Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

By Deniz Gündüz, *Fellow IEEE*, Michèle A. Wigger, *Senior Member IEEE*,
Tze-Yang Tung, *Member IEEE*, Ping Zhang, *Fellow IEEE*, and Yong Xiao, *Senior Member IEEE*

ABSTRACT | Semantic- and task-oriented communication has emerged as a promising approach to reducing the latency and bandwidth requirements of the next-generation mobile networks by transmitting only the most relevant information needed to complete a specific task at the receiver. This is particularly advantageous for machine-oriented communication of high-data-rate content, such as images and videos, where the goal is rapid and accurate inference, rather than perfect signal reconstruction. While semantic- and task-oriented compression can be implemented in conventional communication systems, joint source–channel coding (JSCC) offers an alternative end-to-end approach by optimizing compression and channel coding together, or even directly mapping the source signal to the modulated waveform. Although all digital communication systems today rely on separation, thanks to its modularity, JSCC is known to achieve higher performance in finite blocklength scenarios and to avoid *cliff* and the *leveling-off effects* in time-varying channel scenarios. This article provides an overview of the information theoretic foundations of JSCC, surveys practical JSCC designs over the decades, and discusses the reasons for their limited adoption in practical systems. We then examine the recent resurgence of JSCC, driven by the integration of deep learning techniques, particularly through DeepJSCC, highlighting its many surprising advantages in various scenarios. Finally, we discuss why it may be time to reconsider today's strictly separate architectures and reintroduce JSCC to enable high-fidelity, low-latency communications in critical applications such as autonomous driving, drone surveillance, or wearable systems.

KEYWORDS | Coding theory; deep neural networks; DeepJSCC; joint source–channel coding (JSCC); machine learning (ML); multi-user information theory; semantic communications.

Deniz Gündüz is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT London, U.K. (e-mail: d.gunduz@imperial.ac.uk).
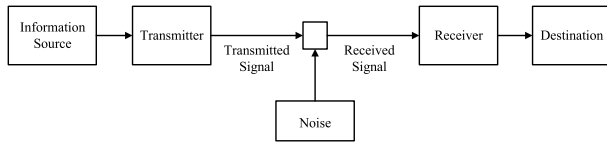
Michèle A. Wigger is with LTCI, Télécom Paris, IP Paris, 91120 Palaiseau, France (e-mail: michele.wigger@telecom-paris.fr).

Tze-Yang Tung was with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT London, U.K. He is now with Not Diamond, San Francisco, CA 94114 USA (e-mail: tze-yang.tung14@imperial.ac.uk).

Ping Zhang is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the Department of Broadband Communication, Pengcheng Laboratory, Shenzhen, Guangdong 518055, China (e-mail: pzhang@bupt.edu.cn).

Yong Xiao is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China, with the Pengcheng Laboratory, Shenzhen, Guangdong 518055, China, and also with the Pazhou Laboratory (Huangpu), Guangzhou, Guangdong 510555, China (e-mail: yongxiao@hust.edu.cn).

I. INTRODUCTION

Shannon [1] defined the fundamental problem of communication as that of "reproducing at one point either exactly or approximately a message selected at another point." A generic point-to-point communication system, reproduced in Fig. 1 from [1], consists of an information source, a transmitter, a communication channel, and a receiver. The information source, which generates the messages to be transmitted, and the channel are typically assumed to be given, while the goal of the system designer is to come up with the transmitter and receiver

Fig. 1. *Schematic of a general communication system according to Shannon (reproduced from [1]).*

mappings. The former, called the *encoding function*, maps the message to a form that is amenable to be transmitted over the channel, while the latter, called the *decoding function*, is a complementary mapping that recovers the message under a prescribed fidelity measure. The fidelity measure, also assumed to be given, maps every message and reconstruction pair to a real number, quantifying the goodness of the reconstruction.

The problem stated in this form is fairly general: the information source can be a sequence of discrete symbols from a prescribed alphabet or a function of time and other variables. In general, the information source is a stochastic process, and the communication system should operate for any source realization as it is not known at the time of design. The communication channel is also modeled as a stochastic kernel from the input alphabet space to the output alphabet space, which depends on the communication medium under consideration. We may assume the statistical properties of the information source and the channel as fixed and known, or they may belong to a certain set of possible statistics [2], [3], or we can also assume that the statistics are fixed yet unknown, but we can sample from these distributions as desired.

The design of a communication system in this full generality is highly challenging. The transmitter and receiver must be designed jointly considering both the source and channel characteristics, that is, their alphabets as well as statistics. Some of the earliest practical telecommunication systems, amplitude/frequency modulation (AM/FM), analog TV broadcasting, and the first generation of mobile wireless networks have all been based on this approach. On the other hand, a fundamental result in Shannon's information theory is the Separation Theorem [1], which states that for a large class of point-to-point communication scenarios characterized by large block lengths, optimal performance can be effectively approached by independently formulating the compression and error correction schemes. Equivalently, the problem of mapping the input source message to the channel input and recovering the input signal from the noisy received channel output can be reformulated into two subproblems of *source coding* and *channel coding* without loss of optimality.

The source coding problem deals with mapping the message to the minimum number of bits so that it can be reconstructed within the desired fidelity. Shannon's source coding theorem identifies the fundamental limit of this problem, characterized by the so-called *rate–distortion function*, considering an arbitrarily long sequence of source samples and an additive distortion measure. This function provides a bound on the minimum number of bits per source sample that must be transmitted to the receiver to achieve the desired fidelity. On the other hand, Shannon's channel coding theorem identifies the fundamental limit of communications over a noisy channel, characterized by the channel capacity. Similar to the source coding theory, channel coding theory assumes infinitely many uses of the channel and identifies the maximum number of bits per channel use that can be conveyed reliably over the channel.

In the point-to-point setting, the Separation Theorem suggests that if the compression rate corresponding to the target distortion of the given source distribution is below the channel capacity of the link, then the desired distortion can be achieved. Conversely, if this rate is above the capacity, there is no coding scheme that can achieve this distortion level over this channel (more rigorous definitions and statements will be provided later in this article). Thanks to the converse theorem, we conclude that the desired distortion level can either be achieved by designing separate source and channel codes, or it is impossible to achieve.

In separate source and channel coding (SSCC), the compression scheme is designed oblivious to the channel statistics, while the channel code is designed to ignore the source characteristics. This separate design principle leads to a natural modular system architecture, allowing a single network infrastructure that can serve a great variety of services, e.g., voice, data, and multimedia. Indeed, almost all contemporary communication systems adhere to such a structured layered architecture, in which compression-related functions are primarily dealt with at the application layer, at the top of the protocol stack, while channel coding functions are taken care of at the link and physical layers, at the bottom of the stack.

This modularity in design combined with the theoretical optimality led to a natural separation among the research communities as well. Researchers in the data compression community have mainly focused on developing high-performance compression algorithms by exploiting properties of particular types of source signals, leading to various standards for text, image, audio, and video compression, among others. In parallel, communication researchers have focused on designing advanced coding and modulation techniques with the goal of turning the noisy wireless medium into a network of noise-free bit pipes.

Although a segregated architecture offers indisputable advantages in facilitating modular system design and enabling the convergence of diverse services on a shared data network infrastructure, there are instances where a joint source–channel coding (JSCC) approach becomes essential. This need arises both in various multiterminal settings where the separated approach proves suboptimal, as well as in standard point-to-point channels. Even in cases where SSCC is asymptotically optimal, the application of independently designed source and channel codes may result in subpar performance under practical conditions characterized by finite block length, low-complexity encoding/decoding, or communication over uncertain channels.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

JSCC has been a topic of continuous research since Shannon; however, there has been a significant surge in interest in recent years. This can be attributed to two main reasons, both are connected to the recent advances in machine learning (ML) technologies. First, as ML tools and applications become increasingly widespread, they are expected to be deployed on mobile edge devices, requiring distributed implementations of training as well as inference tasks [4], [5], [6], [7], [8], [9], [10]. These applications typically have much stricter latency requirements compared to content delivery applications, for which the current network architectures and communication schemes have been designed. The gap between the optimal performance that can be achieved through JSCC and what can be achieved by conventional SSCC tends to increase as the blocklengths become shorter. Moreover, for shorter blocklengths, resources used for channel estimation may become more significant, particularly for highly mobile environments. The modularity that separation provides is extremely valuable for larger network architectures, such as cellular networks; however, modularity can be sacrificed in point-to-point communication scenarios, where the performance gains become critical. For example, for the delivery of drone video footage to a ground station, the wireless connection of an augmented reality (AR)/virtual reality (VR) headset to a computing unit, or the exchange of LIDAR or video data among autonomous vehicles for collaborative perception, it may be crucial to go beyond the constraints of a strictly separate design.

Second, despite many years of efforts, the research on practical code design for JSCC has not produced codes that can be applied to a large variety of source–channel pairs with a performance-complexity tradeoff that is better than, or on par with the state-of-the-art separate baselines. This has changed recently with the advances in deep learning-aided JSCC design, called DeepJSCC [11], [12]. Such codes can be easily designed for different source modalities and datasets targeting any desired performance measure [13], [14], [15], [16], [17], providing highly competitive performance results.

Developments in JSCC design have been further encouraged with the recent growing interest in semantic and goal-oriented communication systems [18], [19], [20], [21], [22], [23], [24], [25], [26], driven by AR/VR applications, and the emerging concepts of metaverse [27], tactile/haptic communications [28], the Internet of senses [29], and holographic communications [30], which are foreseen to be the main drivers of 6G and beyond mobile networks. The motivation for semantic and goal-oriented communication comes from taming the growing network data traffic, particularly due to the emerging artificial intelligence (AI) applications involving AR/VR, autonomous vehicles, and machine vision. The efforts on the communication network design side have mainly focused on increasing the capacity of the network; however, gains from physical and network layer modifications (e.g., the introduction of a massive number of antennas or terminals) have become increasingly costly and incremental as we approach the corresponding fundamental limits. Hence, a complementary direction is to reduce the amount of information that must be communicated over the network. In a nutshell, the core idea behind semantic and goal-oriented communication is to transmit only the relevant information for the underlying task to the receiver, which can be considered as a form of compression [18]. Naturally, such an approach can be carried out purely on the application layer by employing the appropriate compression algorithm for the desired goal, reducing the number of bits that need to be conveyed. One can argue that this is already done for most multimedia applications that rely on various compression algorithms although new types of compression algorithms need to be designed for specific inference tasks (e.g., image retrieval [31] or other inference tasks [32], [33], [34], [35], [36]) since, in most AI applications, the goal is not to recover the underlying signal, but to carry out some intelligence on it. However, when the underlying communication channel is noisy, and we want to communicate under extreme bandwidth and/or power limitations, semantic and goal-oriented communications will significantly benefit from JSCC approaches to push the limits of communications for relevant applications [11], [15], [18], [31]. Moreover, many of the aforementioned applications, e.g., haptic communications and metaverse, impose strict delay constraints on communications [37]. While 1-ms round trip delay target over the communication network has been promoted widely for 5G networks to enable such applications [38], [39], these promises have not been realized although ultralow latency communication has been delivered by 5G implementations. This is because coding latency to implement state-of-the-art compression algorithms for high-data-rate contents such as video or holograms is significantly higher than 1 ms, appearing as a major bottleneck. An important promise of the JSCC approach is to reduce the coding latency by combining compression and channel coding into a single operation. Particularly, recent JSCC designs based on neural network architectures report promising results in terms of the end-to-end coding latency [11], [40]. As we will show in the first part of this article through information theoretic formulation, the gains from JSCC are fundamental, that is, they do not rely on a particular suboptimal implementation of separate coding approaches. In the second part, we will show practical code designs that can provide these theoretical gains in practical scenarios with an emphasis on the recent developments using deep learning techniques.

A. Outline and Organization

This article provides a comprehensive overview of JSCC in communication systems, starting from information theoretic foundations to practical designs, with particular emphasis on recent progress on deep learning-aided JSCC

schemes. Section II is dedicated to the information theoretic foundations. This part not only fills an important gap in the literature by providing a comprehensive overview of fundamental information theoretic performance limits and coding ideas on JSCC, but also intends to provide a foundation for the development of new neural network-based codes and protocols that are inspired by these foundations. We start with the introduction of Shannon's Separation Theorem in Section II-A for a point-to-point channel. Section II-C shows that the availability of channel output feedback, while does not have an impact on the optimality of separation, can allow a much simpler and low-delay JSCC scheme. We then introduce a generalized JSCC scheme when there is correlated side information at the receiver. This generalized scheme can specialize to a separation-based approach with explicit source binning, as well as a simpler encoding scheme with joint decoding. We show that this joint encoding scheme can be particularly beneficial when the quality of the channel and the side information can be both uncertain. Section II-E is dedicated to multiuser channels. We start by showing that the Separation Theorem does not hold in multiuser scenarios. Then, we consider multiple access and broadcast channels (BCs) separately and propose various JSCC schemes. In Section II-F, we discuss two different types of separation. In Section II-G, we show that JSCC can also help increase the rate of communication in certain multiuser scenarios. Considerations up until Section II-H follow Shannon's formulation: distortion between two sequences is the average distortion between pairs of symbols. In Section II-H, we consider distortion measures that do not necessarily satisfy this property.

Sections III and IV are dedicated to practical designs for JSCC. The former focuses on classical approaches, while the latter covers recent developments that rely on data-driven deep learning tools to design so-called DeepJSCC solutions. Section IV is organized into various subsections, each dedicated to a different source type. In particular, Section IV-A focuses on DeepJSCC for image transmission, Section IV-D to video transmission, Section IV-E to text transmission, while Section IV-F covers other source and channel distributions. Finally, Section V concludes this article.

II. INFORMATION THEORETIC FOUNDATIONS of JSCC

In this section, we will start with an overview of fundamental information theoretic results on JSCC, starting from Shannon's Separation Theorem, and covering various scenarios in which this theorem fails. In such cases, we also provide alternative coding schemes that can go beyond the performance achieved by pure separation, which also provides guidelines for the design of practical codes.

A. Fundamentals

In the fundamental problem of information theory as studied by Shannon [1], the transmitter wants to transmit
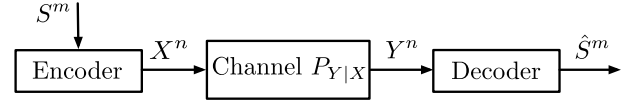
Fig. 2. *Illustration of a JSCC problem over a noisy communication channel.*

the output of an information source to a receiver over a noisy communication channel. In general, the goal of the receiver is to recover the observed source sequence either reliably (i.e., with vanishing error probability), or within some distortion limit under a prescribed distortion measure. Shannon formulated this problem using a statistical model for the source signal and the channel transform, and focused on a block coding framework in the asymptotic infinite blocklength regime. In particular, the model considered by Shannon is illustrated in Fig. 2 and described next in a mathematical form.

Assume that the encoder wants to transmit a sequence of independent source symbols $S^m \in \mathcal{S}^m$ sampled from the source distribution $p_S(s)$, over a memoryless noisy communication channel characterized by the conditional probability distribution $P(Y|X)$, where $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. Let $\hat{S}^m \in \hat{\mathcal{S}}^m$ denote the reconstruction at the receiver based on Y^n. Similar to the rate–distortion theory formulation, the goal is to minimize the distortion between S^m and $\hat{S}^m$ under some given distortion (fidelity) measure, $d : \mathcal{S}^m \times \hat{\mathcal{S}}^m \to [0, \infty)$. More formally, let $f^{m,n} : \mathcal{S}^m \to \mathcal{X}^n$ denote the encoding function, and $g^{m,n} : \mathcal{Y}^n \to \hat{\mathcal{S}}^m$ denote the decoding function. In the case of average distortion criteria, the goal is to identify the encoder and decoder function pairs that minimize $\mathbb{E}[d(S^m, \hat{S}^m)]$, where the expectation is over the source and channel distributions as well as any randomness the encoding and decoding functions may introduce. One can also impose an excess distortion criterion, where the goal is to minimize $\mathbb{P}[d(S^m, \hat{S}^m) > D]$, for some maximum allowable distortion target $D > 0$.

An (m, n) joint source–channel code of rate $r = m/n$ consists of an encoder–decoder pair, where the encoder $f^{(m,n)} : \mathcal{S}^m \to \mathcal{X}^n$ maps each source sequence s^m to a channel input sequence $x^n(s^m)$, and the decoder $g^{(m,n)} : \mathcal{Y}^n \to \hat{\mathcal{S}}^m$ maps the channel output y^n to an estimated source sequence $\hat{s}^m$.

Definition 1: A rate–distortion pair (r, D) is said to be *achievable* if there exists a sequence of $(m, n(m))$ joint source–channel codes with rate r such that $r \cdot n(m) \leq m$, $\forall m$, and

$$\limsup_{m \to \infty} \mathbb{E}\left[d\left(S^m, \hat{S}^m\left(Y^{n(m)} \right) \right) \right] \leq D. \tag{1}$$

Similar to the definition of channel capacity, we can define a source–channel capacity for a given distortion target, which considers both the source and the channel characteristics.
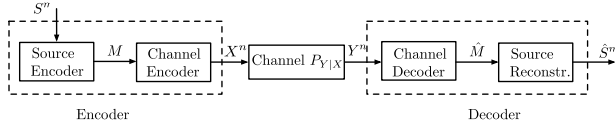
Fig. 3. *Separation-based architecture for the JSCC problem.*

Definition 2: For a given target distortion D, the *source–channel capacity* of a channel is defined as the supremum of rates r among all achievable (r, D) pairs.

Shannon proved his Separation Theorem considering single-letter additive distortion measures, that is,

$$d\left(S^m, \hat{S}^m\right) = \frac{1}{m} \sum_{i=1}^{m} d\left(S_i, \hat{S}_i\right) \qquad (2)$$

for the distortion measure $d(S, \hat{S}) \in [0, \infty)$. The Separation Theorem can be stated as follows, after defining the capacity of a discrete memoryless channel $P_{Y|X}$:

$$C := \sup_{p_X(x)} I\left(X; Y\right) \qquad (3)$$

and the rate–distortion function of a memoryless source p_S

$$R\left(D\right) := \inf_{P_{\hat{S}|S}} I\left(S; \hat{S}\right) \qquad (4)$$

where the infimum is taken only over conditional distributions satisfying $\mathrm{E}[d(S, \hat{S})] \leq D$.

Theorem 1 (Shannon's Separation Theorem, [1]): Given a memoryless source S with distribution p_S and a memoryless channel $p_{Y|X}$ with capacity C, a rate–distortion pair (r, D) is achievable if $rR(D) < C$. Conversely, if a rate–distortion pair (r, D) is achievable, then $rR(D) \leq C$. Then, for a given distortion target D, the source–channel capacity is given by $C/R(D)$.

The theorem states that we can separate the design of the communication system into two subproblems without loss of optimality, the first focusing on compression and the second focusing on channel coding, each of them designed independently of the other. This separation-based architecture is depicted in Fig. 3. However, the optimality of separation holds only in the limit of infinite blocklength and only in terms of capacity. For example, still in the asymptotic regime of infinite blocklengths, the exponential decay rate of the probability of error as well as the second-order coding rates can be improved with a joint source–channel code compared to the best separation-based approach (see [41], [42], [43], and [44]). The same observation also holds in the regime of finite blocklengths, where optimal joint source–channel codes outperform the best separation-based code design.

The suboptimality of separation in the general finite blocklength regime was observed by Shannon [45] in his 1959 article, where he considered a binary source generating independent and equiprobable symbols and

a memoryless binary symmetric channel, and observed that simple uncoded transmission of symbols achieves the optimal distortion with rate $r = 1$ for one particular value of distortion D determined by the error probability of the channel. This observation was later extended by Goblick [46] to Gaussian sources transmitted over Gaussian channels.

In general, uncoded transmission is optimal [47] when the source distribution matches the optimal capacity-achieving input distribution of the channel, and the channel at hand matches the optimal test channel achieving the optimal rate–distortion function of the source. That is, when the source and channel input alphabets match, $\mathcal{X} = \mathcal{S}$ and the source distribution p_S and the channel transition law $P_{Y|X}$ satisfy the following two conditions:

$$I\left(X; Y\right)|_{X \sim p_S} = C \qquad (5)$$

$$I\left(S; \hat{S}\right)|_{\hat{S} \sim P_{Y|X}(\cdot|S)} = R\left(D\right). \qquad (6)$$

However, these conditions are not satisfied for most practical source and channel distributions, and even when they hold, optimality of uncoded transmission fails when the coding rate is not 1, i.e., in the case of bandwidth compression or expansion. On the other hand, the presence of such optimality results, that is, the fact that asymptotically optimal performance, which requires infinite blocklength source and channel codes in general, can be achieved by simple zero-delay uncoded transmission implies that there can be other simple joint coding schemes that can achieve near optimal performance, and outperform separation-based schemes in the finite blocklength regime. Various JSCC schemes that perform well in the finite blocklength regime have been proposed (see [44], [48], [49], and [50]).

