Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA, INFORMATICA E
DELLE TELECOMUNICAZIONI

Ciclo XXVII

Settore Concorsuale di Afferenza: 09/F2 -
TELECOMUNICAZIONI
Settore Scientifico Disciplinare: ING-INF/03 -
TELECOMUNICAZIONI

TITOLO TESI

ENERGY-EFFICIENT COMMUNICATION AND ESTIMATION IN WIRELESS SENSOR NETWORKS

Presentata da: VINCENZO ZAMBIANCHI

Coordinatore Dottorato:
Chiar.mo Prof. Ing. ALESSANDRO VANELLI CORALLI

Relatore:
Chiar.mo Prof. Ing. DAVIDE DARDARI

Correlatore:
Chiar.mo Prof. Ing. MARCO CHIANI

Esame Finale Anno 2015

Keywords

low-power wireless communications
peak power limited input channels
information theory
confidence regions
estimation theory

A chi mi regala sorrisi e a chi continua a farlo da lassú

Contents

Sommario

Questa tesi è centrata sugli aspetti di efficienza energetica nelle reti wireless, sia per quanto riguarda la trasmissione, sia per quanto concerne la diffusione dell'informazione in reti complesse. In particolare, se da un lato si analizza l'efficienza della comunicazione, puntando a ridurre i consumi sul fronte trasmissione, dall'altro non si trascura che i processi e gli algoritmi richiedenti l'accesso al mezzo di comunicazione debbano essi stessi essere efficienti, minimizzando quindi la domanda di traffico.

In tema di trasmissioni energeticamente efficienti si introduce uno schema a riuso di segnali di opportunità, finora mai studiati in letteratura a scopo di comunicazione, con l'obiettivo di arrivare ad un consumo prossimo allo zero. A livello teorico si generalizza il tema dei segnali di trasmissione a bassa potenza considerando modelli di canale con segnali di ingresso limitati in ampiezza. Si analizzano quindi le caratteristiche che tali canali devono possedere affinché la distribuzione di ingresso, che garantisce il raggiungimento della capacità, sia discreta.

Per quanto riguarda il progetto di algoritmi efficienti di diffusione dell'informazione, ci si focalizza su un problema di stima decentralizzata su una rete di sensori wireless, risolto ponendo l'enfasi sull'efficienza energetica delle possibili soluzioni e sulla robustezza delle stesse anche in presenza di perdite (in generale troncamento) nelle comunicazioni.

Résumé

Cette thèse porte sur les aspects d'efficacité énergétique, soit pour ce qui concerne la transmission soit pour ce qui concerne la propagation de l'information sur des reseaux sans fils complexes. En particulier, d'un côté, on analyse l'efficacité de la communication étendue vers la reduction des consommations pour l'émetteur et, de l'autre côté, on pose l'attention sur l'efficacité des algorithmes et des procès qui doivent être eux-mêmes le moins exigeants possible en termes de trafic nécessaire.

Dans le domaine des communications efficaces, on introduit un nouveau méthode exploitant des signaux déjà présents dans l' environnement, dits signaux d'opportunité, qui n'ont jamais été étudiés dans la literature aux fins de la communication. Le scope est d'ensurer des consommations proches de zero. Ensuite, la thématique des signaux de transmission de basse puissance est généralisée: on analyse les conditions sous lesquelles les chanaux de communication, ayant le signal d'entrée limité en puissance maxime, présentent une distribution d'entrée optimale (ainsi atteignant la capacité) discrète.

À l'égard du projet d'algorithmes efficaces pour la propagation de l'information, on se concentre sur un problème d'estimation distribué sur un réseau sans fils, résolu en posant l'accent sur l'efficacité énergétique des solutions possibles et sur leur degré de résistance face à des pertes lors des communications (soit en présence d'une troncature plus en général).

Summary

This thesis focuses on the energy efficiency in wireless networks under the transmission and information diffusion points of view. In particular, on one hand, the communication efficiency is investigated, attempting to reduce the consumption during transmissions, while on the other hand the energy efficiency of the procedures required to distribute the information among wireless nodes in complex networks is taken into account.

For what concerns energy efficient communications, an innovative transmission scheme reusing source of opportunity signals is introduced. This kind of signals has never been previously studied in literature for communication purposes. The scope is to provide a way for transmitting information with energy consumption close to zero. On the theoretical side, starting from a general communication channel model subject to a limited input amplitude, the theme of low power transmission signals is tackled under the perspective of stating sufficient conditions for the capacity achieving input distribution to be discrete.

Finally, the focus is shifted towards the design of energy efficient algorithms for the diffusion of information. In particular, the endeavours are aimed at solving an estimation problem distributed over a wireless sensor network. The proposed solutions are deeply analyzed both to ensure their energy efficiency and to guarantee their robustness against losses during the diffusion of information (against information diffusion truncation more in general).

Introduction

0.1 Motivation

The majority of recent advances in the field of communication theory and distributed sensing is due to the rise of the new challenges posed by the transformation of the one to one paradigm into the many to many paradigm. Multiple input multiple output (MIMO) communications [1–3], spectrum reuse and cognitive radio [4], and distributed detection and estimation algorithms in wireless sensor networks (WSNs) [5,6] are perhaps the most prominent examples of this tendency. Nonetheless, literature on WSNs has seen lots of efforts being attempted in many other directions. For instance, the necessity for improved localization capabilities, network lifetime extension and, in general, energy efficiency in the use of network resources have led to the development of ultrawide bandwidth (UWB) localization and communication techniques [7–9], energy efficient routing [10], and ad hoc channel access protocols [11, 12], just to cite a few. Moreover, since the pioneering work [13] was published, a rising interest has been devoted to collaborative distributed schemes aiming to fulfill tasks shared among the several nodes forming a WSN.

Energy efficiency is on of the most important topic in the WSNs literature. An ever growing attention has been driven by green communication and computation systems, which represent the future of the information and communication technology (ICT) world. This vision is supported by the emergence of Internet of things (IoT) applications [14,15], which require WSNs as an enabling technology under the constraint that devices consumption is really low. This stringent requirement should foster the opportunity to employ energy-harvesting

technologies [16–19] and reduce the strong environmental impact of batteries. Two are the main declensions for the energy efficiency in WSNs. On one hand, new communication paradigms can gain in efficiency by requiring less transmit power. On the other hand, the tasks to which WSNs are dedicated should be less transmissions-eager and they should require a lower computational burden as well.

To summarize, one might well say that the lowering of the energy required by data transmissions is the main way for achieving energy efficiency. The strong interest for lowering power consumption in wireless communications is not limited to its intrinsic benefits but also derives from the positive effects it has on electromagnetic (EM) pollution and on the reduction of interference levels. Specifically, this allows for a better spatial reuse of spectrum as well as for a less troublesome coexistence of primary and secondary systems on the same bandwidth. This means that pushing in the direction of autonomous systems, able to harvest the small required amount of energy from the surrounding environment, is profitable in many aspects and particularly could enforce the massive deployment of IoT applications. As for what concerns currently available technologies, modern radiofrequency identification (RFID) systems do provide tags with very low energy consumption and working without batteries [20]. However if one includes also the RFID reader in the energy budget computation, it could be seen that the overall system efficiency dramatically decreases. In fact, to energize the tags, the reader sends high power (in the order of Watts) interrogation signals but only a fractional portion of such energy is actually captured by tags while the rest is wasted in the environment. This makes the overall energy efficiency of RFID systems much less than the one that would be obtained using batteries in tags. In this context, the present work proposes a solution that lowers the overall system energy consumption by "recycling" signals already present in the scenario as it will be later explained.

One of the most common task that a WSN may be in charge of is the estimation of a (typically vector) parameter, generally representing a spatially distributed physical quantity. Subsequently some decisions are taken based on the acquired estimate. If all the devices in the WSN are potential decision taking entities they all should be ensured with the required estimate. Classically, this issue is solved

by letting a central unit collect all data, compute the estimate and transmit it back to nodes. Unfortunately, centralized solutions have some drawbacks related to reliability and traffic load: The central unit may fail, thus provoking an utter network outage, while the traffic necessary for back transmissions may be overwhelming. This is where the dichotomy between centralized and distributed approaches comes from. Typically, a distributed solution would cancel the central unit failure problem. However, this problem is generally solved mostly at the expenses of efficiency in communications. In fact, the required traffic load should be carefully considered, since, on one hand, it is more costly to distribute information towards many (possibly all) nodes in a network than gathering it at a single node while, on the other hand, distributed solutions do not need any final redistribution of the produced estimate.

In literature many results in distributed estimation theory can be found. Precisely, one may observe how the interest for efficient and robust information diffusion strategies has steadily been increasing in the last two decades. Nonetheless, the need for robustness has perhaps drifted the researchers attention mostly on solving the central unit failure problem. Hence, the preference for peer-to-peer like solutions has arisen. For instance, consensus algorithms perform really well in guaranteeing that all nodes in a WSN are able to fulfill an identical estimation task. However, in literature, the comparison of consensus based algorithms with completely unrouted diffusion strategies is somehow flawed. Generally, one may find comparisons with the backwards transmission of centrally computed quantities (*e.g.*, in [21, 22]) but this is still not enough to claim for efficiency in a satisfying way. Moreover, the literature is not so keen in the investigation of the effects of truncation in distributed procedures. Some issues caused by losses in communications are addressed in, *e.g.*, [23,24], however many solutions to distributed problems are completely missing this aspect. This is an important issue that should always be considered when developing a distributed procedure. Finally, the accent in literature is on determining estimates in a distributed way. Nevertheless, the estimate is not always enough to take a subsequent decision. Sometimes, like in source localization problems, it is more important to have information about the probability with which the parameter of

interest (*e.g.*, the true source position) lies in a region that surrounds the obtained estimate. This gives rise to the characterization of confidence regions. One can obtain a good characterization of confidence regions for an infinite number of measurements by resorting to classical tools such as those based on the Cramér-Rao bound. Unfortunately such tools do not provide useful information (they are too optimistic) when the number of measurements is limited as happens, for example, when delays or energy constraints are present in the network. This aspect is investigated in this thesis where distributed approaches for non-asymptotic confidence regions computation are proposed.

0.2 Thesis Outline

This thesis is mostly focused on energy efficiency in WSNs. This topic was chosen as a consequence of the growing attention driven by the need for green communications in WSNs and more generally in the ICT world, as sketched in the previous section. Energy efficiency is considered both for what concerns transmission and information diffusion.

An innovative physical layer scheme for data transmissions is proposed. This scheme aims at lowering the required power on the transmitter side. The contribution lies in the design and analysis of an almost zero-power transmission approach. The main underlying idea is to exploit signals already available in the surrounding environment as transmission carriers instead of generating ad hoc signals, as done in conventional wireless communication systems. These signals are modulated by means of backscatter techniques, yielding a significant advantage for the transmitter in terms of energy consumption. This principle has already been exploited for localization purposes [25, 26], but it has never been adopted to support data communication. The analysis is carried out in terms of achievable transmission range and bit rate. Such a kind of communication approach might find applications in low power monitoring devices, from environmental to industrial scenarios. Chapter 1 covers the investigation of this innovative communication scheme and reports results that have been published in [27, 28]. The paper [28] received the "F. Carassa" best paper award

at the 2013 annual GTTI (associazione Gruppo nazionale Telecomunicazioni e Tecnologie dell'Informazione) meeting.

The conducted analysis is successively enlarged in Chapter 2 to peak power limited transmission schemes from an information theoretic point of view. A pretty general channel model is defined and the attention is drawn to the investigation of sufficient conditions for the capacity achieving input probability measures to be discrete with a finite number of probability mass points. These conditions are shown to be met in interesting case studies such as the peak power limited Rayleigh fading channel and additive noise peak power limited channels. In fact, all feasible communication schemes are limited in peak power and this clearly sets the ensemble of target applications as really vast. The literature on the topic of peak power limited channels is not so large and covers channel models on a case-by-case basis. Here the approach that is taken aims at abstracting from the particular channel model, and concentrates on the set of conditions that should be met. Results have been published/submitted in [29, 30].

The second aspect of energy efficiency is covered in the last chapter of the thesis. It is worth analyzing how different strategies for diffusing information across a WSN impact on the yielded traffic load. The focus is on lowering the eagerness for transmissions. To this purpose, a distributed problem in estimation theory is presented and solved, accounting for both an analysis of efficiency in terms of required communication efforts and robustness to truncation. Specifically, Chapter 3 investigates the problem of the efficient computation of confidence regions over WSNs. This differs from literature approaches focused on point estimation and also from classical, centralized methods for the computation of confidence regions. A state-of-the-art centralized method for the computation of non-asymptotic confidence regions is extended to the distributed scenario. Several solutions for the diffusion of information across the network are designed and compared for different network topologies and dimensions. The distributed computation of confidence regions may find application in a variety of scenarios, ranging from source localization to metering and monitoring in smart grids as well as all the contexts in which the mere point estimates are inadequate. The results presented in Chapter 3 have been published/submitted in [31–33].

Part I

Energy efficient wireless communications

Chapter 1

Opportunistic almost Zero-Power Communications

1.1 Introduction

In recent years, a great effort and attention has been devoted to backscatter communications. This has grown in importance as the interest in low power, or ideally zero-power, consuming devices has been a driving concern. In particular, after its introduction [20, 34], backscatter modulation has emerged as one of the best strategies to perform identification in low complexity, low power, passive or semi-passive RFID devices. Far more recently, UWB technology has been combined with backscatter modulation to integrate identification with high-definition localization accuracy capabilities [8,35–37]. Parallel to this, a significant impulse to backscatter modulation technology has been driven by sensor applications, in which the influence of the observed physical characteristics on the sensor antenna radiation properties is the feature exploited for information conveying [38].

In current implementations, backscatter based communication requires the presence of an interrogating device (receiver) that provides typically a high power radio frequency (RF) signal used by sensors/tags as carrier for information transmission as well as power source.

The idea investigated in this chapter is to overtake the need for an interrogation signal, *i.e.*, the transmission of an ad hoc radiofre-

quency (RF) carrier, by recycling signals already present in the environment as carrier for sensor-to-receiver data transmissions. WiMax and Wi-Fi systems are two examples of widely diffused systems that are broadcasting RF signals that may be reused to the scope. These signals are referred to as source of opportunity (SoO) signals. The exploitation of SoO signals is not a totally new concept: It has been already studied for the design of infrastructure-less terrestrial positioning systems [25, 26]. However, there is no evidence in literature that SoO signals have ever been considered for communication purposes. The modulation scheme adopted by sensor nodes is the same as the one used for passive tags/sensors in RFID applications, *i.e.*, antenna load modulation causing power level variations of the signal backscattered by sensors (backscatter modulation). Due to the opportunistic nature of the proposed communication technique, it is named "*piggyback communication*".

It is necessary to underline how the here proposed nomenclature is clearly distinct from the previous uses of "piggybacking": In literature, this terminology may be found to denote a network layer/protocol mechanism that achieves overhead minimization in the acknowledgement transmissions by exploiting an overwriting in packets to be sent, while in this context it denotes a physical layer procedure.

An application scenario may be the one depicted in Fig. 1.1, where a certain number of passive or semi-passive wireless sensors, deployed in a small area, transmit their sensed data to a receiver (collector). The sensors can modulate the SoO signal that is reflected back to the receiver, thus creating short-range communication links without additional power emissions. Piggyback communication is expected to lead to prominent improvements in terms of both energetic cost and system security, the latter being a fundamental issue for instance in military applications. In fact, since communication occurs without the emission of dedicated wireless power in the environment, it is not easily detectable. The largely improved energy efficiency is expected to be, on the other hand, a driving factor for the adoption of the proposed communication technique in body area networks (BANs) or WSNs applications. The main concern is clearly the feasibility of piggyback communications in terms of link budget, data rate performance, and interference issues.

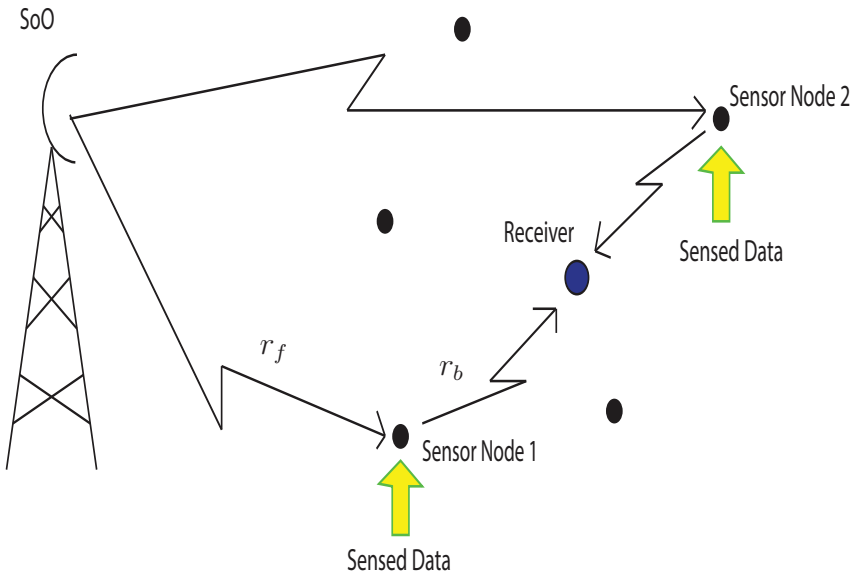

Figure 1.1: Example of application scenario using communications with SoO signals.

Hereby, an investigation of the piggyback communication concept is presented with the purpose to assess its feasibility from a link budget as well as from an information theoretic point of view. In particular, it can be shown that communication using SoO signals leads to the definition of a peculiar discrete-time channel model starting from which some theoretical considerations can be made considering a typical WSN scenario.

1.2 Piggyback Communications

The scope of this section is to describe the working principle of a piggyback communication scheme. Consider the simplified scenario, shown in Fig. 1.2, where a SoO transmitter, a sensor and a receiver are present with respective distances r_f and r_b, typically with $r_f \gg r_b$ (see also Fig. 1.1). Taking without loss of generality as reference time the receiver local time, denote by $V_S\, z(t+\tau_1+\tau_2)$ the complex equivalent baseband signal emitted by the SoO transmitter, where τ_1 and τ_2

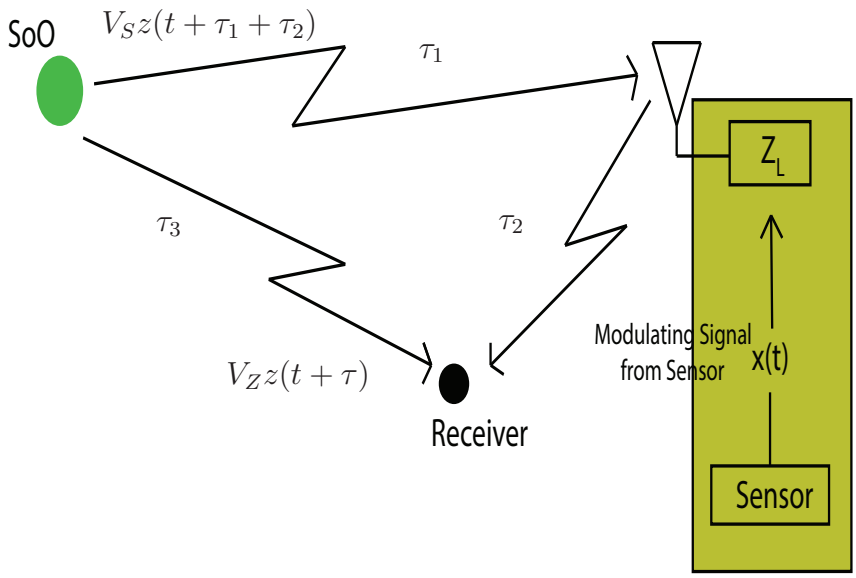

Figure 1.2: The piggyback communication mechanism.

are the SoO-sensor and sensor-receiver signals time-of-flight (TOF), respectively, and where, for further convenience, $z(t)$ is normalized so that $< |z(t)|^2 >= 1$.

The SoO signal can be received directly by the receiver with amplitude V_Z, also depending on the antenna pattern shape, after $\tau = \tau_1 + \tau_2 - \tau_3$ seconds, being τ_3 the TOF between the SoO transmitter and the receiver. In addition, the SoO signal is backscattered by the sensor according to the sensor antenna load conditions. Specifically, extending the approach in [39], the complex equivalent baseband received signal is

$$y(t) = V_0 \left(A_s - \rho_{\text{rcs}}(t)\right) z(t) + V_Z \, z(t+\tau) + n(t) \tag{1.1}$$

where $n(t)$ is the additive white Gaussian noise (AWGN), independent of $z(t)$, with bilateral power spectral density $N_0/2$, A_s is a load-independent coefficient, related to the current induced on the antenna conducting surface by the incident wave, and $\rho_{\text{rcs}}(t)$ is the antenna

load reflection coefficient given by

$$\rho_{\text{rcs}}(t) = \frac{Z_L(t) - Z_A^*}{Z_L(t) + Z_A} \tag{1.2}$$

Z_A and Z_L being the sensor antenna and sensor load impedances, respectively.[1] The constant V_0 is related to the received power as

$$V_0 = \sqrt{\frac{2\, R_A\, P_{R_0}}{A_s^2}}\exp\{j\phi\} \tag{1.3}$$

where ϕ is the phase offset, R_A the antenna resistance, and

$$P_{R_0} = \frac{\mathsf{EIRP}}{4\pi r_f^2}\sigma_0\frac{\lambda^2}{(4\pi)^2 r_b^2}G_R \tag{1.4}$$

the RF received power, in free space conditions, evaluated in matching load condition, *i.e.*, $Z_L = Z_A^*$ ($\rightarrow \rho_{\text{rcs}}(t) = 0$). The parameter σ_0 represents the sensor radar cross section (RCS) under matched load condition, *i.e.*,

$$\sigma_0 = \frac{\lambda^2}{4\pi}\, K\, G_t^2 \tag{1.5}$$

where K is the modulation factor, defined as $K = |A_s - \rho_{\text{rcs}}(t)|^2$ [39] and then coincident with $|A_s|^2$ for a matched load ($\rho_{\text{rcs}}(t) = 0$), and G_t is the sensor antenna gain in the direction of interest. In (1.4) and (1.5), EIRP is the transmitter equivalent isotropic radiated power, G_R is the receiver antenna gain, and λ is the incident wavelength.

To make information transmission possible, the antenna load $Z_L(t)$ changes with time according to the data to be transmitted. As a consequence, also the reflection coefficient $\rho_{\text{rcs}}(t)$ is time-dependent. Denoting by $x(t)$ the modulating signal, *i.e.*, the piggyback channel input carrying data generated by the sensor, in the following analysis consider $x(t) = -\rho_{\text{rcs}}(t)$. Since $|\rho_{\text{rcs}}(t)| \leq 1$, then the channel input results both power and amplitude limited, *i.e.*, $|x(t)| \leq 1$. This derives by the passive nature (no signal amplification) of piggyback communication. For example, in UWB RFID systems proposed in [8], $x(t)$ takes two values, *i.e.*, $\{-1, +1\}$.

[1] The superscript * indicates complex conjugation.

Rearranging, (1.1) may then be written as

$$y(t) = V_0\, z(t)\, x(t) + V_0\, A_s\, z(t) + V_Z\, z(t+\tau) + n(t)\,. \tag{1.6}$$

From (1.6) one can note that the signal backscattered by the sensor antenna is composed of a component (antenna mode) dependent on the antenna load, and hence on the input signal $x(t)$, and of a second component (structural mode) independent of the load and determined by the physical structure of the antenna. Denote by $\eta = \frac{|V_Z|^2}{|V_0|^2}$ the ratio between the direct SoO-receiver signal and the antenna mode signal square amplitudes. Unlike in RFID systems analyzed in [39], here the carrier is not an a priori known sinusoidal wave, but is given by the random SoO signal $z(t)$: Therefore the useful component is determined by the multiplication between $z(t)$ and $x(t)$, assuming the antenna and load reflection properties constant within the SoO signal bandwidth.

1.3 Channel Model

From the presented received signal expression one can derive a discrete time channel model. Considering a sampling time $T_s = 1/B$, with B the RF bandwidth of $z(t)$, the discrete-time channel model associated to (1.6) is

$$y_n = V_0\, z_n\, x_n + w_n \tag{1.7}$$

where $y_n = y(n\,T_s)$, $z_n = z(n\,T_s)$, and all not data dependent components are included in the term $w_n = w(n\,T_s)$, with $w(t) = V_Z\, z_1(t) + V_0\, A_s\, z(t) + n(t)$, $z_1(t) = z(t+\tau)$.

One can model the complex equivalent baseband SoO signal as a circular complex normal (CN) random process with bandwidth B and white spectral density. Indeed, this is known to be an accurate approximation for orthogonal frequency division multiplexing (OFDM) signals [40], that is the most common multiple access technology employed by the earlier mentioned SoO signals. Therefore, assuming the channel input as memoryless, one can overlook the time dependence in (1.7) and write simply

$$Y = V_0\, Z\, X + W \tag{1.8}$$

with $Z \sim CN(0, \sigma_Z^2)$ and $W \sim CN(0, \sigma_W^2)$. Conditioned on the input X, the output Y is also a circular CN distributed random variable (RV) because Z and W are jointly circular CN RVs. As already stated, the channel input X is amplitude constrained, *i.e.*, $|X| \leq 1$.

The derived discrete-time channel model (1.8) shares some similarities with previously studied ones [41–43]. Specifically, [41,42] studied an additive scalar Gaussian channel $Y = X+N$, with input X subject to an amplitude constraint. In that case, the channel capacity is reached by a unique discrete input distribution with a finite number of probability mass points. Besides, in [43] a similar approach is followed in studying the multiplicative Rayleigh fading channel $Y = V\,X+N$ with power constraint, V and N, being independent Gaussian multiplicative and additive RVs, respectively. It was proven that the capacity achieving input distribution is again discrete with a finite number of probability mass points, one of them being located at zero. It was also showed that, at low signal-to-noise ratio (SNR), the capacity achieving input distribution has only two mass points: One located at zero, the other one having amplitude which decreases with the power constraint. On the other hand, at high SNR, more than two points are necessary but the mutual information is not so sensitive to their exact locations, then values close to capacity can be achieved even without a perfect positioning of the input mass points.

In Chapter 2 a wider analysis of sufficient conditions under which peak power limited channels have a capacity achieving real input probability measure, taking values on a finite number of probability mass points, will follow. However, it is not possible to claim that the piggyback channel model is covered by this analysis. What is instead possible to claim is that the here derived channel model is additive and multiplicative but with correlated noises (Z and W). This means that it degenerates into the Faycal, Trott and Shamai channel in the particular case where Z and W are independent. If the channel input is reduced to take values in the real field, then one way to prove that the capacity achieving input distribution is discrete with a finite number of probability mass points is following the same trace as done in [44], since the output conditioned on the input variable is here also Gaussian distributed. This way of proceeding accounts also for the correlation between Z and W. Vice-versa and more generally, if the

input variable is complex, one may prove (as again sketched in [44]) that the relationship between the real and imaginary parts of the input variable is such to give rise to a shell distributed input, (*i.e.*, the input modulus can assume only a finite number of values, while the phase is uniformly distributed in general).

