Signal Processing 107 (2015) 141–152

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Audio coding in wireless acoustic sensor networks ☆

Adel Zahedi [a,*], Jan Østergaard [a], Søren Holdt Jensen [a], Søren Bech [a,b], Patrick Naylor [c]

[a] Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark
[b] Bang & Olufsen, 7600 Struer, Denmark
[c] Electrical and Electronic Engineering Department, London Imperial College, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form
22 July 2014
Accepted 25 July 2014
Available online 7 August 2014

Keywords:
Wireless acoustic sensor network
Distributed source coding
Rate-distortion function
Covariance distortion constraint
Zero-error coding

ABSTRACT

In this paper, we consider the problem of source coding for a wireless acoustic sensor network where each node in the network makes its own noisy measurement of the sound field, and communicates with other nodes in the network by sending and receiving encoded versions of the measurements. To make use of the correlation between the sources available at the nodes, we consider the possibility of combining the measurement and the received messages into one single message at each node instead of forwarding the received messages and separate encoding of the measurement. Moreover, to exploit the correlation between the messages received by a node and the node's measurement of the source, we propose to use the measurement as side information and thereby form a distributed source coding (DSC) problem. Assuming that the sources are Gaussian, we then derive the rate-distortion function (RDF) for the resulting remote DSC problem under covariance matrix distortion constraints. We further show that for this problem, the Gaussian source is the worst to code. Thus, the Gaussian RDF provides an upper bound to other sources such as audio signals. We then turn our attention to audio signals. We consider an acoustical model based on the room impulse response (RIR) and provide simulation results for the rate-distortion performance in a practical setup where a set of microphones record the sound in a standard listening room. Since our reconstruction scheme and distortion measure are defined over the direct sound source, coding and dereverberation are performed in a joint manner.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The fact that sensor nodes in a Wireless Acoustic Sensor Network (WASN) are wireless, necessitates parsimony in spending the resources, namely power and bandwidth [1]. Two possible remedies are to allow the sensor nodes to communicate with neighboring nodes in order to save communication power, and to use source coding to compress the data acquired by the microphones before transmission in order to reduce the required transmission rate, leading to a better power-bandwidth trade-off.

There are three main consumers of power in a typical sensor node, i.e. sensing, signal processing, and communication units, where the last one is the dominant [2]. Electromagnetic wave power is a highly superlinear function of the inverse of distance, and thus fades rapidly as the distance increases. This means that a multihop scenario where instead of directly being sent to a far-end

☆ The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant agreement n° ITN-GA-2012-316969.

* Corresponding author.
E-mail addresses: adz@es.aau.dk (A. Zahedi), jo@es.aau.dk (J. Østergaard), shj@es.aau.dk (S.H. Jensen), sbe@es.aau.dk (S. Bech), p.naylor@imperial.ac.uk (P. Naylor).

http://dx.doi.org/10.1016/j.sigpro.2014.07.021
0165-1684/© 2014 Elsevier B.V. All rights reserved.

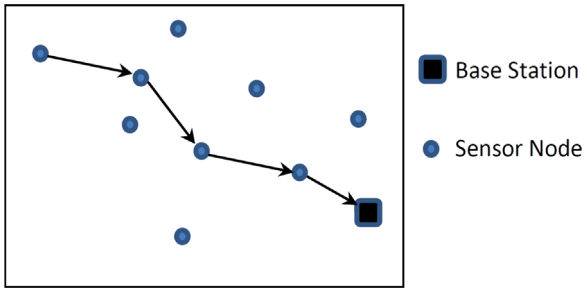

Fig. 1. Message forwarding by multihop communication in a WASN.

destination, the message goes through a sequence of transmissions in short distances, can be more power-efficient. To implement such a scenario, one should assume that neighboring nodes are able to communicate, and thus it is possible to let the nodes which are far from the base station deliver their messages by sequential forwarding through neighboring nodes. This is illustrated in Fig. 1.

To lower the required transmission rate, one may consider source coding for data compression. The audio signals acquired by the microphones are typically highly redundant and can therefore be compressed before transmission. The redundancy is due to correlation in time between the samples of the data acquired by a microphone, as well as spatial correlation between the measurements of the sound field performed by microphones placed at different locations in the environment.

Applying a data compression technique separately to the sequences (the data acquired by the microphones and also those received from neighboring microphones) is equivalent to making use of the time correlation between the samples of a sequence, but ignoring the spatial correlation between different sequences of data.

There are several possibilities to exploit the spatial correlation to further reduce the rate. One possibility is to jointly encode the sequences received by a microphone and the microphone's measurement into one single message instead of separate encoding [3]. Another possibility is to consider the spatial correlation between the message sent from a node and the measurement of the sound field available at the destination node. This latter can be utilized as side information at the decoder to reduce the rate at the encoder. Finally, assuming that nodes transmitting to a neighboring receiving node have the knowledge that which other nodes are sending messages to the same receiving node, it is possible to reduce in the rates of the transmitting nodes due to correlation between their messages by applying e.g. asymmetric or non-asymmetric Slepian–Wolf coding [4]. However, in this work, we do not make this assumption, and thus consider only the two first cases as done in [5].

To exploit the availability of the side information at the encoder for reduction in the rate, we rely on results from distributed source coding (DSC), which was started by Slepian and Wolf in [6]. In particular, it is possible to separately encode two correlated discrete sources at a sum-rate equal to the joint entropy of the two sources and reconstruct the sources at a joint decoder. This result was then extended in [7,8] to the case of continuous sources, assuming that one source is available at the decoder and the other one is to be discretized and encoded for a given distortion. It was shown that for Gaussian sources under an MSE distortion constraint, the rate can be as low as the case where the side information is also available at the encoder.

Measurements made by wireless microphones are generally digitized to discrete-time and discrete-amplitude sequences. This incurs some distortion depending on the quantization step-size. For a given distortion, there is a minimum achievable rate given by the so-called rate-distortion function (RDF). The RDF can be used as a lower bound to assess the performance of any source coding scheme. The interested reader is referred to [9,10] for more information on source coding theory.

In this paper, we derive the local RDF for an arbitrary node in a WASN, assuming that multiple sources at a transmitting node are jointly encoded taking into account (in a distributed sense) the side information available at the receiving node. We will then show in the simulations that local RDFs can be used for rate allocation in the network to achieve the optimal sum-rate. We will also design a coding scheme, which in theory and under certain asymptotical conditions, achieves the theoretical RDF. Vector sources will be considered to allow for modeling of memory in the sources, and distortion constraints are defined in form of covariance matrices for generality. This paper extends our previous work [5] by first proving that Gaussian RDF is an upper bound to other distributions including audio, and then applying our results to Gaussian as well as audio signals. We also provide a complete proof for the RDF, which was only sketched in [5]. In Section 2, the problem is formulated, notations and assumptions are mentioned, and the acoustic channel model used in this paper will be discussed. In Section 3, sufficient statistics are used for joint encoding of correlated sources into a single source in a DSC scenario. Then the RDF will be derived for vector Gaussian sources with noisy measurements under covariance matrix distortion constraints. We further prove that for this setup, Gaussian distribution is the worst case for coding in terms of rate-distortion. In Section 4, we provide simulation results for Gaussian sources and real audio measurements. Section 5 concludes the paper.

2. Preliminaries

2.1. Notation, assumptions, and problem formulation

We denote random vectors by lowercase boldface, matrices by uppercase boldface, and scalars by italic letters. The operations $\mathrm{tr}(\cdot)$, $I(\cdot;\cdot)$, $h(\cdot)$, and $E[\cdot]$ stand for trace of a matrix, mutual information, differential entropy and expectation, respectively. Markov chains are denoted by two-headed arrows; e.g. $\mathbf{x}\leftrightarrow\mathbf{y}\leftrightarrow\mathbf{z}$. Probability density functions are denoted by $f(\cdot)$ and covariance and cross-covariance matrices are denoted by symbol $\mathbf{\Sigma}$ followed by a subscript indicating the random vectors involved and possibly conditioning. We assume that all covariance matrices are of full rank. For two positive semidefinite

matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, $\mathbf{\Sigma}_1 \succcurlyeq \mathbf{\Sigma}_2$ means that $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is positive semidefinite.

The source is assumed to generate independent Gaussian vectors $\mathbf{x} \in \mathbb{R}^n$, i.e. the vectors are mutually independent, but the elements within a vector can be arbitrarily correlated. Node j in the network makes a noisy measurement $\mathbf{z}_j$ of the source. We denote by $\mathbf{y}_j$ a version of the measurement, which will be used for source coding. This might be a preprocessed version of $\mathbf{z}_j$, or simply $\mathbf{z}_j$ itself (see Section 2.2). We refer to both $\mathbf{y}_j$ and $\mathbf{z}_j$ as the measurement at node j when the meaning is clear from the context. Node j in general receives messages from arbitrary nodes $i, l, \ldots$ denoted by $\mathbf{u}_i, \mathbf{u}_l, \ldots$. The structure of such a network is illustrated in Fig. 2.

