Distributed Quantization for Partially Cooperating Sensors Using the Information Bottleneck Method

Steffen Steiner 1,*, Abdulrahman Dayo Aminu 2 and Volker Kuehn 1

1 Institute of Communications Engineering, University of Rostock, 18119 Rostock, Germany; volker.kuehn@uni-rostock.de
2 Pydro GmbH, 18119 Rostock, Germany; aminuabdulrahmandayo@gmail.com
* Correspondence: steffen.steiner@uni-rostock.de

Abstract: This paper addresses the optimization of distributed compression in a sensor network with partial cooperation among sensors. The widely known Chief Executive Officer (CEO) problem, where each sensor has to compress its measurements locally in order to forward them over capacity limited links to a common receiver is extended by allowing sensors to mutually communicate. This extension comes along with modified statistical dependencies among involved random variables compared to the original CEO problem, such that well-known outer and inner bounds do not hold anymore. Three different inter-sensor communication protocols are investigated. The successive broadcast approach allows each sensor to exploit instantaneous side-information of all previously transmitting sensors. As this leads to dimensionality problems for larger networks, a sequential point-to-point communication scheme is considered forwarding instantaneous side-information to only one successor. Thirdly, a two-phase transmission protocol separates the information exchange between sensors and the communication with the common receiver. Inspired by algorithmic solutions for the original CEO problem, the sensors are optimized in a greedy manner. It turns out that partial communication among sensors improves the performance significantly. In particular, the two-phase transmission can reach the performance of a fully cooperative CEO scenario, where each sensor has access to all measurements and the knowledge about all channel conditions. Moreover, exchanging instantaneous side-information increases the robustness against bad Wyner–Ziv coding strategies, which can lead to significant performance losses in the original CEO problem.

Keywords: distributed compression; chief executive officer problem; cooperating sensors; distributed source coding; information bottleneck

1. Introduction

This contribution considers a special case of the distributed source coding problem where each sensor observes the same source signal. In order to forward their measurements over capacity limited links to a common receiver, the sensors have to compress their measurements. In the case where a direct communication among operating sensors is not possible, this problem is termed as the CEO problem. Here, the compression at each sensor is optimized according to the Wyner–Ziv coding principle exploiting only statistical side-information. Within this paper, the term CEO problem always stands for this non-cooperative CEO problem, which means that sensors cannot communicate with each other during runtime.

We extend this scenario and allow sensors to cooperate with each other by exchanging instantaneous side-information. The fully cooperative Chief Executive Officer (fcCEO) problem is obtained if sensors can forward their uncompressed observations over inter-sensor links to all other sensors. The partially cooperative Chief Executive Officer (pcCEO) problem represents a scenario where instantaneous side-information is compressed before it is forwarded to other sensors.
1.1. The CEO Problem

The CEO problem has been investigated for various system assumptions which mainly differ in the distribution of the variable of interest and the applied distortion measure. The quadratic Gaussian CEO problem considers jointly Gaussian signals and the mean squared error (MSE) distortion measure [1–6]. Using an infinite number of encoders, Oohama analytically derived an asymptotic version of the sum-rate distortion in [1]. In [2], the authors investigated the influence of cooperating and non-cooperating encoders on the distortion measure. It turns out that the distortion decreases asymptotically with the reciprocal sum-rate $R$ for non-cooperating encoder. In scenarios with cooperating sensors, the MSE distortion decays exponentially with $2^{-2R}$. The non-asymptotic case was first investigated in [3] where the authors derived an upper bound on the sum-rate distortion function. Moreover, they showed that this bound is tight when each encoder has the same measurement SNR. Prabhakaran et al. [4], Oohama [5] and Wagner et al. [6] characterized the complete rate-region for this quadratic Gaussian CEO problem. Results also exist for multivariate Gaussian relevant processes [7,8]. The CEO problem for arbitrary discrete source distributions and the logarithmic loss distortion measure has been analyzed in [9–11]. For this scenario, Courtade and Weissman completely characterized the CEO rate-region in [9]. In [10,11], asymptotic analyses for an infinite number of sensors have been performed. Using the Hamming distance as a distortion measure, in [10] an inevitable loss in error rate performance due to non-cooperating sensors has been discovered. For arbitrary distortion measures, a scaling law on the sum-rate distortion function has been derived in [11]. In [12,13], the authors developed a variational bound on the optimal trade-off between relevance and complexity of the CEO setup and used neural networks for encoder and decoder to compute this bound.

There exist several algorithmic approaches to solve the CEO problem [8,14–17]. In our previous work [14,15], a Greedy Distributed Information Bottleneck (GDIB) algorithm based on the inner bound of the CEO problem with log-loss distortion measure defined in [9] has been introduced to determine extreme points of the contra-polymatroid solution space. Equivalently to Wyner–Ziv coding, the algorithm optimizes the quantizer of a specific sensor using the mappings of previously designed quantizers as statistical side-information. We showed that the GDIB algorithm outperforms an individual scalar information bottleneck optimization of sensors without Wyner–Ziv coding, especially for larger networks. For asymmetric scenarios, we demonstrated the dependency of the performance on the optimization order, i.e., the Wyner–Ziv coding strategy. Since the memory complexity of this optimization algorithm depends exponentially on the network size, we introduced a way to compress the Wyner–Ziv side-information by means of the information bottleneck principle. However, there still remains a large performance gap between non-cooperative and fully cooperative distributed compression.

1.2. Partially Cooperating Sensors

In order to close the gap between non-cooperative and fully cooperative distributed compression, we consider an extension of the CEO system model allowing partial cooperation among sensors. While a rich literature can be found on the classical CEO problem, less results are known for partially cooperating sensors. Most work has been done for jointly Gaussian signals because they allow an analytical treatment at least in parts. In [18], it is shown that cooperation among sensors can reduce the compression sum-rate except for the quadratic Gaussian CEO problem. In [19], the authors consider estimation problems under communication constraints and propose coding strategies for tree-structured sensor networks. Exploiting the Wyner–Ziv coding principle, the authors developed solutions for general trees and provide particular results for serial and parallel networks. A two sensor system with a Gaussian source was investigated in [20] for two transmission scenarios: orthogonal but rate-limited links between sensors and the common receiver as well as the Gaussian multiple access channel. It was shown that cooperation between the sensors over rate-limited inter-sensor links leads in both cases to substantial gains in terms of the
compression sum-rate. Finally, a simple three node network consisting of encoder, helper and decoder was analyzed in [21]. The authors showed that side-information provided by the helper to the encoder need not exceed that given to the decoder.

1.3. Structure and Notation

This paper is structured as follows: Section 2 gives a short introduction to the information bottleneck principle. Sections 3 and 4 introduce the non-cooperative distributed sensing scenario with the Greedy Distributed Information Bottleneck as an algorithmic solution defined in [14,15] and the fully cooperative distributed sensing scenario being equivalent to a centralized quantization approach, respectively. The main contribution of this paper can be found in Section 5, which introduces the partially cooperative distributed sensing scenario containing three different inter-sensor communication protocols, i.e., successive broadcasting, successive point to point transmission and two-phase transmission. Section 6 concludes this paper.

Throughout this paper, the following notation is used. Calligraphic letters $X, Y, Z$ denote random variables with realizations $x, y, z$, which are elements of the sets $X, Y, Z$ with cardinalities $|X|, |Y|$ and $|Z|$, respectively. Bold letters $y = [y_1 \ldots y_M]^T$ denote vectors while boldface calligraphic letters $Y, Z$ denote multivariate random variables. Note that $Z_{<m}$ covers only the processes $Z_1$ to $Z_{m−1}$. $I(X; Y)$ represents the mutual information between the random variables $X$ and $Y$. Conditional and joint probability mass functions (pmfs) are termed $p(y|x)$ and $p(x, y)$, respectively. The Kullback-Leibler (KL) divergence is given as $D_{KL}[\cdot || \cdot]$. Finally, the expectation of a function $f(x)$ with respect to the random variable $X$ is denoted as $E_X[f(X)]$.

2. The Information Bottleneck Principle

The information bottleneck (IB) principle was first introduced by Tishby et al. in [22,23] and defines a clustering framework based on information theoretic measures. An overview about algorithmic solutions for this basic optimization problem is given in [23,24]. The IB principle finds application in various fields in communications [25–30].

The general IB setup is depicted in Figure 1. It contains the relevant process $X$, a noisy observation $Y$ of $X$ and a compressed version $Z$ of $Y$. The IB approach aims to optimize the mapping $p(z|y)$ in order to preserve as much information about the relevant process $X$ in $Z$ as possible. More precisely, it tries to maximize the relevant mutual information $I(X; Z)$ while fulfilling a rate constraint $I(Y; Z) \leq C$. This general goal is summarized in Figure 2a. The optimization can be formulated as a maximization of the Lagrangian function

$$L_{IB} = I(X; Z) - \beta I(Y; Z).$$

(1)

It turns out to be a non-convex optimization problem, since $I(X; Z)$ and $I(Y; Z)$ are both convex functions of the mapping $p(z|y)$. The parameter $\beta$ is a trade-off parameter steering the focus between the preservation of relevant information and the compression of the observation. In the case of $\beta = 0$, the focus only lies on preservation of relevant information. By increasing $\beta$ the compression becomes more and more important up to the case of $\beta \to \infty$. Here, the functional in (1) becomes maximal if $I(Y; Z) = 0$, which means that all information is compressed to a single cluster. Therefore, the parameter $\beta$ can be used to adjust the compression rate $I(Y; Z)$ in order to fulfill a desired rate constraint $I(Y; Z) \leq C$. Since the compression-rate curve is a monotonic increasing function in $\frac{1}{\beta}$, a simple bisection search can be applied. The optimization problem in (1) can be solved by taking the derivative with respect to the mapping $p(z|y)$ and equating it to zero. It results in the implicit update equation

$$p(z|y) = \frac{e^{-d_p(y,z)}}{\sum_z e^{-d_p(y,z)}}$$

(2)
with
\[ d_{\beta}(y, z) = \frac{1}{\beta} D_{KL}[p(x|y) \| p(x|z)] - \log p(z) \]
(3)
\[ = \frac{1}{\beta} \mathbb{E}_{x|y} \left[ \log \frac{p(x|y)}{p(x|z)} \right] - \log p(z). \]

In (3), \( D_{KL}[p(x|y) \| p(x|z)] \) denotes the Kullback–Leibler divergence. This implicit solution can be solved by an iterative Blahut–Arimoto-like algorithm.

In the case of focusing solely on preservation of relevant information with \( \beta = 0 \), the optimization algorithm yields a deterministic clustering \( p(z|y) \in \{0, 1\} \). For \( \beta > 0 \), the clustering \( p(z|y) \in [0, 1] \) is generally stochastic. The IB method can easily be extended to multiple input values. A graphical tool for visualization are IB graphs [31]. Figure 2b illustrates an example where the observations \( y_1, y_2 \) are compressed into the cluster index \( z \). The trapezoid represents the IB compression with respect to the relevant variable written inside the trapezoid.

