

Double-Layer Video Transmission Over Decode-and-Forward Wireless Relay Networks Using Hierarchical Modulation

Tu V. Nguyen, *Member, IEEE*, Pamela C. Cosman, *Fellow, IEEE*, and Laurence B. Milstein, *Fellow, IEEE*

Abstract—We consider a wireless relay network with a single source, a single destination, and a multiple relay. The relays are half-duplex and use the decode-and-forward protocol. The transmit source is a layered video bitstream, which can be partitioned into two layers, a base layer (BL) and an enhancement layer (EL), where the BL is more important than the EL in terms of the source distortion. The source broadcasts both layers to the relays and the destination using hierarchical 16-QAM. Each relay detects and transmits successfully decoded layers to the destination using either hierarchical 16-QAM or QPSK. The destination can thus receive multiple signals, each of which can include either only the BL or both the BL and the EL. We derive the optimal linear combining method at the destination, where the uncoded bit error rate is minimized. We also present a suboptimal combining method with a closed-form solution, which performs very close to the optimal. We use the proposed double-layer transmission scheme with our combining methods for transmitting layered video bitstreams. Numerical results show that the double-layer scheme can gain 2–2.5 dB in channel signal-to-noise ratio or 5–7 dB in video peak signal-to-noise ratio, compared with the classical single-layer scheme using conventional modulation.

Index Terms—Relay networks, decode-and-forward relaying, layered video transmission, hierarchical modulation, maximal ratio combining.

I. INTRODUCTION

MULTIMEDIA transmission has many challenges, such as the needs for large bandwidth and low latency, and the transmitted signals over wireless networks experience large fading fluctuations. Many techniques have been studied for reliable multimedia delivery over wireless links, such as multiple-input-multiple-output (MIMO), cooperation, and unequal error protection (UEP) schemes. MIMO schemes have been extensively studied in recent years, and widely used to achieve spatial diversity and/or multiplexing gain for data transmission by employing multiple antennas at both transmitter and receiver [1]–[3]. However, in many scenarios,

multiple antennas cannot be deployed, due to the size limitation of the end-user devices. In cooperation schemes, single antenna terminals can achieve spatial diversity by sharing their antennas with each other to create a virtual MIMO system [4]–[11]. Cooperation diversity can improve system performance and increase coverage [10], [11]. A simple and efficient cooperation scheme, namely, choosing the best relay, was analyzed in [12], and shown to achieve full diversity gain while maintaining high spectral efficiency. In this scheme, and other cooperation schemes in general, slow fading is assumed, and instantaneous channel state information (CSI) is generally used for relay selection/cooperation purposes.

For successive refinement sources [13], [14], or layered multimedia bitstreams, bits have different importance in terms of source distortion. Hence, UEP has been used for more reliable transmission. There are several methods providing UEP for layered bitstream transmission, such as rate-compatible punctured convolutional/turbo coding [15], [16] and hierarchical modulation [17], [18]. Hierarchical modulation has been used in commercial standards such as Digital Video Broadcasting [19] and MediaFlo [20] to provide multiple levels of quality of service (QoS).

Combinations of these methods were also considered to gain better transmission quality. In [21]–[24], several UEP methods were proposed for multimedia data over MIMO systems. For example, in [21], closed-loop MIMO with UEP using a scalable video encoder was considered. In [22], channel coding and spatial diversity were exploited to achieve UEP. In [25] and [26], the authors proposed layered video coding and two-hop relaying to deliver video to two groups of users with different QoS, depending on the weakest instantaneous channel signal-to-noise ratio (SNR) in each group. In [27] and [28], source-channel rate optimization for block fading channels was considered.

We consider wireless relay networks with a single source, a single destination, and multiple relays, all of which are equipped with a single antenna. In this work, we assume that instantaneous CSI is only available at the corresponding receiver. Neither packet acknowledgement nor retransmission is used. Also, the relays cannot communicate with each other for cooperation purposes. These assumptions are made in order to reduce system design complexity and extra latency. However, we assume a low-rate channel to feed back the channel SNRs of all links to the source. The source, which can be, e.g., a base station or an access point, optimizes system parameters, such as modulation order, based on the

Manuscript received April 12, 2013; revised September 14, 2013 and December 20, 2013; accepted February 13, 2014. Date of current version March 7, 2014. This work was supported in part by the Center for Wireless Communications of University of California at San Diego, La Jolla, in part by the UC Discovery Grant Program, and in part by the Vietnam Education Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sanghoon Lee.

T. V. Nguyen is with the Mobile and Wireless Group, Broadcom Corporation, San Diego, CA 92127 USA (e-mail: tu.vnguyen@yahoo.com).

P. C. Cosman and L. B. Milstein are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: pcosman@ucsd.edu; milstein@ece.ucsd.edu).

Digital Object Identifier 10.1109/TIP.2014.2308428

channel SNRs, and broadcasts these parameters to all other nodes either in a dedicated packet or in a packet header. A layered video bitstream, encoded using the H.264/AVC reference encoder [29]–[32], is transmitted from the source to the destination using the help of multiple relays.

In our proposed scheme, hierarchical QAM modulation will be used to provide UEP for a layered video bitstream. We assume the layered video bitstream can be partitioned into two layers, called a base layer (BL) and an enhancement layer (EL), where the BL is more important than the EL. The relays adaptively use different modulation schemes, depending on the number of successfully decoded layers, to forward the successfully decoded layers to the destination using frequency-orthogonal channels. The destination receives multiple noisy signals which include either both the BL and the EL or only the BL. To decode the BL and the EL, a linear combining method is used. We will see later that, because different modulation schemes are used at the relays, classical maximal ratio combining (MRC) [33], [34], which is optimal in terms of maximizing combined SNR for uncorrelated signal and uncorrelated noise branches, as well as its extensions [35]–[37], cannot straightforwardly be applied at the destination. In [38] and [39], simple combining methods are used; however, both are suboptimal and perform significantly worse than the optimal.

In this work, we derive the optimal linear combining weight vectors for the BL and the EL by a two-step combining method, where the optimality is in terms of minimizing the uncoded BER. Numerical results show that our proposed double-layer scheme using hierarchical 16-QAM significantly outperforms a classical single-layer scheme using conventional modulation. For example, approximately 2–2.5dB gain in channel SNR, or 5–7dB gain in video peak-signal-to-noise ratio (PSNR), can be observed.

The rest of this paper is organized as follows. In Section II, we present the system model. The novel combining technique and the system performance are presented in Section III. The application to transmission of layered video bitstreams is presented in Section IV, and Section V concludes the paper.

II. SYSTEM MODEL

We consider a single source, a single destination, and N relays, all of which are equipped with a single antenna, as shown in Fig. 1. The relays are half-duplex, i.e., the relays cannot transmit and receive at the same time, and use the decode-and-forward protocol [7]. Due to the half-duplex relay, each transmission requires two time slots. In the first time slot, the source broadcasts a message to all the relays and the destination. We assume the relays are not able to communicate with each other; hence, they do not know if any other relay successfully decodes the message. Rather, if the relays decode the message (or a portion of it) successfully, they forward it to the destination in the second time slot. We assume the relays communicate with the destination using orthogonal channels. Similar to [39], we assume the entire bandwidth is equally divided into N sub-bands. One sub-band is allocated to each relay for communicating with the destination, as shown

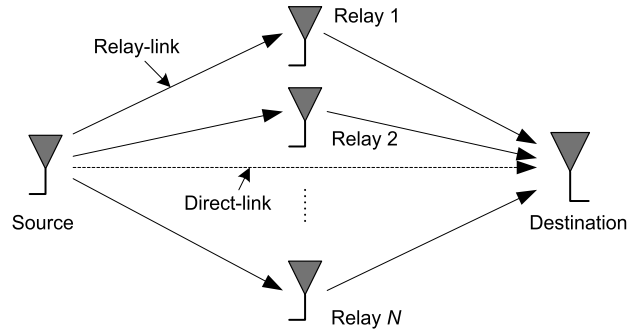


Fig. 1. Relay network: a single source, a single destination, and multiple relays.

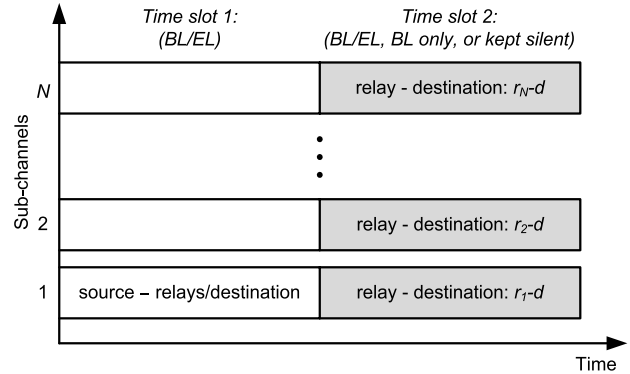


Fig. 2. Relay protocol description: the transmission bandwidth is equally divided into N sub-bands, and the relays use half-duplex protocol.

in Fig. 2. The broadcast channel from the source to the relays can use any one of the sub-bands, since it is transmitting in a different time slot.

A. Channel Model

We assume the channels from the source to the relays and from the relays to the destination experience flat Rayleigh fading, and we use the modified Jakes' model [41] to simulate different fading rates. Due to the spatial separation, we also assume that all the channels from the source to the relays and from the relays to the destination are independent. We assume that the channel gain is constant for each symbol, and that it can be accurately estimated at the receiver. However, the channel gain is assumed to be unknown at the transmitter.

B. Source Model

We consider the transmission of a layered video bitstream, encoded with a H.264/AVC encoder [29]–[32]. The bitstream can be partitioned into two layers, a BL and an EL. The details about the layer partition will be presented in Section IV. The BL is more important than the EL in terms of the quality of the reconstructed source at the receiver. Hence, the BL generally needs higher protection against transmission errors. In this work, we adopt hierarchical 16-QAM modulation [18], [42] to provide unequal error protection.¹

¹We note that the method proposed in this paper can be generalized to higher order modulation schemes such as hierarchical 64-QAM or 256-QAM with 3 or 4 source layers, respectively.

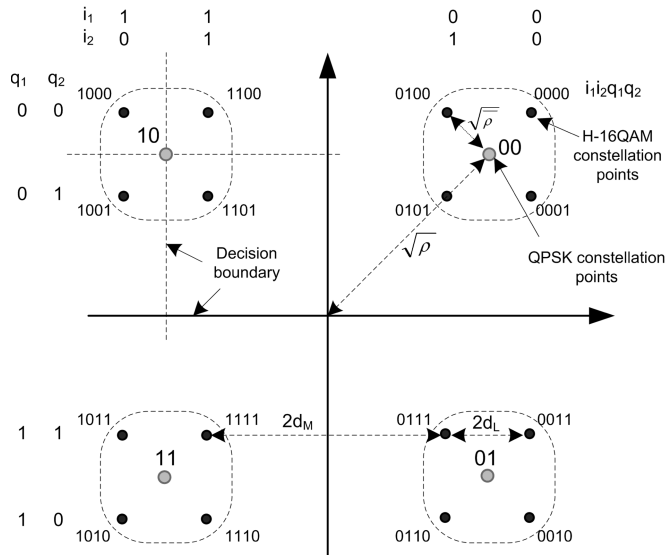


Fig. 3. Hierarchical 16-QAM constellation as superposition mapping of two QPSK constellations.

C. Transmission Schemes

1) *Classical Single-Layer Scheme - Baseline*: For the classical single-layer scheme, we assume the system is unaware of the source's information; hence, it considers the source as an i.i.d. bitstream. Classical 16-QAM with Gray-coded bit-mapping is used, where all bits in a symbol are considered to be of the same importance. This is referred to as the equal error protection (EEP) scheme. In the first time slot, the source encodes and broadcasts a message to all the relays. The relays separately decode the message. If successful, they re-encode and forward it to the destination using the same modulation scheme. At the destination, since the received signals use the same modulation scheme, a maximum-ratio-combining (MRC) receiver can be used to combine all the received signals to decode the message.