In a simple forwarding scenario, the received messages are forwarded to node k as they are, and $\mathbf{y}_j$ is also encoded into a message and transmitted to node k. We propose to go beyond the message forwarding scenarios via the following improvements:

- The measurement made by node k is correlated with the measurement made and messages received by node j. This means that the two nodes have a part of information about the source in common. To avoid transmitting this redundant data, we propose to utilize DSC and make use of each node's measurement as side information for decoding the received messages.
- To reduce the rate by making use of the correlation between the messages received and the measurement made by the same node, instead of encoding and sending the measurement and forwarding the received messages separately, we propose to combine all into one message which contains all the information about the source available in the measurement and the received messages.

We formulate the problem based on the above ideas as follows. By use of the measurement $\mathbf{y}_j$ as the decoder side information, node j is able to decode the received messages and form the estimations $\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_l, \ldots$ of the source. The problem is to compute the minimum rate R_j for joint encoding of $\mathbf{y}_j, \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_l, \ldots$ into a message $\mathbf{u}_j$ to be sent to node k. Moreover, it is required that the estimation $\hat{\mathbf{x}}_j$ of the source using some estimation function $\psi(\mathbf{u}_j, \mathbf{y}_k)$ satisfy the following constraint:

$$E[(\mathbf{x}-\hat{\mathbf{x}}_j)(\mathbf{x}-\hat{\mathbf{x}}_j)^T] \preceq \mathbf{D}_j, \tag{1}$$

where $\mathbf{D}_j$ is the covariance matrix distortion constraint for node j. This problem is illustrated in Fig. 3.

Note that the encoding and decoding process for each message involves two nodes, here indicated by subscripts j and k. The encoding is done at node j subject to a distortion constraint $\mathbf{D}_j$. The reconstruction although done at node k is denoted by $\hat{\mathbf{x}}_j$ to emphasize that the distortion constraint $\mathbf{D}_j$ has to be satisfied according to (1). Also notice that the distortion is defined with respect to the direct sound source $\mathbf{x}$ rather than the measurements (which are corrupted by the room effects), and the reconstructions are approximations of $\mathbf{x}$. Therefore a dereverberation effect is inherent in the encoding and decoding procedure.

After deriving the local RDFs $R_j(\mathbf{D}_j)$ for nodes $j = 1, \ldots, M$, one can consider a constraint for example on the distortion $\mathbf{D}_{bs}$ at the base station and formulate the following sum-rate minimization problem over the set of distortion constraints at the intermediate nodes:

$$\min_{\mathbf{D}_1,\ldots,\mathbf{D}_M} \sum_{j=1}^{M} R_j(\mathbf{D}_j) \quad \text{s.t. } \mathbf{D}_{bs} \preceq \mathbf{D}_0, \tag{2}$$

where $\mathbf{D}_0$ is the desired distortion at the base station. We do not solve this problem analytically due to the complex form of the local RDFs. Instead, in Section 4, we consider a scenario where two nodes transmit to a destination (which might be a third node or a base station). For given distortions at the destination, we experimentally find the optimal choice of the rates for the two transmitting nodes to minimize the sum-rate. (See Figs. 6 and 11.)

2.2. Acoustic channel model

We consider the following model as the relationship between the source $\mathbf{x}$ and the preprocessed measurement $\mathbf{y}_j$:

$$\mathbf{y}_j = \mathbf{A}_j \mathbf{x} + \mathbf{n}_j, \tag{3}$$

where $\mathbf{A}_j$ is called the mixing matrix, and $\mathbf{n}_j$ is the additive noise, which is independent of the source. We consider two possible ways to relate these parameters to the acoustical environment in the sequel.

Assuming that the entire acoustical system is linear, the measurement $\mathbf{z}_j$ made by the microphone at node j can be written as

$$\mathbf{z}_j = \mathbf{h}_j \star \mathbf{x} + \mathbf{v}_j, \tag{4}$$

where $\mathbf{h}_j = [h_{j1} \ldots h_{jm}]$ is the impulse response of the system from the sound source to node j, $\star$ denotes

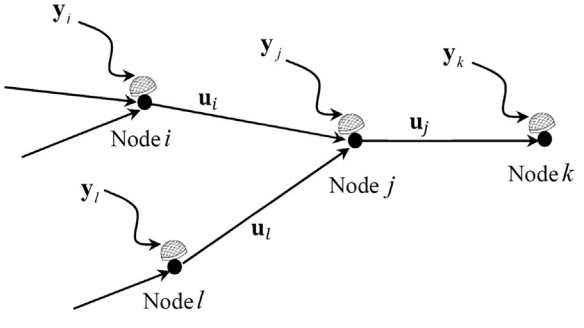

Fig. 2. Schematic diagram of a WASN with the possibility of communication between the nodes [5].

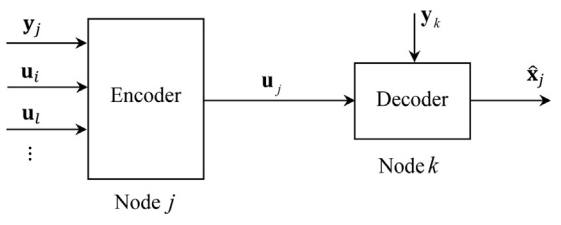

Fig. 3. Joint distributed encoding of the measurement and messages.

convolution, and $\mathbf{v}_j$ is the additive noise. In the steady state, this can be rewritten as follows [11]:

$$\mathbf{z}_j = \begin{bmatrix} h_m & h_{m-1} & \dots & h_2 & h_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & h_m & \dots & h_3 & h_2 & h_1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & h_m & h_{m-1} & \dots & h_2 & h_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \mathbf{v}_j$$

$$= \mathbf{H}_j \mathbf{x} + \mathbf{v}_j. \tag{5}$$

Eq. (5) could be used as the relationship between the sound source and the measurements made by the microphones. In this case, the source coding scheme would be directly applied to the raw measurement $\mathbf{z}_j$. In other words, with this model, we have $\mathbf{y}_j = \mathbf{z}_j$, $\mathbf{A}_j = \mathbf{H}_j$, and $\mathbf{n}_j = \mathbf{v}_j$. However, there are some drawbacks with this model. First, for a normal room with a reverberation time of hundreds of milliseconds, the RIR will be of the order of thousands of samples. This means that the size of the matrix $\mathbf{H}_j$ in (5) will be of order thousands times thousands, and the encoding and decoding procedures should be applied to vectors of dimension thousands. This leads to impractical memory requirements and computational burden for the signal processing unit. Moreover, $\mathbf{A}_j$ is non-square, which leads to different dimensions for the source $\mathbf{x}$ and the measurement $\mathbf{y}_j$. This makes the source coding problem mathematically more complicated. In particular, we need to have the mixing matrices invertible to derive the lower bound on the RDF in Section 3.2.1, which is not the case with this model.

Another possibility is to perform some preprocessing on the measurement $\mathbf{z}_j$. Since the measurement is a noisy version of the convolution of the sound source and the RIR as in (4), one can obtain a reconstruction of the source by deconvolution provided that $\mathbf{h}_j$ is known. In this case, $\mathbf{y}_j$ is the result of deconvolution and can be written as

$$\mathbf{y}_j = \mathbf{x} + \mathbf{n}_j, \tag{6}$$

where $\mathbf{n}_j$ is the resulting noise. In this case we have $\mathbf{A}_j = \mathbf{I}$, which is invertible, and also the dimension of the vectors need not be large. The problem with this model is that the RIR might have zeros (or almost zeros) at certain frequencies, and if the additive noise $\mathbf{v}_j$ in (4) has nonzero components at those frequencies, with a simple deconvolution, the reconstruction in (6) will suffer from strong colored noise. To alleviate this problem, one may consider more sophisticated approaches to deconvolution. (See e.g. [11] for a review of Bayesian approaches.)

3. Distributed source coding for WASN

In this section, we solve the problem formulated in Section 2.1. First we consider Gaussian sources, and reduce the problem to a DSC problem with a single source at the encoder. Then we derive the RDF for this problem. Finally, we will show that for our DSC problem with covariance distortion constraint, a Gaussian distribution is the worst for source coding, i.e. for a given distortion and source covariance, any other distribution requires a lower rate compared to the Gaussian case. The use of this fact is twofold for our audio coding problem. Firstly, it shows that for audio signals the RDF is upperbounded by the Gaussian RDF, and thus can be used as a guideline. Secondly, the coding schemes designed based on the Gaussian assumption can be used for encoding audio signals with a performance not worse that the Gaussian case.

