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 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...
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.a., 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, $a$, with the probability mass function, $p(a)$, takes particular realizations, $a$, from its domain, $A$. The same also applies to the random vector, $a_{1:T} = [a_1, \ldots, a_J]$ with boldface counterparts. $a_j$ excludes $a_j$ from $a_{1:j}$. $H(\cdot)$, $D_{KL}(\cdot|\cdot)$, $I(\cdot;\cdot)$ and $D_{JS}^{(\cdot)}(\cdot|\cdot)$ 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).\footnote{Although 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 descriptions.} For that, the illustrated system model in Fig. 1 has been considered. A data source, $x$, is observed via a discrete memoryless (access) channel with the transition probabilities, $p(y|x)$. The observation, $y$, is then compressed into the signal, $z$, ahead of transmission over another discrete, memoryless (forward) channel with the transition probabilities, $p(t|z)$, 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, $I(x; t)$, which is the MI between the source, $x$, and the forward channel...
output signal, $t$, 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

$$p^*(z|y) = \arg\max_{p(z|y)} I(x; t),$$  \hspace{1cm} (1)

with $0 \leq R \leq \log_2 |Z|$ 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)

$$p^*(z|y) = \arg\max_{p(z|y)} I(x; t) - \lambda I(y; z),$$  \hspace{1cm} (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 $(y, z) \in Y \times Z$)

$$p(z|y) = \frac{p(z)}{\psi(y, \beta)} \exp\left(-\beta \sum_{t \in T} p(t|z) D_{KL}(p(x|y)||p(x|t))\right),$$  \hspace{1cm} (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, $x$, 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, $x$. Analogous to the uniterminal scenario, we presume that the source distribution, $p(x)$, the access channels’ transition probabilities, $p(y_j|x)$, and the forward channels’ statistics, $p(t_j|z_j)$, are available for all branches. We presume that the Markovian independence relation $x \leftrightarrow y_j \leftrightarrow z_j \leftrightarrow t_j$ applies per branch, $j$, and the counterpart signals of every two distinct branches are conditionally independent given the source signal, $x$.

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, $I(x; t_{1:J})$, 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, $P^* = \{p^*(z_1|y_1), \ldots, p^*(z_J|y_J)\}$, where

$$P^* = \arg\max_{P: \forall j \leq \sum_{i} R_i} I(x; t_{1:J}),$$  \hspace{1cm} (4)
with 0 \leq R_j \leq \log_2 |Z_j| \text{ bits, setting an upper-bound on the} 
\text{the } j\text{-th compression rate, } I(y_j;z_j). \text{ Employing the LM} 
\text{method, the design problem (4) can be reformulated as an} 
\text{unconstrained optimization (up to the pertinent mappings’} 
\text{validity)} 

P^* = \arg \max_P I(x; t_{1:j}) - \sum_{j=1}^J \lambda_j I(y_j; z_j), \quad (5) 

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

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

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

if and only if for each pair \((y_j, z_j) \in Y_j \times Z_j\) 

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

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

d_{\text{Par.}}(y_j, z_j) = \beta_j \sum_{z_{1:j}^*} p(z_{1:j}^*|y_j) \sum_{t_{1:j}} p(t_{1:j}|z_{1:j}) 
\times D_{\text{KL}}(p(x|y_j, z_{1:j}^*) || p(x|z_{1:j})), \quad (8) 

with \beta_j = \frac{1}{x_j} \text{ and } p(t_{1:j}|z_{1:j}) = \prod_{j=1}^J p(t_j|z_j), \text{ resulting from} 
\text{presumed Markov properties.} 

The proof has been presented in Appendix A. 

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

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

\textit{Note 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. 

\textbf{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 \text{to the previous setup, from a compression’s perspective, the} 
\text{side-information is exploited when treating the observation, } y_j. \text{ The principal idea behind this scheme is fully} 
\text{aligned with the well-known Wyner-Ziv setup} [10] \text{ for source} 
\text{coding in which a statistically correlated signal is utilized as} 
\text{side-information at the decoder. With the major difference of} 
\text{assuming ideal (rate-limited) forward channels, in [27],} 

\textit{The two arguments on its subscript differentiate this notation from the} 
one that is used for functional derivatives. 

