# Forward-Aware Information Bottleneck-Based Vector Quantization for Noisy Channels

Shayan Hassanpour, *Member, IEEE*, Tobias Monsees, *Graduate Student Member, IEEE*, Dirk Wübben, *Senior Member, IEEE*, and Armin Dekorsy, *Senior Member, IEEE*

*Abstract*— The main focus will be on the indirect *Joint Source-Channel Coding* problem in which a noisy observation of the source has to be quantized ahead of transmission over an *error-prone* forward link to a remote processing unit. To that end, we present here a complete extension to the preliminary *Information Bottleneck* method by providing the formal optimal solution to this newly established *Variational Principle*, together with an algorithm, the *Forward-Aware Vector Information Bottleneck (FAVIB)*, to pragmatically tackle its underlying non-convex design optimization. FAVIB extends the current state-of-the-art approaches via capacitating a full sweep over the entire gamut of the trade-off parameter. Consequently, the trajectory of all achievable points in the *Information-Compression* plane becomes traversable via soft mappings. It will be shown that, by enjoying an inherent error protection, this novel compression scheme can obviate the call for separate channel coding on the forward path.

*Index Terms*— Channel quantization, error-prone forward channel, information bottleneck, mutual information.

## I. INTRODUCTION

**T**HE *Information Bottleneck (IB)* method was first introduced in [1] as a novel stratagem towards *Clustering* [2] which is a pivotal task in the context of *Unsupervised Learning* [3]. In a nutshell, its principal idea is to compress a *Random Variable (RV)* such that its information content w.r.t. a statistically correlated (relevant) variable is mostly retained. This information preservation capability is rather flexible and, in effect, can be tuned through twiddling a parameter that establishes a fundamental trade-off between the *compactness* and *informativity* of its resultant outcome. Actually, the IB method accomplishes this in an entirely symmetric fashion by exploiting the same concept of *Mutual Information (MI)* to quantify both aspects. For acquiring somewhat more concrete insights into this framework with some of the most recent studies on the pertinent design techniques, interested readers are referred to [4]–[16].

Specifically, in [16] a comprehensive discussion has been provided about the learning and information-theoretic aspects of the IB paradigm together with some interesting connections with several classical problems in the context of information theory, e.g., the Wyner-Ziv setup [17] with the logarithmic-loss distortion [18] under the *Common Reconstruction (CR)* constraint [19], the Wyner-Ahlswede-Körner problem [20], [21], and also the efficiency of investment information [22].

The structure of problems the IB method aims to address turns it into an appropriate choice for *Noisy Source Coding (NSC)* [23]–[29] scenarios. Explicitly, pursuing the IB philosophy to compress a noisy observation signal at the output of an imperfect access channel to a remote source brings about a purely statistical design setup. Further, an important special instance of the IB principle focuses on designing the quantizers that intend to maximize the overall transmission rate for a given input statistics that, in general, is an absolutely desirable criterion for all communication systems. In fact, the IB method has already found its path into various aspects of modern data transmission schemes from design of analog-to-digital converters for receiver front ends [30], to construction of polar codes [31] and low-complexity discrete (channel) decoding schemes [32]–[37] with rather promising performance.

In a vast variety of practical situations, the compressed data at the output of the quantizer block(s) should be transmitted over a/some noisy channel(s) to a distant unit for further processing. As several examples of such scenarios, one can think of relaying setups with the *Quantize-and-Forward* approach [38], [39], *Cloud-based Radio Access Networks (C-RANs)* with non-ideal fronthaul channels [40]–[43], distributed inference sensor networks with noisy links to the fusion center [44], [45] and, last but not least, reception schemes with unreliable memories [46], [47]. The aforementioned miscellaneous list of applications can be subsumed under the umbrella of a broader setup, being known as *Joint Source-Channel Coding (JSCC)* [48], [49] in which the impacts of error-prone forwarding of the quantizer output have to be taken into account and brought into the pertinent design problem. Otherwise, it might lead to a substantial performance degradation of the applied compression scheme. In the available literature on this subject some general approaches can be recognized: Among others, multiple techniques are presented to judiciously assign binary codewords to quantizer output (see, e.g., [50], [51]), or focusing on squared-error distortion, some modifications to the conventional Lloyd algorithm [52]

are proposed (see, e.g., [53], [54]). In [55] MI has been chosen as the fidelity criterion to obtain a compression scheme that intends to maximize the end-to-end transmission rate. A rather extensive survey of various relevant methods can be found, e.g., in [56], [57].

**Contributions:** Within the scope of this article, the mathematical framework of the IB method [1] for NSC is completely extended to enable its applicability for the JSCC setup. Specifically, we derive the optimal solution to the IB-based JSCC problem and utilize that to propose a design algorithm, along with its proof of convergence. By this, the trajectory of all achievable points in the *Information-Compression* plane[1] for JSCC becomes traversable via probabilistic mappings. Furthermore, we directly adapt the conducted derivations to the *Vector Quantization (VQ)* setup and introduce a forward-aware vector quantizer that, by enjoying an inherent error protection, can obviate the demand for separate employment of channel coding at the second hop. This approach becomes interesting and crucial in many practical occasions, i.a., in applications where tight latency constraints prohibit the utilization of modern iterative *Forward Error Correction (FEC)* schemes on the error-prone forward channel, or in case of hardware imperfections where the separate implementation of such error-correction schemes comes with a substantial overhead in terms of energy efficiency, leading to an evident waste of resources since they serve solely as a precaution against worst-case conditions [47]. Moreover, it is noteworthy that from a pure theoretical perspective, as the optimality of the Shannon's source and channel separation [58] does not hold in general [59] (e.g., when working in a non-asymptotic regime), devising such joint schemes becomes relevant as well.

**Outline:** To facilitate a seamless segue, after a succinct discussion on the IB method for the NSC in Section II, we introduce our presumed system model and state the quantizer design problem as a constrained optimization in Section III-A. Subsequently, in perfect harmony with the content flow of the seminal work [1], applying *Variational Calculus* we derive a complete formal characterization of the optimal solution to the IB-based JSCC *Variational Principle* in Section III-B and propose an algorithm to pragmatically tackle the pertinent quantization design problem in Section III-C. Furthermore, we present a concise proof for the convergence of the suggested algorithm to a stationary point of its objective functional. We also provide the essential mathematical insights concerning the underlying optimization problem and the structure of the presented solution in Section III-D. As the conducted derivations for the scalar case can be generalized straightforwardly, in Section IV, we turn our focus into the VQ [60]. There, we introduce the <u>F</u>orward-<u>A</u>ware <u>V</u>ector <u>IB</u> (FAVIB) routine which extends the *State-of-the-Art (SotA)* algorithm <u>M</u>utual <u>I</u>nformation-based <u>C</u>hannel-<u>O</u>ptimized <u>V</u>ector <u>Q</u>uantizer (MICOVQ) [55] that is an approach proposed for maximizing the end-to-end transmission rate. This can be interpreted as an extreme instance of *full informativity* from FAVIB's perspective. Finally, we con-

---

[1]It is a plane featuring two sides of the IB trade-off as its ordinate and abscissa, depicting their fundamental interrelations.
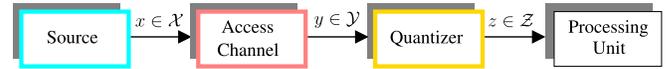


Fig. 1.   The considered system model for Noisy Source Coding. The access channel is presumed to be discrete and memoryless.

sider some typical data transmission setups in Section V to substantiate the effectiveness of the proposed approach, prior to a succinct wrap-up in the end. The respective proofs for the presented theorems are provided in the Appendix.

**Notation:** The RV, $a$, with the probability mass function, $p(a)$, accepts certain realizations, $a$, from its domain, $\mathcal{A}$. With boldface counterparts, the same holds for the random vector, $\mathbf{a}$. Moreover, $H(\cdot)$, $D_{\mathrm{KL}}(\cdot \| \cdot)$ and $I(\cdot\,;\cdot)$ denote the Shannon's entropy, the relative entropy (*Kullback-Leibler (KL)* divergence) and the MI [61], respectively. Also, $D_{\mathrm{JS}}^{\{\cdot,\cdot\}}(\cdot \| \cdot)$ stands for the Jensen-Shannon divergence [62].

## II. INFORMATION BOTTLENECK METHOD FOR NOISY SOURCE CODING (NSC)

This section provides a brief overview on the *Information Bottleneck (IB)* paradigm and its mathematical methodology. In an attempt to bypass the complicated problem of feature selection in pattern recognition to extract a *relevant* summary of data, the authors in [1] came up with the intuitive idea of bringing an additional variable into play to effectively determine the meaning of *relevance*. In a variety of practical applications, e.g., bioinformatics and neural coding, defining such a relevant (target) variable is a fairly simple task with a natural answer that is way easier to be addressed compared to the rather complicated problem of the correct feature selection. The procedure is then to extract the part of information from data that is important for prediction of the target variable. In other words, the information that the data set provides about the target variable is squeezed through a bottleneck being formed by a limited number of clusters, hence the name.

Describing it more tangibly, the IB setup [1] considers the quantization of a given RV, $y$, into the compression variable, $z$, such that it remains highly informative w.r.t. a target variable, $x$. As a straightforward translation into the context of *Noisy Source Coding (NSC)*, the system model illustrated in Fig. 1 is considered. The data source, $x$, is observed through a discrete, memoryless access channel being described by transition probabilities, $p(y|x)$. The noisy observation, $y$, then has to be compressed into the signal, $z$, prior to getting forwarded over an *error-free* channel to a (remote) processing unit. It is presumed that the joint distribution $p(x, y)$ is given and, further, $x \leftrightarrow y \leftrightarrow z$ is a first-order Markov chain. To design the quantizer, $p(z|y)$, then the IB method establishes a fundamental trade-off between the *compactness* and *informativity* of its outcome in a symmetric fashion, employing two MI terms to quantify both aspects. On the one hand, $I(y; z)$ is considered as the term gauging the *compactness* of outcome. Clearly, lower values of this quantity imply higher compression and vice versa. A rather more formal interpretation associates $I(y; z)$ with the maximum number of bits which can be reliably transmitted over the quantizer block, employing the

Fig. 2.   The considered system model for Joint Source-Channel Coding. Both access and forward channels are presumed to be discrete and memoryless.

*Asymptotic Equipartition Property (AEP)* [61]. On the other hand, $I(\mathsf{x};\mathsf{z})$ is chosen as the indicator of the *informativity*. The design problem is then stated as finding the conditional distribution, $p^*(\mathsf{z}|\mathsf{y})$, satisfying

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y}):\, I(\mathsf{x};\mathsf{z}) \geq I}{\arg\min} I(\mathsf{y};\mathsf{z}), \tag{1}$$

with $0 \leq I \leq H(\mathsf{x})$. Applying the method of *Lagrange Multipliers*, the constrained optimization (1) can be transformed into an unconstrained one (up to the validity of the quantizer mapping) over an augmented objective functional

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y})}{\arg\min}\ I(\mathsf{y};\mathsf{z}) - \beta I(\mathsf{x};\mathsf{z}), \tag{2}$$

in which $\beta$ denotes a *non-negative* trade-off parameter. Exploiting the Variational Calculus, the formal characterization of the optimal solution to the pertinent design problem has been derived in [1] for each pair $(y,z) \in \mathcal{Y} \times \mathcal{Z}$ as

$$p^*(z|y) = \frac{p(z)}{\psi(y,\beta)}\exp\big(-\beta d_{\mathrm{IB}}^{\mathrm{NSC}}(y,z)\big), \tag{3}$$

wherein $\psi(y,\beta)$ is a partition (normalization) function, ensuring a valid distribution and the so-called *IB-NSC Relevant Distortion*, $d_{\mathrm{IB}}^{\mathrm{NSC}}(y,z)$, is given as

$$d_{\mathrm{IB}}^{\mathrm{NSC}}(y,z) = D_{\mathrm{KL}}\big(p(\mathsf{x}|y)\|p(\mathsf{x}|z)\big) = \sum_{x \in \mathcal{X}} p(x|y) \log \frac{p(x|y)}{p(x|z)}. \tag{4}$$

This indicates that, principally, the IB-based clustering must gather together those realizations of the signal, y, which provide quite similar information about the source, x. Specifically, for two distinct realizations $y_1 \in \mathcal{Y}$ and $y_2 \in \mathcal{Y}$ to be allocated to the same cluster $z \in \mathcal{Z}$, the KL divergence between a posteriori terms $p(\mathsf{x}|y_1)$ and $p(\mathsf{x}|y_2)$ must be relatively small as it must be the case for both KL divergences among each individual a posteriori term and $p(\mathsf{x}|z)$. Finally, it is noteworthy that a design algorithm, the *iterative IB*, has been provided in [1] as well that exerts the *Fixed-Point Iteration* method [63] on (3) together with a short proof of its convergence to a stationary point of the pertinent objective functional in (2).

