Switching Between Hybrid MIMO Structures for Video Transmission Based on Distortion Model

Martin B. Obando, Walter C. Freitas Jr and Francisco R. P. Cavalcanti
Federal University of Ceará, Wireless Telecommunications Research Group (GTEL)
CP 6005, Campus do Pici, 60455-760, Fortaleza-CE, Brazil
Email:{martin, walter, rodrigo}@gtel.ufc.br

***Abstract*—In this paper, we propose a switching between hybrid Multiple-Input Multiple-Output (MIMO) structures based on a distortion model per Group of Pictures (GOP) to improve the transmission of video streaming over wireless systems. The proposed algorithm uses measurements of the application and physical layers. In the application layer, a distortion model for GOP is used based on the packet loss rate and media data units. In the physical layer, the switching between hybrid MIMO structures is realized by simultaneously taking advantage of multiplexing and diversity gains provided by hybrid MIMO structure. The simulation results presented in this article show that the switching between hybrid MIMO structures based in the distortion model per GOPs is capable of improving the objective and subjective video quality at the end user.**

I. INTRODUCTION

The video streaming transmission consists in the transferring of video data as a continuous stream. Thus, the end user can start displaying the video data before the entire file has been transmitted. In order to accomplish this, bandwidth efficiency and flexibility between video servers and equipment of end-users are important requirements [1].

When wireless communications systems are concerned, the task of video streaming transmission is even more challenging. Wireless links are in general affected by propagation mechanisms such as long and short term fading which degrade the quality of the received signal leading to packet loss. In recent years, the transmission of video over wireless communication systems and their integration with the Internet is a subject of major interest. This is employed in the IP Multimedia Subsystem (IMS) that is standardized by the 3rd. Generation Partnership Project (3GPP) and consists in a system of services for next generation networks [2].

The cooperation between the different layers that compose the protocol stack of wireless communication networks is an alternative that can be used in order to improve the transmission of multimedia services in wireless systems [3]. In the literature, there are several works that use this mechanism to improve the quality of multimedia services, such as video.

In [4] the authors propose an Hybrid MIMO Structure (HMS) to implement Unequal Error Protection (UEP) for video delivery in Multiple-Input Multiple-Output (MIMO) system. The goal of that work was to exploit the diversity gain to provide better protection to the high priority data, while transmitting the low priority data with spatial multiplexing to achieve high data rate. In [5], the authors propose a layered video transmission over MIMO, which is called Adaptive Channel Selection, where each bitstream is switched among multiple antennas periodically. In application layer is adopted a fine grain scalability in order to satisfy both targets bit rate and error robustness.

In [6] a transmission distortion estimation approach for video delivery over hybrid channels in successive frames with the prediction mode was proposed. The method takes into account packet erasures in wired networks and bit errors in wireless links. Using this method, the sender can estimate the transmission distortion of received decoded video for a given time-varying channel conditions. In [7] a link adaptation scheme to improve the Quality of Service (QoS) for video transmission was presented. It works minimizing the distortion of the received video sequence. Using local rate distortion measures and end-to-end distortion model at the video encoder, the scheme estimates the received video distortion at the current rate, as well as on adjacent lower and higher rate, which allows the system to select the link-speed that offers the lowest distortion and to adapt to the channel conditions.

In this work, we propose an algorithm that switches between HMSs based on a distortion model by Group of Pictures (GOP) to improve the video streaming quality service perceived at the end user. The difference among the works above mentioned and our proposed algorithm is that in [4] was proposed a priority bitstream transmission using the HMS and in [5] was proposed a bitstream transmission based in adaptive channel selection. While in our work, a GOP is transmitted using a switching between HMSs based in a distortion model by GOP providing UEP in the physical layer. The main contribution of our proposal is to exploit the HMSs and distortion model by GOP for video streaming transmission and to improve the objective and subjective video quality performance at the end user. Furthermore, the algorithm requires a simple implementation and small amount of feedback.

The remainder of the paper is organized as follows: in section II, we present the system and channel model assumed. In section III, we explain the distortion model estimation considered. Section IV, presents the HMSs and their interference cancellation algorithm used in this work. In section V, we present the proposed algorithm. Finally, section VI presents the performance results and section VII concludes this paper.