1.4 Channel Mutual Information

It is now possible to provide some information theoretic results, concerning the mutual information of the piggyback channel model: In Section 1.5, some consequent considerations will be presented.

Suppose the input X takes M possible complex values $\{x_i\}_{i=1}^{M}$ with $P_i = \Pr\{X = x_i\}$. The mutual information is [42, 43]:

$$\begin{aligned} \mathrm{I}(X;Y) &= h(Y) - h(Y|X) \\ &= \sum_{i=1}^{M} P_i \int_{\mathbb{C}} p_{Y|x_i}(y|x_i) \log \frac{p_{Y|x_i}(y|x_i)}{\sum_{k=1}^{M} p_{Y|x_k}(y|x_k) P_k} \, \mathrm{d}y \,. \end{aligned} \tag{1.9}$$

As previously stated, $p_{Y|x_i}(y|x_i)$ is circular CN distributed with mean

$$\mu_{Y|x_i} = \mathbb{E}\left[Y|x_i\right] = V_0 x_i \mathbb{E}\left[Z\right] + \mathbb{E}\left[W\right] = 0 \tag{1.10}$$

and variance

$$\begin{aligned} \sigma^2_{Y|x_i} &= \mathbb{E}\left[YY^*|x_i\right] \\ &= |V_0 x_i|^2 \underbrace{\mathbb{E}\left[|Z|^2\right]}_{\sigma_Z^2} + 2\Re\{(V_0 x_i)^* \mathbb{E}\left[WZ^*\right]\} + \underbrace{\mathbb{E}\left[|W|^2\right]}_{\sigma_W^2} . \end{aligned} \tag{1.11}$$

One can notice that the conditioned variance of Y is a real positive scalar and that it is completely known, given the conditioning input x_i, the variances of Z and W and their covariance

$$\begin{aligned} Cov(W,Z) &= \mathbb{E}\left[WZ^*\right] \\ &= V_Z \rho \sigma_{Z_1} \sigma_Z + V_0 A_s \sigma_Z^2 \end{aligned} \tag{1.12}$$

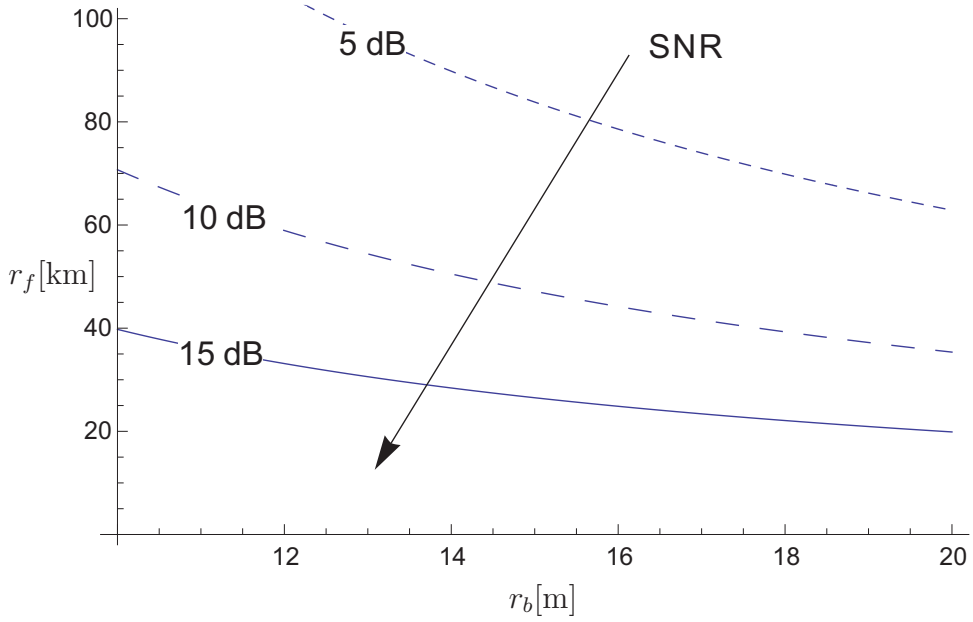

Figure 1.3: Trade off between r_f and r_b for increasing values of the SNR are shown.

where

$$\rho = \frac{\mathbb{E}[Z_{1_X} Z_X]}{\sigma_{Z_{1_X}} \sigma_{Z_X}} = \frac{\mathbb{E}[Z_{1_Y} Z_Y]}{\sigma_{Z_{1_Y}} \sigma_{Z_Y}}, \tag{1.13}$$

The distribution $p_{Y|x_i}(y|x_i)$ is therefore

$$\begin{aligned} & p_{Y|x_i}(y|x_i) \\ & = \frac{1}{\pi|\sigma^2_{Y|x_i}|} \exp\left\{ -\frac{1}{2} \begin{bmatrix} y^* & y \end{bmatrix} \begin{bmatrix} \sigma^2_{Y|x_i} & 0 \\ 0 & (\sigma^2_{Y|x_i})^* \end{bmatrix}^{-1} \begin{bmatrix} y \\ y^* \end{bmatrix} \right\} = \\ & = \frac{1}{\pi\sigma^2_{Y|x_i}} \exp\left\{ -\frac{|y|^2}{\sigma^2_{Y|x_i}} \right\}. \end{aligned} \tag{1.14}$$

Inserting (1.14) in (1.9), the mutual information can be computed in any desired discrete input scenario.

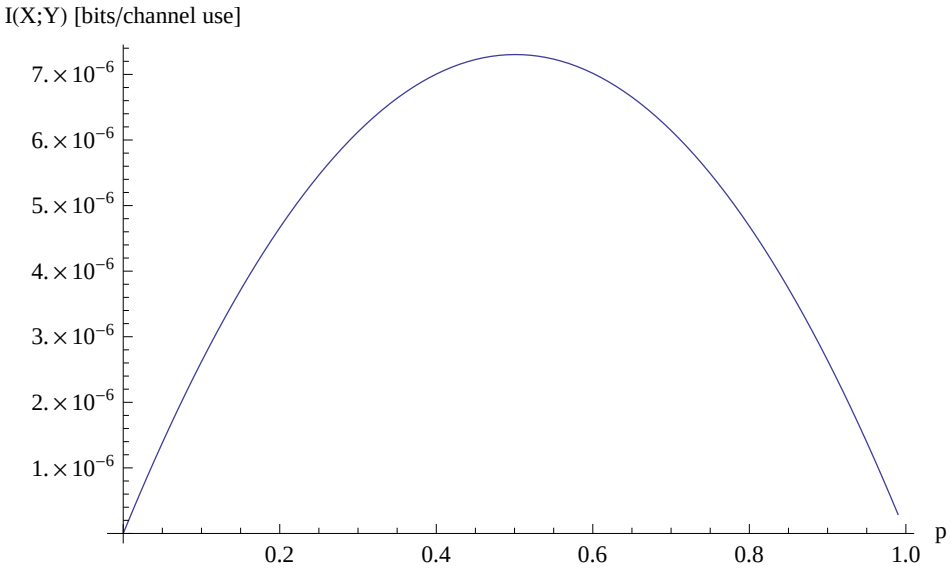

Figure 1.4: Antipodal mutual information vs. input distribution parameter p, $\mathsf{SNR} = 10\,\mathrm{dB}$, $\eta = 2.5\ 10^4$, $\rho = 0.5$.

1.5 Numerical Results

In this section an evaluation of the link budget is aimed at demonstrating the feasibility of the piggyback communication scheme in a typical WSN scenario. The mutual information is also evaluated in one specific case and some conclusions are drawn from an information theoretic perspective. The evaluation of the mutual information allows to fix a precise benchmark for the transmission data rate: Even if it does not represent the achievable maximum data rate, it states a value which, being under channel capacity, guarantees an arbitrary low bit error probability at the receiver.

It has already been stressed how the piggyback communication scheme borrows its carrier signal from another existing application. This imposes to pay particular attention in choosing the SoO. In fact, the exploited signal is required to be as largely diffused as possible. Because of its wide diffusion, a digital video broadcasting-terrestrial (DVB-T) SoO signal, using OFDM modulation, represents an interest-

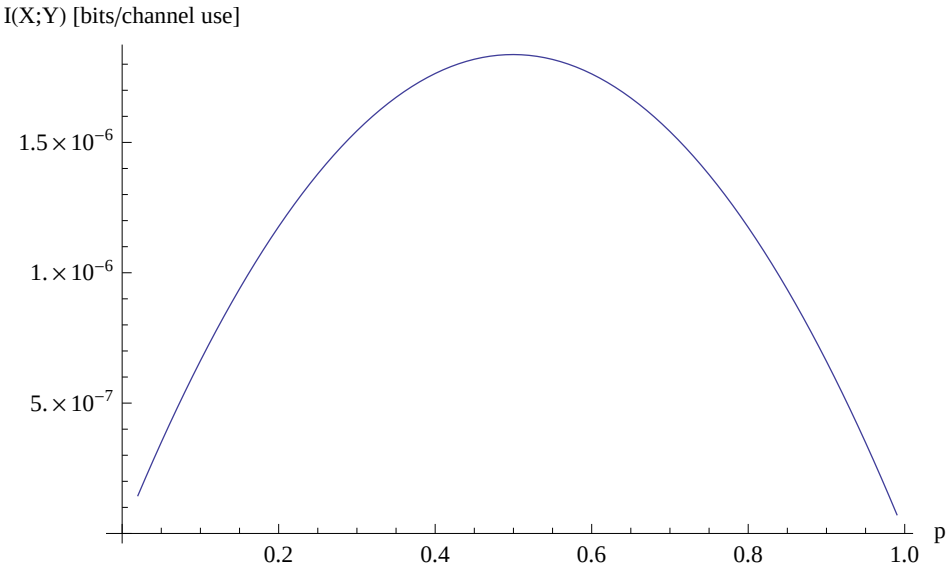

Figure 1.5: On-off mutual information vs. input distribution parameter p, $\mathsf{SNR} = 10\,\mathrm{dB}$, $\eta = 2.5\ 10^4$, $\rho = 0.5$.

ing choice. A typical effective radiated isotropic power (EIRP) value for a DVB-T transmitter may well be 10 kW (see, for example, [45]), serving a single $B = 5\,\mathrm{MHz}$ bandwidth channel. Next, consider the following realistic values for the numerical analysis, having in mind a DVB-T broadcast signal: $\mathsf{EIRP} = 10\,\mathrm{kW}$, $G_t = 2\,\mathrm{dB}$ (a typical dipole antenna gain, assumed to be isotropic in the horizontal plane), $A_s = 1$ [39], $G_R = 9\,\mathrm{dB}$, $\lambda = 0.375\,\mathrm{m}$ ($f_0 = 800\,\mathrm{MHz}$), $R_A = 50\ \Omega$, $N_0 = 1.266\ 10^{-20}\,\mathrm{J}$. In addition, it is possible to define the signal-to-noise ratio

$$\mathsf{SNR} = \frac{|V_0|^2 \mathbb{E}[|X|^2]}{N_0\, B\, R_A} \tag{1.15}$$

and, finally, by taking into account the independence of Gaussian noise and the direct path and backscattered signals, one has

$$\sigma_W^2 = N_0 B R_A + |V_Z|^2 + |V_0|^2 |A_s|^2 + 2\Re\{V_Z V_0^* A_s^* \rho\}. \tag{1.16}$$

Fig. 1.3 shows the trade off involving the SoO transmitter-sensor and sensor-receiver distances r_f and r_b, respectively, by considering an-

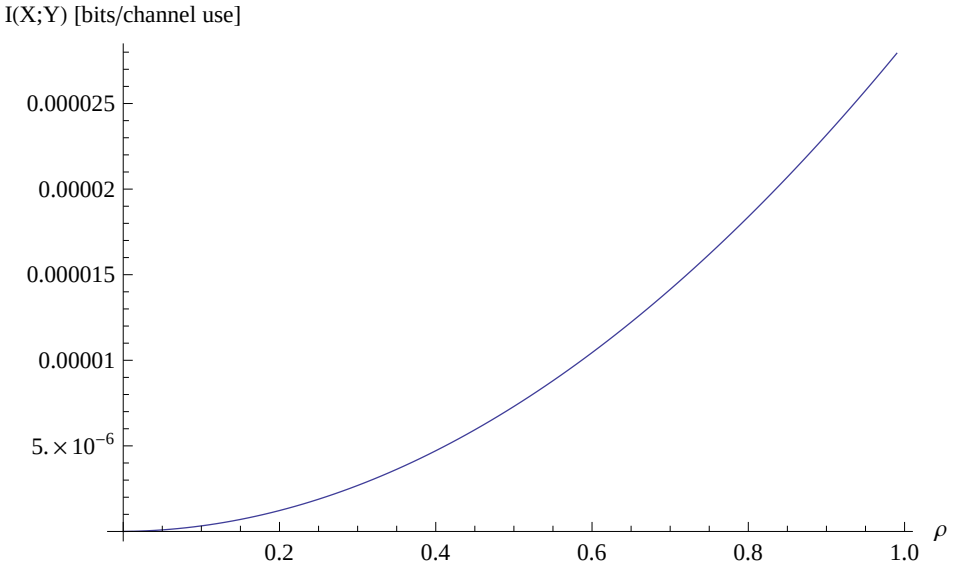

Figure 1.6: Antipodal mutual information vs. correlation parameter ρ, $\mathsf{SNR} = 10\,\mathrm{dB}$, $\eta = 2.5\ 10^4$, $p = 0.5$.

tipodal modulation ($X \in \{-1, 1\}$) and different SNRs. Thanks to the high EIRP of the SoO transmitter, it can be noted that appealing sensor-receiver distances (larger than 10 m) can potentially be achieved with typical values of SNR. This result assesses the possibility to employ this new communication technique for the proposed applications (WSNs and BANs).

As far as the achievable data rate performance is concerned, channel capacity analysis should be considered. Unfortunately, the capacity expression for the piggyback channel model is still unknown. However, the mutual information obtained with a discrete input distribution may be a meaningful term of comparison for capacity itself, as already cited papers [42, 43] show. Hence in a preliminary analysis, one can consider the mutual information resulting from he adoption of simple binary modulation schemes, usually implemented in backscatter modulators. The obtained results constitute a significant starting point: It has already been hinted out how, at low SNR, the not correlated version of the piggyback channel model degenerates into a binary

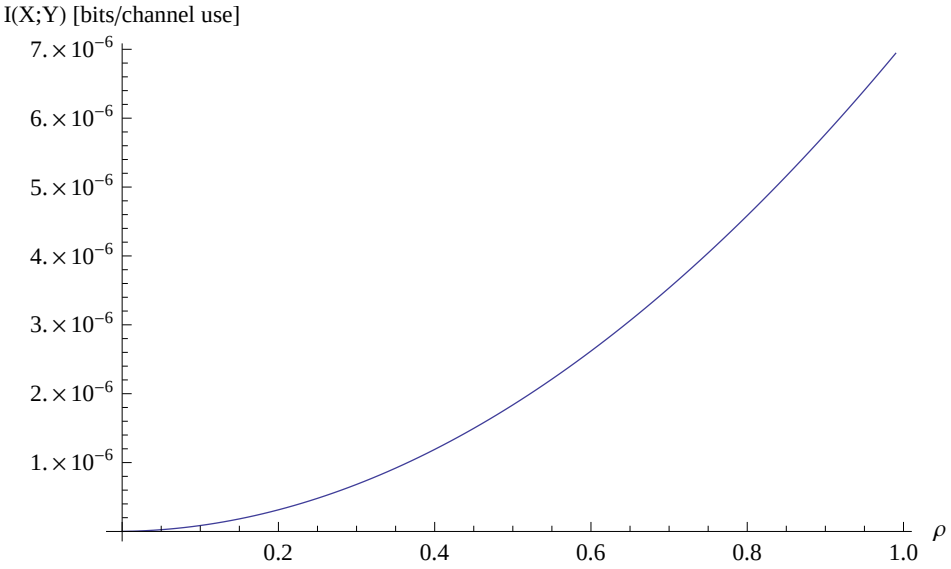

Figure 1.7: On-off mutual information vs. correlation parameter ρ, $\mathsf{SNR} = 10\,\mathrm{dB}$, $\eta = 2.5\ 10^4$, $p = 0.5$.

input capacity achieving channel.

For the above mentioned reasons, consider the cases of input RV X assuming values 1, -1 (antipodal case) and 0, 1 (on-off case) with probabilities $P_1 = 1 - p$ and $P_2 = p$, respectively. Figs. 1.4 and 1.5 show the behavior of the mutual information, as a function of p, while keeping fixed the correlation between the direct path signal and the backscatter path signal to $\rho = 0.5$. A strong direct path is considered in these examples. Results indicate that the antipodal scheme is better under the same boundary conditions.

It is also interesting to investigate the influence of the correlation between the direct path and the backscattered signal on the performance, by fixing $p = 0.5$. The obtained results are shown in Figs. 1.6 and 1.7, where the mutual information increases with the correlation: This is because the receiver can take advantage of the knowledge of the carrier SoO signal, when ρ is close to 1.

Finally, one can study the effect of the SoO direct path intensity with respect to the backscattered signal, accounted for by the parame-

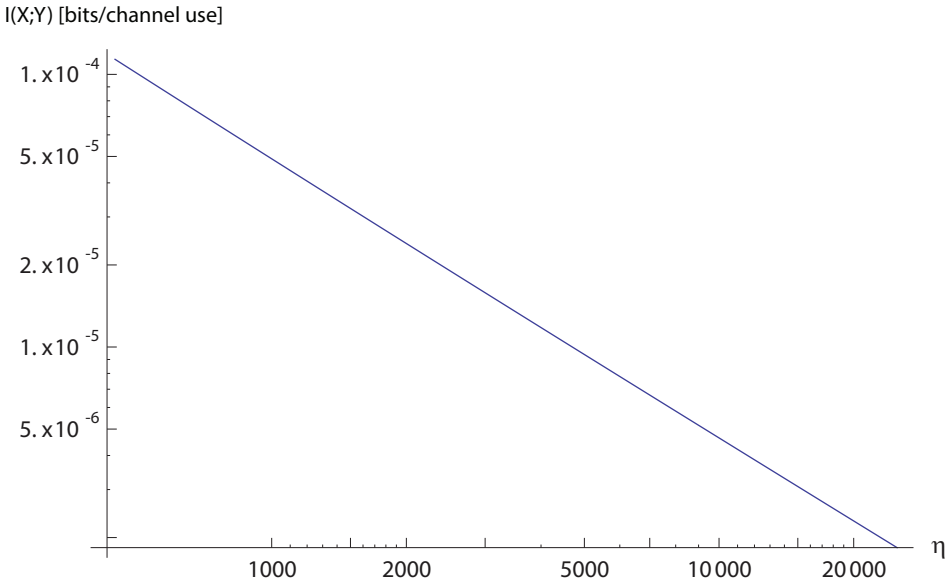

Figure 1.8: On-off mutual information vs. η, $\mathsf{SNR} = 10\,\mathrm{dB}$, $p = 0.5$, $\rho = 0.5$, logarithmic scale

ter η, reflecting different geometrical situations for what concerns the relative transmitter and receiver orientations. Fig. 1.8 shows that, in general, the performance significantly improves as the direct path power decreases, even in the presence of partial correlation between signals. This result indicates that the potential benefit, that can be obtained through the reception of a correlated version of the SoO signal, is hamstrung by the additional strong disturbance present at the receiver.

A direct multiplication of the mutual information by the number of channel uses per second, *i.e.* the signal bandwidth, leads to bit rates in the range of 10 to 500 bits per second, as it can be clearly deduced by Figs. 1.4, 1.5, 1.6, and 1.8. These values are quite encouraging for near zero-power remote sensing applications, which are not expected to be much data consuming.

1.6 Conclusions

Piggyback communication, recycling radio signals as sources of opportunity, has been introduced. The corresponding channel model, where the input symbol is subject to an amplitude constraint due to the passive nature of the communication, has been developed. The link budget analysis, in a typical WSN scenario with DVB-T SoO signal, demonstrates the feasibility of this scheme. In addition, some preliminary considerations in terms of mutual information have brought some insights in the understanding of the effect of signals correlation as well as of the impact of the SoO signal direct path power. Results show that a better performance can be achieved with increased correlation or by decreasing the direct path component power level.

As a further consideration, the interference on the primary user signal is expected to be negligible, since the backscattered signal is orders of magnitude below the directly received signal.

Summarizing the conducted analysis, the proposed communication technique appears to be very promising in view of near zero-power wireless networks, employing passive devices.

Chapter 2

Constrained Peak Power Channels

2.1 Introduction

In recent years, a great interest has been rising in what can be called discrete input channel modeling. This theory takes its first steps from the study of classical (Gaussian) additive noise channels under input constraints. The class of channels with input limitations is important from a practical point of view since feasible systems do always have to deal with input constraints: Peak and average power are necessarily bounded. The first works in this field were the ones by Smith back in the 70's [41, 42]. The forward step, with respect to the work [46] of Shannon, was to consider an additive Gaussian noise channel in which the input is either peak or both peak and average power constrained. The main discovery was that, under both constraints, the capacity achieving input probability measure (p.m.) is discrete with a finite number of probability mass points. This kind of p.m.s will be referred to as *finitely discrete* throughout this chapter. The results in [41, 42] are of notable importance since continuous inputs are not feasible in practice and have to be approximated with finitely discrete inputs.

The finitely discrete feature was demonstrated to be the exact solution for the capacity achieving input p.m. in the constrained additive scalar Gaussian noise channel model. This paved the way to several subsequent studies that, more recently, explored the finite discrete-

ness of capacity achieving input p.m.s for other input constrained channel models, presenting quite disparate characteristics. Among them one may cite [47] and [43], which inspired further works such as [48] and [44]. Concerning the two last mentioned works, the former presents conditions on the p.m. of an additive scalar channel noise, that are sufficient for the optimal bounded input p.m. to have a finitely discrete support. The latter demonstrates that such a support is sparse (see [44] for definition) when the channel conditional output p.m., possibly not scalar, is Gaussian distributed. Subsequent works exploited the finitely discrete nature of the input p.m. in some specific cases (e.g., [49, 50]) but no further generalizations have been developed.

The analysis conducted in this chapter was teased by the findings of Chapter 1. Notwithstanding, here, a general real scalar channel model is considered and sufficient conditions on the conditional output p.m. for the peak power limited capacity achieving input p.m. to be finitely discrete are provided. This result is established without indicating any particular type of conditional output p.m. nor any particular kind of the channel input-output law. Moreover, a class of peak power constrained additive channels as well as the peak power constrained Rayleigh fading channel are proven to fall in the developed framework as particular cases, whereas so far they have always been regarded as distinct categories, necessitating different mathematical treatments. In this respect, the presented conditions extend the theory of peak power limited real input scalar channels.

The chapter is organized as follows. In Section 2.2 all necessary notation and definitions are introduced, while in Section 2.3 the general channel model under consideration is introduced. The main result is stated in Section 2.4. This result is gradually proved in Sections 2.5, 2.6, and 2.7. Some hints about uniqueness of the capacity achieving input p.m. are provided in Section 2.8. The above mentioned examples are analysed in Section 2.9, while conclusions are drawn in Section 2.10. Ancillary results necessary for the proof of the main theorem are deferred to Annexes 2.11.1, 2.11.2, and 2.11.4 while Annex 2.11.3 provides some deeper explanations concerning the earlier discussed examples.

2.2 Notation and Definitions

In this section notation and definitions are presented in coherence with the ones given by previous authors [42, 44].

Throughout this chapter, Y and X represent the real scalar channel output and input RVs, respectively. Moreover, denote by $F(x)$ the input cumulative distribution function (c.d.f.), by $p_X(x)$ the input p.m., and by $p_{Y|X}(y|x)$ the conditional output p.m. The input RV X is assumed to take values in the set $\mathbb{S}$, with $\mathscr{P}$ being the ensemble of possible p.m.s defined on that set. The corresponding class of c.d.f.s is denoted by $\mathscr{F}$. One has

$$\begin{aligned} p_Y(y) &= \int_{\mathbb{S}} p_{Y|X}(y|x)p_X(x)\mathrm{d}x = \int_{\mathbb{S}} p_{Y|X}(y|x)\mathrm{d}F(x) \\ &= p_Y(y;p_X) = p_Y(y;F) \end{aligned} \tag{2.1}$$

where the dependence on $p_X(x)$ of the output p.m. $p_Y(\cdot)$ is kept explicit.[1]

Channel capacity is the supremum over the input p.m. of the mutual information functional [51]

$$\mathrm{I}(X;Y) = \int_{\mathbb{R}}\int_{\mathbb{S}} p_{Y|X}(y|x)\log\frac{p_{Y|X}(y|x)}{p_Y(y;F)}\mathrm{d}F(x)\mathrm{d}y = I(F) \tag{2.2}$$

where $\log(\cdot)$ denotes the base-2 logarithm.[2] Since only meaningless channel structure have zero capacity, one may assume the channel capacity to be strictly positive and denote the capacity achieving (hence optimal) input p.m. as $p_{X_0}(x)$. The mutual information functional can be further developed as

$$I(F) = H(F) - D(F) \tag{2.3}$$

where

$$H(F) \triangleq -\int_{\mathbb{R}} p_Y(y;F)\log p_Y(y;F)\mathrm{d}y$$

[1] Here, and throughout the whole chapter, one of the two equivalent formulations with $p_X(x)$ or $F(x)$ will be freely used as appropriately needed.

[2] In contrast, $\ln(\cdot)$ will denote the natural logarithm.

and

$$D(F) \triangleq -\int_{\mathbb{R}}\int_{\mathbb{S}} p_{Y|X}(y|x)\log p_{Y|X}(y|x)\mathrm{d}F(x)\mathrm{d}y.$$

One can notice how $D(F)$ depends in general on the input c.d.f., as opposed to what happens for an additive Gaussian channel (model in [42]).

