A Near Optimal QoE-Driven Power Allocation Scheme for Scalable Video Transmissions over MIMO Systems

Xiang Chen, Student Member, IEEE, Jenq-Neng Hwang, Fellow, IEEE, Chung-Nan Lee, Shih-I Chen

Abstract—The rapid increasing demands of wireless multimedia applications have boosted the developments of video delivery technologies with cross-layer designs, driven by optimizing quality of experiences (QoEs) of end users. In this paper, a near optimal power allocation scheme, targeting at maximizing QoE, is proposed for transmitting scalable video coding (SVC) based videos over multi-input multi-output (MIMO) systems. Both transmission errors in the physical (PHY) layer and video source coding characteristics in the application (APP) layer are jointly considered in the proposed scheme. A near optimal solution is achieved by decomposing the original optimization problem into several convex optimization sub-problems. Detailed algorithms with corresponding theoretical reasoning are provided. Since forward error corrections (FEC) techniques are widely implemented in modern wireless communication systems, the proposed scheme is further extended to the systems with Reed-Solomon (RS) code and a more practical approach with different modulation and coding schemes (MCSs). The near optimality of our proposed scheme, in terms of measured utilities, is shown by comparing with the exhaustive searched optimal solutions. Simulations with real H.264 SVC video traces demonstrate the effectiveness of our proposed scheme by comparing with other existing schemes in terms of well-accepted video quality assessment methods, such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index.

Index Terms—Power allocation, cross-layer, quality of experience (QoE), scalable video coding (SVC), multi-input multi-output (MIMO), forward error correction (FEC), convex optimization

I. INTRODUCTION

The rapid increasing demands of wireless multimedia applications have boosted the development of modern video delivery technologies over wireless channels [1]. Providing higher quality mobile video services becomes the everlasting endeavors of multimedia service providers [2]. However, the error-prone and band-limited nature of wireless communication environments, which causes high packet loss/error rate, delay and jitter, may lead to tremendous quality degradation of real-time streaming video services [3]. Several techniques can be applied to compensate the effects caused by varying wireless channel qualities. In the application (APP) layer, scalable video coding (SVC) attracts more and more attentions in recent years. Videos can be encoded with different spatial, temporal and quality scalabilities where valid bit streams can still be formed even when parts of the encoded bit streams (higher enhancement layers) are removed [4]. Therefore, SVC provides a capability of adapting to various needs or preferences of end users as well as to varying terminal capabilities or network conditions [5]. In the physical (PHY) layer, the use of multiple antennas at both transmitters and receivers, known as multi-input multi-output (MIMO) wireless, can significantly improve transmission reliability or spectral efficiency by either spatial diversity (SD) or spatial multiplexing (SM) approaches [6], [7].

Network quality of service (QoS) parameters, such as packet delay and packet loss rates, would ultimately contribute to user experiences on streaming video delivery [8]. Since video service providers need to observe and react quickly on quality problems perceived by customers on the delivered videos, the concept of quality of experience (QoE) emerges, combining user perceptions and experiences with non-technical and technical parameters [9]. In order to optimize the overall user experiences during video transmissions in mobile broadband networks, it is necessary to devise generic cross-layer design driven by QoE directly [10].

Plenty of researches have been conducted in transmitting videos over MIMO systems, where both PHY layer structures and APP layer video coding characteristics are considered to improve the quality of received videos. In [11], an optimal power allocation scheme for minimizing visual distortion over MIMO systems is proposed, where H.264/AVC video and theoretical link-capacity are considered. In [12], the scalable video streams are transmitted over a MIMO system by using space-time block codes with equal power allocation. An adaptive channel selection (ACS) scheme is proposed in [13]. In this scheme, MIMO-SM approach is considered and video bit streams with higher priorities (i.e. base layer and lower enhancement layers) are transmitted through spatial channels with relatively higher signal-to-noise ratios (SNRs). A power allocation scheme, targeting on maximum throughput delivery of SVC-based videos over MIMO-SM systems, is proposed in [2]. In this scheme, the theoretical capacity-achieving water-filling (WF) algorithm is improved when discrete modulation levels are considered in a more practical scenario.

The research is based on work supported by the Ministry of Economic Affairs (MOEA) of Taiwan, under Grant number MOEA 102-EC-17-A-03-S1-214.

Xiang Chen and Jenq-Neng Hwang are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: xchen28@uw.edu; hwang@uw.edu).

Chung-Nan Lee is with the Department of Computer Science and Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan (e-mail: cnlee@cse.nsysu.edu.tw).

Shih-I Chen is with the Institute for Information Industry (III), Taiwan (e-mail: shihI@iii.org.tw).
In [14], authors propose another power allocation scheme, which includes different bit-error-rate (BER) targets for unequal error protections (UEPs) of transmitting different SVC video layers. Nevertheless, due to the empirical nature when setting different BER targets on SVC video layers, this scheme is far from optimal. A resource allocation scheme is proposed in [15], where PHY layer parameters such as power and modulation and coding scheme (MCS) are adjusted according to video distortions measured in mean square error (MSE). The power allocation solutions are solved by using Lagrangian method and the convexity of the optimization problem is addressed by experimental studies.

