Generalized Distributed Information Bottleneck for Fronthaul Rate Reduction at the Cloud-RANs Uplink

Shayan Hassanpour, Dirk Wübben, and Armin Dekorsy
Department of Communications Engineering
University of Bremen, 28359 Bremen, Germany
Email: {hassanpour, wuebben, dekorsy}@ant.uni-bremen.de

***Abstract*—The focus is on Wyner-Ziv type distributed fronthaul compression for the uplink of *Cloud Radio Access Networks* with single-hop topology to leverage the correlation among the received signals of neighboring Radio Access Points. For this, we highlight the relation between the problem at hand and the *Chief Executive Officer* source coding under logarithmic-loss distortion and depict that the achievability arguments from the latter verify addressing the postulated optimization. Subsequently, we derive the pertinent optimal solution and utilize that as the backbone of the *Generalized Distributed Information Bottleneck (G-DIB)* routine proposed here to tackle the considered remote source coding problem. As its name suggests, this novel approach in its very core spirit extends the State-of-the-Art *Distributed Information Bottleneck (DIB)* method by enabling individual rate constraints for various fronthaul links.**

I. Introduction

Cloud Radio Access Networks (Cloud-RANs) feature a number of advantages compared to the conventional cellular architecture, incorporating more efficient interference management and traffic handling [1], [2]. These fortes are secured mostly due to the centralized processing of data that is realized by transferring the baseband information to (and from) the cloud using a fronthaul network. However, it is well known that the fronthaul capacity limitations inflict an overwhelming bottleneck on the overall performance. Hence, advanced compression schemes are highly requested to ameliorate the aforementioned situation.

At the uplink, a practical (optimal under obliviousness [3]) strategy to tackle this problem is the *Compress-and-Forward* wherein *Radio Access Points (RAPs)* quantize their baseband signals ahead of forwarding them to the cloud-based *Central Processor (CP)*. *Point-to-point* and *multiterminal* compression are various techniques for realizing such a strategy [4]. In our previous investigation [5] we studied the point-to-point fronthaul compression for the uplink of single-hop Cloud-RAN topology in which all RAPs are directly connected to the CP. To address the system design problem we tailored the flexible structure of *Multivariate Information Bottleneck* [6]. In this work, we focus on the multiterminal compression. Contrary to the point-to-point compression, the multiterminal technique allows for the joint decompression of signals from adjacent RAPs to exploit their correlations. The focal idea behind is to perform the Wyner-Ziv type coding [7], [8] that is based on the consecutive leveraging of side information by decompressors for enabling the compressors to enhance their descriptions' granularity.

Explicitly, first we provide an information-theoretic validation regarding our considered design optimization. To that end, we exploit the single-letter characterization of the achievable *rate-information* region for the *Distributed Information Bottleneck (DIB)* setup presented in [9]. This result is, indeed, based on the study of the *Chief Executive Officer (CEO)* multiterminal source coding problem under *logarithmic-loss* [10]. Subsequently, by exploiting the *Variational Calculus* we derive the formal optimal solution per quantizer mapping of the individual RAPs and utilize its particular structure for devising an iterative algorithm, the *Generalized DIB (G-DIB)*, to tackle the considered design problem. This novel approach directly extends the State-of-the-Art *Blahut-Arimoto (BA)-type* routine [9] by allowing for a set of individual rate constraints for various fronthaul links rather than presuming merely a single constraint on the fronthauls' sum rate. Furthermore, we mathematically discuss the suboptimality of an available heuristic [11] aiming at attacking the same design problem and, later on, through some numerical investigations we substantiate its performance inferiority as well.

Notation: The random variable, a, with the probability mass function, $p(\mathsf{a})$, accepts particular realizations, a, from its domain, $\mathcal{A}$. With boldface counterparts, the same applies to the random vector, $\mathbf{a}$. $H(\cdot)$, $D_{\text{KL}}(\cdot\|\cdot)$, $D_{\text{JS}}^{\{\cdot,\cdot\}}(\cdot\|\cdot)$ and $I(\cdot;\cdot)$ denote Shannon's and relative entropy, Jensen-Shannon divergence and mutual information [12], [13]. By $\mathbf{a}_{1:j}$ is meant $\{\mathsf{a}_1,\cdots,\mathsf{a}_j\}$ and, more generally, by $\mathbf{a}_{\mathcal{J}}$ is meant $\{\mathsf{a}_j\,|j\in\mathcal{J}\}$ for a given set, $\mathcal{J}$, while $\mathbf{a}_{\mathcal{J}}^{-\ell}$ excludes a_ℓ from $\mathbf{a}_{\mathcal{J}}$. Further, $\mathcal{J}^c$ denotes the complement set of $\mathcal{J}$, and $[\cdot]^+=\max\{0,\cdot\}$.

II. Multiterminal Fronthaul Compression

A. System Model & Problem Formulation

Consider the depicted system model in Fig. 1 where a certain *User Equipment (UE)* signal, x, is presumed to be supported by a number, J, of RAPs. The noisy observation signals, $\mathsf{y}_{\pi(j)}$ for $j=1,\cdots,J$, must be compressed prior to getting transmitted over the respective fronthaul channels with limited capacities, $C_{\pi(j)}$, to the CP (with a fixed processing order, π). There, the decompression of the received signals is supposed to be done in a consecutive manner such that at the $\pi(j)$-th decompressor, the already retrieved signals, $\hat{\mathbf{y}}_{\pi(1):\pi(j-1)}$, can be exploited as side information. It is presumed that given the UE signal, x, the signals of different RAPs, $\mathsf{y}_{\pi(j)}$ and $\mathsf{y}_{\pi(\ell)}$ for $j\neq\ell$, are independent.

