Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Underwater video transceiver designs based on channel state information and video content*

Rong-xin ZHANG[1,2], Xiao-li MA[†‡3], De-qing WANG[1,2], Fei YUAN[1,2], En CHENG[1,2]

[1]*MOE Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Xiamen University, Xiamen 361005, China*

[2]*Department of Communication Engineering, Xiamen University, Xiamen 361005, China*

[3]*School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta 30332, USA*

[†]E-mail: xiaoli@gatech.edu

Received Nov. 7, 2017; Revision accepted June 21, 2018; Crosschecked Aug. 9, 2018

Abstract: Underwater hostile channel conditions challenge video transmission designs. The current designs often treat video coding and transmission schemes as individual modules. In this study, we develop an adaptive transceiver with channel state information (CSI) by taking into account the importance of video components and channel conditions. The design is more effective than the traditional ones. However, in practical systems, perfect CSI may not be available. Therefore, we compare the imperfect CSI case with existing schemes, and validate the effectiveness of our design through simulations and measured channels in terms of a better peak signal-to-noise ratio and a higher video structural similarity index.

Key words: Underwater video transmission; Transceiver design; Imperfect channel state information
https://doi.org/10.1631/FITEE.1700767 **CLC number:** TN919.8

1 Introduction

Underwater video transmissions are an important yet challenging academic topic with a wide range of industrial applications, such as marine resource exploration and exploitation, marine ecological monitoring, undersea equipment inspection, and assistant control for inshore cultivation (Vall et al., 2011; Wang et al., 2015). Although underwater acoustic transmissions have drawn significant attention from researchers, few schemes are designed based on video content. Most of the references have not taken into account the type of data source, but instead, dedicated their efforts to tackling hostile channel characteristics, i.e., limited bandwidth, low data rate, and high propagation attenuation (Stojanovic and Preisig, 2009). Recently, orthogonal frequency division multiplexing (OFDM) has been investigated due to its multipath mitigation and high data rate features (Wang et al., 2014; Kuai et al., 2016; Kumar and Kumar, 2016). Stojanovic (2006) proposed an adaptive OFDM receiver design with low complexity along with a non-uniform Doppler compensation algorithm. Different modulation schemes with OFDM technology have been discussed in underwater acoustic communications, such as on-off keying (OOK) (Li et al., 2008b), binary phase shift keying (BPSK) (Santoso et al., 2012), and quadrature phase shift keying (QPSK) and 16-QAM (quadrature amplitude modulation) (Li et al., 2008a). Santoso et al. (2012)

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61571377, 61471308, and 61771412), the Fundamental Research Funds for the Central Universities, China (No. 20720180068), and the Research Fund for the Visiting Scholar Program by the Scholarship Council of China (Nos. 201506310080 and 201506315026)

ORCID: Rong-xin ZHANG, http://orcid.org/0000-0002-0664-294X

proposed an image transmission system with OFDM where the images were converted into a series of binary data after compression; the transmission design was independent of image coding.

Generally, videos have to be compressed before transmission because they include a large amount of data and the bandwidth of acoustic channels is limited. On one hand, videos are sequences of recorded images playing at certain frame rates. Actually, the frame rate that can be achieved using acoustic waves is not sufficient for transmitting an arbitrary video, but it is quite sufficient for transmitting a specific marine surveillance video (Ribas et al., 2011; Vall et al., 2011). Moreover, the contents and motion trajectories of adjacent frames should be similar. As a result, videos have great temporal redundancies. On the other hand, the correlations between adjacent pixels within a frame lead to spatial redundancies. Therefore, a common video compression method often consists of interframe coding and intraframe coding (Hoag et al., 1997). These two coding techniques are used to reduce temporal redundancies and spatial redundancies, respectively. Most of the interframe coding schemes were implemented using local and global motion compensation (Hoag et al., 1997; Negahdaripour and Khamene, 2000; Li et al., 2009). With respect to intraframe coding, MPEG-4 and H.264 compression (Ribas et al., 2011), discrete wavelet transform (DWT) (Hoag et al., 1997; Li et al., 2009; Zhang et al., 2016b), and discrete cosine transform (DCT) (Holla and Geetha, 2015) were commonly used. It was verified in Uma et al. (2014) that the MPEG-4 compression standard satisfied the requirements of real-time underwater transmission with good video quality. Rather than pursuing a high compression ratio, Zhang et al. (2016a) tried to find a balance between coding efficiency and error resiliency. Based on the importance of the contents, a video was split into base and enhancement descriptions. The enhancement frames were selectively discarded based on the current channel conditions. To lighten the burden on underwater encoding nodes, Holla and Geetha (2015) proposed a distributed video coding (DVC) scheme combined with DCT to shift the computational complexity to the onshore receiver.

Jakubczak and Katabi (2010) developed a novel video transmission technique named SoftCast. Three-dimensional (3D) discrete cosine transform (3D-DCT) was adopted in SoftCast because it can simultaneously remove both intraframe and interframe redundancies in video streams. The coefficients obtained by 3D-DCT were grouped into different chunks. Chunks with low variances were dropped and retained chunks were scaled before transmission according to their importance. SoftCast used a raw OFDM structure, which bypasses forward error correction and QAM mapping. Instead, transmitted signal samples were linearly correlated with the original 3D-DCT coefficient values. This feature makes SoftCast more effective on video error protection, compared with traditional codeword mapping technologies (BPSK, QPSK, 16QAM, etc.) where a single bit flip in a codeword can cause a dramatic change in value reconstruction. That is, SoftCast is more resilient to noise and channel fading. SoftCast has been validated to be superior to conventional video coding schemes, like multiple resolution coding (MRC), scalable video coding with hierarchical modulation (SVC-HM), and MPEG-4.

However, the aforementioned designs considered video coding and transmission schemes as two different parts. Little information was exchanged to address the complicated channel conditions. Therefore, the performance of the systems was limited over realistic fading channels. Zhang et al. (2017) proposed ECast, which is an improvement of SoftCast that aims to overcome these problems. Taking the advantage of SoftCast, the ECast coefficient components with higher energies were preferentially allocated to the subcarriers with less subchannel fading, and were then scaled according to both their importance and their subchannel fading coefficients. However, ECast cannot be applied in an underwater environment, because the channel time-variations have not been taken into account.

Zhang et al. (2018) developed an adaptive design for underwater channels when both transmitter and receiver have perfect channel state information (CSI). In this study, we investigate an adaptive design in which both the video content and channel conditions are considered to find the optimal subcarrier matching and power allocation schemes. Compared with Zhang et al. (2018), we provide theoretical analysis. Due to channel estimation error and other uncertainties, CSI is usually imperfect in real systems. Consequently, we develop an adaptive transceiver for the imperfect CSI case.

Joint OFDM subcarrier matching and power allocation schemes have been widely studied for different scenarios, e.g., in a high-speed mobile OFDMA system (Gong et al., 2008), in an uplink relay-enhanced system (Fan et al., 2010), in a one-way relay system (Wang et al., 2008), in a two-way relay system (Vu and Kong, 2012), and in multihop networks (Ramagiri and Jagannatham, 2015). All of these studies were based on symbol-by-symbol detection and focused on maximizing the overall capacity of the system and the transmission rates. Here, we consider a block-by-block detection mechanism for underwater channels and focus directly on demodulation performance.

The notations used throughout the paper are explained as follows. Boldfaced lowercase and uppercase letters denote vectors and matrices, respectively. $\mathbb{E}[\cdot]$ denotes the expectation of the corresponding (multi-)variate. $\mathrm{tr}(\cdot)$ and $|\cdot|$ denote the trace and determinant of a matrix, respectively. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and Hermitian transpose, respectively. The operator $\otimes$ denotes the Kronecker product. $\mathcal{N}(u,\delta^2)$ represents the Gaussian distribution with mean u and variance δ^2. $\mathcal{CN}(u,\delta^2)$ represents the complex Gaussian distribution with mean u and variance δ^2. $\mathcal{CN}_p(\boldsymbol{u}_{p\times 1},\boldsymbol{R})$ represents the complex Gaussian distribution with mean vector $\boldsymbol{u}\in\mathbb{R}^{p\times 1}$ and covariance matrix $\boldsymbol{R}\in\mathbb{R}^{p\times p}$. $\boldsymbol{A}\succeq 0$ means that matrix $\boldsymbol{A}$ is a positive semidefinite matrix.

2 System model

Let us consider a single-user video communication system over underwater acoustic channels. The model is similar to that considered in Scaglione et al. (1999) and is brought into the underwater area here.

2.1 Overview of the design

The design is similar to SoftCast (Jakubczak and Katabi, 2010) and ECast (Zhang et al., 2017). As mentioned previously, SoftCast did not consider channel conditions and ECast did not take into account channel time-variations. Therefore, they cannot be applied in the underwater environment. Fig. 1 depicts the framework of the proposed design. The main components are described as follows.

2.1.1 Video source encoder

Fig. 2 shows a block diagram of the video source encoder. Every $\mathcal{T}$ successive frames are treated as a group. Each group is transformed into a 3D coefficient matrix after the 3D-DCT operations. According to the properties of natural images, the coefficient matrix is sparse and the non-zero components are concentrated in low temporal and spatial frequencies (Jakubczak and Katabi, 2010). The matrix is adjacently segmented into multiple chunks, each of which has the same size. Because the chunks with low energy barely contribute to video reconstruction, they are discarded at the transmitter to lighten the burden on bandwidth resources. Correspondingly, a bitmap is generated as metadata to help the receiver know the locations of the retained chunks. To get zero-mean distributions, all chunks retained are subtracted by their means. Moreover, each of them is rewritten to a row vector in a zig-zag order (the scanning path of the zig-zag order is shown in the lower right area of Fig. 2); these chunks are then tiled together. Through a series of these steps, matrix $\boldsymbol{D}\in\mathbb{R}^{N\times N_{\mathrm{C}}}$ is created, where N is the number of