978-1-4244-3574-6/10/$25.00 ©2010 IEEE

II. CHANNEL MODEL

We consider wireless transceivers equipped with M_{tx} transmit antennas and M_{rx} receive antennas. The wireless channel is assumed to have rich scattering and flat fading that could be achieved with Orthogonal Frequency Division Multiplexing (OFDM), even for a high demanding symbol rate as for video transmission. The channel is quasi-static, i.e., the channel does not change significantly in a single block, but can vary from block to block. In each block we can represent the sampled received signal as in [8]:

$$\mathbf{Y} = \sqrt{\frac{\rho}{M_{\text{tx}}}}\mathbf{H}\mathbf{S} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{C}^{M_{\text{rx}} \times T}$ is the received signal matrix, T is the number of signaling intervals, $\mathbf{S} \in \mathbb{C}^{M_{\text{tx}} \times T}$ is the transmitted signal matrix, ρ is the average signal to noise ratio at each receiver antenna, $\mathbf{H} \in \mathbb{C}^{M_{\text{rx}} \times M_{\text{tx}}}$ is the random channel matrix and $\mathbf{N} \in \mathbb{C}^{M_{\text{rx}} \times T}$ is the additive noise matrix, whose entries are i.i.d. Zero Mean Circularly Symmetric Complex Gaussian (ZMCSCG).

III. DISTORTION ESTIMATION MODELS

In video coding the most important standards as the H.264/AVC and MPEG4 [9] define three main frames types for the compressive video streams, including the I frame (Intra-coded), P frame (Predictive-coded) and B frame (Bidirectionally predictive-coded). The I frames are encoded independently and decoded by themselves. The P frames are encoding using predictions from the preceding I or P frames in the video sequence. The B frames are encoded using predictions from the preceding and succeeding I or P frames.

Expected distortion is an important measure of video quality in packet-based video streaming where video packets may be lost due to random channel errors in wireless systems. The majority of the studies concerning loss-distortion modeling uses Mean Square Error (MSE) as a metric for quality distortion. Given a packet loss probability of p_k, the expected distortion for packet k can be expressed as [10] by

$$E[D_k] = (1 - p_k)E[D_k^{\text{r}}] + p_k E[D_k^{\text{l}}], \tag{2}$$

where $E[D_k^{\text{l}}]$ denotes the expected distortion given the data packet k is lost and $E[D_k^{\text{r}}]$ denotes the expected distortion given the packet is successfully received. Based on Equation (2) the end-to-end distortion for a sequence of video packets can be expressed as function of all the individual packets expected distortion as considered in [10]. The total distortion can be function of both the source coding and the network. The source coding affects the source distortion while the network impacts on the packet loss probability p_k. In this paper, we use the distortion model per GOP (i.e. in which all the video sequence is decomposed into smaller units which are then coded together) and the media data units process in Network Abstraction Layer Units (NALU)s, as it is presented in [11]. We assume this model since it presents a good approach in the estimate distortion using the H.264/AVC standard for video streaming services. In this model the total quality distortion (in terms of MSE) from losing media data units can be modeled with an additive model where the total distortion is the sum of the initial compression distortion D_c (i.e. distortion due to the source coding) and the individual quality distortions D_i derived from each lost NALU. As we are interested in mean distortion at certain packet loss rate p, the average D_{Total} can be estimated as in [12]

$$D_{\text{Total}} = D_c + p \times \sum_{i=0}^{nal} D_i, \tag{3}$$

where nal is the number of NALUs in each GOP. It is known from the literature that the additive distortion model per GOP gives accurate results, since the impacted areas are sufficiently far apart from each other.

IV. HYBRID MIMO SYSTEM

The use of MIMO systems can be exploited generally to achieve two main objectives. The first one is link reliability (i.e. the diversity gain) using for example Orthogonal Space-Time Block Code (OSTBC) schemes, in which the same symbol stream is transmitted from different transmit antennas in appropriate manner to obtain transmit diversity. The second option is to obtain link spectrum efficiency (i.e. spatial multiplexing gain) using for example Vertical Bell Laboratories Layered Space-Time Architecture (VBLAST) schemes, in which different symbol streams are simultaneously transmitted in all transmit antennas.

Another way to exploit the MIMO systems is using the HMS, where the transmission process of a HMS can be divided in layers, somewhat like VBLAST. However, in contrast to VBLAST, in the HMS case a layer may consist of the stream of symbols at the output of a OSTBC, which is sent to a group of antennas, or of an uncoded stream, which is transmitted from a single antenna. Based on this concept of layers, the HMS combine pure diversity schemes (e.g. OSTBC) with pure spatial multiplexing schemes (e.g. VBLAST). In the following we explain the HMSs.

