

Article

Surveying Multidisciplinary Aspects in Real-Time Distributed Coding for Wireless Sensor Networks

Carlo Braccini ^{1,2}, Franco Davoli ^{1,2,*}, Mario Marchese ^{1,2} and Maurizio Mongelli ³

¹ Department of Electrical, Electronic and Telecommunications Engineering, and Naval Architecture (DITEN), University of Genoa, Via Opera Pia 13, 16145 Genoa, Italy;

E-Mails: carlo.braccini@unige.it (C.B.); mario.marchese@unige.it (M.M.)

² National Inter-University Consortium for Telecommunications (CNIT)–University of Genoa Research Unit, Via Opera Pia 13, 16145 Genoa, Italy

³ Institute of Electronics, Computer and Telecommunication Engineering, National Research Council of Italy (IEIIT-CNR), Genoa Site, Area della Ricerca, Via De Marini, 6-16149 Genoa, Italy;

E-Mail: maurizio.mongelli@ieiit.cnr.it

* Author to whom correspondence should be addressed; E-Mail: franco.davoli@unige.it; Tel.: +39-10-353-2732.

Academic Editor: Stefano Mariani

Received: 28 October 2014 / Accepted: 16 January 2015 / Published: 27 January 2015

Abstract: Wireless Sensor Networks (WSNs), where a multiplicity of sensors observe a physical phenomenon and transmit their measurements to one or more sinks, pertain to the class of multi-terminal source and channel coding problems of Information Theory. In this category, “real-time” coding is often encountered for WSNs, referring to the problem of finding the minimum distortion (according to a given measure), under transmission power constraints, attainable by encoding and decoding functions, with stringent limits on delay and complexity. On the other hand, the Decision Theory approach seeks to determine the optimal coding/decoding strategies or some of their structural properties. Since encoder(s) and decoder(s) possess different information, though sharing a common goal, the setting here is that of Team Decision Theory. A more pragmatic vision rooted in Signal Processing consists of fixing the form of the coding strategies (e.g., to linear functions) and, consequently, finding the corresponding optimal decoding strategies and the achievable distortion, generally by applying parametric optimization techniques. All approaches have a long history of past investigations and recent results. The goal of the present paper is to provide

the taxonomy of the various formulations, a survey of the vast related literature, examples from the authors' own research, and some highlights on the inter-play of the different theories.

Keywords: single-hop wireless sensor networks; distributed coding; decision theory; signal processing; optimization

1. Introduction

Multi-terminal source-channel coding in Wireless Sensor Networks (WSNs) arises whenever a number of sensors observe a physical phenomenon represented by one or more random variables, and transmit their measurements over noisy channels to one (or more) sink node(s). The goal of encoding and decoding functions that transform the information emitted by the sources into symbols suitable for transmission over the channels and reconstruct it at the receiving end is twofold: (i) to reduce the possible redundancy intrinsic in the original variables (*source coding*); (ii) to adapt the variables to the channel conditions (*channel coding*). Such operations should be performed in order to minimize the distortion (according to some given measure) between the original and the reconstructed information, under a constraint on the power available for transmission. Information theory aims at finding the fundamental limits attainable by this process, disregarding the possible delay introduced by the encoding-decoding operations and their complexity. Such limits can be attained—at least in the single source-channel case—under Shannon's Separation Theorem [1], by separately performing source and channel coding on the digital representation of the sources.

Still in information theoretic terms, zero-delay (also referred to as “real-time”, “single-letter”, or “instantaneous”) coding is the problem of finding the minimum distortion (according to a given measure), subject to a power constraint, attainable by encoding and decoding functions, with precise limits on delay and complexity. In other words, “the encoder maps every source output symbol separately onto a channel input symbol, and the decoder maps every channel output symbol separately onto a source reconstruction symbol” [2]. In some cases (more specifically, under the conditions stated in [2]), zero-delay coding can indeed be optimal in an unconstrained sense, *i.e.*, even in the class of functions that allow infinite delay and complexity.

The information theoretic approach, though not aiming at finding the optimal coding/decoding strategies, but rather the optimum attainable performance values (minimum average distortion achievable under a given power constraint, or minimum average power to achieve a given distortion), sometimes surprisingly yields an answer to the existence of globally optimal *linear* solutions (*i.e.*, where encoders and decoders are constituted by linear transformations of the observed information). This has been long known for the scalar case of a single Gaussian channel and a single Gaussian source, where the optimum encoder-decoder pair is instantaneous and linear [3]; in [4], Wyner provides a beautiful outline of the reason why this turns out to be so, by equating the rate distortion function $R_{eq}(\beta)$ for a certain distortion β to the channel capacity $C_{eq}(\alpha)$ for a certain average power α . However, once entered the realm of multi-terminal and multi-channel information theory (referred to as *network* information theory), this simple linear (and analog) *joint* source-channel coding (also termed “uncoded” solution, or “Amplify and Forward”—AF), as opposed to the asymptotically

optimal source-channel separation in digital communications of [1], occurs only in some special situations. This aspect was thoroughly investigated by Gastpar, Gastpar and Vetterli, *et al.* [5–10], among others.

The recent widespread diffusion of sensor networks and the evolution toward the Internet of Things (IoT) has given new momentum to the investigation of zero-delay distributed source-channel coding (owing to the limited processing and power capabilities of the sensors), and motivates the present paper. In particular, we are interested here in the Physical Layer of WSNs, disregarding other aspects as routing, considerations of node proximity, system organization, latencies and possible losses introduced by the traversal of queues in relay nodes. To keep the discussion more focused we also limit our consideration to the so-called single-hop cases, where direct communication between sensor nodes and a sink is attained, without the presence of relay nodes that characterizes multi-hop WSNs (though similar considerations can be extended to the presence of relay nodes, as well; see, e.g., [11–13]). The term “network”, therefore, is related here to the presence of multiple distributed terminals (the sensors), communicating with a sink over multiple channels (as opposed to the single source-channel case), where the transmitted pieces of information can combine in different fashions, owing to possible interference.

The approach of information theory tries to determine the achievable optimum, without necessarily looking for the coding/decoding strategies yielding it. From another perspective, however, the ultimate goal of the category of problems considered above would be to find the optimal encoder(s)-decoder(s) pairs that minimize the given distortion function under a certain power constraint. The decision theory approach seeks to determine such coding/decoding strategies. Since the decisional agents (encoder(s) and decoder(s)) or “Decision Makers” (DMs) possess different information, though sharing a common goal, the most natural framework here is the functional optimization of team theory [14–16]. Even though the original team problem is *dynamic* [15], in the sense that the encoders’ decisions influence the information of the other DMs (the decoders), if the distortion function is quadratic the team can be reduced to a *static* one (*i.e.*, where the decision strategies to be derived by each DM do not influence the information of the others), by keeping into account that the decoder will always compute a conditional mean and by expressing the latter as a functional of the encoders’ decision strategies (as will be briefly sketched in Section 5). Nonetheless, the ensuing functional optimization problem still remains formidable. Some insights can be gained by transforming it into a parametric optimization, by means of nonlinear approximating functions (e.g., neural networks [17]).

Between these two visions, a more pragmatic approach consists of fixing the form of the coding strategies and, consequently, finding the decoding strategies and the achievable distortion. In this respect, interesting recent work concerning the application of instantaneous nonlinear mappings at the encoders (not necessarily stemming from a functional optimization problem) regards Shannon-Kotel’nikov (SK) mappings [18]. In the Gaussian Sensor Network (GSN) case, where all random variables (source symbols, measurement and channel noises) have Gaussian distributions, this SK joint analog source-channel coding has been shown to perform better than the linear “uncoded” solution in some cases [19]. Linear solutions optimized in their parameters have been extensively investigated, especially in the Signal Processing literature [20–27], under both GSN and non-Gaussian hypotheses. It is worth noting that, once the coding strategies have been fixed to a linear form, giving rise to a linear conditional mean at the decoder(s), finding the encoders’ coefficients that minimize a

quadratic distortion function under a power constraint is not a trivial problem. In fact, owing to the presence of such coefficients inside the gain matrices of the decoder(s), the optimization problem turns out to be non-quadratic and, in general, even non-convex.

Summing up, WSNs in which analog source symbols (stemming from measurements of a certain physical phenomenon) need to be collected and transmitted to remote sink stations, are a significant example of systems where network information theory, team decision theory, and distributed estimation can be applied to study different aspects of a multi-faceted problem. All approaches have a long history of past investigations and recent results, the problem has undergone a huge number of formulations and possible variants, all of a certain relevance, and sometimes even slight variations can make the difference between finding a feasible optimal solution or encountering formidable difficulties. The goal of the paper is not to introduce new results, but rather to: (i) provide the taxonomy of the various formulations; (ii) highlight the relevance of analog joint source-channel coding to the field of WSNs; (iii) conduct a survey of the vast related literature; and (iv) show the different points of view introduced by the information theoretic, decision/control theoretic and signal processing approaches. Many survey papers can be found on WSNs in general. Some address computational intelligence [28], data collection [29], and data aggregation [30], which all have some points in common with the environment considered here. However, to the best of our knowledge and with the exception of [7], none treats the zero-delay joint analog coding-decoding problem under multiple points of view.

In the next Section we provide a discussion on the relevance of such type of problems in the WSN field. Section 3 contains an introduction to the multi-terminal source-channel coding problem in WSNs, where multiple spatially separated sensors collect noisy measurements of a physical phenomenon and send them in single-hop fashion to a common sink node over noisy communication channels. Though multiple sinks may be present, the essence of the problem is well reflected in the multi-sensor-single-sink case (multiple encoders and a single common decoder), and we limit our consideration to it. We define the taxonomy of the different problem variants, and we highlight: (i) the measurement process; (ii) the encoding functions; (iii) the channel models; (iv) the distortion function; (v) the decoder structure. In Section 4 we examine the information theoretic approaches to the problem, and survey some of the relevant results. Section 5 deals with the much less investigated team decision theory approaches. We briefly outline the team decision problem in this context, point out where the main difficulties arise in the functional optimization, and how they might be circumvented by suboptimal strategies. Finally, in Section 6, we recall the optimal (parametric) solution in the case where the form of the encoding strategies is fixed. In particular, we focus on linear functions (the “uncoded” case of information theory) under quadratic distortion (LQ). The signal processing literature contains many examples of this particular situation, both in the presence of Gaussian (LQG) and non-Gaussian random variables. We provide a tutorial survey of the problem formulations and of the parametric optimization solutions. Section 7 briefly describes an example, based on the authors’ own work, about the non-linearity of the coding/decoding strategies, which bridges, to some extent, the team theoretical and the signal processing aspects. Section 8 contains the conclusions and a classification of the literature surveyed in the different fields.