Received video quality may be degraded by damaged or lost packets [8]. When SVC-based videos are transmitted, decoding errors in the high-priority layers will cause propagation errors in the low-priority layers. Therefore, directly minimizing BER of the system, without considering the video coding structures, does not necessarily ensure that more video frames with higher quality enhancement layers can be received. To optimize QoEs of subscribers, a cross-layer design of SVC-based video delivery scheme, which considers both the PHY layer characteristics and the APP layer video coding structures, is highly desired. In this paper, based on our preliminary work in [16], we propose a near optimal power allocation scheme for SVC-based video transmissions over MIMO-SM systems. In our proposed scheme, both video coding structures in the APP layer and effects of power allocation to BER in the PHY layer are jointly considered. Due to its high complexity, we decompose the original optimization problem into several sub-problems which can then be solved by classical convex optimization methods. Detailed algorithms for searching the optimal solutions and its corresponding theoretical reasoning are provided. The near optimality of our proposed scheme is shown by comparing with optimal solutions through exhaustive searches. Several simulations with real H.264 SVC video traces demonstrate the effectiveness of our proposed scheme by comparing with other existing schemes in terms of well-accepted video quality assessment methods such as peak signal-to-noise ratio (PSNR) and structure similarity (SSIM).

Different from our previous work in [16], we extend the algorithm by taking into account the forward error correction (FEC) codes in the system. In practice, FEC codes are often applied to improve the reliability of the communication systems [18]. Using different kind of codes with different coding rate leads to different error performances. Among plenty of error correction codes, Reed-Solomon (RS) codes are widely used in practice and can achieve Singleton bound with its maximum distance separable (MDS) characteristic [18], [19]. Also, RS code is used in IEEE 802.16 standard [20]. In this paper, a general solution with traditional RS code and a more practical solution with different MCSs are provided. Moreover, a new low-complexity algorithm with little performance reduction is also provided in this paper.

This paper is organized as follows. In the next section, descriptions of the system, including introductions about SVC-based video and MIMO-SM are provided. In Section III, problem formulations, including the original problem and its solving strategy: decomposing into several sub-problems, are given. Conditions of obtaining the optimal solutions of sub-problems are described in Section IV. In Section V, Our proposed algorithms of solving each sub-problem is provided. Simulation results and conclusion remarks are given in Section VI and VII respectively.

Notations: Upper (lower) boldface letters are used for matrices (column vectors). diag(\(h\)) is a diagonal matrix with the elements of \(h\) sitting on the diagonal. \(||.||_1\) denotes the l-1 norm of a vector. \((.)^T\) means the transposes. \((.)^H\) means the Hermitian. \(I_N\) denotes the \(N \times N\) identity matrix. \(E[.]\) denotes the expectation. \(GF(.)\) stands for Galois field.

\section{System Overview}

A MIMO-SM system used for transmitting SVC-based videos is shown in Fig. 1. A video sequence is encoded into \(L\)
layers of bit streams, with one base layer and $L-1$ enhancement layers. After encoded with FECs, the bit streams are fed into a MIMO system with $N_t \times N_r$ transmitter antennas and $N_r$ receiver antennas. An adaptive channel selection (ACS) module [2], [13] is applied so that bit streams with higher priorities are transmitted through the spatial channels with higher SNRs. The power allocation module allocates appropriate power to modulated symbols with cross-layer video information and channel state information (CSI) fed back from the receiver side. The data symbols are then transmitted through the wireless channel after precoding. At the receiver side, a channel estimation module sends CSIs back to the transmitter side for both power allocation and precoding processes. In this paper, we assume the accurate estimated CSIs with full channel knowledge are fed back without error and delay. Moreover, the channel selection sequences, modulation and FEC coding schemes are known at the receiver side through the control channel. After decoding, detection, demodulation, channel selection and FEC decoding, the received bit streams are fed into the SVC decoder for video reconstruction. For any video frame, the bit stream of video layer $l$ is dropped if any single bit error is detected after FEC correction or the required lower layers (e.g., from base layer to the $l$-th layer) are not successfully decoded.

A. SVC-Based Video

A video bit stream is called scalable when parts of the stream can be removed in a way that the resulting sub-stream forms another valid bit stream for some target decoder [4], [5]. An SVC encoded video consists of one base layer and several enhancement layers in a hierarchical dependency structure, where the base layer and the lower enhancement layers are required in order to decode the higher enhancement layers. The decoded video quality is progressively improved when more enhancement layers are successfully decoded [21]. SVC can support all of the temporal (frame rates), spatial (picture resolutions) and quality (image fidelity) scalabilities. In this paper, we design our proposed scheme based on videos with quality scalability only. However, similar idea can be applied to videos with temporal and spatial scalabilities.

As we will try to maximize the overall QoEs at user ends, which are normally measured in utility values [22], we adopt a simple perceptual quality model for SVC-based videos with quality scalabilities [23]:

$$u_l = \begin{cases} e^{(1-q_l/q_{max})}, & l = 1 \\ e^{(1-q_l/q_{max})} - e^{(1-q_{l-1}/q_{max})}, & l \geq 2 \end{cases},$$

where $u_l$ denotes the utilities for layer $l$ with quality scalability; $c$ is video dependent model parameters; $q_l$ is the quantization stepsize of the $l$th quality layer; $q_{max}$ is the minimum quantization stepsize corresponding to the video layer with the highest quality. Note that the video quality model developed in [23] is specifically based on videos with CIF (352×288) resolution. Therefore, we adopt videos with CIF resolution in this work. However, similar concept can be applied to videos in other resolutions with appropriate video quality models.

B. MIMO System Model

The equation of an $N_t \times N_r$ MIMO system can be described as:

$$y = Hx + n,$$

where $y$ is $N_t \times 1$ received (complex) signal vector. $x$ is $N_r \times 1$ complex transmitted symbol vector with $E[xx^H] = \text{diag}(p)$, subject to normalized power $\|x\|=1$ and each element in $p$ is not less than 0. $n$ is $N_t \times 1$ independent and identically distributed (i.i.d.) complex additive white Gaussian noise (AWGN) vector with covariance matrix $N_0 I$. $H$ is $N_t \times N_r$ channel matrix in which all elements are i.i.d. circularly symmetric complex Gaussian (ZMCSG) random variables with zero mean and variance $1$, i.e., $CN(0,1)$. Therefore, the average SNR of the system $\rho = \|x\|/\|n\| = I/N_0$.

