

Deep Joint Source-Channel Coding for Image Transmission in Hardware-Impaired OFDM Systems

Pei-Jung Chen, Po-Hung Cheng, and Chia-Han Lee

National Yang Ming Chiao Tung University, Taiwan

Abstract—Deep joint source-channel coding (DeepJSCC) has recently become the state-of-the-art for wireless image transmission, but its practical deployment faces two challenges. Firstly, conventional DeepJSCC transmissions are analog, which is incompatible with standard orthogonal frequency-division multiplexing (OFDM) systems. Secondly, it is unclear whether DeepJSCC is robust to hardware impairments. To solve these two problems, we incorporate a dedicated quantizer and constellation mapper between the DeepJSCC encoder and the OFDM modulator. Furthermore, we conduct an investigation into the impact of hardware impairments, including clipping, I/Q imbalance, phase noise, power amplifier nonlinearity, carrier frequency offset, and sampling clock offset, on system performance. Two state-of-the-art DeepJSCC architectures, SwinJSCC and MambaJSCC, are adopted for evaluations. Experimental results show that at low signal-to-noise ratio (SNR), both quantized SwinJSCC and MambaJSCC significantly outperform the traditional BPG+LDPC no matter whether the OFDM system has hardware impairments or not. Moreover, at high SNR, both quantized SwinJSCC and MambaJSCC outperform the traditional BPG+LDPC when the OFDM system is hardware-impaired. This demonstrates the robustness of DeepJSCC under the impact of hardware impairments in OFDM systems.

I. INTRODUCTION

With the rapid evolution of deep learning, deep joint source-channel coding (DeepJSCC) has recently become the state-of-the-art for wireless image transmission [1]. However, the practical integration of DeepJSCC into orthogonal frequency-division multiplexing (OFDM) systems remains challenging. The first challenge is that conventional DeepJSCC transmissions are analog, which is incompatible with standard OFDM systems. To facilitate a digital implementation of DeepJSCC in OFDM systems, in this paper we incorporate a dedicated quantizer and constellation mapper to transform the analog encoder latent representations into discrete digital symbols for OFDM modulation. To ensure end-to-end differentiability, we employ a soft-to-hard quantization strategy, utilizing hard quantization in the forward pass for realistic digital signaling and soft quantization in the backward pass to maintain gradient flow. The second challenge is that it is unclear whether DeepJSCC is robust to hardware impairments, particularly when the output of DeepJSCC is quantized for bridging standard OFDM systems. Most existing studies evaluate performances of OFDM

and/or DeepJSCC systems under idealized channel conditions or focus on single hardware non-ideality. For example, most works of OFDM systems investigate the impact of either power amplifier (PA) nonlinearity [2], clipping distortion [3], or I/Q imbalance and carrier frequency offset (CFO) [4]. For DeepJSCC-OFDM systems, the papers [5] and [6] only address non-linear signal clipping. While some recent literature has begun to investigate the simultaneous effects of multiple hardware impairments on DeepJSCC [7], the performance analysis of hardware-impaired OFDM systems with quantized DeepJSCC is still unexplored.

In this paper, we conduct a comprehensive evaluation of hardware-impaired quantized DeepJSCC OFDM systems with two state-of-the-art DeepJSCC architectures, SwinJSCC [8] and MambaJSCC [9]. SwinJSCC integrates the Swin Transformer [10] into the DeepJSCC framework, constructing hierarchical feature maps and leveraging shifted window-based self-attention to capture multi-scale visual features with linear computational complexity. More recently, MambaJSCC, which is based on Mamba [11], a state space model (SSM), utilizes a bidirectional scan mechanism to achieve a global receptive field and superior long-range dependency modeling, while maintaining linear scaling relative to the image size. Furthermore, multiple hardware non-idealities, including clipping, I/Q imbalance, phase noise, PA nonlinearity, CFO, and sampling clock offset (SCO), are simultaneously considered.