978-1-7281-8298-8/20/$31.00 ©2020 IEEE

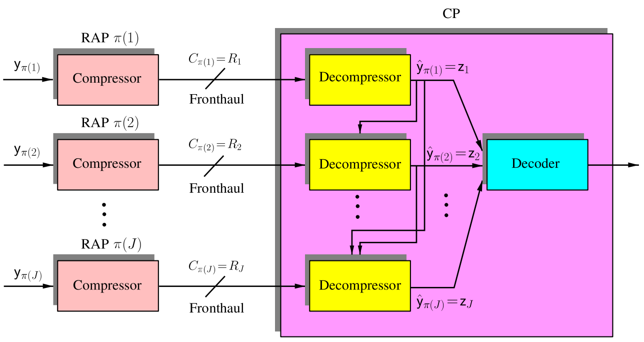


Fig. 1. Multiterminal fronthaul compression at the uplink of Cloud-RAN

Following the information-theoretic arguments [14] and under purely *Gaussian assumptions*, the relation between each RAP signal, $y_{\pi(j)}$, and its compressed counterpart, $\hat{y}_{\pi(j)}$, can be modeled as [4]

$$\hat{y}_{\pi(j)} = y_{\pi(j)} + q_{\pi(j)}, \tag{1}$$

wherein the quantization noise, $q_{\pi(j)}$, that is independent of $y_{\pi(j)}$ is considered as a complex Gaussian random variable with zero mean and the variance, $\Omega_{\pi(j)}$. A classical information-theoretic result ensures the attainability of the quantization error variance, $\Omega_{\pi(j)}$, utilizing the Wyner-Ziv type compression if the fronthaul capacity, $C_{\pi(j)}$, satisfies the inequality

$$I(y_{\pi(j)}; \hat{y}_{\pi(j)} | \hat{\mathbf{y}}_{\pi(1):\pi(j-1)}) \leq C_{\pi(j)} \ . \tag{2}$$

The respective system design problem is then formulated as a constrained optimization in which the overall transmission rate, $I(x; \hat{\mathbf{y}}_{\pi(1):\pi(J)})$, is intended to be maximized under a set of side constraints stipulated in (2) for a fixed π (that is subject to optimization). This maximization is performed over the set of all quantization error variances, $\mathcal{W} = \{\Omega_{\pi(1)}, \cdots, \Omega_{\pi(J)}\}$. Explicitly, one has to address the following problem

$$\mathcal{W}^* = \underset{\mathcal{W}:\, \forall j\ (2)\text{ applies}}{\text{argmax}} \ I(x; \hat{\mathbf{y}}_{\pi(1):\pi(J)}) \ . \tag{3}$$

Aside from Gaussian assumptions, in practice, the UE signal, x, is usually chosen from a *finite constellation*, $\mathcal{X}$. Moreover, as the received signals of the individual RAPs are first discretized and sampled, the same can also be presumed for $y_{\pi(j)}$, i.e., having a finite alphabet, $\mathcal{Y}_{\pi(j)}$. Now, the relation between each RAP signal, $y_{\pi(j)}$, and its compressed counterpart, $\hat{y}_{\pi(j)}$, can be described by a (possibly soft) mapping, $p(\hat{y}_{\pi(j)}|y_{\pi(j)})$. As a direct ramification, at the very first step the design problem (3) shall be modified such that now the aforementioned constrained maximization should be performed over the set of all quantizer mappings, $\mathcal{Q} = \{p(\hat{y}_{\pi(1)}|y_{\pi(1)}), \cdots, p(\hat{y}_{\pi(J)}|y_{\pi(J)})\}$. A basic question arising at this point is about the information-theoretic verification for this generalized discrete arrangement. In other words, one has to ponder whether it still makes sense to consider the same optimization for system design, i.e.,

$$\mathcal{Q}^* = \underset{\mathcal{Q}:\, \forall j\ (2)\text{ applies}}{\text{argmax}} \ I(x; \hat{\mathbf{y}}_{\pi(1):\pi(J)}) \ . \tag{4}$$

The clear answer to this basic question lies in the achievability arguments of the CEO problem under logarithmic-loss [10].

B. Apposite Achievability Arguments

In [10], the authors provided the single-letter characterization of the achievable *rate-distortion* region for the general setup of CEO problem [15] featuring J encoders under logarithmic-loss distortion. The roadmap there was to provide an inner- and a matching outer-bound for the achievable region and to show that the outer-bound is, indeed, a subset of the inner-bound, indicating that both are tight. This brilliant result was, later on, leveraged in [9] for single-letter characterization of the achievable *rate-information* region of the considered *Distributed Information Bottleneck* setup. The connecting trick was to depict that the relation between the distortion upper-bound, D, in the former and the information lower-bound, I, in the latter will be set through $D = H(x) - I$, with $H(x)$ denoting the entropy of the source (UE) signal, x. In this part, we minimally recap the main results of [9], [10] to demonstrate that both the inner- and outer-bounds from the achievability arguments are in full accordance with the stated optimization (4) regarding the system design for multiterminal compression at the uplink of Cloud-RAN (under obliviousness assumption at RAPs, i.e., no knowledge on the UE codebook).

Denoting by $\mathcal{RD}^{\text{log}}_{\text{CEO}}$ the achievable rate-distortion region for the CEO problem with J encoders under logarithmic-loss [10], the following propositions apply (ruling out time-sharing):

Proposition 1 (*Berger-Tung Inner-Bound* [16], [17]): Define $\mathcal{RD}^{\text{log, inn.}}_{\text{CEO}}$ as the set of all rate-distortion tuples $(D, R_1, \cdots, R_J)$ satisfying (symbol-wise description length, R_j, for j-th encoder)

$$I(\mathbf{y}_{\mathcal{J}}; \mathbf{z}_{\mathcal{J}} | \mathbf{z}_{\mathcal{J}^c}) \leq \sum_{j \in \mathcal{J}} R_j \tag{5a}$$

$$H(x | \mathbf{z}_{1:J}) \leq D, \tag{5b}$$

with $\mathcal{J} \subseteq \{1, \cdots, J\}$ for the source signal, x, the observations, $\mathbf{y}_{1:J}$, and a set of auxiliary random variables, $\mathbf{z}_{1:J}$, such that the Markov chain $z_j \leftrightarrow y_j \leftrightarrow x \leftrightarrow y_\ell \leftrightarrow z_\ell$ is formed for every pair $j, \ell \in \{1, \cdots, J\}$ and $j \neq \ell$. It holds $\mathcal{RD}^{\text{log, inn.}}_{\text{CEO}} \subseteq \mathcal{RD}^{\text{log}}_{\text{CEO}}$.

