and interference mitigating. Energy efficiency tailored to each user as well as characterizing the whole system can be met with joint channel assignment which strongly implies statistical correlation. Nevertheless, the independence model can provide performance bounds and limitations when compared to the more realistic correlation assumption. Regarding the feedback addressing channel distortion minimization, channel estimation with the energy efficiency constraint impinges on complexity with the additional issue of outdated measurement history, which heavily degrades channel estimation for rapidly varying channels. As a means of reducing hardware complexity in a multiple antenna setting, antenna selection involves zeroing power allocation, an approach which is

based on uniform power spreading and must be investigated by considering mutual coupling effect regarding empowered antennas. A very interesting approach of achieving capacity for a multi-antenna fading channel concerns the asymptotically cancelling of fading randomness as transmit antennas tend to infinity. In a mMIMO setting, this rule could establish useful bounds to be achieved together with employing CS optimization theory. It is already evident that for practical finite antennas number spectral efficiency is improved. As for the mathematical formalism supporting massive connectivity, CS random matrix theory could provide reduced complexity as the channel matrix formulation in the performance optimization problem generally depends on the product of the numbers of transmit and receive antennas, which is a cumbersome indication. Energy efficiency in a heterogeneous network is a function of many parameters. The mere definition of heterogeneity directly implies the different energy budget such as macro or smaller cells and the query of whether the total consumed energy needs to be coordinated by clustering or, in general, a central BS. In such architectures, randomized models could address deficiencies and derive the optimal topology for ensuring the desired connectivity. The quantification of uncertainty given asymmetric computations and asynchronous communication is another indicative manner of modeling.

Concluding with 5G oriented applications of CS and results of this thesis to 5G network paradigm we refer to the mobile edge caching schemes in a mobile perspective [181],[182] as to address the data exponential growth which constitutes the critical burden on 5G system parameter design and evaluation. Edge computing essentially reduces latency induced by distance between users and data center in the cloud computing paradigm. Hence, MEC has arisen as a vital 5G network technology. Furthermore, it is the capability of MEC to assess channel conditions and mobility patterns that render the entire analysis of this thesis applicable by adopting a statistical model exploiting similarities and abilities so that an optimization problem with practical constraints can be formulated. In other words, additional knowledge can be leveraged to obtain the caching solution that optimally fits into the caching application. As briefly commented in Chapter 4, mobility caching is a strategy that can provide significant improvement of content caching and delivery depending on file popularity and velocity of mobile provider. Thus, classification of popularity and request frequency by means of entropy could further characterize the mobile caching scheme adopted. Bandwidth requirements as a function of file popularity, channel conditions and dynamic data traffic constitutes an extension viable to the WSN analysis in Chapter 3. As MEC corresponds to real time applications, a probabilistic analysis comparing different distributions, such as in this thesis, could provide the solution to low latency and connectivity. Moreover, latency reduction could be traded with energy efficiency. Regarding cellular network architecture distinction between centralized and distributed, the crucial differentiation relates to global information in the former compared to localized knowledge confined to the user neighborhood. This contributes to shaping the caching problem in terms of file segments selection to be cached and extent to which the caching strategy considered adequately serves the target set of users requesting content. To that end, cooperative caching can further enhance the efficiency of edge caching scheme at the expense of increased overhead. The latter can be reduced in a manner accounting for correlation of cached content which as deduced in the WSN

analysis in Chapter 3, generates redundancy and thus reduces cached, requested and delivered content. It is worth noting that the aforementioned patterns should be exploited in order to contribute to the tradeoff between randomness resulting from statistical models adopted and structure. The latter favors implementation and is the main property reflecting CS sparsity based optimization. As a comment to the latter, the small dimensionality of frequently requested content compared to the total files cached in mobile providers could assume sparsity property and provide practical bounds on caching efficiency. It is also imperative to stress the merit of cached content prior to request or after the specific file has been requested. This raises the concerns of caching efficiency and past request history in addition to need for adequate storage capacity in mobile caches. Caching efficiency is also dependent on the alignment of mobility pattern and file request pattern. This direction subtly implies that CS could be applied in terms of these patterns in order to alleviate complexity in a randomized model that jointly encompasses the above. In a post content delivery context, update phase relates to finite storage capacity and optimal cached content tailored to each mobile provider. Sharing of cached content is feasible while reducing energy consumption. From a mathematical point of view, statistical correlation can be employed in terms of a weighted expression of the cached files either in a single user's cache or in a group of mobile users confined in an area inside the network. Thus, correlation induced redundancy supporting a decentralized network topology, as in our analysis in Chapter 3, further promotes caching efficiency. Caching efficiency also conveys a tradeoff with respect to interference, which necessitates utilization of beamforming techniques to limit cached content delivery to the desired target users. However, cooperation which promises availability of requested content in a localized manner also amplifies interference. Clearly, if a file request can be served by redundant mobile providers, the increase in throughput and mitigation of interference is feasible. To that end, a well fitted probabilistic model just as in our analysis by including fading and noise for channel conditions could contribute not only to low complexity file delivery but also optimize resource allocation such as bandwidth or power and reduce delay. Moreover, integration of content sharing and time varying non-uniform file popularity in an information theoretic framework should be considered to address practical edge caching scenarios while adopting probabilistic models that accurately predict dynamic file requests. The effectiveness of such models when integrated with our comparisons made on the basis of the three distinct distributions will surely sharpen the accuracy of our approach and lead to optimal prediction of future file requests and deliveries to the intended users with the aid of mobility. Another interesting point to be commented relates to the content placement problem which, in the absence of extra knowledge, could lead to a computationally demanding problem where CS theory can be applied, the latter exploiting features such a limited number of most popular files requested and salient correlation of file requests. Concluding this topic, dynamic file requests and deliveries entail optimal exploitation of both time and space along with all dependencies that regulate these domains. Probabilistic models along with uncertainty quantification portray interesting approaches that randomness can be used to compensate for deficiencies arising from rapid variations of file requests and channel conditions specifically interference.

6. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

6.1 Conclusions

The current wireless communication systems have naturally evolved in the era of exponential data growth that does not cease to increase in dimensions as well as heterogeneity and diversity. This data growth emigrates to wireless communication system performance and design requirements with a vast range of conflicting objectives that formed what is referred to by the term tradeoffs. As the type of information that are transmitted wirelessly and reconstructed at receiver side fall into the category of multimedia streaming which is characterized as resource demanding when exchanged through wireless systems, signal processing techniques that are of low complexity while still preserving information content are more needed than ever. To this end, current wireless communication technology continues to evolve based on thorough understanding of current system limitations and providing solution to overcome the latter.

This is where CS steps into the stage to offer low complexity information processing which in terms of sampling and acquisition outperforms the traditional Shannon-Nyquist sampling theorem and thus provides promising results for low complexity and simplified hardware decoding at receiver end. The mere requirement for achieving the above goals is structure at information translating to the notion of sparsity. This structure existence is verified in most scientific areas as well as wireless communications. Along with statistics and information theory, modeling of wireless communication systems and performance assessment along with CS principles applied this thesis has contributed to performance evaluation of wireless system performance with the practical diverse fading and additive Gaussian noise considerations. The fundamental statistical assumption upon which the analysis of this thesis is founded is statistical independence. As an assumption, the latter statistical property is far from realistic as is perfectly illustrated in the findings of noisy fading channels modeling and performance. Indicatively, information theoretic derivation of average code description length for the wireless channel exhibited less samples when the additive noise was considered as a practical assumption of the channel. Moreover, Rician fading accounting for an optical LOS link between transmitter and receiver was found to require more bits than the Rayleigh fading channel which relates to severe fading and NLOS conditions. All these are due to the prism of independence. A feasible interpretation to the above states that channel with more uncertainty i.e. most informative require less bits under the independent assumption. Moreover, the analysis conveyed that uncertainty can be coupled with CS reconstruction accuracy with reference to the fading channel.

