Harmonizing Efficiency and Precision in Semantic-Bit Coexisting Communication Systems

Biqian Feng*‡, Xue Han*, and Chenyuan Feng‡, *Member, IEEE*
* Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang 200240, China
‡ Eurecom, Sophia Antipolis 06410, France
E-mails: {fengbiqian, han.xue}@sjtu.edu.cn, Chenyuan.Feng@eurecom.fr

***Abstract*—As semantic communication advances, the next generation (6G) communication needs to intelligently manage the emerging semantic transmission alongside existing bit transmission to ensure flexibility and efficiency. This study concentrates on the coexistence of semantic- and bit- communications within multi-cell multi-user multiple-input multiple-output (MIMO) systems, aiming to coordinate the efficiency of emerging semantic communication with the bit-level accuracy of traditional digital communication. To this end, a multi-objective optimization framework is established to bolster the efficacy of both communication paradigms while adhering to the constraints of maximum transmit power and permissible transmission delay. To simplify the complexity of this problem, we apply the ϵ-constraint method, effectively converting the multi-objective problem into a single-objective one. Subsequently, we implement a block coordinate descent (BCD) algorithm to alternately design the transceiver beamformer and the compression ratio of semantic communication alternately. Numerical results demonstrate the superiority of the proposed approach compared to existing methods. Our research sheds light on the coexistence and upgrading of future semantic-bit coexisting communications.**

***Index Terms*—Semantic communications, bit-level transmission, multi-cell multi-user MIMO, ϵ-constraint method, block coordinate descent.**

I. Introduction

With the rapid development of intelligent devices and explosive growth of multimedia applications, e.g., augmented reality/virtual reality (AR/VR), intelligent transportation, and the Internet of Everything, continue to demand faster, more reliable, and more efficient wireless communication networks. Semantic communication, as one of the promising technologies in the upcoming sixth-generation (6G) wireless communications by leveraging cutting-edge deep learning (DL) algorithms, is anticipated to satisfy the system demands for high bandwidth and ultra-low latency [1]–[3].

Different from bit-level communication which focuses on the bit-level precision between transmitters and receivers, semantic communication is dedicated to conveying the essential meaning delivery implied in the source messages. Semantic communication is more effective than bit communication in resource-limited scenarios and low signal-to-noise ratio (SNR) regimes [4]–[6]. By designing the effective joint source-channel coding (JSCC) framework which directly maps the input source signal to channel symbols, semantic communication has the potential to ease spectrum limitation by reducing transmission overhead [7].

Recent years have witnessed significant progress in semantic communication from the standpoint of wireless resource management [8]. Xia *et al.* [9] developed a bit-rate-to-message-rate transformation mechanism and a semantic-aware metric to jointly optimize user association and bandwidth allocation problems in both perfect and imperfect knowledge matching scenarios. Feng *et al.* [10] considered resource blocks and power allocation in tandem for dual-mode base station (BS), aiming to minimize energy consumption while taking into account the channel state information, maximum delay, load, and user applications. Zhang *et al.* [11] introduced a deep reinforcement learning (DRL) based dynamic resource allocation model for task-oriented semantic communication networks, and designed a deep deterministic policy gradient (DDPG) agent to maximize the long-term transmission efficiency. Han *et al.* [12] investigated a semantic-aware resource allocation problem to balance the trade-off between transmission latency and task performance in terms of transceiver beamformer design, the quantity of transmitted semantic symbols, and channel assignment. Wang *et al.* [13] modeled the text semantic information by a knowledge graph and formulated an optimization problem to maximize the total semantic similarity by optimizing the resource allocation policy and determining the partial semantic information to be transmitted. Nevertheless, it is worth noting that although these pioneering works demonstrate promising performance results, their adoption in practical systems is challenging as they require new standardized protocols and hardware design.

This motivates researchers to investigate a more practical and gradual transition approach to technology updates by exploiting the coexistence of both semantic communication and bit communication. Mu *et al.* [14] exploited a heterogeneous semantic and bit transmission framework and proposed three multiple access schemes, where an access point (AP) simultaneously sends the semantic and bit streams. Yu *et al.* [15] inserted the encoded semantic information into bit information for transmission by utilizing the same spectrum resources, and conducted experiments in a pluto-based software defined radio (SDR) platform in a real wireless channel. Zhang *et*

This work was supported in part by the National Natural Science Foundation of China under Grant 62301328.
Corresponding author: Chenyuan Feng.

al. [16] employed orthogonal/nonorthogonal multiple access techniques for serving multiple users and proposed channel & user demand-based transmission protocols where a joint optimization mode selection problem is formulated. Xia *et al.* [17] jointly investigated user association, semantic or bit mode selection, and bandwidth allocation problems to maximize the overall message throughput.

Furthermore, some researchers have proposed a semantic-bit coexistence framework, which benefits from high efficiency and low complexity [18]. The following reasons drive our proposal of this framework: First, semantic communication demonstrates remarkable performance in low and moderate SNR regimes, while bit communication outperforms in high SNR regimes [19]. It is advantageous to investigate a hybrid network that fully utilizes bit communication as well as semantic communication given their distinct characteristics and proficiencies. Second, the practical channel conditions and user requirements may change rapidly and drastically in the upcoming 6G communications. The coexistence of semantic-bit communication paradigms can provide users with flexible and enhanced service quality in resource-limited systems.

