WAVELET-BASED DISTRIBUTED SOURCE CODING OF VIDEO

James E. Fowler[†], Marco Tagliasacchi[‡], and Béatrice Pesquet-Popescu[*]

[†]Dept. of Electrical & Computer Engineering
GeoResources Institute, Mississippi State ERC
Mississippi State University, MS USA

[‡]Dipartimento di Elettronica e Informazione
Politechnico di Milano, Italy

[*]ENST Paris
Paris, France

ABSTRACT

A wavelet-based video coder built on the principles of distributed source coding is described. The encoder employs a syndrome-based encoding strategy for intercoded coefficients while other coefficients are intracoded using an embedded wavelet-based coder designed for the coding of arbitrarily shaped image objects. The decoder uses a reference frame in the domain of a redundant wavelet transform to search for blocks matching the syndrome received from the encoder in order to decode intercoded coefficients. Experimental results indicate that, due to improved intraframe coding, the proposed wavelet-based algorithm significantly outperforms a similar technique constructed with JPEG-like intraframe coding.

1. INTRODUCTION

Whereas video coding has traditionally followed the so-called "broadcast" model in which a complex encoder codes video and distributes it to a large number of cheap and simple decoders, the "mobile-video" model often requires the inverse paradigm—cheap and simple encoders transmitting to possibly complex decoders. In essence, newly emerging mobile-video techniques combine distributed source-coding methods with traditional video-coding systems so as to shift the bulk of computation—the estimation of and compensation for frame-to-frame motion—to the decoder. Information-theoretic principles that describe source coding with side information known only to the decoder [1] are the theoretical underpinnings for these techniques that, in practice, rely on well-established error-correcting codes and related channel-coding implementations.

Current work in the area of mobile video has employed relatively simple image- and video-coding techniques akin to traditional JPEG and MPEG coding. However, wavelet-based methods are currently the preferred paradigm for still-image coding due to their outstanding coding performance. Moreover, recent wavelet-based video coders (e.g., [2]) exploit the overcompleteness of a redundant discrete wavelet transform (RDWT) in order to provide shift invariance, which, in turn, permits wavelet-domain motion estimation and compensation. Finally, recent algorithms for the wavelet-based coding of arbitrarily shaped image objects (e.g., [3]) can provide efficient, shape-adaptive (SA) coding of blocks in intraframe or interframe modes while ignoring blocks skipped due to their similarity to co-located blocks in the previous frame.

In this paper, we assemble these techniques into a system in which the JPEG-like intraframe-coding mechanisms employed in the Power-efficient, Robust, hIgh compression, Syndrome-based Multimedia coding (PRISM) [4, 5] algorithm are replaced with embedded wavelet-based coding. To this end, the discrete cosine transform (DCT) employed by the PRISM encoder is replaced with a discrete wavelet transform (DWT); JPEG-like runlength-Huffman coding is replaced with embedded bitplane coding; and a SA intraframe coder [3] is used to code around skipped blocks. In the resulting Wavelet-based Syndrome Coding (WSC) system, the encoder employs the syndrome-coding strategy pioneered in PRISM

[4, 5], while the motion search in the WSC decoder takes place in the wavelet domain using an RDWT of the reference frame. Experimental results indicate that, due to improved intraframe coding, the proposed wavelet-based WSC system significantly outperforms the DCT-based PRISM architecture.

2. PRISM

The PRISM framework is a a practical implementation of the paradigm of source coding with side information originally considered by Wyner and Ziv [1]. In the original Wyner-Ziv framework, the encoder compresses source X while the decoder has access to side information, source Y, which is unknown to the encoder, yet correlated with X. For the MSE distortion measure, and $X = Y + N$ where N is Gaussian, it is known that the rate-distortion performance of the coding of X is the same whether or not the encoder has access to Y. While these results are non-constructive and asymptotic in nature, a number of constructive methods to solve this problem have been proposed recently (e.g., [6–8]). In these approaches, a source codebook is partitioned into cosets of a channel code.