Another point relating CS to fading channel distributions in the additive noise regime, was inference based on preserving largest in magnitude channel gains as a measure of compressibility. This inference resulted in very indicative results that reflect the fact that CS does not increase complexity if statistical independence is assumed to hold. Three distinct fading distributions were considered and the analysis conducted revealed that CS based inference resulted not only in error performance gains but also in zero additional complexity. From a mathematical point of view, the above results are

consequences of the fact that the convolution operation in terms of statistical independence creates mathematical expressions that produce these findings. All these results admit the aforementioned interpretations. Moreover, in order to adopt an approach that resembles CS linear projections a Taylor polynomial representation was leveraged with the assumption of second degree polynomial as the minimum order capturing curvature verified by fading distributions curves. Performance tradeoffs were stated relative to convenient property of polynomial being differentiated and integrated in a trivial manner. This property was thus traded with performance penalty. It must also be noted that the results from the Taylor approximated distribution are all consequences of the combined approximation quality as a function of expansion point and polynomial degree as well as statistical independence. Hence, CS inferred and Taylor approximation cases provided notable insights under this statistical assumption.

Proceeding to the CR technology, the additive noise cases were only accounted for as more realistic. The diverse fading distributions were considered and the results obtained categorized CS optimization case in terms of LRT statistics in favor of PU presence. Independence also impacted the derived results. Related threshold and CLT cases abiding by the Gaussian assumption served as optimistic measures of skewed non-Gaussian fading distributions considered.

The thesis analysis also proceeded in the consideration of temporal, spatial and spatiotemporal correlation in a WSN setting with the vital assumption of Gaussian distribution i.e. symmetric and adequately described by mean and variance. Moreover, the zero correlation cases investigated directly implied independence by exploiting the relative property uniquely characterizing Gaussian statistics. Reconstruction error and energy estimation errors were shown to demand conflicting correlation magnitudes for each of the aforementioned errors. Specifically, reconstruction error was found to improve with increasing correlation while lowest energy estimation errors resulted from low correlation values. It thus became intuitive that, according to the WSN scenario at hand, optimal correlation value must be chosen as to provide a balance between these two error metrics. However, the main concept behind this analysis is its applicability to monitoring networks where abrupt readings are not received as opposed to other scenarios where real time operation is imperative and delay as well as dynamic readings are received.

The results of this thesis further encompassed content caching in the case of network exchanging frequently accessed data retrieved from static or mobile providers. A probabilistic model is adopted and the well-established Zipf-like distribution is compared to the uniform and Gaussian distributions. The results obtained were interpreted in terms of file popularity assumptions considered to being consistent. Finally, the CS based file segmentation scenario was adopted and the cases of small fraction, in terms of total file segments, as well as number of segments closely approaching the total file number were assumed and the total resulting probabilities produced convergent curves. The analysis briefly underlined the security issues in networks such as the SNs as a Big Data effect. Suggestions on applying CS to the practice of medicine as well as healthcare data management were drawn and led to the conclusion of the requirement on joint progress of many scientific areas in order to reap the optimization benefits of CS theory and practice. Finally, a smart education paradigm

was described in terms of an augmented reality interactive classroom that provides advantages as an integration of many technologies.

The results of all above chapters were directly fitted into the performance and limitations of 5G systems that constitute the currently deployed wireless systems generation. Potential for evolving in 5G and beyond are apparent under the prism of statistical models that address the overcoming of current limitations and compensate for the challenges of the information technologies relative to Big Data deluge.

6.2 Discussion and Future Research Directions

The major guideline for future investigation is the impractical assumption of statistical independence. Hence, although this thesis provided useful insights, comparison with the much more realistic case of correlation is imperative in order to approach a thorough understanding of performance bounds and limitations that shape today's wireless communication technology. In particular, correlation of fading channel realization as well as correlation of additive noise and fading define a practical scenario as opposed to independence.

To this direction, CS problem formulation by means of matrix algebra and l_p norms enhanced with correlation assumption being already under intense research indicates an approach that despite progress made can tremendously benefit from an information theoretic point of view not only by our channel coding consideration but with scrutinous research for efficient and capacity achieving coding schemes. The numerous interpretations that coding and information sources have taken point towards a significant gap between the concept of coding and the practical constraints that describe performance of wireless communication systems. All the above from a correlation point of view.

Another fundamental assumption upon which our analysis is founded is second order statistics. Hence, apart from the Gaussian distribution case in the WSN case where mean and variance are adequate to fully define the distribution, fading distribution require higher order moments given their asymmetric property. However, second order statistics are proven adequate to conduct the analysis in this thesis as variance being a measure of spread of the distribution with respect to its peak. Variance was considered as a measure of uncertainty given a specific distribution quantifying the average number of bits required for describing the channel while relative entropy is another measure of uncertainty quantifying additional required bits by approaching the assumed distribution by an approximating one. Another fact that reflects the use of second order statistics is the use of Gaussian related parameters such as threshold for our proposed ED method for CR as well as CS required number of samples. It is thus straightforward that extension of the analysis in this thesis considers higher order moments and the consequences in channel modeling accuracy in this extended case. It must also be noted that non-Gaussian distributions are not only confined to fading channel models but also to non-Gaussian noise as well as modeling other problem parameters. In summary, examining the distributions included in this thesis from an information theoretic point of view some remarks indicate future research directions. The uniform distribution is the most uncertainty containing distribution while Gaussian distribution entropy is only

a function of its variance. This also deduces that second order statistics does not degrade entropy calculations.

It is also worth noting that CS theory essentially deals with underdetermined problems where structure is necessary in order to narrow down solutions with reference to the initial combinatorial problem. Hence, it is necessary to relate the above to the wireless communication system degrees of freedom, namely, time, frequency, space and code. Throughout this thesis the time and space domains were considered while frequency and code were not accounted for except for the entropy based average code length derivations for the latter code dimension. Regarding frequency domain, the findings of this thesis are most representative and extension to the frequency domain such as power spectrum analysis, spectral efficiency along with capacity achieving coding schemes can be made. In the context of 5G communication systems, effective bandwidth for employing CS based notion of sparsity is another interesting direction. Moreover, correlation in the frequency domain is also an interesting consideration if integrated with the twofold bandwidth allocation issues: spectrum scarcity and spectrum underutilization. Compressive spectrum sensing is also an already mature scientific area, relative to which the computational complexity in the UWB regime is drastically reduced by sub-Nyquist sampling technique. Concerning code, the vast range of efficient coding schemes with the pursuit of achieving Shannon capacity in terms of the specific application formulates various extensions of our work which suggests statistical independence and correlation contrast.