### 3. Non-Cooperative Distributed Sensing System

Figure 3 illustrates the CEO system model without communication among operating sensors. Here, \( M \) sensors observe noisy versions \( y_m \) of the same relevant signal \( x \). The measurement processes can be modeled as statistically independent memoryless channels (MCs).

Exemplarily, a measurement \( y_m = x + w_m \) represents the relevant signal \( x \) corrupted by zero mean white Gaussian measurement noise \( w_m \) with measurement signal-to-noise-ratio (SNR) \( \gamma_m = \frac{\sigma_x^2}{\sigma_{w_m}^2} \), where \( \sigma_x^2, \sigma_{w_m}^2 \) denote signal and noise variances, respectively. In order to be able to forward the measurements over capacity limited links with capacities \( C_1, \ldots, C_M \), each sensor has to compress its observations using a specific encoding process. More precisely, each sensor compresses its observations \( y_m \) to a cluster index \( z_m \) using the mapping \( p(z_m|y_m) \) leading to the Markov property:

\[ p(z, y, x) = \prod_{m=1}^{M} p(z_m|y_m)p(y_m|x)p(x). \]

The encoding process contains a second lossy compression step if the mapping \( p(z_m|y_m) \) is stochastic and lossless entropy coding if the mapping \( p(z_m|y_m) \) is deterministic. Therefore, a compressed version of the index \( z_m \) is transmitted without any further loss to the common receiver. The optimization of \( p(z_m|y_m) \) for each sensor is done offline.
The mathematical analysis of the CEO problem and the structure of its rate-region for discrete input alphabets and the log-loss distortion measure was presented in [9] and exploits (4). It was proved that the extreme points of the contra-polymatroid solution space can be determined by greedy algorithms as the one described next. Since the communication among sensors during run-time is not possible in this approach, the solution represents a lower bound on the performance of cooperative distributed compression in this paper.

![Figure 3. Non-cooperative distributed sensing system with M sensors, a common receiver and individual link capacities C_m.](image)

An algorithmic solution to solve the CEO problem has previously been proposed in [14,15] as the so called Greedy Distributed Information Bottleneck (GDIB) algorithm. It is based on the inner bound of the CEO rate-region for the logarithmic loss distortion measure [9] and optimizes the quantization at the sensors successively. Replacing the logarithmic loss function \( H(X|Z) \) by the relevant mutual information \( I(X;Z) \) delivers the optimization problem

\[
\max_P I(X;Z) \text{ s.t. } I(Y_S;Z_S|Z_S) \leq \sum_{m \in S} C_m \quad \forall \ S \subseteq \{1,2,\ldots,M\}. \quad (5)
\]

The set \( P = [p(z_1|y_1) \cdots p(z_M|y_M)] \) defines the set of all mappings. According to [9], the compression rates \( I(Y_S;Z_S|Z_S) \) are supermodular set functions with respect to the sets \( S \) [32], while the relevant information \( I(X;Z) \) does not depend on \( S \). Therefore, the greedy optimization structure of the GDIB algorithm is optimal and finds the extreme points of the solution space. It has to be emphasized that since the GDIB algorithm is based on the inner bound of the rate-region, it does not find the complete rate-region of the CEO problem. Following this approach, \( M \) IB related Lagrangian optimization problems are obtained, one for each sensor.

\[
L_{\text{GDIB}}^{(1)} = I(X;Z_1) - \beta_1 I(Y_1;Z_1)
\]

\[
L_{\text{GDIB}}^{(M)} = I(X;Z_M|Z_{<M}) - \beta_M I(Y_M;Z_M|Z_{<M})
\]

Obviously, the optimization problem of the first sensor resembles the optimization problem for the scalar IB problem given in (1) since there is no predecessor. Subsequent sensors exploit the mappings of previously designed quantizers as statistical side-information leading to the well-known Wyner–Ziv coding strategy. Naturally, each Lagrange multiplier \( \beta_m \) has to be chosen such that the corresponding compression rate fulfills the individual rate requirements.
constraint $I(Y_m; Z_m | Z_{<m}) \leq C_m$. The objectives in (6) and (7) can be solved by equating the derivative with respect to the mapping $p(z_m | y_m)$ to zero delivering the update rule

$$p(z_m | y_m) = \frac{e^{-d_{\beta_m}(y_m, z_m)}}{\sum_{z_m} e^{-d_{\beta_m}(y_m, z_m)}}$$

(8)

with the exponent

$$d_{\beta_m}(y_m, z_m) := E_{Z_{<m} | y_m} \left[ \frac{1}{\beta_m} D_{KL}[p(x | y_m, Z_{<m}) \| p(x | z_m)] - \log p(z_m | Z_{<m}) \right].$$

(9)

Similar to the scalar IB optimization, the implicit expression in (8) can be solved using a Blahut–Arimoto like algorithm, providing local optimal solutions. It has to be mentioned that for asymmetric scenarios, this optimization has to be performed for all $M!$ possible permutations of the optimization order to find the best solution.

A detailed derivation and performance analysis of this algorithm can be found in [14,15]. If the capacity is equally distributed over all sensors in the network, e.g., sensors share the same channel in an orthogonal way and a round robin fashion, numerical results demonstrate that the GDIB algorithm outperforms an individual scalar IB optimization at each sensor. However, there is still a large gap to the performance of a fcCEO scenario, which is defined in Section 4. Moreover, in asymmetric scenarios, the performance highly depends on the optimization order. Although no clear conclusion about the optimal Wyner–Ziv coding strategy can be drawn, a good solution can be expected when starting the optimization with the best forward channel conditions, i.e., the lowest compression (highest compression rate).

### 4. Fully Cooperative Distributed Sensing—A Centralized Quantization Approach

This section introduces the fcCEO scenario, which considers distributed sensors being able to forward their uncompressed observations to all other sensors in the network over ideal noiseless inter-sensor links. In this case, sensors can perfectly exchange their measurements $y_m$ before they jointly compress the received signals taking into account the rate constraints of all individual forward channels. Naturally, the exchange has to be done by a two-phase transmission protocol, consisting of a cooperation phase and a transmission phase. During the cooperation phase sensors exchange information until every sensor knows measurements $y = [y_1 \ldots y_M]^T$ of all $M$ sensors. The actual forwarding of the compressed observations to the common receiver is performed during the transmission phase. This full cooperation is equivalent to a single central quantizer having access to all measurements $y$ as depicted in Figure 4. Applying the IB principle, this central quantizer can be designed in order to compress the vector $y$ onto a cluster index $z$ using the mapping $p(z | y)$, which motivates the name centralized IB (CIB) for the algorithmic solution in a fcCEO scenario. The optimization problem can be formulated as the maximization of

$$L_{CIB} = I(\mathbf{X}; \mathbf{Z}) - \beta I(\mathbf{Y}; \mathbf{Z})$$

(10)

and is solved using update Equation (2) with (3) substituting the scalar $y$ by vector $y$. The number of output clusters $|Z|$ has to be chosen to $|Z| = \prod_{m=1}^{M} |Z_m|$ while the single link from the imaginary central quantizer to the receiver in Figure 4 has a channel capacity of $C_{\text{sum}} = \sum_{m=1}^{M} C_m$. The actual transmission over the $M$ links has to be coordinated such that each sensor $m$ transmits a specific part of the bits corresponding to its link capacity $C_m$. 
In the special case of the measurement process being modeled as additive noise, the algorithm can be simplified to a scalar optimization problem where maximum ratio combining of all inputs $y_m$ delivers a scalar sufficient statistics

$$\bar{y} = \sum_{m=1}^{M} \gamma_m \cdot y_m$$

of the desired relevant signal $x$ with an overall SNR $\gamma = \sum_{m} \gamma_m$. The solution of the fcCEO scenario serves as an upper bound in this paper.

5. Partially Cooperative Distributed Sensing

In order to investigate how the gap between non-cooperative and fully-cooperative distributed compression can be reduced, partially cooperating sensors shall now be considered. Partial cooperation means a limited exchange of instantaneous side-information among the sensors during runtime due to a rate-limitation of inter-sensor links. Non-cooperative CEO and fully-cooperative CEO problems represent the extreme cases for zero rate and unlimited rate inter-sensor links, respectively. The rate limitation requires the compression of instantaneous side-information before forwarding it to other sensors.

In this paper, only deterministic mappings are considered for this compression, while indexes $z_m$ are still obtained by stochastic mappings. This is motivated by the fact that deterministic mappings do not require further lossy compression and the resulting side-information indices $s_m$ can be exploited at other sensors by choosing a particular mapping $p(z_m|y_m)$ from a list of possible mappings designed offline in advance. As a consequence, the compression rates for instantaneous side-information can only be adjusted by changing the cardinalities $|S_m|$. For all results presented below, inter-sensor links are modeled as bit pipes being able to deliver $s_m$ reliably.

The GDIB algorithm to solve the non-cooperative CEO problem is based on the inner bound (5) of the CEO rate-region. Moreover, a greedy optimization approach is optimal due to the supermodularity of the compression rates in (5). Both require the Markovian structure in (4). However, cooperation among sensors changes the Markovian structure and implies different statistical dependencies among involved random variables. As (4) does not hold anymore in pcCEO scenarios, the inner bound on the rate-region in (5) cannot be utilized to find solutions of the pcCEO scenario. To the knowledge of the authors, tight bounds on the rate-region are not available for the cooperative case. Therefore, a heuristic approach based on the greedy optimization structure of the GDIB algorithm will be applied to solve the pcCEO scenario, which is not proven to be optimal. Nevertheless, the numerical evaluation of the found solutions demonstrate their usefulness. However, the computation of required pmfs becomes more challenging and results in recursive calculations given in Appendices A.1 and A.2 because the Markovian structure of (4) does not hold anymore in pcCEO scenarios.

This paper introduces three different inter-sensor communication protocols for exchanging this instantaneous side-information: successive broadcasting, a successive point-to-point transmission and a two-phase transmission. The first two protocols perform the exchange of instantaneous side-information $s_m$ with other sensors and the forwarding of
compressed versions of $z_m$ to the common receiver in the same time slot. Contrarily, the two-phase transmission protocol separates the exchange of instantaneous side-information among sensors and the communication with the common receiver into two distinct phases. The latter starts after the exchange among sensors has been completed such that all sensors have (approximately) the same amount of side-information.