The PRISM framework is a variation on the source-coding-with-side-information problem in which there is uncertainty inherent in the side information. For PRISM-based video coding [4, 5], this uncertainty comes from the decoder using motion estimation to determine the side information. Specifically, the encoder divides the current frame into non-overlapping spatial blocks. The source $\mathbf{X}$ of the Wyner-Ziv framework is then the current block to be encoded, while the side-information $\mathbf{Y}$ is the best (motion-compensated) predictor for $\mathbf{X}$ in the previous frame, where it is assumed that $\mathbf{X} = \mathbf{Y} + \mathbf{N}$. The encoder quantizes $\mathbf{X}$ and then performs syndrome encoding on the resulting quantized codeword; i.e., the encoder finds a channel code that is matched to the "correlation noise" $\mathbf{N}$ and uses that channel code to partition the source codebook into cosets of the channel code. The encoder transmits the syndrome (indicating the coset for $\mathbf{X}$) as well as a CRC calculated on the quantization indices. In contrast to traditional, hybrid video coding, it is the task of the decoder perform the motion search, as it is the decoder that searches over the space of candidate predictors, one-by-one, seeking a block from the coset labeled by the syndrome. When the decoded block matches the CRC, decoding is declared to be successful. In essence, the decoder tries successive versions of side information $\mathbf{Y}$ until it finds one that permits successful decoding. Thus, the computational burden of motion estimation is shifted from the encoder to the decoder, so that the encoder is on the same order of complexity as frame-by-frame intraframe coding. Indeed, a large part of the PRISM encoder resembles the frame-by-frame application of traditional JPEG coding (i.e., "motion JPEG").

In this paper, we use the implementation of PRISM described in [9] which is built using the open-source QccPack library [10]. In this implementation, the PRISM encoder partitions the current frame into 8×8 blocks. Blocks with little difference from the co-located blocks in the previous frame are labeled as `skip` and not coded. The remaining blocks are coded in `intra` or `inter` modes. `intra`-mode blocks are essentially JPEG coded—an 8×8 DCT is applied, followed by runlength-Huffman coding. For the `inter` blocks, a combination of syndrome coding and intracoding is used—the lowest-frequency DCT coefficients of an `inter` block

are fed into the syndrome/CRC coding framework as described above while the other coefficients are coded with traditional, JPEG-like runlength-Huffman coding. For further detail on the PRISM approach and its implementation, see [4, 5, 9].

3. WAVELET-BASED SYNDROME CODING (WSC)

Modern embedded wavelet-based image coders are well-known to dramatically outperform traditional JPEG image coding in a rate-distortion sense. Consequently, the motivation of the proposed WSC system is to replace the JPEG-like mechanisms of the PRISM framework with techniques deriving from embedded wavelet-based coders. Specifically, we base WSC on the Binary Set Splitting with k-d Trees (BISK) [3] image coder which was designed specifically for the coding of image objects with arbitrary shape.

3.1 Encoder

The WSC encoder, depicted in Fig. 1(a), operates as follows. A 3-scale DWT is applied to the current frame, and the resulting DWT coefficients are partitioned into 8×8 cross-scale blocks as illustrated in Fig. 2. Each is block is then classified into one of three block types—skip blocks, intra blocks, and inter blocks. This classification is based on the mean squared error (MSE) between the current block and the co-located block in the previous frame. If the MSE is small, the block is a skip block. If the MSE is large, the block is classified as an intra block. Otherwise, the block type is set to inter. Preset thresholds determine this block-type classification, and the block type for each block is passed to the decoder in the bitstream.

For skip blocks, none of the wavelet coefficients of the block are coded. On the other hand, for intra blocks, all coefficients in the block are coded using embedded intraframe coding as described below. Finally, for inter blocks, the coefficients of the block are partitioned into inter coefficients and intra coefficients as illustrated in Fig. 2. The inter coefficients—the 16 low-frequency coefficients of the block—are quantized and syndrome coded as described below, while the remaining high-frequency intra coefficients are processed by the embedded intraframe coder in the same manner as the intra coefficients from intra blocks.