Interesting phenomena have also been proved for JSCC setups that are not memoryless, in particular for the special case where the source has to be reconstructed perfectly. While the Separation Theorem has been shown to continue to hold for a large class of sources and channels [51], [52], [53], Vembu et al. [54] and Han [55] have identified source–channel pairs where separation does not hold. In fact, they characterized matching necessary and sufficient conditions for a source to be transmittable over a channel for a large class of source distributions and channel law. Uncoded transmission, and thus JSCC, can also be advantageous under channel uncertainty, for example, over fading channels, where in certain cases (e.g., for the transmission of Gaussian sources over a Gaussian fading channel with receiver channel state information), it may allow attaining the optimal distortion under any given channel condition. SSCC scheme can either be in excess and thus result in huge distortion under bad channel conditions, or has to rely on more advanced and complex coding schemes (like the multilayer broadcast approach of [56], [57], and [58]).
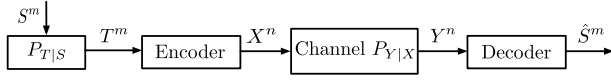
Fig. 4. *Illustration of a JSCC problem with a remote source S^m that has to be reconstructed at the decoder.*

B. JSCC With a Remote Source

A variation of the standard source coding problem that is particularly relevant to semantic communication is the setup of remote JSSC depicted in Fig. 4. In this setup, the encoder does not directly have access to the source sequence itself but only to a related observation T^m. The remote source could, for example, represent a sequence of features of the image T^m. The main leitmotif of semantic communication is that the decoder does not wish to reconstruct the entire image but only the sequence of features. In the remote JSCC problem, the decoder indeed attempts to reconstruct the remote (hidden) source sequence S^m and not the sequence observed at the encoder.

Optimality of source–channel separation for the remote JSCC has been established independently by Dobrushin and Tsybakov [59] and by Wolf and Ziv [60]. More recently, extensions of this problem have been considered in [61] and [62], in which both the remote and the observed sources are reconstructed at the receiver. For this setup, Theorem 1 has to be adapted by employing the *remote rate–distortion function*

$$R_{\text{remote}}(D) = \inf_{p_{\hat{S}|T}} I(T; S) \qquad (7)$$

where the infimum is taken over all laws $p_{\hat{S}|T}$ satisfying the distortion constraint $\mathrm{E}[d(S, \hat{S})] \leq D$.

Theorem 2 (Separation Theorem With a Remote Source): Given a remote memoryless source S with distribution p_S, a memoryless observation T related to the source by the transition law $p_{T|S}$, and a memoryless channel $p_{Y|X}$ with capacity C, a rate–distortion pair (r, D) is achievable if $rR_{\text{remote}}(D) < C$. Conversely, if a rate–distortion pair (r, D) is achievable, then $rR_{\text{remote}}(D) \leq C$. Thus, for given distortion target D, the source–channel capacity is given by $C/R_{\text{remote}}(D)$.

The optimal SSCC scheme can be described as follows. The encoder attempts to reconstruct the hidden features S^m from the observed sequence T^m with desired accuracy and then uses a capacity-achieving channel code to send this information reliably (i.e., with a vanishing probability of error) to the decoder, which first decodes the transmitted information and then reproduces the encoder's reconstruction $\hat{S}^m$.

C. JSCC Over Feedback Channels

Optimality of source–channel separation (Theorem 1) extends also to setups with feedback, i.e., to setups where the transmitter observes the past channel outputs $Y_1, \ldots, Y_i$ (or noisy or imperfect versions thereof) before producing the time-i channel input X_i. Thus, also with feedback, a rate–distortion pair (r, D) is achievable if, and only if, $rR(D) < C$. Notice here that, for memoryless point-to-point channels, feedback does not increase the capacity; therefore, the best achievable source–channel coding rate remains the same with feedback.

JSCC, however, allows to reduce the delay and obtain simplified schemes. We illustrate this on the well-known example of sending a Gaussian source over a Gaussian channel with a bandwidth mismatch factor r, whose inverse r^{-1} is a positive integer. The proposed scheme is based on the idea of Schalkwijk and Kailath [63]. A similar scheme can be designed for arbitrary discrete sources and discrete memoryless channels using the "posterior matching" idea [64].

Let S^m be a sequence of independent and identically distributed (i.i.d.) centered Gaussian samples with variance σ^2. Consider further an additive white Gaussian noise (AWGN) channel with input–output relation

$$Y_i = X_i + Z_i, \quad i = 1, \ldots, n \qquad (8)$$

where $\{Z_i\}$ is an i.i.d. sequence of centered Gaussian noise variables of variance N. We assume perfect channel output feedback, in which case X_i can depend on the past channel outputs Y^{i-1}. Channel inputs are subject to an expected average block-power constraint $(1/n) \sum_{i=1}^{n} \mathbb{E}[X_i^2] \leq P$, for a given $P > 0$. For this Gaussian-quadratic setup distortion D is achievable if, and only if[1]

$$\max\left\{\frac{1}{2} \log\left(\frac{\sigma^2}{D}\right), 0\right\} \leq \frac{1}{2} \log\left(1 + \frac{P}{N}\right) \qquad (9)$$

where notice that the left-hand side of the above inequality is the Gaussian-quadratic rate–distortion function and the right-hand side is the capacity of the power-constrained Gaussian channel.

The following low-delay scheme achieves distortion D whenever (9) is satisfied. The encoder communicates each source S_t over r^{-1} consecutive channel uses. We describe the transmission of S_1 during the first r^{-1} channel uses. The other transmissions are similar.

The encoder first sends

$$X_1 = \sqrt{\frac{P}{\sigma^2}} S_1 \qquad (10)$$

and subsequently observes the output symbol $Y_1 = X_1 + Z_1 = ((P/\sigma^2))^{1/2} S_1 + Z_1$. From this output, it computes the minimum mean-squared error (mmse) estimate of source symbol S_1

$$\tilde{S}_{1,1} = \frac{\sqrt{P\sigma^2}}{P+N} Y_1 = \frac{P}{P+N} S_1 + \frac{\sqrt{P\sigma^2}}{P+N} Z_1 \qquad (11)$$

---

[1]Since the right-hand side of (9) is nonnegative, the left-hand side can be simplified to $(1/2) \log(\sigma^2/D)$.

and the corresponding mmse

$$\epsilon_1 = \tilde{S}_{1,1} - S_1. \tag{12}$$

It, then, scales this error so as to meet the input power constraint and transmits this error over the next channel input

$$X_2 = \sqrt{\frac{P}{\alpha_1}}\epsilon_1 \tag{13}$$

for $\alpha_1 := \mathbb{E}[\epsilon_1^2]$. After reception of the feedback output Y_2, the encoder computes the mmse estimate of ϵ_1

$$\tilde{\epsilon}_1 = \frac{\sqrt{P\alpha_1}}{P+N}Y_2 = \frac{P}{P+N}S_2 + \frac{\sqrt{P\alpha_1}}{P+N}Z_2 \tag{14}$$

and uses it to update its estimate on S_1

$$\tilde{S}_{1,2} = \tilde{S}_{1,1} - \tilde{\epsilon}_1. \tag{15}$$

The new estimation error then becomes

$$\epsilon_2 = \tilde{S}_{1,2} - S_1 = \tilde{\epsilon}_1 - \epsilon_1 \tag{16}$$

and is sent over the channel with the next input X_3. These steps are repeated until r^{-1} channel inputs are sent. Specifically, for any $\ell = 2, 3, \ldots, r^{-1}$, after sending the latest estimation error

$$X_\ell = \sqrt{\frac{P}{\alpha_{\ell-1}}}\epsilon_{\ell-1}, \quad \ell = 2, 3, \ldots, r^{-1} \tag{17}$$

at time ℓ, where

$$\alpha_{\ell-1} := \mathbb{E}\left[\epsilon_{\ell-1}^2\right] \tag{18}$$

the encoder uses the feedback $Y_\ell = X_\ell + Z_\ell$ to calculate an estimate

$$\tilde{\epsilon}_{\ell-1} = \frac{\sqrt{P\alpha_{\ell-1}}}{P+N}Y_{\ell-1} \tag{19}$$
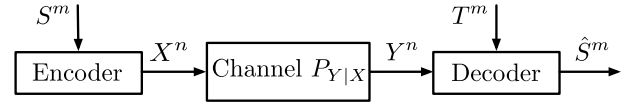
which is then used to update the estimate $\tilde{S}_{1,\ell}$ of S_1

$$\tilde{S}_{1,\ell} = \tilde{S}_{1,\ell-1} - \tilde{\epsilon}_{\ell-1} = S_1 + \epsilon_{\ell-1} - \tilde{\epsilon}_{\ell-1}. \tag{20}$$

Subsequently, a scaled version of the estimation error

$$\epsilon_\ell := \tilde{S}_{1,\ell} - S_1 = \epsilon_{\ell-1} - \tilde{\epsilon}_{\ell-1} \tag{21}$$

is sent in the following transmission $X_{\ell+1}$.

Fig. 5. *JSCC problem in the presence of correlated side information at the receiver.*

After receiving the first r^{-1} channel outputs $Y_1, \ldots, Y_{r^{-1}}$, the receiver sets

$$\hat{S}_1 = \tilde{S}_{1,r^{-1}} \tag{22}$$

where it computes $\tilde{S}_{1,r^{-1}}$ as described for the encoder side.

To derive the distortion achieved by (22), notice that the mmse between $\tilde{S}_{1,\ell}$ and S_1 can recursively be calculated as

$$\mathbb{E}\left[\left(\tilde{S}_{1,\ell} - S_1\right)^2\right] = \alpha_\ell = \frac{N}{P+N}\alpha_{\ell-1} \tag{23}$$

and thus, $\alpha_\ell = \sigma^2((N/P+N))^\ell$. Thus, the proposed coding scheme matches mmse distortion constraint d whenever

$$\left(\frac{N}{P+N}\right)^\ell \leq \frac{D}{\sigma^2} \tag{24}$$

which is equivalent to (9) when specialized to $\ell = r^{-1}$ because $(N/P + N) < 1$ and because of the monotonicity of the log function.

Compared to an SSCC scheme, the above scheme is much simpler and has a reduced delay, that is, each sample can be decoded after $1/r$ channel uses, instead of waiting for the transmission of all the channel symbols to decode all the source symbols at once.

D. JSCC With Side Information: To Bin or Not To Bin?

Next, we consider the scenario in which the receiver has access to correlated side information (see Fig. 5), that is, samples of (S_i, T_i) come from a joint distribution $p(s, t)$. Here, we will focus on lossless transmission of S^m to highlight the core idea of binning; the extension to lossy transmission follows similarly. Different from the scenario in Section II-A, here, the decoder has access to correlated side information T^m, so the decoding function is now given by $g^{(m,n)} : \mathcal{Y}^n \times \mathcal{T}^m \to \mathcal{S}^m$. From a source coding perspective, this is known as the Slepian–Wolf coding problem [65].

This problem was originally studied by Shamai and Verdú [66], where it is shown that a source–channel code rate r is achievable if there exists an input distribution $p(x)$ for which $r < (I(X;Y)/H(S|T))$; and conversely, if r is achievable, then there exists an input distribution $p(x)$ for which $r \leq (I(X;Y)/H(S|T))$. Equivalently, the source–channel capacity of this system is given by

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

$(C/H(S|T))$, where C is the capacity of the underlying channel. It is not difficult to see that the Separation Theorem continues to hold in this setting; that is, any source–channel code rate less than the source–channel capacity, $(C/H(S|T))$, can be achieved by first applying source coding and then transmitting the compressed source bits over the channel using a capacity-achieving channel code.

The source encoder in the case of separation employs Slepian–Wolf coding, which exploits *binning* of the source outcomes. The source encoder randomly distributes all possible source output sequences S^m into $2^{mH(S|T)}$ bins; that is, it independently assigns an index uniformly distributed over $[1 : 2^{mH(S|T)}]$ to each possible source sequence. This constitutes the compression codebook. We also generate a channel codebook of the same size, consisting of length-n channel codewords, each generated in an i.i.d. fashion from $\prod_{t=1}^{n} p(x_t)$, where $p(x)$ is the capacity-achieving input distribution. Then, to transmit the bin index to the decoder, the corresponding channel codeword is transmitted over the channel. Since $r < (C/H(S|T))$, the transmitted channel codeword will be decoded correctly with high probability. Having decoded the channel codeword, the decoder recovers the bin index and outputs the source sequence in the corresponding bin that is jointly typical with its side-information sequence, T^m. It is shown in [65] that the decoder can recover the correct source sequence with high probability since the number of bins is at least $2^{mH(S|T)}$.

Next, we present a coding scheme that generalizes SSCC [67]. In this generalized coding scheme, we randomly distribute all S^m sequences into $M = 2^{mR}$ bins, where R is not necessarily equal to $H(S|T)$. Let $\mathcal{B}(i)$ be the set of sequences allocated to bin i. Then, we generate M i.i.d. length-n channel codewords with distribution $\prod_{t=1}^{n} p(x_t)$ and enumerate these codewords as $x^n(w)$ for $w = 1, \ldots, M$. This constitutes the only codebook in the system. Encoding is done similar to the separation-based scheme. The transmitter finds the index i of the bin s^m belongs to and transmits the corresponding codeword $x^n(i)$.

Note that, in the separation approach described above, the channel decoder decides a single channel codeword index, which conveys the bin index that is then used for source decoding. In the separation approach, for the channel decoding to be successful, it is better to reduce the number of possible indices to be transmitted. In the generalized scheme, we consider a joint source–channel decoder, following the approach in [68]. The decoder looks for an index i for which $x^n(i)$ and Y^n are jointly typical; and, at the same time, there exists exactly one source sequence $\hat{s}^m$ in bin i that is jointly typical with its side information, T^m.

In the joint decoding scheme, we have an error if there exists no or more than one such bin index i, or if there exists more than one jointly typical sequence within bin i. The probability that there is no bin index satisfying

the joint typicality condition vanishes as n grows. The probability of having no jointly typical source sequence within the correct bin also vanishes since S^m and T^m are jointly typical with high probability as m grows. Using the classical arguments on typical sets [69], the probability of having another jointly typical source sequence in the same bin as S^m can be bounded by

$$
\left| \mathcal{B}(i) \bigcap A_\epsilon^m(S) \right| 2^{-m(I(S;T)-3\epsilon)}
$$
$$
\leq 2^{m(H(S)+\epsilon)} 2^{-mR} 2^{-m(I(S;T)-3\epsilon)} \quad (25)
$$

where $A_\epsilon^m(S)$ denotes the set of ϵ-typical m-tuples according to P_S. We can see that (25) goes to zero if $R \geq H(S|T)$.

We also have an error if there exists another bin index j satisfying the joint typicality conditions. The probability of this event can be bounded by

$$
2^{mR} 2^{-n(I(X;Y)-3\epsilon)} \left| \mathcal{B}(i) \bigcap A_\epsilon^m(S) \right| 2^{-m(I(S;T)-3\epsilon)}
$$
$$
\leq 2^{-n(I(X;Y_1)-3\epsilon)} 2^{m(H(S|T)-2\epsilon)} \quad (26)
$$

which goes to zero if $mH(S|T) < nI(X;Y_1)$. Hence, any rate r satisfying $r < (I(X;Y)/H(S|T))$ is achievable.

We have obtained a set of coding schemes each with a different number of bins, that is, with different R values satisfying $R \geq H(S|T)$. The "joint" decoding operation considered in the generalized scheme can equivalently be viewed as a separate source and channel decoding scheme, in which the channel decoder is a list decoder, which outputs the list of bin indices i for which $x^n(i)$ and Y^n are jointly typical. This list decoding approach includes SSCC as a special case with $R = H(S|T)$, in which case, with high probability, there is a single element in the list, i.e., there exists only a single bin index whose channel codeword is typical with the channel output.

Please also note that, in the other extreme, this generalized scheme works without any binning at all. We can generate an independent channel codeword for each possible source outcome, i.e., $R = \log|\mathcal{S}|$. From a practical point of view, this can be interpreted as transferring the complexity of binning from the encoder to the decoder, which now needs to apply joint decoding or list decoding instead of separate source and channel decoding steps. From a theoretical point of view, since the decoder only outputs typical source sequences, there is no point in having more than $2^{m(H(S)+\epsilon)}$ bins as, otherwise, we are creating bins without any typical source sequence in them; and thus, they will never be output. Therefore, we can set $R \leq H(S)$ without loss of generality.

In the case of a point-to-point channel, the only difference between SSCC with binning and joint decoding with no-binning is the operation at the encoder and the decoder. However, binning or no-binning can lead to different protocols, or even different performances in the case of multiuser networks or nonergodic settings as we will see in Section II-E. Even in the point-to-point scenario considered
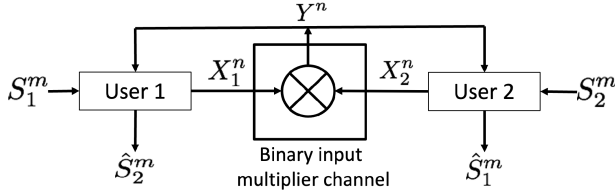
Fig. 6. *JSCC problem over the binary multiplier two-way communication channel.*

here, if the channel and the side information are time-varying, and if the transmitter knows only the distributions of these time-varying processes, it is shown in [70] that the no-binning scheme improves the end-to-end performance compared to separate binning and channel coding, as the former naturally adapts to the source and channel states without targeting specific source and channel realizations.

E. Multiuser JSCC

1) Suboptimality of SSCC With Correlated Sources: Shannon's Separation paradigm (Theorem 1) extends to a certain class of problems where multiple sources are transmitted over a network with multiple transmitters and/or multiple receivers. In particular, source–channel separation is optimal for the following:
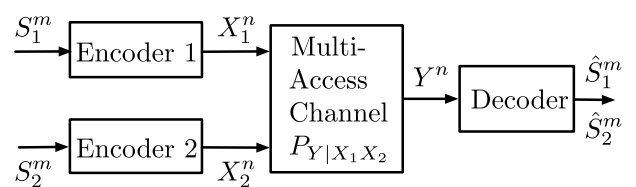
1) all problems where all source sequences are memoryless and independent of each other and each source sequence has to be reconstructed at a single receiver;
2) all problems where the network can be modeled as a set of noninterfering (orthogonal) point-to-point links.

The paradigm, however, does not extend to general multiterminal networks with arbitrarily correlated sources. In general, for these networks, JSCC achieves a better performance. A particular advantage of JSCC in the multiuser context is that it allows to transfer the correlation between the sources observed at different terminals to their channel inputs, which is not possible via SSCC, since the latter relies on transmitting independently generated channel codewords.

In the literature, the proof of the suboptimality of separation is commonly attributed to the work by Cover et al. [71], which studies the transmission of correlated sources over a multiple access channel (MAC). However, Shannon [72] already made the observation that JSCC can outperform separation in the context of two-way communication. Shannon [72] proposed inner and outer bounds on the set of achievable rate pairs over a general two-way communication channel, whose capacity remains an open problem to this day. Shannon [72] particularly considered the multiplier channel, suggested to him by David Blackwell, as an example, which is illustrated in Fig. 6. In this channel, both transmitters have binary inputs, i.e., $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, while the common channel output is also binary and given by $y = x_1\, x_2$. Shannon first

showed numerically that his inner and outer bounds do not match for the multiplier channel. Then, by considering correlated source sequences S_1^m and S_2^m available at the two encoders, where $p_{S_1 S_2}(s_1, s_2)$ satisfies $p(0, 0) = 0$, $p(0, 1) = p(1, 0) = 0.275$, and $p(1, 1) = 0.45$, Shannon showed that it is possible to reach the outer bound on the equal rate point. Note that, with this particular source distribution, no coding is needed, and by simply sending the source realizations over the channel, each user can recover the other user's source symbol reliably, and thus, a source–channel code rate of $r = 1$ source sample per channel use is achievable. Shannon did not explicitly show that SSCC cannot achieve this source–channel rate performance. Transmitting these sources at the same source–channel code rate with SSCC would require the two users to simultaneously transmit at rates $H(S_1|S_2) = H(S_2|S_1) = 0.6942$ bits per channel use. A tighter outer bound on the capacity region of the two-way channel was proposed by Hekstra and Willems [73]. According to the dependence-balance bound derived in [73], the symmetric rate values larger than $R_1 = R_2 = 0.64628$ bits per channel use are not achievable. Therefore, we can conclude that the correlated sources S_1 and S_2 specified above, which can be easily transmitted over the multiplier two-way channel without any coding, cannot be reliably transmitted by any channel code. This result shows that the Separation Theorem does not generalize to arbitrary multiuser channels.

