IEEE Open Journal of the Communications Society

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/OJCOMS.2026.XXXXXXX

Learning to Compress over Arbitrary Channels: Information Bottleneck-Based Distributed Joint Source-Channel Coding via Deep MARL

SHAYAN HASSANPOUR (Member, IEEE), DIRK WÜBBEN (Senior Member, IEEE), AND ARMIN DEKORSY (Senior Member, IEEE)

Department of Communications Engineering, University of Bremen, 28359 Bremen, Germany

CORRESPONDING AUTHOR: S. HASSANPOUR (e-mail: hassanpour@ant.uni-bremen.de)

This research was partly supported by the German Federal Ministry of Research, Technology and Space (BMFTR) within the project Open6GHub+ under grant number 16KIS2408.

ABSTRACT In this article, we introduce a Reinforcement Learning (RL)–driven framework for Information Bottleneck (IB)–based distributed Joint Source–Channel Coding (JSCC) that operates reliably when the forward channels are unknown, non-differentiable, stochastic, or entirely black-box. Current state-of-the-art deep variational IB methods require differentiable end-to-end models and full knowledge of the forward channels' statistics, which renders them ineffective in many practical scenarios involving hidden channel states, human-in-the-loop setups, or proprietary simulators. To overcome these limitations, we reformulate the IB-based JSCC design problem as a sequential decision-making task. This enables the use of deep Multi-Agent Reinforcement Learning (MARL). We first revisit the single-terminal setup and show that the encoder can be treated as a policy in a contextual bandit whose sampled reward is set as the decoder reconstruction term, while the compression penalty is applied directly as an analytic actor regularizer derived from a variational surrogate of the IB objective. This yields a model-free compressor that learns to preserve relevance while respecting the rate constraint, without requiring gradients through the channel. We then extend the framework to the multiterminal setting, where multiple encoders observe noisy versions of a common source and must coordinate implicitly through the environment. Two retrieval strategies are considered: a parallel scheme, which ignores the side-information at the decoder, and a successive scheme, which exploits the side-information at the decoder via conditional priors. For both scenarios, we derive RL-compatible variational lower-bounds on the original IB objectives, enabling Centralized Training with Decentralized Execution (CTDE). By this, we generalize state-of-the-art distributed data-driven IB-based JSCC schemes to arbitrary forward channels while retaining the scalability and sample efficiency. As the main highlight, this work demonstrates that MARL provides a principled foundation for learning distributed compressors in environments where model- or gradient-based approaches are fundamentally inapplicable.

INDEX TERMS 6G, deep multi-agent reinforcement learning, information bottleneck, joint source-channel coding, centralized training with decentralized execution, black-box channels, distributed compression.

I. INTRODUCTION

THE original formulation of the Information Bottleneck (IB) method [1] was rooted in Shannon's seminal work on lossy source coding [2]. To quantify the fundamental limits of the inherent "complexity–precision" trade-off, the Rate–Distortion (RD) function was introduced as a central concept. The IB principle applies an intuitive twist to the single-letter characterization of this RD function by lower-bounding a mutual information term with respect to a target (relevant) variable, rather than upper-bounding an expected distortion term. This modification is particularly compelling because, in many real-world compression tasks, identifying the variable whose information must be preserved is far more natural than specifying an appropriate distortion function. With the specific choice of logarithmic loss distortion [3], it was later shown (see, e.g., [4]) that solving the IB constrained optimization problem yields the boundary of the achievable rate–distortion region for a *remote / indirect* source coding problem [5]–[9]. For a broader perspective on this variational principle from both information-theoretic and learning-theoretic viewpoints, the interested readers are referred to [10]–[12].

The IB method is connected to several classical problems, including the Wyner–Ahlswede–Körner problem [13], [14], the efficiency of investment information [15], the privacy funnel [16]–[18], and the distributed functional compression [19]–[22]. Beyond its theoretical significance, the IB method has found practical relevance in modern communication systems. Several applications include the construction of polar codes [23], [24], discrete channel decoding [25]–[27], analog-to-digital conversion for receiver front-ends [28], and semantic or task-oriented communication schemes [29]–[31].

Distributed extensions of the IB principle have also been studied extensively in the relevant literature (see, e.g., [32]–[38]). These works typically consider scenarios in which several noisy observations of a common source must be compressed—potentially at different rates—before being forwarded to a remote processing unit through multiple rate-limited channels. Depending on whether the forward links are assumed error-free or error-prone, the resulting design problems correspond to remote source coding or Joint Source–Channel Coding (JSCC), respectively. An overview of recent developments in both the theory and applications of the IB method can be found in [39].

The recent rise of machine learning-based approaches has brought a fresh perspective to IB-based compression. Leveraging the representational power of Deep Neural Networks (DNNs), several works have proposed data-driven architectures to efficiently optimize the IB trade-off [40]–[42]. These methods operate on finite sample sets and thus avoid the need for full prior knowledge of the joint statistics of the input signals. Moreover, they scale naturally to high-dimensional and continuous data, making them attractive alternatives to classical model-based algorithms in many practical settings.

Motivated by these developments, a deep variational approach to the IB-based JSCC has recently been introduced in [43], demonstrating that generative latent-variable models can effectively address both point-to-point and distributed scenarios when the forward channels can be modeled explicitly and incorporated into a differentiable end-to-end pipeline. However, this reliance on parametric channel knowledge and differentiability limits the applicability of such methods in many practical settings where the forward links are unknown, non-differentiable, stochastic, or entirely black-box. These situations arise, for example, in systems with hidden channel states, human-in-the-loop setups, or proprietary simulators that do not expose gradients or transition probabilities. This gap motivates the present work, in which we move beyond the variational paradigm and reformulate the IB-based JSCC as a sequential decision-making problem, enabling the use of deep Multi-Agent Reinforcement Learning (MARL) to learn distributed compressors directly from interaction with arbitrary forward channels.

A. CONTRIBUTIONS

The principal contributions of this work are summarized as follows:

1) **RL formulation of IB-based JSCC.** We reformulate the design problem of IB-based JSCC as a sequential decision-making task. This enables learning directly from interactions with arbitrary forward channels, without requiring differentiability or explicit channel models.
2) **Single-terminal deep RL-IB compressor.** For the point-to-point setting, we show that the encoder can be viewed as a policy in a contextual bandit whose sampled reward is the decoder reconstruction term, while the compression penalty is enforced through a direct actor regularizer derived from the same IB surrogate. This yields a model-free compressor that preserves relevance under rate constraints while operating over unknown or black-box channels.
3) **Distributed extension via deep MARL.** We extend the RL-IB formulation to multiterminal scenarios in which multiple encoders observe noisy versions of a common source. By leveraging deep MARL, we construct a distributed IB-based JSCC framework supporting both parallel and successive retrieval strategies.
4) **RL-compatible lower-bounds for distributed IB.** For both retrieval schemes, we derive learning-friendly lower-bounds on the relevance term together with local or conditional complexity penalties. These bounds enable stable optimization under the Centralized Training with Decentralized Execution (CTDE) paradigm.
5) **Generalization beyond variational IB.** The proposed MARL-driven approach removes the differentiability and model-knowledge assumptions inherent to variational IB methods, thereby extending state-of-the-art data-driven JSCC schemes to unknown, stochastic, or non-differentiable forward channels.
6) **Unified perspective on learning distributed compressors.** Overall, this work establishes RL—and in particular MARL—as a principled and scalable foundation for learning distributed IB-based compressors in environments where model-based or gradient-based approaches are fundamentally inapplicable.

To further underscore the generality of the considered setup here, we note that similar distributed compression and communication architectures arise across a wide range of contemporary wireless systems envisioned for 5G and 6G networks. Representative examples include distributed inference over sensor networks with imperfect links to a fusion center [44], Cloud-based Radio Access Networks (Cloud-RANs) with rate-limited and error-prone fronthaul channels [45], [46], (user-centric) Cell-Free massive Multiple-Input Multiple-Output (CF-mMIMO) architectures operating under constrained fronthaul channels [47]–[49], and cooperative relaying schemes based on "quantize-and-forward" strategy [50], [51]. These scenarios all involve multiple terminals observing correlated signals and forwarding over constrained channels, making them a natural application domain for the framework developed in this article.

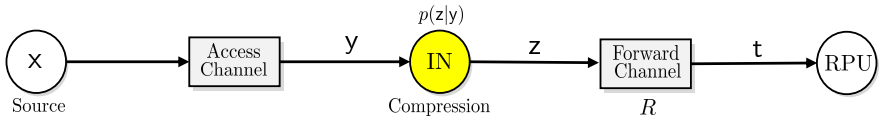


FIGURE 1. **The system model for point-to-point joint source-channel coding. IN, and RPU stand for Intermediate Node, and Remote Processing Unit, respectively. The Markov chain $\mathsf{x} \leftrightarrow \mathsf{y} \leftrightarrow \mathsf{z} \leftrightarrow \mathsf{t}$ applies.**

B. OUTLINE

The remainder of the paper is organized as follows. Section II focuses on the point-to-point setting and revisits the classical IB formulation for JSCC. We introduce the system model, formalize the underlying optimization problem, and develop its RL reinterpretation, which leads to a contextual-bandit framework for model-free encoder design over unknown, non-differentiable, stochastic, or entirely black-box channels. Section III extends these ideas to the multiterminal scenario. After presenting the distributed system model and problem formulation, we construct a MARL framework that enables the coordinated encoders' design under the CTDE paradigm. There, both the parallel and successive retrieval strategies are examined, together with their associated learning-compatible bounds. Section IV provides numerical simulation results that corroborate the effectiveness of the proposed methods on an exemplary stochastic forward link with hidden-state dynamics. Finally, Section V concisely summarizes the key contributions of this work.

C. NOTATIONS

We use lowercase letters $a \in \mathcal{A}$ to denote realizations of a (discrete) random variable a, whose distributions are written as p, q, or r, depending on the context. The same conventions apply to random vectors: for $\mathbf{a}_{1:J} = \{\mathsf{a}_1, \ldots, \mathsf{a}_J\}$, we use boldface symbols for both the vector and its realizations. For convenience, we write $\mathbf{a}_{1:J}^{-j} = \mathbf{a}_{1:J} \setminus \{\mathsf{a}_j\}$ to denote the collection of all components except the jth one. The operator $\mathbb{E}_{\bullet}\{\cdot\}$ denotes expectation w.r.t. the distribution indicated in the subscript. We use $D_{\mathrm{KL}}(\cdot\|\cdot)$ for the Kullback-Leibler (KL) divergence, $H(\cdot)$ for Shannon entropy, and $I(\cdot\,;\cdot)$ for mutual information [52]. Finally, the notation $\{\cdot\}_{j=1}^{J}$ represents a set of J elements indexed by j.

II. IB-BASED JSCC: THE POINT-TO-POINT SETTING

We begin our technical discussion by examining the point-to-point setting, which serves as the foundational building block for the distributed framework developed later in the article. In this section, we introduce the classical IB-based JSCC formulation together with the corresponding system model and optimization objective. We then show how this formulation can be reinterpreted through the lens of RL, paving the way for a model-free encoder design that operates directly over unknown forward channels. The resulting single-terminal perspective provides the conceptual and methodological groundwork upon which the multiterminal extensions of Section III are constructed.

A. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the general point-to-point setup depicted in Fig. 1. An Intermediate Node (IN) observes a noisy version y of an underlying source signal x and must compress this observation into a representation z prior to transmission over a *rate-limited* and *error-prone* forward channel. The Remote Processing Unit (RPU) receives a corrupted version t of this representation. At this level of generality, no structural assumptions are imposed on either the access channel or the forward channel; both may be stochastic, time-varying, or otherwise difficult to model analytically.

The encoder design problem can then be posed directly in terms of the mutual-information trade-off central to the IB framework. In particular, the goal is to maximize the relevant information $I(\mathsf{x};\mathsf{t})$ subject to the compression-rate constraint $I(\mathsf{y};\mathsf{z}) \le R$, where R specifies the rate-limit of the forward channel. This leads to the optimization problem

$$p^*(\mathsf{z}|\mathsf{y}) = \underset{p(\mathsf{z}|\mathsf{y}):\, I(\mathsf{y};\mathsf{z}) \le R}{\operatorname{argmax}} I(\mathsf{x};\mathsf{t}). \tag{1}$$

Using the method of Lagrange multipliers [53], this design problem is reformulated as

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

where $\lambda \ge 0$ is associated with the rate-limit R and may be determined, e.g., via a bisection search over a finite interval.

Closed-form model-based solution

Classical IB-based JSCC formulations impose additional structure by assuming that both the access and forward links are Discrete Memoryless Channels (DMCs) with known transition probabilities. Under these assumptions, the joint distribution $p(\mathsf{x},\mathsf{y},\mathsf{z},\mathsf{t})$ is fully specified through the known source distribution $p(\mathsf{x})$, access-channel transition probabilities $p(\mathsf{y}|\mathsf{x})$, and forward-channel transition probabilities $p(\mathsf{t}|\mathsf{z})$, together with the chosen encoder mapping $p(\mathsf{z}|\mathsf{y})$.