## III. Extension of Information Bottleneck Method to Joint Source-Channel Coding (JSCC)

In this part, after introducing the system model and the respective design problem we present the optimal solution to IB-based *Joint Source-Channel Coding (JSCC)* quantization. We also devise an algorithm to pragmatically tackle the non-convex design optimization and provide a theorem regarding the convergence of the proposed algorithm to a stationary point of its objective functional. Furthermore, we analyze in full detail various mathematical aspects of the design problem itself together with the features of the presented solution for the sake of a solid and crisp understanding of the method.

### A. System Model and Problem Formulation

Consider the system model depicted in Fig. 2. The main difference compared to the original IB setup (cf. Fig. 1) is to relax the *ideal* transmission presumption of the quantizer output, z, to the (remote) processing unit by adding an *error-prone*, discrete memoryless forward channel with transition probabilities, $p(\mathsf{t}|\mathsf{z})$, to the previous system model. Presumably, the statistical description of the source, $p(\mathsf{x})$, as well as the access $p(\mathsf{y}|\mathsf{x})$ and the forward $p(\mathsf{t}|\mathsf{z})$ channels are available (fixing an upper-bound on the quantizer's output cardinality, $|\mathcal{Z}|$) and, further, $\mathsf{x} \leftrightarrow \mathsf{y} \leftrightarrow \mathsf{z} \leftrightarrow \mathsf{t}$ forms a first-order Markov chain. Henceforth, the main focus will be on the quantizer design, $p(\mathsf{z}|\mathsf{y})$. For that, in perfect harmony with the foundational idea of the IB method [1], here we introduce the mathematical formulation of the required quantization through establishing a fundamental trade-off between two MI terms. On the one hand, the MI between the quantizer's input and output signals, $I(\mathsf{y};\mathsf{z})$, denoted as the *compression rate*, is the term indicating the *compactness* of outcome. On the other hand, the MI between the relevant variable, i.e., the source, x, and the output signal, t, of the forward channel, $I(\mathsf{x};\mathsf{t})$, that is denominated as the *relevant information*, is treated as the term signifying the *informativity*. Naturally, one then aims for such solutions which, despite being relatively compact ($|\mathcal{Z}| \leq |\mathcal{Y}|$), are still highly informative. Analogous to (1), this goal is mathematically formulated as the following constrained optimization

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y}):\, I(\mathsf{x};\mathsf{t}) \geq I}{\arg\min} I(\mathsf{y};\mathsf{z}), \tag{5}$$

wherein for the lower-bound of relevant information, $I$, it applies $0 \leq I \leq H(\mathsf{x})$. [2] As before, making use of the method of *Lagrange Multipliers*, the constrained optimization in (5) can be transformed into an unconstrained one (up to the validity of the quantizer mapping) over an augmented objective functional

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y})}{\arg\min}\ I(\mathsf{y};\mathsf{z}) - \beta I(\mathsf{x};\mathsf{t}), \tag{6}$$

with the *non-negative* trade-off parameter, $\beta$, as the counterpart of the lower-bound, $I$, in (5) and $p(\mathsf{z}|\mathsf{y})$ being a valid conditional distribution, i.e., for each $y \in \mathcal{Y}$, it applies $\sum_{z \in \mathcal{Z}} p(z|y) = 1$.

### B. Optimal Quantizer Mapping

The following theorem provides a complete characterization of the formal optimal solution to the IB-based JSCC quantization design problem (6).

---

[2]Besides this one-shot formulation, the pertinent (asymptotic) coding problem calls for a formal multi-letter description.

*Theorem 1 (Optimal Solution for IB-JSCC):* Assume the joint distribution, $p(\mathsf{x},\mathsf{y})$, the forward transition probabilities, $p(\mathsf{t}|\mathsf{z})$, and $\beta$ are given. The conditional distribution, $p^*(\mathsf{z}|\mathsf{y})$, is a stationary point of the IB-JSCC functional

$$\mathcal{L}_{\text{IB}}^{\text{JSCC}} = I(\mathsf{y};\mathsf{z}) - \beta I(\mathsf{x};\mathsf{t}) \tag{7}$$

if and only if for each pair $(y,z) \in \mathcal{Y} \times \mathcal{Z}$

$$p^*(\mathsf{z}|\mathsf{y}) = \frac{p(z)}{\psi(y,\beta)}\exp\big(-\beta d_{\text{IB}}^{\text{JSCC}}(y,z)\big), \tag{8}$$

in which $\psi(y,\beta)$ is a normalization function, ensuring the validity of the pertinent conditional distribution and the so-called *IB-JSCC Relevant Distortion*, $d_{\text{IB}}^{\text{JSCC}}(y,z)$, is given as

$$d_{\text{IB}}^{\text{JSCC}}(y,z) = \sum_{t \in \mathcal{T}} p(\mathsf{t}|\mathsf{z}) D_{\text{KL}}\big(p(\mathsf{x}|\mathsf{y})\|p(\mathsf{x}|\mathsf{t})\big). \tag{9}$$

The proof is provided in Appendix A.

Theorem 1 can, actually, be interpreted as the direct generalization of Theorem 4 in [1] to incorporate the noisy forward channel as well. The *Backward Compatibility* of the provided solution here to the one derived for the primary IB setup can be checked by assuming an *ideal* forward link, i.e., $p(\mathsf{t}|\mathsf{z}) = \delta_{z,t}$ (with the *Kronecker Delta* denotation [3]). For that, (8) and (9) boil down to the given optimal solution at Theorem 4 in [1] that has been restated in (3) and (4).

### C. Forward-Aware Iterative IB Algorithm

It shall be noted that (8) has an *implicit* form as $p(z)$ and $p(\mathsf{x}|t)$ on its right-hand side are functions of its left-hand side, $p^*(\mathsf{z}|\mathsf{y})$, through self-consistency relations below

$$p(z) = \sum_{y \in \mathcal{Y}} p(\mathsf{y})p^*(\mathsf{z}|\mathsf{y}) \tag{10a}$$

$$p(\mathsf{x}|t) = \frac{\sum\limits_{y \in \mathcal{Y}} \sum\limits_{z \in \mathcal{Z}} p(\mathsf{x},\mathsf{y})p^*(\mathsf{z}|\mathsf{y})p(\mathsf{t}|\mathsf{z})}{\sum\limits_{z' \in \mathcal{Z}} p(z')p(\mathsf{t}|z')}. \tag{10b}$$

Therefore, the iterative IB routine in [1] can be straightforwardly adapted to a novel algorithm which implements the *Fixed-Point Iteration* method [63] for the IB-based JSCC. To that end, the generalized version of Theorem 5 in [1] will be presented here, leading to an alternating optimization procedure analogous to the well-established *Blahut-Arimoto (BA)* method [64], [65] and the *Expectation-Maximization (EM)* algorithm [66].

*Theorem 2 (Iterative Algorithm):* The self-consistent equations (10a), (10b) and (8) will be satisfied simultaneously at the minima of the functional $\mathcal{F} = -\mathbb{E}_{\mathsf{y},\mathsf{z}}\{\log \psi(y,\beta)\}$, wherein the minimization is done independently over the convex sets of normalized distributions $p(z)$, $p(\mathsf{z}|\mathsf{y})$, and $p(\mathsf{x}|t)$. This minimization is performed by the following convergent, alternating

iterations: With $(i)$ denoting the iteration counter, it applies

$$p^{(i)}(z) = \sum_{y \in \mathcal{Y}} p(\mathsf{y})p^{(i)}(\mathsf{z}|\mathsf{y}) \tag{11a}$$

$$p^{(i)}(\mathsf{x}|t) = \frac{\sum\limits_{y \in \mathcal{Y}} \sum\limits_{z \in \mathcal{Z}} p(\mathsf{x},\mathsf{y})p^{(i)}(\mathsf{z}|\mathsf{y})p(\mathsf{t}|\mathsf{z})}{\sum\limits_{z' \in \mathcal{Z}} p^{(i)}(z')p(\mathsf{t}|z')} \tag{11b}$$

$$p^{(i+1)}(\mathsf{z}|\mathsf{y}) = \frac{p^{(i)}(z)\exp\Big(-\beta\sum\limits_{t \in \mathcal{T}} p(\mathsf{t}|\mathsf{z}) D_{\text{KL}}\big(p(\mathsf{x}|\mathsf{y})\|p^{(i)}(\mathsf{x}|\mathsf{t})\big)\Big)}{\psi^{(i+1)}(y,\beta)}. \tag{11c}$$

The partition (normalization) function [4], $\psi^{(i+1)}(y,\beta)$, is evaluated per iteration.

The proof is provided in Appendix B.

### D. Supplementary Mathematical Discussion

In this part, we provide an in-depth and comprehensive discussion on various mathematical aspects of the proposed extended approach (IB-based JSCC). For that, a number of important issues have to be elucidated at this point:

1) The precise relation between the auxiliary functional $\mathcal{F}$ introduced in Theorem 2 and the IB-JSCC functional $\mathcal{L}_{\text{IB}}^{\text{JSCC}}$ of Theorem 1 must be determined. For that, the constituent terms of $\mathcal{F}$ shall be analyzed more critically. It applies

$$\mathcal{F} = \sum_{y,z} p(\mathsf{y})p(\mathsf{z}|\mathsf{y}) \log \frac{p(z|y)}{p(z)}$$
$$+ \beta \sum_{y,z} p(\mathsf{y})p(\mathsf{z}|\mathsf{y}) \sum_{t} p(\mathsf{t}|\mathsf{z}) D_{\text{KL}}\big(p(\mathsf{x}|\mathsf{y})\|p(\mathsf{x}|\mathsf{t})\big). \tag{12}$$

The first summand in (12) equals $I(\mathsf{y};\mathsf{z})$ by definition. Further, its second summand is equal to $\beta\big(I(\mathsf{x};\mathsf{y}) - I(\mathsf{x};\mathsf{t})\big)$. To realize that, applying the presumed Markovian properties and through expansion of the respective KL divergences, it holds

$$\sum_{y,z} p(\mathsf{y})p(\mathsf{z}|\mathsf{y}) \sum_{t} p(\mathsf{t}|\mathsf{z}) D_{\text{KL}}\big(p(\mathsf{x}|\mathsf{y})\|p(\mathsf{x}|\mathsf{t})\big) + I(\mathsf{x};\mathsf{t}) \tag{13a}$$

$$= \sum_{x,y,z,t} p(x,y,z,t)\Big(\log \frac{p(x|y)}{p(x|t)} + \log \frac{p(x|t)}{p(x)}\Big) \tag{13b}$$

$$= \sum_{x,y,z,t} p(x,y,z,t) \log \frac{p(x|y)}{p(x)} = I(\mathsf{x};\mathsf{y}). \tag{13c}$$

Consequently, for a certain trade-off parameter, $\beta$, since $I(\mathsf{x};\mathsf{y})$ is fixed via the given joint distribution, $p(\mathsf{x},\mathsf{y})$, it is deduced that the functionals $\mathcal{F}$ and $\mathcal{L}_{\text{IB}}^{\text{JSCC}}$ differ only in a constant term and, thus, converging to the minima of either of them directly corresponds to converging to the minima of the other one too.