One can also define the marginal information density and the marginal entropy density as

$$i(x;F) \triangleq \int_{\mathbb{R}} p_{Y|X}(y|x)\log\frac{p_{Y|X}(y|x)}{p_Y(y;F)}\mathrm{d}y$$

and

$$h(x;F) \triangleq -\int_{\mathbb{R}} p_{Y|X}(y|x)\log p_Y(y;F)\mathrm{d}y$$

respectively. These two densities are related as

$$i(x;F) = h(x;F) - d(x)$$

where

$$d(x) \triangleq -\int_{\mathbb{R}} p_{Y|X}(y|x)\log p_{Y|X}(y|x)\mathrm{d}y.$$

It is straightforward to show that the following three equations also hold:

$$I(F) = \int_{\mathbb{S}} i(x;F)\mathrm{d}F(x) \tag{2.4}$$

$$H(F) = \int_{\mathbb{S}} h(x;F)\mathrm{d}F(x) \tag{2.5}$$

and

$$D(F) = \int_{\mathbb{S}} d(x)\mathrm{d}F(x). \tag{2.6}$$

In this chapter, (2.4), (2.5), and (2.6) are well-defined since $h(\cdot)$, $i(\cdot)$, and $d(\cdot)$ are finitely bounded under the conditions enunciated in Section 2.3, as proven in Appendix 2.11.1.

2.3 Channel Model

The memoryless real scalar channel of interest is considered to be governed by a general input-output relationship in the form

$$Y = f(X, \underline{\Theta}) \tag{2.7}$$

where X is the input RV and $\underline{\Theta}$ a vector of nuisance parameters. No further conditions are imposed on the input-output channel law $f(\cdot)$, which may be linear or nonlinear, additive in noise or multiplicative or both, with independent or correlated noises.

Throughout the chapter, consider a peak power constrained input RV X taking values in the bounded set (see Fig. 2.1)

$$\mathbb{S} = [-A, A] \cap \mathbb{A}$$

where $[-A, A]$ is the compact real interval of radius A and $\mathbb{A}$ represents an open subset of the complex extended input plane on which the conditional output p.m. $p_{Y|X}(y|x)$ is analytic (hence continuous) in the input variable.

The fundamental conditions on which the analysis relies may be summarized as follows:

1. The conditional output p.m. can be analytically extended to complex inputs, *i.e.*, there exists an open set $\mathbb{A} \subseteq \mathbb{C}$ such that

   $$x \mapsto p_{Y|X}(y|x)$$

   is an analytic map over $\mathbb{A}$, while

   $$(x, y) \mapsto p_{Y|X}(y|x)$$

   is a continuous function over $\mathbb{A} \times \mathbb{R}$.

2. There exist two functions $q(y)$ and $Q(y)$, both nonnegative, and bounded above, and integrable, such that $\forall x \in \mathbb{S}$ one has

   $$0 \leq q(y) \leq p_{Y|X}(y|x) \leq Q(y) \leq K < +\infty, \ \forall y \in \mathbb{R} \tag{2.8}$$

   and the map

   $$y \mapsto Q(y) \log q(y)$$

   is integrable in y.

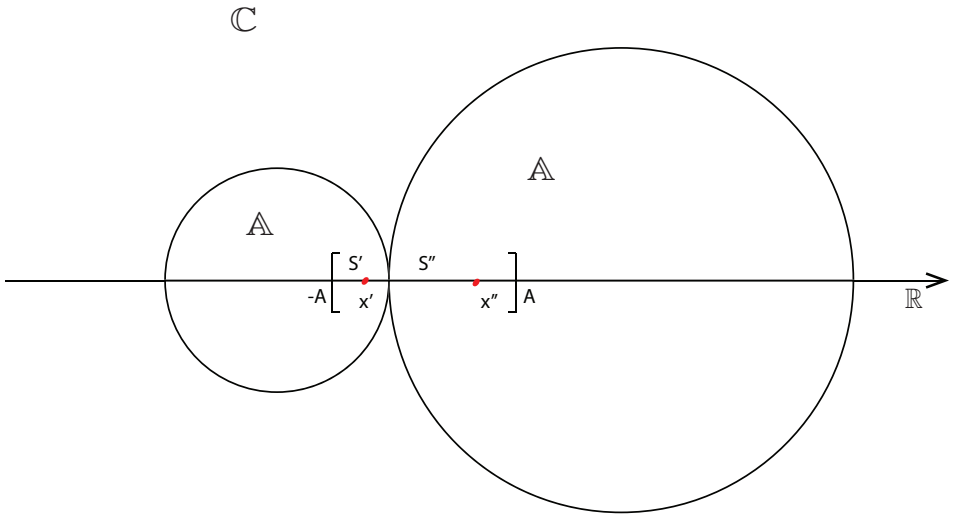

Figure 2.1: Pictorial representation of the set $\mathbb{S} = [-A, A] \cap \mathbb{A}$ on which the input RV takes its values.

3. The two integrals

$$\int_{\mathbb{R}} p_{Y|X}(y|w) \log p_{Y|X}(y|w) \mathrm{d}y$$
$$\int_{\mathbb{R}} p_{Y|X}(y|w) \log p_{Y}(y; p_X) \mathrm{d}y$$

are uniformly convergent (see [52] for definition) $\forall w \in \mathbb{D}$, for some $\mathbb{D}$ such that $\mathbb{S} \subset \mathbb{D} \subseteq \mathbb{A}$.[3]

4. For each of the maximally extended connected regions forming $\mathbb{S}$ (one can call them $S', S'', \ldots$), one of the following three conditions holds:

   (a) there exist $x', x'', \ldots \in S', S'', \ldots$ (see Fig. 2.1) and corresponding c.d.f.s $F', F'', \ldots$ with

   $$\log p_{Y|X}(y|x') - \log q(y) < I(F'), \ \forall y \in \mathbb{R} \tag{2.9}$$

[3]For the sake of clarity, here and elsewhere in the chapter a generic input value is denoted by x or w whenever the input is considered strictly real or complex extended, respectively.

and analogously for the other regions, where $I(F')$ is the mutual information between the output and input variable when the input is distributed according to $F'(x)$.

(b) for all real input p.m.s $p_X(x)$, there exist $x', x'', \ldots \in S'$, $S'', \ldots$ (see Fig. 2.1) such that $p_{Y|X}(y|x')$ is the unique conditional output p.m. satisfying

$$\min_{x \in S'} D_{KL}(p_{Y|X}(y|x)||p_Y(y; p_X)) = \\ = D_{KL}(p_{Y|X}(y|x')||p_Y(y; p_X))$$

and analogously for the other regions, where D_{KL} denotes the Kullback-Leibler divergence.

(c) for all real input p.m.s $p_X(x)$, there exist pairs of distinct points $(x'_1, x'_2), (x''_1, x''_2), \ldots \in S' \times S', S'' \times S'', \ldots$ such that

$$D_{KL}(p_{Y|X}(y|x'_1)||p_Y(y; p_X)) \neq D_{KL}(p_{Y|X}(y|x'_2)||p_Y(y; p_X))$$

and analogously for the other regions.

Remark 1. *The here stated conditions do not impose any peculiar kind of conditional output p.m., as it was the case in [41, 42, 44], nor any particular channel law, as it was done in [48]. The first three conditions are mild regularity conditions on the conditional output p.m. while the fourth one states that it should actually depend on the input value that is assigned as a condition. Moreover, the input set compactness, deeply exploited in [44], is* not *a required condition here. Examples, considered in Section 2.9, further show the presented theory to extend the previously known treatments.*

2.4 Main Result

The statement of the main result of this chapter hereby follows.

Theorem 1. Every real scalar and peak power constrained input channel, whose conditional output p.m. fulfills the aforementioned conditions 1 to 4, has a finitely discrete capacity achieving input p.m.

The remainder of this chapter is devoted to prove Theorem 1. The proof requires some intermediate steps: In particular, Section 2.5 proves that the capacity achieving input p.m. exists and also states, as a corollary, the Kuhn-Tucker conditions on the marginal information density (defined in Section 2.2) for an input p.m. to be optimal. Section 2.6 proves the analyticity of the marginal information density which is exploited in Section 2.7, alongside the corollary statement, to finally prove the finitely discrete nature of the capacity achieving input p.m. support. Besides, Section 2.8 hints in the direction of proving uniqueness of the optimal input p.m..[4]

2.5 Existence of a Capacity Achieving Probability Measure

Following the approach in [41, 42], in this section it is proven that an *optimal* input p.m. exists and that the Kuhn-Tucker conditions are necessary and sufficient for its optimality. Some basic results in optimization theory are first reviewed [41, 42, 53].

A map $f : \Omega \mapsto \mathbb{R}$, where Ω is a convex space, is said to be weakly differentiable in Ω if, for $\theta \in [0, 1]$ and $x_0 \in \Omega$, the map $f'_{x_0} : \Omega \to \mathbb{R}$, defined as

$$f'_{x_0}(x) = \lim_{\theta \to 0} \frac{f[(1-\theta)x_0 + \theta x] - f(x_0)}{\theta}$$

exists for all x and x_0 in Ω. Besides, f is said to be concave if, for all $\theta \in [0, 1]$ and for all x and x_0 in Ω,

$$f[(1-\theta)x_0 + \theta x] \geq (1-\theta)f(x_0) + \theta f(x).$$

Theorem 2 (Optimization Theorem [53])**.** Let f be a continuous, weakly differentiable, and concave map from a compact, convex topological space Ω to $\mathbb{R}$, and define

$$C \triangleq \sup_{x \in \Omega} f(x)\,.$$

Then:

[4] Uniqueness was not proven in general neither in [48] nor in [44].

1. $C = \max_{x\in\Omega} f(x) = f(x_0)$ for some $x_0 \in \Omega$;
2. $f(x_0) = C$ if and only if $f'_{x_0}(x) \leq 0 \ \forall x \in \Omega$.

Exploiting the above results from optimization theory, one has the following proposition.

Proposition 1. Let $I(F)$ be the mutual information functional between X and Y, as defined in (2.2). Then, under an input peak power constraint and conditions 1 and 2 of Sec. 2.3, there exists an $F_0 \in \mathscr{F}$ (equivalently a $p_{X_0} \in \mathscr{P}$) such that

$$C = I(F_0) = \max_{F\in\mathscr{F}} I(F).$$

Moreover, a necessary and sufficient condition for the input c.d.f. F_0 to maximize $I(F)$, *i.e.*, to achieve capacity, is

$$\int_{\mathbb{S}} i(x; F_0)\mathrm{d}F(x) \leq I(F_0), \quad \forall F \in \mathscr{F}. \tag{2.10}$$

Proof. As from Theorem 2, it suffices to show that $\mathscr{F}$ is convex and compact in some topology and that $I : \mathscr{F} \mapsto \mathbb{R}$ is continuous, concave and weakly differentiable. The necessary and sufficient condition (2.10) also follows from Theorem 2, as it will be shown.

2.5.1 Convexity and Compactness

The convexity of $\mathscr{F}$, *i.e.* the fact that

$$F_\theta(x) = (1-\theta)F_1(x) + \theta F_2(x)$$

still belongs to $\mathscr{F}$ for each F_1, F_2 in $\mathscr{F}$ and for each $\theta \in [0,1]$, is immediate. The compactness of $\mathscr{F}$ in the Lèvy metric[5] topology (as defined in [41] and recalled in Appendix 2.11.5) follows from the Helly Weak Compactness Theorem (see Appendix 2.11.4) and from the fact that convergence in the Lèvy metric is equivalent to complete convergence [54], which on a bounded interval is equivalent to weak convergence.

[5]The corresponding distance is here indicated with $d(\cdot,\cdot)$.

2.5.2 Continuity

The continuity of functional $I(F)$ descends from the Helly-Bray Theorem (see Appendix 2.11.4), according to which $d(F_n, F) \underset{n}{\to} 0$ implies $I(F_n) \underset{n}{\to} I(F)$, provided the boundedness and continuity in x of $i(x;F)$. The latter two properties are demonstrated in Appendix 2.11.1 (continuity of $i(x;F)$ is a consequence also of analyticity discussed in Section 2.6).

2.5.3 Concavity

For what concerns $I(F)$ being concave, one can note how

$$\begin{aligned} p_Y(y;F_\theta) &= p_Y(y;(1-\theta)F_1+\theta F_2) \\ &= \int_{\mathbb{S}} p_{Y|X}(y|x)[(1-\theta)\mathrm{d}F_1(x)+\theta \mathrm{d}F_2(x)] \\ &= (1-\theta)p_Y(y;F_1)+\theta p_Y(y;F_2) \end{aligned}$$

and

$$\begin{aligned} D\left((1-\theta)F_1\right. &\left.+\theta F_2\right) = \\ &= -\int_{\mathbb{R}}\int_{\mathbb{S}} p(y|x)\log p(y|x)[(1-\theta)\mathrm{d}F_1(x)+\theta \mathrm{d}F_2(x)]\mathrm{d}y \\ &= (1-\theta)D(F_1)+\theta D(F_2). \end{aligned} \tag{2.11}$$

Hence, one has that

$$I((1-\theta)F_1+\theta F_2) \geq (1-\theta)I(F_1)+\theta I(F_2)$$

is equivalent, from (2.3) and (2.11), to

$$H((1-\theta)F_1+\theta F_2) \geq (1-\theta)H(F_1)+\theta H(F_2). \tag{2.12}$$

Inequality (2.12) may be proved as follows:

$$\begin{aligned} H((1-\theta)F_1+\theta F_2) &= -\int_{\mathbb{R}} p_Y(y;(1-\theta)F_1+\theta F_2)\log p_Y(y;F_\theta)\mathrm{d}y \\ &= -\int_{\mathbb{R}} [(1-\theta)p_Y(y;F_1)+\theta p_Y(y;F_2)]\log p_Y(y;F_\theta)\mathrm{d}y \\ &\overset{(a)}{\geq} -(1-\theta)\int_{\mathbb{R}} p_Y(y;F_1)\log p_Y(y;F_1)\mathrm{d}y \\ &\quad -\theta\int_{\mathbb{R}} p_Y(y;F_2)\log p_Y(y;F_2)\mathrm{d}y \\ &= (1-\theta)H(F_1)+\theta H(F_2) \end{aligned}$$

where (a) exploits the Gibbs inequality [51], which states that for any two random variables, Z_1 and Z_2, one has

$$-\int_{\mathbb{R}} p_{Z_1}(z)\log p_{Z_1}(z)\mathrm{d}z \leq -\int_{\mathbb{R}} p_{Z_1}(z)\log p_{Z_2}(z)\mathrm{d}z$$

with equality if and only if

$$p_{Z_1}(z) = p_{Z_2}(z).$$

Hence, concavity of $I(\cdot)$ is proven and equality holds if and only if $p_Y(y;F_1) = p_Y(y;F_2)$.

2.5.4 Weak Differentiability

As proven in Appendix 2.11.2, for arbitrary F_1 and F_2 in $\mathscr{F}$ one has

$$\lim_{\theta\to 0}\frac{I((1-\theta)F_1+\theta F_2)-I(F_1)}{\theta} = \int_{\mathbb{S}} i(x;F_1)\mathrm{d}F_2(x) - I(F_1). \tag{2.13}$$

The proof of weak differentiability is completed by observing that $i(x;F)$ is finitely bounded (Appendix 2.11.1), which guarantees the existence of the integral in the right-hand side of (2.13).

Since all hypotheses of Theorem 2 are satisfied, the optimal input p.m. exists in $\mathscr{P}$. Furthermore, from (2.13), it is immediate to derive the necessary and sufficient condition (2.10). □

The following corollary of Proposition 1 states the Kuhn-Tucker conditions that will be used in Section 2.7 to prove the final result.

Corollary 1 (Kuhn-Tucker Conditions)**.** Let p_{X_0} be an arbitrary p.m. in $\mathscr{P}$. Let S_0 denote the set of mass points of p_{X_0} on $\mathbb{S}$.[6] Then p_{X_0} is optimal if and only if

$$\begin{cases} i(x;p_{X_0}) \leq I(p_{X_0}), & \forall x \in \mathbb{S} \\ i(x;p_{X_0}) = I(p_{X_0}), & \forall x \in S_0 \end{cases}$$

The proof of this corollary is taken from [41, 42].

Proof. If both conditions hold p_{X_0} is optimal because it satisfies the condition stated in Proposition 1. The converse holds in the subsequent way: Assume p_{X_0} is optimal but the first equation is not valid. Then an $x_1 \in \mathbb{S}$ exists such that $i(x_1;p_{X_0}) > I(p_{X_0})$. Choose $p_X(x) = \delta(x - x_1)$. Then

$$\int_{\mathbb{S}} i(x;p_{X_0})p_X(x)\mathrm{d}x = i(x_1;p_{X_0}) > I(p_{X_0}). \tag{2.14}$$

This contradicts Proposition 1, hence the first statement has to be true. Now, assume the second one is not true. Then, because of the first equation,

$$i(x;p_{X_0}) < I(p_{X_0}), \tag{2.15}$$

on a set $E' \subseteq S_0$. This implies

$$\int_{E'} p_{X_0}(x)\mathrm{d}x = \delta > 0,$$
$$\int_{S_0 \setminus E'} p_{X_0}(x)\mathrm{d}x = 1 - \delta,$$

[6] The set S_0 is defined independently of the discreteness or continuity of the input p.m..

with $i(x;p_{X_0}) = I(p_{X_0})$ on $S_0 \setminus E'$. In the end one can write

$$I(p_{X_0}) = \int_{\mathbb{S}} i(x;p_{X_0})p_{X_0}(x)\mathrm{d}x = \int_{S_0} i(x;p_{X_0})p_{X_0}(x)\mathrm{d}x =$$
$$= \int_{E'} i(x;p_{X_0})p_{X_0}(x)\mathrm{d}x + \int_{S_0\setminus E'} i(x;p_{X_0})p_{X_0}(x)\mathrm{d}x <$$
$$< \delta I(p_{X_0}) + (1-\delta)I(p_{X_0}) = I(p_{X_0}), \tag{2.16}$$

which is a clear absurd. Thus also the second statement holds. □

2.6 Analyticity of the Mutual Information Density

In this section it is proven that $i(x;p_X)$ can be analytically extended to $i(w;p_X)$, $\forall w \in \mathbb{D}$. This step is necessary as a starting point for the capacity achieving input p.m. characterization in Section 2.7.

First, one has to extend $i(x;p_X)$ to the analyticity region $\mathbb{A}$ of $x \mapsto p_{Y|X}(y|x)$ as

$$i(w;p_X) \triangleq \int_{\mathbb{R}} p_{Y|X}(y|w)\log\frac{p_{Y|X}(y|w)}{p_Y(y;p_X)}\mathrm{d}y$$

$\forall w \in \mathbb{A}$ where convergence holds.[7] Now apply the Differentiation Lemma (see Appendix 2.11.4, with $I = \mathbb{R}$, $U = \mathbb{D}$), to the functions

$$f_1(w,y) = p_{Y|X}(y|w)\log p_{Y|X}(y|w),$$
$$f_2(w,y) = p_{Y|X}(y|w)\log p_Y(y;p_X).$$

The two functions are continuous (see Section 2.3) over $\mathbb{D}\times\mathbb{R}$.[8] Moreover, from conditions in Section 2.3, they are uniformly integrable over $\mathbb{R}$ and, being compositions of analytic functions, they are analytic. The difference of the two analytic (from Differentiation Lemma)

[7] Convergence is guaranteed inside $\mathbb{S}$, as proven in Appendix 2.11.1.

[8] $\mathbb{D}$ has to exclude the possibility for $p_{Y|X}(y|w)$ to be real negative valued, this to ensure continuity of the principal value complex logarithm.

integral functions

$$\int_{\mathbb{R}} f_1(y,w)\mathrm{d}y - \int_{\mathbb{R}} f_2(y,w)\mathrm{d}y$$

is analytic on $\mathbb{D}$. This means that $i(w;p_X)$ is an analytic function over $\mathbb{D}$.

2.7 A Finitely Discrete Capacity Achieving Probability Measure

In this section, the finite discreteness of the capacity achieving input p.m. is finally proven .
Define $v(w)$ as[9]

$$\begin{aligned} v(w) &\triangleq \int_{\mathbb{R}} p_{Y|X}(y|w)\left[-\log\left(\frac{p_Y(y;p_{X_0})}{p_{Y|X}(y|w)}\right) - I(p_{X_0})\right]\mathrm{d}y \\ &= i(w;p_{X_0}) - I(p_{X_0}) \end{aligned} \tag{2.17}$$

where $p_{X_0}(x)$ is a capacity achieving input p.m.. Recall from Section 2.3 that $S', S'', \ldots$ are the maximally extended connected regions forming $\mathbb{S}$, while $S'_0, S''_0, \ldots$ is the corresponding decomposition for S_0 (the support of $p_{X_0}(x)$), *i.e.*, S'_0 is the set of points of S_0 in S', S''_0 is the set of points of S_0 in S'', and so on. Note that, if each of the optimal input domain decomposition sets were not finitely discrete, then, for the Bolzano-Weierstrass Theorem, it would have an accumulation point in the corresponding connected subregion of $\mathbb{S}$ and thus, by the identity principle of analytic functions and Corollary 1, $v(w) = 0$ in that entire subregion. From (2.17), $v(w) = 0$ means

$$-\int_{\mathbb{R}} p_{Y|X}(y|w)\log\left(\frac{p_Y(y;p_{X_0})}{p_{Y|X}(y|w)}\right)\mathrm{d}y - I(p_{X_0}) = 0.$$

[9]Recall that a generic input value is respectively denoted by x and w, depending on whether the input is considered strictly real or complex extended.

In the following, for notation convenience, suppose to consider the S' subregion of $\mathbb{S}$.

In case one of the first two options 4a, 4b presented in Section 2.3 is verified and since $v(w) = 0$ on the entire considered subregion, one must have:

$$v(x') = -\int_{\mathbb{R}} p_{Y|X}(y|x') \log\left(\frac{p_Y(y;p_{X_0})}{p_{Y|X}(y|x')}\right) \mathrm{d}y - I(p_{X_0}) = 0$$

also for the corresponding particular value x', whose existence was supposed in Section 2.3. However this is in clear contradiction with either

$$\begin{aligned} v(x') &= \int_{\mathbb{R}} p_{Y|X}(y|x') \log\left(\frac{p_{Y|X}(y|x')}{p_Y(y;p_{X_0})}\right) \mathrm{d}y - I(p_{X_0}) \\ &\leq \int_{\mathbb{R}} p_{Y|X}(y|x') \underbrace{\Big(\log p_{Y|X}(y|x') - \log q(y)\Big)}_{<I(F') \text{ see eq.(2.9)}} \mathrm{d}y - I(p_{X_0}) \\ &< I(F') - I(F_0) \leq 0. \end{aligned}$$

or

$$\begin{aligned} v(x') &= \int_{\mathbb{R}} p_{Y|X}(y|x') \log\left(\frac{p_{Y|X}(y|x')}{p_Y(y;p_{X_0})}\right) \mathrm{d}y - I(p_{X_0}) \\ &= D_{KL}(p_{Y|X}(y|x')||p_Y(y;p_{X_0})) - I(p_{X_0}) \\ &< D_{KL}(p_{Y|X}(y|x)||p_Y(y;p_{X_0})) - I(p_{X_0}) = 0. \end{aligned}$$

If vice versa the third option 4c holds, it follows

$$\begin{aligned} v(x'_1) &= D_{KL}(p_{Y|X}(y|x'_1)||p_Y(y;p_{X_0})) - I(p_{X_0}) \\ &\neq D_{KL}(p_{Y|X}(y|x'_2)||p_Y(y;p_{X_0})) - I(p_{X_0}) = 0 \end{aligned}$$

and again a contradiction occurs.

This finally proves that the hypothesis to have an infinite set of probability mass points S_0 is incorrect and, hence, the input RV X can take only on a finitely discrete set of values.

2.8 Hints about Uniqueness

The so far developed conditions on the capacity achieving input p.m. do not guarantee also its uniqueness. In this direction, a further property that all possibly optimal input p.m.s must satisfy, differently to any other capacity achieving p.m., can be outlined.
Consider all the optimal input p.m.s[10] and denote the i-th of them by $p_{X_i}(x)$. Then, the following proposition holds.

Proposition 2. All the optimal input p.m.s of a channel model satisfying conditions 1-4 in Section 2.3, must fulfil the condition

$$i(x; p_{X_0}) = I(p_{X_0}), \quad \forall x \in S_i$$

$S_i \subset \mathbb{S}$ being the support of $p_{X_i}(x)$.

Proof. Let $p_{X_0}(x)$ and $p_{X_1}(x)$ be two optimal input p.m.s (whose existence is guaranteed by Proposition 1), both with a finitely discrete support. Then also $(1-\theta)p_{X_1}(x) + \theta p_{X_0}(x)$ is capacity achieving, since the mutual information functional is concave (see Theorem 6 in Appendix 2.11.4). This fact yields the weak derivative $I'_{p_{X_0}}(p_{X_1})$ to be null. Recall the probability mass points in S_0 and S_1 x_m and x_n, and the correspondent probability b_m and a_n, respectively. In addition suppose that the condition enunciated in Proposition 2 is not verified, *i.e.*, $i(x; p_{X_0}) < I(p_{X_0})$ for at least one of the $x_n \in S_1$, where the order relation is imposed by Corollary 1. The cited weak derivative expression becomes

$$\begin{aligned}\int_{\mathbb{S}} i(x; p_{X_0}) \left[p_{X_1}(x) - p_{X_0}(x)\right] \mathrm{d}x &= \sum_n a_n i(x_n; p_{X_0}) - \sum_m b_m i(x_m; p_{X_0}) \\ &< I(p_{X_0}) \sum_n a_n - I(p_{X_0}) \sum_m b_m = 0.\end{aligned}$$

A contradiction has arisen since $I'_{p_{X_0}}(p_{X_1}) = 0$ and $I'_{p_{X_0}}(p_{X_1}) < 0$, which completes the proof. □

[10] In the previous sections, it was proven that they belong to $\mathscr{P}'$, the restriction of $\mathscr{P}$ to the class of finitely discrete generalized functions defined on a finite number of probability mass points in the input support $\mathbb{S}$.

Proposition 2 does not provide uniqueness of the capacity achieving input p.m., nevertheless it tightens the conditions for an input p.m. to be optimal.

2.9 Examples

This section is divided in two subsections. The first one proves that any peak power constrained channel with additive noise satisfies condition 4, stated in Section 2.3 and, therefore, it belongs to the general class of channels treated in this chapter upon fulfilling also conditions 1, 2, and 3.[11] The second one proves that the Rayleigh fading channel undergoes all the conditions in Section 2.3.

2.9.1 Additive Channels

Consider an additive channel model $Y = X + N$, where N is the noise RV. The marginal information density can be rewritten as

$$\begin{aligned} i(x;p_X) &= \int_{\mathbb{R}} p_N(y-x)\log p_N(y-x)\mathrm{d}y - \int_{\mathbb{R}} p_N(y-x)\log p_Y(y;p_X)\mathrm{d}y \\ &= k - \int_{\mathbb{R}} p_N(y-x)\log p_Y(y;p_X)\mathrm{d}y \end{aligned}$$

where k is constant as it can be easily shown with an ordinary variable substitution. The second term is in the form of convolution and admits Fourier Transform (FT) since $p_N(\cdot)$ is integrable on $\mathbb{R}$ and $\log p_Y(y;p_X) = u(y)$ is locally integrable hence transformable at least in the sense of distributions. Now assume the marginal information density is equal to a constant c_1: Its FT would then be

$$\Psi_N(2\pi f)U(f) = c_1\delta(f)$$

where $\Psi_N(\cdot)$ denotes the characteristic function of the RV N, defined as

$$\Psi_N(f) = \mathbb{E}[\exp\{jxf\}] = \int_{\mathbb{R}} p_N(x)\exp\{jxf\}\mathrm{d}x.$$

[11] The fulfillment of conditions 1-3 must be checked case by case, but it is expected to be a simple verification.