A. Hybrid MIMO Structure:G2+1+1

This structure consists of four antennas arranged at different parallel layers. The first two antennas form the first layer of the HMS. This layer uses OSTBC known as Alamouti (G2) [13] and provides diversity gains. The third and fourth antennas comprise the second and third layers and provide spatial multiplexing gain. In this way, this HMS is referred as G2+1+1 in [14], the transmission matrix $\mathbf{S}$ of this structure is given

$$\mathbf{S}[T_1, T_2] = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \\ s_3 & s_4 \\ s_5 & s_6 \end{bmatrix}, \tag{4}$$

in which $K = 6$ symbols (i.e. s_1, s_2, s_3, s_4, s_5 and s_6) are transmitted in $T = 2$ (i.e. T_1 and T_2) consecutive signaling intervals. Therefore, its effective spectral efficiency is given by

$$\eta = (K/T) \cdot \log_2(\mu) = 3 \cdot \log_2(\mu), \tag{5}$$

where μ is considered the cardinality of the modulation scheme.

B. Hybrid MIMO Structure: G2+G2

In this structure the first two antennas form the first layer and the remaining antennas form the second layer of the HMS. In this structure two OSTBC scheme also known as Alamouti (G2) are configured in each layer. This structure is referred as G2+G2 in [14] and provides both multiplexing and diversity gain. The transmission matrix $\mathbf{S}$ of this structure is given by:

$$\mathbf{S}[T_1, T_2] = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \\ s_3 & -s_4^* \\ s_4 & s_3^* \end{bmatrix}, \tag{6}$$

in which $K = 4$ symbols (i.e. s_1, s_2, s_3 and s_4) are transmitted in $T = 2$ (i.e. T_1 and T_2) consecutive signaling intervals providing an effective spectral efficiency of

$$\eta = (K/T) \cdot \log_2(\mu) = 2 \cdot \log_2(\mu). \tag{7}$$

C. Interference Cancellation Algorithm for the Hybrid MIMO Structures

Since HMS combine OSTBC with VBLAST in parallel, all HMS have at least two layers, at least one of which is space-time block coded. Further, they all employ OSTBC, whose maximum-likelihood detection involves simple linear operations in the receiver. We consider a receiver for the HMS that combines the simplicity of Succesive Interference Cancellation (SIC) [15] algorithm with the simplicity of decoding of OSTBC. In fact, we adapt the Interference Cancellation (IC) algorithm in such a way that the orthogonal structure of the space-time code is preserved as much as possible in its output signal. The general structure of the receiver is shown in Fig. 1

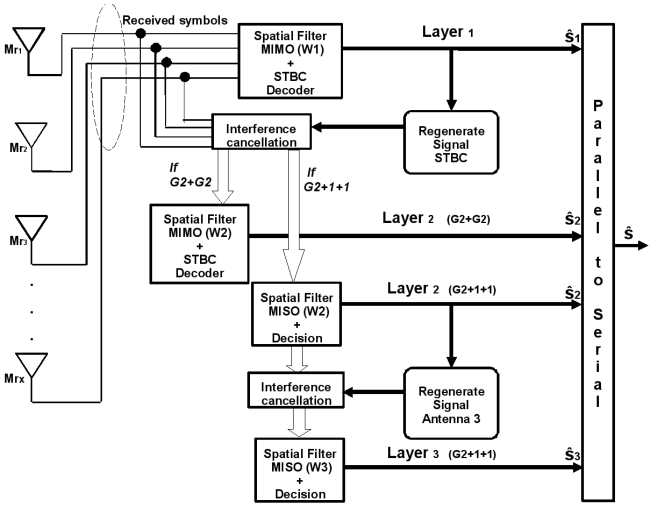

Fig. 1. Interference Cancellation Algorithm used with the HMS.

