Soft Delivery: Survey on A New Paradigm for Wireless and Mobile Multimedia Streaming

Fujihashi, Takuya; Koike-Akino, Toshiaki; Watanabe, Takashi

TR2023-102   August 18, 2023

Abstract

The increasing demand for video streaming services is the key driver of modern wireless and mobile commu- nications. Although many studies have designed digital-based delivery schemes to send video contents over wireless and mobile networks, significant quality degradation, known as cliff and leveling effects, often occurs owing to fluctuating channel characteristics. In this paper, we present a comprehensive summary of soft delivery, which is a new paradigm for wireless and mobile video streaming and discuss the future directions of soft delivery. Existing studies found that introducing multi-dimensional cosine transform, human vision system, and graph signal processing can make soft delivery schemes more effective in untethered immersive experiences, including virtual reality and volumetric media, than digital-based delivery schemes. In addition, this study finds that soft delivery has the potential to be a new standard to deliver deep neural network models and tactile information over wireless and mobile networks.

ACM Computing Surveys 2023

Soft Delivery: Survey on A New Paradigm for Wireless and Mobile Multimedia Streaming

TAKUYA FUJIHASHI, Osaka University, Japan

TOSHIAKI KOIKE-AKINO, Mitsubishi Electric Research Laboratories (MERL), USA

TAKASHI WATANABE, Osaka University, Japan

The increasing demand for video streaming services is the key driver of modern wireless and mobile communications. Although many studies have designed digital-based delivery schemes to send video contents over wireless and mobile networks, significant quality degradation, known as cliff and leveling effects, often occurs owing to fluctuating channel characteristics. In this paper, we present a comprehensive summary of soft delivery, which is a new paradigm for wireless and mobile video streaming and discuss the future directions of soft delivery. Existing studies found that introducing multi-dimensional cosine transform, human vision system, and graph signal processing can make soft delivery schemes more effective in untethered immersive experiences, including virtual reality and volumetric media, than digital-based delivery schemes. In addition, this study finds that soft delivery has the potential to be a new standard to deliver deep neural network models and tactile information over wireless and mobile networks.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Information systems** → **Multimedia streaming**; • **Networks** → *Mobile networks*.

Additional Key Words and Phrases: soft delivery, hybrid digital-analog delivery, extended reality

1 INTRODUCTION

Video streaming over wireless and mobile networks is one of the major applications in wireless environments. According to Cisco Visual Networking Index studies, approximately four-fifths (82%) of the world's mobile data traffic will be video content by 2022 [1]. The explosive growth of data traffic, especially video traffic, poses a huge challenge to wireless and mobile networks. In recent years, immersive content, such as virtual reality (VR), augmented reality (AR), and mixed reality (MR), have shown very good potential to be the next important applications for networks. The growth of such immersive applications is rapidly increasing together with the development of fifth generation (5G) technology and smart wearable devices, which can enable technology for all extended reality (XR) applications.

In general, wireless video streaming systems transmit images and video signals with different channel characteristics to single or multiple users. For high-quality video streaming applications,

Authors' addresses: Takuya Fujihashi, Osaka University, Graduate School of Information Science and Technology, Suita, Osaka, 565-0871, Japan, tfuji@ist.osaka-u.ac.jp; Toshiaki Koike-Akino, Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, 02139, USA, koike@merl.com; Takashi Watanabe, Osaka University, Graduate School of Information Science and Technology, Suita, Osaka, 565-0871, Japan, watanabe@ist.osaka-u.ac.jp.

the main challenge is the difficulty in fully utilizing each user's channel capacity and providing each user with the best video quality possible under his/her channel conditions. Solving this challenge will provide users with an improved quality of experience (QoE). To address this challenge, conventional streaming systems, which consist of video coding and transmission technologies, have been proposed based on digital-based solutions. For video coding, the H.265/High-Efficiency Video Coding (HEVC) standard [2], which has been standardized by the Joint Collaborative Team on Video Coding (JCT-VC), can be used to encode VR/360-degree videos. As the successor of HEVC, the future video coding standard, H.266/Versatile Video Coding (VVC), has been developed by the Joint Video Experts Team (JVET) The VVC standard takes camera-view video, high dynamic range video, and VR/360-degree video into account. In addition, video- and geometry-based point cloud coding [3, 4] have been standardized by the Motion Picture Experts Group (MPEG) for volumetric video encoding and decoding. For video transmission, the source bits are channel-coded with time interleaving to exhibit robustness against a certain level of channel errors. The channel-coded bits are then mapped into the transmit data symbols corresponding to arbitrary modulation schemes, such as binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), or quadrature amplitude modulation (QAM). To choose an appropriate source and channel coding rate according to the user's channel condition, the channel statistics are generally required to be known at the time of source and channel coding. Once both the source and channel coding processes are completed, the conventional systems work optimally only for a specific channel condition and have performance limitations in noisy time-varying channels [5, 6].

If the observed channel quality, i.e., the channel signal-to-noise ratio (SNR), falls below a threshold, the decoding process tends to break down completely. This phenomenon is called the cliff effect [7]. In contrast, if the observed channel quality increases beyond the threshold, it does not improve the performance unless an adaptive rate control of the source and channel coding is performed in real-time according to the rapid fading channels. This phenomenon is known as the leveling effect.

Thus, accurate channel estimation and real-time rate control of the source and channel coding are desired for conventional streaming systems. However, the channel conditions of wireless and mobile networks may vary drastically and unpredictably, resulting in imperfect channel estimation and rate control. Conventional streaming systems tend to utilize the channel capacity conservatively to prevent cliff and leveling effects, considering the fact that rate control may be inaccurate. Scalable Video Coding (SVC) [8] and Dynamic Adaptive Streaming over HTTP (DASH) [9, 10] are typical standardized systems that utilize channel capacity without cliff and leveling effects. SVC encodes the video frames into multiple layers to progressively improve the video quality according to the number of received layers. DASH encodes video frames at multiple quality levels, and stores all encoded video frames on the server. The main difference between SVC and DASH is the sender-dependent and receiver-dependent schemes. SVC determines how many layers to send based on the estimated channel quality. DASH retrieves the appropriate quality of video frames from the server based on the estimated channel quality. However, the conservative strategy in SVC and DASH systems under imperfect channel estimation and rate control will cause quality degradation.

A new paradigm of wireless video delivery, namely, soft delivery [11–13], has been proposed to fully utilize the instantaneous channel capacity without cliff and leveling effects. In contrast to SVC and DASH systems, soft delivery does not require channel estimation to utilize the instantaneous channel capacity. It is essentially a scheme with "lossless compression and lossy transmission." The compression stage is solely a transform to decorrelate the image and video signals into frequency-domain coefficients, leaving out the conventional quantization and entropy coding. The transmission stage skips digital-based channel coding. Instead, it scales each transform coefficient individually and modulates it directly to a dense constellation for transmission inspired by the advantage of analog transmission with linear coding [5, 14, 15]. Here, the scaling operation serves the

Soft delivery

- **Energy compaction**
  - Decorrelation in time domain → 3D-DCT [11-13], MCTF [41-46], Effect of decorrelations [47,48], Adaptive GoP size [49]
  - Protect large energy chunks → ASoftCast [50], OPA-SoftCast [51,52]
  - Integration w/ digital coding → Digital encoding of BL layer [53-58], DCast [61-64], Pseudo/Layered coset coding [67-74]

- **Power allocation**
  - Channel-aware allocation → AWGN [11], Fading [75], OFDM [76, 77], impulse [84], MIMO [78-80], MIMO-OFDM [81-83], NOMA [85, 86], Underwater [87], UAV [88], mmWave [89]
  - Perception-aware allocation → SSIM-based [90], Foveation-based [91], Saliency-based [92], Foreground and background-based [93]

- **Bandwidth utilization**
  - Insufficient bandwidth → Discarding high frequency chunks [11,96], Shannon-Kotelnikov mapping [97], Compressive sensing [98-105], Data-assisted delivery [106-110]
  - Extra bandwidth → Analog channel coding [116,117], MCast [119]
  - Bandwidth variation → Progressive pseudo-analog transmission [120,121]

- **Packet loss resilience**
  - Compressive sensing → Block-wise compressive sensing [122], Multi-scale block-wise compressive sensing [123], adaptive block-wise compressive sensing [124]

- **Overhead reduction**
  - Sender-side reduction → Four-parameter fitting [126], log-linear fitting [127], L-shape chunk division [128], Lorentzian fitting [129]
  - Receiver-side reduction → Metadata estimation from received symbols [130], Blind data detection [131]

- **Hardware implementation**
  - Empirical evaluations → Software defined radio platform [11, 103, 117, 135], FPGA [136-138], MU-MIMO prototype [139], LTE prototype [140]

- **Extension for immersive experiences**
  - Free viewpoint video → FreeCast [151,152], Bandwidth and power allocation [153], 3DV SoftCast [154], Depth map transmission [155]
  - Virtual reality → Viewport-adaptive soft delivery [161], OmniCast [162], 360Cast [163], 360Cast+ [164]
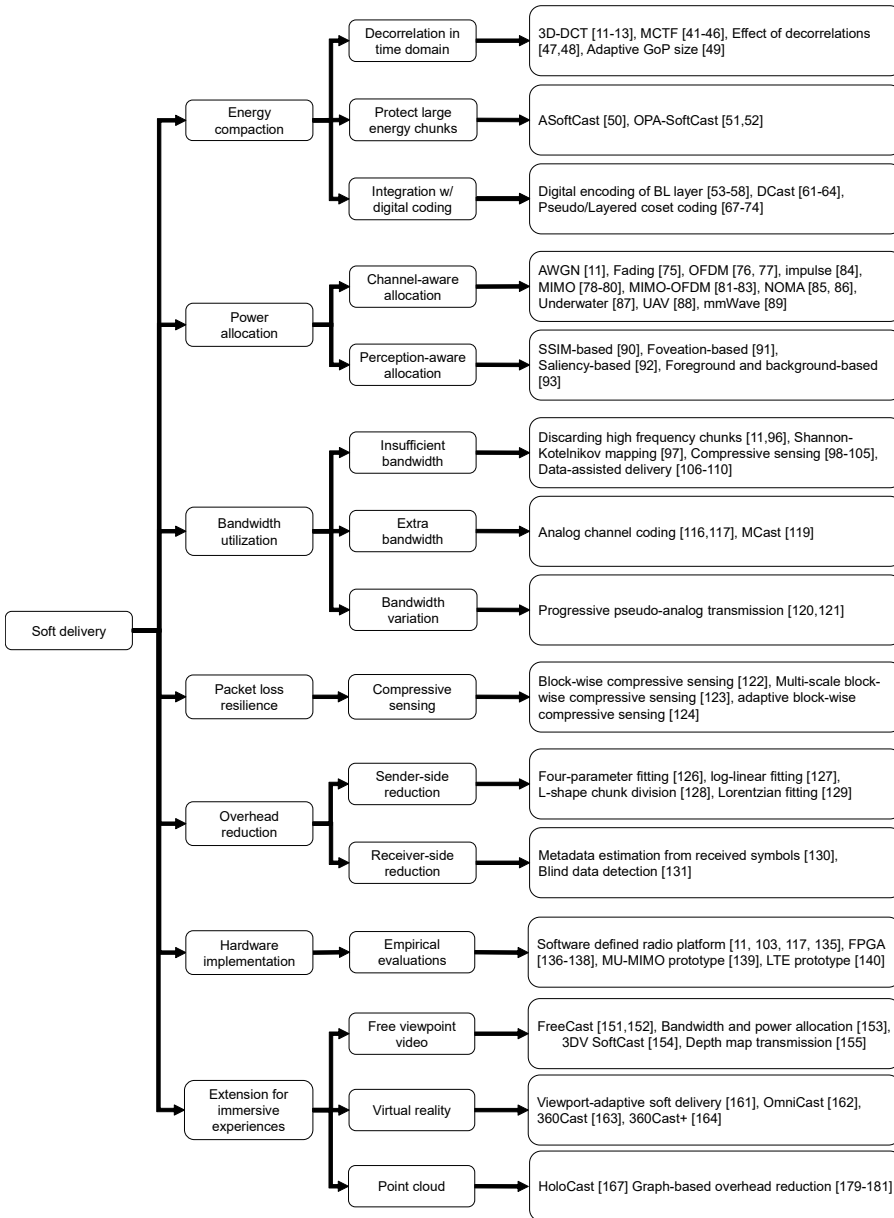 - Point cloud → HoloCast [167] Graph-based overhead reduction [179-181]

Fig. 1. Taxonomy of the studies on soft delivery schemes.

purposes of both power allocation and unequal signal protection against channel noises and fading effects to maximize the reconstruction quality. At the receiver end, the image and video signals are reconstructed by demodulating the received signals and inverting the scaling and transform operations. The soft delivery scheme was shown to not only provide a graceful performance transition in a wide channel SNR range but also achieve competitive performance compared with the conventional digital-based delivery schemes.

1.1 Contributions

This work provides a comprehensive survey of soft delivery schemes, including an overview on existing techniques, extension for immersive experiences, and future research directions. Some existing studies have focused on soft delivery schemes, which are shown in Fig. 1, with a brief description of the related topics and key techniques. Although there are some survey papers [9, 16–20] related to the video delivery over wireless channels, all of the survey papers deal with the digital-based approaches. To the best of our knowledge, this survey is the first to introduce methodologies and approaches for soft and hybrid digital-analog delivery to transmit high-quality image and video signals via unstable and diverse wireless and mobile channel environments. The main contributions of this study are summarized as follows:

- We present an overview of the conventional digital-based and soft delivery schemes, as well as the benefits of the soft delivery schemes.
- The existing soft delivery techniques, such as energy compaction, power allocation, bandwidth utilization, overhead reduction, and packet loss resilience, are surveyed. In this context, the abstraction and key contributions of these techniques are reviewed and summarized.
- We present the extensions of the soft delivery for immersive experiences, which are mainly contributed by our prior research. We summarize the key ideas of the extensions and benefits in contrast to the digital-based delivery schemes.
- In addition, we review the future research directions of the soft delivery: hybrid digital-analog delivery, AI-empowered soft delivery, soft delivery for AI, and tactile Internet. We carry out some evaluations to discuss the various directions.

This survey identified that soft delivery works particularly well in untethered immersive experiences including free viewpoint video, VR, and point cloud. Soft delivery yields better reconstruction quality compared with HEVC-based delivery schemes by integrating the energy compaction, power allocation, and overhead reduction techniques discussed in the existing studies. In addition, soft delivery can realize high-quality and adaptive delivery of deep neural network (DNN) models and tactile information over wireless channels. Such delivery will help to realize future services including federated learning and untethered XR applications.

1.2 Survey Structure

The remainder of this paper is organized as follows:

- Section 2 describes an overview of conventional digital-based delivery schemes and their issues, such as the cliff, leveling, and staircase effects.
- Section 3 presents the basic principles of the pioneering work on soft delivery to solve the aforementioned effects in wireless and mobile video streaming applications.
- Section 4 presents a review of the existing techniques on soft delivery. We classify these techniques into energy compaction, power allocation, bandwidth utilization, packet loss resilience, overhead reduction, extension for immersive experiences, and discuss their implementation, as well as their contributions.
- Section 5 suggests the future directions of the soft delivery approach. In addition to image and video signals, the soft delivery approach has a potential to realize high-quality DNN model and tactile delivery over wireless channels.
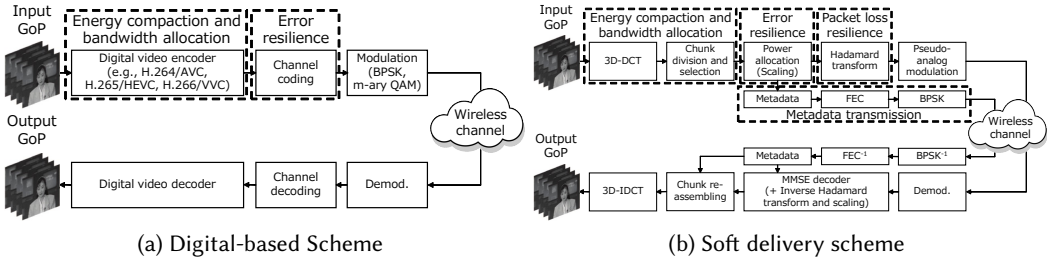- Section 6 concludes the paper.

Fig. 2. Framework of the conventional digital-based delivery and soft delivery schemes.

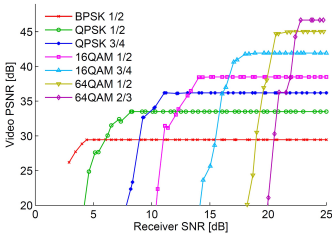## 2 CONVENTIONAL DIGITAL-BASED DELIVERY

2.1 Overview

One of the major issues in wireless video delivery is sending high-quality videos within the considerably limited capacity of wireless links. For this purpose, standardized digital video compression is carried out for video frames in conventional video delivery schemes [21–23], as shown in Fig. 2 (a) to remove redundancy among video frames. In particular, H.264/Advanced Video Coding (AVC) [24], H.265/HEVC, and H.266/VCC standards are typical video coding standards for generating a compressed bitstream from video frames. In such video encoders, pixel values in each video frame are divided into blocks and transformed into frequency domain coefficients by using the discrete cosine transform (DCT) or discrete sine transform (DST), and then non-uniformly quantizing the coefficients according to a quantization parameter (QP). A large QP indicates a larger quantization step, leading to a smaller bit rate. Finally, the quantized coefficients are compressed by an entropy coder, which removes statistical redundancy in the coefficients. Variable length coding (VLC) is widely deployed for entropy coding because of its efficiency and simplicity.

After passing through the digital video compression, the bitstream is then passed to the wireless transmission system in sequence. There are typically two ways to deliver the bitstream: Internet Protocol (IP) and non-IP networks. For the IP networks, the sender uses DASH [9, 10] or Realtime Transport Protocol (RTP) [25] for the bitstream. Although the IP-based protocols can deliver the content via the deployed IP networks, they cause transmission delays to packetize and de-packetize the bitstream [26]. For the non-IP networks, the sender passes the bitstream to the physical layer for transmissions. The existing studies related to the non-IP network schemes reported that such schemes reduce the delay compared with the IP-based schemes, and thus the non-IP schemes are applicable for low-delay applications including untethered XR experiences and cooperative and competitive gaming.