The only case for this to hold is $u(y)$ being a constant itself: This is however contradictory since $u(y) = c_2$ implies $p_Y(y; p_X) = 2^{c_2}$, $\forall y \in \mathbb{R}$, which is clearly an absurd, and hence condition 4c stands.

2.9.2 The Rayleigh Fading Channel

Consider the Rayleigh fading channel conditional output p.m., as defined in [43],

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{1}{1+x^2} \exp\left\{-\frac{y}{1+x^2}\right\} \\ &= s \exp\{-ys\} \end{aligned}$$

and assume the channel input X is subject to a peak power constraint A as defined in Section 2.3. Since this conditional p.m. derives from normalizations of the original input and output modules, U and V in [43], this is a real scalar memoryless channel whose output takes values in $[0, +\infty)$.
It is now possible to assess that the four conditions stated in Section 2.3 are fulfilled.

(a) It is immediate to verify that condition 1 holds over the set $\mathbb{A} = \mathbb{C} \setminus \{-j, j\}$.

(b) Concerning condition 2, let us define

$$Q(y) = \begin{cases} 1, & 0 \leq y \leq c(A^2+1) \\ \frac{1}{y^{1+\gamma}}, & y > c(A^2+1) \end{cases}$$

where parameter γ fulfils $\gamma < 1$ and c is a constant such that $c > 2$ (the details are provided in Appendix 2.11.3). Moreover, let us define

$$q(y) = \begin{cases} \frac{1}{1+A^2} \exp\left\{-\frac{y}{1+A^2}\right\}, & 0 \leq y \leq \frac{(1+A^2)\ln(1+A^2)}{A^2} \\ \exp\{-y\}, & y > \frac{(1+A^2)\ln(1+A^2)}{A^2} \end{cases}$$

where $y_2 = \frac{(1+A^2)\ln(1+A^2)}{A^2}$ is the solution of $\frac{1}{1+A^2} \exp\left\{-\frac{y}{1+A^2}\right\} = \exp\{-y\}$. The two functions $q(\cdot)$ and $Q(\cdot)$ satisfy inequality

(2.8), as rigorously proven in Appendix 2.11.3. Furthermore, both of them are nonnegative, superiorly bounded, and integrable over the output domain $[0, +\infty)$. Besides $Q(y)\log q(y)$ is integrable over $[0, +\infty)$, which may be shown by analysing integrability over the tail.[12] One has

$$\begin{aligned}\int_{y_3}^{+\infty} Q(y)\log q(y)\mathrm{d}y &= \int_{y_3}^{+\infty} \frac{1}{y^{1+\gamma}}\exp\{-y\}\mathrm{d}y \\ &= \left[-\frac{y^{-\gamma}}{\gamma}\exp\{-y\}\right]_{y_3}^{+\infty} - \int_{y_3}^{+\infty}\frac{y^{-\gamma}}{\gamma}\exp\{-y\}\mathrm{d}y\end{aligned}$$

which is finite. The considered y_3 is sufficiently large to guarantee that the expressions employed for $Q(y)$ and $q(y)$ are the proper ones.

(c) Consider now the condition 3. The integral

$$\int_0^{+\infty} p_{Y|X}(y|w)\log p_Y(y;p_X)\mathrm{d}y$$

is uniformly convergent on

$$\mathbb{D} = \left\{w : \Re\{\frac{1}{1+w^2}\} \geq a_1, |\frac{1}{1+w^2}| \leq a_2\right\},$$

with strictly positive a_1 and a_2, and with a_1 ensuring that $\mathbb{S} \subset \mathbb{D}$. Uniform convergence holds since, for each $w \in \mathbb{D}$, given ϵ, there exist $B_0 < B_1 < B_2$ such that

$$\begin{aligned}&\Big|\int_{B_1}^{B_2} p_{Y|X}(y|w)\log p_Y(y;p_X)\mathrm{d}y\Big| \\ &\qquad \leq \int_{B_1}^{B_2}\big|p_{Y|X}(y|w)\log p_Y(y;p_X)\big|\,\mathrm{d}y\end{aligned}$$

[12] $Q(y)\log q(y)$ is locally integrable since it is continuous.

$$
\begin{aligned}
&= \int_{B_1}^{B_2} \left|\frac{1}{1+w^2}\right| \left|\exp\left\{-\frac{y}{1+w^2}\right\} \log p_Y(y;p_X)\right| \mathrm{d}y \\
&= \int_{B_1}^{B_2} \left|\frac{1}{1+w^2}\right| \left|\exp\left\{-y\Re\left\{\frac{1}{1+w^2}\right\}\right\} \log p_Y(y;p_X)\right| \mathrm{d}y \\
&\leq -\int_{B_1}^{B_2} \frac{1}{y^3} \log q(y) \mathrm{d}y < \epsilon
\end{aligned}
$$

as $\left|\frac{1}{1+w^2} \exp\left\{-y\Re\left\{\frac{1}{1+w^2}\right\}\right\}\right|$ is minor in a definitive manner in y than $1/y^3$ regardless of $w \in \mathbb{D}$.[13] To prove the result it is also necessary to employ (2.21) in Appendix 2.11.1 and to choose B_0 in such a way that $B_0 > y_2$ and $\frac{1}{B_0} < \epsilon$. The choice for $\mathbb{D}$ is dictated by the necessity to guarantee the existence of a uniform upper bound for $|p_{Y|X}(y|w) \log p_Y(y;p_X)|$. Analogously, also

$$
\int_0^{+\infty} p_{Y|X}(y|w) \log p_{Y|X}(y|w) \mathrm{d}y
$$

is uniformly convergent on $\mathbb{D}$. In fact, for each $w \in \mathbb{D}$, given ϵ, there exist $B_0 < B_1 < B_2$ such that

$$
\begin{aligned}
&\left| \int_{B_1}^{B_2} p_{Y|X}(y|w) \log p_{Y|X}(y|w) \mathrm{d}y \right| \\
&\quad \leq \int_{B_1}^{B_2} \left| p_{Y|X}(y|w) \log p_{Y|X}(y|w) \right| \mathrm{d}y \\
&\quad \leq \int_{B_1}^{B_2} \left|\frac{1}{1+w^2}\right| \left|\exp\{-\frac{y}{1+w^2}\} \log \frac{1}{1+w^2}\right| \mathrm{d}y \\
&\qquad + \int_{B_1}^{B_2} \left|\frac{1}{1+w^2}\right| \left|\exp\left\{-\frac{y}{1+w^2}\right\} \log\left(\exp\left\{-\frac{y}{1+w^2}\right\}\right)\right| \mathrm{d}y \\
&\quad \leq \int_{B_1}^{B_2} \frac{1}{y^2} \mathrm{d}y < \epsilon
\end{aligned}
$$

where again B_0 is chosen to ensure $1/B_0 < \epsilon$.

[13] This is guaranteed by the existence of a maximum for $|\frac{1}{1+w^2}|$ and a non zero minimum for $\Re\left\{\frac{1}{1+w^2}\right\}$ on $\mathbb{D}$.

(d) One should finally address condition 4. Consider

$$\begin{aligned}\int_0^{+\infty} p_{Y|X}(y|x)\log p_{Y|X}(y|x)\mathrm{d}y - \int_0^{+\infty} p_{Y|X}(y|x)\log p_Y(y;p_X)\mathrm{d}y \\ = \int_0^{+\infty} s\exp\{-ys\}\log\left(s\exp\{-ys\}\right)\mathrm{d}y \\ - \int_0^{+\infty} s\exp\{-ys\}\log p_Y(y;p_X)\mathrm{d}y \\ = \log s - \frac{1}{\ln 2} - \int_0^{+\infty} s\exp\{-ys\}\log p_Y(y;p_X)\mathrm{d}y.\end{aligned} \tag{2.18}$$

The third term dependence[14] on s cannot be logarithmic since

$$\lim_{s\to+\infty} -\int_0^{+\infty} s\exp\{-ys\}\log p_Y(y;p_X)\mathrm{d}y = 0$$

where exchange between integral and limit is licit since when $s \to +\infty$ it can be supposed greater than 1, this ensuring the existence of an integrable upper bound of $|s\exp\{-ys\}\log p_Y(y;p_X)|$, much as previously done for integrability of $Q(y)\log q(y)$. Hence the difference between the first and third term of (2.18) cannot be constant on $\mathbb{S}$, this proving condition 4c to hold.

2.10 Conclusions

This chapter has proposed general conditions on the conditional output p.m. under which real scalar channel models, with input peak power constraints, show to have capacity achieving p.m.s which are finitely discrete. These conditions represent a step towards the full understanding of the basic channel characteristics that determine the capacity achieving input p.m. to be finitely discrete under peak power

[14]Dependence on variable s is the same independently of the s considered: It is thus possible to consider values for s even outside the region dictated by the particular channel capacity problem which is under consideration.

constraints. The here presented theory of peak power limited channels unifies under a same framework several channel models that were previously investigated using separated approaches, as shown by the provided examples.

Particular attention will be paid to whether all of the posed conditions are strictly necessary. The feeling is that some of those conditions are not negotiable, while other ones may not be as fundamental as they appear to be.

As last but not least consideration, the deep belief that only *real* scalar peak power limited channels can have a finitely discrete capacity achieving input p.m. has matured.

2.11 Annexes

2.11.1 Boundedness and Continuity of the Marginal Information Density

The existence and boundedness of the upper and lower bounds on $p_{Y|X}(y|x)$, postulated in Section 2.3 is sufficient to prove the existence and boundedness of $p_Y(y; p_X)$. In fact, one can write

$$\begin{aligned} q(y) = \int_{\mathbb{S}} q(y) p_X(x) \mathrm{d}x &\leq \int_{\mathbb{S}} p_{Y|X}(y|x) p_X(x) \mathrm{d}x \\ &\leq \int_{\mathbb{S}} Q(y) p_X(x) \mathrm{d}x = Q(y) \end{aligned}$$

that is

$$0 \leq q(y) \leq p_Y(y; p_X) \leq Q(y) \leq K, \quad \forall y \in \mathbb{R} \text{ and } \forall\, p_X(x) \in \mathscr{P}. \tag{2.19}$$

An equally useful inequality, immediately descending from the previous one, is the following:

$$-\log Q(y) \leq -\log p_Y(y; p_X) \leq -\log q(y), \quad \forall y \in \mathbb{R} \text{ and } \forall\, p_X(x) \in \mathscr{P}. \tag{2.20}$$

Moreover, consider the pair of functions $f(y)$ and $g(y)$, respectively nonnegative and positive, such that $g(y) \leq K < +\infty$. The next inequality holds:

$$\begin{aligned} |f(y)\log g(y)| &\leq -f(y)\log\frac{g(y)}{K} + f(y)|\log K| \\ &\leq -f(y)\log g(y) + 2f(y)|\log K|. \end{aligned} \tag{2.21}$$

Besides

$$G(y) = -Q(y)\log q(y) + 2Q(y)|\log K|$$

is integrable on $\mathbb{R}$. Proof for this is an immediate consequence of the conditions in Sec. 2.3.

The next step consists in showing that $h(x;p_X)$ and $i(x;p_X)$ are bounded $\forall x \in \mathbb{S}$ and $\forall\, p_X(x) \in \mathscr{P}$. In fact one has

$$\begin{aligned} |h(x;p_X)| &= \left| \int_{\mathbb{R}} p_{Y|X}(y|x)\log p_Y(y;p_X)\mathrm{d}y \right| \\ &\leq \int_{\mathbb{R}} \left| p_{Y|X}(y|x)\log p_Y(y;p_X) \right| \mathrm{d}y \\ &\leq \int_{\mathbb{R}} p_{Y|X}(y|x)\big[-\log p_Y(y;p_X) + 2|\log K| \big]\mathrm{d}y \\ &\leq \int_{\mathbb{R}} Q(y)[-\log q(y) + 2|\log K|]\mathrm{d}y \\ &= \int_{\mathbb{R}} G(y)\mathrm{d}y < +\infty \end{aligned}$$

having used (2.8), (2.19), (2.20) and (2.21). Moreover, one has

$$\begin{aligned} |d(x)| &= \left| -\int_{\mathbb{R}} p_{Y|X}(y|x)\log p_{Y|X}(y|x)\mathrm{d}y \right| \\ &\leq \int_{\mathbb{R}} \left| p_{Y|X}(y|x)\log p_{Y|X}(y|x) \right| \mathrm{d}y \end{aligned}$$

$$\begin{aligned}
&\leq \int_{\mathbb{R}} p_{Y|X}(y|x) \left[-\log p_{Y|X}(y|x) + 2|\log K| \right] \mathrm{d}y \\
&\leq \int_{\mathbb{R}} Q(y) \left[-\log q(y) + 2|\log K| \right] \mathrm{d}y \\
&= \int_{\mathbb{R}} G(y) \mathrm{d}y < +\infty
\end{aligned}$$

where again (2.8), (2.19), (2.20) and (2.21) are exploited. One may then conclude that $i(x; p_X) = h(x; p_X) - d(x)$ is bounded, as it is the difference between two quantities fulfilling the same finite boundedness property.

Continuity of $i(x; p_X)$ can be demonstrated in an almost identical way since, $\forall x \in \mathbb{S}$, it is possible to exchange the continuity limit with the integral in the definition of $i(\cdot)$, this being guaranteed by integrability of $G(y)$, and continuity of the integrand functions being an immediate evidence.

2.11.2 Proof of Equation (2.13)

The weak derivative can be developed as

$$\begin{aligned}
&\lim_{\theta \to 0} \frac{I((1-\theta)F_1 + \theta F_2) - I(F_1)}{\theta} \\
&= \lim_{\theta \to 0} \Bigg\{ \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p(y; (1-\theta)F_1 + \theta F_2)} \\
&\qquad [(1-\theta)\mathrm{d}F_1(x) + \theta \mathrm{d}F_2(x)] \mathrm{d}y \\
&\qquad - \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y; F_1)} \mathrm{d}F_1(x) \mathrm{d}y \Bigg\} \\
&= \lim_{\theta \to 0} \Bigg\{ \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \left[-\log p_Y(y; F_\theta) + \log p_Y(y; F_1) \right] \mathrm{d}F_1(x) \mathrm{d}y \\
&\qquad + \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{(1-\theta) p_Y(y; F_1) + \theta p_Y(y; F_2)} \\
&\qquad [\mathrm{d}F_2(x) - \mathrm{d}F_1(x)] \mathrm{d}y \Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} \lim_{\theta\to 0}\left\{\frac{1}{\theta}\int_{\mathbb{R}}\int_{\mathbb{S}} p_{Y|X}(y|x)\left[\log p_Y(y;F_1)\right.\right.\\
&\quad\left.\left.-\log\left((1-\theta)p_Y(y;F_1)+\theta p_Y(y;F_2)\right)\right]\mathrm{d}F_1(x)\mathrm{d}y\right\}\\
&\quad+\int_{\mathbb{S}} i(x;F_1)\mathrm{d}F_2(x)-I(F_1)\\
&\overset{(b)}{=} \lim_{\theta\to 0}\left\{\frac{1}{\theta}\int_{\mathbb{R}}\int_{\mathbb{S}} p_{Y|X}(y|x)\Big[\log p_Y(y;F_1)-\log p_Y(y;F_1)\right.\\
&\quad\left.-\frac{1}{\ln 2}\theta\frac{-p_Y(y;F_1)+p_Y(y;F_2)}{p_Y(y;F_1)}\Big]\mathrm{d}F_1(x)\mathrm{d}y\right\}\\
&\quad+\int_{\mathbb{S}} i(x;F_1)\mathrm{d}F_2(x)-I(F_1)\\
&=\frac{1}{\ln 2}\int_{\mathbb{R}}\int_{\mathbb{S}} p_{Y|X}(y|x)\left(1-\frac{p_Y(y;F_2)}{p_Y(y;F_1)}\right)\mathrm{d}F_1(x)\mathrm{d}y\\
&\quad+\int_{\mathbb{S}} i(x;F_1)\mathrm{d}F_2(x)-I(F_1)\\
&=\int_{\mathbb{S}} i(x;F_1)\mathrm{d}F_2(x)-I(F_1)\\
&\quad+\frac{1}{\ln 2}\int_{\mathbb{R}} p_Y(y;F_1)\left(1-\frac{p_Y(y;F_2)}{p_Y(y;F_1)}\right)\mathrm{d}y\\
&=\int_{\mathbb{S}} i(x;F_1)\mathrm{d}F_2(x)-I(F_1)
\end{aligned}
$$

where the exchange between limit and integral in (a) follows from the Lebesgue Dominated Convergence Theorem. In fact, $\forall\theta\in[0,1]$

$$
\begin{aligned}
f_\theta(y,x)p_{X_2}(x) &= p_{Y|X}(y|x)p_{X_2}(x)\log\frac{p_{Y|X}(y|x)}{p_Y(y;F_\theta)}\\
&\le p_{X_2}(x)\left(\left|p_{Y|X}(y|x)\log p_{Y|X}(y|x)\right|\right.\\
&\quad\left.+\left|p_{Y|X}(y|x)\log p_Y(y;F_\theta)\right|\right)
\end{aligned}
$$

which is integrable on $\mathbb{R}\times\mathbb{S}$,[15] by integrability of $G(y)$, and then also on $\mathbb{S}\times\mathbb{R}$ via Tonelli and Fubini Theorems and due to the fact

[15]Integration on $\mathbb{R}$ produces $p_{X_2}(x)i(x;F_\theta)$ that is integrable on $\mathbb{S}$ due to the boundedness of $i(\cdot)$.

that $f_\theta(y,x)$ converges, for $\theta \to 0$, to $f(y,x) = p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y;F_1)}$. Moreover, (b) follows from the first order McLaurin Series

$$\log(a(1-x)+bx) = \log(a) + \frac{x(-a+b)}{a}\frac{1}{\ln 2} + o(x).$$

2.11.3 An upper and lower bound for the Rayleigh fading conditional output p.m.

In this appendix inequality (2.8) is rigorously proven to be satisfied in case the considered conditional output p.m. and correspondent $Q(y)$ and $q(y)$ are the ones introduced in Section 2.9.2. Concerning the upper bound, one should show that there exist a parameter γ such that

$$\frac{1}{y^{1+\gamma}} > \frac{1}{1+x^2}\exp\left\{-\frac{y}{1+x^2}\right\}, \quad \forall x : 0 \le |x| \le A \tag{2.22}$$

is valid for $y > c(A^2+1)$, where $c > 2$. The considered inequality can be reformulated as follows

$$\frac{y}{(1+x^2)(1+\gamma)} + \frac{\ln(1+x^2)}{1+\gamma} > \ln y.$$

To guarantee the inequality to be fulfilled even in the worst case, the left hand side (x is confined in it) can be studied, for each fixed y, to find out that $\sqrt{y-1}$ is its minimum in x, provided $y \ge 1$. Moreover, if $y \ge A^2+1$ the minimum becomes $x = A$, since x is bounded and $\sqrt{y-1}$ is unreachable in this case. The minimum expression for $y \ge A^2+1$ is

$$\frac{y}{(1+A^2)(1+\gamma)} + \frac{\ln(1+A^2)}{1+\gamma}$$

which has constant derivative in y, equalling the derivative of $\ln y$ taken at $y = (1+A^2)(1+\gamma)$. Now, consider $y_1 = c(1+A^2)$: If constants c and γ are chosen such that $c > 2$ and $\gamma < 1$, then $y_1 > (1+A^2)(1+\gamma)$ and

$$\frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{y}{(1+A^2)(1+\gamma)} + \frac{\ln(1+A^2)}{1+\gamma}\right]\Bigg|_{y_1} > \frac{\mathrm{d}}{\mathrm{d}y}\ln y\Bigg|_{y_1}.$$

This ensures the derivative of $\frac{y}{(1+A^2)(1+\gamma)} + \frac{\ln(1+A^2)}{1+\gamma}$ to be greater than the one of $\ln y$, which is decreasing, for $y > y_1$. If, finally, it is possible to derive a condition on γ to provide that

$$\left(\frac{y}{(1+A^2)(1+\gamma)} + \frac{\ln(1+A^2)}{1+\gamma} \right) \Bigg|_{y_1} > \ln y|_{y_1} \tag{2.23}$$

the original assertion (2.22) would be satisfied. This is indeed possible since (2.23) becomes

$$\frac{c}{1+\gamma} + \frac{\ln(1+A^2)}{1+\gamma} > \ln(1+A^2) + \ln c$$

which is satisfied for $\gamma < \frac{c - \ln c}{\ln[c(1+A^2)]}$. Any choice of γ such that

$$\gamma < \min \left\{ 1, \frac{c - \ln c}{\ln \left[c(1+A^2) \right]} \right\}$$

would fulfil the scope. Consequently the definition

$$Q(y) = \begin{cases} 1, & 0 \leq y \leq c(A^2+1) \\ \frac{1}{y^{1+\gamma}}, & y > c(A^2+1) \end{cases}$$

is well posed since it guarantees the right hand side of inequality (2.8) to be respected.

Concerning the lower bound $q(y)$, one has to prove that it coincides with the output p.m. conditioned by the maximum input up to $y_2 = \frac{(1+A^2)\ln(1+A^2)}{A^2}$ and that it coincides with the output p.m. conditioned by the minimum input after that same y_2. To do that, consider the intersection between $\frac{1}{1+x^2} \exp\left\{-\frac{y}{1+x^2}\right\}$ and $\exp\{-y\}$ which is given by

$$y(x) = \frac{(1+x^2)\ln(1+x^2)}{x^2}.$$

This intersection is non decreasing in x for $0 \leq x \leq A$ (only positive values are admissible for x, deriving from normalization in [43]) since

$$\frac{\mathrm{d}y(x)}{\mathrm{d}x} = \frac{2}{x^3} \left[x^2 - \ln(1+x^2) \right] \geq 0$$

this meaning that it is maximum for $x = A$. This duly proves that the output p.m. conditioned by the maximum input lies under all the other conditional output p.m. up to its intersection with $\exp\{-y\}$ in $y_2 = \frac{(1+A^2)\ln(1+A^2)}{A^2}$, while afterwards the same role is taken by $\exp\{-y\}$. This finally proves that

$$q(y) = \begin{cases} \frac{1}{1+A^2}\exp\left\{-\frac{y}{1+A^2}\right\}, & 0 \le y \le \frac{(1+A^2)\ln(1+A^2)}{A^2} \\ \exp\{-y\}, & y > \frac{(1+A^2)\ln(1+A^2)}{A^2} \end{cases}$$

is also well posed, fulfilling the left hand side of (2.8).

2.11.4 Useful Theorems

This appendix provides a collection of theorem statements (along with the appropriate references) that are used throughout this chapter.

Theorem 3 (Helly Weak Compactness Theorem [55])**.** Every sequence of c.d.f.s is weakly compact.[16]

Theorem 4 (Helly-Bray Theorem [55])**.** If g is continuous and bounded on $\mathbb{R}^n$, then $F_n \xrightarrow[n]{c} F$ up to additive constants implies $\int g\,\mathrm{d}F_n \to \int g\,\mathrm{d}F$.

This theorem is formulated in terms of complete convergence, but complete convergence is equivalent to Lèvy convergence in $\mathscr{F}$.

Theorem 5 (Differentiation Lemma [52])**.** Let I be an interval of real numbers, eventually infinite, and U be an open set of complex numbers. Let $f = f(t,z)$ be a continuous function on $I \times U$. Assume 1) for each compact subset $\mathbb{K}$ of U the integral $\int_I f(t,z)\mathrm{d}t$ is uniformly convergent for $z \in \mathbb{K}$, 2) for each t, the function $z \mapsto f(t,z)$ is analytic, then the integral function $F(z) = \int_I f(t,z)\mathrm{d}t$ is analytic on U.

[16] Recall that a set is said to be *compact*, in the sense of a type of convergence, if every infinite sequence in the set contains a subsequence which is convergent in that same sense [55].

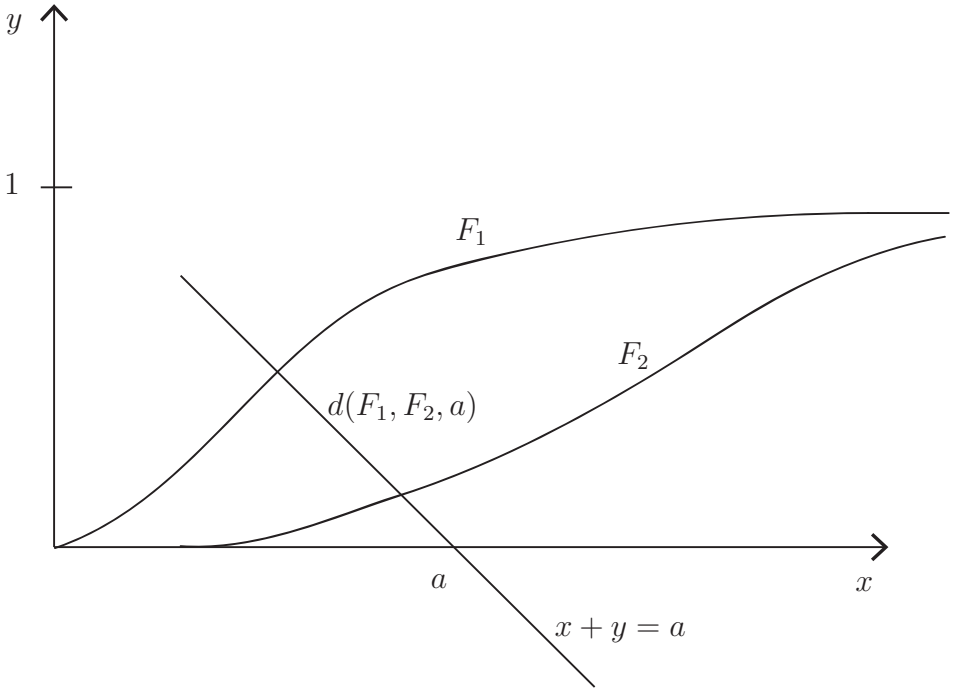

Figure 2.2: Illustrative example for the definition of the Lèvy metric

Theorem 6 ([53], Proposition 1, Chapter 7.8)**.** Let f be a concave functional defined on a convex subset C of a normed space. Let $\mu = \sup_{x \in C} f(x)$. Then

1. The subset Ω of C where $f(x) = \mu$ is convex.

2. If x_0 is a local maximum of $f(\cdot)$, then $f(x_0) = \mu$ and, hence x_0 is a global maximum.

2.11.5 Definition of the Lèvy metric

This section provides the definition for the Lèvy metric, which is taken from [41]. In general it is possible to define this metric on the space of all c.d.f.s. Here, it is anyway restricted to the class $\mathscr{F}$. Consider any two c.d.f.s in $\mathscr{F}$, F_1, F_2 and any $a \in \mathbb{R}$. Denote $d(F_1, F_2, a)$ the a-dependent slant distance between the intersections of the line $x + y = a$ and F_1 and F_2, respectively (see also Fig. 2.2). The Lèvy

distance between F_1 and F_2 is defined as

$$d(F_1, F_2) = \frac{1}{\sqrt{2}} \sup_{a \in \mathbb{R}} d(F_1, F_2, a). \tag{2.24}$$

Part II

Energy efficient distributed estimation

Chapter 3

Distributed Computation of Exact Non-Asymptotic Confidence Regions

3.1 Introduction

A WSN consists of energy-limited sensing devices deployed to collaborate in performing a common task. Examples may be the monitoring of an environmental parameter (*e.g.* temperature or pressure [5, 56, 57]), the detection of a binary event [58], the estimation of a spatial field [6], the estimation of the coordinates of a signal source [59], etc.