The stationary solution of (2) was derived in closed form in [54] by means of variational calculus. The resulting expression

$$p(z|y) = \frac{p(z)}{\omega(y,\beta)} \exp\left(-\beta \sum_{t\in\mathcal{T}} p(t|z)\, D_{\mathrm{KL}}\big(p(\mathsf{x}|y) \,\|\, p(\mathsf{x}|t)\big)\right), \tag{3}$$

with $\beta = 1/\lambda$ and $\omega(y,\beta)$ denoting a normalization function to ensure the validity of the encoder mapping, forms the basis of the *Forward-Aware Vector Information Bottleneck (FAVIB)* algorithm, which iteratively computes a fixed point of (3) and is guaranteed to converge to a stationary point of the IB Lagrangian [54].

Variational, data-driven solution

A data-driven extension of this model-based approach was later introduced in [43], where a deep variational architecture was proposed to approximate the IB objective using latent-variable models. This *Deep FAVIB* framework enables learning from finite samples $\{(x^{(i)}, y^{(i)})\}_{i=1}^{M}$ and does not require explicit knowledge of the access-channel statistics. However, it still relies on the restrictive presumption that the forward link is a known DMC, since its transition probabilities should be incorporated into the end-to-end differentiable training pipeline obtained by applying the *Gumbel-Softmax / Concrete Distribution* reparametrization trick [55], [56]. When the forward link is unknown, non-differentiable, or black-box, Deep FAVIB cannot evaluate its objective function, cannot propagate gradients, and therefore cannot be applied.

Toward a model-free RL solution

The present work moves beyond this limitation by considering forward channels that are not only unknown but may also be non-differentiable, stochastic, or entirely black-box. In such settings, neither the closed-form solution in (3) nor variational methods that rely on differentiable channel models remain applicable. This naturally motivates a fundamentally different perspective: we reinterpret the IB-based JSCC problem as a sequential decision-making task and develop an RL framework that learns the encoder directly through interaction with the forward channel, without requiring any parametric model or gradient information.

B. RL REINTERPRETATION

We now develop a model-free RL reinterpretation of the IB-based point-to-point JSCC design problem. Contrary to variational approaches, which rely on differentiable channel models and explicit likelihood expressions, the RL viewpoint enables learning directly from interaction with an unknown forward channel. To construct such a framework, we derive a variational surrogate of the IB Lagrangian in (2), whose relevance term supplies the sampled reward and whose compression term is handled as a direct actor regularizer.

Lower-bound on the relevance term

For the relevance $I(\mathsf{x};\mathsf{t})$, we introduce a decoder distribution $q(\mathsf{x}|\mathsf{t})$ and apply the standard variational lower-bound:

$$I(\mathsf{x};\mathsf{t}) = \underbrace{H(\mathsf{x})}_{\geq 0} - H(\mathsf{x}|\mathsf{t}) \tag{4a}$$

$$\geq \underbrace{\sum_{t\in\mathcal{T}} p(t) D_{\mathrm{KL}}\big(p(\mathsf{x}|t)\|q(\mathsf{x}|t)\big)}_{\geq 0} + \sum_{x\in\mathcal{X},\, t\in\mathcal{T}} p(x,t)\log q(x|t) \tag{4b}$$

$$\geq \mathbb{E}_{\mathsf{x},\mathsf{t}}\{\log q(\mathsf{x}|\mathsf{t})\}, \tag{4c}$$

wherein, from (4a) to (4b), the non-negativity of entropy (for the discrete source signal x), and, from (4b) to (4c), the non-negativity of KL divergence, also known as the *information inequality*, has been applied [52]. Therefore, the log-likelihood $\log q(x|t)$ provides a sample-based surrogate for the relevance term.

Upper-bound on the compression term

For the compression rate $I(\mathsf{y};\mathsf{z})$, we introduce a reference prior $r(\mathsf{z})$ and apply the standard variational upper-bound:

$$I(\mathsf{y};\mathsf{z}) = \sum_{y\in\mathcal{Y},\, z\in\mathcal{Z}} p(y,z)\log\frac{p(z|y)}{r(z)} - \underbrace{D_{\mathrm{KL}}\big(p(\mathsf{z})\|r(\mathsf{z})\big)}_{\geq 0} \tag{5a}$$

$$\leq \mathbb{E}_{\mathsf{y},\mathsf{z}}\left\{\log\frac{p(\mathsf{z}|\mathsf{y})}{r(\mathsf{z})}\right\}. \tag{5b}$$

where the inequality follows from the non-negativity of $D_{\mathrm{KL}}\big(p(\mathsf{z})\|r(\mathsf{z})\big)$. Hence, the log-ratio $\log\frac{p(z|y)}{r(z)}$ provides an upper-bound surrogate for the compression cost.

Utility decomposition

Combining the lower-bound on relevance and the upper-bound on compression rate yields a tractable surrogate for the IB Lagrangian, leading to the following scalar utility function:

$$\mathsf{u}_{\mathrm{P2P}}(x,y,z,t) = \log q(x|t) - \lambda\log\frac{p(z|y)}{r(z)}, \tag{6}$$

where the first term encourages informative reconstructions and is used as the sampled reward, whereas the second penalizes high-rate encodings and is applied directly in the encoder update. So, we keep this scalar utility, but interpret it as the sum of a sampled reconstruction reward and an explicit actor-side compression regularizer. Importantly, both terms are computable from samples (x,y,z,t) without requiring knowledge of the forward-channel transition law.

Expected utility as the RL objective

Under the encoder policy $p(\mathsf{z}|\mathsf{y})$, the expected utility is

$$\mathcal{J}_{\mathrm{P2P}}(p) = \mathbb{E}_{\mathsf{x},\mathsf{y},\mathsf{z},\mathsf{t}}\big\{\mathsf{u}_{\mathrm{P2P}}(\mathsf{x},\mathsf{y},\mathsf{z},\mathsf{t})\big\}, \tag{7}$$

where the expectation is taken over the empirical source-observation distribution, the encoder policy, and the forward-channel dynamics. Maximizing $\mathcal{J}_{\mathrm{P2P}}(p)$ therefore maximizes a sample-based lower-bound on the relevance term while minimizing a sample-based upper-bound on the compression term—precisely mirroring the structure of the original IB objective for point-to-point JSCC.

RL interpretation

This construction provides a principled RL decomposition: the encoder acts as a stochastic policy, the forward channel serves as the environment, the decoder log-likelihood in (6) supplies the sampled reward used for return estimation, and the compression term is injected directly into the actor objective as an analytic regularizer. Although the interaction consists of a single decision per episode, contextual bandits are formally part of the sequential decision-making family in RL. Crucially, this approach remains valid even when the forward channel is non-differentiable, stochastic, or entirely unknown, thereby enabling a fully model-free learning procedure for IB-based JSCC design problem.

C. MODEL-FREE TRAINING VIA PPO

We now operationalize the RL reinterpretation by developing a model-free training based on *Proximal Policy Optimization (PPO)*, a state-of-the-art policy-gradient method known for its stability and sample efficiency (in the standard RL sense of effective learning from interactions) [57]. We presume access to a finite dataset $\mathcal{D}=\{(x^{(i)},y^{(i)})\}_{i=1}^{M}$ generated by the access link, while the forward link is treated as a black-box environment, returning samples $t\sim p(\mathsf{t}|\mathsf{z})$ upon query.

Parameterization

We consider the following parameterized components:

- **Encoder (policy):** $p_{\boldsymbol{\theta}}(\mathsf{z}|\mathsf{y})$, a stochastic policy that is implemented by a DNN with parameters $\boldsymbol{\theta}$.
- **Decoder:** $q_{\boldsymbol{\phi}}(\mathsf{x}|\mathsf{t})$, a DNN that reconstructs x from t.
- **Latent prior:** $r_{\boldsymbol{\psi}}(\mathsf{z})$, a parametric prior over latent representations.
- **Value function:** $V_{\boldsymbol{\zeta}}(\mathsf{y})$, a baseline, which estimates the expected sampled reconstruction return under the encoder's observation y.

Given these, the sampled reward used for trajectory evaluation and critic learning is

$$\mathsf{r}_{\text{P2P}}(x,t)=\log q_{\boldsymbol{\phi}}(x|t)\,. \tag{8}$$

The compression term is handled separately inside the encoder update through

$$\mathsf{c}_{\text{P2P}}(y,z)=\log\frac{p_{\boldsymbol{\theta}}(z|y)}{r_{\boldsymbol{\psi}}(z)}. \tag{9}$$

Trajectory generation

Training proceeds by repeatedly sampling mini-batches from $\mathcal{D}$ and interacting with the forward channel:

1) Sample a mini-batch $\{(x^{(i)},y^{(i)})\}_{i=1}^{B}$.
2) For each $y^{(i)}$, sample an action $z^{(i)}\sim p_{\boldsymbol{\theta}}(\mathsf{z}|y^{(i)})$.
3) Query the forward channel to obtain $t^{(i)}\sim p(\mathsf{t}|z^{(i)})$.
4) Compute sampled rewards $\mathsf{r}_{\text{P2P}}^{(i)}$ using (8).
5) Compute value estimates $V_{\boldsymbol{\zeta}}(y^{(i)})$.

These tuples $(x^{(i)},y^{(i)},z^{(i)},t^{(i)},\mathsf{r}_{\text{P2P}}^{(i)})$ constitute a batch of experience for PPO, while the sampled pairs $(y^{(i)},z^{(i)})$ also determine the actor-side compression regularizer through (9).

Advantage estimation

The advantage function is obtained from the sampled reconstruction reward through a one-step return approximation:

$$A_{\text{P2P}}^{(i)}=\mathsf{r}_{\text{P2P}}^{(i)}-V_{\boldsymbol{\zeta}}(y^{(i)}). \tag{10}$$

More sophisticated estimators (e.g., GAE [58]) are possible, but the one-step estimate suffices in this work. This choice is not an implementation preference; it follows directly from the contextual-bandit structure here, whose horizon is one.

PPO clipped objective

Let

$$\rho^{(i)}=\frac{p_{\boldsymbol{\theta}}(z^{(i)}|y^{(i)})}{p_{\boldsymbol{\theta}_{\text{old}}}(z^{(i)}|y^{(i)})} \tag{11}$$

denote the likelihood ratio between the current and previous policies. The PPO term for the encoder is

$$\mathcal{L}_{\text{P2P}}^{\text{PPO}}(\boldsymbol{\theta})=\frac{1}{B}\sum_{i=1}^{B}\min\Big(\rho^{(i)}A_{\text{P2P}}^{(i)},\,\text{clip}(\rho^{(i)},1-\epsilon,1+\epsilon)\,A_{\text{P2P}}^{(i)}\Big), \tag{12}$$

where $\epsilon>0$ is the clipping parameter. This term prevents large, destabilizing policy updates while still encouraging improvement. To enforce the IB trade-off directly at the actor, we augment it with the analytic compression regularizer and optimize

$$\mathcal{L}_{\text{P2P}}^{\text{Act}}(\boldsymbol{\theta})=\mathcal{L}_{\text{P2P}}^{\text{PPO}}(\boldsymbol{\theta})-\frac{\lambda}{B}\sum_{i=1}^{B}\mathsf{c}_{\text{P2P}}(y^{(i)},z^{(i)}). \tag{13}$$

Decoder, prior, and value-function updates

The decoder and prior parameters optimize the two terms of the IB surrogate, but through separate updates: the decoder maximizes the sampled reconstruction reward, while the prior tightens the reference term appearing in the actor regularizer. For a mini-batch, the corresponding sample-based gradient estimates are

$$\mathbf{g}_{\boldsymbol{\phi}}=\frac{1}{B}\sum_{i=1}^{B}\nabla_{\boldsymbol{\phi}}\log q_{\boldsymbol{\phi}}(x^{(i)}|t^{(i)}), \tag{14}$$

$$\mathbf{g}_{\boldsymbol{\psi}}=\frac{1}{B}\sum_{i=1}^{B}\lambda\,\nabla_{\boldsymbol{\psi}}\log r_{\boldsymbol{\psi}}(z^{(i)}). \tag{15}$$

The value function is trained by minimizing the squared error

$$\mathcal{L}_{\text{P2P}}^{\text{VF}}(\boldsymbol{\zeta})=\frac{1}{B}\sum_{i=1}^{B}\big(V_{\boldsymbol{\zeta}}(y^{(i)})-\mathsf{r}_{\text{P2P}}^{(i)}\big)^2. \tag{16}$$

Parameter updates

The parameters are updated via gradient ascent/descent:

$$\boldsymbol{\theta}\leftarrow\boldsymbol{\theta}+\eta_\theta\,\nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{P2P}}^{\text{Act}}, \tag{17}$$

Algorithm 1 Deep RL-FAVIB

Initialize parameters $\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\zeta}$.
repeat
  Sample a mini-batch $\{(x^{(i)},y^{(i)})\}_{i=1}^{B}$.
  for $i=1,\dots,B$ **do**
    Sample action $z^{(i)}\sim p_{\boldsymbol{\theta}}(\mathsf{z}|y^{(i)})$.
    Query forward channel: $t^{(i)}\sim p(\mathsf{t}|z^{(i)})$.
    Compute sampled reward $\mathsf{r}_{\text{P2P}}^{(i)}$ using (8).
    Evaluate value estimate $V_{\boldsymbol{\zeta}}(y^{(i)})$.
    Compute advantage $A_{\text{P2P}}^{(i)}$ using (10).
  end for
  Compute actor objective $\mathcal{L}_{\text{P2P}}^{\text{Act}}$ using (13).
  Update encoder parameters $\boldsymbol{\theta}$ via actor-regularized PPO.
  Update decoder parameters $\boldsymbol{\phi}$ using $\mathbf{g}_{\boldsymbol{\phi}}$.
  Update prior parameters $\boldsymbol{\psi}$ using $\mathbf{g}_{\boldsymbol{\psi}}$.
  Update value-function parameters $\boldsymbol{\zeta}$ by minimizing $\mathcal{L}_{\text{P2P}}^{\text{VF}}$.
until convergence

$$\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} + \eta_{\phi}\, \mathbf{g}_{\boldsymbol{\phi}}, \tag{18}$$

$$\boldsymbol{\psi} \leftarrow \boldsymbol{\psi} + \eta_{\psi}\, \mathbf{g}_{\boldsymbol{\psi}}, \tag{19}$$

$$\boldsymbol{\zeta} \leftarrow \boldsymbol{\zeta} - \eta_{\zeta}\, \nabla_{\boldsymbol{\zeta}} \mathcal{L}^{\mathrm{VF}}_{\mathrm{P2P}}. \tag{20}$$

Algorithm summary

The complete training procedure is summarized in Alg. 1. This PPO-based training algorithm provides a stable and fully model-free mechanism for optimizing the IB-based JSCC encoder. It leverages clipped policy updates driven by reconstruction-based advantages, together with an explicit actor-side rate regularizer and a learned value baseline, to ensure robust learning even when the forward channel is non-differentiable, stochastic, or entirely unknown. We refer to this overall learning framework as *Deep RL-FAVIB*.