The performance of quantized DeepJSCC systems is compared to that of the traditional separation-based scheme with BPG image compression [12] and low-density parity-check (LDPC) code [13]. Experimental results show that at low signal-to-noise ratio (SNR), both quantized SwinJSCC and MambaJSCC significantly outperform the traditional BPG+LDPC no matter whether the OFDM system is affected by hardware impairments or not. Moreover, at high SNR, both quantized SwinJSCC and MambaJSCC outperform BPG+LDPC when the OFDM system is hardware-impaired. This demonstrates the robustness of DeepJSCC under the impact of hardware impairments in OFDM systems.

II. QUANTIZED DEEPJSCC OFDM SYSTEM WITH HARDWARE IMPAIRMENTS

The DeepJSCC OFDM system considered in this paper is illustrated in Fig. 1. The DeepJSCC encoder takes an input image and generates complex-valued latent representations. These symbols are subsequently quantized and mapped onto

P.-J. Chen, P.-H. Cheng, and C.-H. Lee are with the Institute of Communications Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan (e-mail: chiahan@nycu.edu.tw). This work is supported in part by MediaTek and the National Science and Technology Council (NSTC), Taiwan under grants 114-2622-8-A49-010 and 112-2221-E-A49-078-MY3.

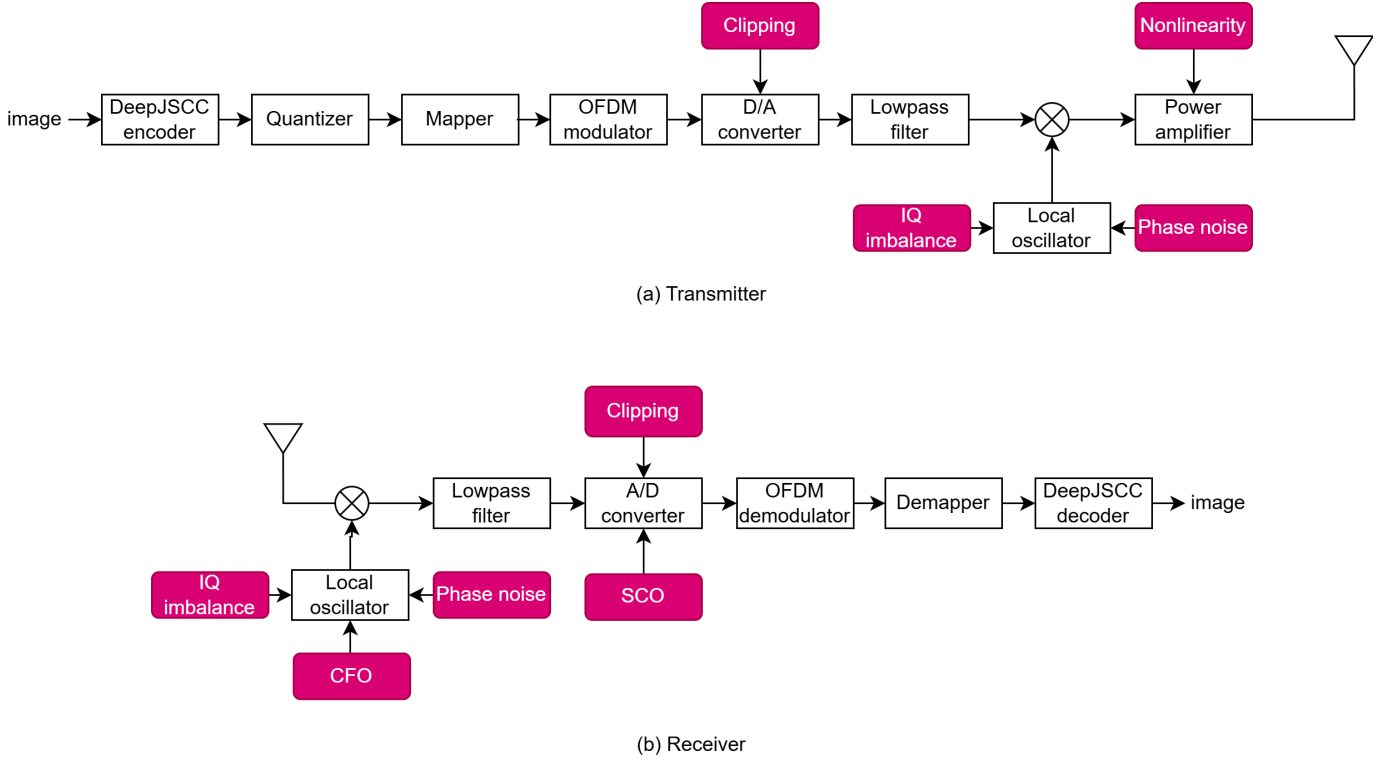


Fig. 1: The quantized DeepJSCC-based OFDM system with hardware impairments.

N OFDM subcarriers, followed by front-end processing before transmission. At the receiver, after front-end processing, OFDM demodulation, and demapping, the DeepJSCC decoder reconstructs the image. Note that the front-ends experience hardware impairments, as illustrated in the figure.

A. Quantization and Constellation Mapping

In order to be compatible with the standard OFDM modulation, the latent vector generated by the DeepJSCC encoder is processed by a quantization layer and subsequently mapped onto a set of discrete constellation points. To maintain end-to-end differentiability during the training phase, we adopt a soft-to-hard quantization strategy [14]. This approach utilizes a differentiable soft approximation for gradient estimation in the backward pass, while employing hard quantization in the forward pass to reflect realistic digital signaling.

