A Joint Source-Channel Distortion Model for JPEG Compressed Images

Muhammad F. Sabir, *Student Member, IEEE,* Hamid R. Sheikh, *Member, IEEE,* Robert W. Heath Jr., *Member, IEEE,* Alan C. Bovik, *Fellow, IEEE,*

Abstract

The need for efficient joint source-channel coding (JSCC) is growing as new multimedia services are introduced in commercial wireless communication systems. An important component of practical JSCC schemes is a distortion model that can predict the quality of compressed digital multimedia such as images and videos. The usual approach in the JSCC literature for quantifying the distortion due to quantization and channel errors is to estimate it for *each* image using the statistics of the image for a given signal to noise ratio (SNR). This is not an efficient approach in the design of real-time systems because of the computational complexity. A more useful and practical approach would be to design JSCC techniques that minimize *average* distortion for a large set of images based on some distortion model rather than carrying out *per-image* optimizations. However, models for estimating average distortion due to quantization and channel bit errors in a combined fashion for a large set of images are not available for practical image or video coding standards employing entropy coding and differential coding. This paper presents a statistical model for estimating the distortion introduced in progressive JPEG compressed images due to quantization and channel bit errors in a joint manner. Statistical modelling of important compression techniques such as Huffman coding, differential pulse coding modulation (DPCM), and run-length coding are included in the model. Examples show that the distortion in terms of peak signal to noise ratio (PSNR) can be predicted within a 2 dB maximum error over a variety of compression ratios and bit error rates. To illustrate the utility of the proposed model, we present an unequal power

Muhammad. F. Sabir is affiliated with the Laboratory for Image and Video Engineering, Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA, email: mfsabir@ece.utexas.edu

Hamid R. Sheikh is affiliated with Texas Instruments Inc, Dallas, TX, USA. He was previously affiliated with the Laboratory for Image and Video Engineering, Department of Electrical & Computer Engineering, The University of Texas at Austin, USA. Phone: (469) 467-7947, email: hamid.sheikh@ieee.org

Robert W. Heath Jr. is affiliated with the Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA, Phone: (512) 232-2014, email:rheath@ece.utexas.edu

Alan. C. Bovik is affiliated with the Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA, Phone: (512) 471-5370, email:bovik@ece.utexas.edu

This work was supported by the Texas Advanced Technology Program Grant No. $003658 - 0380 - 2003$.

allocation scheme as a simple application of our model. Results show that it gives a PSNR gain of around 6.5 dB at low SNRs as compared to equal power allocation.

Index Terms

Joint Source-Channel Coding, Distortion Model, JPEG, Unequal Power Allocation.

I. INTRODUCTION

With the introduction of real-time imaging in third generation wireless systems, it is clear that high bandwidth multimedia will be one of the main applications in next generation wireless systems. These multimedia sources require high bandwidth to be transmitted with high fidelity. However, in practical systems there is always a constraint on resources, especially bandwidth. The most common way of reducing the bandwidth requirements of these sources is to use source coding, also known as data compression. Source coding can either be lossless or lossy. Lossy compression not only introduces distortion in the coded source in the form of quantization errors, but also makes the resulting data stream highly sensitive to bit errors when transmitted over noisy/fading channels. These bit errors could introduce large amounts of perceptual distortion in the transmitted data. Channel coding (error protection) is the most common method of protecting the transmitted data against channel bit errors by introducing redundancy, thereby causing bandwidth expansion. Therefore, we are faced with contrasting goals of minimizing the bandwidth as well as the distortions due to quantization and transmission errors.

