Analytical Rate-Distortion-Complexity Modeling of Wavelet-based Scalable Video Coders

Brian Foo*, Yiannis Andreopoulos, and Mihaela van der Schaar

Abstract

Analytical modeling of the performance of video coders is essential in a variety of applications, such as power constrained processing, complexity-driven video streaming, etc., where information concerning rate, distortion or complexity (and their interrelation) is required. In this work, we present a novel rate-distortion-complexity (R-D-C) analysis for state-of-the-art wavelet video coding methods by explicitly modeling several aspects found in operational coders, i.e. embedded quantization, quadtree decompositions of block significance maps and context-adaptive entropy coding of subband blocks. This work achieves two main goals. Firstly, unlike existing R-D models for wavelet video coders, the proposed derivations reveal for the first time the expected coding behavior of specific coding algorithms (e.g. quadtree decompositions, coefficient significance and refinement coding) and therefore can be used for a variety of coding mechanisms incorporating some or all the coding algorithms discussed. Secondly, the proposed modeling derives for the first time analytical estimates of the expected number of operations (complexity) of a broad class of wavelet video coding algorithms based on stochastic source models, the coding algorithm characteristics and the system parameters. This enables the formulation of an analytical model characterizing the complexity of various video decoding operations. As a result, this work complements prior complexity-prediction research that is based on operational measurements. The accuracy of the proposed analytical R-D-C expressions is justified against experimental data obtained with a state-of-the-art motion-compensated temporal filtering based wavelet video coder, and several new insights are revealed on the different tradeoffs between rate-distortion performance and the required decoding complexity.

Keywords: *rate-distortion, complexity, wavelet-based video coding, analytical modeling*

EDICS: HDW-AAOP, DSP-ALGO, MLT-APPL

I. Introduction

Energy consumption is an important issue in mobile devices. In the case of multimedia, the battery life of such devices has been shown to be directly linked to the complexity of coding algorithms [3] [4] [8]. For this reason, recent advances in scalable coding algorithms that provide schemes enabling a variety of rate-distortion-complexity (R-D-C) tradeoffs with state-of-the-art performance [2] are very appealing frameworks for such resource constrained systems. This flexibility in video encoding and decoding is also

* Corresponding author. The authors are with the Dept. of Electrical Engineering (EE), University of California Los Angeles (UCLA), 66-147E Engineering IV Building, 420 Westwood Plaza, Los Angeles, CA, 90095-1594, USA, Phone: +1-310-825-5843, Fax: +1-310-206-4685, E-mail: { bkungfoo, yandreop, mihaela } @ee.ucla.edu

very suitable for the increasing diversity of multimedia implementation platforms based on embedded systems or processors that can provide significant tradeoffs between video coding quality and energy consumption [3] [4]. In order to select the optimal operational point for a multimedia application in a particular system, accurate modeling of the source, algorithm *and system (implementation) characteristics* is required. Such modeling approaches are important because they can also serve as the driving mechanism behind the design of future complexity-scalable coders.

Two methods have been used to determine the rate-distortion and the complexity characteristics of operational video coders. The first is an empirical approach, where analytical formulations are fitted to experimental data to derive an operational model suitable for a particular class of video sequences and a particular instantiation of a compression algorithm in a fixed implementation architecture; see [17] [18] for such examples of R-D models and [3]-[7] for such examples of complexity modeling. While this modeling approach is simple, the obtained rate-distortion-complexity (R-D-C) expressions cannot be generalized because their dependency on the sequence, algorithm and system parameters is not explicitly expressed via the model. As a result, while current state-of-the-art multimedia compression algorithms and standards provide profiles for rate control [1] [4] [6], they lack analytical methods to determine the complexity tradeoffs between different coding operations that can be exploited for different systems.

The second approach is a theoretical approach, where stochastic models are used for pixels or transform coefficients. Using this approach, analytical expressions can be derived for the R-D-C behavior of a particular system or class of systems processing a broad category of input sources in function of the sources' statistics; see [9] [12] [13] for such examples of R-D models and [10] [11] [14] for complexity modeling using operational source statistics and off-line or on-line training to estimate (learn) the algorithm and system parameters. We remark that, although there is a significant volume of work in modeling of transform-domain statistics [26] [27] [28] and also in the efficiency analysis of coding mechanisms [17], there is significantly less literature on rate-distortion modeling for state-of-the-art operational video coders, and (to the best of the authors' knowledge) scarcely any work exists on *complexity modeling* for such systems in function of *stochastic source models* and *algorithm characteristics*. We emphasize that, while the derived theoretical expressions of such approaches are typically more complex than the expressions derived from the first category, the dependencies on the source and system modeling parameters are explicitly indicated via the derived analytical framework. This is of great importance to several cross-layer or resource optimization problems [3] [8] [10] that need to judicially balance the network or system resources in order to accommodate the viewer preferences in the most efficient manner. In addition, the explicit dependency of the derived R-D-C estimation on source, algorithm and system parameters facilitates the application of the derived framework for a variety of input video source classes. Moreover, various algorithms and systems of interest are accommodated in this way. Finally, a rigorous, analytical R-D-C formulation methodology can be easily extended to model properties of various other coding algorithms based on input source and algorithm parameters. In this way, analytical comparisons of the R-D-C efficiency for particular algorithms can accompany experimental testing in order to facilitate system design decisions and options.

For these reasons, we follow the second category of approaches and provide a unified R-D-C modeling framework for motion-compensated temporal filtering (MCTF) based wavelet video coders [2] [22] [25] [29]. Two aspects are typically required for unified R-D-C modeling of video coding: *i*) modeling of the temporal prediction process [15] [16]; *ii*) modeling of the quantization and coding process [9] [12]. Since the motion-compensation complexity of the MCTF process has been studied in detail in prior work [7] [11], we focus on the second part and assume that the transform-domain statistics of the intra and error frames produced by the temporal decomposition are available. Unlike the existing theoretical work [9] [12] in this area, the proposed R-D-C model is based on a thorough analysis of *different coding operations* (quadtree coding, coefficient significance and refinement coding) and as a result can encompass many state-of-the-art wavelet video coders found in the literature.

Consequently, this work extends prior R-D modeling of block-based wavelet video coders to a broader class of coding mechanisms. Perhaps more importantly, this work proposes an analytical derivation of complexity estimates for the entropy decoding and the inverse spatial transform, thereby complementing our prior work on complexity estimation [7] [10] [11].

Based on the derived theoretical results and their experimental validation, we explore the R-D-C space of achievable operational points for state-of-the-art wavelet video coders and derive several interesting properties for the interrelation of rate-distortion performance and the associated decoding and inverse spatial transform complexity.

The paper is organized as follows. Section II introduces the types of quantization and coding schemes analyzed in this paper. Some important nomenclature is also provided. Section III presents the utilized wavelet coefficient models and derived probability estimates for a variety of coding/decoding operations. These probabililities will be used to determine the average rate, distortion and complexity (Sections IV-VI, respectively) for decoding a video sequence. Section VII displays theoretical and experimental R-D-C results that validate the proposed models and discusses several interesting R-D-C properties of operational video coders. Section VIII concludes the paper.

II. Overview of Wavelet Video Coders

In this section, we introduce a basic overview of state-of-the-art wavelet coding schemes analyzed in this paper. They involve temporal decomposition via MCTF, spatial discrete wavelet transform (DWT) decomposition, embedded quantization, and the entropy coding process.

A. Temporal Decomposition

Recent state-of-the-art scalable video coding schemes are based on motion compensated temporal filtering [2]. During MCTF, the original video frames are filtered temporally in the direction of motion [2] [15], prior to performing the spatial transformation and coding. Video frames are filtered into L (low-frequency or average) and H (high-frequency or error) frames [2]. The process is applied initially in a group of pictures (GOP) and also to all the subsequently-produced L frames thereby forming a total of T_{MCTF} temporal levels. After the temporal decomposition, the derived L and H temporal frames are spatially decomposed in a hierarchy of *spatio-temporal subbands*. Quantization and entropy coding are applied to these subbands to form the final compressed bitstream.

B. Embedded Quantization

An important category of quantizers used in image and video coding is the family of embedded double-deadzone scalar quantizers [19]. For this family, each input wavelet coefficient x is quantized to:

$$Q_b(x) = \left\{ \mathrm{sign}(x) \cdot \left\lfloor \tfrac{|x|}{2^b \Delta} \right\rfloor, \text{ if } \frac{|x|}{2^b \Delta} \geq 1;\ 0, \text{otherwise} \right\}, \tag{1}$$

where $\lfloor a \rfloor$ denotes the integer part of a; $\Delta > 0$ is the basic quantization step size (basic partition interval size) of the quantizer family; $b \in \mathbb{Z}_+$ indicates the quantizer level (granularity), with higher values of b indicating coarser quantizers. In general, b is upper bounded by a value $B_{\max}$, selected to cover the dynamic range of the input signal. The signal reconstruction is performed by:

$$Q_b^{-1}(Q_b(x)) = \left\{ \mathrm{sign}(Q_b(x)) \cdot (|Q_b(x)| + \tfrac{1}{2}) 2^b \Delta,\ \text{if } Q_b(x) \neq 0;\ 0 \text{ if } Q_b(x) = 0 \right\} \tag{2}$$

where the reconstructed value $Q_b^{-1}(Q_b(x))$ is placed in the middle of the corresponding uncertainty interval (partition cell), and $Q_b(x)$ is the partition cell index, which is bounded by a predefined value for each quantizer level. For example, $0 \leq Q_b(x) \leq M_b - 1$, for each b, with $M_{B_{\max}} = \ldots = M_0 = 2$ and $\Delta = 1$ for the popular case of successive approximation quantization (SAQ) [19]. If the b least-significant bits of $Q_0(x)$ are not available, one can still dequantize at a lower level of quality using the inverse quantization formula given in (2). SAQ can be implemented via thresholding, by applying a monotonically decreasing set of thresholds of the form $T_{b-1} = T_b / 2$, with $B_{\max} \geq b \geq 1$ and

$T_{B_{\max}} = \alpha_{\text{quant}} \cdot x_{\max}$, where $x_{\max}$ is the highest coefficient magnitude in the input wavelet decomposition, and α_{quant} is a constant that is taken as $\alpha_{\text{quant}} > 1/2$. By using SAQ, the significance of the wavelet coefficients with respect to any threshold T_b is indicated in a corresponding binary map, called the *significance map*. Coding of $B_{\max},\ldots,B_{\min}$ significance maps corresponds to coding the $B_{\max} - B_{\min} + 1$ most significant bitplanes of each wavelet coefficient x.

C. Coding of the Significance Maps and Coefficients

In all state-of-the-art wavelet coders [19]-[25], the coding process exploits intra-band dependencies following a block-partitioning process within each transform subband. This coding process is performed for every bitplane b. As indicated in Figure 1, several coding passes that identify coefficient significance ("Significance Pass") or refine wavelet coefficients ("Refinement Pass") with respect to the current SAQ threshold are performed either within quadtree coding [22] [23] or within block coding [19]. Several state-of-the-art embedded image coders invoke both approaches, i.e. the quadtree coding partitions the input subbands until a minimum block size, which is then coded with the block coding module [20] [21].

We analyze such intra-band coders that use quadtrees to decompose subbands into non-overlapping blocks of dyadically-decreasing sizes followed by block coding for the blocks of the maximally decomposed quadtree [20] [21]. In particular, the initial subbands are hierarchically split in K quadtree levels using several coding passes, with blocks at quadtree level K having the smallest size. The significance information (i.e. whether the block contains significant coefficients) is encoded using depth-first-search along the quadtree, where the significance of a block at quadtree level k is encoded only if its parent block at quadtree level $k-1$ is found to be significant. For the blocks found significant at the bottom of the quadtree (level K), the block coding is invoked. Block coding performs raster scan to obtain the significance of each coefficient. The coefficients found significant are then placed in a refinement list to be refined at the next finer quantization level.

The produced symbols from each coding pass, from block significance information to coefficient significance, refinement, and sign information, are then encoded using context-based adaptive arithmetic coding [33] [34]. This technique exploits the dependencies between the symbols to be encoded and the neighboring symbols (the context) [33]. Context conditioning reduces the entropy and improves the coding performance. An example of context-based entropy coding is to use several arithmetic coder models with different initial probabilities to encode coefficients based on the significance of their neighbors, since a coefficient with significant neighboring coefficients has a larger probability to be significant than coefficients with insignificant neighboring coefficients. Using these separate arithmetic coder models for different "contexts," context-based coding schemes achieve better performance than simply compressing all symbols using a single arithmetic coder [33] [34].

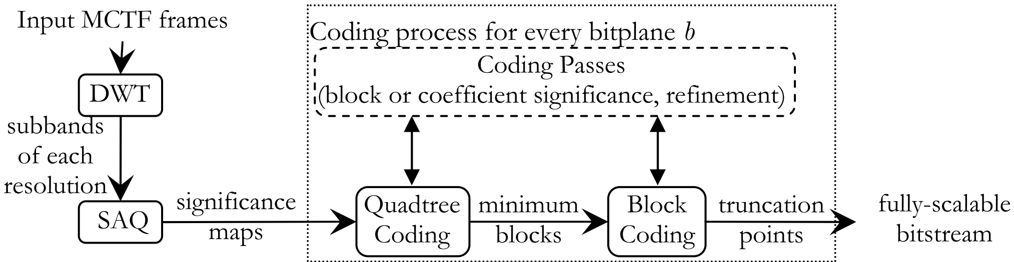

Figure 1 Block diagram of intra-band coding process of state-of-the-art wavelet-based coders encompassing quadtree and block coding of the significance maps.