Depending on the specific task requirements (fault tolerance, privacy issues, energy constraints), either a centralized or a distributed approach can be adopted: In the former a central unit is needed, that collects all the information and completes the objective task, whereas in the latter all nodes accomplish the objective task on the basis of the information previously exchanged among them. In the centralized scenario the adoption of efficient routing schemes is of capital importance. Contributions in this sense are the energy-efficient adaptive clustering proposed in [60] and the routing protocols in [10, 61–63], aimed at extending network lifetime.

One of the most studied topic in the WSN literature is the estimation of physical parameters. The literature is mostly focused on the development of some specific estimation techniques, both for

the centralized and distributed approaches. Classical maximum likelihood (ML) or least squares (LS) estimators [64] work under the hypothesis of having all the required observations available at one central unit. The scarce robustness to central unit failures and poor network scalability have brought to consideration of distributed approaches. For instance, [65,66] address recursive weighted LS estimation, alongside a consensus-based algorithm that allows to incorporate information from neighbor nodes in the local estimate. A similar approach is taken within the Bayesian framework in [67–69], where consensus-based distributed Kalman filtering is proposed.

The computation of confidence regions has been less considered: In some applications, however, (*e.g.,* in source localization) the derivation of the confidence region is as important as the determination of the estimate. Classical Cramér-Rao-like bounds have been proposed, for instance, in [70–73]. Confidence regions can also be derived as a by-product of the application of Kalman filtering [68,69]. However, strong assumptions on measurement noise (typically Gaussian) are necessary and a good characterization of confidence regions is only possible for a large number of measurements (asymptotic regime).

If attention is restricted the to the centralized setup, the derivation of confidence regions in the non-asymptotic regime has been proved to be possible using, for example, the results in [74–77]. Specifically, [74,75] and [76] respectively propose the Leave-out Sign-dominant Correlation Regions (LSCR) and the sign perturbed sums (SPS) methods. These algorithms allow a central unit to derive a confidence region, from a finite set of measurements, and obtain the *exact* probability that the true parameter value falls within it. In [77], an efficient centralized computation of confidence regions is obtained using interval analysis techniques. Differently from Cramér-Rao-like bounds, the SPS algorithm does not require precise statistical knowledge of the noise, and works under very mild assumptions on its distribution.

3.1.1 Main Contributions

This chapter aims at developing distributed versions of the SPS algorithm. The main advantage of the distributed computation lies in its reliability with respect to central unit failures. Moreover, the com-

putation of a confidence region allows an assessment of quality in the estimation procedure that the point estimate alone cannot guarantee. The last benefit, that the reader should bear in mind, is that the here treated confidence regions are non-asymptotic, hence derived from a finite set of measurements.

To ensure that the confidence region computed by each node is similar in shape to the one that would be evaluated in a centralized setup, nodes have to share their local information with one another. The way of diffusing information drastically impacts on the amount of data exchanged, that has to be kept as low as possible. For this reason, several information diffusion strategies are analyzed and compared in the following. A novel information diffusion strategy, named tagged and aggregated sums (TAS), is presented. It exploits the peculiarities of the SPS algorithm, allowing a reduction of the amount of information to be exchanged among nodes. Its performance is compared to that of established information diffusion strategies, such as flooding [56, 78] and consensus algorithms [67], in terms of generated traffic load as well as confidence region volume/traffic trade-off. Performance predictions and simulation results are provided for various topologies. The chapter presents the results that were treated in [31–33].

Constraints on traffic load may lead to information diffusion truncation: Certain nodes might hence compute a confidence region with partial data. However, it is here proven that consistent non-asymptotic confidence regions can be computed, even starting from an incomplete set of measurements. This constitutes a further contribution of this chapter, hinting in the direction of ensuring the robustness of the proposed distributed approach.

The remainder is organized as follows. Section 3.2 defines the measurement model and recalls the SPS algorithm. Section 3.3 presents several information diffusion strategies. The computation of non-asymptotic confidence regions, from an incomplete set of measurements, is analyzed in Section 3.4. Information diffusion techniques are compared on various network topologies in Sections 3.5 and 3.6. Conclusions are drawn in Section 3.7.

3.1.2 Notation

In this chapter, RVs are indicated with capital roman or greek letters. Their realizations are denoted by the corresponding lowercase letters. Vectors are denoted by bold letters, being lowercase or uppercase according to their random or deterministic nature, while matrices are indicated with bold capital letters.

3.2 Recalls

3.2.1 Measurement model

This section introduces the measurement model on which the following analysis relies. Consider some spatial field described by the parametric model [79]

$$y_{\mathrm{m}}(\mathbf{x}, \mathbf{p}) = \boldsymbol{\varphi}^T(\mathbf{x})\,\mathbf{p}, \tag{3.1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ represents some vector of experimental conditions (time, location, ...) under which the field is observed, $\boldsymbol{\varphi}(\mathbf{x})$ is some regressor function, and $\mathbf{p}$ is the vector of unknown parameters, belonging to the parameter space $\mathbb{P} \subset \mathbb{R}^{n_p}$.

Measurements are taken by a network of N sensor nodes, spread at random locations $\mathbf{x}_i \in \mathbb{R}^{n_x}$, $i = 1, \ldots, N$. Each sensor collects its scalar measurement y_i according to the local measurement model

$$Y_i = y_{\mathrm{m}}(\mathbf{x}_i, \mathring{\mathbf{p}}) + W_i = \boldsymbol{\varphi}_i^T \mathring{\mathbf{p}} + W_i, \tag{3.2}$$

where $\boldsymbol{\varphi}_i^T = \boldsymbol{\varphi}^T(\mathbf{x}_i)$ is the regressor vector at $\mathbf{x}_i$, assumed to be known at the corresponding node i, $\mathring{\mathbf{p}}$ is the true value of the parameter vector and W_i is a random variable representing the measurement noise. The only assumption on W_i's is that they are independent from node to node with a distribution, whichever its shape, symmetric with respect to zero.

3.2.2 The SPS algorithm

As starting point, for the distributed computation of confidence regions, the centralized SPS algorithm [76] is recalled. SPS assumes

all measurements and regressors to be known to a central processing unit and returns the exact confidence region around the least squares estimate $\widehat{\mathbf{p}}$ of $\mathring{\mathbf{p}}$, obtained as the solution of the normal equations $\sum_{i=1}^{N} \boldsymbol{\varphi}_i \left(y_i - \boldsymbol{\varphi}_i^T \mathbf{p}\right) = \mathbf{0}$. Specifically, [76] introduces the unperturbed sum

$$\mathbf{S}_0(\mathbf{p}) = \sum_{i=1}^{N} \boldsymbol{\varphi}_i \left(Y_i - \boldsymbol{\varphi}_i^T \mathbf{p}\right) \tag{3.3}$$

and the $m-1$ sign-perturbed sums, for some m, with $2 \leq m \leq N$,

$$\mathbf{S}_j(\mathbf{p}) = \sum_{i=1}^{N} A_{j,i} \boldsymbol{\varphi}_i \left(Y_i - \boldsymbol{\varphi}_i^T \mathbf{p}\right), \; j = 1, \ldots, m-1 \tag{3.4}$$

where $A_{j,i} \in \{\pm 1\}$ are independent random signs.[1] Introducing

$$Z_j(\mathbf{p}) = ||\mathbf{S}_j(\mathbf{p})||_2^2, \qquad j = 0, \ldots, m-1, \tag{3.5}$$

one may define the set

$$\begin{aligned} \boldsymbol{\Sigma}_q &= \{\mathbf{p} \in \mathbb{P} | Z_0(\mathbf{p}) \text{ is not among the } q \text{ largest } Z_j(\mathbf{p})\} \\ &= \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \sum_{j=1}^{m-1} \mathbb{I}(Z_j(\mathbf{p}) - Z_0(\mathbf{p})) \geq q \right\}, \end{aligned} \tag{3.6}$$

where $\mathbb{I}(\cdot)$ is the indicator function on positive reals. In [76], it was proven that

$$\text{Prob}(\mathring{\mathbf{p}} \in \boldsymbol{\Sigma}_q) = 1 - \frac{q}{m}. \tag{3.7}$$

As a consequence $\boldsymbol{\Sigma}_q$ is a non-asymptotic (*i.e.* derived from a finite set of measurements) confidence region with confidence level $1 - q/m$. As a clarification example, consider the one depicted in Fig. 3.1.[2] There it is considered the computation of the confidence region for a one dimensional parameter. One may observe as the comparison of the computed Z_j leads to the definition of the confidence region. Two confidence regions with different confidence levels are illustrated.

In the following, the distributed computation of $\boldsymbol{\Sigma}_q$ will be addressed considering different information diffusion strategies.

[1] A random sign is a symmetric ± 1 valued random variable taking both values with the same probability.

[2] Thanks to Professor M. Kieffer for having kindly conceded the use of this picture.

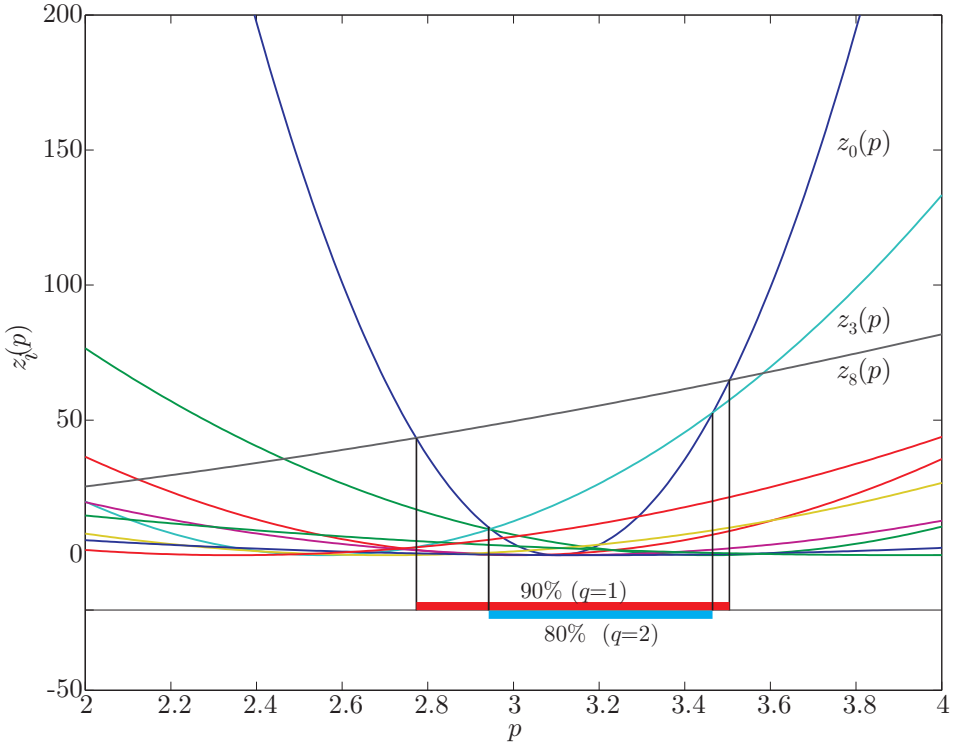

Figure 3.1: An illustrative example of the working principle of the SPS algorithm.

3.3 Information Diffusion Algorithms

This section describes concurrent procedures for information diffusion adapted to SPS. The purpose is to let each node capable of collecting the largest amount of measurements y_i and regressors φ_i possibly with the lowest amount of data exchanged in the network. This section starts by describing two state-of-the-art information diffusion algorithms (Plain Flooding (PF) and consensus algorithms) while the remainder of it is dedicated to novel contributions.

3.3.1 State-of-the-Art Algorithms

This subsection is devoted to providing descriptions of two state-of-the-art algorithms for information diffusion.

3.3.1.1 Plain Flooding (PF)

When adopting this simple information diffusion strategy [31, 56, 78], the generic node i initially broadcasts its own state $\mathbf{x}_i^{(0)} = \left[\boldsymbol{\psi}_i^T, y_i\right]$ at time 0. As next step, it collects the states from its neighbors, then forwards a new data packet containing the aggregated state $\mathbf{x}_i^{(1)} = \left[\left(\boldsymbol{\psi}_i^T, y_i\right), \left\{\left(\boldsymbol{\psi}_j^T, y_j\right)\right\}_{j\in\mathcal{N}_i}\right]$, where $\left\{\left(\boldsymbol{\psi}_j^T, y_j\right)\right\}_{j\in\mathcal{N}_i}$ denotes the set of regressor and measurement pairs collected from nodes j in the neighborhood $\mathcal{N}_i$ of node i. As successive step, the i-th node transmits all the data received in the previous communication rounds. This process is repeated until each node in the network has collected the state from all nodes. Afterwards, each node is able to compute the perturbed and unperturbed sums in (3.3) and (3.4) for any $\mathbf{p}$, and hence derive the confidence region. No transmission of $\alpha_{j,i}$ is necessary, provided that all nodes agree on seeds for their random generators. Note that each node has to transmit a packet containing $D_{\mathrm{f}}^{(0)} = n_p + 1$ values during the first communication round and $D_{\mathrm{f}}^{(\mathrm{last})} = N(n_p + 1)$ values when the last communication occurs. This strategy is the most trivial one but does not result to be particularly efficient on lossless networks, as later explained.

3.3.1.2 Consensus Algorithms

In (3.3) and (3.4), one can see that the computation of $\mathbf{s}_0(\mathbf{p})$ and $\mathbf{s}_j(\mathbf{p})$ does not necessarily require the knowledge of each local quantity, but rather of aggregated values. This suggests the adoption of consensus strategies that are suited for the distributed computation of sums or averages.

A consensus scheme may be viewed as the following discrete-time evolving system [67, 80–83]

$$\mathbf{x}^{(k+1)} = \mathbf{W}\,\mathbf{x}^{(k)} \tag{3.8}$$

where $\mathbf{x}^{(k)} = \left[\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \ldots, \mathbf{x}_N^{(k)}\right]^T$ is the global state at time k, whose i-th entry is the local state at node i. $\mathbf{W}$ is the system dynamics matrix that depends on the chosen consensus algorithm as well as on the network topology. The properties of $\mathbf{W}$ are such to asymptotically

lead to a global state whose entries are all equal to the average of the initial quantities $\mathbf{x}^{(0)}$ [67, 81]. In particular, the convergence of a consensus algorithm is assured by the following three necessary and sufficient conditions:

$$\mathbf{1}^T\mathbf{W} = \mathbf{1}^T, \quad \mathbf{W}\mathbf{1} = \mathbf{1}, \quad \rho(\mathbf{W} - \mathbf{1}\mathbf{1}^T/N) < 1 \tag{3.9}$$

where ρ denotes the spectral radius [81]. Two possible consensus matrices are presented: For a more detailed comparison of convergence speed (and convergence analysis in general) the reader may refer to [67, 81]. The first presented matrix is the Metropolis matrix whose entries are given by

$$w_{i,j} = \begin{cases} \frac{1}{1+\max\{d_i,d_j\}}, & \text{if } (i,j) \in \mathcal{E}, \\ 1 - \sum\limits_{(i,q)\in\mathcal{E}} w_{i,q}, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \tag{3.10}$$

where $d_i = |\mathcal{N}_i|$ denotes the degree of node i (*i.e.*, the number of neighbors) and $\mathcal{E}$ is the set of edges in the network topology. An alternative choice is represented by setting $\mathbf{W}$ as a Perron matrix,

$$\mathbf{W} = \mathbf{I} - \epsilon\mathbf{L}, \tag{3.11}$$

where $\mathbf{I}$ denotes the identity matrix, while $\mathbf{L}$ is the network graph Laplacian matrix given by

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \tag{3.12}$$

with $\mathbf{D}$ being the diagonal matrix $\mathbf{D} = (d_1, d_2, \ldots, d_N)$ and $\mathbf{A}$ representing the network graph adjacency matrix whose entries are

$$a_{i,j} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.13}$$

The constant ϵ should be set such that it fulfills the condition

$$0 < \epsilon < \frac{1}{\Delta}, \tag{3.14}$$

where $\Delta = \max_i d_i$.

According to (3.8) and independently of the chosen $\mathbf{W}$, each node performs its local state update as follows,

$$\mathbf{x}_i^{(k+1)} = \sum_{j=1}^{N} w_{i,j}\mathbf{x}_j^{(k)} . \tag{3.15}$$

To apply the consensus algorithm to the problem of the distributed computation of (3.3) and (3.4), consider the following averages that do not depend on $\mathbf{p}$

$$\mathbf{b}_0 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\psi}_i y_i \qquad \mathbf{b}_j = \frac{1}{N}\sum_{i=1}^{N} \alpha_{j,i}\boldsymbol{\psi}_i y_i \tag{3.16}$$

$$\mathbf{A}_0 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \qquad \mathbf{A}_j = \frac{1}{N}\sum_{i=1}^{N} \alpha_{j,i}\boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \tag{3.17}$$

for $j = 1, 2, \ldots, m-1$.

The consensus algorithm is launched on all $\mathbf{A}_j$'s and $\mathbf{b}_j$'s (including $\mathbf{A}_0$ and $\mathbf{b}_0$). At step $k = 0$, the local state at node i is given by $\mathbf{x}_i^{(0)} = \left[(\boldsymbol{\psi}_i y_i)^T, \left\{\boldsymbol{\psi}_i\boldsymbol{\psi}_i^T\right\}, \left\{(\alpha_{j,i}\boldsymbol{\psi}_i y_i)^T\right\}_j, \left\{\alpha_{j,i}\boldsymbol{\psi}_i\boldsymbol{\psi}_i^T\right\}_j\right]^T$, with $j = 1, 2, \ldots, m-1$, that is, the single addends in (3.16) and (3.17). The matrices entries should be indexed so that $\mathbf{x}_i^{(0)}$ results to be a column vector. At each successive step the ith node updates its own state according to (3.15). Once a consensus on $\mathbf{A}_j$ and $\mathbf{b}_j$ is reached, each node is able to locally evaluate (3.3) and (3.4) for any value $\mathbf{p}$ in the parameter search space. It is worth noting that no particular value of the parameter $\mathbf{p}$ has to be transmitted and that getting averages, instead of the true sums, does not affect the SPS algorithm since the comparison of rescaled norms or norms gives the same ordering in (3.6). Therefore the algorithm works also without the knowledge of N. Note that the state is composed of $D_c = m(3n_p + n_p^2)/2$ values, where symmetry of $\boldsymbol{\psi}_i\boldsymbol{\psi}_i^T$ is exploited. The state dimension is constant during the entire running of the algorithm but is larger than the one initially required by flooding.

3.3.2 Mixed Flooding+Consensus Approach

As observed, the pure average consensus algorithm requires an amount of data to be transmitted that is initially strictly larger than that required by flooding. However, in few iterations the amount of data transmitted with flooding exceeds D_{c} due to data accumulation at nodes. This fact suggested the introduction of a third mixed strategy conceived as follows: flooding, as described in Section 3.3.1.1, works until at least one node i experiments an amount of data that exceeds D_{c}, that is $n_i^{(k)} D_{\mathrm{f}}^{(0)} > D_{\mathrm{c}}$, where $n_i^{(k)}$ is the number of distinguished received data at node i, at iteration k. When this happens, all nodes switch to the consensus strategy described in Section 3.3.1.2. The correct initialization, for the consensus state $\mathbf{x}_i^{(0)}$, is to be set as the average of the quantities received during the initial flooding. Numerical results will show that, when using this strategy, a benefit is possible. Reason for this is that, for the first iterations, the mixed approach behaves like flooding, that is always initially advantageous, and then behaves like consensus, that, even if asymptotically worse than flooding, ensures a gain in performance when the number of data that can be exchanged is limited (and not almost asymptotic).

3.3.3 Modified Flooding (MF)

For the last two algorithms to be described, the evolution of the amount of information available at a node k can be represented by means of a table $\mathbf{R}^{(k)}$. The construction of $\mathbf{R}^{(k)}$ and the transmission of information depend on the considered procedure.

The main difference between MF and PF is that, in the former, an information already transmitted by a node is never transmitted again by the same node. This kind of behavior is certainly efficient in terms of amount of data to be transmitted on lossless links.

The MF algorithm generates at runtime a table of contents available at nodes. An example for this is depicted in Table 3.1. This table gathers the information collected at node $k = 1$ in a network composed of 7 nodes. Each row r in $\mathbf{R}^{(k)}$ contains an available information $\mathbf{D}_r^{(k)}$ and its related tag, indicating the originating node. When performing the MF algorithm, the generic node k initially fills the first line of $\mathbf{R}^{(k)}$ with its own local information, *i.e.*, $\mathbf{D}_1^{(k)} = \left[\boldsymbol{\varphi}_k^T, y_k\right]$ and the

$\mathbf{R}^{(1)} =$

$\mathbf{D}_1^{(1)}$	1	0	0	0	0	0	0
$\mathbf{D}_2^{(1)}$	0	0	1	0	0	0	0
$\mathbf{D}_3^{(1)}$	0	0	0	0	0	1	0
$\mathbf{D}_4^{(1)}$	0	1	0	0	0	0	0
$\mathbf{D}_5^{(1)}$	0	0	0	1	0	0	0
$\mathbf{D}_6^{(1)}$	0	0	0	0	0	0	1
$\mathbf{D}_7^{(1)}$	0	0	0	0	1	0	0

Table 3.1: Table $\mathbf{R}^{(1)}$ of available information at node $k = 1$ when MF is used for information diffusion.

corresponding tag[3] $\mathbf{t}_1^{(k)}$ having a single 1 at the k-th entry. It then broadcasts $\left\{\mathbf{D}_1^{(k)}, \mathbf{t}_1^{(k)}\right\}$ and marks the line as already transmitted. As next step, it collects the data coming from neighbors and inserts in the table this new acquired information, thus creating a set of rows corresponding to its set of neighbor nodes, denoted as $\mathcal{N}_k$. Then it forwards a new data packet containing the data of all lines in $\mathbf{R}^{(k)}$ which were not marked as already transmitted. This means that the second message that node k transmits contains $\left\{\mathbf{D}_j^{(k)}, \mathbf{t}_j^{(k)}\right\}_{j\in\mathcal{N}_k}$. All rows whose data have been transmitted are then marked. The iteration of the procedure yields, at the next transmission step, a message to be transmitted containing only information never previously transmitted. This process terminates when each node in the network has collected the information from all nodes. Section 3.4 analyses the case when all data cannot be gathered at all nodes due, e.g., to delay/traffic constraints.

Afterwards, each node is able to compute the perturbed and unperturbed sums in (3.3) and (3.4) for any $\mathbf{p}$, and hence derive the confidence region. During the first iteration, each node has to transmit a packet containing

$$d_{MF} = n_p + 1 \tag{3.18}$$

[3]The tag matrix is denoted by $\mathbf{T}^{(k)}$ and its r-th row by $\mathbf{t}_r^{(k)}$.

real values. The dimension of successive data packets is an integer multiple of this value, possibly zero.

Remark 2. *If all nodes agree on their random generators seed, the computed confidence regions are the same at all nodes without any need for transmission of $A_{j,i}$. In case this agreement is lacking, still transmission of $A_{j,i}$ can be avoided, but the shape of confidence regions computed at different nodes may differ.*

3.3.4 Tagged and aggregated sums (TAS) algorithm

Before coming to the detailed description of the TAS algorithm, a preliminary consideration is needed. Expanding a realization of (3.3) and (3.4) one gets,

$$\mathbf{s}_0(\mathbf{p}) = \sum_{k=1}^{N} \boldsymbol{\varphi}_k y_k - \left(\sum_{k=1}^{N} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T \right) \mathbf{p} \tag{3.19}$$

$$\mathbf{s}_j(\mathbf{p}) = \sum_{k=1}^{N} a_{j,k} \boldsymbol{\varphi}_k y_k - \left(\sum_{k=1}^{N} a_{j,k} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T \right) \mathbf{p}, \quad j = 1, \ldots, m-1. \tag{3.20}$$

The evaluation of (3.19) and (3.20) for any value of $\mathbf{p} \in \mathbb{P}$ does not necessarily require the availability of each individual term in the sums but rather of

$$\left\{ \sum_{k=1}^{N} \boldsymbol{\varphi}_k y_k, \sum_{k=1}^{N} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T, \left\{ \sum_{k=1}^{N} a_{j,k} \boldsymbol{\varphi}_k y_k \right\}_{\forall j}, \left\{ \sum_{k=1}^{N} a_{j,k} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T \right\}_{\forall j} \right\}. \tag{3.21}$$

Therefore, at each information diffusion step, the available information can be composed into an aggregated sum, reducing the traffic load. This is the peculiarity of the SPS algorithm that can be exploited by both TAS and consensus algorithms. The main difficulty lies in avoiding the same term to appear more than once in each sum, independently of network topology. This consideration led to the formulation of the TAS algorithm whose details follow.

The TAS algorithm consists of six phases, namely, i) initialization, ii) reception, iii) distillation, iv) aggregation, v) transmission, and vi) wrap-up.

i) *Initialization phase.* During the initialization phase each node $k \in \{1, ..., N\}$ creates and transmits a data packet which consists of the first row of its table $\mathbf{R}^{(k)}$. This first row is composed of:

- a *data set* $\mathbf{D}_1^{(k)} = \left\{ \boldsymbol{\varphi}_k y_k, \left\{ \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T \right\}, \left\{ a_{j,k} \boldsymbol{\varphi}_k y_k \right\}_{\forall j}, \left\{ a_{j,k} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T \right\}_{\forall j} \right\}$, corresponding to the local quantities related to node k. This set consists of

$$d_{\text{TAS}} = m \left(n_p + n_p \frac{n_p + 1}{2} \right) \tag{3.22}$$

  real values, taking into account the symmetry of $\boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T$.