This paper investigates the optimization design of semantic-bit coexisting communications over multi-cell multi-user multiple-input multiple-output (MIMO) systems. The main contributions are summarized as follows:

- We formulate a multi-objective optimization problem to maximize the semantic similarity for semantic communication and minimize expected mean-square error (MSE) for bit communication simultaneously, subject to the maximum transmit power and maximum allowable transmission delay constraints.
- To make the problem tractable, we utilize the ϵ-constraint method to transform it into a single-objective problem, and then employ the block coordinate descent (BCD) method to optimize all variables alternately. Specifically, we propose a constrained successive convex approximation (SCA)-based algorithm to optimize the transceiver beamformer and the exhaustive search method to search the optimal length of extracted features of semantic communication.
- To validate the effectiveness of the proposed strategy and algorithm, we conducted extensive experiments using two real-world datasets for image transmission tasks. The results demonstrate that our semantic-bit coexisting communications strategy for multi-cell multi-user MIMO outperforms traditional image compression and pure JSCC-based semantic transmission strategies in terms of transmission efficiency and accuracy.

Notations: We adopt x, $\mathbf{x}$, and $\mathbf{X}$ to denote a scalar, vector, and matrix, respectively. Superscripts $*, T$, and H stand for the conjugate, transpose, and conjugate transpose, respectively. $\mathbb{C}^{m\times n}$ denotes an m by n dimensional complex space. $\mathrm{Re}(\cdot)$ denotes the real part of complex vectors or matrices. $\mathcal{CN}(\mu,\sigma^2)$ denotes the circularly symmetric complex Gaussian (CSCG) distribution with mean μ and variance σ^2.

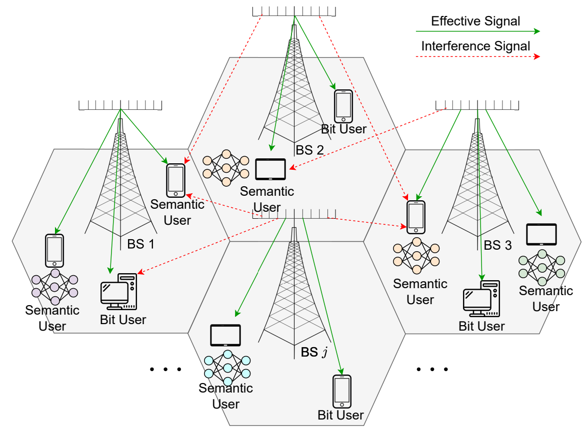


Fig. 1. Example of the proposed semantic-bit coexisting communication for the multi-cell multi-user MIMO system.

II. System Model

In this section, we describe the transmission model of the considered semantic-bit coexisting communication system. Then, a multi-objective optimization problem is formulated to maximize semantic similarity for the semantic communication paradigm and minimize the expected MSE for the bit communication paradigm simultaneously.

A. Semantic-Bit Coexisting Communication Systems

As illustrated in Fig. 1, we consider an L-cell multi-user MIMO wireless communication system, where the BS $l, l = 1, 2, \cdots, L$ equipped with N transmit antennas serves $K_l = K_l^1 + K_l^2$ single-antenna users in cell l, including K_l^1 semantic users and K_l^2 bit users. Let us define $\mathcal{K}^1$ and $\mathcal{K}^2$ to be the set of all semantic users and the set of all bit users, respectively. Then, we have $|\mathcal{K}^1| = \sum_{l=1}^{L} K_l^1$ and $|\mathcal{K}^2| = \sum_{l=1}^{L} K_l^2$. Let k_l denote the k-th user in cell l. Without loss of generality, we assume that the size of the information sources $\mathbf{f}_{k_l}$ is identical for all semantic users and for all bit users, represented by M^1 and M^2 respectively.

As shown in Fig. 2, the main process of both communication paradigms contains three parts: transmitter, MIMO channel, and receiver. Next, we will introduce each part.

Transmitter: *1) Semantic Communication Paradigm:* The variable-length JSCC encoder, $E_\theta^1(\cdot) : \mathbb{R}^{M^1\times 1} \mapsto \mathbb{C}^{M_{k_l}\times 1}$, maps the source information into the channel input symbols, i.e., $\mathbf{s}_{k_l} = E_\theta^1(\mathbf{f}_{k_l}), \forall k_l \in \mathcal{K}^1$, where M_{k_l} represents the size of extracted semantic symbol vectors for user k_l and θ denotes the trainable parameters. The varying values of M_{k_l} correspond to distinct levels of system compression ratio. Let $\mathcal{M}$ denote the set of all possible lengths of M_{k_l}.

2) Bit Communication Paradigm: The conventional separate source-channel coding (SSCC) encoder and the modulator, $E^2(\cdot) : \mathbb{R}^{M^2\times 1} \mapsto \mathbb{C}^{M\times 1}$, maps the source information into the channel input symbols, i.e., $\mathbf{s}_{k_l} = E^2(\mathbf{f}_{k_l}), \forall k_l \in \mathcal{K}^2$, where M represents the size of transmit symbol vectors after modulator.