III. IB-BASED JSCC: MULTITERMINAL EXTENSIONS

In this section, we extend the IB-based JSCC framework from the point-to-point setting to a general multiterminal architecture with distributed observations and multiple error-prone forward channels. We begin by formalizing the system model. Building on this model, first we revisit the classical DMC-based solutions [38], which provide analytical encoder designs under fully known channel transition probabilities. We then discuss the variational, data-driven extensions introduced in [43], which replace the mutual-information terms with tractable bounds and enable learning over an end-to-end differentiable training pipeline. Finally, motivated by the limitations of both classical and variational approaches, we develop a model-free MARL reinterpretation of the multiterminal IB-based JSCC problems, paving the way for a fully interaction-driven solution that remains applicable even when the forward channels are non-differentiable, stochastic, or entirely unknown.

A. SYSTEM MODEL AND PROBLEM FORMULATION

We consider the multiterminal IB-based JSCC architecture illustrated in Fig. 2, consisting of J distributed branches. Each branch j observes a random variable y_j generated from an underlying source x through an access channel. The branches operate without inter-encoder communication and produce compact representations z_j through encoder distributions $p(\mathsf{z}_j|\mathsf{y}_j)$. These representations are then transmitted over J independent *rate-limited* and *error-prone* forward channels. The RPU receives the forward channels' outputs $\mathbf{t}_{1:J} \triangleq \{\mathsf{t}_1, \ldots, \mathsf{t}_J\}$. At this level of generality, no structural assumptions are imposed on either the access channels or the forward channels; both may be stochastic, time-varying, or otherwise difficult to model analytically. This generic architecture captures a broad class of multiterminal communication problems, including distributed sensing, multi-view compression, and cooperative JSCC.

Within this general architecture, the manner in which decoding is performed at the RPU gives rise to two distinct retrieval strategies, each inducing a different optimization problem. In the *parallel retrieval* setting, all forward-channel outputs are processed jointly in parallel, and no sequential side-information from other branches is exploited. In contrast, in the *successive retrieval* setting, the RPU leverages the side-information obtained from previous branches, in a manner reminiscent of the Wyner–Ziv paradigm for source coding [59], where statistically correlated signals serve as decoder side-information. These two distinct modes correspond to fundamentally different information-processing structures and therefore require separate formulations of the underlying IB-based JSCC design problem.

1) Parallel Retrieval

In the parallel retrieval setting, all forward-link outputs $\mathbf{t}_{1:J}$ are processed jointly at the RPU, and no sequential side-information is exploited. The design objective is therefore to maximize the relevance of the aggregated channel outputs for estimating x, while constraining the individual compression rates of the local encoders. This leads to the constrained optimization problem

$$P^* \triangleq \{p^*(\mathsf{z}_j|\mathsf{y}_j)\}_{j=1}^{J} = \underset{P:\, \forall j\, I(\mathsf{y}_j;\mathsf{z}_j)\le R_j}{\arg\max} I(\mathsf{x};\mathbf{t}_{1:J}), \tag{21}$$

where R_j denotes the rate-limit of the forward channel in branch j. Introducing the Lagrange multipliers $\{\lambda_j \ge 0\}_{j=1}^{J}$, this problem can be recast into the following unconstrained maximization (up to the validity of the encoder mappings)

$$P^* = \underset{P}{\operatorname{argmax}}\; I(\mathsf{x};\mathbf{t}_{1:J}) - \sum_{j=1}^{J} \lambda_j I(\mathsf{y}_j;\mathsf{z}_j). \tag{22}$$

Closed-form model-based solution

In [38], a closed-form stationary solution to (22) was derived under the additional assumptions that (i) each access channel is a DMC with known transition probabilities $p(\mathsf{y}_j|\mathsf{x})$, and (ii) each forward link is also a DMC with known transition probabilities $p(\mathsf{t}_j|\mathsf{z}_j)$. Under these conditions, the stationary solution of (22) for each local encoder has been derived as

$$p(z_j|y_j) = \frac{p(z_j)}{\omega^{\mathrm{Par}}_{\mathsf{z}_j}(y_j,\beta_j)} \exp\Big(-d_{\mathrm{Par}}(y_j,z_j)\Big), \tag{23}$$

where $\omega^{\mathrm{Par}}_{\mathsf{z}_j}(y_j,\beta_j)$ is a normalization function ensuring the validity of the pertinent encoder mapping, and $d_{\mathrm{Par}}(y_j,z_j)$ is a distortion term that depends explicitly on the statistics of the forward links through

$$d_{\mathrm{Par}}(y_j,z_j) = \beta_j \sum_{\boldsymbol{z}^{-j}_{1:J}} p(\boldsymbol{z}^{-j}_{1:J}|y_j) \times \sum_{\boldsymbol{t}_{1:J}} p(\boldsymbol{t}_{1:J}|\boldsymbol{z}_{1:J}) D_{\mathrm{KL}}\big(p(\mathsf{x}|y_j,\boldsymbol{z}^{-j}_{1:J}) \| p(\mathsf{x}|\boldsymbol{t}_{1:J})\big), \tag{24}$$

with $\beta_j = \frac{1}{\lambda_j}$, $p(\boldsymbol{t}_{1:J}|\boldsymbol{z}_{1:J}) = \prod_{j=1}^{J} p(t_j|z_j)$. The resulting *Multivariate Forward-Aware Vector Information Bottleneck (MFAVIB-Parallel)* algorithm [38] iteratively applies the fixed-point updates in (23) (going through all $j=1$ to J) and is guaranteed to converge to a stationary point of the IB Lagrangian in (22). This closed-form solution that lies at the heart of the MFAVIB-Parallel, however, critically relies on prior knowledge of all forward channels' transition laws.

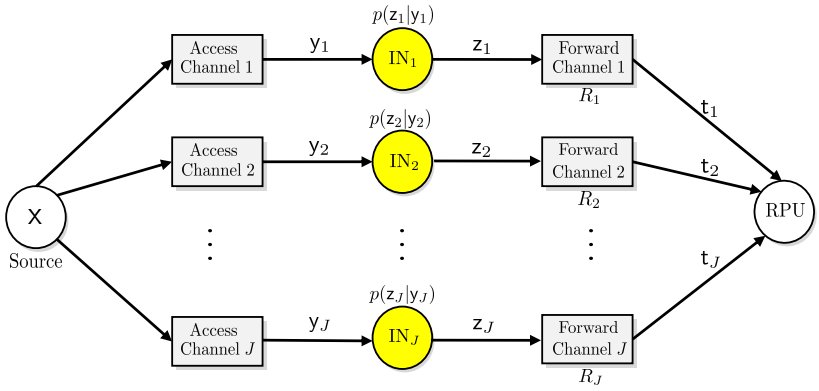


FIGURE 2. **The considered system model for distributed joint source-channel coding. The Markov chain $\mathsf{x} \leftrightarrow \mathsf{y}_j \leftrightarrow \mathsf{z}_j \leftrightarrow \mathsf{t}_j$ applies per branch j. Moreover, given the source x, the counterpart signals of different branches are presumed to be statistically independent.**

Variational, data-driven solution

To overcome the reliance of the model-based approach on full prior knowledge of the joint input statistics $p(\mathsf{x}, \mathbf{y}_{1:J})$, the variational multiterminal IB framework introduced in [43] adopts a data-driven solution. There, a deep variational architecture approximates the IB objective through latent-variable models, enabling learning directly from finite samples $\{(x^{(i)}, \boldsymbol{y}_{1:J}^{(i)})\}_{i=1}^{M}$ without requiring explicit access-channel distributions. This *Deep MFAVIB-Parallel* methodology thus removes the need for model-based statistical knowledge and supports an end-to-end differentiable training pipeline. However, it still presumes that each forward channel is a known DMC, since its transition probabilities must be embedded into the differentiable computation graph. As in the point-to-point setting, this is achieved by applying the *Gumbel–Softmax / Concrete Distribution* reparameterization trick [55], [56] for the underlying categorical distributions. When the forward links are unknown, non-differentiable, or black-box, Deep MFAVIB-Parallel cannot evaluate its objective function, cannot propagate gradients, and therefore cannot be applied.

Toward a model-free MARL solution

Despite their flexibility, variational methods still require differentiability of the forward channels or reparameterizable latent variables. In scenarios where the forward channels are non-differentiable, stochastic, or entirely unknown, these assumptions may fail. Motivated by this, in Subsection III.B.1 we develop a fully model-free MARL reinterpretation of the parallel retrieval problem, enabling encoders to learn directly from interaction with the environment without requiring any knowledge of the forward-link transition laws.

2) Successive Retrieval

In the successive retrieval setting, the RPU processes the forward-channel outputs in a prescribed order and leverages the side-information obtained from previous branches. This structure follows the Wyner–Ziv principle for source coding [59], where statistically correlated signals are exploited as decoder side-information. As a result, the compression constraint for branch j becomes *conditional*, reflecting the fact that the RPU has already observed $\mathbf{t}_{1:j-1}$ when processing t_j. The corresponding design problem is

$$P^* = \underset{P:\, \forall j\, I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{t}_{1:j-1}) \le R_j}{\arg\max} I(\mathsf{x};\mathbf{t}_{1:J}), \tag{25}$$

where R_j denotes the rate-limit of the forward channel in branch j. As in the classical successive-refinement literature, the processing order generally affects performance and must be optimized (e.g., via a brute-force search). By introducing a set of Lagrange multipliers $\{\lambda_j \ge 0\}_{j=1}^{J}$, the constrained design problem in (25) can be rewritten as the following unconstrained maximization

$$P^* = \arg\max_{P} I(\mathsf{x};\mathbf{t}_{1:J}) - \sum_{j=1}^{J} \lambda_j\, I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{t}_{1:j-1}), \tag{26}$$

up to the validity of the encoder mappings. Note that the major difference between the successive and parallel retrieval shows itself in the consideration of side-information at the compression rates to further leverage the correlations in the output signals of local encoders.

Closed-form model-based solution

In [38], a closed-form stationary solution to (26) was derived under the same model-based assumptions as in the parallel case: (i) each access channel is a DMC with known transition probabilities $p(\mathsf{y}_j|\mathsf{x})$, and (ii) each forward channel is also a DMC with known transition probabilities $p(\mathsf{t}_j|\mathsf{z}_j)$. Under these assumptions, the stationary solution of (26) for each local encoder has been derived as

$$p(z_j|y_j) = \frac{p(z_j)}{\omega_{\mathsf{z}_j}^{\mathrm{Suc}}(y_j,\beta_j)} \exp\big(-d_{\mathrm{Suc}}(y_j,z_j)\big), \tag{27}$$

where $\omega_{\mathsf{z}_j}^{\mathrm{Suc}}(y_j,\beta_j)$ is a normalization function ensuring the validity of the pertinent encoder mapping, and $d_{\mathrm{Suc}}(y_j,z_j)$ is a distortion term that extends its parallel counterpart by

incorporating the effect of side-information through

$$d_{\text{Suc}}(y_j, z_j) = \beta_j \sum_{\boldsymbol{z}_{1:J}^{-j}} p(\boldsymbol{z}_{1:J}^{-j}|y_j)\times \sum_{\boldsymbol{t}_{1:J}} p(\boldsymbol{t}_{1:J}|\boldsymbol{z}_{1:J}) D_{\text{KL}}\big(p(\mathsf{x}|y_j, \boldsymbol{z}_{1:J}^{-j})\|p(\mathsf{x}|\boldsymbol{t}_{1:J})\big) - \sum_{\boldsymbol{t}_{1:j-1}} p(\boldsymbol{t}_{1:j-1}|y_j)\log p(\boldsymbol{t}_{1:j-1}|z_j) - \beta_j \sum_{k=j+1}^{J} \beta_k^{-1} \sum_{\boldsymbol{t}_{1:k-1},\, z_k} p(t_j|z_j) p(\boldsymbol{t}_{1:k-1}^{-j}, z_k|y_j)\log p(z_k|\boldsymbol{t}_{1:k-1}), \tag{28}$$

with $\beta_j = \frac{1}{\lambda_j}$, $p(\boldsymbol{t}_{1:J}|\boldsymbol{z}_{1:J}) = \prod_{j=1}^{J} p(t_j|z_j)$. Specifically, the distortion here includes two additional terms arising from conditioning the compression rate on $\mathbf{t}_{1:j-1}$, reflecting the fact that the decoder already possesses partial information about x when processing branch j. The resulting *MFAVIB-Successive* algorithm [38] iteratively applies the fixed-point updates in (27) (going through all $j=1$ to J) and is guaranteed to converge to a stationary point of the IB Lagrangian in (26). As in the parallel case, this solution critically relies on full knowledge of all forward channels' transition laws.