- a *tag vector* $\mathbf{t}_1^{(k)}$, that is an all-zero vector except for the k-th entry where a 1 is located.

After initialization, the reception, distillation, aggregation, and transmission phases are sequentially repeated until a termination condition is met. Typically this conditions can be set as the reaching of a maximum allowed number of iterations. Within each cycle, new rows $\{\mathbf{D}_r^{(k)}, \mathbf{t}_r^{(k)}\}$ are possibly added to $\mathbf{R}^{(k)}$, with $r > 1$ representing the row number.

For $r > 1$, the r-th *data set* $\mathbf{D}_r^{(k)}$ can either contain the local quantities related to another node or the sum of quantities related to several nodes, as specified in $\mathbf{t}_r^{(k)}$. The dimension of data sets obtained as sums of initial data sets does not vary and stays equal to d_{TAS}.

ii) *Reception phase.* During this phase each node collects *messages* transmitted by its neighbors. The message $\mathbf{m}^{(n)}$ coming from node n, with $n \neq k$, consists of a *data set* and a *tag vector*.

iii) *Distillation phase.* At the end of the reception phase, at each node, the received *tag vectors* and those already stored in $\mathbf{R}^{(k)}$ are compared to detect whether the received data contains new information. More precisely, the received *tag vector* of each incoming *message* is compared to all the already available *tag vectors* $\mathbf{t}_r^{(k)}$ contained in

$\mathbf{D}_1^{(1)}$	1	0	0	0	0	0	0
$\mathbf{D}_2^{(1)}$	0	0	1	0	0	0	0
$\mathbf{D}_3^{(1)}$	0	0	0	0	0	1	0
$\mathbf{D}_4^{(1)}$	0	1	0	0	0	0	1
$\mathbf{D}_5^{(1)}$	0	0	0	1	0	0	0
$\mathbf{D}_6^{(1)}$	0	0	0	0	1	0	1
$\mathbf{D}_7^{(1)}$	0	1	0	0	1	0	0

Table 3.2: Table of available information at node $k = 1$ when information diffusion is done via the TAS algorithm.

$\mathbf{R}^{(k)}$. If a received *message* does not contain any new contribution, it is discarded, otherwise a new row is added to $\mathbf{R}^{(k)}$, containing the new information contribution, that is, the received *message* (*data set+tag vector*) duly polished of already available information.

Example 1: if node k receives a *message* containing the sum of quantities originating from nodes 1, 2, 7, 8, 11 and if it has already rows in $\mathbf{R}^{(k)}$ containing the information relative to node 1 and to the sum of local quantities of nodes 2 and 7, it can successfully detect the sum of quantities related to nodes 8 and 11 and insert them in the table. Only this distilled information, composed of a new *data set* and its corresponding *tag vector* (having the 8th and 11th bits set to 1) is added to $\mathbf{R}^{(k)}$.

iv) *Aggregation phase*. To form the next packet to transmit, each node aggregates the information contained in $\mathbf{R}^{(k)}$, summing the available *data sets* and merging the related *tag vectors*. The merge is done as follows. When the aggregation phase takes place for the first time, each node k initializes a *temporary data set* $\mathbf{D}_T^{(k)}$ and the related N elements *temporary tag vector* $\mathbf{t}_T^{(k)}$ with the content of the first row of $\mathbf{R}^{(k)}$ and marks this row as already merged. Then, the node checks whether the next row contains some information that is already accounted for in $\mathbf{t}_T^{(k)}$. If not, it marks it as already merged and sums its corresponding *data set* to $\mathbf{D}_T^{(k)}$ and updates $\mathbf{t}_T^{(k)}$. All rows are then progressively examined. Successive aggregation phases initialize $\mathbf{D}_T^{(k)}$

and $\mathbf{t}_T^{(k)}$ as the content of the first never merged row, starting the search from the first row.

Example 2: Consider $\mathbf{R}^{(1)}$ reported in Table 3.2. Node $k = 1$ has to compose the first *message* that it should transmit. It starts from the first row and initializes $\mathbf{D}_T^{(1)}$ and $\mathbf{t}_T^{(1)}$ with the content of the first row. It then marks the first row as already merged. The second row is then examined. As it contains only new information, with respect to the content of $\mathbf{t}_T^{(1)}$, its corresponding *data set* is added to $\mathbf{D}_T^{(1)}$ and its *tag vector* is merged with $\mathbf{t}_T^{(1)}$, resulting in $\mathbf{t}_T^{(1)} = (1, 0, 1, 0, 0, 0)$ and $\mathbf{D}_T^{(1)} = \mathbf{D}_1^{(1)} + \mathbf{D}_2^{(1)}$. The same happens for the third, fourth and fifth rows, that are then all marked as already merged. The sixth row contains, instead, information relative to node 7: Node 7 is already contributing to the current $\mathbf{t}_T^{(1)}$, thus, $\mathbf{D}_6^{(1)}$ is not added to $\mathbf{D}_T^{(1)}$. Afterwards, $\mathbf{D}_T^{(1)}$ and $\mathbf{t}_T^{(1)}$ are transmitted as definitive message, when all rows of the table have been traversed. When the next aggregation phase takes place, $\mathbf{D}_T^{(1)}$ and $\mathbf{t}_T^{(1)}$ are initialized as the content of the first row that has never been merged in the previous aggregation phases: In the proposed example, this happens for the sixth row.

v) *Transmission phase*. The message obtained at the end of the aggregation phase is broadcasted to all neighbor nodes.

The information diffusion stops after a fixed number of transmission phases: On random networks the limit can be set equal to the diameter of the network (as would be the case for any flooding approach).

vi) *Wrap-up phase*. Once the information diffusion expires, the objective, for any node k, is the computation of (3.21), which is then used to evaluate (3.5). This means finding a strategy to combine the rows in $\mathbf{R}^{(k)}$ to obtain the aggregated data in (3.21). Two cases are possible: Either $rank\left(\mathbf{T}^{(k)}\right) = N$ and then a perfect reconstruction of (3.21) is possible, since each appearing term can be individually retrieved, or $rank\left(\mathbf{T}^{(k)}\right) < N$ and node k will try to close as much as possible on (3.21). This can be realized performing a linear combination of the rows of $\mathbf{R}^{(k)}$, aiming at maximizing the amount of data taken into account.

Each node k will evaluate a linearly weighted sum $\mathbf{D}_F^{(k)} = \sum_r \widehat{b}_{k,r}\mathbf{D}_r^{(k)}$, where $\widehat{\mathbf{b}}_k$ is the solution of the following constrained optimization

problem

$$\widehat{\mathbf{b}}_k = \arg\max_{\mathbf{b}_k} \mathbf{b}_k \mathbf{T}^{(k)} \mathbf{1}, \tag{3.23}$$

$$\text{s.t. } 0 \leq \sum_r b_{k,r} t_{r,i}^{(k)} \leq 1, \quad i = 1, \ldots, N. \tag{3.24}$$

Here, $t_{r,i}^{(k)}$ are the elements of $\mathbf{T}^{(k)}$, with r and i denoting the row and column indexes. The solution of (3.23)-(3.24) is obtained by linear programming.

The term $c_{k,i} = \sum_r b_{k,r} t_{r,i}^{(k)}$ in (3.24) represents the weight of the quantities related to node i. Since local quantities in (3.4) cannot contribute more than once, to keep independence among all terms intervening in (3.4), then it must be $0 \leq c_{k,i} \leq 1$, that determines the constraints (3.24).

Remark 3. *The TAS algorithm takes some inspiration from network coding techniques [84–86]. However, the main difference is that each node does not need to decode, by means of Gaussian elimination, all the individual messages transmitted by the other nodes, but rather the decoding of their sum (possibly of an* incomplete *sum) suffices.*

The performance of the TAS algorithm will be investigated in Sections 3.5 and 3.6.

3.4 Analysis of Information Diffusion Truncation

In this section, the effect of truncation of information diffusion is discussed. The objective is to prove that consistent non-asymptotic confidence regions can still be computed via SPS, at all nodes, even when the information diffusion process is stopped before each node has gathered all data.

To achieve this objective, the truncated expressions of (3.3) and (3.4) are provided first. Then, some other preliminary definitions and recalls are outlined. Last, a theorem closes the section.

Truncating the information diffusion algorithm entails that (3.3) and (3.4) are estimated taking into account only the data actually

received by each node. Hence, at node k, the following quantities are evaluated from the available data

$$\tilde{\mathbf{S}}_{k,0}(\mathbf{p}) = \sum_{i=1}^{N} c_{k,i}\boldsymbol{\varphi}_i \left(Y_i - \boldsymbol{\varphi}_i^T \mathbf{p}\right) \tag{3.25}$$

$$\tilde{\mathbf{S}}_{k,j}(\mathbf{p}) = \sum_{i=1}^{N} c_{k,i}A_{j,i}\boldsymbol{\varphi}_i \left(Y_i - \boldsymbol{\varphi}_i^T \mathbf{p}\right), \tag{3.26}$$

where $j = 1, \ldots, m-1$, and $c_{k,i} \in \{0, 1\}$. The coefficients $c_{k,i}$ reckon with the availability or absence of the i-th measurement, due to truncation, at node k.[4]

Note that (3.25) is the set of normal equations that would be obtained in a centralized context, considering a weighted least-squares estimator, with a diagonal weight matrix $\mathbf{C}_k = \mathrm{diag}\,(c_{k,1}, \ldots, c_{k,N})$. Similarly, (3.26) is the sign perturbed sum that would be obtained when considering weighted least-squares. It will be shown that the confidence region, obtained considering (3.25) and (3.26) in (3.6), is still a non-asymptotic confidence region. Reaching completion of the information diffusion algorithm entails that the $c_{k,i}$ are all equal to one, thus ensuring equivalence with the centralized scenario. In case of truncation, instead, the $c_{k,i}$ fall in the interval $[0, 1]$, their values depending on the applied information diffusion procedure: In case that the TAS or a consensus approach are applied they might take any value in $[0, 1]$, otherwise, with flooding, only 0 and 1 are possible values.

Taking the squared norms of (3.25) and (3.26), respectively named $\tilde{Z}_0(\mathbf{p})$ and $\tilde{Z}_j(\mathbf{p})$, for $j = 1, \ldots, m-1$, allows to define the confidence region that is obtained at node k when truncation occurs, that is,

$$\widetilde{\boldsymbol{\Sigma}}_{q,k} = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \sum_{j=1}^{m-1} \mathbb{I}(\tilde{Z}_j(\mathbf{p}) - \tilde{Z}_0(\mathbf{p})) \geq q \right\}. \tag{3.27}$$

In order to characterize the consistency of $\widetilde{\boldsymbol{\Sigma}}_{q,k}$, that relies on an incomplete set of measurements, it is necessary to recall some definitions taken from [76].

[4]The $c_{k,i}$ differ from the weighting coefficients appearing in [76, Section 2.2]. The difference is that, here, they depend on the measurement index i. This makes the two forms of weighting completely unrelated.

Definition 1 (Symmetric Random Variables)**.** Given a probability space $(\Omega, \mathcal{F}, \mathbb{\Pi})$, Ω being the sample space, $\mathcal{F}$ the σ-algebra of events, and $\mathbb{\Pi}$ the probability measure, a real (possibly $\mathbb{R}^d$-valued) RV X is said to be symmetric around the origin 0 (possibly origin vector $\mathbf{0}$) if

$$\forall A \in \mathcal{F} : \mathbb{\Pi}(X \in A) = \mathbb{\Pi}(-X \in A). \tag{3.28}$$

The following property recalls [76, Lemma 2].

Property 1. Let $A, B_1, \ldots, B_k$ be independent, identically distributed (i.i.d.) random signs. Then $A, AB_1, \ldots, AB_k$ are also i.i.d. random signs.

Definition 2 (Uniformly Ordered Variables)**.** A finite set of real-valued RVs $Z_0, Z_1, \ldots, Z_{m-1}$ is said to be uniformly ordered if for all permutations $i_0, i_1, \ldots, i_{m-1}$ of indexes $0, 1, \ldots, m-1$, one has

$$\mathbb{P}(Z_{i_0} < Z_{i_1} < \ldots < Z_{i_{m-1}}) = \frac{1}{m!}. \tag{3.29}$$

Definition 2 states that all orderings are equiprobable. A direct consequence is that, for a set of uniformly ordered RVs $Z_0, Z_1, \ldots, Z_{m-1}$, each variable Z_i takes any position in the ordering with probability $1/m$.

With the purpose to formulate Lemma 1, introduced in the following, another few more considerations are needed. Let $h(Z_0, Z_1, \ldots, Z_{m-1}) : \mathbb{R}^m \to \mathbb{N}_0^{m-1}$ be a function of m real variables, with $\mathbb{N}_0^{m-1}$ denoting the set of naturals from 0 to $m-1$. The function provides a permutation $i_0, i_1, \ldots, i_{m-1}$, such that $Z_{i_0} \leq Z_{i_1} \leq \ldots \leq Z_{i_{m-1}}$. In case of ties between input variables, the permutation is uniquely determined by applying the following rule. Suppose that n variables are tied: Thus $n!$ orderings are possible. Then $h(\cdot)$ provides a reordering choosing among the possible $n!$ with uniform distribution. Having premised this, when $h(\cdot)$ takes RVs as inputs, it can be considered as a discrete random variable with $m!$ possible outcomes, *i.e.*, as many as the number of possible permutations of m integers.

Lemma 1 (Uniform Ordering Lemma)**.** Let $Z_0, Z_1, \ldots, Z_{m-1}$ be real-valued, i.i.d. RVs. Then they are uniformly ordered.

Proof. Consider $h(\cdot)$, as previously defined. Since $Z_0, Z_1, \ldots, Z_{m-1}$ are i.i.d. the distribution of $h(Z_{i_0}, Z_{i_1}, \ldots, Z_{i_{m-1}})$ is the same for all permutations. Permutations are in number of $m!$, hence each of the outcome of $h(\cdot)$ has probability $1/m!$, since the mechanism, by which $h(\cdot)$ is defined, guarantees that all outcomes are equally possible. This is equivalent to saying that the variables are uniformly ordered. □

Lemma 1 is a generalization, to both continuous and discrete RVs, of [76, Lemma 4], which does not hold for discrete RVs. The need for this extension will appear in the proof of Theorem 7.

Now, one can state the following theorem.

Theorem 7. Under the assumption of measurement noises being symmetric RVs and independent across nodes, the confidence level with which the true parameter value $\mathring{\mathbf{p}}$ falls in the region $\widetilde{\boldsymbol{\Sigma}}_{q,k}$, yielded at node k, is

$$\text{Prob}(\mathring{\mathbf{p}} \in \widetilde{\boldsymbol{\Sigma}}_{q,k}) = 1 - \frac{q}{m}, \tag{3.30}$$

for every $k = 1, \ldots, N$.

Proof. Following a similar approach as in [76], the evaluation of (3.25) and (3.26) for $\mathring{\mathbf{p}}$ gives

$$\tilde{\mathbf{S}}_{k,0}(\mathring{\mathbf{p}}) = \sum_{i=1}^{N} c_{k,i} \boldsymbol{\varphi}_i W_i \tag{3.31}$$

and

$$\tilde{\mathbf{S}}_{k,j}(\mathring{\mathbf{p}}) = \sum_{i=1}^{N} c_{k,i} A_{j,i} \boldsymbol{\varphi}_i W_i, \tag{3.32}$$

with $j = 1, \ldots, m-1$. The truncation results in a rescaling of measurement noise terms W_i, since it only depends on the communication links effectively traversed during the information diffusion phase. This rescaling preserves independence as well as symmetry of noise distributions. Consider, further, that from (3.31) and (3.32), one can derive

$$\tilde{Z}_{k,0}(\mathring{\mathbf{p}}) = \left\| \sum_{i=1}^{N} c_{k,i} \boldsymbol{\varphi}_i W_i \right\|_2^2, \tag{3.33}$$

and

$$\tilde{Z}_{k,j}(\mathring{\mathbf{p}}) = \left\| \sum_{i=1}^{N} c_{k,i} A_{j,i} \boldsymbol{\varphi}_i W_i \right\|_2^2 . \tag{3.34}$$

These last two expressions may be rewritten highlighting the independent random measurement noise terms $W_1, \ldots, W_N$, *i.e.*,

$$\tilde{Z}_{k,0}(\mathring{\mathbf{p}}) = f(c_{k,1} W_1, \ldots, c_{k,N} W_N), \tag{3.35}$$

$$\tilde{Z}_{k,j}(\mathring{\mathbf{p}}) = f(c_{k,1} A_{j,1} W_1, \ldots, c_{k,N} A_{j,N} W_N), \tag{3.36}$$

As already pointed out, each $c_{k,i} W_i$ has a symmetric distribution. By applying Lemma 1 from [76] to the variables in the collection $\{c_{k,i} W_i\}_{i=1}^N$ and introducing the set of random signs $\{B_i\}_{i=1}^N$ one can write $c_{k,i} W_i = B_i (B_i c_{k,i} W_i) = B_i V_i$, where B_i and $V_i = B_i c_{k,i} W_i$ are independent $\forall i$ [76, Lemma 1]. It is possible to compact (3.35) and (3.36) in the single expression

$$\tilde{Z}_{k,j}(\mathring{\mathbf{p}}) = f(D_{j,1} V_1, \ldots, D_{j,N} V_N) \tag{3.37}$$

for $j = 0, \ldots, m-1$, with $D_{0,i} \triangleq B_i$ and $D_{j,i} \triangleq A_{j,i} B_i$ for $j = 1, \ldots, m-1$. The set of RVs $\{D_{j,i}\}_{i=1,\, j=0}^{N,\, m-1}$ is also a collection of i.i.d. random signs, this deriving from Property 1 applied to the i.i.d. random signs B_i and $\{A_{j,i}\}_{j=1}^{m-1}$. Now fix a realization for $\{V_i\}_{i=1}^N$, indicated as $\{v_i\}_{i=1}^N$. Conditioning on $\{V_i\}_{i=1}^N = \{v_i\}_{i=1}^N$, $\left\{ \tilde{Z}_{k,j}(\mathring{\mathbf{p}}) \middle| \{V_i\}_{i=1}^N = \{v_i\}_{i=1}^N \right\}_{j=0}^{m-1}$ is a collection of discrete, real-valued, and i.i.d. RVs, since $\{D_{j,i}\}_{i=1,\, j=0}^{N,\, m-1}$ is a collection of i.i.d. random signs. Applying Lemma 1 to $\left\{ \tilde{Z}_{k,j}(\mathring{\mathbf{p}}) \middle| \{V_i\}_{i=1}^N = \{v_i\}_{i=1}^N \right\}_{j=0}^{m-1}$ leads to the consideration that these variables are uniformly ordered. This implies that the RV $\tilde{Z}_{k,0}(\mathring{\mathbf{p}}) \big| \{V_i\}_{i=1}^N = \{v_i\}_{i=1}^N$ takes each position in the ordering with probability $1/m$. The conclusion is that it is not among the q largest $\tilde{Z}_{k,j}(\mathring{\mathbf{p}}) \big| \{V_i\}_{i=1}^N = \{v_i\}_{i=1}^N$, $j = 0, \ldots, m-1$, with probability $1 - q/m$. Since this probability value is independent of the particular realization of $\{V_i\}_{i=1}^N$, one can apply [76, Lemma 3] to say that Prob($\mathring{\mathbf{p}} \in$

$\widetilde{\mathbf{\Sigma}}_{q,k}) = \text{Prob}\left(\tilde{Z}_{k,0}(\mathring{\mathbf{p}}) \text{ is not among the } q \text{ largest } \tilde{Z}_{k,j}(\mathring{\mathbf{p}})\right) = 1 - \frac{q}{m}$ is valid also when not conditioning on noise realizations. This concludes the proof. □

Remark 4. *Lemma 1 is introduced to prove that i.i.d. discrete RVs are uniformly ordered. From [76, Lemma 4], one can draw this conclusion only for continuous variables. Lemma 1 generalizes [76, Lemma 4]. This copes with the discrete RVs, that are appearing when noise realizations are fixed, as done in the proof.*

Remark 5. *When at sensor node k there is only a single non-zero coefficient, $c_{k,k} = 1$, meaning that truncation in information diffusion occurred before node k could gain knowledge about any other sensor than itself, then, the $m \times N$ matrix formed by all random signs $A_{j,i}$ participating in the confidence region computation at node k has only one column filled with values $\{-1, 1\}$, while all the remaining ones are filled with zeros. Its rank is hence equal to 1 and the norms $\tilde{Z}_{k,j}(\mathbf{p})$ are all equal independently of j and for any value of $\mathbf{p}$. This is certainly the case for which the highest number of ties occurs, nevertheless, choosing at random for the reordering, yields a random confidence region, covering a percentage equal to $1 - q/m$ of the initial search space. The computed confidence region keeps again the same level of confidence $1 - \frac{q}{m}$, as stated by Theorem 7. This observation gives an insight on the reason why the shape of confidence regions is affected by information availability.*

3.5 Analysis of Traffic Load on Generic Topologies

In order to fairly compare different information diffusion strategies, the network traffic burden has to be characterized. The algorithms are compared on specific topologies, such as random trees, with binary trees as a special case, and clustered networks, that are the most commonly used in practical applications [5]. In Section 3.6, completely random networks will also be considered.

Before entering into the details of the analytical investigation, recall that d_{TAS} and d_{MF}, respectively given by (3.22) and (3.18), denote

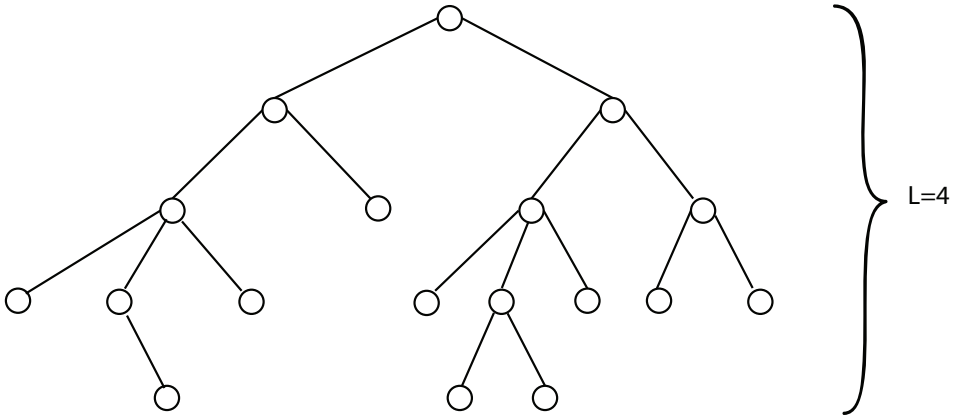

Figure 3.2: A random tree topology. Some of the random variables, that describe it, take values $\Lambda(0) = 1$, $\Lambda(1) = 2$, $\Lambda(2) = 4$, $\bar{\Lambda}(2) = 1$, $\Lambda(3) = 8$, $\bar{\Lambda}(3) = 6$, etc.

the numbers of real-valued scalars that a single data is composed of when the TAS or the MF algorithm are considered. Only these two information diffusion strategies will be analytically compared. This is due to the fact that PF is outperformed by MF on lossless networks while the amount of data that should be transmitted by a consensus algorithm strictly depends on its convergence threshold and is therefore not a-priori predictable if only the network topology is given. The behavior of these strategies will nevertheless be simulated in Section 3.6.

The remainder of this section is divided into as many subsections as the considered topologies.

3.5.1 Random Trees

Consider a random tree topology, *i.e.*, a tree where each node has a random number of sons, possibly zero. The number of nodes forming the network is considered equal to N. The levels in the tree are indicated by ℓ, with $\ell \geq 0$. L denotes the lowest level in the tree. The set of nodes in the ℓ-th level of the tree is denoted as $\mathbb{L}_\ell$, having cardinality $\Lambda(\ell)$, which is a RV. Nevertheless, $\Lambda(0) = 1$ and is not random, since the tree is single rooted. Moreover, the set of nodes, in

the ℓ-th level in the tree, that are not parents to any nodes, is denoted by $\bar{\mathbb{L}}_\ell$ and its cardinality is a RV denoted by $\bar{\Lambda}(\ell)$.

3.5.1.1 TAS algorithm

The TAS algorithm of Section 3.3.4 does not assume any ordering in the network on which it should run. On a random tree, however, it is possible to simplify it making nodes transmit much less frequently than required on an unstructured random topology.

During each transmission phase a single level of the tree is active. Only nodes in this level can transmit. Starting from level L, each node in $\mathbb{L}_L$ has to broadcast its own local data. Then, parent nodes distil and aggregate the received quantities with their own ones and broadcast. The process is repeated until the tree root is reached. The tree is then traveled backwards, from Level 0 to Level L, making the complete sum available to all nodes. This way of operating ensures that an exact retrieval of the entire sum is possible and that no truncation occurs, if the procedure is completed. Nodes participate only in the (at most two) rounds of transmission involving the level they belong to. The number of data that must be transmitted when employing the TAS algorithm is a discrete RV given by

$$N_{\text{TAS}}^{\text{RT}} = \sum_{\ell=0}^{L} \Lambda(\ell) d_{\text{TAS}} + \sum_{\ell=1}^{L-1} \Lambda(\ell) d_{\text{TAS}} - \sum_{\ell=1}^{L-1} \bar{\Lambda}(\ell) d_{\text{TAS}}. \tag{3.38}$$

$N_{\text{TAS}}^{\text{RT}}$ consists of the number of data transmitted when traversing the tree from level L to the root, included, plus the amount of data required by the backwards travel. The last term in (3.38) is related to nodes without sons, which do not transmit anything when the backwards travel is performed.