This survey mainly discusses the content delivery over non-IP networks. In such non-IP schemes, channel coding is first used for the bitstream to protect against channel errors. For example, binary convolutional codes and low-density parity-checks are widely used as forward error correction in Wi-Fi systems. The coded bitstream is then mapped onto in-phase and quadrature (I and Q) components using digital modulation formats such, as QPSK and m-ary QAM. In both wireless and mobile networks, a combination of modulation formats and different channel coding rates, for example, 1/2 and 3/4, is defined in the modulation and coding scheme (MCS). According to the measured wireless channel SNRs, the sender adapts its MCS value to maximize the link data rate. At the receiver end, bit errors may occur in channel-coded bits owing to effective noise and/or fading effects. The receiver then tries to reconstruct video frames from the received bits using inverse procedures, i.e., demodulation, channel decoding, and video decoding.

Table 1. Critical issues regarding video quality in wireless and mobile video streaming

Phenomenon	Effect on video quality	Cause	Solution in soft delivery
Cliff effect	Sudden degradation	All-or-nothing behavior in entropy and channel codings	Skip entropy and channel codings to prevent all-or-nothing behavior
Leveling effect	Constant irrespective of channel quality	Unrecoverable quantization error in video coding	Skip quantization and adopt pseudo-analog modulation for recoverable errors at the receiver end
Staircase effect	Step function of channel quality	All-or-nothing behavior and quantization error in layered coding	Skip hierarchical operations and adopt pseudo-analog modulation for linear video quality

(a) Cliff and leveling effects in digital-based schemes
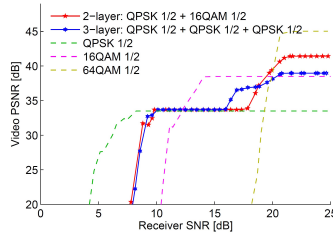
(b) Staircase effect in layered coding with hierarchical modulation schemes
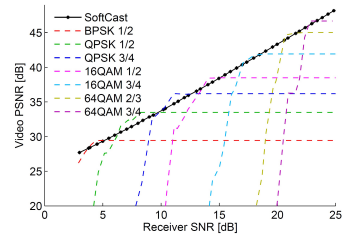
(c) Soft delivery scheme

Fig. 3. Video quality of the conventional digital-based schemes and SoftCast scheme via wireless networks [11]: (a) Cliff and leveling effects in H.264/AVC over 802.11 under different MCSs. (b) Staircase effect in two-layered video coding and three-layered video coding shown in red and blue, respectively. For reference, the dashed lines are the three-equivalent single-layer H.264/AVC videos. (c) Performance of SoftCast versus single-layer H.264/AVC.

2.2 Critical Issues on Quality

If the measured wireless channel quality is stable during video transmission, conventional digital-based schemes can provide high-quality video frames for users. However, the channel quality of each user fluctuates over time owing to a combination of user mobility, multipath propagation, and obstacle shadow. Table 1 lists three critical issues regarding the video quality of the digital-based schemes because of the channel quality fluctuations: cliff, leveling, and staircase effects.

2.2.1 Cliff Effect. Digitally encoded bits are known to be susceptible to errors during wireless transmission. Because entropy coding schemes have an all-or-nothing behavior, even a single bit error can cause the loss of entire data [27]. As mentioned earlier, channel coding schemes are adopted to correct burst and random bit errors. However, they generally exhibit an all-or-nothing behavior for error correction. When the instantaneous channel quality, i.e., the SNR, falls below a certain threshold, possible errors that occur in the bitstream during wireless communications will disable video decoding.

A collapsed signal reconstruction causes a cliff effect. The cliff effect is a phenomenon whereby the quality of the received information abruptly drops as soon as the channel quality falls below the threshold, as shown in Fig. 3 (a). For example, the video quality of the BPSK modulation format with 1/2-rate channel coding drops below the wireless channel SNR of 4 dB. In modern network environments (e.g., content delivery, mobile, and wireless networks), the cliff effect becomes a major impediment when video frames are transmitted over diverse channel conditions to heterogeneous

users. In this case, users whose channel quality is below the critical point receive unwatchable video frames.

Some solutions have addressed the cliff effect associated exclusively with channel coding, such as hybrid automatic repeat request, and rateless coding schemes [28–32]. They adapt the number of transmissions to changing channel conditions for error prevention. However, these schemes are not well suited for streaming multiple users under diverse channel conditions. In addition, they do not reduce the quantization error at the video encoder and thus, the leveling effect still occurs in video quality.

2.2.2 Leveling Effect. Once the channel quality surpasses the threshold, the video quality remains constant as shown in Fig. 3 (a). As mentioned earlier, the cliff effect is caused when the receiver SNR is below 4 dB in the BPSK modulation format with 1/2-rate channel coding, whereas the channel gain does not reflect on the video quality above the wireless channel SNR of 4 dB. Digital-based schemes determine the parameters of the video coding and wireless transmission parts based on the channel estimation. If the instantaneous channel quality is better than the estimated one, no additional gain can be obtained because the distortion of the video coding cannot be reconstructed for each user.

2.2.3 Staircase Effect. To mitigate the cliff and leveling effects, some layered coding schemes, referred to as schemes with SVC [8] with a combination of hierarchical modulation (HM) [33], have been proposed for wireless and mobile video streaming [34, 35]. These layered coding schemes encode video frames into one base layer (BL) and several enhancement layers (ELs). The BL is used to ensure that all the users in the target channel SNR range can receive the baseline quality of video frames, whereas the ELs are used to enhance the video quality of the users in high-channel SNRs. Each SVC layer is then mapped onto the corresponding HM layer. Notably, HM provides unequal error protection to the transmitted video frames according to their relative importance. However, SVC with HM cannot completely remove the cliff effect; it only divides one big cliff into multiple stairs according to the number of layers, as shown in Fig. 3 (b). In addition, because the assigned transmission power to each layer is lower than that of the single-layer coding schemes, the cliff shifts to higher wireless channel SNRs.

3 SOFTCAST: A PIONEER WORK ON SOFT DELIVERY

3.1 Overview

To prevent the cliff, leveling, and staircase effects in wireless video delivery, a pioneer soft delivery work, namely, SoftCast was proposed in [11–13]. The block diagram of SoftCast is illustrated in Fig. 2 (b). The design of SoftCast is based on a simple principle that ensures that the transmitted signal samples are linearly related to the original pixel values. This principle naturally enables a sender to satisfy multiple receivers with diverse channel qualities, as well as a single receiver, where different packets experience different channel qualities.

The sender first takes a group of pictures (GoP) and uses a full-frame 3D-DCT as the decorrelation transform. The DCT frames are then divided into N small rectangular blocks of transformed coefficients called chunks. The coefficients in each chunk are then scaled to match the transmission power constraints. Specifically, the scaling coefficients are chosen to minimize the reconstruction mean square error (MSE). A Walsh–Hadamard transform (WHT) is then applied to the scaled chunks for power normalization across the chunks to provide packet loss resilience. This process transforms the chunks into slices. Each slice is a linear combination of all scaled chunks. Finally, the coefficients in the slices are directly mapped to the I and Q components in a pseudo-analog manner for transmission. Here, channel coding operations are skipped for the coefficients.

(a) 16-QAM [36]

(b) SoftCast [11]

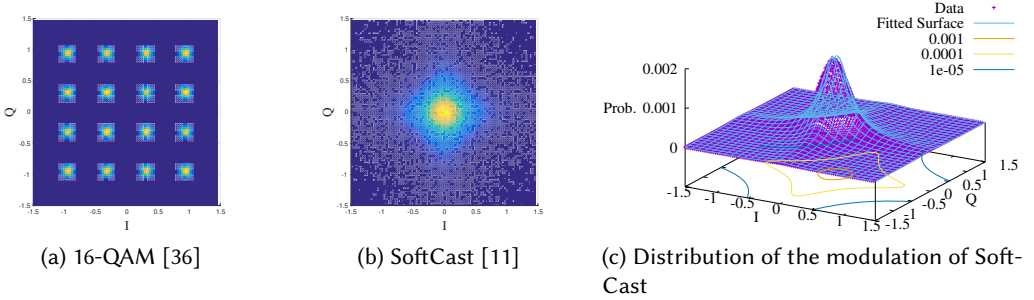(c) Distribution of the modulation of Soft-Cast

Fig. 4. Mapping coded video to I/Q components of the transmitted signal. (a) Traditional 16-QAM maps a bit sequence to the complex number corresponding to the point labeled with that sequence. (b) By contrast, PHY of SoftCast treats pairs of coded values as the real and imaginary parts of a complex number. (c) We find the modulation of SoftCast follows a Lorentzian distribution.

Figs. 4 (a) and (b) show the conventional digital-based modulation, i.e., 16-QAM, and pseudo-analog modulation proposed in SoftCast. Conventional modulation modulates channel-coded bits to produce real-value digital samples that are transmitted to the channel. For example, 16-QAM modulation takes sequences of four bits and maps each sequence to a complex I/Q number, as shown in Fig. 4 (a). After modulation, the wireless physical layer (PHY) of the sender transmits the mapped complex numbers to the receiver. Because of the broadcast nature of the wireless medium, multiple receivers hear the transmitted samples but with different noise levels. A receiver with a low channel SNR can distinguish only the quadrant of the transmitted sample, and hence, can decode only the two bits of the transmitted sample. In this case, these bit errors may cause a collapsed signal reconstruction during digital video decoding.

In contrast to the existing modulation design, SoftCast outputs the real values of the DCT coefficients that are already coded for error protection. The pseudo-analog modulation in Fig. 4 (b) directly maps pairs of the scaled DCT coefficients to the I and Q of the digital signal samples. Fig. 4 (c) shows the distribution of the analog-modulated symbols of the test video sequence of "Akiyo" with the resolution of common intermediate format (CIF) [37]. As shown in Fig. 4 (c), we find the pseudo-analog modulation of the DCT coefficients follows a bivariate Lorentzian distribution as follows:

$$f(x, y) = a \frac{1}{\pi^2 \left(b^2 + y^2 + x^2 + \frac{x^2 y^2}{b^2} \right)}, \tag{1}$$

where a and b represent the fitting parameters. From least-squares fitting based on the distribution of pseudo-analog symbols of the test video sequence and the bivariate Lorentzian function, the best fitting parameters are a of 0.001 and b of 0.24. As mentioned earlier, multiple receivers hear the transmitted samples under different channel SNRs. Although the transmitted samples are distorted according to their SNR, the receiver regards the received samples as scaled DCT coefficients. The sender does not need to estimate the channel condition, and the noise level in the received samples faithfully reflects the instantaneous channel condition [38], and thus pseudo-analog modulation ensures that the received video quality is proportional to the instantaneous channel quality. Consequently, this process avoids all cliff, leveling, and staircase effects.

In parallel, SoftCast sends an amount of data, referred to as metadata, for signal reconstruction. These metadata consist of the mean and variance of each transmitted chunk as well as a bitmap. The mean of each chunk is used to obtain the chunks' approximate zero-mean distributions by
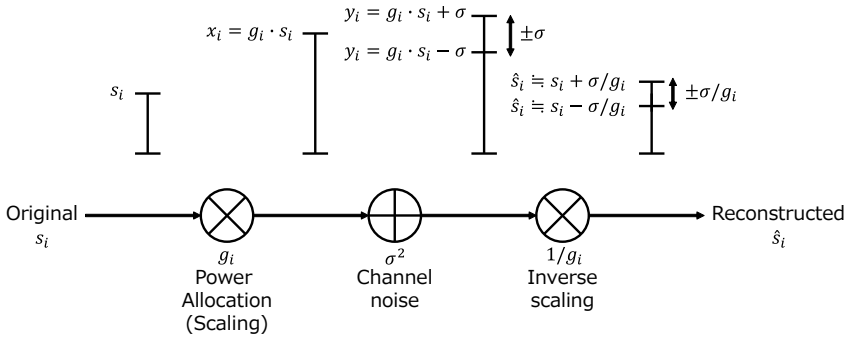
Fig. 5. Scaling and inverse scaling operations in soft delivery schemes.

subtracting the mean of all pixels in each chunk [39]. The variance of each chunk is used to find the per-chunk scaling factors that minimize the reconstruction error. The bitmap indicates the positions of the discarded chunks into the GoP. When the available channel bandwidth for SoftCast is less than the required bandwidth, SoftCast discards chunks with less energy. Specifically, when the available and required bandwidths for SoftCast are M chunks and $N(> M)$ chunks, respectively, SoftCast discards lower-energy $M - N$ chunks to meet the bandwidth requirement. On the receiver side, these discarded chunks are replaced by null values. The discarded chunks are registered as a bitmap and then compressed using run-length encoding. Metadata are strongly protected and transmitted in a robust way (e.g., BPSK modulation format with a low-rate channel code) to ensure correct delivery and decoding.

At the receiver side, a minimum MSE (MMSE) decoder is used to estimate the content of the chunks to counteract channel noise. The MMSE provides a high-quality estimate of the DCT coefficients by leveraging the knowledge of the statistics of the DCT coefficients, i.e., chunk variance, as well as the statistics of the channel noise. Using the metadata, the denoised chunks are properly reassembled and undergo an inverse 3D-DCT, thereby providing the corresponding GoP. In principle, the above design and performance do not affect the content types, i.e., on-demand or live content. Regardless of the content type, SoftCast provides adaptive video delivery based on the channel quality between the sender and each receiver.

Table 2 shows the strengths, limitations, and use cases of the digital-based and soft delivery schemes. The digital-based delivery schemes are well suited for point-to-point communications over time-invariant channels. In addition, the buffering cost is relatively low compared to the soft delivery schemes because the required storage size for the compressed bitstream is small. Soft delivery schemes perform well in time-varying and diverse channels. In addition, soft delivery schemes are low delay because they do not need to perform an expensive motion search for compression, making soft delivery schemes preferable for delay-sensitive applications. On the other hand, they require a high cost for modulation and demodulation in the PHY layer because modifying the PHY layer in both sender and receiver is required for the pseudo-analog modulation.

In summary, the typical use cases of digital-based delivery schemes are streaming services over wide area networks and HTTP adaptive streaming. In contrast, soft delivery schemes are well-suited for delay-sensitive applications such as video broadcasting, vehicle-to-everything (V2X) communication, and real-time video surveillance.

3.2 Details of Scaling and Inverse Scaling Operations

Table 2. Strengths, limitations, and use cases for digital-based and soft delivery schemes

Schemes	Channel types	Communication types	Latency	Buffering cost	Modem cost	Use cases
Digital-based delivery	Time-invariant	Point-to-point	High	Low	Low	Video streaming over wide-area networks, HTTP adaptive streaming
Soft delivery	Time-variant	Multicast Broadcast	Low	High	High	Broadcasting, V2X communication, Video surveillance

In SoftCast, a full-frame 3D-DCT is carried out for the video frames in each GoP to compact the energy of the video signals, and the resulting DCT coefficients are transmitted to the receivers using pseudo-analog modulation. Here, the transmitted analog-modulated symbols are degraded over wireless channels at each receiver. SoftCast should minimize the MSE between the received and transmitted DCT coefficients under the wireless channel SNR to reconstruct the highest-quality video frames at each receiver. For this purpose, SoftCast must design MSE-minimized power allocation and denoising filters, i.e., scaling and inverse scaling operations, for analog-modulated symbols. Figure 5 illustrates the procedures for obtaining the reconstructed DCT coefficients at the receiver end. SoftCast implements chunk-wise power allocation and filter operations according to the statistics of the chunks and channel conditions. Let x_i denote the ith analog-modulated symbol. Each analog-modulated symbol is scaled by g_i for noise reduction:

$$x_i = g_i \cdot s_i. \tag{2}$$

Here, s_i is the ith DCT coefficient, and g_i is the scale factor for the coefficient power allocation. The sender performs optimal power control for g_i to achieve the highest video quality. Specifically, the best g_i is obtained by minimizing the MSE under the power constraint with the total power budget P as follows:

$$\min \quad \text{MSE} = \mathbb{E}\left[(x_i - \hat{x}_i)^2\right] = \sum_i^N \frac{\sigma^2 \lambda_i}{g_i^2 \lambda_i + \sigma^2} \tag{3}$$

$$\text{s.t.} \quad \frac{1}{N} \sum_i^N g_i^2 \lambda_i = P,$$

where $\mathbb{E}[\cdot]$ denotes the expectation, $\hat{x}_i$ is an estimate of the transmitted symbol, λ_i is the power of the ith DCT coefficient, N is the number of DCT coefficients, and σ^2 is the receiver noise variance. The near-optimal solution is expressed as:

$$g_i = \lambda_i^{-1/4} \sqrt{\frac{P}{\sum_j \lambda_j}}. \tag{4}$$

After transmission over the wireless channel, each symbol at the receiver end can be modeled as $y_i = x_i + n_i$, where y_i is the ith received symbol and n_i is an effective noise with a variance of σ^2. The receiver extracts DCT coefficients from the I and Q components and reconstructs the coefficients using the MMSE filter [11] as follows:

$$\hat{s}_i = \frac{g_i \lambda_i^2}{g_i^2 \lambda_i^2 + \sigma^2} \cdot y_i. \tag{5}$$

The receiver then obtains the corresponding video sequence using the inverse 3D-DCT for the filter output $\hat{s}_i$.