Additionally, spatial diversity exploitation being the definition of multiple antenna systems and encompassing the reason for achieving improved performance already verifies intense scientific research by means of CS optimization theory, a crucial observation being the multipath clustering and sparse dominant clusters admitting CS based statistical modeling. One step further in the 5G wireless system paradigm being mMIMO in higher frequency ranges is a plausible extension of CS and realistic assumption of correlation. Multiuser setting also poses interesting cooperation and performance issues in the 5G cellular network paradigm. Summarizing the effect of CS on multiple antennas at receiver as well as transmitter, the angle domain reflects the AOAs of multipaths at the receiver. Combined with geometrical models of multipath scattering environment and geometric perspective of CS norm minimization theory in terms of the optimal value of parameter p for expressing the l_p norm also poses a challenging optimization problem.

Due to power domain consideration, the NOMA oriented pursuit of fairness implying a uniform power allocation and in general resource allocation must also be further investigated. Contrary to this strategy, the approach of selecting users with for instance better channel conditions or equivalently more remaining energy as in our WSN correlation based analysis is also a viable solution. The two aforementioned strategies could be simultaneously leveraged in a heterogeneous network setting by carefully contemplating performance tradeoffs.

Moving on to a mathematical formalism oriented issue, the contrast of discreteness and continuity is also prominent in this thesis. Hence, it is observed that entropy calculations abide by the discrete equivalent definition leading to quantification of average code

length for the channel. However, variance calculations in terms of first and second order moments, the latter being included in the Lagrange multiplier setting in the combined entropy and CS distribution reconstruction scenarios, all assume integral derivations thus continuous assumptions. Among the most important results in this thesis that contributed to the distinction between statistical independence and correlation concerns the additive noise setting where moment integrals derive different results as opposed to the independence based variance being the variance increased by one due to unit variance additive noise. Clearly, the fading distributions considered are continuous while channel gains generated are discrete. This discreteness is evident in CS theory both relating to sparsity as well as compressibility.

Concluding this thesis, we boldly stress the fact that assessing wireless communication system performance either by CS theory or other optimization tools is all about formulating an optimization problem along with an efficient algorithm modeling the problem and adopted approach. This optimization problem as such includes the objective function to be minimized in terms of cost or penalty or maximizing the efficiency of a specific parameter. CS relates to unknown sparse vector to be minimized by employing l_p norm algebra. The next ingredient includes the constraints which compensate for mathematical properties to be fulfilled for instance the probabilities values taking values between zero and one as well as their sum being equal to one in terms of a complete probabilistic optimization approach. Constraints also extent to incorporating the strategy adopted in resource allocation in a wireless communication system and other practical constraints that will jointly define the optimization problem fitness to the specific wireless communication system at hand. Thus, what must be highlighted is that constraints are the key ingredients of an optimization problem that could even decide whether the problem has a solution or solution uniqueness can be sorted out by a feasible optimization region that could bear a number of solutions. This is exactly the concept behind CS that due to the underdetermined problem nature of the initial problem, sparsity enters the stage thus narrowing down solutions and under circumstances leads to the optimal solution by addressing a problem that is more solvable and of low complexity, in the sense that we depart from a combinatorial problem. Hence, by properly controlling the decision variables, the optimization problem can accurately model a practical communication scenario and quantify uncertainty due to the inherent stochastic nature of information transfer through the wireless channel. From the above, it becomes evident that optimization problem constraints and also properly calculated weighted quantities that are assigned to problem variables ultimately answer the question of whether the problem admits a solution departing from a computationally demanding context to a simpler one or inversely if the problem is rendered insolvable due to the very insertion of a specific constraint that nevertheless reflects the practicality of the problem at hand. Following the above, algorithmic design must firstly consider the optimization problem above and also provide convergence speed, and accuracy, an issue that may be proven cumbersome due to conflicting objectives i.e. performance tradeoffs. CS already exhibits a variety of effective algorithms that ensure the low complexity in terms of computations required.

As the 5G and beyond become a reality, optimization CS tools continue to evolve as most effective tools for performance evaluation and specifically contribute to accurately modeling the wireless channel along with the distortion and impairments that it causes to the transmitted signal. Thus, randomness and uncertainty are properties bound to information wirelessly transmitted signals and also encourage the adoption of a generic stochastic model for describing the information source in order to exploit randomness compensating for system design limitations. In plain words, statistical independence and correlation both bear the potential of providing performance insights. The main characteristics of the next generation wireless communication technologies being orders of magnitude greater amounts of resource expensive data transmission, processing and analysis as well as energy efficiency aligned with higher data rates achieved, performance evaluation problem modeling has already shown to be flexible and adaptive to encompass diverse requirements and desired goals. CS along with statistics and information theory will emerge as the key optimization tools for advancing into a new era of massive connectivity and increased capabilities offered by future communication technologies.

REFERENCES

[1] A. Goldsmith, “Wireless Communications”, Cambridge University Press, 2005.

[2] J. G. Proakis and M. Salehi, “Digital Communications”, McGraw-Hill International Edition, 5th Edition, 2008.

[3] D. Tse and P. Viswanath, “Fundamentals of Wireless Communication”, Cambridge: Cambridge University Press, 2005.

[4] L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," in IEEE Transactions on Information Theory, vol. 50, no. 5, pp. 748-767, May 2004.

[5] J. D. Parsons, “The Mobile Radio Propagation Channel”, John Wiley & Sons Ltd, 2nd Edition, 2000.

[6] M. K. Simon and M. Alouini, “Digital communication over fading channel: A unified approach to performance analysis”, John Wiley & Sons Ltd, 2000.

[7] M. Pätzold, “Mobile Fading Channels”, John Wiley & Sons Ltd, 2002.

[8] B. Kumbhani and R. S. Kshetrimayum, “MIMO wireless communications over generalized fading channels”, CRC Press, 2017.

[9] E. Biglieri, R. Calderbank, A. Constadinides, A. Goldsmith, A. Paulraj and H. Poor, “MIMO Wireless Communications”, Cambridge: Cambridge University Press, 2007.

[10] K. Vasudevan, “Digital communications and signal processing”, CRC Press, 2008.

[11] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek and R. S. Thoma, "Capacity of MIMO systems based on measured wireless channels," in IEEE Journal on Selected Areas in Communications, vol. 20, no. 3, pp. 561-569, April 2002.

[12] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity limits of MIMO channels," in IEEE Journal on Selected Areas in Communications, vol. 21, no. 5, pp. 684-702, June 2003.

[13] J. G. Proakis, "Adaptive equalization for TDMA digital mobile radio," in IEEE Transactions on Vehicular Technology, vol. 40, no. 2, pp. 333-341, May 1991.

[14] F. E. Abdel-Samie, F. S. Al-kamali, A. Y. Al-nahari and M. I. Dessouky, “SC-FDMA for mobile communications”, CRC Press, 2017.

[15] R. Prasad, “OFDM for Wireless Communications Systems”, Artech, 2004.

[16] K. Sh. Zigangirov, “Theory of Code Division Multiple Access Communication”, John Wiley and Sons Inc, 2004.

[17] T. M. Cover and J. A. Thomas, “Elements of information theory”, John Wiley & Sons Inc, 2nd Edition, 2005.

[18] D. J. C. Mackay, “Information theory, inference and learning algorithms”, Cambridge: Cambridge University Press, 2003.