Proposition 2 (*Courtade-Weissman Outer-Bound* [10]): Define $\mathcal{RD}^{\text{log, out.}}_{\text{CEO}}$ as the set of all rate-distortion tuples $(D, R_1, \cdots, R_J)$ satisfying (symbol-wise description length, R_j, for j-th encoder)

$$\Big[\sum_{j \in \mathcal{J}} I(y_j; z_j | x) + H(x | \mathbf{z}_{\mathcal{J}^c}) - D\Big]^+ \leq \sum_{j \in \mathcal{J}} R_j \tag{6a}$$

$$H(x | \mathbf{z}_{1:J}) \leq D, \tag{6b}$$

with $\mathcal{J} \subseteq \{1, \cdots, J\}$ for the source signal, x, the observations, $\mathbf{y}_{1:J}$, and a set of auxiliary random variables, $\mathbf{z}_{1:J}$, such that the Markov chain $z_j \leftrightarrow y_j \leftrightarrow x \leftrightarrow y_\ell \leftrightarrow z_\ell$ is formed for every pair $j, \ell \in \{1, \cdots, J\}$ and $j \neq \ell$. It holds $\mathcal{RD}^{\text{log}}_{\text{CEO}} \subseteq \mathcal{RD}^{\text{log, out.}}_{\text{CEO}}$.

Proposition 3 (*Tightness* [10]): It holds $\mathcal{RD}^{\text{log, out.}}_{\text{CEO}} \subseteq \mathcal{RD}^{\text{log, inn.}}_{\text{CEO}}$, indicating both are tight, i.e., $\mathcal{RD}^{\text{log}}_{\text{CEO}} = \mathcal{RD}^{\text{log, inn.}}_{\text{CEO}} = \mathcal{RD}^{\text{log, out.}}_{\text{CEO}}$.

Substituting D by $H(x) - I$, the respective inner- and outer-bounds for $\mathcal{RI}_{\text{DIB}}$, i.e., the achievable rate-information region for the DIB problem [9], are directly obtained. Explicitly, using $I(x; \mathbf{z}_{1:J}) = H(x) - H(x|\mathbf{z}_{1:J})$, (5b) and (6b) will be replaced by

$$H(x|\mathbf{z}_{1:J}) \leq H(x) - I \ \Rightarrow \ I \leq I(x; \mathbf{z}_{1:J}) \ . \tag{7}$$

Analogously, (6a) will be rewritten as

$$\Big[\sum_{j\in\mathcal{J}} I(\mathsf{y}_j;\mathsf{z}_j|\mathsf{x}) - I(\mathsf{x};\mathbf{z}_{\mathcal{J}^c}) + I\Big]^+ \leq \sum_{j\in\mathcal{J}} R_j\ . \tag{8}$$

Focusing on the presented inner-bound, for the special case of $\mathcal{J}=\{1,\cdots,J\}$ the DIB-adapted version of (5) boils down to

$$I(\mathbf{y}_{1:J};\mathbf{z}_{1:J}) \leq \sum_{j=1}^{J} R_j \tag{9a}$$

$$I \leq I(\mathsf{x};\mathbf{z}_{1:J})\ . \tag{9b}$$

Through the substitutions $\mathsf{z}_j = \hat{\mathsf{y}}_{\pi(j)}$ and $R_j = C_{\pi(j)}$ and with the application of the chain rule for mutual information, the sum over all individual constraints in (2) brings about (9a). Naturally, the goal is then to maximize the (achievable) upper-bound in (9b). This is fully aligned with the stated design optimization (4).

Now, we consider the presented outer-bound. Rewriting (8) as

$$I \leq \sum_{j\in\mathcal{J}} \big(R_j - I(\mathsf{y}_j;\mathsf{z}_j|\mathsf{x})\big) + I(\mathsf{x};\mathbf{z}_{\mathcal{J}^c}), \tag{10}$$

with two extreme choices of $\mathcal{J}=\varnothing$ and $\mathcal{J}=\{1,\cdots,J\}$, it is directly inferred that

$$I \leq \min\left\{ I(\mathsf{x};\mathbf{z}_{1:J}), \sum_{j=1}^{J} R_j - \sum_{j=1}^{J} I(\mathsf{y}_j;\mathsf{z}_j|\mathsf{x}) \right\}\ . \tag{11}$$

This, in turn, justifies the definition of *information-rate function* in [9] for the DIB setup. From the Markov chain $\mathsf{x}\leftrightarrow\mathbf{y}_{1:J}\leftrightarrow\mathbf{z}_{1:J}$ it can be deduced that the following holds

$$\begin{aligned}\sum_{j=1}^{J} I(\mathsf{y}_j;\mathsf{z}_j|\mathsf{x}) &= I(\mathbf{y}_{1:J};\mathbf{z}_{1:J}|\mathsf{x}) \\ &= I(\mathbf{y}_{1:J};\mathbf{z}_{1:J}) - I(\mathsf{x};\mathbf{z}_{1:J})\ .\end{aligned} \tag{12}$$

Through the same substitutions as previously mentioned and by assuming individual rate constraints in the form of (2) it is directly inferred that the second operand at the right-side of (11) becomes greater or equal than the first one. Consequently, (11) boils down to (9b) and, once again, it is straightly deduced that the provided outer-bound is in full accordance with the stated design optimization (4). All in all, the conducted analysis yields the information-theoretic verification regarding the validity of (4) when considering a purely discrete model. In the next part, employing the *Variational Calculus* we characterize the formal optimal solution per quantizer mapping of the individual RAPs. The obtained result will be utilized as the backbone of the devised iterative algorithm, the G-DIB, to tackle the design problem (4).

C. Characterization of the Optimal Solution

Henceforth, for simplicity we will presume $\pi(j)=j$ and also adhere to the aforementioned substitutions. Applying the method of *Lagrange Multipliers*, the constrained optimization (4) can be transformed into an unconstrained one (up to the validity of the pertinent quantizer mappings) by augmenting the objective as

$$\mathcal{Q}^* = \underset{\mathcal{Q}}{\operatorname{argmax}}\ I(\mathsf{x};\mathbf{z}_{1:J}) - \sum_{j=1}^{J} \lambda_j I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{z}_{1:j-1}), \tag{13}$$

with the *non-negative* Lagrange multiplier, λ_j, representing the counterpart of C_j in the original formulation (4). The following theorem provides a complete characterization of the optimal solution to the multiterminal compression problem (13).