Variational, data-driven solution

To eliminate the need for full prior knowledge of the joint input statistics $p(\mathsf{x}, \mathbf{y}_{1:J})$, the variational multiterminal IB framework of [43] extends naturally to the successive setting. The proposed deep variational architecture approximates the conditional IB objective in (26) using latent-variable models, enabling learning directly from finite samples $\{(x^{(i)}, \boldsymbol{y}_{1:J}^{(i)})\}_{i=1}^{M}$ without requiring explicit access-channel distributions. This *Deep MFAVIB-Successive* approach supports end-to-end differentiable training and captures the benefits of side-information in a data-driven manner. However, as in the parallel retrieval, it still presumes that each forward channel is a known DMC, since its transition probabilities must be embedded into the differentiable computation graph through the same reparameterization trick for the underlying categorical distributions. When the forward links are unknown, non-differentiable, or black-box, Deep MFAVIB-Successive cannot evaluate its objective function, cannot propagate gradients, and therefore cannot be applied.

Toward a model-free MARL solution

Analogous to the parallel retrieval, variational methods for successive retrieval still rely on differentiability of the forward channels or reparameterizable latent variables. These assumptions break down when the forward links are non-differentiable, stochastic, or entirely black-box. Motivated by this limitation, in Subsection III.B.2 we develop a completely model-free MARL reinterpretation of the successive retrieval problem as well, enabling encoders to learn directly from interaction with the environment while exploiting side-information at the RPU.

B. MARL REINTERPRETATION

Having established both the classical model-based and the variational data-driven solutions for multiterminal IB-based JSCC, we now turn to a fundamentally different perspective—the one that removes the need for any structural assumptions or differentiability requirements on the forward channels. The key idea is to reinterpret the multiterminal IB-based JSCC objective as a *cooperative* MARL problem, in which each encoder acts as an autonomous agent interacting with an unknown environment, while the decoder and latent priors define a shared training signal whose sampled reward is the joint reconstruction term and whose compression terms act as direct actor regularizers. This viewpoint enables learning directly from samples generated by the true forward channels, regardless of whether they are stochastic, non-differentiable, or analytically intractable. In what follows, we first develop the MARL formulation for the parallel retrieval setting, and then extend it to the successive retrieval as well.

1) Parallel Retrieval

In the parallel retrieval, all J encoders act simultaneously and the RPU processes the aggregated forward-link outputs $\mathbf{t}_{1:J}$ in a single stage. This induces a natural cooperative MARL structure: each encoder j behaves as an autonomous agent with policy $p(\mathsf{z}_j|\mathsf{y}_j)$, the collection of forward channels forms the environment, and the decoder together with the latent priors define a shared objective whose sampled reward is common across agents while the compression terms are enforced directly in the actors. This viewpoint enables learning directly from samples generated by the true forward channels, without requiring differentiability or explicit knowledge of their transition probabilities. To construct this MARL framework, we derive a sample-based utility by bounding the two information terms in the multiterminal IB Lagrangian for parallel retrieval in (22).

Lower-bound on the relevance term

For the mutual-information term $I(\mathsf{x}; \mathbf{t}_{1:J})$, we introduce a decoder distribution $q(\mathsf{x}|\mathbf{t}_{1:J})$ and apply the standard variational lower-bound:

$$I(\mathsf{x}; \mathbf{t}_{1:J}) = H(\mathsf{x}) - H(\mathsf{x}|\mathbf{t}_{1:J}) \geq \mathbb{E}_{\mathsf{x},\mathbf{t}_{1:J}}\big\{\log q(\mathsf{x}|\mathbf{t}_{1:J})\big\}, \tag{29}$$

where, similar to derivations in (4), the inequality follows from the non-negativity of $D_{\text{KL}}\big(p(\mathsf{x}|\boldsymbol{t}_{1:J})\|q(\mathsf{x}|\boldsymbol{t}_{1:J})\big)$. Hence, it is directly inferred that the log-likelihood $\log q(x|\boldsymbol{t}_{1:J})$ provides a sample-based surrogate for the relevance term.

Upper-bound on the compression terms

For each encoder j, the compression rate satisfies

$$I(\mathsf{y}_j; \mathsf{z}_j) = \mathbb{E}_{\mathsf{y}_j,\mathsf{z}_j}\left\{\log \frac{p(\mathsf{z}_j|\mathsf{y}_j)}{p(\mathsf{z}_j)}\right\}. \tag{30}$$

Introducing a reference prior $r(\mathsf{z}_j)$ yields the upper-bound

$$I(\mathsf{y}_j; \mathsf{z}_j) \leq \mathbb{E}_{\mathsf{y}_j,\mathsf{z}_j}\left\{\log \frac{p(\mathsf{z}_j|\mathsf{y}_j)}{r(\mathsf{z}_j)}\right\}, \tag{31}$$

where, similar to derivations in (5), the inequality follows from $D_{\mathrm{KL}}\big(p(\mathsf{z}_j)\|r(\mathsf{z}_j)\big)\geq 0$. Thus, each log-ratio $\log\frac{p(z_j|y_j)}{r(z_j)}$ provides a sample-based surrogate for the compression cost of branch j.

Utility decomposition

Combining the relevance lower-bound and the compression upper-bounds yields the multiterminal utility

$$\mathsf{u}_{\mathrm{Par}}(x,\boldsymbol{y}_{1:J},\boldsymbol{z}_{1:J},\boldsymbol{t}_{1:J})=\log q(x|\boldsymbol{t}_{1:J})-\sum_{j=1}^{J}\lambda_j\log\frac{p(z_j|y_j)}{r(z_j)}. \tag{32}$$

Its first term is a shared sampled reward, while the second term is a sum of local compression regularizers applied directly in the encoder updates. Importantly, the complete surrogate is computable from samples $(x,\boldsymbol{y}_{1:J},\boldsymbol{z}_{1:J},\boldsymbol{t}_{1:J})$ without requiring any knowledge of the forward-channel transition laws.

Expected utility as the MARL objective

Under the joint encoder policy

$$p(\mathbf{z}_{1:J}|\mathbf{y}_{1:J})=\prod_{j=1}^{J}p(\mathsf{z}_j|\mathsf{y}_j), \tag{33}$$

the expected utility is

$$\mathcal{J}_{\mathrm{Par}}(P)=\mathbb{E}_{\mathsf{x},\mathbf{y}_{1:J},\mathbf{z}_{1:J},\mathbf{t}_{1:J}}\big\{\mathsf{u}_{\mathrm{Par}}(\mathsf{x},\mathbf{y}_{1:J},\mathbf{z}_{1:J},\mathbf{t}_{1:J})\big\}. \tag{34}$$

Maximizing $\mathcal{J}_{\mathrm{Par}}(P)$ thus maximizes a sample-based lower-bound on relevance while minimizing a sample-based upper-bound on the sum of compression rates—precisely mirroring the structure of the original IB objective for multiterminal JSCC with parallel retrieval.

MARL interpretation

This construction yields a clean MARL interpretation:

- **Agents:** the J encoders, each with policy $p(\mathsf{z}_j|\mathsf{y}_j)$;
- **Environment:** the set of forward links $\{p(\mathsf{t}_j|\mathsf{z}_j)\}_{j=1}^{J}$;
- **State:** the observations $\mathbf{y}_{1:J}$;
- **Actions:** the compact representations $\mathbf{z}_{1:J}$;
- **Shared reward:** the reconstruction term $\log q(x|\boldsymbol{t}_{1:J})$ in (32);
- **Actor regularizer:** the local penalties $\log\frac{p(z_j|y_j)}{r(z_j)}$ aggregated in (32);
- **Transition:** the environment produces $\mathbf{t}_{1:J}$.

Since the reconstruction reward is shared and all agents are optimized against compatible actor-side compression regularizers, the agents cooperate to maximize a common objective—a standard CTDE MARL setting. Crucially, this formulation remains valid even when the forward channels are non-differentiable, stochastic, time-varying, or entirely unknown, enabling a fully model-free learning procedure for multiterminal IB-based JSCC with parallel retrieval.

Model-free training via PPO (parallel retrieval)

We now operationalize the RL reinterpretation by developing a model-free multi-agent training procedure based on PPO. In contrast to the point-to-point case, the distributed setting involves multiple encoders, each observing a local noisy version of the source and acting independently, while the relevance term couples all agents through the joint forward-channel outputs. This naturally motivates a CTDE paradigm, in which each encoder is trained using a centralized critic driven by a shared reconstruction reward together with an explicit actor-side rate regularizer, but executes solely based on its own local observation. As before, we assume access to a finite dataset $\mathcal{D}=\{(x^{(i)},\boldsymbol{y}_{1:J}^{(i)})\}_{i=1}^{M}$ generated by the access channels, while each forward channel is treated as a black-box environment that returns samples $t_j\sim p(\mathsf{t}_j|\mathsf{z}_j)$ upon query. No differentiability or parametric knowledge of the forward channels is required.

Parameterization

For the parallel retrieval scheme, we consider the following parameterized components:

- **Local encoder policies:** each agent $j\in\{1,\dots,J\}$ implements a stochastic encoder $p_{\boldsymbol{\theta}_j}(\mathsf{z}_j|\mathsf{y}_j)$, parameterized by a DNN with parameters $\boldsymbol{\theta}_j$. Execution is fully decentralized: encoder j only observes y_j.
- **Joint decoder:** a centralized decoder $q_{\boldsymbol{\phi}}(\mathsf{x}|\mathbf{t}_{1:J})$ reconstructs the source from all forward-channel outputs.
- **Latent priors:** each agent has an independent prior $r_{\boldsymbol{\psi}_j}(\mathsf{z}_j)$, to approximate the local complexity penalty.
- **Centralized value function:** a critic $V_{\boldsymbol{\zeta}}(\mathbf{y}_{1:J})$ estimates the expected shared reconstruction return under the joint observation tuple. This critic is used only during training.

Given these components, the shared sampled reward for a single sample is defined as

$$\mathsf{r}_{\mathrm{Par}}(x,\boldsymbol{t}_{1:J})=\log q_{\boldsymbol{\phi}}(x|\boldsymbol{t}_{1:J}). \tag{35}$$

The local actor-side compression regularizer for agent j is

$$\mathsf{c}_{\mathrm{Par},j}(y_j,z_j)=\log\frac{p_{\boldsymbol{\theta}_j}(z_j|y_j)}{r_{\boldsymbol{\psi}_j}(z_j)}. \tag{36}$$

The first term encourages team relevance, while the second provides the local complexity penalty used directly in each actor update.

Trajectory generation

Training proceeds by repeatedly sampling mini-batches from $\mathcal{D}$ and interacting with the forward channels:

1) Sample a mini-batch $\{(x^{(i)},y_1^{(i)},\dots,y_J^{(i)})\}_{i=1}^{B}$.
2) For each agent j and each sample i, draw an action $z_j^{(i)}\sim p_{\boldsymbol{\theta}_j}(\mathsf{z}_j|y_j^{(i)})$.
3) Query each forward link: $t_j^{(i)}\sim p(\mathsf{t}_j|z_j^{(i)})$.
4) Compute the shared sampled reward $\mathsf{r}_{\mathrm{Par}}^{(i)}$ using (35).
5) Compute the centralized value estimate $V_{\boldsymbol{\zeta}}(\boldsymbol{y}_{1:J}^{(i)})$.

These tuples $(x^{(i)},\boldsymbol{y}_{1:J}^{(i)},\boldsymbol{z}_{1:J}^{(i)},\boldsymbol{t}_{1:J}^{(i)},\mathsf{r}_{\mathrm{Par}}^{(i)})$ constitute a batch of experience for multi-agent PPO, while each sampled pair $(y_j^{(i)},z_j^{(i)})$ contributes to the local actor regularizer in (36).

Advantage estimation

We use a one-step advantage estimator based on the shared reconstruction reward:

$$A_{\text{Par}}^{(i)} = \mathsf{r}_{\text{Par}}^{(i)} - V_{\boldsymbol{\zeta}}(\boldsymbol{y}_{1:J}^{(i)}). \tag{37}$$

This advantage is shared across all agents, since the sampled reward is global.