[19] C. E. Shannon, "A mathematical theory of communication," in The Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, July 1948.

[20] C. Rush, A. Greig and R. Venkataramanan, "Capacity-Achieving Sparse Superposition Codes via Approximate Message Passing Decoding," in IEEE Transactions on Information Theory, vol. 63, no. 3, pp. 1476-1500, March 2017.

[21] D. S. Lun, M. Medard and I. C. Abou-Faycal, "On the performance of peaky capacity-achieving signaling on multipath fading channels," in IEEE Transactions on Communications, vol. 52, no. 6, pp. 931-938, June 2004.

[22] D. Baron, M. F. Duarte, M. Wakin, S. Sarvotham and R. Baraniuk, “Distributed Compressed Sensing”,ArXiv, abs/0901.3403, January 2009.

[23] D. L. Donoho, "Compressed sensing," in IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306, April 2006.

[24] Y.C. Eldar and G. Kutyniok, “Compressed Sensing: Theory and Applications”, Cambridge: Cambridge University Press, 2012.

[25] S. Foucart and H. Rauhut, “A Mathematical Introduction to Compressive Sensing”, Springer New York, 2013.

[26] M. Mishali and Y. C. Eldar, "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals," in IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 375-391, April 2010.

[27] C. E. Shannon, "Communication in the Presence of Noise," in Proceedings of the IRE, vol. 37, no. 1, pp. 10-21, Jan. 1949.

[28] J. Lu, N. Verma and N. K. Jha, "Compressed Signal Processing on Nyquist-Sampled Signals," in IEEE Transactions on Computers, vol. 65, no. 11, pp. 3293-3303, Nov. 2016.

[29] Y. Tang, X. Li, Y. Liu, J. Wang and Y. Xu, "Sparse dimensionality reduction based on compressed sensing," 2014 IEEE Wireless Communications and Networking Conference (WCNC), 2014, pp. 3373-3378.

[30] D. Baron and M. F. Duarte, "Universal MAP estimation in compressed sensing," 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2011, pp. 768-775.

[31] S. D. Apte, “Random Signal Processing”, CRC Press, 2017.

[32] H. Shi, H. Zhang, G. Li and X. Wang, "Stable Embedding of Grassmann Manifold via Gaussian Random Matrices," in IEEE Transactions on Information Theory, vol. 61, no. 5, pp. 2924-2941, May 2015.

[33] K. Usman, H. Gunawan and A. B. Suksmono, “Compressive Sensing Reconstruction Algorithm Using L1-norm minimization via L2-norm minimization”, International Journal on Electrical Engineering and Informatics, vol. 10, no. 1, pp. 37-50, March 2018.

[34] N. Nguyen, D. Needell and T. Woolf, "Linear Convergence of Stochastic Iterative Greedy Algorithms With Sparse Constraints," in IEEE Transactions on Information Theory, vol. 63, no. 11, pp. 6869-6895, Nov. 2017.

[35] S. A. Razavi, E. Ollila and V. Koivunen, "Robust greedy algorithms for compressed sensing," 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 969-973.

[36] Z. Tan, P. Yang and A. Nehorai, "Joint Sparse Recovery Method for Compressed Sensing With Structured Dictionary Mismatches," in IEEE Transactions on Signal Processing, vol. 62, no. 19, pp. 4997-5008, Oct.1, 2014.

[37] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 21-30, March 2008.

[38] J. Zhang and S. Zhang, "Null Space Property of L1-2 Minimization With Prior Support Information," in IEEE Signal Processing Letters, vol. 28, pp. 1779-1783, Aug. 2021.

[39] A. M. Tillmann and M. E. Pfetsch, "The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing," in IEEE Transactions on Information Theory, vol. 60, no. 2, pp. 1248-1259, Feb. 2014.

[40] A. M. Tillmann, "On the Computational Intractability of Exact and Approximate Dictionary Learning," in IEEE Signal Processing Letters, vol. 22, no. 1, pp. 45-49, Jan. 2015.

[41] Y. Ding and B. D. Rao, "Dictionary Learning-Based Sparse Channel Representation and Estimation for FDD Massive MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 17, no. 8, pp. 5437-5451, Aug. 2018.

[42] V. K. Amalladinne, J. -F. Chamberland and K. R. Narayanan, "A Coded Compressed Sensing Scheme for Unsourced Multiple Access," in IEEE Transactions on Information Theory, vol. 66, no. 10, pp. 6509-6533, Oct. 2020.

[43] F. E. Rezagah, S. Jalali, E. Erkip and H. V. Poor, "Compression-Based Compressed Sensing," in IEEE Transactions on Information Theory, vol. 63, no. 10, pp. 6735-6752, Oct. 2017.

[44] S. Feizi, M. Médard and M. Effros, "Compressive sensing over networks," 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2010, pp. 1129-1136.

[45] J. W. Choi, B. Shim, Y. Ding, B. Rao and D. I. Kim, "Compressed Sensing for Wireless Communications: Useful Tips and Tricks," in IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1527-1550, thirdquarter 2017.

[46] Z. Gao, L. Dai, S. Han, C. -L. I, Z. Wang and L. Hanzo, "Compressive Sensing Techniques for Next-Generation Wireless Communications," in IEEE Wireless Communications, vol. 25, no. 3, pp. 144-153, June 2018.

[47] M. Leinonen, M. Codreanu and G. Giannakis, “Compressed Sensing with Applications in Wireless Networks”, Foundations and Trends in Signal Processing, vol. 13, no. 1-2, pp. 1-282, 2019.

[48] W. U. Bajwa, J. Haupt, A. M. Sayeed and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," in Proceedings of the IEEE, vol. 98, no. 6, pp. 1058-1076, June 2010.

[49] J. L. Paredes, G. R. Arce and Z. Wang, "Ultra-Wideband Compressed Sensing: Channel Estimation," in IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 3, pp. 383-395, Oct. 2007.

[50] K. Kim and G. Shevlyakov, "Why Gaussianity?," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 102-113, March 2008.

[51] T. Xifilidis and K. E. Psannis, "Fading Channel Coding Based on Entropy and Compressive Sensing," 2020 3rd World Symposium on Communication Engineering (WSCE), 2020, pp. 44-48.

[52] A. Dogandzic and J. Jin, "Estimating statistical properties of MIMO Rician fading channels," Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002, pp. 149-153.

[53] A. Dogandzic and Jinghua Jin, "Estimating statistical properties of MIMO fading channels," in IEEE Transactions on Signal Processing, vol. 53, no. 8, pp. 3065-3080, Aug. 2005.

[54] M. Debbah and R. R. Muller, "MIMO channel modeling and the principle of maximum entropy," in IEEE Transactions on Information Theory, vol. 51, no. 5, pp. 1667-1690, May 2005.

[55] E. Biglieri, J. Proakis and S. Shamai, "Fading channels: information-theoretic and communications aspects," in IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2619-2692, Oct. 1998.

[56] S. Verdú, "Mismatched Estimation and Relative Entropy," in IEEE Transactions on Information Theory, vol. 56, no. 8, pp. 3712-3720, Aug. 2010.

[57] M. Chen and J. Lafferty, "Mismatched estimation and relative entropy in vector Gaussian channels," 2013 IEEE International Symposium on Information Theory, 2013, pp. 2845-2849.

