Spatially Scalable Video Broadcasting in Multiple Antenna Systems

Arash Vosoughi[1], Seok-Ho Chang[2], Sang-Hyo Kim[3], Pamela C. Cosman[4] and Laurence B. Milstein[4]

[1]LG Electronics Mobile Research Lab, San Jose, CA 95131, USA
[2]CSE Dept., Dankook University, Yongin, Korea
[3]COICE, Sungkyunkwan University, Suwon, Korea
[4]ECE Dept., University of California at San Diego, La Jolla, CA 92093, USA

***Abstract*—We propose an efficient multiple-input multiple-output (MIMO) video broadcasting technique with low computational complexity that well serves different types of users residing inside the service area of a base station. We consider a video broadcasting scenario in which heterogeneous users with different display resolutions, different operating bit rates, and different numbers of receive antennas are present. Our proposed scheme adopts spatially scalable video coding, and makes use of both spatial diversity and spatial multiplexing techniques. We compare the performance of our proposed scheme with that of a non-scalable video transmission scheme. Simulation results show that our proposed video broadcasting scheme significantly outperforms the non-scalable video broadcasting strategy. The performance improvement is observed for both different types of users we consider.**

I. Introduction

Cross-layer optimization of wireless multimedia communications [1]–[2] has been motivated by the increasing demand for mobile multimedia services. Multimedia scalable sources, such as scalable video or progressive images [3]–[4], have a desirable feature that the quality of the decoded source improves as the number of successfully received bits increases. Such advances in source codecs, however, have made the source bit streams very susceptible to the impairments of mobile fading channels.

Multiple-input multiple-output (MIMO) channels offer large gains in terms of link reliability and data rate. Spatial diversity techniques, such as orthogonal space-time block codes (OSTBC) [5], [6], extract diversity gain to combat signal fading from the channels, and obtain reliability. OSTBC is an important class of linear STBC, in that it achieves the full diversity of channels with a very simple linear receiver. Spatial multiplexing techniques use a layered approach to increase data rate [7], [8]. One popular example is the vertical Bell Laboratories layered space-time (V-BLAST) architecture, where independent data signals are transmitted over multiple antennas to increase the data rate. Although spatial multiplexing increases the data rate, it does not usually achieve full spatial diversity.

In mobile video broadcasting systems such as digital video broadcasting (DVB) [9], there exist various kinds of user equipment in the service area, and they in general have different numbers of antennas. As an example, a tiny, low-end mobile phone may have a single antenna, due to its limited hardware space. On the other hand, a device such as a tablet or notebook computer usually has more than one antenna, since it retains large hardware space as well as high computing capability. We assume that users with different numbers of receive antennas and different screen resolutions are present within the service area of a base station. As a simple example, this paper considers a MIMO video broadcasting system where a base station incorporates two transmit antennas, and two different types of user devices reside in the service area: i) a *big user* (i.e., a user of large hardware space) with two antennas and a higher-resolution screen, ii) a *small user* (i.e., a user of small hardware space) with a single antenna and a lower-resolution screen. For this setup, we propose an efficient video broadcasting scheme which combines spatial diversity and spatial multiplexing. The *base layer* (BL) of the spatially scalable bit stream is encoded by spatial diversity techniques such as the Alamouti code, while the *enhancement layer* (EL) is encoded by spatial multiplexing techniques such as V-BLAST. Therefore, a small user is capable of decoding only the BL (since it has only one receive antenna and is not able to decode a V-BLAST coded data), while a big user is able to decode both the BL and EL (since it has two receive antennas and is able decode both Alamouti and V-BLAST coded data).

We compare the performance of the proposed scheme with that of a *baseline* scheme, where video data is compressed using a non-scalable codec and broadcast using only spatial diversity.

We show that adoption of the proposed scheme significantly improves the video quality for the big user compared to the baseline scheme. We also show that the proposed scheme leads to a PSNR (peak signal-to-noise ratio) loss for the small user, but this loss happens in the range of very high PSNRs that is perceptually unnoticeable.