3.5.1.2 MF algorithm

For the MF, one instead gets,

$$N_{\text{MF}}^{\text{RT}} = \Lambda(L) d_{\text{MF}} + (\Lambda(L) + \Lambda(L-1)) d_{\text{MF}} + \ldots + \sum_{\ell=0}^{L} \Lambda(\ell) d_{\text{MF}}$$

$$
\begin{aligned}
&+ N\left(\Lambda(1) - \bar{\Lambda}(1)\right) d_{\mathrm{MF}} - \sum_{\ell=2}^{L} \Lambda(\ell) d_{\mathrm{MF}} - \left(\Lambda(1) - \bar{\Lambda}(1)\right) d_{\mathrm{MF}} \\
&+ \ldots + N\left(\Lambda(L-1) - \bar{\Lambda}(L-1)\right) d_{\mathrm{MF}} \\
&- \Lambda(L) d_{\mathrm{MF}} - \left(\Lambda(L-1) - \bar{\Lambda}(L-1)\right) d_{\mathrm{MF}} \\
= \Lambda(L) d_{\mathrm{MF}} &+ (\Lambda(L) + \Lambda(L-1)) d_{\mathrm{MF}} + \ldots + \sum_{\ell=0}^{L} \Lambda(\ell) d_{\mathrm{MF}} \\
&+ (N-1)\left(\Lambda(1) - \bar{\Lambda}(1)\right) d_{\mathrm{MF}} - \sum_{\ell=2}^{L} \Lambda(\ell) d_{\mathrm{MF}} \\
&+ \ldots + (N-1)\left(\Lambda(L-1) - \bar{\Lambda}(L-1)\right) d_{\mathrm{MF}} - \Lambda(L) d_{\mathrm{MF}} \\
= \sum_{\ell=0}^{L} \Lambda(\ell) d_{\mathrm{MF}} &+ \Lambda(L) d_{\mathrm{MF}} + N \sum_{\ell=1}^{L-1} \left(\Lambda(\ell) - \bar{\Lambda}(\ell)\right) d_{\mathrm{MF}} \\
&+ \sum_{\ell=1}^{L-1} \bar{\Lambda}(\ell) d_{\mathrm{MF}}, \qquad (3.39)
\end{aligned}
$$

where $\Lambda(L)d_{\mathrm{MF}} + (\Lambda(L) + \Lambda(L-1))d_{\mathrm{MF}} + \ldots + \sum_{\ell=0}^{L} \Lambda(\ell) d_{\mathrm{MF}}$ is the amount of data transmitted in the forward travel, $N\left(\Lambda(1) - \bar{\Lambda}(1)\right) d_{\mathrm{MF}}$ is the amount of data (proportional to N) that nodes, with sons, in level 1 would transmit when the tree is traveled backwards, as if no forward travel was ever performed, and $\sum_{\ell=2}^{L} \Lambda(\ell) d_{\mathrm{MF}} + \left(\Lambda(1) - \bar{\Lambda}(1)\right) d_{\mathrm{MF}}$ is the amount of data that has to be subtracted from the previous one, since these data have already been transmitted in the forward travel. The other terms can be similarly explained.

3.5.1.3 Comparison

TAS is more efficient than MF when $N_{\mathrm{TAS}}^{\mathrm{RT}} < N_{\mathrm{MF}}^{\mathrm{RT}}$, *i.e.*, when

$$
\begin{aligned}
&\sum_{\ell=0}^{L} \Lambda(\ell)(d_{\mathrm{TAS}} - d_{\mathrm{MF}}) - \sum_{\ell=1}^{L-1} \bar{\Lambda}(\ell)(d_{\mathrm{TAS}} + d_{\mathrm{MF}}) - \Lambda(L) d_{\mathrm{MF}} \\
&\qquad - N \sum_{\ell=1}^{L-1} \left(\Lambda(\ell) - \bar{\Lambda}(\ell)\right) d_{\mathrm{MF}} + \sum_{\ell=1}^{L-1} \Lambda(\ell) d_{\mathrm{TAS}} < 0 \\
&N(d_{\mathrm{TAS}} - d_{\mathrm{MF}}) - \Lambda(L) d_{\mathrm{MF}} + (N - \Lambda(L) - 1) d_{\mathrm{TAS}}
\end{aligned}
$$

$$- N(N - \Lambda(L) - 1)d_{\mathrm{MF}} + \sum_{\ell=1}^{L-1} \bar{\Lambda}(\ell)\left((N-1)d_{\mathrm{MF}} - d_{\mathrm{TAS}}\right) < 0, \tag{3.40}$$

that is,

$$\begin{aligned}(d_{\mathrm{MF}} + d_{\mathrm{TAS}} - Nd_{\mathrm{MF}})\,\Lambda(L) > N(d_{\mathrm{TAS}} - d_{\mathrm{MF}}) + (N-1)d_{\mathrm{TAS}} \\ - N(N-1)d_{\mathrm{MF}} + \sum_{\ell=1}^{L-1} \bar{\Lambda}(\ell)\left((N-1)d_{\mathrm{MF}} - d_{\mathrm{TAS}}\right).\end{aligned} \tag{3.41}$$

In case N is finite, (3.41) is satisfied with a probability that is not easily evaluated. Nevertheless, an asymptotic consideration can be done. First, when $N \to \infty$, assume that $L \to \infty$. This is the case when the area on which nodes are deployed is increasing with N, due to coverage extension purposes, or if the communication range d_{comm} is diminishing with N, due to interference mitigation purposes. Moreover, one has $\sum_{\ell=1}^{L-1} \bar{\Lambda}(\ell) < N - \Lambda(L) - L$, since the number of nodes without sons, in all levels except 0 and L, cannot exceed the total number of nodes deprived of the number of nodes at level L and of at least one node for each of the L levels from 0 to $L-1$ (if no nodes are present in a level, there cannot be any further levels). Then, passing to the limit for $N \to +\infty$, if

$$\begin{aligned}\lim_{N\to\infty} \{&N(d_{\mathrm{TAS}} - d_{\mathrm{MF}}) + (N-1)d_{\mathrm{TAS}} - N(N-1)d_{\mathrm{MF}} \\ &+ \left((N-1)d_{\mathrm{MF}} - d_{\mathrm{TAS}}\right)(N - \Lambda(L) - L) \\ &- (d_{\mathrm{MF}} + d_{\mathrm{TAS}} - Nd_{\mathrm{MF}})\,\Lambda(L)\} < 0\end{aligned} \tag{3.42}$$

holds, *i.e.*,

$$\lim_{N\to\infty} (d_{\mathrm{TAS}} - d_{\mathrm{MF}}) + \frac{(L-1)}{N} d_{\mathrm{TAS}} - \frac{L(N-1)}{N} d_{\mathrm{MF}} < 0, \tag{3.43}$$

then also (3.41) holds asymptotically. Since L also goes to $+\infty$, (3.43) is verified and thus (3.41) holds for all values of $\Lambda(L)$, hence with probability 1. Moreover, this is true for all values of the problem dimensions, *i.e.*, n_p and m.

3.5.2 Binary Trees

A deterministic *complete binary* tree topology, that is a tree where each node has exactly two sons apart from nodes in level L that do not have any, is now considered. Assuming that the binary tree consists of L levels entails that $N = 2^{L+1} - 1$.

3.5.2.1 TAS algorithm

For the TAS algorithm, the total number of required data communications is deduced from (3.38)

$$\begin{aligned} N_{\mathrm{TAS}}^{\mathrm{BT}} &= \left(2^L + 2^{L-1} + \ldots + 2^1 + 1 + 2^1 + \ldots + 2^{L-1}\right) d_{\mathrm{TAS}} \\ &= \left(2^{L+1} - 2 + \frac{1}{2}\left(2^{L+1} - 2\right)\right) d_{\mathrm{TAS}} = \left(\frac{3N-3}{2}\right) d_{\mathrm{TAS}}. \end{aligned} \quad (3.44)$$

3.5.2.2 MF algorithm

The total number of data communications required by the MF algorithm is deduced from (3.39)

$$N_{\mathrm{MF}}^{\mathrm{BT}} = \left(2^{L+1} - 1\right) d_{\mathrm{MF}} + 2^L d_{\mathrm{MF}} + \left(2^{L+1} - 1\right) \sum_{i=1}^{L-1} 2^i d_{\mathrm{MF}} = \frac{N^2+1}{2} d_{\mathrm{MF}}. \quad (3.45)$$

3.5.2.3 Comparison

On a binary tree, TAS is more efficient than MF when

$$\frac{3N-3}{2} d_{\mathrm{TAS}} < \frac{N^2+1}{2} d_{\mathrm{MF}}. \quad (3.46)$$

Using (3.22) and (3.18) one obtains the following condition

$$\left(N^2+1\right) k_1 - 3N + 3 > 0, \quad (3.47)$$

where $k_1 = \frac{n_p+1}{\left(n_p + n_p \frac{n_p+1}{2}\right) m}$. For sufficiently large N, (3.47) is always satisfied, disregarding n_p and m. Moreover, and unlike in the random

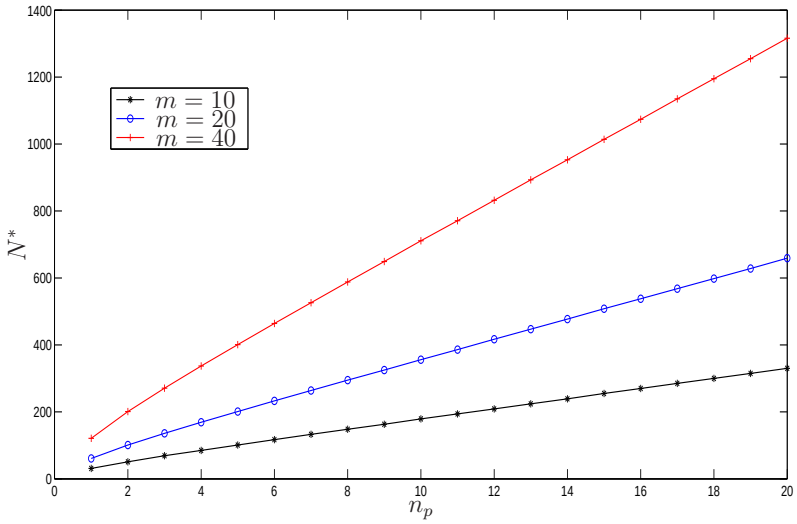

Figure 3.3: Critical value N^*, as a function of n_p, on binary trees, for several values of m.

tree case, given n_p and m, it is possible to derive the value

$$N^* = \frac{3 + \sqrt{9 - 4k_1\,(3 + k_1)}}{2k_1}, \tag{3.48}$$

for which TAS is more efficient than MF. Fig. 3.3 represents N^* as a function of n_p, considering $m = 10, 20, 40$. The behavior is not exactly linear, as it can be easily verified by derivation of (3.48), but rapidly tends to be such: In fact, when n_p grows large, $k_1 \approx \frac{2}{n_p m}$ and

$$N^* \approx \frac{n_p m}{4}\left[3 + \sqrt{9 - \frac{8}{n_p m}\left(3 + \frac{2}{n_p m}\right)}\right] \approx \frac{3}{2} m n_p. \tag{3.49}$$

3.5.3 Clustered Networks

Consider a clustered network, formed by N nodes, structured on a single level of hierarchy (see Fig. 3.4). The network is hence assumed to be divided in n_c clusters. The i-th cluster comprises a random number of nodes N_i^c, including the clusterhead, that is the special node

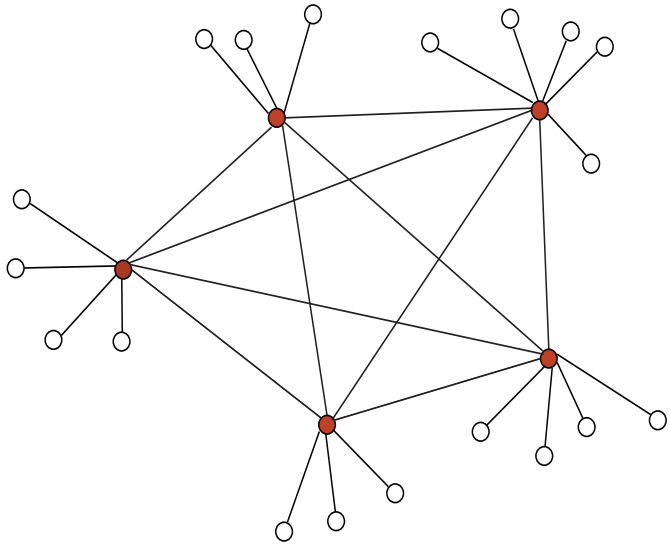

Figure 3.4: A clustered topology. Clusterheads are indicated in red.

responsible for aggregating the local data of its sons. The subnetwork formed by clusterheads is considered to be fully connected: Clusterheads can directly communicate to one another. Moreover, each node in a cluster is assumed to directly communicate with its clusterhead (and vice-versa).

3.5.3.1 TAS algorithm

On this topology, the TAS algorithm transmission phases can be organized as follows. At the beginning, all nodes, with the exception of clusterheads, transmit their local data to the clusterheads. Then each clusterhead aggregates the local data of all nodes in its cluster. Successively, clusterheads transmit to all other clusterheads their aggregated data. Since the network of clusterheads is fully connected, a single broadcast transmission for each of the clusterheads suffices for all clusterheads being capable to construct the completely aggregated data. The amount of scalar data, that has to be transmitted, is thus

$$\begin{aligned} N_{\mathrm{TAS}}^{\mathrm{cc}} &= ((N - n_c) + n_c + n_c)\, d_{\mathrm{TAS}} \\ &= (N + n_c) d_{\mathrm{TAS}}. \end{aligned} \tag{3.50}$$

This accounts for the initial $N - n_c$ transmissions and the subsequent actions of clusterheads, that should broadcast to each other the partially aggregated data and then broadcast, towards nodes forming their cluster, the completely aggregated data.

3.5.3.2 MF algorithm

All nodes in a cluster can overhear broadcast transmissions operated by the corresponding clusterhead. Therefore, the amount of data to be transmitted when employing the MF algorithm is

$$\begin{aligned} N_{\mathrm{MF}}^{\mathrm{cc}} &= \left((N - n_c) + N + (n_c - 1)\, N\right) d_{\mathrm{MF}} \\ &= (N - n_c + n_c N)\, d_{\mathrm{MF}}. \end{aligned} \tag{3.51}$$

This is because all nodes, apart from clusterheads, initially transmit their local information to clusterheads, giving rise to $(N - n_c)d_{\mathrm{MF}}$ transmitted scalar data. Then clusterheads broadcast the received data and their own, this forming a total flow of Nd_{MF} scalar data. At this point, all nodes in each cluster are completely informed about data related to their respective cluster. Finally, there is a backwards transmission during which each clusterhead is transmitting towards its cluster all the Nd_{MF} scalar data except the ones that it previously transmitted, this being equivalent to further $(n_c - 1)\, Nd_{\mathrm{MF}}$ transmitted scalars, composed of n_c clusterheads transmitting not N, but $(N - N_i^{\mathrm{c}})d_{\mathrm{MF}}$ scalar data, *i.e.*, a total of $\sum_{i=1}^{n_c} (N - N_i^{\mathrm{c}})\, d_{\mathrm{MF}} = (n_c N - N)\, d_{\mathrm{MF}}$.

3.5.3.3 Comparison

TAS is better than MF when $N_{\mathrm{TAS}}^{\mathrm{cc}} < N_{\mathrm{MF}}^{\mathrm{cc}}$, *i.e.*, when

$$\begin{aligned} (N - n_c + n_c N)\, d_{\mathrm{MF}} - (N + n_c)\, d_{\mathrm{TAS}} &> 0 \\ \left(1 + \frac{n_c(N-2)}{N + n_c}\right) \frac{d_{\mathrm{MF}}}{d_{\mathrm{TAS}}} &> 1. \end{aligned} \tag{3.52}$$

Here n_c is the degree of freedom, in lieu of L in the tree topologies. Assuming that, due to coverage extension or interference mitigation

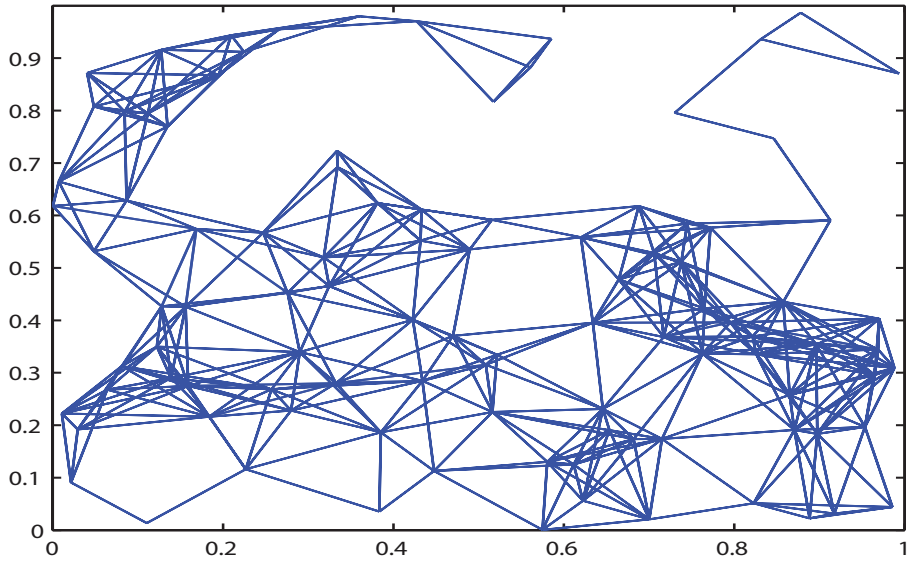

Figure 3.5: Realization of an unstructured random network composed of $N = 100$ nodes.

purposes, n_c grows to ∞ with N going to ∞, one has

$$\lim_{N\to+\infty}\left(1+\frac{n_cN-2n_c}{N+n_c}\right)\frac{d_{\mathrm{MF}}}{d_{\mathrm{TAS}}}=\lim_{N\to+\infty}n_c\frac{d_{\mathrm{MF}}}{d_{\mathrm{TAS}}}=+\infty, \qquad (3.53)$$

independently on n_p and m. Thus, TAS is asymptotically better than MF.

3.6 Numerical Results

In this section, all simulations results have been obtained assuming lossless links while confidence regions are evaluated with the interval analysis techniques described in [77].[5] The Intlab package [87] is employed for intervals computations.

[5]These techniques allow for the numerical computation of tight outer approximations of confidence regions via contraction of the initial search space. The contraction halting criterion may be set such that single box outer approximations are obtained, instead of multiple boxes outer approximations. For the sake of simplicity and with abuse of terminology, in the remainder, 'confidence regions'

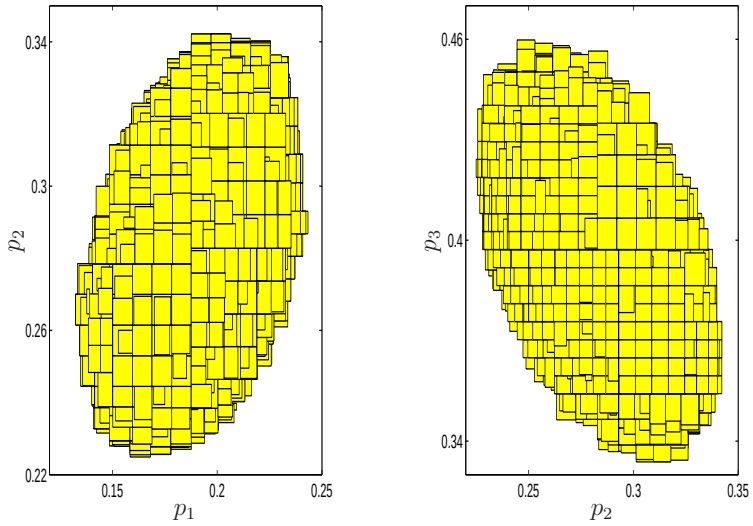

Figure 3.6: Projections of the 90% confidence region computed at node 1 after 4 consensus iterations. A random unstructured network of 100 nodes is considered.

3.6.1 Effect of truncation

Firstly, a numerical investigation of the effect of truncation in information diffusion on the shape of the confidence region is performed. To this purpose, a random unstructured network of $N = 100$ nodes (reported in Fig. 3.5), uniformly distributed over a unit area, is instantiated and a true parameter value $\mathring{\mathbf{p}} = [p_1, p_2, p_3] = [0.2, 0.3, 0.4]$ is considered. The inter-node communication range is set to $d_{\text{comm}} = \sqrt{\frac{\log_2 N}{2N}}$. According to [88], this range guarantees almost sure connectivity of a network of N nodes, deployed on a finite area. White Gaussian measurement noise is considered and the regressors are composed of random equiprobable and independent elements with values in $\{-1, 1\}$. No unit of measurement is specified since it is not necessary to restrict $\mathring{\mathbf{p}}$ to any specific domain. A truncated Metropolis consensus algorithm [31, 81, 82] is run for the distributed computa-

is used in lieu of 'outer approximations of confidence regions', if not otherwise specified.

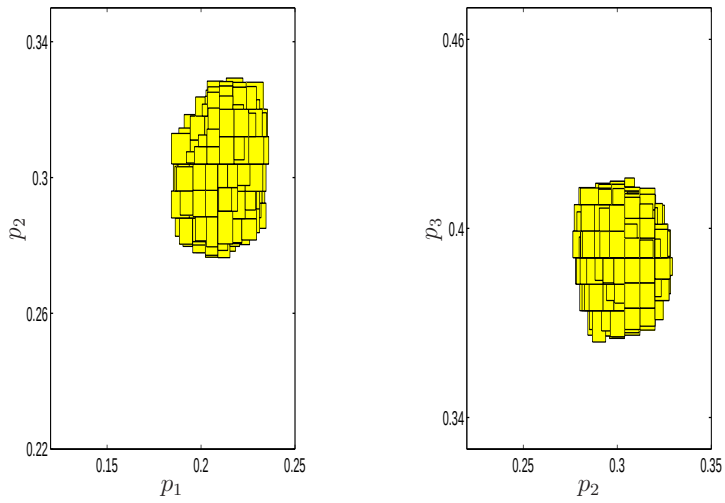

Figure 3.7: Projections of the 90% confidence region computed at node 1 after 30 consensus iterations. A random unstructured network of 100 nodes is considered.

tion of confidence regions. Similar results may be obtained also for the other information diffusion strategies, described in Section 3.3. Figs. 3.6 and 3.7 show the confidence region computed at node 1 after 4 and 30 iterations, respectively, projected over the dimensions of the parameter space. The reduction in terms of volume is quite evident in the second case, while it is to be underlined that the confidence level is the same, independently of when the truncation occurs. In fact, according to the analysis in Section 3.4, any information diffusion truncation does not change the level of confidence but only the shape of the region.

3.6.2 Comparison of information diffusion algorithms

The initial step is to compare state-of-the-art algorithms, hence PF and consensus algorithms, to the mixed approach. Consensus algorithms need all nodes to transmit and update almost synchronously

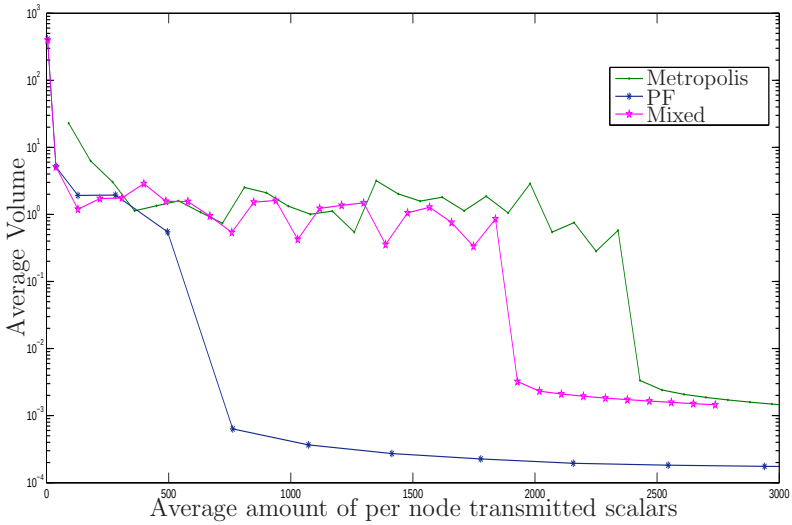

Figure 3.8: Average volume, across nodes, of the 90% confidence region as a function of the average amount of data transmitted by each node. A random unstructured network of $N = 100$ nodes is considered.

their local state and thus are not able to exploit peculiarities of structured networks. This is the reason why the comparison is carried out only on random unstructured topologies. Afterwards, the MF and TAS algorithm will be compared on random trees and clustered networks. Finally their performance will be investigated also on random unstructured networks, where a full comparison with state-of-the-art consensus algorithms will be meaningful.

Figs. 3.8 and 3.9 show the average volume of the 90% confidence regions, obtained running the Metropolis consensus algorithm, the PF, and the mixed approach, as a function of the amount of data transmitted on average by each node. Fig. 3.8 was obtained for the topology shown in Fig. 3.5, while Fig. 3.9 was obtained for a network of 150 nodes, shown in Fig. 3.10. As it can be observed, the confidence region achieved with flooding is smaller than that obtainable with consensus when the same amount of transmitted data is considered, but a gain is possible when applying the mixed approach. In fact, in both cases

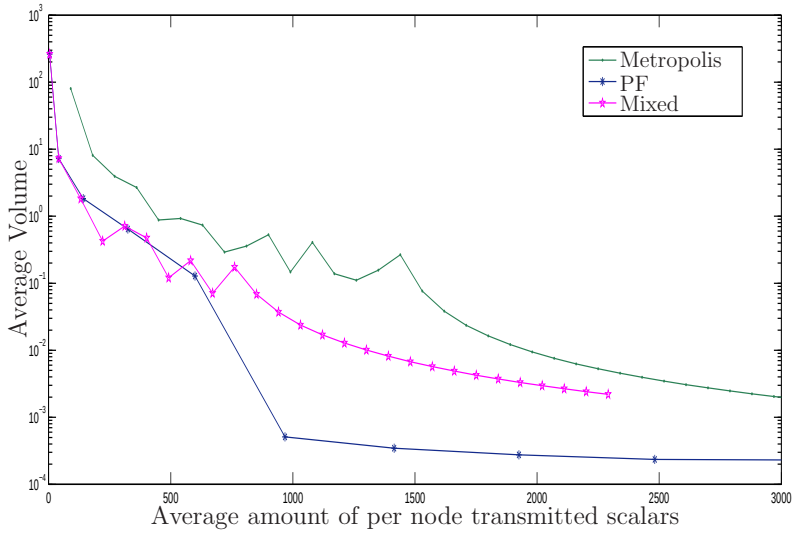

Figure 3.9: Average volume, across nodes, of the 90% confidence region as a function of the average amount of data transmitted by each node. A random unstructured network of $N = 150$ nodes is considered.

there is a range of amount of transmitted data wherein the mixed approach slightly outperforms flooding, which is, however, the best solution when no limitation on data exchanges is present (see the asymptotic behavior in Figs. 3.8 and 3.9). The evidenced behavior of the mixed approach, getting an advantage over the flooding for low data amounts, is always guaranteed to be present, given the definition of the mixed approach itself, for which the initial performance coincides with the one of flooding and, subsequently, improves on it, when that becomes possible. The improvement is related to the number of measurements taken into account: This number is progressively higher and, when applying a mixed approach, this increment comes at a communication cost, in terms of number of data to be transmitted, that is lower than that required with flooding. Anyway, after a certain amount of iterations, depending on network size and connectivity, flooding regains its advantage. This is due to the fact that the coefficients, with which measurements are considered in the perturbed

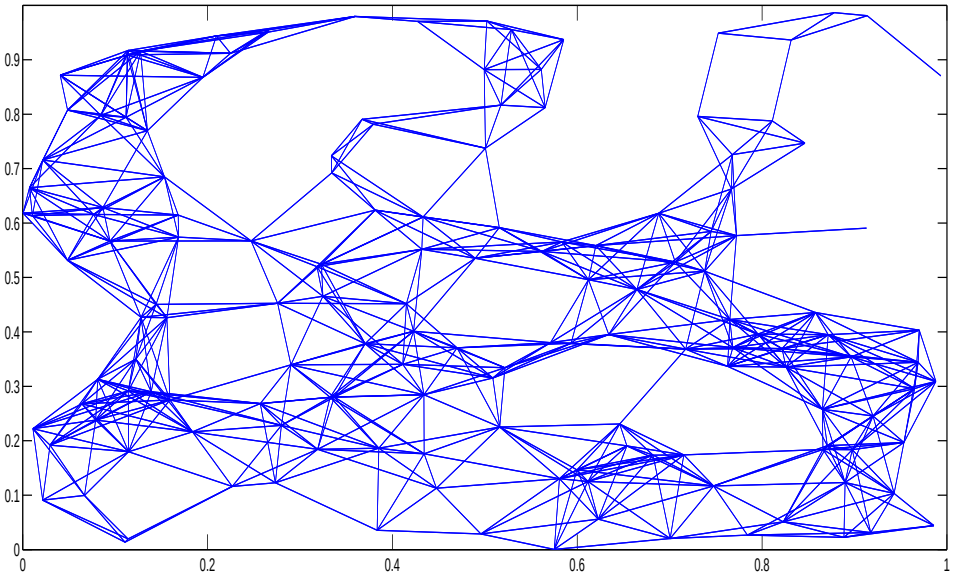

Figure 3.10: Realization of an unstructured random network composed of $N = 150$ nodes.

and unperturbed sums, are equal to 1, while with the mixed approach algorithm they are between 0 and 1. This plays a favorable role in the confidence region volume reduction, as evidenced by the simulation results.