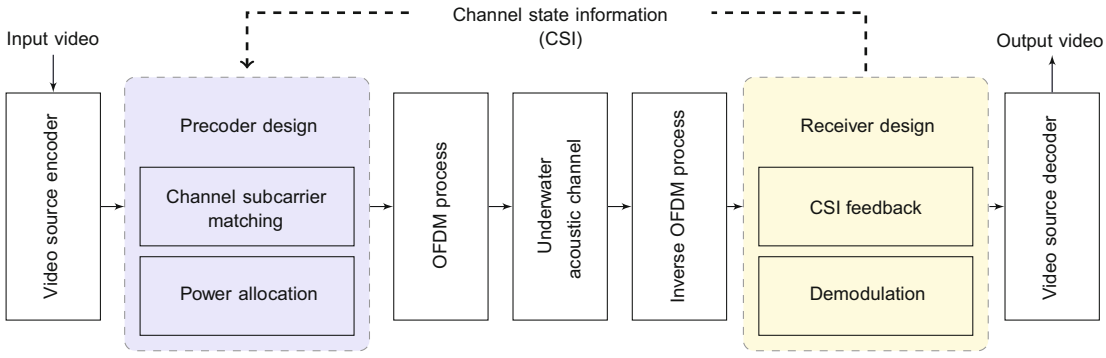

Fig. 1 Framework of our underwater video transmission system

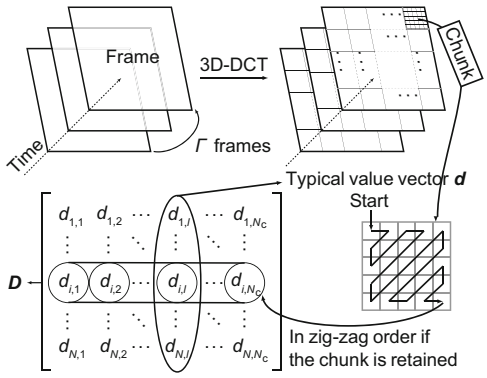

Fig. 2 Video source encoder ($d_{i,\ell}$ denotes the entry in the i^{th} row and ℓ^{th} column of D)

chunks retained and N_{C} is the number of elements in a chunk. The i^{th} row of $\boldsymbol{D}$ is the i^{th} chunk retained. The details of the procedure are given in Appendix. How to choose the group size $\mathcal{T}$ is outside the scope of this paper.

2.1.2 Metadata

The role of metadata is to assist the receiver in decoding the received signals. Metadata include a bitmap that indicates the locations of the retained chunks together with their means and variances. Metadata are coded with powerful correction capabilities (e.g., forward error correction), and are transmitted through a low rate channel with traditional communication schemes (e.g., BPSK modulation).

2.1.3 Precoder and receiver designs

Our precoder design strives to distribute different chunks to different subcarriers, and optimally schedule the power allocation based on their importance and the channel conditions. When the channel is time-invariant, the subcarrier matching problem degrades to the results in Zhang et al. (2017). At the receiver, data are decoded based on the minimum mean square error (MMSE) criterion. Because the problem is a joint transceiver design, the precoder and receiver designs are presented together in Section 2.3.

2.1.4 Video source decoder

The video source decoder receives chunk data demodulated from the receiver and performs the operations of the encoder in reverse to reconstruct the video streams with the help of metadata.

2.2 Channel model and input-output relationship

Denote an $N\times 1$ vector $\boldsymbol{d}$ (as depicted in Fig. 2), where the i^{th} entry of $\boldsymbol{d}$ is from the i^{th} chunk retained. The precoder is denoted as an $N\times N$ matrix $\boldsymbol{F}$. The output of the precoder

$$\boldsymbol{x}=\boldsymbol{F}\boldsymbol{d} \tag{1}$$

is fed to OFDM, where the i^{th} entry of $\boldsymbol{x}$ is denoted as x_i.

Suppose the baseband discrete-time channel impulse response (CIR) is causal with finite memory L. Denote the time index, the delay tap index, and the corresponding CIR coefficient as n, ℓ, and $h(n;\ell)$, respectively. After OFDM operations at both the transmitter and receiver, the u^{th} received symbol takes the form:

$$y_u=H_{u,u}x_u+\sum_{\substack{v=0\\v\neq u}}^{N-1}H_{u,v}x_v+w_u, \tag{2}$$

where w_u is the Fourier transform of the corresponding additive noise and the channel transfer function $H_{u,v}$ indicates the interference coming from the v^{th} symbol or the v^{th} subcarrier, which can be expressed as follows (Molisch, 2012):

$$H_{u,v}=\frac{1}{N}\sum_{n=0}^{N-1}\sum_{\ell=0}^{L-1}h(n;\ell)\exp\left(\mathrm{j}\frac{2\pi}{N}(v(n-\ell)-un)\right). \tag{3}$$

Collectively forming a block of size N symbols, the matrix-vector form of Eq. (2) is expressed as

$$\begin{bmatrix}y_0\\y_1\\\vdots\\y_{N-1}\end{bmatrix}=\begin{bmatrix}H_{0,0}&\dots&H_{0,N-1}\\H_{1,0}&\dots&H_{1,N-1}\\\vdots&&\vdots\\H_{N-1,0}&\dots&H_{N-1,N-1}\end{bmatrix}\begin{bmatrix}x_0\\x_1\\\vdots\\x_{N-1}\end{bmatrix}+\begin{bmatrix}w_0\\w_1\\\vdots\\w_{N-1}\end{bmatrix},$$

$$\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{w}, \tag{4}$$

where $\boldsymbol{H}$ is pre-estimated as the channel transfer matrix.

The equalizer is denoted as $\boldsymbol{G}$ and the output at the receiver can be expressed as

$$\hat{\boldsymbol{d}}=\boldsymbol{G}\boldsymbol{y}=\boldsymbol{G}\boldsymbol{H}\boldsymbol{F}\boldsymbol{d}+\boldsymbol{G}\boldsymbol{w}. \tag{5}$$

Before we introduce the transceiver design, we adopt the following assumptions:

Assumption 1 Metadata are delivered correctly to the receiver (this is reasonable according to Section 2.1.2).

Assumption 2 Data from different chunks are mutually uncorrelated; i.e., $\boldsymbol{R_d} \triangleq \mathbb{E}[\boldsymbol{d}\boldsymbol{d}^{\mathrm{H}}]$ is diagonal.

3D-DCT operation uncorrelates the original pixel values because it projects the video components on orthogonal bases, and the non-zero coefficients after the transform tend to be spatially distributed. Therefore, it is reasonable to make the assumption described in Assumption 2.

Assumption 3 The elements of $\boldsymbol{w}$ obey the complex Gaussian distribution $\mathcal{CN}(0,\delta_{\boldsymbol{w}}^2)$, and the covariance matrix of $\boldsymbol{w}$ is $\boldsymbol{R_w} \triangleq \mathbb{E}[\boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}] = \delta_{\boldsymbol{w}}^2\boldsymbol{I}$.

The noise is assumed to be Gaussian and white, and its statistical properties should remain constant across the OFDM frequency band.

Assumption 4 The video content and the noise are mutually uncorrelated; i.e., $\mathbb{E}[\boldsymbol{d}\boldsymbol{w}^{\mathrm{H}}] = \boldsymbol{0}$.

Because $\boldsymbol{w}$ is Gaussian white with zero mean, the linear combination $\boldsymbol{d}\boldsymbol{w}^{\mathrm{H}}$ is also Gaussian white with zero mean.

Assumption 5 Channel matrix $\boldsymbol{H}$ is estimated at the receiver and fed back to the transmitter.

2.3 Transceiver designs and problem formulation

According to OFDM channel conditions, our goal is to achieve better performance by allocating better and more resources to the important chunks. The resources include transmit power and subcarriers with different channel attenuations. Based on this, we define $\boldsymbol{F}$ as

$$\boldsymbol{F} = \boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{P}, \tag{6}$$

where $\boldsymbol{V}$ is a unitary precoder, $\boldsymbol{\Phi}$ is a diagonal matrix for power allocation, and $\boldsymbol{P}$ is a permutation matrix for subcarrier location matching. The entry in the i^{th} row and ℓ^{th} column of $\boldsymbol{P}$ is denoted as $P_{i\ell}$.

Without loss of generality, we assume that the number of channel subcarriers is N, which is the same as that of the chunks retained. In matrix $\boldsymbol{P}$, $P_{i\ell}=1$ implies that the ℓ^{th} chunk is allocated to the i^{th} subcarrier, and $P_{i\ell}=0$ means the opposite.

According to Eqs. (5) and (6), the error vector at the receiver is

$$\boldsymbol{e} = \hat{\boldsymbol{d}} - \boldsymbol{d} = (\boldsymbol{G}\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{P} - \boldsymbol{I})\boldsymbol{d} + \boldsymbol{G}\boldsymbol{w}. \tag{7}$$

Therefore, the mean square error (MSE) is calculated as

$$\begin{aligned}\xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}) &= \mathrm{tr}(\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^{\mathrm{H}}])\\ &= \mathrm{tr}\big((\boldsymbol{G}\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{P}-\boldsymbol{I})\boldsymbol{R_d}(\boldsymbol{G}\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{P}-\boldsymbol{I})^{\mathrm{H}}\big)\\ &\quad + \mathrm{tr}(\boldsymbol{G}\boldsymbol{R_w}\boldsymbol{G}^{\mathrm{H}}).\end{aligned} \tag{8}$$

Taking into account the limited transmit power, i.e.,

$$\mathrm{tr}(\boldsymbol{F}\boldsymbol{R_d}\boldsymbol{F}^{\mathrm{H}}) = \mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R_d}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}) \le P_0, \tag{9}$$

the problem can be formulated as follows:

Problem 1

$$\begin{aligned}\min_{\boldsymbol{V},\boldsymbol{\Phi},\boldsymbol{P},\boldsymbol{G}} \quad & \xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G})\\ \text{s.t.} \quad & \mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R_d}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}) \le P_0.\end{aligned} \tag{10}$$

The objective is to minimize the MSE by finding an optimal combination of subcarrier matching order $\boldsymbol{P}$, power allocation scheme $\boldsymbol{\Phi}$, unitary precoder $\boldsymbol{V}$, and equalizer $\boldsymbol{G}$. The constraint is that the total power allocated at the transmitter side is bounded.

The problem can be solved using the Lagrangian method as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}) &= \xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G})\\ &\quad + \mu\big(\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R_d}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}) - P_0\big),\end{aligned} \tag{11}$$

where μ is the Lagrangian multiplier.

3 Perfect channel state information case

We first develop our design in the case where the channel is perfectly known at the transmitter and the receiver.