The rest of this paper is organized as follows. We present the proposed MIMO video broadcasting scheme in Section II. We provide simulation results in Section III, and conclude this paper in Section IV.

II. MIMO Video Broadcasting

The MIMO video broadcasting system we consider in this paper consists of a base station with two transmit antennas. We also assume that two kinds of user devices are residing in the service area of the base station: a *big user* with two receive antennas and a higher-resolution screen, and a *small user* with a single receive antenna and a lower-resolution screen. A baseline scheme for such a MIMO broadcasting system is described in subsection II-A, and the proposed scheme is detailed in subsection II-B.

A. Baseline MIMO Video Broadcasting Scheme

The baseline scheme we consider in this study adopts non-scalable coding for video compression. It broadcasts a video bit stream using only spatial diversity (in particular, the Alamouti code) so that the coded video bit stream is decodable by both the small and big users. Note that the transmitter is not allowed to use spatial multiplexing, because then a small user with only one receive antenna is not able to decode the coded bit stream. Fig. 1 depicts a block diagram of this broadcasting scheme. The system produces a non-scalable bit stream that is then converted into a sequence of channel codewords with error detection and correction capability. The coded bit stream is then mapped to constellation symbols. The constellation symbols are encoded by the Alamouti code and transmitted from two transmit antennas afterward.

B. Proposed MIMO Video Broadcasting Scheme

Fig. 2 depicts the proposed MIMO video broadcasting scheme. It employs spatially scalable video coding for video compression, where the scalable bit stream has a BL and an EL. A nice feature of a scalable bit stream is that it naturally enables the use of unequal error protection (UEP) so that the more important layer (i.e., BL) is protected more compared to the less important layer (i.e., EL). According to Fig. 2, the proposed broadcasting scheme applies different space-time codes to the BL and EL. In particular, it applies the Alamouti code to the BL and V-BLAST to the EL. The resulting BL and EL symbols are multiplied by the transmit gains G_{BL} and G_{EL}, respectively, and then are superposed to yield the final symbol stream which is transmitted from two transmit antennas.

Note that the alphabet size of the constellation symbols of the BL is in general different from that of the EL. The alphabet sizes of the BL and EL will be determined based on the lengths of the coded BL and EL bit streams. We also note that the UEP is achieved by the unequal transmit gains placed in the front end of the transmitter. The transmit gain ratio $\beta = G_{BL}/G_{EL}$ (> 1) determines how much more the BL is protected compared to the EL against channel errors.

In Fig. 2, s_i^{BL} $(i = 1, 2, 3, \ldots)$ denotes the BL constellation symbols, and x_i^{BL} and y_i^{BL} denote the BL symbols which are transmitted from antennas x and y, respectively. Since x_i^{BL} and y_i^{BL} are encoded by the Alamouti code, we have

$$\begin{aligned} x_{2i-1}^{BL} &= s_{2i-1}^{BL}, \quad x_{2i}^{BL} = -(s_{2i}^{BL})^* \\ y_{2i-1}^{BL} &= s_{2i}^{BL}, \quad y_{2i}^{BL} = (s_{2i-1}^{BL})^*, \end{aligned} \tag{1}$$

where $(\cdot)^*$ denotes the complex conjugate operation. s_i^{EL} denotes the EL constellation symbols, and x_i^{EL} and y_i^{EL} denote the EL symbols that are transmitted from antennas x and y, respectively. If the EL is encoded by V-BLAST, x_i^{EL} and y_i^{EL} are given by

$$x_i^{EL} = s_{2i-1}^{EL}, \quad y_i^{EL} = s_{2i}^{EL}. \tag{2}$$

The final transmit symbols are given by

$$\begin{bmatrix} G_{BL}\, s_{2i-1}^{BL} + G_{EL}\, s_{4i-3}^{EL} & -G_{BL}\,(s_{2i}^{BL})^* + G_{EL}\, s_{4i-1}^{EL} \\ G_{BL}\, s_{2i}^{BL} + G_{EL}\, s_{4i-2}^{EL} & G_{BL}\,(s_{2i-1}^{BL})^* + G_{EL}\, s_{4i}^{EL} \end{bmatrix}, \tag{3}$$

where each row corresponds to a transmit antenna and each column corresponds to a time symbol.