2) *Proposed Double-Layer Scheme*: In the first time slot, the source encodes and broadcasts a message to all the relays and the destination. The first half of the message contains the BL and the second half of the message contains the EL. We use hierarchical 16-QAM as the modulation scheme, where the BL is mapped into the most significant bits (MSBs) and the EL is mapped into the least significant bits (LSBs), as shown in Fig. 3. The relays separately decode the message. If the BL is decoded successfully, an attempt is made to decode the EL. Depending on the channel quality from the source to the relays, and the power allocation parameter for the hierarchical 16-QAM, a relay can successfully decode both the BL and the EL (the BL/EL), only the BL, or neither. In the second time slot, the relays re-encode and forward any successfully decoded layers to the destination. The hierarchical 16-QAM modulation scheme is used if both layers were successfully received. If, however, only the BL is successfully received at a relay, it will transmit only the BL to the destination using conventional QPSK. Lastly, the relay remains silent if no layer was successfully decoded. In this

paper, we propose combining methods to first detect the BL and then to detect the EL, which minimizes the uncoded BER.

We note that because of the 16-QAM modulation, we need the same number of bits in the two priority classes. In general case, we do not have the same number of bits in the two classes, so some method is needed to form symbols. If the two classes have differing numbers of bits, we can use zero padding to lengthen the shorter stream to equal the longer stream. Alternatively, we can use the double-layer approach for as many bits as are paired between the two streams, and then revert to a single layer transmission for the excess bits. In this paper, we partition the single stream into two halves which contain roughly equal numbers of bits, and then we use zero padding to force the two streams to be the same.

D. Signal Model at the Relays and the Destination

Since the mathematical representation of the single-layer scheme with conventional 16-QAM is straightforward, in the following, we focus on the double-layer scheme with hierarchical 16-QAM modulation.

We consider hierarchical 16-QAM modulation using Gray-coded bit mapping, as shown in Fig. 3. We can express a hierarchical 16-QAM symbol, denoted by A_l , as the weighted sum, or superposition, of two QPSK symbols as follows [43]:

$$A_l = \sqrt{\rho}b_l + \sqrt{\bar{\rho}}e_l, \quad (1)$$

where b_l and e_l denote two QPSK-modulated symbols, which depend on the MSBs and the LSBs, respectively, of the hierarchical 16-QAM symbol. We use $\rho \in (0.5, 1]$ to denote the normalized power allocated to the MSB signal b_l , and $\bar{\rho} \triangleq 1 - \rho$ to denote the normalized power allocated to the LSB signal e_l .²

In the first time slot, the source broadcasts the layered video bitstream including the BL/EL to all the relays using the hierarchical 16-QAM scheme using a fixed sub-band. The received signal sample at the n -th relay (at the output of a matched filter) can be written as

$$y_{n,l}^{(r)} = \alpha_{n,l}^{sr} \sqrt{2E_s} (\sqrt{\rho}b_l + \sqrt{\bar{\rho}}e_l) + z_{n,l}^{(r)}, \quad n=1, 2, \dots, N, \quad (2)$$

where $\alpha_{n,l}^{sr}$ is the channel gain at the sampling time lT_s , which is assumed to be real-valued and non-negative (i.e., a coherent receiver is assumed), with T_s being the sampling period. In (2), E_s denotes the transmitted symbol energy at the source, b_l and e_l denote the BL and the EL modulated signals at time lT_s , respectively, each of which is a QPSK-modulated symbol with unity power, and ρ is the power allocation parameter (see Fig. 3). The terms $\{z_{n,l}^{(r)}\}$ are assumed to be independently and identically distributed (i.i.d.) and circularly symmetric complex Gaussian noise $\mathcal{CN}(0, 2N_0)$. Similarly, the received signal at the destination in the first time slot is given by

$$y_{0,l}^{(d)} = \alpha_l^{sd} \sqrt{2E_s} (\sqrt{\rho}b_l + \sqrt{\bar{\rho}}e_l) + z_{0,l}^{(d)}. \quad (3)$$

²Relative to the power allocation ratio in [18], we have $\alpha = d_M/d_L = (\sqrt{\rho} - \sqrt{\bar{\rho}})/\sqrt{\bar{\rho}}$ (where $\bar{\rho} = 1 - \rho$), that is, e.g., if $\rho = 0.70$ then $\alpha \approx 0.528$, if $\rho = 0.80$ then $\alpha = 1.0$ (i.e., conventional constellation), and if $\rho = 0.90$ then $\alpha = 2.0$.

In the second time slot, depending on the number of successfully decoded layers at the relays, the destination can receive different signals in each sub-band. If the n -th relay transmitted the BL/EL using hierarchical 16-QAM, the received signal in the n -th sub-band at the destination is given by

$$y_{n,l}^{(d)} = \alpha_{n,l}^{rd} \sqrt{2E_r} (\sqrt{\rho} b_l + \sqrt{\bar{\rho}} e_l) + z_{n,l}^{(d)}, \quad (4)$$

or, if only the BL is transmitted (using QPSK),

$$y_{n,l}^{(d)} = \alpha_{n,l}^{rd} \sqrt{2E_r} b_l + z_{n,l}^{(d)}, \quad (5)$$

where E_r denotes the average transmit symbol energy at the relays. Note that all the power is allocated to the BL when QPSK is used. In (4) and (5), similarly, we assume the $\{z_{n,l}^{(d)}\}$ are i.i.d. and complex Gaussian noise $\mathcal{CN}(0, 2N_0)$. For combining and detection at the destination, in practice, each relay needs to send a signalling message (of two bits) to inform the destination which type of signal it will send. For simplicity, we assume the signalling message is received at the destination error-free, and ignore the signalling overhead.

Due to the spatial separation, we assume the fading gains $\alpha_{n,l}^{sr}$, $\alpha_{n,l}^{rd}$, and α_l^{sd} are independent, with second moments Ω_n^{sr} , Ω_n^{rd} , and Ω^{sd} , respectively. For notational simplicity, we also consider the source itself as relay 0, and denote $\alpha_{0,l}^{rd} = \sqrt{E_s/E_r} \alpha_l^{sd}$, and thus $\Omega_0^{rd} = \Omega^{sd} E_s/E_r$, such that the received signal at the destination in the first time slot, as shown in (3), can be written as

$$y_{0,l}^{(d)} = \alpha_{0,l}^{rd} \sqrt{2E_r} (\sqrt{\rho} b_l + \sqrt{\bar{\rho}} e_l) + z_{0,l}^{(d)}, \quad (6)$$

which is (4) for $n = 0$.

Since the destination can receive two types of signals, one of which includes the BL/EL using hierarchical 16-QAM and the other includes only the BL using QPSK, the classical MRC receiver cannot be applied straightforwardly. In [39], the received signals that include only the BL, if at least one is available, are combined to detect the BL. If no such signal is available, that is, all the received signals include both layers, then an MRC receiver is used to combine the received signals to detect the BL (and the EL). Clearly, this is a suboptimal receiver, because not all of the available received signals were exploited. In [38], another combining method was used which is equivalent to an MRC receiver when all the transmitted signals are identical, i.e., all are either hierarchical 16-QAM or QPSK. This is again a suboptimal method.

III. PROPOSED COMBINING METHODS AT THE DESTINATION

Let Θ_k and Ψ_k , $k = 1, 2, \dots, 3^N$, be all possible pairs of subsets of the received signals at the destination which include only the BL or both the BL/EL, respectively. The received signals are indexed by an integer between 0 and N which corresponds to the transmit terminal indices. That is, when the direct link is not used, (Θ_k, Ψ_k) are subsets of the set of all relay indices $\{1, 2, \dots, N\}$, with $\Theta_k \cap \Psi_k = \emptyset$.³ Alternatively,

³For example, for $N = 2$, and the direct-link is not used, all the possible pairs (Θ_k, Ψ_k) , for $k = 1, 2, \dots, 9$, are given as follows: (\emptyset, \emptyset) , $(\emptyset, \{1\})$, $(\emptyset, \{2\})$, $(\emptyset, \{1, 2\})$, $(\{1\}, \emptyset)$, $(\{1\}, \{2\})$, $(\{2\}, \emptyset)$, $(\{2\}, \{1\})$, and $(\{1, 2\}, \emptyset)$.

when the direct link is used, (Θ_k, Ψ_k) are subsets of the set $\{0, 1, 2, \dots, N\}$, with $\Theta_k \cap \Psi_k = \emptyset$ and $0 \in \Psi_k$, where we use index 0 to denote the source. In both cases, the index k is in the range of $1, 2, \dots, 3^N$. Note that those received signals at the destination in the sets Θ_k and Ψ_k are given in (5) and (4), respectively.

In the following, we consider both the optimal and a suboptimal method to combine the received signals in the sets Θ_k and Ψ_k to decode the BL/EL signals. Optimality is defined in terms of minimizing the uncoded bit error rate (BER). We note that if either Θ_k or Ψ_k is an empty set, i.e., only one kind of signal is received, the classical MRC receiver can be applied, and it is optimal. Thus, in the following, we consider the case that both sets are nonempty.

A. Combined Signals and Optimization Problem

Given the received signals $y_{n,l}^{(d)}$ for $n \in \Theta_k$, as shown in (5), and $y_{n,l}^{(d)}$ for $n \in \Psi_k$, as shown in (4), we need to combine them to detect the BL and the EL. Generally, the optimal weights for detecting the BL and the EL are different. We first consider the combining for the BL (the combining for the EL is done in a similar manner). Let w_n be the weight corresponding to the received signal from the n -th relay. The combined signal for the BL at the destination is given by

$$y_{l,bl}^{(d)} = \sum_{n \in \Theta_k \cup \Psi_k} w_n y_{n,l}^{(d)} = \sum_{n \in \Theta_k} w_n y_{n,l}^{(d)} + \sum_{n \in \Psi_k} w_n y_{n,l}^{(d)}, \quad (7)$$

where, as noted above, $\Theta_k \cap \Psi_k = \emptyset$. Substituting $y_{n,l}^{(d)}$ from (5) and (4) for $n \in \Theta_k$ and $n \in \Psi_k$, respectively, we can write the combined signal $y_{l,bl}^{(d)}$ as follows:

$$y_{l,bl}^{(d)} = \sqrt{2E_r} [C_{\Theta_k}(\mathbf{w}_{\Theta_k}) + C_{\Psi_k}(\mathbf{w}_{\Psi_k}) \sqrt{\rho}] b_l + \sqrt{2E_r} [C_{\Psi_k}(\mathbf{w}_{\Psi_k}) \sqrt{\bar{\rho}}] e_l + N_{\Theta_k}(\mathbf{w}_{\Theta_k}) + N_{\Psi_k}(\mathbf{w}_{\Psi_k}), \quad (8)$$

where \mathbf{w}_{Θ_k} and \mathbf{w}_{Ψ_k} denote the vectors whose elements are w_n for $n \in \Theta_k$ and $n \in \Psi_k$, respectively, and similarly for α_{Θ_k} , α_{Ψ_k} , \mathbf{z}_{Θ_k} , and \mathbf{z}_{Ψ_k} . Also, we defined

$$C_{\Theta_k}(\mathbf{w}_{\Theta_k}) \triangleq \mathbf{w}_{\Theta_k}^T \alpha_{\Theta_k}, \quad C_{\Psi_k}(\mathbf{w}_{\Psi_k}) \triangleq \mathbf{w}_{\Psi_k}^T \alpha_{\Psi_k}, \quad (9)$$

$$N_{\Theta_k}(\mathbf{w}_{\Theta_k}) \triangleq \mathbf{w}_{\Theta_k}^T \mathbf{z}_{\Theta_k}, \quad N_{\Psi_k}(\mathbf{w}_{\Psi_k}) \triangleq \mathbf{w}_{\Psi_k}^T \mathbf{z}_{\Psi_k}, \quad (10)$$

where the superscript T denotes transpose. From (10), we note that $N_{\Theta_k}(\mathbf{w}_{\Theta_k}) \sim \mathcal{CN}(0, 2N_0 \|\mathbf{w}_{\Theta_k}\|^2)$ and $N_{\Psi_k}(\mathbf{w}_{\Psi_k}) \sim \mathcal{CN}(0, 2N_0 \|\mathbf{w}_{\Psi_k}\|^2)$, where we use $\|\mathbf{x}\| \triangleq \sqrt{\mathbf{x}^T \mathbf{x}}$ to denote the Euclidean norm of a column vector \mathbf{x} . Since $\Theta_k \cap \Psi_k = \emptyset$, N_{Θ_k} and N_{Ψ_k} are independent.