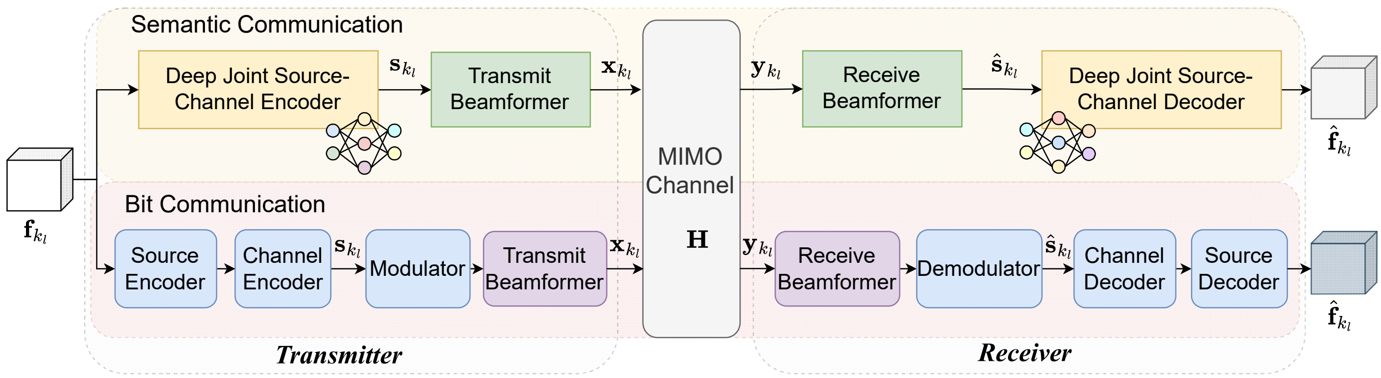


Fig. 2. Comparison between the semantic communication paradigm and the bit communication paradigm.

MIMO Channel: Let $\mathbf{v}_{k_l} \in \mathbb{C}^{N\times 1}$ denote the transmit beamformer at BS l to convey the semantic or bit signal $\mathbf{s}_{k_l}$. It is important to note that we allocate only a single data stream s_{k_l}, which is one of the elements of $\mathbf{s}_{k_l}$, for transmission at each time slot. Consequently, the transmit signal at BS l is

$$\mathbf{x}_l = \sum_{k=1}^{K_l} \mathbf{v}_{k_l} s_{k_l}, \tag{1}$$

where s_{k_l} has zero mean and $\mathbb{E}[s_{k_l} s_{k_l}^H] = 1$, the symbols desired for different users are independent of each other [16]. Thus, the discrete-time signal received at user k_l is

$$\begin{aligned} y_{k_l} &= \sum_{j=1}^{L} \mathbf{h}_{j,k_l}^H \mathbf{x}_j + n_{k_l} = \mathbf{h}_{l,k_l}^H \mathbf{v}_{k_l} s_{k_l} \\ &+ \underbrace{\mathbf{h}_{l,k_l}^H \sum_{i=1, i\neq k}^{K_l} \mathbf{v}_{i_l} s_{i_l}}_{\text{intracell interference}} + \underbrace{\sum_{j=1, j\neq l}^{L} \mathbf{h}_{j,k_l}^H \sum_{i=1}^{K_j} \mathbf{v}_{i_j} s_{i_j}}_{\text{intercell interference}} + n_{k_l}, \end{aligned} \tag{2}$$

where $\mathbf{h}_{l,k_l} \in \mathbb{C}^{N\times 1}$ denotes the channel state information (CSI) between the BS l and user k_l; $n_{k_l} \sim \mathcal{CN}(0, \sigma^2)$ models the additive noise with zero mean and variance σ^2. Consider the linear receive beamforming strategy so that the estimated signal at user k_l is given by

$$\hat{s}_{k_l} = u_{k_l}^* y_{k_l}, \tag{3}$$

where u_{k_l} represents the receive beamformer. Based on (2), the signal-to-interference-plus-noise ratio (SINR) experienced at user k_l is obtained as

$$\gamma_{k_l} \triangleq \frac{|\mathbf{h}_{l,k_l}^H \mathbf{v}_{k_l}|^2}{\sum\limits_{(i,j)\neq(k,l)} |\mathbf{h}_{j,i_j}^H \mathbf{v}_{i_j}|^2 + \sigma^2}. \tag{4}$$

Furthermore, the achievable rate and the transmission delay at user k_l are, respectively, given by

$$R_{k_l} = \log_2(1+\gamma_{k_l}), \tag{5a}$$

$$T_{k_l} = \begin{cases} \frac{M_{k_l}}{BR_{k_l}}, & k_l \in \mathcal{K}^1, \\ \frac{M}{BR_{k_l}}, & k_l \in \mathcal{K}^2, \end{cases} \tag{5b}$$

where B is the bandwidth in the system.

Then, the problem of interest is to optimize the hybrid semantic and bit transmission scheme detailed later, subject to the maximum transmit power $P_{\max}, \forall l$ and maximum transmission delay $T_{k_l}^{\max}, \forall k_l$:

$$\sum_{k=1}^{K_l} \|\mathbf{v}_{k_l}\|^2 \leq P_{\max}, \forall l, \tag{6a}$$

$$T_{k_l} \leq T_{k_l}^{\max}, \forall k_l. \tag{6b}$$

Combining Eqs. (4), (5b), and (6b), we can derive that the maximum transmission delay constraint is equivalent to

$$\gamma_{k_l} \geq \gamma_{k_l}^{\min} \triangleq \begin{cases} 2^{M_{k_l}/(BT_{k_l}^{\max})} - 1 & k_l \in \mathcal{K}^1, \\ 2^{M/(BT_{k_l}^{\max})} - 1 & k_l \in \mathcal{K}^2. \end{cases} \tag{7}$$

Receiver: After receiving all $\hat{s}_{k_l}$ and combining them into $\hat{\mathbf{s}}_{k_l}$, the decoding process of both communication paradigms is, respectively, described as follows:

1) Semantic Communication Paradigm: The variable-length JSCC decoder, $D_\theta^1(\cdot) : \mathbb{C}^{M_{k_l}\times 1} \mapsto \mathbb{R}^{M^1\times 1}$, maps the channel output symbols into the reconstructed information, i.e., $\hat{\mathbf{f}}_{k_l} = D_\theta^1(\hat{\mathbf{s}}_{k_l}), \forall k_l \in \mathcal{K}^1$, where θ denotes the trainable parameters at the receiver.