We note that the BL, to which the Alamouti code is applied, can be decoded by either a small or a big user. On the other hand, the EL, which is encoded by V-BLAST, is decodable only by a big user. The motivation for this is as follows. (i) A big user needs to decode many source bits to meet the quality of its higher-resolution screen. On the other hand, the lower-resolution screen of a small user can produce satisfactory quality with a small number of source bits. Thus, for a small user, decoding of only the BL of spatially scalable video is sufficient. (ii) When the EL is encoded by spatial multiplexing instead of spatial diversity, the video quality for a big user is significantly enhanced. This is because the size of the EL is typically greater than that of the BL, and thus a high spectral efficiency is employed for the EL. Note that spatial multiplexing outperforms spatial diversity at high spectral efficiencies in terms of error probabilities [10], [11]. The reason is that, when both spatial diversity and spatial multiplexing transmit at the same spectral efficiency, spatial multiplexing can use a smaller symbol alphabet size on each stream compared to the spatial diversity. Therefore, spatial multiplexing benefits from a larger minimum Euclidean distance compared to the spatial diversity, and hence outperforms spatial diversity at low SNRs. Since spatial diversity has a much smaller minimum distance than spatial multiplexing for high spectral efficiency, the advantage of spatial multiplexing over spatial diversity becomes more significant. The mathematical proof of this argument is presented in [10]. Note that in our proposed scheme, β needs to be greater than unity for unequal error protection of BL and EL. On the other hand, the result regarding the tradeoff between spatial multiplexing and spatial diversity does not depend on how large β is. Thus, the tradeoff between spatial multiplexing and spatial diversity is valid in our proposed scheme.

III. Results

We evaluate the performance of the proposed MIMO video broadcasting scheme in this section. The decoded video quality

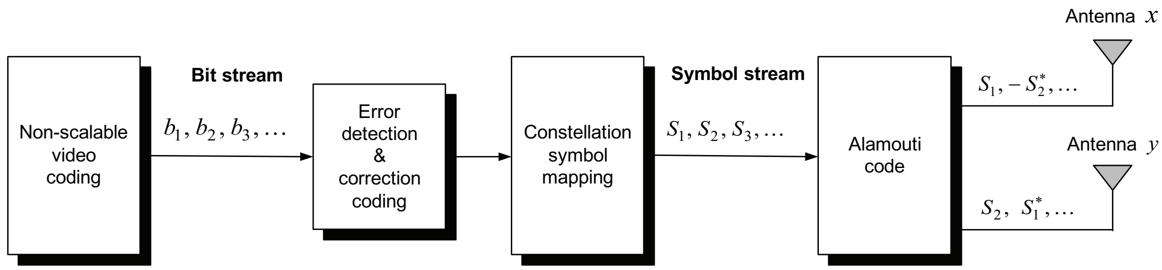

Fig. 1. A baseline MIMO video broadcasting scheme with non-scalable video and uniformly-spaced signal constellation. b_i denotes the compressed bit of the non-scalable video, and S_i denotes the constellation symbol. $(\cdot)^*$ denotes the complex conjugate operation.