3.1 Optimal equalizer $\boldsymbol{G}$

Differentiating $\mathcal{L}$ in Eq. (11) with respect to $\boldsymbol{G}$ and setting it to zero, we obtain

$$\begin{aligned}\boldsymbol{G}_{\mathrm{opt}} &= \boldsymbol{R_d}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\\ &\quad\cdot(\boldsymbol{R_w} + \boldsymbol{H}\boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R_d}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}})^{-1}.\end{aligned} \tag{12}$$

By using Lemma II in Lienart (2015), Eq. (12) becomes

$$\boldsymbol{G}_{\mathrm{opt}} = (\boldsymbol{R_d}^{-1} + \boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R_w}^{-1}, \tag{13}$$

where $\boldsymbol{Z} = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{w}}^{-1}\boldsymbol{H}\boldsymbol{V}$. The eigenvalue decomposition of the positive semidefinite matrix $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{w}}^{-1}\boldsymbol{H}$ is represented as

$$\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{w}}^{-1}\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}, \tag{14}$$

and then $\boldsymbol{Z} = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V}$.

According to Lemma I in Lienart (2015), the inverse term in Eq. (13) is equivalent to

$$\begin{aligned}&(\boldsymbol{R}_{\boldsymbol{d}}^{-1} + \boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi}\boldsymbol{P})^{-1}\\ &= \boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}} + \boldsymbol{Z}^{-1})^{-1}\\ &\quad\cdot \boldsymbol{Z}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}})^{-1}.\end{aligned} \tag{15}$$

Thus, $\boldsymbol{G}_{\mathrm{opt}}$ can be rewritten as

$$\begin{aligned}\boldsymbol{G}_{\mathrm{opt}} &= \boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}} + \boldsymbol{Z}^{-1})^{-1}\\ &\quad\cdot \boldsymbol{Z}^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{w}}^{-1}.\end{aligned} \tag{16}$$

3.2 Optimal precoder $\boldsymbol{F}$

Substituting Eq. (16) into Eq. (11) and taking the Woodbury formula, we arrive at

$$\begin{aligned}&\mathcal{L}(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}_{\mathrm{opt}})\\ &= \xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}_{\mathrm{opt}}) + \mu(\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}) - P_0),\end{aligned} \tag{17}$$

where

$$\xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}_{\mathrm{opt}}) = \mathrm{tr}((\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi} + \boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}^{-1}\boldsymbol{P}^{\mathrm{H}})^{-1}). \tag{18}$$

3.2.1 Choice of precoder $\boldsymbol{V}$

Proposition 1 The solution of the optimization problem:

$$\begin{aligned}\boldsymbol{V}_{\mathrm{opt}} &= \arg\min_{\boldsymbol{V}}\ \xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}_{\mathrm{opt}})\\ \text{s.t.}\quad &\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}) \le P_0\end{aligned} \tag{19}$$

is given by $\boldsymbol{V}_{\mathrm{opt}} = \boldsymbol{U}$.

Proof $\xi(\boldsymbol{P},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{G}_{\mathrm{opt}})$ is abbreviated as $\xi(\boldsymbol{V})$ for convenience. Because the power constraint has nothing to do with the choice of $\boldsymbol{V}$, the problem is equivalent to finding the optimal $\boldsymbol{V}$ that minimizes $\xi(\boldsymbol{V})$.

According to Theorem 9 in Choudhury and Sivakumar (2017), $\xi(\boldsymbol{V})$ is upper-bounded by

$$\xi(\boldsymbol{V}) < \mathrm{tr}((\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi})^{-1}), \tag{20}$$

where the upper bound will never be reached because $\boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}^{-1}\boldsymbol{P}^{\mathrm{H}}$ will never be a zero matrix. Furthermore, the upper bound is rewritten as

$$\mathrm{tr}((\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi})^{-1}) = \mathrm{tr}((\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V})(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}})^{-1}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V})^{\mathrm{H}}\boldsymbol{\Lambda}^{-1}), \tag{21}$$

by plugging in $\boldsymbol{Z} = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V}$.

The diagonal entries of $\boldsymbol{\Lambda}$ are assumed to be arranged in non-increasing order. Based on the water-filling principle, the order of the diagonal entries of $\boldsymbol{\Phi}$ is assumed to be opposite that of $\boldsymbol{\Lambda}$. This means that the diagonal entries of $(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}})^{-1}$ are also arranged in non-increasing order. Therefore, the upper bound is minimized when $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V} = \boldsymbol{I}$ or $\boldsymbol{V} = \boldsymbol{U}$ (Jafar et al., 2001).

By Theorem 17.9.4 in Cover and Thomas (2006), $\xi(\boldsymbol{V})$ is lower-bounded by

$$\begin{aligned}\xi(\boldsymbol{V}) &\ge N\left|(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi} + \boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}^{-1}\boldsymbol{P}^{\mathrm{H}})^{-1}\right|^{1/N}\\ &= N\left(\frac{1}{\left|\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi} + \boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}^{-1}\boldsymbol{P}^{\mathrm{H}}\right|}\right)^{1/N}.\end{aligned} \tag{22}$$

According to Hadamard's inequality, $\left|\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{\Phi} + \boldsymbol{P}\boldsymbol{R}_{\boldsymbol{d}}^{-1}\boldsymbol{P}^{\mathrm{H}}\right|$ is maximized when the matrix is diagonal. Therefore, when $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{U}$ is a permutation matrix, the lower bound is minimized. In this case, the choice of $\boldsymbol{V}$ is not unique, and $\boldsymbol{V} = \boldsymbol{U}$ is surely one of them.

Corollary 1 Precoder $\boldsymbol{V}_{\mathrm{opt}}$ given by Proposition 1 maximizes the mutual information of the system.

Proof When the symbols within $\boldsymbol{d}$ are Gaussian distributed, the mutual information of the system can be written as

$$\begin{aligned}&I(\hat{\boldsymbol{d}},\boldsymbol{x})\\ &= \log_2\frac{\left|\boldsymbol{G}(\boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{R}_{\boldsymbol{w}})\boldsymbol{G}^{\mathrm{H}}\right|}{|\boldsymbol{G}\boldsymbol{R}_{\boldsymbol{w}}\boldsymbol{G}^{\mathrm{H}}|}\\ &= \log_2\left|\boldsymbol{G}\boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{d}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}(\boldsymbol{G}\boldsymbol{R}_{\boldsymbol{w}}\boldsymbol{G}^{\mathrm{H}})^{-1} + \boldsymbol{I}\right|.\end{aligned} \tag{23}$$

For any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with compatible dimensions (Scaglione et al., 2002),

$$|\boldsymbol{A}\boldsymbol{B} + \boldsymbol{I}| = |\boldsymbol{B}\boldsymbol{A} + \boldsymbol{I}|. \tag{24}$$

Therefore, we obtain

$$\begin{aligned} I(\hat{\boldsymbol{d}},\boldsymbol{x}) &= \log_2\left|\boldsymbol{HFR_dF^{\mathrm{H}}H^{\mathrm{H}}R_w^{-1}}+\boldsymbol{I}\right| \\ &= \log_2\left|\boldsymbol{HV\Phi PR_dP^{\mathrm{H}}\Phi^{\mathrm{H}}V^{\mathrm{H}}H^{\mathrm{H}}R_w^{-1}}+\boldsymbol{I}\right| \\ &= \log_2\left|\boldsymbol{\Phi PR_dP^{\mathrm{H}}\Phi^{\mathrm{H}}V^{\mathrm{H}}H^{\mathrm{H}}R_w^{-1}HV}+\boldsymbol{I}\right| \\ &= \log_2\left|\boldsymbol{PR_dP^{\mathrm{H}}\Phi^{\mathrm{H}}Z\Phi}+\boldsymbol{I}\right| \\ &= \log_2\Big(\left|\boldsymbol{PR_dP^{\mathrm{H}}}\right|\cdot\left|\boldsymbol{\Phi^{\mathrm{H}}Z\Phi}+\boldsymbol{PR_d^{-1}P^{\mathrm{H}}}\right|\Big). \end{aligned} \tag{25}$$

In the proof of Proposition 1, we know that when $\boldsymbol{V}=\boldsymbol{U}$, $\left|\boldsymbol{\Phi^{\mathrm{H}}Z\Phi}+\boldsymbol{PR_d^{-1}P^{\mathrm{H}}}\right|$ is maximized. Therefore, the mutual information is maximized.

3.2.2 Optimal power allocation matrix $\boldsymbol{\Phi}$ and subcarrier matching matrix $\boldsymbol{P}$

$\boldsymbol{PR_dP^{\mathrm{H}}}$ and $\boldsymbol{PR_d^2P^{\mathrm{H}}}$ are diagonal matrices. By choosing $\boldsymbol{V}_{\mathrm{opt}}=\boldsymbol{U}$, we have $\boldsymbol{Z}=\boldsymbol{\Lambda}$, which is also diagonal. Denoting the i^{th} diagonal entry of $\boldsymbol{PR_dP^{\mathrm{H}}}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Lambda}$ by ρ_i', ϕ_i, and λ_i respectively, Eq. (17) is equivalent to

$$\begin{aligned} &\min_{\phi_i}(\mathcal{L}) \\ &=\min_{\phi_i}\left(\sum_i\frac{\rho_i'}{1+|\phi_i|^2\rho_i'\lambda_i}+\mu\Big(\sum_i|\phi_i|^2\rho_i'-P_0\Big)\right). \end{aligned} \tag{26}$$

Similar derivations can be found in Scaglione et al. (1999). Differentiating $\mathcal{L}$ in Eq. (26) with respect to $|\phi_i|^2$ and setting it to zero, we arrive at

$$|\phi_i|^2=\sqrt{\frac{1}{\mu\rho_i'\lambda_i}}-\frac{1}{\rho_i'\lambda_i}. \tag{27}$$

Setting $\sum_i|\phi_i|^2\rho_i'$ to be P_0, we obtain the Lagrangian multiplier as

$$\sqrt{\mu}=\frac{\mathrm{tr}(\boldsymbol{\Lambda}^{-1/2}\boldsymbol{PR_d^{1/2}P^{\mathrm{H}}})}{P_0+\mathrm{tr}(\boldsymbol{\Lambda}^{-1})}. \tag{28}$$

Accordingly, we define

$$\phi_i|_{\mathrm{opt}}=\begin{cases}\alpha\sqrt{\gamma_i}, & \text{if } 0\le\gamma_i\le\mathrm{Th},\\ 0, & \text{otherwise},\end{cases} \tag{29}$$

where $\gamma_i\triangleq 1/\sqrt{\mu\rho_i'\lambda_i}-1/(\rho_i'\lambda_i)$, Th is a preset threshold, and α is a positive scalar making the equality $\sum_i\phi_i^2\rho_i'=P_0$ hold.