III. Approximation of Block Significance Probabilities in Quadtree Decompositions of the Significance Maps

In this section, we introduce the utilized stochastic source model for wavelet coefficients. We then derive probabilities of significance for quadtree decompositions over quantized spatio-temporal subbands. These probabilities form the core of the rate and complexity estimation derived in the remaining sections of this paper as they provide the means of establishing the percentage of blocks that are expected to be coded or decoded at a given distortion bound, expressed by the terminating SAQ threshold $T_{B_{\min}}$. In addition, the percentage of significant areas within the spatio-temporal subbands along with the percentage of non-zero coefficients are the two features (or "decomposition functions" [11]) that express the complexity of the inverse DWT.

A. Source Models for Low and High-frequency Wavelet Coefficients

The R-D characteristics of low-frequency wavelet coefficients are typically modeled using the high-rate uniform quantization assumption [9] [19] for independent zero-mean Gaussian random variables. This model will be accurate if the low-frequency coefficients exhibit sufficiently low correlation. We investigate this in Table 1, which displays the ratio of the average correlation between neighboring coefficients to the average coefficient variance. In Figure 2 we validated that the Gaussian distribution for low-frequency spatio-temporal subband coefficients was accurate.

While low-frequency spatio-temporal subbands account for a large percentage of the video coding rate, the high-frequency spatio-temporal subbands also contribute a significant amount to the overall coding rate and complexity [13]. Thus, accurate modeling of the high-frequency spatio-temporal subband statistics is also very important for precise R-D-C modeling of wavelet video coders. When applied to image or residual frame data, typical wavelet filters tend to produce decorrelated coefficients in the high-frequency subbands. However, dependencies remain among coefficients within the same scale and across different scales [26]. Certain highly-popular wavelet filter-banks, such as the Daubechies 9/7 filter-pair, have further properties that can reduce most of the interscale dependencies, leaving only dependencies among neighboring coefficients within the same subband [26].

	Foreman	*Coastguard*	*Silent*	*Mobile*
Autocorrelation Coefficient	0.5340	0.6733	0.5100	0.3763

Table 1 Ratio of correlation between neighboring coefficients to the average coefficient variance for the LL subband of L-frames of *Foreman*, *Coastguard*, *Silent*, and *Mobile* after a 2 temporal level-4 spatial level decomposition.

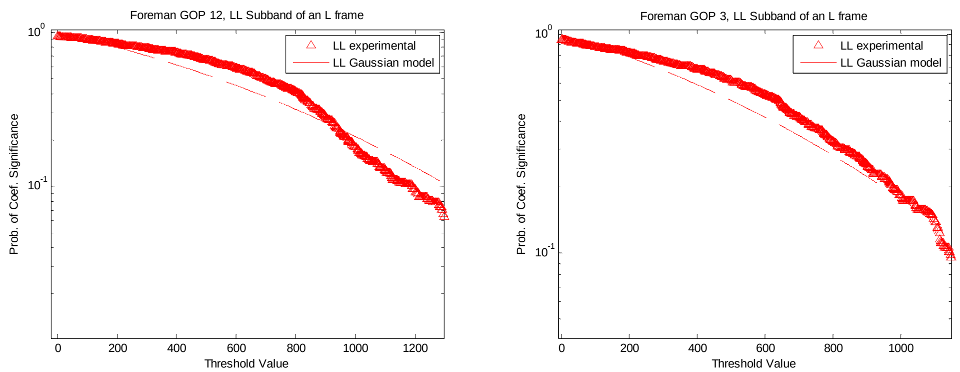

Figure 2. Experimental examples that the Gaussian assumption for the low-frequency wavelet coefficients of the MCTF-based decomposition (with four spatio-temporal levels) is accurate.

In order to capture the experimentally-observed heavy-tailed non-Gaussian distribution of wavelet coefficients within each subband, in this paper high-frequency wavelet coefficients are modeled as a doubly-stochastic process, i.e. a Gaussian distribution parameterized by Θ, which is exponentially distributed with parameter σ^2:

$$\Theta \sim p(\theta) = \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \tag{3}$$

In this case, each high-frequency wavelet coefficient x can be modeled by a random variable X with marginally Laplacian distribution and variance σ^2 [36]:

$$X \sim p(x) = \int_0^\infty p(x \mid \theta) p(\theta) d\theta = \int_0^\infty \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}x^2} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} d\theta = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x|} \tag{4}$$

where $p(x)$ indicates the probability density function (PDF). Figure 3 demonstrates the accuracy of the doubly-stochastic model of (4) for different spatio-temporal high-frequency subbands. In addition, Table 2 presents the change in the subband statistics for different spatio-temporal levels across the MCTF decomposition and the corresponding rate for terminating the coding of the wavelet coefficients of each spatio-temporal level at several bitplanes. The coder of [29] was used for the examples of this section. While we focused on 4x4 blocks in Table 2, similar results can be shown for larger block sizes (e.g. 8x8), since, for natural images or error frames, wavelet coefficients within such small areas of high-frequency subbands are modeled accurately with the same local parameter θ .

Based on the results of Table 2 we conclude that there is significant variation in the rate associated with each spatio-temporal level, ranging from 0 bpp to almost 1 bpp for low-rate coding ($B_{\min} = 7$ in Table 2) and from 0.15 bpp to almost 5 bpp for medium and high rate coding ($B_{\min} = 3$ in Table 2). Furthermore, the higher (coarser) spatio-temporal high-frequency subbands exhibit significant variance and the correlation of the subband statistics (parameter θ) varies significantly as well. Consequently, there is a significant portion of the coding rate attributed to them for a variety of quantization thresholds; thus, accurate modeling of the rate-distortion-complexity characteristics of high-frequency spatio-temporal subbands is important for predicting the overall R-D-C behavior. Finally, although the results of Table 2 reveal certain trends between the spatio-temporal subband rate and the model parameters (σ^2 and θ), the overall rate contribution of each subband depends not only on the statistics of the input but also on the details of the invoked coding algorithm. Hence, a detailed theoretical analysis of the coding operations as a function of the source model is of paramount importance for precise R-D-C estimations, and intuitive models based on the source statistics and experimental observations do not suffice.

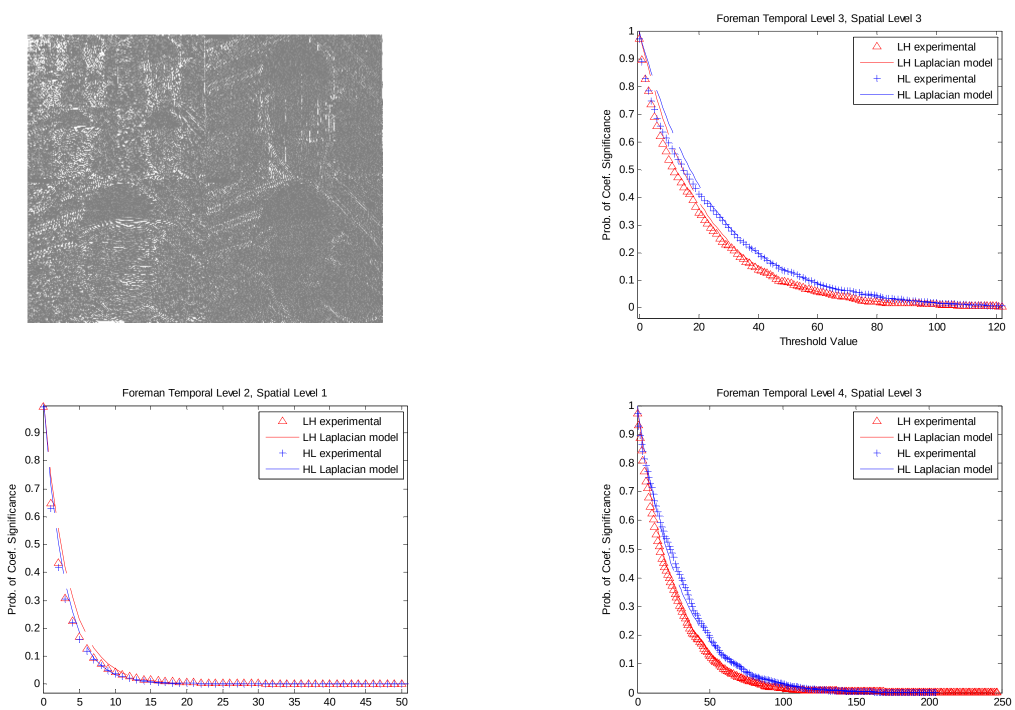

Figure 3. The discrete wavelet transform of an H frame of temporal level two (top left) and plots of the doubly-stochastic (Laplacian) model and simulation data for several H frames of different spatio-temporal resolutions.

Temporal(T)-Spatial(S)	Subband variance σ^2, [variance of θ]			Rate (bpp) for various decoded bitplanes $B_{\min}$				
	LH	HL	HH, LL (if exists)	7	6	5	4	3

level								
1T-1S	3.18, [12.1]	1.87, [9.1]	1.61. [7.8]	0	0	0.0021	0.0261	0.1489
2T-2S	6.59, [37.5]	5.55, [22.8]	4.46, [25.7]	0.0017	0.0219	0.1216	0.3910	1.0257
3T-3S	18.2, [60.3]	14.8, [70.1]	11.7, [63.1]	0.1178	0.4209	0.9630	1.7071	2.7121
4T-4S	39.7, [241]	33.2, [496]	{26.0,[144]}, {53.1, 690]}	0.9697	1.7323	2.6717	3.7929	4.9343

Table 2. Examples of subband variances as well as the variance of the correlation θ (for block sizes of 4×4) formed across the spatio-temporal MCTF subbands of sequence Foreman, along with the corresponding bitrates for several values of $B_{\min}$.

We denote the minimum decoded bitplane threshold level as $T_{B_{\min}} = 2^{B_{\min}}$. In addition, we define the following parameters for all bitplanes b:

$$\upsilon_b = \frac{T_b}{\sigma} \tag{5}$$

$$\rho_b = e^{-\sqrt{2}\upsilon_b} \tag{6}$$

where υ_b describes the ratio of the threshold of bitplane b to the variance of each wavelet coefficient, and ρ_b is the probability of significance of a wavelet coefficient under a certain υ_b under the model of (4).

B. Probability of Block Significance at Bitplane b

We begin this section by introducing some notation. We define the significance test of a block of n coefficients with respect to a threshold T_b as $\mathrm{sig}(T_b, n) \in \{0,1\}$, where $\mathrm{sig}(T_b, n) = 1$ if at least one coefficient within the block is found significant with respect to the threshold T_b, and $\mathrm{sig}(T_b, n) = 0$ otherwise. We also define the newly significance test as $\mathrm{newsig}(T_b, n) = \{0,1\}$, which returns one if the block was found to be significant at bitplane b and insignificant at bitplane $b+1$, i.e. $\mathrm{sig}(T_b, n) = 1$ and $\mathrm{sig}(T_{b+1}, n) = 0$. For notational abbreviation, the *probability* of a block being significant or newly-significant at bitplane b is indicated by $\chi_{\upsilon_b,n}^{\mathrm{band}}$ and $\delta_{\upsilon_b,n}^{\mathrm{band}}$, respectively, with $\mathrm{band} = \{\mathrm{low,high}\}$ indicating the frequency subband that the block belongs to. Note that these metrics depend on υ_b, which is a function of the bitplane b as well as the variance of subband coefficients, σ^2. Let us first consider a high-frequency spatio-temporal subband, which may be any subband of an error (H) frame, or any high-frequency subband of an L frame. The probability that a block of n wavelet coefficients is found significant during (or before) the significance pass at bitplane b, $\chi_{\upsilon_b,n}^{\mathrm{high}}$, is no more than the probability that at least one coefficient in the block is found significant when compared to $T_b = 2^b$. Thus, we have:

$$\chi_{\upsilon_b,n}^{\mathrm{high}} = \Pr\{\mathrm{sig}(T_b, n) = 1\} = 1 - \Pr\{|\,\mathbf{X}\,|_\infty \le T_b\} \tag{7}$$

where $\mathbf{X} = (X_1, ..., X_n)$ is a length-n random vector of variables X_i ($1 \le i \le n$) for coefficients in the same block, and $|\bullet|_\infty$ is the $L^{\inf}$ norm. Considering that block sizes are generally small enough to capture local variances, we follow the doubly stochastic model in equation (3). Given Θ, the conditional joint distribution of $\mathbf{X}$ is then a uncorrelated Gaussian random vector. Given that the variance of the subband coefficients is σ^2, the probability density function of $\mathbf{X}$ is:

$$p(\mathbf{x}) = \int_{0+}^{\infty} p(\theta)p(\mathbf{x} \mid \theta)d\theta = \int_{0+}^{\infty} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \cdot \frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}(x_1^2+x_2^2+\ldots+x_n^2)} d\theta \tag{8}$$

where $\mathbf{x} = (x_1, x_2, ..., x_n)$ is a vector of coefficient values, and $p(\mathbf{x})$ is the n-dimensional PDF of $\mathbf{X}$.