Theorem 1 (*Optimal Solution per Mapping*): Assume the joint distribution, $p(\mathsf{x},\mathbf{y}_{1:J})$, and λ_j are given for all $j=1,\cdots,J$. The (local) quantizers, $\{p(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$, make a stationary point of $\mathcal{L}=I(\mathsf{x};\mathbf{z}_{1:J})-\sum_{j=1}^{J}\lambda_j I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{z}_{1:j-1})$ if and only if for each pair $(y_j,z_j)\in\mathcal{Y}_j\times\mathcal{Z}_j$

$$p(z_j|y_j) = \frac{p(z_j)}{\psi_{\mathsf{z}_j}(y_j,\beta_j)} \exp\Big(-d(y_j,z_j)\Big), \tag{14}$$

wherein $\psi_{\mathsf{z}_j}(y_j,\beta_j)$ is a partition function ensuring the validity of the pertinent conditional distribution and $\beta_j=\frac{1}{\lambda_j}$. Further, the relevant distortion, $d(y_j,z_j)$, is calculated as

$$\begin{aligned} d(y_j,z_j) = {} & \beta_j \sum_{\boldsymbol{z}_{1:J}^{-j}} p(\boldsymbol{z}_{1:J}^{-j}|y_j) D_{\mathrm{KL}}\big(p(\mathsf{x}|y_j,\boldsymbol{z}_{1:J}^{-j})\|p(\mathsf{x}|z_j,\boldsymbol{z}_{1:J}^{-j})\big) \\ & -\sum_{\boldsymbol{z}_{1:j-1}} p(\boldsymbol{z}_{1:j-1}|y_j) \log p(\boldsymbol{z}_{1:j-1}|z_j) \\ & -\beta_j \sum_{k=j+1}^{J} \frac{1}{\beta_k} \sum_{\boldsymbol{z}_{1:k}^{-j}} p(\boldsymbol{z}_{1:k}^{-j}|y_j) \log p(z_k|\boldsymbol{z}_{1:k-1})\ . \end{aligned} \tag{15}$$

Proof: Introducing a Lagrange multiplier, λ_{y_j}, per realization, $y_j\in\mathcal{Y}_j$, of the observation (RAP) signal, y_j, all the pertinent mappings' validity conditions are incorporated into the overall Lagrangian, $\mathcal{L}_{\text{Ov.}}$, being defined as

$$\mathcal{L}_{\text{Ov.}} = \mathcal{L} + \sum_{j=1}^{J}\sum_{y_j\in\mathcal{Y}_j} \lambda_{y_j}\Big(\sum_{z_j\in\mathcal{Z}_j} p(z_j|y_j) - 1\Big)\ . \tag{16}$$

Since $\mathcal{L}_{\text{Ov.}}$ is, in fact, a functional of all individual conditional distributions $\{p(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$, to come into a stationary point of it, its derivative w.r.t. every quantizer mapping, $p(\mathsf{z}_j|\mathsf{y}_j)$, must be equated to zero. For that, fixing $\{p(\mathsf{z}_\ell|\mathsf{y}_\ell)\ |\ell\neq j\}$, to obtain the functional derivative of $\mathcal{L}_{\text{Ov.}}$ w.r.t. $p(\mathsf{z}_j|\mathsf{y}_j)$, the pertinent derivatives of its individual components have to be calculated. Applying the chain rule for mutual information to decompose $I(\mathsf{x};\mathbf{z}_{1:J})$, it is discerned that

$$\frac{\delta I(\mathsf{x};\mathbf{z}_{1:J})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)} = \frac{\delta I(\mathsf{x};\mathsf{z}_j|\mathbf{z}_{1:J}^{-j})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}, \tag{17}$$

as fixing $\{p(\mathsf{z}_\ell|\mathsf{y}_\ell)\ |\ell\neq j\}$ also fixes the other appearing terms in the conducted decomposition. Hence, it applies

$$\begin{aligned} \frac{\delta I(\mathsf{x};\mathsf{z}_j|\mathbf{z}_{1:J}^{-j})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)} &= \frac{\delta\Big(H(\mathsf{z}_j|\mathbf{z}_{1:J}^{-j}) - H(\mathsf{z}_j|\mathsf{x},\mathbf{z}_{1:J}^{-j})\Big)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)} \\ &= p(y_j)\left[\sum_{\boldsymbol{z}_{1:J}^{-j}} p(\boldsymbol{z}_{1:J}^{-j}|y_j) \log\frac{1}{p(z_j|\boldsymbol{z}_{1:J}^{-j})} - 1\right] - \\ &\quad p(y_j)\left[\sum_{\boldsymbol{z}_{1:J}^{-j}} p(\boldsymbol{z}_{1:J}^{-j}|y_j) \sum_{x} p(x|y_j,\boldsymbol{z}_{1:J}^{-j}) \log\frac{1}{p(z_j|x,\boldsymbol{z}_{1:J}^{-j})} - 1\right] \\ &= p(y_j)\sum_{\boldsymbol{z}_{1:J}^{-j}} p(\boldsymbol{z}_{1:J}^{-j}|y_j) \sum_{x} p(x|y_j,\boldsymbol{z}_{1:J}^{-j}) \log\frac{p(x|\boldsymbol{z}_{1:J})}{p(x|\boldsymbol{z}_{1:J}^{-j})}\ . \end{aligned} \tag{18}$$

Next, it has to be noticed that

$$\frac{\delta\left(\sum_{\ell=1}^{J}\lambda_\ell I(\mathsf{y}_\ell;\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1})\right)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}=\frac{\delta\left(\sum_{\ell=j}^{J}\lambda_\ell I(\mathsf{y}_\ell;\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1})\right)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}, \quad (19)$$

as the first $j-1$ appearing summands in the numerator do not depend on $p(\mathsf{z}_j|\mathsf{y}_j)$. Further, noting the presumed Markov chain and via some intermediate reformulations, it holds

$$\frac{\delta I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{z}_{1:j-1})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}=\frac{\delta\Big(H(\mathsf{z}_j|\mathbf{z}_{1:j-1})-H(\mathsf{z}_j|\mathbf{z}_{1:j-1},\mathsf{y}_j)\Big)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}=$$

$$p(y_j)\left[\log\frac{p(z_j|y_j)}{p(z_j)}-\sum_{\boldsymbol{z}_{1:j-1}}p(\boldsymbol{z}_{1:j-1}|y_j)\log\frac{p(\boldsymbol{z}_{1:j-1}|z_j)}{p(\boldsymbol{z}_{1:j-1})}\right], \quad (20)$$

and in case of $j<\ell\leq J$

$$\begin{aligned}\frac{\delta I(\mathsf{y}_\ell;\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}&=\frac{\delta H(\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}-\frac{\delta H(\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1},\mathsf{y}_\ell)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}\\&=\frac{\delta H(\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}-\underbrace{\frac{\delta H(\mathsf{z}_\ell|\mathsf{y}_\ell)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}}_{0}\\&=\frac{\delta H(\mathsf{z}_\ell|\mathbf{z}_{1:\ell-1})}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}\\&=p(y_j)\sum_{\boldsymbol{z}_{1:\ell}^{\text{-}j}}p(\boldsymbol{z}_{1:\ell}^{\text{-}j}|y_j)\log\frac{1}{p(z_\ell|\boldsymbol{z}_{1:\ell-1})}\,.\end{aligned} \quad (21)$$

