Deep Joint Source-Channel Coding for CSI Feedback: An End-to-End Approach

Jialong Xu, *Student Member, IEEE*, Bo Ai, *Fellow, IEEE*, Ning Wang, *Member, IEEE*, and Wei Chen, *Senior Member, IEEE*

***Abstract*— The increased throughput brought by MIMO technology relies on the knowledge of channel state information (CSI) acquired in the base station (BS). To make the CSI feedback overhead affordable for the evolution of MIMO technology (e.g., massive MIMO and ultra-massive MIMO), deep learning (DL) is introduced to deal with the CSI compression task. In traditional communication systems, the compressed CSI bits is treated equally and expected to be transmitted accurately over the noisy channel. While the errors occur due to the limited bandwidth or low signal-to-noise ratios (SNRs), the reconstruction performance of the CSI degrades drastically. As a branch of semantic communications, deep joint source-channel coding (DJSCC) scheme performs better than the separate source-channel coding (SSCC) scheme—the cornerstone of traditional communication systems—in the limited bandwidth and low SNRs. In this paper, we propose a DJSCC based framework for the CSI feedback task. In particular, the proposed method can simultaneously learn from the CSI source and the wireless channel. Instead of truncating CSI via Fourier transform in the delay domain in existing methods, we apply non-linear transform networks to compress the CSI. Furthermore, we adopt an SNR adaption mechanism to deal with wireless channel variations. The extensive experiments demonstrate the validity, adaptability, and generality of the proposed framework.**

***Index Terms*— CSI feedback, deep joint source-channel coding, autoencoder, deep learning.**

Manuscript received 7 April 2022; revised 2 July 2022; accepted 7 September 2022. Date of publication 17 November 2022; date of current version 19 December 2022. This work was supported in part by the Natural Science Foundation of China under Grant 62122012, Grant 61961130391, Grant U1834210, and Grant 62171380; in part by the Beijing Natural Science Foundation under Grant L202019 and Grant L211012; in part by the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2021ZZ004; in part by the Fundamental Research Funds for the Central Universities under Grant 2022JBQY004; and in part by the Royal Society Newton Advanced Fellowship under Grant NA191006. *(Corresponding authors: Bo Ai; Wei Chen.)*

Jialong Xu is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China, and also with the Frontiers Science Center for Smart High-Speed Railway System, Beijing Jiaotong University, Beijing 100044, China (e-mail: jialongxu@bjtu.edu.cn).

Bo Ai is with the State Key Laboratory of Rail Traffic Control and Safety, the Beijing Engineering Research Center of High-Speed Railway Broadband Mobile Communications, and the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China, also with the Peng Cheng Laboratory, Research Center of Networks and Communications, Shenzhen 518055, China, and also with the Henan Joint International Research Laboratory of Intelligent Networking and Data Analysis, Zhengzhou University, Zhengzhou 450001, China (e-mail: boai@bjtu.edu.cn).

Ning Wang is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: ienwang@zzu.edu.cn).

Wei Chen is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China, and also with the Key Laboratory of Railway Industry of Broadband Mobile Information Communications, Beijing Jiaotong University, Beijing 100044, China (e-mail: weich@bjtu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3221963.

Digital Object Identifier 10.1109/JSAC.2022.3221963

I. Introduction

SINCE Shannon capacity limit is gradually approached by the advanced channel coding schemes (e.g., LDPC and Polar), it is difficult to get coding gain when considering the technical level of communications, where bits are treated equally and expected to be transmitted accurately. Beyond the technical level is the semantic level, which transmits the meaning of the information rather than the bitstream of the information in the communication process [1]. Some pioneering works build semantic communications upon the logical probability, which limits semantic communication mainly in the text domain [2], [3], [4]. Driven by the artificial intelligence (AI) technologies, especially the deep learning (DL) technology [5], semantic communications are compatible with various sources, e.g., image sources and video sources, with the advantages of reducing communication bandwidth and improving communication robustness [6].

To serve the emerging intelligent applications with tremendous communication throughput, e.g., holographic communication, metaverse, and brain-to-computer interaction, the sixth generation (6G) communication system should support ultra-large bandwidth and ultra-high spectral efficiency (SE). As the exploitation of the frequencies beyond mmWave in 6G [7], the available bandwidth further increases. However, propagation losses become more severe at higher millimeter wave (mmWave) and terahertz (THz) frequencies [8], which will cause either more power consumption or a smaller coverage radius. To improve SE, Multiple-input multiple-output (MIMO) and its successors, i.e., massive MIMO and ultramassive MIMO, have been and will be main technologies in 4G, 5G and 6G, respectively.

A typical massive MIMO scenario is that the base station (BS) with a large number of antennas can simultaneously serve multiple user equipments (UEs). To sufficiently exploit a large number of antennas, the BS needs to have the knowledge of the instantaneous downlink channel state information (CSI). In time-division duplexing (TDD) mode, the BS estimates the uplink CSI through the pilot signals transmitted by the UE and then the downlink CSI can be inferred from the uplink CSI by using channel reciprocity. In frequency-division duplexing

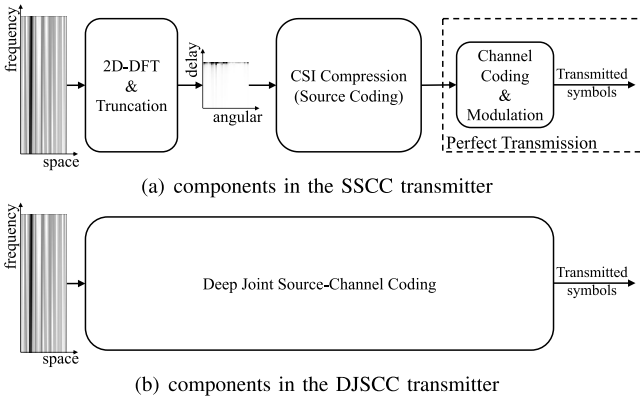


Fig. 1. Overviews of the separate source-channel coding (SSCC) scheme and the deep joint source-channel coding (DJSCC) scheme.

(FDD) mode, since the uplink and downlink work on different frequencies, channel reciprocity is no longer satisfied. A three-step interaction is applied to acquire the downlink CSI in FDD: (1) the BS first transmits pilot signals to the UE, then (2) the UE estimates the downlink CSI according to the pilot signals, and lastly (3) the UE feeds back the estimated downlink CSI to the BS. This CSI feedback mechanism inevitably occupies part of uplink resources and reduces the amount of remaining resources for data transmission.

To reduce the feedback overhead, the codebook-based method which feeds back the codebook index instead of downlink CSI is adopted in 5G (Type I and enhanced Type II) [9]. However, the increasing number of MIMO antennas from 5G to beyond 5G or 6G (e.g., from massive MIMO to ultra-massive MIMO) expands the codebook space and simultaneously aggravates the feedback overhead. It is key to compress CSI to an affordable quantity by exploiting the channel characteristics. The compressive sensing (CS) based methods transform the CSI to a sparse representation in some bases and feed back the sparse representation to reduce the feedback overhead [10], [11], [12]. While the sparse assumption—the prerequisite of CS-based method—is not strictly satisfied in the practical system. Moreover, the iterative approaches in the existing CS-based reconstruction algorithms are usually time-consuming.

Although DL has been also introduced to CSI compression [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], most of these works are regarded as the source coding module in the separate source channel coding (SSCC) scheme, which is adopted by traditional communication systems. As shown in Fig. 1(a), the UE transform the CSI from the spatial-frequency domain to the angular-delay domain and truncates the transformed CSI to reduce CSI feedback overhead at first. Then the UE compresses the truncated CSI in the angular-delay domain into codewords by the encoder. The channel coding module and modulation module in the SSCC scheme are assumed to be perfect, i.e., adaptable according to the feedback channel quality, and all feedback codewords can be successfully transmitted by the UE. In the BS side, the BS decodes the codewords and reconstructs the CSI in the spatial-frequency domain. However, there are several drawbacks. Firstly, this SSCC scheme has been demonstrated inferior to the joint source-channel coding (JSCC) scheme in the finite blocklength regime in theory [24]. Secondly, the SSCC scheme has the "cliff effect" [25] in the real wireless scenario. That means the reconstruction quality of the CSI drops drastically, if the real feedback channel condition is worse than expected, and beyond the capability of the applied channel coding scheme. In this case, the recovered CSI at the BS is useless for the subsequent process. The JSCC scheme can provide a graceful performance degradation even the real channel condition becomes worse than the expected channel condition, which makes the recovered CSI still valuable for the subsequent process. Lastly, even though the hybrid automatic repeat request (HARQ) mechanism [26] can compensate for errors of channel decoding caused by channel condition mismatch, HARQ inevitably increases the additional feedback overhead and brings the latency problem for the CSI feedback task. In all, the inferior performance, the "cliff effect", and the additional feedback overhead in variational channels of the SSCC scheme promote researchers to embrace the JSCC scheme.

The channel characteristic of MIMO depends on the scattering environment, the frequency, the speed of the UE and the BS, etc. It is too complicated to model the CSI source and the feedback channel accurately, which hinders the use of classical JSCC schemes. As an end-to-end paradigm of semantic communications and task-oriented communications [27], DL based JSCC (DJSCC) methods have shown reduced signal distortion in comparison to the SSCC methods with the use of the same amount of spectrum and energy resources [28], [29], [30], [31], [32], [33], [34]. In this paper, we design a general DJSCC framework for CSI feedback as shown in Fig. 1(b). The proposed framework focuses on efficiently executing the CSI feedback task over a noisy communication channel instead of considering accurately transmitting bits in traditional SSCC based communication systems. The major contributions are summarized as follows:

- To deal with the CSI feedback task, we propose a general DJSCC based framework consisting of the nonlinear transform module, the DJSCC module and the SNR adaption module, which can overcome the problem of "cliff effect", avoid the latency brought by channel mismatch, and most importantly, lead to considerable performance.
- We design two DL-based transform networks—one for converting the CSI from the spatial-frequency domain to a transform domain and the other for converting the CSI from a transform domain to the spatial-frequency domain—for the nonlinear transform module, which can reduce the distortion brought by dimensionality reduction in the existing SSCC based CSI feedback methods.
- We design a unified DJSCC module, where the architectures of existing CSI compression networks are easily upgraded to a DJSCC based network. Benefit from the end-to-end training, the encoder network and the decoder network simultaneously learn from the CSI source and the wireless channel.

- We introduce an SNR adaption strategy to match channel variations for DJSCC based CSI feedback. Instead of training multiple DJSCC networks with various SNRs, this strategy only needs to train a single DJSCC network with various SNRs. During network deployment in real wireless communication systems, it can save the storage space and reduce the computational complexity at the UE side and the BS side.