2) Analogous to the content of the discussion in the original work [1] about the structure of the derived solution, it shall be noted that each specific $\beta$ value results in certain values of compression rate, $I(\mathsf{y};\mathsf{z})$, and relevant information, $I(\mathsf{x};\mathsf{t})$, for

---

[3]The *Kronecker Delta* denotation shall not be confused with the one which will be used for functional derivatives. Besides the context, the clear distinction is that the former has two arguments on its subscript while the latter does not have any.

[4]Per definition, for each access channel output realization, $y \in \mathcal{Y}$, the sum of the calculated terms in (11c) (ignoring $\psi$) over all output clusters, $z \in \mathcal{Z}$, acts as the partition (normalization) function, $\psi(y,\beta)$, to ensure $\sum_{z \in \mathcal{Z}} p(\mathsf{z}|\mathsf{y}) = 1$.

every choice of the quantizer output cardinality, $|\mathcal{Z}|$. Further, the variational principle implies

$$\frac{\delta I(\mathsf{x};\mathsf{t})}{\delta I(\mathsf{y};\mathsf{z})} = \beta^{-1} > 0 . \tag{14}$$

As will be demonstrated in Section V-B with a relevant example, in cases where the forward transition probabilities, $p(\mathsf{t}|\mathsf{z})$, are output semi-symmetric (i.e., having an identical distribution to all outputs regardless of the particular choice of input), analogous to the original IB setup, this brings about a *Deterministic Annealing* behavior [67]. Through increasing the value of the trade-off parameter, $\beta$, one can move along *concave* curves in the *Information-Compression* plane. These curves exist for each choice of the quantizer output cardinality, $|\mathcal{Z}|$. The solution of self-consistent equations in (11) will then correspond to a family of such annealing curves, all starting from the origin with an infinite slope and getting parameterized by $\beta$. Every two curves in this family then bifurcate at some finite (critical) value of $\beta$ via a second order phase transition. These transitions constitute a hierarchy of quantizations with different granularity, from an extreme case of $|\mathcal{Z}| = 1$ up to the other extreme where $|\mathcal{Z}| = |\mathcal{Y}|$.

3) It is important to note that the IB-based JSCC problem (6) is, basically, a *non-convex* optimization in its most general form. To discern this, it has to be noticed that both MI terms involved in (7) are convex functionals of the quantizer mapping, $p(\mathsf{z}|\mathsf{y})$. Indeed, for a given $p(\mathsf{y})$, the compression rate, $I(\mathsf{y};\mathsf{z})$, is convex w.r.t. $p(\mathsf{z}|\mathsf{y})$ [61]. Analogously, for a given $p(\mathsf{x})$, the relevant information, $I(\mathsf{x};\mathsf{t})$, is convex w.r.t. $p(\mathsf{t}|\mathsf{x})$. Since the relation between $p(\mathsf{t}|\mathsf{x})$ and $p(\mathsf{z}|\mathsf{y})$ is established through

$$p(t|x) = \sum_{y \in \mathcal{Y}} p(y|x) \sum_{z \in \mathcal{Z}} p(t|z) p(z|y), \tag{15}$$

which is an *affine* transform preserving convexity, $I(\mathsf{x};\mathsf{t})$ will also be convex w.r.t. $p(\mathsf{z}|\mathsf{y})$. Thus, the IB-JSCC functional (7) that is a difference of two convex functionals is non-convex in general, leading to a non-convex (more precisely, a *Difference of Convex (DC)*) optimization [68, Ch. 4]. The aforementioned proposition implies the fact that the auxiliary functional $\mathcal{F}$ is not *jointly* convex w.r.t. all of its arguments. As an immediate ramification, the random choice of initialization, $p^{(0)}(\mathsf{z}|\mathsf{y})$, for the proposed iterative algorithm has a direct influence on the quality of final upshot, as convergence to different local minima may occur. A common workaround is then to rerun this routine with multiple distinct starting points, $p^{(0)}(\mathsf{z}|\mathsf{y})$, and retain the best outcome. It is noteworthy that for finite trade-off values, $\beta$, usually a stochastic (soft) mapping $0 \leq p(z|y) \leq 1$ is resulted that associates each realization, $y \in \mathcal{Y}$, with a certain probability to every output cluster, $z \in \mathcal{Z}$.

3) Considering (8), it can be discerned that the value of $\beta$ determines how diffused the conditional distribution, $p(\mathsf{z}|\mathsf{y})$, is. Small $\beta$ values imply high diffusion as, in those cases, $\beta$ reduces the difference among the calculated IB-JSCC relevant distortion, $d_{\mathrm{IB}}^{\mathrm{JSCC}}(y, z)$, for different candidates, $z \in \mathcal{Z}$. In the limit of letting $\beta \to 0$, there would be maximal diffusion and each certain realization, $y \in \mathcal{Y}$, will be allotted to all output levels, $z \in \mathcal{Z}$, equiprobably. By that, the input and output

of the quantizer block become statistically independent and the compression rate, $I(\mathsf{y};\mathsf{z})$, achieves its global minimum of zero. Obviously, this case is not of interest as no relevant information is kept at all. In contrast, by increasing the value of $\beta$ for each realization, $y \in \mathcal{Y}$, most of the probability mass gets assigned into the very bin showing the least relevant distortion, $d_{\mathrm{IB}}^{\mathrm{JSCC}}$, among all output candidates. Naturally, in the limit of letting $\beta \to \infty$ full concentration is achieved wherein a particular bin, $z^*(y)$, contains all the probability mass, i.e., each $y \in \mathcal{Y}$, is allotted to one and only one output bin, $z^*(y)$.

5) Finally, it is worth exclusively discussing the extreme case of letting $\beta \to \infty$ in more details. For that, the design focus is solely on the preservation of relevant information, $I(\mathsf{x};\mathsf{t})$. Thus, practically, the only effective limit on compression rate, $I(\mathsf{y};\mathsf{z})$, will be stipulated by fixing the cardinality, $|\mathcal{Z}|$, of the quantizer's output alphabet. For $\beta \to \infty$, the second term of $\mathcal{L}_{\mathrm{IB}}^{\mathrm{JSCC}}$ in (7) predominates entirely. Hence, via substituting the minimization by maximization through dropping the minus sign, (6) boils down to

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y})}{\operatorname{argmax}} I(\mathsf{x};\mathsf{t}) . \tag{16}$$

This is a *concave* optimization task, i.e., *maximizing* a *convex* function over a closed convex set. Resorting to a well-known proposition which asserts a convex function attains its global maximum at an *extreme point* of its search space [68, Th. 1.19], it can then be realized that the focus will be on deterministic mappings, i.e., $p(\mathsf{z}|\mathsf{y}) \in \{0, 1\}$. To comprehend this, one may note that the validity condition, $\sum_z p(z|y) = 1$, for each realization, $y \in \mathcal{Y}$, introduces a *Probability Simplex*. Consequently, the overall search space in (16) will be a closed and convex polytope which is engendered by the Cartesian product of its constituent probability simplices [69]. The extreme points of this polytope happen at its corners. Furthermore, each corner corresponds to the Cartesian product of different corners of the individual probability simplices (all being deterministic), leading to a hard mapping for each extreme point.

Focusing on the *assignment phase* (quantizer input realizations to its output clusters) of the proposed algorithm that is performed by applying (8) per iteration, also reveals (as already discussed) that, in line with the discussion above, the resultant mapping will be deterministic. This is simply an effect of the present normalization function, $\psi$, which, for a given $y \in \mathcal{Y}$, concentrates all of the probability mass into the very cluster with the least relevant distortion among all candidates. Mathematically, denoting by $z^*(y)$ the chosen bin for $y \in \mathcal{Y}$

$$z^*(y) = \underset{z \in \mathcal{Z}}{\operatorname{argmin}} \ d_{\mathrm{IB}}^{\mathrm{JSCC}}(y, z) . \tag{17}$$

**Remark:** Interestingly, the criterion in (17) is identical to the one already provided in [55] regarding the SotA algorithm *Mutual Information-based Channel-Optimized Vector Quantizer (MICOVQ)* which delivers a vector quantizer aiming at maximizing the overall transmission rate in the same setup as the one illustrated in Fig. 2 with the main difference of considering vector-valued signals **x** and **y**. This becomes

more surprising when noting the fact that the mathematical deduction applied in [55] is completely different from the one presented here and does not involve the Variational Calculus. To realize that, the line of derivation in [55] is briefly discussed in the next section for the generalized VQ case. Finally, it is noteworthy that in [44] an algorithm has been proposed for multiterminal extension of the presumed system model in Fig. 2 wherein a number of observations from a source, x, are quantized locally with the aim of maximizing the overall transmission rate. The algorithmic equivalence of the approach presented in [44] and that of MICOVQ [55] has been established in [70] for the overlapping case of scalar quantization of a single observation.

## IV. IB-BASED JSCC VECTOR QUANTIZATION

In this part, we focus on a certain VQ problem and, after a brief discussion on the SotA routine MICOVQ [55] that only considers the special case of assuming an asymptotically large trade-off parameter ($\beta \to \infty$), we extend that by introducing the *Forward-Aware Vector IB (FAVIB)* algorithm which also allows for finite trade-off values. For explicit adaptation to the general case of $K$-dimensional VQ, in this study, we consider the length-$K$ source sequences, $\boldsymbol{x} = [x_1, \cdots, x_K] \in \mathcal{X}^K$, with the available distribution, $p(\mathbf{x})$, accompanied by the length-$K$ (noisy) observation sequences, $\boldsymbol{y} = [y_1, \cdots, y_K] \in \mathcal{Y}^K$, at the output of the discrete memoryless access channel depicted at the presumed system model in Fig. 2. This immediately implies that $p(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{K} p(\mathsf{y}_k|\mathsf{x}_k)$. Thus, knowing the transition probabilities, $p(\mathsf{y}|\mathsf{x})$, of the access channel indicates that the joint distribution, $p(\mathbf{x}, \mathbf{y})$, is also available. The aim is then to obtain a suitable quantizer, $p^*(\mathsf{z}|\mathsf{y})$, that (probabilistically) maps each vector, $\boldsymbol{y} \in \mathcal{Y}^K$, to every scalar, $z \in \mathcal{Z}$, as the solution of the following optimization

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y}):\, I(\mathbf{x};\mathsf{t}) \geq I}{\operatorname{argmin}} I(\mathbf{y};\mathsf{z}), \qquad (18)$$

where for the lower-bound of the relevant information, $I$, it applies $0 \leq I \leq H(\mathbf{x})$. Like before, making use of the method of *Lagrange Multipliers*, the constrained optimization in (18) can be transformed into an unconstrained one (up to the validity of the quantizer mapping) over an augmented objective functional (with *non-negative* trade-off parameter, $\beta$)

$$\mathcal{L}_{\mathrm{VQIB}}^{\mathrm{JSCC}} = I(\mathbf{y};\mathsf{z}) - \beta I(\mathbf{x};\mathsf{t}) . \qquad (19)$$

It is noteworthy that the upcoming discussion also applies to the memoryless access channels with continuous outputs via a prediscretization step to the required precision. The pertinent extension for the continuous-output forward channels is more demanding as the corresponding transition probabilities, $p(\mathsf{t}|\mathsf{z})$, may depend on the quantizer output probabilities, $p(\mathsf{z})$, (e.g., through an average power constraint) that are not known a priori.