Moreover, it holds

$$\frac{\delta\Big(\sum_{\ell=1}^{J}\sum_{y_\ell\in\mathcal{Y}_\ell}\lambda_{y_\ell}\big(\sum_{z_\ell\in\mathcal{Z}_\ell}p(z_\ell|y_\ell)-1\big)\Big)}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}=\lambda_{y_j}\,. \quad (22)$$

Due to the positivity of $p(y_j)$ and by application of the required optimality condition, i.e., $\frac{\delta\mathcal{L}_{\text{Ov.}}}{\delta p(\mathsf{z}_j|\mathsf{y}_j)}=0$, it is directly deduced that

$$\begin{aligned}&-\sum_{\boldsymbol{z}_{1:J}^{\text{-}j}}p(\boldsymbol{z}_{1:J}^{\text{-}j}|y_j)D_{\text{KL}}\big(p(\mathsf{x}|y_j,\boldsymbol{z}_{1:J}^{\text{-}j})\|p(\mathsf{x}|z_j,\boldsymbol{z}_{1:J}^{\text{-}j})\big)\\&-\lambda_j\log\frac{p(z_j|y_j)}{p(z_j)}+\lambda_j\sum_{\boldsymbol{z}_{1:j-1}}p(\boldsymbol{z}_{1:j-1}|y_j)\log p(\boldsymbol{z}_{1:j-1}|z_j)\\&+\sum_{k=j+1}^{J}\lambda_k\sum_{\boldsymbol{z}_{1:k}^{\text{-}j}}p(\boldsymbol{z}_{1:k}^{\text{-}j}|y_j)\log p(z_k|\boldsymbol{z}_{1:k-1})+\tilde{\lambda}_{y_j}=0,\end{aligned} \quad (23)$$

with $\tilde{\lambda}_{y_j}$ being equal to

$$\begin{aligned}&-\lambda_j\sum_{\boldsymbol{z}_{1:j-1}}p(\boldsymbol{z}_{1:j-1}|y_j)\log p(\boldsymbol{z}_{1:j-1})+\frac{\lambda_{y_j}}{p(y_j)}\\&+\sum_{\boldsymbol{z}_{1:J}^{\text{-}j}}p(\boldsymbol{z}_{1:J}^{\text{-}j}|y_j)D_{\text{KL}}\big(p(\mathsf{x}|y_j,\boldsymbol{z}_{1:J}^{\text{-}j})\|p(\mathsf{x}|\boldsymbol{z}_{1:J}^{\text{-}j})\big)\,.\end{aligned} \quad (24)$$

Bringing the second summand in (23) to the other side of the equality, multiplying both sides by $\beta_j=\frac{1}{\lambda_j}$, exponentiating them and, eventually, multiplying by $p(z_j)$, one obtains

$$p(z_j|y_j)=p(z_j)\exp\Big(-d(y_j,z_j)+\beta_j\tilde{\lambda}_{y_j}\Big)\,. \quad (25)$$

Enforcing the validity condition, $\sum_{z_j}p(z_j|y_j)=1$, and noting that $\tilde{\lambda}_{y_j}$ does not depend on z_j, one can treat $\exp(-\beta_j\tilde{\lambda}_{y_j})$ as the respective normalization (partition) function, $\psi_{\mathsf{z}_j}(y_j,\beta_j)$, to come into the form given in (14). ∎

D. The G-DIB Algorithm

In this part, based on the particular format of the derived optimal solution per quantizer mapping of the individual RAPs, we propose the G-DIB algorithm to jointly design the local quantizers for the considered problem (13). It should be noticed that the provided solution (14) has an *implicit* form as the relevant distortion, $d(y_j,z_j)$, on its right-hand side is a functional of all involved quantizer mappings, $\{p(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$. This indicates that, mathematically, (14) can be viewed as (f_j denoting a functional)

$$p(\mathsf{z}_j|\mathsf{y}_j)=f_j\big(p(\mathsf{z}_1|\mathsf{y}_1),\cdots,p(\mathsf{z}_J|\mathsf{y}_J)\big), \quad (26)$$

for the j-th RAP. Going through all different RAPs then leads to the following non-linear system of equations

$$\begin{cases}p(\mathsf{z}_1|\mathsf{y}_1)=f_1\big(p(\mathsf{z}_1|\mathsf{y}_1),\cdots,p(\mathsf{z}_J|\mathsf{y}_J)\big)\\p(\mathsf{z}_2|\mathsf{y}_2)=f_2\big(p(\mathsf{z}_1|\mathsf{y}_1),\cdots,p(\mathsf{z}_J|\mathsf{y}_J)\big)\\\quad\vdots\\p(\mathsf{z}_J|\mathsf{y}_J)=f_J\big(p(\mathsf{z}_1|\mathsf{y}_1),\cdots,p(\mathsf{z}_J|\mathsf{y}_J)\big)\end{cases}, \quad (27)$$

directly extending the structure of the *Multivariate Fixed-Point* system [18] to the field of functionals wherein the functions of multiple variables are substituted by the functionals of multiple mappings. Consequently, the conventional iterative methods can be directly applied to solve (27) as well. Following the key idea behind the *Gauss-Seidel* method, here, we propose establishing an *asynchronous* (sequential) iterative update procedure wherein individual quantizer mappings of different RAPs are actualized consecutively. Therefore, to update the j-th quantizer mapping, $p(\mathsf{z}_j|\mathsf{y}_j)$, one can directly apply the already available updated versions of the quantizer mappings from its preceding RAPs.

