

Forward-Aware Information Bottleneck-Based Vector Quantization: Multiterminal Extensions for Parallel and Successive Retrieval

Shayan Hassanpour^{ID}, *Member, IEEE*, Dirk Wübben^{ID}, *Senior Member, IEEE*,
and Armin Dekorsy^{ID}, *Senior Member, IEEE*

Abstract—Consider the following setup: Through a *joint* design, multiple observations of a remote data source shall be *locally* compressed before getting transmitted via several *error-prone*, rate-limited forward links to a (distant) processing unit. For addressing this specific instance of *multiterminal Joint Source-Channel Coding* problem, in this article, the foundational principle of the *Information Bottleneck* method is fully extended to obtain purely statistical design approaches, enjoying the *Mutual Information* as their fidelity criterion. Specifically, the forms of stationary points for two types of distributed compression schemes are characterized here. Subsequently, those acquired solutions are utilized as the centerpiece of the proposed generic, iterative algorithm, termed the *Multiterminal Forward-Aware Vector Information Bottleneck (M-FAVIB)*, for addressing the design optimizations. Leveraging an *unfolding* trick, it will be proven that both distributed compression schemes fall into the category of *Successive Upper-Bound Minimization*, ensuring their convergence to a stationary point. Eventually, the effectiveness of the proposed compression schemes will be substantiated as well by means of numerical investigations over some typical transmission scenarios.

Index Terms—CEO problem setup, distributed quantization, information bottleneck, mutual information, noisy channels.

I. INTRODUCTION

THE *Information Bottleneck (IB)* method was introduced in [1] as a task-based compression technique, and its initial applications are traced back to the context of *Unsupervised Learning*, where it was utilized as an information-theoretic approach towards *Cluster Analysis* [2]. To put it briefly, its main idea is to extract a *relevant* summary of data by bringing a statistically correlated variable into play that determines the meaning of *relevance*. The data set is then summarized in such a fashion that its information content w.r.t. that target variable is mostly retained. The inevitable trade-off between the *informativity* and *compactness* of the obtained result is established by employing two *Mutual Information (MI)* terms for gauging each aspect. Interested readers are

referred to [3]–[6] to gain further insights into this *Variational Principle*. Particularly, [5] provides an exhaustive discussion on the information-theoretic and learning features of the IB method and points out relevant connections to some classical problems like the Wyner-Ahlsvede-Körner problem [7], [8], the efficiency of investment information [9], and also the Wyner-Ziv setup [10] under the *Common Reconstruction (CR)* constraint [11] with the logarithmic-loss distortion [12].

The underlying setup of the IB method can be immediately exploited to address the problem of *Noisy Source Coding (NSC)* [13]–[16]. Specifically, utilizing the IB paradigm to compress an imperfect observation from a remote data source yields a purely statistical design approach. Besides, a certain instance of this framework tackles the problem of designing such quantizers that maximize the end-to-end transmission rate for a fixed source statistics, a criterion that is highly requested in all modern transmission systems. The footprints of the IB method are already traceable in various fields of advanced data communication schemes. Those applications encompass the design of receiver front ends’ *Analog-to-Digital* converters [17], polar codes construction [18], and efficient discrete decoding concepts [19], [20] with excellent performance, to mention a few.

The multiterminal extensions of the original IB method have been considered as well (see, e.g., [21]–[27]). The underlying scenario for such distributed schemes usually sets about dealing with several noisy observations from a (set of) remote source signal(s) that have to be compressed (with possibly different rates) following specific strategies such that, collectively, the retrieved signals after decompression preserve as much information as possible about the (set of) source signal(s). In practice, it occurs rather frequently that the compressed signals at the quantizers’ outputs have to be transmitted over second *error-prone* hops to be fed into a (distant) processing unit. Several instances for such scenarios incorporate the distributed inference sensor networks with imperfect connections to the fusion center [28], [29], the *Cloud-based Radio Access Networks (Cloud-RANs)* with noisy fronthaul links [30], [31], the cooperative relaying setups with the *Quantize-and-Forward* strategy [32], [33] and also devices with unreliable memories [34], [35]. The aforementioned assorted applications can be subsumed under a more general framework, the *Joint Source-Channel Coding (JSCC)* [36], [37], wherein the impacts of the imperfect forwarding of the quantizers’ outputs are taken into account

Manuscript received October 22, 2020; revised February 19, 2021 and May 26, 2021; accepted July 2, 2021. This work was partly funded by the German ministry of education and research (BMBF) under grants 16KIS0720 (TACNET 4.0) and 16KIS1180K (FunKI). The associate editor coordinating the review of this article and approving it for publication was C. Tian. (Corresponding author: Shayan Hassanpour.)

The authors are with the Department of Communications Engineering, University of Bremen, 28359 Bremen, Germany (e-mail: hassanpour@ant.uni-bremen.de; wuebben@ant.uni-bremen.de; dekorsy@ant.uni-bremen.de).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCOMM.2021.3097142>.

Digital Object Identifier 10.1109/TCOMM.2021.3097142



Fig. 1. The considered system model for uniterminal Joint Source-Channel Coding. Access and forward channels are assumed to be discrete and memoryless.

within the quantization design formulation. A closer look at the pertinent literature on this topic reveals some general approaches: Among others, several techniques have been proposed for judicious assignments of some binary codewords to the quantizers' output signals [38], [39] or, concentrating on the squared-error distortion, specific modifications have been suggested [40], [41] for adapting the conventional Lloyd algorithm [42] to the JSCC setup. Furthermore, in [28], [43], [44], as the fidelity criterion, the MI has been considered to acquire quantization schemes, maximizing the end-to-end transmission rate. An extensive review of different approaches has been provided in [45], [46]. Also, [47], [48] survey some distributed JSCC schemes over independent, multiple access, and broadcast channels.

Contributions: The full-format extension of the original IB method for NSC to *uniterminal JSCC* has been developed in [49], where an iterative algorithm, termed *Forward-Aware Vector Information Bottleneck (FAVIB)* has been proposed to tackle the underlying non-convex design problem. Herein, we concentrate on a particular instance of the *multiterminal JSCC* problem in which several observations from a given data source shall be compressed *locally* before getting transmitted over multiple noisy forward links to a (distant) processing unit. This can be reckoned as a straightforward extension of the well-known *Chief Executive Officer (CEO)* problem setup [50], in which the communications between the deployed agents and the CEO happen over several *error-prone* channels. As one tangible example out of many practical applications of such a distributed setup, one can think of achieving a highly reliable wireless transmission scheme by leveraging the spatial diversity obtained by the *joint* processing of the incoming signals from a densely employed network of radio access points with the overlapping coverage areas. Pursuing the IB philosophy, then we introduce the parallel and successive compression schemes, both enjoying a *joint* design of local quantizers and utilizing specific MI terms to quantify the *informativity* and *compactness* of their resultant outcomes. Subsequently, with a similar argument to [49], we characterize the form of stationary solutions regarding the considered distributed compression problems. Particularly, we find the stationary points obtained by the Lagrangian relaxation of the constrained optimization modeling of such problems. Thereupon, we propose a generic algorithm, the *Multiterminal FAVIB (M-FAVIB)*, that leverages the derived stationary points to address both design optimizations. Applying an unfolding trick, we also prove that both schemes fall into the *Successive Upper-Bound Minimization (SUM)* framework [51], ensuring that they converge to a stationary point. In a variety of practical occasions, the proposed methods become crucial, i.e., in applications where the separate employment of the iterative, modern error-correcting techniques on the noisy forward links will be prohibited by stringent latency constraints, or when dealing with the hardware imperfections

in which the separate utilization of those techniques results in a substantial overhead regarding the energy efficiency, giving rise to a plain waste of resources since they are utilized merely to take precautions against worst-case situations [35]. Further, it should be mentioned that from a sheer theoretical viewpoint, since Shannon's source and channel separation's optimality [52] does not hold in general [53], e.g., when working in a non-asymptotic blocklength regime, devising such JSCC schemes becomes relevant.

Outline: The uniterminal IB-based JSCC will be briefly discussed in Section II as a prelude to the multiterminal extensions. In Section III, a specific distributed setup that frequently appears in a diversity of applications is considered, and, over that, the parallel and successive compression schemes are introduced. Subsequently, the stationary points are characterized for both cases, and, in Section IV, a generic algorithm, the M-FAVIB, is presented to address the design optimizations together with the convergence proofs to the stationary points. The effectiveness of the devised compression schemes is corroborated in Section V by means of numerical investigations over typical transmission setups. Finally, a bare concise wrap-up is given in Section VI. The proofs have been relegated to the Appendix.

Notation: The discrete random variable, \mathbf{a} , with the probability mass function, $p(\mathbf{a})$, takes particular realizations, a , from its domain, \mathcal{A} . The same also applies to the random vector, $\mathbf{a}_{1:J} = [a_1, \dots, a_J]$ with boldface counterparts. $\mathbf{a}_{1:J}^j$ excludes a_j from $\mathbf{a}_{1:J}$. $H(\cdot)$, $D_{\text{KL}}(\cdot\|\cdot)$, $I(\cdot;\cdot)$ and $D_{\text{JS}}^{\{\cdot,\cdot\}}(\cdot\|\cdot)$ stand for Shannon's entropy, the *Kullback-Leibler (KL)* divergence [54], the MI, and the *Jensen-Shannon (JS)* divergence [3].

II. IB-BASED JSCC: A BRIEF OVERVIEW OF THE UNITERMINAL CASE

In [49], the *Noisy Source Coding (NSC)* scenario of the *Information Bottleneck (IB)* method [1] has been extended to the *Joint Source-Channel Coding (JSCC)* setup for quantization of a single noisy observation (scalar/vector-valued).¹ For that, the illustrated system model in Fig. 1 has been considered. A data source, \mathbf{x} , is observed via a discrete memoryless (access) channel with the transition probabilities, $p(\mathbf{y}|\mathbf{x})$. The observation, \mathbf{y} , is then compressed into the signal, \mathbf{z} , ahead of transmission over another discrete, memoryless (forward) channel with the transition probabilities, $p(\mathbf{t}|\mathbf{z})$, to a (distant) processing unit. To formulate the quantizer design problem for this setup and fully aligned with the fundamental principle of the IB paradigm, two *Mutual Information (MI)* terms are employed. On the one hand, the *relevant information*, $I(\mathbf{x}; \mathbf{t})$, which is the MI between the source, \mathbf{x} , and the forward channel

¹Although the entire study in this article can be directly applied to the *Vector Quantization (VQ)* case as already presented and discussed in [49], solely for the sake of simplicity, we will stick to the scalar-valued setup in the upcoming discussions.

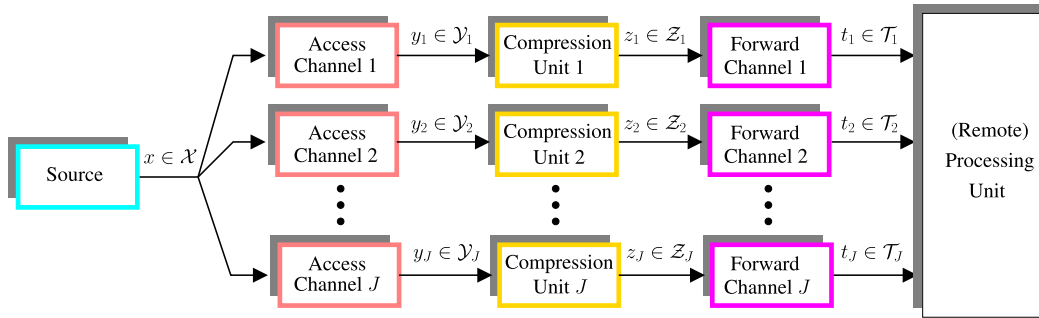


Fig. 2. Considered system model for multiterminal Joint Source-Channel Coding. All access and forward channels are assumed to be discrete and memoryless.

output signal, \mathbf{t} , is considered as the natural choice for gauging the *informativity* of the resultant outcome. On the other hand, the MI between the input and output of the compression unit, $I(y; z)$, called the *compression rate*, is chosen as the term that quantifies its *compactness*. Lower values of $I(y; z)$ indicate higher compression and vice versa. The quantizer design problem is then formulated by establishing a fundamental trade-off between both aspects. The goal is to derive the conditional distribution, $p^*(z|y)$, for which it applies

$$p^*(z|y) = \underset{p(z|y): I(y; z) \leq R}{\operatorname{argmax}} I(x; \mathbf{t}), \quad (1)$$

with $0 \leq R \leq \log_2 |\mathcal{Z}|$ bits, stipulating an upper-bound on the compression rate. By employing the method of *Lagrange Multipliers (LM)*, one can reformulate the design problem (1) as an unconstrained optimization (up to the pertinent mapping's validity)

$$p^*(z|y) = \underset{p(z|y)}{\operatorname{argmax}} I(x; \mathbf{t}) - \lambda I(y; z), \quad (2)$$

with $\lambda \geq 0$, denoting the counterpart relaxation of the upper-bound, R , in (1). Utilizing *Variational Calculus*, it has been shown in [49] that for stationary points of the objective functional in (2), it holds (for each pair $(y, z) \in \mathcal{Y} \times \mathcal{Z}$)

$$p(z|y) = \frac{p(z)}{\psi(y, \beta)} \exp\left(-\beta \sum_{t \in \mathcal{T}} p(t|z) D_{\text{KL}}(p(x|y) \| p(x|t))\right), \quad (3)$$

with $\beta = \frac{1}{\lambda}$ and $\psi(y, \beta)$, being a normalization function, ensuring the resultant mapping's validity. Further, an iterative routine, termed *Forward-Aware Vector Information Bottleneck (FAVIB)*, has been proposed in [49] that, indeed, carries out the *Fixed-Point Iteration* method on (3). The convergence of FAVIB to a local optimum of its objective functional has been proven as well, and it has been shown that this compression scheme enjoys inherent error-protection, capable of obviating the need for channel coding on the noisy forward link.