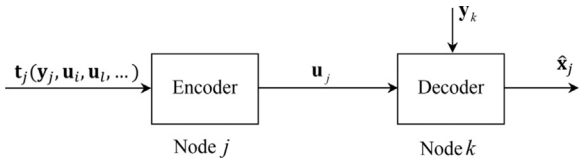

Fig. 4. The equivalent DSC problem after combining the sources into a sufficient statistic.

3.1. Sufficient statistics

We show how multiple sources $\mathbf{y}_j, \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_l, \dots$ which are to be encoded jointly at node j as shown in Fig. 3 can be losslessly represented by a single source. This is suggested by the following theorem.

Theorem 1. *Given side information $\mathbf{y}_k$ at the decoder, the RDF for joint encoding of multiple sources $\mathbf{y}_j, \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_l, \dots$ coincides with the RDF for encoding the single source*:

$$\mathbf{t}_j = \mathbf{A}_j^T \boldsymbol{\Sigma}_j^{-1} \mathbf{y}_j + \mathbf{D}_i^{-1} \hat{\mathbf{x}}_i + \mathbf{D}_l^{-1} \hat{\mathbf{x}}_l + \cdots \tag{7}$$

where $\mathbf{D}_i$ is the target distortion for node i.

Proof. See [5]. □

Theorem 1 is based on the concept of sufficiency. $\mathbf{t}_j$ in (7) is sufficient for representing the sources $\mathbf{y}_j, \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_l, \dots$, in sense that given $\mathbf{y}_k$, all the information about $\mathbf{x}$ in those sources is also available in $\mathbf{t}_j$. This means that the problem formulated in Fig. 3 can be simplified to a single source remote source coding problem with side information as shown in Fig. 4.[1]

As shown in [12], a sufficient statistic can be used for joint encoding of multiple sources in a non-distributed source coding scenario. For our DSC problem, we required a modification of the definition of sufficient statistics, which we refer to as conditional sufficient statistics, see [5] for more details.

3.2. Rate-distortion functions

After reducing the problem to the one with a single source as shown in Fig. 4, we derive a lower bound on the RDF. Then we design a coding scheme which achieves this lower bound. This proves that the lower bound is indeed tight, and also shows how the source should be coded to achieve the RDF.

The RDF for the scalar case of this problem was derived in [13]. For the vector case with mean-squared error distortion constraint, the RDF was found in [14], while the case of covariance distortion was not treated in those works. We derive the RDF for the covariance distortion

[1] This could be equivalently done by putting the measurements and estimations from the received messages into a vector, and then encoding the aggrandized vector. We have preferred using a sufficient statistic to avoid vectors with large dimensions.

constraint assuming that the mixing matrices are invertible.

Without loss of generality, using linear estimation, we can write $\mathbf{x}$ and $\mathbf{t}_j$ as

$$\mathbf{t}_j = \mathbf{A}\mathbf{y}_k + \mathbf{B}\mathbf{u}_j + \boldsymbol{\upsilon}_1, \tag{8}$$

$$\mathbf{x} = \mathbf{C}\mathbf{t}_j + \mathbf{G}\mathbf{y}_k + \boldsymbol{\upsilon}_2 = (\mathbf{G}+\mathbf{C}\mathbf{A})\mathbf{y}_k + \mathbf{C}\mathbf{B}\mathbf{u}_j + \mathbf{C}\boldsymbol{\upsilon}_1 + \boldsymbol{\upsilon}_2, \tag{9}$$

$$\mathbf{t}_j = \boldsymbol{\Gamma}\mathbf{y}_k + \boldsymbol{\upsilon}_3, \tag{10}$$

where $\boldsymbol{\upsilon}_1$, $\boldsymbol{\upsilon}_2$, and $\boldsymbol{\upsilon}_3$ are estimation errors with covariance matrices $\boldsymbol{\Sigma}_{\mathbf{t}_j|\mathbf{y}_k,\mathbf{u}_j}$, $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}$, and $\boldsymbol{\Sigma}_{\mathbf{t}_j|\mathbf{y}_k}$, respectively, and $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G}$ and $\boldsymbol{\Gamma}$ are the coefficients of linear estimation and depend only on the covariance and cross-covariance matrices of $\mathbf{x},\mathbf{t}_j,\mathbf{y}_k$ and $\mathbf{u}_j$. As an example, $\mathbf{C}$ is given by

$$\mathbf{C}\boldsymbol{\Sigma}_{\mathbf{t}_j} = \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{t}_j} + (\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{t}_j}\boldsymbol{\Sigma}_{\mathbf{t}_j}^{-1}\boldsymbol{\Sigma}_{\mathbf{t}_j\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_k})(\boldsymbol{\Sigma}_{\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{t}_j\mathbf{y}_k}^T\boldsymbol{\Sigma}_{\mathbf{t}_j}^{-1}\boldsymbol{\Sigma}_{\mathbf{t}_j\mathbf{y}_k})^{-1}\boldsymbol{\Sigma}_{\mathbf{t}_j\mathbf{y}_k}^T. \tag{11}$$

The rest of the coefficients can be derived using symmetry or by setting a variable equal to zero. The following equalities, both resulting from (9), will be used later for the derivation of the RDF:

$$\hat{\mathbf{x}}_j = E[\mathbf{x}|\mathbf{u}_j,\mathbf{y}_k] = (\mathbf{G}+\mathbf{C}\mathbf{A})\mathbf{y}_k + \mathbf{C}\mathbf{B}\mathbf{u}_j, \tag{12}$$

$$\mathbf{C}\boldsymbol{\Sigma}_{\mathbf{t}_j|\mathbf{y}_k}\mathbf{C}^T = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}. \tag{13}$$

We will show that if the mixing matrices $\mathbf{A}_j$ in (3) are invertible, then for $\boldsymbol{\Sigma}_{x|\mathbf{t}_j,\mathbf{y}_k} \prec \mathbf{D}_j \preceq \boldsymbol{\Sigma}_{x|\mathbf{y}_k}$ the RDF for node j is given by

$$R_j(\mathbf{D}_j) = \frac{1}{2}\log\left(\frac{|\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}|}{|\mathbf{D}_j - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}|}\right). \tag{14}$$

The RDF (14) generalizes the RDF derived in [13] for the scalar Gaussian case.

3.2.1. Lower bound

To derive the lower bound, we need the following lemma:

Lemma 1. *If the mixing matrices $\mathbf{A}_j, j = 1,2,\ldots,$ are invertible, the matrix $\mathbf{C}$ in (9) is also invertible.*

Proof. See the appendix in [5]. □

Starting from $R_j(\mathbf{D}_j) \geq I(\mathbf{t}_j;\mathbf{u}_j|\mathbf{y}_k)$, and using (13) and Gaussianity of the source and the measurements one can show that if $\mathbf{C}$ is invertible, the RDF can be lower-bounded by (14). See [5] for more details.

3.2.2. Coding scheme

We will show that the lower bound we derived in 3.2.1 is achievable by a coding scheme, and is indeed the RDF. Let us assume that $\mathbf{t}_j$ is encoded into $\mathbf{u}_j$ according to some coding scheme. Due to the availability of the side information $\mathbf{y}_k$ at the decoder, it is possible with high probability to noiselessly encode $\mathbf{u}_j$ in blocks of sufficiently large length N with a rate arbitrarily close to $H(\mathbf{u}_j|\mathbf{y}_k)$ [6]. This conditional entropy can be rewritten as

$$H(\mathbf{u}_j|\mathbf{y}_k) = I(\mathbf{t}_j;\mathbf{u}_j|\mathbf{y}_k) + H(\mathbf{u}_j|\mathbf{t}_j,\mathbf{y}_k) = I(\mathbf{t}_j;\mathbf{u}_j|\mathbf{y}_k). \tag{15}$$

Based on the above argument, to show that (14) is achievable, one should design a scheme for which $I(\mathbf{t}_j;\mathbf{u}_j|\mathbf{y}_k)$ coincides with (14), and the reconstruction error at the decoder satisfies the distortion constraint. We will show that this is possible by using the following scheme:

$$\mathbf{u}_j = \mathbf{U}\mathbf{C}\mathbf{t}_j + \boldsymbol{\nu}_j, \tag{16}$$

where the eigenvalue decomposition of $(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})$ is assumed to be $\mathbf{U}^T\boldsymbol{\Lambda}\mathbf{U}$, i.e.

$$\boldsymbol{\Lambda} = \mathbf{U}(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})\mathbf{U}^T \tag{17}$$

and the covariance $\boldsymbol{\Sigma}_{\boldsymbol{\nu}_j}$ of the coding noise $\boldsymbol{\nu}_j$ is given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\nu}_j} = \mathbf{U}(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \mathbf{D}_j)^{-1}(\mathbf{D}_j - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})\mathbf{U}^T. \tag{18}$$

This implies that $\mathbf{t}_j$ is first transformed by the linear map $\mathbf{U}\mathbf{C}$ and then quantized, which amounts to adding the noise $\boldsymbol{\nu}_j$. We assume the existence of a high dimensional vector quantizer in which the coding noise $\boldsymbol{\nu}_j$ becomes Gaussian distributed, thus $\mathbf{u}_j$ belongs to a continuous alphabet and $h(\mathbf{u}_j)$ is well-defined [15].

