Optimal Power Allocation and Rate Adaptation for Scalable Video over Multi-User MIMO

Jounsup Park, Xiang Chen, Jenq-Neng Hwang
Department of Electrical Engineering, Box 352500,
University of Washington, Seattle, WA 98195, USA
Email: {jsup517, xchen28, hwangj}@uw.edu

Abstract - In this paper, an optimal power allocation and rate adaptation scheme is proposed to maximize the user Quality of Experience (QoE) of scalable video streaming over multi-user MIMO (MU-MIMO) networks. The proposed scheme utilizes the wireless resources to accommodate multiple users consuming different video contents more efficiently under limited propagation power and spectrum. The optimal power allocation scheme is derived based on convex optimization and an effective algorithm is proposed to select the best modulation set to control the transmission rate of each scalable video layer. Simulation results show that the proposed scheme can achieve more favorable performance than water-filling algorithm and any other power allocation scheme in terms of the sum-utility measure and QoE.

Keywords – scalable video coding, multi-user MIMO, convex optimization, modulation

I. INTRODUCTION

As increasing data traffic being over wireless networks, efficient utilization of the wireless resources, such as propagation power, spectrum, and time, is getting more important. Since the display panels in mobile devices are getting bigger and having higher resolution, video streaming over wireless network becomes a more challenging problem. For streaming video, Quality-of-Experience (QoE) has to be considered while most of wireless systems consider only the throughput. Therefore, the channel allocation schemes were introduced to serve the users with better user satisfaction or QoE within the limited bandwidth [1]. Multi-antenna, such as multi-input and multi-output (MIMO), technology is one of the most promising technologies to achieve higher spectral efficiency using spatial multiplexing techniques or get better channel quality using diversity techniques [2][3]. To enhance the communication capabilities, a single-user MIMO system considers a single multi-antenna transmitter communicating with a single multi-antenna receiver, while a multi-user MIMO (MU-MIMO) system applies an extended version of space-division multiple access (SDMA) to allow multiple transmit antenna to send separate signals and multiple receivers to receive separate signals simultaneously in the same band [3].

The H.264 scalable video coding (SVC) is an extension of the H.264/AVC video compression standard [4]. The layered coding scheme of SVC is used to achieve a wide range of spatial, temporal and Signal-to-Noise Ratio (SNR) scalability. The redundancy between different layers is exploited by different inter-layer prediction techniques that include prediction mechanisms for motion parameters as well as texture data. The layered structure is defined as the combination of a base layer and several enhancement layers. The base layer corresponds to the lowest supported video performance, which is obtained by the transform coding similar to that of H.264/AVC, whereas the enhancement layers allow for the refinement of the aforementioned base layer. The adaptation is based on a combination within the set of selected strategies for the spatial, temporal and quality scalability.

For the reliable video streaming over wireless channel, the MIMO and SVC combined technologies were suggested. Chen et al, found the near-optimal power allocation [5] and rate adaptation [6] of multiple transmit antenna to transmit single SVC video over the wireless channel. Khalek et al [7] optimized loss visibility by ordering the video packets which have different priority for real-time video transmission over MIMO system. Liu et al [8] derived the modified water-filling algorithm based on the desired SNR of each video layer. However, these researches only focused on single video transmission over MIMO channel. Wei et al [9] proposed the scheduling algorithm in MU-MIMO networks for mobile video application using equal power allocation scheme. Li et al [10] proposed the resource allocation algorithm that iteratively allocates the power and subcarrier to video layers in MU-MIMO-Orthogonal Frequency Division Multiplexing (OFDM) system.

In this paper, both MU-MIMO and SVC technologies are adopted by the physical layer and the application layer respectively, and the optimal power allocation and rate adaptation solution for maximizing users’ QoE within shared spectrum is found by using a convex optimization technique. This work is different from finding optimal resource allocation to maximize the sum-rate of systems [11][12][13], instead our goal is finding the optimal solution for user experience, which can be achieved by maximizing the sum-utility, highly dependent on the video quality and cannot be measured only by sum-rate.