According to Shannon's classical separation theorem [1], the source coding and the channel coding can be performed separately and sequentially while still maintaining optimality. However, this is true only for the case of asymptotically long block lengths, which is not the case in most practical situations. Hence, in practical coding and transmission systems, a joint design and optimization of source and channel codes may reduce distortion while providing a bandwidth efficient solution. This approach is known as joint source-channel coding (JSCC). In JSCC, the goal is to optimize the use of available resources for source and channel coding while minimizing the end-to-end distortion. A significant reduction in transmission bandwidth requirements can be achieved without introducing a large amount of perceptual distortion in the received data by using JSCC methods. JSCC has gained significant research attention during the last decade, particularly since the Internet revolution. A large proportion of this work has been devoted to digital images and videos due to the large bandwidth required to achieve a reasonable quality of service. There are many different approaches to JSCC. One of the most common approaches is to divide the available bit rate amongst the source and the channel encoders in an optimized manner, such that the

distortion is minimized [2], [3]. Another well-known approach to JSCC is that of unequal error protection (UEP) for different parts of the source coded stream based on their sensitivity to channel errors [4], [5]. In another common approach to JSCC, the encoder/decoder pair is optimized to minimize the distortion in the presence of a noisy channel [6].

A major step in designing a JSCC scheme is to model the distortion introduced in the received image due to quantization and channel errors. This distortion is then used either as an objective function to be minimized or as a constraint in the design of a JSCC scheme. We formulate one such distortion model in this paper that can predict the amount of distortion introduced in JPEG [7] compressed images due to quantization and channel errors. In order to motivate the need for such a model, we discuss a few key JSCC techniques for image and video transmission, and their corresponding distortion metrics and models in the next section.

A. Previous Work

The works by Modestino *et al.* for JSCC of two-dimensional (2D) differential pulse coded modulation (DPCM) coded images [2], and block cosine transform (BCT) coded images [3], are considered to be classical techniques for JSCC of digital images. They estimated the distortion due to quantization and channel errors in terms of mean squared error (MSE) using the probability density functions of the DPCM or BCT coded source, the quantizer step size, and the channel probability of error, which were all assumed to be known. Using this estimate of MSE, and the input source energy, selective error protection was provided to those bits that contributed the most to image reconstruction. Rate-distortion (RD) bounds were derived, and it was demonstrated that through channel coding, the reconstructed image quality was significantly improved without sacrificing transmission bandwidth, and the performance achieved approached the RD bounds. In another JSCC scheme described in [8], the authors have described how quantization and channel errors affect the overall distortion for different source coding schemes. They derived an expression for distortion (MSE) as a function of the number of quantization bits used, and the specific channel codes employed for transmission over an additive white Gaussian noise (AWGN) channel at specific values of SNR. Individual bit sensitivities to channel errors were computed, and joint optimization of source and channel coding was carried out using these bit sensitivities. In [9], the authors provided UEP to a discrete cosine transform (DCT) and variable length coding (VLC) based source coder using Blokh-Zyablov (BZ) channel coding. An expression representing the sensitivity of individual bits to a bit error in a coded block was developed, which was obtained by systematically placing an error on each bit position of all coded frames. These bit sensitivities were then used to provide UEP to

different bits. In another bit sensitivity based UEP method, Comstock and Gibson [10] demonstrated the use of Hamming codes to protect the most important bits for transmission of DCT compressed images over noisy channels. A closed form expression for MSE, based on individual bit sensitivity was derived, which was then used to determine how many (and which) bits to protect in each block.

In their classical paper, Farvardin and Vaishampayan [6] presented joint design of an optimal encoder decoder pair for transmission over noisy channels. A locally optimum quantizer and block coder that minimizes MSE was designed iteratively. It was shown that this design technique offered substantial improvement for generalized Gaussian (GG) distributions over Lloyd-Max quantization using natural and folded binary codes. In [11], Tanabe and Farvardin proposed subband coding techniques for noisy and noiseless channels employing uniform threshold quantization (UTQ), Huffman coding, and error control coding (ECC). A simulation based procedure was described for determining the best (UTQ, HC, ECC) triple for encoding a memoryless source over a binary symmetric channel (BSC) at a given encoding rate. Distortion rate curves were constructed, which were then used for optimum bit allocation between different subbands. In [12], Chande and Farvardin proposed a JSCC scheme for progressive transmission of embedded source coded images using embedded error correcting and detecting channel codes over noisy channels. Average distortion (MSE), average peak signal-to-noise ratio (PSNR), and average useful source coding rate were considered as the cost functions, which were then used to develop algorithms for optimum allocation of source and channel coding bits.