Table 3. Brief introduction to typical energy compaction techniques for soft delivery schemes

Techniques	Features	Pros	Cons
2D-DCT/ 2D-DWT	Take DCT/DWT operation for each video frame	Reduce spatial redundancy	No temporal filter
3D-DCT	Take DCT operation for each GoP	Reduce both spatial and temporal redundancy	Weak temporal filter
MCTF	Take wavelet transform for temporal filtering	Further reduce temporal redundancy	Computational cost for temporal filtering
Component protection	Send lower frequency coefficients as metadata	Distribute power to higher frequency coefficients	Communication overhead Significant degradation due to metadata error
Layered operation	Divide video frames into BL and ELs and send them in digital and pseudo-analog ways, respectively	Provide baseline quality via the BL while enhancing the quality via the ELs	ELs will be meaningless if bit errors occur in BL
Coset coding	Partition coefficients into several cosets and transmit the coset residual codes	Bring lower entropy according to a coset step	Accuracy of coset step and side information is crucial for reconstruction

4 TECHNICAL SOLUTIONS FOR SOFT DELIVERY

Because SoftCast skips nonlinear digital-based encoding and decoding operations corresponding to motion estimation, quantization, and entropy coding, it realizes a linear quality improvement associated with channel quality improvement. In particular, SoftCast has shown outstanding performance compared with the conventional digital-based delivery schemes when receivers are highly diverse and/or the channel condition of each receiver varies drastically. Conversely, the design of SoftCast is simplistic, so there remains much scope for improvement in adopting soft delivery in practical scenarios, including stable channel conditions, band-limited, and/or error-prone environments. For this purpose, many studies have been conducted to improve the performance of soft delivery. The existing works on soft delivery schemes can be classified into seven types, as shown in Fig. 1: energy compaction, optimal scaling, bandwidth utilization, resilience to packet loss, overhead reduction, hardware implementation, and extension for immersive experiences.

4.1 Energy Compaction of Source Signals

In soft delivery schemes via linear mapping (from source signals to channel signals), the reconstruction quality greatly depends on the performance of the energy compaction technique for the source signals. Specifically, the study in [40] clarified that the performance of soft delivery schemes degrades as the ratio of maximum energy to minimum energy of the source component increases. To yield better quality under both stable and unstable channel conditions, existing studies have adopted different energy compaction techniques listed in Table 3 for the source signals.

Typical solutions are to adopt wavelet-based signal decorrelation methods. Specifically, some studies [41–46] have adopted a motion-compensated temporal filter (MCTF), which is a temporal wavelet transform method, to remove inter-frame redundancy by realizing motion compensation in soft delivery. The MCTF recursively decomposes video frames into low- and high-frequency frames according to a predefined level. For example, WaveCast [44] adopted a 3D-discrete wavelet transform (DWT), i.e., the integration of 2D-DWT and MCTF, to remove temporal and spatial redundancy. Although SoftCast exploits a full-frame 3D-DCT to remove the intra- and inter-frame redundancy for energy compaction, WaveCast can further improve the reconstruction quality by fully exploiting the inter-frame redundancy using motion compensation. A detailed discussion on the effects of other decorrelation methods is presented in [47, 48]. [49] also utilized inter-frame redundancy by designing an adaptive GOP size mechanism. It adaptively controlled the GoP size based on shot changes and the spatio-temporal characteristics of the video frames. It then used a full-frame 3D-DCT for energy compaction across the video frames in one GoP.

Another typical solution is to send large energy coefficients as metadata, and thus prevent the transmission of such coefficients using pseudo-analog modulation. [50] designed Advanced SoftCast (ASoftCast) to send low-frequency coefficients as the metadata. ASoftCast decomposed the original images into frequency components using 2D-DWT. The frequency component was then divided into two parts: the lowest frequency sub-band and other sub-bands. The wavelet coefficients in the lowest-frequency sub-band are processed by run-length coding. They are then channel-coded and digitally modulated for additional metadata transmissions. The optimized power allocation for the SoftCast scheme in [51] selected and sent high-energy coefficients as the metadata to reduce the energy of the analog-modulated symbols. These results can assign a high transmission power to low-energy coefficients to improve the received quality. Here, determining the high-energy coefficients for each GoP is computationally complex owing to the use of an exhaustive search. To reduce the computational complexity, [52] adopted a zigzag scan to select the side information. Other studies in [53–58] divided the video into BL and ELs, which were coded and sent in digital and pseudo-analog ways, respectively. For example, the base layer in gradient-based image SoftCast (G-Cast) [57] sent the DC and low-frequency coefficients of the image, while the enhancement layer extracted and sent an image gradient, which represents the edge portion of the image, using a gradient transform. The receiver then created a final estimation of the image via a gradient-based reconstruction (GBR) procedure, utilizing both the image gradient at the enhancement layer and the low-frequency coefficients provided by the base layer.

Other solutions adopted a nonlinear encoder and decoder for source signals to decrease the ratio of the maximum to the minimum energy of the analog-modulated symbols. The typical solution for soft delivery is to introduce coset coding [59, 60], which is a typical technique in distributed source coding. Coset coding partitions the set of possible source values into several cosets and transmits the coset residual codes to the receiver. With the received coset codes and the predictor, the receiver can recover the source value in the coset by choosing the one closest to the predictor. DCast [61–64] first introduced coset coding for the soft delivery of inter frames. The coset coding in DCast divides each frequency domain coefficient s_i by a coset step q and obtains the coset residual code l_i as $l_i = s_i - \left\lfloor \frac{s_i}{q} + \frac{1}{2} \right\rfloor q$, where $\left\lfloor \frac{s_i}{q} + \frac{1}{2} \right\rfloor$ represents the coset index. At this time, the sender only needs to transmit the coset residual code for energy compaction. At the user side, with the received coset residual code $\hat{l}_i$ and the side information $\bar{s}_i$ (i.e., the predicted DCT coefficient obtained from the reference video frame), the receiver reconstructs the DCT coefficients by coset decoding. Given the coset residual code l_i, there are multiple possible reconstructions of s_i that form a coset $C = \{\hat{l}_i, \hat{l}_i \pm q, \hat{l}_i \pm 2q, \hat{l}_i \pm 3q, \ldots\}$. DCast then selects the coset C that is nearest to the side information $\bar{s}_i$ as the reconstruction of the DCT coefficient. In this case, the value of each coset step q is crucial for the coding performance of DCast. The value of q is calculated by estimating the noise at the receiver end shown in [63, 64]. However, the reconstruction quality of DCast also depends on the side information quality. If the side information $\bar{s}_i$ is error-prone, the receiver may make wrong decisions with a smaller q. [65] introduced a side information refinement (SIR) algorithm [66] to refine the side information for the quality enhancement of DCast.

The concept of coset coding has been widely applied in other studies on soft delivery for the same purpose. For example, [67–71] utilized pseudo-coset coding for lower frequency components and sent the coset index using the digital framework. Here, the residuals in the lowest-frequency components and other frequency components are sent using pseudo-analog modulation. The main difference between coset coding and pseudo-coset coding is the sending of the coset index as additional metadata. The layered coset coding and adaptive coset coding were applied to the soft delivery scheme in [69] and [70], respectively. LayerCast in [69] introduced layered coset coding to simultaneously accommodate heterogeneous users with diverse SNRs and bandwidths. The

Table 4. Overview of power allocation techniques for soft delivery schemes

Categories	Papers	Channel Consideration	Optimization Metric
Channel-aware power allocation	[11]	AWGN	MSE
	[75]	Fading	MSE
	[76, 77]	OFDM	MSE
	[78−80]	MIMO	MSE
	[81−83]	MIMO-OFDM	MSE
	[84]	Impulse noise	MSE
	[85, 86]	NOMA	MSE
	[87]	Underwater acoustic networks	MSE
	[88]	UAV-enabled networks	MSE
	[89]	mmWave lens MIMO	MSE
Perception-aware power allocation	[90]	AWGN	SSIM
	[91]	AWGN and MIMO	FWD
	[92]	AWGN	EQMSE
	[93]	AWGN	Foreground and background distortions

layered coset coding used large to small coset steps to obtain coarse to fine layers from each chunk. The coarse layer, i.e., BL, is sufficient to reconstruct a low-quality DCT chunk for narrowband users, whereas each fine layer, i.e., EL, provides refinement information of the DCT chunk for wideband users. [72−74] utilized the coset coding for cooperative soft delivery systems, i.e., a three-node relay network. A sender broadcasts the DCT coefficients obtained from the video frames using pseudo-analog modulation to the relay node and the destination node. If the channel quality between the sender and the destination node is higher than a threshold, the destination node reconstructs the video frames from the soft-delivered DCT coefficients. If the channel condition is lower than the threshold, the relay node sends the coset residual code to the destination node, and then the destination node reconstructs the video frames using the received coset residual code and the side information obtained from the softly delivered DCT coefficients from the sender.

4.2 Channel-Aware and Perception-Aware Power Allocation

As mentioned in Section 3.2, the power allocation in SoftCast minimizes the MSE between the original and reconstructed video signals over additive white Gaussian noise (AWGN) channels. There are several drawbacks toward adopting SoftCast in practical scenarios: 1) practical wireless channels have more complex characteristics, e.g., fading caused by multipath and impulse noise, than the AWGN channels, and 2) MSE is not an effective index for describing the perceptual fidelity of images/videos. To address the drawbacks related to the power allocation, the existing studies in Table 4 propose the power allocation for practical wireless channels and perceptual considerations.

For the first drawback, the existing studies redesigned the power allocation for practical wireless channels, including fading [75] and frequency-selective fading, i.e., orthogonal frequency-division multiplexing (OFDM) [76, 77], impulse noise [84], multiple-input and multiple-output (MIMO) [78−80], and MIMO-OFDM channels [81−83]. In [75], the authors designed an optimal power allocation for fading channels. In fading channels, a fading effect, i.e., multiplicative noise, will degrade the reconstruction quality. Although SoftCast assumes that multiplicative noise can be canceled with exact channel estimation at the receiver end, no algorithm can guarantee an error-free channel estimation. In addition to the power allocation design, the authors analyzed the effect of the channel estimation error on the reconstruction quality at the receiver end.

For frequency-selective fading channels, such as OFDM and MIMO-OFDM channels, the key issue is how to match the analog-modulated symbols to the independent subcarriers/subchannels for high-quality image/video reconstruction. In [81, 82], they observed similarities between the source

and channel characteristics and exploited the similarities for subcarrier/subchannel matching. Par-Cast [81] and the extended version of ParCast+ [82] assigned the more important DCT coefficients to higher gain channel components and allocated power weights for each DCT coefficient with joint consideration of the source and channel for video unicast systems. ECast in [83] extended the source and channel matching and power allocation for video multicast systems. For multicast systems, it is necessary to deal with the large overhead of channel feedback from multiple receivers. In ECast, multiple users simultaneously send tone signals for the channel feedback, and the sender receives the superposition of multiple tone signals. Although the sender cannot distinguish each of the channel gains, the weighted harmonic means of channel gains can be obtained from the superposed tone signals; thus, ECast utilizes the channel gain for the source and channel matching and power allocation.

Other studies solved power allocation problems in modern wireless systems, including non-orthogonal multiple access (NOMA) [85, 86], underwater acoustic OFDM [87], unmanned aerial vehicle (UAV)-enabled [88], and mmWave lens MIMO systems [89]. For example, in NOMA systems, source signals are coded into BL and ELs and then transmitted simultaneously through superposition coding (SC). With successive interference cancellation (SIC), near users with strong channel gains can decode both BL and EL signals, whereas far users with weak channel gains may only decode BL signals. In the existing studies, both BL and ELs are analog-coded in [85], whereas BL and ELs are digital- and analog-coded, respectively, in [86]. They solved the power allocation across the BL and ELs to minimize the distortion for all receivers with heterogeneous channel conditions. In underwater acoustic OFDM [87] and mmWave lens MIMO systems [89], the error behavior differed substantially across channel components, and the channel characteristics showed a similar tendency. They solved the source and channel matching and power allocation problems, which are also discussed in frequency-selective fading channels, to minimize the distortion at the receiver end.

For the second drawback, some studies [90–93] also redesigned the power allocation with perceptual considerations, including structural similarity (SSIM) [90], foveation [91], and saliency [92]. In these studies, determining the perception-aware weights for each source component is challenging. Specifically, in SoftCast, the scaling factor for each coefficient is obtained from its power information to minimize the MSE: $g_i \propto \lambda_i^{-1/4}$. These studies considered the perception-aware weight for the ith coefficient w_i in the scaling factor to minimize the perceptual distortion as: $g_i \propto w_i^{1/4} \lambda_i^{-1/4}$. For this purpose, [90] demonstrated the relationship between the MSE in the DCT coefficients and the SSIM distortion to obtain the weight for the ith DCT coefficients of all the chunks w_i. They found that the weight for the high-frequency coefficients was larger than that for the low-frequency coefficients, which was consistent with the characteristics of the human visual system (HVS). FoveaCast in [91] introduced the foveation-based HVS [94] and the corresponding HVS-based visual perceptual quality metric, called foveated weighted distortion (FWD), for the optimization objective. For a given foveation point (f_x, f_y) in the pixel and frequency domains, the error sensitivity for each pixel/frequency coefficient at location (x, y) can be defined in the foveation-based HVS. FoveaCast regarded the error sensitivity in the DWT domains as the weight w_i and performed foveation-aware power allocation. In [92], visual saliency maps were introduced for the perception-aware power allocation. Saliency maps represent the attended regions in an image when a user watches the image owing to the visual attention mechanism of the human brain. In this case, the weight for the ith pixel w_i is based on the normalized visual saliency defined from any arbitrary visual saliency model, such as the Itti–Koch–Niebur model [95]. Based on the weight, it allocates considerable transmission power to salient regions to minimize the eye-tracking weighted MSE (EQMSE).

Table 5. Brief introduction to the existing soft delivery techniques for band-limited channels

Papers	Techniques	Pros	Cons
[11]	Discarding low-energy chunks	Simple algorithm	Additional metadata for the location of discarded chunks and coefficients
[96]	Adaptive chunk division	Fully utilize available bandwidth by discarding small high-frequency chunks	Improper power allocation in low-frequency chunks
[97]	SK mapping	High reconstruction quality in middle and high channel SNRs	Low reconstruction quality in a low SNR regime
[98–105]	Compressive sensing	Recover discarded coefficients using a reconstruction algorithm	Computational cost for the algorithm
[106–110]	Data-assisted communication	Reduce traffic by utilizing related images in a cloud	Same or correlated images are available in a cloud

4.3 Bandwidth Utilization

The source bandwidth of soft delivery schemes depends on the number of transmitted analog-modulated symbols every second, i.e., baud rate. In the aforementioned designs, the source bandwidth is mainly considered sufficient to send all the transmitted non-zero analog-modulated symbols over the wireless medium. However, when the channel bandwidth is lower than the source bandwidth, some analog-modulated symbols are discarded at the receiver side. Hence, the loss of the important coefficients, i.e., the low-frequency coefficients, may have a significant impact on the reconstruction quality. Specifically, the expected distortions in soft delivery schemes for single and multiple contents owing to the bandwidth constraint under the transmission power constraint are discussed in [111] and [112, 113], respectively. Some existing studies have adopted different techniques listed in Table 5 to meet the bandwidth constraint. The typical method is to selectively discard the chunks in higher frequency components to fill the bandwidth [11, 96]. When the sender discards some chunks, the receiver regards all the coefficients in the discarded chunks as zeros. Because it needs to send the locations of the discarded chunks to the receiver, SoftCast sends the location information as a bitmap. Although SoftCast assumes equal-size chunks across low- to high-frequency components, [96] adopted smaller chunk sizes in high-frequency components to realize a fine-grained control to meet the bandwidth limitation. Another study in [97] used bandwidth-reducing Shannon-Kotelnikov (SK) mappings to increase the number of chunks transmitted over bandwidth-constrained channels. The SK mappings are typical N:1 bandwidth-reducing or 1:M bandwidth-expanding non-linear mappings. In this study, 2:1 SK mappings were used to encode several pairs of chunks with less energy to send more chunks with medium energy within the channel bandwidth.

Other studies [98–105] introduced compressive sensing (CS) techniques [114, 115] for soft delivery over bandwidth-constrained wireless channels. Notably, CS is a sampling paradigm that allows the simultaneous measurement and compression of signals that are sparse or compressible in some domains. In general, recovering source signals from compressed signals is impossible because the system is underdetermined. However, if the source signals are sufficiently sparse in some domains, the CS theory indicates that the source signals can be reconstructed from the compressed signals by solving the ℓ_1 minimization problem. The advantage of CS-based soft delivery is the recovery of chunks in high-frequency coefficients using CS-based signal reconstruction algorithms, such as approximate message passing (AMP) and iterative thresholding, even though the chunks are discarded at the sender's end. For high-quality reconstruction, adaptive rate control and reconstruction algorithms are mainly adopted for CS-based soft delivery. For instance, [100] adaptively controlled the compression rate based on visual attention, i.e., both the texture complexity and visual saliency, to satisfy the bandwidth constraint while maintaining better perceptual quality. [104] adaptively selected reliable columns from the measurement matrix and compressed source

signals using the selected columns. In view of the reconstruction algorithm, [101] designed an adaptive transform for noisy measurement signals to obtain sparser transform coefficients for clean reconstruction. [102] and [103] designed grouping methods for measurement signals to utilize the similarity between video frames for the reconstruction.

Other studies utilized stored images/videos on the cloud to reduce the bandwidth requirement in soft delivery. Specifically, data-assisted communication of mobile image (DAC-Mobi) [106], data-assisted cloud radio access network (DaC-RAN) [107], and knowledge-enhanced mobile video broadcasting (KMV-Cast) schemes [108–110], which are referred to as data-assisted soft delivery schemes, have been proposed for high-quality image/video transmission. The main contributions of the data-assisted soft delivery schemes are 1) a sender sends a limited number of analog-modulated symbols and 2) the receiver reconstructs images/videos using correlated images, i.e., side information, obtained from a cloud.

In DaC-Mobi [106], successive coset encoders were introduced to divide the DCT coefficients into three layers of bit planes: most significant bits (MSBs) in low-frequency coefficients, MSBs in other frequency coefficients and middle bits, and least significant bits (LSBs). Here, MSBs in low-frequency coefficients and LSBs were transmitted to the receiver in digital and pseudo-analog manners, respectively, whereas MSBs in other frequency coefficients and middle bits were discarded. Based on the received MSB in the low-frequency coefficients, the receiver reconstructs a down-sampled image to retrieve correlated images in the cloud. The retrieved correlated images were used as side information to resolve ambiguity due to discarded bits and reconstruct the entire image. DaC-RAN [107] and the extended version of KMV-Cast [108–110] adopted Bayesian reconstruction algorithms that utilize correlated images/videos. in the cloud as prior information to reduce the required bandwidth for soft delivery. The main difference between the DaC-RAN and KMV-Cast schemes is that the former assumes that the same images/videos exist in the cloud, whereas the latter does not require that the same images/videos exist at the receiver end by designing prior knowledge broadcasting in a digital manner.