\textit{Note that the choice of forward channel is arbitrary, and the BSC has been} 
taken here just as an exemplary error-prone channel.
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, \( P^* = \{ p^* (z_1 | y_1), \cdots, p^* (z_J | y_J) \} \), where

\[
P^* = \arg \max_{P: \forall y_j} I (x; t_{1:j}), \quad (10)
\]

with \( 0 \leq R_j \leq \log_2 |Z_j| \) bits, stipulating an upper-bound on the \( j \)-th conditional compression rate, \( I (y_j; z_j | t_{1:j-1}) \). 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)

\[
P^* = \arg \max_P I (x; t_{1:j}) - \sum_{j=1}^{J} \lambda_j I (y_j; z_j | t_{1:j-1}), \quad (11)
\]

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 \( p(x, y_{1:j}) = p(x) \prod_{j} p(y_j | x) \), the forward channels’ statistics, \( p(t_j | z_j) \), and \( \lambda_j \) are available for all \( j = 1 \) to \( J \). The set of local quantizers, \( \{ p(z_j | y_j) \} \), is a stationary point of the successive IB-based JSCC Lagrangian

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

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

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

with \( \psi_{\text{Suc.}} (y_j, \beta_j) \), 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

\[
d_{\text{Suc.}} (y_j, z_j) = \beta_j \sum_{z_j^{t_{1:j}}} p (z_j^{t_{1:j}} | y_j) \sum_{t_{1:j}} p (t_{1:j} | z_j^{t_{1:j}})
\]

\[	imes \mathcal{D}_{KL} (p (x | y_j, z_j^{t_{1:j}}) \| p (x | t_{1:j})) - \beta_j \sum_{n=j+1}^{J} \beta_n \sum_{z_n} p (t_{1:n-1} | z_n) \log p (z_n | t_{1:n-1}),
\]

(14)

with \( \beta_j = \frac{1}{\lambda_j} \) and \( p (t_{1:j} | z_{1:j}) = \prod_{j=1}^{J} p (t_j | z_j) \), 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 \( I (x; z_{1:j}) \), under the adapted set of constraints on individual conditional compression rates, i.e., for the \( j \)-th branch it must apply \( I (y_j; z_j | t_{1:j-1}) \leq R_j \). By 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 denotation), the calculated relevant distortion in (14) boils down to

\[
d_{\text{Suc.}}^{\text{dec.}} (y_j, z_j) = \beta_j \sum_{z_j^{t_{1:j}}} p (z_j^{t_{1:j}} | y_j) \mathcal{D}_{KL} (p (x | y_j, z_j^{t_{1:j}}) \| p (x | z_j^{t_{1:j}})) - \beta_j \sum_{n=j+1}^{J} \beta_n \sum_{z_n} p (t_{1:n-1} | z_n) \log p (z_n | t_{1:n-1}),
\]

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, $z_{\pi}(j)$, up to the retrieved one, $t_{\pi}(j)$, 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, $t_{\pi(1):\pi(j-1)}$, can be utilized as an available side-information when applying the pertinent demapping to retrieve $t_{\pi(j)}$, since all demapper blocks are located in the (remote) processing unit. 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

$$p(z_j|y_j) = \frac{p(z_j)}{\psi^*_j(y_j, \beta_j)} \exp \left( -d_r(y_j, z_j) \right),$$

with $r \in \{\text{Par}, \text{Suc}\}$. The right side of (16) can be interpreted as a functional featuring all the individual quantizer mappings, $\{p(z_j|y_j)\}$, 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)

$$p(z_j|y_j) = \mathcal{F}_j(p(z_j|y_1), \ldots, p(z_j|y_J)),$$

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, $\{p^{(0)}(z_j|y_j)\}$, for each pair, $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$, the updates are executed (till convergence by $\varepsilon \ll 1$) through

$$p^{(i+1)}(z_j|y_j) = \frac{p^{(i)}(z_j)}{\psi^*_j(z_j, \beta_j)} \exp \left( -d_r(y_j, z_j) \right),$$