III. IB-BASED JSCC: PARALLEL & SUCCESSIVE MULTITERMINAL EXTENSIONS

We consider the illustrated system model in Fig. 2 for multiterminal extensions of the uniterminal JSCC setup. Explicitly, a data source, \mathbf{x} , is observed independently via J discrete, memoryless (access) channels. The outputs of these channels have to be *locally* compressed (yet following a *joint* design) and then transmitted over J discrete, memoryless (forward) links to a (distant) processing unit

that should figure out the source signal, \mathbf{x} . Analogous to the uniterminal scenario, we presume that the source distribution, $p(\mathbf{x})$, the access channels' transition probabilities, $p(y_j|\mathbf{x})$, and the forward channels' statistics, $p(\mathbf{t}_j|z_j)$, are available for all branches. We presume that the Markovian independence relation $\mathbf{x} \leftrightarrow y_j \leftrightarrow z_j \leftrightarrow \mathbf{t}_j$ applies per branch, j , and the counterpart signals of every two distinct branches are conditionally independent given the source signal, \mathbf{x} .

To formulate the design problem(s) through an analogous approach to the original IB framework, one should specify the responsible terms regarding both the *informativity* and *compactness* of the resultant outcomes. In this work, we naturally choose the end-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:J})$, as the term gauging the information preservation. In contrast, there is no natural, unique choice for the other side of the trade-off, and, indeed, different meaningful expressions can be applied. Herein, we consider two distinct constraint sets for stipulating the compactness of outcomes (extending the raised ideas in [31]). Then, for each specific choice, we characterize the form of stationary points regarding all the individual quantizer mappings. Those solutions will be utilized as the backbone of the proposed M-FAVIB algorithm to address both design optimizations for this multiterminal IB-based JSCC setup. We also provide the convergence proofs to the stationary points of the objective functionals.

A. Parallel Retrieval: Compression Without Side-Information

As the first choice of the imposed constraint set, we consider the case where individual branches are subject to different rate limitations, and from a compression's perspective, no side-information will be utilized when treating each particular observation, y_j , allowing for the pure parallel processing across branches. This fully discrete scenario has been analyzed in [25] for an identical setup to the one depicted in Fig. 2 with the major difference of presuming *ideal* (rate-limited) forward links. There, for characterizing the form of stationary points and tackling the design problem, the *Multivariate Information Bottleneck (MIB)* [21] framework was aptly tailored to the considered scenario. The conducted derivations here extend the provided results in [25] to this parallel IB-based JSCC setup. The design problem is then formulated as looking for the optimal set, $P^* = \{p^*(z_1|y_1), \dots, p^*(z_J|y_J)\}$, where

$$P^* = \underset{P: \forall j I(y_j; z_j) \leq R_j}{\operatorname{argmax}} I(\mathbf{x}; \mathbf{t}_{1:J}), \quad (4)$$

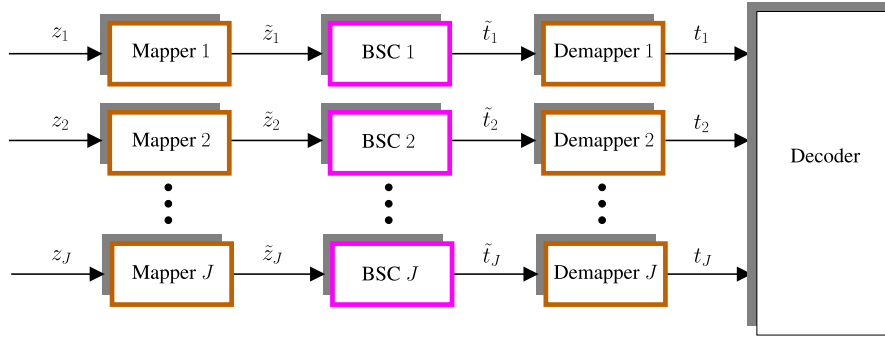


Fig. 3. An illustrative example of parallel retrieval: Independent processing flow across branches from the codeword, z_j , to the retrieved one, t_j . Mappers are located in the compression units and demappers in the (remote) processing unit.

with $0 \leq R_j \leq \log_2 |\mathcal{Z}_j|$ bits, setting an upper-bound on the j -th compression rate, $I(y_j; z_j)$.² Employing the LM method, the design problem (4) can be reformulated as an unconstrained optimization (up to the pertinent mappings' validity)

$$P^* = \operatorname{argmax}_P I(\mathbf{x}; \mathbf{t}_{1:J}) - \sum_{j=1}^J \lambda_j I(y_j; z_j), \quad (5)$$

with $\lambda_j \geq 0$, denoting the counterpart relaxation of the upper-bound, R_j , in (4). Aligned with the performed analysis in [1], the following theorem characterizes the form of stationary points³ of the objective functional in (5).

Theorem 1 (Parallel IB-based JSCC): Assume the joint distribution, $p(\mathbf{x}, \mathbf{y}_{1:J}) = p(\mathbf{x}) \prod_j p(y_j | \mathbf{x})$, the forward channels' statistics, $p(t_j | z_j)$, and λ_j are available for all $j = 1$ to J . The set of local quantizers, $\{p(z_j | y_j) | j\}$, is a stationary point of the parallel IB-based JSCC Lagrangian

$$\mathcal{L}_{\text{Par.}} = I(\mathbf{x}; \mathbf{t}_{1:J}) - \sum_{j=1}^J \lambda_j I(y_j; z_j), \quad (6)$$

if and only if for each pair $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$

$$p(z_j | y_j) = \frac{p(z_j)}{\psi_{z_j}^{\text{Par.}}(y_j, \beta_j)} \exp\left(-d_{\text{Par.}}(y_j, z_j)\right), \quad (7)$$

with $\psi_{z_j}^{\text{Par.}}(y_j, \beta_j)$, being a partition (normalization) function, ensuring the pertinent quantizer mapping's validity, and for the relevant distortion, $d_{\text{Par.}}(y_j, z_j)$, it holds

$$d_{\text{Par.}}(y_j, z_j) = \beta_j \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j} | y_j) \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J} | \mathbf{z}_{1:J}) \times D_{\text{KL}}(p(\mathbf{x} | y_j, \mathbf{z}_{1:J}^{-j}) \| p(\mathbf{x} | \mathbf{t}_{1:J})), \quad (8)$$

with $\beta_j = \frac{1}{\lambda_j}$ and $p(\mathbf{t}_{1:J} | \mathbf{z}_{1:J}) = \prod_{j=1}^J p(t_j | z_j)$, resulting from presumed Markov properties.

The proof has been presented in Appendix A.

The provided solution in Theorem 1 incorporates the *ideal* forwarding scenario as a specific instance. Explicitly, for the case in which the forward channels are presumed to

be rate-limited but *error-free*, the design problem is then to maximize the overall transmission rate that now gets reduced to $I(\mathbf{x}; \mathbf{z}_{1:J})$, under the same set of constraints on individual compression rates. Through substituting, $p(t_j | z_j) = \delta_{t_j, z_j}$, for each pair, $(z_j, t_j) \in \mathcal{Z}_j \times \mathcal{T}_j$ (with *Kronecker Delta* denotation⁴), the acquired relevant distortion in (8) boils down to

$$d_{\text{Par.}}^{\text{ide.}}(y_j, z_j) = \beta_j \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j} | y_j) \times D_{\text{KL}}(p(\mathbf{x} | y_j, \mathbf{z}_{1:J}^{-j}) \| p(\mathbf{x} | \mathbf{z}_{1:J})), \quad (9)$$

which is identical to the one provided in [25]. This verifies the *Backward Compatibility* of our deduced solution for this (parallel) IB-based JSCC problem to its NSC counterpart.

Coding Perspective: To acquire a crisp image about possible applications of this setup, one can consider a conventional coding scenario. For that, as depicted in Fig. 3, the processing flow from the compressed signal, z_j , up to the retrieved one, t_j , is presumed to be realized independently across branches. Each mapper allocates a bit-tuple, \tilde{z}_j , to the codeword, z_j , and, conversely, each demapper retrieves a codeword, t_j , from the received bit-tuple, \tilde{t}_j at the output of a *Binary Symmetric Channel (BSC)*.⁵ Contrary to the case of ideal forwarding, it must be noted that, here, the apposite choice of bit-tuple mappings (applied in the compression units) becomes relevant in terms of the end-to-end information preservation. The respective demappings will be performed independently across branches in the (remote) processing unit.

B. Successive Retrieval: Compression With Side-Information

As the second choice regarding the imposed set of constraints, again we consider the case where individual branches are subject to different rate limitations but here, contrary to the previous setup, from a compression's perspective, the side-information is exploited when treating the observation, y_j . The principal idea behind this scheme is fully aligned with the well-known Wyner-Ziv setup [10] for source coding in which a statistically correlated signal is utilized as side-information at the decoder. With the major difference of presuming *ideal* (rate-limited) forward channels, in [27],

²Aside from one-shot formulations, the (asymptotic) coding problems of Section III call for a formal multi-letter description.

³Note that, with the presumption of smoothness, all internal local extrema are stationary points as well. Thus, from this perspective, the derived forms can be regarded as necessary conditions for local optimality.

⁴The two arguments on its subscript differentiate this denotation from the one that is used for functional derivatives.

⁵Note that the choice of forward channel is arbitrary, and the BSC has been taken here just as an exemplary error-prone channel.