The rest of this paper is organized as follows. Section II presents related work on DJSCC and DL based CSI feedback. Section III introduces the system model for the end-to-end CSI feedback. Section V overviews the separate source-channel coding framework based on the existing CSI methods and proposes a DJSCC based CSI feedback framework. In Section VI, the proposed method is evaluated to demonstrate its validity, adaptability, and generality. Finally, Section VII concludes our work.

Notations: Vectors and Matrices are denoted by boldface lower- and upper-case letters, respectively. Without special clarification, the vector is regarded as a column vector by default in this paper. The i-th element in the vector $\mathbf{x}$ is represented by x_i. Similarly, $\mathbf{x}_i$ (or $\mathbf{x}_u^i$) denotes the i-th column vector of the matrix $\mathbf{X}$ (or $\mathbf{X}_u$). $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively. $\|\cdot\|_2$ is the Euclidean norm. A circular symmetric Gaussian random vector $\mathbf{x}$ with covariance matrix $\mathbf{\Sigma}$ is denoted as $\mathbf{x} \sim \mathbb{CN}(0, \mathbf{\Sigma})$. Finally, $(\cdot)^T$ and $(\cdot)^H$ denote the transpose operation and the conjugate transpose operation, respectively.

II. Related Work

A. Deep Joint Source-Channel Coding

Compared with SSCC methods, DJSCC methods have shown remarkable reconstruction performance with the use of the same transmission resource for variant task. For the text transmission task, the DJSCC method outperforms the traditional SSCC method in terms of word error rate under the erasure channel [28]. For the image transmission task, the DJSCC methods proposed in [29] and [30] show better reconstruction performance than the JPEG/JPEG2000&LDPC method and the BPG&LDPC method, respectively. For the video transmission task, the proposed DJSCC method in [32] outperforms the H.264/H.265&LDPC method in terms of the multi-scale structural similarity index measure (MS-SSIM). Inspired by resource assignment strategies in traditional JSCC [35], an attention based DJSCC method is proposed to deal with different SNR conditions with a single neural network, which balances the performance and the storage requirement in real wireless scenarios [33]. Considering information privacy and confidentiality in using DJSCC, a novel source-channel coding method is proposed to protect the visual content of the plain image transmitted by the user [34]. The feasibility of combining DJSCC with orthogonal frequency division multiplexing (OFDM) system is demonstrated in [36].

B. Deep Learning Based CSI Feedback

The CSI of massive MIMO can be regarded as a high-dimensional and low-rank image. The problem of the CSI feedback task can be seen as a problem of the image compression task. In this view, a series of research based on autoencoder architecture were proposed to perform CSI compression [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. Wen et al. first proposed CSINet, which uses the architecture of autoencoder in DL [13]. The proposed CSINet outperforms the CS-based methods (e.g., LASSO [10], BM3D-AMP [11], TVAL3 [12]) by a large margin. From then on, autoencoder based architectures were proposed to explore the temporal correlation [14], [15], [16], improve the reconstruction accuracy [17], [18], [19], [20], or balance the storage overhead, computational complexity and reconstruction accuracy [21], [22], [23] in the truncated angular-delay domain. However, owing to the compressed CSI represented by 32-bit float values in the neural network, the number of feedback bits is 32 times the size of the output of the encoder, which is still a large amount for CSI feedback. To this end, novel quantization methods are proposed to further reduce CSI feedback overhead [17], [37], [38], [39].

CSI feedback with noisy channels have been considered in [17], [40], and [41]. The proposed method in [17] is trained for a specific SNR and evaluated in different SNRs to show its robustness. The proposed method in [40] introduces a noise extraction unit for denoising the received codeword at the BS and a joint training strategy for improving the reconstruction performance. In [41], a convolutional neural network (CNN)-based analog feedback method is proposed for the CSI in the spatial-frequency domain with a multi-carrier feedback channel. However, [17] and [40] are trained in the truncated angular delay domain, which discard some useful CSI information. And all of them need to train and test at the same SNR to get the optimal performance. Multiple models are needed to cover a range of SNRs, which leads to a significant model storage and dissemination overhead.

III. System Model

We consider a single cell FDD massive MIMO-OFDM system with $N_t \gg 1$ antennas at a BS and a single antenna at a UE. Both the uplink and the downlink are over N_c subcarriers. The downlink CSI and the uplink CSI in the spatial-frequency domain are represented by $\mathbf{H}_d \in \mathbb{C}^{N_c \times N_t}$ and $\mathbf{H}_u \in \mathbb{C}^{N_c \times N_t}$, respectively.

The CSI feedback system with known SNR both in the UE and in the BS is shown in Fig. 2. We assume that perfect uplink CSI and downlink CSI can be acquired by the pilot based training in the BS and in the UE, respectively. The methods CSINet, CSINet+ and CRNet also assume perfect CSI [13], [17], [18]. The joint source-channel encoder in the UE encodes the downlink CSI information $\mathbf{H}_d$ and the known SNR $\mu \in \mathbb{R}$ into a vector $\mathbf{s} \in \mathbb{C}^k$, which is expressed as:

$$\mathbf{s} = f_{en}(\mathbf{H}_d, \mu) \in \mathbb{C}^k, \tag{1}$$

where $f_{en} : \mathbb{C}^{N_c \times N_t} \times \mathbb{R} \rightarrow \mathbb{C}^k$ represents the encoding function and $k \leq N_c \times N_t$ is the size of the encoded vector $\mathbf{s}$. The uplink feedback channel SNR can be estimated at the BS and send to the UE in the downlink. The encoded vector $\mathbf{s} = [s_1, s_2, \cdots, s_k]^T$ is mapped to k subcarriers by

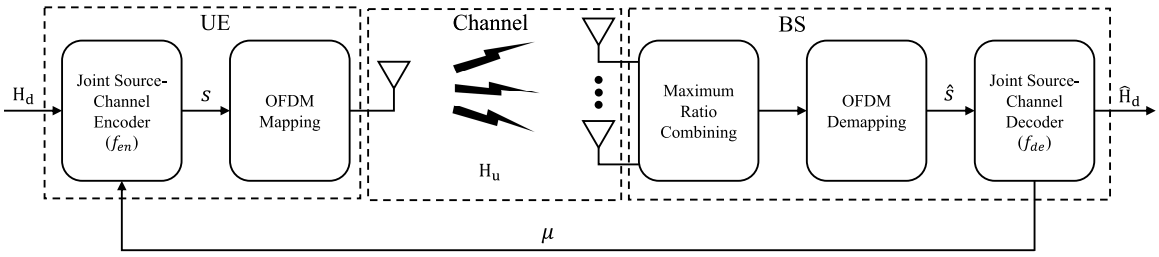


Fig. 2. The CSI feedback system with known SNR.

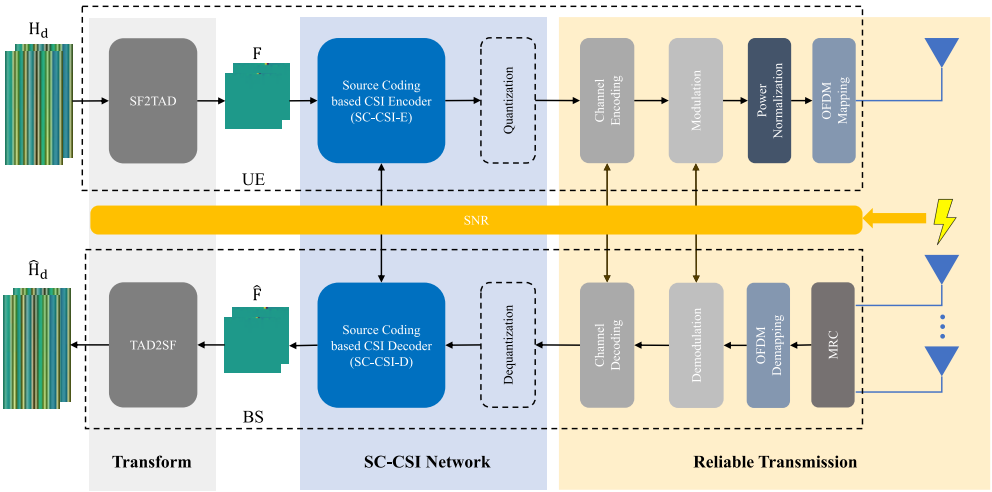


Fig. 3. The architecture of the SSCC based framework for the CSI feedback task. This architecture consists of three parts: (1) Transform, (2) SC-CSI network, and (3) Reliable transmission. SF2TAD converts the CSI from the spatial-frequency domain to the truncated angular-delay domain. Conversely, TAD2SF converts the CSI from truncated angular-delay domain to the spatial-frequency domain.

OFDM mapping and then transmitted by the UE, where s_i is the symbol carried on the i-th subcarrier. Assume the average power of a symbol over a subcarrier is 1, then the encoded vector $\mathbf{s}$ should be imposed a power normalization $\frac{1}{k}\mathbb{E}(\mathbf{s}\mathbf{s}^*) = 1$ to satisfy the power constraint at the UE.

The received i-th subcarrier of the feedback signal at the BS can be expressed as:

$$\mathbf{y}_i = \mathbf{h}_u^i s_i + \mathbf{z}_i, \tag{2}$$

where $\mathbf{y}_i \in \mathbb{C}^{N_t}$ is the received vector containing the received symbols at the BS antennas over the i-th subcarrier, and $\mathbf{z}_i \in \mathbb{C}^{N_t}$ consists of independent and identically distributed (i.i.d) samples with the distribution $\mathbb{CN}(0, \sigma^2\mathbf{I})$. σ^2 represents the noise power.

After receiving multiple noisy replicas of s_i $(i = 1, 2, \cdots, k)$ at the BS antennas, the recovered $\hat{s}_i$ over the i-th subcarrier is acquired by executing maximum ratio combining (MRC) at the BS, which is expressed as:

$$\hat{s}_i = \mathbf{w}_i^H \mathbf{y}_i, \tag{3}$$

For MRC, the combing vector $\mathbf{w}_i$ is $\mathbf{h}_u^i / \|\mathbf{h}_u^i\|_2$. After the MRC process, the received symbols are converted to the vector $\hat{\mathbf{s}} = [\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_k]^T$, which represents the reconstruction of the vector $\mathbf{s}$. The joint source-channel decoder at the BS uses a decoding function $f_{de} : \mathbb{C}^k \times \mathbb{R} \to \mathbb{C}^{N_c \times N_t}$ to map $\hat{\mathbf{s}}$ and the SNR μ to the recovered CSI $\hat{\mathbf{H}}_d$, which is expressed as:

$$\hat{\mathbf{H}}_d = f_{de}(\hat{\boldsymbol{s}}, \mu) \in \mathbb{C}^{N_c \times N_t}. \tag{4}$$

The feedback bandwidth is defined as k, which is the number of uplink subcarriers used for CSI feedback.