To verify that (16) achieves (14), we can write

$$I(\mathbf{t}_j;\mathbf{u}_j|\mathbf{y}_k) = h(\mathbf{u}_j|\mathbf{y}_k) - h(\mathbf{u}_j|\mathbf{t}_j,\mathbf{y}_k) \tag{19}$$

$$= h(\mathbf{U}\mathbf{C}\mathbf{t}_j + \boldsymbol{\nu}_j|\mathbf{y}_k) - h(\boldsymbol{\nu}_j) \tag{20}$$

$$= h(\mathbf{U}\mathbf{C}\boldsymbol{\Gamma}\mathbf{y}_k + \mathbf{U}\mathbf{C}\boldsymbol{\upsilon}_3 + \boldsymbol{\nu}_j|\mathbf{y}_k) - h(\boldsymbol{\nu}_j) \tag{21}$$

$$= h(\mathbf{U}\mathbf{C}\boldsymbol{\upsilon}_3 + \boldsymbol{\nu}_j) - h(\boldsymbol{\nu}_j)$$
$$= \frac{1}{2}\log\left(\frac{|\mathbf{U}\mathbf{C}\boldsymbol{\Sigma}_{\mathbf{t}_j|\mathbf{y}_k}\mathbf{C}^T\mathbf{U}^T + \boldsymbol{\Sigma}_{\boldsymbol{\nu}_j}|}{|\boldsymbol{\Sigma}_{\boldsymbol{\nu}_j}|}\right) \tag{22}$$

$$= \frac{1}{2}\log\left(\frac{|\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}|}{|\mathbf{D}_j - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}|}\right), \tag{23}$$

where (20) is the result of substituting (16) in (19), (21) follows from (10), and (23) is the result of substituting (13) and (18) in (22) and simplifying the result.

Now we should show that the coding scheme (16) leads to a reconstruction error at the decoder which satisfies the distortion constraint (1). Let us define $\mathbf{P} \triangleq \mathbf{G}+\mathbf{C}\mathbf{A}$, $\mathbf{Q} \triangleq \mathbf{C}\mathbf{B}$, and $\boldsymbol{\upsilon}_4 \triangleq \mathbf{C}\boldsymbol{\upsilon}_1 + \boldsymbol{\upsilon}_2$. Then from (9) we have

$$\mathbf{x} = \mathbf{P}\mathbf{y}_k + \mathbf{Q}\mathbf{u}_j + \boldsymbol{\upsilon}_4. \tag{24}$$

From (24) it follows that $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{u}_j|\mathbf{y}_k} = \mathbf{Q}\boldsymbol{\Sigma}_{\mathbf{u}_j|\mathbf{y}_k}$, or

$$\mathbf{Q} = \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{u}_j|\mathbf{y}_k}\boldsymbol{\Sigma}_{\mathbf{u}_j|\mathbf{y}_k}^{-1}. \tag{25}$$

From (16) and (13) we have

$$\boldsymbol{\Sigma}_{\mathbf{u}_j|\mathbf{y}_k} = \boldsymbol{\Lambda} + \boldsymbol{\Sigma}_{\boldsymbol{\nu}_j}. \tag{26}$$

Using (16) we can also write

$$\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{u}_j|\mathbf{y}_k} = \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{t}_j|\mathbf{y}_k}\mathbf{C}^T\mathbf{U}^T \tag{27}$$

$$= \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{t}_j|\mathbf{y}_k}\mathbf{C}^T\mathbf{U}^T \tag{28}$$

$$= (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})\mathbf{U}^T, \tag{29}$$

where (28) and (29) follow from (9) and (13), respectively. Combining the above equations, we can write the covariance matrix of the reconstruction error as

$$E[(\mathbf{x}-\hat{\mathbf{x}}_j)(\mathbf{x}-\hat{\mathbf{x}}_j)^T] = \boldsymbol{\Sigma}_{\boldsymbol{\upsilon}_4} \tag{30}$$

$$= \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \mathbf{Q}\boldsymbol{\Sigma}_{\mathbf{u}_j|\mathbf{y}_k}\mathbf{Q}^T \tag{31}$$

$$= \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{u}_j|\mathbf{y}_k} \boldsymbol{\Sigma}_{\mathbf{u}_j|\mathbf{y}_k}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{u}_j|\mathbf{y}_k}^{T} \tag{32}$$

$$= \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})\mathbf{U}^T \times (\boldsymbol{\Lambda} + \boldsymbol{\Sigma}_{\boldsymbol{\nu}_j})^{-1} \mathbf{U} (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}) \tag{33}$$

$$= \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}) \times [\mathbf{I} + (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \mathbf{D}_j)^{-1} (\mathbf{D}_j - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k})]^{-1} \tag{34}$$

$$= \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}) \times [\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}]^{-1} (\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \mathbf{D}_j) = \mathbf{D}_j$$

where (30) follows from (24) and (12), (31) follows from (24) and (32) is result of substituting (25) in (31), (33) follows from (26) and (29), and (34) results from substituting (17) and (18) in (33).

Before concluding this subsection, let us mention our motivation to consider the covariance matrix distortion constraints. Suppose that $p = f(\mathbf{D})$ is a parameter of interest, representing some performance criterion for the network. Assume also that a constraint in form of $p \le p_0$ is set on this parameter. One can then consider the following optimization problem:

$$\min_{\mathbf{D}} R(\mathbf{D}) \quad \text{s.t. } f(\mathbf{D}) \le p_0, \tag{35}$$

where $R(\mathbf{D})$ is the RDF. Noting that from (18) and (16) each target distortion matrix $\mathbf{D}$ leads to specific coding scheme properties (e.g. quantization step-size), the above problem means that for a desired parameter p, one can find the optimal coding scheme in sense that yields $p \le p_0$ with the minimum rate. This would be a direct application of coding for a given purpose e.g. enhancement. As an example, it was shown in a recent work that in the special case where p is the mean-squared error between the direct sound $\mathbf{x}$ and the reconstructed source at the receiver and $f(\mathbf{D}) = \mathrm{tr}(\mathbf{D})$, the optimization problem (35) yields the RDF for the mean- squared error constraint [16]. Study of other parameters is the scope of our future work.

3.3. Non-Gaussian sources

In this subsection, we show that in our distributed vector source coding framework with covariance distortion constraint, if the source and measurements have arbitrary distribution, any distribution requires lower rate compared to the Gaussian distribution for the same target distortion and joint source-measurement covariance. The proof is based on the fact that a Gaussian distribution maximizes the differential entropy, and follows similar ideas as in existing works on worst case source coding cf. [17,18].

To avoid baffling notation, we drop the subscripts used in previous subsections to indicate the nodes, i.e. we denote by $\mathbf{t}$, $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{D}$ the variable to be encoded, the side information, the encoded message, and the target distortion, respectively.

Let us assume that the source $\mathbf{x}$ has an arbitrary distribution with finite variance and full-rank joint covariance matrices. Also, let us define the set $\mathcal{Q}_{\mathbf{D}}$ as the set of all conditional probability density functions $q(\mathbf{u}|\mathbf{t},\mathbf{y})$ for which there exists some decoding function $\phi(\mathbf{u},\mathbf{y})$ that yields a reconstruction of the source satisfying the distortion constraint, i.e.

$$\mathcal{Q}_{\mathbf{D}} = \{q(\mathbf{u}|\mathbf{t},\mathbf{y}) | E[(\mathbf{x} - \phi(\mathbf{u},\mathbf{y}))(\mathbf{x} - \phi(\mathbf{u},\mathbf{y}))^T] \preceq \mathbf{D}\}. \tag{36}$$

The RDF is then defined as

$$R_{\mathbf{x}}(\mathbf{D}) = \min_{\mathcal{Q}_{\mathbf{D}}} I(\mathbf{t};\mathbf{u}|\mathbf{y}). \tag{37}$$

We also denote the corresponding RDF for the Gaussian case in (14) by $R_G(\mathbf{D})$. Let us pick an encoding scheme defined as follows:

$$\mathbf{u}^* = \mathbf{U}\mathbf{C}\mathbf{t} + \boldsymbol{\nu}, \tag{38}$$

where $\mathbf{U}$ and $\mathbf{C}$ are defined in (17) and (9), respectively, and $\boldsymbol{\nu}$ is chosen to be Gaussian noise with covariance matrix given in (18). The decoding function is defined similar to (12) as

$$\phi^*(\mathbf{u}^*,\mathbf{y}) = (\mathbf{G} + \mathbf{C}\mathbf{A})\mathbf{y} + \mathbf{C}\mathbf{B}\mathbf{u}^*. \tag{39}$$