2. Relevance to WSNs

Though the majority of WSNs adopt digital transmission [31] (commonly used standards are IEEE 802.15.4/ZigBee [32], IEEE 802.15.1/Bluetooth [33], ISA100/WirelessHart [34]), a number of solutions based on analog modulation are emerging ([35–41], among others). Although the communication problem presented here is formulated for analog transmission, the impact on WSN applications may be relevant, because fully digital and fully analog architectures may be intrinsically inefficient for a WSN [42]. In multi-hop WSNs with relaying, analog and digital solutions have been compared in [43]. In general, besides the optimality and scalability properties that it exhibits in some cases, analog zero-delay coding (and, in some cases, processing) appears to be convenient where very low power consumption and computational complexity are required.

We summarize here some applications that may be put in relation with the problem addressed by the paper. A set of applications belongs to the family of analog transmission, in particular, when considering hybrid analog-digital architectures. Another one deals with more complex operations performed by the sink, for example, involving a classification task (e.g., target tracking in video surveillance). For the latter case, the following works are relevant. Reference [44] exploits analog joint source-channel coding to drive power allocation while addressing a classification problem at the sink. It highlights how the joint problem of communication and classification needs more sophisticated analytical and numerical tools, as similarly outlined in this paper. Reference [45] addresses the same objective (classification at the sink), by deriving an optimal trade-off between classification accuracy and energy preservation. Again in the detection field, the authors of [43] compare the efficacy of digital vs. analog relaying in a sensor network and show that the superiority of digital relaying actually depends on the signal to noise ratio. In a similar hybrid digital–analog (HDA) context, the acoustic sensor network of [35] shows how HDA systems may supersede purely digital transmission in dependence of the radio channel quality. Another analog case is that of [41], in which an AF strategy is applied, together with cooperative coding for the sake of interference mitigation. An analog scatter-radio WSN is presented in [37–39] for environmental monitoring purposes. Extremely low power consumptions over transmission ranges of tens of meters have been achieved.

Analog processing is also adopted for computation. The analog computation for data fusion in [46,47] has also similarities with the framework proposed here, in particular with respect to approaching the problem through functional optimization techniques. In [48], analog signal processing is integrated in a sensor node to simplify the digital computation tasks, thus increasing energy-efficiency; the considered application is vehicle classification. Reference [49] adopts analog computation of Fourier transform coefficients for lower power consumption.

3. Taxonomy of WSN Zero-Delay Coding/Decoding Problems

We are interested in zero-delay coding and decoding functions that minimize a given distortion functional (usually a quadratic one) under given constraints on transmission power (or minimize power under a given distortion constraint). As coding and decoding strategies operate on a single realization of the source random variables, our problem has no dynamics over time. To fix ideas, we introduce the problem in the basic Gaussian case. Whenever we deal with non-Gaussian random variables we will

state it explicitly. The basic setting we consider comprises a number of sensors that observe a physical phenomenon, whose output can be represented by Gaussian random variables (r.v.'s). The observations are to be transmitted to a sink over one or more noisy channels with some power constraint, and the task of the sink is to provide an estimation of the original variables under a quadratic distortion criterion.

Let $\mathbf{x} \in R^n$, $\mathbf{x} \sim N(0, \Sigma_x)$, be the original unknown vector, which we suppose 0-mean and with covariance matrix Σ_x , and let $\mathbf{y} \in R^m$ be the vector of variables observed at the sink (whatever the observation channel, the action performed by the transmitters, and the transmission channel). Let:

$$\hat{\mathbf{x}} = \gamma_2(\mathbf{y}) \tag{1}$$

be the estimation performed at the sink. Then, since (letting $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$):

$$\begin{aligned} \gamma_2^*(\cdot) &= \underset{\gamma_2(\cdot)}{\operatorname{argmin}} E \left\{ \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \right\} = \underset{\gamma_2(\cdot)}{\operatorname{argmin}} E \left\{ E \left\{ \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \mid \mathbf{y} \right\} \right\} = \\ &= \underset{\hat{\mathbf{x}}}{\operatorname{argmin}} E \left\{ \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \mid \mathbf{y} \right\} = E \left\{ \mathbf{x} \mid \mathbf{y} \right\} \end{aligned} \tag{2}$$

the estimation is always given by the conditional mean (Actually, by Sherman's Theorem [50], the cost function that yields the conditional mean as optimal estimator can be more general. Sherman's Theorem is as follows: Let \mathbf{x} be a random vector with mean $\boldsymbol{\mu}$ and density $f_x(\cdot)$. Let $L(e)$, $e = \mathbf{x} - \hat{\mathbf{x}}$, be a loss function such that $L(\mathbf{0}) = 0$ and $\rho(x_1) \geq \rho(x_2) \geq 0 \Rightarrow L(x_1) \geq L(x_2) \geq 0$, where $\rho(\cdot)$ is a nonnegative and convex distance function (e.g., the Euclidean norm). If $f_x(\cdot)$ is symmetric about $\boldsymbol{\mu}$ and unimodal (i.e., it has only one peak), then $\hat{\mathbf{x}} = \boldsymbol{\mu}$ minimizes $E[L(e)]$. Applied to our conditional mean problem, Sherman's Theorem is: Let $f_{x|y}(\cdot)$ be symmetric about its mean $E\{\mathbf{x}|\mathbf{y}\}$ and unimodal, and let $L[e(\mathbf{y})]$ be a cost function defined as above; then, $\hat{\mathbf{x}} = E\{\mathbf{x}|\mathbf{y}\}$ minimizes $E\{L[e(\mathbf{y})]\}$).

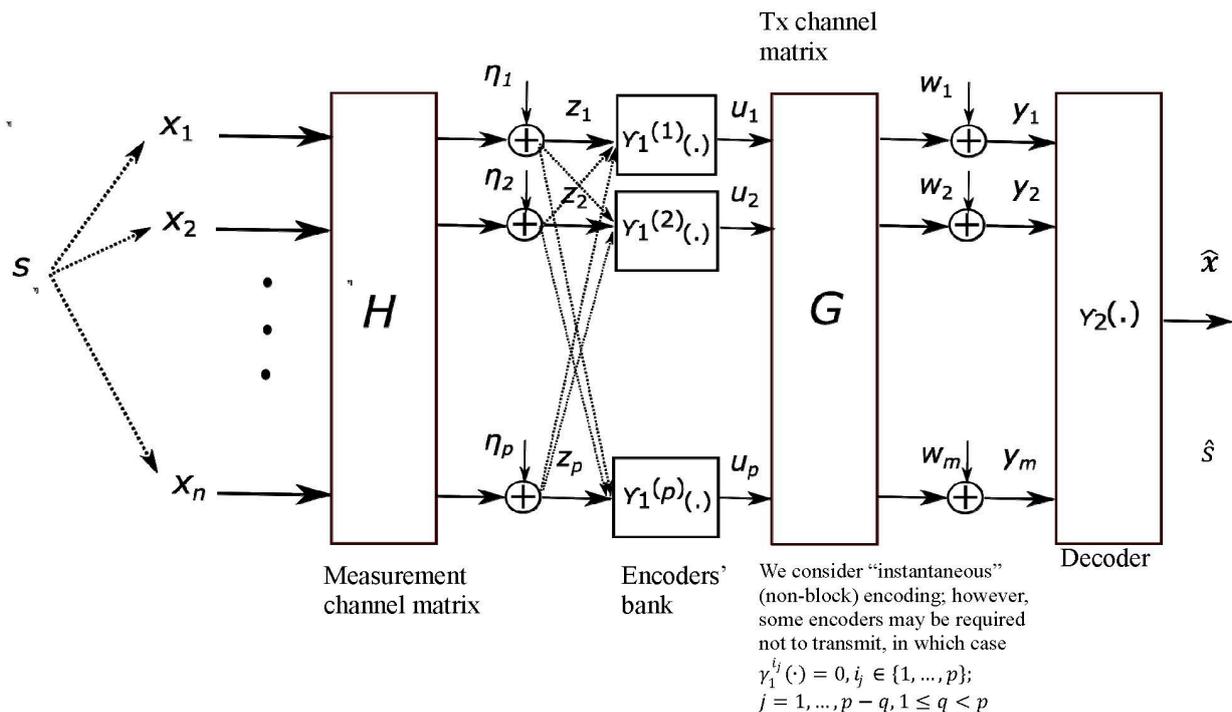


Figure 1. General structure of the WSN zero-delay coding/decoding problem.

In the following, we will introduce the taxonomy of a number of variations of this problem. However, whatever the structure of the problem, if the random vectors \mathbf{x} and \mathbf{y} are Gaussian and the relation between them is linear, then the conditional mean in Equation (2) will be a linear function of \mathbf{y} (or, more generally, if we consider \mathbf{x} to be non-zero-mean, an affine function). We will return on this point in Section 6. We can depict the general configuration of our WSN problem as in Figure 1, along with the basic elements that will be discussed in this classification. With reference to Figure 1, we note the following settings and possible variations of the problem, according to different aspects being considered.

3.1. Observation and Measurement Noises

All random vectors \mathbf{x} , $\boldsymbol{\eta}$ and \mathbf{w} are considered mutually independent.

3.2. Original Random Variables and Distortion Functions

There may be an additional “original” variable s representing the physical phenomenon, which vector \mathbf{x} may be related to (by mutual correlation, or more generally via a joint probability distribution function; the relation is indicated by dashed lines between s and the components of \mathbf{x} in Figure 1). Then, a reconstruction of s directly, rather than of the components of \mathbf{x} , may be required. It must be noted that in this case the quadratic cost function is the expectation of the square error between two scalar r.v.’s, rather than some quadratic norm of the error between two random vectors. In these cases, the mutual influence between the “source” s and the related vector’s components that are observed is usually specified in terms of their joint distribution (*i.e.*, in the case of zero-mean jointly Gaussian r.v.’s, by their mutual correlations). There may be, for instance, some distance-dependent correlation function, which characterizes the mutual correlation between the source and each component of \mathbf{x} and between two different components of \mathbf{x} (which may correspond to measurement points spread around the physical phenomenon of interest described by the source s). This is the situation considered, among others, in [51–56]. The distortion measures corresponding to the two cases in which the variable of interest is either \mathbf{x} or s are, respectively:

$$J_x = E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\} \quad (3)$$

$$J_s = E\{(s - \hat{s})^2\} \quad (4)$$

where $\|\cdot\|$ is a suitable norm (e.g., Euclidean).