The remainder of this paper is organized as follows: Section II defines the system architecture for wireless video streaming assumed for our proposed research. Section III addresses the problem formulation based on the utility model and MIMO channel; Section IV offers the problem solving strategy with simulation results given in Section V, followed by the conclusion in Section VI.
where model is improved as the number of decoded layers increase and several e

similar idea can be applied to videos with temporal and spatial (picture resolutions) or quality of video under MU

considered in this paper, the scalability is adopted in [15] as a capacity achieving scheme, but because of its higher complexity than linear precoding schemes, ZFBF [3], the linear pre-coding scheme, is getting the most attention because of its simplicity in implementation and reasonable performance. Especially, when ZFBF is applied, the receiver can receive the desired signal without interference among the users and does not need to do additional computation for reducing interference, which is a desirable advantage because most receivers are power hungry mobile devices.

In the MU-MIMO systems, received vector \( \mathbf{r} \in \mathbb{C}^{N_r \times 1} \), whose elements are received symbols corresponding to each user, can be described as,

\[
\mathbf{r} = \mathbf{D}\mathbf{s} + \mathbf{n}
\]  

(2)

where \( \mathbf{D} \) is the \( N_r \times N_r \) diagonal matrix whose diagonal elements are corresponding to \( \sqrt{\alpha_n} \), where \( \alpha_n \) depicts the path loss between the base-station and user \( n \), \( \mathbf{H} \in \mathbb{C}^{N_t \times N_r} \) is the normalized channel matrix, \( \mathbf{s} \in \mathbb{C}^{N_s \times 1} \) is the transmit vector, and \( \mathbf{n} \in \mathbb{C}^{N_r \times 1} \) is additive white Gaussian noise (AWGN) whose mean is 0 and the variance is \( \sigma^2 \). The transmit vector \( \mathbf{s} \) is summation of multiple symbols generated by multiplying the beam-forming vector \( \mathbf{w}_n \in \mathbb{C}^{N_r \times 1} \) and square root of allocated power \( p_n \) to the \( M \)-ary QAM modulated symbols \( \mathbf{x}_n \) as,

\[
\mathbf{s} = \mathbf{w}_1 \sqrt{p_1} \mathbf{x}_1 + \mathbf{w}_2 \sqrt{p_2} \mathbf{x}_2 + \cdots + \mathbf{w}_N \sqrt{p_N} \mathbf{x}_N.
\]  

(3)

Beam-forming weight vectors \( \mathbf{w}_n \) are the normalized weights derived from pseudo-inverse of the channel matrix \( \mathbf{H} \). It can be expressed as,

\[
\mathbf{w}_n = \begin{bmatrix} w_{n1} \\ w_{n2} \\ \vdots \\ w_{nN_r} \end{bmatrix} = \frac{\begin{bmatrix} \mathbf{w}_1^H H^H \end{bmatrix}_n \begin{bmatrix} \mathbf{H}^H \end{bmatrix}_n \begin{bmatrix} \mathbf{H}^H \end{bmatrix}_n}{\begin{bmatrix} \mathbf{H}^H \end{bmatrix}_n \begin{bmatrix} \mathbf{H}^H \end{bmatrix}_n^2}.
\]  

(4)

where \( \mathbf{g}^H \) denotes the Hermitian of the matrix, and \( \mathbf{g}_n \) denotes the \( n \)-th column of the matrix.

Selecting the weight vectors to make interference zero at the receiver side, every receiver can thus receive the interference free signal \( r_n \), with the channel gain \( \lambda_n \),
\[
\mathbf{r} = \begin{bmatrix}
\sqrt{\alpha_1 \lambda_1} & 0 & 0 & 0 \\
0 & \sqrt{\alpha_2 \lambda_2} & 0 & 0 \\
0 & 0 & 0 & \mathbf{M} \\
0 & 0 & 0 & \sqrt{\alpha_{\mathbf{N}} \, \lambda_{\mathbf{N}}} \\
\end{bmatrix}
\begin{bmatrix}
\sqrt{p_1 \, x_1} \\
\sqrt{p_2 \, x_2} \\
\mathbf{P} \\
\sqrt{p_{\mathbf{N}} \, x_{\mathbf{N}}} \\
\end{bmatrix}
+ \mathbf{n},
\]

where, \( \mathbf{h}_n \in \mathbb{C}^{1 \times N} \) are the row vectors of channel matrix \( \mathbf{H} \).