MIMO channel matrix $H$ can be decomposed by the singular value decomposition (SVD):

$$H = UVH^H,$$

where $U$ and $V$ are unitary matrices and $\Lambda$ is a diagonal matrix specified as:

$$\Lambda = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_k}, 0, \ldots, 0).$$

where $R \leq \min(N_t, N_r)$ is the rank of channel matrix $H$, and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_R \geq 0$ are eigenvalues of $HH^H$. With accurate full channel knowledge at both transmitter and receiver side, a precoder $V$ and a decoder $U^H$ can be applied so that the MIMO system can be expressed as:

$$\tilde{y} = U^HHVx + U^Hn = \Lambda x + \tilde{n}.$$  

Since $U$ is a unitary matrix, the elements in $\tilde{n} = U^Hn$ are still i.i.d. complex Gaussian distributed, i.e., $CN(0, I)$. Obviously, by using precoder and decoder, a MIMO system can be decomposed into $R$ independent single-input single-output (SISO) channels [2]. The SNR on the $l$th channel can be expressed as $\text{SNR} = \rho \lambda_l \rho$. 

III. PROBLEM FORMULATION

The error prone nature of wireless channels will cause transmission errors in terms of bit errors or symbol errors, which further lead to video frame decoding errors and degradation of received video qualities (i.e. system utilities). Therefore, in order to maximize QoE at the user end, we define our optimization problem as:

$$Q : \max_{\rho} \sum_{l=1}^{L} u_l \tilde{f}_l(p)$$

subject to $p_k \geq 0, \forall k; \sum_{k=1}^{L} p_k = 1,$

where $u_l$ is the utility of layer $l$. $L$ is the total number of video layers. $p=[p_1, p_2, \ldots, p_l, \ldots, p_L]^T$. Note that $\tilde{f}_l(p)$ is the frame correction rate described as:

$$\tilde{f}_l(p) = \prod_{k=1}^{l} (1 - Pe_k(p_k))^s_k,$$

where $Pe_k(p_k)$, which is the bit or symbol error rate of the $k$th layer, is a decreasing function of power $p_k$; and $s_k$ is the total amount of bits or symbols left in the buffer for transmitting the $k$th layer of a single group of pictures (GOP). In this paper, we
only consider the objective function as the expected system utility for simplicity. The actual utility model can vary in different scenarios such as applying APP level error concealment techniques. However, the overall utility of a system should be an increasing function of frame correction rate. Thus, optimizing frame correction rate is a more general solution, which leads to the sub-problems discussed later.

A. Bit or Symbol Error Rates for Different MCSs

Since $P_{e_0}(p_i)$ is different under different MCSs and power applied on the $k^{th}$ spatial channel, we briefly describe its expressions used in this paper with three cases: 1) pure quadrature amplitude modulation (QAM) with constellation size $M$ (M-QAM), 2) M-QAM with RS codes and 3) approximations of MCSs in real applications.

1) Pure M-QAM

The receiver bit error rate (BER) of M-QAM can be approximated as [19]:

$$
P_{e_k}(p_i) \approx \frac{2(1-M^{-0.5})}{\log_2(\sqrt{M})} Q\left(\sqrt{\frac{3\log_2(\sqrt{M})}{M-1}} \frac{2E_b}{N_0}\right),
$$

where $Q(.)$ is the complementary error function and $E_b/N_0$ is the average bit energy to average noise power ratio. Since SNR can be calculated from $E_b/N_0$, i.e. SNR=$\log_2(M) \times E_b/N_0$ [2], in our proposed MIMO-SM system, BER of the $k^{th}$ channel can be derived as:

$$
P_{e_k}(p_i) \approx \frac{2(1-M^{-0.5})}{\log_2(\sqrt{M})} \left(1-\Phi\left(\sqrt{\frac{3}{M-1} \rho_k^2 p_i}\right)\right),
$$

where $\Phi(.)$ is the cumulative distribution function (CDF) of the standard normal distribution.

2) M-QAM with RS Codes

In this paper, our proposed scheme can be also designed based on RS codes. For an $(N, K, N-K+1)$ RS code over GF($2^m$), which has correction capability $t=(N-K)/2$, the decoded symbol error probability is given by [18]:

$$
P_{e_k}(p_i) = \frac{1}{N} \sum_{i=1}^{N} \left(\begin{array}{c} N \\ i \end{array}\right) P_{M_i}(p_i) \left(1-P_{M_i}(p_i)\right)^{N-i},
$$

where $P_{M_i}(p_i)$ is the RS symbol error rate before decoding. In this paper, we assume the correction capability $t$ is an integer (i.e., $N-K$ is multiple of 2). When M-QAM modulation is applied with $\log_2(M) \leq n$, $P_{M_i}(p_i)$ can be derived as:

$$
P_{M_i}(p_i) = \left(1 - 2 \left(1-M^{-0.5}\right) Q\left(\sqrt{\frac{3\rho_k^2 p_i}{M-1}}\right)^{2n/\log_2(M)}\right),
$$

3) Approximations of MCSs in Real Applications

In real application standards, such as 3GPP, HIPERLAN/2, IEEE 802.11a and IEEE 802.16, the BER expressions of each MCS can be approximated by [24]:

$$
P_{e_k}(p_i) = a_k e^{-b_k \rho_k^2 p_i},
$$

where $a_k$ and $b_k$ are coefficients depending on the MCS of layer $k$. Possible coefficients and their corresponding MCSs are listed in Table I.