2) Correlated Sources Over a Multiple Access Channel (MAC): As we have mentioned above, a better known example illustrating the suboptimality of separation in multiuser scenarios considers the transmission of correlated sources over a MAC (see Fig. 7). Cover et al. [71] showed the suboptimality of separation through an example very similar to the one considered by Shannon for the two-way channel. They considered a binary-input adder channel, where the channel output is given by $Y = X_1 + X_2$, $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, and $\mathcal{Y} = \{0, 1, 2\}$, and a source distribution $p_{S_1 S_2}(s_1, s_2)$ characterized by $p(0, 0) = 0$, $p(0, 1) = p(1, 0) = p(1, 1) = 1/3$. Similar to Shannon's example, uncoded transmission of the source samples trivially achieves a source–channel rate of $r = 1$. On the other hand, unlike the two-way channel example of Shannon, the capacity region of this MAC is known, and the maximum sum rate that can be achieved by independent channel inputs is 1.5 bits per channel use, while the joint

Fig. 7. *Transmission of correlated sources over a MAC.*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

entropy of the two sources is $H(S_1, S_2) = \log 3 = 1.58$ bits per sample. Hence, the maximum source–channel rate that can be achieved through separation is given by $1.5/1.58 \sim 0.949$ channel uses per sample.

JSCC strategies and achievability results for general discrete memoryless sources and MACs were reported for the lossless case [71] as well as for the lossy case [74], [75]. Information theoretic converses were established in [76], [77], and [78]. In the finite blocklength regime, i.e., for fixed and nonasymptotic n, JSCC schemes were proposed in [79].

For Gaussian sources with squared-error distortions and Gaussian MACs without bandwidth expansion ($r = 1$), Lapidoth and Tinguely [80] presented both JSCC strategies and corresponding achievability results, as well as a (partially matching and generally very tight) converse result. We describe the results for the Gaussian setup in [80] in more detail. Consider the communication setup in Fig. 7, where two users observe correlated jointly Gaussian source sequences S_1^n and S_2^n, which they wish to communicate to a common receiver over n uses of a Gaussian MAC with input–output relation $Y_t = X_{1,t} + X_{2,t} + Z_t$. Here, Z_t are the independent samples from a zero-mean Gaussian distribution with variance N, and the average block-power constraints of P_1 and P_2 are imposed on the channel inputs of the users. We, thus, have $m = n$ and no bandwidth mismatch in this scenario, i.e., $r = 1$. Based on the observed sequence of channel outputs Y^n, the receiver has to produce reconstruction sequences $\hat{S}_1^n$ and $\hat{S}_2^n$ that satisfy the squared-error distortion constraints $\mathbb{E}[\|\hat{S}_k^n - S_k^n\|^2] \leq D_k$, for $k \in \{1, 2\}$.

Consider a symmetric setting where the symbols $(S_{1,t}, S_{2,t})$ are assumed i.i.d. jointly Gaussian of variances σ^2 and correlation coefficient ρ, and we set the power constraints equal, i.e., $P_1 = P_2 = P$. It is shown in [80] that, for large correlation factors ρ satisfying $(\rho/1 - \rho) \geq (P/N)$, uncoded transmission achieves the minimum possible distortion values D_1 and D_2, i.e., the best encodings are to choose $X_{1,t} = ((P/\sigma^2))^{1/2} S_{1,t}$ and $X_{2,t} = ((P/\sigma^2))^{1/2} S_{2,t}$ and optimal decoding is to produce the linear mmse (LMMSE) of $S_{1,t}$ and $S_{2,t}$ based on channel output Y_t, for each $t = 1, \ldots, n$. Separation-based schemes can be shown to achieve significantly larger distortions, and thus worse performance.

For correlation coefficients ρ satisfying $(\rho/1 - \rho) \geq (P/N)$, the hybrid coding scheme in Fig. 8 can achieve smaller distortions than uncoded transmission. The idea of hybrid coding is that each encoder first locally quantizes its observed source sequence and then sends a linear combination of the original source sequence and the quantized sequence over the channel. The receiver first decodes the transmitted quantized sequences and then combines them with its channel outputs to estimate the two source sequences S_1^n and S_2^n. Hybrid coding, thus, includes uncoded transmission and source–channel separation as special cases. For the MAC setup under consideration, hybrid coding is the best scheme proposed so far and in

fact, performs close to the fundamental limit of minimal distortions for all parameter values ρ, σ^2, P, and N.

Hybrid coding or other JSCC schemes have been proposed for a variety of other multiterminal networks (see [75], [79], [81], [82], [83], and [57] for a few examples). A canonical setup is the transmission of correlated sources over a BC, for which both achievability [84] and converse results [85], [86], [87] have been proposed. Special attention was given for the transmission of correlated Gaussian sources over a Gaussian BC, for which hybrid coding was first proved to generally improve over both uncoded transmission and SSCC [88], and shortly thereafter shown to be optimal under a quadratic distortion constraint in all regimes of source correlations and source and channel noise variances and for all admissible distortions [89].

Related to the problem of sending correlated sources over a MAC is the joint CEO-MAC problem in Fig. 9. Communication again takes place over a MAC, but the two encoders cannot directly observe the source that is of interest to the receiver. Instead, each encoder only observes a (correlated) noisy version of a hidden source S^m, which the decoder aims to reconstruct by the sequence $\hat{S}^m$. This setup was first introduced and studied in the special case in which the MAC degenerates to two rate-limited but noiseless links [90], [91], [92], [93]. In this case, the problem turns into a pure source-coding problem. The variation with a Gaussian MAC was studied for example in [94] and [95]. The results in [94] show that in some cases uncoded transmission of the sources is optimal. In general, however, JSCC schemes strictly improve over separation-based architectures that combine an optimal CEO code with an optimal MAC code. The setup in which the source of interest cannot be directly observed at the encoders is of great interest for semantic communication, as it models the desired feature (or class in a classification task) that is not directly observed at any of the terminals but can only be accessed indirectly through the data sample, e.g., an image.

Related to the joint CEO-MAC problem and to the problem of sending correlated sources over a MAC is also the problem of function computation over MACs [96]. In this problem, the setup is again similar to the one in Fig. 7, except that the decoder wishes to reconstruct a symbolwise function of the two source sequences $f^{\otimes m}(S_1^m, S_2^m)$. As shown in [96], lattice-based JSCC schemes outperform separation-based schemes in this scenario.

3) BC With Side Information: Let us first consider the generalization of the problem with side information, studied in Section II-D, to the BC scenario, as illustrated in Fig. 10. Here, the encoder transmits the outcome of a single source sequence S^m to multiple receivers with side information $T_1^n, \ldots, T_K^m$, respectively, over a discrete memoryless BC $P_{Y_1, \ldots, Y_K|X}$. This scenario is named as Slepian–Wolf coding over BCs by Tuncel [68].

The generalized coding scheme introduced in Section II-D can be directly adopted in this scenario [68].
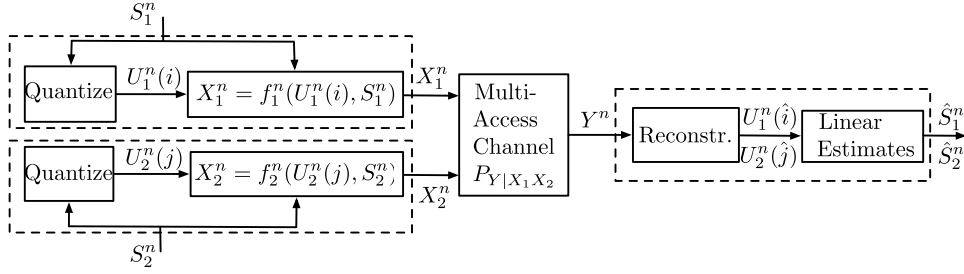
Fig. 8. *Hybrid coding for transmitting a bivariate source over a MAC.*

The encoder maps each source outcome to a different channel codeword, each of these codewords is generated in an i.i.d. fashion from a channel input distribution p_X. After observing the channel outputs Y_k^n and the source side-information sequence T_k^m, decoder k employs either the joint decoding or the list decoding scheme presented in Section II-D.

It can be shown that the error probability, i.e., the probability that $\hat{S}_k^m \neq S^m$ tends to 0 as n tends to infinity and $m = r \cdot n$, whenever
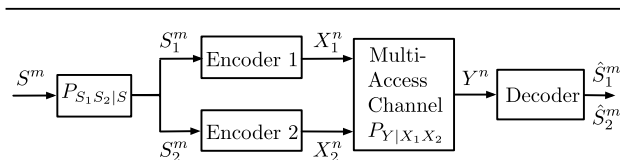
$$r \cdot H\left(S|T_k\right) < I\left(X; Y_k\right), \quad k = 1, \ldots, K. \tag{27}$$

Following the standard converse steps in the single-user Slepian–Wolf coding problem, it can be shown that (27) (with nonstrict inequality) has to hold for some probability mass function (pmf) P_X. This leads to the following theorem, which is a slightly modified expression of [68, Th. 6].

Theorem 3: The source–channel coding capacity for the Slepian–Wolf coding problem over the BC is given by

$$\sup_{p(x)} \min_{k=1, \ldots, K} \frac{I\left(X; Y_k\right)}{H\left(S|S_k\right)}. \tag{28}$$

Above theorem shows that each user can compensate bad side information with good channel conditions and vice versa. For the reconstruction capability of a decoder, only the global quality of the side information and the channel matter, but not the individual qualities. In this problem, if we were to apply an SSCC approach, we would need to deliver $H(S|S_k)$ bits per source sample to decoder k. Hence, we need an achievable tuple $(R_1, \ldots, R_K)$ within the capacity 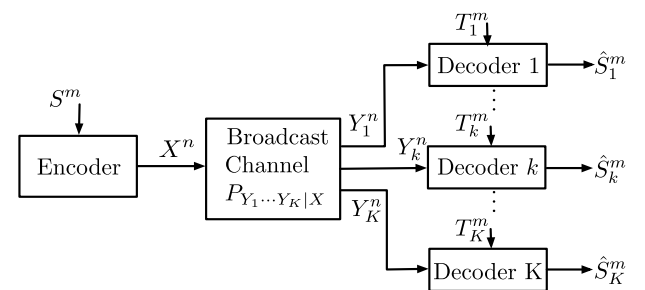region of the underlying BC, which maximizes $R_k/H(S|S_k)$ over k. Tuncel [68] showed through an example that this rate can be strictly below the source–channel capacity.

In the described code construction, the encoder behaves like in SSCC. It is, however, important that the decoder jointly reconstructs the channel and source codewords. The proposed scheme is, thus, still a JSCC scheme, even if encoding is performed separately. An alternative coding scheme which applies SSCC that can still achieve the optimal source–channel capacity is also possible. In this alternative coding scheme, source samples are compressed separately for each user such that each user receives enough information to decode the source sequence when combined with its own side-information sequence. Then, the source sequence is divided into blocks, and the bin indices for these blocks are transmitted to each receiver at a different channel block, and the users employ backward decoding. We will illustrate this coding scheme considering a network with two users.

We fix $p(x)$ such that (27) holds. Then, we consider a sequence of (m, n) pairs such that m/n is less than (28), and its limit is r, i.e., $\lim_{m,n\to\infty} m/n = r$. For each (m, n) pair, a total of Bm source samples will be transmitted over $(B + 1)n$ channel uses. This corresponds to a source–channel code rate of $Bm/(B + 1)n$, which becomes arbitrarily close to m/n as $B \to \infty$.

Source Code Generation: Corresponding to each decoder i, $i = 1, 2$, we consider $M_i = 2^{mR_i}$ bins, called the T_i bins. All possible source outcomes $s^m \in \mathcal{S}^m$ are partitioned randomly and uniformly into these bins, independently of

Fig. 9. *Joint CEO-MAC problem.*

Fig. 10. *Slepian–Wolf coding over a BC.*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

Block 1	Block 2	$\cdots$	Block B	Block B+1
$x^n(w_{1,1},1)$	$x^n(w_{2,1},w_{1,2})$	$\cdots$	$x^n(w_{B,1},w_{B-1,2})$	$x^n(1,w_{B,2})$

Fig. 11. *Channel codeword assignment for separation-based encoding and backward decoding scheme for Slepian–Wolf coding over BC with K = 2 users. B source blocks are transmitted in (B + 1) channel blocks.*

each other, i.e., the distribution into M_1 bins is independent of the distribution into M_2 bins. The ith bin index of the source sequence represents the information that will be provided to user i. This bin assignment, which corresponds to source compression, is made available to all the users.

Channel Code Generation: For the channel codebook, generate at random $M_1 M_2$ channel codewords i.i.d. with probability $p(x^n(i,j)) = \Pi_{t=1}^n p(x_t)$, and index them as $x^n(ji,j)$ with $(i,j) \in [1:M_1] \times [1:M_2]$.

Encoding: Consider a source sequence s^{Bm} of length Bm. Partition this sequence into B portions, s_b^m, $b = 1, \ldots, B$. Similarly, partition the side-information sequences into B length-m blocks $t_i^{Bm} = [t_{i,1}^m, \ldots, t_{i,B}^m]$ for $i = 1, 2$. The bin index of the jth block of the source output sequence s_j^m with respect to T_i bins is denoted by $w_{j,i}$. The estimate of $w_{j,i}$ at user k, $k = 1, 2$, $k \neq j$, is denoted by $\hat{w}_{j,i}^k$. See Fig. 11 for an illustration of the encoding scheme.

In block 1, transmitter observes s_1^m and finds the corresponding T_1 bin index $w_{1,1} \in [1 : M_1]$. It transmits the channel codeword $x^n(w_{1,1}, 1)$. In block 2, it transmits the channel codeword $x^n(w_{2,1}, w_{1,2})$ corresponding to the T_1 bin index of the second source block and the T_2 bin index of the first source block. In the following blocks $b = 2, \ldots, B$, it transmits the channel codewords $x^n(w_{b,1}, w_{b-1,2})$ where $w_{b,i} \in [1 : M_i]$ for $i = 1, 2$. In block $B + 1$, it transmits $x^n(1, w_{B,2})$.

The first user estimates the source block s_{b-1}^m at the end of block $b - 1$, denoted by $\hat{s}_{1,b-1}^m$, and finds the corresponding bin index $\hat{w}_{b-1,2}^1 \in [1 : M_2]$. When the condition in (27) holds, it is possible to show that the user will recover each source block successfully with vanishing error probability as $m, n \to \infty$.

On the other hand, user 2 will wait until the end of the last block to start decoding and will decode the source blocks backward. At the end of channel block $B + 1$, user 2 decodes the first source block s_1^m, denoted by $\hat{s}_{2,B}^m$, and finds the corresponding bin index $\hat{w}_{B,1}^2 \in [1 : M_1]$. Having estimated this bin index, it can then decode the next source block $B - 1$, and so on so forth. Similarly, each source block will be decoded by the second user with high probability if (27) holds.

The idea of Slepian–Wolf coding over BCs is extended to lossy broadcasting scenario in [97]. This coding principle is also very useful in cache-aided networks, where it was termed joint cache-channel coding [98], [99], [100]. In cache-aided networks, users store contents at cache memories close to end users, which then serves as side information during the actual communication phase. The situation is, thus, similar to the one described above, but

we face a data communication problem where different decoders wish to learn different parts of the data.

F. Types of Source–Channel Separation

The above results and coding schemes beg further discussion on source and channel separation in multiuser networks. In [101], two types of separations are proposed: the first type of separation is a direct generalization of Shannon's Separation Theorem to multiuser networks, and answers the question "Can we determine the source–channel capacity by simply comparing the source coding rate region with the capacity region of the underlying sources and channels?" This is trivially the case in point-to-point channels. This also holds for a MAC when there is no multiuser interference, that is, the users have orthogonal channels to the receiver [102]. In such a case, the source–channel capacity for a given target distortion tuple $(D_1, \ldots, D_K)$ and input cost constraint tuple $(P_1, \ldots, P_K)$ is given by $\sup_{(R_1^s, \ldots, R_K^s) \in \mathcal{R}_{comp}, (R_1^c, \ldots, R_K^c) \in \mathcal{R}_{chan}} \min_{k=1,\ldots,K} R_k^c / R_k^s$, where $\mathcal{R}_{comp}$ denotes the source coding rate region for the considered sources for target distortion tuple $(D_1, \ldots, D_K)$ and $\mathcal{R}_{chan}$ denotes the capacity region of the underlying channel with cost constraints $(P_1, \ldots, P_K)$.

Note that the two regions are characterized completely independently of each other; that is, the optimal source coding rate region is ignorant of the channels, and the channel capacity region is ignorant of the source statistics. Moreover, the coding schemes that achieve the operating points in these two rate regions are also oblivious to each other, apart from exchanging bits. On the other hand, the coding scheme achieving the optimal source–channel capacity in Theorem 3 points to a different type of separation. The source–channel capacity can still be determined by comparing two rate regions that are statistically independent of each other; however, the coding schemes to achieve these rate regions are not oblivious to each other, and moreover, the channel coding rate region does not correspond to the capacity region of the underlying channel. In [68], these two types of separations are called "informational separation" and "operational separation," respectively.

G. JSCC for Channel Coding

So far, we have focused on JSCC for transmitting source signals to be recovered at the decoder under a prescribed distortion measure. However, JSCC in general can also be instrumental in increasing the achievable coding rates in multiterminal networks. Treating received signals as sources to be forwarded to other terminals to help their decoding of the underlying message was first considered in [103] in the context of a relay channel (illustrated in Fig. 12). Here, the goal of the encoder is to convey message $W \in \{1, \ldots, 2^{nR}\}$ to the decoder. The relay terminal can overhear the transmission and would like to help the decoder by conveying extra information about the
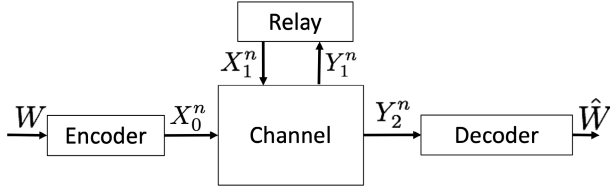
Fig. 12. *Relay channel.*

transmitted message. One option is for the relay to first decode W and forward it to the decoder in collaboration with the encoder. However, if the relay's channel is not strictly better than that of the decoder, requiring correct decoding at the relay can become a bottleneck instead. An alternative scheme is to treat the signal received at the relay, Y_1^n, as a source signal, which is to be forwarded to the decoder with as high fidelity as possible. Cover and El Gamal [103] treated this problem as a source and channel coding problem. Since it is a point-to-point channel, separation is known to be optimal in this case, and thus, the relay can first compress Y_1^n and forward it to the decoder using channel coding. Note, however, that, the decoder's received signal, Y_2^n, is correlated with that of the relay and, thus, can act as correlated side information, as in Wyner–Ziv coding. Let $\hat{Y}_1^n$ denote the compressed version of relay's received signal. To forward it to the decoder, the following condition should hold: $I(Y_1; \hat{Y}_1|X_1, Y_2) \leq I(X_1; Y_2)$. Then, the decoder can combine the compressed version of relay's signal with its own received signal to decode the message transmitted by the encoder. The achievable rate with this compressed-and-forward scheme is then given by $I(X_0; \hat{Y}_1, Y_2|X_1)$, where the joint distribution of the random variables is given by $p(x_0)p(x_1)p(\hat{y}_1|x_1, y_1)p(y_1, y_2|x_0, x_1)$.

In the case of multiuser relay channels, the relay will need to broadcast its received signal to multiple receivers for an extension of the compress-and-forward scheme. Note, however, that there is no Separation Theorem in this case since the goal is to send a lossy version of the relay's signal to multiple receivers each with a different correlated side information (similar to [97]). The simplest scenario requiring the broadcast of a relay signal to multiple users would be the separated two-way relay channel model. In this model, two users try to communicate with each other with the help of a relay terminal. It is called separated since there are no direct links between the users, and thus, they rely on the relay terminal for communications. The relay receives the superposition of the signals transmitted by each user. In general, it may not be feasible for the relay to decode the users' messages, and instead, it can broadcast its received signal to both users. Since each user already knows its own transmitted signal, it may be easier for them to cancel their own interference and decode the other user's message. In this case, user's own signals act as correlated side information when the relay's received signal is broadcast to the users. Two achievable

rate regions are proposed in [104] relying on JSCC of relay's signal using Wyner–Ziv coding over the BC from the relay to the users. A single compressed version of the relay's signal is transmitted to both users in the first scheme, while a refinement layer is sent to one of the users in the second scheme. Further gains are possible by introducing a hybrid relaying scheme, which exploits amplify-and-forward relaying together with the proposed digital broadcasting approaches, similar to [105]. While these schemes generally lead to achievable rates or rate regions and are not capacity-achieving, they are shown to achieve the optimal diversity-multiplexing tradeoff for MIMO relay and two-way relay channels in [106] and [107], respectively.