PPO clipped objective (per agent)

For each agent j, define the likelihood ratio

$$\rho_j^{(i)} = \frac{p_{\boldsymbol{\theta}_j}(z_j^{(i)}|y_j^{(i)})}{p_{\boldsymbol{\theta}_{j,\text{old}}}(z_j^{(i)}|y_j^{(i)})}. \tag{38}$$

The PPO term for agent j is

$$\mathcal{L}_{\text{Par},j}^{\text{PPO}}(\boldsymbol{\theta}_j) = \frac{1}{B}\sum_{i=1}^{B} \min\Big(\rho_j^{(i)} A_{\text{Par}}^{(i)}, \text{clip}(\rho_j^{(i)}, 1-\epsilon, 1+\epsilon)\, A_{\text{Par}}^{(i)}\Big), \tag{39}$$

where $\epsilon > 0$ is the clipping parameter. The actor-regularized objective for agent j is then

$$\mathcal{L}_{\text{Par},j}^{\text{Act}}(\boldsymbol{\theta}_j) = \mathcal{L}_{\text{Par},j}^{\text{PPO}}(\boldsymbol{\theta}_j) - \frac{\lambda_j}{B}\sum_{i=1}^{B} \mathsf{c}_{\text{Par},j}(y_j^{(i)}, z_j^{(i)}). \tag{40}$$

Each encoder updates its parameters independently, but all encoders share the same advantage signal.

Decoder, priors, and value-function updates

The decoder and priors optimize two components of the IB surrogate, but via separate updates: the decoder maximizes the shared reconstruction reward, while the priors tighten the local reference terms used in the actor regularizers:

$$\mathbf{g}_{\boldsymbol{\phi}} = \frac{1}{B}\sum_{i=1}^{B} \nabla_{\boldsymbol{\phi}} \log q_{\boldsymbol{\phi}}(x^{(i)}|\boldsymbol{t}_{1:J}^{(i)}), \tag{41}$$

$$\mathbf{g}_{\boldsymbol{\psi}_j} = \frac{1}{B}\sum_{i=1}^{B} \lambda_j \nabla_{\boldsymbol{\psi}_j} \log r_{\boldsymbol{\psi}_j}(z_j^{(i)}). \tag{42}$$

The centralized critic is trained by minimizing

$$\mathcal{L}_{\text{Par}}^{\text{VF}}(\boldsymbol{\zeta}) = \frac{1}{B}\sum_{i=1}^{B} \big(V_{\boldsymbol{\zeta}}(\boldsymbol{y}_{1:J}^{(i)}) - \mathsf{r}_{\text{Par}}^{(i)}\big)^2. \tag{43}$$

Parameter updates

The parameters are updated via gradient ascent/descent:

$$\boldsymbol{\theta}_j \leftarrow \boldsymbol{\theta}_j + \eta_\theta \nabla_{\boldsymbol{\theta}_j} \mathcal{L}_{\text{Par},j}^{\text{Act}}, \tag{44}$$

$$\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} + \eta_\phi \mathbf{g}_{\boldsymbol{\phi}}, \tag{45}$$

$$\boldsymbol{\psi}_j \leftarrow \boldsymbol{\psi}_j + \eta_\psi \mathbf{g}_{\boldsymbol{\psi}_j}, \tag{46}$$

$$\boldsymbol{\zeta} \leftarrow \boldsymbol{\zeta} - \eta_\zeta \nabla_{\boldsymbol{\zeta}} \mathcal{L}_{\text{Par}}^{\text{VF}}. \tag{47}$$

Algorithm summary

The resulting training procedure constitutes a multi-agent extension of the point-to-point PPO framework. Each encoder acts as an independent RL agent, all agents share the same reconstruction-based reward signal, and a centralized critic stabilizes training by providing advantage estimates, while each actor is regularized by its local rate penalty. Execution remains fully decentralized, as each encoder uses only its local observation y_j. We refer to this distributed PPO-based training framework—presented in Alg. 2—as *Deep MARL-MFAVIB-Parallel*.

Algorithm 2 Deep MARL-MFAVIB-Parallel

Initialize parameters $\{\boldsymbol{\theta}_j\}_{j=1}^J$, $\boldsymbol{\phi}$, $\{\boldsymbol{\psi}_j\}_{j=1}^J$, $\boldsymbol{\zeta}$.
repeat
  Sample a mini-batch $\{(x^{(i)}, y_1^{(i)}, \ldots, y_J^{(i)})\}_{i=1}^B$.
  for $i = 1, \ldots, B$ **do**
    for $j = 1, \ldots, J$ **do**
      Sample action $z_j^{(i)} \sim p_{\boldsymbol{\theta}_j}(\mathsf{z}_j|y_j^{(i)})$.
      Query forward channel: $t_j^{(i)} \sim p(\mathsf{t}_j|z_j^{(i)})$.
    end for
    Compute shared sampled reward $\mathsf{r}_{\text{Par}}^{(i)}$ using (35).
    Evaluate value estimate $V_{\boldsymbol{\zeta}}(\boldsymbol{y}_{1:J}^{(i)})$.
    Compute advantage $A_{\text{Par}}^{(i)}$ using (37).
  end for
  for $j = 1, \ldots, J$ **do**
    Compute actor objective $\mathcal{L}_{\text{Par},j}^{\text{Act}}$ using (40).
    Update encoder parameters $\boldsymbol{\theta}_j$ via actor-regularized PPO.
    Update prior parameters $\boldsymbol{\psi}_j$ using $\mathbf{g}_{\boldsymbol{\psi}_j}$.
  end for
  Update decoder parameters $\boldsymbol{\phi}$ using $\mathbf{g}_{\boldsymbol{\phi}}$.
  Update value-function parameters $\boldsymbol{\zeta}$ by minimizing $\mathcal{L}_{\text{Par}}^{\text{VF}}$.
until convergence

2) Successive Retrieval

In the successive retrieval, the RPU processes the forward-link outputs in a fixed order and exploits the side-information from previous branches when treating the current branch. This induces a *sequential* information-processing structure: encoder j produces z_j without access to side-information, but the decoder evaluates its relevance conditioned on the previously received outputs $\mathbf{t}_{1:j-1}$. This setting naturally leads to a cooperative MARL formulation with *conditional* complexity penalties. Each encoder j behaves as an autonomous agent with policy $p(\mathsf{z}_j|\mathsf{y}_j)$, the forward links again form the environment, and the decoder together with a set of *conditional* priors define a shared objective whose sampled reward is common across agents while the *conditional* compression terms act as actor-side regularizers. To construct this MARL framework, we derive a sample-based utility by bounding the two information terms in the multiterminal IB Lagrangian for successive retrieval in (26).

Lower-bound on the relevance term

For the mutual-information term $I(\mathsf{x}; \mathbf{t}_{1:J})$, the same variational argument as in the parallel case applies. Thus, without repeating the whole derivation here again, we only remind that by introducing a decoder distribution $q(\mathsf{x}|\mathbf{t}_{1:J})$, the log-likelihood $\log q(x|\boldsymbol{t}_{1:J})$ provides a sample-based surrogate for the relevance term.

Upper-bound on the conditional compression terms

In successive retrieval, the compression rate of encoder j is conditional:

$$I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{t}_{1:j-1}) = \mathbb{E}_{\mathsf{y}_j,\mathsf{z}_j,\mathbf{t}_{1:j-1}}\left\{\log\frac{p(\mathsf{z}_j|\mathsf{y}_j)}{p(\mathsf{z}_j|\mathbf{t}_{1:j-1})}\right\}. \quad (48)$$

Since the true conditional prior $p(\mathsf{z}_j|\mathbf{t}_{1:j-1})$ is unknown and generally intractable, we introduce a *conditional* reference prior $r(\mathsf{z}_j|\mathbf{t}_{1:j-1})$, which yields the upper-bound

$$I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{t}_{1:j-1}) \le \mathbb{E}_{\mathsf{y}_j,\mathsf{z}_j,\mathbf{t}_{1:j-1}}\left\{\log\frac{p(\mathsf{z}_j|\mathsf{y}_j)}{r(\mathsf{z}_j|\mathbf{t}_{1:j-1})}\right\}, \quad (49)$$

as it applies

$$\begin{aligned}\mathbb{E}_{\mathsf{y}_j,\mathsf{z}_j,\mathbf{t}_{1:j-1}}\left\{\log\frac{p(\mathsf{z}_j|\mathsf{y}_j)}{p(\mathsf{z}_j|\mathbf{t}_{1:j-1})}\right\} - \mathbb{E}_{\mathsf{y}_j,\mathsf{z}_j,\mathbf{t}_{1:j-1}}\left\{\log\frac{p(\mathsf{z}_j|\mathsf{y}_j)}{r(\mathsf{z}_j|\mathbf{t}_{1:j-1})}\right\}\\ = -\mathbb{E}_{\mathbf{t}_{1:j-1}}\left\{D_{\mathrm{KL}}\big(p(\mathsf{z}_j|\mathbf{t}_{1:j-1}) \parallel r(\mathsf{z}_j|\mathbf{t}_{1:j-1})\big)\right\} \le 0.\end{aligned} \quad (50)$$

So, it is inferred that each log-ratio $\log\frac{p(z_j|y_j)}{r(z_j|\boldsymbol{t}_{1:j-1})}$ provides a sample-based surrogate for the *conditional* compression cost of branch j.

Utility decomposition

Combining the relevance lower-bound with the conditional compression upper-bounds yields the multiterminal utility

$$\mathsf{u}_{\mathrm{Suc}}(x,\boldsymbol{y}_{1:J},\boldsymbol{z}_{1:J},\boldsymbol{t}_{1:J}) = \log q(x|\boldsymbol{t}_{1:J}) - \sum_{j=1}^{J}\lambda_j\log\frac{p(z_j|y_j)}{r(z_j|\boldsymbol{t}_{1:j-1})}. \quad (51)$$

Its first term is a shared sampled reward, while the second term is a sum of conditional compression regularizers applied directly in the encoder updates. Importantly, similar to the previous case of parallel retrieval, here as well the complete surrogate is computable from samples $(x,\boldsymbol{y}_{1:J},\boldsymbol{z}_{1:J},\boldsymbol{t}_{1:J})$ without requiring any knowledge of the forward-channel transition laws or the true conditional priors.

Expected utility as the MARL objective

Under the joint encoder policy

$$p(\mathbf{z}_{1:J}|\mathbf{y}_{1:J}) = \prod_{j=1}^{J}p(\mathsf{z}_j|\mathsf{y}_j), \quad (52)$$

the expected utility is

$$\mathcal{J}_{\mathrm{Suc}}(P) = \mathbb{E}_{\mathsf{x},\mathbf{y}_{1:J},\mathbf{z}_{1:J},\mathbf{t}_{1:J}}\left\{\mathsf{u}_{\mathrm{Suc}}(\mathsf{x},\mathbf{y}_{1:J},\mathbf{z}_{1:J},\mathbf{t}_{1:J})\right\}. \quad (53)$$

Maximizing $\mathcal{J}_{\mathrm{Suc}}(P)$ thus maximizes a sample-based lower-bound on relevance while minimizing a sample-based upper-bound on the *conditional* compression rates—precisely mirroring the structure of the original IB objective for multiterminal JSCC with successive retrieval.

MARL interpretation

As in the parallel retrieval setting, the interaction remains a single-step contextual bandit under CTDE. The shared sampled reward is again the reconstruction term, while the actor regularizer now uses *conditional* priors, which introduce a dependence on previous forward-channel outputs. Apart from this modification in the regularizer structure, all MARL elements—agents, actions, environment, and training protocol—remain unchanged.

Model-free training via PPO (successive retrieval)

We now extend the PPO-based MARL training framework to the successive-retrieval setting, where the decoder exploits side-information from earlier branches. In contrast to the parallel case, each encoder j is regularized according to the *conditional* mutual-information term $I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{t}_{1:j-1})$. The overall structure of the training pipeline remains identical to the parallel scheme—CTDE MARL—while the actor regularizer and priors are modified to incorporate the conditional dependence on previously received forward-channel outputs and the critic continues to track the shared reconstruction reward. As before, we assume access to a finite dataset $\mathcal{D} = \{(x^{(i)},\boldsymbol{y}_{1:J}^{(i)})\}_{i=1}^{M}$ generated by the access channels, and each forward channel is treated as a black-box environment that returns samples $t_j \sim p(\mathsf{t}_j|\mathsf{z}_j)$ upon query.

Parameterization

The overall parameterization mirrors the parallel retrieval case, and we therefore only emphasize the modification required by the successive formulation. The local encoder policies $p_{\boldsymbol{\theta}_j}(\mathsf{z}_j|\mathsf{y}_j)$, the centralized decoder $q_{\boldsymbol{\phi}}(\mathsf{x}|\mathbf{t}_{1:J})$, and the centralized value function $V_{\boldsymbol{\zeta}}(\mathbf{y}_{1:J})$ retain exactly the same structure and roles as in the parallel scheme. Execution remains fully decentralized at the encoders, and the critic is used only during training.