5.1. Successive Broadcasting Protocol

The system model for the successive broadcasting protocol is illustrated in Figure 5. In the same time slot, sensor $m - 1$ not only forwards a compressed version of the quantization index $z_{m-1}$ to the common receiver, but also broadcasts instantaneous side-information $s_{m-1}$ to all other sensors. However, due to the greedy optimization structure, only subsequent sensors can exploit this instantaneous side-information. Thus, sensor $m$ can exploit indices $s_{<m}$ of all previously transmitting sensors in order to select its quantization index $z_m$ as well as a new instantaneous side-information index $s_m$. This scenario leads to the Markov model

$$p(x, y, z, s) = \prod_{m=1}^{M} p(z_m | y_m, s_{<m}) p(s_m | y_m, s_{<m}) p(y_m | x) p(x).$$

Figure 5. Partially cooperative CEO scenario using broadcast exchange of side-information among sensors.

The indices $s_m$ and $z_m$ are obtained by deterministic mappings $p(s_m | y_m, s_{<m}) \in \{0, 1\}$ and stochastic mappings $p(z_m | y_m, s_{<m}) \in [0, 1]$, respectively. The design of these mappings can be performed offline leveraging the IB principle as illustrated in Figure 6. It combines the observation $y_m$ and the instantaneous side-information $s_{<m}$ to indexes $s_m$ and $z_m$ while maintaining as much information as possible about the relevant signal $x$.

Figure 6. Graphical illustration of IB fusion of involved inputs to determine instantaneous side-information $s_m$ (a) and the quantizer sensor output $z_m$ (b) for a broadcast exchange of instantaneous side-information.
5.1.1. Generation of Broadcast Side-Information

The design of $p(s_m|y_m, s_{<m})$ is inspired by the general GDIB algorithm, i.e., the optimization is done in a greedy manner. Again, there emerges one optimization problem for each sensor:

$$L^{(1)}_{BC-SIDE} = I(X; S_1) - \beta I(Y_1; S_1)$$

$$\vdots$$

$$L^{(M-1)}_{BC-SIDE} = I(X; S_{M-1}|S_{<M-1}) - \beta I(Y_{M-1}; S_{M-1}|S_{<M-1}) .$$

The optimization problem of the first sensor equals the individual scalar optimization without any side-information at all, as described in Section 2. Subsequent sensors combine the instantaneous side-information of all previously transmitting sensors $s_{<m}$ with its observation $y_m$. The relevant mutual information and the compression rate of sensor $m$ are conditioned on $S_{<m}$ since broadcasting instantaneous side-information ensures all successive sensors to have access to $s_{<m}$ allowing Wyner–Ziv coding for generating $s_m$. Each optimization problem given in (12) and (13) can be solved by taking the derivative with respect to the mapping $p(s_m|y_m, s_{<m})$ and equating it to zero. This results in the implicit update equation

$$p(s_m|y_m, s_{<m}) = \frac{e^{-d_{\beta_m}(y_m, s_m, s_{<m})}}{\sum_{s_m} e^{-d_{\beta_m}(y_m, s_m, s_{<m})}}$$

with

$$d_{\beta_m}(y_m, s_m, s_{<m}) := \frac{1}{\beta_m} D_{K L}[p(x|y_m, s_{<m})||p(x|s_{<m})] - \log p(s_m|s_{<m}) .$$

As in the general GDIB algorithm, the implicit update equation in (14) can be solved using a Blahut–Arimoto like algorithm resulting in local optimal solutions.

Algorithmic pcCEO Solution for the Successive Broadcasting Protocol

After designing the mapping for the instantaneous side-information, the mapping $p(z_m|y_m, s_{<m})$ can be optimized, again by means of the IB principle. Therefore, the original GDIB algorithm is modified to exploit the broadcasted instantaneous side-information, defining the GDIB-BC algorithm. The optimization problem for each sensor is given as

$$L^{(1)}_{GDIB-BC} = I(X; Z_1) - \beta_1 I(Y_1; Z_1)$$

$$\vdots$$

$$L^{(M)}_{GDIB-BC} = I(X; Z_M|Z_{<M}) - \beta_M I(Y_M, S_{<M}; Z_M|Z_{<M}).$$

The main difference to the original GDIB optimization problem in (6) and (7) lies in the definition of the compression rate $I(Y_m, S_{<m}; Z_m|Z_{<m})$ which emerges from the combination of the observation $y_m$ and the instantaneous side-information $s_{<m}$. Taking the derivative of the optimization problem for sensor $m$ with respect to the mapping $p(z_m|y_m, s_{<m})$ and equating it to zero delivers

$$p(z_m|y_m, s_{<m}) = \frac{e^{-d_{\beta_m}(y_m, z_m, s_{<m})}}{\sum_{z_m} e^{-d_{\beta_m}(y_m, z_m, s_{<m})}}$$
with
\[
d_{\beta_m}(y_m, z_m, s_{<m}) := E_{Z_{<m}|y_m,s_{<m}} \left[ \frac{1}{\hat{p}_m} \right].
\]
\[
D_{KL}[p(x|y_m, s_{<m}, z_{<m}) \| p(x|z_{<m})] - \log p(z_m|z_{<m}) \right].
\]

\[
(19)
\]

Again, the implicit update equation in (18) can be solved using a Blahut–Arimoto like algorithm. The extended Blahut–Arimoto like algorithm to design the mapping \( p(z_m|y_m, s_{<m}) \) of sensor \( m \) for a specific Lagrange parameter \( \beta_m \) and instantaneous side-information \( s_{<m} \) is given in Algorithm 1. The input pmf \( p(y_m-1, s_{<m-1}, z_{<m-1}, x) \) can be computed during the optimization of previous sensors. Lines 3 to 5 determine the required pmfs for the calculation of the KL-divergence of (19) in lines 6 to 9. The statistical distance of (19) is determined in lines 10 to 14. It is used to update the quantizer mapping \( p(z_m|y_m, s_{<m}) \) of sensor \( m \). This procedure is repeated until no significant changes of the desired mappings occur anymore. The algorithm returns the updated mapping \( p(z_m|y_m, s_{<m}) \) as well as the pmf \( p(y_m, s_{<m}, z_{<m}, x) \), which is used as an input for the successive sensor.

**Algorithm 1:** Extended Blahut–Arimoto algorithm for broadcast cooperating sensors.

**input**
\( m, p^{init}(z_m|y_m, s_{<m}), p(y_i, x), \beta_m, \epsilon \)

\( p(z_i|y_i, s_{<i}) \) \( \forall i \leq m \)

**recursively calculated input from previous sensor optimizations:**
\( p(y_m-1, s_{<m-1}, z_{<m-1}, x) \)

**output**
\( p(z_m|y_m, s_{<m}) \) \in \( [0,1] \), \( p(y_m, s_{<m}, z_{<m}, x) \)

1 **begin**

**initialization:**
\( p(z_m|y_m, s_{<m})^{(0)} \leftarrow p^{init}(z_m|y_m, s_{<m}), \)

\( l \leftarrow 1 \)

2 **do**

3 // calculate required pmfs (see Appendix A.1)

4 \( p(y_m, s_{<m}, z_{<m}, x) = \sum_{y_m-1} p(z_{m-1}|y_m-1, s_{<m-1})p(y_m|x) \)

5 \( \cdot p(y_m-1|y_m-1, s_{<m-1})p(y_m-1, s_{<m-1}, z_{<m-1}, x) \)

6 \( p(z_{<m}, x) = \sum_{z_{<m}} \sum_{y_m} p(z_m|y_m, s_{<m})p(y_m, s_{<m}, z_{<m}, x) \)

7 // KL-Divergence \( D_{KL} \) of (19)

8 \( p(x|y_m, s_{<m}, z_{<m}) = \frac{p(y_m, s_{<m}, z_{<m}, x)}{\sum_{x} p(y_m, s_{<m}, z_{<m}, x)} \)

9 \( p(x|z_{<m}) = \frac{\sum_{y_m} p(y_m, s_{<m}, z_{<m}), x)}{\sum_{x} p(y_m, s_{<m}, z_{<m}, x)} \)

10 \( D_{KL} = \sum_{x} p(x|y_m, s_{<m}, z_{<m}) \cdot \log \frac{p(x|y_m, s_{<m}, z_{<m})}{p(x|z_{<m})} \)

11 // distance \( d_{\beta_m}(y_m, z_m, s_{<m}) \) (19)

12 \( p(z_{<m}|y_m, s_{<m}) = \frac{\sum_{x} p(y_m, s_{<m}, z_{<m}, x)}{\sum_{x} p(y_m, s_{<m}, z_{<m}, x)} \)

13 \( p(z_{<m}|x) = \frac{\sum_{y_m} p(y_m, s_{<m}, z_{<m}, x)}{\sum_{x} p(y_m, s_{<m}, z_{<m}, x)} \)

14 \( p(z_m|z_{<m}) = \frac{\sum_{y_m} p(y_m, s_{<m}, z_{<m}, x)}{\sum_{x} p(y_m, s_{<m}, z_{<m}, x)} \)

15 // update quantizer \( p(z_m|y_m, s_{<m}) \)

16 \( p(z_m|y_m, s_{<m})^{l} = e^{-\frac{d_{\beta_m}(y_m, z_m, s_{<m})}{\sum_{x} e^{-d_{\beta_m}(y_m, z_m, s_{<m})}}} \)

17 \( l \leftarrow l + 1 \)

18 **while** \( D_{KL}[p^{l}(z_m|y_m, s_{<m}) \| p^{(l-1)}(z_m|y_m, s_{<m})] < \epsilon \)
The parameter $\beta_m$, which determines the compression rate at sensor $m$, has to be adjusted such that $I(Y_m;S_{\leq m};Z_m|Z_{<m}) \leq C_m$ is fulfilled. Similar to the original GDIB algorithm, the GDIB-BC algorithm has to be performed for each sensor and all possible optimization orders.

5.1.2. Evolution of Instantaneous Side-Information

Figure 7 illustrates the amount of instantaneous side-information available at the different sensors in a network of size $M = 6$ considering the broadcast of side-information. It depicts the relevant mutual information $I(X;S_{\leq m})$ versus the sensor number $m$ for different cardinalities $|S_m|$ and SNRs $\gamma_m$. The relevant signal is chosen to be a uniformly distributed 4-ASK signal leading to $|X| = 4$. As expected, the amount of available instantaneous side-information increases with each additional sensor for all $|S_m|$ and $\gamma_m$. To be more specific, the resolution and the quality of instantaneous side-information available at sensor $m$ increases with growing $m$. In the considered symmetric scenario, the amount of information $I(X;S_m|S_{<m})$ a sensor can contribute to $I(X;S_{\leq m})$ gets smaller for each additional sensor and the slopes of the curves decrease. Since one bit is not enough to represent the information of $|X| = 4$, the largest gain can be observed between $|S_m| = 2$ and $|S_m| = 4$. Increasing the cardinality further to $|S_m| = 8$ results only in a small additional improvement. Certainly, this observation depends on the relevant signal $X$ and can not be generalized.