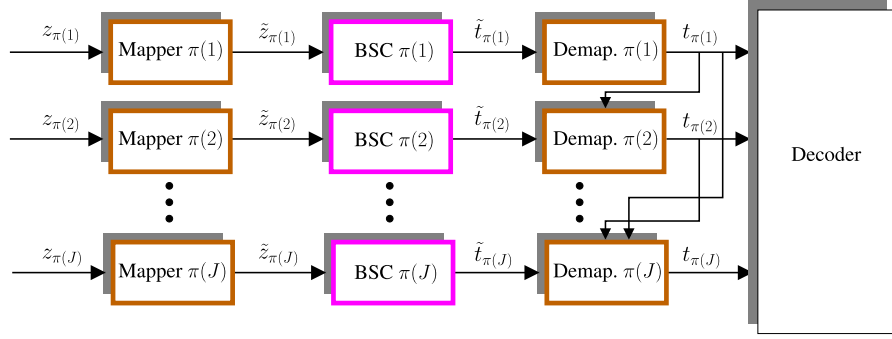


Fig. 4. An illustrative example of successive retrieval: Interconnected processing flow across branches from the codeword, $z_{\pi(j)}$, to the retrieved one, $t_{\pi(j)}$, in line with the well-known principle of “binning & detection” (see, e.g., [55]).

a completely discrete scenario has been analyzed for an identical setup to the one illustrated in Fig. 2. There, to address the design problem, the proposed algorithm in [24] for the *Distributed Information Bottleneck (DIB)* has been generalized to allow for various rate constraints across branches. The conducted derivations here extend the provided results in [27] to this successive IB-based JSCC setup. The design problem is then formulated as looking for the optimal set, $P^* = \{p^*(z_1|y_1), \dots, p^*(z_J|y_J)\}$, where

$$P^* = \underset{P: \forall j I(y_j; z_j | \mathbf{t}_{1:j-1}) \leq R_j}{\operatorname{argmax}} I(\mathbf{x}; \mathbf{t}_{1:J}), \quad (10)$$

with $0 \leq R_j \leq \log_2 |\mathcal{Z}_j|$ bits, stipulating an upper-bound on the j -th conditional compression rate, $I(y_j; z_j | \mathbf{t}_{1:j-1})$. Here as well, by employing the LM method, the design problem (10) can be restated as an unconstrained optimization (up to the pertinent mappings’ validity)

$$P^* = \underset{P}{\operatorname{argmax}} I(\mathbf{x}; \mathbf{t}_{1:J}) - \sum_{j=1}^J \lambda_j I(y_j; z_j | \mathbf{t}_{1:j-1}), \quad (11)$$

with $\lambda_j \geq 0$, being associated with the upper-bound, R_j , in (10). Aligned with the performed analysis in [1], the following theorem characterizes the form of stationary points of the objective functional in (11).

Theorem 2 (Successive IB-based JSCC): Assume the joint distribution $p(\mathbf{x}, \mathbf{y}_{1:J}) = p(\mathbf{x}) \prod_j p(y_j | \mathbf{x})$, the forward channels’ statistics, $p(\mathbf{t}_j | z_j)$, and λ_j are available for all $j = 1$ to J . The set of local quantizers, $\{p(z_j | y_j) | j\}$, is a stationary point of the successive IB-based JSCC Lagrangian

$$\mathcal{L}_{\text{Suc.}} = I(\mathbf{x}; \mathbf{t}_{1:J}) - \sum_{j=1}^J \lambda_j I(y_j; z_j | \mathbf{t}_{1:j-1}), \quad (12)$$

if and only if for each pair $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$

$$p(z_j | y_j) = \frac{p(z_j)}{\psi_{z_j}^{\text{Suc.}}(y_j, \beta_j)} \exp\left(-d_{\text{Suc.}}(y_j, z_j)\right), \quad (13)$$

with $\psi_{z_j}^{\text{Suc.}}(y_j, \beta_j)$, being a partition (normalization) function, ensuring the pertinent quantizer mapping’s validity, and for the relevant distortion, $d_{\text{Suc.}}(y_j, z_j)$, it holds

$$d_{\text{Suc.}}(y_j, z_j) = \beta_j \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j} | y_j) \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J} | \mathbf{z}_{1:J})$$

$$\begin{aligned} & \times D_{\text{KL}}(p(\mathbf{x} | y_j, \mathbf{z}_{1:J}^{-j}) \| p(\mathbf{x} | \mathbf{t}_{1:J})) \\ & - \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1} | y_j) \log p(\mathbf{t}_{1:j-1} | z_j) - \beta_j \sum_{n=j+1}^J \frac{1}{\beta_n} \\ & \times \sum_{\mathbf{t}_{1:n-1}, z_n} p(\mathbf{t}_j | z_j) p(\mathbf{t}_{1:n-1}^{-j}, z_n | y_j) \log p(z_n | \mathbf{t}_{1:n-1}), \quad (14) \end{aligned}$$

with $\beta_j = \frac{1}{\lambda_j}$ and $p(\mathbf{t}_{1:J} | \mathbf{z}_{1:J}) = \prod_{j=1}^J p(\mathbf{t}_j | z_j)$, resulting from presumed Markov properties.

The proof has been presented in Appendix B.

A closer look at the obtained relevant distortion in (14) reveals that, basically, it extends the derived solution in (8) for the case of parallel retrieval by two extra terms appearing due to the consideration of side-information for the compression rates of individual branches. In the upcoming section, the present common structure in the obtained results for both parallel and successive compression schemes will be leveraged for devising a generic algorithm, the M-FAVIB, to address the respective design optimizations.

The *ideal* forwarding scenario is encompassed as a specific instance in the derived solution in Theorem 2. Explicitly, in the case where the forward channels are presumed to be rate-limited but *error-free*, the design problem would be to maximize the overall transmission rate that now gets reduced to $I(\mathbf{x}; \mathbf{z}_{1:J})$, under the adapted set of constraints on individual conditional compression rates, i.e., for the j -th branch it must apply $I(y_j; z_j | \mathbf{z}_{1:j-1}) \leq R_j$. By substituting, $p(\mathbf{t}_j | z_j) = \delta_{\mathbf{t}_j, z_j}$, for each pair $(z_j, \mathbf{t}_j) \in \mathcal{Z}_j \times \mathcal{T}_j$ (with *Kronecker Delta* denotation), the calculated relevant distortion in (14) boils down to

$$\begin{aligned} & d_{\text{Suc.}}^{\text{ide.}}(y_j, z_j) \\ & = \beta_j \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j} | y_j) D_{\text{KL}}(p(\mathbf{x} | y_j, \mathbf{z}_{1:J}^{-j}) \| p(\mathbf{x} | \mathbf{z}_{1:J})) \\ & - \sum_{\mathbf{z}_{1:j-1}} p(\mathbf{z}_{1:j-1} | y_j) \log p(\mathbf{z}_{1:j-1} | z_j) \\ & - \beta_j \sum_{n=j+1}^J \frac{1}{\beta_n} \sum_{\mathbf{z}_{1:n}^{-j}} p(\mathbf{z}_{1:n}^{-j} | y_j) \log p(z_n | \mathbf{z}_{1:n-1}), \quad (15) \end{aligned}$$

which is the same as the one provided in [27]. This confirms the *Backward Compatibility* of our deduced solution for this (successive) IB-based JSCC problem to its NSC counterpart.

Coding Perspective: As a tangible example of possible applications for this setup, one can consider a conventional coding scenario. Denoting by $\pi(\cdot)$ a specific demapping order, as illustrated in Fig. 4, the processing flow from the compressed signal, $\mathbf{z}_{\pi(j)}$, up to the retrieved one, $\mathbf{t}_{\pi(j)}$, is presumed to be realized in an interconnected fashion across branches. Therefore, the major difference compared to the previous setup is that the demappings are performed in a sequential manner across branches such that the already retrieved signals, $\mathbf{t}_{\pi(1):\pi(j-1)}$, can be utilized as an available side-information when applying the pertinent demapping to retrieve $\mathbf{t}_{\pi(j)}$, since all demapper blocks are located in the (remote) processing unit.⁶ In this fashion, contrary to the parallel setup, the present correlation between the signals of different branches will be leveraged as well.

IV. A GENERIC DESIGN METHOD

A. Proposed Algorithmic Approach

In this section, we develop a generic, iterative algorithm to tackle both distributed compression design problems. To that end, it has to be noted that irrespective of the chosen compression scheme, the derived stationary solution for every pair $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$ of each individual quantizer mapping has the following *implicit* form

$$p(z_j|y_j) = \frac{p(z_j)}{\psi_{z_j}^r(y_j, \beta_j)} \exp(-d_r(y_j, z_j)), \quad (16)$$

with $r \in \{\text{Par.}, \text{Suc.}\}$.⁷ The right side of (16) can be interpreted as a functional featuring all the individual quantizer mappings, $\{p(z_j|y_j) \mid j\}$, as its input arguments since they come into play when calculating the relevant distortion, $d_r(y_j, z_j)$. This indicates that (16) can be viewed as (with \mathcal{F}_j denoting a specific functional)

$$p(z_j|y_j) = \mathcal{F}_j(p(z_1|y_1), \dots, p(z_J|y_J)), \quad (17)$$

for the j -th branch. Going through all branches then yields a non-linear system that extends the structure of *Multivariate Fixed-Point* systems [56] to the field of functionals wherein the functions of multiple variables are substituted by the functionals of multiple mappings. Hence, the conventional iterative procedures can be applied for solving the obtained system as well. Herein, we propose an iterative procedure with the *synchronous* updating rule in line with the standard *Jacobi* method for solving linear systems [56].

The proposed algorithm, termed Multiterminal Forward-Aware Vector IB (M-FAVIB), with the pseudo-code presented in Alg. 1 proceeds as follows: Commencing with a set of random (yet valid) mappings, $\{p^{(0)}(z_j|y_j) \mid j\}$, for each pair, $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$, the updates are executed (till convergence

⁶It shall be mentioned that, generally, the processing order, $\pi(\cdot)$, affects the obtainable performance. Thus, it is subject to optimization. Herein, for the sake of simplicity, we presume that the optimal ordering is set by $\pi(j) = j$.

⁷Note that the form of the solution was to be expected since the problem characterization and the solving technique here are the same as the original IB method [1], but for a more complex setting.

Alg. 1 Multiterminal Forward-Aware Vector IB (M-FAVIB)

Input: $p(x, \mathbf{y}_{1:J})$, $\beta_j > 0$, $p(\mathbf{t}_{1:J}|\mathbf{z}_{1:J})$, conv. parameter $\varepsilon > 0$

Output: A (soft) partition \mathbf{z}_j of \mathcal{Y}_j into (at most) $|\mathcal{Z}_j|$ clusters

Initialization: $i = 0$, random mappings $\{p^{(0)}(z_j|y_j) \mid j\}$

while True **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 the i -th update for distributions in $d_r(y_j, z_j)$ from

$$p^{(i)}(x, \mathbf{y}_{1:J}, \mathbf{z}_{1:J}, \mathbf{t}_{1:J}) = p(x, \mathbf{y}_{1:J}) \prod_{\ell=1}^J p^{(i)}(z_\ell|y_\ell) \times p(\mathbf{t}_{1:J}|\mathbf{z}_{1:J})$$

 • $p^{(i+1)}(z_j|y_j) \leftarrow \frac{p^{(i)}(z_j)}{\psi_{z_j}^{r(i+1)}(y_j, \beta_j)} \exp(-d_r^{(i)}(y_j, z_j))$

end for

if $\forall j, \forall y_j : D_{\text{JS}}^{\{\frac{1}{2}, \frac{1}{2}\}}(p^{(i+1)}(z_j|y_j) \| p^{(i)}(z_j|y_j)) \leq \varepsilon$ **then**

 Break

else

$i \leftarrow i + 1$

end if

end while

by $\varepsilon \ll 1$) through⁸

$$p^{(i+1)}(z_j|y_j) = \frac{p^{(i)}(z_j)}{\psi_{z_j}^{r(i+1)}(y_j, \beta_j)} \exp(-d_r^{(i)}(y_j, z_j)), \quad (18)$$

with, i , representing the iteration counter. The quantizer output probability, $p^{(i)}(z_j)$, and the relevant distortion, $d_r^{(i)}(y_j, z_j)$, in (18) are calculated through exerting the current versions of all the individual quantizer mappings, $\{p^{(i)}(z_j|y_j) \mid j\}$, to marginalize the joint distribution, $p^{(i)}(x, \mathbf{y}_{1:J}, \mathbf{z}_{1:J}, \mathbf{t}_{1:J})$. The updates are performed *synchronously*, i.e., at iteration $i + 1$, all the individual quantizer mappings of different branches, $\{p^{(i+1)}(z_j|y_j) \mid j\}$, will be updated based on the previous configuration of the same set, i.e., $\{p^{(i)}(z_j|y_j) \mid j\}$. In the following part, it will be proven that irrespective of the chosen compression scheme (i.e., parallel or successive retrieval), the M-FAVIB algorithm converges to a stationary point of the objective functional. Hence, as a commonly used workaround for avoiding poor results, the aforementioned procedure can be repeated with different starting points, $\{p^{(0)}(z_j|y_j) \mid j\}$, with the aim of retaining the best outcome. It should be reminded that, as its name suggests, the M-FAVIB algorithm is directly applicable to the case of *Vector Quantization (VQ)*, following the same line of argumentation that already has been provided in [49] for its uniterminal counterpart.

