Distributed Estimation and Coding: A Sequential Framework Based on a Side-Informed Decomposition

Chao Yu, *Student Member, IEEE*, and Gaurav Sharma, *Senior Member, IEEE*

Abstract—We propose a sequential framework for the distributed multiple-sensor estimation and coding problem that decomposes the problem into a series of side-informed source coding problems and enables construction of good codecs. Our construction relies on a separation result for the simplified scenario where information from one sensor is to be sent to the CP that already has information regarding the desired signal. We show that the optimal encoder decoder combination, in this setting, can be decomposed, without loss in performance, into two steps: A first preprocessing step to extract *relevant information* from the indirect observation with consideration of the side information, followed by a second step of side-informed encoding of the preprocessed output. A recursive exploit of the decomposition coupled with side-informed transform coding allows us to construct encoders by reusing scalar Wyner–Ziv codecs. We develop a numerical procedure for obtaining bounds delineating the best achievable performance for the proposed sequential framework and construct and demonstrate a practical codec in the proposed framework that achieves empirical performance close to the bound. Furthermore, we also compare the bounds for the sequential scheme against bounds for a number of alternate schemes—for some of which codec constructions are obvious and others for which codec constructions are inobvious. Our results show that, in most cases, our codec exceeds the performance bound of schemes offering obvious constructions. The achievable bound for our sequential framework is also shown to be close to a general nonconstructive distributed bound that does not impose the sequential constraint indicating that the sequential approach may not cause a significant performance compromise.

Index Terms—Distributed estimation, Gaussian CEO, side-informed coding, side-informed estimation, Wyner-Ziv coding.

I. INTRODUCTION

THE increasing popularity of wireless sensor networks made up of nodes of devices with sensing, communication, and computation capabilities has recently motivated
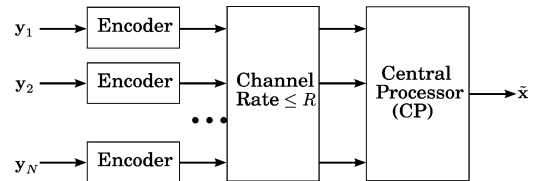
C. Yu is with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627-0126 USA (e-mail: chyu@ece.rochester.edu; devincyu@gmail.com; website: http://www.ece.rochester.edu/~chyu/).

G. Sharma is with the Electrical and Computer Engineering Department, the Department of Biostatistics and Computational Biology, and the Department of Oncology, University of Rochester, Rochester, NY 14627-0126 USA (e-mail: gaurav.sharma@rochester.edu; g.sharma@ieee.org; website: http://www.ece.rochester.edu/~gsharma/).

Fig. 1. N sensors collect observations $\mathbf{y}_i$, $i = 1, 2, \ldots, N$ regarding a desired signal $\mathbf{x}$. Each observation $\mathbf{y}_i$ is encoded locally and sent to the CP, where an estimate $\tilde{\mathbf{x}}$ is achieved.

intense research on distributed signal processing techniques [2]. In some sensor network applications, it is desirable to directly collect the observations recorded by the sensor nodes at a central processor (CP) for the purposes for archival and further processing. On the other hand, in other application scenarios, the CP is not interested in the raw sensor observations but only an aggregated estimate obtained from them. An example of this type is a visual sensor network where a CP may wish to create a desired view of a scene based on a number of images captured by individual cameras (none of which may be exactly at the desired viewpoint). In this application, the individual camera views and the desired view represent lower dimensional projections of the underlying $3D$ scene. Motivated by this scenario, we consider a setup as illustrated in Fig. 1, where a group of geographically dispersed nodes (sensors) are deployed to collect observations pertaining to a desired signal, which are indirect and noisy. Each sensor communicates with the CP via a shared rate-constrained channel. Individual sensor observations are processed and encoded locally before being communicated to the CP in order to meet the channel rate constraint. The CP utilizes the received data to compute an estimate of the desired signal. We consider the design of encoders at these sensors and the decoder at the CP with a view to minimizing the mean-square error (MSE) for the estimate.

Our investigation is motivated by the fact that direct encoding of local observations using existing source coding techniques yields suboptimal performance. The reasons for the failure of direct encoding are twofold. First, the observations are indirect and contain irrelevant data and noise. Second, observations collected by different sensors are often correlated in a dense network, resulting in redundancy among these sensor observations. In this paper, we develop a decomposition and an encoder implementation that addresses these two sources of inefficiency. First, under Gaussian statistics and the MSE distortion metric, we establish that *without incurring a rate-distortion penalty*, the optimal encoder can be decomposed into two stages comprising of a first preprocessing step that extracts *relevant information* from local observations, referred to as a "side-informed estimation", followed by side-informed coding of the preprocessed data. The

processing at the CP is also correspondingly decomposed into two steps: a first side-informed decoding step followed by estimation of the desired signal using the side information and the received data. Next, we consider the general scenario involving multiple sensors, and develop a sequential approach that recursively exploits our decomposition, each sensor preprocesses and encodes its observation for transmission considering as side-information the information already available at the CP from the sensors preceding it, thereby reducing the distributed estimation and coding problem into a sequence of preprocessing and side-informed coding problems.

The decomposition paves the way for the development of constructive encoders for transmission of indirect observations with side information, by leveraging source coding techniques developed for scenarios where the desired signal is available at the encoder in its pristine form. To demonstrate the utility of this approach, we develop a side-informed codec in a transform coding framework [3], using the conditional Karhunen–Loéve transform (KLT) [4], followed by scalar side-informed codes [5], [6], i.e., Wyner–Ziv codes. We also compute an achievable rate-distortion (R-D) bound for our proposed sequential approach and demonstrate, via simulations, that our codec construction offers performance close to this R-D bound. The codec's R-D performance in simulations exceeds bounds delineating the best R-D performance achievable with a number of alternate side-informed and non-side-informed codec constructions, thereby validating the proposed methodology. Compared to a more general distributed achievable R-D bound [7] that does not impose the sequential constraint, only a small compromise in performance is seen for the sequential bound, further indicating that the overall approach is effective and that the sequential approach may not incur a significant R-D penalty.

Related work exists in several contexts. For the non-side-informed scenario, a separation result analogous to ours was established by Sakrison [8]. This part of our work can be considered a generalization of the result of Sakrison. Although derived for the case of rate-constrained channels, our decomposition is readily applicable for reduced-dimensionality channels considered in [9]–[12]. The optimal linear preprocessing in these cases can be obtained from the decomposition and the conditional KLT [4]. Our work extends this body of research in distributed estimation to address more realistic rate-constrained channels. For rate constrained channels, an achievable rate distortion bound for the distributed estimation and coding problem is obtained in [7]. The analysis therein (particularly in Section IV of [7]), arrives at conclusions analogous to, though not identical to, our proposed decomposition. Our analysis methodology and outcome also differ from [6] because of the difference in our motivation. Particularly, the decomposition at the decoder side required for our construction is unnecessary for computing R-D bounds and therefore missing from [7].

Also related is the problem of distributed estimation using quantized observations under a rate constraint for estimating unknown deterministic parameters [13], [14] and for random signals [15]–[17]. For these problems, it has been observed that the strong statistical dependency among sensor observations, once taken into account, can significantly reduce the data rate required to communicate the local observation. For this pur-

pose, either a feedback channel from the CP to sensors is used [18] or inter-sensor communications are utilized [15], [16], so that the sensor can calculate the *innovation* in its local observation, and only the quantized innovation is sent to the CP, where the quantizer enforces the rate constraint. Nested quantizers are used as a simple form of side-informed coding technique to exploit the side information at the CP [17] without feedback of the side information. Under the assumption of high rate quantization, a separative structure is also suggested in [19] for the asymptotically optimal quantization of noisy sources with side information.

The general distributed estimation problem, without the sequential constraints that we consider in our framework, is a vector version of the Gaussian CEO problem [20]. A complete characterization of the rate-distortion region for this general scenario remains an open problem though several papers have addressed this in the past for the Gaussian CEO problem [21], [22] . Alternate related work [23] establishes that, a successive-refinement scheme can achieve any point in the achievable rate-distortion region of the quadratic Gaussian CEO problem. The proof however, is nonconstructive, relying on random binning, and does not lead to practical coding schemes.

The rest of the paper is organized as follows. Section II presents the decomposed structure for the optimal encoder of a local observation when the CP has side information. The sequential approach for the multisensor setup is considered in Section III. We next briefly present in Section IV our Wyner–Ziv codec design. Simulation results are presented in Section V. Sections VI and VII conclude the paper with a discussion and a summary of the conclusions.

II. OPTIMAL SIDE-INFORMED ESTIMATION UNDER A RATE CONSTRAINT: A DECOMPOSITION

Fig. 1 illustrates the general setup, where a set of N sensors are deployed to estimate the real-valued random vector $\mathbf{X} \in \mathbb{R}^L$. The $i^{\text{th}}(i = 1, 2, \ldots, N)$ sensor collects a p_i dimensional observation $\mathbf{y}_i \in \mathbb{R}^{p_i}$ pertaining to this desired signal $\mathbf{X}$. Each observation $\mathbf{y}_i$ is separately encoded and transmitted to the CP, over a channel that is constrained to a total rate R (over all sensors) but is otherwise error-free. The CP utilizes the messages received from the sensors to obtain an estimate of $\mathbf{X}$. We focus on the scenario where the desired signal $\mathbf{X}$ and measurements $\{\mathbf{Y}\}_{i=1}^N$ are jointly Gaussian, and assume the joint statistics of these vector random variables are available at the CP and at each of the sensors. For notational simplicity, we assume all signals are zero mean, and denote by $\boldsymbol{\Sigma}_{\mathbf{ab}} = \mathsf{E}[\mathbf{AB}^T]$ the cross-covariance of random vectors $\mathbf{A}$, $\mathbf{B}$ and by $\boldsymbol{\Sigma}_{\mathbf{a}} = \mathsf{E}[\mathbf{AA}^T]$ the autocovariance of vector $\mathbf{A}$.

In this section, we consider a simplified setup as illustrated in Fig. 2(a), where an observation $\mathbf{y}_1$ needs to be encoded and transmitted to the CP that already has information $\mathbf{y}_0$ available from other sensor nodes and possibly from direct local observations. The sensor encodes $\mathbf{y}_1$ as a message $m \in \mathcal{M}$ where $\mathcal{M} = \{1, 2, \ldots, 2^R\}$ denotes the output space of the encoder—defined so as to meet the channel rate constraint. The CP utilizes the local information $\mathbf{y}_0$ and the message m that it receives from the sensor in a "decoder" to obtain an estimate
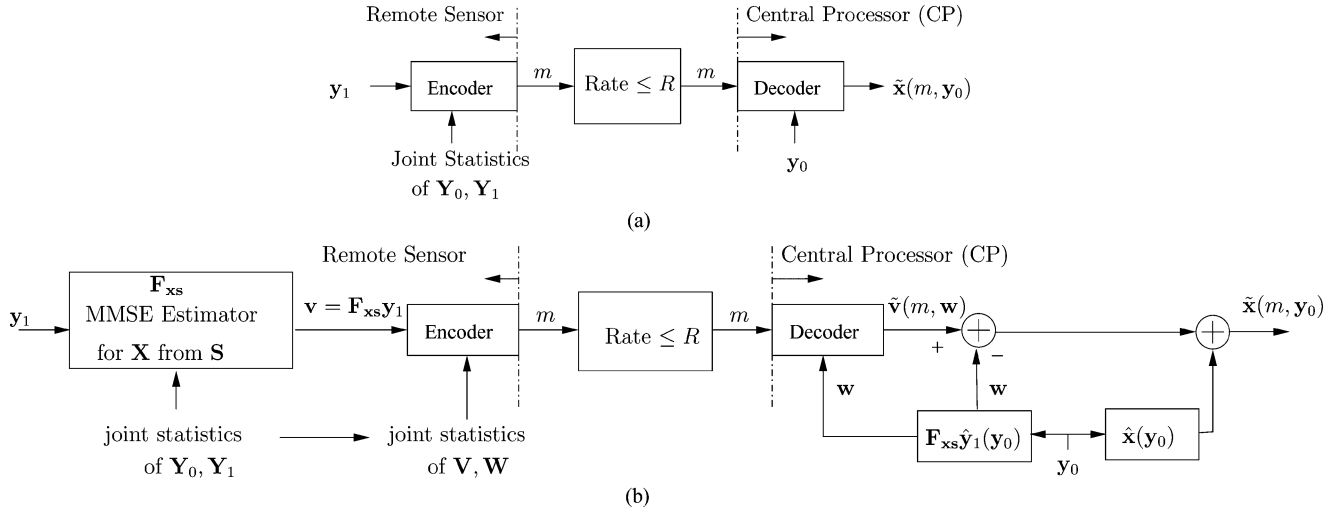
Fig. 2. Transmission of an indirect observation with side information. (a) Single-stage encoder and decoder. (b) Encoder and decoder are decomposed into side-informed estimation and side-informed coding stages.