B. Hardware Impairments

The following hardware impairments are considered.

Clipping: Due to the limited dynamic range of the analog-to-digital converter (ADC) and digital to analog converter (DAC), the signal $x[n]$ undergoes non-linear clipping when its amplitude exceeds a threshold $A_{\max}$, with the output $y[n]$:

$$y[n] = \begin{cases} x[n], & |x[n]| \leq A_{\max}, \\ A_{\max} \text{sign}(x[n]), & |x[n]| > A_{\max}. \end{cases} \quad (1)$$

Power Amplifier Nonlinearity: The PA is susceptible to saturation, leading to gain compression. For a third-order nonlinearity, the lowpass complex equivalent output $y(t)$ is

modeled as $y(t) = x(t) (a_1 + \frac{3a_3}{4}|x(t)|^2)$, where $x(t)$ is the input, a_1 is the linear gain, and a_3 is a negative coefficient characterizing the compression effect.

I/Q Imbalance: Caused by gain and phase mismatches (ϵ, ϕ) between the in-phase and quadrature branches, the impaired signal $y(t)$ is expressed as $y(t) = K_1 x(t) + K_2 x^*(t)$, where $K_1 = \frac{1}{2}(1 + (1 + \epsilon)e^{-j\phi})$, $K_2 = \frac{1}{2}(1 - (1 + \epsilon)e^{j\phi})$, and $x^*(t)$ is the complex conjugate of the input $x(t)$. This distortion destroys signal orthogonality, resulting in inter-carrier interference (ICI).

Carrier Frequency Offset: Mismatches between the transmitter and receiver local oscillators lead to a frequency shift Δf . The baseband equivalent effect is modeled as a continuous phase rotation, i.e., $y(t) = x(t)e^{j2\pi\Delta f t}$.

Phase Noise: Random phase fluctuations $\phi(t)$ arise from instabilities in the frequency synthesizer, and cause the constellation rotation in the quadrature direction. This effect is modeled as $y(t) = x(t)e^{j\phi(t)}$.

Sampling Clock Offset: Timing drifts δ between the sampling frequencies of the DAC and ADC result in a cumulative time-dependent offset. The received signal $y[n]$ is modeled as $y[n] = y(nT + n\delta T)$, $0 \leq n \leq N-1$, where T is the sampling period and $n\delta T$ is the increasing timing delay.