![Figure 7. Available mutual information $I(X;S_{\leq m})$ for sensor $m$ in a network with $M = 6$ sensors and different cardinalities $|S_m|$ using the successive broadcasting protocol; $|X| = 4$, $|Y_m| = 64$.](image)

5.1.3. Performance for Different Network Sizes

Figures 8 and 9 illustrate the overall performance of the GDIB-BC approach when broadcasting instantaneous side-information for different network sizes. The relevant mutual information $I(X;Z)$ is depicted versus the number of sensors $M$ in the network. The gray colored area represents the non-achievable region, since $I(X;Z)$ cannot exceed $I(X;Y)$ due to the data-processing inequality. Both figures consider a scenario where all sensors in the network share the same channel to the common receiver with a fixed sum-rate $C_{\text{sum}}$ in an orthogonal way and a round robin fashion. Consequently, larger network sizes correspond to smaller individual capacities $C_m = \frac{C_{\text{sum}}}{M}$ for each forward link. The performance of partially cooperating sensors broadcasting instantaneous side-information (pcCEO-BC) is compared to the non-cooperative case (CEO) of Section 3 and the fully cooperative case (fcCEO) of Section 4. As already mentioned, these two scenarios provide upper and lower bounds. In general, it can be observed that increasing the number of sensors in the network also increases the overall relevant mutual information $I(X;Z)$. This
holds even for the case without cooperation, since each sensor applies Wyner–Ziv coding and exploits the mapping of previously designed quantizers as statistical side-information [14]. Independent of the cardinality $|S_m|$, the performance of the pcCEO-BC scenario is superior to the case without cooperation among sensors. This difference grows for larger network sizes because the amount of information $s_m$ has about the relevant variable $x$ increases. As expected from Figure 7, increasing the cardinality $|S_m|$ not only improves the relevant information $I(X; S_m)$, but also the overall performance measured by $I(X; Z)$. However, it can be observed that even for large $|S_m|$ there remains a gap to the fcCEO upper bound, especially for smaller network sizes or lower SNRs. This gap can be explained by the successive transmission protocol resulting in a gradually increasing amount of instantaneous side-information at the sensors. For instance, the first sensors does not profit at all from the partial cooperation in contrast to the fcCEO scenario where all sensors exploit almost the same amount of side-information.

Figure 8. Relevant mutual information $I(X; Z)$ versus the network size for a fixed sum-rate of $C_{\text{sum}} = 2.5 \text{ bit/s/Hz}$ and $C_m = \frac{C_{\text{sum}}}{M}$ using the successive broadcasting protocol with different cardinalities $|S_m|; |X| = 4, |Y_m| = 64, |Z_m| = 4$.

Figure 9. Relevant mutual information $I(X; Z)$ versus the network size for a fixed sum-rate of $C_{\text{sum}} = 4 \text{ bit/s/Hz}$ and $C_m = \frac{C_{\text{sum}}}{M}$ using the successive broadcasting protocol with different cardinalities $|S_m|; |X| = 4, |Y_m| = 64, |Z_m| = 4$. 
Considering the pmfs in Algorithm 1, it becomes obvious that larger networks might suffer from the curse of dimensionality. More precisely, pmfs like \( p(y_m, s_{<m}, z_{<m}, x) \) can become very large during the optimization for larger network sizes. Moreover, the mapping \( p(z_m|y_m, s_{<m}) \) also depends on the network size, i.e., this problem does not only occur during the optimization, but also when storing the already optimized mapping. This numerical issue is the reason why there is no result for \( |S_m| = 8 \) and a network size of \( M = 6 \) in Figures 8 and 9. In this case, it requires 2024 GiB (1 GiB = 1024 MiB, 1 MiB = 1024 KiB, 1 KiB = 1024 byte) just for storing a single instance of the pmf \( p(y_m, s_{<m}, z_{<m}, x) \).

### 5.2. Successive Point-to-Point Protocol

For larger network sizes, broadcasting side-information might not be feasible anymore, since the dimensions of the mappings \( p(z_m|y_m, s_{<m}) \) and \( p(s_m|y_m, s_{<m}) \) as well as intermediate pmfs used within the optimization become huge. In order to relax this curse of dimensionality, the successive way of cooperation is exploited and the instantaneous side-information of sensor \( m \) shall only be forwarded to the direct successor \( m + 1 \) as depicted in Figure 10. Hence, a sequential chain is established from the first to the last sensor leading to the Markov Model:

\[
p(x, y, z, s) = \prod_{m=1}^{M} p(z_m|y_m, s_{m-1}) p(s_m|y_m, s_{m-1}) p(y_m|x) p(x). \tag{20}
\]

Again, the instantaneous side-information is obtained by a deterministic mapping optimized by means of the information bottleneck principle, illustrated in Figure 11. With each step in the sequential chain, the information \( s_m \) has about the relevant signal \( x \) increases.

![Figure 10. Partially cooperative CEO scenario using successive point-to-point transmission of side-information.](image)

![Figure 11. Graphical illustration of IB fusion of two inputs to determine instantaneous side-information \( s_m \) (a) and the quantizer sensor output \( z_m \) (b) for a successive point-to-point transmission of instantaneous side-information.](image)
5.2.1. Generation of Point-to-Point Side-Information

Similar to the broadcast case, the design of \( p(s_m|y_m, s_{m-1}) \) is inspired by the original GDIB algorithm. The optimization problem can be formulated in a greedy manner as

\[
I^{(1)}_{\text{PTP-SIDE}} = I(x; S_1) - \beta I(y_1; S_1) \\
I^{(M-1)}_{\text{PTP-SIDE}} = I(x; S_{M-1}) - \beta I(y_{M-1}, S_{M-2}; S_{M-1}),
\]

where Equation (21) equals the individual scalar optimization without any side-information. Subsequent sensors combine the instantaneous side-information \( s_{m-1} \) sent by the previous sensor with its observation \( y_m \). In contrast to the broadcast case, the relevant mutual information is not conditioned on \( S_{<m} \) as in (12) and (13) because sensor \( m \) will only have access to \( s_{m-1} \) and not to indices of any other sensor. Therefore, Wyner-Ziv coding cannot be applied for exchanging instantaneous side-information with the successive point-to-point protocol. The optimization problems can be solved by taking the derivative with respect to the mapping \( p(s_m|y_m, s_{m-1}) \) and equating it to zero, resulting in the implicit update equation

\[
p(s_m|y_m, s_{m-1}) = \frac{e^{-d_{\beta m}(y_m, s_m, s_{m-1})}}{\sum_{s_m} e^{-d_{\beta m}(y_m, s_m, s_{m-1})}}
\]

with

\[
d_{\beta m}(y_m, s_m, s_{m-1}) := \frac{1}{\beta_m} \cdot D_{\text{KL}}[p(x|y_m, s_m)\|p(x|y_m)] - \log p(s_m).
\]

As in the broadcast case, using a Blahut-Arimoto like algorithm to solve the update Equation (23) results in local optimal solutions.

5.2.2. Algorithmic pcCEO Solution Applying the Successive Point-to-Point Protocol

After the optimization of the mapping for instantaneous side-information, the mapping \( p(z_m|y_m, s_{m-1}) \) can be designed by means of the information bottleneck principle. Inspired by the original GDIB algorithm, the optimization problem can be formulated as

\[
L^{(1)}_{\text{GDIB-PTP}} = I(x; Z_1) - \beta_1 I(y_1; Z_1) \\
\vdots \\
L^{(M)}_{\text{GDIB-PTP}} = I(x; Z_M|Z_{<M}) - \beta_M I(y_M, S_{M-1}; Z_M|Z_{<M}).
\]

The main difference to the original GDIB optimization problem in (6) and (7) now lies in the compression rate \( I(y_m, S_{m-1}; Z_m|Z_{<m}) \), which emerges from the combination of the instantaneous side-information \( s_{m-1} \) of the previous sensor \( m-1 \) and the observation \( y_m \) of sensor \( m \). Taking the derivative with respect to the mapping \( p(z_m|y_m, s_{m-1}) \) and equating it to zero, the optimization problem for sensor \( m \) can be solved, leading to the implicit update equation

\[
p(z_m|y_m, s_{m-1}) = \frac{e^{-d_{\beta m}(y_m, z_m, s_{m-1})}}{\sum_{z_m} e^{-d_{\beta m}(y_m, z_m, s_{m-1})}}
\]

with
The specific algorithm for a given sensor \( m \) and a Lagrange parameter \( \beta_m \) is given in Algorithm 2. The input pmfs \( p(z_i|y_{m}, s_{i-1}) \) for all \( i < m \) and \( p(s_i|z_{<i}, x) \) for all \( i < m \) as well as \( p(z_{<m}, x) \) are calculated in advance by previous sensor optimizations. Lines 3 to 7 calculates required pmfs as given the Appendix A.2. The KL-divergence is calculated in lines 8 to 11. Using this, the statistical distance \( d_{\beta_m}(z_m|y_m, s_{m-1}) \) and \( D_{\text{KL}}[p(x|y_{m}, s_{m-1}, z_{<m})] - \log p(z_m|z_{<m}) \) can be optimized using a Blahut–Arimoto-like algorithm.

Thus, the mapping \( p(z_m|y_m, s_{m-1}) \) can be optimized using a Blahut–Arimoto-like algorithm. The specific algorithm for a given sensor \( m \) and a Lagrange parameter \( \beta_m \) is given in Algorithm 2. The input pmfs \( p(z_i|y_{m}, s_{i-1}) \) for all \( i < m \) and \( p(s_i|z_{<i}, x) \) for all \( i < m \) as well as \( p(z_{<m}, x) \) are calculated in advance by previous sensor optimizations. Lines 3 to 7 calculates required pmfs as given the Appendix A.2. The KL-divergence is calculated in lines 8 to 11. Using this, the statistical distance \( d_{\beta_m}(z_m|y_m, s_{m-1}) \) of (28) can be calculated in lines 12 to 16, which is then used to update the quantizer mapping \( p(z_m|y_m, s_{m-1}) \). The algorithm stops if this mapping does not change significantly anymore during subsequent iterations. Finally, the output pmfs \( p(s_m|z_{<m}, x) \) and \( p(z_{<m}, x) \) need to be calculated in lines 21 to 25 for their usage in the optimization of the next sensor.

Similar to the original GDB algorithm, the optimization needs to be done for all possible optimization orders. A simple bisection search can be applied to find the rate-fulfilling parameter \( \beta_m \), such that \( I(Y_m, S_{m-1}; Z_m|Z_{<m}) \leq C_m \) holds.

5.2.3. Evolution of Instantaneous Side-Information

Figure 12 illustrates the amount of instantaneous side-information \( I(\mathcal{X}; S_m) \) at a specific sensor \( m \) in a network of size \( M = 6 \) using the successive point-to-point transmission protocol for different cardinalities \( |S_m| \). Obviously, \( I(\mathcal{X}; S_m) \) increases with each further sensor. The main difference to the broadcast case is that the instantaneous side-information provided to sensor \( m \) is represented by a single highly compressed index \( s_{m-1} \) with cardinality \( |S_{m-1}| \). While the resolution \( |S_{<m}| \) of the available instantaneous side-information \( s_{<m} \) increases with \( m \) in the broadcast case, it remains the same for the successive point-to-point protocol. Therefore, a higher cardinality \( |S_m| \) is required compared to the broadcast case to avoid additional compression losses.