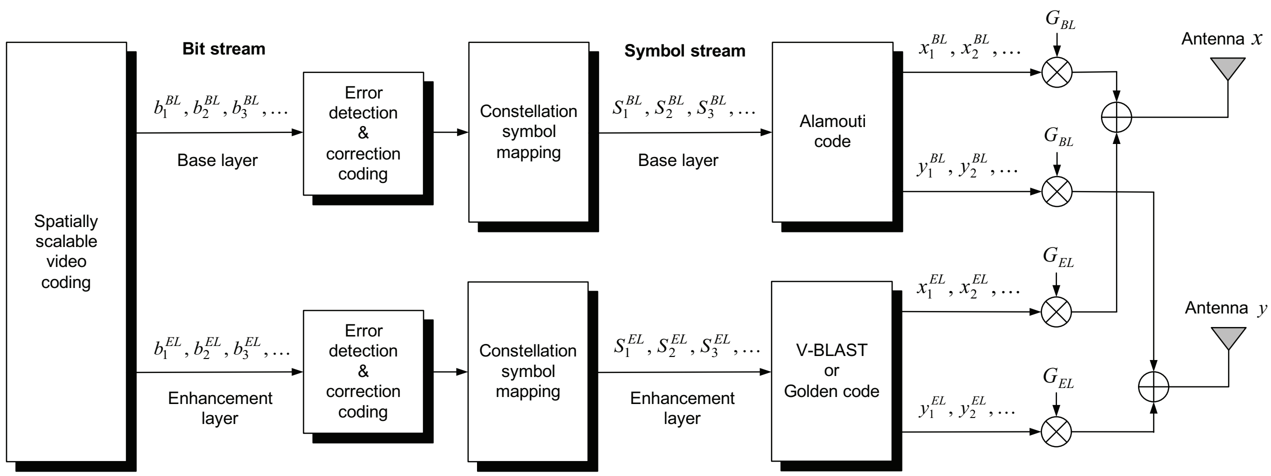

Fig. 2. The suggested MIMO video broadcasting scheme with spatially scalable video and superposed space-time codes. b_i^{BL} and b_i^{EL} denote the compressed bits for the base and ELs, respectively. s_i^{BL} and s_i^{EL} denote the constellation symbols for the base and ELs, respectively. x_i^{BL} and y_i^{BL} denote the symbols for the BL which are transmitted from antennas x and y, respectively; x_i^{EL} and y_i^{EL} denote the symbols for the EL that are transmitted from antennas x and y, respectively. G_{BL} and G_{EL} denote the transmit gains multiplied to the base and ELs, respectively.

is measured using the peak-signal-to-noise-ratio (PSNR). In our simulations, we use QPSK for BL, and 16-QAM for EL. This choice implies that, during one symbol time period, the proposed scheme transmits 2 bits for the BL, and 8 bits for the EL. This is because the spatial multiplexing rate of the Alamouti code is one, and that of V-BLAST for two transmit antennas is two. For the non-scalable baseline scheme (Fig. 1), we employ the 1024-QAM constellation so that the date rate of this scheme (which is 10 bits for one symbol time period) is equal to that of the proposed scheme.

We use the optimal maximum likelihood (ML) decoding for the non-scalable baseline scheme. For the proposed scheme, we use a *successive* decoding algorithm [12] as detailed below:

1) Alamouti decoding is performed on the received signal to decode the symbols to which the BL is mapped.
2) The decoded symbols are subtracted from the received signal.
3) ML decoding for V-BLAST is performed on the residual signal to decode the symbols to which the EL is mapped.

Although the above successive decoding algorithm has a suboptimal performance, it offers much lower computational complexity compared to the ML decoding of the entire received signal.

We use the H.264/SVC (JSVM software) to generate a spatially scalable video bit stream. The system performance is evaluated for video sequence 'Foreman' with resolution 352×288 and frame rate 30 fps. For this choice, decoding both the BL and EL by a big user yields a full-resolution reconstruction of 352×288. Note that a small user is only able to decode the BL which yields a lower-resolution reconstruction of 176×144. The PSNR is computed between the reconstructed video at the decoder and a reference uncompressed video. For the big user, the uncompressed video with original resolution 352×288 is used as the reference, while, for the small user, the original 352×288 reference video is lowpass filtered and downsampled, and the resulting 176×144 video is used as the reference. We use a hierarchical B-frame group of pictures (GOP) structure, where each GOP has 16 frames. We assume that the transmitted video signal experiences a slow fading wireless channel such that channel coefficients are nearly constant over a GOP. We also assume perfect channel estimation at the receiver.