3.3. “Expansion” and “Refinement” in Estimation

Two opposite settings are highlighted in [7] with respect to the case of the estimation of \mathbf{x} . The first one, called “Expanding Sensor Network” corresponds to the case where the observations (the components of \mathbf{z} , or of \mathbf{x} if no observation noise is present) are relatively independent, \mathbf{H} is the identity matrix \mathbf{I} (which also implies $p = n$), and matrix \mathbf{G} has a rank of the order of n . The term “expanding” derives from the fact that if new sensors were to be added, each new sensor practically brings in new data of interest. In the second situation, called “Refining Sensor Network”, the matrix \mathbf{H}

is such that $p > n$ (in other words, each sensor measures a noisy combination of a relatively small number of variables of interest), so that each sensor adds something to the knowledge of the same group of variables. A case of interest here is that of a relatively “poor” communication infrastructure, with a Multi-Input Multi-Output (MIMO) channel represented by a matrix \mathbf{G} with low rank (with respect to q). We will return on these situations in the discussion further on.

3.4. Noisy Observations/Multiple Hops

The noisy observation channel may be present or not, mainly to account for measurement uncertainty or “garbled” measurements. It is worth noting that, whereas in the centralized coding case it was shown in [57] (generalizing the earlier result of [58]) that computing the conditional mean of the variables of interest (conditioned to the observation) and using it as the argument of the coding function is optimal, this needs no longer be true in the informationally decentralized coding situation, as noted in [8]. Moreover, in practical applications, the sink might not be the final destination of the information, but rather an intermediate point forwarding the measurements to a processing center. Consider, for example, the case where the sink is a cluster-head collecting measurements from a number of sensors, which should be forwarded to a distant processing center via a satellite link [59]. In these scenarios, involving tandem links, it would be very useful to adopt the definition of link Mean Square Error (MSE) introduced in [60] (as the MSE between the conditional mean estimators of the original signal at the input and output of that link), which allows summing the MSEs of the individual links to obtain the overall MSE.

3.5. Power Constraints

The transmission of the encoded variables u_i , $i = 1, \dots, p$, is subject to a power constraint. There are two possibilities.

- An overall power constraint:

$$E\left\{\sum_{i=1}^p u_i^2\right\} \leq P \quad (5)$$

- Individual power constraints on each transmitted variable:

$$E\{u_i^2\} \leq P_i, \quad i = 1, \dots, p \quad (6)$$

3.6. Zero-Delay Coding

We consider only “instantaneous” (“single-letter”) coding, whereby the coding functions—generally nonlinear—at the sensors are applied to a single measurement individually for a single channel use, rather than to a block of measurements collected over time. The reason for this is essentially to limit the complexity of the code and the ensuing processing burden at the sensors. However, the dashed lines toward the encoders in Figure 1 indicate possible communication among sensors to exchange their measurements, with the arguments of encoding functions $\gamma_1^{(i)}(\cdot)$, $i = 1, \dots, p$, changing accordingly. This yields the possibility to consider different encoding strategies, from completely decentralized

(no measurement exchange) to fully centralized ones. The latter case corresponds to the centralized encoding of a Gaussian vector source over a Gaussian vector channel, a problem solved long ago in the linear case (*i.e.*, when the encoder is constrained to be linear) in [61,62]. It is worth noting that the linearly constrained solutions, in the presence of an overall power constraint as in Equation (5), imply that some variables might not be transmitted (from the application of Karush-Kuhn-Tucker optimality conditions to the choice of the optimal coding matrix \mathbf{A}), and hence give rise to $q \leq p$ encoders.

3.7. Transmission Channels' Structure

As already implied by one of the previous points, the structure of the matrix \mathbf{G} (and of the noise covariance matrix Σ_w) can model very different types of MIMO channels. $\mathbf{G} = \mathbf{I}$ and diagonal Σ_w represent parallel independent channels without interference, also called Orthogonal Multiple Access Channel (MAC) [63]; \mathbf{G} equal to a row vector of all 1's corresponds to the classical MAC (the receiver sees the sum of all channel inputs).

3.8. Distortion/Power Minimization

Though still without entering into details, we have outlined so far the situation in which the encoders want to minimize the average distortion under a power constraint. Obviously, the reverse situation is also meaningful to consider, namely, the minimization of average power by using the distortion as a constraint (see, e.g., [24]).

4. Information Theory Approaches

As we already noted in the Introduction, Information Theory, though not aiming at finding the optimal strategies, but rather the optimum attainable performance values, sometimes surprisingly yields an answer to the existence of globally optimal linear solutions. In particular, it was shown in [9] that uncoded transmission is strictly optimal in the following case (with reference to Figure 1): a single source, observed n -fold ($x_i = s, \forall i$), $\mathbf{H} = \mathbf{I}$ ($p = n$), $\Sigma_\eta = \sigma_\eta^2 \mathbf{I}$, $\mathbf{G} = [1, 1, \dots, 1]$ ($m=1$, *i.e.*, a Multiple Access Channel), distortion as in Equation (4), and constraints as in Equation (6), with $P_i = P, \forall i$. The instantaneous linear solution here is not only optimal among single-letter codes, but globally optimal among all arbitrarily long block codes. In this and other situations, uncoded transmission has been proven to scale exponentially better asymptotically with the number of sensors, with respect to digital communication with separate source and channel coding. The result of [9] has been generalized to the asymmetric case of different power constraints and noise, and by considering also the sum power constraint, in [64–66]. Inhomogeneous measurement and transmission channels are considered in [67]. Liu and Ulukus [68,69] determined bounds and an asymptotic scaling law for dense sensor networks transmitting over a cooperative MAC with noise feedback, where, contrarily to the previous cases, separation is order optimal when the underlying random process satisfies some general conditions. In [70], the optimality of uncoded transmission schemes is investigated for two correlated random sources over the Gaussian MAC, whereas [71] characterizes the distortion pairs that are simultaneously achievable on the two source components of a bivariate Gaussian source transmitted to a common receiver by two separate transmitters over an average power-constrained Gaussian MAC,

and proves the optimality of uncoded transmission for low signal-to-noise ratio (SNR); the same problem in the presence of perfect causal feedback from the receiver to each transmitter is analyzed in [72]. Rajesh and Sharma consider discrete alphabet sources [73] and the presence of side information over the Gaussian MAC [74]; they also compare three different joint source-channel coding schemes. The same authors study the Orthogonal MAC in [75,76], and the fading Gaussian MAC in [77]. Several types of multiuser channels, not necessarily Gaussian, with correlated source side information are studied in [78], and optimality of separation is proved in those cases for the joint source-channel code, though not yielding the same codes of the source and channel coding problems considered separately.

Recently, Jain *et al.* [79] studied the minimum achievable transmission energy under a distortion constraint. For two correlated Gaussian sources communicated over a Gaussian multiple-access channel, they obtain inner and outer bounds on the energy-distortion region, also showing that uncoded transmission outperforms separation-based schemes in many different conditions. Still in the context of Information Theory, a situation of interest is that of pure source coding, disregarding transmission and channel noise. In a distributed setting like the one we have introduced, this is the framework of the Chief Executive Officer (CEO) problem [80–84].

5. The WSN as a Team Decision Problem

The decision theory approach looks for the determination of the strategies, and, since the decisional agents (encoders and decoder) or “Decision Makers” (DMs) possess different information, though sharing a common goal, it must be necessarily in the framework of team theory [14–16,85]. If, to fix ideas, we focus on the minimization of distortion under a power constraint, and we suppose to work with the objective function Equation (3) and constraint Equation (5), the decision problem is:

$$\min_{\gamma_1^{(1)}(\cdot), \gamma_1^{(2)}(\cdot), \dots, \gamma_1^{(p)}(\cdot); \gamma_2(\cdot)} E \left\{ \left\| \mathbf{x} - \gamma_2(\mathbf{y}) \right\|^2 \right\} \quad (7)$$

subject to:

$$E \left\{ \sum_{i=1}^p \left[\gamma_1^{(i)}(z_i) \right]^2 \right\} \leq P \quad (8)$$

Though this problem falls in the category of dynamic teams [86] (the decisions of the encoders influence the information available to the decoder), it can actually be reduced to a *static* one, by remembering that, indeed, the optimal strategy of the decoder is to compute the conditional mean:

$$\begin{aligned} \hat{\mathbf{x}} &= \gamma_2(\mathbf{y}) = E \{ \mathbf{x} | \mathbf{y} \} = \int_{-\infty}^{+\infty} \boldsymbol{\xi} f_{\mathbf{x}|\mathbf{y}}(\boldsymbol{\xi} | \mathbf{y}) d\boldsymbol{\xi} = \\ &= \int_{-\infty}^{+\infty} \boldsymbol{\xi} \frac{f_{\mathbf{y}|\mathbf{x}}(\mathbf{y} | \boldsymbol{\xi}) f_{\mathbf{x}}(\boldsymbol{\xi})}{f_{\mathbf{y}}(\mathbf{y})} d\boldsymbol{\xi} = \int_{-\infty}^{+\infty} \boldsymbol{\xi} \frac{f_{\mathbf{x}}(\boldsymbol{\xi}) \int_{-\infty}^{+\infty} f_w[\mathbf{y} - \mathbf{G}\boldsymbol{\gamma}_1(\mathbf{H}\boldsymbol{\xi} + \boldsymbol{\zeta})] f_{\eta}(\boldsymbol{\zeta}) d\boldsymbol{\zeta}}{f_{\mathbf{y}}(\mathbf{y})} d\boldsymbol{\xi} \end{aligned} \quad (9)$$

where we have defined $\boldsymbol{\gamma}_1(\cdot) = \text{col}[\gamma_1^{(1)}(\cdot), \gamma_1^{(2)}(\cdot), \dots, \gamma_1^{(p)}(\cdot)]$.

Substitution of Equation (9) into the part to be minimized in Equation (7) yields an expression that is a functional of $\gamma_1(\cdot)$ only. The reduction to a static team is actually always possible, as was noted long ago by Witsenhausen [87]. Though this fact is sometimes useful and has actually been exploited to specify properties of the solution [88], the still extremely complex form of the expression obtained in our WSN case renders the functional optimization problem formidable, unless some restrictions are imposed on the structure of the encoding functions.