B. Proof of Convergence

In this part, it will be shown that the design optimizations for parallel and successive schemes can be addressed by an alternating minimization w.r.t. the set of all quantizer mappings, P , and another set of apposite auxiliary probability distributions, Q , by introducing a tight *variational upper-bound*, $\bar{\mathcal{F}}_r(P, Q)$,

⁸Per iteration, for every realization, $y_j \in \mathcal{Y}_j$, of the access channel output the sum of calculated terms in (18) (by ignoring $\psi_{z_j}^r$) over all output bins, $z_j \in \mathcal{Z}_j$, acts as the normalization (partition) function, $\psi_{z_j}^r(y_j, \beta_j)$, to ensure $\sum_{z_j \in \mathcal{Z}_j} p(z_j|y_j) = 1$.

on the respective objective functional, $\mathcal{F}_r(P)$. The focal update step of M-FAVIB can then be interpreted as merging together the updates of P and Q . Through this unfolding trick, one principally shows that M-FAVIB falls into the category of *Successive Upper-Bound Minimization (SUM)*⁹ [51], ensuring that it converges to a stationary point. Similar approaches have been presented in [24] and [26] for convergence proof of their proposed algorithms.

Parallel Retrieval: The design optimization in (5) can be reformulated as minimizing the functional

$$\mathcal{F}_{\text{Par.}}(P) = H(\mathbf{x}) - \mathcal{L}_{\text{Par.}} = \sum_{j=1}^J \lambda_j I(y_j; z_j) + H(\mathbf{x}|\mathbf{t}_{1:J}), \quad (19)$$

over P . By defining $Q = \{q(\mathbf{x}|\mathbf{t}_{1:J})\}$ and the functional

$$\bar{\mathcal{F}}_{\text{Par.}}(P, Q) = \sum_{j=1}^J \lambda_j I(y_j; z_j) - \mathbb{E}_{\mathbf{x}, \mathbf{t}_{1:J}} \{\log q(\mathbf{x}|\mathbf{t}_{1:J})\}, \quad (20)$$

the following Lemmas hold:

Lemma 1: It applies

$$\min_P \mathcal{F}_{\text{Par.}}(P) = \min_P \min_Q \bar{\mathcal{F}}_{\text{Par.}}(P, Q). \quad (21)$$

Proof: For the difference of $\bar{\mathcal{F}}_{\text{Par.}}(P, Q)$ and $\mathcal{F}_{\text{Par.}}(P)$, it holds

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Par.}}(P, Q) - \mathcal{F}_{\text{Par.}}(P) &= \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J}) \\ &\quad \times D_{\text{KL}}(p(\mathbf{x}|\mathbf{t}_{1:J}) \| q(\mathbf{x}|\mathbf{t}_{1:J})) \geq 0, \quad (22) \end{aligned}$$

where the equality holds iff $q(\mathbf{x}|\mathbf{t}_{1:J}) = p(\mathbf{x}|\mathbf{t}_{1:J})$.

Lemma 2: $\bar{\mathcal{F}}_{\text{Par.}}(P, Q)$ is separately convex in P and Q .

Proof: It follows directly from the application of log-sum inequality [54].

Lemma 3: For a fixed P , there exists a unique Q that minimizes $\bar{\mathcal{F}}_{\text{Par.}}(P, Q)$, given by

$$q^*(\mathbf{x}|\mathbf{t}_{1:J}) = p(\mathbf{x}|\mathbf{t}_{1:J}), \quad (23)$$

where $p(\mathbf{x}|\mathbf{t}_{1:J})$ is calculated from P .

Proof: It follows directly from the proof of Lemma 1.

Lemma 4: For a fixed Q , there exists a P that minimizes $\bar{\mathcal{F}}_{\text{Par.}}(P, Q)$, given by

$$\begin{aligned} p^*(z_j|y_j) &= \frac{p(z_j)}{\psi_{z_j}^{\text{Par.}}(y_j, \beta_j)} \exp\left(-\bar{d}_{\text{Par.}}(y_j, z_j)\right) \\ &\quad \forall (y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j, \quad (24) \end{aligned}$$

with $\psi_{z_j}^{\text{Par.}}(y_j, \beta_j)$, being a partition function, ensuring the pertinent quantizer mapping's validity and the relevant distortion, $\bar{d}_{\text{Par.}}(y_j, z_j)$, is calculated by

$$\begin{aligned} \bar{d}_{\text{Par.}}(y_j, z_j) &= \beta_j \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j}|y_j) \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J}|\mathbf{z}_{1:J}) \\ &\quad \times D_{\text{KL}}(p(\mathbf{x}|y_j, \mathbf{z}_{1:J}^{-j}) \| q(\mathbf{x}|\mathbf{t}_{1:J})). \quad (25) \end{aligned}$$

⁹The principal idea of the SUM is to optimize a sequence of approximate objective function(al)s (satisfying certain mild assumptions [51]), rather than directly optimizing the original non-convex and/or non-smooth objective function(al).

Proof: It follows from the same line of reasoning as in the proof of Theorem 1, noting that

$$\begin{aligned} \frac{\delta(\mathbb{E}_{\mathbf{x}, \mathbf{t}_{1:J}} \{\log q(\mathbf{x}|\mathbf{t}_{1:J})\})}{\delta p(z_j|y_j)} &= p(y_j) \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j}|y_j) \\ &\quad \times \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J}|\mathbf{z}_{1:J}) \sum_x p(x|y_j, \mathbf{z}_{1:J}^{-j}) \log q(x|\mathbf{t}_{1:J}). \quad (26) \end{aligned}$$

Merging together the optimal results for Q and P from the last two Lemmas, one directly obtains the focal update step (7)-(8) of M-FAVIB regarding any of the involved quantizer mappings. Consequently, it is ensured by [51, Thm 1] that the M-FAVIB will converge to a stationary point of the objective functional as $\bar{\mathcal{F}}_{\text{Par.}}(P, Q)$ and $\mathcal{F}_{\text{Par.}}(P)$ satisfy [51, Prop. 1].

Successive Retrieval: The design optimization in (11) can be reformulated as minimizing the functional

$$\begin{aligned} \mathcal{F}_{\text{Suc.}}(P) = H(\mathbf{x}) - \mathcal{L}_{\text{Suc.}} &= \sum_{j=1}^J \lambda_j I(y_j; z_j) \\ &\quad - \sum_{j=2}^J \lambda_j I(z_j; \mathbf{t}_{1:j-1}) + H(\mathbf{x}|\mathbf{t}_{1:J}), \quad (27) \end{aligned}$$

over P . Defining the set of auxiliary probability distributions $Q = \{q(z_2|\mathbf{t}_1), \dots, q(z_J|\mathbf{t}_{1:J-1}), q(\mathbf{x}|\mathbf{t}_{1:J})\}$ and the functional

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Suc.}}(P, Q) &= \sum_{j=1}^J \lambda_j I(y_j; z_j) - \sum_{j=2}^J \lambda_j \\ &\quad \times \mathbb{E}_{z_j, \mathbf{t}_{1:j-1}} \left\{ \log \frac{q(z_j|\mathbf{t}_{1:j-1})}{p(z_j)} \right\} - \mathbb{E}_{\mathbf{x}, \mathbf{t}_{1:J}} \{\log q(\mathbf{x}|\mathbf{t}_{1:J})\}, \quad (28) \end{aligned}$$

the following Lemmas hold:

Lemma 5: It applies

$$\min_P \mathcal{F}_{\text{Suc.}}(P) = \min_P \min_Q \bar{\mathcal{F}}_{\text{Suc.}}(P, Q). \quad (29)$$

Proof: For the difference of $\bar{\mathcal{F}}_{\text{Suc.}}(P, Q)$ and $\mathcal{F}_{\text{Suc.}}(P)$, it holds

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Suc.}}(P, Q) - \mathcal{F}_{\text{Suc.}}(P) &= \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J}) D_{\text{KL}}(p(\mathbf{x}|\mathbf{t}_{1:J}) \| q(\mathbf{x}|\mathbf{t}_{1:J})) \\ &\quad + \sum_{j=2}^J \lambda_j \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1}) \\ &\quad \times D_{\text{KL}}(p(z_j|\mathbf{t}_{1:j-1}) \| q(z_j|\mathbf{t}_{1:j-1})) \geq 0, \quad (30) \end{aligned}$$

where the equality holds iff $q(z_j|\mathbf{t}_{1:j-1}) = p(z_j|\mathbf{t}_{1:j-1})$ for $j = 2$ to J and $q(\mathbf{x}|\mathbf{t}_{1:J}) = p(\mathbf{x}|\mathbf{t}_{1:J})$.

Lemma 6: $\bar{\mathcal{F}}_{\text{Suc.}}(P, Q)$ is separately convex in P and Q .

Proof: It follows directly from the application of log-sum inequality [54].

Lemma 7: For a fixed P , there exists a unique Q that minimizes $\bar{\mathcal{F}}_{\text{Suc.}}(P, Q)$, given by

$$\begin{aligned} q^*(z_j|\mathbf{t}_{1:j-1}) &= p(z_j|\mathbf{t}_{1:j-1}) \text{ for } j = 2 \text{ to } J, \\ q^*(\mathbf{x}|\mathbf{t}_{1:J}) &= p(\mathbf{x}|\mathbf{t}_{1:J}), \quad (31) \end{aligned}$$

where $p(z_j|\mathbf{t}_{1:j-1})$ for $j = 2$ to J and $p(\mathbf{x}|\mathbf{t}_{1:J})$ are calculated from P .

Proof: It follows directly from the proof of Lemma 5.

Lemma 8: For a fixed Q , there exists a P that minimizes $\bar{\mathcal{F}}_{\text{Suc.}}(P, Q)$, given by

$$p^*(z_j|y_j) = \frac{p(z_j)}{\psi_{z_j}^{\text{Suc.}}(y_j, \beta_j)} \exp\left(-\bar{d}_{\text{Suc.}}(y_j, z_j)\right) \quad \forall (y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j, \quad (32)$$

with $\bar{\psi}_{z_j}^{\text{Suc.}}(y_j, \beta_j)$, being a partition function, ensuring the pertinent quantizer mapping's validity and the relevant distortion, $\bar{d}_{\text{Suc.}}(y_j, z_j)$, is calculated by

$$\begin{aligned} \bar{d}_{\text{Suc.}}(y_j, z_j) &= \beta_j \sum_{\mathbf{z}_{1:j}^{-j}} p(\mathbf{z}_{1:j}^{-j}|y_j) \sum_{\mathbf{t}_{1:j}} p(\mathbf{t}_{1:j}|\mathbf{z}_{1:j}) \\ &\quad \times D_{\text{KL}}(p(\mathbf{x}|y_j, \mathbf{z}_{1:j}^{-j})||q(\mathbf{x}|\mathbf{t}_{1:j})) \\ &\quad - \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1}|y_j) \log \frac{q(z_j|\mathbf{t}_{1:j-1})}{p(z_j)} - \beta_j \sum_{n=j+1}^J \frac{1}{\beta_n} \\ &\quad \times \sum_{\mathbf{t}_{1:n-1}, z_n} p(\mathbf{t}_{1:n-1}|z_j) p(\mathbf{t}_{1:n-1}^{-j}, z_n|y_j) \log q(z_n|\mathbf{t}_{1:n-1}). \quad (33) \end{aligned}$$

Proof: It follows from the same line of reasoning as in the proof of Theorem 2, noting that

$$\begin{aligned} \delta \left(\mathbb{E}_{z_j, \mathbf{t}_{1:j-1}} \left\{ \log \frac{q(z_j|\mathbf{t}_{1:j-1})}{p(z_j)} \right\} \right) & \\ \delta p(z_j|y_j) & \\ = p(y_j) \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1}|y_j) \log \frac{q(z_j|\mathbf{t}_{1:j-1})}{p(z_j)}, & \quad (34) \end{aligned}$$

and in the case of $n > j$

$$\begin{aligned} \delta \left(\mathbb{E}_{z_n, \mathbf{t}_{1:n-1}} \left\{ \log \frac{q(z_n|\mathbf{t}_{1:n-1})}{p(z_n)} \right\} \right) & \\ \delta p(z_j|y_j) & \\ = p(y_j) \sum_{\mathbf{t}_{1:n-1}, z_n} p(\mathbf{t}_{1:n-1}|z_j) p(\mathbf{t}_{1:n-1}^{-j}, z_n|y_j) \log \frac{q(z_n|\mathbf{t}_{1:n-1})}{p(z_n)}. & \quad (35) \end{aligned}$$

The focal update step (13)-(14) of M-FAVIB regarding any of the involved quantizer mappings is directly obtained by merging together the optimal results for Q and P from the last two Lemmas.¹⁰ Hence, the convergence of M-FAVIB to a stationary point of the objective functional is ensured by [51, Thm 1] as $\bar{\mathcal{F}}_{\text{Suc.}}(P, Q)$ and $\mathcal{F}_{\text{Suc.}}(P)$ satisfy [51, Prop. 1].