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, $\{p^{(i)}(z_j|y_j)\}$, to marginalize the joint distribution, $p^{(i)}(x,y_{1:J},z_{1:J},t_{1:J})$. The updates are performed synchronously, i.e., at iteration $i + 1$, all the individual quantizer mappings of different branches, $\{p^{(i+1)}(z_j|y_j)\}$, will be updated based on the previous configuration of the same set, i.e., $\{p^{(i)}(z_j|y_j)\}$. 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, $\{p^{(0)}(z_j|y_j)\}$, 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, $\mathcal{F}_r(P, Q)$.
on the respective objective functional, $\mathcal{F}_P(P)$. 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

$$
\mathcal{F}_{\text{Par}}(P) = H(x) - L_{\text{Par}} = \sum_{j=1}^{J} \lambda_j I(y_j ; z_j) + H(x_{t_1:j}) \tag{19}
$$

over $P$. By defining $Q = \{q(x|t_{1:j})\}$ and the functional

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

the following Lemmas hold:

**Lemma 1:** It applies

$$
\min_{P} \mathcal{F}_{\text{Par}}(P) = \min_{P} \min_{Q} \mathcal{F}_{\text{Par}}(P,Q). \tag{21}
$$

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

$$
\mathcal{F}_{\text{Par}}(P,Q) - \mathcal{F}_{\text{Par}}(P) = \sum_{t_{1:j}} p(t_{1:j}) \tag{22}
$$

$$
x \times \mathbb{D}_{\text{KL}}(p(x|t_{1:j}) || q(x|t_{1:j})) \geq 0, \tag{22}
$$

where the equality holds iff $q(x|t_{1:j}) = p(x|t_{1:j})$.

**Lemma 2:** $\mathcal{F}_{\text{Par}}(P,Q)$ 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 $\mathcal{F}_{\text{Par}}(P,Q)$, given by

$$
q^*(x|t_{1:j}) = p(x|t_{1:j}), \tag{23}
$$

where $p(x|t_{1:j})$ 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 $\mathcal{F}_{\text{Par}}(P,Q)$, given by

$$
p^*(z_j)|y_j) = \frac{p(z_j)}{\hat{\psi}_{y_j}(y_j, \beta_j)} \exp \left( \frac{\bar{d}_{\text{Par}}(y_j, z_j)}{1} \right) \tag{24}
$$

where $\hat{\psi}_{y_j}(y_j, \beta_j)$, 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

$$
\bar{d}_{\text{Par}}(y_j, z_j) = \beta_j \sum_{z_{1:j}^j} p(z_{1:j}^j|y_j) \times \mathbb{D}_{\text{KL}}(p(x|y_j, z_{1:j}^j) || q(x|t_{1:j})). \tag{25}
$$

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

$$
\frac{1}{\delta p(z_j|y_j)} \sum_{t_{1:j}} p(t_{1:j}|z_{1:j}) \times \sum_{t_{1:j}} p(t_{1:j}|z_{1:j}) \times \mathbb{E}_{x,y_j} \{ \log q(x|t_{1:j}) \} - \mathbb{E}_{x,y_j} \{ \log q(x|t_{1:j}) \}. \tag{26}
$$

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 $\mathcal{F}_P(P,Q)$ and $\mathcal{F}_{\text{Par}}(P)$ satisfy [51, Prop. 1].