Then, the received signal at the \( n \)-th user can be derived as,

\[
r_n = \sqrt{\alpha_n \lambda_n} \, p_n \, x_n + n_n
\]

The received SNR for each user \( n \) is

\[
\gamma_n = \frac{\alpha_n \lambda_n p_n}{\sigma^2}.
\]

C. Modulations and Power Allocation

Let \( x_n \) be the \( M \)-ary QAM modulated symbol, where constellation size \( M \) could be 4, 16, or 64, which corresponds to modulation of QPSK, 16-QAM, and 64-QAM. To achieve the adaptive modulation capability, constellation sizes can be different by users and video layers. Therefore, the constellation size vector \( \mathbf{M}_n \) is defined as,

\[
\mathbf{M}_n = \left[ M_n^{(1)} \ M_n^{(2)} \ L \ M_n^{(L)} \right],
\]

where \( M_n^{(l)} \in \{4,16,64\} \),

for the \( l \)-th video layer of the \( n \)-th video. The \( M_n^{(l)} \) can be one of the 4, 16, or 64 value. Let also \( p_n \) denote the power allocated to the \( n \)-th video stream, and the sum of allocated power should be 1, since all wireless communication systems have their power constraint and users have to share the power.

III. PROBLEM FORMULATION

The maximum sum-utility problem under limited power and bandwidth can be defined as,

\[
Q : \text{maximize} \sum_{n=1}^{N} U_n \left( p_n, \mathbf{M}_n \right),
\]

subject to \( \sum_{n=1}^{N} p_n = 1 \), \( p_n \geq 0 \),

and \( M_n^{(l)} \in \{4,16,64\} \)

In (10), \( U_n \left( p_n, \mathbf{M}_n \right) \) denotes the utility function of the \( n \)-th video corresponding to the \( n \)-th user, where \( p_n \) is the allocated power to this video stream and \( \mathbf{M}_n \) is the constellation vector defined in the Section II.

The utility function \( U_n \left( p_n, \mathbf{M}_n \right) \) is defined as a summation of correction rate of all transmitted video layers multiplied by their utility defined in (1), which is the ratio of the video the receiver can possibly decode from the original video.

\[
U_n \left( p_n, \mathbf{M}_n \right) = \sum_{l=1}^{L} C_n^{(l)} \, U_n^{(l)} \, f_l \left( p_n, \mathbf{M}_n^{(l)} \right),
\]

In (12), \( C_n^{(l)} \in \{0,1\} \) denotes the Boolean logic to indicate whether the layer-\( l \) of the \( n \)-th video can be transmitted or not. \( U_n^{(l)} \) is the utility weight constant of layer-\( l \) of the \( n \)-th video defined in (1).

\[
f_l \left( p_n, \mathbf{M}_n^{(l)} \right) = \prod_{i=1}^{L} \left( 1 - P_e \left( p_n, \mathbf{M}_n^{(l)} \right) \right)^{s_n^{(l)}}.
\]

The \( f_l \left( p_n, \mathbf{M}_n^{(l)} \right) \) is the correction probability of the \( l \)-th layer which has the power \( p_n \), and constellation vector \( \mathbf{M}_n \) [16]. The \( P_e \left( p_n, \mathbf{M}_n^{(l)} \right) \) is probability of symbol error under SNR \( \gamma_n = \frac{\lambda_n p_n}{\sigma^2} \), and constellation size \( M_n^{(l)} \) [2].

\[
P_e \left( p_n, M_n^{(l)} \right) = 2 \left( 1 - \frac{1}{\sqrt{M_n^{(l)}}} \right) \frac{3 \alpha_n \lambda_n p_n}{\sigma^2 \left( M_n^{(l)} - 1 \right)}.
\]

The lower layer’s correction rate cumulatively affects the correction rate of the higher layers because of the layer dependency characteristic of SVC. Note that,

\[
s_n^{(l)} = \frac{P_e^{(l)}}{\log_2 M_n^{(l)}}
\]

The \( s_n^{(l)} \) defines the symbol rates to be transmitted for the \( l \)-th layer of the \( n \)-th video, where \( P_e^{(l)} \) denotes the bit-rate needed to send the \( l \)-th video layer of \( n \)-th video. Since the bandwidth (BW) is limited, the symbol rate \( s_n \), which is corresponding to the sampling rate of receiver matched filter output, is bounded by the BW of the wireless channel,

\[
s_n = \sum_{l=1}^{L} s_n^{(l)} \leq \text{BW}.
\]

Therefore, it can be shown that the utility of the \( n \)-th layer is sum of correction rate multiplied by utility of each video layer actually transmitted.