In order to compare the TAS and the MF algorithms, consider random trees, clustered networks, and random unstructured topologies, for the same order of magnitude in terms of number of nodes.

For what concerns the analysis on random trees, consider to build a spanning tree on top of a random unstructured network, setting d_{comm} as earlier done. For each N (see the horizontal axis in Fig. 3.11), 100 connected network realizations are instantiated. TAS and MF are compared in terms of the required number of data to be transmitted on each network realization. The success rate of TAS is the percentage of network realizations that proved favorable to TAS and it is shown in Fig. 3.11 as a function of N, for several values of n_p. As foreseen in the theoretical analysis in Section 3.5, there always exists a threshold value of N, depending on n_p, above which the TAS outperforms the MF algorithm, *i.e.*, the percentage closes to 100%.

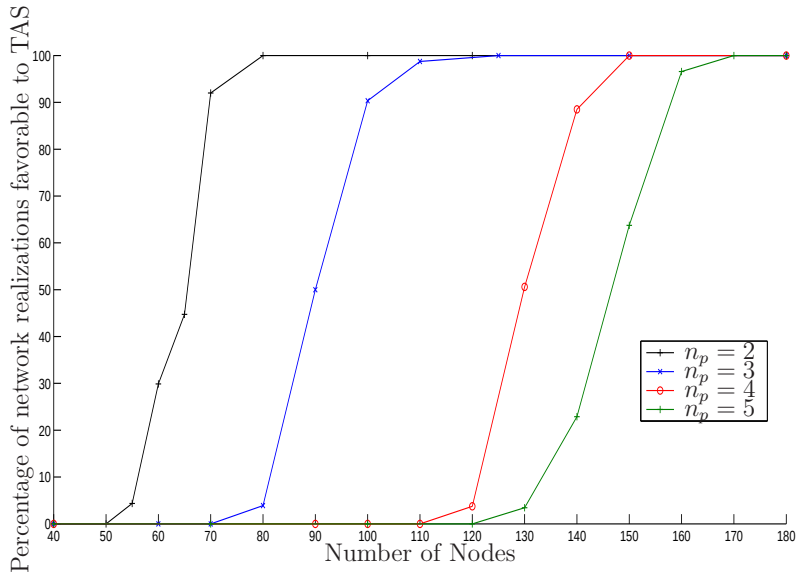

Figure 3.11: Percentage of network realizations favorable to TAS, in terms of required data exchanges, compared to MF, as a function of the number of nodes forming a random tree topology for different values of n_p. 100 random tree realizations are considered for each value of N.

Now the investigation of the trade-off between the confidence region volume and the amount of data transmitted by each node follows. Fig. 3.12 shows the average volume of the 90% confidence region as a function of the average amount of data that needs to be communicated by each node. The volume and data amount are averaged across all nodes and across 100 random tree realizations, while simulation parameters are set to $n_p = 2$, $q = 1$, $N = 100$ and $m = 10$. Fig. 3.12 allows to know which is the amount of data that needs to be transmitted by each node on average to obtain a given confidence region average volume. Each pair of coordinates corresponds to one different transmission round, hence, only partial information could be available at the generic node. However, this affects only the confidence region volume, but does not compromise the level of confidence, which, according to the analysis in Section 3.4, remains the same as if all the information had already been gathered. The TAS algorithm outper-

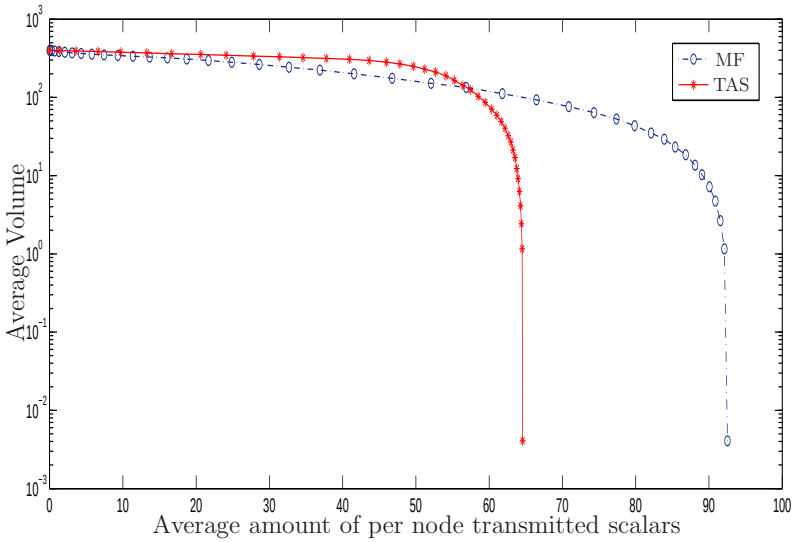

Figure 3.12: Average volume, across nodes and 100 random tree realizations, of the 90% confidence region. Simulation parameters are set to $N = 100, n_p = 2, q = 1$, and $m = 10$.

forms the MF to achieve meaningful small volume values, in terms of the average amount of data transmitted by each node.

Similar results can be obtained on clustered networks. The number of clusters is set to $n_c = 20$ and the average number of per cluster nodes is set to $\mathbb{E}[N_i^c] = 7$ (the parameter dimension is $n_p = 2$, while $q = 1$ and $m = 10$). In particular, Fig. 3.13 shows the average volume of the confidence region, across nodes and 100 clustered network realizations. Here the number of computed pairs volume-amount of data is much lower than that of random trees, due to the fewer transmission rounds. The average amount of data transmitted by each node, needed to obtain meaningful small volumes, is lower when employing the TAS algorithm, as it was on random trees.

Finally, consider a random unstructured network, setting $N = 100$ and $n_p = 3$. As shown in Fig. 3.14, the MF algorithm behaves better than TAS, providing lower volume values for the same amount of data. For comparison, it is also shown how both the MF and the TAS algorithm outperform the state-of-the-art consensus algorithms,

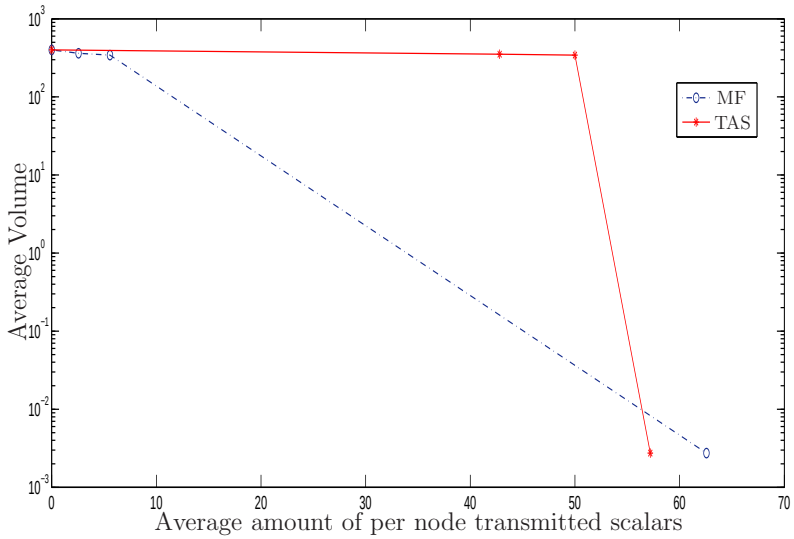

Figure 3.13: Average volume, across nodes and 100 clustered network realizations, of the 90% confidence region. Simulation parameters are set to $n_p = 2, q = 1, n_c = 20$, and $m = 10$.

independently of the considered consensus matrix (Metropolis [81] or Perron [67]).

This section confirms the general behavior that was highlighted in Section 3.5: On structured topologies, such as random trees and clustered networks, there is an advantage in employing the TAS algorithm when the network dimension is sufficiently large, and this independently of n_p. On unstructured networks of comparable size, the MF produces the best results, but, in any case, the absolute amount of data transmitted by each node is much larger than in structured networks. This suggests the adoption of structured networks, together with the TAS algorithm for the distributed computation of confidence regions, when the network traffic load for data diffusion is particularly critical.

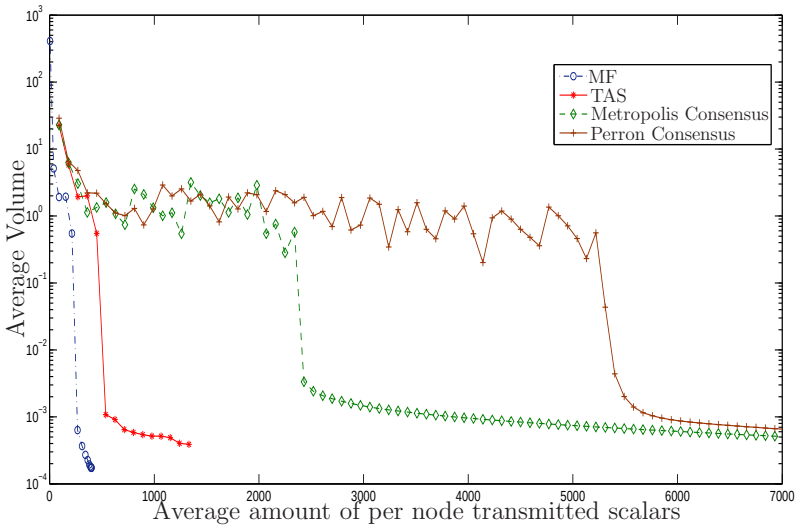

Figure 3.14: Average volume, across nodes, of the 90% confidence region. A random unstructured network of 100 nodes is considered.

3.7 Conclusions

This chapter has investigated the distributed evaluation of non- asymptotic confidence regions at each node in WSNs. Several state-of-the-art information diffusion algorithms are compared to innovative schemes in terms of the required traffic burden. A particular attention is devoted to robustness both guaranteeing that central unit failures are avoided and providing an analysis of the effects of truncation in the information diffusion procedure. The TAS algorithm is presented and its comparison with the other information diffusion algorithms on structured and unstructured topologies is carried out. The TAS algorithm has been designed to efficiently exploit the peculiarities of the distributed evaluation of confidence regions via SPS. This chapter demonstrates that, even in presence of truncated information diffusion, the level of confidence remains the same as in the centralized not truncated case. Simulation results provide a characterization of the trade-off for the achievable average confidence region volume as a function of the required amount of data that each node should trans-

mit on average. The contributions nicely concur at showing that, on structured networks, the proposed TAS algorithm is able to outperform the MF, when the network dimension is sufficiently high, this independently of the specific dimension of the parameter space, as investigated in the theoretical and numerical analyses.

Conclusions

Almost zero-power wireless communications are a key enabler for future energy autonomous systems. The very low power consumption can ensure a much improved device lifetime, paving the way for applications in hostile environments as well as for everyday life IoT applications. The understanding of achievable communication ranges and bit rates is of fundamental importance for the effective deployment of new communication paradigms.

This thesis has proposed an innovative scheme and has theoretically and numerically assessed its performance. The channel characteristics are important features that have to be studied as well: The limitations in peak power availability translate into precise properties of the optimal input distribution. These topics have been investigated in this thesis leading to advances in the knowledge of low power communication systems: Finding a set of sufficient conditions for the capacity-achieving input distribution to be discrete has been one of the major achievements.

The efficiency in communications should pair with the efficiency of the procedures that provide information diffusion in WSNs. The importance of guaranteeing energy savings discloses the meaning of Green ICT: The technology should on one side reduce its intrinsic energy consumption, while on the other one it should contribute to the reduction of the applications consumption. This is the main reason why this thesis has not been limited to low power communications but has also been dedicated to the design and analysis of efficient procedures for the communication of data required by tasks distributed over WSNs. Specifically, the distributed computation of non-asymptotic confidence regions has been considered to avoid the drawbacks of centralized solutions such as strong sensitivity to failures. The problem

has been tackled by investigating the consistency and robustness of the distributed solutions in presence of truncation in data diffusion caused by energy and/or traffic constraints. Contrary to existing studies based on asymptotic bounds, the here obtained results allow the computation of confidence regions even when finite and limited data are available.

For what concerns future advancements, it is worth noting that the experimental assessment of the comparison of information diffusion algorithms is particularly attractive. In fact, the specific medium access control (MAC) issues, like collisions and collisions recovery, that are usually overlooked in numerical verification can be reckoned with. These issues are currently being investigated in an experimental campaign: The results of this campaign are expected to complete the investigation of the topic of information diffusion related to the computation of confidence regions.

More in general, and to further deepen the perspective open by this thesis, it is remarkable that low power communication systems can rely on external carriers. This can ensure deployment of systems on a larger scale, a lower complexity of the employed devices and higher communication security as well. In fact spoofing of these opportunistic communications is much more complicated than detecting ad-hoc generated carriers. Moreover, the low complexity required in the generation of discrete inputs is winking in the direction of establishing low power devices as legacy systems. Finally, the possibility to spare energy by means of clever information diffusion strategies is an important feature that next future systems should be provided with. To say it in a nutshell, this thesis has hopefully paved the way for interesting advancements in communication and estimation theory, having a possibly strong impact on next-future communication systems.

List of Tables

List of Figures

Bibliography

[1] A. Paulraj, D. Gore, R. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, February 2004.

[2] M. Chiani, M. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. on Inf. Theory,*, vol. 49, no. 10, pp. 2363–2371, Oct 2003.

[3] M. Chiani, M. Win, and H. Shin, "MIMO networks: The effects of interference," *IEEE Trans. on Inf. Theory,*, vol. 56, no. 1, pp. 336–349, Jan 2010.

[4] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb 2005.

[5] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, in *Wireless Sensor and Actuator Networks: technologies, analysis and design.* Elsevier Ltd, London, 2008.

[6] J. Matamoros, F. Fabbri, C. Anton-Haro, and D. Dardari, "On the estimation of randomly sampled 2D spatial fields under bandwidth constraints," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4184–4192, Dec 2011.

[7] M. Z. Win, D. Dardari, A. F. Molisch, W. Wiesbeck, and Z. Jinyun, "History and applications of UWB," *Proc. of IEEE, Special Issue on UWB Technology & Emerging Applications*, vol. 97, no. 2, pp. 198–204, Feb 2009.

[8] D. Dardari, R. D'Errico, C. Roblin, A. Sibille, and M. Z. Win, "Ultrawide bandwidth RFID: The next generation?" *Proc. IEEE, Special Issue on RFID - A Unique Radio Innovation for the 21st Century*, vol. 98, no. 9, pp. 1570 –1582, Sep 2010.

[9] D. Jourdan, D. Dardari, and M. Z. Win, "Position error bound for UWB localization in dense cluttered environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 2, pp. 613–628, Apr 2008.

[10] M. Yu, K. Leung, and A. Malvankar, "A dynamic clustering and energy efficient routing technique for sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3069–3079, August 2007.

[11] V. Rajendran, K. Obraczka, and J. Garcia-Luna-Aceves, "Energy-efficient, collision-free medium access control for wireless sensor networks," *ACM Journal on Wireless Networks*, vol. 12, no. 1, pp. 63–78, February 2006.

[12] I. Demirkol, C. Ersoy, and F. Alagoz, "MAC protocols for wireless sensor networks: A survey," *IEEE Communications Magazine*, vol. 44, no. 4, pp. 115–121, April 2006.

[13] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachussets Inst. of Tech. Cambridge Lab for Information and Decision Systems, 1984.

[14] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787 – 2805, 2010.

[15] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, "A survey on facilities for experimental internet of things research," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 58–67, November 2011.

[16] J. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," *IEEE Pervasive Computing*, vol. 4, no. 1, pp. 18–27, Jan 2005.

[17] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.

[18] *Proc. IEEE, Special Issue: Energy Harvesting and Scavenging*, vol. 102, no. 11, November 2014.

[19] L. Berbakov, C. Antón-Haro, and J. Matamoros, "Optimal transmission policy for cooperative transmission with energy harvesting and battery operated sensor nodes," *Signal Processing*, vol. 93, no. 11, pp. 3159 – 3170, 2013.

[20] K. Finkenzeller, in *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification. Second Edition.* Wiley, 2004.

[21] O. Hlinka, O. Sluiak, F. Hlawatsch, P. Djuric, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, Aug. 2012.

[22] F. Cattivelli, C. Lopes, and A. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.

[23] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. on Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan 2009.

[24] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. on Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[25] M. Rabinowitz and J. J. J. Spilker, "A new positioning system using television synchronization signals," *IEEE Trans. Broadcasting*, vol. 51, no. 1, pp. 51–61, 2005.

[26] C. Yang, T. Nguyen, D. Venable, L. M. White, and R. Siegel, "Cooperative position location with signals of opportunity," *Proc. IEEE 2009 National Aerospace & Electronics Conf.*, pp. 18–25, Jul. 2009.

[27] V. Zambianchi, E. Paolini, and D. Dardari, "Information transmission via source of opportunity signals: Piggyback communications," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 5415–5419.

[28] ——, "A piggyback communication scheme for reuse of source of opportunity signals," in *Proceedings of the GTTI annual meeting*, June 2013, **"F. Carassa" best paper award**.

[29] ——, "Peak power limited channels: Analysis of capacity achieving probability measures," in *IEEE International Conference on Communications (ICC)*, June 2014, pp. 1984–1989.

[30] ——, "Capacity achieving peak power limited probability measures: Sufficient conditions for finite discreteness," 2014. [Online]. Available: http://arxiv.org/abs/1411.2364v1

[31] V. Zambianchi, M. Kieffer, F. Bassi, G. Pasolini, and D. Dardari, "Distributed SPS algorithms for non-asymptotic confidence region evaluation," in *Proc. of European Conference on Networking and Communication, EUCNC 2014*, June 2014.

[32] V. Zambianchi, M. Kieffer, G. Pasolini, F. Bassi, and D. Dardari, "Information diffusion for the distributed computation of non-asymptotic confidence regions," 2015, submitted to IEEE International Conference on Communications (ICC).

[33] ——, "Efficient distributed non-asymptotic confidence regions computation over wireless sensor networks," 2014, submitted to IEEE Transactions on Wireless Communications. [Online]. Available: http://arxiv.org/abs/1409.8585

[34] R. C. Hansen, "Relationships between antennas as scatters and as radiators," *Proc. IEEE*, vol. 77, no. 5, pp. 659–662, May 1989.

[35] D. Dardari, F. Guidi, C. Roblin, and A. Sibille, "Ultra-wide bandwidth backscatter modulation: Processing schemes and performance," *EURASIP J. Wireless Commun. and Networking*, vol. 2011, no. 1, July 2011.

[36] N. Decarli, F. Guidi, and D. Dardari, "A novel joint rfid and radar sensor network for passive localization: Design and performance bounds," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 1, pp. 80–95, Feb 2014.

[37] F. Guidi, N. Decarli, S. Bartoletti, A. Conti, and D. Dardari, "Detection of multiple tags based on impulsive backscattered signals," *IEEE Trans. on Communications*, vol. 62, no. 11, pp. 3918–3930, Nov 2014.

[38] G. Marrocco, L. Mattioni, and C. Calabrese, "Multiport sensor RFIDs for wireless passive sensing of objects — basic theory and early results," *IEEE Trans. Antennas and Propagation*, vol. 56, no. 8, pp. 2691–2702, Aug. 2008.

[39] F. Fuschini, C. Piersanti, F. Paolazzi, and G. Falciasecca, "Analytical approach to the backscattering from UHF RFID transponder," *IEEE Antennas and Wireless Propagation Letters*, vol. 7, pp. 33–35, 2008.

[40] D. Dardari, V. Tralli, and A. Vaccari, "A theoretical characterization of nonlinear distortion effects in OFDM systems," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1755–1754, Oct 2000.

[41] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," *Ph.D. dissertation, University of California, Berkeley*, 1969.

[42] ——, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, 1971.

[43] I. Abou-Faycal, M. Trott, and S. Shamai, "The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290 –1301, May 2001.

[44] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-Achieving Probability Measure for Conditionally Gaussian Channels With Bounded Inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.

[45] "Digital Terrestrial Television Broadcasting Planning Handbook," 2005. [Online]. Available: www.acma.gov.au/webwr/aba/tv/licence/digitaltv/planning/documents

[46] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, Jul./Oct. 1948.

[47] S. Shamai and I. Bar-David, "The Capacity of Average and Peak-Power-Limited Quadrature Gaussian Channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.

[48] A. Tchamkerten, "On the Discreteness of Capacity-Achieving Distributions," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2773–2778, Nov. 2004.

[49] A. Feiten and R. Mathar, "Capacity-Achieving Discrete Signaling over Additive Noise Channels," in *IEEE International Conference on Communications (ICC)*, Jun. 2007, pp. 5401–5405.

[50] E. Leitinger, B. C. Geiger, and K. Witrisal, "Capacity and Capacity-Achieving Input Distribution of the Energy Detector," in *IEEE Int. Conf. on Utra-Wideband*, Sep. 2012, pp. 57–61.

[51] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY, 10158: John Wiley & Sons, Inc., 1991.

[52] S. Lang, *Complex Analysis*. New York: Springer-Verlag, 1999.

[53] D. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley & Sons, 1969.

[54] P. Moran, *An Introduction to Probability Theory*. Oxford: Clarendon Press, 1968.

[55] M. Loève, *Probability Theory I*, 4th ed. New York, NY: Springer-Verlag, 1977.

[56] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, March 2002.

[57] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, ser. WSNA '02. New York, NY, USA: ACM, 2002, pp. 88–97.

[58] T. Quek, D. Dardari, and M. Win, "Energy efficiency of dense wireless sensor networks: to cooperate or not to cooperate," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 459–470, February 2007.

[59] G. Mao, B. Fidan, and B. D. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, no. 10, pp. 2529 – 2553, 2007.

[60] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct 2002.

[61] R. Madan and S. Lall, "Distributed algorithms for maximum lifetime routing in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2185–2193, Aug 2006.

[62] Y. Yu, V. Prasanna, and B. Krishnamachari, "Energy minimization for real-time data gathering in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3087–3096, November 2006.

[63] S. Kwon and N. Shroff, "Energy-efficient unified routing algorithm for multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3890–3899, November 2012.

[64] S. M. Kay, *Fundamentals of Statistical Signal Processing-Estimation Theory*. Prentice Hall, 2013.

[65] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. on Signal Process.*, vol. 57, no. 6, pp. 2365–2382, June 2009.

[66] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Trans. on Signal Process.*, vol. 57, no. 11, pp. 4583–4588, November 2009.

[67] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, January 2007.

[68] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *IEEE Conference on Decision and Control*, Dec 2007, pp. 5492–5498.

[69] ——, "Kalman-consensus filter: Optimality, stability, and performance," in *IEEE Conference on Decision and Control*, Shanghai, China, December 2009, pp. 7036–7042.

[70] B. Yang and J. Scheuing, "Cramér-Rao bound and optimum sensor array for source localization from time differences of arrival," in *Proceedings of IEEE ICASSP*, March 2005, pp. 961–964.

[71] X. Sheng and Y.-H. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 44–53, Jan 2005.

[72] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses, and N. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005.

[73] D. Jourdan, D. Dardari, and M. Win, "Position error bound for UWB localization in dense cluttered environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 2, pp. 613–628, April 2008.

[74] M. C. Campi and E. Weyer, "Guaranteed non-asymptotic confidence regions in system identification," *Automatica*, vol. 41, no. 10, pp. 1751–1764, October 2005.

[75] M. C. Campi, S. Ko, and E. Weyer, "Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics," *Automatica*, vol. 45, no. 10, pp. 2175–2186, October 2009.

[76] B. C. Csàji, M. C. Campi, and E. Weyer, "Non-asymptotic confidence regions for the least-squares estimate," in *Proc. IFAC SYSID*, Brussels, Belgium, 2012, pp. 227–232.

[77] M. Kieffer and E. Walter, "Guaranteed characterization of exact non-asymptotic confidence regions as defined by LSCR and SPS," *Automatica*, vol. 50, no. 2, pp. 507–512, February 2014.

[78] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ser. MobiCom '99. New York, NY, USA: ACM, 1999, pp. 174–185.

[79] A. Nordio, C. Chiasserini, and E. Viterbo, "Performance of linear field reconstruction techniques with noise and uncertain sensor locations," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3535–3547, Aug 2008.

[80] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. of Information Processing in Sensor Networks, IPSN 2005. Fourth International Symposium on*, April 2005, pp. 63–70.

[81] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, September 2004.

[82] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, 2006.

[83] L. Xiao, S. Boyd, and S. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributing Computing*, vol. 67, no. 1, pp. 34–46, 2007.

[84] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb 2003.

[85] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, Oct 2003.

[86] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct 2006.

[87] S. Rump, "INTLAB - INTerval LABoratory," in *Developments in Reliable Computing*, T. Csendes, Ed. Dordrecht: Kluwer Academic Publishers, 1999, pp. 77–104. [Online]. Available: http://www.ti3.tu-harburg.de/rump/

[88] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar 2000.