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

$$
\mathcal{F}_{\text{Suc.}}(P,Q) = H(x) - L_{\text{Suc.}} = \sum_{j=1}^{J} \lambda_j I(y_j ; z_j) \tag{27}
$$

$$
- \sum_{j=2}^{J} \lambda_j I(z_j ; t_{1:j-1}) + H(x|t_{1:j}), \tag{27}
$$

over $P$. Defining the set of auxiliary probability distributions $Q = \{q(z_1|t_1), \ldots, q(z_j|t_{1:j-1}), q(x|t_{1:j})\}$ and the functional

$$
\bar{\mathcal{F}}_{\text{Suc.}}(P,Q) = \sum_{j=1}^{J} \lambda_j I(y_j ; z_j) - \sum_{j=2}^{J} \lambda_j \tag{28}
$$

the following Lemmas hold:

**Lemma 5:** It applies

$$
\min_{P,Q} \mathcal{F}_{\text{Suc.}}(P,Q) = \min_{P} \mathcal{F}_{\text{Suc.}}(P,Q). \tag{29}
$$

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

$$
\mathcal{F}_{\text{Suc.}}(P,Q) - \mathcal{F}_{\text{Suc.}}(P) = \sum_{t_{1:j}} p(t_{1:j}) \tag{30}
$$

$$
\times \mathbb{D}_{\text{KL}}(p(x|t_{1:j}) || q(x|t_{1:j})) \tag{30}
$$

$$
+ \sum_{j=2}^{J} \lambda_j \sum_{t_{1:j-1}} p(t_{1:j-1}) \times \mathbb{D}_{\text{KL}}(p(z_j|t_{1:j-1}) || q(z_j|t_{1:j-1})) \geq 0, \tag{30}
$$

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

**Lemma 6:** $\mathcal{F}_{\text{Suc.}}(P,Q)$ 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 $\mathcal{F}_{\text{Suc.}}(P,Q)$, given by

$$
q^*(z_j|t_{1:j-1}) = p(z_j|t_{1:j-1}) \tag{31}
$$

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

---

\(^9\)The principal idea of the SUM is to optimize a sequence of approximate objective functional(s) (satisfying certain mild assumptions [51]), rather than directly optimizing the original non-convex and/or non-smooth objective functional(s).
Proof: It follows directly from the proof of Lemma 5.

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

$$p^*(z_j|y_j) = \frac{p(z_j)}{\psi_{\text{Suc}}(y_j, \beta_j)} \exp \left( - \bar{d}_{\text{Suc}}(y_j, z_j) \right)$$

for all $(y_j, z_j) \in \mathcal{Y}_j \times \mathcal{Z}_j$, with $\bar{d}_{\text{Suc}}(y_j, z_j)$ defined in (32).

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

$$\delta \left( \mathbb{E}_{z_j} \{ \log \frac{q(z_j)}{p(z_j)} \} \right) \propto \frac{\partial}{\partial p(z_j)} \mathbb{E}_{z_j} \{ \log q(z_j) \} = p(y_j) \sum_{t_{1:j-1}} p(t_{1:j-1}|y_j) \log \frac{q(z_j)}{p(z_j)},$$

and in the case of $n > j$

$$\delta \left( \mathbb{E}_{z_n, t_{1:n-1}} \{ \log \frac{q(z_n)}{p(z_n)} \} \right) \propto \frac{\partial}{\partial p(z_n)} \mathbb{E}_{z_n, t_{1:n-1}} \{ \log q(z_n) \} = p(y_j) \prod_{t_{1:1}} p(t_{1:1}|y_j) \log \frac{q(z_n)}{p(z_n)}.$$

The focal update steps (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. Hence, the convergence of M-FAVIB to a stationary point of the objective functional is ensured by [51, Thm 1] as $\mathcal{F}_{\text{Suc}}(P,Q)$ 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 $\beta_j \to 0$, the design problem (5) boils down to minimizing the $j$-th compression rate, $I(y_j;z_j)$ (presuming fixed $p(z_j|y_j)$ and $\lambda_{\ell}$ for all $\ell = 1$ to $J$ and $\ell \neq j$) w.r.t. the $j$-th quantizer 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 $\beta_j$ values, stochastic (soft) mappings, $p(z_j|y_j)$, are engendered in general, while in the asymptotic case of letting $\beta_j \to \infty$, the partition function, $\psi_{\text{Par}}^j(y_j, \beta_j)$, for each realization, $y_j$, allocates all the probability mass into the specific bin that reveals the minimum $\Delta_{\text{par}}$ 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 $p(z_j|y_j)$ and $\lambda_{\ell}$ for all $\ell = 1$ to $J$ and $\ell \neq j$, and by letting $\beta_j \to \infty (\lambda_j \to 0)$, 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. $p(t_{1:j}|x)$ for a fixed $p(x)$ [54]. Furthermore, $p(t_{1:j}|x)$ and $p(t_j|x)$ are related through