![Figure 12](image-url)
5.2.4. Performance for Different Network Sizes

Figures 13 and 14 illustrate the overall performance of the pcCEO system with point-to-point exchanged instantaneous side-information where all sensors share the same channel to the common receiver with a fixed sum-rate $C_{\text{sum}} = \sum_{m=1}^{M} C_m$ in an orthogonal way and
a round robin fashion. Again, the black curves represent the non-cooperative CEO scenario and the fcCEO scenario. Hence, they serve as lower and upper bound, respectively. In general, the curves are very similar to those for broadcasting instantaneous side-information in Figures 8 and 9. Independent of the SNR or the sum-rate \( C_{\text{sum}} \), the relevant mutual information \( I(X; Z) \) increases for larger networks and even a single bit as instantaneous side-information \( |S_m| = 2 \) leads to slight improvements compared to the non-cooperative case. However, there still remains a gap to the fcCEO scenario even for large \( |S_m| \), which results from the successive communication strategy, since sensors at the beginning of the optimization chain can exploit no or little instantaneous side-information.

### Figure 13.
Relevant mutual information \( I(X; Z) \) versus the network size for a fixed sum-rate of \( C_{\text{sum}} = 2.5 \text{ bit/s/Hz} \) and \( C_m = \frac{C_{\text{sum}}}{M} \) using the successive point-to-point transmission protocol with different cardinalities \( |S_m|; |X| = 4, |Y_m| = 64, |Z_m| = 4 \).

### Figure 14.
Relevant mutual information \( I(X; Z) \) versus the network size for a fixed sum-rate of \( C_{\text{sum}} = 4 \text{ bit/s/Hz} \) and \( C_m = \frac{C_{\text{sum}}}{M} \) using the successive point-to-point transmission protocol with different cardinalities \( |S_m|; |X| = 4, |Y_m| = 64, |Z_m| = 4 \).

5.2.5. Performance for Different Sum-Rates

Figure 15 illustrates the influence of the sum-rate \( C_{\text{sum}} = \sum_{m=1}^{M} C_m \) for a scenario with \( M = 5 \) sensors. Naturally, larger sum-rates correlate with higher individual link
capacities. Again, CEO and fcCEO scenarios provide lower and upper bounds, respectively. For a cardinality of $|S_m| = 2$, only a small gain compared to the non-cooperative CEO scenario can be observed. However, the gain gets more and more significant with increasing $|S_m|$. Comparing the results to the upper fcCEO bound illuminates the loss due to limited available side-information at early transmitting sensors. The largest difference can be observed for sum-rates between $2 \leq C_{\text{sum}} \leq 4 \text{ bit/s/Hz}$.

![Figure 15](image-url)

**Figure 15.** Relevant mutual information $I(X; Z)$ versus sum-rate $C_{\text{sum}}$ with $C_m = \frac{C_m}{\gamma_m}$ using the successive point-to-point transmission protocol with different cardinalities $|S_m|$; $M = 5$, $|X| = 4$, $|Y_m| = 64$, $|Z_m| = 4$.

5.2.6. Asymmetric Scenarios

A very important part is the investigation of asymmetric scenarios. As the achievable relevant information $I(X; Z)$ of a non-cooperative CEO scenario is very sensitive to the optimization order, i.e., the Wyner–Ziv coding strategy, in asymmetric scenarios [14] the question arises if the exchange of instantaneous side-information can improve the robustness against bad optimization orders. Therefore, the same two asymmetric setups as in [14] are analyzed. Scenario 1 considers the case where sensors with low SNRs $\gamma_m$ have low link capacities $C_m$ while sensors with high SNRs $\gamma_m$ have high link capacities $C_m$. Scenario 2 considers the opposite case, where sensors with low SNRs have high link capacities and vice versa.

Figure 16 illustrates the relevant mutual information $I(X; Z)$ for all $M! = 24$ sensor permutations for a network of $M = 4$ sensors. The dots represent the results from [14] for a non-cooperative CEO scenario. Blue dots show Scenario 1 while the red dots represent Scenario 2. The results for the pcCEO scenario with successive point-to-point side-information exchange is depicted as bars.

Comparing the non-cooperative case with the successive point-to-point exchange of side-information for Scenario 1, we observe a slight increase of the overall relevant mutual information $I(X; Z)$ for partial cooperation and this particular scenario. Moreover, the influence of the Wyner–Ziv coding strategy (optimization order) becomes smaller due to cooperation. The performance for Scenario 2 is worse than the performance for Scenario 1, again for both the cooperative and the non-cooperative case. In this scenario, accurate measurements have to be strongly compressed in order to forward them to the common receiver while unreliable measurements cannot contribute much to the overall performance although they can be forwarded to the common receiver at high rates. However, the loss due to bad optimization orders is much lower for partial cooperation. A sensor with a bad forward channel and a high SNR can still forward its information to the next sensor, which
might have a better forward channel. Therefore, exchanging instantaneous side-information can improve the robustness against bad optimization orders.

### Figure 16.

Relevant mutual information $I(X; Z)$ for non-symmetric scenario with $M = 4$ sensors, SNRs $\gamma_m = [2, 4, 6, 8]$ dB and $|X| = 4, |Y_m| = 64, |Z_m| = 4$ using the successive point-to-point transmission protocol with $|S_m| = 8$.

#### 5.3. Two-Phase Transmission Protocol with Artificial Side-Information

Previous subsections revealed that partial cooperation by exchanging instantaneous side-information improves the overall performance. However, a gap to the fcCEO scenario still remains, and we claimed that the successive exchange of instantaneous side-information is the reason for this difference. Due to the sequential forwarding protocols considered so far, early sensors have no or little instantaneous side-information. They hardly profit from the cooperation as opposed to the full cooperation case where all sensors have access to the complete information. In order to substantiate this statement, a third transmission protocol consisting of two phases is considered. Inspired by the fcCEO scenario, the first cooperation phase is used to exchange instantaneous side-information between all sensors, while the transmission phase is used to forward the information to the common receiver in the usual way. The difference to the fcCEO scenario is that only compressed versions of the observations can be exchanged during the cooperation phase.

For simplicity, we assume that each sensor obtains the same instantaneous side-information represented by $s^*$, independent of its position in the optimization chain, see Figure 17. Moreover, we pursue the EXIT chart philosophy [33], where extrinsic information is artificially created to analyze the information exchange between decoders in concatenated coding schemes. In the pcCEO context, the artificial side-information can be interpreted as extrinsic information about the relevant signal $x$ being generated by adding AWGN to $x$. The noise variance is adapted to obtain a specific SNR $\gamma_{\text{extr}}$ or equivalently a desired mutual side-information $I(X; S^*)$. It has to be emphasized that $\gamma_{\text{extr}}$ can be chosen independently from the measurement SNRs at the sensors in order to obtain general conclusions. Since the instantaneous side-information is created artificially, $s^*$ is assumed to be independent of the indexes $y_m$, given the relevant signal $x$, i.e., $p(y_m, s^* | x) = p(y_m | x) p(s^* | x)$ holds. This simplifies the Markovian structure of the optimization problem which equals the one of the original CEO problem. With the same argumentation as in the original CEO problem,
we claim that the supermodularity holds and the greedy optimization structure is optimal. This model leads to the modified optimization problem

\[
L^{(1)}_{\text{GDB-TP}} = I(X; Z_1) - \beta_1 I(Y_1; S^*; Z_1) 
\]

\[
\vdots
\]

\[
L^{(M)}_{\text{GDB-TP}} = I(X; Z_M|Z_{<M}) - \beta_M I(Y_M; S^*; Z_M|Z_{<M}).
\]

The optimization problem can be solved using the same strategy as described in previous Sections 5.1 and 5.2. This leads to the implicit update equation

\[
p(z_m|y_m, s^*) = \frac{e^{-d_{\beta_m}(y_m, z_m, s^*)}}{\sum_{z_m} e^{-d_{\beta_m}(y_m, z_m, s^*)}}
\]

with

\[
d_{\beta_m}(y_m, z_m, s^*) := E_{z_{<m}|y_m, s^*} \left[ \frac{1}{\beta_m} \right] - D_{\text{KL}}[p(x|y_m, s^*; z_{<m})] - \log p(z_m|z_{<m}).
\]

Figure 17. Partially cooperative CEO scenario using the two-phase transmission protocol.

5.3.1. Performance of Two-Phase Transmission

Figure 18 illustrates the same experiment as in Figure 8 or Figure 13, but for the two-phase transmission protocol. The extrinsic information is chosen independent of the measurement SNR and has its own SNR $\gamma_{\text{ext}}$ represented by different colors in Figure 18. The cardinality of the extrinsic information is chosen as $|S^*| = 512$ to not introduce any compression losses. As before, the black dashed line represents the fcCEO scenario. The curve for $\gamma_{\text{ext}} = \gamma_m = 8$ dB represents the case where each sensor forwards instantaneous side-information whose quality corresponds to its measurement SNR. We observe the same performance as for the fcCEO scenario. This demonstrates that the remaining performance gap to the fcCEO scenario disappears completely for appropriate cooperation among sensors. Naturally, decreasing the SNR of the extrinsic information $\gamma_{\text{ext}}$ or equivalently $I(X; S^*)$ leads to a lower overall performance $I(X; Z)$. 
5.3.2. Influence of Extrinsic Information

The influence of extrinsic information is depicted in Figure 19 for $\gamma_m = 8$ dB and $\gamma_m = 3$ dB. Therefore, the overall relevant mutual information $I(\mathcal{X}; \mathcal{Z})$ is depicted versus the mutual information of the extrinsic information $I(\mathcal{X}; \mathcal{S}^*)$ for different network sizes. As before, all sensors share the same forward channel with $C_{\text{sum}} = 2.5$ bit/s/Hz and $C_m = \frac{C_{\text{sum}}}{M}$. Providing no extrinsic information, i.e., $I(\mathcal{X}; \mathcal{S}^*) = 0$ delivers the same result as the non-cooperative CEO scenario of Section 3. Naturally, enhancing the quality of the extrinsic information increases the overall relevant mutual information $I(\mathcal{X}; \mathcal{Z})$ up to the maximum of 2 bit/s/Hz.

Figure 18. Relevant mutual information $I(\mathcal{X}; \mathcal{Z})$ versus the network size for a fixed sum-rate of $C_{\text{sum}} = 2.5$ bit/s/Hz and $C_m = \frac{C_{\text{sum}}}{M}$ using a two-phase transmission protocol for artificially decoupled extrinsic information with different $\gamma_{\text{ext}}$. $\gamma_m = 8$ dB, $|\mathcal{X}| = 4$, $|\mathcal{Y}_m| = 64$, $|\mathcal{Z}_m| = 4$, $|\mathcal{S}^*| = 512$.