2) Bit Communication Paradigm: The conventional SSCC decoder and the demodulator, $D^2(\cdot) : \mathbb{C}^{M\times 1} \mapsto \mathbb{R}^{M^2\times 1}$, maps the source information into the channel input symbols, i.e., $\hat{\mathbf{f}}_{k_l} = D^1(\hat{\mathbf{s}}_{k_l}), \forall k_l \in \mathcal{K}^2$.

B. System Performance Metrics

To evaluate the performance of semantic and bit communication, we adopt different metrics:

1) Semantic Similarity for Semantic Communication: The semantic similarity between the source information $\mathbf{f}_{k_l}$ and reconstructed information $\hat{\mathbf{f}}_{k_l}$ is defined as [19]:

$$\begin{aligned} \xi_{k_l} &= \mathbb{E}\left[\frac{B_\lambda(\mathbf{f}_{k_l}) B_\lambda(\hat{\mathbf{f}}_{k_l})^T}{\|B_\lambda(\mathbf{f}_{k_l})\| \|B_\lambda(\hat{\mathbf{f}}_{k_l})\|}\right] \\ &\approx \frac{1}{I}\sum_{i=1}^{I} \frac{B_\lambda(\mathbf{f}_{k_l}^{(i)}) B_\lambda(\hat{\mathbf{f}}_{k_l}^{(i)})^T}{\|B_\lambda(\mathbf{f}_{k_l}^{(i)})\| \|B_\lambda(\hat{\mathbf{f}}_{k_l}^{(i)})\|}, \end{aligned} \tag{8}$$

where $B_\lambda(\cdot)$ represents the downstream task execution module, and we utilize the cosine similarity of the inference outcomes to quantify the feature semantic similarity. The greater the semantic similarity, the more task-relevant information is retained in the reconstructed features. However, calculating the true feature semantic similarity is often impractical in real-world scenarios. Instead, this can be effectively approximated with the Sample Average Approximation (SAA) method. Here, I denotes the total number of samples, and the pair of $\mathbf{f}_{k_l}$ and $\hat{\mathbf{f}}_{k_l}$ represents an individual sample within the dataset.

Remark 1: Most neural networks can be trained with gradient-based methods as the gradients with respect to the network's parameters can be effectively computed with automatic differentiation tools. That implies $\frac{\partial \xi_{k_l}}{\partial u_{k_l}}$ and $\frac{\partial \xi_{k_l}}{\partial \mathbf{v}_{k_l}}$ can be effectively computed.

2) Expected MSE for Bit Communication: The conventional bit communication is concentrated on recovering the accuracy of information, which is usually measured by MSE [20]

$$
\begin{aligned}
e_k &\triangleq \mathbb{E}\left[|\hat{s}_{k_l} - s_{k_l}|^2\right] \\
&= \sigma^2 |u_{k_l}|^2 + 1 + \sum_{i,j} |u_{k_l}^* \mathbf{h}_{j,k_l}^H \mathbf{v}_{i_j}|^2 - 2\mathrm{Re}(u_{k_l}^* \mathbf{h}_{l,k_l}^H \mathbf{v}_{k_l}).
\end{aligned} \tag{9}
$$

Remark 2: The MSE function exhibits element-wise convexity with respect to $u_{k_l}^*$ and $\mathbf{v}_{k_l}$, yet it is not jointly convex.

C. Problem Formulation

In this paper, we aim to maximize the semantic similarity for semantic users and minimize the expected MSE for bit users by jointly designing the transceiver beamformer and the compression ratio, subject to the SINR requirements and the transmit power constraints. Let $\mathbf{v} \triangleq (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{K_L})$, $\mathbf{u} \triangleq (u_1, u_2, \cdots, u_{K_L})$, $\mathbf{M} \triangleq (M_1, M_2, \cdots, M_{K_L})$, $\boldsymbol{\xi} \triangleq (\xi_1, \xi_2, \cdots, \xi_{K_L})$, and $\mathbf{e} \triangleq (e_1, e_2, \cdots, e_{K_L})$. The joint performance maximization problem is formulated as

$$
\begin{aligned}
\max_{\mathbf{u},\mathbf{v},\mathbf{M}} \quad & \{\boldsymbol{\xi}, \mathbf{e}\}, && \text{(10a)} \\
\text{s.t.} \quad & \sum_{k=1}^{K_l} \|\mathbf{v}_{k_l}\|^2 \le P_{\max}, \forall l, && \text{(10b)} \\
& \gamma_{k_l} \ge \gamma_{k_l}^{\min}, \forall k_l, && \text{(10c)} \\
& M_{k_l} \in \mathcal{M}, \forall k_l \in \mathcal{K}^1. && \text{(10d)}
\end{aligned}
$$

Transforming a multi-objective optimization problem into a single-objective one is a priori approach, essentially creating a single-objective optimization problem whose optimal solutions correspond to the Pareto optimal solutions of the original multi-objective problem. In this study, we utilize the ϵ-constraint method as our scalarization technique [21]:

$$
\begin{aligned}
\max_{u_{k_l},\mathbf{v}_{k_l},M_{k_l}} \quad & \sum_{k_l \in \mathcal{K}^1} \xi_{k_l}, && \text{(11a)} \\
\text{s.t.} \quad & \sum_{k=1}^{K_l} \|\mathbf{v}_{k_l}\|^2 \le P_{\max}, \forall l, && \text{(11b)} \\
& \gamma_{k_l} \ge \gamma_{k_l}^{\min}, \forall k_l, && \text{(11c)} \\
& M_{k_l} \in \mathcal{M}, \forall k_l \in \mathcal{K}^1, && \text{(11d)} \\
& e_{k_l} \le e_{\max}, \forall k_l \in \mathcal{K}^2, && \text{(11e)}
\end{aligned}
$$

where the weight $e_{\max}$ represents the maximum allowable MSE. The problem is challenging mainly due to the nonconvexity of the objective function (11a) and the constraints (11c)-(11e), which is an obstacle for the development of an optimal solution.