Proposition 1: The probability that a block of size n is significant compared to threshold T_b can be approximated by:

$$\chi_{\upsilon_b,n}^{\text{high}} \cong \exp\left\{\frac{\upsilon_b^2}{\ln(n)^{1.296} + 0.166}\right\} \tag{9}$$

*Proof***:** See Appendix A. ■

For low-frequency subbands, assuming sufficiently-decorrelated Gaussian distributed coefficients, the probability of block significance is simply the n-dimensional Gaussian tail probability along one of the orthogonal axes:

$$\chi_{\upsilon_b,n}^{\text{low}} = \left[\text{erfc}\left(\frac{\upsilon_b}{\sqrt{2}}\right)\right]^n \tag{10}$$

where $\text{erfc}\left(\frac{\upsilon_b}{\sqrt{2}}\right) = 1 - \text{erf}\left(\frac{\upsilon_b}{\sqrt{2}}\right)$, and $\text{erf}\left(\frac{\upsilon_b}{\sqrt{2}}\right)$ can be piecewise approximated by (11):

$$\text{erf}\left(\frac{\upsilon_b}{\sqrt{2}}\right) \approx \begin{cases} 0.2\upsilon_b(4.4 - \upsilon_b) & ,0 \le \upsilon_b \le 2.2 \\ .98 & ,2.2 < \upsilon_b < 2.6 \\ 1 & ,\upsilon_b \ge 2.6 \end{cases} \tag{11}$$

in order to avoid numerical integrations during the model calculation.

C. The Probability a Block is found Newly Significant at Bitplane b

In order to model the number of operations performed during the significance pass at each bitplane, it is necessary to derive the probability that a block is found significant at bitplane b, but not at any higher bitplanes. This is due to the fact that in all coding algorithms using quadtrees of wavelet coefficients, once a block is found significant at bitplane b, it is moved into the refinement list and its significance is not encoded at the subsequent bitplanes $b-1, \ldots, B_{\min}$.

Proposition 2: The probability that a block of n coefficients in a high-frequency subband is found significant at bitplane b, but it is insignificant at bitplane $b+1, \ldots, B_{\max}$, is:

$$\delta_{\upsilon_b,n}^{\text{high}} = \Pr\{\text{newsig}(T_b, n) = 1\} \cong \chi_{\upsilon_b,n}^{\text{high}}(1 - \chi_{\upsilon_b,n}^{\text{high}}) \tag{12}$$

*Proof***:** See Appendix B. ■

We note that our proof only specifies the existence of a large enough n for the approximation above, but it does not give the exact lower bound for n. To verify the accuracy of this estimate for typical blocks during the quadtree significance passes, we did a qualitative comparison of plotted curves for

$\mathrm{erf}(x)/\mathrm{erf}(2x)$ raised to various powers. For the minimum block size $n = 16$ used in practical coders [20] [24] [25] the match is approximately equal (Figure 4). The fit only improves for a larger n.

For low-frequency subbands, we will assume decorrelated coefficients. Our result is given below.

Proposition 3: The probability that a block of n coefficients in a low-frequency subband is found significant at bitplane b, but it is insignificant at bitplane $b+1,\ldots,B_{\max}$ is:

$$\begin{aligned}\delta_{\upsilon_b,n}^{\mathrm{low}} &= \left[\mathrm{erfc}\left(\frac{\upsilon_b}{\sqrt{2}}\right)\right]^n - \left[\mathrm{erfc}\left(\sqrt{2}\upsilon_b\right)\right]^n = \left[\mathrm{erf}\left(\sqrt{2}\upsilon_b\right)\right]^n - \left[\mathrm{erf}\left(\frac{\upsilon_b}{\sqrt{2}}\right)\right]^n \\ &\cong \left[\mathrm{erf}\left(\sqrt{2}\upsilon_b\right)\right]^n \left[1 - \left[\mathrm{erfc}\left(\frac{\upsilon_b}{\sqrt{2}}\right)\right]^n\right] = (1-\chi_{\upsilon_{b+1},n}^{\mathrm{low}})\chi_{\upsilon_b,n}^{\mathrm{low}}\end{aligned} \tag{13}$$

Proof: The approximation of (13) is a straightforward result of Lemma 3 in Appendix B. ■

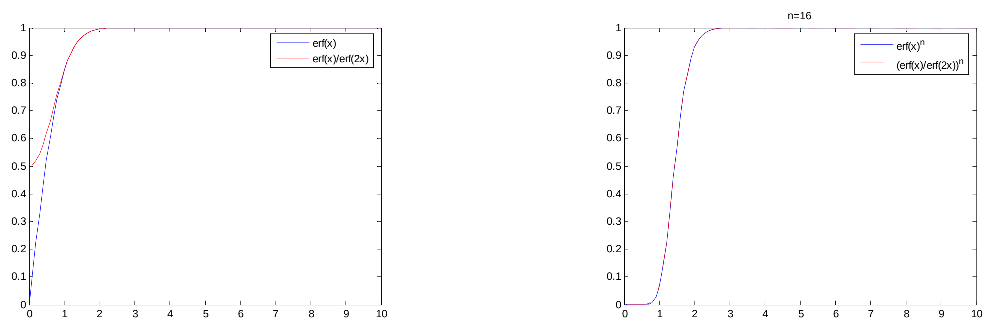

Figure 4. Plot of: $\mathrm{erf}(x)$ vs. $\mathrm{erf}(x)/\mathrm{erf}(2x)$ (left), and $[\mathrm{erf}(x)]^{16}$ vs. $[\mathrm{erf}(x)/\mathrm{erf}(2x)]^{16}$ (right).

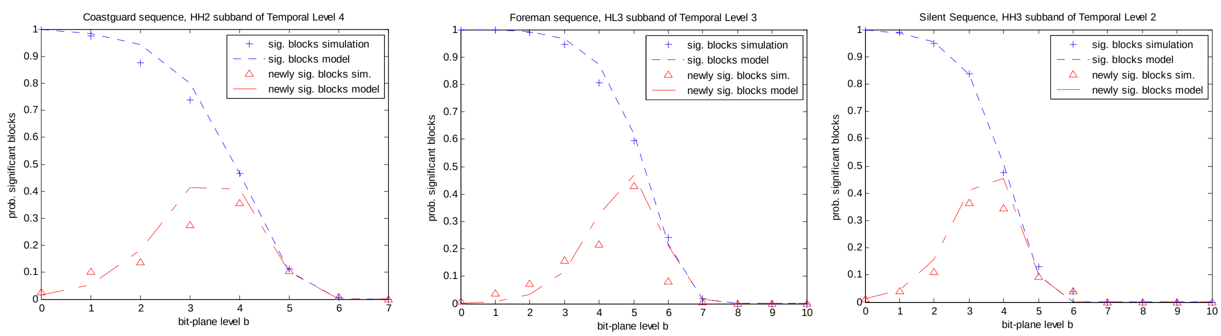

Figure 5. Simulation and model prediction of significance and newly-significance of 4×4 blocks in various high-frequency spatio-temporal subbands.

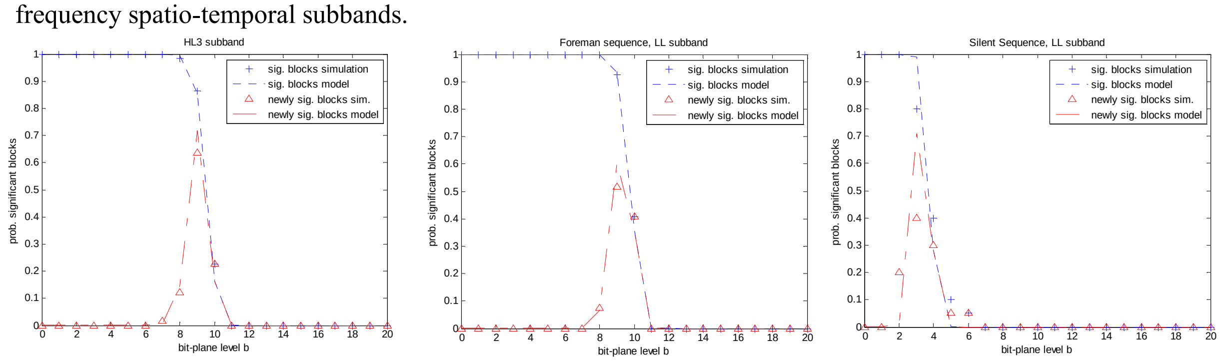

Figure 6. Simulation and model prediction of significance and newly-significance of 4×4 blocks in LL subbands of L frames.

To verify the accuracy of the proposed model of (9) and (12), Figure 5 demonstrates the model prediction of the probability of block significance and newly-significance (with $n = 16$) for several high-frequency subbands belonging to various temporal levels in MCTF-decomposed frames of video sequences. Similarly, we plot several examples that validate the proposed model of (10) and (13) (low-frequency spatio-temporal subbands) in Figure 6. The experimentally-derived significance and newly-significance for low and high-frequency spatio-temporal subbands are in agreement with the theoretical derivations for a large variety of cases, as shown by these experiments. In addition, the model validation reveals several interesting properties. Firstly, the approximation (13) for the significance probability of blocks in the low-frequency spatio-temporal subbands suggests that most of the blocks within the subband will be found newly-significant at the same bitplane, or at most at two consecutive bitplanes. This observation can be intuitively explained due to the removal of high-frequency (detail) information based on the repetitive application of the low-pass analysis filter. Experimental validation is given in Figure 6, where most of the blocks become significant within bitplanes $b = \{10, 9\}$. Secondly, the high-frequency spatio-temporal subbands exhibit a heavy-tail distribution based on the doubly-stochastic model of (4) and therefore the probability of significance (and newly-significance) is more skewed, as seen in Figure 5.

IV. Rate Approximation of Intra-band Embedded Coding

In this section, we derive rate estimations for an embedded coding scheme using a quadtree decomposition structure followed by block-coding for the maximum depth of the quadtree. This follows the system model outlined in Figure 1. Using the high-frequency and low-frequency wavelet coefficient models and the previously-derived probability estimates from Proposition 1-Proposition 3 in Section III, we derive the rate of quadtree coding, block coding, and coefficient refinement coding. The derived rate estimates can be modified to fit a variety of wavelet coding schemes consisting of subsets of the general structure of Figure 1, i.e. quadtree-based coders [22] [23], and block-based coders [19] [24] [25].

A. Rate of Quadtree Significance-map Coding for High-frequency Spatio-temporal Subbands

In most video frames, the probability of block significances at various levels in the quadtree varies considerably depending on the spatio-temporal subband statistics and the block size, since small blocks encapsulate small areas while large blocks represent large areas within each subband. For most practical wavelet video coders, we can simplify our analysis by assuming that the significance map encoding of quadtree structures remains virtually uncompressed. An example that strongly suggests this property is given in Table 3, where operational measurements from the coder of [29] were used for a variety of input video sequences. Note that while the rate is not identically 1 bit per quadtree symbol encoded, the difference in size between quadtrees leads to a distribution that cannot be well compressed, especially for lower bitplanes. In the remainder of this paper, we assume that the rate of significance-map coding for blocks in the quadtree structure is approximately equal to the number of times significance-map symbols are encoded. We additionally assume that coefficients in blocks of all sizes are sufficiently correlated so that the i.i.d. joint Gaussian distribution with a fixed local variance θ can be used to model them.

The significance of a block in the quadtree decomposition may be encoded in two cases: *i*) If the block is found newly significant at bitplane b, its significance will be encoded at that moment and it will never be encoded again; *ii*) if the block's parent is found to be significant at bitplane b even though the block itself is non-significant, it will be coded continuously until the block is found newly significant. Condition (*ii*) is added in most state-of-the-art coders to exploit intra-band spatial correlation of wavelet coefficients.

Under the above-stated two conditions, we now derive the probability of block significance, which corresponds to the rate of block significance coding. In general, if a block at quadtree level k, $2 \le k \le K$, has n coefficients, its parent block at level $k-1$ has $4n$ coefficients[2]. The number of symbols (or rate) used to encode the significance of a block of size n found significant at bitplane b depends on the probability that its parent is found significant at higher bitplanes $b+r$, $r \ge 0$ (which means that the block significance will be coded a total of $r+1$ times). Given the subband variance σ^2, this can be formulated as:

$$R_{\text{block\_newsig}}(\upsilon_b, n) = \sum_{r=0}^{B_{\max}-b} \Pr\{\text{sig}(T_{b+r}, 4n) = 1 \mid \text{newsig}(T_b, n) = 1\} \tag{14}$$

Averaging the rates over all bitplanes $B_{\min}, \ldots, B_{\max}$, we get the following rate estimate:

$$\begin{aligned} R_{\text{block\_sig}}(\upsilon_{B_{\min}}, n) &= \sum_{b=B_{\min}}^{B_{\max}} \Pr\{\text{newsig}(T_b, n) = 1\} R_{\text{block\_newsig}}(\upsilon_b, n) \\ &= \sum_{b=B_{\min}}^{B_{\max}} \delta_{\upsilon_b, n}^{\text{band}} \sum_{r=0}^{B_{\max}-b} \Pr\{\text{sig}(T_{b+r}, 4n) = 1 \mid \text{newsig}(T_b, n) = 1\} \end{aligned} \tag{15}$$

where $\text{band} \in \{\text{low,high}\}$ depending on the type of frequency subband we are interested in. The probability within the summation of (14) can be estimated by obtaining the local distribution of parameter θ given the newly-significant child block at bitplane b, and then determining the probability of a significant coefficient (in the other 3 child blocks) at bitplane $b+r$, thereby deriving the conditional probability of significance for the parent block of $4n$ coefficients:

[2] In the subsequent derivations of this paper, whenever blocks of size n and $4n$ appear in the same expression, it is implied that the first is the child block at quadtree level k while the second is the parent block at quadtree level $k-1$.