B. Problem Solving Strategy: Solving Sequence of Sub-Problems

Due to different possible relations between utilities and frame correction rate, a general solution is to optimize the frame correction rate directly when different numbers of SVC layers are considered. Therefore, we decompose the original problem $Q$ into $L$ sub-problems $Q_l$, where $l=1,2,\ldots,L$. More specifically, when up to the $l^{th}$ SVC layer is allowed to be transmitted, the corresponding frame correction rate of layer $l$ can be optimized by solving the following sub-problem:

$$
Q_l: \min_{p_i} \sum_{s=1}^{L} s_k \log(1-P_{e_k}(p_i)) \quad \text{subject to } p_k \geq 0, \forall k; \sum_{k=1}^{l} p_k = 1.
$$

Note that $p_{l+1}=p_{l+2}=\ldots=p_{L}=0$ are implied since the layers higher than $l$ are not allowed to be transmitted. If $p^*$ denotes the solution of the $l^{th}$ sub-problem in Eq. (13), the optimal solution of $Q$ is found by:

$$
p^* = \arg\max_{p_{1},p_{2},\ldots,p_{L}} \sum_{k=1}^{L} u_k \tilde{f}_k\left(p^*_k\right).
$$

Since the original problem is solved by choosing the best solution among the $L$ candidate solutions from the corresponding $L$ sub-problems, the solution found in Eq. (14) is not global optimal. In section VI, we will demonstrate the near-optimality of our proposed scheme by comparing with the global optimal points obtained by exhaustive searches.

C. Convexity of Sub-problems

In this subsection, we will show that the objective function of $Q_l$ in Eq. (13) is convex. Thus, with linear equality and inequality constraints, the sub-problem $Q_l$ can be solved by classical convex optimization.

1) Pure M-QAM

For pure M-QAM case without any FECs, in Eq. (9), $\Phi(.)$ is concave and non-decreasing function for any non-negative input argument. And the argument inside $\Phi(.)$ is concave with respect to non-negative variable $p_i$. $P_{e_k}(p_i)$ in Eq. (9) is convex due to the composition rule: $f(x)=h(g(x))$ is concave if $h$ is concave and non-decreasing, and $g$ is concave; and the property: $f(x)$ is convex if $f(x)$ is concave [25]. The above composition rule can be used again to show $\log(1-P_{e_k}(p_i))$ is concave since $\log(.)$ is concave and non-decreasing for any non-negative input. Therefore, the objective function of $Q_l$ is convex since non-negative weighted sums preserve convexity [25].

---

**Table I. Possible Coefficients for Different MCSs**

<table>
<thead>
<tr>
<th>Type (m)</th>
<th>Modulation</th>
<th>Code Rate</th>
<th>$a_m$</th>
<th>$b_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BPSK</td>
<td>1/2</td>
<td>1.1369</td>
<td>7.5556</td>
</tr>
<tr>
<td>2</td>
<td>QPSK</td>
<td>1/2</td>
<td>0.3531</td>
<td>3.2543</td>
</tr>
<tr>
<td>3</td>
<td>QPSK</td>
<td>3/4</td>
<td>0.2197</td>
<td>1.5244</td>
</tr>
<tr>
<td>4</td>
<td>16 QAM</td>
<td>9/16</td>
<td>0.2081</td>
<td>0.6250</td>
</tr>
<tr>
<td>5</td>
<td>16 QAM</td>
<td>3/4</td>
<td>0.1936</td>
<td>0.3484</td>
</tr>
<tr>
<td>6</td>
<td>64 QAM</td>
<td>3/4</td>
<td>0.1887</td>
<td>0.0871</td>
</tr>
</tbody>
</table>
2) M-QAM with RS Codes

When the RS symbol error rate $P_{e_1}(p_{k_1})$ is low enough, which is the general case in practice, the decoded symbol error rate in Eq. (10) can be approximated as (see Appendix A):

$$P_e(p_k) = \left\{ \begin{array}{ll}
\frac{(N-1)!}{(N-k-1)!} P_{e_1}(p_{k_1}), & 0 \leq k \leq 2 \\
\sum_{j=0}^{N-k-1} \frac{(N-j-1)!}{(N-k-j-1)!} P_{e_1}(p_{k_1})^j, & j \geq 3
\end{array} \right. $$

(15)

The de-modulated symbol error rate, denoted by $P_s(p_k)$:

$$P_s(p_k) = 2 \left( 1 - M_k^{-0.5} \right) Q \left( \frac{3 \rho \lambda_k}{M_k - 1} \right),$$

(16)

is convex by similar proof in pure M-QAM case discussed above. $P_{M_k}(p_k)$ in Eq. (11) can be expressed as:

$$P_{M_k}(p_k) = 1 - (1 - P_s(p_k))^{2 \frac{2}{\log_2(M_k)}} \approx 2 n P_s(p_k) \left( \frac{p_k}{M_k - 1} \right)$$

by first order Taylor series approximation since $P_s(p_k)$ is much smaller than 1 in practice. Therefore, by composition rule: $f(g(x))$ is convex if $f$ is convex and non-decreasing, and $g$ is convex, ($P_{M_k}(p_k)$) is convex for $p_k \geq 1$. Therefore, $P_{M_k}(p_k)$ is convex since it is non-negative sums of convex functions. Therefore the objective function of $Q(t)$ is convex for M-QAM with RS codes.

3) Approximations of MCSs in Real Standards

$P_{M_k}(p_k)$ is convex with positive $a_k$ and $b_k$ in Eq. (12). Therefore, with similar proof in the pure M-QAM case, the objective function of $Q(t)$ is convex with the BER approximation equation for different MCSs.