### A. MICOVQ (MI-Based Channel-Optimized Vector Quantizer)

Attempting to achieve an end-to-end rate maximizing vector quantizer for the IB-JSCC setup, the authors in [55] have

focused on the extreme case of having an asymptotically large trade-off parameter, i.e., $\beta \to \infty$. For that, it is rather straight to discern that minimizing (19) boils down to maximizing the relevant information, $I(\mathbf{x};\mathsf{t})$, as the second term in (19) predominates entirely and the minimization is substituted by maximization through dropping the minus sign. Applying the chain rule for MI, it holds

$$I(\mathbf{x};\mathbf{y},\mathsf{t}) = I(\mathbf{x};\mathbf{y}) + I(\mathbf{x};\mathsf{t}|\mathbf{y}) = I(\mathbf{x};\mathsf{t}) + I(\mathbf{x};\mathbf{y}|\mathsf{t}) . \quad (20)$$

The presumed Markov chain implies that $I(\mathbf{x};\mathsf{t}|\mathbf{y}) = 0$ (given the observation y, the source, x, and the forward channel output signal, t, are statistically independent). Hence, the relevant information can be rewritten as

$$I(\mathbf{x};\mathsf{t}) = I(\mathbf{x};\mathbf{y}) - I(\mathbf{x};\mathbf{y}|\mathsf{t}), \qquad (21)$$

which immediately indicates that the maximization of $I(\mathbf{x};\mathsf{t})$ corresponds to the minimization of $I(\mathbf{x};\mathbf{y}|\mathsf{t})$ as $I(\mathbf{x};\mathbf{y})$ is fixed. Through the definition $C(\boldsymbol{y},t) = D_{\mathrm{KL}}(p(\mathbf{x}|\boldsymbol{y})\|p(\mathbf{x}|t))$, it holds

$$I(\mathbf{x};\mathbf{y}|\mathsf{t}) = \mathbb{E}_{\mathbf{y}}\{\mathbb{E}_{\mathsf{t}}\{C(\mathbf{y},\mathsf{t})|\mathbf{y}\}\}, \qquad (22)$$

in which the stated conditional expectation is calculated as

$$\mathbb{E}_{\mathsf{t}}\{C(\mathbf{y},\mathsf{t})|\mathbf{y}=\boldsymbol{y}\} = \sum_{z \in \mathcal{Z}} p(z|\boldsymbol{y}) \sum_{t \in \mathcal{T}} p(t|z) C(\boldsymbol{y},t) . \quad (23)$$

Since the inner sum in (23) is constant for each $z \in \mathcal{Z}$, to minimize the conditional expectation (23) for every $\boldsymbol{y} \in \mathcal{Y}^K$, the quantizer mapping must be chosen as $p(z|\boldsymbol{y}) = \delta_{z,z^*(\boldsymbol{y})}$, where

$$z^*(\boldsymbol{y}) = \underset{z}{\operatorname{argmin}} \sum_{t \in \mathcal{T}} p(t|z) C(\boldsymbol{y},t) . \qquad (24)$$

In this manner, the conditional information term in (22) will also be minimized for the given $C(\boldsymbol{y},t)$. To establish an iterative algorithm, the authors in [55] then suggest to commence with a random (yet valid) initialization, $C^{(0)}(\boldsymbol{y},t)$ and obtain the quantizer mapping, $p^{(i)}(\mathsf{z}|\mathsf{y})$, with the help of (24). This mapping is then utilized to update the distribution, $p^{(i+1)}(\mathbf{x}|t)$, and, subsequently, $C^{(i+1)}(\boldsymbol{y},t)$, putting into effect the relations implied by the presumed Markov chain. It should be mentioned that, contrary to the conventional vector quantizers, depending on forward channel transition probabilities, $p(\mathsf{t}|\mathsf{z})$, some elements of the quantizer's output alphabet *may not be utilized at all* [55], i.e., it might happen that $p(z) = 0$ for some $z \in \mathcal{Z}$.

### B. FAVIB (Forward-Aware Vector IB)

One shall note that, since the dimensionality of the source as well as the observation signal sequences does not play a role in the derivation of the optimal solution (as long as their joint distribution is available), all of the provided discussion in Section III-B for *scalar* case can be immediately translated to the $K$-dimensional *vector* case without posing further problems. Thus, the proposed algorithm in Theorem 2 is directly adaptable to the $K$-dimensional VQ scenario. This fact brings about a generalized routine that not only tallies with MICOVQ for $\beta \to \infty$, but also enables a full sweep through the entire gamut of $\beta$ values. Consequently, all the

---

**Algorithm 1**      <u>F</u>orward-<u>A</u>ware <u>V</u>ector <u>IB</u> (FAVIB)

---

**Input:** $p(\mathbf{x}, \mathbf{y})$, $\beta > 0$, stopping parameter $\varepsilon > 0$, $K > 0$, $p(\mathsf{t}|\mathsf{z})$

**Output:** A (soft) partition $\mathsf{z}$ of $\mathcal{Y}^K$ into (at most) $|\mathcal{Z}|$ bins

**Initialization:** $i = 0$, random mapping $p^{(0)}(\mathsf{z}|\mathbf{y})$

  **while** True **do**

    • $p^{(i)}(z) = \sum\limits_{\boldsymbol{y} \in \mathcal{Y}^K} p(\boldsymbol{y}) p^{(i)}(z|\boldsymbol{y})$   $\forall z \in \mathcal{Z}$

    • $p^{(i)}(\boldsymbol{x}|t) = \dfrac{\sum\limits_{\boldsymbol{y} \in \mathcal{Y}^K} \sum\limits_{z \in \mathcal{Z}} p(\boldsymbol{x},\boldsymbol{y}) p^{(i)}(z|\boldsymbol{y}) p(t|z)}{\sum\limits_{z' \in \mathcal{Z}} p^{(i)}(z') p(t|z')}$   $\forall \boldsymbol{x} \in \mathcal{X}^K, \forall t \in \mathcal{T}$

    • $p^{(i+1)}(z|\boldsymbol{y}) = \dfrac{p^{(i)}(z) \exp\left(-\beta \sum\limits_{t \in \mathcal{T}} p(t|z) D_{\mathrm{KL}}\big(p(\mathbf{x}|\boldsymbol{y}) \| p^{(i)}(\mathbf{x}|t)\big)\right)}{\psi^{(i+1)}(\boldsymbol{y}, \beta)}$

    • $i \leftarrow i+1$

    **if**   $\forall \boldsymbol{y} \in \mathcal{Y}^K$: $D_{\mathrm{JS}}^{\{\frac{1}{2}, \frac{1}{2}\}}\big(p^{(i)}(\mathsf{z}|\boldsymbol{y}) \| p^{(i-1)}(\mathsf{z}|\boldsymbol{y})\big) \leq \varepsilon$ **then**

     Break

    **end if**

  **end while**

---

achievable points in the respective *Information-Compression* plane become traversable via stochastic mappings resulted from choosing finite $\beta$ values. For each certain choice of the output cardinality, $|\mathcal{Z}|$, a concave curve is achieved. This curve starts from the origin ($\beta \to 0$, corresponding to *full diffusion* of the quantizer mapping, $p(\mathsf{z}|\mathbf{y})$) and by increasing $\beta$, other points on the graph are swept. This conforms to a gradual shift from high compression towards preservation of relevant information. In the extreme case of letting $\beta \to \infty$ (corresponding to *full concentration* of the quantizer mapping, $p(\mathsf{z}|\mathbf{y})$) this curve gets saturated at the maximum possible overall transmission rate, $I(\mathbf{x}; \mathsf{t})$, that can be supported by at most $|\mathcal{Z}|$ output levels. By this, for each choice of $|\mathcal{Z}|$, the achievable *Information-Compression* region is obtained as the set of points lying on and beneath the relevant curve. As discussed, (14) indicates that the slope of the tangent line at any point on the curve equals the inverse of pertinent $\beta$ value.

Regarding the claimed algorithmic equivalence among MICOVQ and FAVIB for $\beta \to \infty$, while the equivalence of *assignment phases* has been explicitly discussed (cp. (24) and (17)), the one concerning the *update phases*, i.e., recalculating $p(\mathbf{x}|t)$, can be realized by noting the fact that both algorithms, indeed, marginalize $p(\mathbf{x}, \mathbf{y}, \mathsf{z}, \mathsf{t})$, employing the relations induced by the presumed Markov chain. Exploiting the conventional vector quantization approaches (see, e.g., [71]) in the JSCC setup usually calls for separate consideration of the NP-hard quantizer output labeling problem as these methods mostly require fixed labels in advance. Contrarily, the FAVIB algorithm implicitly addresses this issue by directly working on forward transition probabilities, $p(\mathsf{t}|\mathsf{z})$, in the process of designing the quantizer that is a clear advantage.

Before concluding this section, we present the pseudo-code of the FAVIB routine in Alg. 1 that yields a $K$-dimensional vector quantizer extending the foundational idea of the IB method in its *full format* to the JSCC scenario. As already argued, FAVIB is a direct generalization of the provided iterative algorithm in Section III-C to the considered VQ setup. Explicitly, after a random initialization, it executes the convergent, alternating iterations over the adapted version of (11). These iterations are perpetuated until the bounded and symmetric Jensen-Shannon divergence between the obtained quantizer mappings from two consecutive iterations equals or falls below a threshold, $\varepsilon$. One may note, that the *a posteriori* term, $p(\mathbf{x}|\mathsf{t})$, for every $t \in \mathcal{T}$ appears directly as a potentially beneficial by-product to be utilized in the preplanned post-processing phase, e.g., efficient lookup table-based discrete channel decoding schemes with quite comparable performance to the conventional floating-point decoder (see, e.g., [72]).

## V. SIMULATION RESULTS

In this section, we specify some typical data transmission setups over which we perform a comprehensive analysis on the FAVIB algorithm to substantiate its effectiveness. To that end, we consider an equiprobable source signaling from a *Quadrature Phase Shift Keying (QPSK)* constellation with $\sigma_{\mathsf{x}}^2 = 1$ over an *Additive White Gaussian Noise (AWGN)* access channel model that is characterized by the noise variance, $\sigma_{\mathsf{n}}^2$. To obtain a discrete access channel, as will be explicitly mentioned, either we generate some samples out of the continuous model (the first two scenarios), or we perform a prediscretization over a fine grid (the last scenario). Generally, there is no specific restriction on the structure of the forward channel when utilizing the FAVIB algorithm. Indeed, throughout our numerical investigations, we will consider three various forward channel models: *a)* A square symmetric model being characterized by the reliability parameter, $e$, and the allowed number of output levels, $N$; *b)* A non-square model being characterized by the reliability parameter, $e$, and the forward channel output cardinality, $|\mathcal{T}|$; *c)* A square *Binary Symmetric Channel (BSC)*-based bit-pipe model being delineated by the bit-flip probability, $P_e$, as well as the allocated number of the forward bits, $M = \log_2 N$. The pertinent detailed descriptions will be provided in the upcoming respective subsections.