III. Block Coordinate Descent Method for Semantic-Bit Coexisting Systems

The BCD method decomposes the optimization variables into multiple blocks. During each iteration, it optimizes one block while keeping the other blocks fixed.

A. Transmit Beamformer Design

According to the principle of BCD, we first investigate the optimization of transmit beamformer $\mathbf{v}_{k_l}$ for fixed M_{k_l} and receive beamformer u_{k_l}. The transmit beamformer design problem accordingly reduced to

$$
\begin{aligned}
\max_{\mathbf{v}_{k_l}} \quad & \sum_{k_l \in \mathcal{K}^1} \xi_{k_l}, && \text{(12a)} \\
\text{s.t.} \quad & \sum_{k=1}^{K_l} \|\mathbf{v}_{k_l}\|^2 \le P_{\max}, \forall l, && \text{(12b)} \\
& \gamma_{k_l} \ge \gamma_{k_l}^{\min}, \forall k_l, && \text{(12c)} \\
& e_{k_l} \le e_{\max}, \forall k_l \in \mathcal{K}^2. && \text{(12d)}
\end{aligned}
$$

Next, we propose to leverage the constrained SCA [22] technique to update the transmit beamformer. The key to the success of constrained SCA lies in constructing a surrogate function satisfying gradient consistency for the objective function (12a) and a surrogate function satisfying the upper bound property for constraint (12c). On one hand, we adopt the first-order expansion to approximate (12a), and the surrogate function is constructed at $\mathbf{v}_{k_l} = \mathbf{v}_{k_l}^{(n)}$ as follows:

$$
\tilde{f}(\mathbf{v}_{k_l}, \mathbf{v}_{k_l}^{(n)}) \triangleq \sum_{k_l \in \mathcal{K}^1} -\nabla \xi_i^{(n)} \left(\mathbf{v}_{k_l} - \mathbf{v}_{k_l}^{(n)}\right). \tag{13}
$$

On the other hand, the constraint (12c) is equivalent to

$$
g(\mathbf{v}_{k_l}) \triangleq \sum_{(i,j) \ne (k,l)} |\mathbf{h}_{j,i_j}^H \mathbf{v}_{i_j}|^2 + \sigma^2 - \gamma_{k_l}^{\min} |\mathbf{h}_{l,k_l}^H \mathbf{v}_{k_l}|^2 \le 0, \tag{14}
$$

and we adopt partial linearization [23, Eq. (10)] to approximate it, i.e.,

$$
\begin{aligned}
\tilde{g}(\mathbf{v}_{k_l}, \mathbf{v}_{k_l}^{(n)}) \triangleq & \sum_{(i,j) \ne (k,l)} |\mathbf{h}_{j,i_j}^H \mathbf{v}_{i_j}|^2 + \sigma^2 \\
& - 2\gamma_{k_l}^{\min} \mathbf{Re}((\mathbf{v}_{k_l}^{(n)})^H \mathbf{h}_{l,k_l} \mathbf{h}_{l,k_l}^H \mathbf{v}_{k_l}),
\end{aligned} \tag{15}
$$

Therefore, the surrogate problem is given by

$$\hat{\mathbf{v}}_{k_l}^{(n)} \triangleq \arg\min_{\mathbf{v}_{k_l}} \quad \tilde{f}(\mathbf{v}_{k_l}, \mathbf{v}_{k_l}^{(n)}) \tag{16a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K_l} \|\mathbf{v}_{k_l}\|^2 \le P_{\max}, \forall l, \tag{16b}$$

$$\tilde{g}_{k_l}(\mathbf{v}_{r_l}, \mathbf{v}_{r_l}^{(n)}) \le 0, \forall k_l \in \mathcal{K}^1, \tag{16c}$$

$$e_{k_l} \le e_{\max}, \forall k_l \in \mathcal{K}^2, \tag{16d}$$

and the transmit beamformer $\mathbf{v}_{k_l}^{(n+1)}$ is updated according to

$$\mathbf{v}_{k_l}^{(n+1)} = \mathbf{v}_{k_l}^{(n)} + \gamma_{k_l}(\hat{\mathbf{v}}_{k_l}^{(n)} - \mathbf{v}_{k_l}^{(n)}), \tag{17}$$

where γ_{k_l} is the iterative step size for updating $\mathbf{v}_{k_l}$. It is evident that the problem is also a convex quadratic programming (QP) problem, which can be effectively solved by the CVXPY.