$\tilde{\mathbf{x}}(m, \mathbf{y}_0)$ of the signal $\mathbf{X}$. The resulting MSE distortion for the decoder is $D = \mathsf{E}[\|\mathbf{X} - \tilde{\mathbf{x}}(M, \mathbf{Y}_0)\|^2]$, where $\mathsf{E}[\cdot]$ represents the expectation operator, and upper-case symbols represent random variables (r.v.s) for which the corresponding lower-case symbols are sample realizations, both being bold when these correspond to vectors. The MSE depends on the encoder mapping $\mathbf{y}_1 \rightarrow m \stackrel{\text{def}}{=} m(\mathbf{y}_1)$, and the decoder mapping defined by $(m, \mathbf{y}_0) \rightarrow \tilde{\mathbf{x}} \stackrel{\text{def}}{=} \tilde{\mathbf{x}}(m, \mathbf{y}_0)$. Our overall objective is to construct encoder and decoder mappings for an indirect observation $\mathbf{y}_1$ using $\mathbf{y}_0$ as side information, i.e., $m^*(\mathbf{y}_1)$ and $\tilde{\mathbf{x}}^*(m, \mathbf{y}_0)$, that minimize the MSE distortion D.

In this section, we establish an equivalence that allows us to decompose the above problem of encoding and decoding from indirect observations into a preprocessing and postprocessing step combined with the problem of encoding and decoding direct observations. The latter problem has been extensively addressed in literature and good codes have been developed for the scalar observation scenario [5], [6], [24]. Note also that our objective is to develop a constructive algorithm by exploiting the equivalence we obtain, and not the determination of approximate or exact rate-distortion bounds.

We begin by considering the distortion D. Denoting by $\hat{\mathbf{x}}(\mathbf{y}_0, \mathbf{y}_1)$ the MMSE estimate of $\mathbf{X}$ obtained from $(\mathbf{y}_0, \mathbf{y}_1)$, where $\hat{\mathbf{x}}(\mathbf{y}_0, \mathbf{y}_1) = \mathsf{E}[\mathbf{X} \mid \mathbf{y}_0, \mathbf{y}_1]$ [25], we obtain (1), shown at the bottom of the page, where in the final step we have used the orthogonality property of the MMSE estimators that ensures the prediction error $(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1))$ is orthogonal to any function of the predictor inputs $(\mathbf{Y}_0, \mathbf{Y}_1)$, hence $\mathsf{E}\left[(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1))^T \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1)\right] = 0$

and $\mathsf{E}\left[(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1))^T \tilde{\mathbf{x}}(m(\mathbf{Y}_1), \mathbf{Y}_0)\right] = 0$. Observing that the first term in (1) is a constant independent of the encoder and decoder mappings, we see that the optimal mappings can be designed to only minimize the second term, viz. $D_1 = \mathsf{E}\left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1) - \tilde{\mathbf{x}}(m, \mathbf{Y}_0)\|^2\right]$, without any loss of rate-distortion performance. Now, rewriting D_1 as

$$D_1 = \int \mathsf{E}\left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) - \tilde{\mathbf{x}}(m, \mathbf{Y}_0)\|^2 \mid \mathbf{y}_1\right] f_{\mathbf{Y}_1}(\mathbf{y}_1) d\mathbf{y}_1 \quad (2)$$

where $f_{\mathbf{Y}_1}(\cdot)$ denotes the probability density function (p.d.f) of r.v. $\mathbf{Y}_1$, we see that for a fixed decoder mapping $\tilde{\mathbf{x}}(\cdot)$, the optimal encoding for the observation $\mathbf{y}_1$ is

$$m^*(\mathbf{y}_1) = \arg\min_{m \in \mathcal{M}} \mathsf{E}\left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) - \tilde{\mathbf{x}}(m, \mathbf{Y}_0)\|^2 \mid \mathbf{y}_1\right]. \quad (3)$$

We assume without loss of generality that this encoder is utilized at the sensor in the system of Fig. 2(a) and show next that we can construct a corresponding partitioned encoder–decoder pair in the structure of Fig. 2(b), where the latter has an expected distortion that is no larger than D.

We define $\mathbf{S} \stackrel{\text{def}}{=} \mathbf{y}_1 - \hat{\mathbf{y}}_1(\mathbf{Y}_0)$ as the *innovation* in $\mathbf{y}_1$ given $\mathbf{Y}_0$, where $\hat{\mathbf{y}}_1(\mathbf{Y}_0) = \mathbf{\Sigma}_{\mathbf{y}_1 \mathbf{y}_0} \mathbf{\Sigma}_{\mathbf{y}_0}^{-1} \mathbf{Y}_0$, and obtain

$$\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) = \hat{\mathbf{x}}(\mathbf{Y}_0) + \hat{\mathbf{x}}(\mathbf{S}). \quad (4)$$

Equation (4) relies on the additive property of the MMSE estimates for zero-mean jointly Gaussian signals [25, Sec. 11.4]: the MMSE estimate of a signal from two uncorrelated observations can be obtained as the sum of the two MMSE estimates obtained from each of these two observations, respectively. Equation (4)

$$D = \mathsf{E}\left[\|\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1) + \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1) - \tilde{\mathbf{x}}(m(\mathbf{Y}_1), \mathbf{Y}_0)\|^2\right]$$
$$= \mathsf{E}\left[\|\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1)\|^2\right] + \mathsf{E}\left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{Y}_1) - \tilde{\mathbf{x}}(m(\mathbf{Y}_1), \mathbf{Y}_0)\|^2\right] \quad (1)$$

can also be interpreted as a two-stage sequential MMSE estimation: an initial estimate $\hat{\mathbf{x}}(\mathbf{Y}_0)$ is first obtained from $\mathbf{Y}_0$, which is further refined by the innovation $\mathbf{S}$ contained in $\mathbf{y}_1$.

Under the assumption of joint Gaussian statistics, the MMSE estimator is a linear operator. We denote by $\mathbf{F}_{\mathbf{xs}} = \Sigma_{\mathbf{xs}}\Sigma_{\mathbf{s}}^{-1}$ the MMSE estimator of $\mathbf{X}$ from $\mathbf{S}$, where $\Sigma_{\mathbf{xs}} = \Sigma_{\mathbf{xy}_1} - \Sigma_{\mathbf{xy}_0}\Sigma_{\mathbf{y}_0}^{-1}\Sigma_{\mathbf{y}_0\mathbf{y}_1}$ and $\Sigma_{\mathbf{s}} = \Sigma_{\mathbf{y}_1} - \Sigma_{\mathbf{y}_1\mathbf{y}_0}\Sigma_{\mathbf{y}_0}^{-1}\Sigma_{\mathbf{y}_0\mathbf{y}_1}$. $\mathbf{F}_{\mathbf{xs}}$ can also be viewed as a "side-informed" estimator of $\mathbf{X}$ from $\mathbf{y}_1$ given $\mathbf{Y}_0$. We further rewrite $\hat{\mathbf{x}}(\mathbf{S}) = \mathbf{F}_{\mathbf{xs}}\mathbf{y}_1 - \mathbf{F}_{\mathbf{xs}}\hat{\mathbf{y}}_1(\mathbf{Y}_0)$. Next, we rewrite the objective function in (3) as

$$
\mathsf{E}\left[\left\|\hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{y}_1) - \tilde{\mathbf{x}}(m,\mathbf{Y}_0)\right\|^2 \mid \mathbf{y}_1\right]
$$
$$
= \mathsf{E}\left[\left\|\hat{\mathbf{x}}(\mathbf{Y}_0) + \mathbf{F}_{\mathbf{xs}}\mathbf{y}_1 - \mathbf{F}_{\mathbf{xs}}\hat{\mathbf{y}}_1(\mathbf{Y}_0) - \tilde{\mathbf{x}}(m,\mathbf{Y}_0)\right\|^2 \mid \mathbf{y}_1\right]
$$
$$
\overset{\text{def}}{=} \mathsf{E}\left[\left\|\mathbf{v} - \tilde{\mathbf{v}}_-(m,\mathbf{Y}_0)\right\|^2 \mid \mathbf{y}_1\right] \tag{5}
$$

where $\mathbf{v} \overset{\text{def}}{=} \mathbf{F}_{\mathbf{xs}}\mathbf{y}_1$, $\hat{\mathbf{v}}(\mathbf{y}_0) = \mathbf{F}_{\mathbf{xs}}\hat{\mathbf{y}}_1(\mathbf{y}_0)$ denotes the MMSE estimate of $\mathbf{V}$ from $\mathbf{y}_0$, and

$$
\tilde{\mathbf{v}}_-(m,\mathbf{y}_0) \overset{\text{def}}{=} \tilde{\mathbf{x}}(m,\mathbf{y}_0) - \hat{\mathbf{x}}(\mathbf{y}_0) + \hat{\mathbf{v}}(\mathbf{y}_0). \tag{6}
$$

The decoding rule $\tilde{\mathbf{v}}_-(\cdot)$ can be viewed as a side-informed decoder for $\mathbf{v}$, using $\mathbf{Y}_0$ as side information, induced by the corresponding side-informed decoding rule of $\hat{\mathbf{x}}(\cdot)$. Conversely, any decoding rule $\tilde{\mathbf{v}}_-(m,\mathbf{y}_0)$ for the (side-informed) decoding of $\mathbf{V}$ induces the corresponding decoding rule $\tilde{\mathbf{x}}(m,\mathbf{y}_0) \overset{\text{def}}{=} \tilde{\mathbf{v}}_-(m,\mathbf{y}_0) + \hat{\mathbf{x}}(\mathbf{y}_0) - \hat{\mathbf{v}}(\mathbf{y}_0)$ for the decoding of $\mathbf{X}$ which also satisfies (5). Since (5) establishes that the error criteria for the optimal encoding rules for these two scenarios, viz. encoding of $\mathbf{x}$ with $\mathbf{y}_0$ as side information and encoding of $\mathbf{v}$ with $\mathbf{y}_0$ as side information, are identical, we see that the optimum encoding rule for $\mathbf{v}$ also defines an optimal encoding rule for $\mathbf{y}_1$, both with $\mathbf{y}_0$ as side information. Equation (6) also defines a decomposed structure for realizing the encoder–decoder pair where $\mathbf{v}$ is obtained at the remote sensor and encoded, and correspondingly at the decoder, first the decoder for $\mathbf{v}$ recovers $\tilde{\mathbf{v}}_-(m,\mathbf{y}_0)$, which is then transformed into a decoded value for $\mathbf{x}$ using (6).

Therefore, for the problem of Fig. 2(a), we can equivalently consider the problem of side-informed encoding of $\mathbf{v}$ with $\mathbf{Y}_0$ as side information. Now $\mathbf{W} \overset{\text{def}}{=} \hat{\mathbf{v}}(\mathbf{Y}_0)$ is trivially a sufficient statistic of $\mathbf{Y}_0$ for estimating $\mathbf{v}$, thus given the encoding rule, the optimal decoder for $\mathbf{v}$ utilizing $\mathbf{w}$, has exactly the same distortion as the optimal decoder for $\mathbf{v}$ utilizing $\mathbf{y}_0$. Defining, $\tilde{\mathbf{v}}(m,\mathbf{w}) \overset{\text{def}}{=} \tilde{\mathbf{x}}(m,\mathbf{y}_0) - \hat{\mathbf{x}}(\mathbf{y}_0) + \mathbf{w}$ and noting that for $\mathbf{w} = \hat{\mathbf{v}}(\mathbf{y}_0)$ we have $\tilde{\mathbf{v}}(m,\mathbf{w}) = \tilde{\mathbf{v}}_-(m,\mathbf{y}_0)$, we see that the right-hand side of (5) can also be written as

$$
\mathsf{E}\left[\left\|\mathbf{v} - \tilde{\mathbf{v}}_-(m,\mathbf{Y}_0)\right\|^2 \mid \mathbf{y}_1\right] = \mathsf{E}\left[\left\|\mathbf{v} - \tilde{\mathbf{v}}(m,\mathbf{W})\right\|^2 \mid \mathbf{y}_1\right]. \tag{7}
$$

As a result, the optimal encoding and decoding rules for encoding of $\mathbf{v}$ with $\mathbf{W}$ as side information also define optimal encoding and decoding rules for $\mathbf{v}$ with $\mathbf{Y}_0$ as side information. This establishes that for the optimal encoder–decoder instantiation, Fig. 2(a) can be decomposed as shown in Fig. 2(b) without incurring any rate distortion penalty.

Remark 1: The preprocessing is useful for two reasons. First, the preprocessing successfully extracts relevant information $\mathbf{v}$ from $\mathbf{y}_1$. Second, the preprocessing reduces an indirect encoding problem into a direct encoding problem. Using our decomposition, $\mathbf{V}$ in (7) is directly computable at the encoder, and a codec can be designed correspondingly to minimize $\mathsf{E}[\|\mathbf{V} - \tilde{\mathbf{V}}\|^2]$. In contrast, $\mathbf{X}$ is not available at the encoder, thus we can not readily design a codec to minimize $\mathsf{E}[\|\mathbf{X} - \tilde{\mathbf{X}}\|^2]$. Practical code constructions with good performance have been demonstrated recently for the direct observation scalar scenario with side information, where the receiver seeks to reconstruct the one-dimensional observation "itself" [6], [26], [27]. The decomposition presented here allows us to directly exploit these existing coding schemes for our indirect vector observation situation[1].

Note also that the side informed estimation stage can operate on individual observation vectors and does not introduce delay, unlike the side-informed coding codec which requires that the encoding be performed over arbitrarily long blocks of i.i.d. observations for better R-D performance.