In Fig. 1 the MIMO spatial filter mitigates the interference from other layers, so that its output signal consists of a single space-time-coded signal or of a single uncoded stream. In our SIC approach, we will first detect the G2 layer, which is more robust. To that end, we first employ a linear Minimum Mean Square Error (MMSE) filter to eliminate the interference from the uncoded layer. Clearly, we can see in this figure that the layers are processed successively, in a two stage process in which

1) first a nulling of the interference from the undetected layers is made, then, the output signal goes through a decoder for the Space Time Block Code (STBC) used in this layer;
2) finally, the received space-time coded signal corresponding to this layer is regenerated and its impact is cancelled from the received signal.

V. PROPOSED ALGORITHM FOR VIDEO STREAMING TRANSMISSION

In the following we present the algorithm specification for video streaming transmission in wireless systems. The input data necessary for initialization of our algorithm, where the video coding process generate a GOP structure of ten frames in the form (IBBPBBPBBP). It is worth to notice that the order of NALUs is altered due to the fact that B frames require the succeeding anchor frame to be encoded first; therefore, the actual order of NALUs in the bitstream is (IPBBPBBPBB). The number of NALUs per GOP (i.e. nal) is defined by the coding rate of the video sequence and the size of the NALU (i.e. $Size_{\text{\_NAL}}$) assumed for video streaming transmission in a wireless system.

The estimation of the packet loss rate p is performed in the receiver side. This parameters can be transmitted using some feedback mechanism or using the encapsulation process.

In the transmitter side there are two HMSs configured with different spectral efficiencies. Thus, estimate p can be based on the effective information received C^{e} which is given by

$$C^e = C^{\text{e}}_{\text{delay}} \times (1 - PER_{\text{MIMO}}) \times BW_{\text{user}}, \tag{8}$$

where $C^{\text{e}}_{\text{delay}}$ represents the effective information received given the delay presented in the HMS itself, BW_{user} represents the bandwidth per user and $(1 - PER_{\text{MIMO}})$ represents the effective information received given the Packet Error Rate (PER) in the HMS, where for the PER_{MIMO} estimation is assumed that each bit inside the packet has the same Bit Error Rate (BER) and bit errors are uncorrelated, then the PER_{MIMO} can be related to the BER through [17]

$$PER_{\text{MIMO}} = 1 - (1 - BER)^L, \tag{9}$$

for a packet containing L bits. Finally the p expression is given by

$$p = 1 - \left(\frac{\frac{C^{\text{e}}}{Size_{\text{\_NAL}}} \cdot 100}{nal} \right). \tag{10}$$

The estimation of packet error rate in the deactivate HMS (i.e. p_{ds}) is done in the transmitter side. For estimate the distortion due to source coding (D_c) the MSE between the original and not transmitted decode GOP is used.

Then, the distortion model in Equation (3) is used to calculate the average distortion by GOP (i.e. D_{GOP}) and the distortion in the deactivate HMS over a GOP (i.e. $deactD_{\text{GOP}}$). If the estimated distortion D_{GOP} is higher than $deactD_{\text{GOP}}$, then a switching between the HMS structures is performed. For the remaining GOPs in the video sequence the same process is repeated.

A summary of the proposed algorithm is presented in Algorithm 1.

Algorithm 1 Proposed Algorithm

```
p ← 0;
D_GOP ← 0;
minD_GOP ← 1;
for all GOP ∈ VideoSequence do
    nal ← numNalus;
    p ← packetLostRate;
    p_ds ← packetLostRateDeactivedStructure();
    Dc ← calculeDistortionSource();
    D_GOP ← calculateDistortion(p,GOP,nal,Dc);
    deactD_GOP ← calculateDistortion(p_ds,GOP,nal,Dc);
    if D_GOP > deactD_GOP then
        HMS_G2+G2 Tx(GOP).
    else
        HMS_G2+1+1 Tx(GOP).
    end if
end for
```

VI. RESULTS

In this section, the performance of the proposed algorithm is evaluated by using Monte Carlo simulation.

In order to assess the performance of our proposed solution we simulated the HMS G2+G2, the HMS G2+1+1 and furthermore, the HMS with UEP as in [4]. We assume a MIMO system with 4 transmit and 4 receive antennas. We considered an 8-Phase Shift Keying (8PSK) modulation and a SIC detector for the HMSs. The Initial HMS used in the proposed algorithm is the G2+1+1 structure. We assumed the $C^{\text{e}}_{\text{delay}}$ per GOP in the transmitter side of 0.9 and 1 for HMSs G2+G2 and G2+1+1, respectively. Since the two HMSs have different spectral efficiency.