IV. CONDITIONS OF OPTIMAL SOLUTIONS FOR SUB-PROBLEMS

The Lagrangian of the $i^{th}$ sub-problem $Q_{i}$, in Eq. (13), can be derived as:

$$L_{i}(p_i, \xi, \nu) = -\sum_{k=1}^{l} \xi_k \log(1 - P_{e_k}(p_k)) - \sum_{k=1}^{l} \xi_k p_k + \nu \left( \sum_{k=1}^{l} p_k - 1 \right),$$

(18)

where $\xi$ and $\nu$ are Lagrange multipliers associated with the inequality constraints and equality constraint respectively. For each $k=1, \ldots, l$, the Karush-Kuhn-Tucker (KKT) conditions can be expressed as:

1. Primal feasible: $p^*_k \geq 0, \sum_{k=1}^{l} p^*_k = 1$.
2. Dual feasible: $\xi^*_k \geq 0$.
3. Complementary slackness: $\xi^*_k p^*_k = 0$.
4. Gradient of Lagrangian slackness:

$$\frac{\partial L(p, \xi, \nu)}{p_k} \bigg|_{p_k^*} = \frac{s_k P_{e_k}'}{1 - P_{e_k}(p_k^*)} - \xi_k^* + \nu^* = 0,$$

(19)

where $P_{e_k}'(p_k)$ is the first derivative of $P_{e_k}(p_k)$. For convex optimization problems, if any point satisfies the KKT conditions, it is primal and dual optimal with zero duality gap [25]. The above KKT conditions imply:

$$v^* - \frac{s_k P_{e_k}'}{(1 - P_{e_k}(p_k^*))} = 0, \quad$$

(20)

and

$$\frac{s_k P_{e_k}'}{1 - P_{e_k}(p_k^*)} + v^* = 0.$$ 

(21)

Here, we only consider the case when $p_k^* > 0$ since the case $p_k^* = 0$ can be solved by the $(k-1)^{th}$ sub-problem. Therefore, Eq. (21) implies:

$$\frac{1}{v^*} = h_k(p_k^*) - \frac{s_k P_{e_k}'}{(1 - P_{e_k}(p_k^*))}.$$

(22)

For pure M-QAM case,

$$P_{e_k}'(p_k^*) \approx -\left( 1 - M_k^{-0.5} \right) e^{\frac{3 \rho \lambda_k}{2(M_k - 1)}} \frac{3 \rho \lambda_k}{\log_2(M_k)} \sqrt{2 \pi (M_k - 1) p_k}.$$ 

(23)

Hence, the optimality conditions of the $i^{th}$ sub-problem $Q_i$ is:

1. $h_k(p_k^*) = l/v^*$ for $k=1,2, \ldots, l$; and
2. $\sum_{k=1}^{l} p_k^* = 1$ ; and
3. $p_k^* > 0$ for $k=1,2, \ldots, l$ and $p_k^* = 0$ for $k=l+1, l+2, \ldots, L$.

V. ALGORITHM

In this paper, we adopt a simple but effective bisection search algorithm to obtain the optimal points of the sub-problems. A low complexity algorithm is also proposed, which can achieve the near optimal solutions of the sub-problems.

A. Monotonicity of Function $h(.)$

The first derivative of $h_k(p_k)$ can be derived as:

$$h_k'(p_k) = \frac{1}{s_k} + \left( \frac{1 - P_{e_k}(p_k^*)}{s_k^2 (P_{e_k}'(p_k^*))^2} \right) P_{e_k}''(p_k^*).$$

(28)

Since $P_{e_k}(p_k)$ is convex as proved above, its second derivative $P_{e_k}''(p_k^*) \geq 0$. Also, $P_{e_k}(p_k)$ is a probability of error, which is between 0 and 1, the term $(1-P_{e_k}(p_k))$ is non-negative. Thus,
$h_k(p_k)$ is positive for any power $p_k$ (positive for the $k^{th}$ sub-problem) and positive $s$. Therefore, function $h_k(p_k)$ is monotonically increasing.

**B. Proposed Bisection Search Algorithm**

Since $h_k(p_k)$ is monotonically increasing and $0<p_k\leq 1$, the optimal point of the $k^{th}$ sub-problem is between 0 and $\min(h_k(1))$ for $k=1,2,...,l$. In practice, we use function $g_k(p_k)=\log(1+h_k(p_k))$ in our algorithm to avoid numerical limitations without sacrificing the optimality of the $k^{th}$ sub-problem $Q_k$. Therefore, the first optimal condition of $Q_k$ becomes:

$$g_k\left(p_k\right) = \mu^* - \log\left(1 + 1/\nu^*\right). \quad (29)$$

The proposed bisection search algorithm, used to find the optimal point $p^*_k$ of the $k^{th}$ sub-problem $Q_k$, is shown in the following table:

**Algorithm 1 (A1): Bisection Search Algorithm**

1. $upper = \min\left(g_k(1)\right)$, for $k = 1,2,...,l$
2. $lower = 0$
3. $p_k = 0$, for $k = 1,2,...,l$
4. while ($\sum p_k < 1$) > $\varepsilon$
5. $\mu = (upper + lower)/2$
6. $p_k = g_k^{-1}(\mu)$, for $k = 1,2,...,l$
7. if ($\sum p_k < 1$)
8. $lower = \mu$
9. else
10. $upper = \mu$
11. end if
12. end while
13. $p^*_k = p_k/\sum p_k$, for $k = 1,2,...,l$
14. $p^*_k = [p^*_1, p^*_2, ..., p^*_l]^T$

Here, $\varepsilon$ is a small positive number for the stopping criterion. The first step is used to determine the upper bound of searching region, which is found by the minimum value of each $g_k(p_k)$ when the power reaches its constraint, i.e., $p_k=1$. Step 13 is used to satisfy the power constraint without any numerical error.