The goal of this research is to find the optimal solution that maximizes sum-utility, subject to the limited power and the bandwidth. However, the optimum solution for the defined utility function is difficult problem because its shape is not a convex nor a concave with respect to allocated power. Nonetheless, the shape of (13) is sigmoidal, and Sigmoidal programming is also an NP-hard problem [17]. Therefore, a new problem solving strategy is proposed in the next section.
IV. PROBLEM SOLVING STRATEGY

The correction rate defined in (13) is a Sigmoidal function, which makes the optimization problem very complicated [16]. Therefore, approximation techniques are applied to the correction rate (13) to be concave rather than sigmoidal, and it is verified that the approximation can still be used to find the optimal solution.

The approximated correction rate of $l$-th video layer of $n$-th user can be shown as,

$$
\hat{P}_l(p_n, M_n^{\text{max}}) @ (1 - P_e(p_n, M_n^{\text{max}})) \sum_{i=1}^{l} s_i P_e(p_n, M_n^{\text{max}}),
$$  \hspace{1cm} (17)

where $M_n^{\text{max}} = \text{max}(M_n)$.

First, the lower bound (17) of correction rate is computed by applying the largest modulation size to all video layers. This lower bound is very tight because the layer which has the largest modulation serves as a bottleneck and therefore the overall correction rate is dominated by this layer. Second, with the low-error rate approximation, the first order of Taylor series of the lower bound of the correction rate is taken (18).

Finally, the approximation of correction rate has the concave form, because $P_e(p_n, M_n^{\text{max}})$ is convex with respect to $p_n$ and will remain concave after the summation processes.

This approximation holds in low-error rate region, but not in high-error rate region. The gap between the approximation and the real objective function in high-error rate region can be compensated by applying the adaptive modulation scheme to avoid high-error rate region. The modulation vector which has lower modulation size will be chosen when it is in the high-error rate region so as to make the objective function stay in the low-error rate region.

Now the original problem $Q$ in (10) can be described as a standard convex optimization problem, which minimizes the objective function (19) subject to power constraint (20) by incorporating approximation of correction rate (18) into the utility-function, rather than using the real correction rate (13).

The new optimization problem $\hat{Q}$ can be depicted as follow.

$$
\hat{Q}: \begin{aligned}
\text{minimize } & - \sum_{n=1}^{N} \hat{Q}_n(p_n, M_n), \\
\text{subject to } & \sum_{n=1}^{N} p_n = 1, \quad p_n \geq 0.
\end{aligned}
\hspace{1cm} (19)

where $\hat{Q}_n(p_n, M_n) = \sum_{l=1}^{L} c_n^{(l)} u_n^{(l)} \frac{\hat{P}_l(p_n, M_n^{\text{max}})}{p_n}$.

By using the Lagrangian method, the objective function of the new problem $\hat{Q}$ is shown as,

$$
L(p,v,\zeta) = - \sum_{n=1}^{N} \sum_{l=1}^{L} c_n^{(l)} u_n^{(l)} \frac{1 - \sum_{j=1}^{l} s_j P_e(p_n, M_n^{\text{max}})}{p_n} + v \left( \sum_{n=1}^{N} p_n - 1 \right) - \sum_{n=1}^{N} \zeta_n p_n.
$$  \hspace{1cm} (22)

where $v$ and $\zeta_n$ are the Lagrangian multipliers of equality constraint and inequality constraint respectively. The gradient of Lagrangian can now be derived:

$$
\frac{\delta L(p,v,\zeta)}{\delta p_n} = \sum_{l=1}^{L} c_n^{(l)} u_n^{(l)} \sum_{j=1}^{l} s_j P_e(p_n, M_n^{\text{max}}) + v - \zeta_n.
$$  \hspace{1cm} (23)

A. Optimality Condition

The optimality condition is derived by using Karush-Kuhn-Tucker (KKT) conditions.