As another interesting approach for JSCC, Sherwood and Zeger [13] demonstrated the use of concatenated channel codes for transmission of 'progressively' coded images over noisy channels. They also presented an UEP scheme based on product channel codes for progressive image transmission over memoryless and fading channels in [14]. As an extension to this, Cosman *et al.* [15] combined the concatenated forward error control (FEC) approach with a zerotree wavelet packetization method [16] to provide more robust performance. In [17], Goldsmith and Effros presented an iterative joint design technique for channel-optimized vector quantizer (COVQ), and rate compatible punctured convolutional (RCPC) codes [18]. Kozintsev and Ramchandran [19] proposed a method for joint design of multiresolution source codebook, multiresolution constellation, and decoding strategy for a wavelet image decomposition based source coding model. A Lagrangian based algorithm was developed to accomplish the joint optimization. A JSCC scheme for run-length coded subbands was presented in [20]. In addition, the authors have derived an analytical expression for estimating the distortion introduced in run-length coded data due to channel errors using Bose-Chaudhuri-Hochquenghem (BCH) codes over a BSC channel.

In [21], [22], two methods of bit allocation between source and channel coders for video transmission

over noisy channels are discussed. A novel technique for minimizing the energy required to transmit video under distortion and delay constraints is presented in [23]. Additional relevant JSCC techniques for image and video transmission are discussed in [5], [24]–[32].

B. Limitations of Existing Schemes

All the above mentioned JSCC schemes provide significant coding gains and improved image and video quality over schemes that do not optimize the source and the channel coders in a joint manner. However, there are a few limitations of these schemes that we would like to point out.

- Most of these JSCC techniques do not take into account entropy coding, with a few exceptions ([9], [11], [20]). Entropy coding is an integral part of all image and video coding standards such as JPEG, JPEG2000, MPEG-4, etc. Because of entropy coding, a bit error in the coded bitstream is propagated, and the exact contribution of a bit error to the overall distortion is hard to determine exactly at the time of encoding.
- Another limitation of these JSCC techniques is their high dependence on image statistics and channel conditions. Since the optimizations in these JSCC techniques are performed on a *per-image* basis, the parameters derived for a particular image can only be used to transmit that image. The results of using the same parameters to transmit any other image could be poor. Similarly, in scenarios where the channel is varying, such as a fading channel, the results could be extremely poor due to channel mismatch, as most of these schemes are designed for a *particular* SNR. One example of such a scenario can be real-time imaging applications, such as new camera cell phones, where pictures are captured and transmitted in real-time over fading channels.
- Since the optimization procedures associated with these JSCC schemes must be carried out for different images and varying channel conditions, the computational complexity of these schemes becomes prohibitive for real-time applications. The design process is expensive to perform in real-time with the currently available processing power of low power digital signal processors.

Note that although the JSCC schemes designed on a per-image basis would result in lower end-to-end distortion as compared to schemes minimizing average distortion over a set of images, they cannot be used for practical real-time image and video communication systems due to their computational complexity. For practical real-time communication systems, JSCC schemes with low computational complexity are needed. JSCC schemes that minimize average distortion ‘off-line’ over a ‘set of images’ for different channel conditions, and then simply switch between different coding parameters in real-time, depending

on the channel condition, would have low computational complexity as no real-time optimization would be needed. Hence such JSCC schemes would be well-suited for real-time image and video communication.