Hybrid coding (which we used to send correlated sources over a MAC) can also improve the performance of data communication problems. Consider, for example, the diamond relay network in Fig. 13, where the goal is to send a message from the encoder to the decoder with the help of two relays. The best known coding scheme (in terms of achievable rates) uses hybrid coding at the relays, i.e., the relays quantize their channel outputs Y_1^n or Y_2^n into source codewords U_1^n or U_2^n, and then sends a symbolwise function of these source codewords and the observed channel outputs. Similar to the MAC with correlated sources, hybrid coding allows to transfer part of the correlation between the outputs Y_1^n and Y_2^n to the inputs X_1^n and X_2^n, which for certain MACs allows for a better communication performance than using independent inputs.

JSCC plays a prominent role also in interactive data communication scenarios when terminals wish to communicate messages and not source sequences. In fact, interactive data communication schemes often (see [72], [108], [109], [110], [111], and [112]) build on strategies where information is transmitted in a given block and refined in subsequent blocks. From a receiver's point of view, during the refinement blocks, the data bits are no longer independent bit sequences because they are correlated with the symbols observed during the previous blocks. In this sense, the communication problem turns into a JSCC problem with side information, and JSCC generally outperforms separation-based schemes.

The refinement idea is, for example, exploited in [110] on a two-user BC with rate-limited feedback, i.e., on a single-transmitter two-receiver channel where both receivers can send back feedback information over rate-limited communication links. A simple but very efficient scheme [110] for this setup is that, after each communication block, the receiver with the worse channel
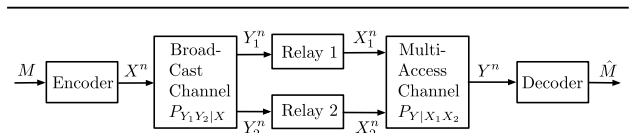
Fig. 13. *Data communication over the diamond relay channel.*
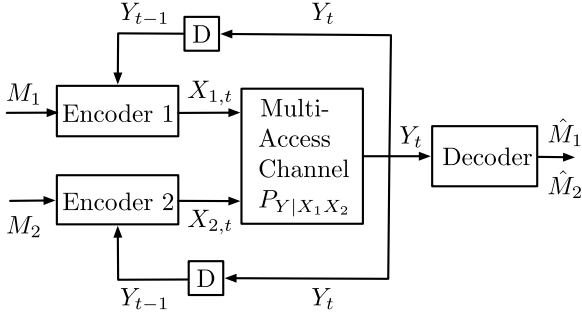
Fig. 14. *Data communication over the two-user MAC with feedback.*

condition quantizes its observed output signal and sends back the quantization information over the rate-limited link. In the next block, this information is sent from the single transmitter to the receiver with the better channel as part of the *cloud center*, i.e., as part of the information that has to be decoded by both receivers. The weaker receiver will not be bothered by this additional transmission in the cloud center since it already knows the quantization information (it generated itself) and can simply ignore its presence. The stronger receiver has sufficiently good channel conditions to be able to decode the quantization information and reconstruct a quantized version of the signal observed at the weaker receiver in the previous block. This second output signal allows the stronger receiver to decode the previous block with a single-input multioutput (SIMO) decoding based on two outputs. On an abstract level, the described scheme establishes a communication path (over the feedback and forward links) from the weaker receiver to the stronger receiver, which allows to exchange information about the output signals essentially without using resources on the forward link.

The described scheme is a JSCC scheme because it does not treat the quantization information as an isolated sequence of bits, but, instead, takes into account that it is information obtained from the previous block's output at the weak receiver, and thus known to it. Without taking into account this knowledge, the proposed scheme does not yield any benefit compared to a simple nonfeedback scheme. However, after accounting for this side information, the scheme improves over the optimal no-feedback scheme on a large class of channels as long as the the feedback rate is nonzero [110]. Incidentally, JSCC schemes that improve capacity with feedback for Gaussian BCs with noisy feedback are proposed in [113], which leads to a complete characterization when capacity improvements are possible and when not.

Another canonical instance of interactive communication is the MAC with feedback problem in Fig. 14, where two encoders communicate to a single decoder sequentially, in the sense that the time-t channel inputs $X_{1,t}$ and $X_{2,t}$ are functions of the respective messages M_1 or M_2, as well as the past channel outputs Y^{t-1}. For this setup, a block-Markov coding scheme is proposed in [108],

where in each block, JSCC is used to send refinement information about the previously transmitted codewords. This refinement information is calculated based on the observed feedback outputs and can be highly correlated across the two transmitters. Sending such correlated side information in the next block allows to correlate the channel inputs from the two transmit terminals (without resorting to trivial inputs) and to achieve higher communication rates.

An elegant coding scheme for the two-user MAC with feedback in the Gaussian case was proposed in [114]. Each of the two transmitters first maps its message M_k, $k \in \{1, 2\}$ to one of the 2^{nR_k} equally-spaced points on a grid over the interval $[-1/2, 1/2)$, i.e.,

$$\Theta_k = \frac{M_k}{2^{nR_k}} - \frac{1}{2} \qquad (29)$$

and then sends this message point during the first channel use by transmitting $X_{k,1} = (12P_k)^{1/2}\Theta_k$, where the scaling is performed to satisfy the input power constraint P_k. In subsequent channel uses, Tx $k \in \{1, 2\}$ calculates the LMMSE $\hat{\Theta}_{k,t}$ of Θ_k based on past feedback outputs $Y_1, \ldots, Y_{t-1}$ and sends a scaled version of the LMMSE error $\epsilon_{k,t}$, as in the single-user scheme explained in Section II-C. Following again the idea in Section II-C, the two transmitters then send a scaled version of their estimation errors so as to satisfy the power constraint, while one of the two transmitters also modulates its inputs with a ± 1 sequence to ensure that at each time instance the inputs of the two transmitters are positively correlated and thus profit from each other. The receiver calculates the LMMSE estimates $\hat{\Theta}_{1,n}$ and $\hat{\Theta}_{2,n}$ of both message points and declares the messages M_1 and M_2 associated with the grid points that are closest to these estimates.

The described scheme is not only particularly simple to implement but achieves the sum capacity of the two-user Gaussian MAC with output feedback. It has been extended to MACs with more than two users [115], for which it was shown to achieve capacity [115], [116], to interference and BCs [83], [117], [118], [119], and to MACs and BCs with noisy feedback [113], [120], [121].

H. Beyond Average Distortion

The classical JSCC problem considers an average distortion measure [see (2)], where each summand depends only on the pair of source and reconstruction symbols at a given time i, which is key to the proof of Shannon's Separation Theorem. Depending on the application, however, more general performance criteria have to be applied.

Consider, for example, the distributed hypothesis testing problem in Fig. 15 without bandwidth mismatch, i.e., $m = n$, where a sensor (transmitter) observes source sequence S^n and communicates to a decision center over a noisy communication channel. Based on the outputs of the channel and its own observation T^n, the receiver has to guess the joint distribution that underlies the distributed
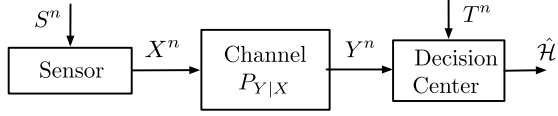
Fig. 15. *Distributed hypothesis testing scenario.*

observations S^n and T^n. A binary hypothesis is assumed where under the null hypothesis

$$\mathcal{H} = 0: \quad (S^n, T^n) \quad \text{i.i.d.} \sim P_{ST} \tag{30}$$

and under the alternative hypothesis

$$\mathcal{H} = 1: \quad (S^n, T^n) \quad \text{i.i.d.} \sim Q_{ST} \tag{31}$$

for two given pmfs P_{ST} and Q_{ST}.

Mathematically speaking, the sensor maps the observed source sequence s^n to channel inputs $x^n(s^n)$, and the decision center maps the pair of channel outputs y^n and observed side information t^n to a decision $\hat{\mathcal{H}}(y^n, t^n)$.

An exponent θ is said achievable, if for sufficiently large blocklengths n, there exist encoding and decoding functions such that the decision center's probabilities of error

$$\alpha_n := \Pr\left[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\right] \tag{32}$$

$$\beta_n := \Pr\left[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\right] \tag{33}$$

satisfy

$$\lim_{n \to \infty} \alpha_n = 0 \tag{34}$$

and

$$-\overline{\lim_{n \to \infty}} \frac{1}{n} \log_2 \beta_n \geq \theta. \tag{35}$$

The goal is to maximize the achievable type-II error exponent θ.

For certain families of source distributions P_{ST} and Q_{ST}, for example, for testing against independence, where $Q_{ST} = P_S P_T$, a separation-based scheme [122] achieves the optimal error exponent θ. In particular, for testing against independence, the optimal exponent is $\theta^\star = \max I(U; T)$, where the maximization is over all auxiliary random variables U such that the tuple $(U, S, T) \sim P_{U|S} P_{ST}$ satisfies $I(U; T) \leq C$, with C denoting the capacity of the communication channel as before.

In general, however, an improved performance can be attained by a JSCC scheme. The best JSCC schemes known to date are the schemes in [123] and [124], which combine either hybrid coding or joint decoding à la Tuncel [68] with

unequal error protection (UEP) [125]. A slightly simplified version of the coding scheme in [123] based on hybrid coding is sketched in Fig. 16 and described in the following. The full scheme in [123] is obtained by additionally introducing coded time-sharing and dependence between various code components.

We pick a conditional pmf $P_{U|S}$ and a function f mapping (U, S) to the channel input X. The joint pmf is given by $P_{USTXY} = P_{U|S} P_{ST} P_{X|US} P_{Y|X}$, where $P_{Y|X}$ indicates the channel law and $P_{X|US}$ indicates the input distribution obtained when applying the chosen function f to (U, S). We assume that $I(U; S) > R > I(U; S) - I(U; T) = I(U; S|T)$ with respect to the joint pmf P_{UST}. We then generate a random codebook $\{u^n(\ell): \ell = 1, \ldots, 2^{nR'}\}$, by picking all entries i.i.d. according to the marginal P_U.

Sensor: Given that it observes the source sequence $S^n = s^n$, the remote sensor looks for an index ℓ so that the codeword $u^n(\ell)$ is jointly typical with the observed source sequence s^n according to the joint pmf P_{US}. It then applies the function f componentwise to the chosen codeword $u^n(\ell)$ and the source sequence s^n, i.e., $x^n = f^n(u^n(\ell), s^n)$, and sends the result over the channel. If no typical codeword can be found, then the sensor sends a special input sequence x_0^n over the channel. In other words, if the typicality test succeeds, the hybrid encoding steps from [75] are used, and otherwise, the UEP mechanism from [125] is employed.

The idea is to choose the sequence x_0^n to be well-distinguishable at the receiver from any of the codewords $u^n(\ell)$, thus implementing a UEP code. This is crucial for the present hypothesis testing scenario because it allows to specially protect the message indicating that none of the codewords has empirical statistics close to P_{US}, which is the statistics that we can expect under $\mathcal{H} = 0$ for at least one of the codewords if we choose the codebook of the indicated size. The absence of such a codeword, thus, hints to the hypothesis $\mathcal{H} = 1$ and the sensor applies a UEP code to enforce high reliability on the transmission of this information, because in our asymmetric setup, we have more stringent requirement on missing hypothesis $\mathcal{H} = 1$ (we require exponentially decreasing β_n) than on missing hypothesis $\mathcal{H} = 0$ (α_n can decay arbitrarily slowly.

Decision Center: Assume that $T^n = t^n$ and $Y^n = y^n$. The receiver first looks for an index $\ell' \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ such that the following two conditions are simultaneously satisfied:

$$\left(u^n\left(\ell'\right), y^n, t^n\right) \in \mathcal{T}_\mu^n\left(P_{UYT}\right) \tag{36}$$

and

$$H_{\text{tp}(u^n(\ell'), t^n)}\left(U | T\right) = \min_{\tilde{\ell}} H_{\text{tp}\left(u^n(\tilde{\ell}), y^n\right)}\left(U | T\right) \tag{37}$$

where here $\mathcal{T}_\mu^n(P_{UYT})$ denotes the set of triples (u^n, y^n, t^n) with empirical statistics equal to P_{UYT} within a small
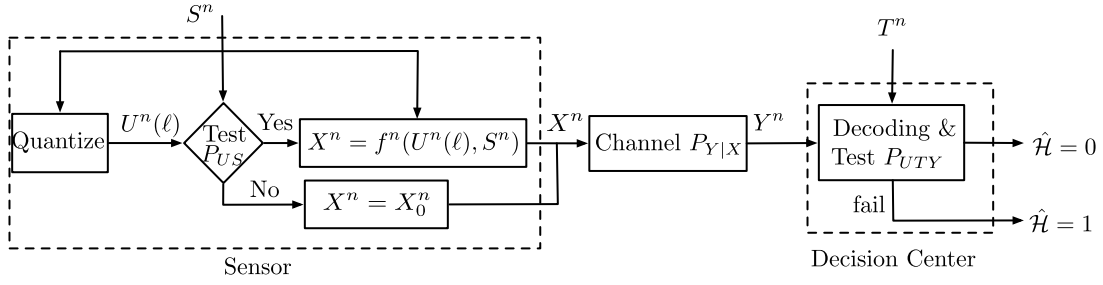
Fig. 16. *Distributed detection scheme based on JSCC (hybrid coding) and UEP.*

positive $\mu > 0$ margin, and $H_{\text{tp}(u^n, y^n)}(U|T)$ denotes the conditional entropy of random variables U given T whose joint pmf is given by the empirical statistics (the type) of the pair (u^n, y^n). If the tests in (36) and (37) are simultaneously successful for the same index ℓ', then the decision center declares $\hat{\mathcal{H}} = 0$, and otherwise, $\hat{\mathcal{H}} = 1$.

The empirical conditional entropy test in (37) is inspired by the distributed hypothesis testing scheme over a noiseless communication channel in [126]. Recently, an improved test for the noiseless setup has been proposed [127], and it can be expected that using their scheme in this setup with a noisy communication channel will also lead to an improved performance.

The study of distributed hypothesis testing over noiseless links has received significant attention in the literature, with several remarkable results [124], [128], [129], [130], [131], [132], [133], [134], [135]. More complicated setups over noisy networks have been considered for example in [136], [137], and [138].

III. PRACTICAL CODE DESIGN FOR JSCC: CLASSICAL APPROACHES

The design of practical JSCC schemes has been a longstanding challenge in information and coding theory. The efforts can be grouped into three general categories. In the first group are joint code designs for idealistic i.i.d. or Markov source distributions. These approaches aim at extending some of the known code designs to exploit the redundancy in the source distribution. In the second group are joint designs for practical sources, such as images or videos, which typically consider two separate codes for compression and channel coding, whose parameters are optimized jointly. The last group is closer to a fully joint design, inspired by the optimality of uncoded/linear transmission for Gaussian sources over Gaussian channels, and employs linear encoding schemes with some hand-designed nonlinearities for the transmission of practical image or video sources. Next, we present overviews of codes under these three categories.

A. Practical JSCC Code Design

There have been many practical coding designs for JSCC based on the classical coding schemes, such as Turbo, low-density parity-check (LDPC), and polar codes. In his

1948 article, Shannon [1] already wrote that "...any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise." He also provided a simple example in [45] showing that the uncoded transmission of a nonuniform binary source over a binary erasure channel can meet the optimal Shannon lower bound achieved by SSCC with infinite blocklength. Various designs have been proposed for general source and channel distributions to benefit from the source redundancy. Optimality of linear codes for lossless coding was observed in [139]. It was also shown in [140] that linear codes cannot achieve optimal performance when lossy transmission is considered. Most follow-up studies on explicit code construction focused on lossless transmission. Hagenauer [141] proposed a modified version of the Viterbi algorithm that can benefit from the a priori or a posteriori information about the source bits. A joint trellis representation of the decoder is proposed in [142] for i.i.d. and Markov sources that can allow Viterbi-like joint decoding of variable-length source codes with convolutional and trellis channel codes. Another trellis representation is suggested in [143] for variable-length Huffman coding combined with a convolutional code, and the Turbo principle is applied at the decoder to iterate between the source and channel decoders. The idea to exploit the residual redundancy of the compression scheme is also considered in [144] and [145]. The former combines differential pulse code modulation (DPCM) with a nonbinary convolutional channel code and employs a Viterbi decoder. In the latter, Lempel–Ziv coding is shown to correct a few errors over the channel by exploiting the redundant bits that remain after compression. Residual redundancy in Huffman coding is exploited in [146] when transmitting a Markov source over a noisy channel, with both a Viterbi and a Turbo channel decoder. A first-order binary Markov source is considered in [147] with a Turbo channel code, and both constituent decoders are modified to optimize the performance. JSCC of more general hidden Markov source models is studied in [148] and [149] with Turbo decoders, where a universal Turbo decoder is used in the latter, which means that the source statistics need not be known in advance.

JSCC of a first-order Markov source using an LDPC code is considered in [150]. A convergence analysis and a general method for the optimization of irregular LDPC codes for joint source–channel receivers are presented in [151]. A lossless JSCC scheme based on two concatenated LDPC codes is proposed in [152]. An LDPC-based source encoder is concatenated with an LDPC channel encoder at the transmitter side, while a single joint belief propagation (BP) decoder is employed at the receiver. The asymptotic performance of the proposed scheme is theoretically analyzed using extrinsic information transfer (EXIT) charts. To reduce the implementation complexity of the JSCC system and to improve its performance in the waterfall region, double protograph LDPC (DP-LDPC) codes are proposed in [153]. In general, the optimal channel code designed for the underlying channel is not optimal in the JSCC context due to the joint decoding operation. A joint protograph extrinsic information transfer (JPEXIT) analysis is proposed in [154] to calculate the decoding threshold of the joint protograph matrix, and the channel code in the DP-LDPC system is redesigned to improve the overall performance in the waterfall region. Further variants of DP-LDPC codes have been introduced to improve the performance [155], [156], [157]. A graph-theoretic construction method is proposed in [158] to transform the parity-check matrices of the source and channel codes into an interset constraint problem. Spatially coupled LDPC (SC-LDPC) codes are proposed in [159] to implement JSCC by concatenating two SC-LDPC coding modules for source and channel coding, respectively, at the transmitter with a joint BP-based decoder. A joint source and channel anytime coding scheme with partial joint extending window decoding is proposed in [160] based on spatially coupled repeat-accumulate (SC-RA) codes. It is shown that the proposed approach provides low latency and high-reliability transmission with respect to [159] at the cost of higher decoding complexity.

A double polar code (D-Polar)-based scheme is proposed to implement JSCC in [161], where the source compression is implemented by a polar encoder before the channel error correction is performed by a systematic polar code. At the receiver side, a turbo-like BP channel and source decoders are introduced to perform channel and source decoding, respectively. A joint BP decoding approach is instead applied in [162], together with a biased extrinsic information transfer (B-EXIT) convergence analysis.

Rateless codes, such as Raptor codes, have also been applied to implement JSCC [163]. In particular, the rateless property of Raptor encoders can be utilized during the explicit entropy coding stage to reduce the channel postdecoding residual errors.

For lossy JSCC of memoryless sources, Kurtenbach and Wintz [164] studied the problem of optimum quantizer design when the quantizer's output is transmitted over a noisy channel. This is further studied in [165] considering jointly the code assignment problem. Farvardin [166] and Farvardin and Vaishampayan [167] later extended their study to vector quantizers. Farvardin [166] employed simulated annealing for the codeword assignment problem and showed that when some statistical information about the channel is available, it is possible to improve the performance of the vector quantizer through a joint design of the codebook and the binary codeword assignment.

Dunham and Gray [168] proved that for stationary and ergodic sources, there exist optimal joint source and channel trellis encoding systems over discrete memoryless channels, which consist of concatenating a source tree coding system with a channel tree coding system, and perform arbitrarily close to the distortion rate function evaluated at the capacity of the underlying channel. While this is an existence result, Ayanoglu and Gray [169] applied the generalized Lloyd algorithm to design joint source and channel trellis codes. The experimental results on independent and autoregressive Gaussian sources over binary symmetric channels under absolute and squared error distortion measures indicate that the joint system can outperform a separately optimized tandem system. A JSCC scheme using combined trellis-coded quantization (TCQ) and trellis-coded modulation (TCM) is proposed in [170] for communication over an AWGN channel. TCQ of memoryless sources is developed in [171] for transmission over a binary symmetric channel. It is shown to achieve the same performance as the one proposed in [169], but with a much lower complexity. A joint design of TCQ/TCM is also proposed and is shown to outperform the cascade of separately designed TCQ and TCM systems, particularly at low signal-to-noise ratios (SNRs). Further improvements in moderate to high SNR regimes are obtained in [172], [173], and [174] by introducing more advanced schemes at the expense of increased computational complexity.

B. JSCC Code Designs for Realistic Sources

In this second class, we can consider JSCC designs that combine compression algorithms for particular source types, e.g., audio, image, or video, with channel codes, where the parameters of the two are jointly optimized to maximize the end-to-end performance. Modestino and Daut [175] combined 2-D DPCM for image coding with a convolutional code. This was later extended to a 2-D discrete cosine transform (DCT) in [175]. Instead, Comstock and Gibson [176] combined 2D-DCT with Hamming codes. Around the same time, there was also significant interest in JSCC for speech signals. Goodman and Sundberg [177] studied the combined effects of quantization and transmission errors on the performance of embedded DPCM. Later, they applied these ideas to embedded DPCM speech encoding and punctured convolutional codes in [178]. Moore and Gibson [179] combined DPCM speech encoding with self-orthogonal convolutional codes.