From (10), we note that if we change the sign of an element of the weight vector \mathbf{w}_{Θ_k} (or \mathbf{w}_{Ψ_k}), the statistics of the combining noise $N_{\Theta_k}(\mathbf{w}_{\Theta_k})$ (or $N_{\Psi_k}(\mathbf{w}_{\Psi_k})$) do not change. Since all the channel gains $\{\alpha_{n,l}^{rd}\}$ are non-negative, from (9), we observe that the optimal weights $\{w_n\}$ must be non-negative, otherwise we can change the signs of the negative weights to increase the signal component by increasing either $C_{\Theta_k}(\mathbf{w}_{\Theta_k})$ or $C_{\Psi_k}(\mathbf{w}_{\Psi_k})$, as shown in (8), while not changing the noise statistics. Hence, the optimal weight vectors \mathbf{w}_{Θ_k} and \mathbf{w}_{Ψ_k} must have all non-negative elements, which we denote by $\mathbf{w}_{\Theta_k} \geq 0$ and $\mathbf{w}_{\Psi_k} \geq 0$. As a result, both $C_{\Theta_k}(\mathbf{w}_{\Theta_k})$ and

$C_{\Psi_k}(\mathbf{w}_{\Psi_k})$ are non-negative. Further, we note that if either $\mathbf{w}_{\Theta_k} = \mathbf{0}$ or $\mathbf{w}_{\Psi_k} = \mathbf{0}$, the combined weight vectors cannot be optimal because we do not use all the received signals for detection. Hence, in the following, we assume both $\mathbf{w}_{\Theta_k} \neq \mathbf{0}$ and $\mathbf{w}_{\Psi_k} \neq \mathbf{0}$ when solving for the optimal weight vectors. For \mathbf{w}_{Θ_k} , $\mathbf{w}_{\Psi_k} \succeq 0$, and which are not equal to the zero vector, we can write

$$\mathbf{w}_{\Theta_k} = a_{\Theta_k} \tilde{\mathbf{w}}_{\Theta_k}, \quad \mathbf{w}_{\Psi_k} = a_{\Psi_k} \tilde{\mathbf{w}}_{\Psi_k}, \quad (11)$$

for $a_{\Theta_k}, a_{\Psi_k} > 0$, $\tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k} \succeq 0$, and $\|\tilde{\mathbf{w}}_{\Theta_k}\| = \|\tilde{\mathbf{w}}_{\Psi_k}\| = 1$. We will refer to $\tilde{\mathbf{w}}_{\Phi}$ as the vector direction, and to a_{Φ} as vector length of the vector $\mathbf{w}_{\Phi} = a_{\Phi} \tilde{\mathbf{w}}_{\Phi}$. As we will see later, the change of variables in (11) allows us to solve the optimization problem in two steps.

Note that the combined signal in (8) is a noisy hierarchical 16-QAM symbol. The conditional (uncoded) BER expressions, conditioned on the channel gains, for the BL as follows [18], [42]:

$$\begin{aligned} BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k}) &\triangleq BER_{BL}(a_{\Theta_k} \tilde{\mathbf{w}}_{\Theta_k}, a_{\Psi_k} \tilde{\mathbf{w}}_{\Psi_k}) \\ &= \frac{1}{2} Q \left(\frac{\sqrt{E_r} a_{\Theta_k} C_{\Theta_k}(\tilde{\mathbf{w}}_{\Theta_k}) + \sqrt{E_r} a_{\Psi_k} C_{\Psi_k}(\tilde{\mathbf{w}}_{\Psi_k}) (\sqrt{\rho} - \sqrt{\bar{\rho}})}{\sqrt{(a_{\Theta_k}^2 + a_{\Psi_k}^2) N_0}} \right) \\ &\quad + \frac{1}{2} Q \left(\frac{\sqrt{E_r} a_{\Theta_k} C_{\Theta_k}(\tilde{\mathbf{w}}_{\Theta_k}) + \sqrt{E_r} a_{\Psi_k} C_{\Psi_k}(\tilde{\mathbf{w}}_{\Psi_k}) (\sqrt{\rho} + \sqrt{\bar{\rho}})}{\sqrt{(a_{\Theta_k}^2 + a_{\Psi_k}^2) N_0}} \right) \end{aligned} \quad (12)$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-u^2/2) du$, and noting from (9) that $C_{\Phi}(\mathbf{x}) = C_{\Phi}(\alpha \tilde{\mathbf{x}}) = \alpha C_{\Phi}(\tilde{\mathbf{x}})$ for $\Phi \in \{\Theta_k, \Psi_k\}$.

A similar procedure can be applied to combine the signal for the EL. The final conditional BER expression, conditioned on the channel gains, for the EL is given by [18], [42]

$$\begin{aligned} BER_{EL}(b_{\Theta_k}, b_{\Psi_k}, \tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k}) &\triangleq BER_{EL}(b_{\Theta_k} \tilde{\mathbf{v}}_{\Theta_k}, b_{\Psi_k} \tilde{\mathbf{v}}_{\Psi_k}) \\ &= Q \left(\frac{\sqrt{E_r} b_{\Psi_k} C_{\Psi_k}(\tilde{\mathbf{v}}_{\Psi_k}) \sqrt{\bar{\rho}}}{\sqrt{(b_{\Theta_k}^2 + b_{\Psi_k}^2) N_0}} \right) \\ &\quad + \frac{1}{2} Q \left(\frac{\sqrt{E_r} [2b_{\Theta_k} C_{\Theta_k}(\tilde{\mathbf{v}}_{\Theta_k}) + b_{\Psi_k} C_{\Psi_k}(\tilde{\mathbf{v}}_{\Psi_k}) (2\sqrt{\rho} - \sqrt{\bar{\rho}})]}{\sqrt{(b_{\Theta_k}^2 + b_{\Psi_k}^2) N_0}} \right) \\ &\quad - \frac{1}{2} Q \left(\frac{\sqrt{E_r} [2b_{\Theta_k} C_{\Theta_k}(\tilde{\mathbf{v}}_{\Theta_k}) + b_{\Psi_k} C_{\Psi_k}(\tilde{\mathbf{v}}_{\Psi_k}) (2\sqrt{\rho} + \sqrt{\bar{\rho}})]}{\sqrt{(b_{\Theta_k}^2 + b_{\Psi_k}^2) N_0}} \right) \end{aligned} \quad (13)$$

where for $b_{\Theta_k}, b_{\Psi_k} > 0$, $\tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k} \succeq 0$, and $\|\tilde{\mathbf{v}}_{\Theta_k}\| = \|\tilde{\mathbf{v}}_{\Psi_k}\| = 1$.

Since the constraints $a_{\Theta_k}, a_{\Psi_k} > 0$ and $\tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k} \succeq 0$, $\|\tilde{\mathbf{w}}_{\Theta_k}\| = \|\tilde{\mathbf{w}}_{\Psi_k}\| = 1$ are separate constraints, we can first optimize over $\tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k}$ and then over a_{Θ_k}, a_{Ψ_k} . Our optimization problem for the BL is formally written as follows:

$$BER_{BL}^* = \min_{\substack{a_{\Theta_k} > 0 \\ a_{\Psi_k} > 0}} \min_{\substack{\tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k} \succeq 0 \\ \|\tilde{\mathbf{w}}_{\Theta_k}\| = \|\tilde{\mathbf{w}}_{\Psi_k}\| = 1}} BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k}). \quad (14)$$

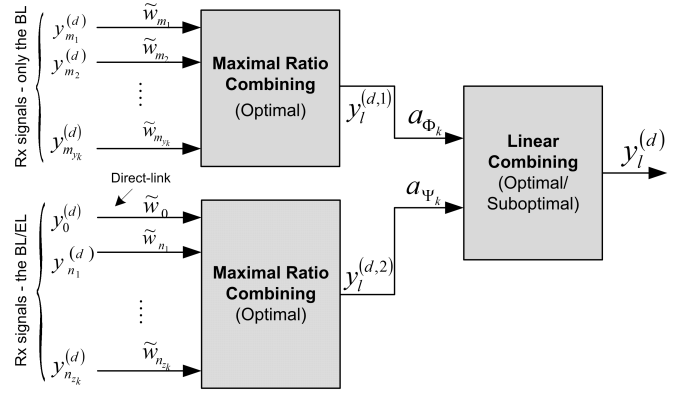


Fig. 4. Two-step combining procedure for the BL (similarly for the EL).

Similarly, for the EL,

$$BER_{EL}^* = \min_{\substack{b_{\Theta_k} > 0 \\ b_{\Psi_k} > 0}} \min_{\substack{\tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k} \succeq 0 \\ \|\tilde{\mathbf{v}}_{\Theta_k}\| = \|\tilde{\mathbf{v}}_{\Psi_k}\| = 1}} BER_{EL}(b_{\Theta_k}, b_{\Psi_k}, \tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k}). \quad (15)$$

B. The First Combining Step: Optimal for Both BL/EL

In the following, we consider the inner optimization problems, as shown in (14) and (15). In Appendix A, we show that

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k} \succeq 0 \\ \|\tilde{\mathbf{w}}_{\Theta_k}\| = \|\tilde{\mathbf{w}}_{\Psi_k}\| = 1}} BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k}) \\ = BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\mathbf{w}}_{\Theta_k}^*, \tilde{\mathbf{w}}_{\Psi_k}^*), \end{aligned} \quad (16)$$

where

$$\begin{aligned} (\tilde{\mathbf{w}}_{\Theta_k}^*, \tilde{\mathbf{w}}_{\Psi_k}^*) &= (\tilde{\alpha}_{\Theta_k}, \tilde{\alpha}_{\Psi_k}) \\ &\triangleq (\alpha_{\Theta_k} / \|\alpha_{\Theta_k}\|, \alpha_{\Psi_k} / \|\alpha_{\Psi_k}\|). \end{aligned} \quad (17)$$

Similarly, we have

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k} \succeq 0 \\ \|\tilde{\mathbf{v}}_{\Theta_k}\| = \|\tilde{\mathbf{v}}_{\Psi_k}\| = 1}} BER_{EL}(b_{\Theta_k}, b_{\Psi_k}, \tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k}) \\ = BER_{EL}(b_{\Theta_k}, b_{\Psi_k}, \tilde{\mathbf{v}}_{\Theta_k}^*, \tilde{\mathbf{v}}_{\Psi_k}^*), \end{aligned} \quad (18)$$

where

$$(\tilde{\mathbf{v}}_{\Theta_k}^*, \tilde{\mathbf{v}}_{\Psi_k}^*) = (\tilde{\alpha}_{\Theta_k}, \tilde{\alpha}_{\Psi_k}). \quad (19)$$

Both of the inner optimization problems, as shown in (14) and (15), have the same optimal solutions $(\tilde{\mathbf{w}}_{\Theta_k}^*, \tilde{\mathbf{w}}_{\Psi_k}^*) = (\tilde{\mathbf{v}}_{\Theta_k}^*, \tilde{\mathbf{v}}_{\Psi_k}^*) = (\tilde{\alpha}_{\Theta_k}, \tilde{\alpha}_{\Psi_k})$, which depend on the channel gains, but not on either $(a_{\Theta_k}, a_{\Psi_k})$ or $(b_{\Theta_k}, b_{\Psi_k})$. Also, the optimal solutions have the same form as the MRC solutions for combining the received signals in the sets Θ_k and Ψ_k . Thus, we can use two MRC receivers for combining the received signals in the sets Θ_k and Ψ_k , separately, as illustrated in Fig. 4.