As regards the presence of the stochastic constraint Equation (8), it is worth noting that it can be handled by using Lagrangian duality, in a similar fashion as done in [89] in a different setting. Indeed, one can consider the minimization:

$$\min_{\gamma_1^{(1)}(\cdot), \gamma_1^{(2)}(\cdot), \dots, \gamma_1^{(p)}(\cdot); \gamma_2(\cdot)} \left[E \left\{ \|\mathbf{x} - \gamma_2(\mathbf{y})\|^2 + \lambda \sum_{i=1}^p [\gamma_1^{(i)}(z_i)]^2 \right\} \right] \quad (10)$$

and then determine the multiplier λ as:

$$\max_{\lambda \geq 0} \left[E \left\{ \|\mathbf{x} - \gamma_2^*(\mathbf{y})\|^2 + \lambda \sum_{i=1}^p [\gamma_1^{(i)*}(z_i)]^2 \right\} - \lambda P \right] \quad (11)$$

Though posing the problem in a team theory setting is so hard, some simplifications are possible by restricting the form of the strategies. As we have already seen in Section 1, restricting the coding strategies to be linear immediately yields a linear form of the conditional mean at the decoder. As an alternative, it would be interesting to investigate the effect of going the other way round, *i.e.*, forcing the conditional mean to be linear, and trying to find the coding functions that minimize the distortion under the given power constraint in this case. The ensuing static team problem is one with linear information structure, (seemingly) quadratic cost, and Gaussian primitive random variables (LQG). LQG static teams are known to have a globally optimal linear solution [14,85,86]. This can be found by writing the so-called “person-by-person satisfactoriness” (p.b.p.s.) conditions, *i.e.*, the necessary conditions for optimality of the problem in strategic form, and then conditioning expectations to the available information for each agent. For example, in the case of two decisional agents, with common cost to be minimized $J(\gamma_1, \gamma_2)$, and admissible strategy sets Γ_1 and Γ_2 , one would write the conditions:

$$\begin{aligned} E\{J(\gamma_1^*, \gamma_2^*)\} &\leq E\{J(\gamma_1, \gamma_2^*)\}, \quad \forall \gamma_1 \in \Gamma_1 \\ E\{J(\gamma_1^*, \gamma_2^*)\} &\leq E\{J(\gamma_1^*, \gamma_2)\}, \quad \forall \gamma_2 \in \Gamma_2 \end{aligned} \quad (12)$$

which can be transformed into ordinary minimizations by conditioning expectations:

$$\begin{aligned} \min_{u_1} E\{J(u_1, \gamma_2^*)|z_1\} \\ \min_{u_2} E\{J(\gamma_1^*, u_2)|z_1\} \end{aligned} \quad (13)$$

In the LQG case (where the observations z_1 and z_2 depend linearly on the primitive r.v.’s), by guessing linear strategies, substituting them in the minimization problems and computing expectations leads to the solution of a linear system (in the unknown parameters that constitute the matrices of the linear strategies); the solution turns out to be unique and hence, given the convexity of the cost and the symmetry of the probability distributions, it is also the globally optimal one.

Now, going back to our case, though it is true that all the assumptions are there, the dependence of the cost on the parameters of the matrices representing the encoders' strategies would be non-quadratic, since these parameters would appear within the expression of the gain matrix at the decoder (matrix \mathbf{B} in Equation (17) of Section 6 below), which is required to compute the linear Minimum Mean Square Error (MMSE) estimator. This (besides the presence of the power constraint) gives rise to a non-classical structure of the static team optimization problem, which would not necessarily imply the existence of globally optimal linear solutions. Forcing also the encoding strategies to be linear (which implies that matrix \mathbf{B} assumes the form of Equation (22) in Section 6), would lead to the same non-convex optimization problem considered and solved in [23].

6. Signal Processing and Parametric Optimization

Whatever the structure of the problem, if the random vectors \mathbf{x} and \mathbf{y} are Gaussian and the relation between them is linear, then the conditional mean in Equation (2) will be a linear function of \mathbf{y} (or, more generally, if we consider \mathbf{x} to be non-zero-mean, an affine function):

$$\hat{\mathbf{x}} = E\{\mathbf{x}|\mathbf{y}\} = \gamma_2^*(\mathbf{y}) = \mathbf{B}\mathbf{y} + \mathbf{b} \quad (14)$$

In this case, the constant vector \mathbf{b} and matrix \mathbf{B} are easily determined by the condition of having an unbiased estimate and by the orthogonality principle [50], respectively:

$$\begin{aligned} E\{\hat{\mathbf{x}}\} = E\{\mathbf{x}\} = \bar{\mathbf{x}} &\Rightarrow \mathbf{B}\bar{\mathbf{y}} + \mathbf{b} = \bar{\mathbf{x}} \Rightarrow \mathbf{b} = \bar{\mathbf{x}} - \mathbf{B}\bar{\mathbf{y}} \\ \hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{B}(\mathbf{y} - \bar{\mathbf{y}}) \end{aligned} \quad (15)$$

Since Equation (15) shows that we can always work with zero-mean vectors if we consider the shifted variables $\mathbf{x} - \bar{\mathbf{x}}$, $\hat{\mathbf{x}} - \bar{\mathbf{x}}$, $\mathbf{y} - \bar{\mathbf{y}}$, we will consider zero-mean r.v.'s for the sake of simplicity. The orthogonality principle states that the estimation error is orthogonal to the data:

$$\begin{aligned} E\{(\mathbf{x} - \hat{\mathbf{x}})\mathbf{y}^T\} &= 0 \\ E\{(\mathbf{x} - \mathbf{B}\mathbf{y})\mathbf{y}^T\} &= 0; \quad \mathbf{B}E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{x}\mathbf{y}^T\} \end{aligned} \quad (16)$$

so that, if the covariance matrix of \mathbf{y} is positive definite:

$$\mathbf{B} = E\{\mathbf{x}\mathbf{y}^T\}E\{\mathbf{y}\mathbf{y}^T\}^{-1} \quad (17)$$

The MMSE linear estimator defined by Equation (14) (where we now consider $\mathbf{b} = 0$) and Equation (17) is a discrete Wiener filter.

The calculation of the covariance matrices depends on the (linear) structure of the observation (measurement by the sensors) and transmission (from the sensors to the sink) channels.

Let:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta} \quad (18)$$

represent the measurement process, with $\mathbf{z} \in R^p$, $\boldsymbol{\eta} \in R^p$, $\boldsymbol{\eta} \sim N(0, \boldsymbol{\Sigma}_\eta)$. In general, we can suppose each sensor to observe a noisy linear combination of the variable(s) of interest. Moreover, let:

$$\mathbf{y} = \mathbf{G}\mathbf{A}\mathbf{z} + \mathbf{w} \quad (19)$$

where $\mathbf{w} \in R^m$, $\mathbf{w} \sim N(0, \Sigma_w)$ and the matrices $\mathbf{A} \in R^{qp}$ and $\mathbf{G} \in R^{mq}$ represent the linear coding and the effect of the transmission channels, respectively (in general, a linear combination of transmitted variables represents interference). The matrix \mathbf{A} would be diagonal (or block-diagonal, if the individual sensors' observations are allowed to be vectors) if no sensor cooperation (by exchanging measurements) is allowed. Given such structure, and supposing all noise vectors to be mutually independent and independent of \mathbf{x} , then:

$$\begin{aligned} E\{\mathbf{y}\mathbf{y}^T\} &= E\left\{(\mathbf{GAH}\mathbf{x} + \mathbf{GA}\boldsymbol{\eta} + \mathbf{w})\left(\mathbf{x}^T \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T + \boldsymbol{\eta}^T \mathbf{A}^T \mathbf{G}^T + \mathbf{w}^T\right)\right\} = \\ &= \mathbf{GAH}\Sigma_x \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T + \mathbf{GA}\Sigma_\eta \mathbf{A}^T \mathbf{G}^T + \Sigma_w \end{aligned} \quad (20)$$

and:

$$E\{\mathbf{x}\mathbf{y}^T\} = E\left\{\mathbf{x}\left(\mathbf{x}^T \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T + \boldsymbol{\eta}^T \mathbf{A}^T \mathbf{G}^T + \mathbf{w}^T\right)\right\} = \Sigma_x \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T \quad (21)$$

so that:

$$\mathbf{B} = \Sigma_x \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T \left(\mathbf{GAH}\Sigma_x \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T + \mathbf{GA}\Sigma_\eta \mathbf{A}^T \mathbf{G}^T + \Sigma_w\right)^{-1} \quad (22)$$

Sometimes, Equation (22) is written in a different form, which is derived by using the Matrix Inversion Lemma ($\mathbf{W}\mathbf{Y}^T(\mathbf{X} + \mathbf{Y}\mathbf{W}\mathbf{Y}^T)^{-1} = (\mathbf{W}^{-1} + \mathbf{Y}^T \mathbf{X}^{-1} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X}^{-1}$):

$$\mathbf{B} = \left[\Sigma_x^{-1} + \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T (\mathbf{GA}\Sigma_\eta \mathbf{A}^T \mathbf{G}^T + \Sigma_w)^{-1} \mathbf{GAH}\right]^{-1} \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T (\mathbf{GA}\Sigma_\eta \mathbf{A}^T \mathbf{G}^T + \Sigma_w)^{-1} \quad (23)$$

If the source statistics is unknown, the Best Linear Unbiased Estimator (BLUE) can be used instead of the MMSE estimator. In this case, the second expression is readily useful as:

$$\mathbf{B}_{BLUE} = \left[\mathbf{H}^T \mathbf{A}^T \mathbf{G}^T (\mathbf{GA}\Sigma_\eta \mathbf{A}^T \mathbf{G}^T + \Sigma_w)^{-1} \mathbf{GAH}\right]^{-1} \mathbf{H}^T \mathbf{A}^T \mathbf{G}^T (\mathbf{GA}\Sigma_\eta \mathbf{A}^T \mathbf{G}^T + \Sigma_w)^{-1} \quad (24)$$

We further note that, in the linear cases considered in this section, and still assuming 0-mean variables, the estimation error would be given by:

$$\begin{aligned} E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\} &= E\{\|\mathbf{x} - \mathbf{B}\mathbf{y}\|^2\} = E\{(\mathbf{x} - \mathbf{B}\mathbf{y})^T (\mathbf{x} - \mathbf{B}\mathbf{y})\} = \text{tr}E\{(\mathbf{x} - \mathbf{B}\mathbf{y})(\mathbf{x} - \mathbf{B}\mathbf{y})^T\} = \\ &= \text{tr}\left[E\{(\mathbf{x} - \mathbf{B}\mathbf{y})\mathbf{x}^T\} - E\{(\mathbf{x} - \mathbf{B}\mathbf{y})\mathbf{y}^T\}\mathbf{B}^T\right] = E\{\mathbf{x}^T (\mathbf{x} - \mathbf{B}\mathbf{y})\} = \\ &= \text{tr}\Sigma_x - \text{tr}E\{\mathbf{B}\mathbf{y}\mathbf{x}^T\} = \text{tr}\Sigma_x - \text{tr}E\{\mathbf{B}(\mathbf{GAH}\mathbf{x} + \mathbf{GA}\boldsymbol{\eta} + \mathbf{w})\mathbf{x}^T\} = \\ &= \text{tr}\Sigma_x - \text{tr}\mathbf{B}\mathbf{GAH}\Sigma_x \end{aligned} \quad (25)$$

The above relations, which are classical ones in linear estimation theory, can be easily modified to the case where the original phenomenon of interest is represented by a single source observed by multiple sensors, which we also considered in Section 3, as depicted in Figure 1.