The syndrome coding of the inter coefficients in WSC is essentially the same as in the PRISM encoder of [9]. First, an inter block is classified into one of several inter-block classes. This classification is again based on the MSE between the inter block and the co-located block in the previous frame. The inter-block class, which is passed to the decoder in the bitstream, determines the step-size to be used for uniform scalar quantization of the inter coefficients. As in PRISM, these stepsizes are determined via an off-line training procedure designed to match stepsizes to the expected "correlation noise" for each class. Specifically, the stepsizes are chosen according to the "inverse-waterfilling" algorithm as applied to the correlation-noise coefficients as dictated by [11]. After quantization, codewords for the 16 inter coefficients are calculated as the quantization indices modulo 4. A 16-bit syndrome is then produced by convolving the mod-4 codewords with the parity-check matrix of a trellis code. Finally, the coefficients are reconstructed and subtracted from the original inter coefficients. The resulting residual coefficients are then passed to the embedded intraframe coder to be coded in the same manner as the intra coefficients.

The intraframe coder codes the intra coefficients from intra and inter blocks, as well as the residual coefficients from inter blocks, following the embedded bitplane-coding methodology common to wavelet-based image coders. However, not all coefficients of the frame are processed by the intraframe coder, as skip blocks are not coded. Consequently, we employ a SA intraframe coder designed to code only those coefficients within an arbitrarily shaped image "object" while ignoring regions not belonging to the object. In the context of the WSC coder, the image "object" to be coded is the set of intra and residual coefficients, while the skip coefficients constitute the non-object regions of the image. A mask constructed from the block types indicates the object and non-object regions to the SA intraframe coder. We use the BISK algorithm [3] which has been shown to yield consistently state-of-the-art SA coding performance as the intraframe coder in WSC.[1]

3.2 Decoder

The WSC encoder, depicted in Fig. 1(b), operates as follows. The bitstream is parsed, and a mask is generated from the extracted block types. This mask is fed into the intraframe BISK decoder which reconstructs the intra and residual coefficients for the intra and inter blocks.

For each inter block, a half-pixel motion search searches for a matching block within the RDWT of the previous frame. This search starts at the $(0, 0)$ motion vector and "spirals" outward with ever-increasing radius until a matching block is found or the entire search window is covered. Each candidate matching block is fed into a Viterbi decoder using soft decoding. The Viterbi decoder takes as input the 16-bit syndrome sequence from the bitstream as well as the 16 inter coefficients of the candidate block and outputs the 16 4-mod codewords of the closest codeword sequence with the desired syndrome. Then, the 16 mod-4 codewords from the Viterbi decoder are used in conjunction with the 16 inter coefficients from the candidate block to recover the 16 quantization indices for the current block. To do so, for each coefficient, the reconstructed value with the proper mod-4 label that is closest to the corresponding coefficient from the candidate block is found. Then the 16-bit CRC for the quantization indices is generated and compared to the CRC transmitted by the encoder. If the CRCs match, the motion search terminates, and the resulting match coefficients (reconstructed from the matching block) are passed to the DWT-reconstruction procedure. If the CRCs do not match, the motion search moves on to the next candidate block. In the event that no matching block is found in the search window, the coefficients are labeled as unmatched.

The pyramid of DWT coefficients for the current frame is reconstructed from the decoded coefficients before performing the inverse DWT. Specifically, for intra coefficients, the coefficients reconstructed by the BISK decoder are used. For skip and unmatched coefficients, the corresponding co-located coefficients from the previous, reconstructed frame are simply copied into the DWT of the current frame. Meanwhile, the match coefficients produced by the syndrome-decoding and motion-search procedure are added to the residual coefficients from the BISK decoder to produce inter coefficients. Finally, the inverse DWT is applied, and the reconstructed frame is produced.

4. EXPERIMENTAL RESULTS

We compare the performance of the proposed WSC system to that of the PRISM implementation of [9] which is constructed using QccPack [10]. We use one 16-frame GOP from grayscale SIF sequences (352×240, 30 Hz). Both systems use 8×8 blocks; the 128-state, rate-1/2 trellis code from [13]; and a 16-bit CRC. WSC uses 3-scale DWTs and RDWTs employing the popular 9/7 biorthogonal filters.