The *Generalized DIB (G-DIB)* algorithm with the provided pseudo-code in Alg. 1 proceeds as follows: Commencing with a set of random (valid) mappings, $\{p^{(0)}(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$, for each pair, $(y_j,z_j)\in\mathcal{Y}_j\times\mathcal{Z}_j$, the updates are executed (till convergence by $\varepsilon\ll 1$, or fulfillment of a stopping criterion by $i_{\max}$) via

$$p^{(i+1)}(z_j|y_j)=\frac{p^{(i)}(z_j)}{\psi_{\mathsf{z}_j}^{(i+1)}(y_j,\beta_j)}\exp\big(-d^{(i)}(y_j,z_j)\big), \quad (28)$$

with, i, denoting the running index. The output probability, $p^{(i)}(z_j)$, and the pertinent relevant distortion, $d^{(i)}(y_j,z_j)$, in (28) are calculated by exerting the current versions of all involved quantizer mappings, $\{p^{(i)}(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$, to suitably marginalize the actualized joint distribution, $p^{(i)}(\mathsf{x},\mathbf{y}_{1:J},\mathbf{z}_{1:J})$, for which the presumed Markov chain implies the following factorization

$$p^{(i)}(\mathsf{x},\mathbf{y}_{1:J},\mathbf{z}_{1:J})=p(\mathsf{x},\mathbf{y}_{1:J})\prod_{j=1}^{J}p^{(i)}(\mathsf{z}_j|\mathsf{y}_j)\,. \quad (29)$$

As discussed, updates are performed asynchronously, i.e., when a certain mapping, $p(\mathsf{z}_j|\mathsf{y}_j)$, is chosen, the update is executed solely for this RAP's mapping and for all $\ell=1:J$ and $\ell\neq j$, $p^{(i+1)}(\mathsf{z}_\ell|\mathsf{y}_\ell)=p^{(i)}(\mathsf{z}_\ell|\mathsf{y}_\ell)$. Doing so, it is apparent that the update of z_j encompasses the implications of recent updates from all of its *preceding* compressed variables, $\mathsf{z}_{\ell'}$, with $\ell'=1:j-1$. To avoid poor local optima, this procedure is repeated with different starting points, $\{p^{(0)}(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$, with the best outcome retained.

Alg. 1 G-DIB for Cloud-RAN Multiterminal Compression

Input: $p(\mathsf{x}, \mathbf{y}_{1:J})$, $\beta_j = \frac{1}{\lambda_j} > 0$, $|\mathcal{Z}_j|$, $\varepsilon > 0$, $i_{\max} > 0$
Output: A (generally soft) partition z_j of $\mathcal{Y}_j$ into $|\mathcal{Z}_j|$ bins
Initialization: $i = 0$, random mappings $\{p^{(i)}(\mathsf{z}_j|\mathsf{y}_j)\ |j\}$
while $i \leq i_{\max}$ **do**
  for $j = 1:J$ **do**
    • $p^{(i)}(z_j) \leftarrow \sum_{y_j} p^{(i)}(z_j|y_j)p(y_j) \quad \forall z_j \in \mathcal{Z}_j$
    • find i-th update for all distributions involved in $d(\mathsf{y}_j, \mathsf{z}_j)$ by marginalizing w.r.t. $p(\mathsf{x}, \mathbf{y}_{1:J})\prod_{j'=1}^{J} p^{(i)}(\mathsf{z}_{j'}|\mathsf{y}_{j'})$
    • $p^{(i+1)}(z_j|y_j) \leftarrow \frac{p^{(i)}(z_j)}{\psi_{\mathsf{z}_j}^{(i+1)}(y_j,\beta_j)} \exp\big(-d^{(i)}(y_j, z_j)\big)$
    • $p^{(i+1)}(\mathsf{z}_\ell|\mathsf{y}_\ell) \leftarrow p^{(i)}(\mathsf{z}_\ell|\mathsf{y}_\ell) \quad \forall \ell = 1:J, \ \ell \neq j$
    • $i \leftarrow i+1$
  end for
  if $\forall j, \forall y_j : D_{\text{JS}}^{\{\frac{1}{2},\frac{1}{2}\}}\big(p^{(i)}(\mathsf{z}_j|y_j)\|p^{(i-J)}(\mathsf{z}_j|y_j)\big) \leq \varepsilon$ **then**
    Break
  end if
end while

E. Supplementary Mathematical Discussion

The design optimization (4) has already been attacked in [11]. Alas, the proposed solution was suboptimal. The core idea there was to decompose (4) into a series of subproblems regarding individual RAPs and to solve each of those simpler subproblems separately. The suboptimality of this approach lurks in this very separation. Specifically, fixing all the other quantizer mappings, $p(\mathsf{z}_\ell|\mathsf{y}_\ell)$ for $\ell = 1:J$ and $\ell \neq j$, the suggested routine in [11] solves the following problem to acquire the j-th quantizer mapping

$$p^{*}(\mathsf{z}_j|\mathsf{y}_j) = \underset{p(\mathsf{z}_j|\mathsf{y}_j):\, I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{z}_{1:j-1}) \leq R_j}{\operatorname{argmax}} I(\mathsf{x}; \mathbf{z}_{1:J}) \ . \tag{30}$$

The underlying problem with this formulation is the fact that, when considering the j-th RAP quantizer mapping, $p(\mathsf{z}_j|\mathsf{y}_j)$, although the other RAP mappings are presumed to be fixed, still this does not imply that one can simply skip those constraints regarding $I(\mathsf{y}_k; \mathsf{z}_k|\mathbf{z}_{1:k-1})$ for $k > j$ as these terms are not fixed and, indeed, depend on $p(\mathsf{z}_j|\mathsf{y}_j)$. Obviously, this deviation from the optimal solution becomes more severe when dealing with larger numbers of RAPs. Contrarily, the conducted derivation here does not skip any of the active constraints when treating various RAPs. Thus, it correctly addresses the pertinent problem.

A special case of the design optimization (4) in which instead of stipulating J rate constraints for individual RAPs, only one constraint over the sum rate, i.e., $I(\mathbf{y}_{1:J}; \mathbf{z}_{1:J})$ is imposed has already been considered in [9]. Specifically, introducing

$$\mathcal{F} = H(\mathsf{x}|\mathbf{z}_{1:J}) + s\sum_{j=1}^{J} I(\mathsf{y}_j; \mathsf{z}_j) + H(\mathsf{x}|\mathsf{z}_j), \tag{31}$$

the heart of the proposed BA-type algorithm there was to solve

$$\mathcal{Q}^{*} = \underset{\mathcal{Q}}{\operatorname{argmax}} \ -\mathcal{F}, \tag{32}$$

with the *non-negative* parameter, s. After some reformulations, it can be shown that (32) is equivalent to

$$\mathcal{Q}^{*} = \underset{\mathcal{Q}}{\operatorname{argmax}} \ I(\mathsf{x}; \mathbf{z}_{1:J}) - \left(\frac{s}{1+s}\right) I(\mathbf{y}_{1:J}; \mathbf{z}_{1:J}) \ . \tag{33}$$