Next, we minimize the BER expressions of the BL and the EL, as shown on the right hand side of (16) and (18), over $(a_{\Theta_k}, a_{\Psi_k})$ and $(b_{\Theta_k}, b_{\Psi_k})$, respectively (see also (12) and (13) for the detailed BER expressions). In the following, we solve these problems separately.

C. The Second Combining Step

1) Combining Methods for the BL:

a) *Optimal solution:* In Appendix B, we show that the optimal weights in the second step are

$$a_{\Theta_k}^* = 1, \quad a_{\Psi_k}^* = \tan(\phi^*), \quad (20)$$

where ϕ^* is the solution of the following convex optimization problem:

$$\phi^* = \arg \min_{\phi \in (0, \pi/2)} \frac{1}{2} [Q(A \cos \phi + B \sin \phi) + Q(A \cos \phi + C \sin \phi)] \quad (21)$$

where $A \geq 0$ and $C \geq B \geq 0$ are dependent on the channel gains, as defined in (B.2).

b) *Suboptimal solution – closed-form:* Since the optimal solution was not obtained in closed form, in the following, we present a suboptimal solution which is obtained in closed form. We will see later that the suboptimal method performs very close to the optimal one.

In Appendix C, we obtain suboptimal weights ($a_{\Theta_k}^\dagger, a_{\Psi_k}^\dagger$), which minimize an upper bound of the BER of the BL. Note that the superscript \dagger denotes a suboptimal value for the corresponding quantity. The result is given as follows

$$a_{\Theta_k}^\dagger = \|\alpha_{\Theta_k}\|, \quad a_{\Psi_k}^\dagger = \|\alpha_{\Psi_k}\|(\sqrt{\rho} - \sqrt{\bar{\rho}}). \quad (22)$$

Substituting the suboptimal weights from (22) into (B.1), the BER of the BL is given by

$$\begin{aligned} BER_{BL}^\dagger &= \frac{1}{2} Q \left(\sqrt{\frac{E_r (C_{\Theta_k}^*)^2}{N_0} + \frac{E_r (C_{\Psi_k}^*)^2 (\sqrt{\rho} - \sqrt{\bar{\rho}})^2}{N_0}} \right) \\ &+ \frac{1}{2} Q \left(\sqrt{\frac{E_r [(C_{\Theta_k}^*)^2 + (C_{\Psi_k}^*)^2 (\rho - \bar{\rho})]^2}{[(C_{\Theta_k}^*)^2 + (C_{\Psi_k}^*)^2 (\sqrt{\rho} - \sqrt{\bar{\rho}})^2] N_0}} \right), \quad (23) \end{aligned}$$

where $C_{\Theta_k}^*$ and $C_{\Psi_k}^*$ are given in (A.1).

For comparison purposes, we define the effective output (instantaneous) SNR to be the square of the argument of the Q -function of the dominant term of BER. From (23), we have

$$SNR_{BL,1}^{(eff)} = \frac{E_r \cdot (C_{\Theta_k}^*)^2}{N_0} + \frac{E_r \cdot (C_{\Psi_k}^*)^2 (\sqrt{\rho} - \sqrt{\bar{\rho}})^2}{N_0}. \quad (24)$$

We note that if both $C_{\Theta_k}^*$ and $C_{\Psi_k}^*$ are not equal to zero, our suboptimal method results in a strictly higher effective SNR than that of the simple combining technique in [39], which results in only the first term in (24). Also, the combining method in [38] uses $w_n = \alpha_{n,l}^{r_d}$ for all n , i.e., $\mathbf{w}_{\Theta_k} = \alpha_{\Theta_k}$, $\mathbf{w}_{\Psi_k} = \alpha_{\Psi_k}$, as opposed to (C.4). Using $\mathbf{w}_{\Theta_k} = \alpha_{\Theta_k}$, $\mathbf{w}_{\Psi_k} = \alpha_{\Psi_k}$ in (12), we can find the corresponding BER expression for the BL. The effective SNR is given by

$$SNR_{BL,2}^{(eff)} = \frac{E_r [(C_{\Theta_k}^*)^2 + (C_{\Psi_k}^*)^2 (\sqrt{\rho} - \sqrt{\bar{\rho}})^2]}{[(C_{\Theta_k}^*)^2 + (C_{\Psi_k}^*)^2] N_0}, \quad (25)$$

which can be shown to be strictly less than that in (24) for $\rho < 1$ and both $C_{\Theta_k}^*, C_{\Psi_k}^* \neq 0$ (which happens with probability one).

2) *Combining Methods for the EL:* Similar to the BL case, from (13) and (18), we can show that the optimal weights in the second step for the EL are given by

$$b_{\Theta_k}^* = \tan \theta^*, \quad b_{\Psi_k}^* = 1, \quad (26)$$

where

$$\theta^* = \arg \min_{\theta \in (0, \pi/2)} BER_{EL}^{(a)}(\theta), \quad (27)$$

in which

$$\begin{aligned} BER_{EL}^{(a)}(\theta) &\triangleq Q(D \cos \theta) + \frac{1}{2} Q(E \sin \theta + F \cos \theta) \\ &- \frac{1}{2} Q(E \sin \theta + G \cos \theta) \quad (28) \end{aligned}$$

denotes the BER of the EL, as a function of θ , and $D \triangleq \sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* \sqrt{\rho}$, $E \triangleq 2 \sqrt{\frac{E_r}{N_0}} C_{\Theta_k}^*$, $F \triangleq \sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* (2\sqrt{\rho} - \sqrt{\bar{\rho}})$, and $G \triangleq \sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* (2\sqrt{\rho} + \sqrt{\bar{\rho}})$. We note, however, that the optimization problem in (27) is, in general, not convex. Therefore, a numerical search using, e.g., the Newton method [40], generally results in a local minimum.

Because of the difficulty in searching for a local (or global) minimum, we use a suboptimal combining method by letting $\theta^\dagger = 0$, so $b_{\Theta_k}^\dagger = 0$ to reduce the computational complexity [39]. That is, we only use the combined signals, which include the BL/EL for detecting the EL [39]. We show in Subsection III-D that this suboptimal method gives results very close to the optimal one. By substituting $b_{\Theta_k} = b_{\Theta_k}^\dagger = 0$, $b_{\Psi_k} = b_{\Psi_k}^\dagger = 1$ into (13), the corresponding BER expression for the EL is given by

$$\begin{aligned} BER_{EL}^\dagger &= Q \left(\sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* \sqrt{\rho} \right) + \frac{1}{2} Q \left(\sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* (2\sqrt{\rho} - \sqrt{\bar{\rho}}) \right) \\ &- \frac{1}{2} Q \left(\sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* (2\sqrt{\rho} + \sqrt{\bar{\rho}}) \right). \quad (29) \end{aligned}$$

For $b_{\Theta_k}^\dagger = 0, b_{\Psi_k}^\dagger = 1$, the suboptimal weights for the EL in both steps are given by (see (19) for the optimal weight vectors in the first step)

$$\mathbf{v}_{\Theta_k}^\dagger = \mathbf{0}, \quad \mathbf{v}_{\Psi_k}^\dagger = \alpha_{\Psi_k}, \quad (30)$$

where we have multiplied all the weights by $\|\alpha_{\Psi_k}\|$.

Since the Q -function monotonically decreases, and $G \geq F \geq 0$, from (28), we have a lower bound for the BER of the EL as follows:

$$BER_{EL}^{(a)}(\theta) \geq Q(D \cos \theta) \geq Q(D) = Q \left(\sqrt{\frac{E_r}{N_0}} C_{\Psi_k}^* \sqrt{\rho} \right), \quad (31)$$

for all $\theta \in (0, \pi/2)$. The lower bound will be used to compare with the suboptimal BER performance and will be shown to yield very close results.

Lastly, we note that for ease of implementation, we can design the optimal and suboptimal combining receivers by two steps as shown in Fig. 4. In the first step, we use two MRC receivers for the received signals which include only the BL and the BL/EL separately. This step is optimal individually for both the BL and the EL. In the second step, we combine

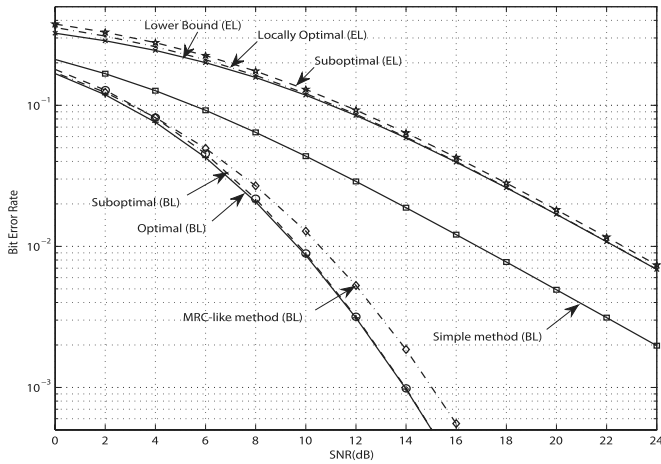


Fig. 5. Comparison of uncoded BER performance for the BL and the EL using various combining methods with power allocation parameter $\rho = 0.72$.

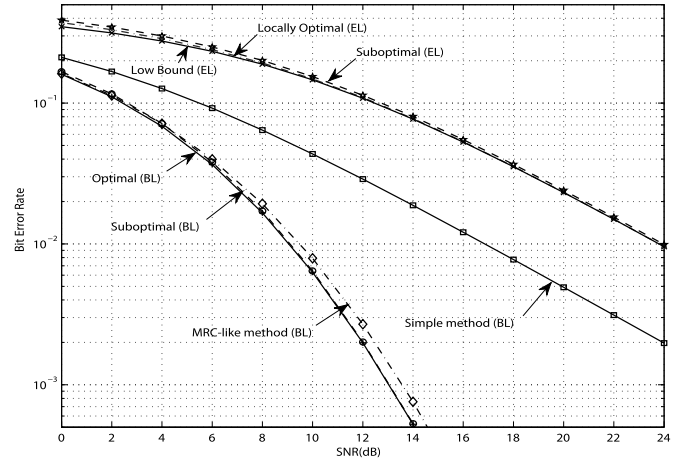


Fig. 6. Comparison of uncoded BER performance for the BL and the EL using various combining methods with power allocation parameter $\rho = 0.80$.

the two resulting signals to detect the BL and the EL, using either the optimal or the suboptimal weights.

D. Numerical BER Performance Comparison for the BL/EL

In the following, we compare the uncoded BER performances of the BL and the EL for the different combining methods above. For simplicity, we assume there are two relays. One is always sending the BL using QPSK, and the other is always sending the BL/EL using hierarchical 16-QAM. We assume the relays experience independent flat Rayleigh fading. The channel state information is perfectly known at the receiver. For the BL, we plot the uncoded BER performances for the combining methods in [39] and [38], and our optimal and suboptimal methods. For the EL, we plot our simple combining method, the locally suboptimal method, and the lower bound.

First, we consider the performances for the BL, as shown in Fig. 5, for the power allocation parameter $\rho = 0.72$, a value which will be of interest below. We note that the combining method in [39] does not perform as well as the other methods since it only uses the QPSK received signal to detect the BL, and thus results in just the QPSK BER performance (with diversity order 1). Our suboptimal combining method of minimizing an upper-bound BER significantly outperforms the combining method used in [38]. We can observe approximately 1dB gain in the medium and high SNR region, i.e., say, channel $SNR \geq 8$ dB. Our suboptimal method performs almost as well as the optimal in the medium and high SNR region. For a higher value of ρ , say, $\rho = 0.8$ (i.e., conventional constellation), as shown in Fig. 6, our suboptimal combining method performs almost identical to the optimal one over the range of SNR in the plot. The gain compared to the combining method in [38] reduces for a high value of ρ , but the proposed method is still much better than the method in [39], as can be seen in Fig. 6.

Next, we consider the performance for the EL. In Fig. 5, we observe that the suboptimal combining method performs very close to the locally optimal and the lower bound performances. For smaller ρ , the performance loss slightly increases.