1. Primal feasibility : $\sum_{n=1}^{N} p_n = 1, \quad p_n \geq 0$.
2. Dual feasibility : $\zeta_n \geq 0$.
3. Complementary slackness : $\zeta_n p_n^* = 0$.
4. The gradient of Lagrangian vanishes when

$$
\sum_{n=1}^{L} c_n^{(l)} u_n^{(l)} \sum_{j=1}^{l} s_j P_e(p_n, M_n^{\text{max}}) + v - \zeta_n = 0
$$  \hspace{1cm} (24)

Thus, the optimal $p_n$ can be found by solving these optimality conditions. The dual feasibility implies that

$$
\zeta_n = \left( v^* + A P_e(p_n^*, M_n^{\text{max}}) \right) \geq 0
$$  \hspace{1cm} (25)

where $A = \sum_{l=1}^{L} c_n^{(l)} u_n^{(l)} \sum_{j=1}^{l} s_j$,

and the complementary slackness implies that,

$$
(v^* + A P_e(p_n^*, M_n^{\text{max}})) p_n^* = 0,
$$  \hspace{1cm} (27)

$$
v^* = - A P_e(p_n^*, M_n^{\text{max}}) = AB \sqrt{\frac{1}{p_n} - \frac{3\pi^2}{2} \frac{3e\lambda p_n^*}{M_n^{\text{max}} - 1}}
$$  \hspace{1cm} (28)

Where $B = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1}{M_n^{\text{max}}} - \frac{3e\lambda p_n^*}{M_n^{\text{max}} - 1}}$.

From the above derivations, the $g(p_n)$ is defined as,
there are video to propagated to the higher layers. of modulation because errors made in Second can send the video layers within the limited bandwidth. However, testing all of the modulation of three modulation size transmitter can send is close to optimal value. As switching the modulation, the number of layers $t$ is also changed that dominates the overall performance. Therefore, in this paper, a simple while effective algorithm to choose the best modulation vector is proposed. Since the single carrier system is assumed, all layers in the $n$-th video stream have the same power and channel gain, which means that the layer which has the largest modulation size (i.e., has the large-error rate) becomes a bottleneck that dominates the overall performance. Therefore, the first principle for choosing the modulation vector is that the lowest modulation size has to be chosen while transmitter can send the video layers within the limited bandwidth. Second principle is that lower layers need to have lower size of modulation because errors made in the lower layers are also propagated to the higher layers. By using the above two principles, the best modulation vectors can be decided for each video to transmit $l$ layers respectively, where $l = \{0, 1, L \}$, because the best modulation vector for each video is deterministic when the number of layers is defined. Since there are $L+1$ candidate modulation vectors, which include the case not transmitting any layer, for one video, resulting in $(L+1)^N$-times of trials to find the best modulation vector.

Table 1. Bisection Search Algorithm

| 1. $upper = \min g_n(1)$, for $n = 1, 2, \ldots, N_r$ |
| 2. $lower = 0$ |
| 3. $p_n = 0$, for $n = 1, 2, \ldots, N_r$ |
| 4. while($\sum_{n=1}^{N_r} (p_n - 1) > 0$) |
| 5. $\mu = (upper + lower)/2$ |
| 6. $p_n = g_n^{-1}(\mu)$, for $n = 1, 2, L, N_r$ |
| 7. if $\sum_{n=1}^{N_r} p_n - 1 < 1$) lower = $\mu$ |
| 8. else upper = $\mu$; end if |
| 9. end while |
| 10. $p_n = p_n/\sum p_n$, for $n = 1, 2, \ldots, N_r$ |
| 11. $p^* = [p_1^*, p_2^*, L, p_N^*]$ |

The optimal $\{ p_n^* \}$ for all $n$ users satisfying above optimality condition can now be found by using the bisection search algorithm shown in Table 1, since the $g(p_n)$ is a monotonically increasing function. In the Table 1, $U$ is a small positive number to stop the searching process when $p^*$ closes to optimal value.