C. DeepJSCC Architectures

Two state-of-the-art DeepJSCC architectures are considered: SwinJSCC and MambaJSCC. SwinJSCC utilizes Transformer-based blocks as its primary computational engine. SwinJSCC focuses on extracting multi-scale visual features by em-

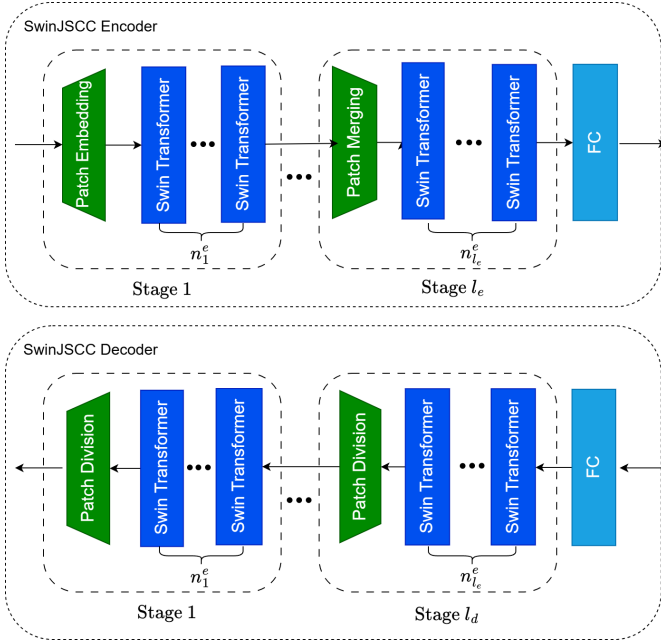


Fig. 2: Architecture of SwinJSCC.

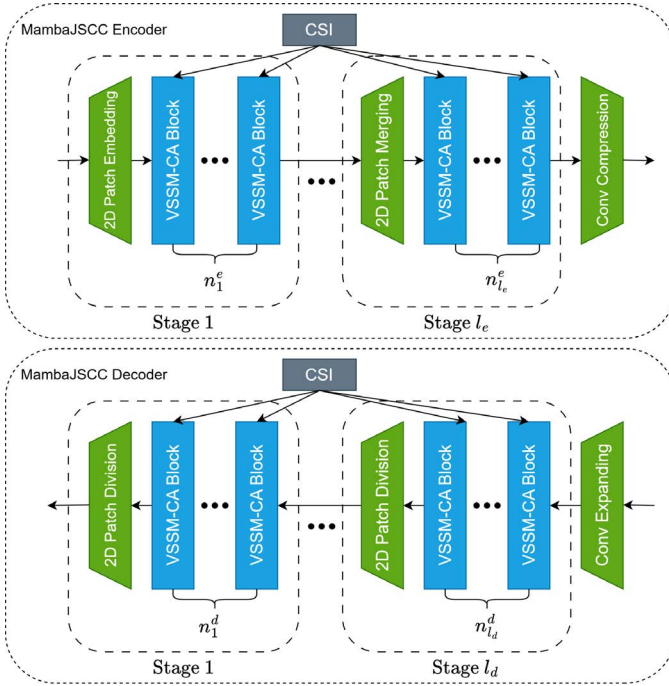


Fig. 3: Architecture of MambaJSCC.

employing shifted-window self-attention mechanisms to balance computational efficiency with effective feature representation. As shown in Fig. 2, the SwinJSCC encoder facilitates the transformation of the input image into latent representations through l_e coding stages. In stage 1, the image is partitioned into patches via a patch embedding layer and subsequently refined by n_1^e Swin Transformer blocks. Subsequent stages utilize patch merging layers to progressively reduce spatial resolution while increasing channel depth. After the final stage

(stage l_e), which comprises $n_{l_e}^e$ blocks, a fully connected (FC) layer projects these high-dimensional hierarchical features into the latent space for transmission. The SwinJSCC decoder reconstructs the image by reversing the encoding process. An initial FC layer projects the received latent features back into the hierarchical dimensions, followed by l_d reconstruction stages. Stage i employs n_i^d Swin Transformer blocks. Between these stages, patch division layers are employed to restore the spatial resolution.