For higher values of ρ , say $\rho = 0.8$, the performance loss is almost negligible, as seen in Fig. 6. We note that in all cases, the locally optimal performance is almost identical to the lower bound for, say, channel $SNR \geq 8$ dB, which suggests that the local optimum is very close, if not identical, to the global optimum in this SNR region.

In summary, for the BL, our suboptimal combining method performs very close to the optimal. For the EL, the suboptimal combining method performs very well compared to the local optimal and the lower-bound BER performance. Hence, in the numerical simulation in Section IV, we will use the proposed suboptimal combining methods that have the closed-form weights, instead of numerically solving for the optimal (or locally optimal) weights to reduce the computational complexity.

IV. APPLICATION TO VIDEO TRANSMISSION: SIMULATION RESULTS

In this section, we apply the transmission protocol and our proposed suboptimal and closed-form combining methods at the destination for transmitting an encoded video bitstream over the relay network.

A. Simulation Setup

In the simulation, we assume there are four relays, and the direct link is so weak that it will be ignored unless otherwise stated. All the channels are independent, with flat Rayleigh fading with normalized Doppler frequency $f_{dn} = 10^{-3}$ [41]. The average channel SNRs of all links from the source to the relays and from the relays to the destination are the same. We will first obtain the system performance in terms of the packet error rate (at the physical layer) for an i.i.d bitstream, and then PSNR of an encoded video bitstream for various average channel SNR values and power allocation parameters $\rho \in [0.6, 0.9]$. Our proposed suboptimal and closed-form combining methods at the destination will be used.

For the channel coding, we use a convolutional code of rate $r_{fec} = 1/2$ for the forward error correction (FEC) for

all links. The convolutional code has constraint length 7 and generator polynomial (133,171) in octal. We use soft decoding for the Viterbi decoder; hence, soft-demodulated bits will be computed [44] from the test statistics given in (8) with the suboptimal weights in (C.4) and (30) for the BL and the EL, respectively. The BL signal is considered as a noisy QPSK-modulated signal, thus the resulting soft-demodulated bits are equivalent to the soft bits calculation in [45].

We encode the Akiyo, Foreman, and Soccer video sequences, which exhibit low, medium, and fast motion, respectively. These video sequences have CIF resolution (352×288) and frame rate of 30fps. These test video sequences are encoded with an H.264/AVC encoder with group of pictures (GoP) size = 16 pictures, and use the full hierarchical B structure [29], [32], [46], where the I-frame occurs once per GoP. For the hierarchical B structure, the frames in decreasing order of importance are given as follows: $I_{0/16}, B_8, B_4, B_{12}, B_2, B_6, B_{10}, B_{14}, B_1, B_3, \dots, B_{15}$. For convenience at the physical layer, we choose a slice to have a fixed size of 376 bytes (or less). Each network abstraction layer (NAL) unit [29] contains one slice of size 383 bytes (or less), including the header. For each GoP of 16 frames, we assign the more important frames to the BL and the others to the EL (at the frame level) such that the difference in the number of NAL units between the two layers, from the first GoP up to the current encoding GoP, is smallest. For example, suppose the total number of NAL units of the BL and the EL up to the last encoded GoP are 100 and 102, respectively. The number of NAL units of frames in the current GoP are 3, 3, 3, 2, 2, 2, 1, 1, \dots , 1 (the frames are in decreasing order of importance), with 25 NAL units in total. If we assign the first 6 frames to the BL and the last 10 frames to the EL, the total numbers of NAL units of the BL and the EL will be 115 and 112, respectively. If we assign the first 5 frames to the BL and the remaining to the EL, then the total numbers of NAL units of the BL and the EL will be 113 and 114, respectively. Our method will select the latter, because it makes the difference in the number of NAL units between the two layers smallest. Using this method, the numbers of NAL units of the BL and the EL for the first 10 GoPs of the Akiyo sequence are 177 and 179, respectively. Those corresponding numbers for the Foreman sequence are 412 and 408, and for the Soccer sequence are 649 and 646. Zero-padding will be used to make the two layers have exactly the same number of NAL units.

In this work, we use the H.264/AVC encoder, not SVC, because H.264/AVC is much more widely used. The scalable layers come from the temporal layers available in H.264/AVC. Note that in the typical terminology for temporal BLs and ELs, the hierarchical-B structure with GOP length 16 would be said to have intra frames 0 and 16 in the BL, predicted frame 8 in the first EL, predicted frames 4 and 12 in the next EL, predicted frames 2, 6, 10 and 14 in the next EL after that, and finally all the odd-numbered frames as the non-reference B-frames in the least important EL. The BL and various ELs in this case might have very different numbers of bits. In our situation, we want to define only two classes of bits, and we want the two classes to have very close to the same number

of bits because we use hierarchical 16-QAM modulation and need to form symbols by joining bits from one class with the same number of bits from the other. Therefore we simply order the frames by importance, and consider the BL to consist of as many of the important frames as yields closest to half of the bits of the GoP. So our terminology on BL and EL differs from the usual for a hierarchical-B structure. We note that the proposed methods are not specific to H.264/AVC. For example, an H.264/SVC MGS coder could be used, where the division of transform coefficients would be chosen so as to make the two layers come out roughly even in number of bits.

At the physical layer, we use a frame length of $2 \times 400/r_{\text{fec}} = 1600$ bytes, which is assumed to be sufficient to transmit a pair of NAL units and the physical frame header. Note that each pair of NAL units consists of one from the BL and one from the EL, where the BL NAL unit is mapped into the MSBs and the EL NAL unit is mapped into the LSBs of the hierarchical 16-QAM symbols. In this simulation, we use a pair of block bit interleavers of size 80×80 bits each (i.e., 800 bytes)⁴ for the BL and the EL separately to partially decorrelate the channel fading correlation.

We repeatedly send the first 10 GoPs (i.e., $N_{\text{frm}} = 160$ frames) 100 times over the relay network for each pair of power allocation parameter ρ and channel SNR. The received bitstreams are then decoded and the average PSNR is computed. For the m -th decoded video sequence for $m = 1, 2, \dots, 100$, the PSNR is computed as follows:

$$PSNR_m = 10 \log_{10} \frac{255^2}{MSE_m}, \quad (32)$$

where

$$\overline{MSE}_m = \frac{1}{N_{\text{frm}}} \sum_{k=1}^{N_{\text{frm}}} MSE_{m,k}, \quad (33)$$

and $MSE_{m,k}$ denotes the mean square error between the k -th original frame and the corresponding received frame of the m -th sequence. The final average PSNR is averaged over 100 realizations $\{PSNR_m\}$. We use motion-copy error concealment for the H.264/AVC decoder [32]; however, if a whole frame is lost, it will be copied from the previous one.

The simulation is also repeated for the single-layer scheme with the classical 16-QAM modulation, where all the bits in the bitstream are given equal priority.

B. Packet Error Rate for an i.i.d. Bitstream

In Fig. 7, we plot the packet error rate (PER) for both the single-layer and double-layer scheme, where an i.i.d. bitstream was sent. The abscissa is the power allocation parameter, ρ , for the double-layer scheme. The PERs of the BL and the EL packets are plotted separately. As the single-layer scheme uses the classical 16-QAM with no distinction among the input bits, the single-layer scheme only depends on channel SNR, and not ρ . It is plotted on the right hand edge as a single point for each channel SNR for comparison.

⁴The block bit interleaver writes the input bitstream in the rows and reads the bitstream out from the columns.

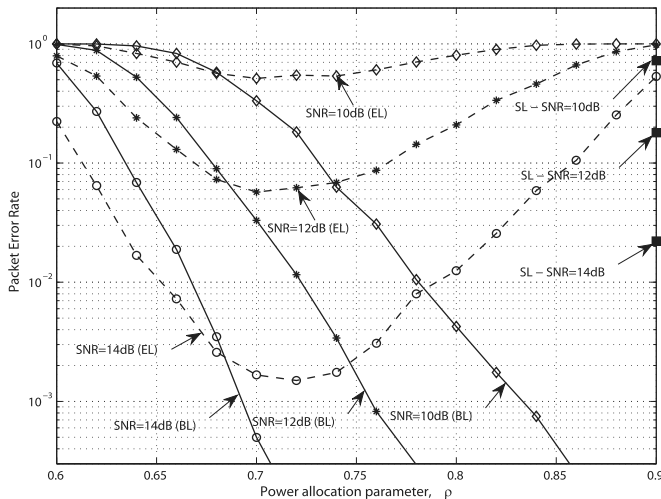


Fig. 7. Coded packet error rate for an i.i.d. bitstream using the double-layer and single-layer schemes.

For each channel SNR, we observe that the PER of the BL monotonically decreases as the power allocation parameter ρ increases (see Fig. 7). This is because for a higher ρ , the BL is allocated more power, while the EL is allocated less power, i.e., higher BL signal power and less interference caused by the EL. For the EL performance, the plot shows that the PERs of the EL are not a monotonic function of ρ . This can be explained as follows: For $\rho \geq 0.72$, the BL is successfully decoded most of the times, and so the EL performance mainly depends on the power allocated to it; thus, the PER of the EL monotonically increases as $\bar{\rho} = 1 - \rho$ decreases. In contrast, for $\rho \leq 0.70$, as ρ decreases, the relays are not able to reliably detect the BL (because of the decreased SINR of the BL), and thus they frequently keep silent. At some point, the destination does not receive enough signal power because too few signals are being relayed; hence, both PER of the BL and PER of the EL become worse. For sufficiently low values of ρ (say $\rho \leq 0.64$), the EL PER can be better than the BL PER due to the fact that the LSBs (i.e., the EL) of hierarchical 16-QAM are more reliable than the MSBs (i.e., the BL). In practice, these small values of ρ should not be used.

For the double-layer scheme, the relays have more flexibility of forwarding packets to the destination than the single-layer scheme. Specifically, a relay that might fail to decode a packet during a deep fade if the single-layer scheme is used, might still be able to decode the BL portion if the double-layer scheme is used, and thus might be able to successfully forward it to the destination to enhance the overall system performance. We note that if only the BL is successfully decoded at a given relay, it is forwarded to the destination using QPSK with full transmit power, which enhances the probability that the BL is successfully decoded at the destination. In Fig. 7, the numerical results show that for $\rho = 0.8$ (equal distance constellation), the performance of the BL is several orders of magnitude better than the EL. In this case, the system performance is limited by the EL PER. By reducing ρ , i.e., increasing the power allocated to the EL, the EL PER gets better at the cost of increasing the BL PER. We can see that

the best value of ρ , in terms of minimizing the average PER, is $\rho \approx 0.70$. In this case, the average PER of both the BL and the EL is better than that of $\rho = 0.8$, and is also better than the PER of the single-layer scheme.

C. PSNR Performances for Layered Video Sequences

1) *Decodable Bitstreams*: Firstly, we note that there are a few received bitstreams at low channel SNR and small values of ρ , which are invalid bitstreams in the sense that the H.264/AVC decoder does not produce any output picture. We refer to these invalid bitstreams as undecodable. If no packet is successfully received, the null bitstream is also considered undecodable. In Fig. 8, we plot the percentage of decodable bitstreams over all 100 received bitstreams, which corresponds to 100 channel realizations for all the three video sequences. We found that there were some undecodable bitstreams only at low channel SNR ≤ 11 dB and small power allocation parameter $\rho \leq 0.62$. For an undecodable bitstream, we compute the PSNR relative to the mean value of each frame. We plot the average PSNR values, averaging over all the received bitstreams, including the undecodable bitstreams. We also plot the average PSNR values, averaging over only the decodable bitstreams. We found that these curves are visually undistinguishable (the plot is not shown here). Hence, in the following, we present the PSNR computed only from the decodable bitstreams.