As a final general note, we recall that the orthogonality principle is related to the quadratic distortion measure, irrespectively of Gaussianity in the underlying r.v.'s; therefore, all relations we have summarized remain valid with respect to *linear* estimation, *i.e.*, when the optimal encoders and decoder are restricted to be linear functions of their observations. In the Gaussian case, the linear

optimal decoder coincides with the conditional mean, under linear (AF) encoding functions. However, as we have seen in Section 4, even under Gaussian hypotheses, linear encoders and decoders turn out to be globally optimal only in some special cases. Very recently, necessary and sufficient conditions have been derived for linearity of (centralized) optimal mappings, given a general noise (or source) distribution, and a specified power constraint [90].

In the general setting that we have described so far, much work has been done in the case of linear (AF) coding and linear decentralized estimation (combined optimization of decoder and coders' matrices), in both Gaussian and non-Gaussian cases. Energy optimized AF is considered in [20,21] with the BLUE estimator for a scalar source with K -fold observations transmitted over the Orthogonal MAC; [21] also derives the optimal power allocation with the sum power constraint and the minimum power under zero-outage estimation distortion constraint. Here too, as in the centralized cases of [61] and [62], the application of Karush-Kuhn-Tucker conditions in the optimal power allocation with the sum power constraint can lead to completely turn off "bad" sensors (with poor channels and observation quality). Owing to the adoption of the BLUE estimator, the source probability distribution can be unspecified. In a different approach, reference [22] supposes sensors' observations to be quantized, and finds the optimum quantization levels and transmit power levels under an MSE constraint. Reference [23] studies the optimum linear decentralized estimation problem under coherent (sum) and orthogonal MAC, in the cases of scalar and vector observations, under general noise. In the scalar case, they derive the optimum power allocation for the coherent MAC and also compare the result to the orthogonal MAC (interestingly, in the coherent MAC case the optimum solution does not turn off any sensors). In the vector case, the linear optimization problem for the orthogonal MAC was shown to be NP-hard in [25] (we recall that this is due to the non-convexity of the optimization problem derived by substituting the conditional mean—see Equation (9)—in the quadratic error function, even in the linear case). The optimal solution in the absence of channel noise is derived analytically (in closed form) in [23] for the coherent MAC case; under noise, the problem of finding the optimal modulators' matrices is formulated as a Semi-Definite Programming (SDP) one, and the effect of power and bandwidth constraints is investigated. In [26], by adopting similar channel models with power and bandwidth constraints as in [23], the optimum linear modulator matrices are found that minimize the MMSE gap between the system over a noisy channel and the one over a noiseless channel, giving rise to a water-filling solution. Always in the general case, optimum linear estimators are derived in [27]. Ribeiro and Giannakis treat both the Gaussian [91] and the general case of unknown probability densities [92]. A complete network problem, considering different protocol layers, is treated in [93].

Back to the GSN case, [94,95] consider distributed analog linear coding of multiple correlated multivariate Gaussian sources over the coherent MAC. Chaudhary and Vandendorpe [96] address the optimization of AF gain coefficients (*i.e.*, the power allocation) under two different settings: (i) exact Channel State Information (CSI), where the fading attenuation coefficients of the transmission channels are known at the encoders and decoder; (ii) imperfect CSI, where the fading coefficients are estimated. They derive an original algorithm based on the successive approximation of the linear MMSE distortion, which is computationally efficient and exhibits very good convergence properties. In [97], the same authors consider a similar problem, but under quantization of the transmitted variables, rather than analog transmission. Their goal is the design of joint quantization and power

allocation to minimize MSE distortion for a given total network power consumption. The 1-bit quantization for decentralized estimation of an unknown parameter in the presence of an orthogonal MAC is treated in [98] for both Gaussian and Cauchy channel noise.

The lifetime maximization is studied in [99] in TDMA and interference limited non-orthogonal multiple access (NOMA) channels as a joint power, rate and time slot (for TDMA) allocation problem under various constraints.

Very interesting recent work concerning the application of instantaneous nonlinear mappings at the encoders (not necessarily stemming from an optimization problem) regards Shannon-Kotel'nikov mappings [18,19,100–103]. The GSN case with joint analog source-channel coding is considered in [19], and it is shown to perform better than the uncoded solution.

7. An Example of Non-Linear Coding/Decoding Strategies

An example of non-linear coding/decoding strategies is outlined now, in order to emphasize the complexity of the problem even under a small number of sensors, and to highlight the contribution from the fields of team decision theory and signal processing. The 1:2 bandwidth expanding system of [18] is used to address cost Equation (4) with respect to the source estimation with $p=2$. Each source measurement z is mapped by two sensors on the double Archimedes' spirals $\mathbf{u} \in R^2$ as follows:

$$\mathbf{u}(z) = \pm \frac{\Delta}{\pi} \varphi(\alpha z) [\cos(\varphi(\alpha z))\mathbf{i} + \sin(\varphi(\alpha z))\mathbf{j}] \quad (26)$$

where α is a gain factor, $\varphi(\cdot)$ is the inverse curve length approximation $\varphi(\xi) = \pm \sqrt{\frac{\xi}{0.16\Delta}}$ (subsection III.B of [18]) and Δ is the (radial) distance between the two spiral arms. The outputs of the sensors are the components of vector \mathbf{u} , subject to constraint Equation (6). The coding operation consists of a bandwidth expansion, because the source $s \in R$ is mapped into $\mathbf{u} \in R^2$. More specifically, s defines the angle of the spiral through the application of $\varphi(\cdot)$ and the Cartesian coordinates of the spiral \mathbf{u} are sent over the channels. The decoding operation at the sink consists of finding (in polar coordinates) the angle corresponding to the point on the spiral closest to the received signal \mathbf{y} . The parameters α and Δ are free variables defining the shape of the spirals; they are set to minimize distortion, while satisfying the power constraint. Under Gaussian hypotheses, the setting may be driven by closed-form expressions. In more general environments, numerical approximation should be provided. The technique belongs to the already mentioned category of SK mappings.

In order to allow more general nonlinear coding-decoding strategies and compare them with the results of SK, we consider here the approach of [56], which we briefly summarize. Coding and decoding strategies for the estimation of s are of the form:

$$u_i = \hat{f}_i(x_i, \boldsymbol{\omega}_i), \quad i = 1, \dots, p; \quad \hat{s} = \hat{g}(\mathbf{y}, \boldsymbol{\omega}_g) \quad (27)$$

where we choose $\hat{f}_i(\cdot)$ and $\hat{g}(\cdot)$ to be neural networks (NNs) depending on the choice of the basis functions (e.g., sigmoid) of each layer; $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_g$ are vectors of parameters activating the basis functions [104]. Among the various possible fixed-structure for coding-decoding functions, NNs have been chosen for their powerful approximation capabilities. Vectors $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_g$ should be numerically

optimized in a joint process in virtue of the team structure of the problem [105]. The technique is identified by the acronym NN in the following.

Some additional remarks may be useful to clarify the basic differences between the two nonlinear strategies. The SK coding is centralized, since the projection operation onto the spirals needs the joint knowledge of the two sensors' inputs. The NN coding process may work in two ways. In the centralized approach, each sensor i knows the input vector \mathbf{x} of all the sensors. In the decentralized one, it knows only its own input x_i as evidenced in Equation (27). SK defines coding functions in polar coordinates, while NN defines coding functions in Cartesian coordinates. SK has limitations in the number of coding components, *i.e.*, $p = 2, 3$; an insightful discussion on how a further generalization may be tricky is presented in section V of [106]. The NN may be scalable to any p ; the most severe limitations to scalability reside in the complexity of the minimization process in terms of (offline) computational effort, in the choice of an appropriate setting of the several parameters affecting the numerical process, and in the presence of local minima.

All the available strategies (linear, SK, centralized and decentralized NN) have been tested in [105] with respect to a bimodal distribution in s . The distribution consists of the superposition (with equal probabilities) of two uniform distributions in $[-4.5, -3.5]$ and $[3.5, 4.5]$. No measurement noise is considered, channel noises are independent normal distributions with unit variance, channel gains are set to 1. The total available power is 11. In SK, $\alpha^* = 3.31$, $\Delta^* = 1.32$, which have been found numerically. To help SK in the decoding phase, the angle interval (in polar coordinates) over which we search for the spiral point closest to the received point \mathbf{y} has been restricted to the one generated just before the addition of the channel noise. Otherwise, \mathbf{y} is projected back to the wrong spiral arm and the effect on distortion is dramatic. This corresponds to a-priori deleting the *threshold* distortion component of SK (subsection II.C of [18]). As far as the NN is considered, the hyperbolic tangent is used as activation function with 30 hidden neurons in the sensors and 20 hidden neurons in the sink (a single coder with 30 hidden neurons is used in the centralized version). In contrast to our previous tests in reference [105], here we have chosen to also introduce constraints of type Equation (6) in the NN (with the total power divided equally between the two encoders), in order to force the strategy to conform to the bandwidth expansion in all cases. In fact, when this is not done, both NN approaches tend to turn off one sensor and apply a constant amplification factor to the input signal on the other sensor (this is quite evident under the centralized NN in [105]), whereas a larger coding range is generated by SK. Figures 2 and 3 show the coders and decoder, respectively, under the different strategies. The distortion is 1.33 under the linear approach (in [105] we reported a value of 1.247, because we used a slightly higher power allowance in the linear case, to be fair with respect to the NN approaches, where the global power constraint is accounted for by a penalty function that gives rise to a looser enforcement; this is not needed here, as the penalty functions on individual power constraints in the NNs appear to be respected more sharply). The values corresponding to the other strategies are: 0.117 with SK, 0.866 with centralized NN and 0.140 with decentralized NN (the surprisingly higher value for the centralized case is probably due to the enforcement of the additional individual constraints that further reduce the degrees of freedom; the corresponding values obtained in [105] without the additional constraints were 0.133 and 0.155, respectively, but tended to deviate significantly from the bandwidth expansion). The nonlinear curve trend of the SK and NN decoders is quite evident. Despite their different coding behaviors, the nonlinear decoding strategies are quite

similar. The further generalization of the results to the presence of measurement noise η shows a higher robustness of the NN with respect to SK [105].