Fig. 3 depicts the PSNR performance of a big user when the proposed scheme is employed. Recall that parameter β controls how much more strongly the BL is protected compared to the EL. For any given β in Fig. 3, we observe that when

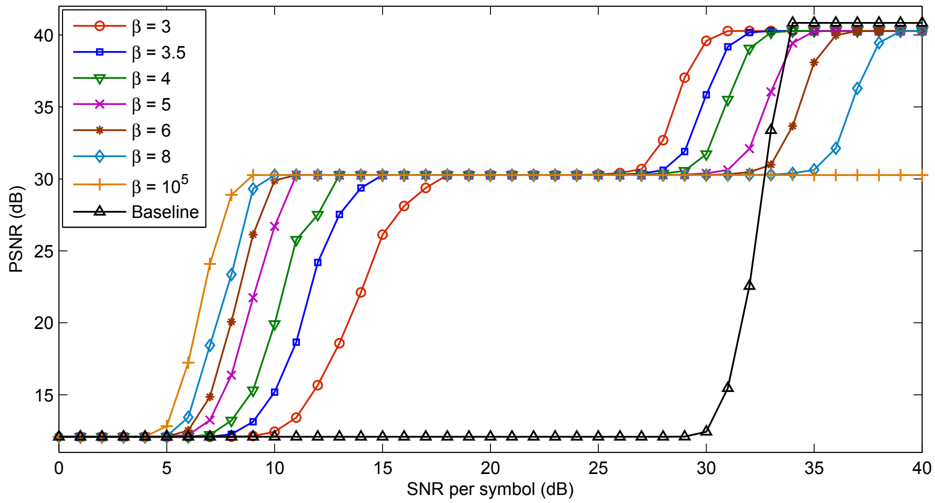

Fig. 3. SNR performance of a big user. The performance of the non-scalable baseline is shown together with that of the proposed scheme.

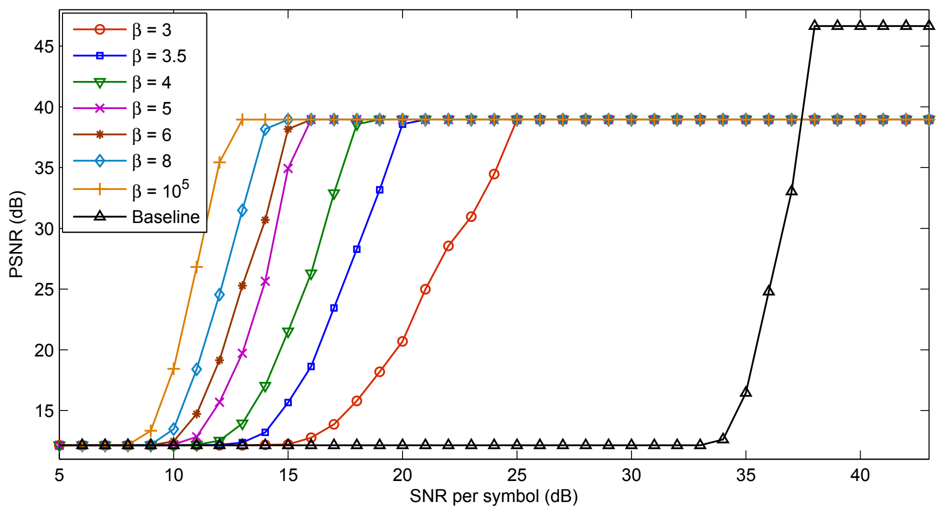

Fig. 4. PSNR performance of a small user. The performance of the non-scalable baseline is shown together with that of the proposed scheme.

SNR is low, the PSNR reaches a plateau of about 30 dB. The reason is that for low SNR values, only the BL is decodable, since it has been protected more strongly compared to the EL. When only the BL is decoded, a lower-resolution video is reconstructed which needs to be upsampled in order to be displayed on the higher-resolution screen of a big user. The PSNR plateau of 30 dB is due to upsampling the BL. When channel SNR increases, the receiver gradually becomes able to decode the EL, and thus the quality of the decoded video gradually improves until it reaches the maximum PSNR value of about 40 dB. Another observation is that, for any given low SNR value, when β increases, the performance in the range of low PSNRs improves. For a fixed channel SNR in that range of PSNRs, increasing β corresponds to providing a stronger protection for the BL, which leads to a better reconstruction quality. However, when β increases, PSNR decreases in the range of high PSNRs. The reason is that a higher β translates to providing a weaker protection for the EL, which implies that we need a higher SNR to achieve the same reconstruction quality with a larger β.