Intuitively, it is better to concentrate the limited power on the most important chunks when the channel noise becomes stronger. Therefore, we set Th to be $P_0\gamma_0/\sqrt{\delta_{\boldsymbol{w}}^2}$ empirically, where γ_0 is the minimum positive value of γ_i. The optimal Th will not be discussed here because it is not the core issue in our design.

Plugging Eqs. (27) and (28) into the MSE part of Eq. (26) yields

$$\frac{(\mathrm{tr}(\boldsymbol{\Lambda}^{-1/2}\boldsymbol{PR_d^{1/2}P^{\mathrm{H}}}))^2}{P_0+\mathrm{tr}(\boldsymbol{\Lambda}^{-1})}. \tag{30}$$

Based on Eq. (30), the optimal $\boldsymbol{P}$ is the matrix that causes the following condition to hold:

$$\rho_i'\ge\rho_\ell',\ \text{if}\ \lambda_i\ge\lambda_\ell,\ \forall\, i,\ \ell\in[1,N]. \tag{31}$$

This is because when condition (31) is met, $\mathrm{tr}(\boldsymbol{\Lambda}^{-1/2}\boldsymbol{PR_d^{1/2}P^{\mathrm{H}}})$ is minimized according to the rearrangement inequality, and therefore the MSE in Eq. (30) is minimized. Table 1 summarizes the design parameters.

4 Imperfect channel state information case

For practical systems, the channel cannot be estimated perfectly.

We now consider the case where channel estimation error exists. The feedback channel is assumed to be ideal. In this scenario, the receiver and the transmitter have the same estimate of the real channel with

$$\hat{\boldsymbol{H}}=\boldsymbol{H}+\boldsymbol{\Xi}, \tag{32}$$

Table 1 Design parameters for the perfect channel state information case

Precoder $\boldsymbol{F}$	$\boldsymbol{V}$	$\boldsymbol{V}_{\mathrm{opt}}=\boldsymbol{U}$	$\boldsymbol{U}$ is given by Eq. (14)
	$\boldsymbol{\Phi}$	$\phi_i\|_{\mathrm{opt}}=\begin{cases}\alpha\sqrt{\gamma_i}, & \text{if } 0\le\gamma_i\le\mathrm{Th},\\ 0, & \text{otherwise.}\end{cases}$	γ_i and α are given by Eq. (29)
	$\boldsymbol{P}$	$\boldsymbol{P}_{\mathrm{opt}}$ is a permutation matrix satisfying condition (31)	
Equalizer $\boldsymbol{G}$	$\boldsymbol{G}_{\mathrm{opt}}=\boldsymbol{R_dP}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{opt}}^{\mathrm{H}}\Big(\boldsymbol{\Phi}_{\mathrm{opt}}\boldsymbol{P}_{\mathrm{opt}}\boldsymbol{R_dP}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{opt}}^{\mathrm{H}}+\boldsymbol{\Lambda}^{-1}\Big)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{R_w^{-1}}$		$\boldsymbol{\Lambda}$ is given by Eq. (14)

where $\hat{\boldsymbol{H}}$ is the estimated one, $\boldsymbol{\Xi}=[\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\dots,\boldsymbol{\epsilon}_N]$ is the corresponding error which is independent of $\hat{\boldsymbol{H}}$, and $\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\dots,\boldsymbol{\epsilon}_N$ are channel estimation errors of size N from $\mathcal{CN}_N(\boldsymbol{0}_{N\times 1},\boldsymbol{R}_{\boldsymbol{\epsilon}})$.

Before we discuss the design for the imperfect CSI case, let us introduce some notations and a theorem, all of which can be found in Gupta and Nagar (1999).

Definition 1 For matrix $\boldsymbol{A}\in\mathbb{R}^{P\times Q}$, $\mathrm{vec}(\boldsymbol{A})$ is a $PQ\times 1$ vector and is defined as

$$\mathrm{vec}(\boldsymbol{A})=\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_Q\end{bmatrix},\tag{33}$$

where $\boldsymbol{a}_q$ is the q^{th} column of $\boldsymbol{A}$.

Definition 2 Random matrix $\boldsymbol{A}\in\mathbb{R}^{P\times Q}$ is said to have a matrix variate normal distribution with mean matrix $\boldsymbol{M}\in\mathbb{R}^{P\times Q}$ and covariance matrix $\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}$, where $\boldsymbol{\Sigma}\in\mathbb{R}^{P\times P}\succeq 0$ and $\boldsymbol{\Psi}\in\mathbb{R}^{Q\times Q}\succeq 0$ if $\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}})\sim\mathcal{CN}_{PQ}(\mathrm{vec}(\boldsymbol{M}^{\mathrm{T}}),\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$. Equivalently, we say $\boldsymbol{A}\sim\mathcal{CN}_{PQ}(\boldsymbol{M},\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$.

Theorem 1 For random matrix $\boldsymbol{A}\in\mathbb{R}^{P\times Q}\sim\mathcal{CN}_{PQ}(\boldsymbol{M},\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$, the following equation holds (Gupta and Nagar, 1999):

$$\mathbb{E}_{\boldsymbol{A}}[\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{\mathrm{H}}]=\mathrm{tr}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{\Psi})\boldsymbol{\Sigma}+\boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^{\mathrm{H}},\tag{34}$$

where $\boldsymbol{B}$ is a $Q\times Q$ matrix.

From Definitions 1 and 2, we have

$$\boldsymbol{\Xi}\sim\mathcal{CN}_{NN}(\boldsymbol{0}_{N\times N},\boldsymbol{R}_{\boldsymbol{\epsilon}}\otimes\boldsymbol{I}_N).\tag{35}$$

In this section, the precoder and equalizer matrices are marked with the hat symbol '^' to differentiate the imperfect CSI case from the perfect CSI case. Plugging Eq. (32) into Eq. (8) yields

$$\begin{aligned}&\xi(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}})\\&=\mathrm{tr}((\hat{\boldsymbol{G}}(\hat{\boldsymbol{H}}-\boldsymbol{\Xi})\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}-\boldsymbol{I})\boldsymbol{R}_{\boldsymbol{d}}(\hat{\boldsymbol{G}}(\hat{\boldsymbol{H}}-\boldsymbol{\Xi})\\&\quad\cdot\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}-\boldsymbol{I})^{\mathrm{H}})+\mathrm{tr}(\hat{\boldsymbol{G}}\boldsymbol{R}_{\boldsymbol{w}}\hat{\boldsymbol{G}}^{\mathrm{H}}).\end{aligned}\tag{36}$$

Taking the expectation of Eq. (36) over $\boldsymbol{\Xi}$, we arrive at Eq. (37) (see the bottom of this page), where step (a) is based on Theorem 1 (i.e., $\mathbb{E}_{\boldsymbol{\Xi}}\big[\boldsymbol{\Xi}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\boldsymbol{\Xi}^{\mathrm{H}}\big]$ is equal to $\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})\boldsymbol{R}_{\boldsymbol{\epsilon}}$), and step (b) is obtained by denoting $\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})\boldsymbol{R}_{\boldsymbol{\epsilon}}+\boldsymbol{R}_{\boldsymbol{w}}$ as $\hat{\boldsymbol{R}}_{\boldsymbol{w}}$. Here, we fully make use of the power, i.e., $\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})=P_0$, and then $\hat{\boldsymbol{R}}_{\boldsymbol{w}}=P_0\boldsymbol{R}_{\boldsymbol{\epsilon}}+\boldsymbol{R}_{\boldsymbol{w}}$.

Problem 1 (described in Eq. (10)) is adjusted for the imperfect CSI case. Then we arrive at the following problem formulation:

Problem 2

$$\begin{aligned}\min_{\hat{\boldsymbol{V}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{P}},\hat{\boldsymbol{G}}}\quad&\mathbb{E}_{\boldsymbol{\Xi}}\big[\xi(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}})\big]\\\text{s.t.}\quad&\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})\le P_0.\end{aligned}\tag{38}$$

The corresponding Lagrangian function can be expressed as

$$\begin{aligned}\hat{\mathcal{L}}(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}})&=\mathbb{E}_{\boldsymbol{\Xi}}\big[\xi(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}})\big]\\&\quad+\hat{\mu}\big(\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})-P_0\big),\end{aligned}\tag{39}$$

where $\hat{\mu}$ is the Lagrangian multiplier.

Using derivations similar to those for the perfect CSI case, we obtain the optimal equalizer $\hat{\boldsymbol{G}}$ with imperfect CSI, which is expressed as

$$\begin{aligned}\hat{\boldsymbol{G}}_{\mathrm{opt}}&=\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}+\hat{\boldsymbol{Z}}^{-1})^{-1}\\&\quad\cdot\hat{\boldsymbol{Z}}^{-1}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{R}}_{\boldsymbol{w}}^{-1},\end{aligned}\tag{40}$$

where $\hat{\boldsymbol{Z}}=\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{U}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{U}}^{\mathrm{H}}\hat{\boldsymbol{V}}$, $\hat{\boldsymbol{U}}$, and $\hat{\boldsymbol{\Lambda}}$ are the matrices resulting from the eigenvalue decomposition:

$$\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{R}}_{\boldsymbol{w}}^{-1}\hat{\boldsymbol{H}}=\hat{\boldsymbol{U}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{U}}^{\mathrm{H}}.\tag{41}$$

Substituting Eq. (40) into Eq. (39) yields

$$\begin{aligned}\hat{\mathcal{L}}(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}}_{\mathrm{opt}})&=\mathbb{E}_{\boldsymbol{\Xi}}\big[\xi(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}}_{\mathrm{opt}})\big]\\&\quad+\hat{\mu}\big(\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})-P_0\big),\end{aligned}\tag{42}$$