$$p(t_{1:j}|x) = p(t_j|x) \prod_{\ell=1:J, \ell \neq j} p(t_{\ell}|x),$$

which is an affine transform since fixing $p(z_{\ell}|y_{\ell})$ implies that $p(t_{\ell}|x)$ 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. $p(t_j|x)$. Moreover, $p(t_j|x)$ and $p(z_j|y_j)$ are related through

$$p(t_j|x) = \sum_{z_j} \sum_{y_j} p(t_j|z_j)p(z_j|y_j)p(y_j|x),$$

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 $\lambda_{\ell}$ for all $\ell = 1$ to $J$ and $\ell \neq j$) and by letting $\beta_j \to 0 (\lambda_j \to \infty)$, the design problem (11) boils down to minimizing the $j$-th conditional compression rate, $I(y_j;z_j|t_{1:j-1})$, w.r.t. the $j$-th quantizer mapping, $p(z_j|y_j)$, for each realization, $y_j \in \mathcal{Y}_j$, is allocated 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, $I(y_j;z_j|t_{1:j-1})$, since it is non-negative and upper-bounded by $I(y_j;z_j)$. Similarly, for finite $\beta_j$ values, stochastic mappings, $p(z_j|y_j)$, are engendered in general, while
in the asymptotic case of letting $\beta_j \to \infty$, as before, the partition function, $\psi^{\text{Soc}}_\ell(y_j, \beta_j)$, 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 $\lambda_\ell$ for all $\ell = 1$ to $J$ and $\ell \neq j$, and by letting $\beta_j \to \infty$ ($\lambda_j \to 0$), the design optimization (11) boils down to maximizing

$$I(x; t_{1:J}) + \sum_{n=j+1} J \lambda_n I(z_n; t_{1:n-1}), \quad (38)$$

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 $I(z_n; t_{1:n-1})$ is convex w.r.t. $p(z_j|y_j)$ since $\lambda_n$ is non-negative and the sum of several convex functions will be convex as well. $I(z_n; t_{1:n-1})$ is a convex function of $p(t_{1:n-1}|z_n)$ for a fixed $p(z_n)$ [54]. Noting that the relation between $p(t_{1:n-1}|z_n)$ 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, $\sigma_n^2$. 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, $\theta$, 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 $\theta$ values indicate less reliable transmissions and vice versa.\textsuperscript{12}

Further, note that for a specific reliability value, $\theta$, the transition probabilities, $p(t_j|z_j)$, will be influenced by the particular choice of $N$. We consider a totally symmetric setup (having the same access channel noise variance, $\sigma_n^2$, and the forward channel reliability value, $\theta$, across branches), and set all $\beta_j$ values (for $j = 1, 2, 3$) to 100. Then, the end-to-end transmission rate, $I(x; t_{1:3})$, 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 $I(x; z_{1:3})$. 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 $I(x; y_{1:3}) \approx 4$.

\textsuperscript{11}This can be realized by noting that from Markovian relations, it is inferred $I(y_n; z_n|t_{1:n-1}) = I(y_n; z_n) - I(z_n; t_{1:n-1})$.