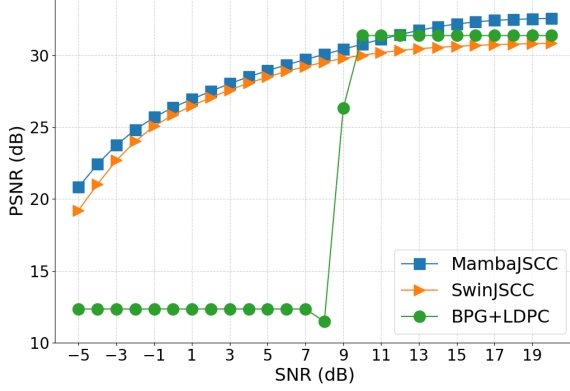
As shown in Fig. 3, the MambaJSCC encoder transforms the input image into channel-ready signals through l_e encoding stages. In the initial stage, the image is processed by a two-dimensional (2D) patch embedding layer to generate latent feature patches, which are then refined by n_1^e visual state space model with channel adaptation (VSSM-CA) blocks. To capture multi-scale semantic information, subsequent stages employ 2D patch merging to progressively reduce spatial resolution while increasing channel depth. In the final encoder stage (stage l_e), which consists of $n_{l_e}^e$ blocks, a convolutional compression module is utilized to compress the feature patches into the designated number of channel uses for transmission. The MambaJSCC decoder reconstructs the image by reversing the encoding hierarchy across l_d stages. A convolutional expanding module first restores the signal to the required patch dimensions, and then n_i^d VSSM-CA blocks are employed for stage i . Between these stages, 2D patch division layers are applied to progressively recover the spatial resolution until the final image is reconstructed. To enhance robustness across diverse channel environments, channel state information (CSI), such as SNR, is injected into each VSSM-CA block.

III. SIMULATION RESULTS

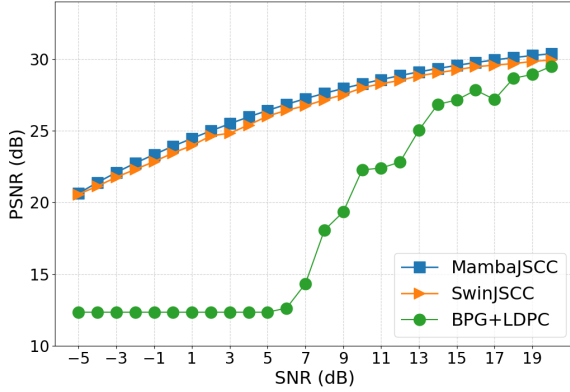
In this section, we compare the performances of quantized SwinJSCC-based and MambaJSCC-based OFDM systems with the traditional separation-based scheme BPG+LDPC under scenarios without and with hardware impairments.

A. Training and Simulation Setup

We consider an OFDM system consisting of $N = 64$ sub-carriers with a cyclic prefix (CP) length of 4. Both SwinJSCC and MambaJSCC utilize a symmetric 4-stage hierarchical configuration with identical embedding dimensions of [128, 192, 256, 320] and depths of [2, 2, 6, 2]. The DeepJSCC models are trained for 150 epochs using mean-square error (MSE) as the loss function. The batch size is 4, the initial learning rate is 10^{-4} , and the cosine scheduler is adopted as the learning rate annealing strategy. The AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay = 10^{-4} is used. For MambaJSCC, the cross-scan strategy is used. To enhance the models' adaptability to varying channel qualities, the SNR during the training phase is randomly sampled for each image from the discrete set $\{0, 5, 10, 15, 20\}$ dB. Both DeepJSCC architectures are trained under ideal channel conditions without hardware impairments. The experiments are conducted using the CelebA-HQ dataset [15], with images resized to 256×256 . For the baseline scheme, we employ BPG for image