Figs. 3(a) and (b) give the rate-distortion performance for the "Football" and "Susie" sequences, respectively. In these figures, we compare WSC and PRISM as well as intraframe versions of WSC and PRISM: WSC-Intra and PRISM-Intra, respectively. In these intraframe coders, all blocks are restricted to be coded as intra blocks, thereby representing the straightforward and computationally lightweight intraframe coding of video.[2]

---

[1]We note that, in the usual methodology of SA image coding, one employs a SA DWT [12] to transform the image object into the wavelet domain before applying the embedded SA coder. In WSC, the mask and object coefficients are constructed directly in the wavelet domain, and so no SA DWT is employed.

[2]PRISM-Intra thus corresponds essentially to "motion JPEG;" WSC-Intra would therefore be "motion BISK."
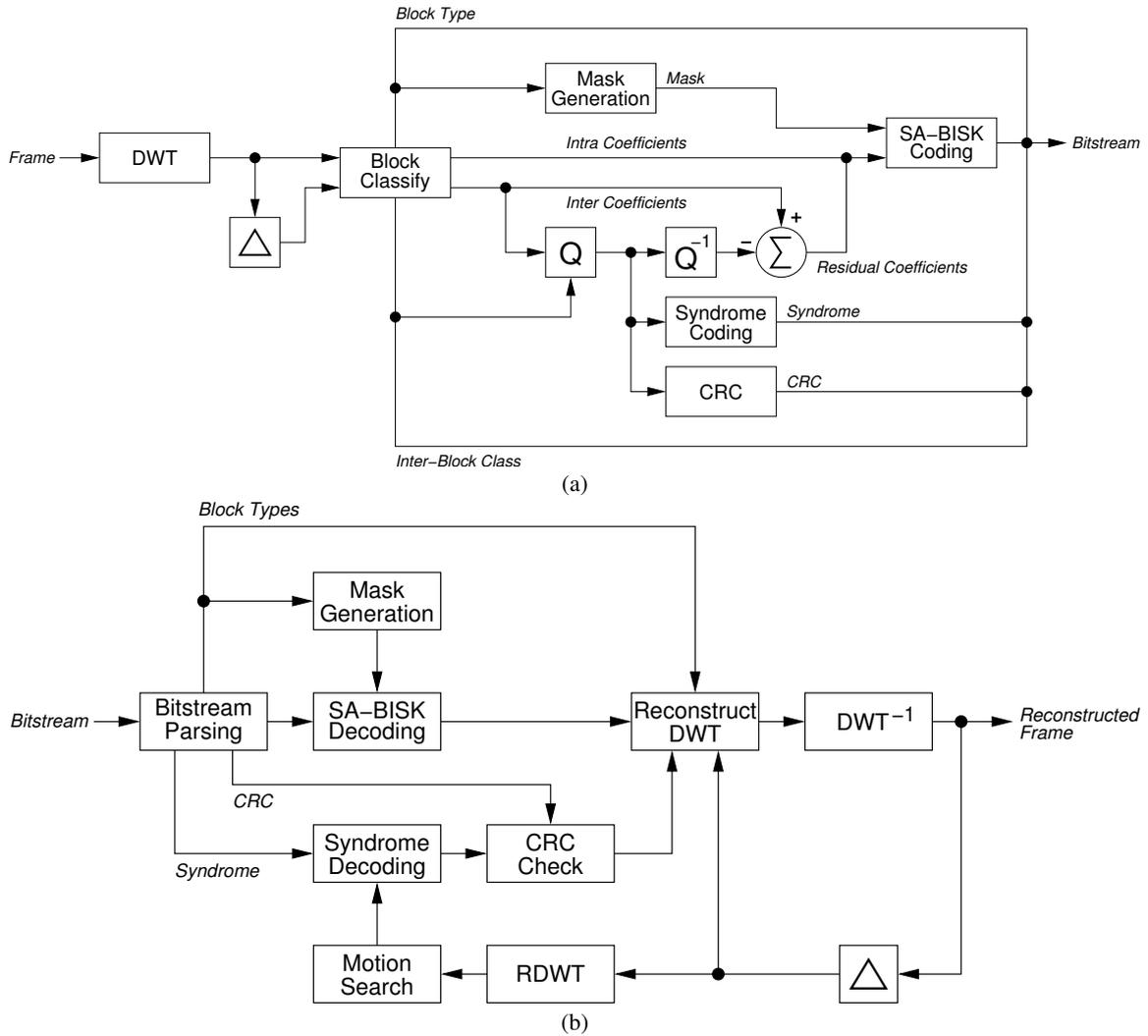
Figure 1: Block diagram of the WSC system, (a) encoder, (b) decoder. Δ indicates a frame delay; Q indicates uniform scalar quantization.
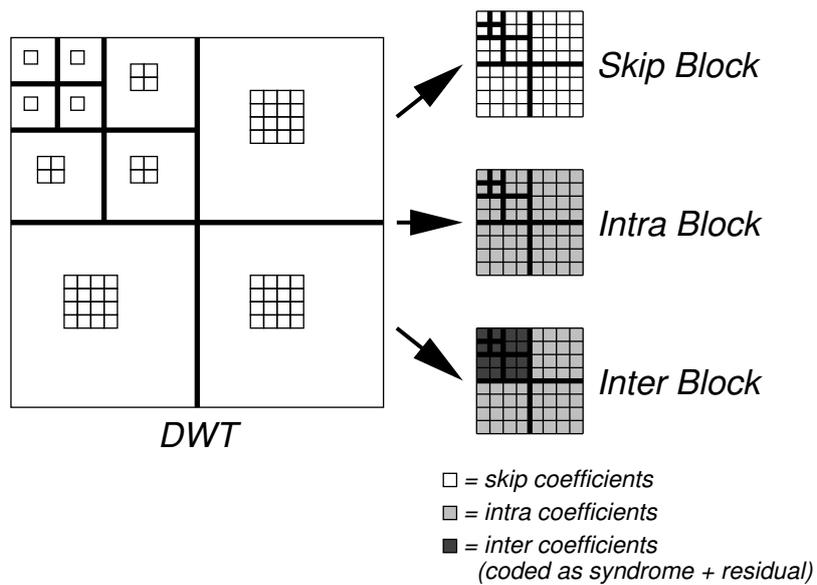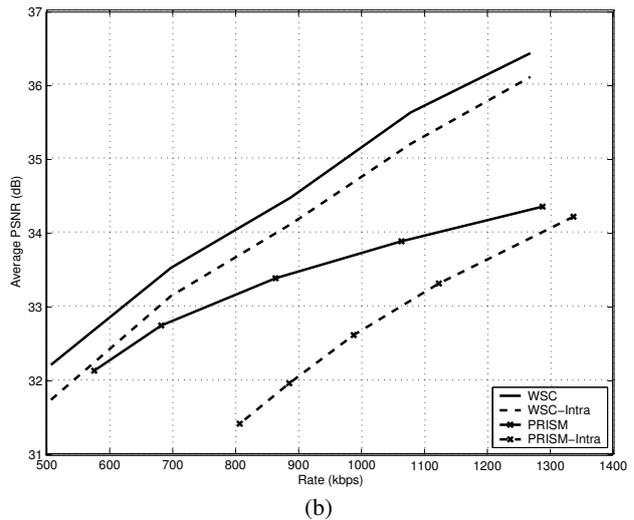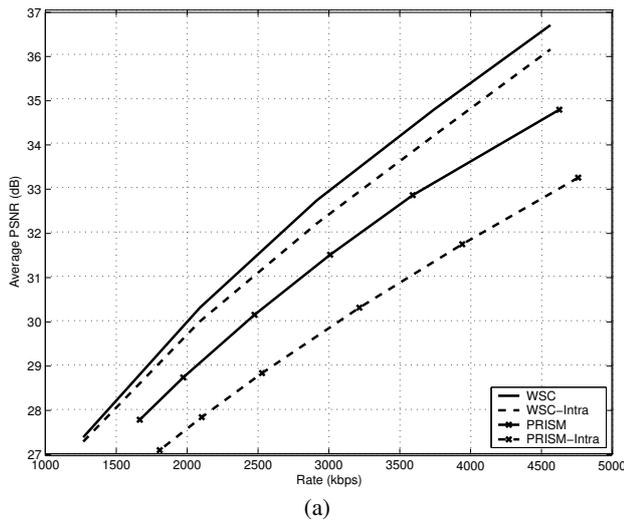

Figure 2: Extraction of blocks from the DWT.