The provided optimal solution in [9] for (33) follows the same format as (14) with

$$d(y_j, z_j) = D_{\text{KL}}\big(p(\mathsf{x}|y_j)\|p(\mathsf{x}|z_j)\big) + \frac{1}{s}\sum_{\boldsymbol{z}_{1:J}^{\bar{j}}} p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) D_{\text{KL}}\big(p(\mathsf{x}|y_j, \boldsymbol{z}_{1:J}^{\bar{j}})\|p(\mathsf{x}|z_j, \boldsymbol{z}_{1:J}^{\bar{j}})\big) \ . \tag{34}$$

The solution above is in full accordance with (15) when having only one $\lambda = \beta^{-1} = \frac{s}{1+s}$. To clearly perceive this, we start with rewriting (15) for this special case. It applies

$$\begin{aligned} d(y_j, z_j) = \beta &\sum_{\boldsymbol{z}_{1:J}^{\bar{j}}} p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) D_{\text{KL}}\big(p(\mathsf{x}|y_j, \boldsymbol{z}_{1:J}^{\bar{j}})\|p(\mathsf{x}|z_j, \boldsymbol{z}_{1:J}^{\bar{j}})\big) \\ &- \sum_{\boldsymbol{z}_{1:J}^{\bar{j}}} p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) \log p(\boldsymbol{z}_{1:j-1}|z_j) \\ &- \sum_{\boldsymbol{z}_{1:J}^{\bar{j}}} p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) \sum_{n=j+1}^{J} \log p(z_n|\boldsymbol{z}_{1:n-1}), \end{aligned} \tag{35}$$

where, compared to (15), the summations in the second and third terms are augmented as the corresponding log expressions do not depend on the added terms. Further, since all β_j for $j = 1, \cdots, J$ are required to be the same, they cancel out each other. After some intermediate reformulations, one obtains

$$\begin{aligned} d(y_j, z_j) = \beta &\sum_{\boldsymbol{z}_{1:J}^{\bar{j}}} p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) D_{\text{KL}}\big(p(\mathsf{x}|y_j, \boldsymbol{z}_{1:J}^{\bar{j}})\|p(\mathsf{x}|z_j, \boldsymbol{z}_{1:J}^{\bar{j}})\big) \\ &+ D_{\text{KL}}\big(p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j)\|p(\boldsymbol{z}_{1:J}^{\bar{j}}|z_j)\big) + H(\mathbf{z}_{1:J}^{\bar{j}}|y_j) \ . \end{aligned} \tag{36}$$

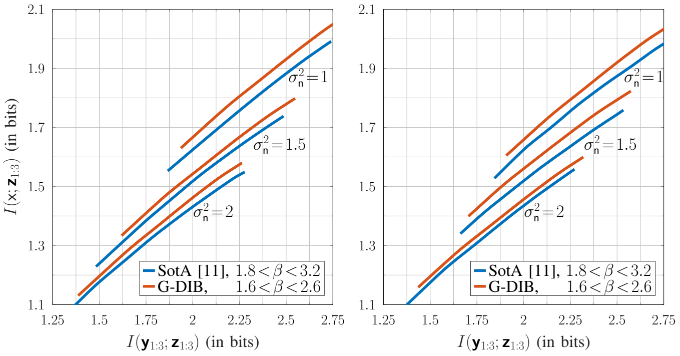
Note that the last term in (36) does not depend on z_j. Thus, it can be ignored (as it gets absorbed into the partition function). Hence, all one has to do at this point is to show that (34) is equivalent to (36) when ignoring its last summand. To that end, replacing $\frac{1}{s}$ by $\beta - 1$ in (34), the following has to be proven

$$\begin{aligned} D_{\text{KL}}\big(p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j)\|p(\boldsymbol{z}_{1:J}^{\bar{j}}|z_j)\big) = &D_{\text{KL}}\big(p(\mathsf{x}|y_j)\|p(\mathsf{x}|z_j)\big) \\ &- \sum_{\boldsymbol{z}_{1:J}^{\bar{j}}} p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) D_{\text{KL}}\big(p(\mathsf{x}|y_j, \boldsymbol{z}_{1:J}^{\bar{j}})\|p(\mathsf{x}|z_j, \boldsymbol{z}_{1:J}^{\bar{j}})\big) \ . \end{aligned} \tag{37}$$

Note that in (37) the term on the left, as well as the first term on the right, do not depend on $\mathbf{z}_{1:J}^{\bar{j}}$. Thus, the summation (including the prefactor) on its second term (right side of (37)) can be ignored as one can presume the same summation (including the prefactor) for the first term as well as the term on the left. Hence, one has to show that the KL divergence on the left side equals the difference of two KL divergences on the right side. The presumed Markov chain, $\mathsf{x} \leftrightarrow \mathbf{y}_{1:J} \leftrightarrow \mathbf{z}_{1:J}$, implies the following relations

$$\begin{aligned} p(\mathsf{x}|y_j)p(\boldsymbol{z}_{1:J}^{\bar{j}}|\mathsf{x}) &= p(\mathsf{x}|y_j, \boldsymbol{z}_{1:J}^{\bar{j}})p(\boldsymbol{z}_{1:J}^{\bar{j}}|y_j) \\ p(\mathsf{x}|z_j)p(\boldsymbol{z}_{1:J}^{\bar{j}}|\mathsf{x}) &= p(\mathsf{x}|z_j, \boldsymbol{z}_{1:J}^{\bar{j}})p(\boldsymbol{z}_{1:J}^{\bar{j}}|z_j) \ . \end{aligned} \tag{38}$$

Applying $D_{\text{KL}}(m_1 m_2\|n_1 n_2) = D_{\text{KL}}(m_1\|n_1) + D_{\text{KL}}(m_2\|n_2)$ on both sides of (38), indeed, concludes the proof of equivalence between the provided solution here and the one given in [9] for the special case of stipulating merely one constraint on the sum rate. This clearly indicates that the proposed approach here, in its core spirit, generalizes the SotA BA-type algorithm [9] by allowing for individual fronthaul rate constraints per RAP rather than imposing a single constraint on the compression sum rate.