C. Supplementary Mathematical Discussion

For the *parallel retrieval*, applying (7)-(8) as the central update step of the M-FAVIB, in the limit of letting $\beta_j \rightarrow 0$, the design problem (5) boils down to minimizing the j -th compression rate, $I(y_j; z_j)$ (presuming fixed $p(z_\ell|y_\ell)$ and λ_ℓ for all $\ell = 1$ to J and $\ell \neq j$) w.r.t. the j -th quantizer

¹⁰Regarding the second component in (33), replacing $q(z_j|\mathbf{t}_{1:j-1})$ by $p(z_j|\mathbf{t}_{1:j-1}) = \frac{p(\mathbf{t}_{1:j-1}, z_j)}{p(\mathbf{t}_{1:j-1})}$, the log term inside can be rewritten through $\log \frac{p(\mathbf{t}_{1:j-1}, z_j)}{p(\mathbf{t}_{1:j-1})p(z_j)} = \log p(\mathbf{t}_{1:j-1}|z_j) - \log p(\mathbf{t}_{1:j-1})$. Furthermore, the term $\sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1}|y_j) \log p(\mathbf{t}_{1:j-1})$ can be ignored since it does not depend on z_j , and, therefore, gets absorbed into the respective normalization function.

mapping, $p(z_j|y_j)$. In that case, each realization, $y_j \in \mathcal{Y}_j$, is allocated to all output bins, $z_j \in \mathcal{Z}_j$, equiprobably (state of *full diffusion*). In this fashion, the input and output of the j -th compression unit become statistically independent, and the respective compression rate, $I(y_j; z_j)$, reaches its global minimum, i.e., zero. Further, for finite β_j values, *stochastic* (soft) mappings, $p(z_j|y_j)$, are engendered in general, while in the asymptotic case of letting $\beta_j \rightarrow \infty$, the partition function, $\psi_{z_j}^{\text{Par.}}(y_j, \beta_j)$, for each realization, y_j , allocates all the probability mass into the specific bin that reveals the minimum $\frac{d_{\text{par.}}}{\beta_j}$ value and, hence, induces the quantizer mapping to become *deterministic* (hard), i.e., $p(z_j|y_j) \in \{0, 1\}$ for each pair $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$ (state of *full concentration*). To rationalize this, by presuming fixed $p(z_\ell|y_\ell)$ and λ_ℓ for all $\ell = 1$ to J and $\ell \neq j$, and by letting $\beta_j \rightarrow \infty$ ($\lambda_j \rightarrow 0$), the design optimization (5) boils down to maximizing the overall transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:J})$, w.r.t. the j -th quantizer mapping, $p(z_j|y_j)$, that is a *convex maximization* task. To discern this, it should be noted that $I(\mathbf{x}; \mathbf{t}_{1:J})$ is convex w.r.t. $p(\mathbf{t}_{1:J}|\mathbf{x})$ for a fixed $p(\mathbf{x})$ [54]. Furthermore, $p(\mathbf{t}_{1:J}|\mathbf{x})$ and $p(\mathbf{t}_j|\mathbf{x})$ are related through

$$p(\mathbf{t}_{1:J}|\mathbf{x}) = p(\mathbf{t}_j|\mathbf{x}) \prod_{\ell=1:J, \ell \neq j} p(\mathbf{t}_\ell|\mathbf{x}), \quad (36)$$

which is an *affine* transform since fixing $p(z_\ell|y_\ell)$ implies that $p(\mathbf{t}_\ell|\mathbf{x})$ is fixed as well. Affine transforms preserve convexity [57, Sec. 3.2]. Hence, $I(\mathbf{x}; \mathbf{t}_{1:J})$ is also convex w.r.t. $p(\mathbf{t}_j|\mathbf{x})$. Moreover, $p(\mathbf{t}_j|\mathbf{x})$ and $p(z_j|y_j)$ are related through

$$p(\mathbf{t}_j|\mathbf{x}) = \sum_{z_j \in \mathcal{Z}_j} \sum_{y_j \in \mathcal{Y}_j} p(\mathbf{t}_j|z_j) p(z_j|y_j) p(y_j|\mathbf{x}), \quad (37)$$

which is also an affine transform. This, in turn, concludes the proof of the claimed proposition. Resorting to a well-known theorem from convex maximization theory asserting that a convex function that is defined over a closed and convex set obtains its global maximum at an *extreme point* of that set [58, Ch. 4], it is directly inferred that it suffices to focus on *deterministic* mappings. To realize this, one may recall that the space of valid mappings, $p(z_j|y_j)$, is a closed, convex polytope generated by the Cartesian product of $|\mathcal{Y}_j|$ probability simplices [59]. The extreme points of this polytope occur at its corners, corresponding to the Cartesian product of the corners of its constituent probability simplices, yielding a *deterministic* mapping per extreme point.

For the *successive retrieval*, analogous behavior is observed as well. Specifically, applying (13)-(14) as the central update step of the M-FAVIB, (presuming fixed $p(z_\ell|y_\ell)$ and λ_ℓ for all $\ell = 1$ to J and $\ell \neq j$) and by letting $\beta_j \rightarrow 0$ ($\lambda_j \rightarrow \infty$), the design problem (11) boils down to minimizing the j -th conditional compression rate, $I(y_j; z_j|\mathbf{t}_{1:j-1})$, w.r.t. the j -th quantizer mapping, $p(z_j|y_j)$. Like before, this is addressed by allotting each realization, $y_j \in \mathcal{Y}_j$ to all output clusters, $z_j \in \mathcal{Z}_j$, equiprobably (state of *full diffusion*), resulting in the global minimum (i.e., zero) of the respective conditional compression rate, $I(y_j; z_j|\mathbf{t}_{1:j-1})$, since it is non-negative and upper-bounded by $I(y_j; z_j)$. Similarly, for finite β_j values, *stochastic* mappings, $p(z_j|y_j)$, are engendered in general, while

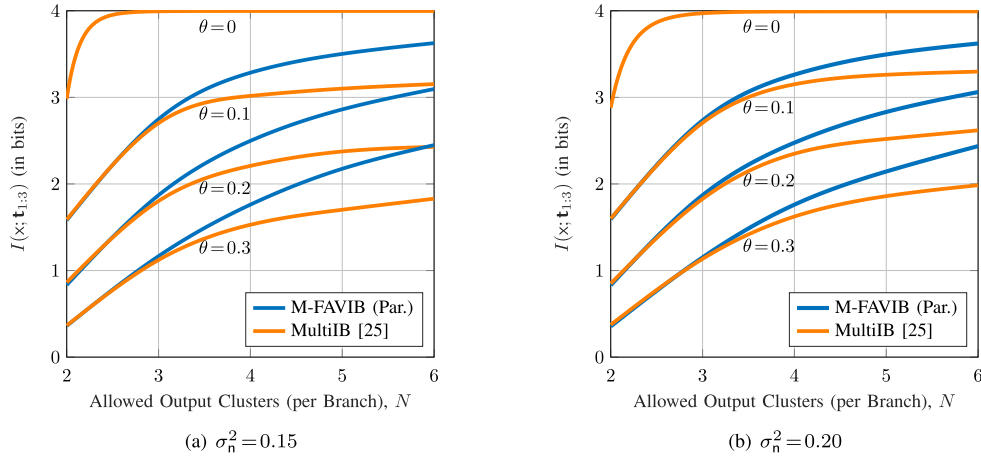


Fig. 5. End-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, vs. allowed number of bins, N , equiprobable 16-QAM signaling ($\sigma_x^2 = 10$), discrete AWGN access channels with the noise variance, a) $\sigma_n^2 = 0.15$ and b) $\sigma_n^2 = 0.20$, symmetric $N \times N$ forward channels with the varying reliability values, $\theta = 0, 0.1, 0.2, 0.3$, $\beta_j = 100$ for $j = 1, 2, 3$, the convergence parameter, $\varepsilon = 10^{-3}$.

in the asymptotic case of letting $\beta_j \rightarrow \infty$, as before, the partition function, $\psi_{z_j}^{\text{Suc}}(y_j, \beta_j)$, induces the quantizer mapping, $p(z_j|y_j)$, to become *deterministic* (state of *full concentration*). To justify this, a similar line of reasoning as in the previous case is also applicable here. Specifically, by presuming fixed $p(z_\ell|y_\ell)$ and λ_ℓ for all $\ell = 1$ to J and $\ell \neq j$, and by letting $\beta_j \rightarrow \infty$ ($\lambda_j \rightarrow 0$), the design optimization (11) boils down to maximizing¹¹

$$I(\mathbf{x}; \mathbf{t}_{1:J}) + \sum_{n=j+1}^J \lambda_n I(\mathbf{z}_n; \mathbf{t}_{1:n-1}), \quad (38)$$

w.r.t. the j -th quantizer mapping, $p(z_j|y_j)$, that can be shown to be a *convex maximization* task. To that end, it only suffices to demonstrate that $I(\mathbf{z}_n; \mathbf{t}_{1:n-1})$ is convex w.r.t. $p(z_j|y_j)$ since λ_n is non-negative and the sum of several convex functions will be convex as well. $I(\mathbf{z}_n; \mathbf{t}_{1:n-1})$ is a convex function of $p(\mathbf{t}_{1:n-1}|\mathbf{z}_n)$ for a fixed $p(\mathbf{z}_n)$ [54]. Noting that the relation between $p(\mathbf{t}_{1:n-1}|\mathbf{z}_n)$ and $p(z_j|y_j)$, established by (39), shown at the bottom of the page, is also an affine transform, in turn, concludes the proof of the claimed proposition.

V. SIMULATION RESULTS

A. (Forward-) Awareness Vs. Unawareness

Regarding the depicted system model in Fig. 2, we presume equiprobable source signals from a standard 16-QAM (Quadrature Amplitude Modulation) constellation ($\sigma_x^2 = 10$) over $J = 3$ branches. The access link between the source and each compression unit has been modeled as a discrete memoryless channel, approximating a discrete-time discrete-input continuous-output Additive White Gaussian Noise (AWGN) channel with the noise variance, σ_n^2 . To acquire the transition

probability matrices, rather than a prequantization of the output signal, 160 samples have been generated per branch, following a pure Monte Carlo approach. Denoting by N the allowed number of quantizers' output clusters, we consider a symmetric $N \times N$ forward channel model per branch that is characterized by the reliability parameter, θ , in the following fashion: Each input symbol is received correctly with the probability $1 - \theta$ and erroneously (to every other output symbols) with the probability $\frac{\theta}{N-1}$. Therefore, higher θ values indicate less reliable transmissions and vice versa.¹² Further, note that for a specific reliability value, θ , the transition probabilities, $p(\mathbf{t}_j|\mathbf{z}_j)$, will be influenced by the particular choice of N . We consider a totally symmetric setup (having the same access channel noise variance, σ_n^2 , and the forward channel reliability value, θ , across branches), and set all β_j values (for $j = 1, 2, 3$) to 100. Then, the end-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, is calculated as the performance indicator.

The required quantization is applied by the proposed M-FAVIB (parallel retrieval) and the baseline MultiIB algorithm from [25] to check whether integrating the forward channels' effects into the quantizers' design problem brings about some performance gain or not in comparison with the case, wherein the *error-prone* forward channels are totally neglected, and one simply aims at maximizing $I(\mathbf{x}; \mathbf{z}_{1:3})$. Since these algorithms are randomly initialized, to be fair, we choose an identical set of starting points, $\{p^{(0)}(z_j|y_j) | j\}$, for both approaches. To avoid poor results, each method has been repeated 100 times, and the best outcome has been retained. The obtained curves have been illustrated in Fig. 5.