Fig. 3 also depicts the PSNR performance of the non-scalable baseline scheme. We see that the maximum PSNR the non-scalable baseline scheme can achieve is slightly larger than the maximum PSNR achieved by the proposed scheme. The reason is that, at very high SNRs, there are no channel errors and compression efficiency solely determines the quality of the received video. It is well known that the compression efficiency of non-scalable coding is slightly higher than that of the scalable coding [3]. Disregarding this minor performance loss, which only pertains to the highest SNRs and PSNRs, we see that the proposed scheme significantly outperforms the non-scalable baseline scheme.

Fig. 4 depicts the PSNR performance of a small user when the proposed scheme is employed. Recall that the proposed scheme employs V-BLAST to encode the EL, and that a small user with a single receive antenna cannot decode data that is encoded by V-BLAST. Thus, the small user is able to decode only the BL and it can achieve a maximum PSNR value of about 39 dB. We also observe that, similar to the case of a big user, as β increases, the performance for low PSNRs improves. Fig. 4 includes the PSNR performance of the non-scalable baseline scheme. The non-scalable baseline provides a slightly higher maximum PSNR compared to the other schemes. However, the proposed scheme outperforms the non-scalable baseline scheme over the entire range of PSNRs we are interested in.

Figs. 3 and 4 tell us that the proposed scheme outperforms the baseline scheme both for the big user and small user, except in the range of PSNRs which are too high to be of interest.

IV. Conclusions

An efficient video broadcasting scheme was proposed for MIMO communication systems, where users with different display resolutions and different numbers of receive antennas reside in the service area. In our design, we used spatially scalable video coding, and employed both spatial diversity and spatial multiplexing techniques. In particular, we adopted the widely used Alamouti code and V-BLAST to encode the BL and EL symbol streams, respectively, and multiplied the two coded symbol streams by different transmit gains to unequally protect the BL and EL against channel errors. For both big and small users that we considered in this study, it was shown that our proposed scheme significantly outperforms the video broadcasting scheme, which employs non-scalable video coding. We pointed out that the proposed broadcasting scheme suffers a performance loss only in the range of highest SNR and PSNR values for the small user. However, we showed that the proposed scheme still yields a high enough PSNR value such that the PSNR loss is almost not perceivable by the human visual system.

Acknowledgment

This work was partially supported by the Army Research Office under Grant #W911NF-14-1-0340, by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2065143), and by 'The Cross-Ministry Giga KOREA Project' grant from the Ministry of Science, ICT and Future Planning, Korea.

References

[1] Z. Yang and X. Wan, "Scalable video broadcast over downlink MIMO-OFDM systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, pp. 212–223, 2013.

[2] S.-H. Chang, "Joint optimization of physical and application layers for wireless multimedia communications," Ph.D. dissertation, University of California, San Diego, 2010.

[3] 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, 2007.

[4] D. Taubman and M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice. Norwell, MA: Kluwer, 2001.

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.

[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[7] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[8] R. Louie, M. McKay, and I. Collings, "Open-loop spatial multiplexing and diversity communications in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 57, pp. 317–344, 2011.

[9] "Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television," ETSI EN 300 744 V1.5.1, Nov. 2004.

[10] S.-H. Chang, P. C. Cosman, and L. B. Milstein, "Optimal transmission of progressive sources based on the error probability analysis of SM and OSTBC," *IEEE Trans. Veh. Technol.*, vol. 63, pp. 94–106, 2014.

[11] H. Yang, "A road to future broadband wireless access: MIMO-OFDM based air interface," *IEEE Commun. Mag.*, vol. 43, pp. 53–60, 2005.

[12] S.-H. Chang, M. Rim, P. C. Cosman, and L. B. Milstein, "Superposition MIMO coding for the broadcast of layered sources," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3240–3248, 2011.