B. Receive Beamformer Design

For fixed M_{k_l} and transmit beamformer $\mathbf{v}_{k_l}$, the problem of designing receive beamformer accordingly reduces to

$$\max_{u_{k_l}} \quad \sum_{k_l \in \mathcal{K}^1} \xi_{k_l} \tag{18a}$$

$$\text{s.t.} \; e_{k_l} \le e_{\max}, \forall k_l \in \mathcal{K}^2. \tag{18b}$$

Similar to transmit beamformer design, the surrogate function for objective function is constructed as $u_{k_l} = u_{k_l}^{(n)}$ as follows:

$$\tilde{f}(u_{k_l}, u_{k_l}^{(n)}) \triangleq \sum_{k_l \in \mathcal{K}^1} -\nabla \xi_i^{(n)} \left(u_{k_l} - u_{k_l}^{(n)} \right), \tag{19}$$

Then, the surrogate problem is given by

$$\hat{u}_{k_l}^{(n)} \triangleq \arg\min_{\mathbf{v}_{k_l}} \quad \tilde{f}(u_{k_l}, u_{k_l}^{(n)}) \tag{20a}$$

$$\text{s.t.} \;\; e_{k_l} \le e_{\max}, \forall k_l \in \mathcal{K}^2, \tag{20b}$$

and the receive beamformer $u_{k_l}^{(n+1)}$ is updated according to

$$u_{k_l}^{(n+1)} = u_{k_l}^{(n)} + \gamma_{k_l}(\hat{u}_{k_l}^{(n)} - u_{k_l}^{(n)}), \tag{21}$$

where γ_{k_l} is the iterative step size for updating u_{k_l}. It is evident that the problem is also a convex QP problem, which can be effectively solved by the CVXPY.

C. Compression Ratio Design

For fixed transceiver beamformer u_{k_l} and $\mathbf{v}_{k_l}$, the problem of designing compression ratio accordingly reduces to

$$\max_{M_{k_l}} \quad \sum_{k_l \in \mathcal{K}^1} \xi_{k_l}, \tag{22a}$$

$$\text{s.t.} \; \gamma_{k_l} \ge \gamma_{k_l}^{\min}, \forall k_l \in \mathcal{K}^1, \tag{22b}$$

$$M_{k_l} \in \mathcal{M}, \forall k_l \in \mathcal{K}^1. \tag{22c}$$

We utilize an exhaustive search algorithm to address the above problem. It is important to highlight that the computational complexity is acceptable, attributed to the finite set $\mathcal{M}$. This approach ensures a meticulous examination of all potential solutions, thereby delivering optimal performance without incurring excessive computational demands.

The details of the proposed algorithm are presented in Algorithm 1.

Algorithm 1 BCD Method for Semantic-Bit Coexisting Communication Systems

Input: All channel matrices $\mathbf{h}_{l,k_j}$, $\gamma_{k_l}^{\min}$, and $e_{\max}$. **Output:** $\mathbf{v}_{k_l}$, u_{k_l}, and M_{k_l}.

```
for κ = 1, 2, ... do
    for n = 1, 2, ... do
        Given v_{k_l}^{(n)} and construct the surrogate functions w.r.t. v_{k_l} according to Eq. (19).
        Solve the surrogate problem (20).
        Update transmitter beamformer v_{k_l}^{(n+1)} according to Eq. (21).
    end for
    for n = 1, 2, ... do
        Given u_{k_l}^{(n)} and construct the surrogate functions w.r.t. u_{k_l} according to Eqs. (13) and (15).
        Solve the surrogate problem (16).
        Update transmitter beamformer u_{k_l}^{(n+1)} according to Eq. (17).
    end for
    Search the optimal compression ratio in the problem (22).
end for
```

D. Convergence and Complexity Analysis

Convergence analysis: Based on the constrained SCA algorithm for transceiver beamformer design and the exhaustive algorithm for compression ratio design, a stationary point is guaranteed for each subproblem. Therefore, the initial equivalent function is guaranteed to increase monotonically, and our proposed algorithm is convergent.

Complexity analysis: The complexity of updating $\mathbf{v}_{k_l}$ and u_{k_l} at each iteration of the subproblem is $\mathcal{O}((|\mathcal{K}^1|+|\mathcal{K}^2|)^2 N^2)$ and $\mathcal{O}((|\mathcal{K}^1|+|\mathcal{K}^2|)^2)$. The complexity of updating M_{k_l} at each iteration of the subproblem is $O(|\mathcal{M}|)$.

IV. NUMERICAL RESULTS

In this section, we numerically evaluate the performance of the proposed algorithm in a 4-cell multi-user MIMO system.

A. Experimental Setups

1) Datasets: For wireless image transmission, we adopt the Cityscapes [24] dataset for training and the CVRG-Pano [25] dataset for testing. Cityscapes is a street scene images dataset that is captured under similar weather and lighting conditions in 50 different locations. CVRG-Pano dataset is a 360° outdoor panoramic image dataset including 600 RGB images with a resolution of 1664×832.

2) Parameters and Deployment Details: Our experiment considers the multi-cell multi-user MIMO system with four users, two transmitter antennas, and two receiver antennas. The path loss is modeled as 128.1 + 37.6 log[d(km)] dB with the standard deviation of shadow fading 6 dB. The power spectral density of Gaussian noise is -174 dBm/Hz and the total bandwidth is 5 MHz. For model training, we utilize a variable learning rate, which decreases step-by-step from 1e-4 to 2e-5. The experiments are conducted on RTX4090 GPUs. The whole framework is optimized with Adam [26]. The batch size is set as 16.