[58] I. Kontoyiannis and S. Verdú, "Optimal Lossless Data Compression: Non-Asymptotics and Asymptotics," in IEEE Transactions on Information Theory, vol. 60, no. 2, pp. 777-795, Feb. 2014.

[59] N. Merhav, "On Empirical Cumulant Generating Functions of Code Lengths for Individual Sequences," in IEEE Transactions on Information Theory, vol. 63, no. 12, pp. 7729-7736, Dec. 2017.

[60] T. Xifilidis, K.E. Psannis, “Wireless fading channels performance based on Taylor expansion and compressed sensing: A comparative approach”, in Wiley International Journal of Communication Systems, vol.34, issue 8, e4794, March 2021.

[61] P. C. Sofotasios, S. Muhaidat, M. Valkama, M. Ghogho and G. K. Karagiannidis, "Entropy and Channel Capacity under Optimum Power and Rate Adaptation over Generalized Fading Conditions," in IEEE Signal Processing Letters, vol. 22, no. 11, pp. 2162-2166, Nov. 2015.

[62] A. Pramanik, S. P. Maity and Z. Farheen, “Compressed Sensing Channel Estimation in Massive MIMO”, in IET Communications, vol. 13, issue 19, pp. 3145-3152, Dec. 2019.

[63] W. Zhang, R. C. de Lamare, C. Pan, M. Chen, B. Wu and X. Bao, "Correlation-driven optimized Taylor expansion precoding for massive MIMO systems with correlated channels," 2017 IEEE International Conference on Communications (ICC), 2017.

[64] J. Munkhammar, L. Mattsson, and J. Rydèn, “Polynomial Probability Distribution Estimation Using the Method of Moments”, in PLoS One, vol. 12, issue 4, pp. 1-14, Apr. 2017.

[65] B. Li, C. Zhao, M. Sun, H. Zhang, Z. Zhou and A. Nallanathan, "A Bayesian Approach for Nonlinear Equalization and Signal Detection in Millimeter-Wave Communications," in IEEE Transactions on Wireless Communications, vol. 14, no. 7, pp. 3794-3809, July 2015.

[66] S. Lv, C. Xing, Z. Zhang and K. Long, "Guard Zone Based Interference Management for D2D-Aided Underlaying Cellular Networks," in IEEE Transactions on Vehicular Technology, vol. 66, no. 6, pp. 5466-5471, June 2017.

[67] W. Li, X. Chen, Q. Zhu, W. Zhong, D. Xu and F. Bai, "A Novel Segment-Based Model for Non-Stationary Vehicle-to-Vehicle Channels With Velocity Variations," in IEEE Access, vol. 7, pp. 133442-133451, Sept. 2019.

[68] C. D. Charalambous, S. Z. Denic and C. Constantinou, "Information Capacity of MIMO Channels with Relative Entropy Constraint," 2006 IEEE International Symposium on Information Theory, 2006, pp. 876-880.

[69] J. Iscar, İ. Güvenç, S. Dikmese and N. Rupasinghe, "Efficient Noise Variance Estimation Under Pilot Contamination for Massive MIMO Systems," in IEEE Transactions on Vehicular Technology, vol. 67, no. 4, pp. 2982-2996, April 2018, doi: 10.1109/TVT.2017.2766226.

[70] R. Gribonval, V. Cevher and M. E. Davies, "Compressible Distributions for High-Dimensional Statistics," in IEEE Transactions on Information Theory, vol. 58, no. 8, pp. 5016-5034, Aug. 2012.

[71] F. Gómez-Cuba and A. J. Goldsmith, "Compressed Sensing Channel Estimation for OFDM With Non-Gaussian Multipath Gains," in IEEE Transactions on Wireless Communications, vol. 19, no. 1, pp. 47-61, Jan. 2020.

[72] A. Chandra and C. Bose, “Error Probability for Coherent Modulations in Rician Fading Channel”, in International Journal of Interdisciplinary Telecommunications and Networking (IJITN), vol. 1, no. 2, pp. 16-27, Apr. 2009.

[73] E. Salahat and I. Abualhaol, "Generalized average BER expression for SC and MRC reciever over Nakagami-m fading channels," 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2013, pp. 3360-3365.

[74] I. Kakalou, D. Papadopoulou, T. Xifilidis, K. E. Psannis, K. Siakavara and Y. Ishibashi, "A survey on spectrum sensing algorithms for cognitive radio networks," 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), 2018, pp. 1-4.

[75] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," in IEEE Personal Communications, vol. 6, no. 4, pp. 13-18, Aug. 1999.

[76] A. Ali and W. Hamouda, "Advances on Spectrum Sensing for Cognitive Radio Networks: Theory and Applications," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1277-1304, Secondquarter 2017.

[77] I. Sobron, P. S. R. Diniz, W. A. Martins and M. Velez, "Energy Detection Technique for Adaptive Spectrum Sensing," in IEEE Transactions on Communications, vol. 63, no. 3, pp. 617-627, March 2015.

[78] S. Dikmese, P. C. Sofotasios, T. Ihalainen, M. Renfors and M. Valkama, "Efficient Energy Detection Methods for Spectrum Sensing Under Non-Flat Spectral Characteristics," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 5, pp. 755-770, May 2015.

[79] S. K. Sharma, E. Lagunas, S. Chatzinotas and B. Ottersten, "Application of Compressive Sensing in Cognitive Radio Communications: A Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1838-1860, thirdquarter 2016.

[80] X. Chen, H. Chen and W. Meng, "Cooperative Communications for Cognitive Radio Networks — From Theory to Applications," in IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1180-1192, Third Quarter 2014.

[81] S. Chaudhari and D. Cabric, "Power Control and Frequency Band Selection Policies for Underlay MIMO Cognitive Radio," in IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 2, pp. 304-317, June 2019.

[82] B. Jabbari, R. Pickholtz and M. Norton, "Dynamic spectrum access and management [Dynamic Spectrum Management]," in IEEE Wireless Communications, vol. 17, no. 4, pp. 6-15, August 2010.

[83] A. El Shafie, T. Khattab and A. Sultan Salem, "Relay-Assisted Primary and Secondary Transmissions in Cognitive Radio Networks," in IEEE Access, vol. 4, pp. 6386-6400, 2016.

[84] M. El Tanab and W. Hamouda, "Resource Allocation for Underlay Cognitive Radio Networks: A Survey," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1249-1276, Secondquarter 2017.

[85] W. Liang, S. X. Ng and L. Hanzo, "Cooperative Overlay Spectrum Access in Cognitive Radio Networks," in IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1924-1944, thirdquarter 2017.

[86] M. C. Filippou, D. Gesbert and G. A. Ropokis, "A Comparative Performance Analysis of Interweave and Underlay Multi-Antenna Cognitive Radio Networks," in IEEE Transactions on Wireless Communications, vol. 14, no. 5, pp. 2911-2925, May 2015.

[87] W. -L. Chin, "On the Noise Uncertainty for the Energy Detection of OFDM Signals," in IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 7593-7602, Aug. 2019.