### A. An $N \times N$ Square Symmetric Forward Model

Here, for simplicity, we consider the scalar quantization case. For that, first we generate 256 access channel output samples. These received points are then clustered (utilizing FAVIB) into a varying number of bins ($N = 16$ to $64$). The forward channel is presumed to be an $N \times N$ symmetric model being purely delineated through the reliability parameter, $e$, in a following manner: For each input symbol, the correct reception occurs with probability $1 - e$ and the erroneous reception to every other output symbol occurs with probability $\frac{e}{N-1}$. Hence, the lower the $e$ value, the more reliable the respective transmissions. Moreover, it shall be noted that, for a certain reliability value, $e$, the choice of $N$ influences the transition probabilities, $p(\mathsf{t}|\mathsf{z})$. This forward channel (which, e.g., has been considered in [41] to model the impacts of a mmWave link) is chosen here in the first part of our numerical investigations solely for the sake of simplicity. We
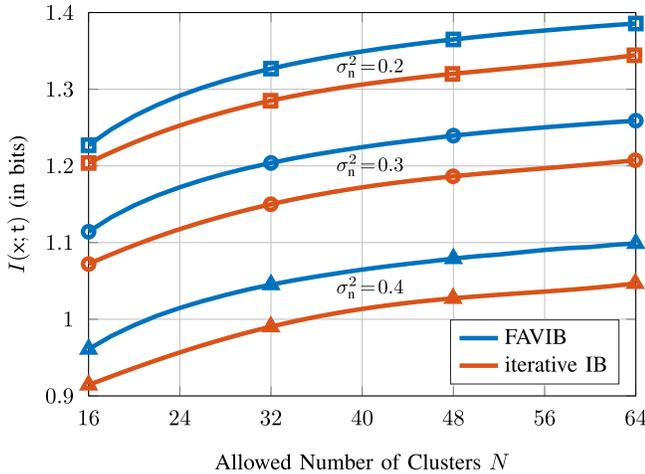
Fig. 3. Overall transmission rate, $I(\mathsf{x};\mathsf{t})$, vs. allowed number of clusters, $N$, equiprobable QPSK signaling ($\sigma_{\mathsf{x}}^2 = 1$), AWGN access channel with varying noise variance, $\sigma_{\mathsf{n}}^2$, $N \times N$ symmetric forward channel with fixed reliability, $e = 0.25$.
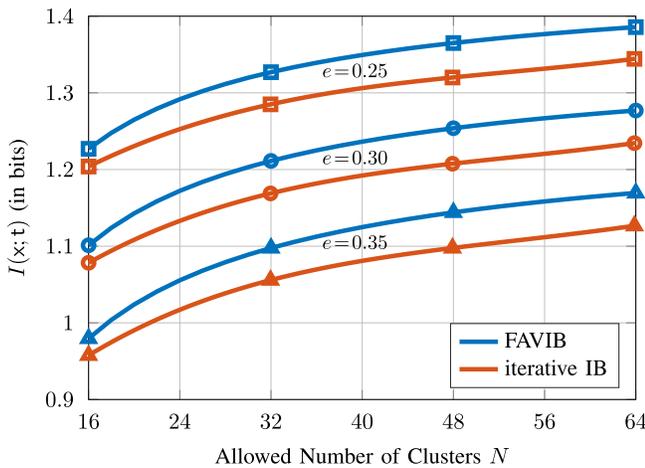


Fig. 4. Overall transmission rate, $I(\mathsf{x};\mathsf{t})$, vs. allowed number of clusters, $N$, equiprobable QPSK signaling ($\sigma_{\mathsf{x}}^2 = 1$), AWGN access channel with fixed noise variance, $\sigma_{\mathsf{n}}^2 = 0.2$, $N \times N$ symmetric forward channel with varying reliability, $e$.

set $\beta$ to 100 and as the performance indicator, we calculate the overall transmission rate, $I(\mathsf{x};\mathsf{t})$. The predescribed scenario is simulated for different noise variances (access channel) and various reliabilities (forward channel) to cover the effects of both present channels at the presumed system model in Fig. 2. The required quantization is performed by the proposed FAVIB algorithm and the baseline iterative IB routine [1] [5] to investigate whether incorporating the forward channel effects into the quantizer design formulation yields some performance enhancement or not (compared to the case where the error-prone forward channel is completely ignored). As these routines are initialized randomly, for the sake of fairness, we use the same starting point, $p^{(0)}(\mathsf{z}|\mathsf{y})$, for both

---

[5] As the iterative IB routine is proven to converge to a stationary point of its objective functional and there is no intention to assess any algorithmic aspect of the provided solution (e.g., the computational complexity, the convergence speed, etc.) in the upcoming numerical investigations, the iterative IB serves as a legitimate baseline.

approaches and to avoid poor local optima, we repeat each method 100 times and retain the best outcome.

Regarding Figs. 3 and 4, in the first case, we fix the reliability parameter of the forward channel to $e = 0.25$ and vary the noise variance, $\sigma_{\mathsf{n}}^2$, of the access channel. In the second case, we fix the access channel by choosing a certain noise variance, $\sigma_{\mathsf{n}}^2 = 0.2$, and change the reliability parameter, $e$, of the forward channel. In both cases we investigate the performance, $I(\mathsf{x};\mathsf{t})$, over a range of allowed output clusters, $N$. It is noteworthy that, for $\beta = 100$, FAVIB yields quite similar results as MICOVQ and, consequently, for the sake of a better readability the corresponding curves for MICOVQ are not drawn. Concerning both plots together, the main observation is that irrespective of the specific choices of the model parameters, i.e., the access channel noise variance, $\sigma_{\mathsf{n}}^2$, the forward channel reliability parameter, $e$, and the allowed number of output clusters, $N$, FAVIB outperforms the algorithm presented for the original IB setup, leading to higher overall transmission rates. This vividly reflects the fact that bringing the forward channel into the quantizer design formulation is beneficial.

Focusing on Fig. 3, it is observed that, expectedly, by increasing the noise variance, $\sigma_{\mathsf{n}}^2$, the end-to-end transmission rate, $I(\mathsf{x};\mathsf{t})$, decreases. To justify this, one shall note that, for a given forward channel, $I(\mathsf{x};\mathsf{t})$ is upper-bounded by the capacity of the access channel, $C_{\mathrm{AC}}$. This can be immediately realized by the direct application of the *Data Processing Inequality* [61] on the presumed Markov chain in our design setup. It is well known that, assuming a fixed input variance, $\sigma_{\mathsf{x}}^2$, the capacity of the discrete-input AWGN channel is reversely related to its noise variance, $\sigma_{\mathsf{n}}^2$ [73]. Hence, the lower the noise variance, the higher the capacity of the access channel, and thus the possibility of attaining higher overall transmission rates.

Regarding Fig. 4, the impact of forward channel on the overall transmission rate, $I(\mathsf{x};\mathsf{t})$, reflects itself on the value of the pertinent reliability parameter, $e$. Explicitly, it can be observed that in case of more reliable forward transmissions, the end-to-end rate increases. This can be justified quite analogously, noting the fact that, for a given access channel, $I(\mathsf{x};\mathsf{t})$ is also upper-bounded by the capacity of the forward channel, $C_{\mathrm{FC}}$. It will be rather undemanding to discern that, presuming a given number of output clusters, $|\mathcal{T}| = N$, the respective capacity of the forward channel, $C_{\mathrm{FC}}$, being calculated as [74]

$$C_{\mathrm{FC}}(N, e) = \log_2 N + (1 - e) \log_2 (1 - e) + e \log_2 \frac{e}{N-1}, \tag{25}$$

increases by decreasing the reliability parameter, $e$, enabling the overall transmission rate to reach higher values. From the discussion above, it is deduced that $I(\mathsf{x};\mathsf{t})$ is upper-bounded by the minimum capacity among the access and forward channels at the presumed system model in Fig. 2. To plainly observe that, one may consider Fig. 4 in case of $N = 32$ as an example. There, even though the capacity of the forward channel, $C_{\mathrm{FC}}$, is calculated as 2.95, 2.63 and 2.33 bits (per channel use) for different values (in ascending order) of the reliability parameter, $e$, the end-to-end transmission rate is
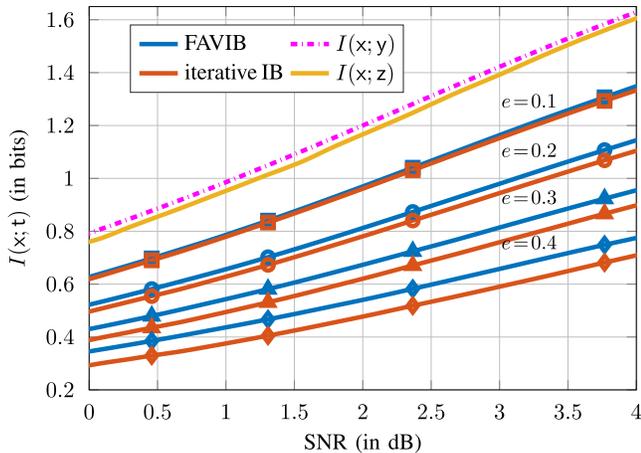
Fig. 5. Overall transmission rate, $I(\mathsf{x};\mathsf{t})$, vs. SNR, equiprobable QPSK signaling ($\sigma_\mathsf{x}^2 = 1$), AWGN access channel with varying noise variance, $\sigma_\mathsf{n}^2$, $N = 32$, $N \times N$ symmetric forward channel with different reliability values, $e = 0.1$ to $0.4$.
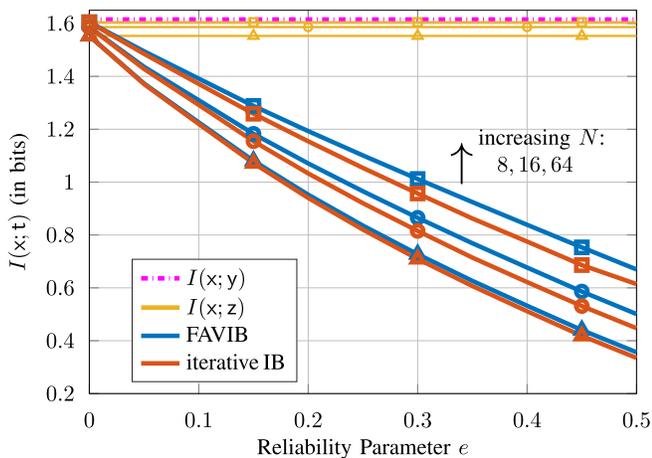


Fig. 6. Overall transmission rate, $I(\mathsf{x};\mathsf{t})$, vs. reliability parameter, $e$, equiprobable QPSK signaling ($\sigma_\mathsf{x}^2 = 1$), AWGN access channel with fixed noise variance, $\sigma_\mathsf{n}^2 = 0.4$, $N \times N$ symmetric forward channel with varying reliability parameter, $e$.

limited by the capacity (more precisely, the input-output MI assuming a uniform input signaling) of the access channel which amounts to 1.9 bits. Obviously, the theoretically optimal rate of 2 bits per channel use for the considered system setup is only achievable by asymptotically large values of *Signal-to-Noise-Ratio (SNR)* for the access channel together with relatively reliable transmissions over the forward channel (i.e., rather small $e$ values). As another important remark, it shall be noticed that, generally, by increasing $e$, corresponding to experiencing less reliable forwarding, the gap between the obtainable performances of the considered algorithms (FAVIB vs. iterative IB) widens. This statement is clearly substantiated in both Figs. 5 and 6.

Particularly, in Fig. 5 the attainable performance is investigated over a certain range of SNR ($\frac{\sigma_\mathsf{x}^2}{\sigma_\mathsf{n}^2}$) values for access channel with a fixed number of the allowed output levels, $N = 32$. In contrast, in Fig. 6 the achieved performance is drawn over a certain range of reliability values, $e$, when assuming a fixed access channel of SNR $\approx 4$ dB. It is commonly observable in

both figures that irrespective of the chosen allowed number of output levels, $N$, by increasing $e$, the outperformance of FAVIB w.r.t. the baseline is steadily more pronounced. Once again, this utterly indicates the key fact that the more critical the forwarding conditions are, the more advantageous the proposed routine becomes. The provided discussion for Figs. 3 and 4 is directly applicable to Fig. 5 as well. Explicitly, it can be noticed that for a given SNR, the more reliable the forward channel, the larger the attainable end-to-end rate. Likewise, for a given reliability value, $e$, the higher the SNR, the better the overall performance. As another interesting observation, it shall be noted that the information loss, i.e., the difference among the access channel's input-output MI (the ultimate upper-bound on transmission rate), $I(\mathsf{x};\mathsf{y})$, and the achieved end-to-end rate, $I(\mathsf{x};\mathsf{t})$, is mostly due to the forward path as the gap between $I(\mathsf{x};\mathsf{y})$ and the obtainable MI at the output of the quantizer block, $I(\mathsf{x};\mathsf{z})$, without considering the forward channel, is almost negligible for the applied compression factor of 8 (from $|\mathcal{Y}| = 256$ output channel samples to $N = 32$ clusters). Naturally, this information loss becomes more substantial for a less reliable forward path.

Concentrating on Fig. 6, it is observed that for an ideal forward channel, i.e., $e = 0$, the gap between the access channel's input-output MI, $I(\mathsf{x};\mathsf{y})$, and the one achieved after quantization, $I(\mathsf{x};\mathsf{z})$, increases with decreasing the allowed number of output levels, $N$. Then, regardless of the chosen value for $N$, through increasing $e$, the forward path becomes less reliable and, as already discussed, the information loss becomes more substantial. It is also noteworthy that for larger allowed output levels, $N$, FAVIB reveals its advantage over the baseline more tangibly, indicating the graver importance of incorporating the error-prone forward channel effects into the quantizer design formulation for such occasions.