3) Comparison Benchmarks: The semantic-bit coexisting system adopts DeepJSCC [27], a pioneering framework of JSCC based on autoencoder for end-to-end wireless image transmission. We compare the proposed design with the following baselines:

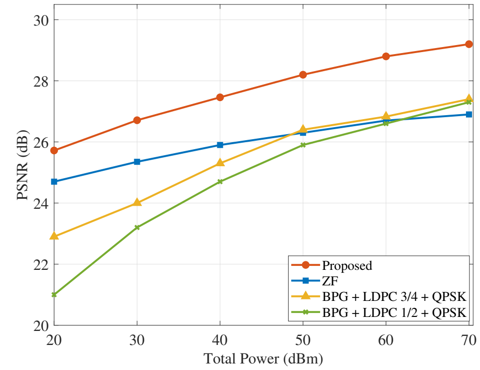
- ZF: This baseline scheme utilizes commonly adopted zero-forcing (ZF) and water-filling algorithm as beamforming, and maintains the DeepJSCC network unchanged.
- BPG + LDPC + QPSK: For the separate source-channel coding scheme, the benchmark employs BPG for source coding, low-density parity-check (LDPC) for channel coding, QPSK modulation, and ZF for beamforming. Here, we consider LDPC with code rates 3/4 and 1/2, which are denoted by "BPG + 3/4LDPC + QPSK" and "BPG + 1/2LDPC + QPSK", respectively.

4) Evaluation Metrics: We utilize the widely used pixel-wise metric peak signal-to-noise (PSNR) and the perceptual metric multi-scale structural similarity index (MS-SSIM) [28] as measurements for the reconstructed image quality. Higher PSNR and MS-SSIM values indicate better reconstruction performance.

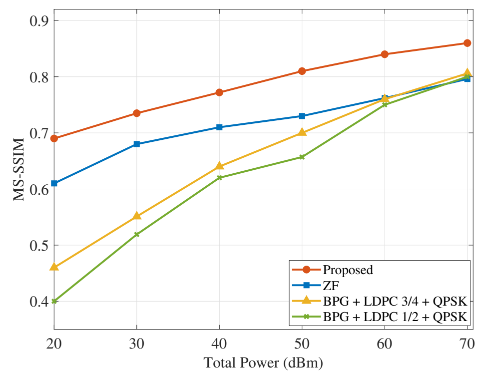
B. Results Analysis

1) Performances for Different Power Budget: We first evaluate the image reconstruction quality of the semantic-bit coexisting scheme and other baselines with 8 users. As shown in Fig. 3(a), it can be observed that PSNR performance improves with the increase of assignable power for all schemes. This is because when sufficient transmission power is available, both semantic and conventional codec schemes can achieve significant coding gain. In addition, for different power levels, the proposed scheme consistently surpasses the benchmarks, which indicates the superiority of the proposed beamforming strategy. For example, when the power = 50 dBm, the proposed scheme achieves about 2 dB performance gains compared to the proposed scheme with the traditional ZF algorithm. In Fig. 3(b), the coexisting scheme achieve better reconstruction results in terms of MS-SSIM compared to the traditional digital coding schemes. This result indicates that the satisfying visual perception quality is ensured through extracting semantics inside the original image signals where the DL-based semantic communication preserves even more high-frequency image details compared to the bit-only scheme.

2) Performances for Different Number of Users: Fig. 4 demonstrates the performance versus different user numbers at a fixed total power $P_{\max} = 60dBm$ to demonstrate a better view of the relationship between proposed algorithm performance and user numbers. As shown in Fig. 4(a), it is observed that the coexisting scheme generally outperforms bit-only schemes in all user numbers for PSNR metric in MIMO transmission scenarios. This is because given a fixed power for image transmission, the semantic-bit coexisting system can provide more resources to the high-quality demand users given the optimized transmit symbol values. Specifically, in Fig. 4(b), the traditional beamforming scheme perform inferior to the proposed algorithm which implies that special design of beamforming for coexisting systems is required. The proposed algorithm outperforms the three benchmark schemes in all user number settings. According to the optimized strategy, the proposed framework can adaptively adjust the source and channel coding rate based on deep JSCC structure, ensuring that it can obtain efficient performance gains. As the user number increases, the coexisting scheme still achieves relatively satisfying performances compared to other benchmarks, which validates the superiority of the proposed algorithm for efficient data compression and transmission.



(a) PSNR versus Total Power

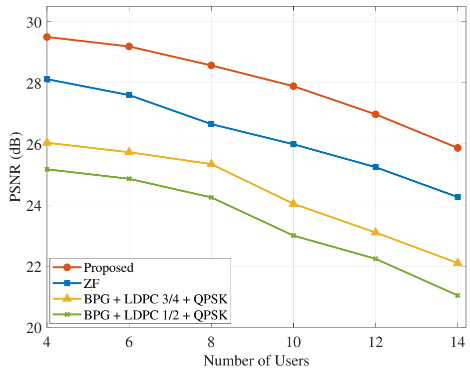


(b) MS-SSIM versus Total Power

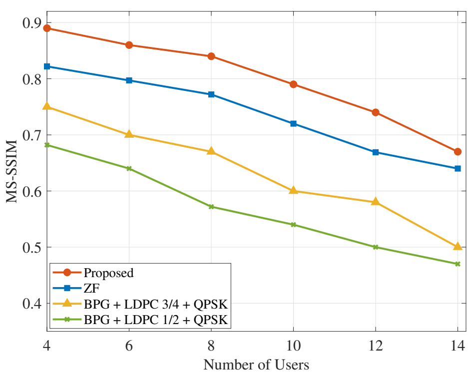
Fig. 3. Performance comparison of the reconstructed image quality under different total power over the CVRG-Pano dataset.