[88] M. Derakhshani, T. Le-Ngoc and M. Nasiri-Kenari, "Efficient Cooperative Cyclostationary Spectrum Sensing in Cognitive Radios at Low SNR Regimes," in IEEE Transactions on Wireless Communications, vol. 10, no. 11, pp. 3754-3764, November 2011.

[89] S. Kapoor, S. Rao and G. Singh, "Opportunistic Spectrum Sensing by Employing Matched Filter in Cognitive Radio Network," 2011 International Conference on Communication Systems and Network Technologies, 2011, pp. 580-583.

[90] M. Iqbal and A. Ghafoor, "Analysis of Multiband Joint Detection Framework for Waveform-Based Sensing in Cognitive Radios," 2012 IEEE Vehicular Technology Conference (VTC Fall), 2012, pp. 1-5.

[91] Y. Zeng and Y. Liang, "Spectrum-Sensing Algorithms for Cognitive Radio Based on Statistical Covariances," in IEEE Transactions on Vehicular Technology, vol. 58, no. 4, pp. 1804-1815, May 2009.

[92] Y. Zeng and Y. -. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," in IEEE Transactions on Communications, vol. 57, no. 6, pp. 1784-1793, June 2009.

[93] H. Hosseini, S. K. B. Syed-Yusof, N. Fisal and A. Farzamnia, "Compressed Wavelet Packet-Based Spectrum Sensing With Adaptive Thresholding for Cognitive Radio," in Canadian Journal of Electrical and Computer Engineering, vol. 38, no. 1, pp. 31-36, winter 2015.

[94] G. Vazquez-Vilar, R. López-Valcarce, C. Mosquera and N. González-Prelcic, "Wideband spectral estimation from compressed measurements exploiting spectral a priori information in Cognitive Radio systems," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 2958-2961.

[95] J. Wang, Q. Li, H. Duan, J. Fang, Z. Ren and H. Li, "Multiantenna-Assisted Wideband Spectrum Sensing Based on Sub-Nyquist Sampling," in IEEE Wireless Communications Letters, vol. 10, no. 4, pp. 795-799, April 2021.

[96] P. Paysarvi-Hoseini and N. C. Beaulieu, "Optimal Wideband Spectrum Sensing Framework for Cognitive Radio Systems," in IEEE Transactions on Signal Processing, vol. 59, no. 3, pp. 1170-1182, March 2011.

[97] B. Adhikari and Jamadagni H. S., "Wideband spectrum sensing receiver for cognitive radio application," 2015 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2015, pp. 1-6.

[98] S. Deshmukh and S. Bhuyan, "Compressive sensing based energy efficient wideband cognitive radio network," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), 2018, pp. 137-142.

[99] M. Hasegawa, H. Hirai, K. Nagano, H. Harada and K. Aihara, "Optimization for Centralized and Decentralized Cognitive Radio Networks," in Proceedings of the IEEE, vol. 102, no. 4, pp. 574-584, April 2014.

[100] K. B. Letaief and W. Zhang, "Cooperative Communications for Cognitive Radio Networks," in Proceedings of the IEEE, vol. 97, no. 5, pp. 878-893, May 2009.

[101] L. Xu, H. Fang and Z. Lin, “Evolutionarily stable opportunistic spectrum access in cognitive radio networks,” in IET Communications, vol. 10, issue 7, pp. 2290-2299, Nov. 2016.

[102] M. Saeedzarandi, "Optimization of multiband sensing-time-adaptive detection in cognitive radio networks using artificial immune algorithm," 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), 2017, pp. 29-34.

[103] M. Bkassiny and S. K. Jayaweera, "Robust, Non-Gaussian Wideband Spectrum Sensing in Cognitive Radios," in IEEE Transactions on Wireless Communications, vol. 13, no. 11, pp. 6410-6421, Nov. 2014.

[104] H. Li et al., "Utility-Based Cooperative Spectrum Sensing Scheduling in Cognitive Radio Networks," in IEEE Transactions on Vehicular Technology, vol. 66, no. 1, pp. 645-655, Jan. 2017.

[105] P. C. Ng, "Optimization of spectrum sensing for cognitive sensor network using differential evolution approach in smart environment," 2015 IEEE 12th International Conference on Networking, Sensing and Control, 2015, pp. 592-596.

[106] X. Huang, F. Hu, J. Wu, H. Chen, G. Wang and T. Jiang, "Intelligent Cooperative Spectrum Sensing via Hierarchical Dirichlet Process in Cognitive Radio Networks," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 5, pp. 771-787, May 2015.

[107] B. Li, S. Li, A. Nallanathan, Y. Nan, C. Zhao and Z. Zhou, "Deep Sensing for Next-Generation Dynamic Spectrum Sharing: More Than Detecting the Occupancy State of Primary Spectrum," in IEEE Transactions on Communications, vol. 63, no. 7, pp. 2442-2457, July 2015.

[108] F. Gao, J. Li, T. Jiang and W. Chen, "Sensing and Recognition When Primary User Has Multiple Transmit Power Levels," in IEEE Transactions on Signal Processing, vol. 63, no. 10, pp. 2704-2717, May15, 2015.

[109] A. Salman, I. M. Qureshi, K. Sultan and S. Saeed, “Joint spectrum sensing for detection of primary users using cognitive relays with evolutionary computing,” in IET Communications, vol. 9, issue 13, pp. 1643-1648, Sept. 2015.

[110] I. Hwang and J. W. Lee, "Cooperative Spectrum Sensing With Quantization Combining Over Imperfect Feedback Channels," in IEEE Transactions on Signal Processing, vol. 65, no. 3, pp. 721-732, 1 Feb.1, 2017.

[111] A. Tsakmalis, S. Chatzinotas and B. Ottersten, "Interference Constraint Active Learning with Uncertain Feedback for Cognitive Radio Networks," in IEEE Transactions on Wireless Communications, vol. 16, no. 7, pp. 4654-4668, July 2017.

[112] H. Wu, F. Yao, Y. Chen, Y. Liu and T. Liang, "Multibit-Quantization-Based Collaborative Spectrum Sensing Scheme for Cognitive Sensor Networks," in IEEE Access, vol. 5, pp. 25207-25216, 2017.

[113] D. Cohen and Y. C. Eldar, "Sub-Nyquist Cyclostationary Detection for Cognitive Radio," in IEEE Transactions on Signal Processing, vol. 65, no. 11, pp. 3004-3019, 1 June1, 2017.

[114] D. Cohen and Y. C. Eldar, "Sub-Nyquist Sampling for Power Spectrum Sensing in Cognitive Radios: A Unified Approach," in IEEE Transactions on Signal Processing, vol. 62, no. 15, pp. 3897-3910, Aug.1, 2014.

[115] K. Cichoń, A. Kliks and H. Bogucka, "Energy-Efficient Cooperative Spectrum Sensing: A Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1861-1886, thirdquarter 2016.

[116] T. Xifilidis, K. E. Psannis, G. Minopoulos, G. Kokkonis and Y. Ishibashi, "Convolution Based Energy Detection Scheme for Cognitive Radio Systems," 2019 2nd World Symposium on Communication Engineering (WSCE), 2019, pp. 58-62.

[117] M. Abo-Zahhad, S. M. Ahmed, M. Farrag and K. A. BaAli, "Primary User Detection in Cognitive Radio Networks Over Fading Channel Using Compressed Sensing," 2018 International Japan-Africa Conference on Electronics, Communications and Computations (JAC-ECC), 2018, pp. 5-8.