$$\Pr\{\text{newsig}(T_b,n)=1 \mid \text{sig}(T_{b+r},4n)=1\} = \Pr\{\text{newsig}(T_b,n)=1 \mid \Theta\} \tag{16}$$

where:

$$\Theta \sim p(\theta \mid \text{sig}(T_{b+r},4n)) = \frac{\Pr\{\text{sig}(T_{b+r},4n) \mid \theta\}\, p(\theta)}{\Pr\{\text{sig}(T_{b+r},4n)\}} = \frac{\left(1-\left[\text{erf}\left(\frac{T_{b+r+1}}{\sqrt{2\theta}}\right)\right]^{4n}\right)\frac{1}{\sigma^2}e^{-\frac{\theta}{\sigma^2}}}{\chi^{\text{band}}_{\upsilon_{b+r},4n}} \tag{17}$$

Thus, (16) becomes:

$$\begin{aligned}\Pr\{\text{newsig}(T_b,n)=1 \mid \Theta\} &= \int_{\theta=0}^{\infty} \Pr\{\theta \mid \text{newsig}(T_{b+r},4n)\}\cdot \Pr\{\text{newsig}(T_b,n)=1 \mid \theta, \text{newsig}(T_{b+r},4n)\}\, d\theta \\ &= \frac{1}{\delta^{\text{band}}_{\upsilon_b,n}} \int_{\theta=0}^{\infty} \frac{1}{\sigma^2} e^{-\frac{\theta}{\sigma^2}} \left(1-\left[\text{erf}\left(\frac{T_{b+r}}{\sqrt{2\theta}}\right)\right]^{4n}\right)\left(\left[\text{erf}\left(\frac{T_{b+1}}{\sqrt{2\theta}}\right)\right]^{n} - \left[\text{erf}\left(\frac{T_b}{\sqrt{2\theta}}\right)\right]^{n}\right) d\theta\end{aligned} \tag{18}$$

Similar to the derivation used in Appendix A, $[\text{erf}(T/x)]^n$ can be seen as an indicator function. Hence, we approximate the integral of (18) by treating the two multiplicative $\text{erf}(\cdot)$ terms as an indicator and a step function using the approximation of (50) with the parameter γ_n^2 estimated by (54) (see the full derivation in the proof given in Appendix A):

$$1-\left[\text{erf}\left(\frac{T_{b+r}}{\sqrt{2\theta}}\right)\right]^{4n} \cong \text{I}\left(\theta \geq \frac{T_{b+r}^2}{\ln(4n)^{1.296}+0.166}\right) \tag{19}$$

$$\left[\text{erf}\left(\frac{T_{b+1}}{\sqrt{2\theta}}\right)\right]^{n} - \left[\text{erf}\left(\frac{T_b}{\sqrt{2\theta}}\right)\right]^{n} \cong \text{I}\left(\frac{T_b^2}{\ln(n)^{1.296}+0.166} \leq \theta \leq \frac{T_{b+1}^2}{\ln(n)^{1.296}+0.166}\right) \tag{20}$$

where I is the indicator function. Based on (19) and (20), we can approximate (18) by:

$$\gamma^{\text{band}}_{\upsilon_b,n,r} \cong \begin{cases} \frac{1}{\delta^{\text{band}}_{\upsilon_b,n}}\left(\chi^{\text{band}}_{\upsilon_{b+r},4n} - \chi^{\text{band}}_{\upsilon_{b+1},n}\right)^{+} & , \frac{T_b^2}{\ln(n)^{1.296}+0.166} \leq \frac{T_{b+r}^2}{\ln(4n)^{1.296}+0.166} \\ 1 & , \text{otherwise} \end{cases} \tag{21}$$

Combining (15), (19)–(21) together, we obtain the final expression:

$$R_{\text{block\_sig}}(\upsilon_{B_{\min}},n) = \sum_{b=B_{\min}}^{B_{\max}} \delta^{\text{band}}_{\upsilon_b,n} \sum_{r=0}^{B_{\max}-b} \gamma^{\text{band}}_{\upsilon_b,n,r} \tag{22}$$

The average rate per coefficient is $R_{\text{block\_sig}}(T_b,n)/n$. If we let n be the smallest block size, then we must sum up the rates for K levels of the quadtree decomposition in the subbands of each spatial resolution to obtain the total rate for quadtree encoding:

$$R_{\text{quadtree}}(\upsilon_{B_{\min}}) = \sum_{k=0}^{K-1} \frac{R_{\text{block\_sig}}(\upsilon_{B_{\min}},4^k n)}{4^k n} \tag{23}$$

H frames, temporal level 3, spatial level 2	$T_b = 16$		$T_b = 32$		$T_b = 64$	
	Symbols	Rate	Symbols	Rate	Symbols	Rate
Coastguard, HL	2254	0.97	1556	0.85	447	0.73
Foreman, HL	1056	0.87	571	0.79	284	0.79
Silent, HL	1787	0.91	1046	0.81	348	0.79

Table 3. Examples of the number of 8×8 blocks encoded at each bitplane (symbols) and the average rate of encoding each block significance (rate is measured in bits-per-symbol), using the coder of [29]. Note that the rate per symbol increases as the bitplane decreases and approaches one bit-per-symbol.

B. Estimation of Block-coding Rate and Coefficient Refinement-coding Rate in High-frequency Spatio-temporal Subbands

In this subsection, we estimate the rate of encoding both the significance and the refinement of coefficients based on the quantization level (or minimum bitplane level) $T_{B_{\min}}$.

First, we consider the significance rate. In block-based coders like JPEG2000 [19] or ESCOT [25] (i.e. intra-band coders that do not employ quadtree decompositions), a subband is simply divided into blocks, which are then coded independently. In these algorithms, the significance of each coefficient within the block is encoded. If a coefficient is significant, then its refinement bits are also encoded. The significance of a coefficient relative to the threshold $T_{B_{\min}}$ is a binary value. Hence, the rate from independently encoding the significance of each coefficient can be expressed by the binary entropy function $\mathrm{H}(\rho_{B_{\min}})$, where $\rho_{B_{\min}}$ is the probability that the coefficient is significant compared to $T_{B_{\min}}$. However, when context-based entropy decoding is employed, dependencies between neighboring coefficients can be exploited, such that the rate (based on the doubly-stochastic model) is [9]:

$$R_{\mathrm{ZC}}(\upsilon_{B_{\min}}, n) = \int_{\theta=0^+}^{\infty} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta^2} \mathrm{H}\left(\mathrm{erf}\left(\frac{T}{\sqrt{2\theta}}\right)\right) d\theta \cong \mathrm{H}(\rho_{B_{\min}}) - 0.6707 \upsilon_{B_{\min}} e^{-1.407 \upsilon_{B_{\min}}} \tag{24}$$

A common weakness of using only block-coding methods without quadtree coding [19] [24] [25] is that the spatial distribution of local variances in the spatio-temporal subbands is not exploited, since the same context-conditioning scheme is utilized for all blocks. Coders combining quadtree coding and block-coding techniques [20] [21] exploit the spatial correlation of local variances by using the quadtree decomposition which inherently assigns fewer symbols to the insignificant areas: if all coefficients within a certain block are insignificant, quadtree-based coders return a single "0" for that block and do not encode the coefficients within that block. However, for (minimum-sized) blocks that are significant, context coding is used for the coefficients in the block. Hence, the effective significance coding rate for the blocks resulting at the maximum quadtree depth depends on the probability of the smallest blocks being significant, and the context coding rate conditioned on the block being significant:

$$\begin{aligned} R_{\mathrm{ZC,QB}}(\upsilon_{B_{\min}}, n) &= \mathrm{H}(\mathrm{sig}(T_{B_{\min}}, 1) \mid \mathrm{sig}(T_{B_{\min}}, n), \Theta) \\ &= \int_{\theta=0}^{\infty} p(\theta) \cdot \Pr\{\mathrm{sig}(T_{B_{\min}}, n) | \theta\} \cdot \mathrm{H}(\Pr\{\mathrm{sig}(T_{B_{\min}}, 1) \mid \mathrm{sig}(T_{B_{\min}}, n), \theta\}) d\theta \end{aligned} \tag{25}$$

where $\mathrm{sig}(T_{B_{\min}}, 1)$ indicates the significance test at bitplane $B_{\min}$ for an individual coefficient within a wavelet subband. The probability of coefficient significance given block significance and local variance θ can be solved using Bayes rule:

$$\Pr\{\mathrm{sig}(T_{B_{\min}},1)|\mathrm{sig}(T_{B_{\min}},n),\theta\}=\frac{\Pr\{\mathrm{sig}(T_{B_{\min}},n)|\mathrm{sig}(T_{B_{\min}},1),\theta\}\Pr\{\mathrm{sig}(T_{B_{\min}},1)|\theta\}}{\Pr\{\mathrm{sig}(T_{B_{\min}},n)|\theta\}}=\frac{1-\mathrm{erf}\left(\frac{T_{B_{\min}}}{\sqrt{2\theta}}\right)}{1-\left[\mathrm{erf}\left(\frac{T_{B_{\min}}}{\sqrt{2\theta}}\right)\right]^n} \tag{26}$$

The resulting expression is:

$$R_{\mathrm{ZC,QB}}(\upsilon_{B_{\min}},n)=\int_{\theta=0}^{\infty}\frac{1}{\sigma^2}e^{-\frac{1}{\sigma^2}\theta^2}\left(1-\left[\mathrm{erf}\left(\frac{T_{B_{\min}}}{\sqrt{2\theta}}\right)\right]^n\right)\mathrm{H}\left(\frac{1-\mathrm{erf}\left(\frac{T_{B_{\min}}}{\sqrt{2\theta}}\right)}{1-\left[\mathrm{erf}\left(\frac{T_{B_{\min}}}{\sqrt{2\theta}}\right)\right]^n}\right)d\theta \tag{27}$$

Notice that the expression is parametrical to the number of coefficients per smallest block, n . A minimum block size of approximately $n=16$ has been shown to achieve the most savings over pure context-based coding.

Consider now the rate of refinement-coding for significant coefficients. For a given error subband, the rate of quantizing a significant coefficient X based on its neighborhood information $\mathcal{N}X$ can be estimated as [9]:

$$R_{\mathrm{refinement}}(Q_{B_{\min}}(X)\mid\mathcal{N}X)\cong 1-\log_2(\frac{1}{\rho_{B_{\min}}}-1)-\frac{\log_2\rho_{B_{\min}}}{1-\rho_{B_{\min}}}-\frac{0.2988}{(\upsilon_{B_{\min}}+0.9773)^{0.8}} \tag{28}$$

Therefore, the total coding rate can be estimated as:

$$R_{\mathrm{high}}(\upsilon_{B_{\min}})=R_{\mathrm{quadtree}}(\upsilon_{B_{\min}})+R_{\mathrm{ZC,QB}}(\upsilon_{B_{\min}},n)+\rho_{B_{\min}}R_{\mathrm{refinement}}(Q_{B_{\min}}(X)\mid\mathcal{N}X) \tag{29}$$

C. Coding Rate of Low-frequency Spatio-temporal Subbands

Following the independent Gaussian model assumption for the low-frequency subbands of L frames, we derived the following rate estimate for the coding of low-frequency wavelet coefficients.

Proposition 4: The rate of encoding a low-frequency coefficient can be approximated as follows:

$$R_{\mathrm{low}}(\upsilon_{B_{\min}})\cong\mathrm{H}\left(\mathrm{erf}\left(\tfrac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)\right)+\mathrm{erfc}\left(\tfrac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)\log_2\left(\mathrm{erfc}\left(\tfrac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)\frac{\sqrt{2\pi e}}{\upsilon_{B_{\min}}}\right)-\frac{\upsilon_{B_{\min}}}{\sqrt{2\pi}}\exp(-\tfrac{\upsilon^2_{B_{\min}}}{2})\log_2 e \tag{30}$$

*Proof***:** We use the estimation method of Mallat and Falzon [28] for the rate in the low-rate (high distortion) region:

$$\begin{aligned}\mathrm{H}(Q_{B_{\min}}(X))&\cong\mathrm{H}(|X|\geq T_{B_{\min}})+p(|X|\geq T_{B_{\min}})\mathrm{H}(Q_{B_{\min}}(X)\mid(|X|\geq T_{B_{\min}}))\\&=\mathrm{H}(\mathrm{erf}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)+p(|X|\geq T_{B_{\min}})\mathrm{H}(Q_{B_{\min}}(X)\mid(|X|\geq T_{B_{\min}}))\end{aligned} \tag{31}$$

For significant coefficients, we use the high-resolution hypothesis from [19], which is:

$$\mathrm{H}(Q_{B_{\min}}(X))\cong\mathrm{H}(X)-\log_2\upsilon_{B_{\min}} \tag{32}$$

This gives us:

$$
\begin{aligned}
\mathrm{H}(Q_{B_{\min}}(X) \mid (|X| \geq T_{B_{\min}})) &\cong \mathrm{H}(X \mid (|X| \geq T_{B_{\min}})) - \log_2(\upsilon_{B_{\min}}) \\
&= -2 \int_{\upsilon_{B_{\min}}}^{\infty} \frac{\exp(-\frac{x^2}{2})}{\mathrm{erfc}(\frac{\upsilon_{B_{\min}}}{2})\sqrt{2\pi}} \log_2\left(\frac{\exp(-\frac{x^2}{2})}{\mathrm{erfc}(\frac{\upsilon_{B_{\min}}}{2})\sqrt{2\pi}} \right) dx - \log_2(\upsilon_{B_{\min}}) \\
&= \log_2(\mathrm{erfc}(\tfrac{\upsilon_{B_{\min}}}{2}) \frac{\sqrt{2\pi}}{\upsilon_{B_{\min}}}) + \int_{\upsilon_{B_{\min}}}^{\infty} \frac{\exp(-\frac{x^2}{2})}{\mathrm{erfc}(\frac{\upsilon_{B_{\min}}}{2})\sqrt{2\pi}} x^2 \log_2(e) dx \\
&= \log_2(\mathrm{erfc}(\tfrac{\upsilon_{B_{\min}}}{2}) \frac{\sqrt{2\pi}}{\upsilon_{B_{\min}}}) + \log_2(e) \left(\frac{\upsilon_{B_{\min}} \exp(-\frac{\upsilon_{B_{\min}}^2}{2})}{\mathrm{erfc}(\frac{\upsilon_{B_{\min}}}{2})\sqrt{2\pi}} + \int_{\upsilon_{B_{\min}}}^{\infty} \frac{\exp(-\frac{x^2}{2})}{\mathrm{erfc}(\frac{\upsilon_{B_{\min}}}{2})\sqrt{2\pi}} dx \right) \\
&= \log_2(\mathrm{erfc}(\tfrac{\upsilon_{B_{\min}}}{2}) \frac{\sqrt{2\pi e}}{\upsilon_{B_{\min}}}) + \left(\frac{\upsilon_{B_{\min}} \exp(-\frac{\upsilon_{B_{\min}}^2}{2})}{\mathrm{erfc}(\frac{\upsilon_{B_{\min}}}{2})\sqrt{2\pi}} \right) \log_2 e
\end{aligned} \tag{33}
$$

The proof follows from substituting (33) into (31). ■

In order to avoid integrations in (30), we use the approximation for $\mathrm{erf}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)$ given in (11).