In [180], a JSCC approach is proposed for Turbo codes to further improve the error control performance of image and video transmission. In this approach, three feedback schemes are introduced to improve the Turbo decoding performance, including error-free source information

feedback, error-detected source information feedback, and channel soft values (CSVs) feedback for source signal postprocessing. These feedback schemes are implemented by modifying the extrinsic information passed between the constituent maximum a posteriori probability (MAP) decoders in the Turbo decoder. The experimental results show that around 0.6 dB of channel, SNR reduction can be achieved by the proposed JSCC schemes without introducing any extra computational cost.

Wu et al. [181] consider the transmission of encoded images over noisy channels. To provide unequal protection to different bit streams of the image to combat channel noise, the authors consider two rate compatible channel codecs, rate compatible punctured convolutional codes (RCPC) and rate compatible punctured turbo codes (RCPT), to construct a JSCC system. In order to derive the optimal rate distribution, they formulate the optimization problem of minimizing the end-to-end distortion, which is then solved by a dynamic programming-based approach. The experimental results show that the proposed JSCC schemes perform better than the baseline without rate allocation.

Pu et al. [182] consider the rate allocation problem for the transmission of scalable images over parallel channels. A JSCC system consisting of an image source codec and an RCPT channel codec is proposed. Different from the aforementioned formulation of the optimization problem, distortion-based rate allocation problem for parallel channels is generally difficult to solve since the search space can be expanded largely and the additional constraints for different subchannel arise. To tackle this problem, the authors consider a rate-based optimization problem that maximizes the expected length of correctly received data. The rate-optimal solution provides a good approximation to the distortion-optimal solution. The simulation results of the comparison of mean-squared error (mse) and peak signal-to-noise ratio (PSNR) for the transmission of images over parallel Gaussian channels show that the proposed JSCC schemes outperform the traditional approaches.

Different from the above solutions, Zhang et al. [183] propose a JSCC scheme with power-minimized rate allocation for video transmission over wireless channel. The optimization problem is formulated as the minimization of the summation of processing power for the source and channel encoders and the transmission power with the constraints of total rate and end-to-end distortion. The simulation results show that the proposed joint power-control and rate allocation scheme achieves higher power savings compared to the conventional JSCC scheme.

C. Joint Designs

While the aforementioned designs are considered under the JSCC context, they inherently rely on separate source and channel codes that are either jointly decoded or optimized jointly. On the other hand, there have also been many efforts in the literature to provide genuinely joint designs. The linear coding scheme of Hellman [139], mentioned previously, can be considered as an example of a joint design although it is obtained by a direct combination of two linear codes for source and channel coding, respectively.

In the case of ideal Gaussian sources and channels, the optimality of uncoded transmission has been well-known, as highlighted in Section II-A. However, this optimality breaks down when there is a mismatch between the source and channel bandwidths, and mapping a multidimensional source directly to a multidimensional channel has been a challenging open problem for many decades. A common approach is based on the use of space-filling curves for bandwidth compression, originally proposed by Shannon [184] and also studied in depth by Kotelnikov [185]. More recent works include [186], [187], and [188]. Hybrid digital and analog schemes that combine vector quantization with analog mappings were considered in [189] and [190] for robustness. The optimal mapping is identified in [191] when both the encoder and decoder are constrained to be linear.

Inspired by the theoretical optimality of uncoded transmission for certain source and channel statistics, practical JSCC schemes for image and video transmission were proposed in [192] and [193], respectively. This idea was later popularized as SoftCast in [194] when uncoded transmission is directly applied after JPEG compression. SoftCast applies a DCT on the input image and transmits the DCT coefficients directly over the channel using a dense constellation. Compression is obtained by discarding blocks of DCT coefficients whose energy is below a certain threshold. On the other hand, the index of the discarded blocks is sent as meta-data to the receiver for reconstruction—in that sense, SoftCast is a hybrid scheme combining uncoded/analog transmission with digital communication. Since the encoder mapping is linear for the transmitted coefficients, and the coefficients are corrupted by additive noise directly, the resultant PSNR of the reconstruction at the receiver is linearly related to the channel signal-to-noise ratio (CSNR). This resolves the cliff effect problem encountered in SSCC benchmarks.

Many variations and improved versions of SoftCast have been introduced in [188] and [195]. A theoretical analysis of SoftCast is presented in [196], which highlights the importance of a decorrelation transform and the energy modeling of the underlying signals. Wavelet transform is employed in [195] and Karhunen–Loeve transform (KLT) is applied in [197] as the decorrelation transform instead of 2D-DCT. ECast [198] focuses on joint subcarrier matching and power allocation for optimal end-to-end performance of transmission. In [199] and [200], compressed sensing (CS) has been used for wireless video transmission, where l_1 approximation is considered for recovery. While this allows the receiver to employ convex optimization tools, it is still complex for video streaming applications with strict delay constraints. Iterative algorithms have been proposed to approximate the solution faster, which achieve

reconstruction through iterative thresholding, at the cost of increased error [201]. This algorithm applies CS on the pixels directly and requires no meta-data as long as the measurement matrix is agreed a priori between the transmitter and receiver. In [202], after applying 2D-DCT on image blocks and thresholding, a novel grouping of the coefficients is applied, where coefficients of the same frequency component are grouped into vectors. Each vector is then multiplied with a pseudorandom measurement matrix, whose size depends on the sparsity level of the corresponding vector. Finally, a scaling factor is applied to the results of this multiplication, which corresponds to power allocation across different frequency components. The receiver employs a combination of approximate message passing (AMP) and mmse estimation. AMP is a low-complexity iterative thresholding algorithm, which does not need to know the exact positions of the nonzero elements in the sparse vector.

IV. ML-BASED JSCC CODE DESIGN: DEEP LEARNING ERA

As outlined above, there have been ongoing efforts on the design of practical joint source–channel codes over the decades since Shannon, they have not found applications in practical systems, and almost all existing communication systems rely on separate digital codes. While this is partially due to the modularity of separation that makes it attractive from an engineering design point of view, it is also because the existing JSCC designs either did not reach the same level of performance achieved by state-of-the-art separate codes or relied on unwieldy and complex designs that made their implementation in practice infeasible.

On the other hand, there has been significant progress in practical joint source–channel code design over the past several years, thanks to the developments in deep learning technology. In general, a joint source–channel code is a pair of mappings as in Section II-A: the first is the encoder mapping from the source signal space to the channel input space under the channel input cost constraint, while the second is the decoding mapping from the channel output space to the source reconstruction space. Separate source and channel codes, or most JSCC code designs that rely on the adoption and combination of existing hand-crafted source and channel codes, impose certain limitations on the transformations that can be used. In particular, they require mapping of the source signal into a sequence of bits. Such structured codes are more amenable to optimization; however, they are combinatorial problems in essence. This combinatorial nature makes the employment of deep learning techniques in the design of source and channel codes challenging. In the case of source coding, typically, a quantization step must be employed, which is not differentiable, and has to be approximated through various methods (e.g., straight through estimator, stochastic quantization, etc.) [203]. The channel coding problem is also highly challenging as the number of classes grows exponentially with the code length, and the target
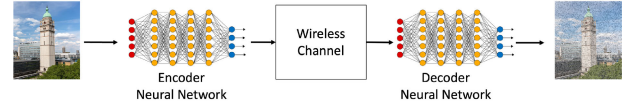
Fig. 17. *DeepJSCC schemes rely on an autoencoder pair trained jointly on a dataset of input signals and a channel model.*

error probability levels are extremely low compared to typical ML tasks, which increases the training time and sample complexity. On the other hand, one can argue that deep learning is a better match for the design of joint source–channel codes, which in essence do not impose any combinatorial constraints, and simply require mapping similar source signals to nearby points in the channel input space in order to limit the reconstruction distortion in the presence of channel noise. In the remainder of this section, we will provide an overview of the recent developments in deep learning-aided design of joint source–channel codes, and sketch some of the challenges in their adoption in practical communication systems.

A. DeepJSCC for Wireless Image Transmission

Deep learning enabled JSCC was first proposed in [11] and applied to wireless image transmission. The scheme, called *DeepJSCC*, uses a trainable autoencoder architecture, with a nontrainable channel between the encoder and decoder, to learn a mapping directly from the image pixels to a set of channel input symbols, and vice versa at the decoder (see Fig. 17). That is, in DeepJSCC, the encoder and decoder mappings are parameterized by neural networks $\boldsymbol{\Theta}_e$ and $\boldsymbol{\Theta}_d$, and can be denoted by $f_{\boldsymbol{\Theta}_e}^{m,n}$ and $g_{\boldsymbol{\Theta}_d}^{m,n}$, respectively. Here, $m = 3 \times h \times w$ denotes the input dimension of the encoder neural network, where 3 represents the R, G, and B components of the colored input image, and h and w denote its height and width, respectively. The corresponding source–channel rate is given by $r = m/n$ as before. In the DeepJSCC literature, the performance is typically quantified in terms of the bandwidth ratio, which is the inverse of the source–channel rate, denoted by $\rho = n/m$.

The two networks are then trained jointly in an end-to-end fashion. The goal is to solve the following optimization problem as before:

$$\min_{\boldsymbol{\Theta}_e, \boldsymbol{\Theta}_d} \mathbb{E}\left[d\left(S^m, \hat{S}^m \right) \right]. \tag{38}$$

However, unlike in information theoretic formulation, for practical image communication problem, we do not have a statistical model of the source S^m, or an additive single-letter distortion measure. Therefore, instead of vying for a closed-form expression, we optimize the neural network parameters using stochastic gradient descent over a dataset of images $\mathcal{D}$

$$\min_{\boldsymbol{\Theta}_e, \boldsymbol{\Theta}_d} \frac{1}{|\mathcal{D}|} \sum_{S^m \in \mathcal{D}} \left[d\left(S^m, \hat{S}^m \right) \right]. \tag{39}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

We note that this formulation is general and can be applied to any channel and distortion measure. Here, the channel can be treated as another layer in between the encoder and decoder neural networks, yet an untrainable one. For the standard additive Gaussian noise channel, the gradient trivially backpropagates through the channel layer. In general, it may be difficult to obtain an explicit gradient for the channel model, or we may not even have a channel model, but simply have access to channel input–output pairs. In that case, we can first train a generative model of the channel [204], [205], [206], which can then be used for training the DeepJSCC encoder/decoder pair.

We would like to highlight that the DeepJSCC approach can be considered as a generalization of the linear/uncoded transmission schemes introduced in Section III-C, which were limited to hand-designed transformations of the input image. Instead, the solution identified through the training operation can be a highly nonlinear mapping from the input source sample to the channel, and in general, its complexity will be dictated by the adopted neural network architecture. Moreover, classical image compression solutions are designed for general natural images and cannot adapt to the statistics of particular types of images. Instead, DeepJSCC encoder/decoder pair can be trained for a certain source modality or a particular dataset (e.g., face images). On the other hand, for the solution to generalize to arbitrary natural images, we would need a richer training dataset.

This end-to-end training approach also allows adopting any arbitrary differentiable loss function between the input and its reconstruction, not only the mse loss, i.e., PSNR. For example, Bourtsoulatze et al. [11] used PSNR as well as the more perceptually aligned structural similarity index measure (SSIM) and showed that DeepJSCC is superior to SSCC under both measures, while the gain under the SSIM loss is even more profound. More recent articles in the literature consider more involved perceptual metrics, such as the learned perceptual image patch similarity (LPIPS) [207], or the Fréchet inception distance (FID) score [208], which are widely used to quantify the perceptual realism of images [209].

In Fig. 18, we present the PSNR performance achieved by DeepJSCC after being trained on a portion of the CIFAR-10 dataset over an AWGN channel in (8). Here, each input is a colored image from the dataset of size 32×32 pixels and is mapped to a channel input codeword of length $n = 256$ symbols. This corresponds to a bandwidth ratio of $\rho = 256/(3 \times 32 \times 32) = 1/12$. Each of the curves denoted by "DeepJSCC" corresponds to an encoder/decoder fully convolutional neural network pair proposed in [40] trained on a different CSNR, $\mathrm{SNR_{train}}$, denoted by $\Theta_e(\mathrm{SNR_{train}})$ and $\Theta_d(\mathrm{SNR_{train}})$, and tested over a range of channel conditions, $\mathrm{SNR_{test}} \in [0, 20]$ dB. Thanks to the fully convolutional architecture, the trained encoder/decoder pair can be used for transmitting images of arbitrary size. The architecture only fixes the source–channel code rate, so that if the input image
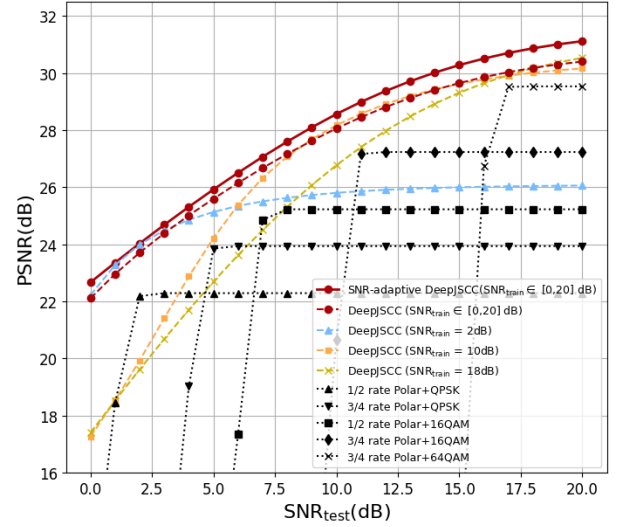
Fig. 18. *Comparison of DeepJSCC performance with that of SSCC on the CIFAR-10 dataset transmitted over an AWGN channel with different SNR values [12].*

size is doubled, the transmitted codeword length will also be doubled.

The performance achieved by a separation-based coding scheme is also included in Fig. 18 as the black lines, which employ better portable graphics (BPG) image compression codec [210] together with polar codes for channel coding at different rates and constellation sizes. Here, we have searched over a range of different code rate and constellation size pairs, and included only the best-performing pairs. As expected, the separation-based scheme has a threshold structure. The code achieves the best performance at a certain $\mathrm{SNR_{test}}$ value, and its performance drops sharply below this threshold, known as the *cliff effect*. Also, the performance remains constant even if the channel quality improves, which is known as the *leveling-off effect*. In contrast to the threshold behavior of the separation scheme, we observe that the DeepJSCC scheme trained for a certain target $\mathrm{SNR_{test}}$ value exhibits a graceful degradation behavior; that is, its performance gracefully degrades as the test SNR drops below the training SNR, and it slowly improves when the SNR increases. This behavior of DeepJSCC is more in line with those of analog modulation schemes, e.g., AM/FM. However, to clarify the distinction, DeepJSCC is still a discrete-time coding scheme, in the sense that, the input signal is converted into discrete samples of potentially continuous real values, which are then mapped to a modulated input waveform [e.g., orthogonal frequency division multiplexing (OFDM)] with discrete-time continuous-amplitude parameters. The main difference compared to conventional modulation schemes is the lack of a discrete constellation set [e.g., BPSK or M-quadratic-amplitude modulation (QAM)], limiting the possible input signals to a finite set of constellation points.

Despite the graceful degradation, we can see from Fig. 18 that the best performance is achieved when the training and test SNRs match. Achieving the convex hull of the performance achieved by all possible such curves

would require training and storing a separate pair of neural networks for each SNR value, or at least for small ranges of SNR values. This is unlikely to be feasible in practical systems, particularly those that need to operate over a wide range of SNR values, due to memory constraints. Moreover, it can quickly become intractable when we go beyond a simple single-input single-output (SISO) system considered here.

One alternative approach is to train a single encoder/ decoder pair to be used over a large range of channel conditions. We see in Fig. 18 that DeepJSCC trained over the whole range of channel SNRs from 0 to 20 dB can achieve competitive performance, not significantly below those achieved by encoder–decoder pairs trained at the target SNR values. We note that this scheme is equivalent to communicating without any channel state information (CSI) at either end of communication, while the channel is time-varying. This is equivalent to a block-fading model, where the channel noise variance remains constant during the transmission of each image, but changes independently from one transmission to the next, taking SNR values uniformly distributed between 0 and 20 dB. This shows that DeepJSCC is capable of learning some form of implicit pilot transmission at the encoder and channel estimation at the decoder, in order to adapt to the varying channel condition. This can be a significant advantage in mobile scenarios when accurate channel estimation can be costly, or even infeasible.

When the system employs channel estimation and channel state feedback, CSI will be available at both the transmitter and the receiver. In that case, rather than using this information to choose the encoder/decoder parameters from among a set of pretrained networks, one can feed this information to the encoder/decoder architectures, and let the networks pick up the appropriate parameter values according to the channel state. For this, an attention feature (AF) module is introduced in [211], which allows adapting the network operations to the CSI, so that in the case of poor channel conditions, only the most important features of the input image are transmitted, but with more protection against noise. The network with the AF model is trained with random channel states so that it learns to adapt its parameters to the CSI. The performance achieved by this SNR-adaptive DeepJSCC architecture is the top curve in Fig. 18, clearly demonstrating that a single network is capable of learning to transmit the input images at any given channel state.

This idea is further extended to OFDM systems in [212] through a double attention mechanism. Note that, when there are many parallel channels available for transmission, each with a different quality, the encoder/decoder networks need to learn not only how much error protection to employ over each channel, but also to which channel each input feature must be mapped. Note that, in the case of classical SSCC over parallel channels, we identify the number of bits that can be transmitted reliably over each channel, and the compressed bits are mapped to different channels according to these numbers; that is, more bits are transmitted over the better channel. In the case of the DeepJSCC architecture proposed in [212], this is learned through the proposed double attention mechanism.

Another limitation of the DeepJSCC architectures in [11], [211], and [212] is that a separate network needs to be trained for each source–channel rate. A bandwidth-agile architecture is proposed in [213], where the image is transmitted in such a way that it can be reconstructed from only a limited portion of the transmitted codeword. This can be considered as a "successive refinement" approach, where the receiver's reconstruction quality increases as it receives more symbols over the channel. A more flexible scheme is considered in [214], where the available bandwidth is dictated by the higher layers (e.g., the medium access control layer) and is given to the encoder and decoder as part of the CSI. The results in [214] show that a single pair of encoder/decoder networks, built upon the Swin transformer architecture [215], can adapt to arbitrary channel SNR and source–channel rate values with only a marginal loss in the performance compared to networks trained for specific channel conditions. This significantly increases the practicality of DeepJSCC, showing that we only need a single network to be deployed on devices for communication over a large variety of channel conditions.

The idea of JSCC in the finite blocklength regimes relies on mapping the source symbols directly to channel symbols. This not only improves the performance by expanding the set of possible transforms compared to separation-based schemes, which impose a digital bottleneck but also provides the graceful degradation with noise. When the channel output is a noisy observation of the input signal, the input can be estimated directly from the channel output, and the estimation error directly depending on the channel noise. In the case of separation, on the other hand, the receiver first needs to detect the channel codeword, which typically has a threshold behavior in the case of coding over long blocklengths. However, relying on the estimation of the source signal at the receiver means that the amount of noise in the reconstructed sequence will be random and depend on the channel realization. This becomes a limitation of DeepJSCC when applied over multihop networks, as it results in noise accumulation. Zhang et al. [216] proposed a recursive training method for DeepJSCC, where a separate DeepJSCC encoder/decoder pair is trained specifically for each hop of the network so that the encoder/decoder pair can adjust to the statistics of the noise images recovered at the relay nodes. This approach improves the performance with respect to using the same encoder/decoder pair throughout the network at the expense of significant increase in complexity; however, it cannot prevent performance degradation as the number of hops increases. An alternative hybrid scheme is proposed in [217] and [218], considering the transmission over a network, where only the first hop is wireless and, thus, time-varying, while

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

the remaining hops are through the core network with limited but fixed capacity. Here, DeepJSCC is used only for the first wireless link, while the received noisy channel output is quantized and forwarded digitally through the core network. This scheme mitigates the problem of noise accumulation over multiple hops, while still providing graceful degradation with the quality of the wireless first hop.

Subsequent works have focused on extending the DeepJSCC architecture in various directions, including to channels with feedback [24], [219], [220], to MIMO channels [221], [222], as well as to multiuser networks [216], [223], [224]. The scheme proposed in [219], called DeepJSCC-f, is inspired by the Schalkwijk–Kailath scheme presented in Section II-C. Each image is transmitted over multiple blocks: after each block, the transmitter first recovers the image reconstructed by the decoder using the feedback signal. This reconstruction is then fed into the encoder together with the original image to generate the next channel block. This is shown to outperform the benchmark obtained by BPG compression followed by a capacity-achieving channel code. However, it comes at the cost of increased complexity as a different encoder–decoder pair is trained for each block. More recently, a transformer-based architecture is proposed in [220], where a single encoder–decoder network pair is trained to refine the receiver's reconstruction at each block. This scheme enjoys both improved performance and reduced complexity compared to [219], emphasizing the importance of combining both information theoretic ideas with careful architecture choice when trying to identify the best solutions in practice.