[118] S. Singh and D. Mitra, "Implementation and evaluation of energy detector performance," 2017 International Conference on Communication and Signal Processing (ICCSP), 2017, pp. 1634-1640.

[119] R. Kishore, C. K. Ramesha, V. Sharma and R. Joshi, "Performance evaluation of energy based spectrum sensing in multipath fading channel for cognitive radio system," 2014 IEEE National Conference on Communication, Signal Processing and Networking (NCCSN), 2014, pp. 1-6.

[120] M. Ashraf, J. Khan, H. Rasheed, F. Ashraf, M. Faizan and M. I. Anis, "Demonstration of energy detector performance and spectrum sensing in Cognitive Radio using AWGN, Rayleigh and Nakagami channels," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), 2017, pp. 1-7.

[121] R. Yadav and S. K. Agrawal, "Performance Analysis of Spectrum Sensing of Cognitive Radio under Different Fading Environment," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 2015, pp. 558-561.

[122] Y. Arjoune, Z. E. Mrabet, H. E. Ghazi and A. Tamtaoui, "Spectrum sensing: Enhanced energy detection technique based on noise measurement," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 828-834.

[123] E. Lagunas, S. K. Sharma, S. Chatzinotas and B. Ottersten, "Compressive sensing based energy detector," 2016 24th European Signal Processing Conference (EUSIPCO), 2016, pp. 1678-1682.

[124] T. Xifilidis and K.E. Psannis, “Correlation-based wireless sensor networks performance: the compressed sensing paradigm,” Springer Cluster Computing, vol. 25, issue 2, pp. 965-981, Nov. 2021.

[125] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless sensor networks: a survey, " Elsevier Computer Networks, vol. 38, issue 4, pp. 393-422, March 2002.

[126] J. Yick, B. Mukherjee and D. Ghosal, "Wireless sensor network survey, "Elsevier Computer Networks, vol. 52, issue 12, pp. 2292-2330, Aug. 2008.

[127] G. Anastasi, M. Conti, M. Di Francesco and A. Passarella, " Energy conservation in wireless sensor networks: A survey, " Elsevier Ad Hoc Networks, vol. 7, issue 3, pp. 537-568, May 2009.

[128] S. N. Das, S. Misra, B. E. Wolfinger and M. S. Obaidat, "Temporal-Correlation-Aware Dynamic Self-Management of Wireless Sensor Networks," in IEEE Transactions on Industrial Informatics, vol. 12, no. 6, pp. 2127-2138, Dec. 2016.

[129] G. A. Shah and M. Bozyigit, "Exploiting Energy-aware Spatial Correlation in Wireless Sensor Networks," 2007 2nd International Conference on Communication Systems Software and Middleware, 2007, pp. 1-6.

[130] R. Maivizhi and P. Yogesh, "Spatial Correlation based Data Redundancy Elimination for Data Aggregation in Wireless Sensor Networks," 2020 International Conference on Innovative Trends in Information Technology (ICITIIT), 2020, pp. 1-5.

[131] C. Liu and G. Cao, "Spatial-Temporal Coverage Optimization in Wireless Sensor Networks," in IEEE Transactions on Mobile Computing, vol. 10, no. 4, pp. 465-478, April 2011.

[132] J. Chen, T. Li, J. Wang and C. W. d. Silva, "WSN Sampling Optimization for Signal Reconstruction Using Spatiotemporal Autoencoder," in IEEE Sensors Journal, vol. 20, no. 23, pp. 14290-14301, 1 Dec.1, 2020.

[133] G. B. Tayeh, A. Makhoul, C. Perera and J. Demerjian, "A Spatial-Temporal Correlation Approach for Data Reduction in Cluster-Based Sensor Networks," in IEEE Access, vol. 7, pp. 50669-50680, 2019.

[134] G. E. P. Kumar, K. Baskaran, R. E. Blessing and M. Lydia," A Comprehensive Review on the Impact of Compressed Sensing in Wireless Sensor Networks," in International Journal on Smart Sensing and Intelligent Systems, vol. 9, no. 2, pp. 818-844, June 2016.

[135] A. S. Alwakeel, M. F. Abdelkader, K. G. Seddik and A. Ghuniem, "Exploiting Temporal Correlation of Sparse Signals in Wireless Sensor Networks," 2014 IEEE 79th Vehicular Technology Conference (VTC Spring), 2014, pp. 1-6.

[136] H. Hu and Z. Yang, "Spatial Correlation-Based Distributed Compressed Sensing in Wireless Sensor Networks," 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), 2010, pp. 1-4.

[137] T. Wimalajeewa and P. K. Varshney, "Robust detection of random events with spatially correlated data in wireless sensor networks via distributed compressive sensing," 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2017, pp. 1-5.

[138] C. Zhang, O. Li, X. Tong, K. Ke and M. Li, "Spatiotemporal Data Gathering Based on Compressive Sensing in WSNs," in IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1252-1255, Aug. 2019.

[139] M. Hooshmand, M. Rossi, D. Zordan and M. Zorzi, "Covariogram-Based Compressive Sensing for Environmental Wireless Sensor Networks," in IEEE Sensors Journal, vol. 16, no. 6, pp. 1716-1729, March15, 2016.

[140] N. Deligiannis, J. F. C. Mota, E. Zimos and M. R. D. Rodrigues, "Heterogeneous Networked Data Recovery From Compressive Measurements Using a Copula Prior," in IEEE Transactions on Communications, vol. 65, no. 12, pp. 5333-5347, Dec. 2017.

[141] R. Seri, "A Tight Bound on the Distance Between a Noncentral Chi Square and a Normal Distribution," in IEEE Communications Letters, vol. 19, no. 11, pp. 1877-1880, Nov. 2015.

[142] S. Xiao, T. Li, Y. Yan and J. Zhuang, " Compressed sensing in wireless sensor networks under complex conditions of Internet of things," in Springer Cluster Computing, vol. 22, issue 6, pp. 14145-14155, Febr. 2018.

[143] F. M. Awaysheh, M. Alazab, S. Garg, D. Niyato and C. Verikoukis, "Big Data Resource Management & Networks: Taxonomy, Survey, and Future Directions," in IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2098-2130, Fourthquarter 2021.

[144] M. A. Maddah-Ali and U. Niesen, "Fundamental Limits of Caching," in IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2856-2867, May 2014.

[145] T. Xifilidis and K. E. Psannis, " Caching hit probability and Compressive Sensing perspective for mobile cellular networks," Elsevier Simulation Modelling Practice and Theory, vol. 87, pp. 92-98, Sept. 2018.

[146] J. Q. Bao and L. Tong, "A performance comparison of CDMA ad-hoc and cellular networks," Globecom '00 - IEEE. Global Telecommunications Conference. Conference Record (Cat. No.00CH37137), 2000, pp. 208-212 vol.1.

[147] H. A. Pedersen, L. Sassatelli and R. Aparicio-Pardo, "Random linear coded distributed caching for video streaming over D2D," 2018 14th Annual Conference on Wireless On-demand Network Systems and Services (WONS), 2018, pp. 1-8.

[148] N. Al-Falahy and O. Y. Alani, "Technologies for 5G Networks: Challenges and Opportunities," in IT Professional, vol. 19, no. 1, pp. 12-20, Jan.-Feb. 2017.