Figure 19. Relevant mutual information $I(\mathcal{X}; \mathcal{Z})$ versus extrinsic mutual information $I(\mathcal{X}; \mathcal{S}^*)$ for different network sizes and a fixed sum-rate of $C_{\text{sum}} = 2.5$ bit/s/Hz and $C_m = \frac{C_{\text{sum}}}{M}$ and $|\mathcal{X}| = 4$, $|\mathcal{Y}_m| = 64$, $|\mathcal{Z}_m| = 4$. 

6. Conclusions

This paper extends the non-cooperative CEO scenario allowing partial cooperation among sensors in the network. Therefore, it extends the algorithmic solution introduced in [14] for three different inter-sensor communication protocols: successive broadcasting, successive point-to-point communication and a two-phase transmission protocol. The first two protocols perform the exchange of instantaneous side-information and forwarding information to the common receiver at the same time step. Therefore, successive broadcasting exploits the instantaneous side-information of all previous sensors within the optimization chain. Since this may cause dimensionality problems during the optimization, the successive point-to-point transmission protocol forwards the instantaneous side-information only to the next sensor. It turns out that allowing this partial communication outperforms the non-cooperative compression where no communication among sensors is possible. Moreover, cooperative compression shows a larger robustness to suboptimal Wyner–Ziv coding strategies in asymmetric scenarios. However, a small performance gap to the fcCEO scenario still remains for the proposed successive broadcasting and successive point-to-point transmission protocols. This gap can be closed by a third protocol separating the cooperation from the forwarding phase and allowing each sensor to access the maximal available side-information. Although no formal conclusion about the optimality of the pcCEO can be drawn, the closeness to the fcCEO scenario in the investigated simulations reveals that solutions found by the proposed greedy algorithms are at least close to optimal.

Author Contributions: Conceptualization, S.S. and V.K.; formal analysis, S.S. and V.K.; investigation, S.S., A.D.A. and V.K.; methodology, S.S.; software, S.S. and A.D.A.; supervision, V.K.; validation, S.S.; visualization, S.S.; writing—original draft, S.S.; writing—review and editing, S.S. and V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the University of Rostock, and in part by the German Research Foundation (DFG) in the funding programme Open Access Publishing under Grant 325496636. Parts of the computation were done by using a compute cluster funded by DFG (grant: 440623123).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Optimization for Broadcasting Side-Information

According to Section 5.1, partial cooperation by broadcasting side-information requires the maximization of

\[ I^{(m)}_{GDB-BC} = I(X; Z_m | Z_{<m}) - \beta_m I(Y_m; S_{<m}; Z_m | Z_{<m}). \]  

(A1)

The variational problem can be solved by taking the derivative with respect to the mapping \( p(z_m | y_m, s_{<m}) \) and equating it to zero. In the following, the derivatives of both mutual information are given.
Appendix A.1.1. Derivative of I(Χ; Zₘ|Z_<ₘ)

The relevant mutual information in (A1) can be rewritten such that the desired mapping occurs explicitly.

\[ I(Χ; Zₘ|Z_<ₘ) = \mathbb{E}_{Χ,Zₘ,Z_<ₘ} \left[ \log \frac{p(zₘ|x,z_<ₘ)}{p(zₘ|z_<ₘ)} \right] \]

\[ = \sum_{zₘ} \sum_{z_<ₘ} \sum_{x,m} \sum_{y,m,s_<m} \frac{p(zₘ|yₘ,s_<m)p(yₘ,s_<m,z_<m,x)}{p(zₘ|z_<m,x)} \]

\[ \cdot \log \sum_{a∈Yₘ} \sum_{b∈S_<ₘ} p(zₘ|a,b)p(a,b|z_<m,x) \]

\[ - \sum_{zₘ} \sum_{z_<ₘ} \sum_{x,m} \sum_{y,m,s_<m} \frac{p(zₘ|yₘ,s_<m)p(yₘ,s_<m,z_<m,x)}{p(zₘ|z_<m,x)} \]

\[ \cdot \log \sum_{a∈Yₘ} \sum_{b∈S_<ₘ} p(zₘ|a,b)p(a,b|z_<m) \]  \hspace{1cm} (A2)

The derivative of (A2) delivers

\[ \frac{∂I(Χ; Zₘ|Z_<ₘ)}{∂p(zₘ|yₘ,s_<m)} \]

\[ = \sum_x \sum_{z_<ₘ} p(yₘ,s_<m,z_<m,x) \cdot \log p(zₘ|x,z_<m) \]

\[ + \sum_x \sum_{z_<ₘ} \left[ \sum_{yₘ,s_<m} \frac{p(zₘ|yₘ,s_<m)p(yₘ,s_<m,z_<m,x)}{p(zₘ|z_<m,x)} \right] \frac{p(yₘ,s_<m|zₘ,x,z_<m)}{p(zₘ|zₘ,x,z_<m)} \]

\[ - \sum_x \sum_{z_<ₘ} p(yₘ,s_<m,z_<m,x) \cdot \log p(zₘ|z_<m) \]

\[ - \sum_x \sum_{z_<ₘ} \left[ \sum_{yₘ,s_<m} \frac{p(zₘ|yₘ,s_<m)p(yₘ,s_<m,z_<m,x)}{p(zₘ|z_<m,x)} \right] \frac{p(yₘ,s_<m|z_<m)}{p(zₘ|z_<m)} \]

\[ = \sum_{z_<ₘ} p(yₘ,s_<m,z_<m) \sum_x p(x|yₘ,s_<m,z_<m) \log \frac{p(zₘ|x,z_<m)}{p(zₘ|z_<m)} . \]  \hspace{1cm} (A3)

Exploiting Bayes’ theorem, the argument of the logarithmic function can be rewritten

\[ \frac{p(zₘ|x,z_<m)}{p(zₘ|z_<m)} = \frac{p(x|z_<m)}{p(x|z_<m)} \cdot \frac{p(x|yₘ,s_<m,z_<m)}{p(x|yₘ,s_<m,z_<m)} \cdot \frac{p(x|yₘ,s_<m,z_<m)}{p(x|yₘ,s_<m,z_<m)} \cdot \frac{p(x|yₘ,s_<m,z_<m)}{p(x|yₘ,s_<m,z_<m)} . \]  \hspace{1cm} (A4)

The last ratio in (A4) can be dropped because it does not depend on \( p(zₘ|yₘ,s_<m) \) and its contribution can be incorporated into the Lagrange multiplier \( βₘ \). The insertion of the first ratio into (A3) yields the contribution of the derivative of the relevant mutual information

\[ \frac{∂I(Χ; Zₘ|Z_<ₘ)}{∂p(zₘ|yₘ,s_<m)} \rightarrow - \sum_{z_<ₘ} p(yₘ,s_<m,z_<m) \sum_x p(x|yₘ,s_<m,z_<m) \log \frac{p(x|yₘ,s_<m,z_<m)}{p(x|z_<m)} \]

\[ = - \sum_{z_<ₘ} p(yₘ,s_<m,z_<m) D_{KL}(p(x|yₘ,s_<m,z_<m)||p(x|z_<m)) . \]  \hspace{1cm} (A5)
Appendix A.1.2. Derivative of $I(Y_m, S_{<m}; Z_m | Z_{<m})$

With the definition of the conditional compression rate

$$I(Y_m, S_{<m}; Z_m | Z_{<m}) = E_{y_m, z_{<m}, S_{<m}} \left[ \log \frac{p(z_m | y_m, S_{<m})}{p(z_m | Z_{<m})} \right]$$

$$= \sum_{z_{<m}} \sum_{s_{<m}} p(z_m | y_m, S_{<m}) p(y_m, s_{<m}, z_{<m}) \log p(z_m | y_m, S_{<m})$$

$$= \sum_{z_{<m}} \sum_{s_{<m}} p(z_m | y_m, S_{<m}) p(y_m, s_{<m}, z_{<m}) \log \sum_{a \in Z_m} \sum_{b \in Z_{<m}} p(z_m | a, b) p(a, b | Z_{<m})$$  \hspace{1cm} (A6)

its derivative becomes

$$\frac{\partial I(Y_m, S_{<m}; Z_m | Z_{<m})}{\partial p(z_m | y_m, S_{<m})}$$

$$= \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) \log p(z_m | y_m, s_{<m})$$

$$+ \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) \frac{p(z_m | y_m, S_{<m})}{p(z_m | y_m, s_{<m})}$$

$$- \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) \log p(z_m | z_{<m})$$

$$- \sum_{z_{<m}} \left[ \sum_{y_{<m}} p(z_m | y_m, s_{<m}) p(y_m, s_{<m}, z_{<m}) \frac{p(y_m, s_{<m}, | Z_{<m})}{p(z_m | Z_{<m})} \right]$$

$$= \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) \log \frac{p(z_m | y_m, s_{<m})}{p(z_m | Z_{<m})} \right) .$$  \hspace{1cm} (A7)

Appendix A.1.3. Fusion of Derived Parts

Combining the result in (A5) and (A7) delivers the complete derivative

$$- \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) D_{KL}[p(x | y_m, s_{<m}, z_{<m}) | p(x | z_{<m})]$$

$$- \beta_m \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) \log p(z_m | y_m, S_{<m})$$

$$+ \beta_m \sum_{z_{<m}} p(y_m, s_{<m}, z_{<m}) \cdot \log p(z_m | Z_{<m}) = 0 .$$  \hspace{1cm} (A8)

Following the idea of Blahut and Arimoto [34], $p(x | z_{<m})$ and $p(z_m | z_{<m})$ are assumed to be independent of $p(z_m | y_m, s_{<m})$. With this trick, (A8) can be resolved with respect to the desired mapping of sensor $m$ leading to the implicit solution

$$p(z_m | y_m, s_{<m}) = \frac{e^{-d_{p_m}(y_m, z_m, s_{<m})}}{\sum_{z_{<m}} e^{-d_{p_m}(y_m, z_m, s_{<m})}}$$  \hspace{1cm} (A9)

with

$$d_{p_m}(y_m, z_m, s_{<m})$$

$$:= \sum_{z_{<m}} p(z_{<m} | y_m, s_{<m}) \left[ \frac{1}{\beta_m} D_{KL}[p(x | y_m, s_{<m}, z_{<m}) | p(x | z_{<m})] - \log p(z_m | Z_{<m}) \right]$$

$$= E_{z_{<m} | y_m, s_{<m}} \left[ \frac{1}{\beta_m} D_{KL}[p(x | y_m, s_{<m}, z_{<m}) | p(x | z_{<m})] - \log p(z_m | Z_{<m}) \right].$$  \hspace{1cm} (A10)
Appendix A.1.4. Calculating Required pmfs