\textsuperscript{12}The capacity of this symmetric $N \times N$ forward channel model is equal to $C_{\text{FC}}(N, \theta) = \log_2 N + (1 - \theta) \log_2(1 - \theta) + \theta \log_2 N$ [60].
bits), the allowed number of output clusters (per branch), \( N \), has been varied from 2 to 6 and the overall transmission rate, \( I(x; t_{1:3}) \), 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, \( \sigma_n^2 \). 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 \( I(x; z_{1:3}) \). 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, \( I(x; t_{1:3}) \), as possible. The saturation (flattening effect) occurring over the baseline curves can also be understood analogously. The lower the access channels’ noise variance, \( \sigma_n^2 \), the lower the number of required quantizers’ output clusters to come quite close to the maximum supportable value of \( I(x; z_{1:3}) \) and, consequently, the sooner the saturation effect and, thus, ending up to lower end-to-end transmission rates, \( I(x; t_{1:3}) \).

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, \( \sigma_n^2 \), for the \( j \)-th branch. The forward links are modeled as \( N \times N \) symmetric channels each characterized by the particular reliability parameter, \( \theta_j \), for the \( j \)-th branch. We set all \( \beta_j \) values (for \( j = 1, 2, 3 \)) to 100 and calculate the end-to-end transmission rate, \( I(x; t_{1:3}) \), 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 \( I(x; t_j) \) for \( j = 1, 2, 3 \), separately.

In Fig. 6a, we considered an asymmetric specification for the access channels through choosing different noise variances
Contrarily, in Fig. 7b, we fixed the reliability value of the forward channels, namely, \( \theta \) value across branches. The quantizers’ allowed output clusters were varied from \( \theta = 0.05 \) to \( \theta = 0.15 \) (with \( \theta = 0.05 \)), approximating a discrete-time discrete-input continuous-output AWGN channel with the noise variance, \( \sigma_n^2 \) (same value across branches), and a symmetric \( N \times N \) forward channel model with the reliability parameter, \( \theta \) (same value across branches). The quantizers’ allowed output clusters are fixed to \( N = 4 \), and the \( \beta_j \) 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 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 \( \sigma_n^2 = 0.25 \) across branches, namely, \( \sigma_{n_1}^2 = 0.3, \sigma_{n_2}^2 = 0.3, \sigma_{n_3}^2 = 1.0 \) (with \( I(x; y_{1:3}) \approx 3.98 \)) and varied the reliability value of the forward channels (the same value, \( \theta \), 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, \( \sigma_n^2 \) across branches), namely, \( \sigma_n^2 = 0.25, 0.40, 0.55 \) (with \( I(x; y_{1:3}) \approx 4, 3.98, 3.91 \) 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, \( I(x; t_{1:3}) \), 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, \( \sigma_n^2 \) (same value across branches), and a symmetric \( N \times N \) forward channel model with the reliability parameter, \( \theta \) (same value across branches). The quantizers’ allowed output clusters are fixed to \( N = 4 \), and the \( \beta_j \) 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 \( \sigma_n^2 = 0.25 \)) and varied the access channels’ noise variance, namely, \( \sigma_n^2 = 0.25, 0.50, 0.75 \) (with \( \sigma_n^2 = 0.25 \)) and varied the access channels’ noise variance, namely, \( \sigma_n^2 = 0.25, 0.50, 0.75 \) (with \( \sigma_n^2 = 0.25 \)).

\( I(x; t_{1:3}) \approx 4, 3.95, 3.75 \) 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 deduce from the current unconditional compression rate or keep it unchanged as it applies

\[
I(y_j; z_j|t_{1:j-1}) = I(y_j; z_j) - I(z_j; t_{1:j-1}),
\]

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 Lagragian, \( \mathcal{L}_{\text{Par.}} \), in (6) is a functional of all the individual local quantizer mappings \( \{p(z_j|y_j)\}_{j} \), 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, \( \lambda_{y_j} \), 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, \( L^{\text{Par}}_{\text{SC}} \), for which it applies

\[
L^{\text{Par}}_{\text{SC}} = I(x; t_{1:j}) - \sum_{j=1}^{J} \lambda_j I(y_j; z_j) + \sum_{j=1}^{J} \sum_{y_j \in \mathcal{Y}_j} \lambda_{y_j} \left( \sum_{z_j \in \mathcal{Z}_j} p(z_j | y_j) - 1 \right). \tag{41}
\]