Specifically, for two distinct values of the access channels' noise variance, namely, $\sigma_n^2 = 0.15, 0.20$, (with $I(\mathbf{x}; \mathbf{y}_{1:3}) \approx 4$

¹¹This can be realized by noting that from Markovian relations, it is inferred $I(\mathbf{y}_n; \mathbf{z}_n | \mathbf{t}_{1:n-1}) = I(\mathbf{y}_n; \mathbf{z}_n) - I(\mathbf{z}_n; \mathbf{t}_{1:n-1})$.

¹²The capacity of this symmetric $N \times N$ forward channel model is equal to $C_{FC}(N, \theta) = \log_2 N + (1 - \theta) \log_2(1 - \theta) + \theta \log_2 \frac{\theta}{N-1}$ [60].

$$p(\mathbf{t}_{1:n-1}|\mathbf{z}_n) = \frac{\sum_{x \in \mathcal{X}} p(x) p(\mathbf{z}_n|x) p(\mathbf{t}_{1:n-1}^j|x) \sum_{z_j \in \mathcal{Z}_j} \sum_{y_j \in \mathcal{Y}_j} p(\mathbf{t}_j|z_j) p(z_j|y_j) p(y_j|x)}{p(\mathbf{z}_n)}, \quad (39)$$

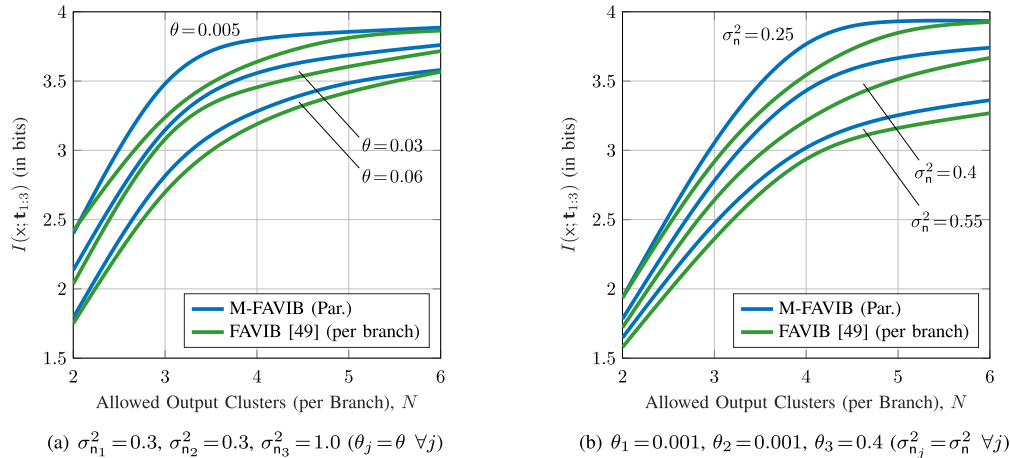


Fig. 6. End-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, vs. allowed number of bins, N , equiprobable 16-QAM signaling ($\sigma_x^2 = 10$), discrete AWGN access channels with the noise variance, $\sigma_{n_j}^2$, for branch j , symmetric $N \times N$ forward channels with the reliability value, θ_j for branch j , $\beta_j = 100$ for $j = 1, 2, 3$, the convergence parameter, $\varepsilon = 10^{-3}$.

bits), the allowed number of output clusters (per branch), N , has been varied from 2 to 6 and the overall transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, has been calculated for four choices of the forward channels' reliability value, namely, $\theta = 0, 0.1, 0.2, 0.3$. For the particular case of $\theta = 0$, corresponding to considering *error-free* forward channels, the M-FAVIB (parallel retrieval) and the MultiIB [25] algorithms engender the same result. For the other three cases, corresponding to having *error-prone* forward channels, it is directly observed that, compared to the MultiIB algorithm that completely ignores the imperfect forwarding of quantizers' output signals, the proposed M-FAVIB provides larger end-to-end transmission rates through an increase in the allowed number of quantizers' output clusters. This clearly substantiates the fact that integrating the forward channels' effects into the design problem of the compression scheme is, indeed, beneficial. This was to be expected, noting that for the M-FAVIB, the design of compressed signals are such that, in addition to capturing well the information from the remote source, they also account for the errors occurring over the imperfect forward links, while for the MultiIB, by ignoring the impacts of the forward channels, they are designed solely to preserve information about the source. Comparing the respective curves of M-FAVIB with the pertinent results for the specific case of *error-free* forward channels, $\theta = 0$, reveals that by increasing the output levels, the occurring information loss over imperfect forward channels is steadily reduced.

Through a more detailed inspection of the provided results in Fig. 5, it is observed that, contrary to the baseline curves, the performance of M-FAVIB remains almost the same for the chosen values of the access channels' noise variance, σ_n^2 . The reason behind this is the fact that in the interplay between the influencing parameters, i.e., the access channels' noise variance and the forward channels' reliability parameter, the predominant confining factor is the reliability of the forward channels. Obviously, this does not happen for the baseline curves as they do not take into account the imperfections of the forward channels. The performance gap of the baseline and the M-FAVIB stems from the point that the baseline totally ignores the forwarding effects and only

tries to maximize $I(\mathbf{x}; \mathbf{z}_{1:3})$. Thus, it does not fully leverage the available resources (the allowed forward rates) in the sense of attempting to preserve as much end-to-end rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, as possible. The saturation (flattening effect) occurring over the baseline curves can also be understood analogously. The lower the access channels' noise variance, σ_n^2 , the lower the number of required quantizers' output clusters to come quite close to the maximum supportable value of $I(\mathbf{x}; \mathbf{z}_{1:3})$ and, consequently, the sooner the saturation effect and, thus, ending up to lower end-to-end transmission rates, $I(\mathbf{x}; \mathbf{t}_{1:3})$.

B. Joint Vs. Separate Design

In this part, we substantiate the fact that the *joint* design of local quantizers is, indeed, beneficial compared to the simplest approach of applying independent (*separately* designed) local quantizers across branches. For that, we consider the same setup as that of the previous part but with different (i.e., asymmetric) parameter specifications. Explicitly, we presume the equiprobable source samples out of a 16-QAM constellation. The access links are modeled as discrete memoryless channels (160 samples per branch), approximating discrete-time discrete-input continuous-output AWGN channels with the particular noise variance, $\sigma_{n_j}^2$, for the j -th branch. The forward links are modeled as $N \times N$ symmetric channels each characterized by the particular reliability parameter, θ_j , for the j -th branch. We set all β_j values (for $j = 1, 2, 3$) to 100 and calculate the end-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, as the performance indicator, when varying the allowed number of quantizers' output clusters, N , from 2 to 6. Here, we perform the same investigations for the M-FAVIB (parallel retrieval) and the case in which, per branch, the proposed FAVIB algorithm in [49] has been employed to maximize $I(\mathbf{x}; \mathbf{t}_j)$ for $j = 1, 2, 3$, separately. To be fair, like before the same set of starting points, $\{p^{(0)}(z_j|y_j) | j\}$, has been used for both approaches, and to avoid poor results, each method has been repeated 100 times with the best outcome retained. The obtained results have been illustrated in Fig. 6.

In Fig. 6a, we considered an asymmetric specification for the access channels through choosing different noise variances

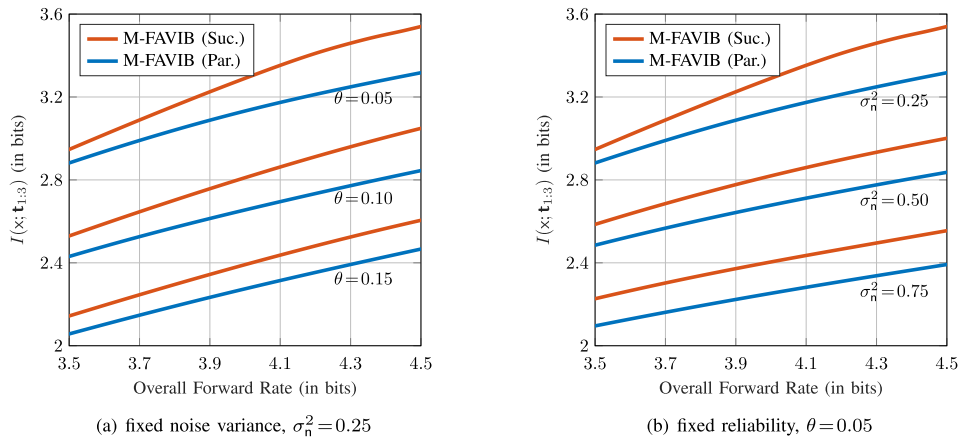


Fig. 7. End-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, vs. overall forward rate, $\sum_j I(y_j; \mathbf{z}_j | \mathbf{t}_{1:j-1})$ successive / $\sum_j I(y_j; \mathbf{z}_j)$ parallel, equiprobable 16-QAM signaling ($\sigma_x^2 = 10$), discrete AWGN access channels with the noise variance, σ_n^2 , allowed output clusters, $N = 4$, symmetric 4×4 forward channels with the reliability value, θ , $1.5 \leq \beta_j \leq 3.5$ for $j = 1, 2, 3$, the convergence parameter, $\varepsilon = 10^{-3}$.

across branches, namely, $\sigma_{n_1}^2 = 0.3, \sigma_{n_2}^2 = 0.3, \sigma_{n_3}^2 = 1.0$ (with $I(\mathbf{x}; \mathbf{y}_{1:3}) \approx 3.98$) and varied the reliability value of the forward channels (the same value, θ , across branches), namely, $\theta = 0.005, 0.03, 0.06$. Contrarily, in Fig. 6b, we considered an asymmetric specification regarding the forward channels by choosing different reliability values across branches, namely, $\theta_1 = 0.001, \theta_2 = 0.001, \theta_3 = 0.4$ and varied the noise variance of the access channels (the same value, σ_n^2 , across branches), namely, $\sigma_n^2 = 0.25, 0.40, 0.55$ (with $I(\mathbf{x}; \mathbf{y}_{1:3}) \approx 4, 3.98, 3.91$ bits, respectively) for capturing the effects of both present channels (access and forward) in the presumed system model. It can be observed that the joint design of local quantizers results in a superior end-to-end transmission rate, $I(\mathbf{x}; \mathbf{t}_{1:3})$, compared to the simplest approach of implementing separately designed quantizers on different branches.

C. Parallel Vs. Successive Retrieval

In the last part, we consider again the same setup as before, but with completely symmetric specifications regarding both the access and forward channels (analogous to the first round of investigations in subsection V-A), i.e., an equiprobable 16-QAM source signaling over $J = 3$ branches, each featuring a discrete memoryless access channel (160 samples per branch), approximating a discrete-time discrete-input continuous-output AWGN channel with the noise variance, σ_n^2 (same value across branches), and a symmetric $N \times N$ forward channel model with the reliability parameter, θ (same value across branches). The quantizers' allowed output clusters are fixed to $N = 4$, and the β_j values (for $j = 1, 2, 3$) are varied from 1.5 to 3.5. The end-to-end transmission rates and the forward (compression) sum rates have been calculated for two cases of utilizing the M-FAVIB with both approaches of the parallel and successive retrieval. The obtained results have been illustrated in Fig. 7.

Explicitly, in Fig. 7a, we fixed the noise variance of the access channels to $\sigma_n^2 = 0.25$, and varied the reliability value of the forward channels, namely, $\theta = 0.05, 0.10, 0.15$. Contrarily, in Fig. 7b, we fixed the reliability value of the forward channels to $\theta = 0.05$, and varied the access channels' noise variance, namely, $\sigma_n^2 = 0.25, 0.50, 0.75$ (with

$I(\mathbf{x}; \mathbf{y}_{1:3}) \approx 4, 3.95, 3.75$ bits, respectively). As the main takeaway, it can be immediately observed from both results that, under identical specifications, the utilization of available side-information can decrease the required overall forward rate for supporting a certain level of the end-to-end transmission rate compared to the parallel scheme wherein the correlations among different forward channels' outputs is totally neglected. This can be clearly rationalized by noting the fact that, due to the presumed Markovian properties, conditioning on the previous forward channels' outputs can either deduct from the current unconditional compression rate or keep it unchanged as it applies

$$I(y_j; \mathbf{z}_j | \mathbf{t}_{1:j-1}) = I(y_j; \mathbf{z}_j) - I(\mathbf{z}_j; \mathbf{t}_{1:j-1}), \quad (40)$$

and the MI is non-negative.