2) *PSNR Performance Versus Power Allocation Parameter*: In Fig. 9, we plot the average PSNR versus the power allocation values ρ , parameterized by channel SNR, for the three video sequences. In general, we observed that for medium to high channel SNR (say channel SNR ≥ 11 dB), either too high or too low values of the power allocation parameter ρ do not result in good PSNR performance. The reasons are similar to the i.i.d. data with PER performance. That is, too high or too low a value of ρ allows either the BL or the EL (or neither) to be successfully received, but usually not both. We observe that at the medium and high SNR region, the best power allocation parameter is around $\rho \approx 0.74$. In the low SNR region, a high value of ρ is preferred to better protect the BL. In the very high SNR region, the PSNR saturates and many values of ρ can approach the maximum PSNR value. The differences between the video sequences will be discussed later.

3) *PSNR Performance Versus Channel SNR*: In Fig. 10, we plot the average PSNR versus channel SNR, parameterized by the power allocation values ρ , for the three video sequences with low, medium, and high motion. We observe that the performance of a video sequence is worse when its motion is higher. For example, at channel SNR = 12 dB, the best PSNR of the Soccer sequence is about 27.7 dB, while those of the Foreman and Akiyo sequences are about 30.1 dB and 34.4 dB, respectively. The main reason is that motion-copy error concealment for a high motion video sequence does not perform as well as for a lower motion video sequence. Also, when the instantaneous channel SNR is low, whole frame loss can occur, and in this case we substitute the previous frame for it. Such copy-frame does not result in a good PSNR for a

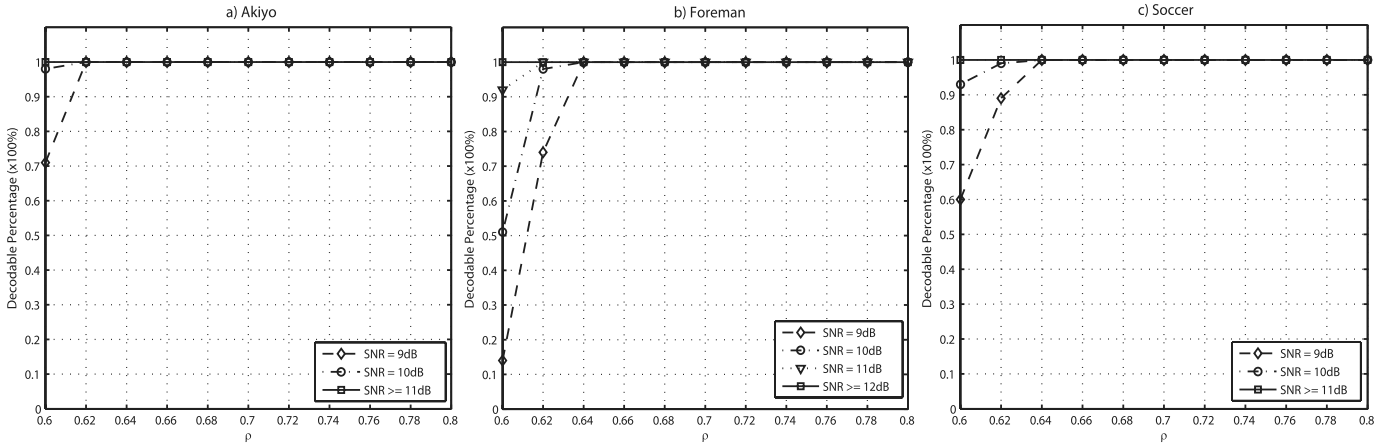


Fig. 8. Percentage of decodable bitstreams over all 100 received bitstreams for various video sequences versus power allocation parameter ρ .

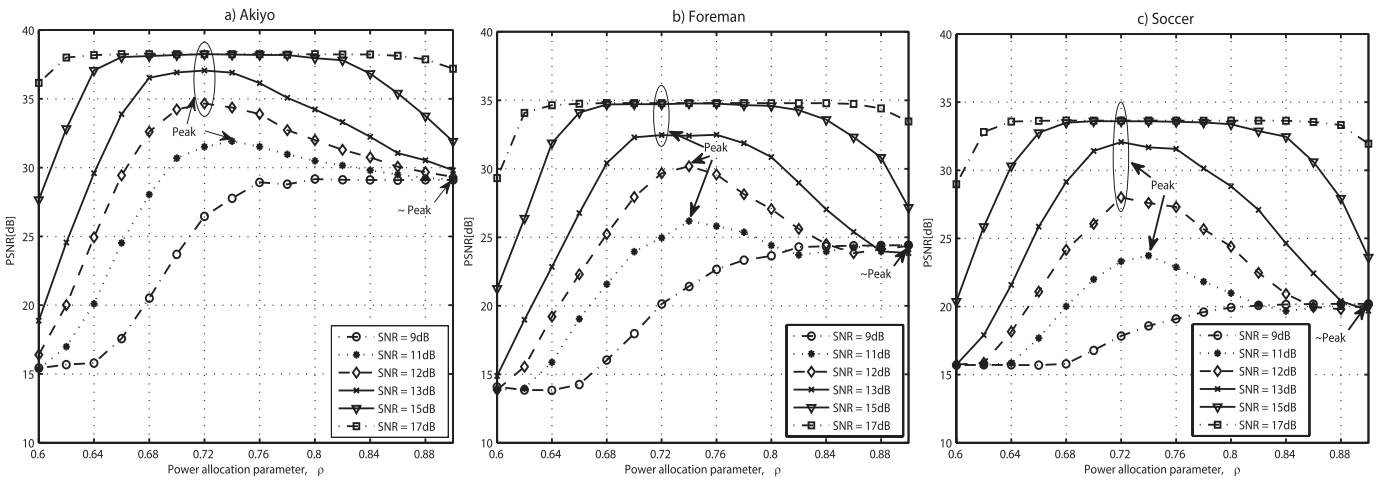


Fig. 9. Average PSNR versus power allocation parameter ρ : the best value of ρ is about 0.74 for medium and high channel SNR, and approaches to 1 for low channel SNR.

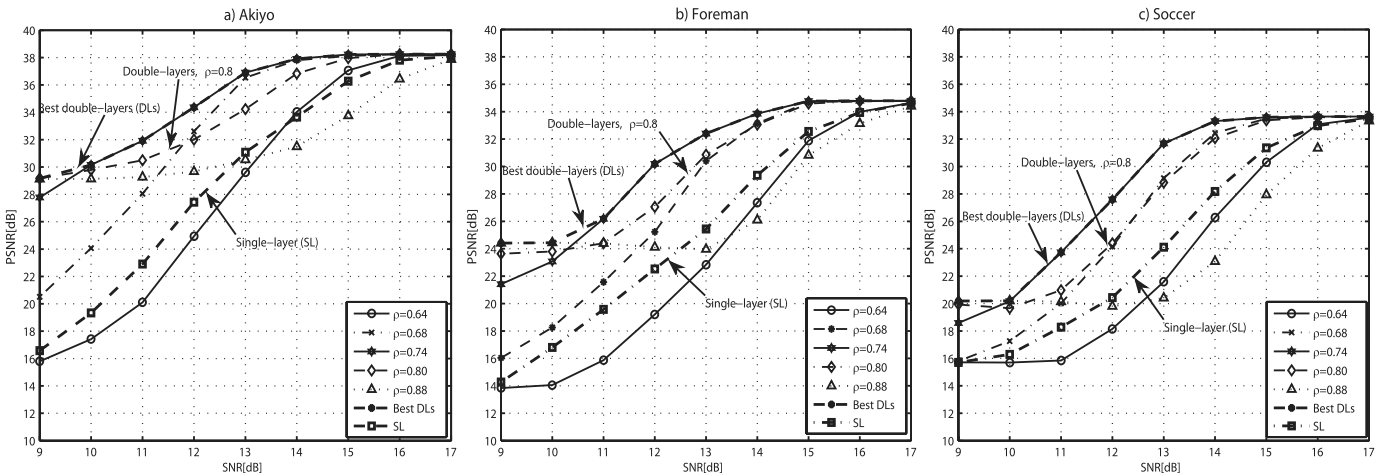


Fig. 10. Average PSNR versus channel SNR: about 2-2.5dB gain in channel SNR, or 5-7dB gain in PSNR can be observed.

high-motion sequence, whereas copy-frame is not too bad for a low or medium motion sequence.

For all sequences, we observe that the ‘best’ double-layer scheme, which is when the system has the capability to

adapt ρ to the average channel SNR, is significantly better than the single-layer scheme over all the SNR region plotted (see Fig. 10). Particularly, about 2–2.5dB gain in channel SNR or 5-7dB gain in the PSNR can be observed. We note

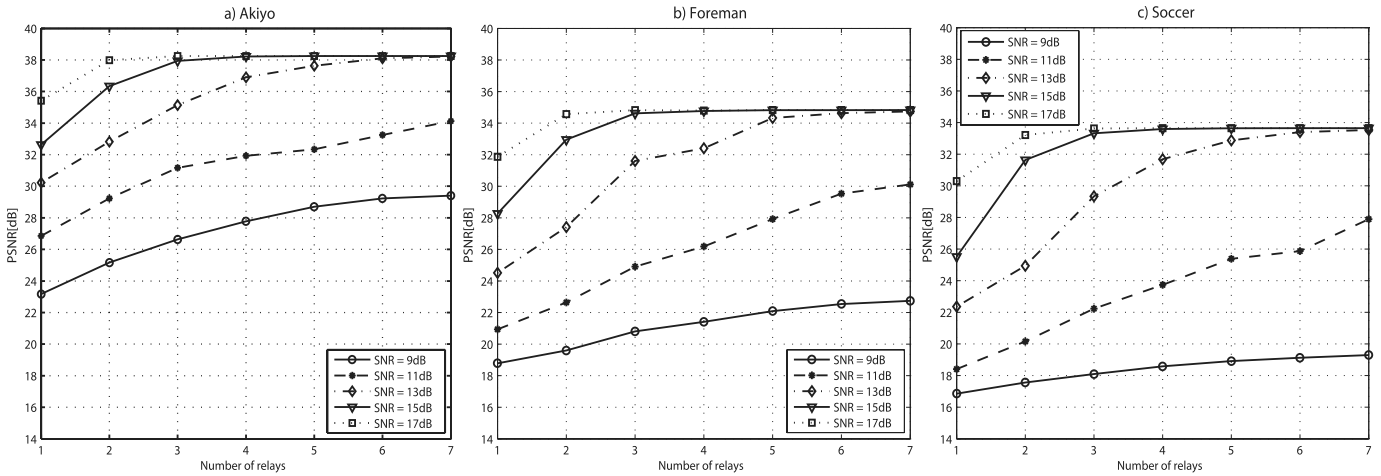


Fig. 11. Average PSNR versus number of relays: for $SNR \geq 15\text{dB}$, we need 2-3 relays to have the PSNR saturated, while for $SNR \leq 13\text{dB}$, we need about 5-6 relays or more.