IV. Separate Source-Channel Coding Based CSI Feedback

Based on the modular design principle (i.e., the SSCC scheme) in wireless communications, the existing DL based methods for the CSI feedback task [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [37], [38] focus on CSI compression (i.e., source coding) with the promise of reliable transmission provided by appropriate channel coding and modulation as shown in Fig. 3. The reliable transmission means that there is no channel decoding error during the wireless transmission. These existing CSI feedback methods are called as source coding based CSI feedback (SC-CSI) methods.

By exploiting the sparsity of the CSI information in the angular-delay domain, a 2 dimensional discrete Fourier transform (2D-DFT) is executed at the UE to convert $\mathbf{H}_d \in \mathbb{C}^{N_c \times N_t}$ from the spatial-frequency domain to the angular-delay domain, which is expressed as:

$$\widetilde{\mathbf{F}} = \mathbf{I}_d \mathbf{H}_d \mathbf{D}_a, \tag{5}$$

where $\widetilde{\mathbf{F}} \in \mathbb{C}^{N_c \times N_t}$ is the CSI information in the angular-delay domain, $\mathbf{I}_d$ is an $N_c \times N_c$ inverse DFT (IDFT) matrix, and $\mathbf{D}_a$ is an $N_t \times N_t$ DFT matrix. Due to the fact that most of the power among multiple paths lies in a particularly limited period, the first N_c' rows which exhibit the distinct non-zero values are reserved as follows:

$$\mathbf{F} = Trun(\widetilde{\mathbf{F}}) \in \mathbb{C}^{N_c' \times N_t}, \tag{6}$$

where $Trun(\cdot)$ represents a truncate operation in the delay dimension and $\mathbf{F}$ is the truncated CSI in the angular-delay domain. These two stages are integrated as the spatial-frequency domain to truncated angular-delay domain (SF2TAD) module in Fig. 3.

The SC-CSI encoder (SC-CSI-E) module at the UE encodes the truncated CSI $\mathbf{F}$ to the vector $\mathbf{c} \in \mathbb{R}^m$ by the encoding function $e_\theta : \mathbb{C}^{N_c' \times N_t} \to \mathbb{R}^m$, which is expressed as:

$$\mathbf{c} = e_\theta(\mathbf{F}) \in \mathbb{R}^m, \tag{7}$$

where θ is the parameter set of the SC-CSI-E. In SC-CSI methods, a reliable transmission is assumed for transmitting the encoded vector $\mathbf{c}$ from the UE to the BS without error. The SC-CSI decoder (SC-CSI-D) module at the BS decodes the vector $\mathbf{c}$ to an approximate reconstruction of the truncated CSI in the angular-delay domain $\hat{\mathbf{F}} \in \mathbb{C}^{N_c' \times N_t}$ by the function $d_\phi : \mathbb{R}^m \to \mathbb{C}^{N_c' \times N_t}$ expressed as:

$$\hat{\mathbf{F}} = d_\phi(\mathbf{c}) \in \mathbb{C}^{N_c' \times N_t}, \tag{8}$$

where ϕ is the parameter set of the SC-CSI-D. The existing SC-CSI methods focus on the designs of the SC-CSI-E and SC-CSI-D networks, where the parameter sets $\{\theta, \phi\}$ are obtained by optimizing the following loss function L_{sc}:

$$L_{sc}(\theta, \phi) = \frac{1}{T} \sum_{i=1}^{T} \| d_\phi(e_\theta(\mathbf{F}^{(i)})) - \mathbf{F}^{(i)} \|_2^2, \tag{9}$$

where T is the number of samples in the training dataset and $\mathbf{F}^{(i)}$ represents the i-th sample in the training dataset in the truncated angular-delay domain.

After the decoding process at the BS, a zero matrix with the size $(N_c - N_c') \times N_t$ is padded at the end of the recovered truncated CSI in the angular-delay domain $\hat{\mathbf{F}}$ to acquire the recovered CSI in the angular-delay domain $\overline{\mathbf{F}}$, which is expressed as:

$$\overline{\mathbf{F}} = Zeropadding(\hat{\mathbf{F}}) \in \mathbb{C}^{N_c \times N_t}. \tag{10}$$

The BS then converts the recovered CSI from the angular-delay domain to the spatial-frequency domain as follows:

$$\overline{\mathbf{H}}_d = \mathbf{D}_f \overline{\mathbf{F}} \mathbf{I}_s, \tag{11}$$

where $\overline{\mathbf{H}}_d \in \mathbb{C}^{N_c \times N_t}$ is the recovered CSI in the spatial-frequency domain, $\mathbf{D}_f$ is an $N_c \times N_c$ DFT matrix, and $\mathbf{I}_s$ is a $N_t \times N_t$ IDFT matrix. These two stages are integrated as the truncated angular-delay domain to spatial-frequency domain (TAD2SF) module in Fig. 3.

The assumption of the reliable transmission simplifies the design of the CSI feedback task based on the SSCC scheme. However, the performance of the SSCC scheme is inferior to that of the JSCC scheme in the finite blocklength regime in theory [24] and easily suffers from the "cliff effect". Once these SC-CSI methods are applied in these wireless communication systems, the adaptive modulation and coding (AMC) strategy should be employed to deal with the varying wireless channel and achieve a high throughput.

V. Deep Joint Source-Channel Coding Based CSI Feedback

The DJSCC methods have been demonstrated to have better performance than SSCC methods for the text transmission task, the image transmission task, and the video transmission task [28], [29], [30], [31], [32], [33], [34]. In this section, we propose a DJSCC based framework with SNR adaption for the CSI feedback task (ADJSCC-CSI), which can exploit the experiences acquired in the existing SC-CSI methods, improve the CSI reconstruction performance, overcome the "cliff effect", and increase the SE for CSI feedback. There are three modules in the ADJSCC-CSI framework as shown in Fig. 4, including nonlinear transform module, deep joint source-channel coding module and SNR adaption module.

A. Nonlinear Transform Module

According to the sparsity assumption of CSI in the angular-delay domain, the CSI information in the spatial-frequency domain $\mathbf{H}_d$ is first converted to the angular-delay domain $\mathbf{F}$ by Eq. (5) and then truncated by Eq. (6) for dimensionality reduction in SC-CSI methods. Fig. 5(a) shows the CSI visualization in the angular domain. Fig. 5(b) shows the reserved part of the CSI for SC-CSI methods. The element with blue color means the value of the element is close to zero. Indeed, Fig. 5(b) contains most of non-zero elements, which is then compressed in SC-CSI methods. However, some useful information about the CSI shown in Fig. 5(c) is discarded and cannot be compensated in the subsequent process, which will lead to some distortion of the recovered CSI in SC-CSI methods. Therefore, it is desired to modify the imperfect truncation adopted in SC-CSI methods.

Here, the nonlinear transform Module in Fig. 4 instead of the transform module in Fig. 3 is exploited for CSI source. In image compression domain, the DL based non-linear transform [42] has been demonstrated superior to the the discrete cosine transform in JPEG [43] and the wavelet transform in JPEG2000 [44]. The nonlinear transforms are implied by using the analysis transform network (ATN) at the UE:

$$\mathbf{T} = A_\alpha(\mathbf{H}_d) \in \mathbb{C}^{N_c' \times N_t}, \tag{12}$$

and the synthesis transform network (STN) at the BS:

$$\hat{\mathbf{H}}_d = S_\beta(\hat{\mathbf{T}}) \in \mathbb{C}^{N_c \times N_t}, \tag{13}$$

where $\mathbf{T} \in \mathbb{C}^{N_c' \times N_t}$ is the transformed CSI at the UE, $\hat{\mathbf{T}} \in \mathbb{C}^{N_c' \times N_t}$ and $\hat{\mathbf{H}}_d \in \mathbb{C}^{N_c \times N_t}$ are the recovered CSI in the transform domain and the recovered CSI in the spatial-frequency domain at the BS, respectively. To simplify the notation and enable fair comparison with SC-CSI methods, the CSI in transform domain $\mathbf{T}$ and the recovered CSI in the

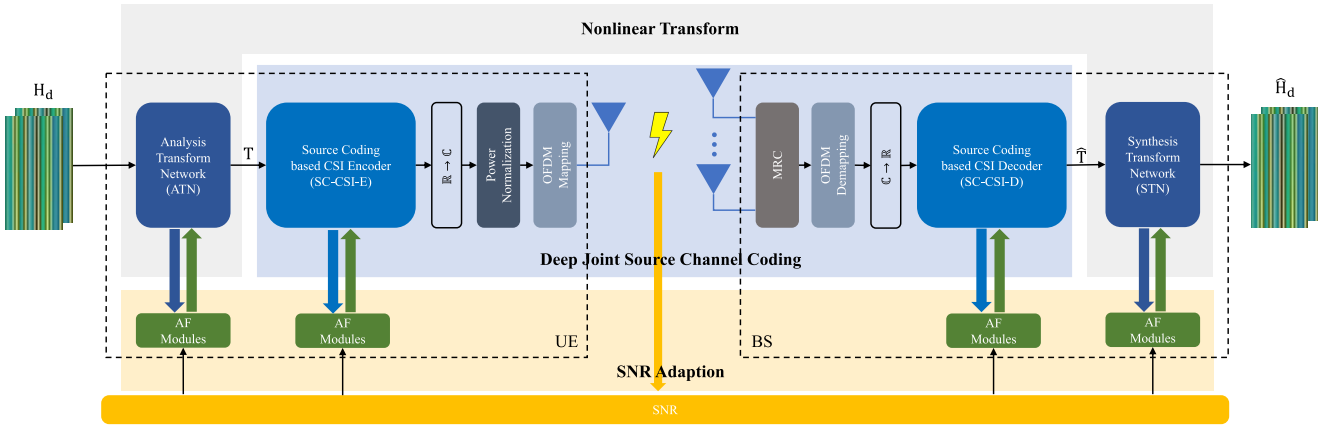


Fig. 4. The architecture of the DJSCC based framework for the CSI feedback task. This architecture consists of three parts: (1) Nonlinear transform, (2) Deep joint source channel coding, and (3) SNR adaption.