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Xi}}\big[\xi(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}})\big]&=\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}})-\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}})-\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}})\\&\quad+\mathrm{tr}\left(\hat{\boldsymbol{G}}\Big(\mathbb{E}_{\boldsymbol{\Xi}}\big[\boldsymbol{\Xi}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\boldsymbol{\Xi}^{\mathrm{H}}\big]\Big)\hat{\boldsymbol{G}}^{\mathrm{H}}\right)+\mathrm{tr}(\hat{\boldsymbol{G}}\boldsymbol{R}_{\boldsymbol{w}}\hat{\boldsymbol{G}}^{\mathrm{H}})\\&\overset{(a)}{=}\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}})-\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}})-\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}})\\&\quad+\mathrm{tr}(\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}})\mathrm{tr}(\hat{\boldsymbol{G}}\boldsymbol{R}_{\boldsymbol{\epsilon}}\hat{\boldsymbol{G}}^{\mathrm{H}})+\mathrm{tr}(\hat{\boldsymbol{G}}\boldsymbol{R}_{\boldsymbol{w}}\hat{\boldsymbol{G}}^{\mathrm{H}})\\&\overset{(b)}{=}\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}})-\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}})\\&\quad-\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\mathrm{H}}\hat{\boldsymbol{\Phi}}^{\mathrm{H}}\hat{\boldsymbol{V}}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\hat{\boldsymbol{G}}^{\mathrm{H}})+\mathrm{tr}(\hat{\boldsymbol{G}}\hat{\boldsymbol{R}}_{\boldsymbol{w}}\hat{\boldsymbol{G}}^{\mathrm{H}}).\end{aligned}\tag{37}$$

where

$$\mathbb{E}_{\boldsymbol{\Xi}}\big[(\hat{\boldsymbol{P}},\hat{\boldsymbol{\Phi}},\hat{\boldsymbol{V}},\hat{\boldsymbol{G}}_{\rm opt})\big]=\mathrm{tr}((\hat{\boldsymbol{\Phi}}^{\rm H}\hat{\boldsymbol{Z}}\hat{\boldsymbol{\Phi}}+\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}^{-1}\hat{\boldsymbol{P}}^{\rm H})^{-1}). \tag{43}$$

Denote the $i^{\rm th}$ diagonal entries of $\hat{\boldsymbol{\Phi}}$, $\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}^{\rm H}$, and $\hat{\boldsymbol{\Lambda}}$ by $\hat{\phi}_i$, $\hat{\rho}'_i$, and $\hat{\lambda}_i$, respectively. Because Eq. (42) is similar to Eq. (17), according to Proposition 1 and the discussion in the perfect CSI case, we have

$$\hat{\boldsymbol{V}}_{\rm opt}=\hat{\boldsymbol{U}}, \tag{44}$$

$$\hat{\phi}_i|_{\rm opt}=\begin{cases}\hat{\alpha}\sqrt{\hat{\gamma}_i}, & \text{if } 0\le\hat{\gamma}_i\le\hat{\rm Th},\\ 0, & \text{otherwise},\end{cases} \tag{45}$$

where $\hat{\alpha}$ is a positive scalar that makes $\sum_i\hat{\phi}_i^2\hat{\rho}'_i=P_0$ hold, $\hat{\gamma}_i\triangleq 1/\sqrt{\hat{\mu}\hat{\rho}'_i\hat{\lambda}_i}-1/(\hat{\rho}'_i\hat{\lambda}_i)$, $\hat{\rm Th}$ is a preset threshold with the value $P_0\hat{\gamma}_0/\sqrt{\delta_{\boldsymbol{w}}^2}$, and $\hat{\gamma}_0$ is the minimum positive value of $\hat{\gamma}_i$.

The Lagrangian multiplier is calculated as

$$\sqrt{\hat{\mu}}=\frac{\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^{-1/2}\hat{\boldsymbol{P}}\boldsymbol{R}_{\boldsymbol{d}}^{1/2}\hat{\boldsymbol{P}}^{\rm H})}{P_0+\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^{-1})}. \tag{46}$$

The optimal $\hat{\boldsymbol{P}}$ is the matrix that causes the following condition to hold:

$$\hat{\rho}'_i\ge\hat{\rho}'_\ell,\ \text{if}\ \hat{\lambda}_i\ge\hat{\lambda}_\ell,\ \forall\, i,\ \ell\in[1,N]. \tag{47}$$

Table 2 summarizes the parameters for the imperfect CSI case.

5 Simulations

In this section, SoftCast is chosen for comparison due to its advantages compared with conventional video coding schemes like MRC, SVC-HM, and MPEG-4, and its advantages compared with the traditional codeword mapping technologies like BPSK and QAM. We use two metrics to evaluate our design: peak signal-to-noise ratio (PSNR) and video structural similarity (VSSIM) index. PSNR is widely used to measure the quality of a specific image/video scheme, and can be calculated as

$$\mathrm{PSNR}=10\lg\Big(\frac{2^M-1}{\mathrm{MSE}}\Big)\ \mathrm{dB}, \tag{48}$$

where MSE is the mean square error between the reconstructed pixels and the original pixels, and M represents the number of bits used to encode each pixel, typically 8 bits for monochrome images/videos and 24 bits for RGB color images/videos.

It is worth mentioning that the MSE term in the PSNR formula is different from that in the system design. The former is related to the pixel data, whereas the latter is related to 3D-DCT coefficients. They are positively correlated without a straightforward relationship.

The VSSIM index uses structural distortion for quality evaluation, which can be calculated as

$$\text{VSSIM index}=\frac{\sum_{\ell=1}^{N_{\rm F}}W_\ell Q_\ell}{\sum_{\ell=1}^{N_{\rm F}}W_\ell}, \tag{49}$$

where $N_{\rm F}$ is the number of frames in the video sequence, Q_ℓ denotes the quality index measure of the $\ell^{\rm th}$ frame, and W_ℓ is the corresponding weight. The calculations of these parameters can be found in Wang et al. (2004). The motion vector needed to compute the VSSIM index is calculated according to Ribas-Corbera and Neuhoff (2001). Compared with the PSNR, the VSSIM index is a measurement closer to human perception. The value range of the VSSIM index is between 0 and 1, where a higher value is better.

The two testing monochrome videos, named 'Tennis' and 'Football', were chosen from the standard reference video library built by the Center for Image Processing Research (http://www.cipr.rpi.edu/resource/sequences/sif.html). As shown in Fig. 3, the two different sets of CIRs, named Channel I and Channel II, were collected from the sea trials depicted in Hu et al. (2016). The transmitter and the receiver were placed at a distance of

Table 2 Design parameters for the imperfect channel state information case

Precoder $\hat{\boldsymbol{F}}$	$\hat{\boldsymbol{V}}$	$\hat{\boldsymbol{V}}_{\rm opt}=\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{U}}$ is given by Eq. (41)
	$\hat{\boldsymbol{\Phi}}$	$\hat{\phi}_i\|_{\rm opt}=\begin{cases}\hat{\alpha}\sqrt{\hat{\gamma}_i}, & \text{if } 0\le\hat{\gamma}_i\le\hat{\rm Th},\\ 0, & \text{otherwise.}\end{cases}$	$\hat{\gamma}_i$ and $\hat{\alpha}$ are given by Eq. (45)
	$\hat{\boldsymbol{P}}$	$\hat{\boldsymbol{P}}_{\rm opt}$ is a permutation matrix satisfying condition (47)	
Equalizer $\hat{\boldsymbol{G}}$	$\hat{\boldsymbol{G}}_{\rm opt}=\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}_{\rm opt}^{\rm H}\hat{\boldsymbol{\Phi}}_{\rm opt}^{\rm H}\Big(\hat{\boldsymbol{\Phi}}_{\rm opt}\hat{\boldsymbol{P}}_{\rm opt}\boldsymbol{R}_{\boldsymbol{d}}\hat{\boldsymbol{P}}_{\rm opt}^{\rm H}\hat{\boldsymbol{\Phi}}_{\rm opt}^{\rm H}+\hat{\boldsymbol{\Lambda}}^{-1}\Big)^{-1}\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{V}}_{\rm opt}^{\rm H}\hat{\boldsymbol{H}}^{\rm H}\hat{\boldsymbol{R}}_{\boldsymbol{w}}^{-1}$	$\hat{\boldsymbol{\Lambda}}$ is given by Eq. (41)	

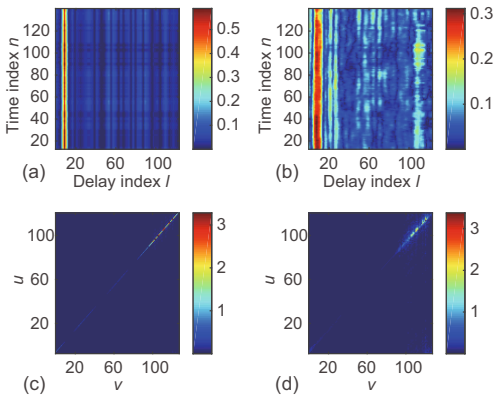

Fig. 3 Channel impulse responses collected at Taiwan Strait in July 2014: (a) Channel I; (b) Channel II; (c) $\boldsymbol{H}(u,v)$ in Channel I; (d) $\boldsymbol{H}(u,v)$ in Channel II. References to color refer to the online version of this figure

30 km, and thus we set P_0 to be 60 dBm (1000 W). We also tested our design and SoftCast under the channels generated by the model presented in Berger et al. (2010), which is expressed as

$$h(\tau,t)=\sum_{p=1}^{N_{\mathrm{d}}}A_p\delta(\tau-(\tau_p-a_pt)), \qquad (50)$$

where N_{d} is the number of dominant discrete paths, A_p is the amplitude of path p, and a_p is the path dependent Doppler rate. Note that the channel conditions deteriorate from the channel generated by Eq. (50) to Channel I, and then to Channel II.

Table 3 summarizes some common parameters.

5.1 Perfect CSI case

As depicted in Fig. 4, for the Tennis video, our design achieves a 6–7 dB gain using the channels generated by Eq. (50), an 8 dB gain under Channel I, and a 9 dB gain under Channel II compared with SoftCast in terms of average PSNR. The average VSSIM index increases by about 0.20, 0.25, and 0.29, respectively.

Table 3 Simulation parameters

Parameter	Value
Number of subcarriers	128
Video resolution	240×352
$\mathcal{T}$	4
Chunk size	30×44
Number of chunks per frame	64
Ratio of chunks retained	50%
N_{d}	15
A_p	$\mathcal{CN}(0,1)$ with tap power decreasing 20 dB in 40 ms and $\sum_p A_p^2=1$
τ_p	Poisson distribution with mean 1×10^3 (within 1 s)
a_p	$\mathcal{N}(1\times10^{-3},6.7\times10^{-4})$

With respect to the Football video, according to Fig. 5, the average VSSIM index increases by about 0.21, 0.27, and 0.27, accompanied by average PSNR gains of about 6 dB, 7 dB, and 9 dB, under the channels generated by Eq. (50), Channel I, and Channel II, respectively.