**C. Proposed Low-Complexity Algorithm**

In practice, we observe that for the $k^{th}$ sub-problem $Q_k$, most of the elements $p_k^*$ of the optimal solution $p^*_k$ obtained from Algorithm 1, are located at the linear region of $g_k(p_k)=\log(1+h_k(p_k))$ for $k=1,2,...,l$. This phenomenon is due to the fact that $h_k(p_k)$, defined in Eq. (22), is dominated by the exponential terms in Eq. (23), Eq. (26) and Eq. (27) for pure M-QAM, M-QAM with RS codes and MCS cases respectively. By taking advantage of this observation, $g_k(p_k)$ can be approximated as: $g_k(p_k)=A_k p_k+B_k$, where $A_k$ and $B_k$ are two constants to be determined. If $A_k$ and $B_k$ are known, the optimal constant $\mu^*$ can be directly calculated:

$$\mu^* = (1+\sum_{k=1}^l B_k A_k^{-1})/(\sum_{k=1}^l A_k^{-1}) \quad (30)$$

And each element of the optimal solution $p_k^*$ can be determined as: $p_k^*=(\mu^*-B_k)/A_k$, for $k = 1,2,...,l$. We propose Algorithm 2 to further reduce the complexity of Algorithm 1 without sacrificing the system performance too much.

**Algorithm 2 (A2): Low-Complexity Algorithm**

1. $p_0=0$, for $k = 1,2,...,l$
2. $\mu = \min\left(g_k(1)\right)$, for $k = 1,2,...,l$
3. $p_k = g_k^{-1}(\mu)$, for $k = 1,2,...,l$
5. $A_k = \mu g_k^{-1}(\mu) - \Delta$, for $k = 1,2,...,l$
6. $B_k = \mu A_k g_k^{-1}(\mu)$, for $k = 1,2,...,l$
7. $p_k^*=(\mu^*-B_k)/A_k$, for $k = 1,2,...,l$
9. $p^*_k = [p^*_1, p^*_2, ..., p^*_l]^T$

Here, step 4 is used to roughly estimate the optimal solution. Step 5 and 6 are used to estimate the constants $A_k$ and $B_k$, where $\Delta$ is a small positive number. Note that there is no iteration process involved in A2.

**VI. SIMULATION RESULTS**

In this section, the near-optimality and the effectiveness of our proposed algorithm are evaluated through plenty of simulations. Video clips “City”, “Foreman” and “Waterfall” with CIF resolutions are encoded by the reference SVC codec JSVM (Joint Scalable Video Model) version 9.19 [26]. Maximum frame rates are both set as 30fps. Both GOP sizes and intra period are set as 8 so that the frame pattern is IBBBBBB in one GOP. There are 161 frames encoded in total so that 20 GOPs with one additional I frame are included. Three additional enhancement layers are encoded with constrained medium-grain scalability (MGS), where motion estimation and compensation are constrained at current layer [23]. The basis quantization parameters of the four layers (i.e., one base layer and three enhancement layers) are set as QP=[48, 42, 36, 30]

<table>
<thead>
<tr>
<th>Video</th>
<th>$c$ [23]</th>
<th>$u_1$</th>
<th>$u_2$</th>
<th>$u_3$</th>
<th>$u_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>0.13</td>
<td>0.4025</td>
<td>0.2745</td>
<td>0.2010</td>
<td>0.1219</td>
</tr>
<tr>
<td>Foreman</td>
<td>0.12</td>
<td>0.4317</td>
<td>0.2660</td>
<td>0.1892</td>
<td>0.1131</td>
</tr>
<tr>
<td>Waterfall</td>
<td>0.15</td>
<td>0.3499</td>
<td>0.2877</td>
<td>0.2231</td>
<td>0.1393</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video</th>
<th>Layer 1</th>
<th>Layer 2</th>
<th>Layer 3</th>
<th>Layer 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>214.9 Kbps</td>
<td>102.3 Kbps</td>
<td>326.4 Kbps</td>
<td>613.4 Kbps</td>
</tr>
<tr>
<td>Foreman</td>
<td>204.3 Kbps</td>
<td>69.2 Kbps</td>
<td>190.6 Kbps</td>
<td>379 Kbps</td>
</tr>
<tr>
<td>Waterfall</td>
<td>549.3 Kbps</td>
<td>77 Kbps</td>
<td>265.7 Kbps</td>
<td>566.3 Kbps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video</th>
<th>Bandwidth (M-QAM)</th>
<th>Bandwidth (M-QAM + RS code)</th>
<th>Bandwidth (MCS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>110KHz</td>
<td>70KHz</td>
<td>180KHz</td>
</tr>
<tr>
<td>Foreman</td>
<td>QPSK</td>
<td>QPSK (255,127,129)</td>
<td>QPSK (MCS)</td>
</tr>
<tr>
<td>Waterfall</td>
<td>QPSK (255,175,83)</td>
<td>QPSK (255,127,129)</td>
<td>BPSK (MCS)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video</th>
<th>Bandwidth</th>
<th>Modulation</th>
<th>Coding Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>Bandwidth</td>
<td>Modulation</td>
<td>Coding Scheme</td>
</tr>
<tr>
<td>Foreman</td>
<td>Bandwidth</td>
<td>Modulation</td>
<td>Coding Scheme</td>
</tr>
<tr>
<td>Waterfall</td>
<td>Bandwidth</td>
<td>Modulation</td>
<td>Coding Scheme</td>
</tr>
</tbody>
</table>
with corresponding uniform quantization stepsizes calculated by \( q=2(QP-4)/6 \) [23]. Based on Eq. (1), the utilities of the four quality layers of “City”, “Foreman” and “Crew” are listed in Table II. The encoded network abstraction layer unit (NALU) is packetized by link layer with packet size as 48 bytes [26] and then transmitted through the PHY layer. A 4x4 MIMO-SM system is used for transmitting SVC videos with 4 layers [2], [12], [13], [15]. The channel bandwidth, modulation and channel coding rates are selected such that the source coding bitrate of each video layer are less than the corresponding throughput of each spatial channel. Table III lists the bit rate of each video layer. The bandwidths, modulation and channel coding schemes are shown in Table IV. The CSIs are fed back every channel coherence time, which is assumed to be 1ms [2]. At the receiver side, control messages such as video coding parameters are assumed to be correctly received. Also, perfect error detection scheme is assumed so that bit or symbol errors are correctly detected at the receiver side. The undecodable NALUs, including erroneous bits caused by channel quality degradation or unsatisfied video layer dependencies, are discarded before passing through the SVC decoder. The missing video frames are concealed by simply copying the previous successfully received frames. In our simulations, we compare our proposed scheme with traditional WF algorithm, simple equal power allocation scheme and modified WF (M-WF) proposed in [14] with BER targets set as \( 10^{-3} \) and \( 10^{-5} \) for base layer and enhancement layers respectively.