V. Conclusion

In this paper, we focus on semantic-bit coexisting communications of multi-cell multi-user MIMO systems and aim to enhance the system performance metrics, i.e., semantic similarity for semantic transmission and MSE for bit transmission. We first establish a multi-objective optimization framework to enhance the performance of both semantic and bit communication paradigms for multi-cell multi-user MIMO systems. Furthermore, we employ the ϵ-constraint method to simplify the



(a) PSNR versus The Number of Users



(b) MS-SSIM versus The Number of Users

Fig. 4. Performance comparison of the reconstructed image quality under different numbers of users over the CVRG-Pano dataset.

multi-objective problem into a single-objective one, facilitating the design of a BCD algorithm for the alternating optimization of the transceiver beamformer and the compression ratio of semantic communication. Simulation results confirmed the huge potential of semantic-bit coexisting communication in improving the performance of future communications systems.

References

[1] M. Kountouris and N. Pappas, ”Semantics-Empowered Communication for Networked Intelligent Systems,” *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 96-102, June 2021.

[2] P. Agheli, N. Pappas and M. Kountouris, “Goal-Oriented Multiple Access Connectivity for Networked Intelligent Systems,” *IEEE Commun. Lett.*, vol. 28, no. 8, pp. 1795-1799, Aug. 2024.

[3] X. Wang, D. Ye, C. Feng, H. H. Yang, X. Chen, and T. Q. S. Quek, “Trustworthy Image Semantic Communication with GenAI: Explainablity, Controllability, and Efficiency,” *Available as ArXiv:2408.03806v1*, Aug. 2024.

[4] Y. Wang, S. Guo, Y. Deng, H. Zhang and Y. Fang, “Privacy-Preserving Task-Oriented Semantic Communications Against Model Inversion Attacks,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 10150-10165, Aug. 2024.

[5] S. Guo, Y. Wang, S. Li and N. Saeed, “Semantic Importance-Aware Communications Using Pre-Trained Language Models,” *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2328-2332, Sept. 2023.

[6] A. Zhang, Y. Wang and S. Guo, ”On the Utility-Informativeness-Security Trade-Off in Discrete Task-Oriented Semantic Communication,” *IEEE Commun. Lett.*, vol. 28, no. 6, pp. 1298-1302, June 2024.

[7] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, 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. 2022.

[8] S. Guo, A. Zhang, Y. Wang, C. Feng, and T. Q. S. Quek, “Trustworthy Semantic-Enabled 6G Communication: A Task-oriented and Privacy-preserving Perspective,” *Available as ArXiv: 2408.04188*, Aug. 2024.

[9] L. Xia, Y. Sun, D. Niyato, X. Li, and M. A. Imran, “Joint user association and bandwidth allocation in semantic communication networks,” *IEEE Trans. Veh. Technol.*, pp. 1–13, 2023.

[10] B. Feng, Z. Liao, Y. Wu, J. Jin, D. W. Kwan Ng, X.-G. Xia, and X. Gong, “Joint millimeter wave and microwave wave resource allocation design for dual-mode base stations,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, 2019, pp. 1–6.

[11] H. Zhang, H. Wang, Y. Li, K. Long, and A. Nallanathan, “DRL-driven dynamic resource allocation for task-oriented semantic communication,” *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 3992–4004, 2023.

[12] X. Han, B. Feng, Y. Shi, Y. Wu, and W. Zhang, “Semantic-aware resource allocation for wireless image transmission,” in *Proc. IEEE Int. Conf. Commun. China (ICCC)*, Hangzhou, China, Jul. 2024, pp. 1–6.

[13] Y. Wang, M. Chen, W. Saad, T. Luo, S. Cui, and H. V. Poor, “Performance optimization for semantic communications: An attention-based learning approach,” in *IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, 2021, pp. 1–6.

[14] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, ”Heterogeneous semantic and bit communications: A semi-NOMA scheme,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 155–169, Jan. 2023.

[15] K. Yu, R. Fan, G. Wu, and Z. Qin, “Hybrid bit and semantic communications,” *Available as ArXiv: 2404.19477*, Apr. 2024.

[16] M. Zhang, R. Zhong, X. Mu, and Y. Liu, “Machine learning enabled heterogeneous semantic and bit Communication,” *IEEE Trans. Wireless Commun.*, early access, May 2024.

[17] L. Xia, Y. Sun, D. Niyato, L. Zhang, and M. A. Imran, “Wireless resource optimization in hybrid semantic/bit communication networks,” *Available as ArXiv: 2404.04162*, Aug. 2024.

[18] W. Yang, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Cao, and K. B. Letaief, “Semantic communication meets edge intelligence,” *IEEE Wireless Commun.*, vol. 29, no. 5, pp. 28–35, Oct. 2022.

[19] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, “Deep learning enabled semantic communication systems,” *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, Apr. 2021.

[20] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[21] K. Miettinen, *Nonlinear Multiobjective Optimization*, Springer, ISBN 978-0-7923-8278-2, May 2012.

[22] G. Scutari, F. Facchinei, and L. Lampariello, “Parallel and distributed methods for constrained nonconvex optimization–part I: Theory,” *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1929–1944, Apr. 2017.

[23] Y. Sun, P. Badu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[24] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 3213–3223.

[25] S. Orhan and Y. Bastanlar, “Semantic segmentation of outdoor panoramic images,” *Signal Image Video Process.*, vol. 16, pp. 643–650, Aug. 2021.

[26] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learn. Rep. (ICLR)*, San Diego, CA, USA, 2015, pp. 1–15.

[27] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, “Deep joint source-channel coding for wireless image transmission,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[28] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *IEEE Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2003, pp. 1398–1402.