Notice that, for high-rate (low distortion) regions where the variance of each coefficient is significantly larger than the quantization stepsize (i.e. $\upsilon_{B_{\min}}$ is small), $\mathrm{erfc}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right) \approx 1$, and the rate estimate of (30) becomes the well-known high-rate approximation [19]:

$$
R_{\mathrm{low}}(\upsilon_{B_{\min}}) \cong \log_2 \sqrt{2\pi e} - \log_2 \upsilon_{B_{\min}} \tag{34}
$$

Table 4 gives an example of the accuracy of (30) as a function of quantization step. We also present the results with the more conventional model of (34) used in prior work [9] in order to indicate the superior approximation achieved with the proposed estimation of (30). Notice from Table 4 that, although the proposed model still remains relatively inaccurate when only the highest 2-3 bitplanes are decoded, it becomes increasingly accurate as the number of decoded bitplanes increases.

Quantization step size (T_b)	Encoded size (bits)	Conventional approximation [9]	Error	Proposed approximation	Error
1024	1336	3289	146.2%	941	-29.6%
512	3208	4513	40.7%	3019	-5.9%
256	5032	5920	17.7%	5128	1.9%
128	6816	7507	10.1%	7053	3.5%
64	8472	9080	7.2%	8856	4.5%
32	10096	10661	5.6%	10595	4.9%

Table 4. Example comparison between the actual encoded rate ("Encoded size" using the coder of [29]), the conventional approximation from prior work, and the proposed approximation of (30) for an 44x38 LL subband of an L-frame in the *Foreman* sequence.

Based on the derived estimations of (29) and (34), the average coding rate per pixel over all subbands of MCTF-based wavelet video coding is:

$$
R_{\mathrm{total}}(\upsilon_{B_{\min}}) = 4^{-J} R_{\mathrm{low},J,0} + \sum_{j=1}^{J} \sum_{m=1}^{3} 4^{-j} R_{\mathrm{high},j,m} \tag{35}
$$

where (band, j, m) indicates which model $\mathrm{band}=\{\mathrm{low},\mathrm{high}\}$ is used to determine the rate of the subband of the mth orientation in the jth scale of the wavelet frame decomposition, with $j = 1$ indicating the finest resolution, and $j = J$ indicating the coarsest resolution (lowest frequency). $R_{\text{low},J,0}$ indicates the coarsest LL subband.

V. Distortion Estimation of Combined Quadtree and Block Coding followed by MCTF Reconstruction

In this section we determine the distortion of the various coding passes mentioned previously for the different subbands, and a general formulation of the average distortion for the combined decoding followed by MCTF reconstruction is derived.

A. Distortion of Combined Decoding followed by Inverse Spatial DWT

As mentioned in Section II, SAQ followed by the accumulation of all coding passes up to any bitplane $B_{\min}$ corresponds to a double-deadzone uniform quantization of wavelet coefficients. In other words, once the produced bitstream is truncated at bitplane $B_{\min}$, we have a quantizer of the form given in (1) with $T_{B_{\min}} = 2^{B_{\min}}\Delta$. The average distortion of a high-frequency subband when a uniform quantizer with this deadzone is applied is [9]:

$$\begin{aligned} D_{\text{high}} &= \mathrm{E}[(X-\hat{X})^2] \\ &= \left\{-\rho_{B_{\min}}\left(\upsilon_{B_{\min}}+1/\sqrt{2}\right)^2+1-\rho_{B_{\min}}{}^2\upsilon_{B_{\min}}{}^2/\left(1-\rho_{B_{\min}}\right)^2\right\}\sigma^2 \end{aligned} \tag{36}$$

where σ^2 is the variance of the Laplacian-distributed coefficients. We now derive the distortion of the low-frequency subband of an L frame.

Proposition 5: The estimated distortion for the low-frequency spatio-temporal subband is:

$$D_{\text{low}} = \left(\operatorname{erf}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\upsilon_{B_{\min}} e^{-\frac{\upsilon_{B_{\min}}^2}{2}} + \frac{\upsilon_{B_{\min}}^2}{12}\operatorname{erfc}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)\right)\sigma^2 \tag{37}$$

Proof: Based on the procedure in [28], we separate the distortion calculation D_{low} into the distortion of non-significant coefficients (deadzone), and the distortion of significant coefficients. The deadzone distortion is the variance of a truncated Gaussian at $T_{B_{\min}}$, which can be shown to be $D_{\text{low,zero}} = 1 - \sqrt{\frac{2}{\pi}}\upsilon_{B_{\min}} e^{-\frac{\upsilon_{B_{\min}}^2}{2}} / \operatorname{erf}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)$ using integration by parts. For significant coefficients, we use the high-rate assumption [19] [28], $D_{\text{low,nonzero}} \cong \frac{\upsilon_{B_{\min}}^2}{12}\sigma^2$. Hence, the total distortion is the weighted sum of the two metrics, or:

$$D_{\text{low}} = p_{\text{low,zero}} D_{\text{low,zero}} + p_{\text{low,nonzero}} D_{\text{low,nonzero}} = \operatorname{erf}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right) D_{\text{low,zero}} + \operatorname{erfc}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right) D_{\text{low,nonzero}} \tag{38}$$

which gives us (37). ■

Notice that, similar to the corresponding rate estimation of (30), for low-distortion (high-rate) regions where the variance of each coefficient is significantly larger than the quantization stepsize (i.e. $\upsilon_{B_{\min}}$ is small), $\operatorname{erfc}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right) \approx 1$, and the distortion estimate of (37) converges to the well-known high-rate approximation of $D_{low} \cong \frac{\upsilon_{B_{\min}}^2}{12}\sigma^2$.

For all the different subbands at all scales of the DWT, we get an average distortion:

$$\mathbf{d} = 4^{-J} G_J \cdot D_{\text{low},J,0} + \sum_{j=1}^{J}\sum_{m=1}^{3} 4^{-j} G_j \cdot D_{\text{high},j,m} \tag{39}$$

where G_j is the synthesis gain of the wavelet filter at the j th scale level, and $D_{\text{band},j,m}$ is the expected distortion of the m th type subband at the j th scale level, with $\text{band}=\{\text{low,high}\}$.

B. Distortion for MCTF Reconstruction

For generalized MCTF filtering, distortion takes on a linear combination of each L and H frame produced by the decomposition [9] [30]:

$$\overline{\mathbf{d}}_L^{(0)} = \sum_{k=1}^{T_{\text{MCTF}}} B^{(k)} \left(\prod_{j=1}^{k-1} A^{(j)}\right) \mathbf{d}_H^{(k)} + \left(\prod_{j=1}^{T_{\text{MCTF}}} A^{(j)}\right) \overline{\mathbf{d}}_L^{(T_{\text{MCTF}})} = [\mathcal{A}, \mathcal{B}_1, \ldots, \mathcal{B}_{T_{\text{MCTF}}}][\overline{\mathbf{d}}_L^{(T_{\text{MCTF}})} \mathbf{d}_H^{(1)}, \ldots, \mathbf{d}_H^{(T_{\text{MCTF}})}]^T \tag{40}$$

The last derivation of (40) is valid because the weight of each H -frame at each temporal level is a function of only the average number of connected pixels in the GOP. However, we note that this approximation can also be applied across several GOPs if the motion between GOPs is similar. In our experiments, linear minimum mean square error (MMSE) fitting is used to determine the weights of L -frames and H -frames and predict the distortion associated with the sequence.

VI. Complexity of Entropy Decoding and IDWT

A. Generic Complexity Modeling for Video Decoding

Since many multimedia decoders today typically reside in a variety of handheld (Video iPod, 3G cellphones, etc) and portable devices (notebooks, PDAs) that have stringent power and processing constraints, they are in general more resource-constrained than encoders. Hence, while a similar complexity estimation framework can be likewise derived for the encoder, we opt to focus on the decoding complexity in this paper. A second (and more algorithm-related) reason is that the encoding complexity is strongly dominated by the motion estimation complexity rather than the coding operations. Hence, accurate modeling of embedded encoding per-se is of a lesser importance for the R-D-C analysis of the encoder, as it is for the decoder.

In order to represent different decoder (receiver) architectures in a generic manner at the encoder (server) side, in our recent work [10] [11] we have deployed a concept that has been successful in the area of computer systems, namely, a virtual machine. The key idea of the proposed paradigm is that the same bitstream will require/involve different resources/complexities on various decoders. We adopt a generic complexity model that captures the abstract/generic complexity metrics (GCMs) of the employed decoding or streaming algorithm depending on the content characteristics and transmission bitrate. GCMs are derived by computing or estimating the average number of times the different operations are executed, such as the number of read symbols during entropy decoding, the number of multiply-accumulate operations performed during inverse transform, the number of motion compensation operations per pixel or coefficient, and the frequency of invocation of fractional pixel interpolation. The value of each GCM may be determined at encoding time for each adaptation unit q (e.g. the q-th video frame, or the q-th macroblock) following experimental or modeling approaches [10]. Our previous work demonstrated that the mapping of the derived GCMs to execution time provides a very accurate and straightforward manner of predicting the real (system-specific) complexity [35]. The added advantage of GCMs however is that they are not system-specific and they are also not restricted to a particular coding structure (predictive or MCTF-based). This makes them applicable for a broad class of motion-compensated video decoders. In this paper, unlike our previous work [10] [11], we focus on the derivation of entropy decoding and inverse transform GCMs based on stochastic models that analytically express the dependencies on the source characteristics and the algorithm operations.

B. Entropy Decoding Complexity

The complexity of decoding the quadtree significance at bitplane b depends on the size of the quadtree before the significance pass. Since the quadtree is virtually uncompressed for the vast majority of cases, the complexity is of the order of the quadtree significance map encoding rate:

$$C_{\text{quadtree}}(\upsilon_{B_{\min}}, n) \cong R_{\text{quadtree}}(\upsilon_{B_{\min}}, n) \tag{41}$$

with $R_{\text{quadtree}}(\upsilon_{B_{\min}}, n)$ given by (23) based on Proposition 1-Proposition 3. This includes both the number of read symbols (RS) associated with quadtree coding, and writing the significances into the quadtree structure.

Concerning block coding, we group together the number of symbols read from significance coding and refinement. Notice that, as long as the coefficient is in a significant block at bitplane b or higher, its significance will be coded, or it will be refined at bitplane b. Summing up all symbols read in the passes until bitplane $B_{\min}$ we have:

$$C_{\text{block}}(\upsilon_{B_{\min}}) = 4^{K-1} n \sum_{b=B_{\min}}^{B_{\max}} \chi_{\upsilon_b,n}^{\text{low}} + 4^{K-1} n \sum_{b=B_{\min}}^{B_{\max}} \chi_{\upsilon_b,n}^{\text{high}} \tag{42}$$

where $4^{K-1}n$ is the number of coefficients in the subband. Notice that the combination of (41) and (42) predicts the number of RS operations during entropy coding/decoding of a low or high-frequency spatio-temporal subband to a certain bitplane $B_{\min}$. Since each subband is encoded independently, the complexity metrics must first be estimated for each subband and then summed in the same weighted fashion as the rate calculation. In other words, for a given frame i, $1 \le i \le N$, we have:

$$C_{\mathrm{op}}^{i}(\upsilon_{B_{\min}}) = 4^{-J} C_{\mathrm{op},J,0}(\upsilon_{B_{\min}}) + \sum_{j=1}^{J}\sum_{m=1}^{3} 4^{-j} C_{\mathrm{op},j,m}(\upsilon_{B_{\min}}) \tag{43}$$

where $\mathrm{op} \in \{\mathrm{quadtree,block}\}$ and $C_{\mathrm{op},j,k}(\upsilon_{B_{\min}})$ is the quadtree and block coding complexity for each subband at spatial resolution j. Having obtained the RS estimates for quadtree and block coding, the expression $\alpha_{\mathrm{quadtree}} C_{\mathrm{quadtree}}^{i}(\upsilon_{B_{\min}}) + \alpha_{\mathrm{block}} C_{\mathrm{block}}^{i}(\upsilon_{B_{\min}})$ derives an estimate of the real complexity for frame i, where α_{op} is an approximate algorithmic (and platform dependent) complexity associated with each symbol used to perform operation op. See our recent work [35] for extended examples of adaptive generation of weighting factors for mapping GCM estimates to platform-specific complexity.