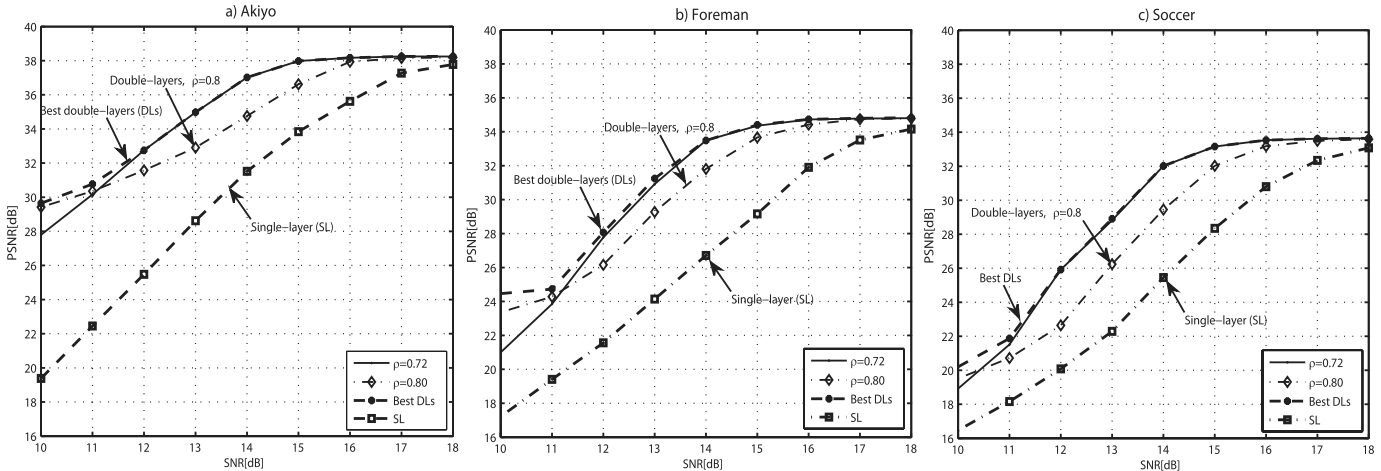


Fig. 12. Average PSNR versus channel SNR for a network with two relays and a direct link.

that in the very high SNR region (say $SNR \geq 16\text{dB}$), the PSNR for the double-layer scheme saturates, and the single-layer scheme performs as well as the double-layer scheme. On the other hand, at very low channel SNR such as $SNR = 9, 10\text{dB}$, the double-layer scheme can reliably deliver at least the BL, and the performance is almost flat at a moderate and acceptable quality, while the performance of the single-layer scheme significantly decreases and becomes useless.

Lastly, we note that if we fix $\rho \in [0.70, 0.80]$, i.e., the system does not adapt ρ to channel SNR, the PSNR performance of the double-layer scheme still significantly outperforms the single-layer scheme over all the range of the SNRs considered (except for a very high SNR value such as 17dB).

4) *PSNR Performance Versus Number of Relays:* In Fig. 11, we plot the PSNR performance versus the number of relays for a typical value of $\rho = 0.74$ for all the video sequences. The plot shows that the scheme is more sensitive to the number of relays N when N is small such as $N \leq 5$. We observe that with a higher channel SNR, fewer relays are needed to get the PSNR saturated. In particular, for $SNR \geq 15\text{dB}$, we need

2–3 relays to have the PSNR saturated, while for $SNR \leq 13\text{dB}$, we need about 5-6 relays or more.

D. PSNR Performances for Layered Video Sequences: Two Relays With the Direct Link

Unlike the previous numerical results, in this subsection, we consider a relaying network of two relays with a direct link. We again assume all the links from the source to the relays and from the relays to the destination have equal average channel SNR. The two relays are located about halfway from the source to the destination. The log-distance path loss model with the path loss exponent of 3.5 is assumed (see, e.g., [47]). Thus, the average channel SNR of the direct link is less than those of the relay links by $10 \log_{10} 2^{3.5} = 10.54\text{dB}$.

In Fig. 12, we plot the PSNR performances versus the average channel SNR. We observe a large improvement of the double-layer scheme, compared to the single-layer scheme. The double-layer scheme achieves the best performance at the power allocation parameter $\rho \approx 0.72$, which has about 2–2.5dB gain in SNR, or 5–7dB gain in PSNR, compared to the classical single-layer scheme.

V. CONCLUSION

In this paper, we consider decode-and-forward wireless relay networks using both hierarchical 16-QAM and QPSK. The source broadcasts a message consisting of two layers to all the relays and the destination. Depending on the number of successfully decoded layers, a relay can use either hierarchical 16-QAM or QPSK to transmit both layers or one layer, respectively, to the destination. By considering a hierarchical 16-QAM symbol as the superposition of two QPSK symbols, we proposed a relaying protocol and novel combining methods for the received signals at the destination. We derived the optimal linear combining solutions in terms of minimizing the uncoded BER. We also presented suboptimal combining methods for both the BL and the EL, which have closed-form solutions and perform very close to the optimal. Both our proposed optimal and suboptimal methods significantly outperform other combining methods in the literature.

We applied the proposed double-layer scheme with our suboptimal combining method to transmit layered video bitstreams through the wireless relay networks. Simulation results showed that the double-layer scheme using hierarchical 16-QAM largely outperforms the classical single-layer scheme using conventional 16-QAM. For example, either about 2-2.5dB gain in channel SNR or 5-7dB gain in the PSNR was observed.

APPENDIX A

THE FIRST COMBINING STEP: OPTIMAL FOR THE BL/EL

First, we note that, similar to the MRC solution, from the definitions in (9), we can show that

$$\begin{aligned} C_{\Theta_k}(\tilde{\mathbf{u}}_{\Theta_k}) &\leq C_{\Theta_k}(\tilde{\boldsymbol{\alpha}}_{\Theta_k}) = \|\boldsymbol{\alpha}_{\Theta_k}\| \triangleq C_{\Theta_k}^*, \text{ for all } \|\tilde{\mathbf{u}}_{\Theta_k}\| = 1, \\ C_{\Psi_k}(\tilde{\mathbf{u}}_{\Psi_k}) &\leq C_{\Psi_k}(\tilde{\boldsymbol{\alpha}}_{\Psi_k}) = \|\boldsymbol{\alpha}_{\Psi_k}\| \triangleq C_{\Psi_k}^*, \text{ for all } \|\tilde{\mathbf{u}}_{\Psi_k}\| = 1, \end{aligned} \quad (\text{A.1})$$

where the optimum normalized weight vectors $\tilde{\boldsymbol{\alpha}}_{\Theta_k}$, $\tilde{\boldsymbol{\alpha}}_{\Psi_k}$ are given in (17).

The BER expressions for the BL and the EL in (12) and (13) have the following forms, respectively,

$$f(x, y) \triangleq \frac{1}{2}Q(ax + by) + \frac{1}{2}Q(ax + cy), \quad (\text{A.2})$$

$$g(x, y) \triangleq Q(dx) + \frac{1}{2}Q(2ax + by) - \frac{1}{2}Q(2ax + cy), \quad (\text{A.3})$$

where $a, b, c, d \geq 0$, $c \geq b$, and x, y correspondingly represent $C_{\Theta_k}(\tilde{\mathbf{w}}_{\Theta_k})$ and $C_{\Psi_k}(\tilde{\mathbf{w}}_{\Psi_k})$. Similar to [18], we can show that both $f(x, y)$ and $g(x, y)$ are non-increasing in (x, y) for $x, y \geq 0$. Therefore, from (A.1), and noting that $C_{\Theta_k}(\tilde{\mathbf{w}}_{\Theta_k})$, $C_{\Psi_k}(\tilde{\mathbf{w}}_{\Psi_k}) \geq 0$ for $\tilde{\mathbf{w}}_{\Theta_k}$, $\tilde{\mathbf{w}}_{\Psi_k} \geq 0$, we have

$$BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\mathbf{w}}_{\Theta_k}, \tilde{\mathbf{w}}_{\Psi_k}) \geq BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\boldsymbol{\alpha}}_{\Theta_k}, \tilde{\boldsymbol{\alpha}}_{\Psi_k}),$$

for all $\tilde{\mathbf{w}}_{\Theta_k}$, $\tilde{\mathbf{w}}_{\Psi_k} \geq 0$ and $\|\tilde{\mathbf{w}}_{\Theta_k}\| = \|\tilde{\mathbf{w}}_{\Psi_k}\| = 1$, and similarly

$$BER_{EL}(b_{\Theta_k}, b_{\Psi_k}, \tilde{\mathbf{v}}_{\Theta_k}, \tilde{\mathbf{v}}_{\Psi_k}) \geq BER_{EL}(b_{\Theta_k}, b_{\Psi_k}, \tilde{\boldsymbol{\alpha}}_{\Theta_k}, \tilde{\boldsymbol{\alpha}}_{\Psi_k}),$$

for all $\tilde{\mathbf{v}}_{\Theta_k}$, $\tilde{\mathbf{v}}_{\Psi_k} \geq 0$ and $\|\tilde{\mathbf{v}}_{\Theta_k}\| = \|\tilde{\mathbf{v}}_{\Psi_k}\| = 1$. That is, we have (16) and (18).

APPENDIX B

THE SECOND COMBINING STEP: OPTIMAL FOR THE BL

From (12), we denote

$$\begin{aligned} BER_{BL}(a_{\Theta_k}, a_{\Psi_k}) &\triangleq BER_{BL}(a_{\Theta_k}, a_{\Psi_k}, \tilde{\mathbf{w}}_{\Theta_k}^*, \tilde{\mathbf{w}}_{\Psi_k}^*) \\ &= \frac{1}{2}Q\left(\frac{a_{\Theta_k}\sqrt{E_r}C_{\Theta_k}^* + a_{\Psi_k}\sqrt{E_r}C_{\Psi_k}^*(\sqrt{\rho} - \sqrt{\bar{\rho}})}{\sqrt{(a_{\Theta_k}^2 + a_{\Psi_k}^2)N_0}}\right) \\ &\quad + \frac{1}{2}Q\left(\frac{a_{\Theta_k}\sqrt{E_r}C_{\Theta_k}^* + a_{\Psi_k}\sqrt{E_r}C_{\Psi_k}^*(\sqrt{\rho} + \sqrt{\bar{\rho}})}{\sqrt{(a_{\Theta_k}^2 + a_{\Psi_k}^2)N_0}}\right), \end{aligned} \quad (\text{B.1})$$

where $\tilde{\mathbf{w}}_{\Theta_k}^*$, $\tilde{\mathbf{w}}_{\Psi_k}^*$ are the optimal weight vectors in the first step. We note that if $(a_{\Theta_k}, a_{\Psi_k})$ minimizes $BER_{BL}(a_{\Theta_k}, a_{\Psi_k})$, so does $(K_c a_{\Theta_k}, K_c a_{\Psi_k})$, for some constant $K_c > 0$. Therefore, without loss of generality, we can assume $a_{\Theta_k} = 1$, and we only need to optimize over a_{Ψ_k} . For further notational simplicity, we define

$$\begin{aligned} A &\triangleq \frac{\sqrt{E_r}C_{\Theta_k}^*}{\sqrt{N_0}}, \quad B \triangleq \frac{\sqrt{E_r}C_{\Psi_k}^*}{\sqrt{N_0}}(\sqrt{\rho} - \sqrt{\bar{\rho}}), \\ C &\triangleq \frac{\sqrt{E_r}C_{\Psi_k}^*}{\sqrt{N_0}}(\sqrt{\rho} + \sqrt{\bar{\rho}}), \end{aligned} \quad (\text{B.2})$$

where $A \geq 0$ and $C \geq B \geq 0$. Then, the BER for the BL, as a function of a_{Θ_k} is given by

$$\begin{aligned} BER_{BL}(a_{\Psi_k}) &\triangleq BER_{BL}(x1, a_{\Psi_k}) \\ &= \frac{1}{2}Q\left(\frac{A + Ba_{\Psi_k}}{\sqrt{1 + a_{\Psi_k}^2}}\right) + \frac{1}{2}Q\left(\frac{A + Ca_{\Psi_k}}{\sqrt{1 + a_{\Psi_k}^2}}\right). \end{aligned} \quad (\text{B.3})$$

The optimization can be written as follows:

$$\min_{a_{\Psi_k} > 0} BER_{BL}(a_{\Psi_k}). \quad (\text{B.4})$$

The function $\frac{A + Ba_{\Psi_k}}{\sqrt{1 + a_{\Psi_k}^2}}$ is generally neither convex nor concave in $a_{\Psi_k} > 0$. Thus, in general, (B.4) is not a convex optimization problem. However, for $a_{\Psi_k} > 0$, define

$$\phi = \tan^{-1}(a_{\Psi_k}), \quad \phi \in (0, \pi/2). \quad (\text{B.5})$$

Substituting $a_{\Psi_k} = \tan(\phi)$ in (B.4), we have

$$\begin{aligned} BER_{BL}^{(a)}(\phi) &\triangleq BER_{BL}(\tan(\phi)) \\ &= \frac{1}{2}Q(A \cos \phi + B \sin \phi) + \frac{1}{2}Q(A \cos \phi + C \sin \phi). \end{aligned} \quad (\text{B.6})$$

The function $(A \cos \phi + B \sin \phi)$ is concave for all $A, B \geq 0$ in $\phi \in (0, \pi/2)$. The function $Q(x)$ is convex and monotonically decreases for $x \geq 0$. Thus, $Q(A \cos \phi + B \sin \phi)$ is convex for all $A, B \geq 0$ [40], and so is the $BER_{BL}^{(a)}(\phi)$. Therefore, we can efficiently use convex optimization programming [40] to solve for ϕ^* , which minimizes $BER_{BL}^{(a)}(\phi)$, as shown in (21).