The aforementioned studies considered the channel bandwidth to be lower than the source bandwidth. If the channel bandwidth is greater than the source bandwidth, the soft delivery schemes become less efficient. In this case, the soft delivery schemes utilize the extra bandwidth by retransmission. [116] and [117] designed an analog channel coding to use theextra channel bandwidth for quality enhancement. For example, [117] proposed a chaotic function-based analog encoding [118] for soft delivery. Although the existing chaotic function-based analog coding is designed for uniformly distributed sources, the analog coding for Gaussian distributed sources significantly amplifies source signals and thus consumes unnecessary transmission power. They designed a chaotic map function for Gaussian distributed source signals to prevent power increments compared to the input power. Mcast in [119] also utilized extra bandwidth for quality improvement. As mentioned earlier, the sender can send the source data multiple times if an extra bandwidth is available. In this case, the utilization of extra time slots for quality improvement is a key issue. To overcome this issue, MCast optimized the assignment of the chunks of the DCT coefficients to available channels in multiple time slots to fully exploit the time and frequency diversities.

In contrast to the aforementioned studies, [120, 121] dealt with bandwidth variations. When the available bandwidth is less than the expected bandwidth at the sender's end, some important chunks will not have the opportunity to be transmitted before the playback deadline. They grouped several chunks into a tile and sent the tile with a large variance and high priority to dispatch important coefficients before the playback deadline.

4.4 Packet Loss Resilience

Even when the channel bandwidth is sufficient to send all the non-zero analog-modulated symbols, some analog-modulated symbols can be discarded at the receiver side owing to loss-prone wireless channels. Specifically, the packet loss owing to strong fading and interference may have a significant impact on the reconstruction quality if important chunks and coefficients are lost. SoftCast used the WHT to redistribute the energy of the source signals across whole packets for resilience against packet loss. However, each packet still contains a large amount of energy, and thus, degradation owing to packet losses remains considerable.

To maintain better reconstruction quality in error-prone wireless channels, some related studies [122–124] have introduced CS techniques, i.e., block-wise CS [125], for packet loss resilience. The CS technique is suitable for wireless transmission with random packet loss owing to its random measurement. Random measurement considers all packets as of equal importance. In contrast to typical CS techniques, block-wise CS can reduce the storage and computational costs of the reconstruction. A pioneering work on packet loss resilience is the distributed compressed sensing-based multicast scheme (DCS-cast) [122]. In the DCS-Cast, each image is first divided into blocks and the coefficients in each block are randomized using the same measurement matrix across the blocks. One coefficient in every block is packetized to normalize the importance across packets. Even though some packets may be lost over loss-prone wireless channels, the receiver obtains noisy pixel values using the same measurement matrix at the sender and reconstructs the lost pixel values using the CS reconstruction algorithm in the DCT/DWT domains. Because the lost pixel values can be recovered from the reconstruction algorithm, DCS-Cast maintains high image/video quality in loss-prone channels. To further improve the reconstruction quality, multi-scale [123] and adaptive [124] block-wise CS algorithms have been adopted for soft delivery. The multi-scale block-wise CS algorithm [123] decomposes each video frame into a multi-level 2D-DWT and then optimizes the sampling rate for each DWT level according to its importance. However, the adaptive block-wise CS algorithm [124] divides several video frames into one reference frame and subsequent non-reference frames and adaptively determines whether direct or predictive sampling should be used for each block in a non-reference frame. Direct sampling randomizes the signals in the block, whereas predictive sampling calculates the residuals between the blocks in the reference and non-reference frames and randomizes residuals to utilize the inter-frame similarity for the reconstruction.

4.5 Overhead Reduction

In soft delivery schemes without chunk division, a sender needs to let the receiver know the power information of all the DCT coefficients to demodulate the signals. For the receiver to carry out the MMSE filtering in Eq. (5), the sender needs to transmit λ_i of all coefficients without errors as metadata, which may constitute a large overhead. For example, when the sender transmits eight video frames with a resolution of 352×288, the sender needs to transmit metadata for all DCT coefficients, i.e., $352 \times 288 \times 8 = 811,008$ variables in total, to the receiver. This overhead may induce performance degradation owing to the rate and power losses in the transmission of analog-modulated symbols. To reduce the overhead, SoftCast divides the DCT coefficients into chunks and carries out chunk-wise power allocation using an MMSE filter. However, overheads are still high, and chunk division causes performance degradation due to improper power allocation.

To achieve better quality under a low overhead requirement, the related studies can be classified into two types, as shown in Figs. 6 (a) and (b): 1) sender-side overhead reduction, and 2) receiver-side overhead reduction. Studies on the sender-side overhead reduction [126–129] designed fitting functions to obtain the power information with fewer parameters. In this case, the sender and

(a) Sender-side Overhead Reduction   (b) Receiver-side Overhead Reduction
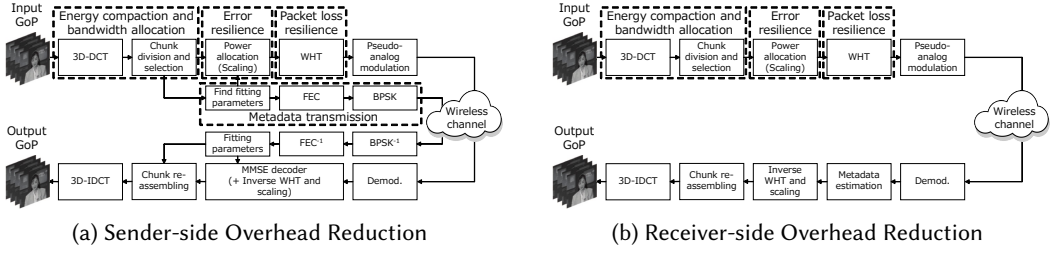
Fig. 6. Block diagram of the sender-side and receiver-side overhead reduction methods. (a) Sender utilizes fitting functions to obtain power information with fewer parameters. (b) Receiver estimates the power information only from the received symbols.

Table 6. Categories of the existing soft delivery schemes in terms of performance evaluation

Evaluations	Papers
Simulations	[11, 12, 38, 39, 42–44, 46–58, 61–65, 67–79, 83–93, 96–102, 104–113, 116, 119, 120, 122–124, 126–133]
Emulations	[41, 45, 81, 82, 121, 134]
Experiments	USRP [13, 103, 135], SOUP [117], Xilinx Virtex7 [136–138], MU-MIMO prototype [139], LTE prototype [140]

receiver share the same fitting function in advance and send the parameters as metadata for overhead reduction. Specifically, [126] designed a fitting function with four parameters for each chunk, and [127] designed a log-linear function with two parameters for each chunk. Another study [128] found that equal-size chunk division was not suitable for chunk-wise fitting, and thus, an adaptive chunk division, i.e., L-shaped chunk division, was designed for an accurate fitting. In addition, [129] exploited a Lorentzian fitting function with seven parameters based on a Gaussian Markov random field (GMRF) for each GoP. The sender-side studies accurately predict the metadata using the fitting function with a limited number of the parameters, i.e., low overhead, whereas the fitting function causes an additional computational cost.

Studies on receiver-side overhead reduction [130, 131] estimate the power information only from the received signals without any additional computational cost at the sender side. [130] is a pioneer work to estimate the power information from the received signals. Blind data detection (BDD) [131] was proposed to decode the received analog-modulated symbols without the power information at the receiver. Specifically, BDD uses a zero-forcing estimator and the sign of the received signals to approximate the source signals. One typical issue of the receiver-side overhead reduction is the reconstruction quality highly depends on the quality of the received signals.

We note that both types of overhead reduction cause quality degradation owing to estimation errors. In [127], the effect of modeling accuracy on the reconstruction quality in soft delivery was analyzed.

4.6 Implementation

The aforementioned studies mainly discussed performance improvements in theoretical analyses and simulations. Table 6 lists the categories of the existing studies in terms of the performance evaluation. The existing studies in [41, 45, 81, 82, 121, 134] used a software defined radio platform SORA [141] to carry out emulations. In contrast to the simulations, the emulations obtain channel fading and noise trace from SORA to evaluate the performance under real wireless environments.

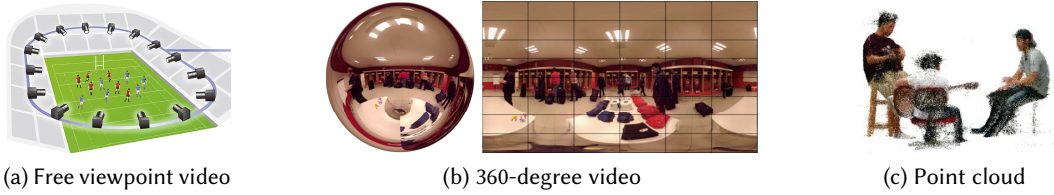(a) Free viewpoint video   (b) 360-degree video   (c) Point cloud

Fig. 7. Typical immersive contents. (a) free-viewpoint video. (b) 360-degree video. (c) point cloud.

Table 7. Typical immersive content and its features

Content	Acquisition	Display	Key issues
Free viewpoint video	Large number of closely spaced RGB and IR camera arrays	Synthesize virtual cameras using rendering and freely switch the viewing camera	Resource allocation for each RGB and IR camera to maximize the viewing quality
360-degree video	360-degree camera	Playback viewport through VR headset	Predict future viewport and allocate resource to the viewport for quality maximization
Point cloud	Laser scanner	Playback 3D points through AR and MR headsets and holographic display	Compress and send numerous and irregular structure of 3D points

Some studies implemented a soft delivery scheme on software-defined radio platform [13, 103, 117, 135] and field-programmable gate array (FPGA) [136–138] to empirically demonstrate the benefits of soft delivery in practical wireless channels. In [13, 103, 135], the authors used Universal Software Radio Peripheral (USRP) 2, USRP NI2900, and USRP X310 and GNU Radio for implementation and evaluated the visual quality of soft delivery, respectively. In addition, the authors in [117] built an experimental system based on OpenAirInterface (OAI) platform and self-developed Software Universal Platform (SOUP). They migrated OAI to self-developed SOUP software defined radio, and implemented the proposed scheme based on OAI eMBMS codes. Conversely, in [136–138], they exploited the Xilinx Virtex7 FPGA for implementation and tested the reconstruction quality as a function of wireless channel SNRs.

Other studies [139, 140] implemented soft delivery on the prototypes of multi-user MIMO (MU-MIMO) and long-term evolution (LTE) systems. For example, in [139], SoftCast is implemented on BUSH, which is a large-scale MU-MIMO prototype that performs scalable beam user selection with hybrid beamforming for phased-array antennas in legacy WLANs. They performed experiments to evaluate the video quality in terms of peak signal-to-noise ratio (PSNR) and SSIM over a lossy MU-MIMO channel.

4.7 Extension for Immersive Experiences

SoftCast and other soft delivery schemes mentioned in the previous sections were designed for conventional images and video signals. In modern wireless and mobile communication scenarios, the streaming of immersive content will be a key application for reconstructing 3D perceptual scenes that provide full parallax and depth information for human eyes. The immersive contents can be applied to various applications, such as three to six degrees-of-freedom (6-DoF) entertainment, remote device operation, medical imaging, vehicular perception, VR/AR/MR, and simulated training. Figures 7 (a) through (c) show the typical immersive contents of free-viewpoint video [142–144], 360-degree video, and point cloud [145] and Table 7 lists the features. Even in immersive contents, the video frames are compressed in a digital manner, and the compressed bitstream is then channel-coded and modulated in sequence. This means that cliff and leveling effects still occur in the streaming of the immersive contents owing to the variation in the channel conditions. To prevent

cliff, leveling, and staircase effects, some studies have extended soft delivery schemes toward immersive content for future wireless multimedia services.

One of the key advantages of soft delivery schemes for immersive content is that they simplify the optimization problem for image and video quality maximization. In immersive content delivery, the main issue is to maximize the image and video quality considering the user's perspective. For example, the view synthesis distortion optimization problem and the viewport optimization problem should be solved in free-viewpoint video and 360-degree video, respectively. In digital-based delivery schemes, a sender needs to find the best bit and transmission power allocation for video frames. However, it is often cumbersome to derive a solution. Soft delivery schemes simplify the optimization problems by reformulating them into a simple power allocation problem since bit allocation for quantization is not required in soft delivery schemes.

4.7.1 Free Viewpoint Video. Free-viewpoint videos enable us to observe a 3D scene from freely switchable angles/viewpoints. Figure 7 (a) shows an example of a free-viewpoint video wherein numerous closely spaced RGB and infrared (IR) camera arrays are deployed to capture the texture and depth frames of a 3D scene, such as a football game. Even though the number of deployed cameras in the field is limited owing to physical constraints, the receiver can synthesize intermediate virtual viewpoints using rendering techniques, e.g., depth image-based rendering [146, 147] to obtain numerous switchable viewpoints. To synthesize intermediate virtual viewpoints using the rendering technique, the sender encodes and transmits the texture and depth frames of two or more adjacent viewpoints, the format of which is known as multi-view plus depth (MVD) [148].

For conventional MVD video streaming over wireless links, digital video compression for MVD video frames, e.g., MVC+D [149] or 3D-AVC [150], fully utilizes the redundancy between the cameras and texture-depth for compression. In this case, the streaming schemes need to solve view-synthesis problems in addition to cliff and leveling effects to yield better video quality even in the synthesized virtual viewpoints. Specifically, the video quality of the virtual viewpoint is determined by the distortion of each texture and depth frame. In digital-based MVD schemes, the distortion depends on the bit and power assignments for each texture and depth frame. It is often cumbersome to achieve the best quality at a target virtual viewpoint using parameter optimization owing to the combinatorial problem with nonlinear quantization.

Some studies [151–155] designed a soft delivery scheme for a free-viewpoint video. Specifically, FreeCast [151, 152] is the first scheme for a free-viewpoint video. Because MVD video frames have redundancy of cameras and texture-depth, FreeCast jointly transforms texture and depth frames using 5D-DCT to exploit inter-view and texture-depth correlations for energy compaction. In addition, FreeCast can simplify the optimization problems of view synthesis by reformulating it into a simple power assignment problem. This is because bit allocation, i.e., quantization, is not required in FreeCast. They found that the power assignment problem for the texture and depth frames can be solved using a quadratic function to yield the best quality at the desired virtual viewpoint. Furthermore, FreeCast introduces a fitting function obtained from multi-dimensional GMRF at the sender and the receiver to obtain the power information with few parameters for the overhead reduction.

3DV SoftCast in [154] focused on the view synthesis problems under the 3D-DCT operations for each camera's texture and depth frames, and designed the power allocation method to solve the problem. The main differences between 3DV SoftCast and FreeCast are that 3DV SoftCast performs 3D-DCT for each camera and controls the transmission power to minimize the view synthesis distortion, while FreeCast performs 5D-DCT for better energy compaction of analog-modulated symbols and multi-dimensional GMRF-based overhead reduction for reconstructing high-quality MVD frames in band-limited environments. The study in [155] designed a soft delivery scheme

for depth video. They found that a block-based DCT performs well on depth video compared to a full-frame DCT because depth video has different characteristics from texture video. Although a different soft delivery scheme is required for texture video, low-distortion depth video in [155] can provide better virtual viewpoint quality in free viewpoint video.

4.7.2 360-Degree Video. Notably, 360-degree video contents build a synthetic virtual environment to mimic the real world with which the users interact. Each user can watch 360-degree videos through a traditional computer-supported VR headset or an all-in-one headset (e.g., Oculus Go). When the user requests the 360-degree video, the sender sends the 360-degree video frames, and the user may play a part of the 360-degree video frames, which is referred to as the viewport, through the user's headset. Here, 360-degree videos are mainly captured by an omnidirectional camera or a combination of multiple cameras and saved in a spherical format. Before transmissions, the sphere frames are mapped onto the 2D plane using a certain projection method, e.g., equirectangular and cube map projections.

In 360-degree video streaming, the major issue is to yield better video quality in the user's viewport by effectively reducing perceptual redundancy within 360-degree video frames. Because each user only watches the viewport via the headset at each time instance, excessive video traffic is created if the sender sends the full resolution of the 2D-projected video frames with an identical quantization parameter. One of the simplest methods to reduce perceptual redundancy is viewport-only streaming [156]. In video playback, the user may move a viewing viewport according to the user's head/eye movement. Based on the movement, the user requests a new viewport from the sender, and the sender sends back the corresponding viewport. Because the sender transmits one viewport at each time instant, viewport-only streaming can mitigate the video traffic. However, the user needs to receive a new viewport from the sender in every viewport switching, which causes a long switching delay. A long switching delay, i.e., approximately 10 ms, may cause *simulator sickness* [157]. Owing to a long delay in the standard Internet, it is difficult for viewport-only streaming schemes to satisfy the switching delay requirements. To prevent simulator sickness, conventional schemes [158] divide 360-degree video frames into multiple tiles and independently encode them with different quantization parameters to yield better viewport quality within the bandwidth constraint.

The studies [159–162] on soft delivery schemes focus on the quality optimization of the user's viewport in addition to cliff and leveling effect prevention. [159] is the first scheme for viewport-aware soft 360-degree video delivery. According to the viewing viewport, the sender first adopts pixel-wise power allocation to reduce the perceptual redundancy in 360-degree video frames and then carries out the combination of one-dimensional DCT (1D-DCT) and spherical wavelet transform (SWT) for decorrelation to utilize the redundancy in the sphere and time domains. Omni-Cast [160] further considers the features of 360-degree videos into quality optimization. Specifically, they analyze the relationship of the distortion between the spherical and projected 2D domains as the spherical distortion for each projection method, and design power allocation to realize the optimal quality in the 2D-projected 360-degree videos. 360Cast [161] and the extended version of 360Cast+ [162] adopt viewport prediction based on linear regression and foveation-aware power allocation within the predicted viewport to further reduce the perceptual redundancy. They evaluate 360Cast+ with the existing digital-based schemes in terms of weighted-to-spherically uniform peak SNR (WS-PSNR) [163]. Here, the digital-based schemes use HEVC Test Model (HM) 16.20 [164] and the modulation format of BPSK. They found 360Cast+ improves the average WS-PSNR performance compared with the digital-based schemes by preventing the cliff effect at low SNR regimes and gradually improving the received video quality with the improvement of the wireless channel quality.