Figure 3: Rate-distortion performance. (a) "Football," (b) "Susie."

Embedded bitplane coding of wavelet coefficients, such as embodied by BISK, is well-known to offer rate-distortion performance dramatically superior to the traditional, block-based-DCT coding of JPEG. Consequently, it comes as no surprise that the WSC-based coders, WSC and WSC-Intra, outperform their PRISM-based counterparts, PRISM and PRISM-Intra. For example, from Figs. 3(a) and (b), we see that WSC-Intra offers average PSNR some 1.5 to 2 dB greater than that of PRISM-Intra. Additionally, both WSC and PRISM improve upon their respective intraframe versions. However, due to the improved intraframe-coding efficiency provided by BISK with respect to the JPEG-like coding in PRISM, one would expect that the `inter` and `skip` modes are less effective in WSC than they are in PRISM. This is verified in Figs. 3(a) and (b), wherein we see that PRISM improves upon PRISM-Intra by 1 to 1.5 dB, while WSC achieves a gain of only 0.5 dB over WSC-Intra.

5. CONCLUSIONS

In this paper, we replaced the JPEG-like coding employed by PRISM for intraframe coding with techniques deriving from embedded wavelet-based image coders. The proposed WSC system uses a bitplane coding of DWT coefficients for intracoded coefficients while intercoded coefficients employ the syndrome-encoding strategy pioneered in PRISM. The SA BISK algorithm codes the intracoded and intercoded coefficients while ignoring skipped blocks. At the decoder, the redundancy of an RDWT representation of the reference frame permits a wavelet-domain motion search. The decoder searches for a wavelet-domain block matching the syndrome received from the encoder such that the computational burden of motion estimation is shifted from the encoder to the decoder. The WSC system thus combines the rate-distortion efficiency of embedded wavelet-based coding with the syndrome-encoding strategy of PRISM to yield a video-coding system with an efficient and lightweight encoder.

REFERENCES

[1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[2] H.-W. Park and H.-S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 577–587, Apr. 2000.

[3] J. E. Fowler, "Shape-adaptive coding using binary set splitting with *k*-d trees," in *Proc. ICIP*, Singapore, Oct. 2004, vol. 2, pp. 1301–1304.

[4] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Urbana-Champaign, IL, Oct. 2002.

[5] R. Puri and K. Ramchandran, "PRISM: A video coding architecture based on distributed compression principles," Tech. Rep. No. UCB/ERL M03/6, ERL, UC Berkeley, Mar. 2003.

[6] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Information Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[7] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Distributed compression of binary sources using conventional parallel and serial concatenated convolutional codes," in *Proc. DCC*, J. A. Storer and M. Cohn, Eds., Snowbird, UT, Mar. 2003, pp. 193–202.

[8] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. DCC*, J. A. Storer and M. Cohn, Eds., Snowbird, UT, Apr. 2002, pp. 252–261.

[9] J. E. Fowler, "An implementation of PRISM using QccPack," Tech. Rep. MSSU-COE-ERC-05-01, Mississippi State ERC, Mississippi State University, Jan. 2005.

[10] J. E. Fowler, "QccPack: An open-source software library for quantization, compression, and coding," in *Applications of Digital Image Processing XXIII*, A. G. Tescher, Ed., San Diego, CA, Aug. 2000, Proc. SPIE 4115, pp. 294–301.

[11] S. S. Pradhan, "On rate-distortion function of Gaussian sources with memory in the presence of side information at the decoder," Tech. Rep., Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Dec. 1998.

[12] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrary shaped visual object coding," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 10, no. 5, pp. 725–743, Aug. 2000.

[13] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Information Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.