VI. SUMMARY

We concentrated on a particular multiterminal *Joint Source-Channel Coding* problem wherein several noisy observations from a remote source are compressed *locally* (yet following a *joint* design) prior to getting forwarded to a (distant) processing unit over multiple *error-prone* channels. Herein, we adapted the fundamental idea of the *Information Bottleneck* method and proposed two distributed compression schemes. For that, we obtained the form of stationary points regarding the individual local quantizers and also presented a generic iterative algorithm, the M-FAVIB, which extends the schemes MIB [25] and FAVIB [49]. Applying an unfolding trick, we linked the proposed algorithm to the *Successive Upper-Bound Minimization*, thereby providing the proof of convergence to a stationary point. After an in-depth analysis of the proposed method, we also substantiated its effectiveness by means of numerical simulations.

APPENDIX

A. Proof of Theorem 1

As the parallel IB-based JSCC Lagrangian, $\mathcal{L}_{\text{Par.}}$, in (6) is a functional of all the individual local quantizer mappings $\{p(\mathbf{z}_j | y_j) | j\}$, to come into a stationary point of it, its derivative w.r.t. every quantizer mapping, $p(\mathbf{z}_j | y_j)$, must be equated

to zero. Associating a Lagrange multiplier, λ_{y_j} , per realization, $y_j \in \mathcal{Y}_j$, of the observation, y_j , one can incorporate the validity conditions into the overall parallel JSCC Lagrangian, $\mathcal{L}_{\text{Par.}}^{\text{Ov.}}$, for which it applies

$$\begin{aligned} \mathcal{L}_{\text{Par.}}^{\text{Ov.}} = & I(x; \mathbf{t}_{1:J}) - \sum_{j=1}^J \lambda_j I(y_j; z_j) \\ & + \sum_{j=1}^J \sum_{y_j \in \mathcal{Y}_j} \lambda_{y_j} \left(\sum_{z_j \in \mathcal{Z}_j} p(z_j|y_j) - 1 \right). \end{aligned} \quad (41)$$

Fixing $\{p(z_\ell|y_\ell) \mid \ell \neq j\}$ and taking the functional derivative w.r.t. $p(z_j|y_j)$, it holds

$$\frac{\delta \left(\sum_{j=1}^J \lambda_j I(y_j; z_j) \right)}{\delta p(z_j|y_j)} = \lambda_j \frac{\delta I(y_j; z_j)}{\delta p(z_j|y_j)} = \lambda_j p(y_j) \log \frac{p(z_j|y_j)}{p(z_j)}, \quad (42)$$

and

$$\frac{\delta \left(\sum_{j=1}^J \sum_{y_j \in \mathcal{Y}_j} \lambda_{y_j} \left(\sum_{z_j \in \mathcal{Z}_j} p(z_j|y_j) - 1 \right) \right)}{\delta p(z_j|y_j)} = \lambda_{y_j}. \quad (43)$$

Therefore, to be able to calculate $\frac{\delta \mathcal{L}_{\text{Par.}}^{\text{Ov.}}}{\delta p(z_j|y_j)}$, the only remaining part is to derive $\frac{\delta I(x; \mathbf{t}_{1:J})}{\delta p(z_j|y_j)}$. By applying the chain rule for MI to decompose $I(x; \mathbf{t}_{1:J})$, it becomes rather undemanding to discern that

$$\frac{\delta I(x; \mathbf{t}_{1:J})}{\delta p(z_j|y_j)} = \frac{\delta I(x; \mathbf{t}_j | \mathbf{t}_{1:J}^{-j})}{\delta p(z_j|y_j)}, \quad (44)$$

as fixing $\{p(z_\ell|y_\ell) \mid \ell \neq j\}$ also fixes the other appearing terms in the conducted decomposition. Noting the present Markov chain, $y_j \leftrightarrow \mathbf{z}_{1:J}^{-j} \leftrightarrow \mathbf{t}_{1:J}^{-j}$, it holds

$$\begin{aligned} \frac{\delta I(x; \mathbf{t}_j | \mathbf{t}_{1:J}^{-j})}{\delta p(z_j|y_j)} &= \frac{\delta \left(H(\mathbf{t}_j | \mathbf{t}_{1:J}^{-j}) - H(\mathbf{t}_j | x, \mathbf{t}_{1:J}^{-j}) \right)}{\delta p(z_j|y_j)} \\ &= \sum_{z_{1:J}^{-j}} p(y_j, z_{1:J}^{-j}) \sum_x p(x|y_j, z_{1:J}^{-j}) \\ &\quad \times \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J} | z_{1:J}^{-j}) \log \frac{p(x|\mathbf{t}_{1:J})}{p(x|\mathbf{t}_{1:J}^{-j})}. \end{aligned} \quad (45)$$

Due to the positivity of $p(y_j)$ and by applying the stationary condition, $\frac{\delta \mathcal{L}_{\text{Par.}}^{\text{Ov.}}}{\delta p(z_j|y_j)} = 0$, it is directly deduced from (42), (43) and (45) that

$$\begin{aligned} \sum_{z_{1:J}^{-j}} p(z_{1:J}^{-j} | y_j) \sum_x p(x|y_j, z_{1:J}^{-j}) \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J} | z_{1:J}^{-j}) \log \frac{p(x|\mathbf{t}_{1:J})}{p(x|\mathbf{t}_{1:J}^{-j})} \\ - \lambda_j \log \frac{p(z_j|y_j)}{p(z_j)} + \frac{\lambda_{y_j}}{p(y_j)} = 0. \end{aligned} \quad (46)$$

Thus, by definition of the KL divergence, it holds

$$\begin{aligned} - \sum_{z_{1:J}^{-j}} p(z_{1:J}^{-j} | y_j) \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J} | z_{1:J}^{-j}) D_{\text{KL}}(p(x|y_j, z_{1:J}^{-j}) \| p(x|\mathbf{t}_{1:J})) \\ - \lambda_j \log \frac{p(z_j|y_j)}{p(z_j)} + \tilde{\lambda}_{y_j}^{\text{Par.}} = 0, \end{aligned} \quad (47)$$

with $\tilde{\lambda}_{y_j}^{\text{Par.}}$, being equal to

$$\begin{aligned} \tilde{\lambda}_{y_j}^{\text{Par.}} = & \frac{\lambda_{y_j}}{p(y_j)} + \sum_{z_{1:J}^{-j}} p(z_{1:J}^{-j} | y_j) \sum_{\mathbf{t}_{1:J}^{-j}} p(\mathbf{t}_{1:J}^{-j} | z_{1:J}^{-j}) \\ & \times D_{\text{KL}}(p(x|y_j, z_{1:J}^{-j}) \| p(x|\mathbf{t}_{1:J}^{-j})). \end{aligned} \quad (48)$$

Bringing the second summand in (47) to the other side of equality, multiplying both sides by $\beta_j = \frac{1}{\tilde{\lambda}_j^{\text{Par.}}}$, exponentiating them, and, eventually, multiplying by $p(z_j)$, it applies

$$p(z_j|y_j) = p(z_j) \exp \left(-d_{\text{Par.}}(y_j, z_j) + \beta_j \tilde{\lambda}_{y_j}^{\text{Par.}} \right). \quad (49)$$

Enforcing the validity condition, $\sum_{z_j} p(z_j|y_j) = 1$, and noting that $\tilde{\lambda}_{y_j}^{\text{Par.}}$ is independent of z_j , one can treat $\exp(-\beta_j \tilde{\lambda}_{y_j}^{\text{Par.}})$ as the partition function, $\psi_{z_j}^{\text{Par.}}$, to come into the form of (7). ■

B. Proof of Theorem 2

The successive IB-based JSCC Lagrangian, $\mathcal{L}_{\text{Suc.}}$, in (12) is a functional of all the individual local quantizer mappings $\{p(z_j|y_j) \mid j\}$. To come into a stationary point of it, its derivative w.r.t. every quantizer mapping, $p(z_j|y_j)$, must be equated to zero. Associating a Lagrange multiplier, λ_{y_j} , for every realization, $y_j \in \mathcal{Y}_j$, of the observation, y_j , one can incorporate the validity conditions into the overall successive JSCC Lagrangian, $\mathcal{L}_{\text{Suc.}}^{\text{Ov.}}$, for which it applies

$$\begin{aligned} \mathcal{L}_{\text{Suc.}}^{\text{Ov.}} = & I(x; \mathbf{t}_{1:J}) - \sum_{j=1}^J \lambda_j I(y_j; z_j | \mathbf{t}_{1:j-1}) \\ & + \sum_{j=1}^J \sum_{y_j \in \mathcal{Y}_j} \lambda_{y_j} \left(\sum_{z_j \in \mathcal{Z}_j} p(z_j|y_j) - 1 \right). \end{aligned} \quad (50)$$

Fixing $\{p(z_\ell|y_\ell) \mid \ell \neq j\}$, to be able to calculate the functional derivative of $\mathcal{L}_{\text{Suc.}}^{\text{Ov.}}$ w.r.t. $p(z_j|y_j)$, only the corresponding derivative of its second term must be determined at this point as the pertinent derivatives of its first and last terms are already given in (45) and (43), respectively. To do so, it has to be noted that

$$\frac{\delta \left(\sum_{\ell=1}^J \lambda_\ell I(y_\ell; z_\ell | \mathbf{t}_{1:\ell-1}) \right)}{\delta p(z_j|y_j)} = \frac{\delta \left(\sum_{\ell=j}^J \lambda_\ell I(y_\ell; z_\ell | \mathbf{t}_{1:\ell-1}) \right)}{\delta p(z_j|y_j)}, \quad (51)$$

since the first $j-1$ appearing summands in the numerator do not depend on $p(z_j|y_j)$. Moreover, noting the presumed Markov properties and via some intermediate manipulations, it holds

$$\begin{aligned} & \frac{\delta I(y_j; z_j | \mathbf{t}_{1:j-1})}{\delta p(z_j|y_j)} \\ &= \frac{\delta \left(H(z_j | \mathbf{t}_{1:j-1}) - H(z_j | \mathbf{t}_{1:j-1}, y_j) \right)}{\delta p(z_j|y_j)} \\ &= \frac{\delta H(z_j | \mathbf{t}_{1:j-1})}{\delta p(z_j|y_j)} - \frac{\delta H(z_j | y_j)}{\delta p(z_j|y_j)} = p(y_j) \\ &\quad \times \left[\log \frac{p(z_j|y_j)}{p(z_j)} - \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1} | y_j) \log \frac{p(\mathbf{t}_{1:j-1} | z_j)}{p(\mathbf{t}_{1:j-1})} \right], \end{aligned} \quad (52)$$

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

$$\begin{aligned} \frac{\delta I(y_\ell; z_\ell | \mathbf{t}_{1:\ell-1})}{\delta p(z_j | y_j)} &= \frac{\delta H(z_\ell | \mathbf{t}_{1:\ell-1})}{\delta p(z_j | y_j)} - \frac{\delta H(z_\ell | \mathbf{t}_{1:\ell-1}, y_\ell)}{\delta p(z_j | y_j)} \\ &= \frac{\delta H(z_\ell | \mathbf{t}_{1:\ell-1})}{\delta p(z_j | y_j)} - \underbrace{\frac{\delta H(z_\ell | y_\ell)}{\delta p(z_j | y_j)}}_0 \\ &= p(y_j) \sum_{\mathbf{t}_{1:\ell-1}, z_\ell} p(t_j | z_j) p(\mathbf{t}_{1:\ell-1}^{-j}, z_\ell | y_j) \\ &\quad \times \log \frac{1}{p(z_\ell | \mathbf{t}_{1:\ell-1})}. \end{aligned} \quad (53)$$