We found that the average PSNR becomes worse when the channel conditions deteriorate (from channels generated by Eq. (50), to Channel I, then to Channel II) under the same $\delta_{\boldsymbol{w}}^2$ for SoftCast. This is because the power loading method in SoftCast does not depend on the channel conditions. This makes SoftCast vulnerable in this atrocious transmission environment, whereas a contrary phenomenon was seen for our design when $\delta_{\boldsymbol{w}}^2\geq40$; i.e., the average PSNR improves when the channel conditions deteriorate. This is because when the channel conditions deteriorate, the chunks with the highest energy are allocated with a greater proportion of the power, which makes them more resilient to channel noise. Although most of the chunks with lower energy are easily buried in the noise in this case, the system performance is improved because those important chunks account for the vast majority of the MSE. However, when $\delta_{\boldsymbol{w}}^2$ is small (the noise is low), the channel conditions are better, more details of the videos (the chunks with lower energy) are retained, and the important chunks are protected. Thus, all designs have similar performance. The aforementioned phenomenon does not apply for the average VSSIM index. This is because our design is derived based on a data-distortion criterion (MMSE), and the PSNR is a metric based on data distortion, whereas VSSIM is more about structural distortion.

Aside from those cold data, let us look directly at the videos. As shown in Fig. 6, the frame quality reconstructed by our design outperforms that reconstructed by SoftCast under the same conditions. From Figs. 6b and 6e, we find that ECast fails to reconstruct the frames under Channel II.

5.2 Imperfect CSI case

Without loss of generality, the channel estimation error is assumed to be additive white. We define

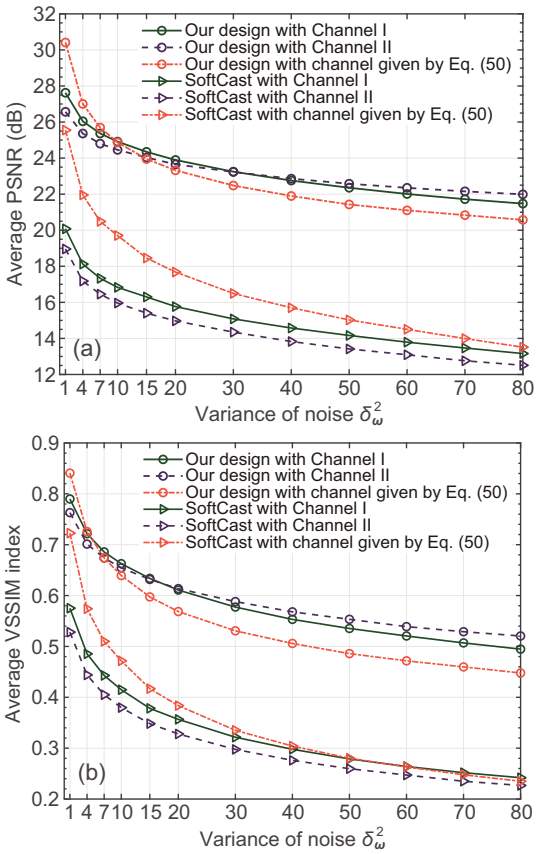

Fig. 4 Performance comparison of the Tennis video: (a) average PSNR; (b) average VSSIM index

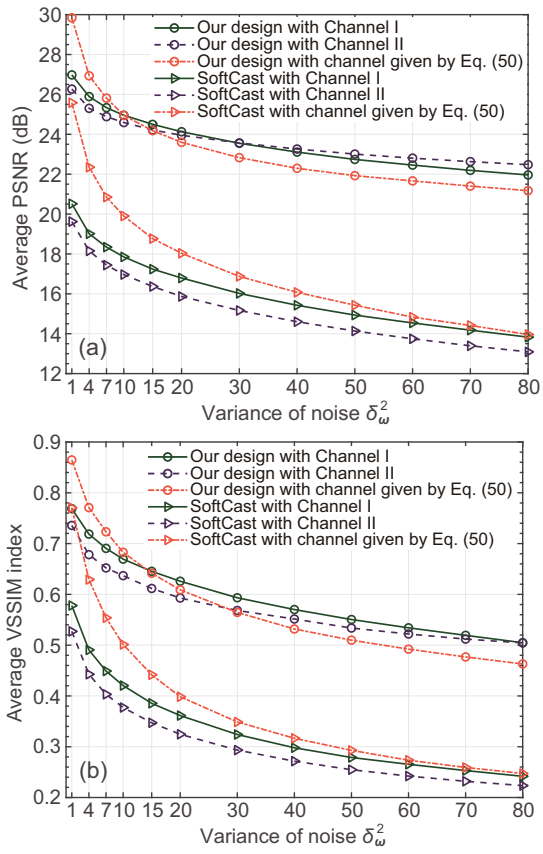

Fig. 5 Performance comparison of the Football video: (a) average PSNR; (b) average VSSIM index

an index called the channel-to-error ratio (CER) (similar to SNR), which can be calculated as

$$\mathrm{CER} = 10\lg\Big(\frac{\text{Average power of } \boldsymbol{H}}{\text{Average power of } \boldsymbol{\Xi}}\Big)\ \mathrm{dB}. \qquad (51)$$

From Figs. 7 and 8, we have the following observations.

For both cases, our design outperforms SoftCast when CER is larger than or equal to 0 dB in terms of PSNR and -4 dB in terms of the VSSIM index.

For both videos, the average PSNR of our design drops below the SoftCast baseline when CER is less than or equal to -2 dB under the channels generated by Eq. (50), -4 dB under Channel I, and -8 dB under Channel II.

For both videos, the average VSSIM index of our design drops below the SoftCast baseline when CER is less than or equal to -6 dB under the channels generated by Eq. (50) and -8 dB under Channel I. With respect to Channel II, when CER is within the range $[-8, +\infty)$ dB, our design has a higher average VSSIM index.

Comparing the curves related to the three different CIRs, the performance of our design improves when the channel conditions deteriorate. As shown in Eq. (37), the imperfect CSI case can be regarded as a prefect CSI case with an effective noise whose covariance matrix is $\hat{\boldsymbol{R}}_{\boldsymbol{w}} = P_0\boldsymbol{R}_{\boldsymbol{\epsilon}} + \boldsymbol{R}_{\boldsymbol{w}}$. Based on our earlier discussion in the perfect CSI case, it is little wonder that we see such a phenomenon.

Let us look at the snapshots of the reconstructed videos. As shown in Figs. 9 and 10, the frames reconstructed by SoftCast and ECast are mottled, whereas the frames reconstructed by our design are smoother. The actual video-viewing experience using our design is better compared with that of SoftCast, even when our PSNR or VSSIM index is lower than the SoftCast baseline. ECast also fails to reconstruct the frames under Channel II when the CSI is imperfect. Figs. 7–10 reflect that, compared with the PSNR, the VSSIM index is more consistent with the characteristics of human vision.

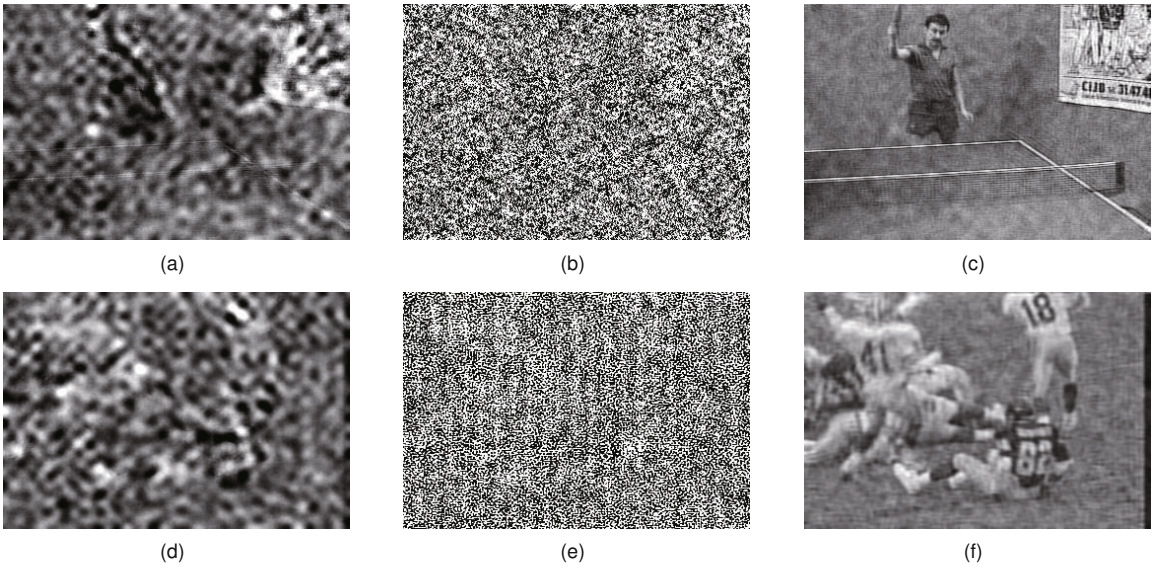

Fig. 6 Snapshots of the videos reconstructed at the receiver (under Channel II, $\delta_w^2 = 50$): the 80^{th} frame of the Tennis video reconstructed by SoftCast (a), ECast (b), and our design (c); the 50^{th} frame of the Football video reconstructed by SoftCast (d), ECast (e), and our design (f)

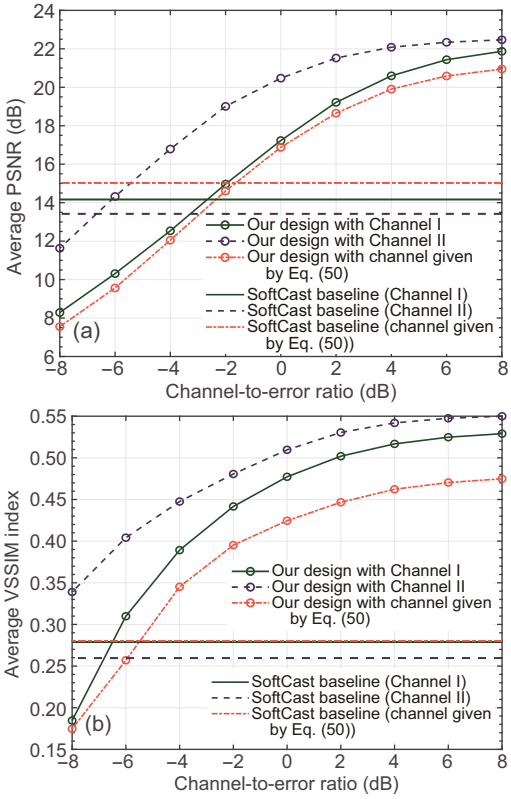

Fig. 7 Performance comparison of the Tennis video between SoftCast and our design with imperfect channel state information ($\delta_w^2 = 50$): (a) average PSNR; (b) average VSSIM index