This section covers the calculation of the required pmfs for the previously described algorithm. The first term in the KL divergence in (A10) is defined by

$$p(x|y_m, s_{<m}, z_{<m}) = \frac{p(x,y_m,s_{<m},z_{<m})}{\sum_x p(x,y_m,s_{<m},z_{<m})} \quad (A11)$$

where $p(x,y_m,s_{<m},z_{<m})$ is determined recursively as

$$p(x,y_m,s_{<m},z_{<m}) = \sum_{y_{m-1}} p(z_{m-1}|y_{m-1},s_{<m-1}) \cdot p(s_{m-1}|y_{m-1},s_{<m-1}) p(y_m|x) p(x,y_{m-1},s_{<m-1},z_{<m-1}) \quad (A12)$$

In (A12), the pmf $p(z_{m-1}|y_{m-1},s_{<m-1})$ represents the quantizer mapping of the previous sensor $m-1$ while $p(s_{m-1}|y_{m-1},s_{<m-1})$ denotes its the mapping for the instantaneous side-information. Hence, both have already been determined when optimizing sensor $m-1$ leading to a recursive computation. The second term in the KL divergence in (A10) is calculated by

$$p(x|z_{\leq m}) = \frac{p(z_{<m}, x)}{\sum_x p(z_{\leq m}, x)} \quad (A13)$$

where $p(z_{\leq m}, x)$ can be calculated by

$$p(z_{\leq m}, x) = \sum_{s_{\leq m}} \sum_{y_m} p(z_m|y_m, s_{<m}) p(x,y_m,s_{<m},z_{<m}) \quad (A14)$$

with $p(z_m|y_m, s_{<m})$ being the quantizer mapping of the current sensor. The joint pmf $p(x,y_m,s_{<m},z_{<m})$ has already been calculated in (A12). The argument of the logarithm in (A10) can be derived as

$$p(z_m|z_{<m}) = \sum_x p(z_{<m}, x) \frac{p(z_{\leq m}, x)}{\sum_{z_{\leq m}} \sum_x p(z_{\leq m}, x)} \quad (A15)$$

Finally, the pmf to calculate the conditional expectation in (A10) is determined as

$$p(z_{<m}|y_m,s_{<m}) = \frac{\sum_x p(x,y_m,s_{<m},z_{<m})}{\sum_{z_{<m}} \sum_x p(x,y_m,s_{<m},z_{<m})} \quad (A16)$$

Note that all above equations simplify to the scalar IB equations given in Section 2 when optimizing the first sensor.

Appendix A.2. Optimization for Point-to-Point Exchange of Side-Information

According to Section 5.2, partial cooperation using a point-to-point communication protocol requires the maximization of

$$L^{(m)}_{\text{GDB-PTP}} = I(X; Z_m|Z_{<m}) - \beta_m I(Y_m, S_{m-1}; Z_m|Z_{<m}). \quad (A17)$$

The variational problem can be solved by taking the derivative with respect to the mapping $p(z_m|y_m,s_{m-1})$ and equating it to zero. In the following, the derivatives of both mutual information are given.
Appendix A.2.1. Derivative of $I(\mathcal{X}; Z_m | Z_{<m})$

The relevant mutual information in (A17) can be rewritten such that the desired mapping occurs explicitly.

$$ I(\mathcal{X}; Z_m | Z_{<m}) = E_{\mathcal{X}, Z_m, Z_{<m}} \left[ \log \frac{p(z_m | x, z_{<m})}{p(z_m | Z_{<m})} \right] $$

$$ = \sum_{z_{<m}} \sum_{y_{m,s_{m-1}}} \sum_{y_{m,s_{m-1}}} p(z_m, y_{m,s_{m-1}}) p(y_{m,s_{m-1}}, z_{<m}, x) $$

$$ \cdot \log \sum_{a \in Y_m} \sum_{b \in S_{m-1}} p(z_m, a, b) p(a, b | z_{<m}, x) $$

$$ - \sum_{z_{<m}} \sum_{y_{m,s_{m-1}}} \sum_{y_{m,s_{m-1}}} p(z_m, y_{m,s_{m-1}}) p(y_{m,s_{m-1}}, z_{<m}, x) $$

$$ \cdot \log \sum_{a \in Y_m} \sum_{b \in S_{m-1}} p(z_m, a, b) p(a, b | z_{<m}) $$

(A18)

The derivative of (A18) delivers

$$ \frac{\partial I(\mathcal{X}; Z_m | Z_{<m})}{\partial p(z_m | y_{m,s_{m-1}})} = \sum_{x} \sum_{z_{<m}} p(y_{m,s_{m-1}, z_{<m}, x}) \cdot \log p(z_m | x, z_{<m}) $$

$$ + \sum_{x} \sum_{z_{<m}} \left[ \sum_{y_{m,s_{m-1}}} p(z_m | y_{m,s_{m-1}}) p(y_{m,s_{m-1}, z_{<m}, x}) \right] \frac{p(y_{m,s_{m-1}} | x, z_{<m})}{p(z_m | x, z_{<m})} $$

$$ = \sum_{x} \sum_{z_{<m}} p(y_{m,s_{m-1}, z_{<m}, x}) \cdot \log p(z_m | x, z_{<m}) $$

$$ - \sum_{x} \sum_{z_{<m}} \left[ \sum_{y_{m,s_{m-1}}} p(z_m | y_{m,s_{m-1}}) p(y_{m,s_{m-1}, z_{<m}, x}) \right] \frac{p(y_{m,s_{m-1}} | z_{<m})}{p(z_m | z_{<m})} $$

$$ = \sum_{x} \sum_{z_{<m}} p(y_{m,s_{m-1}, z_{<m}, x}) \cdot \log \frac{p(x | y_{m,s_{m-1}, z_{<m}, x})}{p(z_m | z_{<m})} $$

(A19)

Exploiting Bayes’ theorem the argument of the logarithmic function can be rewritten

$$ \frac{p(z_m | x, z_{<m})}{p(z_m | z_{<m})} = \frac{p(x | z_{<m})}{p(x | z_{<m})} $$

$$ = \frac{p(x | z_{<m})}{p(x | y_{m,s_{m-1}, z_{<m}, x})} \frac{p(x | y_{m,s_{m-1}, z_{<m}, x})}{p(z_m | z_{<m})}. $$

(A20)

The last ratio in (A20) can be dropped because it does not depend on $p(z_m | y_{m,s_{m-1}})$ and its contribution can be incorporated into the Lagrange multiplier $\beta_m$. The insertion of the first ratio into (A19) yields the contribution of the derivative of the relevant mutual information

$$ \frac{\partial I(\mathcal{X}; Z_m | Z_{<m})}{\partial p(z_m | y_{m,s_{m-1}})} = - \sum_{z_{<m}} p(y_{m,s_{m-1}, z_{<m}}) \sum_{x} p(x | y_{m,s_{m-1}, z_{<m}}) \log \frac{p(x | y_{m,s_{m-1}, z_{<m}})}{p(x | z_{<m})} $$

$$ = - \sum_{z_{<m}} p(y_{m,s_{m-1}, z_{<m}}) D_{KL}[p(x | y_{m,s_{m-1}, z_{<m}}) || p(x | z_{<m})]. $$

(A21)
Appendix A.2.2. Derivative of $I(Y_m, S_{m-1}; Z_m | Z_{<m})$

With the definition of the conditional compression rate

$$I(Y_m, S_{m-1}; Z_m | Z_{<m}) = E_{Y_m, Z_m, Z_{<m}, S_{m-1}} \left[ \log \frac{p(z_m | y_m, s_{m-1}, Z_{<m})}{p(z_m | Z_{<m})} \right]$$

$$= \sum_{z_m} \sum_{z_{<m}} \sum_{s_{m-1}} p(z_m | y_m, s_{m-1}) p(y_m, s_{m-1}, Z_{<m}) \log p(z_m | y_m, s_{m-1})$$

its derivative becomes

$$\frac{\partial I(Y_m, S_{m-1}; Z_m | Z_{<m})}{\partial p(z_m | y_m, s_{m-1})}$$

$$= \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) \log p(z_m | y_m, s_{m-1})$$

$$+ \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) \frac{p(z_m | y_m, s_{m-1})}{p(z_m | y_m, s_{m-1})}$$

$$- \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) \log p(z_m | Z_{<m})$$

$$- \sum_{z_{<m}} \sum_{s_{m-1}} \sum_{y_m} p(z_m | y_m, s_{m-1}) p(y_m, s_{m-1}, Z_{<m}) \frac{p(y_m, s_{m-1} | Z_{<m})}{p(z_m | Z_{<m})}$$

$$= \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) \log \frac{p(z_m | y_m, s_{m-1})}{p(z_m | Z_{<m})}. \quad (A23)$$

Appendix A.2.3. Fusion of Derived Parts

Combining the result in (A21) and (A23) delivers the complete derivative

$$- \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) D_{KL} \left[ p(x | y_m, s_{m-1}, Z_{<m}) \| p(x | Z_{<m}) \right]$$

$$- \beta_m \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) \log p(z_m | y_m, s_{m-1})$$

$$+ \beta_m \sum_{z_{<m}} p(y_m, s_{m-1}, Z_{<m}) \cdot \log p(z_m | Z_{<m}) = 0. \quad (A24)$$

Following the idea of Blahut and Arimoto [34], $p(x | Z_{<m})$ and $p(z_m | Z_{<m})$ are assumed to be independent of $p(z_m | y_m, s_{m-1})$. With this trick, (A24) can be resolved with respect to the desired mapping of sensor $m$ leading to the implicit solution

$$p(z_m | y_m, s_{m-1}) = \frac{e^{-\beta_m (y_m, z_m, s_{m-1})}}{\sum_{z_m} e^{-\beta_m (y_m, z_m, s_{m-1})}} \quad (A25)$$

with

$$\beta_m (y_m, z_m, s_{m-1})$$

$$:= \sum_{z_{<m}} p(z_{<m} | y_m, s_{m-1}) \left[ \frac{1}{\beta_m} D_{KL} \left[ p(x | y_m, s_{m-1}, Z_{<m}) \| p(x | Z_{<m}) \right] - \log p(z_m | Z_{<m}) \right]$$

$$= E_{Z_{<m} | y_m, s_{m-1}} \left[ \frac{1}{\beta_m} D_{KL} \left[ p(x | y_m, s_{m-1}, Z_{<m}) \| p(x | Z_{<m}) \right] - \log p(z_m | Z_{<m}) \right]. \quad (A26)$$
Appendix A.2.4. Calculating Required pmfs

This section covers the calculation of the required pmfs for the previously described algorithm. The first term in the KL divergence in (A26) is defined by

\[
p(x|y_m, s_{m-1}, z_{<m}) = \frac{p(x, y_m, s_{m-1}, z_{<m})}{\sum_x p(x, y_m, s_{m-1}, z_{<m})}
\]  
(A27)

with

\[
p(x, y_m, s_{m-1}, z_{<m}) = p(z_{m-1}, s_{m-1}|x, z_{<m-1}) p(x, y_m, z_{<m-1})
\]  
(A28)