The video encoder used was the H.264/AVC extended profile from [18]. The video encoded bit rate is 128 kbps with GOP structure of ten frames in the form (IBBPBBPBBP). The video sequences used in the simulation were: “Foreman” and “Carphone” from [19]. Each video sequence consists of 240 frames, Quarter Common Intermediate Format (QCIF) of size 176 × 144 pixels per frame with 4:2:0 sampling. The packet size of 500 bytes was assumed which is used in video streaming transmission.

Fig. 2 shows the performance of the BER for both HMSs G2+G2 and G2+1+1. We can observe that the HMS G2+G2 presents better performance in comparison with the HMS G2+1+1. This is due to the fact that this HMS has a higher diversity gain since their two layers are configured with OSTBC G2 from [13]. However, the HMS G2+G2 has a cost in its spectral efficiency which leads to higher packet delays than HMS G2+1+1.

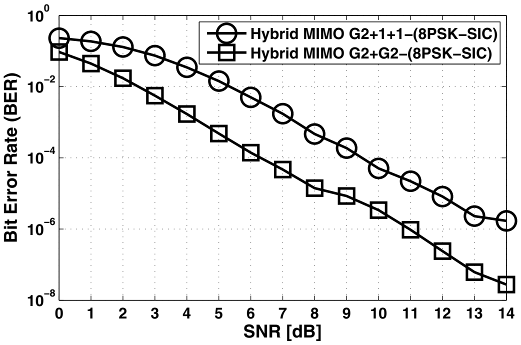

Fig. 2. Performance BER of hybrid MIMO structures.

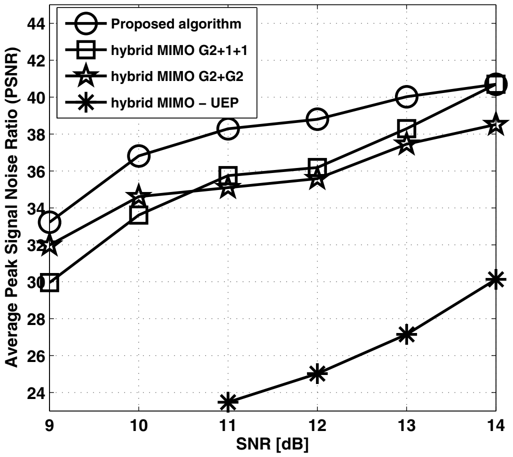

Fig. 3. Objective video quality.

In Fig. 3 we present the performance of the Peak Signal-to-Noise Ratio (PSNR) resulted of the objective evaluation. We can observe that the proposed algorithm in comparison with the HMSs G2+G2 and G2+1+1 presents a gain of approximately 2 dB in the Signal-to-Noise Ratio (SNR) range from 9 dB to 13 dB. This is explained by the fact that our proposed algorithm takes advantage of the distortion model per GOP minimizing the distortion perceived during the transmission. The strategy of switching between HMSs with spatial multiplexing and diversity gains given a measured of distortion per GOP improving the performance of the video received by the end user. The crossing between the two HMSs with respect to the relative measurement PSNR is due to the fact that the HMS G2+G2 presents a delay impact in comparison with G2+1+1 over all the video sequence, although of its BER outperforms the HMS G2+1+1 BER.

The proposed algorithm outperforms the HMS with UEP from a SNR of 9 dB to 14 dB. It is due that HMS with UEP uses the G2+1+1 structure allocating the data priority bitstreams by layers, which become the transmission structure less robustness in diversity gain in comparison with the use of the two HMSs.

TABLE I
Subjective Performance SNR=11 *d*B

Video Sequence	Proposed algorithm	HMS G2+G2	HMS G2+1+1
Foreman	4.1	2.5	2.0
Carphone	3.2	2.0	2.1

TableI presents the average of 30 Human Visual System (HVS) quality assessment for each transmitted video sequence, which consists in presenting four different video sequences of the same video using a web page. The presented videos consist in the original video sequence, the video sequences transmitted by our proposed algorithm, HMS G2+G2 and HMS G2+1+1. The web page used for this subjective evaluation can be found in [20]. Therefore, is verified the effectiveness of the proposed algorithm for a SNR=11 dB.

(a) Proposed algorithm (b) HMS G2+G2

(c) HMS G2+1+1 (d) HMS - UEP

Fig. 4. Examples of subjective quality of the frame 194 from the video sequences "Carphone" transmitted with a SNR = 11dB.