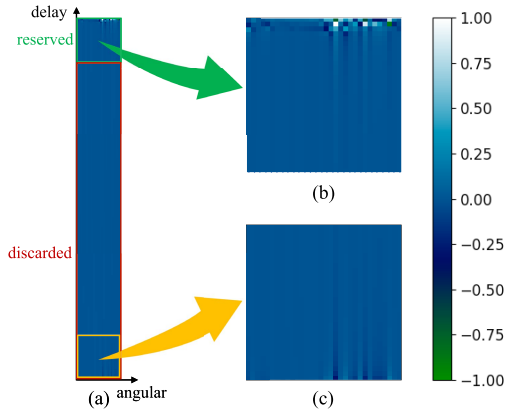


Fig. 5. Visualized images of the CSI sample in the angular-delay domain. The CSI sample is from the dataset generated by QuaDRiGa in Section VI. (a) The whole CSI in the angular-delay domain, (b) The reserved CSI in the angular-delay domain for SC-CSI methods, (c) The last part of the discarded CSI in the angular-delay domain for SC-CSI methods.

transform domain $\hat{\mathbf{T}}$ have the same size with the truncated CSI in the angular-delay domain $\mathbf{F}$. With an end-to-end training strategy, the ATN and the STN can learn more effective transforms than 2D-DFT used in SC-CSI methods.

B. Deep Joint Source-Channel Coding Module

The power of DJSCC comes from both the design of the neural network and the principle of JSCC. A typical DJSCC structure consisting of the joint source-channel encoder, noisy channel, and the joint source-channel decoder is introduced in [29]. The DJSCC encoder maps the input to a low dimensional vector with complex values at first. The real part and the imaginary part of the low dimensional vector is regarded as the I and Q components of the modulated symbols and directly transmitted through the noisy channel. Then the DJSCC decoder receives the noisy symbols and maps the corrupted symbols to an estimation of the original input. A differentiable channel transfer function is necessary for supporting back propagation in the training stage. Noting that the DJSCC encoder is not a deep complex network, the proposed network of the encoder outputs real values and then combines them to complex values, while the real part and imaginary part of the complex values at the receiver are firstly separated into real values and then inputted to the proposed network of the decoder. During the end-to-end training stage, the DJSCC learns effective parameters for the encoder and the decoder by simultaneously perceiving the input source and the noisy channel.

As complex encoded symbols are required in the DJSCC encoder, the output of the SC-CSI-E is first converted to the complex values as follows:

$$\mathbf{s} = R2C(\mathbf{c}) \in \mathbb{C}^k, \tag{14}$$

where $\mathbf{c} \in \mathbb{R}^{2k}$ is the output of the SC-CSI-E, $R2C(\cdot)$ represents the function converting the real vector $\mathbf{c}$ to the k dimensional vector $\mathbf{s}$. According to the power constraint, a power normalization function $PowerNorm(\cdot)$ is applied at the UE:

$$\mathbf{s} = PowerNorm(\mathbf{s}) \in \mathbb{C}^k. \tag{15}$$

In this paper, we adopt average power constraint. Then the symbols are mapped to subcarriers and transmitted by the UE. At the BS side, after the MRC and OFDM demapping, the received complex symbols $\hat{\mathbf{s}} \in \mathbb{C}^k$ is converted to real values as follows:

$$\hat{\mathbf{c}} = C2R(\hat{\mathbf{s}}) \in \mathbb{R}^{2k}, \tag{16}$$

where $C2R(\cdot)$ represents the function that converts the vector $\hat{\mathbf{s}}$ to the $2k$ dimensional vector $\hat{\mathbf{c}}$. Then the SC-CSI-D decodes the vector $\hat{\mathbf{c}}$. Based on these modifications, the SC-CSI network in Fig. 3 can be updated to a DJSCC network in Fig. 4. These conversions endow the encoding network and the decoding network ability to explore the knowledge from both the source and channel by the end-to-end training, which improves the reconstruction performance of the CSI and overcomes the problem of "cliff effect" in the SSCC scheme.

C. SNR Adaption Module

In the training stage, the DJSCC based methods are trained under the specific channel condition to effectively

explore the channel characteristics [29], [30], [31]. Multiple networks trained under different channel conditions are then deployed to match the real channel conditions. The large storage overhead limits its applicability in resource-constrained devices. Inspired by resource assignment strategy in traditional JSCC, our previous work [33] successfully designs a single network to cover different channel conditions by introducing SNR feedback from the decoder to the encoder. Here, we extend our previous work to deal with the CSI feedback task under fading channels. The core module in [33] is Attention Feature (AF) Module, which is a plug and play module. The AF Module can be employed for main layers (e.g., the convolution layer and the transposed convolution layer) of the neural network. Specifically, The inputs of the AF module are the SNR and the output of the main layer, and the output of the AF module is the scaled parameters multiply the output of the main layer. The function of the AF Module is to produce a sequence of scaling parameters to reassign the importance of the features in this layer.

Combining with the AF Modules, the ATN and the SC-CSI-E at the UE are improved as follows:

$$\mathbf{T} = AF_{\gamma}(A_{\alpha}(\mathbf{H}_d), \mu) \in \mathbb{C}^{N_c' \times N_t}, \tag{17}$$

$$\mathbf{c} = AF_{\psi}(e_{\theta}(\mathbf{T}), \mu) \in \mathbb{R}^{m}, \tag{18}$$

where $AF_{\gamma}(\cdot)$ and $AF_{\psi}(\cdot)$ denote that the AF Modules with parameter sets γ and ψ are placed after each main layer in the ATN and the SC-CSI-E as [45], respectively. The SNR $\mu \in \mathbb{R}$ is the feedback SNR from the BS to the UE. The output vector $\mathbf{c}$ with real values is used to generate the normalized symbols with complex values $\mathbf{s}$ according to Eq. (14) and Eq. (15). The complex vector $\mathbf{s}$ is mapped to subcarriers and transmitted through the fading channel, which is modeled in Eq. (2). The received symbols at the BS antennas are combined by MRC as in Eq. (3) and then demapped to the corrupted complex symbols $\hat{\mathbf{s}}$. The recovered vector with real values $\hat{\mathbf{c}}$ is converted according to Eq. (16).

At the BS, the SC-CSI-D and the STN are combined with AF Modules expressed as:

$$\hat{\mathbf{T}} = AF_{\rho}(d_{\phi}(\hat{\mathbf{c}}), \mu) \in \mathbb{C}^{N_c' \times N_t}, \tag{19}$$

$$\hat{\mathbf{H}}_d = AF_{\tau}(S_{\beta}(\hat{\mathbf{T}}), \mu) \in \mathbb{C}^{N_c \times N_t}, \tag{20}$$

where $AF_{\rho}(\cdot)$ and $AF_{\tau}(\cdot)$ denote that the AF Modules with parameter sets ρ and τ are placed after each main layer in the SC-CSI-D and the STN, respectively.

Different from existing SC-CSI methods considering the loss function Eq. (9) in the angular-delay domain, we optimize the parameter sets $\Theta = \{\alpha, \theta, \phi, \beta, \gamma, \psi, \rho, \tau\}$ by minimizing the distortions under a certain bandwidth k as follows:

$$\Theta^* = \arg\min_{\Theta} \mathbb{E}_{p(\mu)} \frac{1}{T} \sum_{i=1}^{T} \|\mathbf{H}_d^{(i)} - \hat{\mathbf{H}}_d^{(i)}\|_2^2, \tag{21}$$

where Θ^* is the optimal parameter set, $p(\mu)$ represents the probability distribution of the SNR, $\mathbf{H}_d^{(i)}$ represents the i-th sample of the training dataset in the spatial-frequency domain, and $\hat{\mathbf{H}}_d^{(i)}$ represents the reconstruction of $\mathbf{H}_d^{(i)}$ at the BS. During the training stage, the proposed ADJSCC-CSI for the CSI feedback task can learn: (1) a pair of transformations separately from the spatial-frequency domain to the transformed domain at the UE and from the transformed domain to the spatial-frequency domain at the BS, (2) a pair of encoder and decoder built upon the SC-CSI methods which learn from both the source and the channel, and (3) an end-to-end network which can adapt against the channel variation.

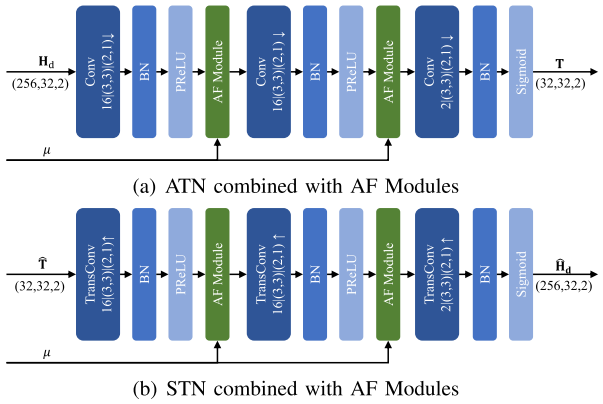


Fig. 6. The architectures of the transform network combined with AF Modules. The notation $C|(F_1, F_2)|(S_1, S_2)$ $\downarrow$ in the convolution layer denotes that it has C filters with size $F_1 \times F_2$. S_1 is the vertical steps for stride operation and S_2 is the horizontal steps for stride operation. $\uparrow$ and $\downarrow$ represents the stride up and down operation, respectively.

VI. Experimental Results

The proposed ADJSCC-CSI is a general framework that can be employed to upgrade many SC-CSI methods to DJSCC based methods. We first demonstrate its effectiveness by using a concrete scheme in this section. Next, we evaluate the ability of SNR adaption under this concrete scheme. Finally, we illustrate the generality of ADJSCC-CSI by exploiting different architectures in other SC-CSI methods.

The uplink CSI and downlink CSI are generated by QuaDRiGa [46] according to the 3rd Generation Partnership Project (3GPP) TR 38.901 [47]. We create an open indoor scenario in which the downlink center frequency is 5.2 GHz and the uplink center frequency is 5.4 GHz. There are 20 clusters and 15 subpaths for each cluster in the open indoor scenario. The BS is positioned at the center of a square area with the size of $20m \times 20m$. A uniform linear array (ULA) with half-wavelength antenna space is deployed at the BS. The number of the antennas at the BS is $N_t = 32$ and the number of the antennas at the UE is $N_r = 1$. Both the antennas in the BS and in the UE are omnidirectional. The heights of the BS and the UE are $3m$ and $1.5m$, respectively. The uplink and the downlink both have $N_c = 256$ subcarriers. The training, validation, and testing sets contain 100,000, 30,000, and 20,000 sample pairs, respectively. A sample pair contains a downlink CSI sample and an uplink CSI sample. Following the setting in the existing SC-CSI methods [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [37], [38], N_c' is set to 32.