The first term on the right hand side in (A28) can be calculated by

\[
p(z_{m-1}, s_{m-1}|x, z_{<m-1}) = \sum_{y_{m-1}, z_{m-2}} p(z_{m-1}|y_{m-1}, s_{m-2}) p(s_{m-1}|y_{m-1} s_{m-2}) p(y_{m-1}|x) p(s_{m-2}|x, z_{<m-1})
\]  
(A29)

where \( p(z_{m-1}|y_{m-1}, s_{m-2}) \) represents the mapping of previously designed quantizer \( m-1 \), \( p(s_{m-1}|y_{m-1}, s_{m-2}) \) denotes its mapping of instantaneous side-information and \( p(y_{m-1}|x) \) statistically describes the corresponding known measurement channel. The last term \( p(s_{m-2}|x, z_{<m-1}) \) has already been calculated when optimizing previous sensors (in this case the pre-predecessor) in a recursive way

\[
p(s_m|x, z_{<m}) = \sum_{y_m, s_{m-1}} p(s_m|y_m, s_{m-1}) p(y_m|x, z_m) p(s_{m-1}|x, z_{<m-1})
\]  
(A30)

with

\[
p(y_m|x, z_m) = \frac{p(y_m, x, z_m)}{\sum_{y_m} p(y_m, x, z_m)}
\]  
(A31)

and

\[
p(y_m, x, z_m) = \sum_{s_{m-1}} p(z_{m}|y_m, s_{m-1}) p(y_m|x) p(s_{m-1}|x) p(x)
\]  
(A32)

The second term on the right hand side in (A28) can be determined by

\[
p(x, y_m, z_{<m-1}) = p(y_m|x) p(z_{<m-1}, x)
\]  
(A33)

Again, the last term in (A33) is obtained as side product when optimizing the previous sensor by

\[
p(z_{<m}, x) = \sum_{y_{m-1}, s_{m-1}} p(z_{m}|y_m, s_{m-1}) p(z_{m-1}, s_{m-1}|x, z_{<m-1}) p(x, y_m, z_{<m-1})
\]  
(A34)

where \( p(z_{m-1}, s_{m-1}|x, z_{<m-1}) \) and \( p(x, y_m, z_{<m-1}) \) are already defined in (A29) and (A33), respectively. The second term in the KL divergence in (A26) is calculated by

\[
p(x|z_{<m}) = \frac{p(z_{<m}, x)}{\sum_x p(z_{<m}, x)}
\]  
(A35)

where \( p(z_{<m}, x) \) is already defined in (A34). The term in the logarithm in (A26) can be expressed as

\[
p(z_{m}|z_{<m}) = \frac{\sum_x p(z_{m}|z_{<m})}{\sum_x \sum_{z_{<m}} p(z_{m}|z_{<m})}
\]  
(A36)
Finally, the required pmf to calculate the conditional expectation in (A26) can be determined by
\[ p(z_{<m} | y_m, s_{m-1}) = \frac{\sum_s p(x, y_m, s_{m-1}, z_{<m})}{\sum_s \sum_{z_{<m}} p(x, y_m, s_{m-1}, z_{<m})} \]  
\[ (A37) \]
with \( p(x, y_m, s_{m-1}, z_{<m}) \) being already defined in (A28). Note that all above equations simplify to the scalar IB equations given in Section 2 when optimizing the first sensor. Moreover, when optimizing the second sensor, there is no pre-predecessor \( m - 2 \) and its impact on the above equations can be omitted.

References
1. Oohama, Y. The rate-distortion function for the quadratic Gaussian CEO problem. *IEEE Trans. Inf. Theory* 1998, 44, 1057–1070. https://doi.org/10.1109/18.669162.
2. Viswanathan, H.; Berger, T. The Quadratic Gaussian CEO Problem. *IEEE Trans. Inf. Theory* 1997, 43, 1549–1559. https://doi.org/10.1109/18.623151.
3. Chen, J.; Zhang, X.; Berger, T.; Wicker, S.B. An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem. *IEEE J. Sel. Areas Commun.* 2004, 22, 977–987. https://doi.org/10.1109/JSAC.2004.830888.
4. Prabhakaran, V.; Tse, D.; Ramachandran, K. Rate region of the quadratic Gaussian CEO problem. In Proceedings of the International Symposium on Information Theory (ISIT 2004), Chicago, IL, USA, 27 June–2 July 2004; p. 119. https://doi.org/10.1109/ISIT.2004.1365154.
5. Oohama, Y. Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder. *IEEE Trans. Inf. Theory* 2005, 51, 2577–2593. https://doi.org/10.1109/TIT.2005.850110.
6. Wagner, A.; Tavildar, S.; Viswanath, P. Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem. *IEEE Trans. Inf. Theory* 2008, 54, 1938–1961.
7. Ugur, Y.; Aguerri, I.E.; Zaidi, A. Vector Gaussian CEO problem under logarithmic loss. In Proceedings of the 2018 IEEE Information Theory Workshop (ITW), Guangzhou, China, 25–29 November 2018; pp. 1–5.
8. Ugur, Y.; Aguerri, I.E.; Zaidi, A. Vector Gaussian CEO Problem Under Logarithmic Loss and Applications. *IEEE Trans. Inf. Theory* 2020, 66, 4183–4202. https://doi.org/10.1109/TIT.2020.2972348.
9. Courtade, T.A.; Weissman, T. Multiterminal Source Coding Under Logarithmic Loss. *IEEE Trans. Inf. Theory* 2014, 60, 740–761.
10. Berger, T.; Zhang, Z.; Viswanathan, H. The CEO Problem [Multiterminal Source Coding]. *IEEE Trans. Inf. Theory* 1996, 42, 887–902. https://doi.org/10.1109/1.490552.
11. Eswaran, K.; Gastpar, M. Remote Source Coding under Gaussian Noise: Dueling Roles of Power and Entropy Power. arXiv 2018, arXiv:1805.06515v2.
12. Zaidi, A.; Aguerri, I.E. Distributed Deep Variational Information Bottleneck. In Proceedings of the 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 26–29 May 2020; pp. 1–5. https://doi.org/10.1109/SPAWC48557.2020.9154315.
13. Aguerri, I.E.; Zaidi, A. Distributed Variational Representation Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 2021, 43, 120–138. https://doi.org/10.1109/TPAMI.2019.2928806.
14. Steiner, S.; Kuehn, V.; Stark, M.; Bauch, G. Reduced-Complexity Optimization of Distributed Quantization Using the Information Bottleneck Principle. *IEEE Open J. Commun. Soc.* 2021, 2, 1267–1278. https://doi.org/10.1109/OJCOMS.2021.3083569.
15. Steiner, S.; Kuehn, V. Distributed Compression using the Information Bottleneck Principle. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. https://doi.org/10.1109/ICC42927.2021.9500324.
16. Estella Aguerri, I.; Zaidi, A. Distributed information bottleneck method for discrete and Gaussian sources. In Proceedings of the International Zurich Seminar on Information and Communication (IZS 2018) Proceedings, Zurich, Switzerland, 21–23 February 2018; ETH Zurich: Zurich, Switzerland, 2018; pp. 35–39.
17. Ugur, Y.; Aguerri, I.E.; Zaidi, A. A generalization of blahut-arimoto algorithm to compute rate-distortion regions of multiterminal source coding under logarithmic loss. In Proceedings of the 2017 IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, 6–10 November 2017; pp. 349–353.
18. Prabhakaran, V.; Ramachandran, K.; Tse, D. On the Role of Interaction Between Sensors in the CEO Problem. In Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2004.
19. Draper, S.; Wornell, G. Side Information Aware Coding Strategies for Sensor Networks. *IEEE J. Sel. Areas Commun.* 2004, 22, 966–976.
20. Simeone, O. Source and Channel Coding for Homogeneous Sensor Networks with Partial Cooperation. *IEEE Trans. Wirel. Commun.* 2009, 8, 1113–1117.
21. Permuter, H.; Steinberg, Y.; Weissman, T. Problems we can solve with a helper. In Proceedings of the 2009 IEEE Information Theory Workshop on Networking and Information Theory, Volos, Greece, 10–12 June 2009; pp. 266–270. https://doi.org/10.1109/ITWNIT.2009.5158584.

22. Tishby, N.; Pereira, F.C.; Bialek, W. The Information Bottleneck Method. In Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 22–24 September 1999; pp. 368–377.

23. Slonim, N. The Information Bottleneck Theory and Applications. Ph.D. Thesis, Hebrew University of Jerusalem, Jerusalem, Israel, 2002.

24. Hassanpour, S.; Wuebben, D.; Dekorsy, A. Overview and Investigation of Algorithms for the Information Bottleneck Method. In Proceedings of the SCC 2017—11th International ITG Conference on Systems, Communications and Coding, Hamburg, Germany, 6–9 February 2017.

25. Lewandowsky, J.; Bauch, G. Information-Optimum LDPC Decoders Based on the Information Bottleneck Method. *IEEE Access* 2018, *6*, 4054–4071.

26. Zeitler, G. Low-precision analog-to-digital conversion and mutual information in channels with memory. In Proceedings of the 48th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 29 September–1 October 2010; pp. 745–752.

27. Meidlinger, M.; Matz, G. On Irregular LDPC Codes with Quantized Message Passing Decoding. In Proceedings of the 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, 3–6 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5. https://doi.org/10.1109/SPAWC.2017.8227780.

28. Romero, F.; Kurkoski, B. LDPC Decoding Mappings That Maximize Mutual Information. *IEEE J. Sel. Areas Commun.* 2016, *34*, 2391–2401. https://doi.org/10.1109/JSAC.2016.2603708.

29. Zeitler, G. Low-Precision Quantizer Design for Communication Problems. Ph.D. Thesis, Technische Universitaet Muenchen, Muenchen, Germany, 2012.

30. Chen, D.; Kuehn, V. Alternating information bottleneck optimization for the compression in the uplink of C-RAN. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–7. https://doi.org/10.1109/ICC.2016.7510694.

31. Lewandowsky, J.; Stark, M.; Bauch, G. Information Bottleneck Graphs for Receiver Design. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 2888–2892. https://doi.org/10.1109/ISIT.2016.7541827.

32. Fujishige, S. *Submodular Functions and Optimization*; Elsevier: Amsterdam, The Netherlands, 2005.

33. ten Brink, S. Convergence Behavior of Iteratively Decoded Parallel Concatenated Codes. *IEEE Trans. Commun.* 2001, *49*, 1727–1737.

34. Cover, T.; Thomas, J. *Elements of Information Theory*, 2nd ed.; Wiley & Sons: New York, NY, USA, 2006.