Figure 2 illustrates the system utilities calculated by the objective function in Eq. (6) when transmitting the first GOP of video clip “City” with pure M-QAM modulations. The optimal curve is obtained by exhaustive searches. As shown in Fig. 2, our proposed algorithms are very close to the optimal solutions. The performance reduction of our proposed low-complexity algorithm (A2) is negligible comparing with that of our proposed bisection search algorithm (A1). Even though the WF algorithm is optimal in terms of the PHY layer capacities, it is no longer optimal in the APP layer utilities. The M-WF scheme is better than the WF scheme at low SNR region since unequal error protection (UEP) is applied on the base layer and enhancement layers for better effect when channel qualities are not good enough. But it becomes worth at high SNR region due to the over protection of high priority layers.

Similar results are plotted in Fig. 3 and Fig. 4. In Fig. 3, transmitting the first GOP of video clip “Foreman” with M-QAM modulations and RS codes over GF(2^n) is considered. In Fig. 4, transmitting the first GOP of video clip “Waterfall” with different MCSs in Table I is considered. It is clear that the near optimality of our proposed scheme still holds when different MCSs are applied on different video layers.

Since the objective of our proposed scheme is to maximize the system utility, there will be no surprises that our proposed scheme can achieve higher utilities than other schemes. Next, we will compare these schemes by using traditional video quality assessment methods such as PSNR and SSIM index.

![Fig. 2. System utility of each power allocation scheme for transmitting the first GOP of video “City” with pure M-QAM modulations.](image1)

![Fig. 3. System utility of each power allocation scheme for transmitting the first GOP of video “Foreman” with M-QAM modulations and RS codes.](image2)

![Fig. 4. System utility of each power allocation scheme for transmitting the first GOP of video “Waterfall” with MCSs in Table I.](image3)

The SVC layer indices of received NALUs of “City” video clip are plotted in Fig. 5. The system average SNR is set as 24dB. Clearly, more video frames with higher layers can be received by using our proposed scheme, which can improve the qualities of the reconstructed videos. The cumulative distribution function (CDF) of the power allocation results obtained from different schemes are illustrated in Fig. 6. One can observe that with empirical settings of BER targets, M-WF scheme allocates more power on the high priority layers (layer 1 and layer 2) while our proposed scheme allocates more power on layer 3. When the system average SNR is 24dB, channel gains of the high priority layers are normally high enough.
Therefore, allocating more power on the high priority layers may cause waste of power. Figure 7 shows the per-frame PSNRs of the reconstructed “City” video clip. It is obvious that our proposed two algorithms outperform the other three, even though our optimization objective function is not PSNR. This is because of the fact that by applying our proposed scheme with reasonable utility functions, more video frames with higher quality layers can be received. The corresponding SSIM indices are plotted in Fig. 8, which still demonstrates the effectiveness of our proposed scheme. Theoretically, no matter what kind of video quality assessment methods are used, as long as they are monotonically increasing functions of video frame correction rates, the proposed algorithms can achieve higher performance than the others since more video frames with higher layers can be received. Decoded sample video frames are shown in Fig. 9, which clearly demonstrates the visual quality advantage of our proposed scheme.

Figure 10 shows the average PSNR and SSIM with respect to system average SNR. The system performance degradation of our proposed low-complexity algorithm (A2) is negligible comparing with the proposed bisection search algorithm (A1). Note that M-WF algorithm has special UEP on each video layer with empirically chosen BER targets, which would provide stronger protection on high priority video layers when channel qualities are bad. However, when the channel qualities are generally good, it would over-protect the high priority layers, which may degrade the QoE of end users since not enough higher enhancement layers can be received.
Fig. 9. Decoded sample video frames of reconstructed video “City” at system average SNR: 24 dB. Top left: proposed A1; Top right: WF; Bottom left: equal; Bottom right: M-WF.

Fig. 10. Average SSIM and PSNR of reconstructed video “City”.

Similar simulations can be repeated on the system with M-QAM and RS codes. The SVC layer indices of received NALUs of “Foreman” video clip are plotted in Fig. 11 where the system average SNR is set as 18 dB. By using our proposed scheme, more video frames with higher quality layers can be received. The CDF of power allocation results obtained by different schemes are plotted in Fig. 12. Figures 13 and 14 show the per-frame PSNRs and SSIM of reconstructed “Foreman” video clip respectively. The better effectiveness of our proposed scheme is again observed comparing with that of other methods. Note that in this case, the differences of either PSNR or SSIM comparisons are smaller. This is due to the fact that the background of “Foreman” is smoother than that of “City”, which causes less difference between each quality layers with different quantization stepsizes. A visual comparison of the reconstructed video frames is shown in Fig. 15. It is still clear that the frame is sharper by applying our proposed scheme. The average PSNR and SSIM with respect to system average SNR are plotted in Fig. 16. Note that in this scenario, the system performance degradation of our proposed low-complexity algorithm (A2) is more obvious comparing with the proposed bisection algorithm (A1). However, it is still better than other schemes.