Fig. 6 illustrates the networks of the ATN and the STN combined with AF Modules. The ATN combined with AF

Modules shown in Fig. 6(a) consists of three convolution layers, each of which is followed by a batch normalization layer and a PReLU [48] activation except the last one. The sigmoid activation follows the last convolution layer. The AF Modules are inserted after the first two activation functions. The notation $C|(F_1, F_2)|(S_1, S_2)\ \downarrow$ in the convolution layer denotes that it has C filters with size $F_1 \times F_2$. $\downarrow$ represents the stride down operation. S_1 and S_2 are the vertical steps and the horizontal steps, respectively. Due to the unsymmetrical steps the stride down operation adopts, the CSI information $\mathbf{H}_d$ of a rectangle shape can be successfully converted to the transformed CSI $\mathbf{T}$ of a square shape, which matches the truncated CSI size in the existing SC-CSI methods. The STN combined with AF Modules is shown in Fig. 6(b). It consists of three transposed convolution layers. Each of the first two transposed convolution layers is followed by a batch normalization layer, a PReLU activation and an AF Module. The activation of the last transposed convolution layer is also sigmoid. A more sophisticated network structure than the proposed one for the ATN and the STN may improve the CSI reconstruction performance. However, in this paper, we focus on the mechanism design rather than the network design.

Besides a new network architecture CSINet+ for CSI feedback, [17] proposes a concrete quantization method for the real values of the CSINet+ encoder. The proposed quantized method provides a more effective representation for the encoded output than its unquantized version in the view of bit compression. We first adopt the CSINet+ encoder and the CSINet+ decoder as the SC-CSI-E and the SC-CSI-D in Fig. 4, respectively. Consistently with the work [17], the batch size is set to 200. Adam optimizer is first initialized with a learning rate of 10^{-3}. When the loss does not decrease in 20 epochs, the learning rate will be decayed by half. The lower bound of the learning rate is set to 10^{-4}. The training epochs are set to 500 for the network convergence. Tensorflow [49] and its high-level Keras are used to implement the following experiments. All of our experiments are performed on a Linux server with four Hygon 7151 16-core CPUs and sixteen NVIDIA RTX A3000 GPU. Each experiment runs on four CPU cores and a GPU. Consistently with existing SC-CSI methods for the CSI feedback task, normalized mean square error (NMSE) is utilized to evaluate CSI reconstruction performance, which is expressed as:

$$\mathrm{NMSE} = \mathbb{E}\|\mathbf{H}_{\mathrm{d}} - \hat{\mathbf{H}}_{\mathrm{d}}\|_2^2 / \|\mathbf{H}_{\mathrm{d}}\|_2^2. \tag{22}$$

The lower NMSE reveals the better performance of CSI feedback.

A. ADJSCC-CSI Validity Experiments

The proposed framework built upon CSINet+ for the CSI feedback task is named as ADJSCC-CSINet+. The ADJSCC-CSINet+ is trained under the uniform distribution of $\mathrm{SNR}_{\mathrm{train}}$ from -10 dB to 10 dB. To demonstrate the validity of the ADJSCC-CSINet+, the SSCC scheme as shown in Fig. 3, which is built upon CSINet+ and 5G NR uplink control Information (UCI) transportation [50], is used as the compared method. The concrete UCI encoding scheme depends on the length of input bits shown in Table I. The modulation scheme of the UCI contains BPSK, QPSK (i.e.,4QAM), 16QAM, 64QAM, and 256QAM. To fairly compare with the ADJSCC-CSINet+, source coding rate, channel coding rate, and modulation scheme should be adjusted to match the feedback bandwidth.

TABLE I
UCI ENCODING SCHEME

Length (L)	Segmentation	CRC Bits	Encoding scheme
1	N/A	N/A	Repetition
2	N/A	N/A	Simplex
3-11	N/A	N/A	Reed-Muller
12-19	N/A	6	Parity-check Polar
20-1706	L $\geq$ 1013	11	Polar

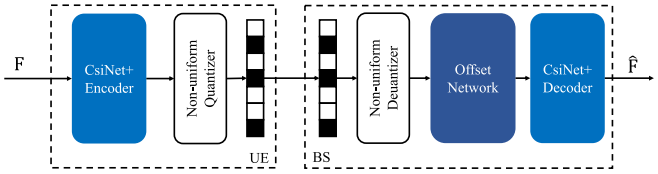


Fig. 7. Proposed bit-level CSINet+ framework proposed by [17]. The original CSI in the angular-delay domain is first compressed at the CSINet+ encoder and then quantized by the non-uniform quantizer to generate a bitstream in the UE. In the BS, the received bitstream are first dequantized and then fed into offset network for quantization distortion compensation. Finally, the CSI in the angular-delay domain is recovered by the CSINet+ decoder.

The framework of bit-level CSINet+ network is shown in Fig. 7. The source coding rate varies for different quantized levels of the CSINet+. The training strategy of the quantized CSINet+ contains the following three steps: (1) The CSINet+ without quantization is trained by an end-to-end approach in the truncated angular-delay domain; (2) The CSINet+ encoder, non-uniform quantizer, and the non-uniform dequantizer are combined to generate the training dataset, which is then used to train the offset network for quantization compensation by minimizing the mean square error (MSE) between the training dataset and the output of the offset network; (3) The parameters of the CSINet+ encoder are fixed and the parameters of the offset network and the CSINet+ decoder are fine tuned to further improve the reconstruction performance. More training Details about bit-level CSINet+ can be found in [17].

Fig. 8 compares the ADJSCC-CSINet+ method with the SSCC methods (i.e., CSINet+ and UCI transmission). The evaluation of CSI reconstruction performance is in the spatial-frequency domain. The label "CSINet+_mQq_r_aQAM" represents that the SSCC method adopts the CSINet+ as the source coding scheme and the concrete UCI encoding scheme as the channel coding, where m, q, r and a represent the output dimension of the CSINet+ encoder, the number of bits for quantizing each dimension of the output of the CSINet+ encoder, the channel coding rate, and the number of constellation points of the QAM, respectively. The CSI feedback bandwidth k for SSCC methods can be

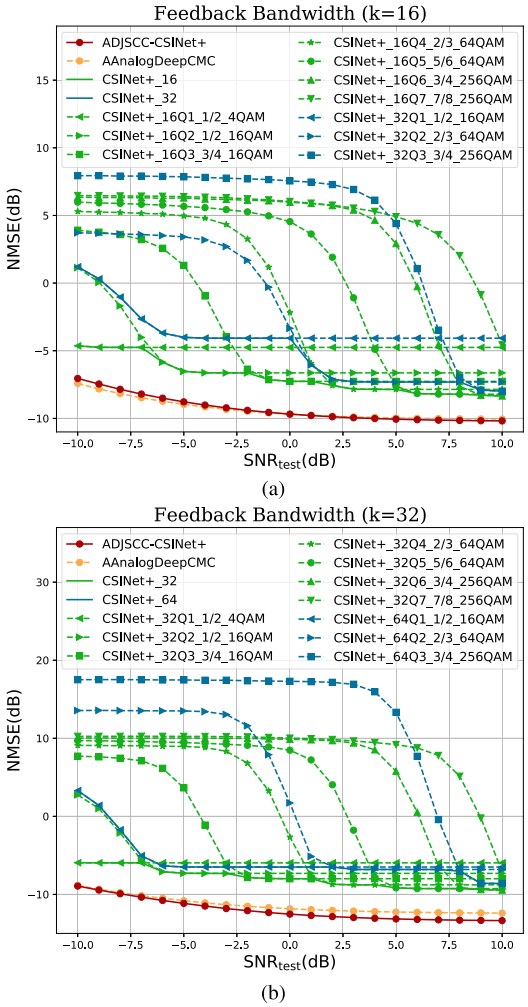


Fig. 8. Performance comparison of ADJSCC-CSINet+ with SSCC based CSINet+. (a) k = 16 and (b) k=32. The SSCC based CSINet+ consists of the CSINet+ as the source coding scheme and 5G NR uplink control Information (UCI) scheme the as channel coding and modulation scheme. The label "CSINet+_16Q2_ 1/2_16QAM" means the CSINet+ compresses the truncated CSI to a 16-dimensional vector, each of which is quantized to 2 bits and the UCI adopts 1/2 channel coding rate and 16QAM modulation scheme.

expressed as:

$$k = \frac{m \times q}{r \times \log_2 a}. \tag{23}$$

In the SSCC method, the fixed source coding rate, channel coding rate, and modulation scheme are designed for the specific SNR. The best performance of the SSCC methods with the same compressed dimensions m of the CSINet+ is simplified as an envelope curve, which is named as CSINet+_m. Specifically, The CSINet+_16, CSINet+_32 and CSINet+_64 represent SNR adaption schemes (i.e., the bit-level CSINet+ combined with the AMC strategy) with 16-, 32- and 64-dimensional output in Fig. 8, respectively. The AnalogDeepCMC proposed by [41] also takes into account the uplink channel explicitly. Here, we insert AF Modules which we have introduced in Section V-C into the AnalogDeepCMC to construct the AAnalogDeepCMC for SNR adaption and plot its performance as reference. In Fig. 8(a) with $k = 16$, the performance of the AJDSCC-CSINet+ is $1.8 \sim 3.4$ dB better than that of the CSINet+_16 and $2.2 \sim 8.3$ dB better than that of the CSINet+_32 in the $\mathrm{SNR}_{\mathrm{test}}$ range from -10 dB to 10 dB. The gap between the performance of ADJSCC-CSINet+ and the performance of the SSCC methods with $k = 32$ in Fig. 8(b) is larger than that with $k = 16$ in Fig. 8(a). Considering the concrete quantization, channel coding and modulation scheme (e.g., "CSINet+_16Q4_2/3_64QAM"), the performance of the ADJSCC-CSINet+ in the target SNR is at least 2 dB better than the performance of the "CSINet+_16Q4_2/3_64QAM". Compared with the AAnalogDeepCMC, the performance of the ADJSCC_CSINet+ is similar with that of the AAnalogDeepCMC in Fig. 8(a) and slightly better than that of the AAnalogDeepCMC in Fig. 8(b).