(a) $|\mathcal{Z}_1|\!=\!|\mathcal{Z}_2|\!=\!|\mathcal{Z}_3|\!=\!4$ (b) $|\mathcal{Z}_1|\!=\!6, |\mathcal{Z}_2|\!=\!4, |\mathcal{Z}_3|\!=\!2$

Fig. 2. Transmission rate, $I(\mathsf{x};\mathbf{z}_{1:3})$, vs. compression sum rate, $I(\mathbf{y}_{1:3};\mathbf{z}_{1:3})$, equiprobable 8-ASK signaling ($\sigma_{\mathsf{x}}^2\!=\!21$), $J\!=\!3$ AWGN access channels with noise variance σ_{n}^2, convergence parameter $\varepsilon\!=\!10^{-4}$

III. Simulation Results

Here, we set about evaluating the performance of the G-DIB routine over a standard digital transmission scenario. For that, we consider an equiprobable signaling from a bipolar 8-*ASK (Amplitude Shift Keying)* constellation with $\sigma_{\mathsf{x}}^2 = 21$ to 3 RAPs. Access channels (UE to RAP) are modeled as *AWGN (Additive White Gaussian Noise)*, presuming identical noise variance, σ_{n}^2, for all RAPs. Specifically, 100 samples per access channel output (RAP signal) have been generated first and then clustered to $|\mathcal{Z}_j|$ bins for $j = 1, 2, 3$, once employing the G-DIB and once more the suggested routine in [11]. As these approaches are initialized randomly, for the sake of fairness, the same starting points are applied for both and the best outcomes are retained out of 100 trials. The obtained trade-offs between the overall transmission rate, $I(\mathsf{x};\mathbf{z}_{1:3})$, and the compression sum rate, $I(\mathbf{y}_{1:3};\mathbf{z}_{1:3})$, are illustrated in Fig. 2, when varying $\beta\!=\!\frac{1}{\lambda}$ over a particular range.

Specifically, for 3 different noise variances, $\sigma_{\mathsf{n}}^2 = 1, 1.5, 2$, the respective trade-off parameter is varied over $1.6 < \beta < 3.2$, once presuming a fully symmetric scenario with $|\mathcal{Z}_j| = 4$ for $j = 1, 2, 3$ and once more considering an asymmetric case wherein different cardinalities are set ($|\mathcal{Z}_1|\!=\!6, |\mathcal{Z}_2|\!=\!4, |\mathcal{Z}_3|\!=\!2$) for compressed signals of the individual RAPs. As a general trend, by increasing β the prefactor $\lambda\!=\!\frac{1}{\beta}$ of the compression sum rate, $I(\mathbf{y}_{1:3};\mathbf{z}_{1:3})$, is diminished and, consequently, the focus leans toward the overall transmission rate, $I(\mathsf{x};\mathbf{z}_{1:3})$. It can be observed that irrespective of the certain choice of model parameters, i.e., the access channel noise variance, σ_{n}^2, and the chosen cardinalities of compressed signals, $|\mathcal{Z}_j|$ for $j = 1, 2, 3$, the result of G-DIB surpasses the one achieved by the algorithm from [11]. This substantiates our concise mathematical discussion on the suboptimality of the suggested algorithm in [11] due to the separate (rather than joint) consideration of individual constraints per RAP's fronthaul rate.

As mathematically shown, in this special case (stipulating one constraint on the compression sum rate) the G-DIB algorithm and the BA-type routine in [9] yield identical results. Yet, once again, it has to be reminded that the G-DIB further extends the BA-type algorithm (its focal optimization) by allowing for J individual constraints corresponding to various RAPs' fronthaul rates rather than stipulating a single constraint on the compression sum rate.

IV. Summary

We focused on the multiterminal fronthaul compression at the uplink of Cloud-RANs with single-hop topology wherein the Wyner-Ziv type coding is presumed to leverage the correlation among signals of adjacent RAPs. For that, after formulating the design problem as a constrained optimization and discussing the apposite achievability arguments, we characterized the pertinent optimal solution. Subsequently, exploiting its specific format we proposed the G-DIB algorithm which, in its core spirit, extends the State-of-the-Art BA-type routine by enabling individual rate constraints per RAP rather than a single constraint on sum rate. Further, we discussed the suboptimality of an available algorithm aiming at attacking the same design problem and via numerical investigations corroborated its performance inferiority as well.

Acknowledgment

This was partly funded by the German ministry of education and research (BMBF) under grant 16KIS0720 (TACNET 4.0).

References

[1] D. Wübben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and Impact of Cloud Computing on 5G Signal Processing," *Special Issue "The 5G Revolution" of the IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 35–44, Nov. 2014.

[2] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-Constrained Cloud Radio Access Networks: Insights and Challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[3] I. E. Aguerri, A. Zaidi, G. Caire, and S. Shamai, "On the Capacity of Cloud Radio Access Networks with Oblivious Relaying," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4575–4596, Feb. 2019.

[4] S. Park, O. Simeone, O. Sahin, and S. Shamai, "Fronthaul Compression for Cloud Radio Access Networks: Signal Processing Advances Inspired by Network Information Theory," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[5] S. Hassanpour, D. Wübben, and A. Dekorsy, "A Novel Approach to Distributed Quantization via Multivariate Information Bottleneck Method," in *IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019.

[6] N. Slonim, N. Friedman, and N. Tishby, "Multivariate Information Bottleneck," *Neural Computation*, vol. 18, no. 8, pp. 1739–1789, Aug. 2006.

[7] 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, Jan. 1976.

[8] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed Source Coding for Sensor Networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–94, Sep. 2004.

[9] I. E. Aguerri and A. Zaidi, "Distributed Information Bottleneck Method for Discrete and Gaussian Sources," in *Int. Zurich Seminar on Information and Communication (IZS)*, Zurich, Switzerland, Feb. 2018.

[10] T. A. Courtade and T. Weissman, "Multiterminal Source Coding Under Logarithmic Loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[11] D. Chen, V. Kuehn, "Alternating Information Bottleneck Optimization for the Compression in the Uplink of C-RAN," in *IEEE Int. Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, July 2016.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., John Wiley & Sons, 2006.

[13] N. Slonim, "The Information Bottleneck: Theory and Applications," Ph.D. dissertation, Hebrew University of Jerusalem, Israel, 2002.

[14] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, 2011.

[15] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO Problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[16] T. Berger, "Multiterminal Source Coding," in *The Information Theory Approach to Communications*, Springer-Verlag, 1977.

[17] S.-Y. Tung, "Multiterminal Source Coding," Ph.D. dissertation, Cornell University, Ithaca, NY, USA, 1978.

[18] J. H. Mathews and K. D. Fink, *Numerical Methods Using MATLAB*, 4th ed., Pearson Prentice Hall Upper Saddle River, NJ, 2004.