4.7.3 Point Cloud. Volumetric content delivery provides highly immersive experiences for users through XR devices. The point cloud [145] is arguably the most popular volumetric data structure for representing 3D scenes and objects on holographic displays [165, 166]. A point cloud typically consists of a set of 3D points, and each point is defined by 3D coordinates, i.e., (X, Y, Z), and color attributes, i.e., (R, G, B). In contrast to conventional 2D images and videos, 3D point cloud data are neither well aligned nor uniformly distributed in space.

The major challenge in volumetric delivery over wireless channels is how to efficiently compress and send numerous and irregular structures of the 3D point cloud within a limited bandwidth. Some compression methods have been proposed for point clouds to deliver 3D data. Specifically, Draco [167] employs kd tree-based compression [168] and a point cloud library (PCL) using octree-based compression [169–171]. To further reduce the amount of data traffic in point cloud delivery, two transform techniques have been proposed for energy compaction of the non-ordered and non-uniformly distributed signals: Fourier-based transform, e.g., graph Fourier transform (GFT) and wavelet-based transform, e.g., region-adaptive Haar transform [172]. For example, recent studies used GFT for the color components [173] and 3D coordinates [174] of graph signals for signal decorrelation. They used quantization and entropy coding for the compression of decorrelated signals.

HoloCast [175] is a pioneering work on soft 3D point-cloud delivery for unstable wireless channels. Specifically, they regard 3D points as vertices in a graph with edges between nearby vertices to deal with the irregular structure of the 3D points motivated by [174, 176]. HoloCast uses GFT for such graph signals to exploit the underlying correlations among adjacent graph signals and directly transmits linear-transformed graph signals as a pseudo-analog modulation over the channel. We have compared HoloCast with the conventional digital-based delivery, which is based on point cloud digital compression used in PCL [169]. HoloCast gradually improves the reconstruction quality with the improvement of wireless channel quality. In addition, the GFT-based HoloCast can achieve better quality compared with the DCT-based HoloCast.

However, it has been found that graph-based coding schemes need to send the graph-based transform basis matrix used in GFT as additional metadata for signal decoding. For example, the sender needs to send N^2 real elements of the graph-based transform basis matrix as the metadata when the number of 3D points is N. In [177–179], Givens rotation [180, 181] was used for GFT basis matrix compression. Givens rotation is used to selectively introduce zeros into a matrix to create an identity matrix from the basis matrix using angle parameters. The angle parameters are uniformly and non-uniformly quantized prior to the metadata transmission for overhead reduction. From the evaluations, Givens rotation with the uniform quantization reduces the overhead up to 89.8% [177] compared with HoloCast without the overhead reduction. In addition, Givens rotation with the non-uniform quantization further reduces the overhead up to 28.6% [178] compared with the uniform quantization.

5 FUTURE DIRECTIONS

The existing soft delivery schemes have been studied to overcome the issues of conventional image and video streaming in modern wireless and mobile networks. In this section, we foresee the future directions of the soft delivery. Table 8 lists the features and challenges of each future direction. Specifically, the integration of digital-based and deep neural network (DNN)-based operations with soft delivery will be further discussed to yield better reconstruction quality. These studies are called hybrid digital-analog (HDA) delivery and AI-empowered soft delivery. In addition, our and other studies find that soft delivery can improve the delivery quality of the DNN architectures and tactile data. The soft delivery-based schemes may become a new standard for such delivery.

Table 8. Future directions and challenges of soft delivery schemes

| Future directions	Advantage	Disadvantage	Challenges
HDA delivery	Compact signal energy by integrating with digital-based operations	Cause cliff effect if digital-coded symbols fail	Discuss trade-off between coding delay and quality
AI-empowered soft delivery	Realize energy compaction and signal reconstruction by using DNN-based architectures	Large computational overhead	Deal with bandwidth heterogeneity, Design an optimal architecture for semantic communication
Soft delivery for AI	Efficiently exchange model parameters by using simultaneous transmission	Require symbol-level synchronization	Design power allocation for low-energy FL
Soft delivery for tactile Internet	Meet strict delay constraint	Consider a single vibrotactile sensor	Deal with multiple vibrotactile sensors, Minimize the distortion of human tactile perception
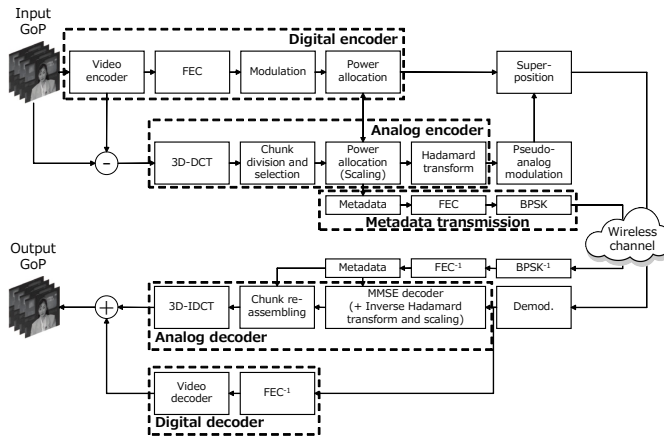
Fig. 8. Typical framework of hybrid digital–analog delivery.

Although the neural network compression and haptic codec are designed for the delivery, it causes low reconstruction quality owing to insufficient energy compaction.

5.1 Hybrid Digital–Analog Delivery

For further quality improvement, the pioneer studies of the HDA delivery [182, 183] integrate low-rate digital-based encoding and decoding into soft delivery. They proposed the superposition of analog-coded and digital-coded symbols to take advantage of conventional digital-based and soft delivery schemes. Specifically, the digital-coded symbols provide the baseline quality of the video frames, while the analog-coded symbols enhance the quality of the video frames according to the wireless channel quality. Here, the low-rate digital-based operations can significantly reduce the signal energy of the analog-coded symbols, e.g., the decrement of the ratio of the maximum energy to the minimum energy of the source component. A theoretical study in [40] clarified that a lower ratio improves the reconstruction quality of the analog-coded symbols. This means that the amount of quality improvement with the improvement of the wireless channel quality in HDA coding schemes becomes more significant compared to the pure soft delivery schemes. Nonetheless, the integration with digital-based encoding has one drawback the cliff effect may occur when the decoding of digital-coded symbols fails.

Figure 8 shows an overview of the HDA delivery schemes. The HDA delivery schemes consist of the digital and analog coding parts. At the sender side, the video frames are first encoded

by the digital video encoder and the digitally-coded bitstream is channel coded, modulated, and
assigned transmission power by the sender. Meanwhile, the residuals are coded, power-assigned,
and modulated by the soft delivery scheme. Both outputs from the digital and analog coding parts
are superposed and transmitted over wireless channels. In this case, the transmitted signal x_i is the
sum of the BPSK-modulated vector signal $x_i^{\langle d \rangle}$ and output vector signal of the soft delivery scheme
$x_i^{\langle a \rangle}$ as follows:

$$x_i = x_i^{\langle d \rangle} + J x_i^{\langle a \rangle}, \tag{6}$$

The BPSK-modulated symbol and the analog-modulated symbol are scaled by P_d and g_i, respectively.

$$x_i^{\langle d \rangle} = \sqrt{P_d} \cdot b_i, \qquad x_i^{\langle a \rangle} = g_i \cdot s_i, \tag{7}$$

where $b_i \in \mathbb{X} = \{\pm 1\}$ is the BPSK-modulated symbol and $J = \sqrt{-1}$ denotes the imaginary unit.
Here, the near-optimal solution of g_i under the transmission power budget P_a is based on Eq. (4).
We note that the budgets of the transmission power for the digital and analog parts need to satisfy
the total power budget P_t, i.e., $P_t = P_a + P_d$.

At the receiver side, it first decodes the digital-modulated symbols and then obtains the analog-
modulated symbols by subtracting the digital-modulated symbols from the received symbols. Finally,
the receiver reconstructs the baseline quality of the video frames from the output of the digital part
and enhances the video quality by adding the output of the analog part.

A key issue in HDA delivery is the assignment of transmission power to the digital and analog
parts [184]. Specifically, the power assigned to the digital part must guarantee the correct decoding
of the symbols. By contrast, the digital decoder treats the superimposed analog-modulated symbols
$x_i^{\langle a \rangle}$ as noise. To achieve better decoding performance, the I component of $x_i^{\langle a \rangle}$ should be kept as
small as possible. In [185], they only select the high-frequency coefficients, which are expected
to be very small values for superposition. The remaining low-frequency coefficients are delivered
using pseudo-analog modulation. The HDA framework in [186] regards the superposed symbols as
three main parts: 1) orthogonal analog symbols, 2) digital symbols, and 3) nonorthogonal analog
symbols superimposed onto digital symbols. They designed resource allocation among these three
parts to achieve a better balance between lowering interference and improving reconstruction
quality. Another study [187] designs a prediction model to describe the relationship between the
variance of residuals and the quantization parameter, and determines the optimal transmission
power for the analog part, which maximizes the reconstruction quality with the correct decoding
of the digital part. The HDA delivery scheme in [188] treats the imperfect decoding of the digital
part and finds the best assignment of the transmission power for the digital and analog parts. This
prevents too much power assignment for the digital part to ensure a low bit error rate (BER). In
contrast to the aforementioned studies, [189] treats the bandwidth of other digital traffic as hidden
resources for HDA video delivery. Specifically, they superimpose the analog-modulated symbols
and digital symbols of the other digital traffic to utilize the hidden resource under the constraint
that the BER requirement of the other digital traffic is not compromised.

Other studies have redesigned the power allocation in HDA delivery for practical wireless
channel environments, including fading [190, 191], OFDM [192, 193], MIMO [194], and relay
networks [195]. For example, the power allocation with perfect channel state information (CSI) is
designed in [190, 192, 194], whereas the power allocation with imperfect CSI is designed in [191, 193].
In view of the packet loss resilience in HDA delivery, the study in [196] introduced compressive
sensing for the residuals.

Other studies [197–199] extend HDA video delivery for immersive contents. Swift in [197]
considers stereo video delivery and designs a zigzag coding structure for the stereo video to utilize
both intra- and inter-view correlations. In the zigzag coding structure, the odd frames in the left view

and the even frames in the right view are encoded digitally, and the rest of the frames are encoded in analog. Here, the reconstructions of the digitally coded frames are used as side information to further remove redundant information from the analog-coded frames. Another study [198] extends HDA delivery for MVD videos and solves the view synthesis optimization to yield the best quality from an intermediate virtual viewpoint. HoloCast+ in [199] designs HDA delivery for point cloud delivery.

In future work, the recent coding standards such as H.266/VVC, learned video compression [200, 201], and point cloud coding can be used for the digital part. Although they have achieved significant energy compaction and can further improve the quality of the analog part, the recent coding standards require a long coding delay for compression. The trade-off between coding delay and reconstructed image and video quality is an open question in HDA delivery.

5.2 AI-Empowered Soft Delivery

Some recent studies integrate DNN architectures for nonlinear encoding and decoding operations of the soft delivery, namely, AI-empowered soft delivery. The AI-empowered soft delivery schemes utilize deep convolutional neural networks (DCNNs) [202, 203] and multi-layer perceptron (MLP) networks for energy compaction, power allocation, and overhead reduction tasks.

The MLP auto-encoder was first adopted to reduce the overhead of soft delivery [204]. Specifically, the proposed encoder obtains a few latent variables from the pixel values, and the proposed decoder decodes the accurate power information from the received latent variables for proper power allocation. The reconstruction quality can be maintained even with only one metadata across one GoP. Another study [205] designs deep joint source channel coding (DJSCC) for the energy compaction of the image and video signals. The DJSCC schemes integrate a DCNN-based auto-encoder into a soft delivery scheme. The proposed encoder directly compresses each image into a limited number of latent variables, and the proposed decoder reconstructs the image from the latent variables. Here, the latent variables are transmitted over wireless channels using pseudo-analog modulation. Even though the latent variables are obtained by nonlinear functions and delivered over wireless channels with a lower SNR, cliff and leveling effects can be prevented via pseudo-analog modulation. Other studies have introduced the DNN architecture for power allocation [206] and decoding operations [207]. The study in [206] uses a you-only-look-once (YOLO) structure [208] to extract the region of interest (ROI) and non-ROI parts from each image and then assign unequal transmission power across ROI and non-ROI parts for perceptual quality enhancement. The proposed scheme in [207] integrates DCNN-based image denoising, specifically, deep-image-prior (DIP) [209], into soft delivery. The DIP finds linear and nonlinear noise effects for reconstructing clean images from noisy images. The proposed scheme can remove fading and noise effects from the received images using DCNN-based image restoration. Another study [210] introduces graph neural networks (GNN) [211] for wireless point cloud delivery. The GNN is a novel model for graph representation learning that allows the analysis of the irregular geometric structure of graph data. GNN-based auto-encoder (GAE) [212, 213] was designed to encode 3D point clouds into a limited number of latent variables. One of the benefits of the GAE is that it allows graph signal reconstruction from a limited number of latent variables without requiring additional metadata.

DJSCC schemes with the latest neural network architectures have been well-studied for further energy compaction in recent years. As a result, they have achieved better image and video quality without cliff, leveling, and staircase effects. However, the existing DJSCC schemes need to deal with the bandwidth heterogeneity among the receivers. Here, how to adaptively improve the image and video quality according to the available bandwidth using the same architecture remains challenging work. In addition, DJSCC schemes have been considered as the fundamental techniques
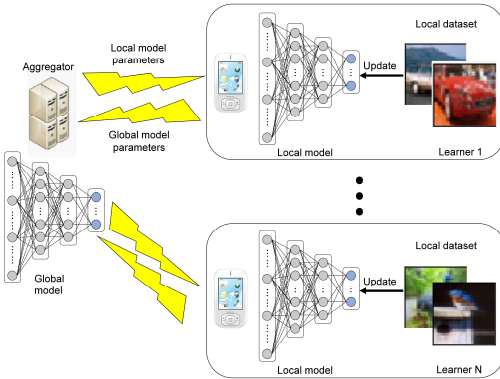
(a) Overview
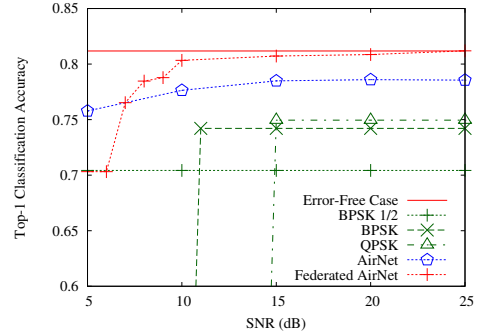
(b) Global model performance over the digital-based, analog-based, and the proposed Federated AirNet schemes

Fig. 9. Federated learning over wireless channels. (a) overview and (b) top-1 classification accuracy of global model over the digital-based, analog-based, and the proposed Federated AirNet schemes as a function of wireless channel quality.

for realizing semantic communication [214, 215] in future wireless and mobile networks. Semantic communication has been envisioned as a new transmission paradigm that delivers semantic meaning rather than a bit stream of transmitted messages. Another challenging issue is to design an optimal DJSCC scheme for the given semantic triple.

5.3 Soft Delivery for AI

Our and other studies find that soft delivery can be utilized to support various AI architectures. Specifically, many AI-based applications need to exchange trained DNN models between the receiver over wireless networks within a short delay including viewport prediction [216] in untethered XR applications, and dead reckoning in autonomous driving [217, 218] and online gaming services [219]. In recent years, digital-based model compression schemes [220, 221] have been designed and standardized for sharing the trained model over the networks. However, the existing studies found that the cliff and leveling effects occur even in the DNN model transmissions. To prevent both effects, analog modulation is effective in the model transmissions. AirNet [222] adopts analog modulation to deliver the DNN model parameters over wireless networks. Specifically, AirNet directly maps the model parameters to the transmission symbols and sends the analog-modulated symbols via wireless channels. This process avoids the issues mentioned above and the model restoration quality faithfully corresponds to the instantaneous channel condition. In addition, AirNet adopts SK mappings [97] to reduce the number of transmission symbols for band-limited channels.

In addition, the model parameter transmission is key technique to realize federated learning (FL) [223, 224] over wireless networks. Figure 9 (a) shows one typical FL example over wireless networks. The FL is a decentralized learning approach, which trains the model over a federation of distributed learners and an aggregator, to obtain an accurate model even with a limited dataset in each distributed learner. Each learner in the federation uses only locally available data for training. For the training over the distributed learners and aggregator, both learners and aggregator exchange the model parameters over wireless channels. The existing studies found that the analog

(a) Soft haptic delivery

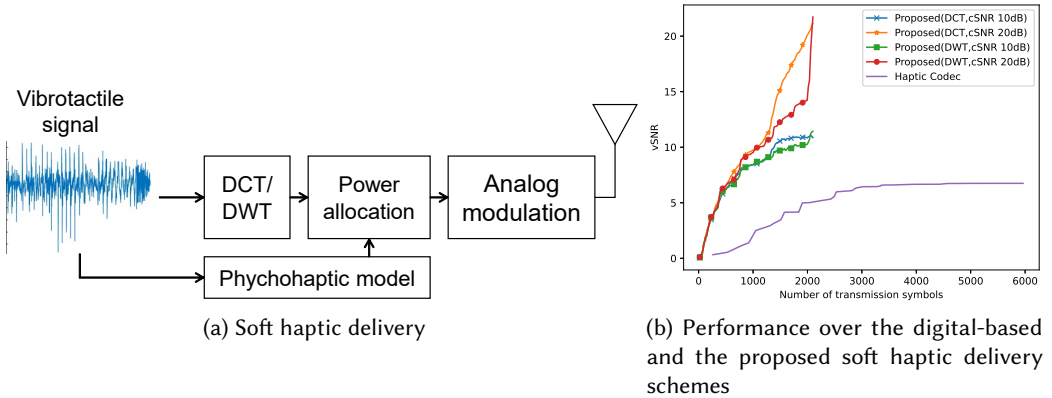(b) Performance over the digital-based and the proposed soft haptic delivery schemes

Fig. 10. Soft haptic delivery. (a) overview and (b) reconstruction quality over the digital-based and the soft haptic delivery schemes as a function of number of transmission symbols.

modulation-based solutions are efficient for exchanging the model parameters for the FL. The analog over-the-air computation (AirComp) [225, 226] is a typical solution for the model parameter transmission in FL. All the learners simultaneously send the analog-modulated parameters with channel inversion to the aggregator, and the aggregator can receive the aggregated model parameters from the superimposed waveforms. Although the simultaneous analog transmission can improve the throughput of the model parameter transmission, AirComp requires a precise symbol-level synchronization between the distributed learners. In [227], we proposed the model parameter transmission in a round-robin manner for quasi-asynchronous FL systems. Specifically, Federated AirNet based on the HDA delivery integrates low-rate model parameter compression and energy-compact analog modulation.