Remark 2: Combining the decomposition of Fig. 2(b), with the Wyner–Ziv result [28], [29] for the Gaussian case, which states that the rate-distortion (R-D) function for the side-informed coding of $\mathbf{v}$ with $\mathbf{w}$ as side information is identical to the R-D function when $\mathbf{w}$ is available at both the encoder and decoder, we can obtain the R-D function for the side-informed estimation as

$$
\mathcal{R}(D) = \begin{cases} \mathcal{R}_g(\Sigma_{\hat{\mathbf{x}}(\mathbf{s})}, D - D_0), & D \geq D_0 \\ \infty, & \text{otherwise} \end{cases} \tag{8}
$$

where $\mathcal{R}_g(\Sigma_{\mathbf{a}}, D)$ denotes the R-D boundary for a multivariate Gaussian signal with covariance matrix $\Sigma_{\mathbf{a}}$ at a distortion value D [30] , and $\Sigma_{\hat{\mathbf{x}}(\mathbf{s})} = \mathbf{F}_{\mathbf{xs}}\Sigma_{\mathbf{s}}\mathbf{F}_{\mathbf{xs}}^T$ denotes the covariance matrix of $\hat{\mathbf{x}}(\mathbf{S})$, and $D_0 = \mathsf{E}\left[\|\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{Y}_1)\|^2\right]$. Achieving the bound of (8), however, requires infinite length codes.

Remark 3: Mirroring the development of the decomposition in (1), we can readily obtain the following (matrix) decomposition, by utilizing the orthogonality principle:

$$
\mathsf{E}\left[(\mathbf{X} - \tilde{\mathbf{x}}(m(\mathbf{Y}_1),\mathbf{Y}_0))(\mathbf{X} - \tilde{\mathbf{x}}(m(\mathbf{Y}_1),\mathbf{Y}_0))^T\right]
$$
$$
= \mathsf{E}\left[(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{Y}_1))(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{Y}_1))^T\right]
$$
$$
+ \mathsf{E}\left[(\hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{Y}_1) - \tilde{\mathbf{x}}(m(\mathbf{Y}_1),\mathbf{Y}_0))(\hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{Y}_1) - \tilde{\mathbf{x}}(m(\mathbf{Y}_1),\mathbf{Y}_0))^T\right]. \tag{9}
$$

Similarly, we can obtain the matrix analog of (5) as

$$
\mathsf{E}\left[(\hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{y}_1) - \tilde{\mathbf{x}}(m,\mathbf{Y}_0))(\hat{\mathbf{x}}(\mathbf{Y}_0,\mathbf{y}_1) - \tilde{\mathbf{x}}(m,\mathbf{Y}_0))^T \mid \mathbf{y}_1\right]
$$
$$
= \mathsf{E}\left[(\mathbf{v} - \tilde{\mathbf{v}}_-(m,\mathbf{Y}_0))(\mathbf{v} - \tilde{\mathbf{v}}_-(m,\mathbf{Y}_0))^T \mid \mathbf{y}_1\right]. \tag{10}
$$

Equations (9) and (10) indicate that the covariance matrix of the estimation error in Fig. 2(b) is the summation of two parts, the first part corresponds to the covariance matrix of the estimation error when the remote observation is also available at the CP, and the second part corresponds to the covariance matrix of the reconstruction error of the preprocessed signal. We shall find these versions useful in the sequential estimation problem described in Section III.

---

[1]The decomposition of Fig. 2(b) is in fact a specific instance of the general side-informed encoder in Fig. 2(a), the former cannot offer an improvement in rate-distortion performance over the latter. Good practical coding methods, however, have not been previously explored for the situation of Fig. 2(a).
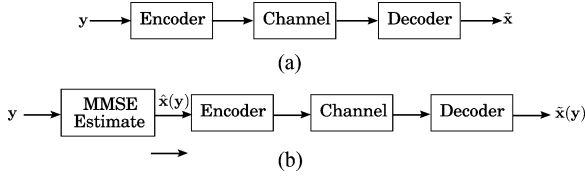
Fig. 3. Transmission of an indirect observation without side information. (a) Single-stage encoder and decoder. (b) Encoder is decomposed into an MMSE estimator and an encoder for the estimate.

A. Special Case: No Side Information

If the CP does not have any side information, the system of Fig. 2(a) reduces to the one in Fig. 3(a), a problem studied by Sakrison [8]. In this setting, (1) simplifies as

$$
\begin{aligned}
\mathsf{E}&\left[\|\mathbf{X} - \tilde{\mathbf{x}}(m(\mathbf{Y}))\|^2\right] \\
&= \mathsf{E}\left[\|\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}) + \hat{\mathbf{x}}(\mathbf{Y}) - \tilde{\mathbf{x}}(m(\mathbf{Y}))\|^2\right] \\
&= \mathsf{E}\left[\|\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y})\|^2\right] + \mathsf{E}\left[\|\hat{\mathbf{x}}(\mathbf{Y}) - \tilde{\mathbf{x}}(m(\mathbf{Y}))\|^2\right] \quad (11)
\end{aligned}
$$

where $\tilde{\mathbf{x}}(m(\mathbf{Y}))$ denotes the decoding rule for $\mathbf{x}$ without side information. Equation (11) immediately implies a decomposed structure for the optimal encoder–decoder chain, as shown in Fig. 3(b): At the encoder, first an MMSE estimate $\hat{\mathbf{x}}(\mathbf{y})$ is computed, which is then transmitted to the decoder over the rate-constrained channel. In this scenario, the innovation $\mathbf{S}$ reduces to the observation $\mathbf{y}$ itself, and the preprocessing $\mathbf{F_{xs}} = \mathbf{\Sigma_{xs}}\mathbf{\Sigma_s^{-1}}$ reduces into the MMSE estimator $\mathbf{\Sigma_{xy}}\mathbf{\Sigma_y^{-1}}$. Thus, Sakrison's result can be viewed as a special case of the separation presented here for the case when no side information is available. We note the Gaussian statistics are not required for the decomposition of Fig. 3(b), and the MMSE estimate $\hat{\mathbf{x}}(\mathbf{y}) = \mathsf{E}(\mathbf{X}|\mathbf{y})$ represents a sufficient statistic for estimating $\mathbf{X}$.

B. Dimensionality Constrained Channels: A Strong Parallel

The decomposition of Fig. 2(b) is obtained for a rate-constrained channel under the joint Gaussian assumption. An identical decomposition can also be obtained for a dimensionality-constrained channel [9], [10] where the $p_1 \times 1$ observations vector $\mathbf{y}_1$ is preprocessed at the encoder via a $r_1 \times p_1$ matrix $\mathbf{G}_1$, where $r_1 \leq p_1$. The resulting $r_1 \times 1$ vector $\mathbf{G}_1\mathbf{y}_1$ is utilized, along with the side information $\mathbf{y}_0$, by a linear decoder to obtain $\tilde{\mathbf{x}} = \mathbf{D}_0\mathbf{y}_0 + \mathbf{D}_1\mathbf{G}_1\mathbf{y}_1$ for $\mathbf{x}$, where $\mathbf{D}_0$ is a $L \times p_0$ matrix and $\mathbf{D}_1$ is a $L \times r_1$ matrix. In this scenario, the MSE optimal encoder–decoder pair $\{\mathbf{G}_1^*, \mathbf{D}_1^*\}$ can be obtained as the concatenation of the side-informed estimation $\mathbf{F_{xs}}$ followed by the optimal side-informed encoding of this estimate. The latter is obtained via the recently developed conditional KLT [4].

This decomposition result can be obtained by utilizing the orthogonality property for *linear* MMSE (LMMSE) estimators by which the estimation error is orthogonal to any *linear* function of the predictor inputs and does not require the assumption of joint Gaussianity. The problem of distributed estimation using reduced-dimensionality observations has been formulated and addressed in [9] and [10], where (among other results) the optimal linear pre and postprocessing matrices $\mathbf{G}_1^*, \mathbf{D}_1^*$ were derived. We illustrate in Appendix I how these optimal matrices

may also be alternatively obtained by combining our separation result with the conditional KLT.

III. MULTI-SENSOR SCENARIO: A SEQUENTIAL APPROACH

In this section, we return to the multisensor network of Fig. 1 and consider a sequential approach for distributed estimation and coding. We assume without loss of generality, that the sensors are indexed such that the CP receives the message transmitted from sensors in the index order, thus the information received from the lower-ordered sensors may be considered as part of the side information for subsequent (higher-ordered) sensors. The estimation and coding at these sensors can therefore be performed sequentially through a recursive exploit of the decomposition of Fig. 2(b), reducing the estimation and coding problem into a sequence of side-informed estimation and side-informed coding problems. In this section, we consider the side-informed estimation component of this framework, deferring discussions of the side-informed encoding to Section IV. In practice, the transmission order may be determined by other considerations or may be arbitrarily selected. In the latter scenario, we assume that the sensors are ordered in decreasing order of the conditional mutual information between the current sensor observation and the desired signal given the preceding observations, i.e., $\mathbf{I}(\mathbf{X}; \mathbf{Y}_i | \mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1}) \geq \mathbf{I}(\mathbf{X}; \mathbf{Y}_k | \mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1})$, for $k = i + 1, \ldots, N$, $i = 1, 2, \ldots, N$.

A. Sequential Estimation and Coding

For our specific development, we assume all observations are modeled as linear measurements corrupted by additive noise $\mathbf{y}_i = \mathbf{H}_i\mathbf{x} + \mathbf{n}_i$ $(i = 1, 2, \ldots, N)$, where $\mathbf{H}_i \in \mathbb{R}^{p_i \times L}$ represents a linear observation matrix, and $\mathbf{n}_i \in \mathbb{R}^{p_i}$ is the additive noise. At the ith sensor, denoted by $\mathcal{S}_i$, the noise vectors $\{\mathbf{n}_i\}_{i=1}^N$ are assumed to be independent of each other and of the signal vector $\mathbf{X}$, and each of these vectors is assumed to be a zero-mean multivariate Gaussian. We denote the allocation of the total rate R among the sensors by a vector $\mathbf{r} = [r_1, r_2, \ldots, r_N]$, where r_i represents the rate allocated for sensor $\mathcal{S}_i$, and $\sum_{i=1}^N r_i = R$. Rate allocation is described in Section III-B, here we consider the sequential preprocessing given an allocation vector.

We next determine the preprocessing matrix $\mathbf{F}_{i+1}$ at sensor $(i + 1)$, when the encoded observation from sensors $1, 2, \ldots, i$ are available as side information at the CP. We denote by $\hat{\mathbf{x}}_i$ the estimate of $\mathbf{X}$ obtained by the CP utilizing the messages from sensors $1, 2, \ldots, i$, and by $\mathbf{K}_i$ the covariance matrix for the corresponding estimation error $(\mathbf{X} - \hat{\mathbf{x}}_i)$. We will also find it notationally convenient to denote $\hat{\mathbf{x}}_0 = \mathsf{E}[\mathbf{X}] = 0$ and $\mathbf{K}_0 = \mathbf{\Sigma_x}$, the "initial values" of these quantities. The first sensor observation $\mathbf{y}_1$ is preprocessed by $\mathbf{F}_1 = \mathbf{\Sigma_{xy_1}}\mathbf{\Sigma_{y_1}^{-1}}$, and yields $\mathbf{v}_1 = \mathbf{F}_1\mathbf{y}_1$. This preprocessing matrix $\mathbf{F}_1$ is obtained from Sakrison's result in Section II-B since no side information pertaining to $\mathbf{y}_1$ is available. Next, $\mathbf{v}_1$ is transmitted using rate r_1 to the CP which obtains an (approximate) MMSE estimate of $\mathbf{v}_1$ using the coded message that it received from the first sensor. As indicated earlier, we defer discussion of the specifics of the coding method to Section IV. To account for the rate restriction on the coding, we model the CP's reconstruction of $\mathbf{v}_1$ as

$\tilde{\mathbf{v}}_1 = \mathbf{v}_1 + \boldsymbol{\epsilon}_1$, where $\boldsymbol{\epsilon}_1$ denotes the error which we assume is zero-mean Gaussian and independent of $\tilde{\mathbf{v}}_1$. The separation[2] represented by (11) indicates that $\tilde{\mathbf{v}}_1$ is also the MMSE estimate of $\mathbf{X}$, thus $\tilde{\mathbf{x}}_1 = \tilde{\mathbf{v}}_1$ and the estimation error covariance matrix $\mathbf{K}_1 = \mathsf{E}[(\mathbf{X} - \tilde{\mathbf{x}}_1)(\mathbf{X} - \tilde{\mathbf{x}}_1)^T]$. Using the matrix decomposition result in (9) and (10), we obtain $\mathbf{K}_1 = \mathbf{M}_1 + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_1}$, where

$$\mathbf{M}_1 \overset{\text{def}}{=} \mathsf{E}[(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_1))(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_1))^T]$$
$$= \mathbf{K}_0 - \mathbf{K}_0 \mathbf{H}_1^T (\mathbf{H}_1 \mathbf{K}_0 \mathbf{H}_1^T + \boldsymbol{\Sigma}_{\mathbf{n}_1})^{-1} \mathbf{H}_1 \mathbf{K}_0 \quad (12)$$

denotes the covariance matrix of $(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_1))$ which represents the MMSE estimation error when $\mathbf{y}_1$ is available at the CP without any distortion. The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_1}$, and also covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_2}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_3}, \ldots, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_N}$ that we require subsequently, are obtained from the codec as described in Section IV.