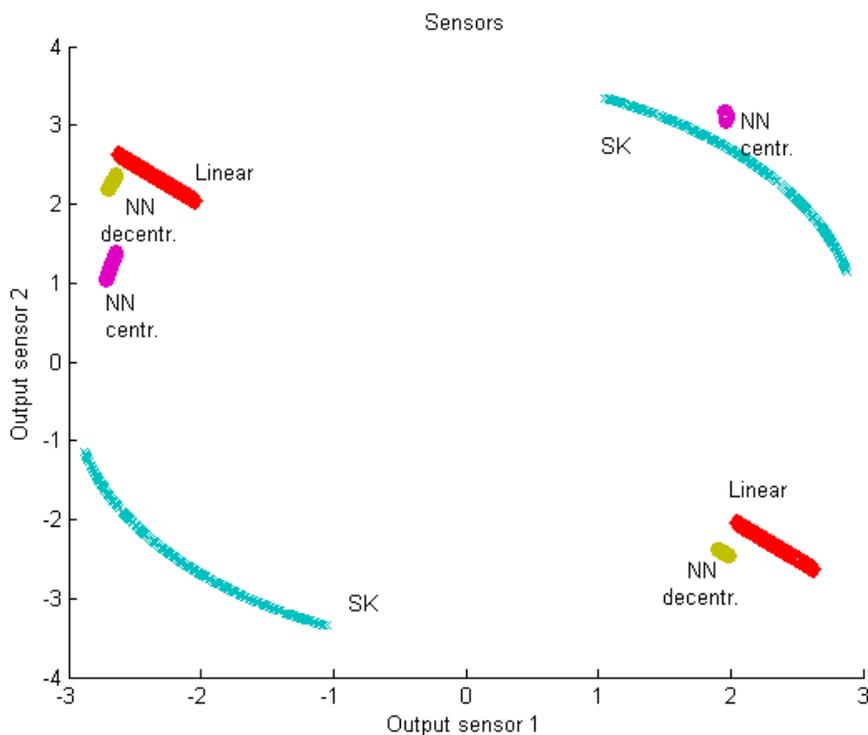


Figure 2. Single source with bimodal distribution: coders.

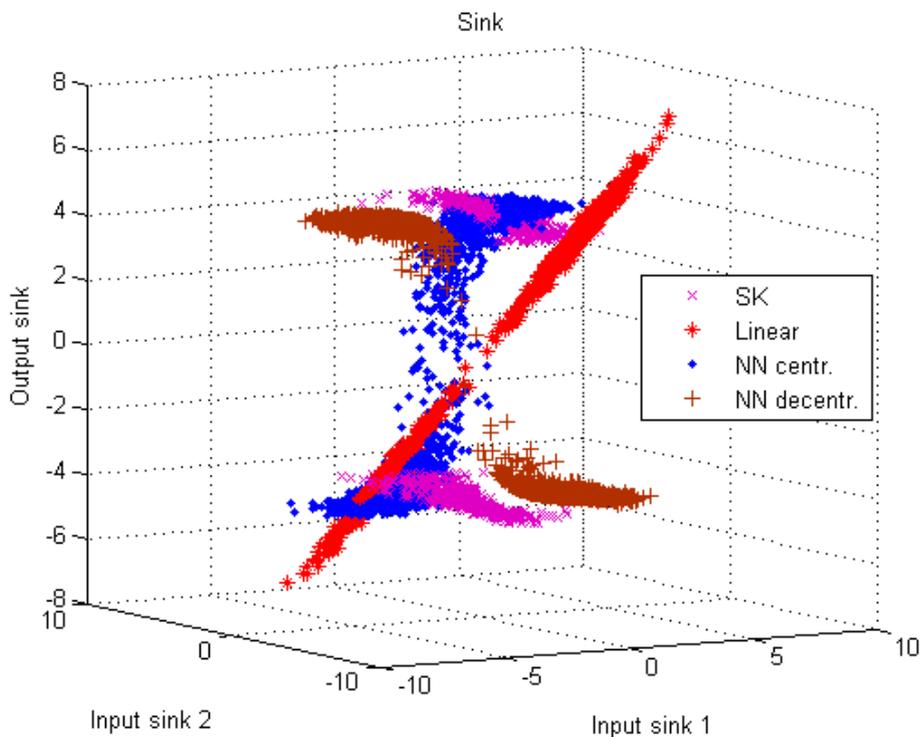


Figure 3. Single source with bimodal distribution: decoder.

8. Conclusions

Despite the apparent simplicity intrinsic in its formulation, zero-delay distributed coding in WSNs is a problem that opens up a surprisingly large spectrum of approaches and interpretations. Different points of view can be perceived by stating it in the framework of disciplines like Information Theory, Team Decision Theory, or Signal Processing. We have attempted to highlight such different viewpoints and formulations in surveying the literature on the topic, in the light of the taxonomy of problem versions introduced in Section 3. The main approaches found can be summarized as in Table 1 below. Besides the general philosophy pertaining to one or the other of the three disciplines considered, we have classified the papers in the literature with respect to the type of transmission channel setting (coherent or orthogonal MAC) and to the distortion measure (scalar or vector). The latter aspect is meaningful to characterize the interest (with respect to the situation depicted in Figure 1) either in the estimation of the random variable representing the very source of the physical phenomenon, or in the estimation of the multiple spatial realizations that are observed by the sensors. As regards the information theoretic formulations, we have distinguished the cases in which fundamental limits are derived and those where the optimality of the AF solution can be proved. Among the Signal Processing methodologies, we have separated the cases regarding: (i) the search of optimal linear encoders and decoder; (ii) the application of nonlinear parametric optimization methods to approximate optimal nonlinear coding and decoding functions (acting on continuous random variables); (iii) the search of optimal quantized encoders.

The optimality of the linear solution has been proved in some cases of the coherent MAC, but no similar results seem to exist for the case of the orthogonal MAC. Structural results exist for problems that are, in principle, substantially more complex than the setting we have considered here, as they involve system dynamics and feedback control systems. The recent book by Yüksel and Başar [85] contains a bulk of results that admirably blend decentralized (team) control theory and information theory and highlight the role of information structures, as summarized also in [107]. We have not gone into details of these aspects; some of the references in [85] are related to our WSN class of problems, and we have indicated it in Table 1. The richest literature referring to the setting represented in Figure 1 appears to be that of Signal Processing, where mostly parametric optimization methods are applied to fixed-form strategies.

Finally, we remark again the relevance of this problem setting for multi-terminal coding in the Physical Layer of WSNs. Notwithstanding the optimality of the zero-delay coding solutions in some cases (often combined with linearity), the application of such analog joint source-channel coding may result particularly attractive in situations where very low power consumption and computational complexity are required, as often happens in WSN deployments in harsh or hardly accessible environments.

Table 1. Classification of different WSN approaches (in the setting studied in this paper) in the literature.

		Coherent MAC	Orthogonal MAC	Scalar Distortion Measure	Vector Distortion Measure
Information Theory	Optimal solution	[9,64–67]		[9,64–67]	
	Fundamental limits	[68–74,76–79]	[75]	[68,69]	[70–79]
Decision Theory			[53–56,85,105] and references therein	[85,105] and references therein	[53–56,85,105] and references therein
Signal Processing	Linear encoders (AF)	[23,26,27,94,95]	[20,21,23,25,27,96]	[20,21,96]	[23,25–27,94,95]
	Nonlinear parametric optimization		[18,19,53–56,100–103,105]	[12,18,105]	[18,19,53–56,100,102,105]
	Quantized encoders		[22,91,92,97,98]	[22,91,97,98]	[92]

Author Contributions

All authors contributed to the survey of the literature. Franco Davoli supervised the overall structure and Carlo Braccini supervised the Information Theory and Signal Processing aspects. Franco Davoli, Mario Marchese and Maurizio Mongelli worked jointly in the NN approaches. Franco Davoli and Maurizio Mongelli contributed to the results of the comparison between SK and NN coders (all numerical results are due to Maurizio Mongelli).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
2. Gastpar, M.; Rimoldi, B.; Vetterli, M. To Code, or Not to Code: Lossy Source-Channel Communication Revisited. *IEEE Trans. Inform. Theory* **2003**, *49*, 1147–1158.
3. Gobblick, T.J., Jr. Theoretical Limitations on the Transmission of Data from Analog Sources. *IEEE Trans. Inform. Theory* **1965**, *11*, 558–567.
4. Wyner, A.D. Another Look at the Coding Theorem of Information Theory—A Tutorial. *IEEE Proc.* **1970**, *58*, 894–913.
5. Gastpar, M.; Vetterli, M. Source-Channel Communication in Sensor Networks. In *IPSN 2003, LNCS 2634*; Zhao, F., Guibas, L., Eds.; Springer: Berlin, Germany, 2003; pp. 162–177.
6. Nazer, B.; Gastpar, M. Structured Random Codes and Sensor Network Coding Theorems. In Proceedings of the International Zurich Seminar on Communications (IZS), Zurich, Switzerland, 12–14 March 2008; pp. 112–115.
7. Gastpar, M.; Vetterli, M.; Dragotti, P.L. Sensing Reality and Communicating Bits: A Dangerous Liaison. *IEEE Signal Proc. Mag.* **2006**, *23*, 70–83.
8. Gastpar, M. Information-Theoretic Bounds on Sensor Network Performance. In *Wireless Sensor Networks: Signal Processing and Communications*; Swami, A., Zhao, Q., Hong, Y.-W., Tong, L., Eds.; Wiley: Hoboken, NJ, USA, 2007.
9. Gastpar, M. Uncoded Transmission Is Exactly Optimal for a Simple Gaussian “Sensor” Network. *IEEE Trans. Inform. Theory* **2008**, *54*, 5247–5251.
10. Gastpar, M.; Vetterli, M. Power, Spatio-Temporal Bandwidth, and Distortion in Large Sensor Networks. *IEEE J. Select. Areas Commun.* **2005**, *23*, 745–754.
11. Karlsson, J. Low-Delay Sensing and Transmission in Wireless Sensor Networks. Licentiate Thesis in Telecommunications, KTH, Stockholm, Sweden, 2008.
12. Karlsson, J.; Skoglund, M. Optimized Low-Delay Source-Channel-Relay Mappings. *IEEE Trans. Commun.* **2010**, *58*, 1397–1404.
13. Karlsson, J.; Skoglund, M. Design and Performance of Optimized Relay Mappings. *IEEE Trans. Commun.* **2010**, *58*, 2718–2724.
14. Marshak, J.; Radner, R. *The Economic Theory of Teams*; Yale University Press: New Haven, CT, USA, 1971.