C. Complexity of the Inverse Spatial DWT

The complexity of the inverse DWT depends on the number of taps of the filter used as well as on the implementation method (convolution or lifting). In our prior work [11], we have modeled the transform-related complexity of a coding system that processes N video frames by expressing it as a decomposition into two functions relating to: *i*) the percentage of non-zero coefficients for a given SAQ threshold T_b (function $\mathcal{T}_{\mathrm{nonzero}}$); *ii*) the sum of run-lengths of zero wavelet coefficients (function $\mathcal{T}_{\mathrm{runlen}}$). The motivation behind (*i*) is that in an input-adaptive implementation, the number of non-zero multiply-accumulate operations in the synthesis filter-bank is directly proportional to the percentage of non-zero coefficients. Moreover, the distribution of the zeros within the transform subbands (as expressed by the sum of run-lengths) affects the number of consecutive filtering operations that can be avoided altogether. Once an estimate of $\mathcal{T}_{\mathrm{nonzero}}$ and $\mathcal{T}_{\mathrm{runlen}}$ is derived, the complexity of the inverse spatial DWT (non-zero MAC operations) is formulated as [11]:

$$\mathrm{FC}^{N} = \mathbf{C}_{\mathrm{nonzero}}^{N} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{nonzero}}^{N} + \mathbf{C}_{\mathrm{runlen}}^{N} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{runlen}}^{N} + \mathbf{C}_{\mathrm{dec\_const}}^{N} \cdot \mathbf{1} \tag{44}$$

with $\boldsymbol{\mathcal{T}}_{\mathrm{nonzero}}^{N}$ and $\boldsymbol{\mathcal{T}}_{\mathrm{runlen}}^{N}$ the N-element vectors of the corresponding functions and the parameter vectors $\mathbf{C}_{\mathrm{nonzero}}^{N}$ and $\mathbf{C}_{\mathrm{runlen}}^{N}$ can be estimated based on linear regression and off-line training [11]. In our current work two main differences exist in the derivation of the non-zero MAC operations of the IDWT in comparison to [11]. Firstly, linear MMSE fitting is used to determine $\mathbf{C}_{\mathrm{nonzero}}^{N}$, $\mathbf{C}_{\mathrm{runlen}}^{N}$ and predict the number of non-zero MAC operations associated with the sequence. This is equivalent to the process performed for the derivation of the final MCTF distortion in (40) (Section V.B). More importantly, in this paper, we present an analytical calculation of the decomposition functions mentioned above based on stochastic source models. In this way, the proposed analytical derivations create a clear link between the source parameters (variances of the distributions) and the derived complexity estimates. The decomposition function $\mathcal{T}_{\text{nonzero}}$ for the high-frequency spatio-temporal subbands is derived by (6), while for the low-frequency spatio-temporal subbands it is derived by:

$$\mathcal{T}_{\text{nonzero}} = \text{erfc}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right) \tag{45}$$

with $\text{erf}\left(\frac{\upsilon_{B_{\min}}}{\sqrt{2}}\right)$ approximated as in (11). In addition, $\mathcal{T}_{\text{runlen}}$ is derived by the percentage of non-significant blocks for a certain SAQ threshold $T_{B_{\min}}$, expressed by:

$$\mathcal{T}_{\text{runlen}} = \Pr\{\text{sig}(T_{B_{\min}}, n) = 0\} = 1 - \chi^{\text{band}}_{\upsilon_{B_{\min}},n} \tag{46}$$

with $\chi^{\text{band}}_{\upsilon_{B_{\min}},n}$ estimated by (9) for the high-frequency temporal subbands and by (10) for the LL subband of the L frames. Following the lifting dependencies of popular wavelet filter-pairs, we set an average of $n = 64$ since a window of 7×7 coefficients and 9×9 coefficients is used in the lifting steps of the inverse DWT for the low and high-frequency subbands [19].

Additionally, note that the number of taps also affects memory usage in the system. While memory usage is another concern in battery-limited devices, in this work we are primarily concerned with time-based complexity, as this more greatly affects the performance of delay-sensitive applications.

VII. Simulation Results

In this section we validate the derived analytical R-D-C expressions of this paper by presenting experiments with three common interchange format (CIF) resolution sequences (*Coastguard*, *Foreman*, *Silent*) that encapsulate a variety of motion and texture characteristics. Apart from validating the theoretical modeling of rate-distortion and complexity-distortion, the interplay of rate and complexity for achieving the same video quality under different coding structures is discussed.

For validation purposes, we utilize the spatial-domain MCTF version of the coder of [29] that performs multihypothesis MCTF decomposition with a variety of temporal filters and intra-band quadtree-based coding of the significance maps, and block-based intra-band coding after a block size of 4×4 coefficients is reached in the quadtree decomposition. Figure 7–Figure 9 present our results for a variety of spatial (S) and temporal (T) decomposition levels. Distortion, as estimated by (36)–(40) in Section V, is converted into peak signal to noise ratio (PSNR). The entropy-decoding complexity is quantified by the number of read symbols per second. For the inverse transform, we plot the number of non-zero MAC operations per second (FC/s). The results demonstrate that the proposed R-D-C modeling predicts the experimental behavior of the advanced MCTF-based wavelet video coder accurately for all the different cases under investigation. Different choices for MCTF (temporal) levels and spatial decomposition levels lead to different tradeoffs in rate and complexity for the same distortion in the decoded video.

Since the proposed models of Sections III–VI enable accurate estimation of rate and complexity for a variety of decoding distortion, the derived framework can be used in a variety of applications where rate or complexity tradeoffs are of paramount importance (see [3] [7] [8] [10]). For example, the R-C curve may be used to optimize post-encoding bitstream shaping, where an encoded bitstream may be truncated and transmitted at a lower rate based on decoder-specified complexity bounds.

There are several interesting aspects to note from our results. For example, for good quality video decoding (PSNR range of 32 dB–40 dB) there is typically an overhead of about 300 to 500 kbps when one uses two temporal levels instead of four and an overhead of about 600 to 900 kbps when one uses two temporal and two spatial levels. Notice that the exact overhead is both sequence and bitrate dependent and the proposed theoretical modeling captures this behavior accurately. Apart from the rate overhead, there is also an increase in the number of entropy decoding operations by about $2.5 \cdot 10^5$ to $5 \cdot 10^5$ RS/s and 10^6 to $2 \cdot 10^6$ RS/s for the "2T-4S" and "2T-2S" cases (respectively) in comparison to the "4T-4S" case. However, concerning the IDWT complexity, Figure 9 demonstrates that a large variation exists in the performance of the different approaches depending on the sequence and bitrate region. The case of "2T-2S" is the best in terms of operations per second, followed by the "4T-4S" case and by the "2T-4S" case, since the two latter require more spatial reconstruction levels. The fact that the "2T-4S" case appears to be worse than the "4T-4S" case can be explained by the increase in the non-zero coefficients due to the fact that the "2T" case includes four times more L frames as compared to the "4T" case, and L frames contain a higher percentage of non-zero coefficients in comparison to H frames (for the same quantization parameters). It is also interesting to notice that, for the low to medium rate coding of the *Coastguard* sequence, the "4T-4S" case is the most efficient both in R-D and C-D performance. The proposed modeling approach agrees with all these observations, a fact that validates the importance of analytical R-D-C modeling methods that adapt based on both source and algorithm statistics.

It is also interesting to note that, if one ignores the coding bitrate and focuses on the complexity-distortion tradeoffs, Figure 8 and Figure 9 reveal that, for the same number of entropy decoding operations, the "4T-4S" case can provide gains of 2 to 8 dB in comparison to the other alternatives. On the other hand, the "2T-2S" case may outperform the other decompositions by 2.5 to 10 dB for the same number of non-zero MAC operations during the IDWT. On different platforms where each entropy decoding operation and IDWT MAC operation may have different respective computational workloads and/or energy consumption levels, the significant tradeoffs between the different types of decoding complexities can be exploited to optimally configure coder parameters to run on a specific system.

Finally, it is interesting to investigate how rate and complexity change for different coding parameters for a higher resolution video, e.g. in sequences of Standard Definition (SD) format. The entropy decoding results for the 720x480 *Mobile* sequence (30 frames/sec) are presented in Figure 10. As seen in the figure, our theoretical approximations are fairly accurate in predicting the large performance gain of 4 temporal levels over 2 temporal levels of decomposition. Interestingly, this gain is not so prominent for CIF sequences, as indicated by Figure 7 and Figure 8. One reason is that the MCTF process can better exploit the correlation between neighboring pixels and coefficients in SD sequences due to the decrease of spatio-temporal aliasing in comparison to CIF sequences. Hence, the percentage of non-zero coefficients in the high-frequency subbands is decreased. Consequently the number of read symbols (entropy decoding complexity) and the required bitrate to encode H-frames are reduced when using more temporal levels.

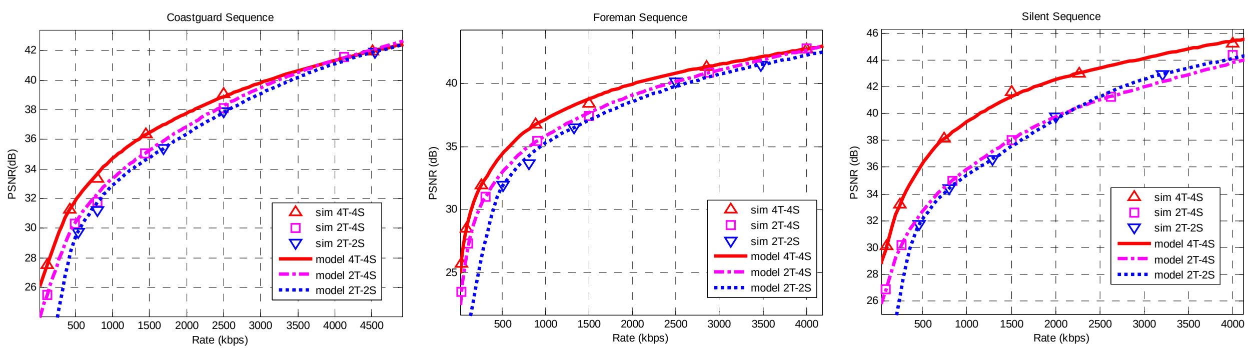

Figure 7. Rate-distortion plots for different configurations of the spatio-temporal decomposition parameters.

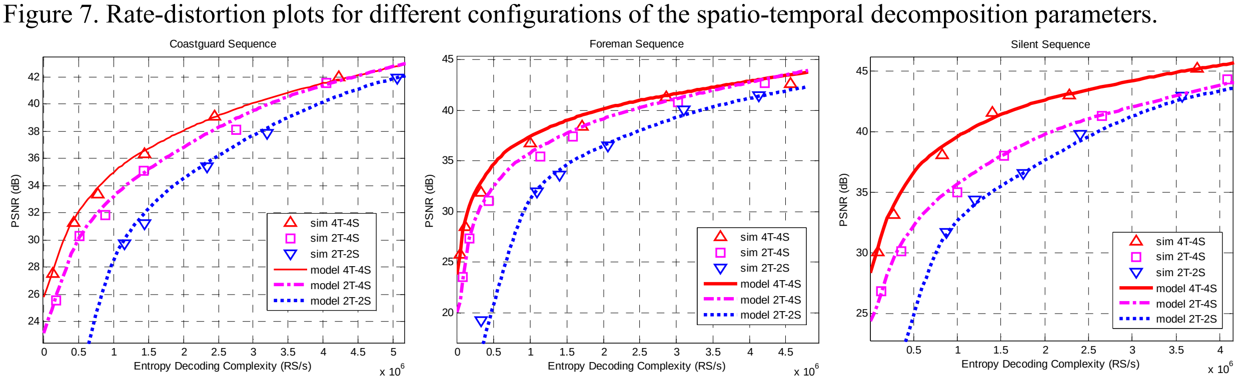

Figure 8. Entropy decoding complexity vs. distortion plots for different spatio-temporal decomposition parameters, where “S” and “T” indicate the number of spatial and temporal levels (respectively).

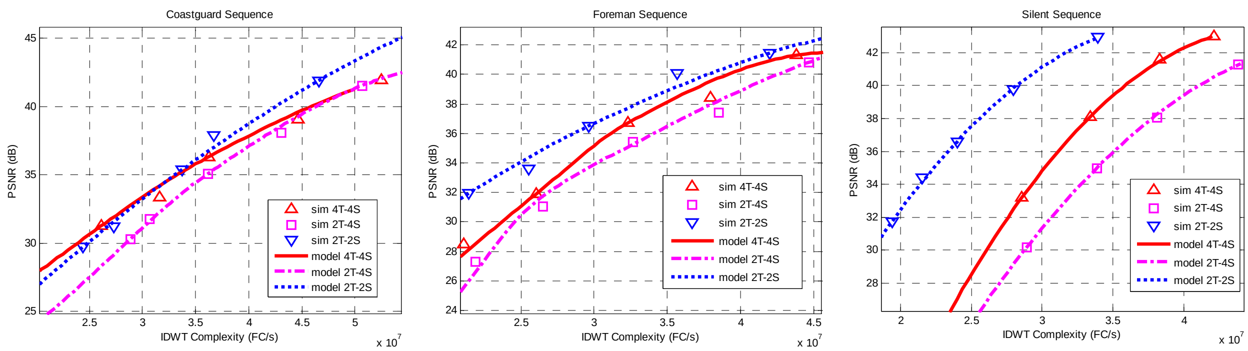

Figure 9. IDWT complexity vs. distortion plots for different spatio-temporal decomposition parameters, where "S" and "T" indicate the number of spatial and temporal levels (respectively).