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

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

and

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

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

\[
\frac{\delta I(x; t_{1:j})}{\delta p(z_j | y_j)} = \frac{\delta I(x; t_j^{1:j})}{\delta p(z_j | y_j)}, \tag{44}
\]

as fixing \( \{p(z_j | y_j) \mid \ell \neq j \} \) also fixes the other appearing terms in the conducted decomposition. Noting the present Markov chain, \( y_j \leftarrow z_{1:j} \leftarrow t_{1:j} \), it holds

\[
\frac{\delta I(x; t_j^{1:j})}{\delta p(z_j | y_j)} = \left( H(t_j | x, t_{1:j}^{1:j}) - H(t_j | x, t_{1:j}^{1:j}) \right) \frac{\delta p(z_j | y_j)}{\delta p(z_j | y_j)} = \sum_{z_j \in \mathcal{Z}_j} p(y_j, z_j^{1:j}) \sum_{t_{1:j}} p(t_{1:j} | z_{1:j}) \log \frac{p(x | t_{1:j})}{p(x | t_{1:j})}, \tag{45}
\]

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

\[
\sum_{z_j \in \mathcal{Z}_j} p(z_j | y_j) \sum_{z_j \in \mathcal{Z}_j} p(x | y_j, z_j^{1:j}) \sum_{t_{1:j}} p(t_{1:j} | z_{1:j}) \log \frac{p(x | t_{1:j})}{p(x | t_{1:j})} - \lambda_j \log \frac{p(z_j | y_j)}{p(z_j)} + \lambda_{y_j} \frac{p(z_j)}{p(y_j)} = 0. \tag{46}
\]

Thus, by definition of the KL divergence, it holds

\[
-\sum_{z_j \in \mathcal{Z}_j} p(z_j | y_j) \sum_{t_{1:j}} p(t_{1:j} | z_{1:j}) D_{KL}(p(x | t_{1:j}) \parallel p(x | t_{1:j})) - \lambda_j \log \frac{p(z_j | y_j)}{p(z_j)} + \tilde{\lambda}_{y_j} = 0, \tag{47}
\]

with \( \tilde{\lambda}_{y_j} \), being equal to

\[
\tilde{\lambda}_{y_j} = \frac{\lambda_{y_j}}{p(y_j)} + \sum_{z_j \in \mathcal{Z}_j} p(z_j | y_j) \sum_{t_{1:j}} p(t_{1:j} | z_{1:j}) \tilde{\lambda}_{y_j} = D_{KL}(p(x | t_{1:j}) \parallel p(x | t_{1:j})). \tag{48}
\]

Bringing the second summand in (47) 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

\[
p(z_j | y_j) = p(z_j) \exp \left( -d_{y_j} p(z_j | y_j) + \beta_j \tilde{\lambda}_{y_j} \right). \tag{49}
\]

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

**B. Proof of Theorem 2**

The successive IB-based JSCC Lagrangian, \( L_{\text{SC}} \), 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, \( \lambda_{y_j} \), 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, \( L^{\text{Par}}_{\text{SC}} \), for which it applies

\[
L^{\text{Par}}_{\text{SC}} = I(x; t_{1:j}) - \sum_{j=1}^{J} \lambda_j I(y_j; z_j | t_{1:j-1}) + \sum_{j=1}^{J} \sum_{y_j \in \mathcal{Y}_j} \lambda_{y_j} \left( \sum_{z_j \in \mathcal{Z}_j} p(z_j | y_j) - 1 \right). \tag{50}
\]

Fixing \( \{p(z_j | y_j) \mid \ell \neq j \} \), to be able to calculate the functional derivative of \( L^{\text{Par}}_{\text{SC}} \) 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

\[
\frac{\delta}{\delta p(z_j | y_j)} \left( \sum_{j=1}^{J} \lambda_j I(y_j; z_j | t_{1:j-1}) \right) = \sum_{\ell=1}^{J} \lambda_j I(y_j; z_j | t_{1:j-1}) \frac{\delta p(z_j | y_j)}{\delta p(z_j | y_j)}, \tag{51}
\]