The sole structural difference lies in the modeling of the complexity term. Instead of unconditional priors $r_{\boldsymbol{\psi}_j}(\mathsf{z}_j)$ used in the parallel retrieval, the successive formulation requires *conditional* priors $r_{\boldsymbol{\psi}_j}(\mathsf{z}_j|\mathbf{t}_{1:j-1})$, which approximate $p(\mathsf{z}_j|\mathbf{t}_{1:j-1})$ and thereby enable estimating the conditional mutual-information penalty $I(\mathsf{y}_j;\mathsf{z}_j|\mathbf{t}_{1:j-1})$. Similar to the previous case of parallel retrieval, in the actor-regularized formulation, the shared sampled reward and the conditional actor-side regularizer are handled separately as

$$\mathsf{r}_{\mathrm{Suc}}(x,\boldsymbol{t}_{1:J}) = \log q_{\boldsymbol{\phi}}(x|\boldsymbol{t}_{1:J}), \quad (54)$$

$$\mathsf{c}_{\mathrm{Suc},j}(y_j,z_j,\boldsymbol{t}_{1:j-1}) = \log\frac{p_{\boldsymbol{\theta}_j}(z_j|y_j)}{r_{\boldsymbol{\psi}_j}(z_j|\boldsymbol{t}_{1:j-1})}. \quad (55)$$

Trajectory generation

The trajectory generation procedure is identical in structure to the parallel retrieval setup; the major difference is that the actor regularizer now depends on the previously realized forward-channel outputs through (55), while the sampled reward remains the shared reconstruction term in (54). All other steps—sampling observations, drawing encoder actions, querying the black-box forward channels, and evaluating the centralized critic—remain unchanged. The resulting tuples $(x^{(i)},\boldsymbol{y}_{1:J}^{(i)},\boldsymbol{z}_{1:J}^{(i)},\boldsymbol{t}_{1:J}^{(i)},\mathsf{r}_{\mathrm{Suc}}^{(i)})$ form a batch of experience for multi-agent PPO.

Advantage estimation

As in the parallel case, we use a one-step estimator based on the shared reconstruction reward:

$$A_{\mathrm{Suc}}^{(i)} = \mathsf{r}_{\mathrm{Suc}}^{(i)} - V_{\boldsymbol{\zeta}}(\boldsymbol{y}_{1:J}^{(i)}). \quad (56)$$

This advantage is shared across all agents.

PPO clipped objective (per agent)

With the same definition of the likelihood ratio $\rho_j^{(i)}$ given in (38), the PPO term for agent j is

$$\mathcal{L}^{\text{PPO}}_{\text{Suc},j}(\boldsymbol{\theta}_j)=\frac{1}{B}\sum_{i=1}^{B}\min\Big(\rho_j^{(i)}A^{(i)}_{\text{Suc}},\,\text{clip}(\rho_j^{(i)},1-\epsilon,1+\epsilon)\,A^{(i)}_{\text{Suc}}\Big). \tag{57}$$

The actor-regularized objective for agent j is then

$$\mathcal{L}^{\text{Act}}_{\text{Suc},j}(\boldsymbol{\theta}_j)=\mathcal{L}^{\text{PPO}}_{\text{Suc},j}(\boldsymbol{\theta}_j)-\frac{\lambda_j}{B}\sum_{i=1}^{B}\mathsf{c}_{\text{Suc},j}(y_j^{(i)},z_j^{(i)},\boldsymbol{t}^{(i)}_{1:j-1}). \tag{58}$$

Decoder, conditional priors, and value-function updates

The decoder and conditional priors optimize the two components of the IB surrogate, but through separate updates: the decoder maximizes the shared reconstruction reward, while the *conditional* priors tighten the reference terms used in the actor regularizers:

$$\mathbf{g}_{\boldsymbol{\phi}}=\frac{1}{B}\sum_{i=1}^{B}\nabla_{\boldsymbol{\phi}}\log q_{\boldsymbol{\phi}}(x^{(i)}|\boldsymbol{t}^{(i)}_{1:J}), \tag{59}$$

$$\mathbf{g}_{\boldsymbol{\psi}_j}=\frac{1}{B}\sum_{i=1}^{B}\lambda_j\,\nabla_{\boldsymbol{\psi}_j}\log r_{\boldsymbol{\psi}_j}(z_j^{(i)}|\boldsymbol{t}^{(i)}_{1:j-1}). \tag{60}$$

The centralized critic is trained by minimizing

$$\mathcal{L}^{\text{VF}}_{\text{Suc}}(\boldsymbol{\zeta})=\frac{1}{B}\sum_{i=1}^{B}\big(V_{\boldsymbol{\zeta}}(\boldsymbol{y}^{(i)}_{1:J})-\mathsf{r}^{(i)}_{\text{Suc}}\big)^2. \tag{61}$$

Parameter updates

The parameters are updated via gradient ascent/descent:

$$\boldsymbol{\theta}_j\leftarrow\boldsymbol{\theta}_j+\eta_\theta\,\nabla_{\boldsymbol{\theta}_j}\mathcal{L}^{\text{Act}}_{\text{Suc},j}, \tag{62}$$

$$\boldsymbol{\phi}\leftarrow\boldsymbol{\phi}+\eta_\phi\,\mathbf{g}_{\boldsymbol{\phi}}, \tag{63}$$

$$\boldsymbol{\psi}_j\leftarrow\boldsymbol{\psi}_j+\eta_\psi\,\mathbf{g}_{\boldsymbol{\psi}_j}, \tag{64}$$

$$\boldsymbol{\zeta}\leftarrow\boldsymbol{\zeta}-\eta_\zeta\,\nabla_{\boldsymbol{\zeta}}\mathcal{L}^{\text{VF}}_{\text{Suc}}. \tag{65}$$

Algorithm summary

The training framework follows the exact same structure as in the parallel retrieval case; the substantive change is that the actor regularizer now depends on *conditional* priors, while the shared reward remains reconstruction-based. All other elements of the PPO-based CTDE framework—independent encoder updates, a shared reconstruction reward, and a centralized critic—remain unchanged. We refer to this training approach—presented in Alg. 3—as *Deep MARL-MFAVIB-Successive*.

Algorithm 3 Deep MARL-MFAVIB-Successive

Initialize parameters $\{\boldsymbol{\theta}_j\}_{j=1}^J$, $\boldsymbol{\phi}$, $\{\boldsymbol{\psi}_j\}_{j=1}^J$, $\boldsymbol{\zeta}$.
repeat
  Sample a mini-batch $\{(x^{(i)},y_1^{(i)},\dots,y_J^{(i)})\}_{i=1}^B$.
  for $i=1,\dots,B$ **do**
    for $j=1,\dots,J$ **do**
      Sample action $z_j^{(i)}\sim p_{\boldsymbol{\theta}_j}(\mathsf{z}_j|y_j^{(i)})$.
      Query forward channel: $t_j^{(i)}\sim p(\mathsf{t}_j|z_j^{(i)})$.
    end for
    Compute shared sampled reward $\mathsf{r}^{(i)}_{\text{Suc}}$ using (54).
    Evaluate value estimate $V_{\boldsymbol{\zeta}}(\boldsymbol{y}^{(i)}_{1:J})$.
    Compute advantage $A^{(i)}_{\text{Suc}}$ using (56).
  end for
  for $j=1,\dots,J$ **do**
    Compute actor objective $\mathcal{L}^{\text{Act}}_{\text{Suc},j}$ using (58).
    Update encoder parameters $\boldsymbol{\theta}_j$ via actor-regularized PPO.
    Update conditional prior parameters $\boldsymbol{\psi}_j$ using $\mathbf{g}_{\boldsymbol{\psi}_j}$.
  end for
  Update decoder parameters $\boldsymbol{\phi}$ using $\mathbf{g}_{\boldsymbol{\phi}}$.
  Update value-function parameters $\boldsymbol{\zeta}$ by minimizing $\mathcal{L}^{\text{VF}}_{\text{Suc}}$.
until convergence

IV. NUMERICAL RESULTS

A. SPECIFICATIONS & IMPLEMENTATION DETAILS

In the following investigations, the source alphabet consists of a normalized 16-QAM constellation. A source symbol is drawn uniformly and transmitted to $J=3$ agents through independent AWGN access channels, each with the noise variance σ_{n}^2. Then, every agent compresses its observation into a discrete index $z_j\in\{0,\dots,N-1\}$. These indices are transmitted through independent two-state Gilbert–Elliott erasure channels. The channel state evolves as Good-to-Bad with different probabilities $(0.01,0.1,0.5)$ and a fixed Bad-to-Good probability 0.10. Conditional on the updated state, the transmitted symbol is erased with probability $P_{\text{good}}=0.05$ in the Good state and $P_{\text{bad}}=0.50$ in the Bad state. Erasures are represented by the dedicated symbol N, so the decoder receives $t_j\in\{0,\dots,N\}$. The decoder always consumes the full tuple (t_1,t_2,t_3).

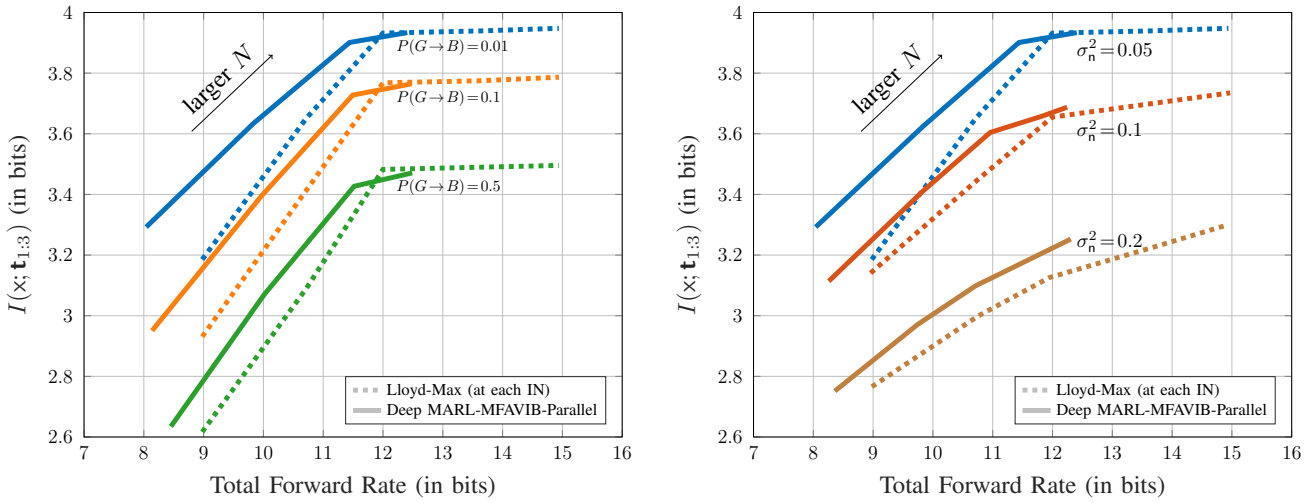
All reported information quantities are estimated using $300{,}000$ Monte Carlo samples, processed in chunks of size $50{,}000$. For each sample, a fresh source symbol is drawn, the access channels are simulated, compression indices are sampled from the learned policies, and the forward channels are implemented using the described Gilbert–Elliott dynamics.

Neural architectures

All learning modules are implemented as fully-connected feed-forward networks (multi-layer perceptrons) with two hidden layers and $\tanh$ activation functions. All weights are initialized i.i.d. from a zero-mean Gaussian with standard deviation 0.1, and all biases are initialized to zero. The following architectures are used throughout:

- **Shared policy:** input dimension 2 (real and imaginary parts of y_j), hidden layers $(256,256)$, output dimension N (logits for $p_{\boldsymbol{\theta}_j}(z_j|y_j)$).
- **Critic:** input dimension 6 (concatenated real and imaginary parts of (y_1,y_2,y_3)), hidden layers $(512,512)$, output dimension 1.

This article has been accepted for publication in IEEE Open Journal of the Communications Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2026.3721698



(a) Fixed access links' noise variance $\sigma_{\mathrm{n}}^2=0.05$ (per branch)

(b) Fixed forward Good-to-Bad probability $P(G\to B)=0.01$

FIGURE 3. The relevant information $I(\mathsf{x};\mathbf{t}_{1:3})$ vs. total forward rate $\sum_j I(\mathsf{y}_j;\mathsf{z}_j)$ for a) fixed access noise variance, and b) fixed forward Good-to-Bad transition probability. Equiprobable (power-normalized) 16-QAM source signaling over AWGN access channels and Gilbert-Elliot forward channels. These results have been obtained for fixed common trade-off $\lambda=0$ by varying the encoders' output cardinality $8\le N\le 32$ (per branch) for both a) & b).

- **Decoder:** input dimension $3(N{+}1)$ (concatenated one-hot encodings of (t_1,t_2,t_3)), hidden layers $(256,256)$, output dimension 16 (logits for $q_{\boldsymbol{\phi}}(x|t_1,t_2,t_3)$).
- **Shared parallel prior:** input dimension 1 (constant scalar), hidden layers $(128,128)$, output dimension N.
- **Successive priors:**
  - $r_{\boldsymbol{\psi}_1}(z_1)$: input dimension 1,
  - $r_{\boldsymbol{\psi}_2}(z_2|t_1)$: input dimension $N{+}1$,
  - $r_{\boldsymbol{\psi}_3}(z_3|t_1,t_2)$: input dimension $2(N{+}1)$,

  each with hidden layers $(128,128)$ and output dimension N.

Optimization and MAPPO setup

All trainable modules are optimized using Adam [60]. Training proceeds in rollout batches of size 2048, with 4 PPO epochs per rollout and minibatch size 512. The clipped PPO surrogate (with $\epsilon=0.2$) is used with one-step advantage estimator that is normalized within each rollout. The decoder is trained by cross-entropy on the true source symbol, and the critic is trained by mean-squared error to the rollout reward. The actor updates include the appropriate KL-based rate penalties—unconditional priors for parallel retrieval and conditional priors for successive retrieval.