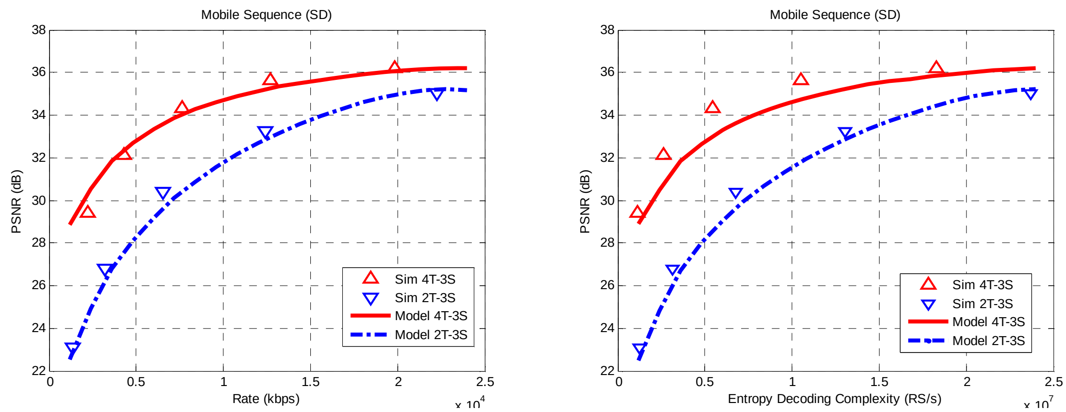

Figure 10 Rate and entropy decoding complexity curves for 720x480 *Mobile* sequence. The cases of 4 and 2 temporal decomposition levels are presented. Both encodings use 3 spatial decomposition levels.

VIII. Conclusions

This paper presents an analytical modeling framework that derives rate, distortion and (decoding) complexity predictions for wavelet-based video coders. Our analysis encapsulates a broad variety of coding techniques found in state-of-the-art coding schemes. By analytically deriving probabilities for block and coefficient significance according to the quantization threshold (for both low and high-frequency temporal subbands), we are able to establish analytical models that approximate well the R-D-C behavior of a state-of-the-art wavelet-based video coder. In this way, this work complements prior work on operational rate-distortion modeling for video coders by extending its applicability to a broader coding paradigm. At the same time, it complements complexity modeling frameworks proposed in earlier work by deriving analytically the input statistics used in these approaches. As such, this work bridges the gap between the operational measurements used in prior complexity modeling work and stochastic estimates common in rate-distortion modeling work.

The theoretical R-D-C analysis presented in this paper may guide the construction of more efficient intra-band coding mechanisms targeting error frames in particular. An open question concerns the efficiency of quadtree coding versus block coding mechanisms (both compression-wise and implementation-wise) and the optimal setting (e.g. minimum block size or combination of coding passes) for a coder that encodes error frames using both schemes in succession. Perhaps more importantly, the proposed R-D-C analysis allows for the efficient exploration of complexity and rate tradeoffs for different video qualities and resolutions. As indicated by our results with CIF and SD-resolution videos, a careful selection of coder parameters is important for optimizing the performance in a complexity-distortion or rate-distortion sense. A thorough investigation of the theoretical R-D-C tradeoffs between different coder parameters for different video resolutions is an interesting topic for future research.

Appendix A: Derivation of Probability of a Significant Block in a High-Frequency Subband

Proof of Proposition 1: First, let us consider the probability that a block with n coefficients is insignificant to a threshold T. According to (8), we are integrating over each component of random vector $\mathbf{X} = (X_1, ..., X_n)$ from $-T$ to T. Hence, we want to find the following probability:

$$\begin{aligned}\iint\limits_{-T\cdot\mathbf{1}\leq\mathbf{x}\leq T\cdot\mathbf{1}} \cdots \int p(\mathbf{x})d\mathbf{x} &= \iiint\limits_{-T\leq x_1,\ldots,x_n\leq T} \left(\int\limits_{0+}^{\infty} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \cdot \frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}(x_1^2+x_2^2+\ldots+x_n^2)} d\theta \right) dx_1 dx_2 \ldots dx_n \\ &= \int\limits_{0+}^{\infty} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \left(\frac{1}{\sqrt{2\pi\theta}} \int\limits_{-T}^{T} e^{-\frac{x^2}{2\theta}} dx \right)^n d\theta\end{aligned} \tag{47}$$

where the final result of (47) is reached by rearranging integrals, and by the fact that the n resulting integrals are equal.

Now, recall the definition of the erf function $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx$. Using substitution of variables, we get the following expression:

$$\frac{1}{\sqrt{2\pi\theta}} \int\limits_{-T}^{T} e^{-\frac{x^2}{2\theta}} dx = 2 \cdot \frac{1}{\sqrt{2\pi\theta}} \int\limits_{0}^{T} e^{-\frac{x^2}{2\theta}} dx = \frac{2}{\sqrt{\pi}} \int\limits_{0}^{\frac{T}{\sqrt{2\theta}}} e^{-x'^2} dx' = \mathrm{erf}\left(\frac{T}{\sqrt{2\theta}}\right) \tag{48}$$

Hence, substituting (48) into (47) gives us:

$$\iint\limits_{-T\cdot\mathbf{1}\leq\mathbf{x}\leq T\cdot\mathbf{1}} \cdots \int p(\mathbf{x})d\mathbf{x} = \int\limits_{0}^{\infty} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \left[\mathrm{erf}\left(\frac{T}{\sqrt{2\theta}}\right) \right]^n d\theta \tag{49}$$

Unfortunately, the integral of (49) has no explicit form. However, $\mathrm{erf}(\ )^n$ resembles an indicator function for large values of n:

$$[\mathrm{erf}(x)]^n \approx \mathrm{I}(x \geq \gamma_n) \tag{50}$$

Or, for our case:

$$\left[\mathrm{erf}\left(\frac{T}{\sqrt{2\theta}}\right)\right]^n \approx \mathrm{I}\left(\frac{T}{\sqrt{2\theta}} \geq \gamma_n\right) = \mathrm{I}\left(\theta \leq \frac{T^2}{2\gamma_n^2}\right) \tag{51}$$

Here, γ_n is an estimate for where the step occurs, given n . Substituting (51) into (49) gives us the following approximation:

$$\int_0^\infty \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \left[\operatorname{erf}\left(\frac{T}{\sqrt{2\theta}}\right)\right]^n d\theta \approx \int_0^\infty \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} \left[\mathrm{I}\left(\theta \le \frac{T^2}{2\gamma_n^2}\right)\right]^n d\theta = \int_0^{\frac{T^2}{2\gamma_n^2}} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta} d\theta = 1 - \exp\left(-\frac{v^2}{2\gamma_n^2}\right) \qquad (52)$$

where $v = T/\sigma$. This suggests that the approximation of (51) is accurate for the approximate computation of the integral of (52). Since (52) approximates the probability that the block is not significant, the probability that the block is significant is of the form:

$$\Pr\{\mathrm{sig}(v,n) = 1\} \approx \exp\left(-\frac{v^2}{2\gamma_n^2}\right) \qquad (53)$$

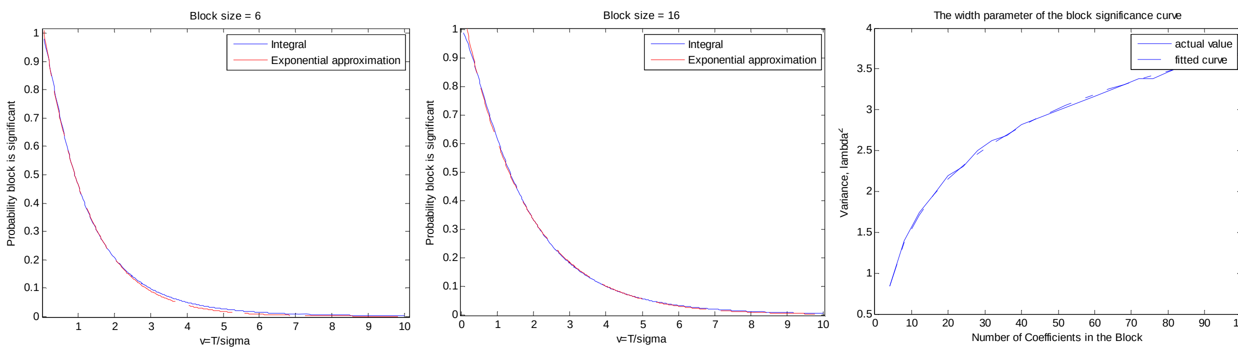

Figure 11. Left two plots: The integrated value of (47) versus the approximation in (53) for different block sizes. Right plot: The variance γ_n^2 in terms of the number of coefficients in the block.

In the left two plots of Figure 11, $\Pr\{\mathrm{sig}(v,n) = 1\}$ and its approximation are plotted over various v and block sizes $n = 6$ and $n = 16$, which are typical minimum block sizes in quadtree-based coders [20]. As can be seen, an approximation of the form in (53) is highly accurate. In order to determine a good estimate for γ_n , γ_n is fitted for blocks of coefficients for which block significance makes a significant contribution to rate or complexity (e.g $n \in [4,100]$). By estimating the mean squared value of v from $\Pr\{\mathrm{sig}(\frac{T}{\sigma},n) = 1\}$, γ_n is found to increase monotonically with the dimension of $\mathbf{X}$, i.e. an approximation of the form n^α is appropriate. We fitted for α by finding the best MSE match that returns a linear plot for $e^{\gamma_n/\alpha}$ in terms of n. The obtained approximation was:

$$\gamma_n^2 \cong \frac{1}{2}\left(\ln(n)^{1.296} - 0.166\right) \qquad (54)$$

The explicit derivation of $\chi_{v_b,k}^{\mathrm{high}}$ given in (9) then follows from (53) with the approximation of γ_n^2 given in (54). A comparison of the estimated (via (54)) versus the experimentally-derived value of γ_n computed with numerical methods is shown in the right plot of Figure 11. ■

Appendix B: Derivation of Probability of a Newly Significant Block in a High-Frequency Subband

We begin this section with several lemmas.

Lemma 1: For all real numbers $n \geq 2$,

$$\tfrac{0.5}{n} \leq 1 - \sqrt[n]{0.5} < \tfrac{1}{n} \tag{55}$$

***Proof*:** Rearranging the inequalities, we get $(1 - \frac{0.5}{n})^n \geq 1 - 0.5 > (1 - \frac{1}{n})^n$. These inequalities can be proven by expanding the binomials [31]. ■

Lemma 2: Let $\delta \in \mathbb{R}^+$ and $\delta \to 0$. There exists $N \in \mathbb{Z}^+$, such that for all integers $n \geq N$ and $x \in \mathbb{R}^+$, if $(\mathrm{erf}(x))^n \geq \delta$ then $(\mathrm{erf}(2x))^n \geq 1 - \delta$.

***Proof*:** We prove the existence of the lower bound $N \in \mathbb{Z}^+$ under the following equivalent condition: if $\mathrm{erf}(x) \geq \sqrt[n]{\delta}$ then $\mathrm{erf}(2x) \geq \sqrt[n]{1-\delta}$.

For $\delta \to 0$, we first bound $\mathrm{erfc}(2x)$ by the following inequality:

$$\mathrm{erfc}(2x) = 1 - \mathrm{erf}(2x) = \sqrt{\frac{2}{\pi}} \int_{2x}^{\infty} e^{-t^2} dt = 2\sqrt{\frac{2}{\pi}} \int_{x}^{\infty} e^{-4t^2} dt = 2\sqrt{\frac{2}{\pi}} \int_{x}^{\infty} e^{-t^2 - 3t^2} dt < 2e^{-3x^2}\, \mathrm{erfc}(x) \tag{56}$$

Let $m > 1$, satisfying: $(1-\delta)^m = \delta$. Choose $N_1 \geq 0$, such that for $x_{\mathrm{th1}} = \mathrm{erf}^{-1}\left(\sqrt[N_1]{\delta}\right)$:

$$e^{-3x_{\mathrm{th1}}^2} = \frac{1}{4m} \tag{57}$$

Combining (56) and (57), we have:

$$\mathrm{erfc}(2x_{\mathrm{th1}}) < 2e^{-3x_{\mathrm{th1}}^2}\left(1 - \mathrm{erf}(x_{\mathrm{th1}})\right) = \frac{1 - \sqrt[N_1]{\delta}}{2m} \tag{58}$$

Now, choose $N_2 \in \mathbb{R}^+$ such that $\sqrt[N_2]{(1-\delta)^m} = 0.5$. Setting $N = \lceil \max(N_1, 2N_2) \rceil$. Applying Lemma 1 and the fact that for any $n > N$, $\frac{n}{N_2} \geq 2$, we have the following two inequalities:

$$1 - \sqrt[n]{(1-\delta)} = 1 - \left(\sqrt[N_2]{(1-\delta)^m}\right)^{\frac{N_2}{mn}} = 1 - 0.5^{\frac{N_2}{mn}} \geq \frac{N_2}{2mn} \tag{59}$$

$$1 - \sqrt[n]{\delta} = 1 - \left(\sqrt[N_2]{(1-\delta)^m}\right)^{\frac{N_2}{n}} = 1 - 0.5^{\frac{N_2}{n}} < \frac{N_2}{n} \tag{60}$$