As we showed in Section 3.2.2, using this scheme, the reconstruction error at the decoder satisfies the distortion constraints. This means that $q(\mathbf{u}^*|\mathbf{t},\mathbf{y}) \in \mathcal{Q}_{\mathbf{D}}$. Since the minimization in (37) is over the whole $\mathcal{Q}_{\mathbf{D}}$ including $q(\mathbf{u}^*|\mathbf{t},\mathbf{y})$, we can write

$$\begin{aligned} R_{\mathbf{x}}(\mathbf{D}) &\le I(\mathbf{t};\mathbf{u}^*|\mathbf{y}) \\ &= h(\mathbf{u}^*|\mathbf{y}) - h(\mathbf{u}^*|\mathbf{t},\mathbf{y}) \\ &= h(\mathbf{u}^*|\mathbf{y}) - h(\boldsymbol{\nu}) \\ &= h(\mathbf{U}\mathbf{C}\mathbf{t} + \boldsymbol{\nu}|\mathbf{y}) - \frac{1}{2}\log|2\pi e \boldsymbol{\Sigma}_{\boldsymbol{\nu}}| \\ &\le \frac{1}{2}\log|2\pi e(\mathbf{U}\mathbf{C}\Sigma_{t|y}\mathbf{C}^T\mathbf{U}^T + \boldsymbol{\Sigma}_{\boldsymbol{\nu}})| - \frac{1}{2}\log|2\pi e \boldsymbol{\Sigma}_{\boldsymbol{\nu}}| \qquad (40) \\ &= R_G(\mathbf{D}), \qquad (41) \end{aligned}$$

where (40) follows from the fact that Gaussian distribution maximizes the differential entropy, and (41) follows from (22) and (23).

4. Simulation results

We apply our results on sufficient statistics and distributed source coding to Gaussian and audio signals separately. To make use of the side information in a DSC manner, we use a suboptimal implementation called zero error coding (ZEC) [19,20], which is based on the fact that due to the correlation between the source at the encoder and the side information, having the knowledge of the side information limits the range of probable values for the source, thus requiring lower rate quantization compared to the non-distributed case. See the appendix in [21] for a brief tutorial on ZEC.

One should note that in a Wyner–Ziv DSC scenario, the side information is exploited in two senses. It is used to reduce the rate as explained above, and also for estimation to reduce the distortion. (See for example [22] for a more detailed discussion.) However, since in our set-up the side information is available at an intermediate node, and the reconstruction of the source at such nodes might not be of

interest, for a fair comparison, we only consider the improvement gained in the first sense, i.e. reduction in the rate due to the availability of the side information.

In all simulations, we set $n=2$, and consider a family of covariance matrix distortion constraints defined as follows:

$$\mathbf{D}_j = \mathbf{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k} + \mu(\mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}_k} - \mathbf{\Sigma}_{\mathbf{x}|\mathbf{t}_j,\mathbf{y}_k}), \tag{42}$$

where $0 < \mu \leq 1$, and the rate can vary from infinity to zero by sweeping over the whole range of values for μ. Due to the matrix form of the distortion constraint, it is not possible to plot the rate versus the target distortion $\mathbf{D}_j$. Instead, we plot the rate as a function of $(1/n)\mathrm{tr}(\mathbf{D}_j)$.

To apply the coding scheme designed in Section 3.2.2, we first calculate the matrices $\mathbf{C}$ and $\mathbf{U}$ in (16). For each target distortion $\mathbf{D}_j$, we calculate the covariance matrix $\mathbf{\Sigma}_{\nu_j}$ of the quantization noise using (18) from which we can calculate the step-size for the scalar quantizer. The quantizer is then applied to produce the message $\mathbf{u}_j$ as in (16). To reconstruct the source at the receiver, we use (12).

4.1. Gaussian data

We consider the model (6) with additive white Gaussian noise. The standard deviation of the noise is assumed to be $\sigma = 0.3$ unless otherwise stated. The original source is also Gaussian with an autoregressive model and the following covariance matrix:

$$\mathbf{\Sigma}_{\mathbf{x}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{43}$$

with $\rho = 0.8$. First, we assume that node j has encoded its measurement $\mathbf{y}_j$ into $\mathbf{u}_j$, and sent it to node k. Node k forwards this message with rate R_1 to the base station and also encodes its own measurement $\mathbf{y}_k$ into $\mathbf{u}_k$, and sends it with rate R_2 to the base station. At the base station, using the two available messages, the source is reconstructed and the covariance matrix of the estimation error (distortion) is estimated. We refer to this scenario as message forwarding. The sum-rate as a function of distortion is plotted in Fig. 5 for the message forwarding scenario.

We compare the message forwarding case with an alternative scenario where the message $\mathbf{u}_j$ is combined with the measurement $\mathbf{y}_k$ at node j into a sufficient statistic $\mathbf{t}_k$, which is then encoded into the message $\mathbf{u}_\mathbf{t}$ and sent to the base station with the rate R. At the base station an estimate of the source is made using $\mathbf{u}_\mathbf{t}$, from which the distortion is estimated. The rate R as a function of the distortion and the theoretical RDF (for the second scenario) are also plotted in Fig. 5. As is seen, using sufficient statistic leads to significant gain in the rate-distortion trade-off.

Next, we consider the case where node k encodes the measurement $\mathbf{y}_k$ into $\mathbf{u}_k$ and sends it to the base station. We refer to this scenario as the single-node case. Then we compare this with another scenario where another node (node j) assists node k by sending the message $\mathbf{u}_j$ to node k with rate R_1. This message is combined with $\mathbf{y}_k$ into a sufficient statistics, which is then encoded and sent to the base station with rate R_2. The sum-rate versus distortion is then compared to the rate-versus distortion in the single-node scenario. We will show that this can lead to significant gains compared to the single-node case, provided that R_1 and R_2 are chosen properly. In particular, some distortion values are achievable using this strategy, which could not be achieved with a finite rate using the single-node case. To illustrate this, we fix R_1 and plot the sum-rate versus distortion, and repeat this for several different values of R_1. The result is shown in Fig. 6. Note that this includes the single-node scenario as a special case with $R_1 = 0$. Since all the points on all the curves in Fig. 6 are achievable rate-distortion pairs, the lower envelope of the curves and the lower convex hull of all the points on the curves are also achievable, from which one can observe the optimal sum-rate as a function of distortion. As is seen in Fig. 6, for example the target distortion value 0.07 is not achievable with a finite rate using the single node scenario,

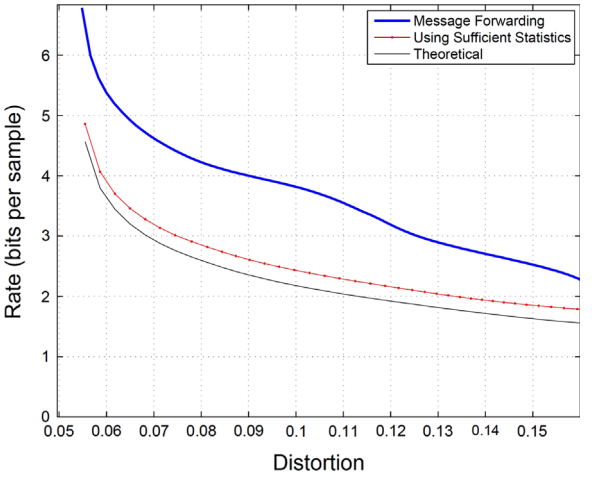

Fig. 5. Rate versus distortion for Gaussian data for message forwarding, encoding scheme with sufficient statistics, and the theoretical function when using the sufficient statistics.