Fig. 11. SVC quality layer indices of received NALUs of video “Foreman” at system average SNR: 18 dB.

Fig. 12. CDF of power allocation results when transmitting video “Foreman” at system average SNR: 18 dB.

Fig. 13. Per-frame PSNR of reconstructed video “Foreman” at system average SNR: 18 dB.
Similar simulations can be repeated for transmitting video clip “Waterfall” with different MCSs in Table I. The SVC layer indices are plotted in Fig. 17 where the system average SNR is set as 16dB. Still, more video frames with higher quality layers can be correctly received by using our proposed scheme. The CDF of power allocation results obtained from different schemes are shown in Fig. 18. Per-frame PSNR and SSIM of reconstructed video frames are illustrated in Fig. 19 and 20 respectively. The visual comparison of the reconstructed video frames is shown in Fig. 21. The average PSNR and SSIM with respect to system average SNR are plotted in Fig. 22.

In this paper, we have proposed a near-optimal QoE-driven power allocation scheme for scalable video delivery over MIMO-SM systems. All three cases including pure M-QAM modulations, M-QAM with RS codes, and MCSs in application standards, are considered. The proposed optimization problem is solved by calculating the optimal solutions of sequence of decomposed sub-problems. We further show that the sub-problems are convex and give the optimal conditions of finding their optimal solutions. A simple bisection search algorithm and a low complexity algorithm are proposed to obtain the optimal solutions of the sub-problems. Plenty of simulation results demonstrate the near-optimality and effectiveness of our proposed scheme with real SVC video traces. Since the end users can receive more error-free video frames with higher layers, our proposed scheme has better performance in terms of PSNR and SSIM.

VII. CONCLUSION

Fig. 14. Per-frame SSIM of reconstructed video “Foreman” at system average SNR: 18dB.

Fig. 15. Decoded sample video frames of reconstructed video “Foreman” at system average SNR: 18dB. Top left: proposed A1; Top right: WF; Bottom left: equal; Bottom right: M-WF.

Fig. 16. Average SSIM and PSNR of reconstructed video “Foreman”.

Fig. 17. SVC quality layer indices of received NALUs of video “Waterfall” at system average SNR: 16dB.

Fig. 18. SVC quality layer indices of received NALUs of video “Waterfall” at system average SNR: 16dB.
Consider Eq. (10), if \( t=0 \), \( P_{e_k}(p_k) = P_{M_k}(p_k) \). If \( t=1 \),

\[
P_{e_k}(p_k) = P_{M_k}(p_k) - \frac{1}{N} N \left( P_{M_k}(p_k) \right) \left( 1 - P_{M_k}(p_k) \right)^{N-1}
\]

with first order Taylor series approximation if \( P_{M_k}(p_k) \ll 1 \).

If \( t=2 \),

\[
P_{e_k}(p_k) \approx (N-1) P_{M_k}(p_k) - \frac{2}{N} \binom{N}{2} P_{M_k}(p_k)^2 \left( 1 - P_{M_k}(p_k) \right)^{N-2}
\]

Similar steps can be followed when \( t=3 \),

\[
P_{e_k}(p_k) \approx \frac{(N-1)! P_{M_k}(p_k)}{(N-3)!} + \frac{(N-1)! P_{M_k}(p_k)^3}{2(N-4)!}.
\]

When \( t=4 \),

\[
P_{e_k}(p_k) \approx \sum_{j=1}^{4} \frac{(N-1)! (j-2)}{(N-j)! (j-1)!} P_{M_k}(p_k)^j + \frac{(N-1)! P_{M_k}(p_k)^5}{6(N-5)!}.
\]
Hypothesis: for any $t \geq 3$,

$$P_e(t) \approx \frac{(N-1)!}{(j-2)(t)!} P_{M_1}(p_j) + \frac{(N-1)!}{(t-1)(N-t-1)!} \cdot (35)$$

When $t = t + 1$,

$$P_e(t) \approx P_e(t) - \frac{t + 1}{N} \left( P_{M_1}(p_j) \right)^{t+1} \left( 1 - P_{M_1}(p_j) \right)^{N-t}$$

$$= P_e(t) - \frac{t + 1}{N} \left( P_{M_1}(p_j) \right)^{t+1} \left( 1 - P_{M_1}(p_j) \right)^{N-t}$$

$$= \sum_{j=1}^{N} \left( \frac{(N-1)!}{(j-2)(t)!} \right) \left( 1 - P_{M_1}(p_j) \right)^{N-t} + \frac{(N-1)!}{(t-1)(N-t-1)!} \cdot (36)$$

Therefore, the hypothesis is true for any $t \geq 3$.

APPENDIX B: DERIVATION OF EQ. (24)

Consider Eq. (10), the first derivative of $P_e(p_j)$ can be derived as:

$$P_e'(p_j) = \frac{P_{M_1}(p_j)}{N} \sum_{i=1}^{N} \int \left( P_{M_1}(p_j) \right)^{i-1} \left( 1 - P_{M_1}(p_j) \right)^{N-i}$$

- $\frac{P_{M_1}(p_j)}{N} \sum_{i=1}^{N} \int \left( P_{M_1}(p_j) \right)^{i-1} \left( 1 - P_{M_1}(p_j) \right)^{N-i}$

$$= \frac{1}{N} \sum_{i=1}^{N} \int \left( P_{M_1}(p_j) \right)^{i-1} \left( 1 - P_{M_1}(p_j) \right)^{N-i}$$

$$= \sum_{j=1}^{N} \left( \frac{(N-1)!}{(j-2)(t)!} \right) \left( 1 - P_{M_1}(p_j) \right)^{N-t} + \frac{(N-1)!}{(t-1)(N-t-1)!} \cdot (37)$$

REFERENCES