Figure 9 (b) shows the average top-1 classification accuracy of the global model as a function of wireless channel SNR when an available number of transmission symbols is at most 6.0 Msymbols. The number of transmission iterations from ten learners to aggregator is 10. We compare Federated AirNet to two state-of-the-art digital and analog approaches, specifically, DeepCABAC and AirNet. In the DeepCABAC scheme, the compressed bitstream is channel coded by a half-rate convolutional code with a constraint length of 8 and digitally modulated by BPSK or 4-QAM formats. The AirNet scheme directly maps each of the model parameters onto a transmission symbol in analog modulation. We found that the AirNet and Federated AirNet schemes prevent the drastic degradation in accuracy since they does not rely on quantization and entropy coding. In addition, the proposed Federated AirNet scheme yields the best accuracy, in higher channel SNR regimes, achieving near error-free performance.

Although the soft delivery schemes have the potential for the realization of FL, they cause a large energy consumption because they consider the transmission power of each distributed learner as identical. Although each learner can limit the transmission power to reduce energy consumption, the limitation may cause a long delay for convergence and low global model performance. The discussion on the transmission power allocation among the distributed learners will be a key issue for the realization of FL with low energy consumption.

5.4   Soft Delivery for Tactile Internet

In addition to visual information, multiple sensorial media (mulsemedia) delivery can enhance the quality of immersive experiences in XR applications. Tactile is one of the typical mulsemedia data. Especially, tactile communications can support the untethered and immersive XR applications. In contrast to visual information, the sampling rate of tactile information is relatively high, i.e., above 1,000 Hz, and the delay requirement for tactile communications is strict. In this case, the sender does not retransmit the tactile under the channel quality fluctuation. Although the haptic codec [228, 229], which is defined in IEEE 1918.1.1 standard, is designed to compress the tactile, e.g., vibrotactile signals, the cliff and leveling effects also occur owing to the channel quality fluctuation. We use the analog modulation with DCT and DWT for the vibrotactile signals to discuss the feasibility of the soft haptic delivery. Figure 10 (a) shows an overview of the proposed soft haptic delivery scheme and Fig. 10 (b) shows the reconstruction quality of the vibrotactile signal over the soft haptic and the digital-based delivery schemes at the channel SNRs of 10 dB and 20 dB under different available bandwidths. Here, the digital-based delivery scheme uses the BPSK modulation format and controls the quantization parameter to fit the transmission symbols into the available bandwidth whereas the soft haptic delivery scheme discards high-frequency coefficients for the same purpose. We used "1spike_Probe_-_aluminumGrid_-_fast" as the reference vibrotactile data provided by IEEE 1918.1.1 standard. From preliminary evaluations, the soft haptic delivery scheme yields better reconstruction quality of the vibrotactile signals irrespective of the available bandwidth.

The soft haptic delivery scheme was designed only for the single vibrotactile sensor and to minimize the MSE between the original and reconstructed vibrotactile signals. On the other hand, vibrotactile signals from multiple sensors should be delivered to provide immersive experiences for users. In addition, a psychohaptic model in [228, 229] demonstrated that each frequency band of vibrotactile signals has unequal sensitivity for humans. Here, how to design energy compression for vibrotactile signals from multiple sensors and power allocation to minimize the distortion considering human tactile perception are challenging issues for the realization of haptic delivery.

6   CONCLUSION

In this paper, we present an exhaustive survey and research outlook of the soft delivery schemes. We first review conventional digital-based video delivery schemes and the critical issues of the schemes, including cliff, leveling, and staircase effects. We then provide an overview of the soft delivery schemes and the taxonomy of the existing schemes from the perspectives of energy compaction, power allocation, bandwidth utilization, packet loss resilience, overhead reduction, and implementation. Some studies adopt the existing techniques of the energy compaction and overhead reduction to the soft delivery for immersive contents and find the reconstruction quality outperforms the digital-based delivery schemes even with the HEVC-based source coding. Finally, we envision the future directions of the soft delivery based on preliminary evaluations. We expect that soft delivery will be essential for sending high-quality model parameters and tactile information over wireless and mobile networks.

REFERENCES

[1] Cisco. Cisco Visual Networking Index: global mobile data traffic forecast update 2017–2022, March 2019.
[2] Dan Grois, Detlev Marpe, Amit Mulayoff, Benaya Itzhaky, and Ofer Hadar. Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders. In *IEEE Picture Coding Symposium*, pages 394–397, 2013.
[3] Sebastian Schwarz, Marius Preda, Vittorio Baroncini, Madhukar Budagavi, Pablo Cesar, Philip A. Chou, Robert A. Cohen, Maja Krivokuca, Sebastien Lassere, Zhu Li, Joan Llach, Khaled Mammou, Rufael Mekuria, Ohji Nakagami, Ernestasia Siahaan, Ali Tabatabai, Alexis M. Tourapis, and Vladyslav Zakharchenko. Emerging MPEG standards for

point cloud compression. *IEEE Journal of Emerging and Selected Topics in Circuits and Systems*, 9(1):133–148, 2019.

[4] Danillo Bracco Graziosi, Ohji Nakagami, Satoru Kuma, Alexandre Zaghetto, Teruhiko Suzuki, and Ali Tabatabai. An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing*, 9:1–17, 2020.

[5] Igor Kozintsev and Kannan Ramchandran. A wavelet zerotree-based hybrid compressed/uncompressed framework for wireless image transmission. In *Thirty-First Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1023–1027, 1997.

[6] Michael Gastpar, Martin Vetterli, and Pier Luigi Dragotti. Sensing reality and communicating bits: a dangerous liaison. *IEEE Signal Processing Magazine*, 23(4):70–83, 2006.

[7] Silvija Kokalj-Filipović, Emina Soljanin, and Yang Gao. Cliff effect suppression through multiple-descriptions with split personality. In *IEEE International Symposium on Information Theory*, pages 948–952, 2011.

[8] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1120, 2007.

[9] Abdelhak Bentaleb, Bayan Taani, Ali C. Begen, Christian Timmerer, and Roger Zimmermann. A survey on bitrate adaptation schemes for streaming media over HTTP. *IEEE Communications Surveys & Tutorials*, 21(1):562–585, 2019.

[10] Tao Wen, Wenxia Cai, Anhong Wang, Jie Liang, and Lijun Zhao. HDA video transmission scheme for DASH. *IEEE Access*, 8:58345–58356, 2020.

[11] Szymon Jakubczak and Dina Katabi. A cross-layer design for scalable mobile video. In *ACM Annual International Conference on Mobile Computing and Networking*, pages 289–300, 2011.

[12] Szymon Jakubczak, John Z Sun, Dina Katabi, and Vivek K Goyal. Performance regimes of uncoded linear communications over AWGN channels. In *45th Annual Conference on Information Sciences and Systems*, pages 1–6, 2011.

[13] Szymon Jakubczak and Dina Katabi. SoftCast: One-size-fits-all wireless video. *Computer Communication Review*, 40(4):449–450, 2010.

[14] Helge Coward. *Joint source-channel coding: Development of methods and utilization in image communications*. PhD thesis, Norwegian University of Science and Technology, 2002.

[15] Kyong-Hwa Lee and D. Petersen. Optimal linear coding for vector channels. *IEEE Transactions on Communications*, 24(12):1283–1290, 1976.

[16] Scott Pudlewski, Nan Cen, Zhangyu Guan, and Tommaso Melodia. Video transmission over lossy wireless networks: A cross-layer perspective. *IEEE Journal of Selected Topics in Signal Processing*, 9(1):6–21, 2015.

[17] Behrouz Jedari, Gopika Premsankar, Gazi Illahi, Mario Di Francesco, Abbas Mehrabi, and Antti Ylä-Jääski. Video caching, analytics, and delivery at the wireless edge: A survey and future directions. *IEEE Communications Surveys & Tutorials*, 23(1):431–471, 2020.

[18] Abid Yaqoob, Ting Bi, and Gabriel-Miro Muntean. A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. *IEEE Communications Surveys & Tutorials*, 22(4):2801–2838, 2020.

[19] Ching-Ling Fan, Wen-Chih Lo, Yu-Tung Pai, and Cheng-Hsin Hsu. A survey on 360 video streaming: Acquisition, transmission, and display. *ACM Computing Surveys (CSUR)*, 52(4):1–36, 2019.

[20] Zhi Liu, Qiyue Li, Xianfu Chen, Celimuge Wu, Susumu Ishihara, Jie Li, and Yusheng Ji. Point cloud video streaming: Challenges and solutions. *IEEE Network*, 35(5):202–209, 2021.

[21] Thomas Stockhammer, Hrvoje Jenkac, and Gabriel Kuhn. Streaming video over variable bit-rate wireless channels. *IEEE Transactions on Multimedia*, 6(2):268–277, 2004.

[22] Zhili Guo, Yao Wang, Elza Erkip, and Shivendra Panwar. Wireless video multicast with cooperative and incremental transmission of parity packets. *IEEE Transactions on Multimedia*, 17(8):1335–1346, 2015.

[23] Saleh Almowuena, Md. Mahfuzur Rahman, Cheng Hsin Hsu, Ahmad AbdAllah Hassan, and Mohamed Hafeeda. Energy-aware and bandwidth-efficient hybrid video streaming over mobile networks. *IEEE Transactions on Multimedia*, 18(1):102–115, 2016.

[24] Wiegand Thomas, Sullivan Gary J, Bjontegaard Gisle, and Luthra Ajay. Overview of the H.264/AVC video coding standard. *IEEE Transactions of Circuits And Systems for Video Technology*, 13(7):560–576, 2003.

[25] Maria Torres Vega, Cristian Perra, and Antonio Liotta. Resilience of video streaming services to network impairments. *IEEE Transactions on Broadcasting*, 64(2):220–234, 2018.

[26] Takumasa Ishioka, Kazuki Aiura, Ryota Shiina, Tatsuya Fukui, Tomohiro Taniguchi, Satoshi Narikawa, Katsuya Minami, Kazuhiro Kizaki, Takuya Fujihashi, Takashi Watanabe, and Shunsuke Saruwatari. Design and prototype implementation of software-defined radio over fiber. *IEEE Access*, 9:72793–72807, 2021.

[27] Qianqian Fan, David J Lilja, and Sachin S Sapatnekar. Adaptive-length coding of image data for low-cost approximate storage. *IEEE Transactions on Computers*, 69(2):239–252, 2020.

[28] Ying Li, Jun Wu, Bin Tan, Min Wang, and Wei Zhang. Compressive spinal codes. *IEEE Transactions on Vehicular Technology*, 68(12):11944–11954, 2019.

[29] Lingyu Liu, Jun Wu, and Jian Wu. COQRC: A rateless video transmission solution. In *International Conference on Computing, Networking and Communications*, pages 463–467, 2018.

[30] Lu Wang, Hailiang Yang, Xiaoke Qi, Jun Xu, and Kaishun Wu. ICast: fine-grained wireless video streaming over internet of intelligent vehicles. *IEEE Internet of Things Journal*, 6(1):111–123, 2019.

[31] Guanhua Wang, Kaishun Wu, Qian Zhang, and Lionel M. Ni. SimCast: Efficient video delivery in MU-MIMO WLANs. In *IEEE Conference on Computer Communications*, pages 2454–2462, 2014.

[32] Siripuram T Aditya and Sachin Katti. FlexCast: Graceful wireless video streaming. In *Annual International Conference on Mobile Computing and Networking*, pages 277–288, 2011.

[33] Tomas Kratochvil and Radim Stukavec. Hierarchical modulation in DVB-T/H mobile TV transmission over fading channels. In *International Symposium on Information Theory and Its Applications*, pages 1–6, 2008.

[34] Cornelius Hellge, Shpend Mirta, Thomas Schierl, and Thomas Wiegand. Mobile TV with SVC and hierarchical modulation for DVB-H broadcast services. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–5, 06 2009.

[35] Mahdi Ghandi and Mohammed Ghanbari. Layered H.264 video transmission with hierarchical QAM. *Journal of Visual Communication and Image Representation*, 17(2):451–466, 2006.

[36] Dongweon Yoon, Kyongkuk Cho, and Jinsock Lee. Bit error probability of M-ary quadrature amplitude modulation. In *Vehicular Technology Conference*, pages 2422–2427, 2000.

[37] Xiph. Xiph.org media.

[38] Anthony Trioux, Giuseppe Valenzise, Marco Cagnazzo, Michel Kieffer, François-Xavier Coudoux, Patrick Corlay, and Mohamed Gharbi. Subjective and objective quality assessment of the SoftCast video transmission scheme. In *IEEE International Conference on Visual Communications and Image Processing*, pages 96–99, 2020.

[39] Anthony Trioux, François Xavier Coudoux, Patrick Corlay, and Mohamed Gharbi. A comparative preprocessing study for SoftCast video transmission. In *International Symposium on Signal, Image, Video and Communications*, pages 54–59, 2018.

[40] Vinod Prabhakaran, Rohit Puri, and Kannan Ramchandran. Hybrid digital-analog codes for source-channel broadcast of Gaussian sources over Gaussian channels. *IEEE Transactions on Information Theory*, 57(7):4573–4588, 2011.

[41] Hao Cui, Ruiqin Xiong, Chong Luo, Zhihai Song, and Feng Wu. Denoising and resource allocation in uncoded video transmission. *IEEE Journal on Selected Topics in Signal Processing*, 9(1):102–112, 2015.

[42] Quan Wang, Xiaocheng Lin, Yu Liu, Lin Zhang, and Xiaofei Wu. A scalable mobile video broadcast scheme using 3D wavelet transform. In *IEEE Vehicular Technology Conference*, pages 3–7, 2014.

[43] Quan Wang, Xiaocheng Lin, Yu Liu, Lin Zhang, and Xiaofei Wu. A scalable framework for mobile video broadcast using MCTF and 2D-DWT. In *International Symposium on Wireless Personal Multimedia Communications*, pages 118–123, 2015.

[44] Xiaopeng Fan, Ruiqin Xiong, Feng Wu, and Debin Zhao. WaveCast: Wavelet based wireless video broadcast using lossy transmission. In *IEEE Visual Communications and Image Processing*, pages 1–6, 2012.

[45] Hao Cui, Zhihai Song, Zhe Yang, Chong Luo, Ruiqin Xiong, and Feng Wu. Cactus: A hybrid digital-analog wireless video communication system. In *ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems*, pages 273–278, 2013.

[46] Xiao Zhao, Hancheng Lu, Chang Wen Chen, and Jun Wu. Adaptive hybrid digital-analog video transmission in wireless fading channel. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(6):1117–1130, 2016.

[47] Ruiqin Xiong, Feng Wu, Jizheng Xu, Xiaopeng Fan, Chong Luo, and Wen Gao. Analysis of decorrelation transform gain for uncoded wireless image and video communication. *IEEE Transactions on Image Processing*, 25(4):1820–1833, 2016.

[48] Ruiqin Xiong, Feng Wu, Jizheng Xu, and Wen Gao. Performance analysis of transform in uncoded wireless visual communication. In *IEEE International Symposium on Circuits and Systems*, pages 1159–1162, 2013.

[49] Anthony Trioux, François Xavier Coudoux, Patrick Corlay, and Mohamed Gharbi. Temporal information based GoP adaptation for linear video delivery schemes. *Signal Processing: Image Communication*, 82:1–17, 2020.

[50] Xiaocheng Lin, Nianfei Fan, Yu Liu, Shuri Cai, and Xiaojing Wang. Soft wireless image/video broadcast based on component protection. In *IEEE International Conference on Network Infrastructure and Digital Content*, pages 84–89, 2014.

[51] Chenfeng He, Huachan Qin, Zhiqiang He, and Kai Niu. Adaptive GoP dividing video coding for wireless broadcast based on power allocation optimization. In *International Conference on Wireless Communications and Signal Processing*, pages 1–5, 2016.

[52] Anthony Trioux, Francois Xavier Coudoux, Patrick Corlay, and Mohamed Gharbi. A reduced complexity/side information preprocessing method for high quality Softcast-based video delivery. In *European Workshop on Visual Information Processing*, pages 205–210, 2019.

[53] Zhihai Song, Ruiqin Xiong, Siwei Ma, Xiaopeng Fan, and Wen Gao. Layered image/video SoftCast with hybrid digital-analog transmission for robust wireless visual communication. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2014.

[54] Jing Zhao, Jiyu Xie, and Ruiqin Xiong. Residual signals modeling for layered image/video Softcast with hybrid digital-analog transmission. *IEEE International Conference on Image Processing*, pages 3284–3288, 2018.

[55] Dongliang He, Chong Luo, Cuiling Lan, Feng Wu, and Wenjun Zeng. Structure-preserving hybrid digital-analog video delivery in wireless networks. *IEEE Transactions on Multimedia*, 17(9):1658–1670, 2015.

[56] Yuanyuan Li, Yu Liu, Yumei Wang, and Zhexin Li. Visual information exploited hybrid digital-analog scheme for wireless video multicast. In *Visual Communications and Image Processing*, pages 1–4, 2016.

[57] Ruiqin Xiong, Hangfan Liu, Siwei Ma, Xiaopeng Fan, Feng Wu, and Wen Gao. G-CAST: gradient based image SoftCast for perception-friendly wireless visual communication. In *Data Compression Conference*, pages 133–142, 2014.

[58] Hangfan Liu, Ruiqin Xiong, Xiaopeng Fan, Debin Zhao, Yongbing Zhang, and Wen Gao. CG-cast: scalable wireless image softcast using compressive gradient. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(6):1832–1843, 2019.

[59] A. Sehgal, A. Jagmohan, and N. Ahuja. Wyner-Ziv coding of video: an error-resilient compression framework. *IEEE Transactions on Multimedia*, 6(2):249–258, 2004.

[60] S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS): design and construction. *IEEE Transactions on Information Theory*, 49(3):626–643, 2003.