Now, let $x = \mathrm{erf}^{-1}(\sqrt[n]{\delta})$. Notice that $x \geq x_{\mathrm{th1}}$ since $\sqrt[n]{\delta} \geq \sqrt[N_1]{\delta}$, and $\mathrm{erf}^{-1}(\bullet)$ is monotonically increasing. Thus, $e^{-3x^2} < e^{-3x_{\mathrm{th1}}^2} = \dfrac{1}{4m}$. Combining (56), (57), (59), and (60), we get the following:

$$\mathrm{erfc}(2x) < 2e^{-3x^2}\,\mathrm{erfc}(x) < \frac{1}{2m}\mathrm{erfc}(x) < \frac{1 - \sqrt[n]{\delta}}{2m} < \frac{N_2}{2mn} \leq 1 - \sqrt[n]{(1-\delta)} \tag{61}$$

Thus $\mathrm{erf}(2x) \geq \sqrt[n]{1-\delta}$. The proof follows for any $x > \mathrm{erf}^{-1}(\sqrt[n]{\delta})$ by the fact that $\mathrm{erf}(2x)$ is an increasing function. ■

Lemma 3: There exists N, $N > 0$, such that for all $n > N$:

$$\left(\frac{\mathrm{erf}(x)}{\mathrm{erf}(2x)}\right)^n \cong (\mathrm{erf}(x))^n \tag{62}$$

*Proof***:** Let $\delta \to 0$, and choose a corresponding N such that Lemma 2 holds, i.e. for $x_{\text{th}} = \text{erf}^{-1}(\sqrt[n]{\delta})$, we have the following relations:

$$(\text{erf}(x_{\text{th}}))^n = \delta \tag{63}$$

$$(\text{erf}(2x_{\text{th}}))^n \geq 1-\delta \tag{64}$$

For all $x \geq x_{\text{th}}$, we have:

$$(\text{erf}(x))^n < \left(\frac{\text{erf}(x)}{\text{erf}(2x)}\right)^n \leq \frac{(\text{erf}(x))^n}{1-\delta} \tag{65}$$

Letting $\delta \to 0$ and applying the squeezing theorem [32], (62) holds for x above the threshold value.

For the $x \leq x_{\text{th}}$ case, we first note that $\left(\frac{\text{erf}(x_{\text{th}})}{\text{erf}(2x_{\text{th}})}\right)^n \leq \frac{\delta}{1-\delta}$, which approaches 0 as $\delta \to 0$. In order to show that $\left(\frac{\text{erf}(x)}{\text{erf}(2x)}\right)^n \to 0$ for all $x \leq x_{\text{th}}$, we need only show that $\frac{\text{erf}(x)}{\text{erf}(2x)}$ is monotonically increasing for all $x > 0$. This corresponds to a non-negative derivative for $\frac{\text{erf}(x)}{\text{erf}(2x)}$:

$$\frac{d}{dx}\left(\frac{\text{erf}(x)}{\text{erf}(2x)}\right) = \frac{\text{erf}(2x)\frac{2}{\sqrt{\pi}}e^{-x^2} - \text{erf}(x)\frac{4}{\sqrt{\pi}}e^{-4x^2}}{(\text{erf}(2x))^2} \geq 0 \tag{66}$$

The rewritten inequality can be easily verified, as shown below:

$$\text{erf}(2x) = \sqrt{\frac{2}{\pi}}\int_0^{2x} e^{-t^2}dt = 2\sqrt{\frac{2}{\pi}}\int_0^{x} e^{-4t^2}dt = 2\sqrt{\frac{2}{\pi}}\int_0^{x} e^{-t^2-3t^2}dt > 2e^{-3x^2}\,\text{erf}(x) \tag{67}$$

Thus for all $0 < x \leq x_{\text{th}}$, $0 < \left(\frac{\text{erf}(x)}{\text{erf}(2x)}\right)^n \leq \left(\frac{\text{erf}(x_{\text{th}})}{\text{erf}(2x_{\text{th}})}\right)^n \leq \frac{\delta}{1-\delta}$, and applying the squeezing theorem [32] once again gives us the desired approximation. ■

We can now conclude with the proof of Proposition 2.

Proof of Proposition 2: Consider a block that is insignificant in comparison to threshold T_{b+1}. The conditional probability that the block is also insignificant compared to threshold T_b is of the form:

$$\Pr\{\text{sig}(\tfrac{T_b}{\sigma},n) = 0 \cap \text{sig}(\tfrac{T_{b+1}}{\sigma},n) = 0\} = \int_0^\infty \frac{1}{\sigma^2}e^{-\frac{1}{\sigma^2}\theta}\left(\frac{\text{erf}(\frac{T_b}{\sqrt{2\theta}})}{\text{erf}(\frac{T_{b+1}}{\sqrt{2\theta}})}\right)^n d\theta \tag{68}$$

where $\left[\frac{\text{erf}(T_b/\sqrt{2\theta})}{\text{erf}(T_{b+1}/\sqrt{2\theta})}\right]^n$ is derived in a similar manner as in Appendix A, and represents the probability that, within a block of n Gaussian-distributed coefficients with variances θ (doubly-stochastic model), all coefficients will be insignificant compared to T_b, given that they were insignificant compared to T_{b+1}. The probability that a coefficient is found newly significant is then:

$$\begin{aligned}
\Pr\{\text{sig}(\tfrac{T_b}{\sigma},n) = 1 \cap \text{sig}(\tfrac{T_{b+1}}{\sigma},n) = 0\} &= \Pr\{\text{sig}(\tfrac{T_{b+1}}{\sigma},n) = 0\}\cdot\Pr\{\text{sig}(\tfrac{T_b}{\sigma},n) = 1 \mid \text{sig}(\tfrac{T_{b+1}}{\sigma},n) = 0\} \\
&= \Pr\{\text{sig}(\tfrac{T_{b+1}}{\sigma},n) = 0\}\cdot(1 - \Pr\{\text{sig}(\tfrac{T_b}{\sigma},n) = 0 \mid \text{sig}(\tfrac{T_{b+1}}{\sigma},n) = 0\} \\
&= \left(\int_0^\infty \frac{1}{\sigma^2}e^{-\frac{1}{\sigma^2}\theta}\left[\text{erf}\left(\tfrac{T_{b+1}}{\sqrt{2\theta}}\right)\right]^n d\theta\right)\left(1 - \int_0^\infty \frac{1}{\sigma^2}\exp(-\tfrac{1}{\sigma^2}\theta)\left[\text{erf}\left(\tfrac{T_b}{\sqrt{2\theta}}\right)\Big/\text{erf}\left(\tfrac{T_{b+1}}{\sqrt{2\theta}}\right)\right]^n d\theta\right)
\end{aligned} \tag{69}$$

For the first multiplicative term in (69), we use the approximation of (52). For the second multiplicative term, we use Lemma 3 under the assumption of appropriately-large n, and then apply the approximation of (52). The final estimate for (69) is:

$$\Pr\{\mathrm{sig}(\tfrac{T_b}{\sigma},n)=1\cap\mathrm{sig}(\tfrac{T_{b+1}}{\sigma},n)=0\}\approx\left[1-\exp\left(-\frac{v^2}{2\gamma_n^2}\right)\right]\exp\left(-\frac{v^2}{2\gamma_n^2}\right)\tag{70}$$

■

References

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circ. Syst. for Video Technol.*, vol. 13, no. 7, July 2003.

[2] J.-R. Ohm, "Advances in scalable video coding," *Proc. of the IEEE*, vol. 93, pp. 42-56, Jan. 2005.

[3] D. G. Sachs, S. V. Adve, and D. L. Jones, "Cross-layer adaptive video coding to reduce energy on general purpose processors," *Proc. IEEE Int. Conf. on Image Proc.*, ICIP 2003, vol. , pp. 25-28, Sept. 2003.

[4] M. Horowitz, A. Joch, F. Kossentini, and A. Hallapuro, "H.264/AVC baseline profile decoder complexity analysis," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 13, no. 7, pp. 704-716, July 2003.

[5] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: Tools, performance, and complexity," *IEEE Circ. and Syst. Mag.*, vol. 4, no. 1, pp. 7-28, Jan. 2004.

[6] J. Valentim, et al, "Evaluating MPEG-4 video decoding complexity for an alternative video verifier complexity model", *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 12, no. 11, pp. 1034-1044, Nov. 2002.

[7] D. S. Turaga, M. van der Schaar, B. Pesquet-Popescu, "Complexity-scalable motion compensated wavelet video encoding," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 15, no. 8, pp. 982-993, Aug. 2005.

[8] W. Yuan and K. Nahrstedt, "Energy-efficient CPU scheduling for multimedia applications," *ACM Trans. on Computer Syst.*, Vol. 24, Issue 3, Aug. 2006.

[9] M. Wang, M. van der Schaar, "Operational rate-distortion modeling for wavelet video coders," *IEEE Trans. on Signal Proc.,* Vol. 54, Issue 9, Sept. 2006.

[10] M. van der Schaar and Y. Andreopoulos, "Rate-distortion-complexity modeling for network and receiver aware adaptation," *IEEE Trans. on Multimedia*, vol. 7, no. 3, pp. 471-479, June 2005.

[11] Y. Andreopoulos and M. van der Schaar, "Complexity-constrained video bitstream shaping," *IEEE Trans. on Signal Processing*, to appear, preprint available at: www.ee.ucla.edu/~yandreop .

[12] Z. He, and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. on Circ. and Syst. for Video Technol.*, vol. 11, no. 12, Dec. 2001.

[13] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application – Part I: Fundamental theory," *IEEE Trans. on Circ. and Syst. for Video Technol.*, vol. 7,no. 2,pp. 287-298,Apr. 1997.

[14] Z. He, Y. Liang, L. Chen, I. Ahmad and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. Circuits and Syst. for Video Technol.*, vol. 15, no. 5, pp. 645-658, May 2005.

[15] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," Signal Processing: Image Communication, vol. 19, no. 7, pp. 561-575.

[16] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Trans. on Image Proc.*, vol. 9, no. 2, pp. 173-183, Feb. 2000.

[17] K. Stuhlmuller, N. Farber, M. Link and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Select. Areas in Comm.*, vol. 18, no. 6, pp. 1012-1032, June 2000.

[18] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. on Circ. and Syst. for Video Technol.*, vol. 6, pp. 12-20, Feb. 1996.

[19] D. Taubman, M. W. Marcellin, JPEG 2000-Image Compression Fundamentals, Standards and Practice, *Kluwer Academic Publishers*, 2002.

[20] P. Schelkens, A. Munteanu, J. Barbarien, M. Galca, X. Giro-Nieto, J. Cornelis, "Wavelet coding of volumetric medical datasets," *IEEE Transactions on Medical Imaging*,vol. 22,no. 3,pp. 441-458,Mar. 2003.

[21] R. Shukla, P. L. Dragotti, M. N. Do, and M. Vetterli, "Rate-distortion optimized tree-structured compression

algorithms for piecewise polynomial images," *IEEE Trans. on Image Proc.*, vol. 14, no. 3, Mar. 2005.

[22] P. Chen and J. W. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Trans. on Circ. and Syst. for Video Technol.*, vol. 14, no. 10, pp. 1183-1194, Oct. 2004.

[23] A. Munteanu, J. Cornelis, G. Van der Auwera and P. Cristea, "Wavelet image compression – the quadtree coding approach," *IEEE Trans. on Information Technol. in Biomed.*, vol. 3, no. 3, pp. 176-185, Sept. 1999.

[24] C. Chrysafis, et al, "SBHP – a low complexity wavelet coder," Proc. IEEE Internat. Conf. on Acoust. Speech and Signal Process., ICASSP '00, vol. 6, pp. 2035-2038, June 2000.

[25] J. Xu, Z. Xiong, S. Li and Y. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *Appl. Comput. Harmon. Anal.*, vol. 10, pp. 290-315, 2001.

[26] J. Liu, P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. on Image Processing*, vol. 10, pp. 164701658, Nov., 2001.

[27] M. K. Mihçak, et al, "Low-complexity Image Denoising based on Statistical Modeling of Wavelet Coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300-303, 1999.

[28] S. Mallat, F. Falzon, "Analysis of low bit-rate image transform coding," *IEEE Trans. on Signal Processing*, vol. 46, pp. 1027-1042, Apr., 1998.

[29] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis and P. Schelkens, "In-band motion compensated temporal filtering," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 653-673, Aug. 2004.

[30] T. Rusert, K. Hanke, and J. Ohm, "Transition filtering and optimization quantization in interframe wavelet video coding," *VCIP, Proc. SPIE*, vol. 5150, pp. 682-693, 2003.

[31] P. Borwein, T. Erdelyi, Polynomials and polynomial inequalities, *Springer*, New York, 1995.

[32] C. Hummel, "Chapter IV: The Squeezing Theorem," in: Gromov's Compactness Theorem for Pseudo-Holomorphic Curves, *Spinger*, Series: Progress in Mathematics, vol. 151, 1997.

[33] D. Marpe, et al, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, Issue 7, July 2003.

[34] C. Chrysafis, A. Ortega, "Efficient context-based entropy coding for lossy wavelet image compression," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, 1997, pp. 221-230.

[35] Y. Andreopoulos and M. van der Schaar, "Adaptive linear prediction for resource estimation of video decoding," *IEEE Trans. on Circuits and Systems for Video Technology*, to appear.

[36] E. Lam, J. Goodman, "A Mathematical Analysis of the DCT Coefficient Distributions for Images," *IEEE Transactions on Image Processing*, Vol. 9, Issue 10, Oct. 2000.