(a) AWGN channel



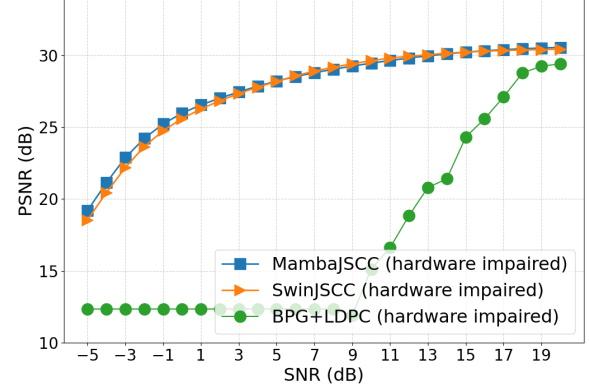
(b) Rayleigh fading channel

Fig. 4: Performances of MambaJSCC, SwinJSCC, and BPG+LDPC OFDM systems without hardware impairments.

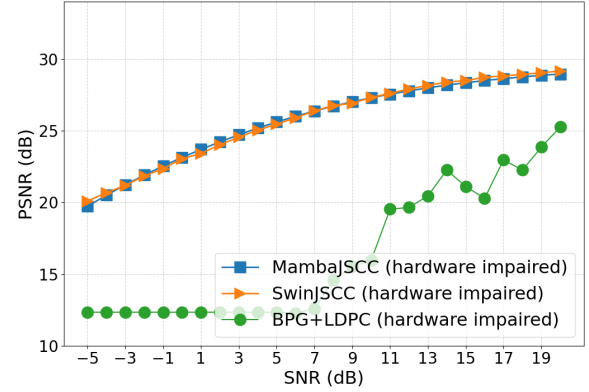
compression, followed by a rate 2/3 LDPC code with 16-QAM modulation. Both the training and the testing are performed on a personal desktop computer equipped with Intel Core i5-13500 @ 2.50 GHz, 96G DRAM, and NVIDIA GeForce RTX 4070 graphics processing unit (GPU).

B. Performance under Ideal Conditions

We first evaluate the peak signal-to-noise ratio (PSNR) performance in a hardware-ideal scenario, as shown in Fig. 4. For the additive white Gaussian noise (AWGN) channel, when the SNR is below 8 dB, the BPG+LDPC baseline scheme suffers from a catastrophic failure, as it is unable to reconstruct the image. In contrast, both DeepJSCC models effectively mitigate the cliff effect, achieving good PSNR at low SNR. In the high SNR regime, the traditional BPG+LDPC scheme achieves a higher PSNR than SwinJSCC, but is surpassed by MambaJSCC. While MambaJSCC and SwinJSCC yield slightly lower PSNR values than BPG+LDPC at SNR = 10 dB, they produce much smoother reconstructions with better preserved visual details, as shown in Fig. 6. It can be seen that the BPG+LDPC scheme produces visual discontinuities and blocky artifacts. For the Rayleigh fading channel,



(a) AWGN channel



(b) Rayleigh fading channel

Fig. 5: Performances of MambaJSCC, SwinJSCC, and BPG+LDPC OFDM systems with hardware impairments.

both MambaJSCC and SwinJSCC outperform BPG+LDPC throughout all SNR regions, and MambaJSCC has a slightly better performance than SwinJSCC.

C. Robustness against Hardware Impairments

The impact of hardware impairments is evaluated in Fig. 5. For the AWGN channel, it is shown that the traditional BPG+LDPC baseline fails to reconstruct images until the SNR exceeds 10 dB. Even at high SNR, the BPG+LDPC scheme does not perform well. In particular, its visual output remains corrupted by random color blocks, particularly in the bottom-right regions of the image, as illustrated in Fig. 6. In contrast, SwinJSCC and MambaJSCC demonstrate remarkable robustness to hardware impairments, as their image reconstruction quality under severe hardware impairments remains close to their performance in ideal conditions. This suggests that (quantized) DeepJSCC architectures have inherent capability to compensate for hardware-induced distortions without requiring explicit impairment-aware training. While MambaJSCC shows a slight advantage over SwinJSCC at low SNR, both models exhibit nearly identical performance in the mid-to-high SNR range, significantly outperforming BPG+LDPC. For the



Fig. 6: Visualization of images samples for different schemes transmitted over the AWGN channel (SNR = 10 dB).