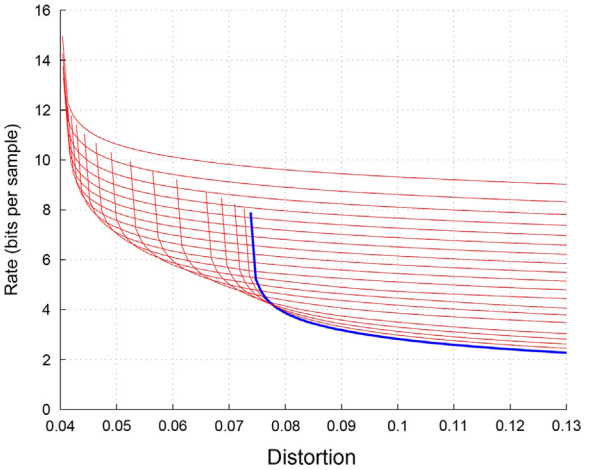

Fig. 6. Sum-rate $R_1 + R_2$ as a function of distortion for different values of R_1. The blue curve is the special case where $R_1 = 0$, which is equivalent to the single-node case. The minimum sum-rate is given by the lower envelope of the curves. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

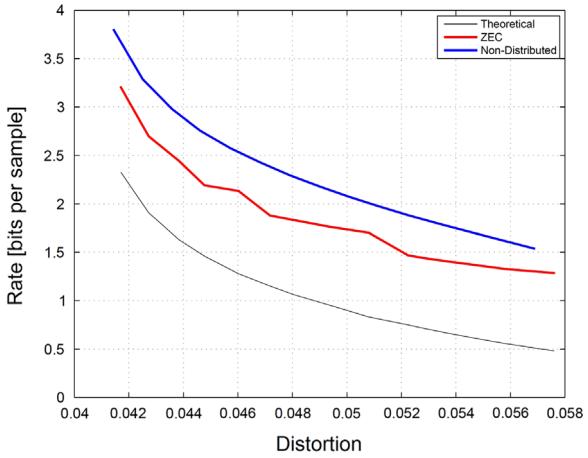

Fig. 7. The RDF for non-distributed source coding and ZEC implementation of the DSC scheme. The theoretical RDF for the distributed case is also plotted for comparison.

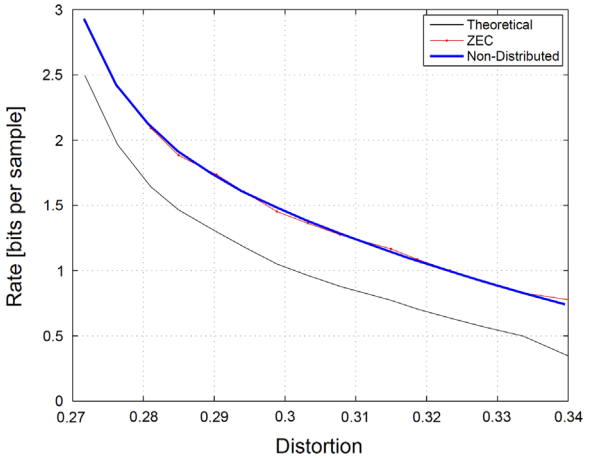

Fig. 8. The RDF for non-distributed source coding, a ZEC implementation of DSC schemes, and theoretical RDF for DSC, when the additive noise standard deviation is $\sigma = 1$.

while using two nodes and sufficient statistics, the required rate is about 5 bits. This is in agreement with the results in [3], where it was shown that a little help from a neighboring node can have a dramatical effect on the resulting sum-rate.

To see the improvement due to the use of side information, we assume that node j encodes its measurement using ZEC while being aware of the presence of side information $\mathbf{y}_k$ at node k. We compare the RDFs for this scenario and the non-distributed case. The result is shown in Fig. 7. The theoretical RDF for the distributed case is also shown in this figure. As is seen, the ZEC implementation leads to a significant improvement, but since it is suboptimal, the resulting rate-distortion performance is far from the theoretical RDF.

To see the effect of correlation on the performance of the DSC scheme, we plot the RDF curves for the case where the standard deviation of the additive noise is $\sigma = 1$. The result is shown in Fig. 8. As is seen, due to less correlation between the sources (because of the strong noise), the theoretical RDF for the DSC case is not far from the non-distributed case. As a result, our practical ZEC implementation leads to no improvement.

4.2. Audio measurements

We consider a setup where two microphones are placed at a fixed location in a standard listening room, which complies with the IEC 268-13 standard. To measure the impulse response for each microphone, we use a logarithmic sweep excitation [23]. The resulting RIR for the first microphone and the magnitude frequency response of the room for the second microphone are plotted in Fig. 9.

We then record several pieces of music of different genres, sampled at 44.1 kHz. We use these measurements for off-line estimation of the covariance matrices as will be explained in the sequel.

Application of the coding schemes to audio signals is more difficult. First, we need to deconvolve the RIRs from the measurements before using the model (6). We use a simple deconvolution in the FFT domain. To prevent the frequency components which do not exist in the RIR but are due to the additive noise, from being unboundedly amplified by the deconvolution operation, we first filter these components. Denoting the FFT operation by $\mathcal{F}$, the deconvolved version $\mathbf{y}_j$ of the measurement $\mathbf{z}_j$ is given by $\mathbf{y}_j = \mathcal{F}^{-1}\{Y_j\}$, where Y_j is defined as

$$Y_j(m) \triangleq \begin{cases} \dfrac{Z_j(m)}{H_j(m)} & \text{if } |H_j(m)| > T \\ 0 & \text{otherwise} \end{cases} \tag{44}$$

where m counts over the set of frequency indices in FFT domain, $H_j \triangleq \mathcal{F}\{\mathbf{h}_j\}$, and $Z_j \triangleq \mathcal{F}\{\mathbf{z}_j\}$, and T is a threshold value chosen experimentally.

The second problem with the audio data is unavailability of the statistics. The cross-covariance matrices between the measurements at different nodes cannot be estimated from the available data at the nodes while encoding or decoding, since each node only has access to its own measurement. One possibility is to use an off-line estimation; i.e. to estimate the covariances using a set of sound sources and their measurements made at the nodes. We consider some music clips of different genres to estimate $\mathbf{\Sigma}_{\mathbf{x}}$, $\mathbf{\Sigma}_{\mathbf{y}_j}$, $\mathbf{\Sigma}_{\mathbf{y}_k}$, $\mathbf{\Sigma}_{\mathbf{x}\mathbf{y}_j}$, $\mathbf{\Sigma}_{\mathbf{x}\mathbf{y}_k}$, and the noise covariance matrices. The remaining conditional and nonconditional covariance and cross-covariance matrices can be derived from these matrices.

Finally, one should note that audio signals are highly nonstationary in general. Although using the above method one can obtain an estimate of the average covariance matrices, these estimates are not accurate, and the accuracy varies over time and across the different music clips.

As the test data, we choose a music clip from the set of clips used for the estimation of statistics and apply the coding schemes. The message forwarding scenario for audio data is compared to the sufficient statistics in Fig. 10, which shows that sufficient statistics yield a significant improvement. Note that although we showed that a Gaussian distribution is worst to code, here the

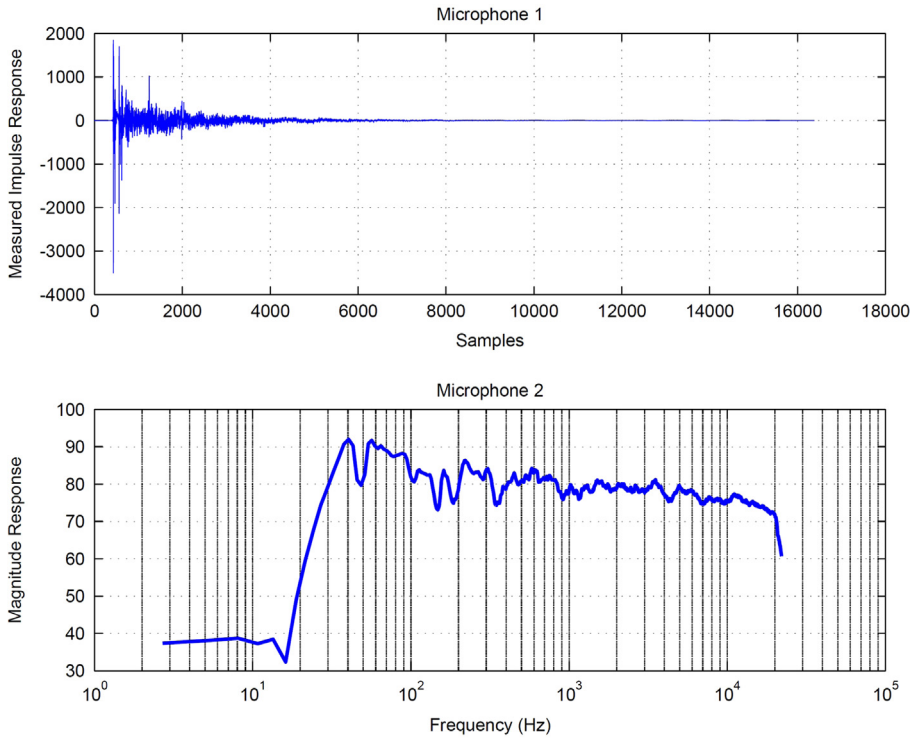

Fig. 9. Measured RIR for the two microphones. The first plot illustrates the measured impulse response at the location of the first microphone. The second plot is the magnitude of the RIR in the frequency domain at the second microphone.