Next we consider the processing for the observation $\mathbf{y}_{i+1}$ at the $(i+1)^{\text{th}}$ sensor. We observe that, for the estimation of $\mathbf{y}_{i+1}$, $\tilde{\mathbf{x}}_i$ is a sufficient statistic of all the information available at the CP from sensors $1, 2, \ldots, i$. The innovation in $\mathbf{y}_{i+1}$ can be represented as $\mathbf{s}_{i+1} = \mathbf{y}_{i+1} - \hat{\mathbf{y}}_{i+1}(\tilde{\mathbf{x}}_i)$, where $\hat{\mathbf{y}}_{i+1}(\tilde{\mathbf{x}}_i) = \mathbf{H}_{i+1}\tilde{\mathbf{x}}_i$ is the MMSE estimate of $\mathbf{y}_{i+1}$ at the CP before $\mathbf{y}_{i+1}$ is encoded and transmitted. The decomposition of Fig. 2(b) indicates that the preprocessing for $\mathbf{y}_{i+1}$ is $\mathbf{F}_{i+1} = \boldsymbol{\Sigma}_{\mathbf{xs}_{i+1}} \boldsymbol{\Sigma}_{\mathbf{s}_{i+1}}^{-1}$, it can be verified that

$$\boldsymbol{\Sigma}_{\mathbf{xs}_{i+1}} = \mathbf{K}_i \mathbf{H}_{i+1}^T, \quad \boldsymbol{\Sigma}_{\mathbf{s}_{i+1}} = \mathbf{H}_{i+1} \mathbf{K}_i \mathbf{H}_{i+1}^T + \boldsymbol{\Sigma}_{\mathbf{n}_{i+1}}. \quad (13)$$

This preprocessed signal $\mathbf{v}_{i+1} = \mathbf{F}_{i+1}\mathbf{y}_{i+1}$ is encoded at rate r_{i+1} using a side-informed codec where $\mathbf{w}_i \overset{\text{def}}{=} \mathbf{F}_{i+1}\hat{\mathbf{y}}_{i+1}(\tilde{\mathbf{x}}_i)$ is available as side information at the CP. The CP's reconstruction of $\mathbf{v}_{i+1}$ via the side-informed decoder is modeled as $\tilde{\mathbf{v}}_{i+1} = \mathbf{v}_{i+1} + \boldsymbol{\epsilon}_{i+1}$, where, as before, we assume $\boldsymbol{\epsilon}_{i+1}$ is zero-mean Gaussian and independent of $\tilde{\mathbf{v}}_{i+1}$. Next, the CP refines its estimate of $\mathbf{X}$ by using $\tilde{\mathbf{v}}_{i+1}$ in the second stage of the decomposition, represented by a variant of (6),

$$\tilde{\mathbf{x}}_{i+1} = \tilde{\mathbf{x}}_i + \tilde{\mathbf{v}}_{i+1} - \mathbf{F}_{i+1}\hat{\mathbf{y}}_{i+1}(\tilde{\mathbf{x}}_i). \quad (14)$$

The estimation error covariance matrix is updated as $\mathbf{K}_{i+1} = \mathbf{M}_{i+1} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{i+1}}$, where

$$\mathbf{M}_{i+1} \overset{\text{def}}{=} \mathsf{E}[(\mathbf{X} - \hat{\mathbf{x}}(\tilde{\mathbf{x}}_i, \mathbf{Y}_{i+1}))(\mathbf{X} - \hat{\mathbf{x}}(\tilde{\mathbf{x}}_i, \mathbf{Y}_{i+1}))^T]$$
$$= \mathbf{K}_i - \mathbf{K}_i \mathbf{H}_{i+1}^T$$
$$\times (\mathbf{H}_{i+1} \mathbf{K}_i \mathbf{H}_{i+1}^T + \boldsymbol{\Sigma}_{\mathbf{n}_{i+1}})^{-1} \mathbf{H}_{i+1} \mathbf{K}_i. \quad (15)$$

Equation (15) corresponds to a simplified version of the recursion in the Kalman filter [25, p. 448]. Now, combining the initialization with the iteration, the recursive processing for the sequentially received sensor observations at the CP is summarized in Algorithm 1 , where for clarity of description we have included the coding and decoding steps which will be described in Section IV. Note that the recursive preprocessing in Algorithm 1 only requires the statistical characterization of the codec in

terms of the covariances $\{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}\}_{i=1}^N$ which, for a given rate allocation vector, can be obtained as described in Section III-B.

Algorithm 1: Recursive Estimation and Coding for Multiple Sensor Observations

Input: Statistics $\boldsymbol{\Sigma}_{\mathbf{x}}, \{\mathbf{H}_i, \boldsymbol{\Sigma}_{\mathbf{n}_i}\}_{i=1}^N$ and a rate vector $\mathbf{r}$. Available at all sensors and CP.

1:    **Initialize**: At CP and each Sensor, set $\tilde{\mathbf{x}}_0 = \mathsf{E}[\mathbf{x}] = \mathbf{0}$, $\mathbf{K}_0 = \boldsymbol{\Sigma}_{\mathbf{x}}$

2:    **for** $i = 1$ to N **do**

3:        Sensor $\mathcal{S}_i$ records the observation $\mathbf{y}_i$ and processes and transmits as follows:

4:          $\mathbf{F}_i = \mathbf{K}_{i-1}\mathbf{H}_i^T (\mathbf{H}_i \mathbf{K}_{i-1} \mathbf{H}_i^T + \boldsymbol{\Sigma}_{\mathbf{n}_i})^{-1}$

5:          $\mathbf{v}_i = \mathbf{F}_i \mathbf{y}_i$

6:        Transmit $\mathbf{v}_i$ with rate r_i

7:       At CP the reconstruction is updated using information from $\mathcal{S}_i$ as follows:

8:        The decoder obtains $\tilde{\mathbf{v}}_i = \mathbf{v}_i + \boldsymbol{\epsilon}_i$.

9:        $\tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_{i-1} + \tilde{\mathbf{v}}_i - \mathbf{F}_i \hat{\mathbf{y}}_i(\tilde{\mathbf{x}}_{i-1})$

10:    At CP and $\mathcal{S}_{i+1}$ update the estimation error covariance matrix:

11:        $\mathbf{K}_i = \mathbf{M}_i + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}$, $\mathbf{M}_i$ calculated by (15) and the evaluation of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}$ is described in Section III-B.

12: **end for**

For comparison, we also consider the scenario where each sensor directly encodes its observation using a KLT codec without applying the side-informed estimator and the CP estimates the desired signal recursively from the received values. The processing is obtained via a minor modification of Algorithm 1, where the sensor observations $\{\mathbf{y}_i\}_{i=1}^N$ are transmitted directly at each sensor without exploiting side information and the CP reconstructs an estimate of $\mathbf{x}$ from these received values. A corresponding test channel provides the required statistics.

B. An Achievable Rate-Distortion Bound and Rate Allocation for Sequential Estimation and Coding

We develop a numerical algorithm to jointly determine the rate allocation among the sensors under the total rate constraint, and an achievable rate-distortion bound under the sequential estimation and coding framework.

For a prescribed rate allocation $\mathbf{r} = [r_1, r_2, \ldots, r_N]$, where $\sum_{i=1}^N r_i = R$, an achievable R-D bound can be calculated in a procedure identical to Algorithm 1 with the actual encoding/decoding steps omitted and the statistics of reconstruction distortion $\boldsymbol{\epsilon}_i$ computed under the assumption that the side information is available at both the encoder and decoder[3]. Under this assumption for the ith channel, the forward test channel model [31, p. 479] specifies the reconstruction distortion $\boldsymbol{\epsilon}_i = \mathbf{t}_i - \tilde{\mathbf{t}}_i$, where $\mathbf{t}_i \overset{\text{def}}{=} \mathbf{v}_i - \hat{\mathbf{v}}_i(\tilde{\mathbf{x}}_{i-1})$ and $\tilde{\mathbf{t}}_i$ represents the reconstruction of $\mathbf{t}_i$ obtained from an optimal encoder–decoder combination with rate r_i.

---

[2]We note that the separation assumes that the encoder–decoder combination is optimal and therefore, strictly speaking, holds only asymptotically for codes operating at the rate distortion bound. The consequence also follows when the rate is high and $\boldsymbol{\epsilon}_1$ can be approximated as uncorrelated with $\mathbf{v}_1$.

[3]This is justified by the Wyner–Ziv rate-distortion result for the Gaussian setting.
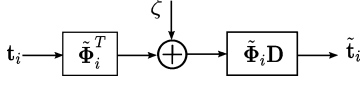
Fig. 4. Test channel model for transmitting a vector signal $\mathbf{t}_i$ under rate constraint.

Fig. 5. The decomposition with a side-informed transform codec.

Denoting $\Sigma_i \overset{\text{def}}{=} \mathsf{E}[\mathbf{t}_i\mathbf{t}_i^T]$, the vector Gaussian test channel (for the ith sensor) is obtained by using the eigendecomposition $\Sigma_i = \bar{\Phi}_i\Lambda_i\bar{\Phi}_i^T$, where Λ_i represents a diagonal matrix whose entries are the eigenvalues of Σ_i, arranged in decreasing order, and $\bar{\Phi}_i$ is the corresponding matrix of eigenvectors. The test channel parameters are then obtained via the reverse water-filling procedure on the diagonal entries in Λ_i [30]. First, the number k of transform channels with nonzero rate allocation and the corresponding Lagrange parameter λ that defines the per-channel distortion for these channels is obtained as the solution to the constraints $[\Lambda_i]_{11} \geq \lambda \geq [\Lambda_i]_{(k+1)(k+1)}$ and $r = \sum_{j=1}^k r_{ij}$, where $r_{ij} = 1/2\log([\Lambda_i]_{jj}/\lambda)$ and $[\Lambda_i]_{(N+1)(N+1)} = 0$ by convention. The corresponding test channel is then depicted in Fig. 4, where $\tilde{\Phi}_i$ is the submatrix obtained by truncating $\bar{\Phi}_i$ to its first k columns, ζ is a $k \times 1$ zero mean multivariate Gaussian independent of $\mathbf{t}_i$, with diagonal covariance matrix Σ_ζ where $[\Sigma_\zeta]_{jj} = \lambda[\Lambda_i]_{jj}/([\Lambda_i]_{jj} - \lambda,)$ and $\mathbf{D}$ is a diagonal matrix with diagonal entries $D_{jj} = ([\Lambda_i]_{jj} - \lambda)/[\Lambda_i]_{jj}$. The covariance matrix of $\epsilon_i = (\mathbf{t}_i - \tilde{\mathbf{t}}_i)$ is obtained from the test channel as

$$\Sigma_{\epsilon_i} = \bar{\Phi}_i\text{diag}(\{\min(\lambda, [\Lambda_i]_{jj})\}_{j=1}^L)\bar{\Phi}_i^T. \qquad (16)$$

Using the achievable R-D bound for a prescribed rate allocation, we develop an algorithm to allocate the total rate R among the multiple sensors as summarized in Algorithm 2. All sensors are initialized with a 0 bit allocation. Available bits, up to the total rate R, are then allocated by sequentially assigning each successive bit to the sensor where this additional allocation would result in the largest reduction in distortion. We note the achievable R-D bound we calculate is limited to the scenario where allocated rates are integers, facilitating practical implementation. The R-D bound corresponding to the optimized allocation is then used as a sequential achievable bound for our problem[4]. For the chosen rate allocation vector, the corresponding test channel of Fig. 4 also provides an estimate of Σ_{ϵ_i} for Algorithm 1.

Algorithm 2: Rate Allocation for Sequential Distributed Estimation and Coding

Input: $\Sigma_\mathbf{x}$, $\{\mathbf{H}_i, \Sigma_{\mathbf{n}_i}\}_{i=1}^N$, total rate R.

Output: $\mathbf{r} = [r_1, r_2, \ldots, r_N]$: rate allocation vector (across sensors)

1: Initialize $r_i = 0$, for all i
2: **for** $j = 1$ to r **do**
3:     **for** $i = 1$ to N **do**
4:         $\Delta_i = \inf$
5:         **if** $(i = 1)$ or $(r_{i-1} > 0)$ **then**
6:             $\mathbf{r}^i = [r_1, \ldots, r_i + 1, r_N]$, {tentatively assign the jth bit to sensor $\mathcal{S}_i$}
7:             calculate $\Delta_i \overset{\text{def}}{=}$ the sequential R-D bound with allocated rate vector $\mathbf{r}^i$ using Algorithm 1
8:         **end if**
9:     **end for**
10:     $k = \arg\min_i \Delta_i$, $r_k = r_k + 1$, {allocate the jth bit to the kth sensor which results in largest reduction in distortion using this bit.}
11: **end for**

IV. SIDE-INFORMED CODEC DESIGN

In order to encode the preprocessed sensor observations $\{\mathbf{v}_i\}_{i=1}^N$ within the corresponding channel rate constraints, we consider a practical side-informed codec construction in a transform coding framework [3], leveraging prior work on scalar side-informed codes for the direct observation setting [5], [6], [24]. An overview of the codec is illustrated in Fig. 5 using the notation of the previous section.