In Fig. 4 we can observe that the use of the proposed algorithm outperform the HMS G2+G2, the HMS G2+1+1 and the HMS with UEP. In this example, the distortion and variable position of the picture in the frame 194 is due to the packet lost rate during the transmission and error concealment strategy configured in the video decoder.

VII. Conclusion

In this paper, we proposed a switching scheme between HMS based on a distortion model per GOP, which involves the use of application and the physical layers. The strategy is to switch between Hybrid MIMO Structure (HMS)s during the transmission time of a Group of Pictures (GOP) based on a distortion model per GOP to improve the quality of video streaming perceived by the end user. Simulation results show that the objective and subjective video quality is improved using this scheme. Therefore, the proposed algorithm is a feasible scheme that can be adopted for video streaming transmission over wireless systems.

Acknowledgment

The author Martin B. Obando is supported by The National Council for Scientific and Technological Development (CNPq).

References

[1] H. Sun., et al., "An Overview of Scalable Video Streaming" *Wireless Communications and Mobile Computing*, vol. 7, pp. 159-172, 2007.

[2] S. Chakraborty., et al., *IMS multimedia Telephony over Cellular Systems*, John Wiley & Sons, Ltd, 2007.

[3] K. M. El defrawy, et al., "Proposal for a Cross-Layer Coordination Framework for Next Generation Wireless Systems," *IWCMC*, Vancouver, British Columbia, Canada, July, 2006.

[4] G. Yang, et al., "Unequal Error Protection for MIMO Systems with Hybrid Structure," *IEEE Circuits and Systems*, pp. 682–685, September 2006.

[5] D. Song, et al., "Scalable H.264/AVC Video Transmission Over MIMO Wireless Systems With Adaptive Channel Selection Based on Partial Channel Information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1218-1226, September 2007.

[6] W. Liu, et al., "Transmission distortion estimation for real-time video delivery over Hybrid channels with bit errors and packet erasures" *ELSEVIER Computer Communications*, vol. 32, pp. 431-438, 2009.

[7] P. Ferré, et al., "Distortion-Based Link Adaptation for Wireless Video Transmission", *EURASIP Journal on Advances in Signal Processing*, vol. 2008, 17 pages, 2008.

[8] A. Paulraj, R. Nabar and D. Gore, *Introduction to Space-Time Wireless Communication*, Cambridge University Press, 2003.

[9] S. Chakraborty, T. Frankkila, J. Peisa and P. Synnergren, *IMS Multimedia Telephony over Cellular Systems*, John Wiley & Sons, Ltd, 2007.

[10] P. V. Pahalawatta, et al., "Review of Content-aware Resource Allocation Schemes for Video Streaming over Wireless Networks" *Wireless Communications and Mobile Computing*, vol. 7, pp. 131-142 , 2007.

[11] J. Korhonen., et al., "Loss-Distortion Estimation for Practical H.264AVC Steams" *IEEE International Conference on Multimedia and Expo*, Mexico, June 28 - July 3, 2009.

[12] J. Chakareski, et al., "Rate Distortion Hint Tracks for Adaptive Video Streaming" *IEEE Trans. on Systems for Video Tech.*, pp. 1257-1269, October 2005.

[13] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal, Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.

[14] W. C. Freitas, et al., "Hybrid Transceiver Schemes for Spatial Multiplexing and Diversity in MIMO Systems," *Journal of Communication and Information Systems*, vol.20, pp. 63–76, 2005.

[15] B. Vucetic and J. Yuan, "Space-time coding," West Sussex, England; Hoboken, NJ : Wiley, c2003. TK5102.92.V82, 2003.

[16] W. C. Freitas, et al., "Hybrid MIMO Transceiver Scheme with Antenna Allocation and Partial CSI at Transmitter Side," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, September 2006.

[17] D. Wu, et al., "Cross-Layer Design for Multimedia Delivery over Wireless Networks" *IEEE Compute Society, International Symposium on Multimedia*, 2006.

[18] "H.264/AVC" *JM Reference Software*, Download:"http://iphome.hhi.de/suehring/tml/download," 2010.

[19] YUV video sequences:"http://trace.eas.asu.edu/yuv/index.html", 2010.

[20] Subjective evaluation:"http://www.gtel.ufc.br/∼martin", 2010.