TABLE I: COMPUTATIONAL COMPLEXITY.

Model	FLOPs	Inference Delay	# Parameters
SwinJSCC	32.72 G	20.56 ms	28.16 M
MambaJSCC	22.68 G	22.49 ms	14.54 M
BPG+LDPC	-	4.628 s	-

Rayleigh fading channel, both MambaJSCC and SwinJSCC significantly outperform BPG+LDPC throughout all SNR regions, while MambaJSCC and SwinJSCC perform similarly.

D. Computational Complexity

The computational complexity of SwinJSCC, MambaJSCC, and BPG+LDPC is compared in TABLE I. MambaJSCC requires approximately 48% fewer parameters compared to SwinJSCC (14.54 M vs 28.162 M). Furthermore, MambaJSCC achieves lower computational complexity (FLOPs) during inference, although with a slightly larger inference delay. Most importantly, both SwinJSCC and MambaJSCC have much shorter inference delay compared to BPG+LDPC.

IV. CONCLUSIONS

In this paper, we have evaluated the performance and robustness of MambaJSCC and SwinJSCC within a hardware-impaired OFDM system. To make the DeepJSCC-based system practical, we integrated a dedicated quantizer and modeled hardware impairments across the transmitter and receiver. Experimental results demonstrate that both quantized SwinJSCC and MambaJSCC architectures achieve higher PSNR and superior perceptual quality compared to the traditional BPG+LDPC baseline scheme in OFDM systems, especially under the influence of hardware impairments. It is thus concluded that both SwinJSCC and MambaJSCC exhibit an inherent ability to compensate for severe hardware impairments.

REFERENCES

- [1] Z. Lyu, G. Zhu, J. Xu, B. Ai, and S. Cui, "Semantic communications for image recovery and classification via deep joint source and channel coding," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8388–8404, Aug. 2024.
- [2] M. Sajedin, A. Ghorbani, and H. R. Amindavar, "Nonlinearity compensation for high power amplifiers based on look-up table method for OFDM transmitters," *Int. J. Adv. Comput. Sci. Inf. Technol. (IJACSIT)*, vol. 3, no. 4, pp. 354–367, 2014.
- [3] S. Liang, J. Tong, and L. Ping, "On iterative compensation of clipping distortion in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 436–439, Apr. 2019.
- [4] J.-Y. Yu, M.-F. Sun, T.-Y. Hsu, and C.-Y. Lee, "A novel technique for I/Q imbalance and CFO compensation in OFDM systems," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2005, pp. 6030–6033.
- [5] 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. (ICC)*, 2021, pp. 1–6.
- [6] M. Yang, C. Bian, and H.-S. Kim, "OFDM-guided deep joint source channel coding for wireless multipath fading channels," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 2, pp. 584–599, June 2022.
- [7] C.-Y. Wu, Y.-K. Lin, C.-K. Wu, and C.-H. Lee, "Deep learning-based end-to-end design for OFDM systems with hardware impairments," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 2468–2482, 2023.
- [8] K. Yang, S. Wang, J. Dai, X. Qin, K. Niu, and P. Zhang, "SwinJSCC: Taming Swin Transformer for deep joint source-channel coding," *IEEE Trans. Cognit. Commun. Netw.*, vol. 11, no. 1, pp. 90–105, Feb. 2025.
- [9] T. Wu *et al.*, "MambaJSCC: Adaptive deep joint source-channel coding with generalized state space model," *IEEE Trans. Wireless Commun.*, vol. 25, pp. 9264–9279, Dec. 2025.
- [10] Z. Liu *et al.*, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 10012–10022.
- [11] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.
- [12] F. Bellard, "BPG image format," 2014. [Online]. Available: <http://bellard.org/bpg/>
- [13] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.
- [14] T.-Y. Tung, D. B. Kurka, M. Jankowski, and D. Gündüz, "DeepJSCC-Q: Constellation constrained deep joint source-channel coding," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 4, pp. 720–733, Dec. 2022.
- [15] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. International Conference on Learning Representations (ICLR)*, 2018.