B. Rate Adaptation

As switching the modulation, the number of layers the transmitter can send is varied, and optimal power for each video is also changed. In total $3^{LN}$ possible combinations of modulation vectors need to be searched to find the best modulation vector that maximizes the sum-utility, because one of three modulation sizes should be chosen for the $L \times N$ layers. However, testing all of the modulation vectors would take too much time and energy. Therefore, in this paper, a simple while effective algorithm to choose the best modulation vector is proposed. Since the single carrier system is assumed, all layers in the $n$-th video stream have the same power and channel gain, which means that the layer which has the largest modulation size (i.e., has the large-error rate) becomes a bottleneck that dominates the overall performance. Therefore, the first principle for choosing the modulation vector is that the lowest modulation size has to be chosen while transmitter can send the video layers within the limited bandwidth. Since the single carrier system is assumed, all layers in the $n$-th video stream have the same power and channel gain, which means that the layer which has the largest modulation size (i.e., has the large-error rate) becomes a bottleneck that dominates the overall performance. Therefore, the first principle for choosing the modulation vector is that the lowest modulation size has to be chosen while transmitter can send the video layers within the limited bandwidth. Second principle is that lower layers need to have lower size of modulation because errors made in the lower layers are also propagated to the higher layers. By using the above two principles, the best modulation vectors can be decided for each video to transmit $l$ layers respectively, where $l = \{0, 1, L \}$, because the best modulation vector for each video is deterministic when the number of layers is defined. Since there are $L+1$ candidate modulation vectors, which include the case not transmitting any layer, for one video, resulting in $(L+1)^N$-times of trials to find the best modulation vector.

$g(p_n) = -\log v^* = -\log AB + \frac{1}{2} \log p_n^* + \frac{3\lambda_n}{2\alpha^2 (M_n^{max} - 1)} p_n^*$ (30)

V. SIMULATION RESULTS

JSVM software [18] is used to generate the trace files of videos and Matlab is used to verify the PHY layer performance using the generated trace files. “City”, “Foreman”, and “Waterfall” videos, encoded by the reference SVC codec JSVM version 9.19 [19], are used as the experimental videos to measure the sum-utility of the proposed scheme. All videos have maximum frame rate as 30fps and GOP sizes are 8. The basis quantization parameter (QP) of one base layer and three enhancement layers are QP\=[48, 42, 36, 30] with corresponding uniform quantization stepsizes calculated by $q=2(QP-4)/4$ [14].

Fig. 2 shows the average sum-utility with respect to average received SNR when “City”, “Foreman”, and “Waterfall” videos are transmitted over the MU-MIMO channel to different users. The results show that the proposed scheme can achieve better average sum-utility than modified water-filling (M-WF) algorithm [8], original water-filling algorithm, equal power allocation scheme, and MAXMIN power allocation scheme. M-WF and WF algorithms can achieve the optimum sum-rate, but it does not guarantee all users have better experience. Fig. 3 represents the average utilities of 3-received videos for 3-different users. The proposed scheme has better utility values than other schemes for all 3-users, and it means that users can receive more number of video layers without error in average. The proposed scheme can achieve better utility, since the proposed scheme allocate more power to the user can increase the utility the best with the smallest power, while the M-WF and WF algorithm allocate more power to the user has better channel condition.

Fig 4 shows more practical measurement of QoE in terms of peak-signal-noise-ratio (PSNR) [20] at 30dB-SNR. The
proposed scheme can achieve better PSNR for all 3-videos, while M-WF, WF, Equal and MAXMIN power allocation schemes cause more frame drops. From the Fig. 4, it is obvious that if users can receive more number of video layers, users will see much similar video from the original video.

VI. CONCLUSION

In this paper, a wireless video streaming algorithm using MU-MIMO and SVC technologies is proposed, to achieve higher sum-utility and QoE. An optimal solution for maximizing sum-utility is provided, which is originally an NP-hard problem, by using a convex optimization technique and rate adaptation scheme. The proposed scheme can allocate the optimal power and choose the best video rates for different videos to maximize sum-utility, and all users can achieve better QoE than maximizing sum-rate.

REFERENCES