15. Witsenhausen, H.S. A Counterexample in Stochastic Optimum Control. *SIAM J. Control* **1968**, *6*, 131–147.
16. Ho, Y.C.; Kastner, M.P.; Wong, E. Teams, Signaling, and Information Theory. *IEEE Trans. Autom. Control* **1987**, *23*, 305–311.
17. Baglietto, M.; Parisini, T.; Zoppoli, R. Numerical Solutions to the Witsenhausen Counterexample by Approximating Networks. *IEEE Trans. Autom. Control* **2001**, *49*, 1471–1477.
18. Hekland, F.; Floor, P.A.; Ramstad, T.A. Shannon-Kotel'nikov Mappings in Joint Source-Channel Coding. *IEEE Trans. Commun.* **2009**, *57*, 94–105.
19. Kim, A.N.; Floor, P.A.; Ramstad, T.A.; Balasingham, I. Delay-Free Joint Source-Channel Coding for Gaussian Network of Multiple Sensors. In Proceedings of the IEEE International Conference Communication (ICC 2011), Kyoto, Japan, 5–9 June 2011.
20. Cui, S.; Xiao, J.-J.; Goldsmith, A.J.; Luo, Z.-Q.; Poor, H.V. Energy-Efficient Joint Estimation in Sensor Networks: Analog vs. Digital. In Proceedings of the ICASSP 2005, Philadelphia, PA, USA, 19–23 March 2005; Volume IV, pp. 745–748.
21. Cui, S.; Xiao, J.-J.; Goldsmith, A.J.; Luo, Z.-Q.; Poor, H.V. Estimation Diversity and Energy Efficiency in Distributed Sensing. *IEEE Trans. Signal Proc.* **2007**, *55*, 4683–4695.
22. Xiao, J.-J.; Cui, S.; Luo, Z.-Q.; Goldsmith, A.J. Power Scheduling of Universal Decentralized Estimation in Sensor Networks. *IEEE Trans. Signal Proc.* **2006**, *54*, 413–422.
23. Xiao, J.; Cui, S.; Luo, Z.-Q.; Goldsmith, A.J. Linear Coherent Decentralized Estimation. *IEEE Trans. Signal Proc.* **2008**, *56*, 757–770.
24. Bahceci, I.; Khandani, A.K. Linear Estimation of Correlated Data in Wireless Sensor Networks with Optimum Power Allocation and Analog Modulation. *IEEE Trans. Commun.* **2008**, *56*, 1146–1156.
25. Luo, Z.-Q.; Giannakis, G.B.; Zhang, S. Optimal Linear Decentralized Estimation in a Bandwidth Constrained Sensor Network. In Proceedings of the IEEE International Symposium on Information Theory, Adelaide, Australia, 4–9 September 2005; pp. 1441–1445.
26. Guo, W.; Xiao, X.; Cui, S. An Efficient Water-Filling Solution for Linear Coherent Joint Estimation. *IEEE Trans. Signal Proc.* **2008**, *56*, 5301–5305.
27. Schizas, I.; Giannakis, G.B.; Luo, Z.-Q. Distributed Estimation Using Reduced-Dimensionality Sensor Observations. *IEEE Trans. Signal Proc.* **2007**, *55*, 4284–4299.
28. Kulkarni, R.V.; Förster, A.; Venayagamoorthy, G.K. Computational Intelligence in Wireless Sensor Networks: A Survey. *IEEE Commun. Surveys Tuts.* **2011**, *13*, 68–96.
29. Wang, F.; Liu, J. Networked Wireless Sensor Data Collection: Issues, Challenges, and Approaches. *IEEE Commun. Surveys Tuts.* **2011**, *13*, 673–687.
30. Rajagopalan, R.; Varshney, P.K. Data Aggregation Techniques in Sensor Networks: A Survey. *IEEE Commun. Surv. Tuts.* **2006**, *8*, 48–63.
31. Lin, H.-C.; Kan, Y.-C.; Hong, Y.-M. The Comprehensive Gateway Model for Diverse Environmental Monitoring Upon Wireless Sensor Network. *IEEE Sens. J.* **2011**, *11*, 1293–1303.
32. IEEE 802.15 WPAN™ Task Group 4 (TG4). Available online: <http://www.ieee802.org/15/pub/TG4.html> (accessed on 21 January 2015).
33. IEEE 802.15.1 Offsite Links Page. Available online: <http://ieee802.org/15/Bluetooth/index.html> (accessed on 21 January 2015).

34. ISA 100 Wireless. Available online: <http://www.isa100wci.org> (accessed on 21 January 2015).
35. Rüngeler, M.; Vary, P. Hybrid Digital-Analog Transmission for Wireless Acoustic Sensor Networks. *Signal Process.* **2015**, *107*, 164–170.
36. Jiang, F.; Chen, J.; Swindlehurst, A.L. Detection in Analog Sensor Networks with a Large Scale Antenna Fusion Center. In Proceedings of the 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2014), A Coruña, Spain, 22–25 June 2014; pp. 245–248.
37. Kampionakis, E.; Kimionis, J.; Tountas, K.; Konstantopoulos, C.; Koutroulis, E.; Bletsas, A. Wireless Environmental Sensor Networking with Analog Scatter Radio and Timer Principles. *IEEE Sens. J.* **2014**, *14*, 3365–3376.
38. Vannucci, G.; Bletsas, A.; Leigh, D. A Software-Defined Radio System for Backscatter Sensor Networks. *IEEE Trans. Wireless Commun.* **2008**, *7*, 2170–2179.
39. Kampionakis, E.; Assimonis, S.D.; Bletsas, A. Network Demonstration of Low-Cost and Ultra-Low-Power Environmental Sensing with Analog Backscatter. In Proceedings of the 2014 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet), Newport Beach, CA, USA, 19–23 January 2014; pp. 61–63.
40. Reise, G.; Matz, G. Optimal Transmit Power Allocation in Wireless Sensor Networks Performing Field Reconstruction. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012), Kyoto, Japan, 25–30 March 2012; pp. 3015–3018.
41. Xu, Y.; Hu, J.; Shen, L. Analog Network Coding Interference Mitigation Methods for Wireless Sensor Networks. In Proceedings of the 2012 International Symposium on Wireless Communication Systems (ISWCS 2012), Paris, France, 28–31 August 2012; pp. 516–520.
42. Sarwate, A.; Gastpar, M. A Little Feedback can Simplify Sensor Network Cooperation. *IEEE J. Select. Areas Commun.* **2010**, *28*, 1159–1168.
43. Thejaswi, C.; Cochran, D.; Zhang, J. A Sufficient Condition for Optimality of Digital vs. Analog Relaying in a Sensor Network. In Proceedings of the 43rd Annual Conference on Information Sciences and Systems (CISS 2009), Baltimore, MD, USA, 18–20 March 2009; pp. 202–206.
44. Alirezaei, G.; Mathar, R. Optimum Power Allocation for Sensor Networks that Perform Object Classification. In Proceedings of the 2013 Australasian Telecommunication Networks and Applications Conference (ATNAC 2013), Christchurch, New Zealand, 20–22 November 2013; pp. 1–6.
45. Danieleto, M.; Bui, N.; Zorzi, M. RAZOR: A Compression and Classification Solution for the Internet of Things. *Sensors* **2013**, *14*, 68–94.
46. Goldenbaum, M.; Boche, H.; Stańczak, S. On Analog Computation of Vector-Valued Functions in Clustered Wireless Sensor Networks. In Proceedings of the 46th Annual Conference on Information Sciences and Systems (CISS 2012), Princeton, NJ, USA, 21–13 March 2012; pp. 1–6.
47. Goldenbaum, M.; Boche, H.; Stańczak, S. Harnessing Interference for Analog Function Computation in Wireless Sensor Networks. *IEEE Trans. Signal Proc.* **2013**, *61*, 4893–4906.
48. Rumberg, B.; Graham, D.W.; Kulathumani, V.; Fernandez, R. Hibernets: Energy-Efficient Sensor Networks Using Analog Signal Processing. *IEEE J. Emerg. Select. Top. Circuits Syst.* **2011**, *1*, 321–334.
49. White, D.J.; William, P.E.; Hoffman, M.W.; Balkir, S. Low-Power Analog Processing for Sensing Applications: Low-Frequency Harmonic Signal Classification. *Sensors* **2013**, *13*, 9604–9623.

50. Speyer, J.L.; Chung, W.H. Stochastic Processes, Estimation, and Control. In *SIAM's Series in Advances in Design and Control*; SIAM: Philadelphia, PA, USA, 2008.
51. Vuran, M.C.; Akan, Ö.B.; Akyildiz, I.F. Spatio-Temporal Correlation: Theory and Applications for Wireless Sensor Networks. *Comput. Netw.* **2004**, *45*, 245–259.
52. Vuran, M.C.; Akyildiz, I.F. Spatial Correlation-Based Collaborative Medium Access Control in Wireless Sensor Networks. *IEEE/ACM Trans. Netw.* **2006**, *14*, 316–329.
53. Davoli, F.; Marchese, M.; Mongelli, M. A Decision Theoretic Approach to Gaussian Sensor Networks. In Proceedings of the IEEE International Conference Communication 2009 (ICC 2009), Ad-Hoc and Sensor Networking Symposium, Dresden, Germany, 14–18 June 2009; pp. 1–5.
54. Davoli, F.; Marchese, M.; Mongelli, M. Energy and Distortion Minimization in “Refining” and “Expanding” Sensor Networks. In Proceedings of the 20th Tyrrhenian International Workshop on Digital Communication, Pula, Sardinia, Italy, 2–4 September 2009; In *The Internet of Things*; Giusto, D., Iera, A., Morabito, G., Atzori, L., Eds.; Springer: New York, NY, USA, 2010; pp. 319–327.
55. Davoli, F.; Marchese, M.; Mongelli, M. Simulation-Based Sensitivity Analysis of Optimal Power Mapping in Gaussian Sensor Networks. In Proceedings of the Australasian Telecommunication Networks and Applications Conference 2010 (ATNAC 2010), Auckland, New Zealand, 31 October–3 November 2010; pp. 7–12.
56. Davoli, F.; Marchese, M.; Mongelli, M. Non-Linear Coding and Decoding Strategies Exploiting Spatial Correlation in Wireless Sensor Networks. *IET Commun.* **2012**, *6*, 2198–2207.
57. Wolf, J.K.; Ziv, J. Transmission of Noisy Information to a Noisy Receiver with Minimum Distortion. *IEEE Trans. Inform. Theory* **1970**, *16*, 406–411.
58. Dobrushin, R.L.; Tsybakov, B.S. Information Transmission with Additional Noise. *IRE Trans. Inform. Theory* **1962**, *8*, 293–304.
59. Celandroni, N.; Ferro, E.; Gotta, A.; Oligeri, G.; Roseti, C.; Luglio, M.; Bisio, I.; Cello, M.; Davoli, F.; Panagopoulos, A.D.; *et al.* A Survey of Architectures and Scenarios in Satellite-Based Wireless Sensor Networks: System Design Aspects. *Int. J. Satell. Commun. Netw.* **2013**, *31*, 1–38.
60. Messerschmitt, D.G. Accumulation of Distortion in Tandem Communication Links. *IEEE Trans. Inform. Theory* **1979**, *25*, 692–698.
61. Pilc, R.J. The Optimum Linear Modulator for a Gaussian Source Used with a Gaussian Channel. *Bell Syst. Tech. J.* **1969**, *48*, 3075–3089.
62. Lee, K.-H.; Petersen, D.P. Optimal Linear Coding for Vector Channels. *IEEE Trans. Commun.* **1976**, *24*, 1283–1290.
63. Xiao, J.-J.; Luo, Z.-Q. Multiterminal Source-Channel Communication over an Orthogonal Multiple-Access Channel. *IEEE Trans. Inform. Theory* **2007**, *53*, 3255–3264.
64. Behroozi, H.; Haghghat, J.; Alajaji, F.; Linder, T. On the Transmission of a Memoryless Gaussian Source over a Memoryless Fading Channel. In Proceedings of the 24th Biennial Symposium on Communication, Kingston, ON, Canada, 1–3 June 2008; pp. 212–215.
65. Behroozi, H.; Soleymani, M.R. On the Optimal Power-Distortion Tradeoff in Asymmetric Gaussian Sensor Network. *IEEE Trans. Commun.* **2009**, *57*, 1612–1617.
66. Behroozi, H.; Alajaji, F.; Linder, T. On the Optimal Performance in Asymmetric Gaussian Wireless Sensor Networks with Fading. *IEEE Trans. Signal Proc.* **2010**, *58*, 2436–2441.