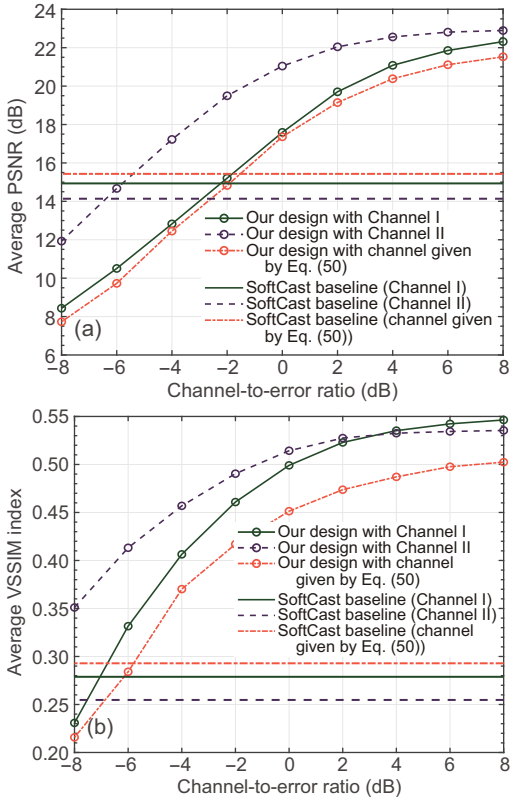

Fig. 8 Performance comparison of the Football video between SoftCast and our design with imperfect channel state information ($\delta_w^2 = 50$): (a) average PSNR; (b) average VSSIM index

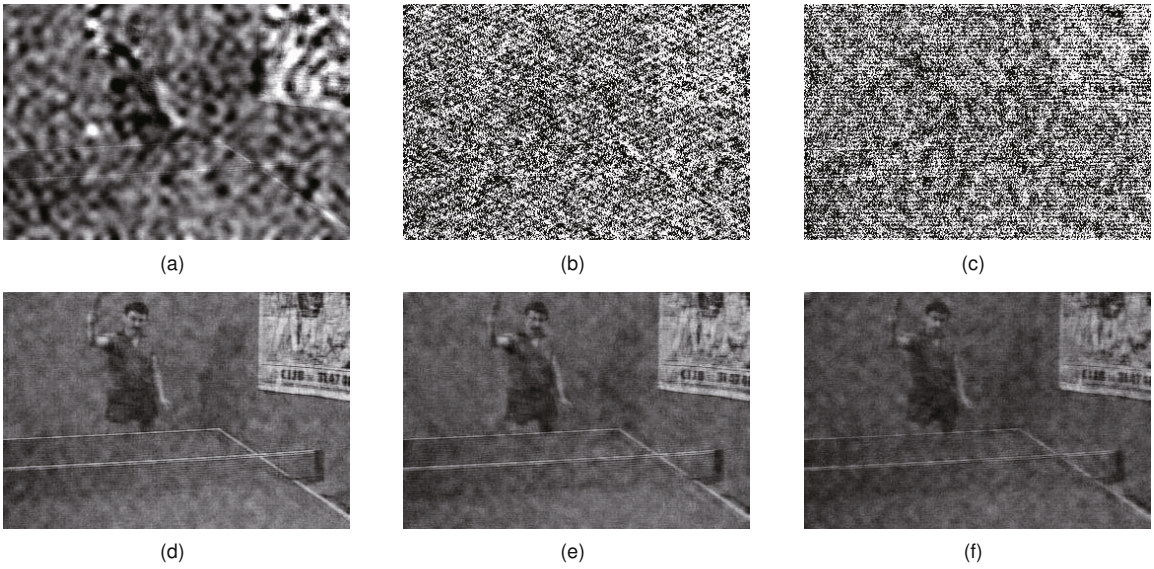

Fig. 9 Direct-vision comparison with imperfect channel state information: snapshots of the 80^{th} frame of the Tennis video reconstructed at the receiver (under Channel II and $\delta_w^2 = 50$) by SoftCast (a), ECast with CER $= -2$ dB (b), ECast with CER $= -4$ dB (c), our design with CER $= -2$ dB (d), our design with CER $= -4$ dB (e), and our design with CER $= -6$ dB (f)

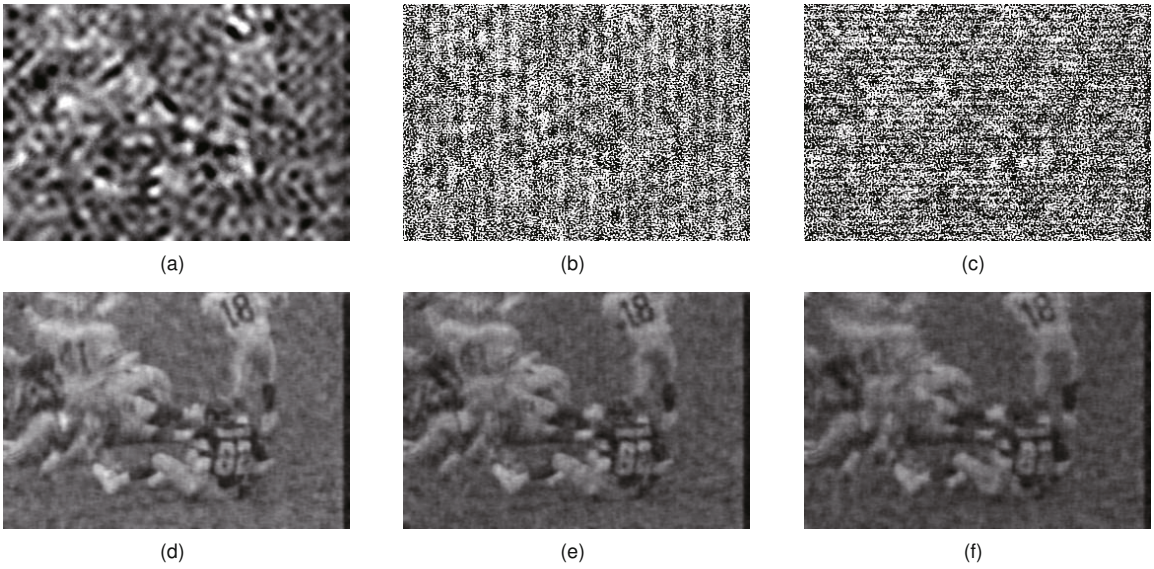

Fig. 10 Direct-vision comparison with imperfect channel state information: snapshots of the 50^{th} frame of the Football video reconstructed at the receiver (under Channel II and $\delta_w^2 = 50$) by SoftCast (a), ECast with CER $= -2$ dB (b), ECast with CER $= -4$ dB (c), our design with CER $= -2$ dB (d), our design with CER $= -4$ dB (e), and our design with CER $= -6$ dB (f)

6 Conclusions

In this paper, we have proposed a new video transceiver over underwater fading channels. The design was first studied based on the perfect CSI case, and then extended to the case of an imperfect CSI case. The optimal subcarrier matching orders and power allocation schemes were similar for both cases by theoretical analysis. Experimental results revealed that, by using the CSI, our design outperforms the existing SoftCast and is resistant to channel estimation error.

References

Berger CR, Zhou S, Preisig JC, et al., 2010. Sparse channel estimation for multicarrier underwater acoustic communication: from subspace methods to compressed sensing. *IEEE Trans Signal Process*, 58(3):1708-1721. https://doi.org/10.1109/TSP.2009.2038424

Choudhury PN, Sivakumar K, 2017. Trace inequalities for positive semidefinite matrices. *Discuss Math Gener Algebra Appl*, 37(1):93-94. https://doi.org/10.7151/dmgaa.1267

Cover TM, Thomas JA, 2006. Elements of Information Theory (2nd Ed). Wiley-Interscience, New York, USA.

Fan B, Wang W, Wu W, et al., 2010. Joint subcarrier and power allocation for uplink relay-enhanced OFDM systems. *Int J Commun Syst*, 23(11):1366-1381. https://doi.org/10.1002/dac.1111

Gong M, Zhang C, Lu J, et al., 2008. Dynamic resource allocation in high speed mobile OFDMA system. IEEE Int Conf on Communications, p.3335-3339. https://doi.org/10.1109/ICC.2008.627

Gupta AK, Nagar DK, 1999. Matrix Variate Distributions. In: Brezis H, Douglas RG, Jeffrey A (Eds.). Chapman and Hall/CRC Press, London, UK.

Hoag DF, Ingle VK, Gaudette RJ, 1997. Low-bit-rate coding of underwater video using wavelet-based compression algorithms. *IEEE J Ocean Eng*, 22(2):393-400. https://doi.org/10.1109/48.585958

Holla S, Geetha K, 2015. Distributed video coding for underwater acoustic channels. IEEE Int Conf on Computer Graphics, Vision and Information Security, p.77-82. https://doi.org/10.1109/CGVIS.2015.7449897

Hu X, Wang D, Lin Y, et al., 2016. Multi-channel time frequency shift keying in underwater acoustic communication. *Appl Acoust*, 103:54-63. https://doi.org/10.1016/j.apacoust.2015.10.009

Jafar SA, Vishwanath S, Goldsmith A, 2001. Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback. IEEE Int Conf on Communications, p.2266-2270. https://doi.org/10.1109/ICC.2001.937059

Jakubczak S, Katabi D, 2010. SoftCast: one-size-fits-all wireless video. *ACM SIGCOMM Comput Commun Rev*, 40(4):449-450. https://doi.org/10.1145/1851275.1851257

Kuai X, Sun H, Zhou S, et al., 2016. Impulsive noise mitigation in underwater acoustic OFDM systems. *IEEE Trans Veh Technol*, 65(10):8190-8202. https://doi.org/10.1109/TVT.2016.2516539

Kumar P, Kumar P, 2016. Performance evaluation of $\pi/4$-DQPSK OFDM over underwater acoustic channels. *Wirel Pers Commun*, 91(3):1137-1152. https://doi.org/10.1007/s11277-016-3517-0

Li B, Zhou S, Huang J, et al., 2008a. Scalable OFDM design for underwater acoustic communications. IEEE Int Conf on Acoustics, Speech and Signal Processing, p.5304-5307. https://doi.org/10.1109/ICASSP.2008.4518857

Li B, Zhou S, Stojanovic M, et al., 2008b. Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts. *IEEE J Ocean Eng*, 33(2):198-209. https://doi.org/10.1109/joe.2008.920471

Li Q, Wang B, Wang W, et al., 2009. An efficient underwater video compression algorithm for underwater acoustic channel transmission. WRI Int Conf on Communications and Mobile Computing, p.211-215. https://doi.org/10.1109/CMC.2009.107

Lienart T, 2015. Matrix Inversion Lemmas. https://www.stats.ox.ac.uk/%7Elienart/blog_linalg_invlemmas.html

Molisch AF, 2012. Wireless Communications (2nd Ed). John Wiley & Sons.