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

\[
\frac{\delta I(y_j; z_j | t_{1:j-1})}{\delta p(z_j | y_j)} = \left( H(z_j | t_{1:j-1}) - H(z_j | t_{1:j-1}, y_j) \right) \frac{\delta p(z_j | y_j)}{\delta p(z_j | y_j)} = \frac{\delta H(z_j | t_{1:j-1})}{\delta p(z_j | y_j)} - \frac{\delta H(z_j | t_{1:j-1})}{\delta p(z_j | y_j)} = p(y_j) \times \sum_{z_{1:j-1}} p(t_{1:j-1} | z_{1:j-1}) \log \frac{p(t_{1:j-1} | z_{1:j-1})}{p(t_{1:j-1})}, \tag{52}
\]

with the notation, \( \tilde{\lambda}_{y_j} \), being equal to

\[
\tilde{\lambda}_{y_j} = \frac{\lambda_{y_j}}{p(y_j)} + \sum_{z_{1:j-1}} p(t_{1:j-1} | z_{1:j-1}) p(t_{1:j-1} | z_{1:j-1}) \times D_{KL}(p(x | t_{1:j}) \parallel p(x | t_{1:j})). \tag{48}
\]

Bringing the second summand in (47) 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

\[
p(z_j | y_j) = p(z_j) \exp \left( -d_{y_j} p(z_j | y_j) + \beta_j \tilde{\lambda}_{y_j} \right). \tag{49}
\]

Enforcing the validity condition, \( \sum_{z_j} p(z_j | y_j) = 1 \), and noting that \( \tilde{\lambda}_{y_j} \) is independent of \( z_j \), one can treat \( \exp(-\beta_j \tilde{\lambda}_{y_j}) \) as the partition function, \( \psi_{y_j} \), to come into the form of (7).
and in the case of \( j < \ell \leq J \)

\[
\frac{\delta f(y_j; z_j|t_{1:\ell-1})}{\delta p(z_j|y_j)} = \frac{\delta H(z_j|t_{1:\ell-1})}{\delta p(z_j|y_j)} - \frac{\delta H(z_j|t_{1:\ell-1}, y_j)}{\delta p(z_j|y_j)}
\]

Due to the positivity of \( \delta I_{pp} \), to come into the form of (13).

(53)

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

\[
- \sum_{z_{1:j}} p(z_{1:j}|y_j) \sum_{t_{1:j}} p(t_{1:j}|z_{1:j}) \times D_{KL}(p(x|y_j, z_{1:j})||p(x|t_{1:j})) - \lambda_j \log \frac{p(z_j|y_j)}{p(z_j)} + \lambda_j \sum_{t_{1:j-1}} p(t_{1:j-1}|y_j) \times \log p(t_{1:j-1}|y_j) + \lambda_{j_{\text{Suc.}}} \sum_{k=j+1}^{J} \lambda_k \sum_{t_{1:k-1}} p(t_{1:k-1}|z_{k-1}) \times \log p(z_k|t_{1:k-1}) = 0,
\]

with \( \lambda_{j_{\text{Suc.}}} \), being equal to

\[
\lambda_{j_{\text{Suc.}}} = -\lambda_j \sum_{t_{1:j-1}} p(t_{1:j-1}|y_j) \times \log p(t_{1:j-1}) + \frac{\lambda_j}{p(y_j)} + \sum_{z_{1:j}} p(z_{1:j}|y_j) \sum_{t_{1:j}} p(t_{1:j}|z_{1:j}) \times D_{KL}(p(x|y_j, z_{1:j})||p(x|t_{1:j})).
\]

(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

\[
p(z_j|y_j) = p(z_j) \exp \left( -d_{\text{Suc.}}(y_j, z_j) + \beta_j \lambda_{j_{\text{Suc.}}} \right).
\]

(56)

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

REFERENCES