67. Wei, S.; Kannan, R.; Iyengar, S.; Rao, N.S. Energy Efficient Estimation of Gaussian Sources over Inhomogeneous Gaussian MAC Channels. In Proceedings of the IEEE Globecom 2008, New Orleans, LO, USA, 30 November–4 December 2008; pp. 1–5.
68. Liu, N.; Ulukus, S. Optimal Distortion-Power Tradeoffs in Gaussian Sensor Networks. In Proceedings of the 2006 IEEE International Symposium Information Theory, Seattle, WA, USA, 9–14 July 2006.
69. Liu, N.; Ulukus, S. Scaling Laws for Dense Gaussian Sensor Networks and the Order Optimality of Separation. *IEEE Trans. Inform. Theory* **2007**, *53*, 3654–3676.
70. Dabeer, O.; Roumy, A.; Guillemot, C. Linear Transceivers for Sending Correlated Sources over the Gaussian MAC. In Proceedings of the 13th National Conference on Communication, Kanpur, India, 26–28 January 2007.
71. Lapidoth, A.; Tinguely, S. Sending a Bivariate Gaussian over a Gaussian MAC. *IEEE Trans. Inform. Theory* **2010**, *56*, 2714–2752.
72. Lapidoth, A.; Tinguely, S. Sending a bivariate Gaussian source over a Gaussian MAC with feedback. *IEEE Trans. Inf. Theory* **2010**, *56*, 1852–1864.
73. Rajesh, R.; Sharma, V. Source-Channel Coding for Gaussian Sources over a Gaussian Multiple Access Channel. In Proceedings of the 45th Annual Allerton Conference, Allerton House, UIUC, IL, USA, 26–28 September 2007; pp. 276–283.
74. Rajesh, R.; Sharma, V. A Joint Source-Channel Coding Scheme for Transmission of Discrete Correlated Sources over a Gaussian Multiple Access Channel. In Proceedings of the International Symposium on Information Theory and Its Applications (ISITA), Auckland, New Zealand, 7–10 December 2008.
75. Rajesh, R.; Sharma, V. Correlated Gaussian Sources over Orthogonal Gaussian Channels. In Proceedings of the International Symposium on Information Theory and Its Applications (ISITA2008), Auckland, New Zealand, 7–10 December 2008.
76. Rajesh, R.; Sharma, V. On Optimal Distributed Joint Source-Channel Coding for Correlated Gaussian Sources over Gaussian Channels. Available online: <http://arxiv.org/abs/1302.3660> (accessed on 21 January 2015).
77. Rajesh, R.; Sharma, V. Transmission of Correlated Sources over a Fading, Multiple Access Channel. In Proceedings of the 46th Annual Allerton Conference, Allerton House, UIUC, IL, USA, 23–26 September 2008; pp. 858–864.
78. Gündüz, D.; Erkip, E.; Goldsmith, A.; Poor, H.V. Source and Channel Coding for Correlated Sources over Multiuser Channels. *IEEE Trans. Inform. Theory* **2009**, *55*, 3927–3944.
79. Jain, A.; Gündüz, D.; Kulkarni, S.R.; Poor, H.V.; Verdú, S. Energy-Distortion Tradeoffs in Gaussian Joint Source-Channel Coding Problems. *IEEE Trans. Inform. Theory* **2012**, *58*, 3153–3168.
80. Berger, T.; Zhang, Z.; Viswanathan, H. The CEO Problem. *IEEE Trans. Inform. Theory* **1996**, *42*, 887–902.
81. Viswanathan, H.; Berger, T. The Quadratic Gaussian CEO Problem. *IEEE Trans. Inform. Theory* **1997**, *43*, 1549–1559.
82. Oohama, Y. The Rate-Distortion Function for the Quadratic Gaussian CEO Problem. *IEEE Trans. Inform. Theory* **1998**, *44*, 1057–1070.

83. Oohama, Y. Distributed Source Coding of Correlated Gaussian Remote Sources. *IEEE Trans. Inform. Theory* **2012**, *58*, 5059–5085.
84. Yang, Y.; Xiong, Z. On the Generalized Gaussian CEO Problem. *IEEE Trans. Inform. Theory* **2012**, *58*, 3350–3372.
85. Yüksel, S.; Başar, T. Stochastic Networked Control Systems—Stabilization and Optimization under Information Constraints, Birkhäuser, New York, NY, USA, 2013.
86. Ho, Y.-C.; Chu, K.-C. Team Decision Theory and Information Structures in Optimal Control Problems—Part I. *IEEE Trans. Autom. Control* **1972**, *17*, 15–22.
87. Witsenhausen, H.S. Equivalent Stochastic Control Problems. *Math. Control Signals Syst.* **1988**, *1*, 3–11.
88. Gnecco, G.; Sanguineti, M. New Insights into Witsenhausen’s Counterexample. *Optim. Lett.* **2012**, *6*, 1425–1446.
89. Lim, A.E.B.; Moore, J.B.; Faybusovich, L. Separation Theorem for Linearly Constrained LQG Optimal Control. *Syst. Control Lett.* **1996**, *28*, 227–235.
90. Akyol, E.; Viswanatha, K.; Rose, K.; Ramstad, T. On Zero Delay Source-Channel Coding. Available online: <http://arxiv.org/pdf/1302.3660v1.pdf> (accessed on 21 January 2015).
91. Ribeiro, A.; Giannakis, G.B. Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks—Part I: Gaussian Case. *IEEE Trans. Signal Proc.* **2006**, *54*, 1131–1143.
92. Ribeiro, A.; Giannakis, G.B. Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks—Part II: Unknown Probability Density Function. *IEEE Trans. Signal Proc.* **2006**, *54*, 2784–2796.
93. Madan, R.; Cui, S.; Lall, S.; Goldsmith, A.J. Modeling and Optimization of Transmission Schemes in Energy-Constrained Wireless Sensor Networks. *IEEE/ACM Trans. Netw.* **2007**, *15*, 1359–1372.
94. Esnaola, I.; Garcia-Frias, J. Distributed Analog Linear Coding of Correlated Gaussian Sources over Multiple Access Channels. In Proceedings of the 6th International Symposium on Wireless Communication Systems (ISWCS’09), Siena, Italy, 7–10 September 2009; pp. 288–292.
95. Esnaola, I.; Garcia-Frias, J. Distributed Analog Linear Coding of Correlated Gaussian Sources. In Proceedings of the 47th Annual Allerton Conference, Allerton House, UIUC, IL, USA, 30 September–2 October 2009; pp. 857–864.
96. Chaudhary, M.H.; Vandendorpe, L. Adaptive Power Allocation in Wireless Sensor Networks with Spatially Correlated Data and Analog Modulation: Perfect and Imperfect CSI. *EURASIP J. Wirel. Commun. Netw.* **2010**, doi:10.1155/2010/817961.
97. Chaudhary, M.H.; Vandendorpe, L. Power Constrained Linear Estimation in Wireless Sensor Networks with Correlated Data and Digital Modulation. *IEEE Trans. Signal Proc.* **2012**, *60*, 570–584.
98. Aysal, T.C.; Barner, K.E. Constrained Decentralized Estimation over Noisy Channels for Sensor Networks. *IEEE Trans. Signal Proc.* **2008**, *56*, 1398–1410.
99. Li, J.C.F.; Dey, S.; Evans, J. Maximal Lifetime Power and Rate Allocation for Wireless Sensor Systems with Data Distortion Constraints. *IEEE Trans. Signal Proc.* **2008**, *56*, 2076–2090.
100. Floor, P.A. On the Theory of Shannon-Kotel’nikov Mappings in Joint Source-Channel Coding. Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2008.

101. Akyol, E.; Rose, K.; Ramstad, T. Optimal Mappings for Joint Source Channel Coding. In Proceedings of the IEEE Information Theory Workshop (ITW 2010), Dublin, Ireland, 30 August–3 September 2010; pp. 150–154.
102. Hu, Y.; Garcia-Frias, J.; Lamarca, M. Analog Joint Source-Channel Coding Using Non-Linear Curves and MMSE Decoding. *IEEE Trans. Commun.* **2011**, *59*, 3016–3026.
103. Fresnedo, O.; Vazquez-Araujo, F.J.; Castedo, L.; Garcia-Frias, J. Low-Complexity Near-Optimal Decoding for Analog Joint Source Channel Coding Using Space-Filling Curves. *IEEE Commun. Lett.* **2013**, *17*, 745–748.
104. Haykin, S. *Neural Networks, a Comprehensive Foundation*; MacMillan Publishing: New York, NY, USA, 1994.
105. Davoli, F.; Mongelli, M. Neural Approximations of Analog Joint Source-Channel Coding. *IEEE Signal Proc. Lett.* **2015**, *22*, 421–425.
106. Floor, P.A.; Ramstad, T.A. Dimension Reducing Mappings in Joint Source-Channel Coding. In Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG 2006), 7–9 June 2006; pp. 282–285.
107. Mahajan, A.; Martins, N.C.; Rotkowitz, M.; Yüksel, S. Information Structures in Optimal Decentralized Control. In Proceedings of the 51st IEEE Conference on Decision and Control (CDC 2012), Maui, HI, USA, 10–13 December 2012; pp. 1291–1306.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).