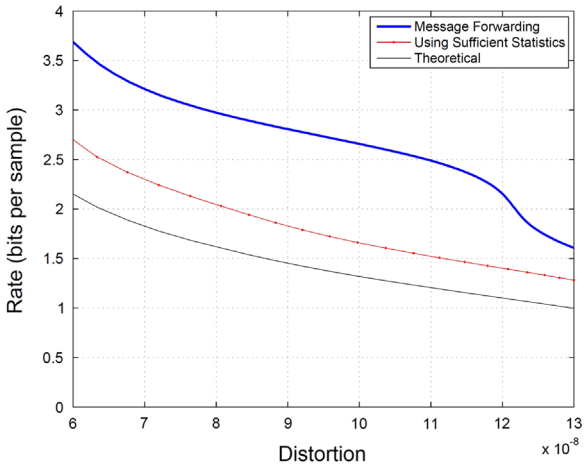

Fig. 10. Rate versus distortion for audio data for message forwarding, encoding scheme with sufficient statistics, and the theoretical function for the Gaussian case when using the sufficient statistics.

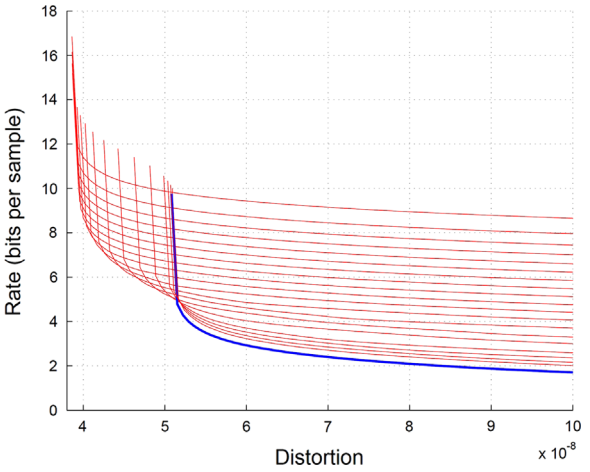

Fig. 11. Sum-rate R_1+R_2 as a function of distortion for audio for different values of R_1. The blue curve shows the rate-distortion performance for the single-node case. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Gaussian RDF is below the rate-distortion performance for the audio signal. This is a result of assuming fixed statistics for the audio signal which is indeed nonstationary. In fact, we observed that depending on the music clip, the rate-distortion curve for audio might fall below or above the theoretical Gaussian RDF.

To compare with the single-node scenario, the sum-rate R_1+R_2 versus distortion curves are plotted in Fig. 11 for different values of R_1. As is seen, again we achieve the similar improvement for the audio measurements as the Gaussian data.

We compare the distributed and non-distributed scenarios for audio data in Fig. 12. The results indicate a slight improvement due to the use of our ZEC implementation.

Finally, to see the reconstructed audio using the data from two nodes, we assume that one node sends its data with the rate 2.5 bits/sample to another node (or the base

station) where another measurement is available as side information. The result is shown in Fig. 13. Note that since the reconstruction is for the direct sound source **x**, it is not perfect, as dereverberation is typically achieved in price of distorting the signal.

4.2.1. Practical considerations

In practice, there are problems like calibration of the microphones and synchronization of the nodes. We have assumed that the sensor nodes are in perfect synchronization and the microphones are calibrated. Here we discuss two other practicalities, namely the dimension of the vector space n, and the block length for the deconvolution at the nodes.

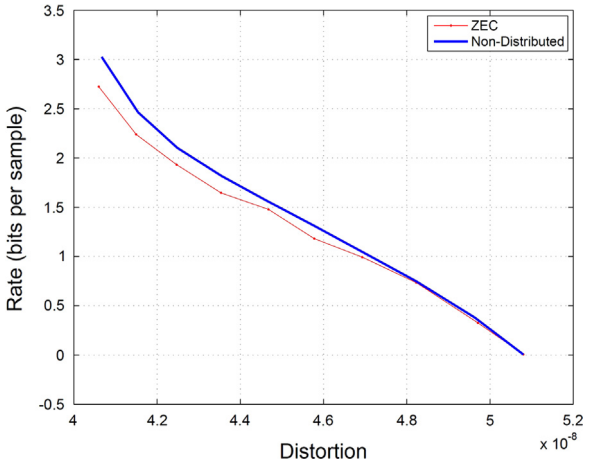

Fig. 12. The rate-distortion performance for non-distributed source coding and a ZEC implementation of the DSC scheme.

The rate-distortion plot for the audio clip is shown in Fig. 14 for three different values of n. Larger n means better use of the time correlation, and thus lower rate for the same distortion. This is achieved in price of higher complexity of the encoder and decoder. Dimensions of much higher than $n=3$ (which is the maximum value we have chosen for n in Fig. 14) are computationally manageable to implement. However, here for the simulation purposes, we have an extra problem with higher dimensions. The error covariance matrix (distortion) is estimated by averaging over individual vector samples. For a data-set with a fixed number of samples, the number of vector samples decreases by increasing n. This leads to an insufficient number of vector samples and thus invalid estimations of the distortion. Note also that for nonstationary audio signals, this problem is more serious, as there is no fixed distortion and one needs much longer data records to guarantee the convergence to the average covariance matrix.

Next, we consider the choice of the block length for the deconvolution of the room impulse response from the audio measurements. Since this block length determines the buffer size, shorter blocks are desirable. However, too short a block might lead to inefficient deconvolution. We do an experiment regarding the choice of the block length. Since the deconvolved measurement is an approximation of the direct source, as a simple criterion for efficiency of the deconvolution, we use the correlation coefficient between the source and the deconvolved signal. We start from an audio clip of length $L=2.5$ million samples, and at each step, split each subsequence from the previous step into

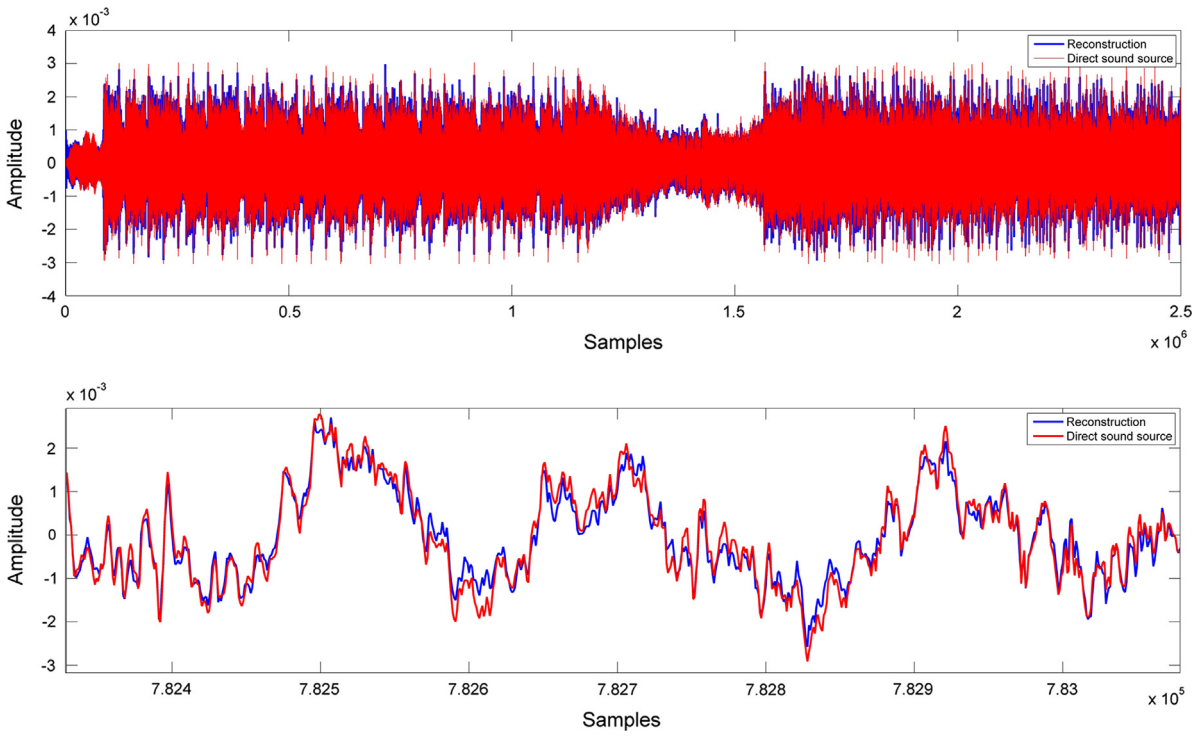

Fig. 13. Direct sound source and the reconstruction, (a) the whole audio clip, (b) zooming in on a part of the clip.