At the ith sensor node, the preprocessed observation $\mathbf{v}_i$ is transformed to obtain a transform vector $\mathbf{g}_i \overset{\text{def}}{=} \bar{\Phi}_i^T\mathbf{v}_i$ and correspondingly, at the CP, first an estimate $\tilde{\mathbf{g}}_i$ is obtained for $\mathbf{g}_i$, from which, $\mathbf{v}_i$ is estimated as $\hat{\mathbf{v}}_i = \bar{\Phi}_i\tilde{\mathbf{g}}_i$. Our transform codec corresponds to the conditional KLT [4], which ensures that the L transform coefficients forming the vector $\mathbf{g}_i$ are conditionally independent (given the side information at the CP). Note that the conditional KLT represents the optimal transform coding in the presence of receiver side information [4], much like the KLT represents the optimal transform coding in the absence of side information [32].[5]

At the CP, the side information $\mathbf{w}_i$ is transformed to obtain $\mathbf{h}_i \overset{\text{def}}{=} \bar{\Phi}_i^T\mathbf{w}_i = \bar{\Phi}_i^T\mathbf{F}_i\mathbf{H}_i\tilde{\mathbf{x}}_{i-1}$. Each of the L coefficients in $\mathbf{g}_i$ is then encoded using a scalar Wyner–Ziv codec, where the jth coefficient g_{ij} uses the corresponding side information h_{ij} in the decoding process. The side information is modeled as $h_{ij} = g_{ij} + \kappa_{ij}$, where κ_{ij} is assumed to be a zero mean Gaussian independent of h_{ij} with variance $\sigma_{\kappa_{ij}}^2 = \mathsf{E}[(g_{ij} - h_{ij})^2] = [\Lambda_i]_{jj}$, which corresponds to the conditional variance of the g_{ij} given the side information h_{ij}. The available rate r_i at the ith sensor

---

[4]The achievable bound we define can be shown to be attainable under random coding—though practical codecs attaining the corresponding performance have not been demonstrated yet.
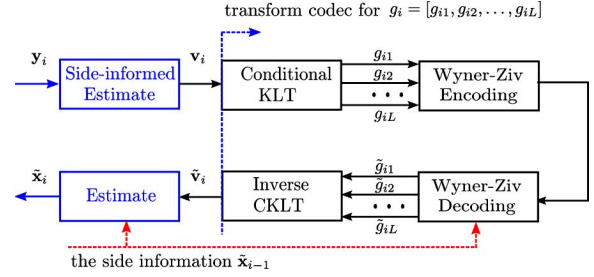
[5]The general optimal vector quantization problem is known to be NP complete [33].
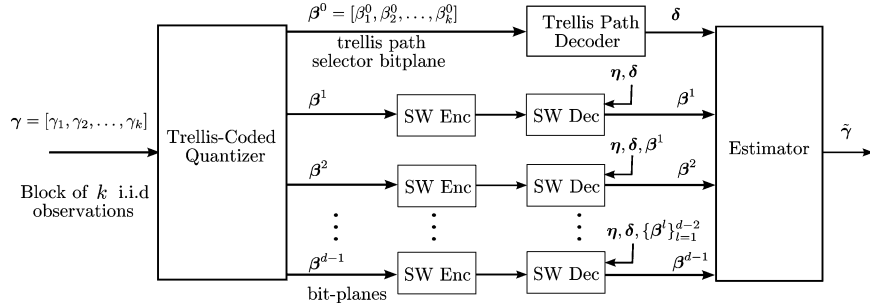
Fig. 6.   The bit-plane sequential scalar Wyner–Ziv codec obtained by concatenating TCQ with a bit plane sequential Slepian–Wolf code [6].

node is distributed among the L coefficients in $\mathbf{g}_i$ using the reverse water-filling algorithm as described in Section III-B to obtain rates $\{r_{ij}\}_{j=1}^{L}$ where r_{ij} is the rate allocated for coefficient g_{ij}. Next, without loss of generality, we consider this scalar Wyner–Ziv encoding for the coefficient g_{ij} with h_{ij} as the CP side information. Furthermore for notational simplicity, we drop the subscripts and denote g_{ij}, h_{ij}, κ_{ij}, and r_{ij}, respectively, by γ, η, κ, and ρ in the remainder of this description.

The scalar side-informed codec for γ is implemented as a trellis-coded quantizer (TCQ) [34], followed by a bit-plane sequential Slepian–Wolf [5], [35] codec,[6] which closely follows [6] and represents the current state-of-the-art in scalar Gaussian Wyner–Ziv coding. Because of space constraints, we include only a brief overview here and refer the reader to [6] and references therein for further details.

The codec operates on a block of k i.i.d observations $\boldsymbol{\gamma} \overset{\text{def}}{=} [\gamma_1, \gamma_2, \ldots, \gamma_k]$ at the sensor node. First, the samples are quantized using a d bit TCQ with 2^{d+1} representation points and then a Slepian–Wolf codec is utilized for communicating the bit planes sequentially where prior bit planes and side information is utilized for the decoding of each bit plane at the receiver, typically requiring fewer than d bits per sample for the transmission because of both the side information and correlations between the bit planes. We describe each of these components next and later in this Section, we describe how d is selected according to the allocated rate ρ.

For TCQ, we assume a specific configuration[7] wherein a quantized representation of the k observations, using d bits per sample, is obtained by using an underlying configuration of 2^{d+1} representation points, which are organized as $2^2 = 4$ cosets with 2^{d-1} points each. The sequence of representation points for the k samples are constrained so that at each time instant, a 2-bit output from a rate 1/2 feedforward binary convolutional code [36] determines the coset of the representation point. As a consequence of this constraint, the coset indexes of the representation points can be encoded as the corresponding sequence of input bits for the convolutional encoder using 1 bit per sample, leaving free the remaining $(d-1)$ bits at each time instant for the selection of the representation point closest to the observed sample, from the 2^{d-1} points within the coset. At the

encoder, a Viterbi [37], [38] search on the convolutional code trellis allows efficient determination of the best convolutional code input sequence $\boldsymbol{\beta}^0 = [\beta_1^0, \beta_2^0, \ldots, \beta_k^0]$ and coset member identifiers for the representation of the block of samples, the latter of which we represent by $(d-1)$ blocks of bit planes proceeding from most significant bit (MSB) to least significant bit (LSB) as $\boldsymbol{\beta}^1, \boldsymbol{\beta}^2 \ldots, \boldsymbol{\beta}^{d-1}$, where $\boldsymbol{\beta}^l = [\beta_1^l, \beta_2^l, \ldots, \beta_k^l]$ represents the lth bit-plane and the bits $[\beta_j^1, \beta_j^2, \ldots, \beta_j^{d-1}]$ indicate the selection of a representation point at time instant j, within the coset determined by the convolutional code output at time j in response to the input sequence $\boldsymbol{\gamma}^0$.

The TCQ module is coupled with a bit-plane sequential Slepian–Wolf codec to obtain our overall scalar Wyner–Ziv codec shown in Fig. 6, which, as previously indicated, is closely based on the construction in [6]. Because $\boldsymbol{\beta}^0$ incorporates the convolutional code trellis memory, it is directly transmitted to the decoder. Given $\boldsymbol{\beta}^0$, using the TCQ convolutional encoder, the decoder obtains the sequence $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_k]$ of 2-bit (or 4-valued) coset index vectors, where δ_j identifies the coset of representation points used at time j. This simplifies the side-informed coding and decoding of remaining bit planes because given δ_j and the side information η_j, the bit planes 1 through $(d-1)$ at time instant j are conditionally independent of corresponding planes and the side information at other time instants. This allows computation of the conditional likelihoods required for the Slepian–Wolf decoding in a memoryless fashion. To provide rate adaptability, we implement the individual bit-plane Slepian–Wolf codecs using low-density parity-check accumulate (LDPCA) codes [27], where the decoder for the lth bit-plane utilizes as side information the coset index $\boldsymbol{\delta}$, the side information $\boldsymbol{\eta}$ at the decoder, and previous bit-planes $\{\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^{l-1}\}$. The LDPCA code represents a rate adaptive version of the LDPC codes [39], [40] that have been used in side-informed coding [26], [27]. Parity information is generated by the LDPCA encoder and sent to the CP, where the coset index vector $\boldsymbol{\delta}$, the parity information and the side information are used by an iterative message-passing decoding [41] to estimate the encoded bit values. Appendix II describes how the side information is used in the decoding algorithm.

The LDPCA encoder provides rate scalability [27] by first transmitting a fraction of the computed parity bits, and progressively transmitting more bits at the request of the decoder if the decoding fails. Given the allocated rate-constraint ρ, $\boldsymbol{\beta}^0$ is first transmitted with rate 1 bit per sample, and the bit-plane $\boldsymbol{\beta}^1$ is

---

[6]Slepian–Wolf coding refers to side-informed coding for discrete alphabet sources as opposed to side-informed coding for continuous valued sources considered in Wyner–Ziv coding.

[7]Using alphabet constrained rate-distortion theory, [34] demonstrates that this configuration suffices for obtaining performance close to the R-D bound in the Gaussian setting.

transmitted next, a fractional bit rate ρ^1 is consumed in this process. The codec continues to transmit additional bit-planes $\boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^{d-1}$ consuming corresponding rates $\rho^2, \ldots, \rho^{d-1}$, respectively. The bit rate actually consumed is $\tilde{\rho} = (1 + \rho^1 + \rho^2 + \cdots + \rho^{d-1})$.

Once the decoded bit-planes are available from the LDPCA Slepian–Wolf decoder, it computes an estimate $\tilde{\boldsymbol{\gamma}}$ for $\boldsymbol{\gamma}$. The optimal estimate can be obtained, by using for each time instant j the conditional mean $\tilde{\gamma}_j = \mathsf{E}[\gamma_j | \delta_j, \beta_j^1, \beta_j^2, \ldots, \beta_j^{d-1}, \eta_j]$ as the estimate for γ_j. While this nonlinear estimate can be evaluated empirically [6], it offers only a minor improvement over a suboptimal linear estimator[8] which first uses the TCQ decoder to obtain an estimate $\boldsymbol{\gamma}^q = [\gamma_1^q, \gamma_2^q, \ldots, \gamma_k^q]$ for $\boldsymbol{\gamma}$ and then uses a linear MMSE estimator to obtain [6], [24]

$$\tilde{\gamma}_j = \frac{\sigma_\kappa^2}{\sigma_\kappa^2 + \sigma_{q,d}^2} \gamma_j^q + \frac{\sigma_{q,d}^2}{\sigma_\kappa^2 + \sigma_{q,d}^2} \eta_j \qquad (17)$$

where $\sigma_{q,d}^2 = \mathsf{E}[\|\gamma_j - \gamma_j^q\|^2]$ is the variance of the TCQ quantizer error, which is observed to be time invariant except for the trellis initialization phase. $\sigma_{q,d}^2$ is empirically estimated for a d-bit TCQ using a unit variance i.i.d. zero mean Gaussian input and scaled by the standard deviation of γ to obtain $\sigma_{q,d}^2$.

The variance of the estimation error for the LMMSE estimator is [25]

$$\mathsf{E}[\|\gamma_j - \tilde{\gamma}_j\|^2] = \frac{\sigma_\kappa^2 \sigma_{q,d}^2}{\sigma_\kappa^2 + \sigma_{q,d}^2}. \qquad (18)$$

In our codec implementation, for each coefficient, bit-plane β^l, $l = 1, 2, \ldots, (d-1)$, we select this TCQ bit depth d for which $0.5 \log(1 + (\sigma_\kappa^2 / \sigma_{q,d}^2))$ is closest to the target rate ρ, where the expression is obtained by considering the R-D function for a scalar Gaussian with variance σ_κ^2 and distortion given by the right-hand side of (18).

We note this codec introduces delay and requires a binary feedback channel.[9] The delay constraint can be avoided by using weaker side-informed codec, e.g., the nested quantizer, at the expense of (practical) rate-distortion performance. The feedback requirement can be eliminated, albeit at prohibitive computational cost, by specifically designing irregular LDPC codes for each individual bit plane customized for the corresponding conditional variance.

V. Simulation

In order to demonstrate the benefit of the decomposition we propose, and to investigate the effect of side-informed codecs, we simulate the sequential estimation and coding setup developed in Sections II and III. All sensor observations are modeled as $\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i$ for $i = 1, \ldots, N$, where the desired signal $\mathbf{x}$ is a multivariate Gaussian generated according to a covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}} \in \mathbb{R}^{L \times L}$. Several different choices of the measurement matrices $\mathbf{H}_i \in \mathbb{R}^{p_i \times L}$ and the signal covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ are explored in the simulations. Each of the p_i coefficients in the vector noise $\mathbf{n}_i$ is generated (independently) according

to the distribution $\mathcal{N}(0, \sigma_{n_i}^2)$. The value of σ_{n_i} is obtained according to a prescribed signal-noise-ratio (SNR) defined as

$$\text{SNR} = \frac{\text{tr}(\mathbf{H}_i \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{H}_i^T)}{\mathsf{E}[\|\mathbf{n}_i\|^2]} = \frac{\text{tr}(\mathbf{H}_i \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{H}_i^T)}{p_i \sigma_{n_i}^2}.$$

SNR values of 10 and 0 dB are utilized in the simulations in order to illustrate system performance in the high and low SNR regime, respectively.

Our Wyner–Ziv code constructions described in Section IV make use of 256 state TCQ using the optimized convolutional code polynomials listed in [6] that are specified by the pair of polynomials with coefficients (515, 362) in hexadecimal format. The 2^{d+1} representation points used in TCQ are chosen as the representation points for a $(d+1)$ bit Lloyd-Max quantizer [42].[10] For the LDPCA codes, we use the implementation of [27], which allow block lengths that are multiples of 66. We choose an encoder block length of $k = 6336$ for compatibility with this constraint while providing long enough blocks to allow both the TCQ and the LDPCA codecs to function efficiently.

We compare the proposed encoder construction, denoted by P-WZ, against the bounds for achievable performance obtained in Section III-B for our construction and against bounds delineating the *best achievable performance* for six other alternative scenarios. In total, we have seven scenarios.

1) Sequential preprocessing with side-informed coding (SEQ-WZ) as described in Section III: As we propose, the encoder applies the side-informed estimator as a preprocessing and uses our side-informed codec. This setting exactly matches our proposed construction. The bound derived in Section III-B, which we denote (SEQ-BD) represents the best performance achievable that practical constructions of SEQ-WZ, in particular, P-WZ, can attain.