Next, to acquire a crisp feeling about the role of the respective trade-off parameter, $\beta$, we generate Fig. 7. There, for a particular access channel noise variance, $\sigma_\mathsf{n}^2 = 0.4$, and also a fixed forward channel reliability value, $e = 0.25$, we vary the trade-off value, $\beta$, and depict the resultant *Information-Compression* pairs for three distinct choices of the allowed quantizer output cardinality, $N$. While MICOVQ can only provide an asymptotic point per choice of $N$, FAVIB enables a full sweep over the entire range of $\beta$. By this, all achievable points in the respective plane become traversable. Furthermore, it is observed that, basically, irrespective of the chosen value for $N$, through increasing $\beta$ the focus tends from compression towards preservation of the relevant information. More specifically, for finite values of $\beta$, both sides of the underlying trade-off (compactness and informativity) are treated via *stochastic* mappings. It is rather imaginable that, depending on the application, these soft decisions may become more beneficial. On the other hand, by letting $\beta \rightarrow \infty$, *hard* mappings merely highlight one side of the trade-off, namely, the informativity. It shall be noted that even in this extreme case, although the focus is solely on preservation of the relevant information, $I(\mathsf{x};\mathsf{t})$, still the compression rate, $I(\mathsf{y};\mathsf{z})$, can not grow arbitrarily large and is upper-bounded by $\log_2 N$ bits that is an ultimate upper-bound on the quantizer output entropy, $H(\mathsf{z})$.
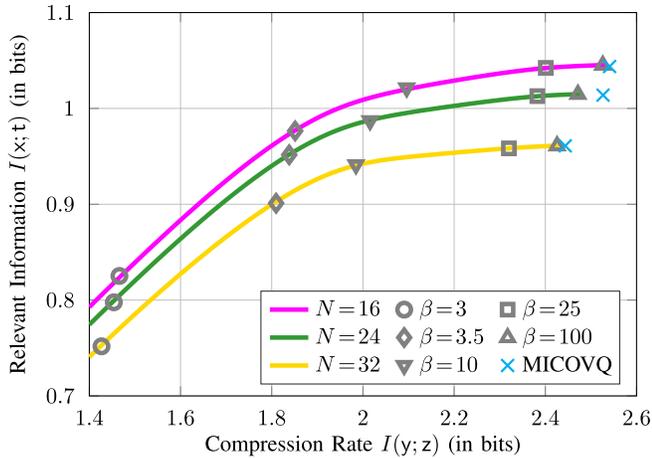
Fig. 7. Relevant information, $I(\mathsf{x};\mathsf{t})$, vs. compression rate, $I(\mathsf{y};\mathsf{z})$, equiprobable QPSK signaling ($\sigma_\mathsf{x}^2 = 1$), AWGN access channel with fixed noise variance, $\sigma_\mathsf{n}^2 = 0.4$, $N \times N$ symmetric forward channel with fixed reliability, $e = 0.25$, varying $\beta$.
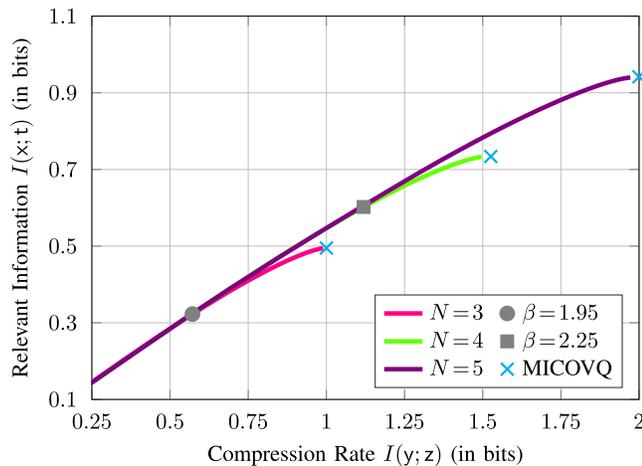


Fig. 8. Relevant information, $I(\mathsf{x};\mathsf{t})$, vs. compression rate, $I(\mathsf{y};\mathsf{z})$, equiprobable QPSK signaling ($\sigma_\mathsf{x}^2 = 1$), AWGN access channel with fixed noise variance, $\sigma_\mathsf{n}^2 = 0.4$, $N \times 16$ forward channel with fixed reliability, $e = 0.25$, varying $\beta$.

### B. An $N \times |\mathcal{T}|$ Non-Square Forward Model

In this part of our numerical investigations, we consider the case in which forward transition probabilities are independent of the allowed quantizer output cardinality, $N$. For that, we fix the forward output cardinality to $|\mathcal{T}| = 16$ and presume an $N \times 16$ forward channel model with reliability parameter, $e$, in the sense that for each input symbol, irrespective of the chosen value of $N$, there will be one transition to a specific output symbol with probability $1-e$ and other transitions are happening with the same probability of $\frac{e}{15}$. The corresponding results for three different choices of the allowed number of output clusters, $N = 3, 4, 5$, are illustrated in Fig. 8. Specifically, by varying the trade-off parameter, $\beta$, three concave curves are attained. These graphs start from the origin and bifurcate at certain (critical) values, $\beta = 1.95, 2.25$. This behavior corroborates the pertinent discussion made in Section III-D about the phase transition phenomena happening at critical $\beta$ values for a family of concave curves pertinent to different choices of the

allowed output cardinality, $N$. It shall be noted that each of these sub-optimal curves splits the Information-Compression plane into two segments of *achievable* (below the curve) and *non-achievable* (above the curve) Information-Compression rate-tuples for every choice of the allowed output cardinality, $N$. Naturally, the overall possible achievable region is then characterized by the trajectory of the curve with full output granularity, i.e., $N = |\mathcal{Y}|$.

### C. A Square BSC-Based Bit-Pipe Forward Model

In this part, we would like to gain some further insights about the engendered *quantization clusters* when applying the *joint* and *individual* compression of both the inphase (i.e., $\mathrm{Re}\{\cdot\}$) and quadrature (i.e., $\mathrm{Im}\{\cdot\}$) signals from a particular source constellation followed by the equivalent joint as well as the individual access and forward channel models for the sake of a fair comparison. Specifically, we consider an equiprobable QPSK source signaling ($\sigma_\mathsf{x}^2 = 1$) over an AWGN access channel with a certain noise variance, $\sigma_\mathsf{n}^2$. To obtain discrete access channel models, we perform a fine equidistant prediscretization of both the real and imaginary part of the pertinent continuous output density function to 256 segments, restricting solely to the region with the border guard interval of $\frac{3\sigma_\mathsf{n}}{\sqrt{2}}$ from data points to ensure 99.7% coverage. Doing so, the individual access channels' matrices (inphase, $p(\mathrm{Re}\{\mathsf{y}\}|\mathrm{Re}\{\mathsf{x}\})$, and quadrature, $p(\mathrm{Im}\{\mathsf{y}\}|\mathrm{Im}\{\mathsf{x}\})$) will be of dimension $2 \times 2^8$ and, accordingly, the equivalent joint access channel matrix, $p(\mathsf{y}|\mathsf{x})$, will be of dimension $2^2 \times 2^{16}$.

Regarding the forward channel, we consider a bit-pipe model in which different bits of the input sequence will be treated separately via a BSC with a particular bit-flip probability, $P_e$. Specifically, the integers $0, \cdots, N-1$ are assigned as labels to individual quantization clusters (for inphase and quadrature signals) and the bit-tuples of length $M = \log_2 N$ which represent these integers are then fed into the forward channels. This indicates that the forward transition probabilities are determined by the Hamming distance, $\alpha$, between the respective input-output sequences. Thus, the individual (inphase and quadrature) forward channels' matrices consist of entries having the form $P_e^\alpha (1-P_e)^{M-\alpha}$ with $0 \leq \alpha \leq M$. By stacking together the binary representations of the quantizer output labels for inphase and quadrature signals, the equivalent joint forward channel matrix is obtained that is of the squared size (i.e., $N^2$) per dimension. In what follows, we set the access channel noise variance to $\sigma_\mathsf{n}^2 = 0.3$ and $N = 4$. Further, we assume the trade-off parameter, $\beta$, to be asymptotically large (therefore obtaining deterministic mappings) and to avoid poor local optima, we rerun the FAVIB algorithm $10^3$ times with various starting points, $p^{(0)}(\mathsf{z}|\mathsf{y})$, and retain the best outcome. The clusters obtained for the joint and individual treatments are illustrated in Figs. 9 and 10, respectively.

The key enabler to understand and analyze the obtained results is, indeed, the concept of *Partial Mutual Information* that indicates the individual contributions of different clusters to the overall transmission rate, $I(\mathsf{x};\mathsf{t})$. Specifically, denoting
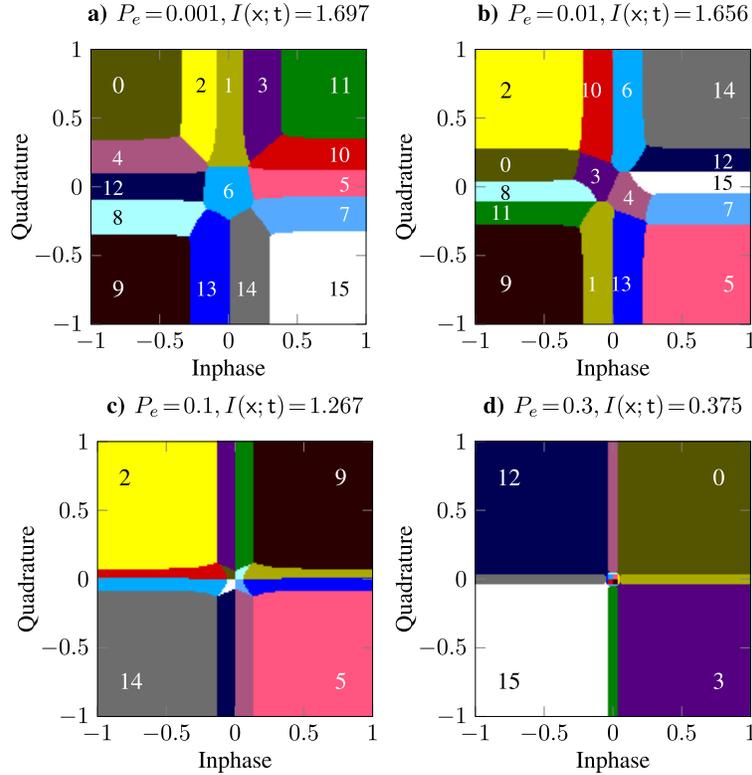
Fig. 9. Quantization regions and integer labels for joint treatment of inphase and quadrature signals, equiprobable QPSK ($\sigma_x^2 = 1$), AWGN access channel with fixed noise variance, $\sigma_n^2 = 0.3$, bit-pipe BSC-based forward channel with varying $P_e$.

TABLE I

PARTIAL MUTUAL INFORMATION, $i(z)$, FOR JOINT TREATMENT OF THE INPHASE AND QUADRATURE SIGNALS, BIT-PIPE BSC-BASED FORWARD CHANNEL WITH VARYING BIT-FLIP PROBABILITY, $P_e$

| | Decimal value of binary label [Right-LSB] | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $P_e$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | $\sum$ |
| .001 | .33 | .016 | .04 | .051 | .041 | .016 | $\approx 0$ | .039 | .043 | .338 | .041 | .313 | .016 | .034 | .037 | .341 | 1.697 |
| .01 | .029 | .021 | .361 | .001 | $\approx 0$ | .361 | .021 | .028 | .011 | .361 | .021 | .024 | .024 | .022 | .361 | .012 | 1.656 |
| .1 | $\approx 0$ | .001 | .312 | .003 | .003 | .311 | .002 | $\approx 0$ | $\approx 0$ | .312 | .001 | .003 | .003 | .002 | .311 | $\approx 0$ | 1.267 |
| .3 | .093 | $\approx 0$ | $\approx 0$ | .093 | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | .093 | $\approx 0$ | $\approx 0$ | .093 | .375 |

by $i(z)$ the partial MI in a certain cluster, $z$, it is implied that for the overall transmission rate, it must apply $I(\mathsf{x}; \mathsf{t}) = \sum_z i(z)$. To analytically derive the partial MI, $i(z)$, it shall be noted that the following holds

$$I(\mathsf{x}; \mathsf{t}) = \sum_z \sum_{y,x,t} p(x,y,z,t) \log \underbrace{\frac{\sum_{z' \in \mathcal{Z}} \sum_{y' \in \mathcal{Y}} p(t|z')p(z'|y')p(y'|x)}{p(t)}}_{i(z)}.$$

(26)

Tables I and II provide in detail the calculated partial MI terms for results depicted in Figs. 9 and 10 regarding the joint and individual treatment of the inphase and quadrature signals.