APPENDIX C
THE FIRST COMBINING STEP:
SUBOPTIMAL FOR THE BL/EL

Similar to Appendix B, we assume $a_{\Theta} = 1$, and find a suboptimal weight a_{Ψ_k} . From the BER expression of the BL in (B.3), and noting that $C \geq B \geq 0$, an upper bound for the BER of the BL is twice the dominant term

$$BER_{BL}(a_{\Psi_k}) \leq Q\left(\frac{A + Ba_{\Psi_k}}{\sqrt{1 + a_{\Psi_k}^2}}\right). \quad (C.1)$$

Note that in high (instantaneous) SNR, the true BER value approaches the dominant term, i.e., a half of the upper bound. Thus, minimizing the upper bound is also minimizing the true BER value in high SNR.

Now, minimizing the upper-bound BER in (C.1) is equivalent to maximizing the argument of the Q function, i.e.,

$$a_{\Psi_k}^{\dagger} \triangleq \arg \max_{a_{\Psi_k} \geq 0} \frac{A + Ba_{\Psi_k}}{\sqrt{1 + a_{\Psi_k}^2}}. \quad (C.2)$$

Using the Cauchy-Schwarz inequality, we can show that

$$a_{\Psi_k}^{\dagger} = B/A = \frac{C_{\Psi_k}^*}{C_{\Theta_k}^*}(\sqrt{\rho} - \sqrt{\bar{\rho}}) = \frac{\|\alpha_{\Psi_k}\|}{\|\alpha_{\Theta_k}\|}(\sqrt{\rho} - \sqrt{\bar{\rho}}), \quad (C.3)$$

where we substituted A and B from (B.2), and $C_{\Theta_k}^*$ and $C_{\Psi_k}^*$ from (A.1). By scaling both $a_{\Theta_k}^{\dagger}$ and $a_{\Psi_k}^{\dagger}$ by $\|\alpha_{\Theta_k}\|$, the system performance does not change, and a pair of suboptimal weights are finally given by (22).

Note that combining with the weights in the first step, as shown in (17), and the weights in (22), closed-form suboptimal weights are given by

$$\mathbf{w}_{\Theta_k}^{\dagger} = \alpha_{\Theta_k}, \quad \mathbf{w}_{\Psi_k}^{\dagger} = (\sqrt{\rho} - \sqrt{\bar{\rho}})\alpha_{\Psi_k}, \quad (C.4)$$

each of which only depends on its corresponding channel gain and the power allocation parameter ρ .

REFERENCES

- [1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 10, pp. 1451–1458, Oct. 1998.
- [2] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.
- [3] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [4] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
- [5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
- [6] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [7] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [8] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3114–3123, Aug. 2007.
- [9] S. S. Ikki and M. H. Ahmed, "On the performance of cooperative diversity networks with the Nth best-relay selection scheme," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3062–3069, Nov. 2010.
- [10] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.
- [11] Y. Tian and A. Yener, "The Gaussian interference relay channel: Improved achievable rates and sum rate upperbounds using a potent relay," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2865–2879, May 2011.
- [12] A. Bletsas, H. Shin, M. Z. Win, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.
- [13] W. H. R. Equitz and T. M. Cover, "Successive re-nement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.
- [14] Z. Chen, G. Barrenetxea, and M. Vetterli, "Distributed successive refinement of multiview images using broadcast advantage," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4581–4592, Nov. 2013.
- [15] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.
- [16] L. H. C. Lee, "New rate-compatible punctured convolutional codes for Viterbi decoding," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3073–3079, Dec. 1994.
- [17] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1234–1248, Jul. 1993.
- [18] S. H. Chang, M. Rim, P. C. Cosman, and L. B. Milstein, "Optimized unequal error protection using multiplexed hierarchical modulation," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5816–5840, Sep. 2012.
- [19] *Digital Video Broadcasting (DVB): Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television*, ETSI, European Standard 300 744, V1.5.1, Nov. 2004.
- [20] M. R. Chari *et al.*, "FLO physical layer: An overview," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 145–160, Mar. 2007.
- [21] D. Song and C. W. Chen, "Scalable H.264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1218–1226, Sep. 2007.
- [22] M. F. Sabir, R. W. Heath, and A. C. Bovik, "An unequal error protection scheme for multiple input multiple output systems," in *Proc. Conf. Rec. 36th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 2002, pp. 575–579.
- [23] Y. Qian, L. Ping, and M. Ronghong, "Unequal error protection scheme for multiple-input and multiple-output wireless systems," in *Proc. IEEE 64th Veh. Technol. Conf.*, Sep. 2006, pp. 1–5.
- [24] G. H. Yang, D. Shen, and V. O. K. Li, "Unequal error protection for MIMO systems with a hybrid structure," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 685–689.
- [25] O. Alay, T. Korakis, W. Yao, and S. Panwar, "Layered wireless video multicast using directional relays," in *Proc. 15th IEEE ICIP*, Oct. 2008, pp. 2020–2023.
- [26] O. Alay, T. Korakis, W. Yao, and S. Panwar, "Layered wireless video multicast using relays," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 8, pp. 1095–1109, Aug. 2010.
- [27] D. Gunduz and E. Erkip, "Source and channel coding for cooperative relaying," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3454–3475, Oct. 2007.
- [28] H. Kim, R. Annavajjala, P. C. Cosman, and L. B. Milstein, "Source-channel rate optimization for progressive image transmission over block fading relay channels," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1631–1642, Jun. 2010.
- [29] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Standard H.264&ISO/IEC 14496-10 AVC, v3, 2005.
- [30] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [31] T. Stutz and A. Uhl, "A survey of H.264 AVC/SVC encryption," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 325–339, Mar. 2012.
- [32] JVT, Dubai, United Arab Emirates. (2013). *H.264/AVC Reference Software* [Online]. Available: <http://iphome.hhi.de/suehring/tml/>
- [33] D. G. Brennan, "Linear diversity combining techniques," *Proc. IRE*, vol. 47, no. 6, pp. 1075–1102, Jun. 1959.
- [34] J. G. Proakis, *Digital Communications*, 3rd ed. New York, NY, USA: McGraw-Hill, 1995.

- [35] X. Dong and N. C. Beaulieu, "Optimal maximal ratio combining with correlated diversity branches," *IEEE Commun. Lett.*, vol. 6, no. 1, pp. 22–24, Jan. 2002.
- [36] C. Siriteanu and S. D. Blostein, "Maximal-ratio eigen-combining for smarter antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 917–925, Mar. 2007.
- [37] B. Holter and G. E. Oien, "The optimal weights of maximum ratio combiner using an eigenfilter approach," in *Proc. IEEE Nordic Signal Process. Symp.*, Hurtigruten, Norway, Jan. 2002, pp. 1–4.
- [38] H. X. Nguyen, H. H. Nguyen, and T. Le-Ngoc, "Signal transmission with unequal error protection in wireless relay networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2166–2178, Jun. 2010.
- [39] H. Kim, P. C. Cosman, and L. B. Milstein, "Superposition coding based cooperative communication with relay selection," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 892–896.
- [40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [41] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, Jun. 2003.
- [42] P. K. Vitthaladevuni and M. S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 297–307, Jan. 2003.
- [43] P. Popovski and E. D. Carvalho, "Improving the rates in wireless relay systems through superposition coding," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4831–4836, Dec. 2008.
- [44] T. Brink, J. Speidel, and Y. Ran-Hong, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Global Telecommun. Conf.*, vol. 1, Nov. 1998, pp. 579–584.
- [45] F. Tosato and P. Bisaglia, "Simplified soft-output demapper for binary interleaved COFDM with application to HIPERLAN/2," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, Jan. 2002, pp. 664–668.
- [46] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [47] S. Y. Seidel and T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multi floored buildings," *IEEE Trans. Antennas Propag.*, vol. 40, no. 2, pp. 207–217, Feb. 1992.



Tu V. Nguyen (S'08–M'13) received the B.S. degree in electrical engineering from the Post and Telecommunication Institute of Technology (PTIT), Hanoi, Vietnam, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, San Diego, La Jolla, CA, USA, in 2004, 2010, and 2013, respectively.

He is currently a Staff Scientist with Broadcom Corporation, where he is involved in WLAN system design and development and 802.11 standardization. In 2012, he was with Nextivity Inc., San Diego, CA,

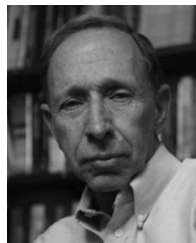
where he was involved in research and developments of relaying systems to improve the 3G and 4G HSPA + network coverage and capacity. In 2010, he was with the VA San Diego Medical Center, where he was involved in analyzing the electrocardiography signals to detect patient's pains. From 2005 to 2007, he was a Teaching Staff Member with the Telecommunications Department, PTIT. His research interests include multimedia transmissions over wireless networks, cross-layer design optimization, image and video processing, information theory, digital signal processing, multiple antenna and multiuser systems, dynamic resource allocation, Wi-Fi, and LTE technology.



Pamela C. Cosman (S'88–M'93–SM'00–F'08) received the B.S. (Hons.) degree in electrical engineering from the California Institute of Technology, Pasadena, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1987, 1989, and 1993, respectively.

She was a National Science Foundation Post-Doctoral Fellow with Stanford University and a Visiting Professor with the University of Minnesota from 1993 to 1995. In 1995, she joined the faculty of the Department of Electrical and Computer Engineering, University of California, San Diego (UCSD), where she is currently a Professor. She was the Director with the Center for Wireless Communications from 2006 to 2008, and is currently the Associate Dean for Students of the Jacobs School of Engineering, UCSD. Her research interests are in the areas of image and video compression and processing, and wireless communications.

Prof. Cosman was an Associate Editor of the IEEE COMMUNICATIONS LETTERS from 1998 to 2001, an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2001 to 2005. She was the Editor-in-Chief from 2006 to 2009 as well as a Senior Editor from 2003 to 2005 and from 2010 to 2013 of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. She is a member of Tau Beta Pi and Sigma Xi.



Laurence B. Milstein (S'66–M'68–SM'77–F'85) received the B.E.E. degree from the City College of New York, New York, NY, USA, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1964, 1966, and 1968, respectively.

From 1968 to 1974, he was with the Space and Communications Group, Hughes Aircraft Company, and from 1974 to 1976, he was a member with the Department of Electrical and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Since 1976, he has been with the Department of Electrical and Computer Engineering, University of California, San Diego (UCSD), La Jolla, where he is the Ericsson Professor of Wireless Communications Access Techniques and former Department Chairman, working in the area of digital communication theory with special emphasis on spread-spectrum communication systems. He has also been a Consultant to both Government and the industry in the areas of radar and communications.

Dr. Milstein was the Vice President for Technical Affairs in 1990 and 1991 of the IEEE Communications Society and is a former Chair of the IEEE Fellows Selection Committee. He was an Associate Editor for Communication Theory of the IEEE TRANSACTIONS ON COMMUNICATIONS, an Associate Editor for Book Reviews of the IEEE TRANSACTIONS ON INFORMATION THEORY, an Associate Technical Editor of the IEEE COMMUNICATIONS MAGAZINE, and the Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was a recipient of the 1998 Military Communications Conference Long-Term Technical Achievement Award, an Academic Senate 1999 UCSD Distinguished Teaching Award, an IEEE Third Millennium Medal in 2000, the 2000 IEEE Communication Society Armstrong Technical Achievement Award, and various prize paper awards, including the 2002 MILCOM Fred Ellersick Award.