It is shown in [225] that DeepJSCC scheme can also benefit from correlated side information at the receiver side, the scenario whose theoretical limits we have presented in Section II-D. In this scenario, the optimality of separation breaks down in the finite blocklength regime, and the scheme presented in [225] outperforms the one that combines the state-of-the-art neural network-based lossy compression scheme that takes the side information into account [226] combined with a capacity-achieving channel code.

DeepJSCC approaches can also benefit from generative models. A generative model is a type of ML model whose goal is to learn the underlying patterns or statistical relations in data in order to generate new unseen samples that are statistically similar to those in the training dataset. Mathematically, the goal of a generative model is to learn a joint distribution of the data features. In the context of JSCC, the availability of a generative model for the underlying source data can be considered as a prior knowledge of the source distribution. In a variational autoencoder (VAE), a sample from the latent distribution is mapped to a sample from the data distribution. Therefore, the goal of the transmitter is to convey the right sample from the latent space to the decoder. This idea was used in [227] and [228] for JSCC over discrete channels, which are

not differentiable, so not convenient for the end-to-end training approach of DeepJSCC. A generative adversarial network (GAN) is used in [229], where the transmitter conveys the Gaussian sample that is then fed into the generator network at the receiver. This approach can also be motivated by the information theoretic optimality of uncoded transmission of Gaussian samples. Therefore, one can expect that if a powerful generative model is available at the decoder, the transmitter can be simplified as opposed to the standard DeepJSCC approach.

Generative models can also be used to extend the performance of DeepJSCC to more specific datasets [229], [230], [231]. Consider a transmitter/receiver pair equipped with DeepJSCC encoder/decoder networks trained for a generic image dataset (e.g., ImageNet). Assume that they are then used to transmit a specific type of image, e.g., face images. A generative model of the transmitted images, if available at the receiver, can be used to improve the quality of the image reconstructed by the DeepJSCC decoder, producing much more visually pleasing results, especially in extremely low bandwidth and noisy channels. In this context, the generative model can be considered as a side information available at the receiver. In particular, diffusion-based generative models have been used in the context of JSCC in [230], [231], and [232].

B. Variable-Length DeepJSCC

Constrained by the computational capabilities of a single encoder and decoder, particularly in the case of convolutional networks, the performance of DeepJSCC architecture diminishes as source dimensionality increases, sometimes even underperforming classic separated schemes. One way to improve the performance is to consider a variable-length coding scheme [233], [234], that is, to dynamically adjust the channel codeword dimension based on the input image. An alternative nonlinear transform coding (NTC)-based variable-length DeepJSCC scheme is proposed in [235], called *NTSCC*. NTSCC first extracts a semantic latent representation $V \in \mathcal{V}$ of the source signal and then introduces a conditional entropy model $p_{V|Z}$ in the latent space, i.e.,

$$p_{V|Z}(V|Z) = \prod_i \left(\mathcal{N}\left(v_i|\mu_i, \sigma_i^2\right) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right) \quad (40)$$

where a hyperprior Z is learned to serve as side information for approximating the Gaussian distribution parameters μ_i and σ_i for each latent element v_i, and "$*$" is a convolutional operation with a standard uniform distribution. The learned entropy model $-\log p_{v_i|Z}(v_i|Z)$ indicates the summation of entropy of each v_i. Thus, the information density distribution of V is captured.

To achieve adaptive rate allocation for DeepJSCC, the hyperprior Z is regarded as a prior for the DeepJSCC codewords. Specifically, the bandwidth cost, such as the number of OFDM subcarriers, k_i for transmitting v_i, can

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

be determined as

$$k_i = Q\left(-\eta \log p_{v_i|Z}\left(v_i|Z\right)\right) \tag{41}$$

where Q represents the scalar quantization, and μ is a scaling factor. Consequently, with the help of the transformer architecture and a rate attention mechanism, NTSCC closely adapts to the source distribution and facilitates source content-aware transmission. Besides, NTSCC mitigates the disparity between the marginal distribution of latent representations for a specific source sample and the marginal distribution intended for the ensemble of source data for which the transmission model was designed. We note here that the k_i values must be reliably conveyed to the receiver for the receiver to be able to recover the latent representation. In that sense, NTSCC is a hybrid scheme; however, the amount of meta-data that needs to be digitally transmitted is relatively small, and thus, it can be transmitted reliably without sacrificing much of the available channel bandwidth.

Similar to the rate–distortion modeling in [236], the optimization problem of NTSCC is formulated within the context of variational modeling, wherein the variational density $q_{Y,Z|S}$ is used to approximate the posterior distribution $p_{Y,Z|S}$. This approximation is achieved by minimizing their Kullback–Leibler (KL) divergence over the data distribution p_S. Accordingly, the optimization of NTSCC system can be formally converted to the minimization of the expected channel bandwidth cost, as well as the expected distortion of the reconstructed data versus the original, which leads to the optimization of the following rate-distortion tradeoff:

$$\mathcal{L}_{\text{BD}} = \mathbb{E}_{S\sim p_S}\left(\lambda\left(-\eta \log p_{V|Z}\left(V|Z\right) + d\left(S, \hat{S}\right)\right)\right) \tag{42}$$

where the weight factor $\lambda > 0$ dictates the tradeoff between the channel bandwidth cost and the distortion. The scaling factor $\eta > 0$ correlates the estimated entropy to the channel bandwidth cost and is associated with the source–channel codec capability.

Results in [235] demonstrate that the NTSCC method yields significantly improved coding gain and rate-distortion performance on well-established perceptual metrics, including PSNR, MS-SSIM, and LPIPS. Furthermore, achieving the same end-to-end transmission performance, the proposed NTSCC method can lead to a great reduction in bandwidth cost when compared to both emerging analog transmission schemes that employ the standard DeepJSCC and classical separation-based digital transmission schemes.

The improved NTSCC architecture in [237] introduces a contextual entropy model to capture spatial correlations among semantic latent features more effectively, enabling more precise rate allocation and contextual JSCC. It also incorporates an online latent feature editing method to allow for more flexible coding rate control aligned with specific semantic guidance. The experimental verification demonstrates that the improved NTSCC system achieves approximately a 16.35% reduction in channel bandwidth compared to the state-of-the-art separation-based baseline combining BPG or versatile video coding (VVC) image compression with 5G LDPC error correction codes.

Adaptive semantic communication (ASC) system [238] emphasizes the practical application of the model based on NTSCC. Specifically, when confronted with disparities between the distribution of test data or channel responses and the conditions encountered during training, the model may exhibit suboptimal performance. The ASC system leverages the overfitting characteristics of deep learning modules and adapts semantic codecs or representations to individual data or channel state instances. The entire ASC system design is formulated as an optimization problem with the objective of minimizing the loss function, which constitutes a three-way tradeoff between data rate, model rate, and distortion terms. While achieving equivalent end-to-end transmission performance based on objective metrics like PSNR, the ASC system can yield up to a 41% reduction in bandwidth costs compared to the state-of-the-art engineered transmission scheme (VVC combined with 5G LDPC coded transmission).

C. DeepJSCC Over Multiuser Channels

One of the simplest multiuser network models is the relay channel shown in Fig. 12. DeepJSCC over the relay channel has been studied in [223], [239], and [240]. A process-and-forward scheme is proposed in [223] and [240], which relies on DeepJSCC encoding/decoding at the source and destination terminals, as well as a deep learning-based processing at the relay. In the case of a half-duplex relay channel, one of the challenges is to decide on the time duration relay listens before starting its transmission. In the case of a decode-and-forward scheme [103], where the relay decodes the message of the source terminal before forwarding it, the relay listen and forward time durations can be adaptive to the source–channel link quality, where the relay listens until it accumulates enough information to guarantee correct decoding of the message. However, this is not possible in the case of DeepJSCC since the relay will always have a noisy estimate of the source signal. This is optimized through a search algorithm in [240]; however, an adaptive mechanism that estimates the quality of relay's estimate can be trained to make this decision adaptively. In the case of a full-duplex relay, on the other hand, the relay can transmit the whole time duration; however, it needs to receive a certain amount of information before starting to forward it. A block coding scheme is proposed in [218], inspired by the relaying schemes proposed in [103]. In this scheme, the source transmits the information in blocks, and the relay decides on its transmitted signal at each block based on what it has received up to that point.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

In general, this would require training a separate relay encoder for each block, which learns what to forward in that particular block. We note that the relay's goal is not only to forward as much relevant information as possible to the destination, but also to do this in a coherent manner with the source signal, as the two are received at the destination simultaneously. While decreasing the number of blocks is beneficial in terms of utilizing the relay (since the relay waits initially until it receives the first block), this would also increase the number of blocks, and thus, the complexity. In [218], a novel transformer-based relaying scheme is used throughout all the blocks. This function is trained in a sequence-to-sequence manner, where the previously received signals and transmitted relay codewords are accumulated at each block to increase the amount of relevant information forwarded by the relay. It is also shown that the relay transmits correlated signals with the source, showing that the scheme learns to act as distributed transmit antennas.

In general, the extension of DeepJSCC to BCs is rather trivial. Indeed, transmission over a fading channel without a CSI available at the transmitter, as studied in [11], can also be considered as communicating over a BC as we can treat each fading channel state as a virtual user with a different channel quality. DeepJSCC over a degraded BC is studied in [241].

DeepJSCC over a MAC is studied in [242]. In the case of a MAC, the challenge is twofold: on the one hand, we would like to reduce the interference among the transmitters. On the other hand, the receiver after recovering the noisy signals should be able to attribute each signal to its correct transmitter. A trivial way to deal with this problem is to train multiple transmitter networks together with the DeepJSCC decoder. However, such a scheme would not scale, as each node will need to have access to all of these encoding functions as it may act as a different encoder at each individual scenario. Instead, Yilmaz et al. [242] propose a scheme in which all the transmitters utilize the exact same DeepJSCC encoder; however, they additionally employ a low-dimensional embedding, which would identify them as a separate transmitter, and allow the receiver to distinguish among different transmitters. This can be considered as a learned code-division multiple access (CDMA) scheme.

D. DL-Aided JSCC for Video Transmission

Video, which accounts for more than 80% of all Internet traffic, is a type of source that is known to be particularly hard to transmit reliably, due to the high-data-rate requirement to achieve high video quality and frame rate. Similar to the image case in Section IV-A, we can design an autoencoder architecture that maps video frames to channel symbols, and vice versa, to learn a JSCC for video transmission. One key difference between images and videos is the temporal correlation between frames in video. In a standard video sequence, motion differences between successive frames are typically very small,

and video sequences can be represented very efficient by identifying intermittent key frames, which are treated as independent images, and storing only the motion and residual information for the remaining frames, thereby exploiting temporal redundancy. Motion information typically refers to translation vectors, which can be used to translate a given frame to subsequent frames. Since motion vectors cannot capture the occlusion and disocclusion of objects, residual information, which captures the difference between the translated frame and the ground truth frame, is also stored.

Existing works on bringing DeepJSCC to video transmission have largely exploited the aforementioned temporal redundancy among frames to achieve efficient bandwidth utilization. Tung and Gündüz [243] train separate autoencoders to transmit the key frames and temporal information (i.e., motion vectors and residual information) and utilize a reinforcement learning-driven policy for allocating channel bandwidth to each frame. By dynamically allocating channel resources, they show that their scheme, called DeepWiVe, can outperform state-of-the-art video compression codecs H.264 and H.265 when transmitting video over AWGN and Rayleigh fading channels.

Wang et al. [244], rather than using reinforcement learning, learn an entropy model of the temporal information and use variable channel uses to dynamically adjust channel resource usage depending on the information content of the frames. The idea is that frames with higher entropy require more channel uses and by using variable channel uses, the average channel use can be lower, similar to variable length coding. They show that their scheme can perform better than H.265 even when assuming capacity-achieving channel coding is used.

E. DL-Aided JSCC for Text Transmission

As the notion of semantics is traditionally related to natural language, or text in general, semantic communication of text can offer interesting insights into the benefits of DeepJSCC. One of the first articles to apply DeepJSCC to text transmission is [15], where the authors proposed DeepSC for text transmission. DeepSC leverages the transformer architecture, which has been so successful at natural language processing (NLP) tasks, to extract semantic features that are useful for JSCC. To measure the distortion, the authors propose to use a pretrained large language model, BERT [245], and compute the cosine similarity between two sentences in the embedding space of BERT. The idea is that, having been trained on a large corpus of text data, BERT has an internal representation of natural language that is consistent with semantic meaning. Therefore, the cosine similarity between the embeddings of two sentences represents the semantic similarity between them.

Similar to what was observed in DeepJSCC for image and video transmission, it is also observed that the

performance of DeepSC, as measured by the bilingual evaluation understudy (BLEU) score [246], gracefully degrades as the channel noise increases, unlike the cliff-edge drop off that SSCC exhibits. Moreover, DeepSC performs much better than SSCC under fading channel conditions. Perhaps more interestingly, DeepSC is more likely to produce coherent sentences that may not be exactly the same as the source, but is semantically similar, showing that not only is the loss function effective at capturing source semantics, it is also a good illustration of how semantic communication is related to JSCC.

Further works have addressed other issues with text transmission. For example, in [247], the spectral efficiency problem is addressed by first defining a semantic spectral efficiency metric, which measures on average how much semantic information is transmitted per unit of bandwidth (usually measured in hertz). The authors define semantic information using the semantic loss used to train DeepSC and maximize the average spectral efficiency over all users. Peng et al. [14] address potential adversarial attacks on the DeepSC transceiver by improving the underlying attention mechanism in the transformer layers. In essence, they train an additional layer at the output of the attention layer to detect which input tokens are potentially corrupted. This information is then passed to the subsequent attention layers to inform it of which tokens to ignore.

F. Other Sources and Channels

Naturally, JSCC can be applied to many other source signals, and the application of neural networks for the design of competitive DeepJSCC schemes allowed researchers to design end-to-end communication systems for a wide variety of source and channel distributions and applications.

One of the promising applications of DeepJSCC is for CSI feedback in massive MIMO systems, where the base station (BS) requires accurate and timely CSI for beamforming and interference management. In a frequency-division duplex (FDD) massive MIMO system, user equipments (UEs) estimate their downlink CSI based on the pilots transmitted by the BS. This CSI estimates then need to be transmitted to the BS over the uplink channel. It is, hence, crucial to design an efficient feedback scheme that provides more accurate CSI to the BS while introducing a limited feedback overhead. There has been significant progress in recent years in the design of efficient CSI compression techniques using neural networks [248], [249], [250]. However, since these rely on digital compression techniques, they need an accurate estimation of the uplink channel state. An alternative analog CSI feedback scheme is proposed in [251], which directly maps the estimated channel coefficients to channel input using DeepJSCC. It is shown in [251] that the joint design can significantly improve the achieved downlink communication rate compared to separation. This approach was later extended in [252] by incorporating SNR adaptivity.

Another source that can benefit from the JSCC approach is point cloud, which refers to a set of points that collectively depict a physical object or a scene [253]. A point cloud conveys not only the geometry of the object, through the coordinates of the points in the space, but also various attributes associated with each point, such as color, intensity, and curvature. It has found extensive applications across different fields in recent years, including robotics, autonomous driving, VR/AR, and metaverse. While there have been significant research efforts for the design of more efficient compression algorithms for point clouds, recent works show that JSCC can provide gains also for this challenging source modality [254], [255], [256].

With the increasing success and popularity of deep learning models for a wide variety of tasks, it is expected that another type of information source that may constitute a significant portion of future wireless data traffic is model weights. While each day new results appear showing that neural networks achieve the state-of-the-art performance in yet another task, it is not reasonable to assume that a mobile device will be capable of storing all possible neural networks locally to be deployed at the time of need. Instead, it is more likely that the model parameters for a desired task will be downloaded on demand from a server when they are needed. This requires transmitting those parameters as quickly and efficiently as possible over wireless channels. Jankowski et al. [16], [257] proposed a joint training and transmission strategy, where the model parameters are mapped directly to the channel inputs, such that a noisy version of the model will be recovered at the receiver depending on the channel conditions.

In addition to different source modalities mentioned above (speech transmission is considered in [13] and [17]), researchers have also studied the design of DeepJSCC solutions for different channels, including image transmission over underwater acoustic channels [258], or image storage in memristive devices [259], [260] or even in DNA sequences [261].

V. FURTHER DISCUSSION AND CONCLUSION

In this article, we have presented a comprehensive overview of the foundations of JSCC. While separation provides modularity for the design of complex communication networks, it is known to be suboptimal in practical scenarios when the blocklengths are finite. Yet, after decades of research, the state-of-the-art SSCC schemes present a benchmark that is hard to beat, particularly when the blocklength is relatively large. On the other hand, in the case of nonergodic communication scenarios, or when the blocklength is extremely limited due to resource and latency constraints, JSCC becomes an attractive alternative. Moreover, in the case of many multiuser communication scenarios, we do not even have the theoretical optimality of separation, and the optimal JSCC scheme is not known for even very simple multiple access

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

and BC models. After a brief review of classical approaches to the practical design of JSCC schemes, we have mainly focused on recent developments in JSCC design benefiting from deep learning techniques for various source and channel combinations. It has been recently shown that DeepJSCC approaches can outperform the state-of-the-art separate designs, making them a very attractive alternative, particularly for mobile communication scenarios where channel state changes rapidly over time, and its accurate estimation would require significant channel resources. In light of these significant developments in DeepJSCC design over the past several years, we believe that it is time to start questioning the strict layered architecture of contemporary communication systems and explore alternative joint approaches.

Despite the fact that almost all current communication systems rely on the SSCC principle, there are many potential immediate applications of JSCC in standalone point-to-point communication systems. An existing successful example is an implementation of JSCC for MIMO systems by Amimon, based on UEP. This technology was adopted in professional wireless cameras that were used by many Hollywood production companies and won an Academy Award for Technical Achievement from the Academy of Motion Picture Arts and Sciences in 2021.

Another potential application of JSCC is the transmission of video signals from a drone to a ground station. Current technology relies on Wi-Fi and results in the loss of video signal beyond few kilometers, particularly in busy urban areas. Adoption of JSCC techniques, e.g., DeepJSCC, for such a point-to-point link can enable constant connectivity, where the quality of received video gracefully degrades depending on the channel condition, but is never completely lost. Moreover, it can also work without channel estimation, which is difficult and costly to obtain in highly mobile scenarios. Another application is AR/VR headsets, which require the transmission of high-rate multimedia content. Most current headsets rely on wired connections or very high-quality stable wireless links, which considerably limits their functionality.

On the other hand, despite their superiority from a purely performance point of view, there are still many practical and fundamental challenges that need to be addressed for the potential adoption of JSCC schemes in 6G or future communication standards. First of all, a full-fledged implementation of JSCC requires significant changes in the network architecture as it relies on the joint optimization of the physical layer coding and modulation together with source compression, which is normally handled at the application layer. This creates a fundamental problem in practical applications in mobile networks beyond architectural changes, as normally the content to be transmitted belongs to the user, and is not accessible by the network provider. Hence, JSCC needs to be implemented without having access to the underlying content at the encoder.

Alternatively, the content providers can apply JSCC directly and convey the coded information to the lower layers, which informs the transmitter what to transmit without seeing the content. This can be done by using DeepJSCC-Q as proposed in [243], which implements DeepJSCC on a finite constellation. However, to achieve the potential benefits of DeepJSCC requires using a large constellation (to sufficiently approximate continuous-amplitude transmission), and this would require a large bandwidth to convey such a code over the core network until it reaches the wireless link.

Another challenge in implementing JSCC in legacy network architectures is the adoption of feedback, in particular how to benefit from automatic repeat request (ARQ) tools that are widely used in current network protocols. In the current framework, since the goal is to recover a channel codebook reliably, a cyclic redundancy check (CRC) is added to the data packet to verify correct decoding, and if the transmission fails, the receiver sends a negative acknowledgment (NACK) signal requesting retransmission. This is repeated until the data packet is correctly decoded, or the maximum number of retransmissions is reached. On the other hand, in the case of JSCC, typically a noisy version of the source data is recovered at the receiver, whose quality depends on the channel quality. Therefore, it is not possible to make a binary decision at the receiver. One alternative could be to make a decision whether a desired quality level is reached or not, but it is also difficult to make such a decision solely based on the reconstructed signal at the receiver, without having access to the original source signal. This may require rethinking the whole feedback system in future communication networks going beyond the current single-bit ARQ framework [220].