B. BASELINE COMPARISON

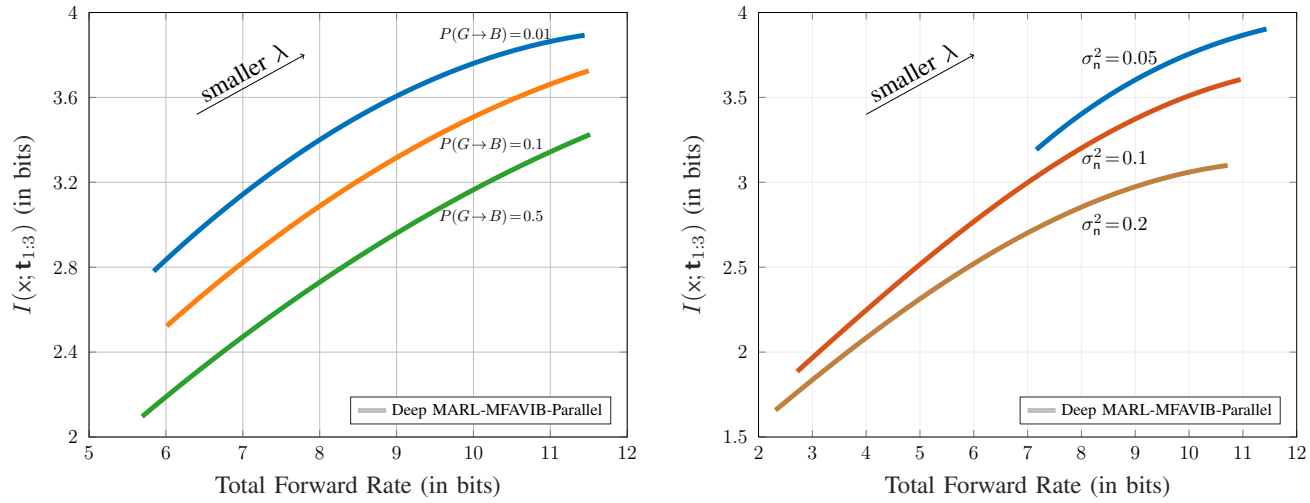
As the first round of investigations, Fig. 3 illustrates the trade-off between the relevant information $I(\mathsf{x};\mathbf{t}_{1:3})$ and the total forward rate $\sum_{j=1}^{3} I(\mathsf{y}_j;\mathsf{z}_j)$ under different reliability conditions of the access and forward links. In all cases, the encoder output cardinality N is varied, which increases the per-agent compression rate and moves the operating point along each curve. As the main trend, it is observed from both results that the Deep MARL-MFAVIB-Parallel method consistently outperforms the popular Lloyd-Max baseline [61] across a wide range of operating points.

(a) Varying forward-channel reliability

Fig. 3(a) depicts the achievable end-to-end relevant information for a fixed access-channel noise variance $\sigma_{\mathrm{n}}^2=0.05$ while varying the Good-to-Bad transition probability of the forward Gilbert-Elliott channels. Three reliability regimes are considered, corresponding to $P(G\to B)\in\{0.01,\,0.1,\,0.5\}$. For each regime, increasing the encoder cardinality N increases the total forward rate and improves the end-to-end information transfer. However, as the forward channels become less reliable (larger $P(G\to B)$ values), the attainable relevant information decreases for any fixed rate, and the curves shift downward. This behavior reflects the fact that more frequent transitions into the Bad state lead to more erasures, thereby degrading the informativity of the tuple (t_1,t_2,t_3) available at the decoder. Despite this degradation, the learned Deep MARL-MFAVIB-Parallel encoders retain a clear advantage over Lloyd-Max, depicting their ability to adapt to the stochastic erasure patterns of the forward links.

(b) Varying access-channel quality

Fig. 3(b) fixes the forward-channel reliability at $P(G\to B)=0.01$ and varies the access-channel noise variance across $\sigma_{\mathrm{n}}^2\in\{0.05,\,0.1,\,0.2\}$. For each noise level, increasing N again increases the total forward rate and improves the relevant information. However, as the access channels become noisier, the curves shift downward: the agents' observations become less informative about the source symbol, and even large encoder cardinalities cannot fully compensate for the reduced quality of y_j. This highlights the fundamental limitation that the decoder cannot recover information that is not present in the agents' initial observations. Across all noise levels, the Deep MARL-MFAVIB-Parallel method consistently achieves higher relevant information than the Lloyd-Max baseline, reflecting its ability to establish superior *information-compression* trade-offs.



(a) Fixed access links' noise variance $\sigma_{\mathrm{n}}^2 = 0.05$ (per branch)

(b) Fixed forward Good-to-Bad probability $P(G \to B) = 0.01$

FIGURE 4. The relevant information $I(\mathrm{x}; \mathbf{t}_{1:3})$ vs. total forward rate $\sum_j I(\mathrm{y}_j; \mathrm{z}_j)$ for a) fixed access noise variance, and b) fixed forward Good-to-Bad transition probability. Equiprobable (power-normalized) 16-QAM source signaling over AWGN access channels and Gilbert-Elliot forward channels. These results have been obtained for fixed output cardinality $N = 16$ by varying the common trade-off parameter a) $0 \leq \lambda \leq 1$ and b) $0 \leq \lambda \leq 0.4$.

Overall interpretation

Taken together, Figs. 3(a) and 3(b) illustrate the complementary roles of the access and forward channels in determining the end-to-end information flow. When the access links are clean but the forward links are unreliable, the information bottleneck is dominated by erasures in the forward channel. Conversely, when the forward links are reliable but the access links are noisy, the bottleneck is dominated by the limited informativity of the observations. In both regimes, the Deep MARL-MFAVIB-Parallel method consistently provides superior performance, demonstrating the benefits of end-to-end MARL-based IB design over classical quantization. Finally, we note that Deep MFAVIB-Parallel [43] cannot be included as a baseline here because its variational objective requires a differentiable and fully known forward-channel model that is incompatible with the non-differentiable Gilbert–Elliott links used in our experiments.

C. INFORMATION-COMPRESSION CURVE SWEEP FOR PARALLEL RETRIEVAL

Fig. 4 illustrates how the achievable relevant information $I(\mathrm{x}; \mathbf{t}_{1:3})$ varies with the total forward rate $\sum_{j=1}^{3} I(\mathrm{y}_j; \mathrm{z}_j)$ when the common trade-off parameter $\lambda = \lambda_j$ for $j = 1, 2, 3$ is swept over a prescribed range. Unlike the previous set of experiments, where the operating points were generated by varying the encoder's output cardinality N, here it is fixed at $N = 16$ and the information-compression balance is controlled solely through λ. Note that classical Lloyd–Max quantization cannot produce a λ-sweep curve because it has no trade-off parameter analogous to λ. Similar to the earlier results, the curves reflect the interplay between the access and forward channels, but the λ-sweep reveals how the learned encoders continuously adjust their operating point along the information-compression frontier.

(a) Fixed access noise variance

Fig. 4(a) fixes the access-channel noise variance at $\sigma_{\mathrm{n}}^2 = 0.05$ and varies the forward-channel reliability through the Good-to-Bad transition probability $P(G \to B) \in \{0.01, 0.1, 0.5\}$. For each reliability regime, decreasing λ moves the operating point toward higher forward rates and higher relevant information, while increasing λ enforces stronger compression and shifts the operating point toward the lower-rate region. The overall shape and ordering of the curves are consistent with the trends observed in the first set of results: more reliable forward channels yield higher $I(\mathrm{x}; \mathbf{t}_{1:3})$ for any given rate, whereas more frequent transitions into the Bad state reduce the informativity of the received tuple. The λ-sweep makes this trade-off explicit by showing how the learned encoders adapt their compression level to the reliability of the forward links.

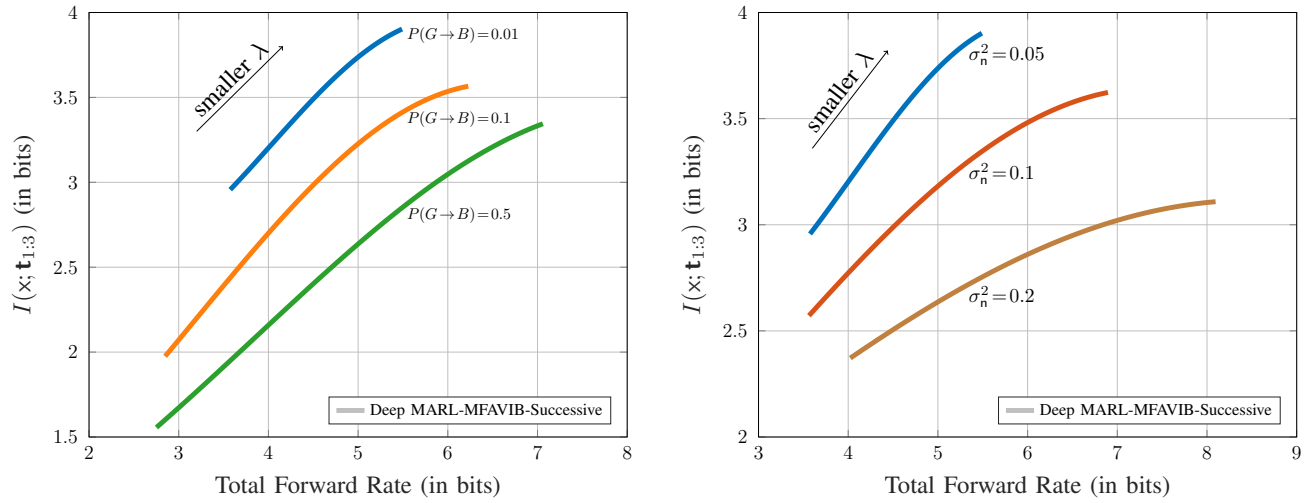
(b) Fixed forward-channel reliability

Fig. 4(b) fixes the forward-channel reliability at $P(G \to B) = 0.01$ and varies the access-channel noise variance across $\sigma_{\mathrm{n}}^2 \in \{0.05, 0.1, 0.2\}$. As λ decreases, the encoders allocate more rate and preserve more information, while larger λ forces stronger compression. The relative ordering of the curves mirrors the behavior seen previously: cleaner access channels support higher relevant information, whereas noisier observations fundamentally limit the amount of information that the decoder can recover, regardless of the compression level. The λ-sweep highlights this limitation by showing that even aggressive rate allocation (small λ) cannot fully compensate for severely degraded access observations.

Overall interpretation

Together, Figs. 4(a) and 4(b) demonstrate that the trade-off parameter λ provides a smooth and effective mechanism for navigating the information-compression frontier. For fixed

This article has been accepted for publication in IEEE Open Journal of the Communications Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2026.3721698



(a) Fixed access links' noise variance $\sigma_{\mathrm{n}}^2 = 0.05$ (per branch)

(b) Fixed forward Good-to-Bad probability $P(G \to B) = 0.01$

FIGURE 5. The relevant information $I(\mathbf{x}; \mathbf{t}_{1:3})$ vs. total forward rate $\sum_j I(\mathbf{y}_j; \mathbf{z}_j | \mathbf{t}_{1:j-1})$ for a) fixed access noise variance, and b) fixed forward Good-to-Bad transition probability. Equiprobable (power-normalized) 16-QAM source signaling over AWGN access channels and Gilbert-Elliot forward channels. These results have been obtained for fixed output cardinality $N = 16$ by varying the common trade-off parameter $0 \leq \lambda \leq 0.2$ for both a) & b).

encoder cardinality, decreasing λ increases the forward rate and improves the preservation of relevant information, while increasing λ enforces stronger compression. The resulting curves reflect the same structural dependencies on access and forward reliability observed in the first set of results, but now reveal how the learned Deep MARL-MFAVIB-Parallel encoders continuously adjust their operating point in response to the imposed relevance-compression trade-off.

D. INFORMATION-COMPRESSION CURVE SWEEP FOR SUCCESSIVE RETRIEVAL

Fig. 5 illustrates the basic information-compression trade-off achieved by the Deep MARL-MFAVIB-Successive method when the common trade-off parameter λ is swept over the range $0 \leq \lambda \leq 0.2$ for a fixed encoder cardinality $N = 16$. As in the previous round of experiments, decreasing λ increases the forward rate and improves the preservation of relevant information, whereas larger λ enforces stronger compression. However, in contrast to the parallel retrieval setting, the successive architecture yields smaller overall forward rates because each agent's compression rate is conditioned on the previously received forward-channel outputs. Under the presumed Markovian structure, this conditioning reduces the per-branch rate according to

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

which directly lowers the total forward rate while retaining the most informative components of the agents' observations.

(a) Fixed access noise variance

Fig. 5(a) fixes the access-channel noise variance at $\sigma_{\mathrm{n}}^2 = 0.05$ and varies the forward-channel reliability across $P(G \to B) \in \{0.01, 0.1, 0.5\}$. The qualitative ordering of the curves mirrors the behavior observed earlier: more reliable forward links support higher relevant information, while more frequent transitions into the Bad state reduce the informativity of the received tuple. The λ-sweep again traces out the achievable frontier, but now at reduced forward rates.