Intuitively, the performance of the CSINet+_64 would be better than that of the CSINet+_32 due to the better performance of the CSINet+ with 64 dimensions than that of the CSINet+ with 32 dimensions. Indeed, once the channel coding is perfect to protect the source coding information, the CSINet+ with more dimensions performs better than that with less dimensions for the same quantization level. For instance, the performance of the "CSINet+_64Q1_1/2_16QAM" is better than that of the "CSINet+_32Q1_1/2_4QAM" when $\mathrm{SNR}_{\mathrm{test}} \geq 6$ dB. For the same quantization bits, however, the performance of the less dimensions quantized with more bits may be better than that of the more dimension quantized with less bits. This phenomenon was also observed in [17]. For instance, the performance of the "CSINet+_32Q2_1/2_16QAM" is better than that of the "CSINet+_64Q1_1/2_16QAM". The performance of CSINet+_32 when $\mathrm{SNR}_{\mathrm{test}} \in [-10, -2]$ dB consists of the performance of the "CSINet+_32Q1_1/2_4QAM" when $\mathrm{SNR}_{\mathrm{test}} \in [-10, -7]$ dB and the performance of the "CSINet+_32Q2_1/2_16QAM" when $\mathrm{SNR}_{\mathrm{test}} \in [-6, -2]$ dB. The performance of CSINet+_64 equals to the performance of the "CSINet+_32Q2_1/2_16QAM" when $\mathrm{SNR}_{\mathrm{test}} \in [-10, -2]$ dB. Since the "CSINet+_64Q2_1/2_16QAM" performs worse than the "CSINet+_32Q1_1/2_4QAM" when $\mathrm{SNR}_{\mathrm{test}} \in$ [-10,-7] dB and the "CSINet+_32Q2_1/2_16QAM" when $\mathrm{SNR}_{\mathrm{test}} \in [-6, -2]$ dB. The performance of the CSINet+_64 is worse than that of the CSINet+_32 when $\mathrm{SNR}_{\mathrm{test}} \in$ [-10,-2] dB. Similar results are revealed when comparing other SSCC schemes with 32 and 64 dimensions in the regime $\mathrm{SNR}_{\mathrm{test}} \in [-2, 10]$ dB. Hence, the performance of CSINet+_64 is worse than that of CSINet+_32.

It worth noting that the CSINet+ in SSCC methods is trained in the truncated angular-delay domain while the ADJSCC-CSINet+ is trained in the spatial-frequency domain. Specifically, the ADJSCC-CSINet+ uses the ATN to convert the CSI in the spatial-frequency domain to the transform domain with the same size as the truncated angular-delay domain in the UE. We infer that the performance gain of ADJSCC-CSINet+ comes from both the non-linear transform and the DJSCC design philosophy. A thorough understanding of the ADJSCC-CSINet+ is necessary to identify which part of the performance gain from the non-linear transform and

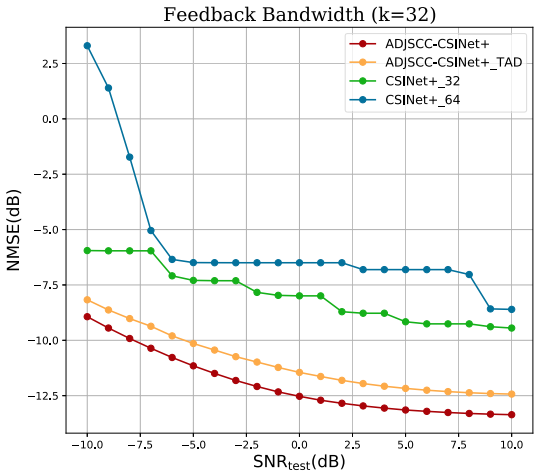


Fig. 9. The ablation study for ADJSCC-CSINet+.

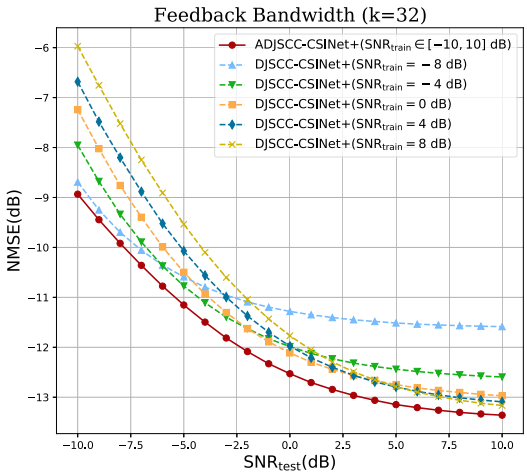


Fig. 10. Performance of ADJSCC-CSINet+ and DJSCC-CSINet+. The curve of ADJSCC-CSINet+ is trained under the uniform distribution of SNR from -10 dB to 10 dB. Each curve of DJSCC-CSINet+ is trained at a specific SNR.

which part of the performance gain from the DJSCC design philosophy. Here, an ablation study is provided as shown in Fig. 9.

The curve "ADJSCC-CSINet+_TAD" represents the proposed ADJSCC-CSINet+ without the ATN and the STN. The "ADJSCC-CSINet+_TAD" is trained in the truncated angular-delay domain. As shown in Fig. 9, the performance of the AJDSCC-CSINet+_TAD is $0.7 \sim 1.1$ dB worse than that of the ADJSCC-CSINet+ in $\mathrm{SNR}_{\mathrm{test}} \in [-10, 10]$ dB, which shows the performance gain of non-linear transform. The DJSCC performance gain is reflected by the performance gap between the ADJSCC-CSINet+ and the CSINet+_m. The performance of the AJDSCC-CSINet+_TAD is $2.2 \sim 3.6$ dB better than that of the CSINet+_32 and $3.4 \sim 11.4$ dB better than that of the CSINet+_64 in the $\mathrm{SNR}_{\mathrm{test}}$ range from -10 dB to 10 dB, respectively.

B. ADJSCC-CSI Adaptability Experiments

Now, we would like to demonstrate the effect of the AF Module only. Here, the DJSCC-CSINet+ is introduced as a compared method. The DJSCC-CSINet+ can be regarded as the ADJSCC-CSINet+ without AF Modules. Fig. 10 compares the ADJSCC-CSINet+ method with the DJSCC-CSINet+ methods with feedback bandwidth $k = 32$. The DJSCC-CSINet+ is trained at specific $\mathrm{SNR}_{\mathrm{train}} = -8, -4, 0, 4, 8$ dB, respectively. Both the ADJSCC-CSINet+ and the DJSCC-CSINet+ are evaluated at specific $\mathrm{SNR}_{\mathrm{test}} \in [-10, 10]$ dB. Although the DJSCC can achieve remarkable performance when $\mathrm{SNR}_{\mathrm{test}}$ is around $\mathrm{SNR}_{\mathrm{train}}$, the performance of the DJSCC trained at specific $\mathrm{SNR}_{\mathrm{train}}$ is not satisfactory when $\mathrm{SNR}_{\mathrm{test}}$ is far from $\mathrm{SNR}_{\mathrm{train}}$. For instance, the performance of the DJSCC-CSINet+ trained at $\mathrm{SNR}_{\mathrm{train}} = -8$ dB has the best performance than other DJSCC-CSINet+ methods in $\mathrm{SNR}_{\mathrm{test}}$ from -10 dB to -7 dB. With the increase of $\mathrm{SNR}_{\mathrm{test}}$ from -6 dB to -3 dB, the performance of the DJSCC-CSINet+ trained at $\mathrm{SNR}_{\mathrm{train}} = -8$ dB is inferior to that of the DJSCC-CSINet+ trained at $\mathrm{SNR}_{\mathrm{train}} = -4$ dB. However, the performance of the ADJSCC-CSINet+ outperforms the lower envelop of the provided DJSCC-CSINet+ methods at any $\mathrm{SNR}_{\mathrm{test}}$, which demonstrates the efficacy of AF Modules dealing with SNR adaption for the CSI feedback task. It worth noting that the ADJSCC-CSINet+ performs better even compared with the DJSCC-CSINet+ when it is trained and tested at the same SNR. The remarkable performance of the ADJSCC-CSINet+ may come from the design of AF modules that utilize not only the SNR information but also the pooling information that extracts effective features [33].

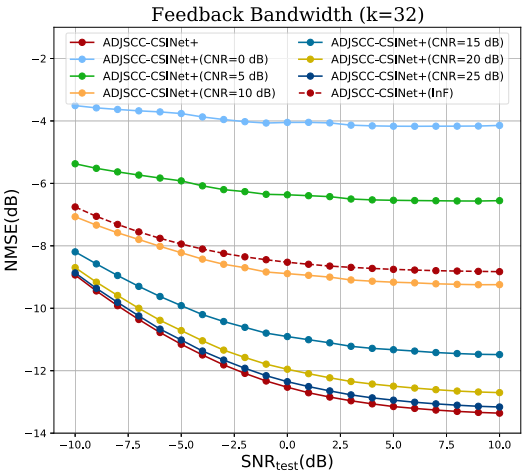


Fig. 11. Dataset mismatch of the ADJSCC-CSINet+. The model is trained by the perfect CSI in the open indoor scenario and evaluated by the imperfect CSI with different noise strength and the mismatched CSI in the indoor factory scenario.

C. ADJSCC-CSI Robustness Experiments

Although the perfect CSI is assumed in Section III, it is difficult to obtain the perfect CSI in the real wireless communication system, where pilot overhead is inefficient and channel estimation errors may exist. In addition, the real scenario is not always consistent with the training scenario. In this part, we evaluate the dataset mismatch and the SNR mismatch to show the robustness of the proposed method.

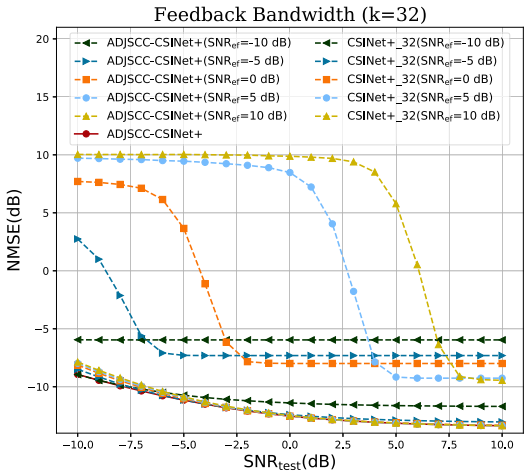


Fig. 12. SNR mismatch of the ADJSCC-CSINet+ and the SSCC based CSI feedback method.

A linear model is used to represent the imperfection of channel estimation as follows:

$$\tilde{\mathbf{H}}_d = \mathbf{H}_d + \mathbf{E}, \tag{24}$$

where $\mathbf{E} \in \mathbb{C}^{N_c \times N_t}$ is the estimation error. The average power of $\mathbf{H}_d$ to that of $\mathbf{E}$ is defined as the channel-to-noise ratio (CNR). Another type of dataset mismatch is scenario mismatch. Here, the CSI dataset of the indoor factory scenario is evaluated. There are 25 clusters and 20 subpaths for each cluster in the indoor factory scenario. The same parameters as the open indoor scenario are used to generate the CSI dataset of the indoor factory scenario. In Fig. 11, the labels "ADJSCC-CSINet+(CNR=25 dB)" and "ADJSCC-CSINet+(InF)" represent the ADJSCC-CSINet+ evaluated by the imperfect CSI dataset with CNR=25 dB and the indoor factory CSI dataset, respectively. The performance of the perfect CSI dataset is plotted as reference. With small channel estimation error, the performance of the ADJSCC-CSINet+(CNR=25 dB) is only 0.3 dB lower than that of the ADJSCC-CSINet+. With the increase of estimation error (i.e., the CNR from 25 dB to 0 dB), the gap between the ADJSCC-CSINet+ with the imperfect CSI and that with perfect CSI becomes larger.