2) Preprocessing with conventional transform coding (PTC): The encoder applies the side-informed estimator of Sections II and III as a preprocessing and uses a transform codec without exploiting the side information. The overall coding procedure to transmit a vector signal $\mathbf{v}_i$ is similar as the transmission of $\mathbf{t}_i$ described in Section III-B. The codec first applies the orthogonal transform matrix $\boldsymbol{\Phi}_{\mathbf{v}_i}^T$ which is specified by the eigendecomposition of $\boldsymbol{\Sigma}_{\mathbf{v}_i}$, i.e., $\boldsymbol{\Sigma}_{\mathbf{v}_i} = \boldsymbol{\Phi}_{\mathbf{v}_i} \boldsymbol{\Lambda}_{\mathbf{v}_i} \boldsymbol{\Phi}_{\mathbf{v}_i}^T$. The codec next allocates the available rate among the coefficients of $\boldsymbol{\Phi}_{\mathbf{v}_i}^T \mathbf{v}_i$ using the reverse water-filling algorithm. Attainable performance bounds are then obtained in a manner similar to the calculation of the achievable rate-distortion bound for the proposed sequential estimation and coding as described in Section III-B. We denote this bound by PTC-BD and it represents the best attainable performance for practical constructions of PTC.

3) Direct side-informed coding (DWZ): Sensor observations are encoded directly using our side-informed codec without the preprocessing but with a transform codec that

---

[8]This is demonstrated in [6] and also independently verified by us.

[9]The binary feedback channel enables rate adaptability by indicating decoding failures for the LDPCA code to the sensor, which serves as a signal for the encoder to send additional parity bits for the scalable LDPCA code.

[10]In our bit-plane sequential side-informed codec, inter bit-plane correlations are exploited in the side-informed decoding at the receiver and negligible performance gain is obtained by using a codebook of representation points optimized specifically for TCQ (for instance, with the methods described in [34] and [43]).

uses a KL transform for $\mathbf{y}_i$ mirroring the description in Part 2 above. Attainable performance bounds are also obtained similarly and denoted by (DWZ-BD).

4) Direct transform coding (DTC): Sensor observations are encoded directly using a conventional transform codec and a corresponding achievable bound (DTC-BD) is computed. The direct coding operation for both DTC-BD and DWZ-BD was outlined at the end of Section III-A.

5) Local MMSE estimation with transform coding (ETC-BD): The encoder obtains a local MMSE estimate $\hat{\mathbf{x}}(\mathbf{y}_i)$, which is sent to the CP. The CP obtains an estimate $\tilde{\mathbf{x}} = \mathsf{E}[\mathbf{x}|\hat{\mathbf{x}}(\mathbf{y}_0), \hat{\mathbf{x}}(\mathbf{y}_1), \ldots \hat{\mathbf{x}}(\mathbf{y}_N)]$. The achievable bound ETC-BD for this setting coincides with the decoupled estimate-compress scheme in [7].

6) General distributed estimation without requiring the sequential constraint (DST): Both preprocessing and side-informed coding can be used in DST and since the sequential constraint is not imposed, the performance at least as good as or better than SEQ-WZ is assured. No practical codec constructions are available as yet for DST, though a computational method for evaluating an achievable bound, which we denote DST-BD, is presented in [7]. We extend the two-sensor case described in [7] to the multisensor case relevant to our simulation settings.

7) Joint estimation from all sensor observations (JNT): JNT is infeasible in practice because it requires all sensor observations at one location. However, for comparison, it is useful to include the rate-distortion function for the transmission of the joint MMSE estimate $\mathsf{E}[\mathbf{x} \mid \{\mathbf{y}_i\}_{i=1}^N]$ represents an absolute lower bound of the rate-distortion function for the distributed estimation problem, which we refer to as the *joint lower bound* (JNT-BD).

Both options 1) and 2) operate according to Algorithm. 1, thus lie within the framework of the decomposition that we propose. In all our plots (unless otherwise specified), the abscissa is the average bit rate at each sensor and the ordinate is the decoder MSE distortion, expressed in decibels as $10\log_{10}\mathsf{E}[||\mathbf{x} - \tilde{\mathbf{x}}||]^2$. All results presented represent an average over ten Monte Carlo simulations. Due to the long block length of our codec, the standard deviation across simulations is quite small and is negligible on the scale of our plots and therefore not included.

A. Two-Sensor Case

We first simulate the two-sensor scenario as illustrated in Fig. 2(a), i.e., $N = 2$. The signal covariance matrix $\boldsymbol{\Sigma}_\mathbf{x}$ is set as $\boldsymbol{\Sigma}_\mathbf{x} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T$, where, prior to commencing the Monte Carlo runs, $\boldsymbol{\Sigma} \in \mathbb{R}^{L \times L}$ and the measurement matrices $\mathbf{H}_i \in \mathbb{R}^{p_i \times L}$ are generated with i.i.d. random coefficients each distributed as $\mathcal{N}(0, 1)$. In this scenario, the information-theoretic rate-distortion function is given by (8), and we benchmark our simulations against this theoretical lower bound (RD-BD), which also coincides with SEQ-BD, DST-BD, and JNT-BD, thereby reducing us to 5 additional scenarios characterized by bounds in addition to the practical P-WZ scheme.

For our first set of simulations, we set $L = 2$, $p_i = 4$ (for $i = 1, 2$) and obtain the results shown in Fig. 7(a) and (b) for the low and high noise regime, respectively. Our first observation from these figures is that our practical codec construc-
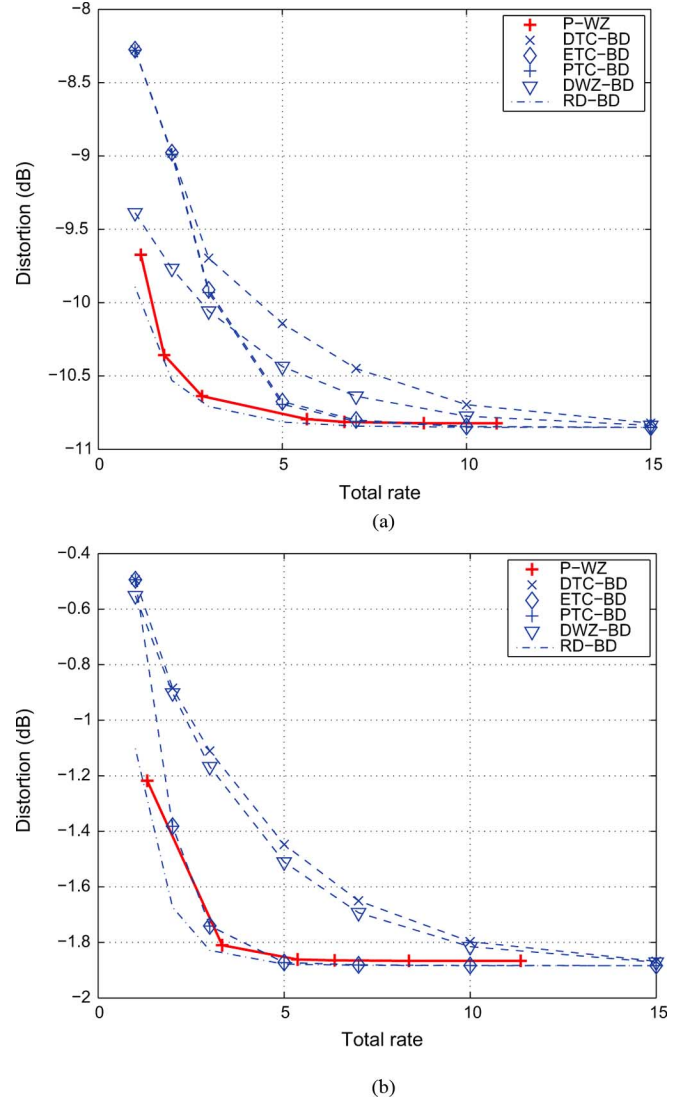

Fig. 7. Comparison of rate-distortion performances of different encoding schemes for the two-sensor case ($p_i = 4$, $L = 2$). (a) Low noise regime SNR = 10 dB. (b) High noise regime, SNR = 0 dB.

tion (P-WZ) does quite well and offers an RD performance approaching the RD-BD except for the very low rate case in the low noise regime.[11] In fact, in the region of interest where the bounds diverge, P-WZ's empirical performance exceeds even the bounds that mark the best attainable performance for all the 4 other alternative encoding scenarios.

A further comparison of the bounds is also instructive. In both the high and low noise regimes, PTC-BD and ETC-BD are quite close. Furthermore, Fig. 7(a) indicates that in a low noise regime the distortion corresponding to DTC-BD is larger compared to PTC-BD and ETC-BD. PTC-BD slightly outperforms ETC-BD. DWZ-BD also significantly outperforms DTC-BD, indicating the effectiveness of Wyner–Ziv coding in the low noise regime, although the cross-over between PTC-BD and DWZ-BD indicates that at high rates the preprocessing is more helpful whereas at low rates the side-informed coding is more

---

[11]At low rates, the direct transmission of the trellis path encoding bit plane introduces inefficiency causing this loss in performance [6].

effective. Fig. 7(b) presents results for the high noise regime. In this setting, DTC-BD yields poor performance as compared to PTC-BD, SEQ-BD and ETC-BD. We observe that PTC-BD has similar performance as ETC-BD. Comparing Fig. 7(a) and (b), we observe that under the low noise regime, the advantage of Wyner–Ziv coding is more significant, and the degradation of side information at high noise regime also decreases the efficiency of side-informed coding.

B. Multi-Sensor Scenario: A Special Case

Next, we consider a special case of the multisensor scenario that is synthesized to highlight the benefit of the proposed decomposition. The signal of interest and the sensor measurements are all derived via linear, potentially noisy, observations of an 6×1 zero-mean multivariate Gaussian vector $\mathbf{z}$ whose elements are i.i.d. with zero mean and unit variance. Specifically, we assume $\mathbf{x} = \mathbf{H}_0\mathbf{z}$, where $\mathbf{H}_0 = \mathrm{diag}(1,1,1,0,0,0)$, and the sensor observations are $\mathbf{y}_i = \mathbf{H}_i\mathbf{z} + \mathbf{n}_i$, where the measurement matrices are[12] $\mathbf{H}_1 = \mathrm{diag}(1,0,0,1,0,0)$, $\mathbf{H}_2 = \mathrm{diag}(1,1,0,0,1,0)$, and $\mathbf{H}_3 = \mathrm{diag}(1,1,1,0,0,1)$. Thus, we have $N = 3$, $L = 3$, and $p_i = 6$ for all i. In this special case, we allocate the total available rate equally among these sensors.

The rate distortion performance obtained in the simulations is summarized in Fig. 8(a) and (b) for the low and high noise regime, respectively. Once again, we see that our P-WZ codec construction offers performance close to the corresponding SEQ-BD. Though the gap between P-WZ and SEQ-BD is larger than the gap in the two sensor case,[13] the P-WZ construction outperforms all other bounds with the exception of the PTC bound in the high noise regime, where the noise makes the side-information ineffective for coding and local preprocessing component common to PTC and SEQ provides most of the gain. We also observe that the SEQ-BD is close to the DST-BD in both scenarios, despite the absence of the sequential constraint for DST. This once again demonstrates the effectiveness of the sequential approach. The three best achievable bounds are DST-BD, SEQ-BD, and PTC-BD in that order and these outperform other achievable bounds by a significant margin in either high or low noise scenarios. Furthermore, PTC-BD significantly outperforms ETC-BD and DTC-BD in both cases by using the preprocessing we propose. DWZ-BD also outperforms DTC-BD in both scenarios, and more significantly in the low noise regime where the Wyner–Ziv coding is more effective.

This special case clearly highlights the effectiveness of the proposed sequential approach, the preprocessing effectively extracts relevant information from the local sensor observation, and the Wyner–Ziv coding explores the side information at the CP to send the preprocessed signal.

[12]If the measurement model formulation is assumed to be in the frequency domain, these measurement matrices correspond to sensors with increasing bandwidth in the order 1, 2, 3, 4 with each of the sensors also covering a band of frequencies that is not of interest for estimation of the vector $\mathbf{x}$ desired at the CP.