Negahdaripour S, Khamene A, 2000. Motion-based compression of underwater video imagery for the operations of unmanned submersible vehicles. *Comput Vis Image Underst*, 79(1):162-183. https://doi.org/10.1006/cviu.2000.0845

Ramagiri VK, Jagannatham AK, 2015. Optimal joint OFDM subcarrier, rate and power allocation for video quality maximization in multihop wireless sensor networks. 10th Int Conf on Information, Communications and Signal Processing, p.1-5. https://doi.org/10.1109/ICICS.2015.7459886

Ribas J, Sura D, Stojanovic M, 2011. Underwater wireless video transmission for supervisory control and inspection using acoustic OFDM. OCEANS, p.1-9. https://doi.org/10.1109/Oceans-Spain.2011.6003396

Ribas-Corbera J, Neuhoff DL, 2001. Optimizing motion-vector accuracy in block-based video coding. *IEEE Trans Circ Syst Video Technol*, 11(4):497-511. https://doi.org/10.1109/76.915356

Santoso TB, Wirawan I, Hendrantoro G, 2012. Image transmission with OFDM technique in underwater acoustic environment. IEEE 7th Int Conf on Telecommunication Systems, Services, and Applications, p.37-41. https://doi.org/10.1109/TSSA.2012.6366017

Scaglione A, Giannakis GB, Barbarossa S, 1999. Redundant filterbank precoders and equalizers I: unification and optimal designs. *IEEE Trans Signal Process*, 47(7):1988-2006. https://doi.org/10.1109/78.771047

Scaglione A, Stoica P, Barbarossa S, et al., 2002. Optimal designs for space-time linear precoders and decoders. *IEEE Trans Signal Process*, 50(5):1051-1064. https://doi.org/10.1109/78.995062

Stojanovic M, 2006. Low complexity OFDM detector for underwater acoustic channels. OCEANS, p.1-6. https://doi.org/10.1109/OCEANS.2006.307057

Stojanovic M, Preisig J, 2009. Underwater acoustic communication channels: propagation models and statistical characterization. *IEEE Commun Mag*, 47(1):84-89. https://doi.org/10.1109/MCOM.2009.4752682

Uma B, Geetha K, Prasanna Kumar S, et al., 2014. Simulation of H.264 based real time video encoder for underwater acoustic channel. *Int J Curr Eng Technol*, 4(3):1715-1718.

Vall LD, Sura D, Stojanovic M, 2011. Towards underwater video transmission. 6th ACM Int Workshop on Underwater Networks, p.1-5. https://doi.org/10.1145/2076569.2076573

Vu HN, Kong HY, 2012. Joint subcarrier matching and power allocation in OFDM two-way relay systems. *J Commun Netw*, 14(3):257-266. https://doi.org/10.1109/JCN.2012.6253086

Wang C, Wang Z, Nooshabadi S, 2014. Signal alignment for secure underwater coordinated multipoint transmissions. IEEE Conf on Communications and Network Security, p.145-150. https://doi.org/10.1109/CNS.2014.6997480

Wang H, Cai W, Yang J, et al., 2015. Design of HD video surveillance system for deep-sea biological exploration. IEEE 16th Int Conf on Communication Technology, p.908-911. https://doi.org/10.1109/ICCT.2015.7399971

Wang W, Yang S, Gao L, 2008. Comparison of schemes for joint subcarrier matching and power allocation in OFDM decode-and-forward relay system. IEEE Int Conf on Communications, p.4983-4987. https://doi.org/10.1109/ICC.2008.934

Wang Z, Lu L, Bovik AC, 2004. Video quality assessment based on structural distortion measurement. *Signal Process Image Commun*, 19(2):121-132. https://doi.org/10.1016/s0923-5965(03)00076-6

Zhang R, Kong Y, Ma X, et al., 2018. Adaptive video transmission designs over underwater acoustic channels. Int Conf on Computing, Networking, and Communication, p.1-5. https://doi.org/10.1109/ICCNC.2018.8390290

Zhang Y, Negahdaripour S, Li Q, 2016a. Error-resilient coding for underwater video transmission. MTS/IEEE OCEANS Monterey, p.1-7. https://doi.org/10.1109/OCEANS.2016.7761300

Zhang Y, Negahdaripour S, Li Q, 2016b. Low bit-rate compression of underwater imagery based on adaptive hybrid wavelets and directional filter banks. *Signal Process Image Commun*, 47:96-114. https://doi.org/10.1016/j.image.2016.06.001

Zhang Z, Liu D, Ma X, et al., 2017. ECast: an enhanced video transmission design for wireless multicast systems over fading channels. *IEEE Syst J*, 11(4):2566-2577. https://doi.org/10.1109/JSYST.2015.2438071

Appendix: Process at the video source encoder

In this appendix, we describe the procedure performed at the video source encoder.

Suppose that the video resolution is $\mathcal{H}\times\mathcal{W}$ pixels per frame. Every $\mathcal{T}$ successive frames are treated as a group. Denote the pixel in row i, column ℓ of the k^{th} frame in the group as $I(i,\ell,k)$. The 3D-DCT coefficients are calculated as

$$I_{3\mathrm{D}}(a,b,c)=\sum_{i=0}^{\mathcal{H}-1}\sum_{\ell=0}^{\mathcal{W}-1}\sum_{k=0}^{\mathcal{T}-1}I(i,\ell,k)f_{\mathcal{H}}(i,a)f_{\mathcal{W}}(\ell,b)f_{\mathcal{T}}(k,c), \tag{A1}$$

where

$$f_{\mathcal{N}}(p,q)=\begin{cases}\frac{1}{\sqrt{\mathcal{N}}}, & \text{if } q=0,\\ \sqrt{\frac{2}{\mathcal{N}}}\cos\left(\frac{\pi(2p+1)}{2\mathcal{N}}q\right), & \text{if } q\neq 0.\end{cases} \tag{A2}$$

Each chunk is denoted as an $\mathcal{H}_{\mathrm{C}}\times\mathcal{W}_{\mathrm{C}}$ matrix with the number of coefficients $N_{\mathrm{C}}=\mathcal{H}_{\mathrm{C}}\mathcal{W}_{\mathrm{C}}$. Suppose that $N_{\mathcal{H}}\triangleq\mathcal{H}/\mathcal{H}_{\mathrm{C}}$ and $N_{\mathcal{W}}\triangleq\mathcal{W}/\mathcal{W}_{\mathrm{C}}$ are positive integers. Therefore, there are $N_{\mathcal{H}}N_{\mathcal{W}}\mathcal{T}$ chunks related to a group. Denote the k^{th} chunk by $\boldsymbol{C}^{(k)}$, where the entry in the m^{th} row and n^{th} column is

$$\begin{aligned}C_{m,n}^{(k)}=I_{3\mathrm{D}}\Big(&\big(g(k,N_{\mathcal{H}})-1\big)N_{\mathcal{H}}+m,\\ &\left(g\left(\left\lceil\frac{k}{N_{\mathcal{H}}}\right\rceil,N_{\mathcal{W}}\right)-1\right)N_{\mathcal{W}}+n,\\ &\left\lceil\frac{\lceil k/N_{\mathcal{H}}\rceil}{N_{\mathcal{W}}}\right\rceil\Big),\end{aligned} \tag{A3}$$

where $\lceil\cdot\rceil$ is the ceiling function that outputs the least integer greater than or equal to the corresponding variable, and

$$g(p,q)=\begin{cases}q, & \text{if } \lfloor\frac{p}{q}\rfloor=\frac{p}{q},\\ p-q\cdot\lfloor\frac{p}{q}\rfloor, & \text{otherwise,}\end{cases} \tag{A4}$$

where p and q are positive integers, and $\lfloor\cdot\rfloor$ is the floor function that outputs the greatest integer less than or equal to the corresponding variable.

The energy and average of the k^{th} chunk are respectively calculated as

$$\mathrm{Eng}(\boldsymbol{C}^{(k)})=\sum_{m=1}^{\mathcal{H}_{\mathrm{C}}}\sum_{N=1}^{\mathcal{W}_{\mathrm{C}}}|C_{m,n}^{(k)}|^2, \tag{A5}$$

$$\bar{\boldsymbol{C}}^{(k)}=\frac{1}{N_{\mathrm{C}}}\sum_{m=1}^{\mathcal{H}_{\mathrm{C}}}\sum_{N=1}^{\mathcal{W}_{\mathrm{C}}}C_{m,n}^{(k)}. \tag{A6}$$

To lighten the burden on bandwidth resources, we retain N chunks with the largest energy. If the k^{th} chunk is the n^{th} chunk retained, we remove the mean from the chunk and obtain a new chunk as

$$\tilde{\boldsymbol{C}}^{(n)}=\boldsymbol{C}^{(k)}-\bar{\boldsymbol{C}}^{(k)}. \tag{A7}$$

The new chunk will be rewritten to a row vector denoted as $\tilde{\boldsymbol{d}}_n^{\mathrm{T}}$ in a zig-zag order (the scanning path of the zig-zag order is given at the lower right corner of Fig. 2). Then we have

$$\boldsymbol{D}=[\tilde{\boldsymbol{d}}_1,\ldots,\tilde{\boldsymbol{d}}_n,\ldots,\tilde{\boldsymbol{d}}_N]^{\mathrm{T}}. \tag{A8}$$