In the real wireless communication system, the input SNR may be inaccurate due to the estimation errors and feedback errors. The SNR estimated by the BS and then fed back to the UE is represented by $\mathrm{SNR}_{\mathrm{ef}}$. To clearly show the effect of SNR mismatch without other disturbance, we evaluate it on the perfect CSI dataset. For the SSCC based CSI feedback scheme, once the $\mathrm{SNR}_{\mathrm{ef}}$ is calculated, the quantization scheme, the channel coding scheme and the modulation scheme are determined. In Fig. 12, the performance of the ADJSCC-CSINet+ ($\mathrm{SNR}_{\mathrm{ef}} = -10, -5, 0, 5, 10$ dB) outperform the CSINet+_32($\mathrm{SNR}_{\mathrm{ef}} = -10, -5, 0, 5, 10$ dB) by a large margin, respectively. The performance of the CSINet+_32($\mathrm{SNR}_{\mathrm{ef}} = -10, -5, 0, 5, 10$ dB) drops drastically caused by the outbreak of channel decoding errors when $\mathrm{SNR}_{\mathrm{test}} < \mathrm{SNR}_{\mathrm{ef}}$. While the performance of the CSINet+_32($\mathrm{SNR}_{\mathrm{ef}} = -10, -5, 0, 5, 10$ dB) remains the same when $\mathrm{SNR}_{\mathrm{test}} > \mathrm{SNR}_{\mathrm{ef}}$. This is the "cliff effect" in SSCC methods. However, the performance of the ADJSCC-CSINet+ ($\mathrm{SNR}_{\mathrm{ef}} = -10, -5, 0, 5, 10$ dB) exhibits a graceful degradation when $\mathrm{SNR}_{\mathrm{test}} < \mathrm{SNR}_{\mathrm{ef}}$ and a gradual increase when $\mathrm{SNR}_{\mathrm{test}} > \mathrm{SNR}_{\mathrm{ef}}$. Even compared with the ADJSCC-CSINet+ with accurate SNR, the most performance loss of the ADJSCC-CSINet+ ($\mathrm{SNR}_{\mathrm{ef}} = -10, -5, 0, 5, 10$ dB) is only 1.7 dB. The ADJSCC-CSINet+ is more robust than the CSINet+_32 in the SNR mismatch condition.

D. ADJSCC-CSI Generality Experiments

Section VI-A reveals that the CSINet+ can be successfully converted to the DJSCC based method ADJSCC-CSINet+ by employing the proposed ADJSCC-CSI framework. In this subsection, we apply the ADJSCC-CSI framework to the other two SC-CSI methods—the CSINet proposed in [13] and the CRNet proposed in [18]—to show the generality of the proposed framework. The CSINet and the CRNet combining with the ADJSCC-CSI framework are named as ADJSCC-CSINet and ADJSCC-CRNet, respectively. It is time consuming to traverse all of the possible combination schemes for the SSCC method to identify the best performance at the specific SNR. Here we employ some additional assumptions to reduce the difficulty in evaluating the SSCC method. Our target is to compare the performance of the DJSCC based method with that of the SSCC based method, rather than to find the practical combination scheme for the SSCC method. Therefore, the channel capacity C—an upper bound for reliable transmission—instead of the AMC strategy is the employed assumption to simplify the evaluation process for the SSCC method. Next, we consider about simplifying the quantization process.

Due to no consideration of the quantization in [13] and [18], the output of the SC-CSI-E at the UE is defaulted to 32-bit float values and then is assumed to be transmitted to the BS without error. The output bit length of the SC-CSI-E is $32 \times m$ in [13] and [18], where m is the output dimension of the SC-CSI-E. However, [17] demonstrates that the feedback of 32-bit float values is inefficient compared with the quantized version. According to the experiment results provided by [17], the performance of the quantized version with $B = 6$ is near the performance of the unquantized version represented by 32-bit float values. Here, we assume each dimension of the unquantized version represented by 5 bits, which is the lower bound of the performance of the SSCC based method. Finally, the chosen dimension of the SC-CSI-E is expressed as:

$$m = \lceil kC/B \rceil, \tag{25}$$

where k is the feedback bandwidth, C is the feedback channel capacity for one channel use, B is the assumed quantized bits, and $\lceil \cdot \rceil$ represents the ceiling operation. Then the performance of the SC-CSI method (e.g., the CSINet and the CRNet) with dimension m is regarded as the performance of the SSCC method built upon the SC-CSI method with feedback bandwidth k. The ceiling operation further improves the lower bound of the performance of the SSCC method.

Fig. 13 compares the ADJSCC-CSI methods with the SSCC methods at feedback bandwidth $k = 32$. Specifically,

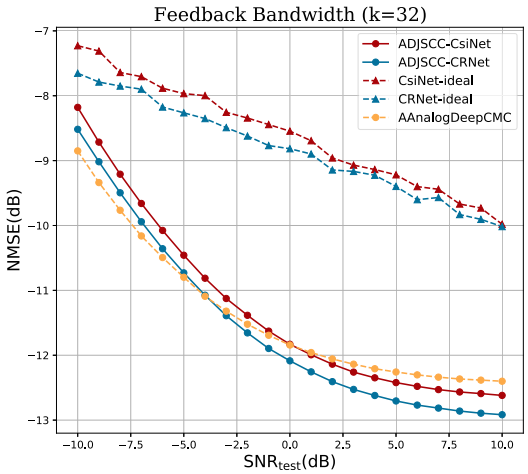


Fig. 13. Performance comparison of ADJSCC and SSCC for CSINet and CRNet.

the labels "CSINet-ideal" and "CRNet-ideal" represent the SSCC methods built upon the CSINet and the CRNet under the three assumptions, respectively. The hyper-parameter setting for the CSINet and the CRNet is the same as that for the ADJSCC-CSINet+. Based on the aforementioned assumptions, the performance of the CSINet-ideal and the CRNet-ideal are the lower bound of the SSCC based CSInet and the SSCC based CRNet, respectively. Nonetheless, the performance of the CSINet-ideal and the CRNet-ideal are still worse than the performance of the ADJSCC-CSINet and the ADJSCC-CRNet, respectively. Specifically, the performance of the CSINet is 0.9 dB worse than the performance of the ADJSCC-CSINet at $\mathrm{SNR}_{\mathrm{test}} = -10$ dB. With the increase of the $\mathrm{SNR}_{\mathrm{test}}$ from -10 dB to 1 dB, the ADJSCC-CSINet brings a gradually increased performance, outperforming the CSINet-ideal by a margin at almost 3.3 dB. With the further increase of the $\mathrm{SNR}_{\mathrm{test}}$ from 1 dB to 10 dB, the gap between the CSINet-ideal and the ADJSCC-CSINet slowly decreases, while is still larger than 2.6 dB. In other words, if $\mathrm{NMSE} = -10$ dB is required at the BS to promise the reconstruction quality, the UE adopting the ADJSCC-CSINet can save at least 14 dB transmission power than that adopting the CSINet-ideal. The comparison between the CRNet-ideal and the ADJSCC-CRNet reveals similar results.

It has been demonstrated that the performance of the CRNet is better than that of the CSINet in [18]. Hence the performance of the CRNet-ideal should be better than that of the CSINet-ideal. This is consistent with the comparison result between the CRNet-ideal and the CSINet-ideal revealed in Fig. 13. Moreover, the comparison between the ADJSCC-CRNet and the ADJSCC-CSINet in Fig. 13 reveals a similar trend. We infer that is because the ability of CRNet is stronger than that of the CSINet. Compared with the AAnalogDeepCMC, the ADJSCC-CSINet is inferior in the $\mathrm{SNR}_{\mathrm{test}}$ range from -10 dB to 0 dB and is superior in in the $\mathrm{SNR}_{\mathrm{test}}$ range from 1 dB to 10 dB. The comparison between the AAnalogDeepCMC and the ADJSCC-CRNet shows similar results except the partition point changed from 0 dB to -4 dB.

TABLE II

STORAGE OVERHEAD AND COMPUTATION COMPLEXITY COMPARISON FOR THE CSI FEEDBACK TASK

Methods	UE		BS	
	params	FLOPs	params	FLOPs
ADJSCC-CSINet	**233.9K**	**13.6M**	**241.3K**	20.4M
ADJSCC-CRNet	234.3K	13.9M	241.6K	**19.6M**
CSINet-ideal	999.3K	18.9M	1.1M	25.7M
CRNet-ideal	999.8K	19.2M	1.1M	24.8M
AAnalogDeepCMC	1.8M	116.7M	8.4M	536.5M

We calculate the parameters and the floating point operations (FLOPs) needed in the UE and in the BS for the aforementioned methods, respectively. In the practical wireless scenario, many CSINet models should be prepared for the CSINet-ideal to cover the SNR range from -10 dB to 10 dB. However, one model is enough for the ADJSCC-CSI method and the AAnalogDeepCMC method due to the SNR adaption mechanism. Assuming one model corresponds to a specific SNR for the CSINet-ideal to keep the optimality. The parameter sizes and the FLOPs of these methods are given in Table II. As we can see, the parameter size of the CSINet-ideal is over $4\times$ heavier than that of the ADJSCC-CSINet both in the UE and in the BS. The parameter size of the AAnalogDeepCMC is over $7\times$ and $34\times$ heavier than that of the ADJSCC-CSINet in the UE and in the BS, respectively. The parameter sizes of the CRNet-ideal and the ADJSCC-CRNet are slightly more than that of the CSINet-ideal and the ADJSCC-CSINet, respectively. Next, we compare the FLOPs. The CSINet-ideal and the CRNet-ideal consist of multiple networks. For simplicity, we only calculate the FLOPs of the 2D-DFT and the minimal FLOPs of the networks in the UE and the BS. The FLOPs of the ADJSCC-CSINet is 5.3M less than that of the CSINet-ideal both in the UE and in the BS. Moreover, the FLOPs of the AAnalogDeepCMC is over $8\times$ and $26\times$ heavier than that of the ADJSCC-CSINet. The FLOPs of the ADJSCC-CRNet is 0.3M more and 0.8M less than that of the ADJSCC-CSINet in the UE and in the BS, respectively. To summarize, the proposed ADJSCC-CSI has lower storage, lower computational complexity, and better performance than the SSCC based method for the CSI feedback task.