(b) Fixed forward-channel reliability

Fig. 5(b) fixes the forward-channel reliability at $P(G \to B) = 0.01$ and varies the access-channel noise variance across $\sigma_{\mathrm{n}}^2 \in \{0.05, 0.1, 0.2\}$. As in the parallel case, cleaner access observations yield higher relevant information, while noisier observations fundamentally limit the decoder's ability to recover the source. The successive structure again produces lower forward rates for all λ, reflecting the reduced per-branch rate contributions.

Overall interpretation

These results demonstrate that the successive retrieval mechanism effectively compresses away redundant information across agents by exploiting the conditional structure induced by previously received signals. This leads to substantially lower total forward rates while maintaining a competitive level of relevant information, thereby offering a more efficient operating regime than the parallel architecture when inter-agent statistical dependencies are present.

V. SUMMARY

In this work, we presented a unified, model-free framework for IB-based JSCC that remains effective even when the forward channels are unknown, stochastic, non-differentiable, or entirely black-box. In contrast to classical IB formulations—which rely on explicit channel models—and variational approaches—which require differentiable end-to-end pipelines and reparameterizable latent variables—we reinterpreted the IB-based JSCC as a sequential decision-making problem, enabling the use of deep MARL to learn distributed compressors directly from interaction with arbitrary links.

Starting from the point-to-point setting, we showed that the encoder can be viewed as a stochastic policy in a contextual bandit, where a variational decomposition of the IB Lagrangian yields a reconstruction-based reward and an analytic actor-side compression regularizer. This leads to a fully model-free, forward-aware compressor trained via PPO without requiring gradients through the channel.

We then extended the framework to the multiterminal scenario. Two retrieval strategies were considered: a *parallel* scheme with unconditional complexity penalties, and a *successive* scheme that exploits decoder side-information via conditional penalties. For both processing schemes, we derived cooperative MARL objectives amenable to CTDE. The resulting algorithms provide scalable, model-free solutions to distributed IB-based JSCC.

Extensive experiments on a representative stochastic forward channel with hidden-state dynamics clearly depicted the effectiveness of the proposed methods. The MARL-based compressors consistently outperformed the popular Lloyd–Max quantization across a wide range of operating points, while the successive retrieval strategy achieved substantially lower forward rates by exploiting conditional structure. These results highlight the robustness and flexibility of the MARL-driven IB-based JSCC, which naturally adapts to the varying access quality, forward reliability, and relevance–compression trade-offs.

In summary, this work establishes deep MARL as a principled and powerful foundation for learning distributed IB-based compressors in complex environments where model- or gradient-based approaches are fundamentally inapplicable. By enabling reliable operation over arbitrary forward links, the proposed approach significantly broadens the applicability of IB-based JSCC and opens new avenues for data-driven communication system design.

REFERENCES

[1] N. Tishby, F. C. Pereira, and W. Bialek, "The Information Bottleneck Method," in *37th Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, USA, Sep. 1999.

[2] C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," *IRE Int. Conv. Rec.*, vol. 7, pp. 142–163, Mar. 1959.

[3] 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.

[4] P. Harremoës and N. Tishby, "The Information Bottleneck Revisited or How to Choose a Good Distortion Measure," in *IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007.

[5] R. Dobrushin and B. Tsybakov, "Information Transmission with Additional Noise," *IRE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[6] D. J. Sakrison, "Source Encoding in the Presence of Random Disturbance," *IEEE Trans. Inf. Theory*, vol. 14, pp. 165–167, Jan. 1968.

[7] 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.

[8] H. Witsenhausen, "Indirect Rate Distortion Problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[9] E. Ayanoglu, "On Optimal Quantization of Noisy Sources," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1450–1452, Nov. 1990.

[10] 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, Art. no. 151, Jan. 2020.

[11] Z. Goldfeld and Y. Polyanskiy, "The Information Bottleneck Problem and Its Applications in Machine Learning," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 19–38, May 2020.

[12] S. Hu, Z. Lou, X. Yan, and Y. Ye, "A Survey on Information Bottleneck," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, p. 5325–5344, Feb. 2024.

[13] A. Wyner, "On Source Coding with Side Information at the Decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.

[14] R. Ahlswede and J. Körner, "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.

[15] E. Erkip and T. M. Cover, "The Efficiency of Investment Information," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1026–1040, May 1998.

[16] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the Information Bottleneck to the Privacy Funnel," in *IEEE Inf. Theory Workshop*, Hobart, TAS, Australia, Nov. 2014.

[17] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, "Information Extraction Under Privacy Constraints," *Information*, vol. 7, no. 1, Art. no. 15, Mar. 2016.

[18] B. Razeghi, P. Rahimi, and S. Marcel, "Deep Variational Privacy Funnel: General Modeling with Applications in Face Recognition," in *2024 IEEE Int. Conf. Acoust., Speech Signal Process.*, Seoul, Republic of Korea, Apr. 2024.

[19] V. Doshi, D. Shah, M. Médard, and M. Effros, "Functional Compression Through Graph Coloring," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3901–3917, Aug. 2010.

[20] S. Feizi and M. Médard, "On Network Functional Compression," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5387–5401, Sep. 2014.

[21] Y. M. Saidutta, A. Abdi, and F. Fekri, "Analog Joint Source-Channel Coding for Distributed Functional Compression using Deep Neural Networks," in *IEEE Int. Symp. Inf. Theory*, Melbourne, Australia, Jul. 2021.

[22] D. Malak and M. Médard, "A Distributed Computationally Aware Quantizer Design via Hyper Binning," *IEEE Trans. Signal Process.*, vol. 71, pp. 76–91, Jan. 2023.

[23] I. Tal and A. Vardy, "How to Construct Polar Codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[24] M. Stark, A. Shah, and G. Bauch, "Polar Code Construction Using the Information Bottleneck Method," in *IEEE Wireless Commun. Netw. Conf. Workshops*, Barcelona, Spain, Apr. 2018.

[25] 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.

[26] M. Stark, L. Wang, G. Bauch, and R. D. Wesel, "Decoding Rate-Compatible 5G-LDPC Codes with Coarse Quantization Using the Information Bottleneck Method," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 646–660, May 2020.

[27] T. Monsees, O. Griebel, M. Herrmann, D. Wübben, A. Dekorsy, and N. Wehn, "Minimum-Integer Computation Finite Alphabet Message Passing Decoder: From Theory to Decoder Implementations towards 1 Tb/s," *Entropy*, vol. 24, no. 10, Art. no. 19, Oct. 2022.

[28] 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.

[29] J. Shao, Y. Mao, and J. Zhang, "Learning Task-Oriented Communication for Edge Inference: An Information Bottleneck Approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 197–211, Jan. 2022.

[30] D. Gündüz, Z. Qin, I. Estella-Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, Jan. 2023.

[31] E. Beck, C. Bockelmann, and A. Dekorsy, "Semantic Information Recovery in Wireless Networks," *Sensors*, vol. 23, no. 14, Art. no. 6347, Jul. 2023.

[32] I. Estella-Aguerri and A. Zaidi, "Distributed Information Bottleneck Method for Discrete and Gaussian Sources," in *Int. Zurich Semin. Inf. Commun.*, Zurich, Switzerland, Feb. 2018.

[33] 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.

[34] Y. Uğur, I. Estella-Aguerri, and A. Zaidi, "Vector Gaussian CEO Problem Under Logarithmic Loss and Applications," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4183–4202, Jul. 2020.

[35] I. Estella-Aguerri and A. Zaidi, "Distributed Variational Representation Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 120–138, Jan. 2021.

[36] S. Hassanpour, D. Wübben, and A. Dekorsy, "A Novel Approach to Distributed Quantization via Multivariate Information Bottleneck Method," in *IEEE Glob. Commun. Conf.*, Waikoloa, HI, USA, Dec. 2019.

[37] ——, "Generalized Distributed Information Bottleneck for Fronthaul Rate Reduction at the Cloud-RANs Uplink," in *IEEE Glob. Commun. Conf.*, Taipei, Taiwan, Dec. 2020.

[38] S. Hassanpour, D. Wübben, and A. Dekorsy, "Forward-Aware Information Bottleneck-Based Vector Quantization: Multiterminal Extensions for Parallel and Successive Retrieval," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6633–6646, Oct. 2021.

[39] J. Lewandowsky and G. Bauch, "Theory and Application of the Information Bottleneck Method," *Entropy*, vol. 26, no. 3, Art. no. 187, Feb. 2024.

[40] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep Variational Information Bottleneck," in *Int. Conf. Learn. Represent.*, Toulon, France, Apr. 2017.

[41] A. Zaidi and I. Estella-Aguerri, "Distributed Deep Variational Information Bottleneck," in *IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Atlanta, GA, USA, May 2020.

[42] Q. Wang, C. Boudreau, Q. Luo, P.-N. Tan, and J. Zhou, "Deep Multi-View Information Bottleneck," in *SIAM Int. Conf. Data Min.*, Calgary, Alberta, Canada, May 2019.

[43] S. Hassanpour, M. Hummert, D. Wübben, and A. Dekorsy, "A Deep Variational Approach to Multiterminal Joint Source-Channel Coding Based on Information Bottleneck Principle," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 4462–4475, May 2025.

[44] S. Movaghati and M. Ardakani, "Distributed Channel-Aware Quantization Based on Maximum Mutual Information," *Int. J. Distrib. Sens. Netw.*, vol. 12, no. 5, Art. no. 3595389, May 2016.

[45] D. Wübben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and Impact of Cloud Computing on 5G Signal Processing: Flexible Centralization through Cloud-RAN," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 35–44, Nov. 2014.

[46] 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.

[47] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-Free Massive MIMO Versus Small Cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[48] E. Björnson and L. Sanguinetti, "Scalable Cell-Free Massive MIMO Systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[49] M. Bashar, P. Xiao, R. Tafazolli, K. Cumanan, A. G. Burr, and E. Björnson, "Limited-Fronthaul Cell-Free Massive MIMO With Local MMSE Receiver Under Rician Fading and Phase Shifts," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1934–1938, Sep. 2021.

[50] 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.

[51] 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.

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

[53] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

[54] S. Hassanpour, T. Monsees, D. Wübben, and A. Dekorsy, "Forward-Aware Information Bottleneck-Based Vector Quantization for Noisy Channels," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7911–7926, Dec. 2020.

[55] E. Jang, S. Gu, and B. Poole, "Categorical Reparameterization with Gumbel-Softmax," *arXiv 2016,* arXiv:1611.01144.

[56] C. J. Maddison, A. Mnih, and Y. W. Teh, "The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables," *arXiv 2016,* arXiv:1611.00712.

[57] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv 2017*, arXiv:1707.06347.

[58] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-Dimensional Continuous Control Using Generalized Advantage Estimation," *arXiv 2015*, arXiv:1506.02438.

[59] 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.

[60] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv 2014*, arXiv:1412.6980.

[61] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

SHAYAN HASSANPOUR (Member, IEEE) works as a Post-Doctoral Researcher with the Department of Communications Engineering at the University of Bremen, Germany. He received the M.Sc. degree in communications engineering (program's award winner) from the Ulm University, Germany, in 2014, and the Dr.-Ing. degree (with summa cum laude) in electrical engineering (communications) from the University of Bremen, Germany, in 2022. He has been prolifically contributing to IEEE top-tier journals and flagship international conferences on his PhD topic, the Information Bottleneck method. His other research interests include applied information theory, MU/MIMO systems, resilient communication networks, statistical signal processing, and applications of modern deep learning in the design of communication systems. He received the 2021's VDE/ITG Best Paper Award and the 2023's OHB Best Doctoral Dissertation Award. His research has contributed to Germany's high-profile 6G initiatives, including 6G-ANNA, Open6GHub, and 6G-TakeOff.

DIRK WÜBBEN (Senior Member, IEEE) received the Dipl.-Ing. (FH) degree in electrical engineering from the University of Applied Science Münster, Germany, in 1998, and the Dipl.-Ing. (Uni.) and Dr.-Ing. degrees in electrical engineering from the University of Bremen, Germany, in 2000 and 2005, respectively. He is currently a Senior Research Group Leader and a Lecturer with the Department of Communications Engineering, University of Bremen. He has published more than 160 papers in international journals and conference proceedings. His research interests include wireless communications, signal processing, multiple antenna systems, cooperative communication systems, channel coding, information theory, and machine learning. He is a Board Member of the Germany Chapter of the IEEE Information Theory Society and a member of VDE/ITG Expert Committee "Information and System Theory." He has been an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS. He received the VDE/ITG best paper award 2021 and the VDE/ITG certificate of honor in 2024.

ARMIN DEKORSY (Senior Member, IEEE) is a Professor with the University of Bremen, where he is the Director of the Gauss-Olbers Space Technology Transfer Center and Heads the Department of Communications Engineering. With over 11 years of industry experience, including distinguished research positions, such as a DMTS with Bell Labs and a Research Coordinator Europe with Qualcomm, he has actively participated in more than 65 international research projects, with leadership roles in 17 of them. He co-authored the textbook *Nachrichtenübertragung* (Release 6, Springer Vieweg), which is a bestseller in the field of communication technologies in German-speaking countries. His research focuses on signal processing and wireless communications for 5G/6G, industrial radio, and 3-D networks. He is a Senior Member of the IEEE Communications and Signal Processing Society and a member of the VDE/ITG Expert Committee on Information and System Theory.