Adoption of DeepJSCC in standards would also require it to be back-compatible and allow its implementation on legacy terminals. Many existing devices use chipsets that already implement current codes and modulation schemes on their hardware, which may be difficult to circumvent, or require JSCC solutions around these existing codes. Another critical challenge is the security. JSCC schemes mostly rely on the correlation between the transmitted signal and the underlying source signal. While this results in graceful degradation in reconstruction quality at the legitimate receiver as the channel quality drops, it also means that a signal received by a potential eavesdropper in the vicinity will leak information about the source signal. Digital transmission allows applying encryption techniques where the transmitted bits can be completely independent of the source bits, e.g., through a one-time pad, whereas it is not clear how to achieve such an encryption on the continuous amplitude source signal. While some initial solutions to the security problem in DeepJSCC have been proposed [262], [263], [264], [265], [266], more significant research is needed before they meet the necessary security guarantees for implementation in mobile networks or other systems communicating sensitive content. ∎

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[2] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 1, pp. 27–34, Jan. 1988.

[3] M. Feder and N. Merhav, "Universal coding for arbitrarily varying sources," in *Proc. Int. Symp. Inf. Theory*, Sep. 1995, p. 16.

[4] D. Gündüz, D. B. Kurka, M. Jankowski, M. M. Amiri, E. Ozfatura, and S. Sreekumar, "Communicate to learn at the edge," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 14–19, Dec. 2020.

[5] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts*, vol. 19, no. 4, pp. 2322–2358, 4th Quart. 2017.

[6] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained IoT devices," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 1–24, Jan. 2022.

[7] Y. Xiao, G. Shi, Y. Li, W. Saad, and H. V. Poor, "Toward self-learning edge intelligence in 6G," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 34–40, Dec. 2020.

[8] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Jan. 2018.

[9] M. Chen et al., "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.

[10] Y. Xiao, Z. Sun, G. Shi, and D. Niyato, "Imitation learning-based implicit semantic-aware communication networks: Multi-layer representation and collaborative reasoning," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 639–658, Mar. 2023.

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

[12] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. D. Gündüz, "Deep joint source-channel coding for semantic communications," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 42–48, Nov. 2023.

[13] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.

[14] X. Peng et al., "A robust deep learning enabled semantic communication system for text," in *Proc. IEEE Global Commun. Conf.*, Dec. 2022, pp. 2704–2709.

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

[16] M. Jankowski, D. Gunduz, and K. Mikolajczyk, "AirNet: Neural network transmission over the air," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 2451–2456.

[17] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, "Semantic-preserved communication system for highly efficient speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 245–259, Jan. 2023.

[18] D. Gündüz et al., "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, Jan. 2023.

[19] M. Kalfa, M. Gok, A. Atalik, B. Tegin, T. M. Duman, and O. Arikan, "Towards goal-oriented semantic signal processing: Applications and future challenges," *Digit. Signal Process.*, vol. 119, Dec. 2021, Art. no. 103134.

[20] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Commun. Mag.*, vol. 59, no. 8, pp. 44–50, Aug. 2021.

[21] E. Uysal et al., "Semantic communications in networked systems: A data significance perspective," *IEEE Netw.*, vol. 36, no. 4, pp. 233–240, Jul. 2022.

[22] T.-Y. Tung, S. Kobus, J. P. Roig, and D. Gündüz, "Effective communications: A joint learning and communication framework for multi-agent reinforcement learning over noisy channels," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2590–2603, Aug. 2021.

[23] Y. Shao, Q. Cao, and D. Gunduz, "A theory of semantic communication," *IEEE Trans. Mobile Comput.*, early access, May 28, 2024, doi: 10.1109/TMC.2024.3406375.

[24] H. Zhang, S. Shao, M. Tao, X. Bi, and K. B. Letaief, "Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 170–185, Jan. 2022.

[25] Y. Xiao et al., "Reasoning over the air: A reasoning-based implicit semantic-aware communication framework," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3839–3855, Apr. 2024.

[26] Y. Xiao, X. Zhang, Y. Li, G. Shi, and T. Basar, "Rate-distortion theory for strategic semantic communication," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Mumbai, India, Nov. 2022, pp. 279–284.

[27] P. Schwenteck, G. T. Nguyen, H. Boche, W. Kellerer, and F. H. P. Fitzek, "6G perspective of mobile network operators, manufacturers, and verticals," *IEEE Netw. Lett.*, vol. 5, no. 3, pp. 169–172, Jun. 2023.

[28] G. P. Fettweis and H. Boche, "6G: The personal tactile internet—And open questions for information theory," *IEEE BITS Inf. Theory Mag.*, vol. 1, no. 1, pp. 71–82, Sep. 2021.

[29] R. Joda et al., "The Internet of Senses: Building on semantic communications and edge intelligence," *IEEE Netw.*, vol. 37, no. 3, pp. 68–75, May 2023.

[30] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.

[31] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "Wireless image retrieval at the edge," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 89–100, Jan. 2021.

[32] P. Tandon et al., "Txt2Vid: Ultra-low bitrate compression of talking-head videos via text," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 107–118, Jan. 2023.

[33] F. Pezone, O. Musa, G. Caire, and S. Barbarossa, "Semantic-preserving image coding based on conditional diffusion models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 13501–13505.

[34] M. C. Anderson, M. S. Atkins, and J. Vaisey, "Task-oriented lossy compression of magnetic resonance images," *Proc. SPIE*, vol. 2707, pp. 239–250, Apr. 1996.

[35] N. Shlezinger, Y. C. Eldar, and M. R. D. Rodrigues, "Asymptotic task-based quantization with application to massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3995–4012, Aug. 2019.

[36] C. Zhang, H. Zou, S. Lasaulce, W. Saad, M. Kountouris, and M. Bennis, "Goal-oriented communications for the IoT and application to data compression," *IEEE Internet Things Mag.*, vol. 5, no. 4, pp. 58–63, Dec. 2022.

[37] A. El Essaili et al., "Holographic communication in 5G networks," *Ericsson Technol. Rev.*, vol. 2022, no. 5, pp. 2–11, May 2022. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/holographic-communication-in-5g-networks

[38] G. P. Fettweis, "The tactile internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[39] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[40] D. B. Kurka and D. Gündüz, "Joint source-channel coding of images with (not very) deep learning," in *Proc. Proc. Int. Zurich Seminar Inf. Commun.*, 2020, pp. 90–94.

[41] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[42] I. Csiszár, "Joint source-channel error exponent," *Problems Control Inf. Theory*, vol. 9, no. 5, pp. 315–328, 1980.

[43] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1450–1468, Apr. 2006.

[44] 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.

[45] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Proc. Inst. Radio Eng., Int. Conv. Rec.*, 1959, pp. 325–350.

[46] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 558–567, Oct. 1965.

[47] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[48] Y. Kochman, O. Ordentlich, and Y. Polyanskiy, "A lower bound on the expected distortion of joint source-channel coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1332–1336.

[49] L. V. Truong and V. Y. F. Tan, "The reliability function of variable-length lossy joint source-channel coding with feedback," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5028–5042, Aug. 2019.

[50] M. Zheng, J. Gu, M. Ma, and C. Ling, "Joint source-channel polar coding for biased Bernoulli sources at short blocklengths," in *Proc. 11th Int. Symp. Topics Coding (ISTC)*, Aug. 2021, pp. 1–5.

[51] R. L. Dobrushin, "A general formulation of the fundamental theorem of Shannon in the theory of information," *Uspekhi Mat. Nauk*, vol. 14, no. 90, pp. 3–104, 1959.

[52] G. Hu, "On Shannon theorem and its converse for sequence of communication schemes in the case of abstract random variables," in *Proc. 3rd Prague Conf. Inf. Theory, Stat. Decis. Functions, Random Processes*, Liblice, Czechoslovakia, Jun. 1962, pp. 285–333.

[53] A. Aghajan, S. J. Zahabi, M. Khosravifard, and T. A. Gulliver, "On the source-channel separation theorem for infinite source alphabets," *Problems Inf. Transmiss.*, vol. 51, no. 2, pp. 200–204, Apr. 2015.

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

[55] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer, 2014.

[56] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[57] D. Gunduz, C. T. K. Ng, E. Erkip, and A. J. Goldsmith, "Source transmission over relay channel with correlated relay side information," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 611–615.

[58] D. Gunduz and E. Erkip, "Joint source–channel

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

codes for MIMO block-fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 116–134, Jan. 2008.

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

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

[61] J. Liu, S. Shao, W. Zhang, and H. V. Poor, "An indirect rate-distortion characterization for semantic sources: General model and the case of Gaussian observation," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 5946–5959, Sep. 2022.

[62] P. A. Stavrou and M. Kountouris, "The role of fidelity in goal-oriented semantic communication: A rate distortion approach," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 3918–3931, Jul. 2023.

[63] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback-I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.

[64] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, Mar. 2011.

[65] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, Sep. 1973.

[66] S. Shamai (Shitz) and S. Verdú, "Capacity of channels with uncoded side information," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 587–600, Sep. 1995.

[67] D. Gunduz, E. Erkip, A. Goldsmith, and H. V. Poor, "Reliable joint source-channel cooperative transmission over relay networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2442–2458, Apr. 2013.

[68] E. Tuncel, "Slepian-wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.

[69] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, NY, USA: Academic, 1981.

[70] I. Estella Aguerri and D. Gündüz, "Joint source-channel coding with time-varying channel and side-information," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 736–753, Feb. 2016.

[71] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[72] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, 1961, pp. 611–644.

[73] A. Hekstra and F. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44–53, Jan. 1989.

[74] M. Salehi, "Multiple-access channels with correlated sources–coding subject to a fidelity criterion," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 1995, p. 198.

[75] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1509–1523, Apr. 2015.

[76] W. Kang and S. Ulukus, "An outer bound for multiple access channels with correlated sources," in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, Mar. 2006, pp. 240–244.

[77] A. Lapidoth, S. S. Bidokhti, and M. Wigger, "Dependence balance in multiple access channels with correlated sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1663–1667.

[78] B. Güler, D. Gündüz, and A. Yener, "Lossy coding of correlated sources over a multiple access channel: Necessary conditions and separation results," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6081–6097, Sep. 2018.

[79] A. Padakandla, "Communicating correlated

sources over MAC and interference channels II: Joint source-channel coding," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3847–3872, Jun. 2021.

[80] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2714–2752, Jun. 2010.

[81] S. I. Bross and Y. Laufer, "Sending a bivariate Gaussian source over a Gaussian MAC with unidirectional conferencing encoders," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1296–1311, Mar. 2016.

[82] W. Liu and B. Chen, "Interference channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 8027–8037, Dec. 2011.

[83] Y. Murin, R. Dabora, and D. Gündüz, "On joint source-channel coding for correlated sources over multiple-access relay channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6231–6253, Oct. 2014.

[84] T. Han and M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 641–650, Sep. 1987.

[85] A. A. Gohari and V. Anantharam, "An outer bound to the admissible source region of broadcast channels with arbitrarily correlated sources and channel variations," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, vol. 2, Sep. 2008, pp. 301–308.

[86] K. Khezeli and J. Chen, "Outer bounds on the admissible source region for broadcast channels with correlated sources," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4616–4629, Sep. 2015.

[87] L. Yu, H. Li, and W. Li, "Distortion bounds for source broadcast problems," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6034–6053, Sep. 2018.

[88] S. I. Bross, A. Lapidoth, and S. Tinguely, "Broadcasting correlated Gaussians," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3057–3068, Jul. 2010.

[89] C. Tian, S. Diggavi, and S. Shamai (Shitz), "The achievable distortion region of sending a bivariate Gaussian source on the Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6419–6427, Oct. 2011.

[90] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[91] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.

[92] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[93] V. Prabhakaran, D. N. C. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. Int. Symp. Inf. Theory*, 2004, p. 117.

[94] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Information Processing in Sensor Networks*, F. Zhao and L. Guibas, Eds., Berlin, Germany: Springer, 2003, pp. 162–177.

[95] Y. Kochman and R. Zamir, "The joint Gaussian CEO/MAC problem and relaying," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 1554–1561.

[96] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[97] J. Nayak, E. Tuncel, and D. Gunduz, "Wyner–Ziv coding over broadcast channels: Digital schemes," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1782–1799, Apr. 2010.

[98] S. S. Bidokhti, M. Wigger, and R. Timo, "Noisy broadcast networks with receiver caching," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6996–7016, Nov. 2018.

[99] M. M. Amiri and D. Gündüz, "Cache-aided content

delivery over erasure broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 370–381, Jan. 2018.

[100] S. S. Bidokhti, M. Wigger, and A. Yener, "Benefits of cache assignment on degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 6999–7019, Nov. 2019.

[101] D. Gunduz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3927–3944, Sep. 2009.

[102] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source–channel communication over an orthogonal multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3255–3264, Sep. 2007.

[103] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[104] D. Gunduz, E. Tuncel, and J. Nayak, "Rate regions for the separated two-way relay channel," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, 2008, pp. 1333–1340.

[105] D. Gunduz, J. Nayak, and E. Tuncel, "Wyner–Ziv coding over broadcast channels using hybrid digital/analog transmission," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1543–1547.

[106] M. Yuksel and E. Erkip, "Multiple-antenna cooperative wireless systems: A diversity–multiplexing tradeoff perspective," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3371–3393, Oct. 2007.

[107] D. Gunduz, A. Goldsmith, and H. V. Poor, "MIMO two-way relay channel: Diversity-multiplexing tradeoff analysis," in *Proc. 42nd Asilomar Conf. Signals, Syst. Comput.*, Oct. 2008, pp. 1474–1478.

[108] R. Venkataramanan and S. S. Pradhan, "A new achievable rate region for the multiple-access channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 8038–8054, Dec. 2011.

[109] R. Venkataramanan and S. S. Pradhan, "An achievable rate region for the broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6175–6191, Oct. 2013.

[110] Y. Wu and M. Wigger, "Coding schemes with rate-limited feedback that improve over the no feedback capacity for a large class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2009–2033, Apr. 2016.

[111] T. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, Jan. 1989.

[112] T. Han, "A general coding scheme for the two-way channel," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 1, pp. 35–44, Jan. 1984.

[113] A. N. Ravi, S. R. B. Pillai, V. M. Prabhakaran, and M. Wigger, "On the capacity enlargement of Gaussian broadcast channels with passive noisy feedback," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6356–6367, Oct. 2021.

[114] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 623–629, Jul. 1984.

[115] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1423–1438, Jun. 2002.

[116] E. Sula, M. Gastpar, and G. Kramer, "Sum-rate capacity for symmetric Gaussian multiple access channels with feedback," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2860–2871, May 2020.

[117] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 667–671, Jul. 1984.

[118] E. Ardestanizadeh, P. Minero, and M. Franceschetti, "LQG control approach to Gaussian broadcast channels with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5267–5278,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

Aug. 2012.

[119] Y. Murin, Y. Kaspi, and R. Dabora, "On the ozarow-leung scheme for the Gaussian broadcast channel with feedback," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 948–952, Jul. 2015.

[120] A. Lapidoth and M. Wigger, "On the AWGN MAC with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5432–5476, Nov. 2010.

[121] C. Tian, J. Chen, S. N. Diggavi, and S. S. Shitz, "Matched multiuser Gaussian source channel communications via uncoded schemes," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4155–4171, Jul. 2017.

[122] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over a noisy channel," in *Proc. IZS*, Feb. 2018, pp. 4150–4182.

[123] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2044–2066, Apr. 2020.

[124] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4150–4182, Jul. 2020.

[125] S. Borade, B. Nakiboglu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5511–5539, Dec. 2009.

[126] H. Shimokawa, T. Sun Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proc. IEEE Int. Symp. Inf. Theory*, Trondheim, Norway, Jul. 1994, p. 114.

[127] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," 2023, *arXiv:2306.14499*.

[128] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 759–772, Nov. 1987.

[129] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[130] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 4, pp. 533–542, Jul. 1986.

[131] T. Sun Han and S. Amari, "Parameter estimation with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1802–1833, Nov. 1995.

[132] H. M. H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, Mar. 1992.

[133] T. Berger, K. Housewright, J. Omura, S. Yung, and J. Wolfowitz, "An upper bound on the rate distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 664–666, Nov. 1979.

[134] N. Weinberger and Y. Kochman, "On the reliability function of distributed hypothesis testing under optimal detection," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4940–4965, Aug. 2019.

[135] S. Watanabe, "On sub-optimality of random binning for distributed hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 2708–2713.

[136] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over multi-access channels," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–5.

[137] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against conditional independence with multiple decision centers," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2409–2420, Jun. 2018.

[138] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 3, pp. 681–694, Nov. 2020.

[139] M. Hellman, "Convolutional source encoding," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 651–656, Nov. 1975.

[140] T. Ancheta, "Bounds and techniques for linear source coding," Ph.D. thesis, Dept. Elect. Eng., Univ. Notre Dame, Notre Dame, IN, USA, 1977.

[141] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sep. 1995.

[142] A. H. Murad and T. E. Fuja, "Joint source-channel decoding of variable-length encoded sources," in *Proc. Inf. Theory Workshop*, 1998, pp. 94–95.

[143] J. Hagenauer and R. Bauer, "The turbo principle in joint source channel decoding of variable length codes," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2001, pp. 33–35.

[144] 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.

[145] S. Lonardi, W. Szpankowski, and M. D. Ward, "Error resilient LZ'77 data compression: Algorithms, analysis, and experiments," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1799–1813, May 2007.

[146] M. Jeanne, J.-C. Carlach, and P. Siohan, "Joint source-channel decoding of variable-length codes for convolutional codes and turbo codes," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 10–15, Jan. 2005.

[147] G.-C. Zhu and F. Alajaji, "Joint source-channel turbo coding for binary Markov sources," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1065–1075, May 2006.

[148] J. Garcia-Frias and J. D. Villasenor, "Combining hidden Markov source models and parallel concatenated codes," *IEEE Commun. Lett.*, vol. 1, no. 4, pp. 111–113, Jul. 1997.

[149] J. Garcia-Frias and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1671–1679, Sep. 2001.

[150] H. Kfir and I. Kanter, "Efficient LDPC codes for joint source-channel coding," in *Proc. 11th IEEE Int. Conf. Electron., Circuits Syst.*, Aug. 2004, pp. 654–657.

[151] C. Poulliat, D. Declercq, C. Lamy-Bergot, and I. Fijalkow, "Analysis and optimization of irregular LDPC codes for joint source-channel decoding," *IEEE Commun. Lett.*, vol. 9, no. 12, pp. 1064–1066, Dec. 2005.

[152] M. Fresia, F. Perez-Cruz, and H. V. Poor, "Optimized concatenated LDPC codes for joint source-channel coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 2131–2135.

[153] J. He, L. Wang, and P. Chen, "A joint source and channel coding scheme base on simple protograph structured codes," in *Proc. Int. Symp. Commun. Inf. Technol. (ISCIT)*, Oct. 2012, pp. 65–69.

[154] Q. Chen, L. Wang, S. Hong, and Z. Xiong, "Performance improvement of JSCC scheme through redesigning channel code," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1088–1091, Jun. 2016.

[155] C. Chen, L. Wang, and S. Liu, "The design of protograph LDPC codes as source codes in a JSCC system," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 672–675, Apr. 2018.

[156] Q. Chen, L. Wang, S. Hong, and Y. Chen, "Integrated design of JSCC scheme based on double protograph LDPC codes system," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 218–221, Feb. 2019.

[157] Z. Xu, L. Wang, S. Hong, and G. Chen, "Design of code pair for protograph-LDPC codes-based JSCC system with joint shuffled scheduling decoding algorithm," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3770–3774, Dec. 2021.

[158] Y. Lv, J. He, W. Xu, and L. Wang, "Design of low-density parity-check code pair for joint source-channel coding systems based on graph theory," *Entropy*, vol. 25, no. 8, p. 1189, Aug. 2023.

[159] A. Golmohammadi and D. G. M. Mitchell, "Concatenated spatially coupled LDPC codes with sliding window decoding for joint source-channel coding," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 851–864, Feb. 2022.

[160] L. Deng et al., "Joint source channel anytime coding based on spatially coupled repeat-accumulate codes," *IEEE Trans. Commun.*, vol. 71, no. 11, pp. 6215–6230, Nov. 2023.

[161] Y. Dong, K. Niu, J. Dai, S. Wang, and Y. Yuan, "Joint source and channel coding using double polar codes," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2810–2814, Sep. 2021.

[162] Y. Dong and K. Niu, "Double polar codes for joint source and channel coding," *Electronics*, vol. 11, no. 21, p. 3557, Oct. 2022.

[163] O. Y. Bursalioglu, G. Caire, and D. Divsalar, "Joint source-channel coding for deep-space image transmission using rateless codes," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3448–3461, Aug. 2013.

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

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

[166] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 799–809, Jul. 1990.

[167] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 155–160, Jan. 1991.