Due to the positivity of $p(y_j)$ and by applying the stationary condition, i.e., $\frac{\delta \mathcal{L}_{\text{Suc}}^{\text{ov}}}{\delta p(z_j | y_j)} = 0$, it is inferred from (43), (45), (52) and (53) that

$$\begin{aligned} & - \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j} | y_j) \sum_{\mathbf{t}_{1:J}} p(\mathbf{t}_{1:J} | \mathbf{z}_{1:J}) \\ & \quad \times D_{\text{KL}}(p(x | y_j, \mathbf{z}_{1:J}^{-j}) \| p(x | \mathbf{t}_{1:J})) \\ & - \lambda_j \log \frac{p(z_j | y_j)}{p(z_j)} + \lambda_j \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1} | y_j) \\ & \quad \times \log p(\mathbf{t}_{1:j-1} | z_j) + \tilde{\lambda}_{y_j}^{\text{Suc}} \\ & + \sum_{k=j+1}^J \lambda_k \sum_{\mathbf{t}_{1:k-1}, z_k} p(t_j | z_j) p(\mathbf{t}_{1:k-1}^{-j}, z_k | y_j) \\ & \quad \times \log p(z_k | \mathbf{t}_{1:k-1}) = 0, \end{aligned} \quad (54)$$

with $\tilde{\lambda}_{y_j}^{\text{Suc}}$, being equal to

$$\begin{aligned} \tilde{\lambda}_{y_j}^{\text{Suc}} &= -\lambda_j \sum_{\mathbf{t}_{1:j-1}} p(\mathbf{t}_{1:j-1} | y_j) \\ & \quad \times \log p(\mathbf{t}_{1:j-1}) + \frac{\lambda_{y_j}}{p(y_j)} \\ & + \sum_{\mathbf{z}_{1:J}^{-j}} p(\mathbf{z}_{1:J}^{-j} | y_j) \sum_{\mathbf{t}_{1:J}^{-j}} p(\mathbf{t}_{1:J}^{-j} | \mathbf{z}_{1:J}^{-j}) \\ & \quad \times D_{\text{KL}}(p(x | y_j, \mathbf{z}_{1:J}^{-j}) \| p(x | \mathbf{t}_{1:J}^{-j})). \end{aligned} \quad (55)$$

Bringing the second component in (54) to the other side of equality, multiplying both sides by $\beta_j = \frac{1}{\lambda_j}$, exponentiating them, and, eventually, multiplying by $p(z_j)$, it applies

$$p(z_j | y_j) = p(z_j) \exp \left(-d_{\text{Suc}}(y_j, z_j) + \beta_j \tilde{\lambda}_{y_j}^{\text{Suc}} \right). \quad (56)$$

Enforcing the validity condition, $\sum_{z_j} p(z_j | y_j) = 1$, and noting that $\tilde{\lambda}_{y_j}^{\text{Suc}}$ is independent of z_j , one can treat $\exp(-\beta_j \tilde{\lambda}_{y_j}^{\text{Suc}})$ as the partition function, $\psi_{z_j}^{\text{Suc}}$, to come into the form of (13). ■

REFERENCES

- [1] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Sep. 1999, pp. 368–377.
- [2] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 2005.
- [3] N. Slonim, "The information bottleneck: Theory and applications," Ph.D. dissertation, Dept. Interdiscipl. Center Neural Comput., Hebrew Univ. Jerusalem, Jerusalem, Israel, 2002. [Online]. Available: <http://www.yaroslavvb.com/papers/slonim-information.pdf>
- [4] S. Hassanpour, D. Wübben, and A. Dekorsy, "Overview and investigation of algorithms for the information bottleneck method," in *Proc. 11th Int. ITG Conf. Syst., Commun. Coding*, Hamburg, Germany, Feb. 2017, pp. 1–6.
- [5] A. Zaidi, I. Estella-Aguerri, and S. Shamai (Shitz), "On the information bottleneck problems: Models, connections, applications and information theoretic views," *Entropy*, vol. 22, no. 2, p. 151, Jan. 2020.
- [6] Z. Goldfeld and Y. Polyanskiy, "The information bottleneck problem and its applications in machine learning," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 19–38, May 2020.
- [7] A. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.
- [8] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.
- [9] E. Erkip and T. M. Cover, "The efficiency of investment information," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1026–1040, May 1998.
- [10] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [11] Y. Steinberg, "Coding and common reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4995–5010, Nov. 2009.
- [12] P. Harremoës and N. Tishby, "The information bottleneck revisited or how to choose a good distortion measure," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 566–570.
- [13] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 5, pp. 293–304, Sep. 1962.
- [14] D. Sakrison, "Source encoding in the presence of random disturbance (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 165–167, Jan. 1968.
- [15] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 406–411, Jul. 1970.
- [16] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 5, pp. 518–521, Sep. 1980.
- [17] G. Zeitler, A. C. Singer, and G. Kramer, "Low-precision A/D conversion for maximum information rate in channels with memory," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2511–2521, Sep. 2012.
- [18] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.
- [19] S. A. A. Shah, M. Stark, and G. Bauch, "Coarsely quantized decoding and construction of polar codes using the information bottleneck method," *Algorithms*, vol. 12, no. 9, Sep. 2019, Art. no. 192.
- [20] M. Stark, L. Wang, G. Bauch, and R. D. Wesel, "Decoding rate-compatible 5G-LDPC codes with coarse quantization using the information bottleneck method," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 646–660, May 2020.
- [21] N. Slonim, N. Friedman, and N. Tishby, "Multivariate information bottleneck," *Neural Comput.*, vol. 18, no. 8, pp. 1739–1789, Aug. 2006.
- [22] 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.
- [23] I. Estella-Aguerri, A. Zaidi, G. Caire, and S. Shamai (Shitz), "On the capacity of cloud radio access networks with oblivious relaying," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4575–4596, Jul. 2019.
- [24] I. Estella-Aguerri and A. Zaidi, "Distributed variational representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 120–138, Jan. 2021.
- [25] S. Hassanpour, D. Wübben, and A. Dekorsy, "A novel approach to distributed quantization via multivariate information bottleneck method," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.
- [26] Y. Uğur, I. Estella-Aguerri, and A. Zaidi, "Vector Gaussian CEO problem under logarithmic loss and applications," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4183–4202, Jul. 2020.
- [27] S. Hassanpour, D. Wübben, and A. Dekorsy, "Generalized distributed information bottleneck for fronthaul rate reduction at the cloud-RANs uplink," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Taipei, Taiwan, Dec. 2020, pp. 1–6.
- [28] S. Movaghati and M. Ardakani, "Distributed channel-aware quantization based on maximum mutual information," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 5, May 2016, Art. no. 3595389.
- [29] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.
- [30] D. Wübben *et al.*, "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 35–44, Nov. 2014.

- [31] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.
- [32] G. Zeidler, G. Bauch, and J. Widmer, "Quantize-and-Forward schemes for the orthogonal multiple-access relay channel," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1148–1158, Apr. 2012.
- [33] I. Avram, N. Aerts, H. Bruneel, and M. Moeneclaey, "Quantize and forward cooperative communication: Channel parameter estimation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1167–1179, Mar. 2012.
- [34] C. Novak, C. Studer, A. Burg, and G. Matz, "The effect of unreliable LLR storage on the performance of MIMO-BICM," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2010, pp. 736–740.
- [35] C. Roth, C. Benkeser, C. Studer, G. Karakostas, and A. Burg, "Data mapping for unreliable memories," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Oct. 2012, pp. 679–685.
- [36] A. Kurtenbach and P. Wintz, "Quantizing for noisy channels," *IEEE Trans. Commun.*, vol. COM-17, no. 2, pp. 291–302, Apr. 1969.
- [37] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source–channel coding," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 827–838, Nov. 1987.
- [38] K. Zeger and A. Gersho, "Pseudo-gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.
- [39] J. B. De Marca and N. S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multi-dimensional quantizer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seattle, WA, USA, Jun. 1987, pp. 1128–1132.
- [40] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Commun. Jpn. I, Commun.*, vol. 67, no. 4, pp. 39–47, 1984.
- [41] E. Ayanoglu and R. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 855–865, Nov. 1987.
- [42] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [43] A. Winkelbauer, G. Matz, and A. Burg, "Channel-optimized vector quantization with mutual information as fidelity criterion," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2013, pp. 851–855.
- [44] S. Hassanpour, D. Wübben, and A. Dekorsy, "A graph-based message passing approach for joint source-channel coding via information bottleneck principle," in *Proc. IEEE 10th Int. Symp. Turbo Codes Iterative Inf. Process. (ISTC)*, Hong Kong, Dec. 2018, pp. 1–5.
- [45] S. B. Zahir Azami, P. Duhamel, and O. Rioul, "Joint source-channel coding: Panorama of methods," in *Proc. CNES Workshop Data Compress.*, Toulouse, France, Nov. 1996, pp. 1232–1254.
- [46] F. Hekland, "A review of joint source-channel coding," Dept. Electron. Telecommun., Norwegian Univ. Sci. Technol., Trondheim, Norway, Tech. Rep., Feb. 2004. [Online]. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.3932&rep=rep1&type=pdf>
- [47] J. Garcia-Frias and Z. Xiong, "Distributed source and joint source-channel coding: From theory to practice," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Philadelphia, PA, USA, Mar. 2005, pp. 1093–1096.
- [48] X. Zhu, Y. Liu, and L. Zhang, "Distributed joint source-channel coding in wireless sensor networks," *Sensors*, vol. 9, no. 6, pp. 4901–4917, Jun. 2009.
- [49] S. Hassanpour, T. Monsees, D. Wübben, and A. Dekorsy, "Forward-aware information bottleneck-based vector quantization for noisy channels," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7911–7926, Dec. 2020.
- [50] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
- [51] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.
- [52] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 4, pp. 142–163, Mar. 1959.
- [53] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.
- [54] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
- [55] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–94, Sep. 2004.
- [56] J. H. Mathews and K. D. Fink, *Numerical Methods Using MATLAB*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.
- [57] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004. [Online]. Available: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf
- [58] R. Horst, P. M. Pardalos, and N. Van Thoai, *Introduction to Global Optimization*, 2nd ed. New York, NY, USA: Springer, 2000.
- [59] S. Hassanpour, D. Wübben, A. Dekorsy, and B. M. Kurkoski, "On the relation between the asymptotic performance of different algorithms for information bottleneck framework," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.
- [60] D. Hankerson, G. A. Harris, and P. D. Johnson, Jr., *Introduction to Information Theory and Data Compression*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2003.



Shayan Hassanpour (Member, IEEE) received the B.Sc. degree in electrical engineering (electronics) from the University of Mazandaran, Babolsar, Iran, in 2011, and the M.Sc. degree in communications engineering from the Ulm University, Germany, in 2014. He is currently pursuing the Ph.D. degree with the University of Bremen, Germany. Over the last four years, he has been prolifically contributing to the international journals and IEEE flagship conferences on his Ph.D. topic, i.e., the information bottleneck method. His other research interests include information theory (source and joint source-channel coding), multiple antenna systems, wireless/mobile communications, statistical signal processing, and the application of machine learning in the design of communication systems.



Dirk Wübben (Senior Member, IEEE) received the Dipl.-Ing. (FH) degree in electrical engineering from the University of Applied Science Münster, Germany, in 1998, and the Dipl.-Ing. (Uni.) and Dr.-Ing. degrees in electrical engineering from the University of Bremen, Germany, in 2000 and 2005, respectively. He is currently a Senior Research Group Leader and a Lecturer with the Department of Communications Engineering, University of Bremen. He has published more than 140 papers in international journals and conference proceedings.

His research interests include wireless communications, signal processing, multiple antenna systems, cooperative communication systems, channel coding, information theory, and machine learning. He is a Board Member of the Germany Chapter of the IEEE Information Theory Society and a member of VDE/ITG Expert Committee "Information and System Theory." He has been an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS.



Armin Dekorsy (Senior Member, IEEE) is currently the Head of the Department of Communications Engineering, University of Bremen. He has more than ten years of industrial experience in leading research positions, such as an DMTS at Bell Labs Europe and the Head of Research Europe Qualcomm, Nuremberg, and by conducting (international) research projects (more than 25 BMBF/BMWI/EU projects) in affiliation with his scientific expertise shown by more than 200 journals and conference publications and holds more than

19 patents. His current research focuses on distributed signal processing, compressed sampling, information bottleneck method, and machine learning leading to the further development of communication technologies for 5G/6G, industrial wireless communications, and NewSpace satellite communications. He investigates new lines of research in wireless communication and signal processing for the baseband of transceivers of future communication systems, the results of which are transferred to the pre-development of industry through political and strategic activities. He is a Senior Member of the IEEE Communications and Signal Processing Society, the Head of VDE/ITG Expert Committee "Information and System Theory," and a member of Executive Board of the Technologie-Zentrum Informatik und Informationstechnik (TZI), University of Bremen.