[13]With the multiple sensors and channels the low rate inefficiencies of the TCQ codec and the scalable LDPCA codes are encountered more frequently.
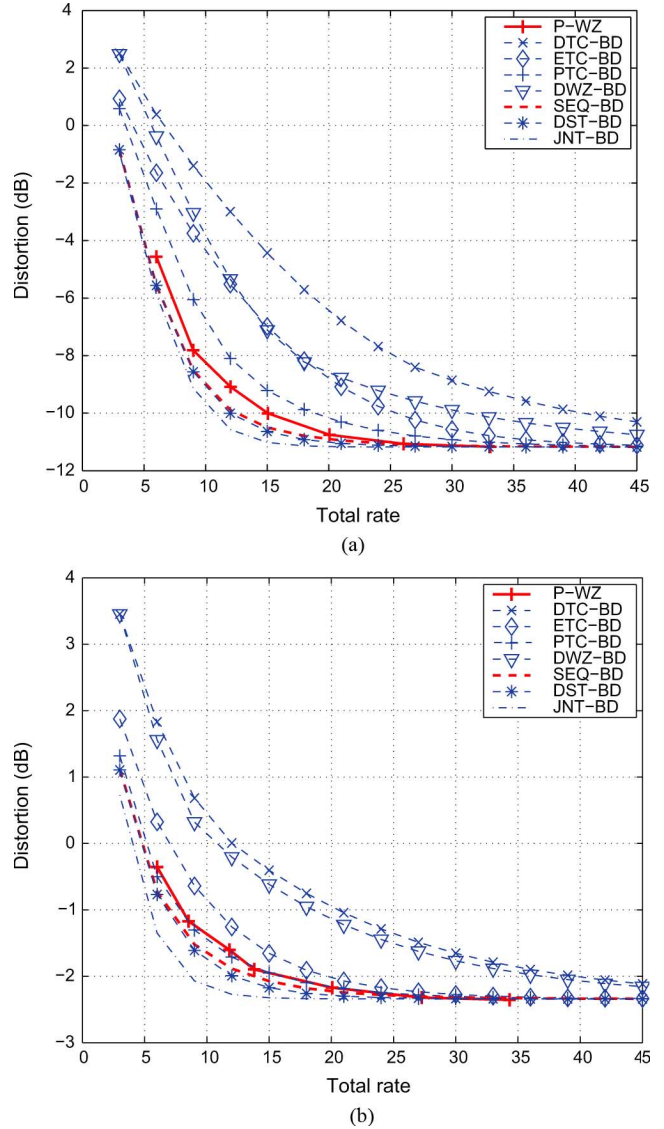

Fig. 8. Comparison of rate-distortion performances of different encoding schemes for the multisensor case with the measurement matrices of Section V-B. (a) Low noise regime SNR = 10 dB. (b) High noise regime, SNR = 0 dB.

C. Multi-Sensor Scenario: The General Case

We next simulate a "generic" multisensor scenario. We consider $N = 4$, $L = 2$, and $p_i = 4$ for all i. In this simulation, the autocovariance matrix $\mathbf{\Sigma}_\mathbf{x} = \mathbf{\Sigma}\mathbf{\Sigma}^T$, where, prior to the commencing the Monte Carlo runs, $\mathbf{\Sigma} \in \mathbb{R}^{L \times L}$ and the measurement matrices $\mathbf{H}_i \in \mathbb{R}^{p_i \times L}$ are generated with i.i.d random coefficients each distributed as $\mathcal{N}(0,1)$.

The simulation results are illustrated in Fig. 9 and similar trends are observed as in Fig. 8. Among the achievable bounds, in both high and low noise scenarios, DST-BD, SEQ-BD represent the best performance in that order with a small compromise in performance for SEQ-BD over DST-BD. Our practical codec construction P-WZ offers performance close to SEQ-BD except for low rate scenarios. P-WZ also outperforms the achievable bounds of other alternative encoding schemes except for low rate scenarios. We also observe that PTC-BD slightly outperforms ETC-BD, both clearly outperform DTC-BD. Furthermore, as expected, in the low noise regime side information
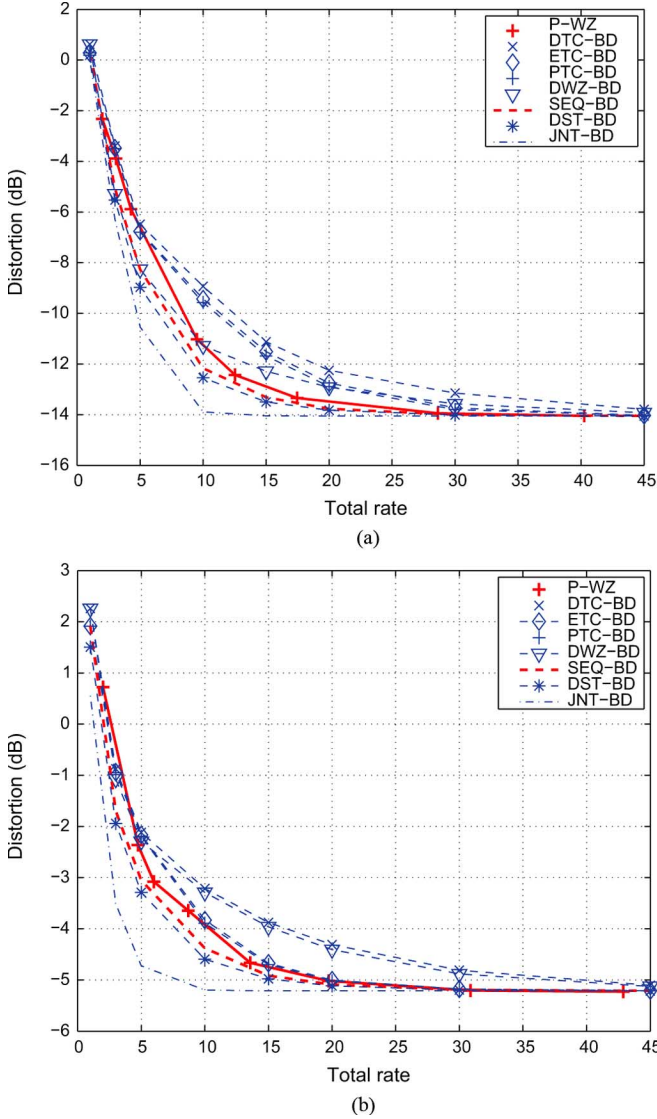
Fig. 9. Comparison of rate-distortion performances of different encoding schemes for the generic multisensor case. (a) Low noise regime SNR = 10 dB. (b) High noise regime, SNR = 0 dB.

is more effective than preprocessing so DWZ-BD outperforms PTC-BD and in the high noise regime the situation is reversed.

VI. DISCUSSION

Our problem setup assumes a distributed setting where the sensors communicate only with the CP and the CP does not transmit any of the information it receives from the sensors. However, signaling from the CP to the sensors is required to establish the network transmissions. Specifically, binary feedback enables rate adaptability for our DSC codes by indicating whether the DSC decoding is successful or not. Note that this is consistent with other DSC schemes [27], [44]. When the block length for the Wyner–Ziv codec is large, this rate for feedback signaling is negligible, when compared to the rate used for data transmission at the sensors, e.g., for our LDPCA codec, transmitting a block of 6336 bits requires at most 66 (on average around 30) bits of binary feedback. The rates consumed in

the feedback change the rate-distortion plots by imperceptible amounts and are therefore not included in the plots.

Our formulation imposes a sequential constraint on the encoding and decoding procedures. For the scenarios we compared in Figs. 8 and 9, we see interestingly that the achievable bounds for the sequential approach (SEQ-BD) and for the general distributed approach (DST-BD) are close. This suggests that the sequential approach may in fact be quite effective and may not levy a significant R-D penalty in a number of scenarios.

The gap between our codec performance and the sequential R-D bound (SEQ-BD) indicates that further improvements are feasible within the proposed sequential framework. In particular, optimized irregular LDPC codes would improve upon our LDPCA code, which provides rate adaptability but levies a moderate rate loss [27]. Such an approach is quite inflexible because individually optimized codes are necessary for specific bit planes and side-information statistics. The low rate performance can also be improved by using trellis coded vector quantization (TCVQ). Minor gains are also feasible by enlarging the number of TCQ states. Each of these approaches further increases computational complexity, which is already significant for our proposed construction. It is worth noting that in the multisensor scenario, deviations of the practical codec from the ideal codec assumptions at a given sensor, create a mismatch with the modeling assumptions for subsequent sensors which also partly explain the greater inefficiency in these scenarios in Figs. 8 and 9.

VII. CONCLUSION

In a sequential framework, the problem of distributed estimation and coding under rate constraints reduces to a sequence of subproblems of side-informed transmission of an indirect (noisy) observation. We show that under jointly Gaussian statistics and for a MSE metric, the optimal encoder–decoder design for indirect observations with side information can be decomposed into a preprocessing "side-informed" estimation step followed by "side-informed" coding of the estimate, without any degradation in rate-distortion performance. The decomposed structure enables the development of practical codec constructions with performance close to the numerical achievable bound for the sequential framework. Furthermore, the achievable bound of our sequential approach is close to the achievable distributed bound without the sequential constraint, implying that the sequential approach, which enables practical codec constructions, may not incur significant performance loss.

APPENDIX I
DISTRIBUTED ESTIMATION FROM DIMENSIONALITY-CONSTRAINED OBSERVATIONS

In this appendix, we consider an alternate instantiation of the distributed estimation and coding problem of Fig. 1 considering dimensionality constrained channels. This alternative instantiation was previously considered in [9], [10], and [12] and is shown in Fig. 10. Here the rate constraints on the channels are replaced by dimensionality constraints and the encoders and
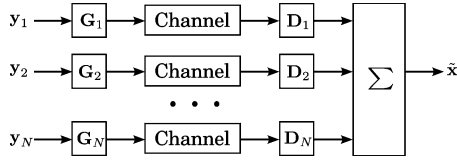
Fig. 10. Dimensionality constrained channels. The CP receives reduced-dimensionality observations and utilizes them in a linear estimator to estimate the desired signal $\mathbf{x}$.
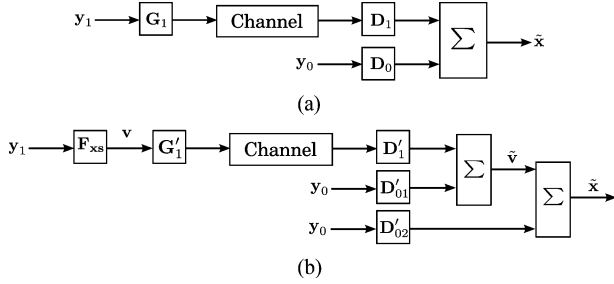
Fig. 11. (a) Estimation of $\mathbf{x}$ from indirect observation $\mathbf{y}_1$ with side information $\mathbf{y}_0$ under dimensionality constraint. (b) A decomposed structure of (a).

decoders are replaced by pre and postprocessing matrices, respectively. Let r_i denotes the dimensionality constraint for the channel connecting $\mathbf{y}_i$ and the CP, matrix $\mathbf{G}_i \in \mathbb{R}^{r_i \times p_i}$ denotes the linear processing for $\mathbf{y}_i$, where $r_i \leq p_i$ thus imposes the dimensionality constraint. At the CP, the preprocessed observation $\mathbf{G}_i \mathbf{y}_i$ is postprocessed by matrix $\mathbf{D}_i \in \mathbb{R}^{L \times r_i}$. The final estimate is obtained by combining the postprocessed outputs as $\tilde{\mathbf{x}} = \sum_{i=1}^{N} \mathbf{D}_i \mathbf{G}_i \mathbf{y}_i$. The optimization of the linear pre and postprocessing matrices $\{\mathbf{G}_i, \mathbf{D}_i\}_{i=1}^{N}$ under a MSE distortion metric can be written as

$$\{\{\mathbf{G}_i^*\}_{i=1}^{N}, \{\mathbf{D}^*\}_{i=1}^{N}\}$$

$$= \arg \min_{\{\mathbf{G}_i\}_{i=1}^{N}, \{\mathbf{D}\}_{i=1}^{N}} E \left\| \mathbf{X} - \sum_{i=1}^{N} \mathbf{D}_i \mathbf{G}_i \mathbf{Y}_i \right\|^2 . \quad (19)$$

We note Gaussian statistics are not required in this setup when only linear processing and linear channels are concerned. We first note in (19) that $\{\mathbf{D}_i\}_{i=1}^{N}$ can be obtained as the corresponding partitions of the LMMSE estimator once $\{\mathbf{G}_i\}_{i=1}^{N}$ are known, thus we focus on the optimization of $\{\mathbf{G}_i\}_{i=1}^{N}$. The analytical solution for the joint optimization of $\{\mathbf{G}_i\}_{i=1}^{N}$ is believed to be intractable [11]. We thus proceed as in prior literature and consider the optimization of $\mathbf{G}_1$, assuming $\{\mathbf{G}_i\}_{i=2}^{N}$ are known. An approximate solution for the joint optimization of $\{\mathbf{G}_i\}_{i=1}^{N}$ can be obtained by using this in an alternating maximization technique [9], where each of $\{\mathbf{G}_i\}_{i=1}^{N}$ is optimized iteratively until convergence. We next consider the optimization of $\mathbf{G}_1$ in the setup of Fig. 11(a) where $\mathbf{y}_0$ denotes all the preprocessed observations available at the CP.
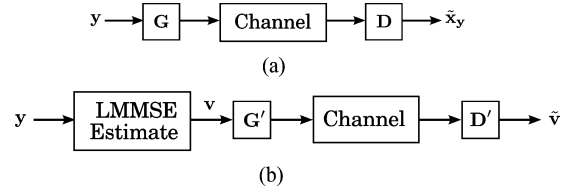
Fig. 12. (a) Estimation of $\mathbf{x}$ from single indirect observation $\mathbf{y}$ under dimensionality constraint. (b) A decomposed structure of (a).