[168] J. Dunham and R. Gray, "Joint source and noisy channel trellis encoding (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 4, pp. 516–519, Jul. 1981.

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

[170] T. R. Fischer, "Joint trellis coded quantization/modulation," *IEEE Trans. Commun.*, vol. 39, no. 2, pp. 172–176, Feb. 1991.

[171] M. Wang and T. R. Fischer, "Trellis-coded quantization designed for noisy channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1792–1802, Nov. 1994.

[172] D. G. Daut and C. A. Sanders, "Joint source/channel coding using trellis-coded CPFSK," in *Proc. Commun., Global Bridge*, Nov. 1994, pp. 212–217.

[173] H. A. Aksu and M. Salehi, "Joint optimization of TCQ-TCM systems," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 529–533, May 1996.

[174] Z. Lin and T. Aulin, "On joint source and channel coding using trellis coded CPM: Analytical bounds on the channel distortion," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3081–3094, Sep. 2007.

[175] J. Modestino and D. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. COM-27, no. 11, pp. 1644–1659, Nov. 1979.

[176] D. Comstock and J. Gibson, "Hamming coding of DCT-compressed images over noisy channels," *IEEE Trans. Commun.*, vol. COM-32, no. 7, pp. 856–861, Jul. 1984.

[177] D. J. Goodman and C. Sundberg, "Transmission errors and forward error correction in embedded differential pulse code modulation," *Bell Syst. Tech. J.*, vol. 62, no. 9, pp. 2735–2764, Nov. 1983.

[178] D. J. Goodman and C.-E. Sundberg, "Combined source and channel coding for variable-bit-rate speech transmission," *Bell Syst. Tech. J.*, vol. 62, no. 7, pp. 2017–2036, Sep. 1983.

[179] C. Moore and J. Gibson, "Self-orthogonal convolutional coding for the DPCM-AQB speech encoder," *IEEE Trans. Commun.*, vol. COM-32, no. 8, pp. 980–982, Aug. 1984.

[180] Z. Peng, Y.-F. Huang, and D. J. Costello, "Turbo codes for image transmission—A joint channel and source decoding approach," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 868–879, Jun. 2000.

[181] Z. Wu, A. Bilgin, and M. W. Marcellin, "An efficient joint source-channel rate allocation scheme for JPEG2000 codestreams," in *Proc. Data*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

Compress. Conf., vol. 36, 2003, pp. 113–122.

[182] L. Pu, M. W. Marcellin, I. Djordjevic, B. Vasic, and A. Bilgin, "Joint source-channel rate allocation in parallel channels," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2016–2022, Aug. 2007.

[183] Q. Zhang, Z. Ji, W. Zhu, and Y.-Q. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 398–410, Jun. 2002.

[184] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[185] V. A. Kotelnikov, *The Theory of Optimum Noise Immunity*. New York, NY, USA: McGraw-Hill, 1959.

[186] T. A. Ramstad, "Shannon mappings for robust communication," *Telektronikk*, vol. 98, no. 1, pp. 114–128, 2002.

[187] F. Hekland, G. E. Oien, and T. A. Ramstad, "Using 2:1 Shannon mapping for joint source-channel coding," in *Proc. Data Compress. Conf.*, 2005, pp. 223–232.

[188] F. Hekland, P. A. Floor, and T. A. Ramstad, "Shannon–Kotel–Nikov mappings in joint source-channel coding," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 94–105, Jan. 2009.

[189] U. Mittal and N. Phamdo, "Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.

[190] 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.

[191] K.-H. Lee and D. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.

[192] J. M. Lervik and T. R. Fischer, "Robust subband image coding for waveform channels with optimum power-and bandwidth-allocation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, Apr. 1997, pp. 3089–3092.

[193] A. Fuldseth, "Robust subband video compression for noisy channels with multilevel signaling," Dept. Electron. Syst., NTNU, Trondheim, Norway, Tech. Rep., 1997.

[194] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 449–450, Aug. 2010.

[195] X. Fan, R. Xiong, F. Wu, and D. Zhao, "WaveCast: Wavelet based wireless video broadcast using lossy transmission," in *Proc. Vis. Commun. Image Process.*, Nov. 2012, pp. 1–6.

[196] R. Xiong, F. Wu, J. Xu, X. Fan, C. Luo, and W. Gao, "Analysis of decorrelation transform gain for uncoded wireless image and video communication," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1820–1833, Apr. 2016.

[197] A. Hagag, X. Fan, and F. E. Abd El-Samie, "HyperCast: Hyperspectral satellite image broadcasting with band ordering optimization," *J. Vis. Commun. Image Represent.*, vol. 42, pp. 14–27, Jan. 2017.

[198] Z. Zhang, D. Liu, X. Ma, and X. Wang, "ECast: An enhanced video transmission design for wireless multicast systems over fading channels," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2566–2577, Dec. 2017.

[199] C. Li, H. Jiang, P. Wilford, and Y. Zhang, "Video coding using compressive sensing for wireless communications," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2011, pp. 2077–2082.

[200] M. B. Schenkel, C. Luo, P. Frossard, and F. Wu, "Compressed sensing based video multicast," *Proc. SPIE*, vol. 7744, May 2010, Art. no. 77441H, doi: 10.1117/12.863531.

[201] W. Yin, X. Fan, Y. Shi, R. Xiong, and D. Zhao, "Compressive sensing based soft video broadcast using spatial and temporal sparsity," *Mobile Netw. Appl.*, vol. 21, no. 6, pp. 1002–1012, Dec. 2016.

[202] T.-Y. Tung and D. Gündüz, "SparseCast: Hybrid digital-analog wireless image transmission exploiting frequency-domain sparsity," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2451–2454, Dec. 2018.

[203] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer, "A survey of quantization methods for efficient neural network inference," 2021, *arXiv:2103.13630*.

[204] T. J. O'Shea, T. Roy, and N. E. West, "Approximating the void: Learning stochastic channel models from observation with variational generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Netw. Commun. (ICNC)*, Aug. 2019, pp. 681–686.

[205] H. Xiao, W. Tian, W. Liu, and J. Shen, "ChannelGAN: Deep learning-based channel modeling and generating," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 650–654, Mar. 2022.

[206] F. Euchner, J. Sanzi, M. Henninger, and S. Ten Brink, "GAN-based massive MIMO channel model trained on measured data," in *Proc. 27th Int. Workshop Smart Antennas (WSA)*, vol. 27, Mar. 2024, pp. 109–116.

[207] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[208] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6626–6637.

[209] E. Agustsson, D. Minnen, G. Toderici, and F. Mentzer, "Multi-realism image compression with a conditional generator," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22324–22333.

[210] (2018). *Better Portable Graphics 0.9.8*. [Online]. Available: https://bellard.org/bpg/

[211] 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.

[212] H. Wu, Y. Shao, K. Mikolajczyk, and D. Gunduz, "Channel-adaptive wireless image transmission with OFDM," *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2400–2404, Nov. 2022.

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

[214] C. Bian, Y. Shao, and D. Gündüz, "DeepJSCC-1++: Robust and Bandwidth-Adaptive Wireless Image Transmission," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Kuala Lumpur, Malaysia, 2023, pp. 3148–3154.

[215] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[216] G. Zhang, Q. Hu, Y. Cai, and G. Yu, "Alleviating distortion accumulation in multi-hop semantic communication," *IEEE Commun. Lett.*, vol. 28, no. 2, pp. 308–312, Feb. 2024.

[217] C. Bian, Y. Shao, and D. Gündüz, "A hybrid joint source-channel coding scheme for mobile multi-hop networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2024, pp. 986–992.

[218] C. Bian, Y. Shao, and D. Gündüz, "A deep joint source-channel coding scheme for hybrid mobile multi-hop networks," 2024, *arXiv:2405.09698*.

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

[220] H. Wu, Y. Shao, E. Ozfatura, K. Mikolajczyk, and D. Gündüz, "Transformer-aided wireless image transmission with channel feedback," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9,

pp. 11904–11919, Sep. 2024.

[221] H. Wu, Y. Shao, C. Bian, K. Mikolajczyk, and D. Gündüz, "Deep joint source-channel coding for adaptive image transmission over MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 15002–15017, Oct. 2024.

[222] C. Bian, Y. Shao, H. Wu, and D. Gunduz, "Space-time design for deep joint source channel coding of images over MIMO channels," 2022, *arXiv:2210.16985*.

[223] C. Bian, Y. Shao, H. Wu, and D. Gündüz, "Deep joint source-channel coding over the relay channel," in *Proc. IEEE Int. Conf. Mach. Learn. Commun. Netw. (ICMLCN)*, May 2024, pp. 1–6.

[224] Y. Yin, L. Huang, Q. Li, B. Tang, A. Pandharipande, and X. Ge, "Semantic communication on multi-hop concatenated relay networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2023, pp. 1–6.

[225] S. F. Yilmaz, E. Ozyilkan, D. Gündüz, and E. Erkip, "Distributed deep joint source-channel coding with decoder-only side information," in *Proc. IEEE Int. Conf. Mach. Learn. Commun. Netw. (ICMLCN)*, May 2024, pp. 139–144.

[226] N. Mital, E. Özyilkan, A. Garjani, and D. Gündüz, "Neural distributed image compression with cross-attention feature alignment," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 2497–2506.

[227] K. Choi, K. Tatwawadi, A. Grover, T. Weissman, and S. Ermon, "Neural joint source-channel coding," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2019, pp. 1182–1192.

[228] Y. Bo, Y. Duan, S. Shao, and M. Tao, "Joint coding-modulation for digital semantic communications via variational autoencoder," *IEEE Trans. Commun.*, vol. 72, no. 9, pp. 5626–5640, Sep. 2024.

[229] E. Erdemir, T.-Y. Tung, P. L. Dragotti, and D. Gündüz, "Generative joint source-channel coding for semantic image transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2645–2657, Aug. 2023.

[230] J. Chen, D. You, D. Gündüz, and P. L. Dragotti, "CommIN: Semantic image communications as an inverse problem with INN-guided diffusion models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 6675–6679.

[231] S. F. Yilmaz, X. Niu, B. Bai, W. Han, L. Deng, and D. Gündüz, "High perceptual quality wireless image delivery with denoising diffusion models," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, May 2024, pp. 1–5.

[232] T. Wu et al., "CDDM: Channel denoising diffusion models for wireless semantic communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11168–11183, Sep. 2024.

[233] M. Yang and H.-S. Kim, "Deep joint source-channel coding for wireless image transmission with adaptive rate control," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 5193–5197.

[234] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, and V. C. M. Leung, "Predictive and adaptive deep coding for wireless image transmission in semantic communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5486–5501, Aug. 2023.

[235] J. Dai et al., "Nonlinear transform source-channel coding for semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2300–2316, Aug. 2022.

[236] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2018, pp. 10771–10780.

[237] S. Wang, J. Dai, X. Qin, Z. Si, K. Niu, and P. Zhang, "Improved nonlinear transform source-channel coding to catalyze semantic communications," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 1022–1037, Sep. 2023.

[238] J. Dai et al., "Toward adaptive semantic

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Gündüz et al.: Joint Source–Channel Coding: Fundamentals and Recent Progress in Practical Designs

communications: Efficient data transmission via online learned nonlinear transform source-channel coding," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2609–2627, Jun. 2023.

[239] B. Tang, L. Huang, Q. Li, A. Pandharipande, and X. Ge, "Cooperative semantic communication with on-demand semantic forwarding," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 349–363, 2024.

[240] C. Bian, Y. Shao, H. Wu, E. Ozfatura, and D. Gunduz, "Process-and-forward: Deep joint source-channel coding over cooperative relay networks," 2024, *arXiv:2403.10613*.

[241] T. Wu, Z. Chen, M. Tao, B. Xia, and W. Zhang, "Fusion-based multi-user semantic communications for wireless image transmission over degraded broadcast channels," 2023, *arXiv:2305.09165*.

[242] S. F. Yilmaz, C. Karamanlı, and D. Gündüz, "Distributed deep joint source-channel coding over a multiple access channel," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 1400–1405.

[243] T. Tung and D. Gündüz, "DeepWiVe: Deep-learning-aided wireless video transmission," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2570–2583, Sep. 2022.

[244] S. Wang et al., "Wireless deep video semantic transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 214–229, Jan. 2023.

[245] M. Peters et al., "Deep contextualized word representations," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, vol. 1, 2018, pp. 1–15.

[246] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, 2001, p. 311.

[247] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1394–1398, Jul. 2022.

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

[249] Y. Liao, H. Yao, Y. Hua, and C. Li, "CSI feedback based on deep learning for massive MIMO systems," *IEEE Access*, vol. 7, pp. 86810–86820, 2019.

[250] 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.

[251] M. B. Mashhadi, Q. Yang, and D. Gündüz, "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.

[252] J. Xu, B. Ai, N. Wang, and W. Chen, "Deep joint source-channel coding for CSI feedback: An end-to-end approach," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 260–273, Jan. 2023.

[253] Y. Shao, C. Bian, L. Yang, Q. Yang, Z. Zhang, and D. Gunduz, "Point cloud in the air," 2024, *arXiv:2401.00658*.

[254] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "HoloCast: Graph signal processing for graceful point cloud delivery," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.

[255] C. Bian, Y. Shao, and D. Gündüz, "Wireless point cloud transmission," in *Proc. IEEE 25th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Lucca, Italy, 2024, pp. 851–855.

[256] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "HoloCast+: Hybrid digital-analog transmission for graceful point cloud delivery with graph Fourier transform," *IEEE Trans. Multimedia*, vol. 24, pp. 2179–2191, 2022.

[257] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "AirNet: Neural network transmission over the air," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 12126–12139, Sep. 2024.

[258] K. Anjum, Z. Qi, and D. Pompili, "Deep joint source-channel coding for underwater image transmission," in *Proc. 16th Int. Conf. Underwater Netw. Syst.*, New York, NY, USA, Nov. 2022, pp. 1–8.

[259] X. Zheng et al., "Error-resilient analog image storage and compression with analog-valued RRAM arrays: An adaptive joint source-channel coding approach," in *IEDM Tech. Dig.*, Dec. 2018, pp. 3.5.1–3.5.4.

[260] R. V. Zarcone et al., "Analog coding in emerging memory systems," *Sci. Rep.*, vol. 10, no. 1, p. 6831, Apr. 2020.

[261] W. Wu, L. Xiang, Q. Liu, and K. Yang, "Deep joint source-channel coding for DNA image storage: A novel approach with enhanced error resilience and biological constraint optimization," 2023, *arXiv:2311.01122*.

[262] T. Marchioro, N. Laurenti, and D. Gündüz, "Adversarial networks for secure wireless communications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8748–8752.

[263] T.-Y. Tung and D. Gündüz, "Deep joint source-channel and encryption coding: Secure semantic communications," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 5620–5625.

[264] S. A. Ameli Kalkhoran, M. Letafati, E. Erdemir, B. H. Khalaj, H. Behroozi, and D. Gündüz, "Secure deep-JSCC against multiple eavesdroppers," in *Proc. IEEE Global Commun. Conf.*, vol. 16, Dec. 2023, pp. 3433–3438.

[265] F. Chen, L. Xiang, H. V. Cheng, and K. Shen, "Deep learning enabled semantic-secure communication with shuffling," in *Proc. IEEE Global Commun. Conf.*, Dec. 2023, pp. 6338–6343.

[266] G. Nan et al., "Physical-layer adversarial robustness for deep learning-based semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2592–2608, Apr. 2023.

ABOUT THE AUTHORS

Deniz Gündüz (Fellow, IEEE) received the B.S. degree in electrical and electronics engineering from METU, Ankara, Turkey, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from the NYU Tandon School of Engineering, Brooklyn, NY, USA, in 2004 and 2007, respectively.

He was a Research Assistant Professor with Stanford University, Palo Alto, CA, USA, from 2007 to 2009. He was a Postdoctoral Researcher from 2007 to 2009 and a Visiting Researcher from 2009 to 2011 with Princeton University, Princeton, NJ, USA. He was a Research Associate with the Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain, from 2009 to 2012. In 2012, he joined the Electrical and Electronic Engineering Department, Imperial College London, London, U.K., where he is currently a Professor of information processing and serves as the Deputy Head of the Intelligent Systems and Networks Group. He was a Visiting Professor with the University of Padova, Padua, Italy, in 2018 and 2020. He was a Part-Time Faculty Member with the University of Modena and Reggio Emilia, Modena, Italy, from 2019 to 2022. He has co-authored several award-winning articles, including the IEEE Communications Society—Young Author Best Paper Award in 2022 and the IEEE International Conference on Communications Best Paper Award in 2023. His research interests lie in the areas of communications, information theory, machine learning, and privacy.

Dr. Gündüz is an Elected Member of the IEEE Signal Processing Society Signal Processing for Communications and Networking (SPCOM) and Machine Learning for Signal Processing (MLSP) Technical Committees, and has been serving as the Chair for the IEEE Information Theory Society U.K. and Ireland Chapter since 2022. He serves as an Area Editor for IEEE TRANSACTIONS ON INFORMATION THEORY and IEEE TRANSACTIONS ON COMMUNICATIONS. He was a recipient of the IEEE Communications Society—Communication Theory Technical Committee (CTTC) Early Achievement Award in 2017, and Starting (2016), Consolidator (2022), and Proof-of-Concept (2023) Grants of the European Research Council (ERC). He received the Imperial College London—President's Award for Excellence in Research Supervision in 2023.

Michèle A. Wigger (Senior Member, IEEE) received the M.Sc. (honors) and Ph.D. degrees in electrical engineering from ETH Zürich, Zürich, Switzerland, in 2003 and 2008, respectively.

In 2009, she was first a Postdoctoral Fellow with the University of California at San Diego, San Diego, CA, USA, and then joined Télécom Paris, Palaiseau, France, where she is currently a Full Professor. She has held visiting professor appointments at the Technion–Israel Institute of Technology, Haifa, Israel, the University of Zurich, Zürich, and ETH Zürich. Her research interests are in multiterminal information theory, in particular in distributed source coding and capacities of networks, and in the limits of distributed hypothesis testing.

Dr. Wigger has previously served as an Associate Editor for IEEE COMMUNICATION LETTERS and is currently serving as an Associate Editor for Shannon Theory for IEEE TRANSACTIONS ON INFORMATION THEORY. From 2016 to 2019, she also served on the Board of Governors of the IEEE Information Theory Society.

Tze-Yang Tung (Member, IEEE) received the B.Eng. degree from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2017, the M.Sc. degree in electrical and computer engineering from the University of Southern California, Los Angeles, CA, USA, in 2019, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, Imperial College London, in 2022.

He is currently the Co-Founder and the CTO with Not Diamond, San Francisco, CA, USA, a start-up that helps developers dynamically determine the best artificial intelligence (AI) model to use for each query. His current work is in developing the interfaces and frameworks to connect machine learning agents in distributed systems. In particular, he is interested in how networks of agents can facilitate not only more powerful but also safer AI.

Ping Zhang (Fellow, IEEE) received the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1990.

He is currently a Professor with the School of Information and Communication Engineering, BUPT, and also the Director of the State Key Laboratory of Networking and Switching Technology. He is also with the Department of Broadband Communication, Pengcheng Laboratory, Shenzhen, China. He is an Academician with the Chinese Academy of Engineering (CAE), Beijing. His research is in the board area of wireless communications.

Dr. Zhang is a member of IMT-2020 (5G) Experts Panel and Experts Panel for China's 6G development. He served as the Chief Scientist for the National Basic Research Program (973 Program), an Expert for the Information Technology Division of the National High-Tech Research and Development Program (863 Program), and a member for the Consultant Committee on International Cooperation of the National Natural Science Foundation of China.

Yong Xiao (Senior Member, IEEE) received the B.S. degree in electrical engineering from China University of Geosciences, Wuhan, China, in 2002, the M.Sc. degree in telecommunication from Hong Kong University of Science and Technology, Hong Kong, in 2006, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2012.
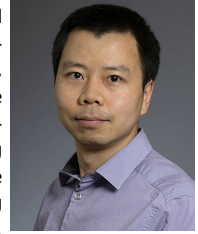
He was a Research Assistant Professor with the Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ, USA, where he was also the Center Manager of the Broadband Wireless Access and Applications Center (BWAC), an NSF Industry/University Cooperative Research Center (I/UCRC) led by The University of Arizona. He is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology (HUST), Wuhan. He is also with the Pengcheng Laboratory, Shenzhen, China, and the Pazhou Laboratory (Huangpu), Guangzhou, China. He is an Associate Group Leader of the Network Intelligence Group of IMT-2030 (6G Promoting Group) and the Vice-Director of the 5G Verticals Innovation Laboratory, HUST. His research interests include machine learning, game theory, distributed optimization and their applications in semantic communications, semantic-aware networking, network artificial intelligence (AI), cloud/fog/mobile edge computing, and green communication networking.