Starting with Fig. 9a (joint treatment of the inphase and quadrature signals with relatively reliable forwarding) and noting the pertinent partial MI terms from Tab. I, three distinct types of clusters can be immediately spotted:

1) Four corner clusters $\{0, 9, 11, 15\}$ lying in the vicinity of different source symbols and substantially contributing to the overall transmission rate ($i(\cdot) \geq 0.31$),
2) Cluster $\{6\}$ locating around the origin and covering the points with similar Euclidean distance to all different input source symbols which provides almost no information about the source ($i(6) \approx 0$) and acts like an *erasure*,
3) Other clusters situating in the middle of the adjacent source symbols with limited contributions to the overall transmission rate ($0.016 \leq i(\cdot) \leq 0.051$).

For quite reliable forwarding, e.g., $P_e = 0.001$, the choice of labeling is rather inconsequential as, obviously, the inter-cluster transitions engendered by forwarding errors are less probable. This clearly explains the fairly arbitrary choice of labeling for different clusters in Fig. 9a.

In contrast, for less reliable forwarding, i.e., larger bit-flip probabilities, $P_e$, the appropriate choice of labeling becomes more crucial as it plays an important role in minimizing
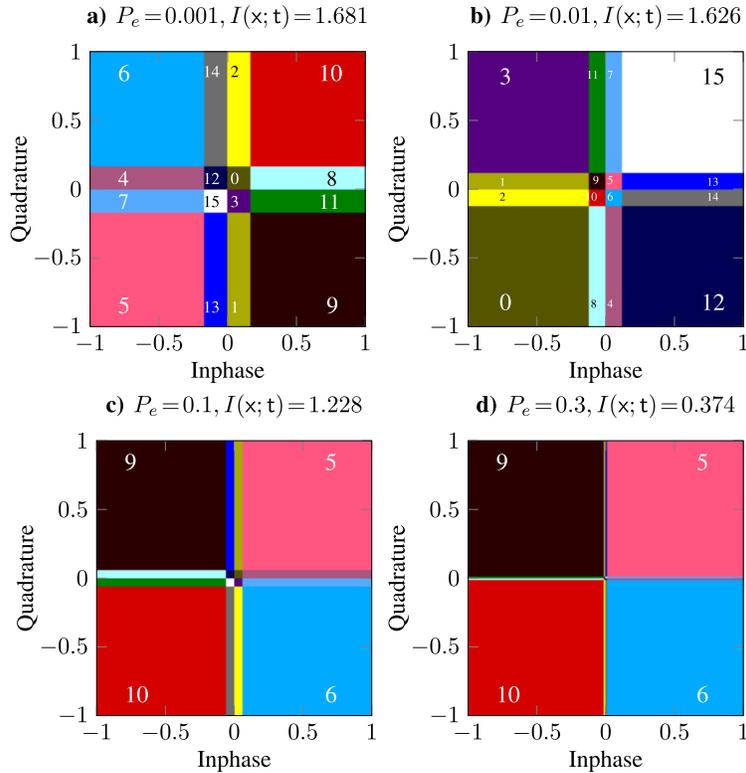
Fig. 10. Quantization regions and integer labels for separate treatment of inphase and quadrature signals, equiprobable QPSK ($\sigma_x^2 = 1$), AWGN access channel with fixed noise variance, $\sigma_n^2 = 0.3$, bit-pipe BSC-based forward channel with varying $P_e$.

TABLE II

PARTIAL MUTUAL INFORMATION, $i(z)$, FOR INDIVIDUAL TREATMENT OF THE INPHASE AND QUADRATURE SIGNALS, BIT-PIPE BSC-BASED FORWARD CHANNEL WITH VARYING BIT-FLIP PROBABILITY, $P_e$

| $P_e$ | Decimal value of binary label [Right-LSB] | | | | | | | | | | | | | | | | $\sum$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| .001 | $\approx 0$ | .016 | .016 | $\approx 0$ | .016 | .387 | .387 | .016 | .016 | .387 | .387 | .016 | $\approx 0$ | .016 | .016 | $\approx 0$ | 1.681 |
| .01 | .386 | .010 | .010 | .386 | .010 | $\approx 0$ | $\approx 0$ | .010 | .010 | $\approx 0$ | $\approx 0$ | .010 | .386 | .010 | .010 | .386 | 1.626 |
| .1 | $\approx 0$ | .002 | .002 | $\approx 0$ | .002 | .303 | .303 | .002 | .002 | .303 | .303 | .002 | $\approx 0$ | .002 | .002 | $\approx 0$ | 1.228 |
| .3 | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | .094 | .094 | $\approx 0$ | $\approx 0$ | .094 | .094 | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | $\approx 0$ | .374 |

the inevitable information loss resulting from inter-cluster transitions due to erroneous forward transmissions. A closer look at the presented results in Fig. 9b-d reveals the following trend: The pertinent labels for the corner clusters carrying the most of information about the source, x, are chosen in the fashion that the Hamming distance, $\alpha$, between two pairs with the point symmetry w.r.t. the origin is getting maximized. That, in turn, minimizes the occurrence probability of such worst-case inter-cluster transitions. This behavior is easily verifiable by considering Fig. 9c-d. While the pairs $\{2, 14\}$ and $\{9, 14\}$ for $P_e = 0.1$ provide the Hamming distance of $\alpha = 3$, both their counterparts $\{12, 3\}$ and $\{0, 15\}$ in case of $P_e = 0.3$ enjoy the maximum protection through the largest possible Hamming distance of $\alpha = 4$. Furthermore, it is observable that for less reliable forwarding the tendency is to allocate more clusters to the erasure-like region in the vicinity of origin. Effectively, these clusters are almost inactive as they do not meaningfully contribute to the process of information preservation ($i(\cdot) \approx 0$). It is noteworthy

that in [41] an analogous behavior has been observed as well.

The aforementioned trend can be nicely interpreted as a type of *inherent error protection* being performed by the applied quantization scheme. In principle, for supporting four input source points, sixteen clusters are available, but through worse forwarding conditions, smaller numbers of clusters are actively utilized (introducing redundancy) and the labels are chosen diligently such that those active clusters get well protected against the inter-cluster transitions.

It is also noteworthy that although the underlying problem setup is completely symmetric, the obtained results are not. In cases of rather small bit-flip probabilities, $P_e$, the joint forward channel capacity, $C_{\text{FC}}$, being calculated as [74] ($M = 4$)

$$C_{\text{FC}}(P_e) = 4 + \sum_{i=0}^{4} \binom{4}{i} P_e^i (1 - P_e)^{4-i} \log_2 P_e^i (1 - P_e)^{4-i},$$

(27)

is relatively large. This fact allows the quantizer to employ most of the available clusters for engendering a finer decomposition at each quadrant that, in turn, leads to an increase in the overall transmission rate, $I(\mathsf{x};\mathsf{t})$. Therefore, only one or two clusters will be allocated around the origin ($P_e = 0.001, 0.01$) and the rest are meaningfully utilized elsewhere. This justifies the noticeable asymmetry in the depicted results at the upper-part of Fig. 9. Contrarily, in case of larger bit-flip probabilities, $P_e$, limited forward channel capacity, $C_{\mathrm{FC}}$, cannot afford the luxury of such fine decompositions. Hence, the quantizer is forced to convey the information with lower number of active clusters. The obtained results in such occasions exhibit more symmetry. This can be observed by the depicted results at the lower-part of Fig. 9.

Considering Fig. 10 for individual treatment of the inphase and quadrature signals, it can be readily noticed that, contrary to the joint treatment and irrespective of the quality of the forward transmission, the symmetry in the outcome is always ensured. This behavior is quite natural as the underlying setup for both the inphase and quadrature signals are exactly the same. Hence, expectedly, the resultant outcomes per dimension of the complex plane become identical as well. This present symmetry can also be easily substantiated by skimming through the calculated partial MI terms which have been provided in Tab. II.

Another interesting observation is about the effect of the quality of forward transmission on the obtained quantization regions. Explicitly, through an increase in the bit-flip probability, $P_e$, corresponding to experiencing less reliable forwarding, the clusters in the middle regions between all pairs of neighboring points of the QPSK source constellation shrink. To explain this, an analogous line of reasoning as before regarding the inherent error protection applies here as well. Basically, per dimension of the complex plane four clusters are available for two input source points. Through less reliable forwarding, two clusters in the middle become gradually inactive to engender the required redundancy for protection of the other two against inter-cluster transitions occurring due to transmission over an erroneous forward path. This clearly indicates that, principally, the applied compression scheme can obviate the need for separate employment of channel coding on the forward path.

Eventually, it is important to observe that performing individual quantization of the inphase and quadrature signals also reveals a kind of proper error protection behavior. The applied labeling on different quantization regions in Fig. 10 clearly exhibits a quite well separation with largest possible Hamming distance of $\alpha = 4$ for all pairs of regions with point symmetry w.r.t. the origin. This observation justifies to some extent the relatively close performance of the applied individual (inphase and quadrature) quantization scheme to the joint compression that can be realized by noting the stated overall transmission rates, $I(\mathsf{x};\mathsf{t})$, in Figs. 9 and 10.

## VI. SUMMARY

We focused on the problem of *Joint Source-Channel Coding*. In that regard, here for the first time, we extended the mathematical framework of the *Information Bottleneck* method in its full format for enabling its applicability to this generalized scenario. Particularly, applying the *Variational Calculus*, we derived the formal solution of the introduced design optimization and utilized that to propose a novel algorithm which extends the SotA methods by capacitating a full sweep through the entire trade-off parameter's range. Further, we provided a brief proof of convergence to a stationary point of its objective functional along with an in-depth analysis on its mathematical aspects. Finally, we corroborated its effectiveness as well via several numerical simulations.