[61] Ailing Zhang, Xiaopeng Fan, Ruiqin Xiong, and Debin Zhao. Distributed soft video broadcast with variable block size motion estimation. In *IEEE International Conference on Visual Communications and Image Processing*, pages 1–5, 2013.

[62] Xiaopeng Fan, Feng Wu, Debin Zhao, Oscar C Au, and Wen Gao. Distributed soft video broadcast (DCAST) with explicit motion. In *Data Compression Conference Proceedings*, pages 199–208, 2012.

[63] Xiaopeng Fan, Feng Wu, and Debin Zhao. D-cast: DSC based soft mobile video broadcast. In *International Conference on Mobile and Ubiquitous Multimedia*, pages 226–235, 2011.

[64] Xiaopeng Fan, Feng Wu, Debin Zhao, and Oscar C. Au. Distributed wireless visual communication with power distortion optimization. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(6):1040–1053, 2013.

[65] Wei Huang, Xiaopeng Fan, and Debin Zhao. Soft mobile video broadcast based on side information refining. In *IEEE International Conference on Visual Communications and Image Processing*, 2013.

[66] Ricardo Martins, Catarina Brites, João Ascenso, and Fernando Pereira. Refining side information for improved transform domain Wyner-Ziv video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(9):1327–1341, 2009.

[67] Mengyang Lv, Yu Liu, and Yumei Wang. Scalable wireless video broadcast based on unequal protection. In *Visual Communication and Image Processing*, pages 1–4, 2017.

[68] Xiaocheng Lin, Yu Liu, and Lin Zhang. Scalable video SoftCast using magnitude shift. In *IEEE Wireless Communications and Networking Conference*, pages 1996–2001, 2015.

[69] Xiaopeng Fan, Ruiqin Xiong, Debin Zhao, and Feng Wu. Layered soft video broadcast for heterogeneous receivers. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(11):1801–1814, 2015.

[70] Mengyang Lv, Yu Liu, and Yumei Wang. Adaptive scalable wireless video coding based on unequal protection and quadtree partition. In *International Conference on Network Infrastructure and Digital Content*, volume 9, pages 214–218, 2017.

[71] Ahmed Hagag, Xiaopeng Fan, and Fathi E. Abd El-Samie. Hyperspectral image coding and transmission scheme based on wavelet transform and distributed source coding. *Multimedia Tools and Applications*, 76(22):23757–23776, 2017.

[72] Jian Shen, Fei Liang, Chong Luo, Houqiang Li, and Wenjun Zeng. Cooperative hybrid digital-analog video transmission in D2D networks. *IEEE International Conference on Image Processing*, pages 3274–3278, 2018.

[73] Yumei Wang, Mengyao Sun, and Yu Liu. Distributed and adaptive analog coding for video broadcast in wireless cooperative system. *Wireless Personal Communications*, 102(3):2287–2306, 2018.

[74] Mengyao Sun, Yumei Wang, Hao Yu, and Yu Liu. Distributed cooperative video coding for wireless video broadcast system. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2015.

[75] Hao Cui, Dian Liu, Yuqi Han, and Jun Wu. Robust uncoded video transmission under practical channel estimation. In *IEEE Global Communications Conference*, pages 1–6, 2016.

[76] Fan Zhang, Anhong Wang, Haidong Wang, Suyue Li, and Xiaoli Ma. Channel-aware video SoftCast scheme. In *IEEE China Summit and International Conference on Signal and Information Processing*, pages 578–581, 2015.

[77] Zhilong Zhang, Danpu Liu, and Xin Wang. Joint carrier matching and power allocation for wireless video with general distortion measure. *IEEE Transactions on Mobile Computing*, 17(3):577–589, 2018.

[78] Jian Wu, Bin Tan, Jun Wu, and Rui Wang. Efficient soft video MIMO design to combine diversity and spatial multiplexing gain. *IEEE Internet of Things Journal*, 6(3):5461–5472, 2019.

[79] S Zheng, M Cagnazzo, and M Kieffer. Precoding matrix design in linear video coding. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1198–1202, 2018.

[80] S. Zheng, M. Antonini, M. Cagnazzo, L. Guerrieri, M. Kieffer, I. Nemoianu, R. Samy, and B. Zhang. Softcast with per-carrier power-constrained channels. In *IEEE International Conference on Image Processing*, pages 2122–2126, 2016.

[81] Xiao Lin Liu, Wenjun Hu, Qifan Pu, Feng Wu, and Yongguang Zhang. ParCast: soft video delivery in MIMO-OFDM WLANs. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 233–244, 2012.

[82] Xiao Lin Liu, Wenjun Hu, Chong Luo, Qifan Pu, Feng Wu, and Yongguang Zhang. ParCast+: Parallel video unicast in MIMO-OFDM WLANs. *IEEE Transactions on Multimedia*, 16(7):2038–2051, 2014.

[83] Zhilong Zhang, Danpu Liu, Xiaoli Ma, and Xin Wang. ECast: an enhanced video transmission design for wireless multicast systems over fading channels. *IEEE Systems Journal*, 11(4):2566–2577, 2015.

[84] Shuo Zheng, Marco Cagnazzo, and Michel Kieffer. Channel Impulsive Noise Mitigation for Linear Video Coding Schemes. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9):3196–3209, 2020.

[85] Xiaoda Jiang, Hancheng Lu, Chang Wen Chen, and Feng Wu. Receiver-driven video multicast over NOMA systems in heterogeneous environments. In *IEEE International Conference on Computer Communications*, pages 982–990. IEEE, 2019.

[86] Jian Wu, Bin Tan, Jun Wu, and Min Wang. Video multicast: Integrating scalability of soft video delivery systems into NOMA. *IEEE Wireless Communications Letters*, 8(6):1722–1726, 2019.

[87] Rongxin Zhang, Yiming Kong, Xiaoli Ma, and Deqing Wang. Adaptive video transmission designs over underwater acoustic channels. In *International Conference on Computing, Networking and Communications*, pages 295–299, 2018.

[88] Xiao-wei Tang, Xin-lin Huang, and Fei Hu. QoE-driven UAV-enabled pseudo-analog wireless video broadcast: A joint optimization of power and trajectory. *IEEE Transactions on Multimedia*, 23:2398–2412, 2021.

[89] Yongqiang Gui, Lu Hancheng, Feng Wu, and Chang Wen Chen. LensCast: Robust wireless video transmission over mmWave MIMO with lens antenna array. *IEEE Transactions on Multimedia*, 24:33–48, 2020.

[90] Jing Zhao, Ruiqin Xiong, Chong Luo, Feng Wu, and Wen Gao. Wireless image and video soft transmission via perception-inspired power distortion optimization. In *IEEE Visual Communications and Image Processing*, pages 1–4, 2018.

[91] Jian Shen, Lei Yu, Li Li, and Houqiang Li. Foveation-based wireless soft image delivery. *IEEE Transactions on Multimedia*, 20(10):2788–2800, 2018.

[92] Hadi Hadizadeh. Saliency-guided wireless transmission of still images using SoftCast. In *International Symposium on Telecommunications*, pages 506–509, 2017.

[93] Yuanyuan Li, Zhexin Li, Yu Liu, and Yumei Wang. SCAST: wireless video multicast scheme based on segmentation and Softcast. In *IEEE Wireless Communications and Networking Conference*, pages 1–6, 2017.

[94] Zhou Wang and Alan Conrad Bovik. Embedded foveation image coding. *IEEE Transactions on Image Processing*, 10(10):1397–1410, 2001.

[95] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[96] Zhexin Li, Yu Liu, and Yumei Wang. Unequal block for low bandwidth adaption in wireless video broadcast. In *International Conference on Network Infrastructure and Digital Content*, pages 386–390, 2017.

[97] Marco Cagnazzo and Michel Kieffer. Shannon-Kotelnikov mappings for softcast-based joint source-channel video coding. In *IEEE International Conference on Image Processing*, pages 1085–1089, 2015.

[98] Yongqiang Gui, Hancheng Lu, Xiaoda Jiang, Feng Wu, and Chang Wen Chen. Compressed pseudo-analog transmission system for remote sensing images over bandwidth-constrained wireless channels. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9):3181–3195, 2020.

[99] Yali Wang, Hancheng Lu, Zexue Li, and Jian Li. Robust satellite image transmission over bandwidth-constrained wireless channels. In *IEEE International Conference on Communications*, pages 1–6, 2017.

[100] Ahmad Shoja Yami and Hadi Hadizadeh. Visual attention-driven wireless multicasting of images using adaptive compressed sensing. In *Artificial Intelligence and Signal Processing*, pages 37–42, oct 2018.

[101] Hadi Hadizadeh and Ivan V. Bajic. Soft video multicasting using adaptive compressed sensing. *IEEE Transactions on Multimedia*, 23:12–25, 2021.

[102] Wenbin Yin, Xiaopeng Fan, Yunhui Shi, Ruiqin Xiong, and Debin Zhao. Compressive sensing based soft video broadcast using spatial and temporal sparsity. *Mobile Networks and Applications*, 21(6):1002–1012, 2016.

[103] Tze Yang Tung and Deniz Gunduz. SparseCast: hybrid digital-analog wireless image transmission exploiting frequency-domain sparsity. *IEEE Communications Letters*, 22(12):2451–2454, 2018.

[104] Siyuan Liu, Kai Niu, and Chao Dong. Channel polarization based block compressive sensing SoftCast system. In *IEEE International Conference on Computer and Communications*, pages 778–783, 2019.

[105] Georgios Angelopoulos, Muriel Medard, and Anantha P. Chandrakasan. AdaptCast: an integrated source to transmission scheme for wireless sensor networks. In *IEEE International Conference on Communications*, pages 2894–2899, 2015.

[106] Jun Wu, Jian Wu, Hao Cui, Chong Luo, Xiaoyan Sun, and Feng Wu. DAC-Mobi: Data-assisted communications of mobile images with cloud computing support. *IEEE Transactions on Multimedia*, 18(5):893–904, 2016.

[107] Jun Wu, Dian Liu, Xin Lin Huang, Chong Luo, Hao Cui, and Feng Wu. DaC-RAN: A data-assisted cloud radio access network for visual communications. *IEEE Wireless Communications*, 22(3):130–136, 2015.

[108] Xin Lin Huang, Jun Wu, and Fei Hu. Knowledge-enhanced mobile video broadcasting framework with cloud support. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(1):6–18, 2017.

[109] Xin Lin Huang, Xiaoning Huan, Jun Wu, Qingquan Sun, and Yingchun Yuan. Performance analysis of KMV-cast with imperfect prior knowledge. In *IEEE Global Communications Conference*, pages 1–5, 2016.

[110] Xin Lin Huang, Xiaowei Tang, Xiaoning Huan, Ping Wang, and Jun Wu. Improved KMV-Cast with BM3D denoising. *Mobile Networks and Applications*, 23(1):100–107, 2018.

[111] Dian Liu, Jun Wu, Hao Cui, Dongdong Zhang, Chong Luo, and Feng Wu. Cost-distortion optimization and resource control in pseudo-analog visual communications. *IEEE Transactions on Multimedia*, 20(11):3097–3110, 2018.

[112] Chaofan He, Yang Hu, Yan Chen, Xiaopeng Fan, Houqiang Li, and Bing Zeng. MUcast: Linear uncoded multiuser video streaming with channel assignment and power allocation optimization. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(4):1136–1146, 2020.

[113] Chaofan He, Yang Hu, Yan Chen, Xiaopeng Fan, Houqiang Li, and Bing Zeng. Exploiting channel assignment and power allocation for linear uncoded multiuser video streaming. In *IEEE International Conference on Communications*, 2019.

[114] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[115] Emmanuel J. Candes and Michael B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[116] Xiaocheng Lin, Yu Liu, and Mengyao Sun. Analog channel coding for wireless image/video SoftCast by data division. In *International Conference on Telecommunications*, pages 353–357, 2015.

[117] Bin Tan, Jun Wu, Ying Li, Hao Cui, Wei Yu, and Chang Wen Chen. Analog Coded SoftCast: A network slice design for multimedia broadcast/multicast. *IEEE Transactions on Multimedia*, 19(10):2293–2306, 2017.

[118] Yang Liu, Jing li, Xuanxuan Lu, Chau Yuen, and Jun Wu. A family of chaotic pure analog coding schemes based on baker's map function. *EURASIP Journal on Advances in Signal Processing*, 2015, 2015.

[119] Chaofan He, Huiying Wang, Yang Hu, Yan Chen, Xiaopeng Fan, Houqiang Li, and Bing Zeng. MCast: high-quality linear video transmission with time and frequency diversities. *IEEE Transactions on Image Processing*, 27(7):3599–3610, 2018.

[120] Cuiling Lan, Dongliang He, Chong Luo, Feng Wu, and Wenjun Zeng. Progressive pseudo-analog transmission for mobile video live streaming. In *IEEE Visual Communications and Image Processing*, pages 1–4, Singapore, dec 2015.

[121] Dongliang He, Cuiling Lan, Chong Luo, Enhong Chen, Feng Wu, and Wenjun Zeng. Progressive pseudo-analog transmission for mobile video streaming. *IEEE Transactions on Multimedia*, 19(8):1894–1907, 2017.

[122] Anhong Wang, Bing Zeng, and Hua Chen. Wireless multicasting of video signals based on distributed compressed sensing. *Signal Processing: Image Communication*, 29(5):599–606, 2014.

[123] Anhong Wang, Qingdian Wu, Xiaoli Ma, and Bing Zeng. A wireless video multicasting scheme based on multi-scale compressed sensing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):1–11, 2015.

[124] Shanshan Liu, Anhong Wang, Haidong Wang, Suyue Li, Meiling Li, and Jie Liang. Adaptive residual-based distributed compressed sensing for soft video multicasting over wireless networks. *Multimedia Tools and Applications*, 76(14):15587–15606, 2017.

[125] Sungkwang Mun and James E. Fowler. Block compressed sensing of images using directional transforms. In *IEEE International Conference on Image Processing*, pages 3021–3024, 2009.

[126] Zhihai Song, Ruiqin Xiong, Xiaopeng Fan, Siwei Ma, and Wen Gao. Transform domain energy modeling of natural images for wireless SoftCast optimization. In *IEEE International Symposium on Circuits and Systems*, pages 1114–1117, 2014.

[127] Ruiqin Xiong, Jian Zhang, Feng Wu, Jizheng Xu, and Wen Gao. Power distortion optimization for uncoded linear transformed transmission of images and videos. *IEEE Transactions on Image Processing*, 26(1):222–236, 2017.

[128] Ruiqin Xiong, Feng Wu, Xiaopeng Fan, Chong Luo, Siwei Ma, and Wen Gao. Power-distortion optimization for wireless image/video SoftCast by transform coefficients energy modeling with adaptive chunk division. In *IEEE International Conference on Visual Communications and Image Processing*, pages 1–6, 2013.

[129] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip V. Orlik. High-quality soft video delivery with GMRF-based overhead reduction. *IEEE Transactions on Multimedia*, 20(2):473–483, 2018.

[130] Jingyu Li, Xiang E. Wen, Huizhu Jia, Xiaodong Xie, and Wen Gao. AnalogCast: Full linear coding and pseudo analog transmission for satellite remote-sensing images. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1362–1366, 2016.

[131] Ticao Zhang and Shiwen Mao. Metadata reduction for soft video delivery. *IEEE Networking Letters*, 1(2):84–88, 2019.

[132] Yongqiang Gui, Hancheng Lu, Feng Wu, and Chang Wen Chen. Robust video broadcast for users with heterogeneous resolution in mobile networks. *IEEE Transactions on Mobile Computing*, 20(11):3251–3266, 2021.

[133] Xiao Lin, Wenjun Hu, Chong Luo, and Feng Wu. Compressive image broadcasting in MIMO systems with receiver antenna heterogeneity. *Signal Processing: Image Communication*, 29(3):361–374, 2014.

[134] Hao Cui, Chong Luo, Chang Wen Chen, and Feng Wu. Scalable video multicast for MU-MIMO systems with antenna heterogeneity. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(5):992–1003, 2016.

[135] Xiao Wei Tang and Xin Lin Huang. A design of SDR-based pseudo-analog wireless video transmission system. *Mobile Networks and Applications*, 25(6):2495–2505, 2020.

[136] Shi Chen, Jun Wu, Haoqi Ren, Jian Wu, Baoye Zhang, and Fusheng Zhu. Hardware implementation of a pseudo-analog wireless video transmission system. In *IEEE International Conference on Communication Technology Proceedings*, pages 519–524, 2019.

[137] Fengxiang Gao, Haoqi Gao, and Jun Wu. A reconfigurable SoC for SoftCast wireless video transmission. In *IEEE International Conference on Industrial Internet*, pages 169–170. IEEE, 2018.

[138] Yao Jiang, Pengfei Xia, Jun Wu, Shi Chen, and Baoye Zhang. Pseudo-analog wireless stereo video transmission in hardware acceleration. In *International Conference on Wireless Communications and Signal Processing*, pages 1–6, 2017.

[139] Zhe Chen, Xu Zhang, Sulei Wang, Yuedong Xu, Jie Xiong, and Xin Wang. Enabling practical large-scale MIMO in WLANs with hybrid beamforming. *IEEE/ACM Transactions on Networking*, 29(4):1605–1619, 2021.

[140] Zhuang Ding, Jun Wu, Wei Yu, Yuqi Han, and Xianghuang Chen. Pseudo analog video transmission based on LTE physical layer. In *IEEE/CIC International Conference on Communications in China*, pages 1–6, 2016.

[141] Kun Tan, He Liu, Jiansong Zhang, Yongguang Zhang, Ji Fang, and Geoffrey Voelker. Sora: high-performance software radio using general-purpose multi-core processors. *Communications of the ACM*, 54(1):99–107, 2011.

[142] Zhe Chen, Xu Zhang, Yuedong Xu, Jie Xiong, Yu Zhu, and Xin Wang. MuVi: Multiview video aware transmission over mimo wireless systems. *IEEE Transactions on Multimedia*, 19(12):2788–2803, 2017.

[143] Masayuki Tanimoto. FTV:free-viewpoint television. *Signal Processing: Image Communication*, 27(6):555–570, 2012.