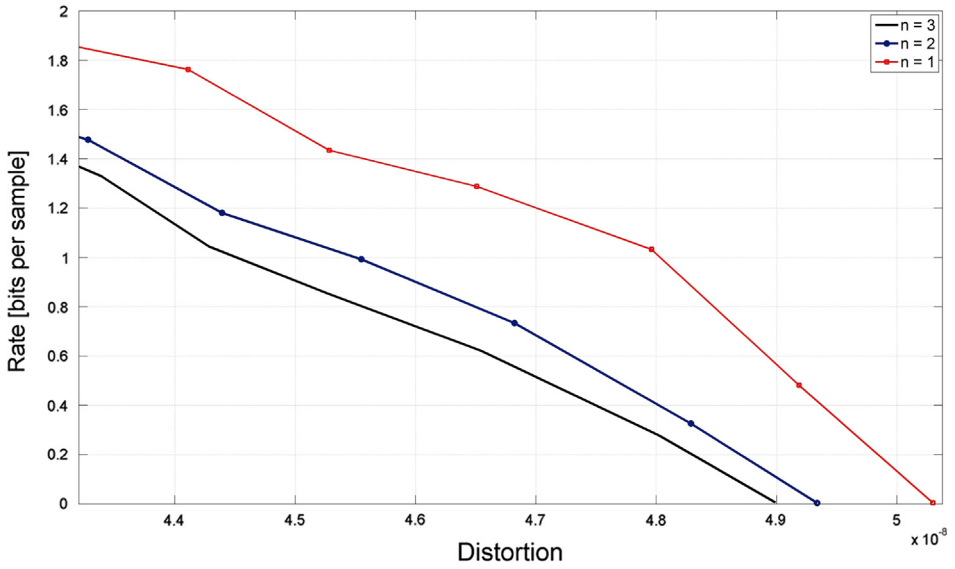

Fig. 14. The rate-distortion performance for audio signals with three different vector space dimensions.

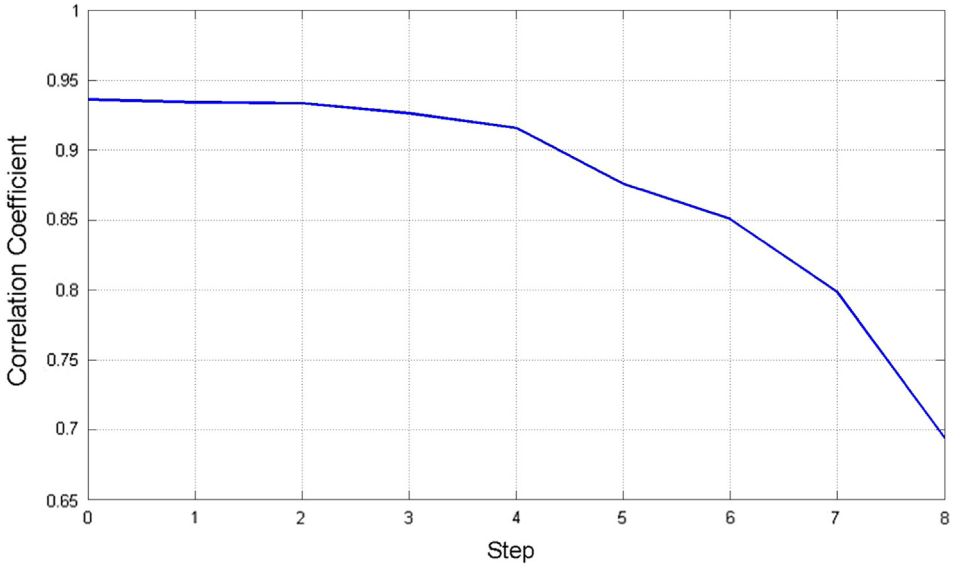

Fig. 15. Correlation coefficient between the deconvolved signal and the sound source versus the counter i. Length of the subsequences at step i is $L/2^i$.

two shorter subsequences of the same length. In other words, at step i, $0 \le i \le 8$, we have 2^i subsequences of length $L/2^i$. We deconvolve the impulse response from each of the subsequences, calculate the correlation coefficient between the deconvolved subsequence and the corresponding subsequence in the direct sound source, and then average over the resulting correlation coefficients. The plot of the correlation coefficient versus i is shown in Fig. 15. A significant degradation in the correlation coefficient is seen after $i=4$, suggesting that the critical block length is between $i=4$ and $i=5$, corresponding to the blocks of length 156 000 and 78 000 samples, respectively. Noting that the effective length of the impulse response is roughly 10 000 samples as seen in Fig. 9, this suggests that as a rule of thumb, the block length should be at least 10 times the length of the impulse response.

5. Conclusions

We considered a source coding problem for wireless acoustic sensor networks with the possibility of communication between the nodes. We used sufficient statistics to losslessly combine the messages and the measurement available in a given node into a single source, which was then encoded and sent to the next node. We also proposed to make use of the measurement available in receiving nodes as side information to reduce the rate. For the resulting distributed source coding problem with the

covariance matrix distortion constraint we derived the rate-distortion function for the case of Gaussian sources. We also showed that other sources are better to code compared to the Gaussian case, in sense that their rate-distortion function is upper-bounded by the one for a Gaussian source. We applied our results to both Gaussian and audio data, and obtained significant gains in the rate-distortion trade-off. Future work includes more efficient implementations of the distributed source coding part instead of using zero-error coding, more advanced techniques for the deconvolution of the audio measurements from the RIRs, and using adaptive estimation of the joint statistics of the sources to model the nonstationarity of audio signals.

Acknowledgments

The authors would like to thank Morten Lydolf for his help in making the audio measurements at Bang & Olufsen, and Jesper Kjær Nielsen for insightful discussions related to our acoustical channel model.

References

[1] A. Bertrand, Applications and trends in wireless acoustic sensor networks: a signal processing perspective, in: Proceedings of IEEE Symposium on Communications and Vehicular Technology (SCVT), Ghent, Belgium, 2011.
[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Comput. Netw. 38 (2002) 393–422.
[3] J. Østergaard, M.S. Derpich, Sequential remote source coding in wireless acoustic sensor networks, in: European Signal Processing Conference (Eusipco), (August) 2012.
[4] P.L. Dragotti, M. Gastpar, Distributed Source Coding, Theory, Algorithms and Applications, 1st ed. Elsevier, USA, 2009.
[5] A. Zahedi, J. Østergaard, S.H. Jensen, P. Naylor, S. Bech, Distributed remote vector Gaussian source coding for wireless acoustic sensor networks, in: IEEE Data Compression Conference, Utah, USA, 2014, arxiv:1401.3945[cs.IT].
[6] D. Slepian, J. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inf. Theory 19 (4) (1973) 4480–4711.
[7] A. Wyner, J. Ziv, The rate-distortion function for source coding with side information at the decoder, IEEE Trans. Inf. Theory 22 (1) (1976) 1–10.
[8] A. Wyner, The rate-distortion function for source coding with side information at the decoder—II: general sources, Inf. Control 38 (1) (1978) 60–80.
[9] R. Gray, Source Coding Theory, Springer, USA, 1990.
[10] J. Thomas, T. Cover, Elements of Information Theory, Wiley, USA, 1991.
[11] J. Idier, Bayesian Approach to Inverse Problems, Wiley, UK, 2008.
[12] K. Eswaran, M. Gastpar, Rate loss in the CEO problem, in: 39th Annual Conference on Information Sciences and Systems, Baltimore, MD, 2005.
[13] S.C. Draper, G.W. Wornell, Side information aware coding strategies for estimation under communication constraints, IEEE J. Sel. Areas Commun. 22 (6) (2004) 1–11.
[14] C. Tian, J. Chen, Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints, IEEE Trans. Inf. Theory 55 (10) (2009) 4676–4680.
[15] R. Zamir, M. Feder, On lattice quantization noise, IEEE Trans. Inf. Theory 42 (4) (1996) 1152–1159.
[16] A. Zahedi, J. Østergaard, S.H. Jensen, P. Naylor, S. Bech, Distributed remote vector Gaussian source coding with covariance distortion constraints, in: Proceedings of IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, July 2014.
[17] T. Berger, Rate-distortion Theory, Wiley Online Library, USA, 2003.
[18] R. Gallager, Information Theory and Reliable Communication, Wiley, New York, 1968.
[19] P. Koulgi, E. Tuncel, S.L. Regunathan, K. Rose, On zero-error coding of correlated sources, IEEE Trans. Inf. Theory 49 (11) (2003) 2856–2873.
[20] N. Alon, A. Orlitsky, Source coding and graph entropies, IEEE Trans. Inf. Theory 42 (9) (1996) 1329–1339.
[21] D. Quevedo, A. Ahlen, Jan Østergaard, Energy efficient state estimation with wireless sensors through the use of predictive power control and coding, IEEE Trans. Signal Process. 58 (9) (2010) 4811–4823.
[22] Y. Steinberg, Coding and common reconstruction, IEEE Trans. Inf. Theory 55 (11) (2009) 4995–5010.
[23] A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, in: 108th AES Convention, Paris, France, 2000.