VII. CONCLUSION

In this work, we propose a novel ADJSCC-CSI framework for the CSI feedback task, which includes three key components, e.g., non-linear transform, DJSCC and SNR adaption. The proposed framework can successfully upgrade the existing SC-CSI method (e.g., CSINet, CSINet+, CRNet) to a DJSCC based method. During the end-to-end training, ADJSCC-CSI framework can learn a pair of transformation networks, a pair of encoder and decoder built upon the SC-CSI method, and SNR adaption strategy.

In the experiments, we compared the ADJSCC-CSINet+ with SSCC based methods and the DJSCC based methods built upon CSINet+ to show its validity and adaptability, respectively. Then we apply the proposed ADJSCC-CSI

framework to the CSINet and the CRNet to demonstrate its generality. In addition, compared with the SSCC based method, the ADJSCC-CSI framework can save more storage and computational resource both in the UE and in the BS.

In future work, a potential direction is to jointly design the channel estimation module and the CSI feedback module with the DJSCC method, which might further reduce the CSI feedback bandwidth on the basis of acceptable CSI reconstruction performance.

REFERENCES

[1] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign, IL, USA: Univ. Illinois Press, 1949.

[2] R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information," Massachusetts Inst. Technol. Cambridge, MA, USA, Tech. Rep., 247, 1952.

[3] J. Barwise and J. Perry, "Situations and attitudes," *J. Philosophy*, vol. 78, no. 11, pp. 668–691, 1981.

[4] L. Floridi, "Outline of a theory of strongly semantic information," *Minds Mach.*, vol. 14, no. 2, pp. 197–221, May 2004.

[5] Y. Bengio, I. Goodfellow, and A. Courville, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2016.

[6] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.

[7] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[8] J. Wang, C.-X. Wang, J. Huang, H. Wang, and X. Gao, "A general 3D space-time-frequency non-stationary THz channel model for 6G ultra-massive MIMO wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1576–1589, Jun. 2021.

[9] *5G; NR; Physical Layer Procedures for Data*, document TS 138 214, 3GPP, Jul. 2020.

[10] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

[11] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5117–5144, Sep. 2016.

[12] C. Li, W. Win, H. Jing, and Y. Zhang, "An efficient augmented Lagrangian method with applications to total variation minimization," *Comput. Optim. Appl.*, vol. 56, no. 3, pp. 507–530, Dec. 2013.

[13] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[14] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, Apr. 2019.

[15] C. Lu, W. Xu, H. Shen, J. Zhu, and K. Wang, "MIMO channel information feedback using deep recurrent network," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 188–191, Jan. 2019.

[16] X. Li and H. Wu, "Spatio–temporal representation with deep neural recurrent network in MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 653–657, May 2020.

[17] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2827–2840, Apr. 2020.

[18] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[19] Z. Hu, J. Guo, G. Liu, H. Zheng, and J. Xue, "MRFNet: A deep learning-based CSI feedback approach of massive MIMO systems," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3310–3314, Oct. 2021.

[20] X. Chen, C. Deng, B. Zhou, H. Zhang, G. Yang, and S. Ma, "High-accuracy CSI feedback with super-resolution network for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 141–145, Jan. 2022.

[21] Z. Cao, W.-T. Shih, J. Guo, C.-K. Wen, and S. Jin, "Lightweight convolutional neural networks for CSI feedback in massive MIMO," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2624–2628, Aug. 2021.

[22] S. Ji and M. Li, "CLNet: Complex input lightweight neural network designed for massive MIMO CSI feedback," *IEEE Commun. Lett.*, vol. 10, no. 10, pp. 2318–2322, Oct. 2021.

[23] H. Tang, J. Guo, M. Matthaiou, C.-K. Wen, and S. Jin, "Knowledge-distillation-aided lightweight neural network for massive MIMO CSI feedback," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2021, pp. 1–5.

[24] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[25] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital–analog source–channel coding for bandwidth compression/expansion," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3757–3763, Aug. 2006.

[26] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem, and E. Hossain, "Hybrid automatic repeat request (HARQ) in wireless communications systems and standards: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2711–2752, 4th Quart., 2021.

[27] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," 2021, *arXiv:2201.01389*.

[28] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2326–2330.

[29] E. Bourtsoulatze, D. B. Kurka, and D. Gunduz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[30] D. B. Kurka and D. Gunduz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 178–193, May 2020.

[31] D. B. Kurka and D. Gunduz, "Bandwidth-agile image transmission with deep joint source-channel coding," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8081–8095, Dec. 2021.

[32] T.-Y. Tung and D. Gündüz, "DeepWiVe: Deep-learning-aided wireless video transmission," 2021, *arXiv:2111.13034*.

[33] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2315–2328, Apr. 2022.

[34] J. Xu, B. Ai, W. Chen, N. Wang, and M. Rodrigues, "Deep joint encryption and source-channel coding: An image visual protection approach," 2021, *arXiv:2111.03234*.

[35] K. Sayood, H. H. Otu, and N. Demir, "Joint source/channel coding for variable length codes," *IEEE Trans. Commun.*, vol. 48, no. 5, pp. 787–794, May 2000.

[36] M. Yang, C. Bian, and H.-S. Kim, "Deep joint source channel coding for wireless image transmission with OFDM," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.

[37] C. Lu, W. Xu, S. Jin, and K. Wang, "Bit-level optimized neural network for multi-antenna channel quantization," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 87–90, Jan. 2020.

[38] Z. Lu, J. Wang, and J. Song, "Binary neural network aided CSI feedback in massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1305–1308, Jun. 2021.

[39] M. B. Mashhadi, Q. Yang, and D. Gündüz, "Distributed deep convolutional compression for massive MIMO CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2621–2633, Apr. 2021.

[40] H. Ye, F. Gao, J. Qian, H. Wang, and G. Y. Li, "Deep learning-based denoise network for CSI feedback in FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1742–1746, Aug. 2020.

[41] M. B. Mashhadi, Q. Yang, and D. Gunduz, "CNN-based analog CSI feedback in FDD MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8579–8583.

[42] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," 2016, *arXiv:1611.01704*.

[43] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[44] M. Rabbani, "JPEG2000: Image compression fundamentals, standards and practice," *J. Electron. Imag.*, vol. 11, no. 2, p. 286, 2002.

[45] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[46] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa-quasi deterministic radio channel generator, user manual and documentation," Fraunhofer Heinrich Hertz Inst., Berlin, Germany, Tech. Rep., v2.6.1, 2021.

[47] *5G; Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document 38.901 V16.1.0, 3GPP, Nov. 2020.

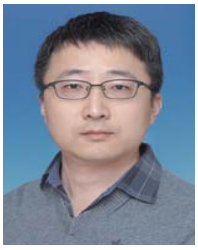
[48] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[49] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
[50] *5G; NR; Multiplexing and Channel Coding*, document TS 138 212, 3GPP, Jul. 2020.

Jialong Xu (Student Member, IEEE) received the B.E. and M.S. degrees from the Engineering University of PAP in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China. His research interests include deep learning, wireless coding, and information theory.

Bo Ai (Fellow, IEEE) received the M.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2002 and 2004, respectively.

He was with Tsinghua University, Beijing, China, where he was an Excellent Post-Doctoral Research Fellow in 2007. He is currently a Professor and an Advisor of Ph.D. candidates with Beijing Jiaotong University, Beijing, where he is also the Deputy Director of the State Key Laboratory of Rail Traffic Control and Safety. He is also with the Engineering College, Armed Police Force, Xi'an. He has authored or coauthored six books and 270 scientific research papers, and holds 26 invention patents in his research areas. His research interests include the research and applications of orthogonal frequency-division multiplexing techniques, high-power amplifier linearization techniques, radio propagation and channel modeling, global systems for mobile communications for railway systems, and long-term evolution for railway systems. He is a fellow of the Institution of Engineering and Technology. He was as the Co-Chair or the Session Chair/the Track Chair for many international conferences, such as the Ninth International Heavy Haul Conference in 2009, the 2011 IEEE International Conference on Intelligent Rail Transportation, HSRCom2011, the 2012 IEEE International Symposium on Consumer Electronics, the 2013 International Conference on Wireless, Mobile and Multimedia, IEEE Green HetNet 2013, and the IEEE 78th Vehicular Technology Conference in 2014. He is an Associate Editor of IEEE TRANSACTIONS ON CONSUMER ELECTRONICS and an Editorial Committee Member of the *Wireless Personal Communications* journal. He has received many awards, such as the Qiushi Outstanding Youth Award by Hong Kong Qiushi Foundation, the New Century Talents by the Chinese Ministry of Education, the Zhan Tianyou Railway Science and Technology Award by the Chinese Ministry of Railways, and the Science and Technology New Star by the Beijing Municipal Science and Technology Commission.

Ning Wang (Member, IEEE) received the B.E. degree in communication engineering from Tianjin University, China, in 2004, the M.A.Sc. degree in electrical engineering from The University of British Columbia, Canada, in 2010, and the Ph.D. degree in electrical engineering from the University of Victoria, Canada, in 2013. He was the Finalist of the Governor General's Gold Medal for Outstanding Graduating Doctoral Student with the University of Victoria in 2013. From 2004 to 2008, he was with the China Information Technology Design and Consulting Institute as a Mobile Communication System Engineer, specializing in planning and design of commercial mobile communication networks, network traffic analysis, and radio network optimization. He was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia, from 2013 to 2015. Since 2015, he has been with the School of Information Engineering, Zhengzhou University, Zhengzhou, China, where he is currently an Associate Professor. He also holds adjunct appointments with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, and the Department of Electrical and Computer Engineering, University of Victoria, Victoria, Canada. He has served on the technical program committees of international conferences, including the IEEE GLOBECOM, IEEE ICC, IEEE WCNC, and CyberC. His research interests include resource allocation and security designs of future cellular networks, channel modeling for wireless communications, statistical signal processing, and cooperative wireless communications.



Wei Chen (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in computer science from the University of Cambridge, Cambridge, U.K., in 2013. He was a Research Associate with the Computer Laboratory, University of Cambridge, from 2013 to 2016. He is currently a Professor with Beijing Jiaotong University, Beijing. His current research interests include sparse representation, Bayesian inference, wireless communication systems, and image processing. He was a recipient of the 2013 IET Wireless Sensor Systems Premium Award and the 2017 International Conference on Computer Vision (ICCV) Young Researcher Award.