[149] S. Kilaru, K. Harikishore, T. Sravani, C. L. Anvesh and T. Balaji, "Review and analysis of promising technologies with respect to Fifth generation networks," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), 2014, pp. 248-251.

[150] A. Gharaibeh, A. Kreishah and I. Khalil, " An O(1)- Competitive Online Caching Algorithm from Content Centric Networking," 2016 35th Annual IEEE International Conference on Computer Communications (INFOCOM), pp. 1-9.

[151] J. Shiral, " A Comparative Study on Cellular Sensor and Ad-Hoc Networks", in International Journal of Advanced Smart Sensor Network Systems, vol. 2, no. 3, pp. 43-50, July 2012.

[152] M. N. Tehrani, M. Uysal and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions," in IEEE Communications Magazine, vol. 52, no. 5, pp. 86-92, May 2014.

[153] R. Wang, J. Zhang, S. H. Song and K. B. Letaief, "Mobility-Aware Caching in D2D Networks," in IEEE Transactions on Wireless Communications, vol. 16, no. 8, pp. 5001-5015, Aug. 2017.

[154] X. Liu, J. Zhang, X. Zhang and W. Wang, "Mobility-Aware Coded Probabilistic Caching Scheme for MEC-Enabled Small Cell Networks," in IEEE Access, vol. 5, pp. 17824-17833, 2017.

[155] S. Krishnan and H. S. Dhillon, "Effect of User Mobility on the Performance of Device-to-Device Networks With Distributed Caching," in IEEE Wireless Communications Letters, vol. 6, no. 2, pp. 194-197, April 2017.

[156] J. Wen, K. Huang, S. Yang and V. O. K. Li, "Cache-Enabled Heterogeneous Cellular Networks: Optimal Tier-Level Content Placement," in IEEE Transactions on Wireless Communications, vol. 16, no. 9, pp. 5939-5952, Sept. 2017.

[157] L. Zhang, M. Xiao, G. Wu and S. Li, "Efficient Scheduling and Power Allocation for D2D-Assisted Wireless Caching Networks," in IEEE Transactions on Communications, vol. 64, no. 6, pp. 2438-2452, June 2016.

[158] L. Breslau, Pei Cao, Li Fan, G. Phillips and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," IEEE INFOCOM '99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the

IEEE Computer and Communications Societies. The Future is Now (Cat. No.99CH36320), 1999, pp. 126-134 vol.1.

[159] C. Stergiou, K. E. Psannis, T. Xifilidis, A. P. Plageras and B. B. Gupta, "Security and privacy of big data for social networking services in cloud," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2018, pp. 438-443.

[160] S. El-Seoud, H. F. El-Sofany and H. Abdelfattah and R. Mohamed, " Big Data and Cloud Computing: Trends and Challenges,", in International Journal of Interactive Mobile Technologies, vol. 11, no. 2, pp. 34-52, April 2017.

[161] H. Gao, J. Hu, T. Huang, J. Wang and Y. Chen, "Security Issues in Online Social Networks," in IEEE Internet Computing, vol. 15, no. 4, pp. 56-63, July-Aug. 2011.

[162] Y. Zhang, P. Wang, L. Fang, X. He, H. Han and B. Chen, "Secure Transmission of Compressed Sampling Data Using Edge Clouds," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6641-6651, Oct. 2020.

[163] T. Xifilidis and K. E. Psannis, "Practicing Medicine: A Compressed Sensing Approach," in Proceedings of New Technologies in Health: Medical, Legal and Ethical Issues Conference, 21-22 November 2019, Thessaloniki, Greece.

[164] A. Siddique, O. Hasan, F. Khalid and M. Shafique, " Approxcs: near-sensor approximate compressed sensing for IOT-healthcare systems," arXiv: Signal Processing, Nov. 2018.

[165] X. Yuan, X. Wang, C. Wang, J. Weng and K. Ren, "Enabling Secure and Fast Indexing for Privacy-Assured Healthcare Monitoring via Compressive Sensing," in IEEE Transactions on Multimedia, vol. 18, no. 10, pp. 2002-2014, Oct. 2016.

[166] C. Stergiou, A. P. Plageras, K. E. Psannis, T. Xifilidis, G. Kokkonis, S. Kontogiannis, K. Tsarava and A. Sapountzi, "Proposed High Level Architecture of a Smart Interconnected Interactive Classroom," 2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM), 2018, pp. 1-6.

[167] Tsarava et al., " Training computational thinking: Game-based unplugged and plugged-in activities in primary school," in European Conference on Games based Learning, pp. 687-695.

[168] S. Valsamidis et al., " A Clustering Methodology of Web Log Data for Learning Management Systems", in Journal of Educational Technology & Society, vol 15, no. 2, pp. 154-167, 2012.

[169] E. Gounopoulos, S. Kontogiannis, I. Kazanidis and S. Valsamidis, " A framework for the evaluation of multilayer web based learning," In Proceedings of the 20th Pan-Hellenic Conference on Informatics, pp. 1-4, 2016.

[170] E. Gounopoulos, S. Kontogiannis, S. Valsamidis and I. Kazanidis, " Blended Learning Evaluation in Higher education courses," KnE Social Sciences, vol. 1, no. 2, pp. 385-399, March 2017.

[171] M. Agiwal, A. Roy and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617-1655, thirdquarter 2016.

[172] J. G. Andrews et al., "What Will 5G Be?," in IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065-1082, June 2014.

[173] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar and L. Hanzo, "Millimeter-Wave Communications: Physical Channel Models, Design Considerations, Antenna Constructions, and Link-Budget," in IEEE Communications Surveys & Tutorials, vol. 20, no. 2, pp. 870-913, Secondquarter 2018.

[174] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai and J. Rodriguez, "Millimeter-Wave Massive MIMO Communication for Future Wireless Systems: A Survey," in IEEE Communications Surveys & Tutorials, vol. 20, no. 2, pp. 836-869, Secondquarter 2018.

[175] M. A. Albreem, M. Juntti and S. Shahabuddin, "Massive MIMO Detection Techniques: A Survey," in IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3109-3132, Fourthquarter 2019.

[176] G. Ding et al., "Spectrum Inference in Cognitive Radio Networks: Algorithms and Applications," in IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 150-182, Firstquarter 2018.

[177] F. Hu, B. Chen and K. Zhu, "Full Spectrum Sharing in Cognitive Radio Networks Toward 5G: A Survey," in IEEE Access, vol. 6, pp. 15754-15776, 2018.

[178] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G," in IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 2294-2323, thirdquarter 2018.

[179] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang and Z. Ding, "Interplay Between NOMA and Other Emerging Technologies: A Survey," in IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 4, pp. 900-919, Dec. 2019.

[180] S. Zhang, Q. Wu, S. Xu and G. Y. Li, "Fundamental Green Tradeoffs: Progresses, Challenges, and Impacts on 5G Networks," in IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 33-56, Firstquarter 2017.

[181] Y. Mao, C. You, J. Zhang, K. Huang and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2322-2358, Fourthquarter 2017.

[182] J. Yao, T. Han and N. Ansari, "On Mobile Edge Caching," in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2525-2553, thirdquarter 2019.