## APPENDIX

### A. Proof of Theorem 1

Introducing a Lagrange multiplier, $\lambda_y$, for each realization, $y \in \mathcal{Y}$, of the observation signal, $\mathsf{y}$, the validity conditions can be incorporated into the overall JSCC Lagrangian, $\mathcal{L}_{\mathrm{O}}$, being defined as

$$\mathcal{L}_{\mathrm{O}} = I(\mathsf{y};\mathsf{z}) - \beta I(\mathsf{x};\mathsf{t}) + \sum_{y \in \mathcal{Y}} \lambda_y \Big( \sum_{z \in \mathcal{Z}} p(z|y) - 1 \Big) . \quad (28)$$

Taking the functional derivative of (28) w.r.t. the present variational variable, i.e., $p(\mathsf{z}|\mathsf{y})$, and equating it to zero yields the required optimality criterion, $\frac{\delta \mathcal{L}_{\mathrm{O}}}{\delta p(\mathsf{z}|\mathsf{y})} = 0$. Based on the definition of MI and also the relations implied by the presumed Markov chain, the following functional derivatives are obtained for three constituent terms in (28)

$$\frac{\delta I(\mathsf{y};\mathsf{z})}{\delta p(\mathsf{z}|\mathsf{y})} = \frac{\delta \big( H(\mathsf{z}) - H(\mathsf{z}|\mathsf{y}) \big)}{\delta p(\mathsf{z}|\mathsf{y})} = p(y) \log \frac{p(z|y)}{p(z)}, \quad (29)$$

$$\frac{\delta I(\mathsf{x};\mathsf{t})}{\delta p(\mathsf{z}|\mathsf{y})} = \frac{\delta \big( H(\mathsf{t}) - H(\mathsf{t}|\mathsf{x}) \big)}{\delta p(\mathsf{z}|\mathsf{y})} \quad (30\mathrm{a})$$

$$= p(y) \Big( \sum_{t \in \mathcal{T}} p(t|z) \sum_{x \in \mathcal{X}} p(x|y) \log \frac{p(x|t)}{p(x)} \Big), \quad (30\mathrm{b})$$

and

$$\frac{\delta \Big( \sum\limits_{y \in \mathcal{Y}} \lambda_y \big( \sum\limits_{z \in \mathcal{Z}} p(z|y) - 1 \big) \Big)}{\delta p(\mathsf{z}|\mathsf{y})} = \lambda_y . \quad (31)$$

Due to the positivity of $p(y)$ and with the direct use of the obtained functional derivatives, it is immediately deduced that the following holds

$$\log \frac{p(z|y)}{p(z)} + \beta \sum_{t \in \mathcal{T}} p(t|z) \sum_{x \in \mathcal{X}} p(x|y) \log \frac{p(x)}{p(x|t)} + \frac{\lambda_y}{p(y)} = 0. \quad (32)$$

Hence, through the definition of KL divergence, it applies

$$\log \frac{p(z|y)}{p(z)} + \beta \sum_{t \in \mathcal{T}} p(t|z) D_{\mathrm{KL}} \big( p(\mathsf{x}|y) \| p(\mathsf{x}|t) \big) + \tilde{\lambda}_y = 0, \quad (33)$$

with

$$\tilde{\lambda}_y = \beta \sum_{x \in \mathcal{X}} p(x|y) \log \frac{p(x)}{p(x|y)} + \frac{\lambda_y}{p(y)} . \quad (34)$$

Bringing both the second and the third summands in (33) to the other side of the equality, exponentiating both sides and finally multiplying by $p(z)$, one obtains

$$p(z|y) = p(z) \exp\left(-\beta \sum_{t \in \mathcal{T}} p(t|z) D_{\mathrm{KL}}\big(p(\mathsf{x}|y)\|p(\mathsf{x}|t)\big) - \tilde{\lambda}_y\right).$$

(35)

Enforcing the validity condition, $\sum_z p(z|y) = 1$, and noting that $\tilde{\lambda}_y$ does not depend on $z$, one can treat $\exp(\tilde{\lambda}_y)$ as the partition function, $\psi(y, \beta)$, to come into the form of (8). ∎

*B. Proof of Theorem 2*

Using (8), for each $(y, z) \in \mathcal{Y} \times \mathcal{Z}$ it holds

$$-\log \psi(y, \beta) = \log \frac{p(z|y)}{p(z)} + \beta \sum_{t \in \mathcal{T}} p(t|z) D_{\mathrm{KL}}\big(p(\mathsf{x}|y)\|p(\mathsf{x}|t)\big).$$

(36)

Applying the expectation operator, $\mathbb{E}_{\mathsf{y},\mathsf{z}}\{f(\cdot)\} = \sum_{y,z} p(y, z) f(\cdot)$, on (36), it can be realized that the (negative) free energy, $\mathcal{F}$, is a sum of KL divergences. Hence, it will be *non-negative* and also *convex* w.r.t. each of its arguments separately [61] as $\beta \geq 0$ and the sum of these convex functions is also convex. Further, it is rather straightforward to verify that equating the functional derivatives of $\mathcal{F}$ w.r.t. each of its arguments (presuming appropriate normalization constraints) to zero engenders exactly the relations in (11) that are expressed in the statement of Theorem 2. Therefore, updating one of these equations while fixing the other two can only reduce $\mathcal{F}$ or keep it unchanged. This is simply due to the separate convexity of $\mathcal{F}$ w.r.t. each of its arguments, along with the fact that all the arguments are constrained to belong to a convex set (projection over a convex set). As $\mathcal{F}$ is bounded from below, the convergence is guaranteed. ∎

## 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] P.-N. Tan, M. Steinbach, and V. Kumar, *Data Mining Cluster Analysis: Basic Concepts Algorithms*. London, U.K.: Pearson, 2013, ch. 8.

[3] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006. [Online]. Available: https://www.springer.com/gp/book/9780387310732

[4] S. Hassanpour, D. Wubben, 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.

[5] 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.

[6] S. Hassanpour, D. Wubben, and A. Dekorsy, "On the equivalence of double maxima and KL-means for information bottleneck-based source coding," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Barcelona, Spain, Apr. 2018, pp. 1–6.

[7] S. Hassanpour, D. Wubben, and A. Dekorsy, "A graph-based message passing approach for noisy source coding via information bottleneck principle," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[8] S. Hassanpour, D. Wubben, 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.

[9] S. Hassanpour, D. Wubben, 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.

[10] N. Slonim, N. Friedman, and N. Tishby, "Multivariate information bottleneck," *Neural Comput.*, vol. 18, no. 8, pp. 1739–1789, Aug. 2006.

[11] 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.

[12] I. Estella Aguerri and A. Zaidi, "Distributed information bottleneck method for discrete and Gaussian sources," 2017, *arXiv:1709.09082*. [Online]. Available: http://arxiv.org/abs/1709.09082

[13] B. Nazer, O. Ordentlich, and Y. Polyanskiy, "Information-distilling quantizers," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 96–100.

[14] Q. Yang, P. Piantanida, and D. Gunduz, "The multi-layer information bottleneck problem," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 404–408.

[15] I. Estella-Aguerri and A. Zaidi, "Distributed variational representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 19, 2020, doi: 10.1109/TPAMI.2019.2928806.

[16] 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.

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

[18] P. Harremoes 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.

[19] Y. Steinberg, "Coding and common reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4995–5010, Nov. 2009.

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

[21] R. Ahlswede and J. Korner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, Nov. 1975.

[22] E. Erkip and T. M. Cover, "The efficiency of investment information," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1026–1040, May 1998.

[23] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IEEE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[24] D. J. Sakrison, "Source encoding in the presence of random disturbance (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 165–167, Jan. 1968.

[25] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 406–411, Jul. 1970.

[26] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 493–501, Sep. 1975.

[27] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[28] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 826–834, Jul. 1988.

[29] E. Ayanoglu, "On optimal quantization of noisy sources," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1450–1452, Nov. 1990.

[30] 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.

[31] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[32] J. Lewandowsky and G. Bauch, "Trellis based node operations for LDPC decoders from the information bottleneck method," in *Proc. 9th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Cairns, QLD, Australia, Dec. 2015, pp. 1–10.

[33] J. Lewandowsky and G. Bauch, "Information-optimum LDPC decoders based on the information bottleneck method," *IEEE Access*, vol. 6, pp. 4054–4071, 2018.

[34] M. Stark, J. Lewandowsky, and G. Bauch, "Information-optimum LDPC decoders with message alignment for irregular codes," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[35] F. J. C. Romero and B. M. Kurkoski, "LDPC decoding mappings that maximize mutual information," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 9, pp. 2391–2401, Sep. 2016.

[36] R. Ghanaatian *et al.*, "A 588-Gb/s LDPC decoder based on finite-alphabet message passing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 26, no. 2, pp. 329–340, Feb. 2018.

[37] T. Monsees, D. Wübben, and A. Dekorsy, "Channel-optimized information bottleneck design for signal forwarding and discrete decoding in cloud-RAN," in *Proc. 12th Int. ITG Conf. Syst., Commun. Coding (SCC)*, Feb. 2019, pp. 1–6.

[38] G. Zeitler, 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.

[39] 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.

[40] 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.

[41] J. Bartelt, L. Landau, and G. Fettweis, "Improved uplink I/Q-signal forwarding for cloud-based radio access networks with millimeter wave fronthaul," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 341–345.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] C. Roth, C. Benkeser, C. Studer, G. Karakonstantis, and A. Burg, "Data mapping for unreliable memories," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 679–685.

[48] A. Kurtenbach and P. Wintz, "Quantizing for noisy channels," *IEEE Trans. Commun.*, vol. 17, no. 2, pp. 291–302, Apr. 1969.

[49] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 827–838, Nov. 1987.

[50] K. Zeger and A. Gersho, "Pseudo-gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.

[51] J. B. De Marca *et al.*, "An algorithm for assigning binary indices to the codevectors of a multi-dimensional quantizer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jan. 1987, pp. 1128–1132.

[52] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[53] 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.

[54] E. Ayanoglu and R. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 855–865, Nov. 1987.

[55] 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.

[56] S. Z. Azami, P. Duhamel, and O. Rioul, "Joint source-channel coding: Panorama of methods," in *Proc. CNES Workshop Data Compress.*, Nov. 1996, pp. 1232–1254.

[57] F. Hekland, "A review of joint source-channel coding," Norwegian Univ. Sci. Technol. (NTNU), Trondheim, Norway, Tech. Rep., Feb. 2004, p. 12. [Online]. Available: https://pdfs.semanticscholar.org/36be/bd540d3626e5c5896e891402cd827ecfed79.pdf

[58] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 7, pp. 142–163, Mar. 1959.

[59] S. Vembu, S. Verdu, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.

[60] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, vol. 159. New York, NY, USA: Springer, 2012. [Online]. Available: https://www.springer.com/gp/book/9780792391814

[61] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[62] N. Slonim, "The information bottleneck: Theory and applications," Ph.D. dissertation, Interdiscipl. Center Neural Comput. (ICNC), Hebrew Univ. Jerusalem, Jerusalem, Israel, 2002. [Online]. Available: https://researcher.watson.ibm.com/researcher/view.php?person=il-NOAMS

[63] J. H. Mathews *et al.*, *Numerical Methods Using MATLAB*, vol. 4. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.

[64] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[65] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[66] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 39, no. 1, pp. 1–22, Sep. 1977.

[67] K. Rose, E. Gurewitz, and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1249–1257, Jul. 1992.

[68] R. Horst, P. M. Pardalos, and N. Van Thoai, *Introduction to Global Optimization*, 2nd ed. New York, NY, USA: Springer, 2000. [Online]. Available: https://www.springer.com/gp/book/9780792365747

[69] T. Gedeon, A. E. Parker, and A. G. Dimitrov, "The mathematical structure of information bottleneck methods," *Entropy*, vol. 14, no. 3, pp. 456–479, Mar. 2012.

[70] S. Hassanpour, D. Wubben, and A. Dekorsy, "On the equivalence of two information bottleneck-based routines devised for joint source-channel coding," in *Proc. 25th Int. Conf. Telecommun. (ICT)*, Jun. 2018, pp. 253–258.

[71] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[72] T. Monsees, D. Wubben, and A. Dekorsy, "Information preserving quantization and decoding for satellite-aided 5G communications," in *Proc. IEEE 2nd 5G World Forum (5GWF)*, Dresden, Germany, Sep. 2019, pp. 516–519.

[73] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[74] 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 University of Ulm, Germany, in 2014. In 2015, he joined the Department of Communications Engineering, University of Bremen, Germany, where he is currently pursuing the Ph.D. degree. His research interests include information theory (source and joint source-channel coding), statistical signal processing, wireless communications, and machine learning.



**Tobias Monsees** (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Applied Sciences, Bremen, Germany, in 2013, and the M.Sc. degree in communication and information technology from the University of Bremen, Germany, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Communications Engineering. His special research interests include information theory, statistical signal processing, and machine learning.

**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, Germany. His research interests include wireless communications, signal processing, multiple antenna systems, cooperative communication systems, channel coding, information theory, and machine learning. He has published more than 140 papers in international journals and conference proceedings. 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 COMMUNICATION 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 (inter)national research projects (more than 25 BMBF/BMWI/EU projects) in affiliation with his scientific expertise shown by more than 200 journal 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 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 Technologiezentrum Informatik und Informationstechnik (TZI) of the University of Bremen. 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.