[144] Olgierd Stankiewicz, Marek Domanski, Adrian Dziembowski, Adam Grzelka, Dawid Mieloch, and Jarosaw Samelak. A free-viewpoint television system for horizontal virtual navigation. *IEEE Transactions on Multimedia*, 20(9):2182–2195, August 2018.

[145] Rufael Mekuria and Lazar Bivolarsky. Overview of the MPEG activity on point cloud compression. In *Data Compression Conference*, page 620, 2016.

[146] Christoph Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Stereoscopic Displays and Virtual Reality Systems*, volume 5291, pages 93–105, 2004.

[147] Shuai Li, Ce Zhu, and Ming-Ting Sun. Hole filling with multiple reference views in DIBR view synthesis. *IEEE Transactions on Multimedia*, 20(8):1948–1959, 2018.

[148] Ying Chen, Miska M. Hannuksela, Teruhiko Suzuki, and Shinobu Hattori. Overview of the MVC+D 3D video coding standard. *Journal of Visual Communciation and Image Representation*, 25(4):679–688, may 2014.

[149] Ana De Abreu, Pascal Frossard, and Fernando Pereira. Optimizing multiview video plus depth prediction structures for interactive multiview video streaming. *IEEE Journal of Selected Topics in Signal Processing*, 9(3):487–500, 2015.

[150] Ying Chen and Sehoon Yen. 3D-AVC draft text 6, document JCT3V-D1002.doc, April 2013.

[151] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip V. Orlik. Soft video delivery for free viewpoint video. In *IEEE International Conference on Communications*, pages 1–7, 2017.

[152] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip V. Orlik. FreeCast: Graceful free-viewpoint video delivery. *IEEE Transactions on Multimedia*, 21(4):1000–1010, 2019.

[153] Ticao Zhang and Shiwen Mao. Joint Power and Channel Resource Optimization in Soft Multi-View Video Delivery. *IEEE Access*, 7:148084–148097, 2019.

[154] Lei Luo, Taihai Yang, Ce Zhu, Zhi Jin, and Shu Tang. Joint texture/depth power allocation for 3-D video SoftCast. *IEEE Transactions on Multimedia*, 21(12):2973–2984, 2019.

[155] Taihai Yang, Lei Luo, Ce Zhu, and Shu Tang. Block DCT based optimization for wireless SoftCast of depth map. *IEEE Access*, 7:29484–29494, 2019.

[156] Feng Qian, Bo Han, Lusheng Ji, and Vijay Gopalakrishnan. Optimizing 360 video delivery over cellular networks. In *5th Workshop on All Things Cellular: Operations, Applications and Challenges*, pages 1–6, 2016.

[157] Jason Moss and Eric Muth. Characteristics of head-mounted displays and their effects on simulator sickness. *The Journal of the Human Factors and Ergonomics Society*, 53(3):308–319, 2011.

[158] Daisuke Ochi, Yutaka Kunita, Akio Kameda, Akira Kojima, and Shinnosuke Iwaki. Live streaming system for omnidirectional video. In *IEEE Virtual Reality*, pages 349–350, 2015.

[159] Takuya Fujihashi, Makoto Kobavashi, Keiichi Endo, Shunsuke Saruwatari, Shinya Kobayashi, and Takashi Watanabe. Graceful quality improvement in wireless 360-degree video delivery. In *IEEE Global Communications Conference*, pages 1–7, 2018.

[160] Jing Zhao, Ruiqin Xiong, and Jizheng Xu. OmniCast: Wireless pseudo-analog transmission for omnidirectional video. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):58–70, mar 2019.

[161] Yujun Lu, Takuya Fujihashi, Shunsuke Saruwatari, and Takashi Watanabe. 360Cast: Foveation-based wireless soft delivery for 360-degree video. In *IEEE International Conference on Communications*, pages 1–6, jun 2020.

[162] Lu Yujun, Takuya Fujihashi, Shunsuke Saruwatari, and Takashi Watanabe. 360Cast+: Viewport adaptive soft delivery for 360-degree videos. *IEEE Access*, 9:52684–52697, 2021.

[163] Yule Sun, Ang Lu, and Lu Yu. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE signal processing letters*, 24(9):1408–1412, 2017.

[164] Heinrich Hertz Institute Fraunhofer Institute for Telecommunications. High efficiency video coding (hevc) reference software hm. https://hevc.hhi.fraunhofer.de/.

[165] P. A. Blanche, A. Bablumian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W. Y. Hsieh, M. Kathaperumal, B. Rachwal, O. Siddiqui, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian. Holographic three-dimensional telepresence using large-area photorefractive polymer. *Nature*, 468(7320):80–83, 2010.

[166] Hyeonseung Yu, KyeoReh Lee, Jongchan Park, and YongKeun Park. Ultrahigh-definition dynamic 3D holographic display by active control of volume speckle fields. *Nature Photonics*, 11(3):186–192, 2017.

[167] Draco 3D data compression, 2022.

[168] Olivier Devillers and Pierre-Marie Gandoin. Geometric compression for interactive transmission. In *Visualization*, pages 319–326, 2000.

[169] Julius Kammerl, Nico Blodow, Radu Bogdan Rusu, Suat Gedikli, Michael Beetz, and Eckehard Steinbach. Real-time compression of point cloud streams. In *IEEE International Conference on Robotics and Automation*, pages 778–785, 2012.

[170] Karsten Muller, Heiko Schwarz, Detlev Marpe, Christian Bartnik, Sebastian Bosse, Heribert Brust, Tobias Hinz, Haricharan Lakshman, Philipp Merkle, Franz Hunn Rhee, Gerhard Tech, Martin Winken, and Thomas Wiegand. 3D is here: Point cloud library (PCL). In *IEEE International Conference on Robotics and Automation*, pages 1–4, 2011.

[171] Ruwen schnabel and Reinhard Klein. Octree-based point-cloud compression. In *Eurographics Symposium on Point-Based Graphics*, pages 111–121, 2006.

[172] De Queiroz, Ricardo L., and Philip A. Chou. Compression of 3D point clouds using a region-adaptive hierarchical transform. *IEEE Transactions on Image Processing*, 25(8):3947–3956, aug 2016.

[173] C. Zhang, D. Florêncio, and C. Loop. Point cloud attribute compression with graph transform. In *IEEE International Conference on Image Processing*, pages 2066–2070, 2014.

[174] Paulo de Oliveira Rente, Catarina Brites, Jao Ascenso, and Fernando Pereira. Graph-based static 3D point clouds geometry coding. *IEEE Transactions on Multimedia*, 21(2):284–299, 2019.

[175] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip Orlik. HoloCast: Graph signal processing for graceful point cloud delivery. In *IEEE International Conference on Communications*, pages 1–7, 2019.

[176] Cha Zhang, Dinei Florencio, and Charles Loop. Point cloud attribute compression with graph transform. *IEEE International Conference on Image Processing*, pages 2066–2070, 2014.

[177] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip Orlik. Overhead reduction in graph-based point cloud delivery. In *IEEE International Conference on Communications*, pages 1–7, 2020.

[178] Soushi Ueno, Takuya Fujihashi, Toshiaki Koike-Akino, and Takashi Watanabe. Overhead reduction for graph-based point cloud delivery using non-uniform quantization. In *IEEE International Conference on Consumer Electronics*, pages 1–6, 2022.

[179] Soushi Ueno, Takuya Fujihashi, Toshiaki Koike-Akino, and Takashi Watanabe. Point cloud soft multicast for untethered xr users. *IEEE Transactions on Multimedia*, PP(99):1–12, 2022.

[180] Mehdi Ansari Sadrabadi, Amir Khandani, and Farshad Lahouti. Channel feedback quantization for high data rate MIMO systems. *IEEE Transactions on Wireless Communications*, 5(12):3335–3338, 2006.

[181] June Chul Roh and Bhaskar D. Rao. Efficient feedback methods for MIMO channels based on parameterization. *IEEE Transactions on Wireless Communications*, 6(1):282–292, 2007.

[182] Lei Yu, Houqiang Li, and Weiping Li. Hybrid digital-analog scheme for video transmission over wireless. In *IEEE International Symposium on Circuits and Systems*, pages 1163–1166, 2013.

[183] Lei Yu, Houquiang Li, and Weiping Li. Wireless scalable video coding using a hybrid digital-analog scheme. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(2):331–345, 2014.

[184] Cuiling Lan, Chong Luo, Wenjun Zeng, and Feng Wu. A practical hybrid digital-analog scheme for wireless video transmission. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(7):1634–1647, 2018.

[185] Zhihai Song, Ruiqin Xiong, Siwei Ma, and Wen Gao. Hybridcast: A wireless image/video Softcast scheme using layered representation and hybrid digital-analog modulation. In *IEEE International Conference on Image Processing*, pages 6001–6005, 2014.

[186] Bin Tan, Jun Wu, Rui Wang, Wenlang Luo, and Jun Liu. An optimal resource allocation for hybrid digital-analog with combined multiplexing. *IEEE Internet of Things Journal*, 6(1):1125–1135, 2019.

[187] Bin Tan, Hao Cui, Jun Wu, and Chang Wen Chen. An optimal resource allocation for superposition coding-based hybrid digital–analog system. *IEEE Internet of Things Journal*, 4(4):945–956, 2017.

[188] Jing Zhang, Anhong Wang, Jie Liang, Haidong Wang, Suyue Li, and Xiong Zhang. Distortion estimation-based adaptive power allocation for hybrid digital-analog video transmission. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(6):1806–1818, 2019.

[189] Fei Liang, Chong Luo, Ruiqin Xiong, Wenjun Zeng, and Feng Wu. Superimposed modulation for soft video delivery with hidden resources. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2345–2358, 2018.

[190] Jian Shen, Lei Yu, and Houqiang Li. Hybrid digital-analog scheme for video transmission over fading channel. In *IEEE International Symposium on Circuits and Systems*, pages 1582–1585, 2016.

[191] Pradeepa Yahampath. Hybrid digital-analog coding with bandwidth expansion for correlated Gaussian sources under Rayleigh fading. *EURASIP Journal on Advances in Signal Processing*, 2017(1):1–16, 2017.

[192] Pradeepa Yahampath. Digital-analog superposition coding for OFDM channels with application to video transmission. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1802–1806, 2018.

[193] Pradeepa Yahampath. Video coding for OFDM systems with imperfect CSI: A hybrid digital–analog approach. *Signal Processing: Image Communication*, 87:1–22, 2020.

[194] Yu Liu, Xiaocheng Lin, Nianfei Fan, and Lin Zhang. Hybrid digital-analog video transmission in wireless multicast and multiple-input multiple-output system. *Journal of Electronic Imaging*, 25(1):1–14, 2016.

[195] Lei Yu, Houquiang Li, and Weiping Li. Wireless cooperative video coding using a hybrid digital-analog scheme. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3):436–450, 2015.

[196] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip V. Orlik. Compressive sensing for loss-resilient hybrid wireless video transmission. In *IEEE Global Communications Conference*, pages 1–5, 2015.

[197] Dongliang He, Chong Luo, Feng Wu, and Wenjun Zeng. Swift: A hybrid digital-analog scheme for low-delay transmission of mobile stereo video. In *ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, pages 327–336, 2015.

[198] Peihao Li, Fengbao Yang, Jing Zhang, Yun Guan, Anhong Wang, and Jie Liang. Synthesis-distortion-aware hybrid digital analog transmission for 3D videos. *IEEE Access*, 8:85128–85139, 2020.

[199] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip V. Orlik. HoloCast+: hybrid digital-analog transmission for graceful point cloud delivery with graph fourier transform. *IEEE Transactions on Multimedia*, 24:2179–2191, 2021.

[200] Guo Lu, Xiaoyun Zhang, Wanli Ouyang, Li Chen, Zhiyong Gao, and Dong Xu. An end-to-end learning framework for video compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3292–3308, 2021.

[201] Oren Rippel, Alexander G. Anderson, K. Tatwawadi, S. Nair, C. Lytle, and L. Bourdev. ELF-VC: efficient learned flexible-rate video coding. In *IEEE/CVF International Conference on Computer Vision*, pages 14459–14468, 2021.

[202] Yunjin Chen and Thomas Pock. Trainable Nonlinear Reaction Diffusion: a flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1256–1272, 2017.

[203] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. MemNet: a persistent memory network for image restoration. In *IEEE International Conference on Computer Vision*, pages 4549–4557, 2018.

[204] Takuya Fujihashi, Toshiaki Koike-Akino, Philip V. Orlik, and Takashi Watanabe. DNN-based overhead reduction for high-quality soft delivery. In *IEEE Global Communications Conference*, pages 1–6, 2019.

[205] Eirina Bourtsoulatze, David Burth Kurka, and Deniz Gunduz. Deep joint source-channel coding for wireless image transmission. *IEEE Transactions on Cognitive Communications and Networking*, 5(3):567–579, 2019.

[206] Xiao-Wei Tang, Xin-Lin Huang, Fei Hu, and Qingjiang Shi. Human-perception-oriented pseudo analog video transmissions with deep learning. *IEEE Transactions on Vehicular Technology*, 69(9):9896–9909, 2020.

[207] Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe, and Philip V Orlik. High-quality soft image delivery with deep image denoising. In *IEEE International Conference on Communications*, pages 1–6, 2020.

[208] Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7263–7271, 2017.

[209] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[210] Takuya Fujihashi, Toshiaki Koike Akino, Siheng Chen, and Takashi Watanabe. Wireless 3D point cloud delivery using deep graph neural networks. In *IEEE International Conference on Communications*, pages 1–6, 2021.

[211] Chi Thang Duong, Thanh Dat Hoang, Ha Hien Dang, Quoc Viet Hung Nguyen, and Karl Aberer. On node features for graph neural networks. *arXiv e-prints*, pages 1–6, November 2019.

[212] Siheng Chen, Chaojing Duan, Yaoqing Yang, Duanshun Li, Chen Feng, and Dong Tian. Deep unsupervised learning of 3D point clouds via graph topology inference and filtering. *IEEE Transactions on Image Processing*, 29:3183–3198, 2020.

[213] Siheng Chen, Sufeng Niu, Tian Lan, and Baoan Liu. PCT: Large-scale 3D point cloud representations via graph inception networks with applications to autonomous driving. In *IEEE International Conference on Image Processing*, pages 4395–4399, 2019.

[214] Huiqiang Xie, Zhijin Qin, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning enabled semantic communication systems. *IEEE Transactions on Signal Processing*, 69:2663–2675, 2021.

[215] Zhenzi Weng and Zhijin Qin. Semantic communication systems for speech transmission. *IEEE Journal on Selected Areas in Communications*, 39(8):2434–2444, 2021.

[216] Dawei Chen, Linda Jiang Xie, BaekGyu Kim, Li Wang, Choong Seon Hong, Li-Chun Wang, and Zhu Han. Federated learning based mobile edge computing for augmented reality applications. In *International Conference on Computing, Networking and Communications*, pages 767–773, 2020.

[217] Yifang Ma, Zhenyu Wang, Hong Yang, and Lin Yang. Artificial intelligence applications in the development of autonomous vehicles: a survey. *IEEE/CAA Journal of Automatica Sinica*, 7(2):315–329, 2020.

[218] Ivar Bjørgo Saksvik, Alex Alcocer, and Vahid Hassani. A deep learning approach to dead-reckoning navigation for autonomous underwater vehicles with limited sensor payloads. In *IEEE OCEANS*, pages 1–9, 2021.

[219] Elias P. Duarte, Aurora T.R. Pozo, and Pamela Beltrani. Smart reckoning: Reducing the traffic of online multiplayer games using machine learning for movement prediction. *Entertainment Computing*, 33:100336, 2020.

[220] Simon Wiedemann, Heiner Kirchhoffer, Stefan Matlage, Paul Haase, Arturo Marban, Talmaj Marinč, David Neumann, Tung Nguyen, Heiko Schwarz, Thomas Wiegand, Detlev Marpe, and Wojciech Samek. DeepCABAC: A universal compression algorithm for deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 14(4):700–714, 2020.

[221] Heiner Kirchhoffer, Paul Haase, Wojciech Samek, Karsten Müller, Hamed Rezazadegan-Tavakoli, Francesco Cricri, Emre Aksu, Miska M. Hannuksela, Wei Jiang, Wei Wang, Shan Liu, Swayambhoo Jain, Shahab Hamidi-Rad, Fabien Racapé, and Werner Bailer. Overview of the neural network compression and representation (NNR) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5):3203–3216, 2022.

[222] Mikolaj Jankowski, Deniz Gündüz, and Krystian Mikolajczyk. AirNet: neural network transmission over the air. In *IEEE International Symposium on Information Theory*, pages 2451–2456, 2022.

[223] Solmaz Niknam, Harpreet S. Dhillon, and Jeffrey H. Reed. Federated learning for wireless communications: Motivation, opportunities, and challenges. *IEEE Communications Magazine*, 58(6):46–51, 2020.

[224] Mohammad Mohammadi Amiri and Deniz Gündüz. Federated learning over wireless fading channels. *IEEE Transactions on Wireless Communications*, 19(5):3546–3557, 2020.

[225] Guangxu Zhu, Yong Wang, and Kaibin Huang. Broadband analog aggregation for low-latency federated edge learning. *IEEE Transactions on Wireless Communications*, 19(1):491–506, 2020.

[226] Kai Yang, Tao Jiang, Yuanming Shi, and Zhi Ding. Federated learning via over-the-air computation. *IEEE Transactions on Wireless Communications*, 19(3):2022–2035, 2020.

[227] Takuya Fujihashi, Toshiaki Koike-Akino, and Takashi Watanabe. Federated AirNet: Hybrid digital-analog neural network transmission for federated learning. *arXiv e-prints*, January 2022.

[228] Andreas Noll, Basak Gülecyüz, Alexander Hofmann, and Eckehard Steinbach. A rate-scalable perceptual wavelet-based vibrotactile codec. In *IEEE Haptics Symposium*, pages 854–859, 2020.

[229] Eckehard Steinbach, Matti Strese, Mohamad Eid, Xun Liu, Amit Bhardwaj, Qian Liu, Mohammad Al-Ja'afreh, Toktam Mahmoodi, Rania Hassen, Abdulmotaleb El Saddik, and Oliver Holland. Haptic codecs for the tactile internet. *Proceedings of the IEEE*, 107(2):447–470, 2019.