A decomposition similar[14] to the one illustrated in Fig. 2(b) applies to the system of Fig. 11(a), and indicates a separative structure illustrated in Fig. 11(b). The processing for $\mathbf{y}_1$ consists of a first side-informed estimation preprocessing and side-informed "coding", the preprocessing is represented by $\mathbf{F}_{\mathbf{xs}}$ which is identical with that in Fig. 2(b), the side-informed "coding" is dimensionality-reduced approximation of $\mathbf{v} = \mathbf{F}_{\mathbf{xs}} \mathbf{y}_1$ with $\mathbf{y}_0$ as the side information. This approximation is obtained by the conditional KLT [4], which indicates $\mathbf{G}_1'^* = [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1}^{T}$, $\mathbf{D}_1'^* = [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1}$ and

$$\mathbf{D}_{01}'^* = (\mathbf{I}_L - [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1} [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1}^{T}) \mathbf{F}_{\mathbf{xs}} \boldsymbol{\Sigma}_{\mathbf{y}_1 \mathbf{y}_0} \boldsymbol{\Sigma}_{\mathbf{y}_0}^{-1}$$

where $\mathbf{I}_L$ represents the $L \times L$ unity matrix, $\mathbf{t} \stackrel{\text{def}}{=} \mathbf{v} - \hat{\mathbf{v}}(\mathbf{y}_0)$ denotes the innovation and $[\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1}$ represents the matrix of the r_1 eigenvectors of $\boldsymbol{\Sigma}_t$ corresponding to the r_1 largest eigenvalues. $\mathbf{D}_{01}'^*$ can be interpreted as the MMSE estimator of $(\mathbf{v} - [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1} [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1}^{T} \mathbf{v})$ from $\mathbf{y}_0$. The CP processing is decomposed similarly, first an approximation of $\mathbf{v}$ is obtained as $\tilde{\mathbf{v}} = \mathbf{D}_1'^* \mathbf{G}_1'^* \mathbf{v} + \mathbf{D}_{01}'^* \mathbf{y}_0$, and then $\tilde{\mathbf{x}}$ is obtained from $\tilde{\mathbf{v}}$ by using a variant of (7), i.e., $\tilde{\mathbf{x}} = \tilde{\mathbf{v}} + \hat{\mathbf{x}}(\mathbf{y}_0) - \mathbf{F}_{\mathbf{xs}} \hat{\mathbf{y}}_1(\mathbf{y}_0)$, indicating

$$\mathbf{D}_{02}'^* = \boldsymbol{\Sigma}_{\mathbf{x} \mathbf{y}_0} \boldsymbol{\Sigma}_{\mathbf{y}_0}^{-1} - \mathbf{F}_{\mathbf{xs}} \boldsymbol{\Sigma}_{\mathbf{y}_1 \mathbf{y}_0} \boldsymbol{\Sigma}_{\mathbf{y}_0}^{-1}.$$

Comparing Fig. 11(a) and (b), we conclude

$$\mathbf{G}_1^* = [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1}^{T} \mathbf{F}_{\mathbf{xs}} \quad (20)$$

$$\mathbf{D}_1^* = [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1} \quad (21)$$

$$\mathbf{D}_0^* = \boldsymbol{\Sigma}_{\mathbf{x} \mathbf{y}_0} \boldsymbol{\Sigma}_{\mathbf{y}_0}^{-1} - [\boldsymbol{\Phi}_{\mathbf{t}}]_{r_1} [\boldsymbol{\Phi}_t]_{r_1}^{T} \mathbf{F}_{\mathbf{xs}} \boldsymbol{\Sigma}_{\mathbf{y}_1 \mathbf{y}_0} \boldsymbol{\Sigma}_{\mathbf{y}_0}^{-1} \quad (22)$$

where we substitute $\mathbf{D}_0^* = (\mathbf{D}_{01}'^* + \mathbf{D}_{02}'^*)$ and reorganize resulting terms. These results are identical with those in [9], however obtained by applying our decomposition and the conditional KLT, both of which are also applicable to other channel models such as rate-constrained channel for jointly Gaussian signals considered in our main development.

Fig. 12(a) illustrates the scenario where no side-information is available (but the channel dimensionality is constrained to a positive integer r). The decomposition in Fig. 3(b) applies and establishes the separative structure in Fig. 12(b) [45]. The encoder first obtains the LMMSE estimate $\mathbf{v} \stackrel{\text{def}}{=} \hat{\mathbf{x}}(\mathbf{y}) = \boldsymbol{\Sigma}_{\mathbf{xy}} \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}$, followed by dimensionality-reduced approximation of $\mathbf{v}$. The latter problem can be reduced to a classic matrix approximation problem, which is readily obtained by the truncated Singular Value Decomposition (SVD)

[14]The decomposition here is slightly different from the one in Section II but can be obtained in an analogous fashion by utilizing the orthogonality property for the LMMSE estimator, that the estimation error is orthogonal to any linear function of the predictor inputs.

[46]: $\mathbf{G}'^* = [\mathbf{\Phi_v}]_r^T$, $\mathbf{D}'^* = [\mathbf{\Phi_v}]_r$. Comparing Fig. 12(a) and (b), we further obtain

$$\mathbf{G}^* = [\mathbf{\Phi_v}]_r^T \mathbf{\Sigma_{xy}} \mathbf{\Sigma_y}^{-1}, \quad \mathbf{D}^* = [\mathbf{\Phi_v}]_r. \tag{23}$$

The decomposition of the preprocessing matrix $\mathbf{G}^*$ as an LMMSE estimator followed by dimensionality reduction has been observed in [9], [10], and [47] after obtaining the expression of $\mathbf{G}^*$ through matrix decomposition or canonical correlation analysis [48]. In contrast, we first establish this decomposition and the solution follows in a straightforward manner.

APPENDIX II
DECODING WITH SIDE INFORMATION

Our side-informed codec uses the iterative message-passing decoding algorithm [41] which has been adapted for side-informed coding [26]. Here, we briefly summarize how the decoder inputs are computed and refer the reader to [27] for details of the Wyner–Ziv codec. This decoding algorithm relies on the parity information sent from the encoder, and the initial log-likelihood ratios (LLR) which are obtained at the decoder from the side information.

After the trellis path selector bit-plane β^0 is received, we obtain the coset index vector δ which is used for decoding the subsequent bit-planes. The decoding of the bit-plane $\beta^l \stackrel{\text{def}}{=} [\beta_1^l, \beta_2^l, \ldots, \beta_k^l]$ use already decoded bit-planes $\beta^1, \ldots, \beta^{l-1}$, together with η and δ, as side information, and requires as input the initial LLRs,

$$q_j^l = \log \frac{\Pr(\beta_j^l = 0 | \beta_j^1, \ldots, \beta_j^{l-1}, \eta_j, \delta_j)}{\Pr(\beta_j^l = 1 | \beta_j^1, \ldots, \beta_j^{l-1}, \eta_j, \delta_j)}. \tag{24}$$

The conditional probabilities in (24) are obtained by marginalization. For example,

$$\Pr(\beta_j^l = 0 | \beta_j^1, \ldots, \beta_j^{l-1}, \eta_j, \delta_j)$$
$$= \sum_{b_{l+1}=0}^1 \cdots \sum_{b_{d-1}=0}^1 \Pr(\beta_j^l = 0, \beta_j^{l+1} = b_{l+1}, \ldots, \beta_j^{d-1}$$
$$= b_{d-1} | \beta_j^1, \ldots, \beta_j^{l-1}, \eta_j, \delta_j). \tag{25}$$

We note that the side information η is not quantized at the decoder and the probability term in the summation on the right-hand side of (25) is obtained by empirical estimation. Specifically, the conditional probability mass function $p(\delta_j | \gamma_j)$ is empirically characterized for the TCQ encoder by assuming time invariance, and the Markov chain relationship $\{\delta_j, \beta_j^1, \ldots, \beta_j^{d-1}\} \rightarrow \gamma_j \rightarrow \eta_j$ is used to obtain the probability in (25). Details can be found in [6].

ACKNOWLEDGMENT

REFERENCES

[1] C. Yu and G. Sharma, "On side-informed coding of noisy sensor observations," in *Proc. 41st Asilomar Conf. Signals, Systems, Computers*, Nov. 2007, pp. 681–685.
[2] "Special issue: Distributed signal processing in sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, Jul. 2006.
[3] V. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 9–21, 2001.
[4] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen–Loeve transform," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.
[5] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–94, 2004.
[6] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner–Ziv coding based on TCQ and LDPC codes," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 376–387, 2009.
[7] I. Schizas, G. Giannakis, and N. Jindal, "Distortion-rate bounds for distributed estimation using wireless sensor networks," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 2, 2008, 748605.
[8] D. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 165–167, 1968.
[9] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.
[10] K. Zhang, X. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion-part VI: Sensor data compression," in *Proc. Int. Conf. Info. Fusion*, Queensland, Australia, 2003, pp. 221–228.
[11] Z. Luo, G. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proc. IEEE Int. Symp. Inf. Theory*, 2005, pp. 1441–1445.
[12] E. Song, Y. Zhu, and J. Zhou, "Sensors' optimal dimensionality compression matrix in estimation fusion," *Automatica*, vol. 41, no. 12, pp. 2131–2139, 2005.
[13] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, 2005.
[14] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: A maximum likelihood approach," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (IEEE ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 3293–3296.
[15] E. Msechu, A. Ribeiro, S. Roumeliotis, and G. B. Giannakis, "Distributed Kalman filtering based on quantized innovations," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (IEEE ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 2733–2736.
[16] A. Ribeiro, G. Giannakis, and S. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4782–4795, 2006.
[17] A. Vosoughi and A. Scaglione, "Sequential source coding with side information for sensor networks," presented at the 8th IEEE Workshop Signal Process. Adv. Wireless Commun., Helsinki, Finland, Jun. 2007.
[18] H. Zhu, I. Schizas, and G. Giannakis, "Power-efficient dimensionality reduction for distributed channel-aware Kalman tracking using wireless sensor networks," in *Proc. IEEE Stat. Signal Process. Workshop*, 2007, pp. 383–387.
[19] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *Signal Process.*, vol. 86, no. 11, pp. 3160–3179, 2006.
[20] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, 1997.
[21] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, 1998.
[22] J. Chen, X. Zhang, T. Berger, and S. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 977–987, 2004.
[23] J. Chen and T. Berger, "Successive Wyner–Ziv coding scheme and its application to the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1586–1603, 2008.
[24] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, 2003.
[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[26] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, 2002.

[27] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Process.*, vol. 86, no. 11, pp. 3123–3130, 2006.

[28] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.

[29] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[30] T. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[31] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[32] J. J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 3, pp. 289–296, Sep. 1963.

[33] M. R. Garey, D. S. Johnson, and H. S. Witsenhausen, "The complexity of the generalized Lloyd-Max problem," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 255–256, Mar. 1982.

[34] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gauss–Markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, 1990.

[35] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jan. 1973.

[36] T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*. New York: Wiley-Interscience, 2005.

[37] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[38] J. Forney and G. D. Forney, Jr, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[39] R. G. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: MIT Press, 1963.

[40] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003 [Online]. Available: http://www.inference.phy.cam.ac.uk/mackay/itila/; http://www.cambridge.org/0521642981

[41] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[42] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 129–137, Mar. 1982.

[43] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Boston, MA: Kluwer Academic, 2002.

[44] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, pp. 71–83, Jan. 2005.

[45] C. Yu and G. Sharma, "Distributed estimation using reduced dimensionality sensor observations: A separation perspective," in *Proc. Conf. Inf. Sci. Sys.*, Princeton, NJ, Mar. 2008, pp. 150–154.

[46] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[47] Y. Hua and W. Liu, "Generalized Karhunen–Loeve transform," *IEEE Signal Process. Lett.*, vol. 5, no. 6, pp. 141–142, 1998.

[48] D. Brillinger, *Time Series: Data Analysis and Theory*, E. Robinson, Ed., Expanded ed. San Francisco: Holden-Day, 1981.

Chao Yu (S'05) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004 and the M.S. degree in electrical and computer engineering from the University of Rochester, Rochester, NY, in 2005.

Since September 2005, he has been working towards the Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Rochester. His research interests lie in the area of signal and image/video processing, and specifically in distributed signal processing, distributed image/video compression and processing, and resource allocation in (visual) sensor networks.

Mr. Yu is a recipient of the best paper awards at the SPIE Visual Communications and Image Processing (VCIP) Conference, San Jose, CA, in 2009.

Gaurav Sharma (S'90–M'90–SM'00) received the B.E. degree in electronics and communication engineering from the Indian Institute of Technology Roorkee (formerly the University of Roorkee), India, in 1990; the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1992; and the M.S. degree in applied mathematics and Ph.D. degree in electrical and computer engineering from North Carolina State University, Raleigh, in 1995 and 1996, respectively.

From August 1996 through August 2003, he was with Xerox Research and Technology, Webster, NY, initially as a Member of Research Staff and subsequently as a Principal Scientist. From 2008 to 2010, he served as the Director for the Center for Emerging and Innovative Sciences (CEIS), a New York State-funded center for promoting joint university-industry research and technology development, which is housed at the University of Rochester. He is currently an Associate Professor at the University of Rochester in the Department of Electrical and Computer Engineering, the Department of Biostatistics and Computational Biology, and the Department of Oncology. He is the Editor of the *Color Imaging Handbook* (Boca Raton, FL: CRC Press, 2003). His research interests include media security and watermarking, color science and imaging, genomic signal processing, and image processing for visual sensor networks.

Dr. Sharma is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, SPIE, IS&T, and the IEEE Signal Processing Society (SPS) and the IEEE Communications Society. He was the 2007 Chair for the Rochester section of the IEEE and served as the 2003 Chair for the Rochester chapter of the IEEE SPS. He currently serves as the Chair for the IEEE Signal Processing Society's Image Video and Multi-Dimensional Signal Processing (IVMSP) technical committee. He is member of the IEEE SPS's Information Forensics and Security (IFS) technical committee and an advisory member of the IEEE Standing Committee on Industry DSP. He is an Associate Editor of the *Journal of Electronic Imaging* and in the past has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.