C. Proposed Work

We address the limitations of existing schemes outlined in the previous section, and propose a distortion model in this paper that predicts the amount of distortion introduced in a 'set of images', rather than a particular image, based on the source coding rate and channel bit error probability when the images are transmitted over noisy/fading channels. The distortion model presented in this paper predicts the distortion due to quantization and channel errors in a combined manner, taking into account the effects of DPCM and entropy coding, thereby modelling the effects of error propagation within the coded coefficients as well.

Although our proposed framework is designed for *progressive* JPEG compressed images, it can be used for any similar coefficient based image or video coding scheme employing DPCM and entropy coding with appropriate modifications. The model works for any memoryless channel such as BSC, AWGN and Rayleigh fading channels. We derive the parameters of our distortion model from a database of randomly selected photographic images. A separate *test* database of randomly selected photographic images is used to verify the performance of our model using simulations. Results show that the values of PSNR predicted using our distortion model fall within 2 dB (and within 1 dB for most practical coding configurations) of those obtained using simulations over the test database. This shows that our model can predict average distortion due to quantization and channel errors with reasonable accuracy.

To our knowledge no model has been developed to date that takes into account such practical coding issues as entropy coding and DPCM coding, and that estimates the average distortion for a set of images. We believe that efficient JSCC schemes could be designed based on this distortion model that do not require real-time optimizations based on individual image statistics or varying channel conditions, as average distortion could be predicted 'off-line' for an ensemble of images and channel conditions. Using this distortion model, the parameters of the JSCC schemes could be obtained off-line for different channel conditions, which in turn could be used for low-complexity real-time coding and transmission. As a very simple example, we present an unequal power allocation scheme based on our distortion model for transmission of JPEG compressed images over Rayleigh fading channels.

In Section II we outline our system model consisting of the source encoder, decoder, and the channel. In Section III, we develop the distortion model, and simulations and results are presented along with some discussion in Section IV. In Section V, an unequal power allocation application for the distortion

model is discussed. We conclude this paper in Section VI.

II. The System Model

Source coding often makes the compressed bitstream highly sensitive to channel errors. A single bit error has the potential to corrupt an entire image or video sequence, especially in the presence of entropy coding. Therefore, error-resilient features are required in multimedia communication systems. In this paper, we derive our model for use with JPEG with certain error-resilient features. Specifically, we use RST (reset) markers in progressive DCT based mode and assume that the headers are not corrupted by any bit error. The headers, which constitute only a small proportion of the overall data, can be transmitted error-free using powerful error correcting codes. These features are important for constructing any practical image communication system based on the JPEG algorithm.

A. The Source Coding Model

In our JPEG encoding, we use the progressive DCT based mode of operation with spectral-selection [7]. In the progressive DCT mode, the data is arranged in different quality layers. In the spectral-selection method, the DCT coefficients are quantized and divided into subbands that are encoded in separate passes. The DC coefficients are DPCM and entropy coded in the first pass, followed by run-length and entropy encoding of AC coefficients for different subbands in subsequent passes. We use Huffman coding for the model presented in this paper, however, this model would also work for arithmetic coding. Due to the presence of entropy coding, the encoded bitstream becomes highly sensitive to bit errors because of error propagation in codewords. Although we derive our model for grayscale JPEG, it could be easily extended for color images by treating errors in the individual components independently.

In the model discussed in this paper, the 64 subbands of DCT coefficients are organized into 64 separate layers: the first one being the DC layer, followed by 63 AC layers. In this way, the resolution and the quality of the decoded image improves as more layers are decoded. RST markers are inserted in each layer regularly. We call the portion of each layer between two consecutive RST markers a *segment*. Note that in AC layers, run-length coding is done across blocks in each segment in each layer. The structure of the JPEG compressed stream is shown in Fig. 1. Decoding is reinitialized whenever a RST marker is encountered [7], and a bit error occurring in a segment only corrupts that segment, and the error is not propagated beyond that segment. We discuss error detection in the following section.

Error detection: Errors can be detected in a compressed image/video bitstream either at the transport coder, or at the source decoder [33]. A common way to perform error detection at the transport coder

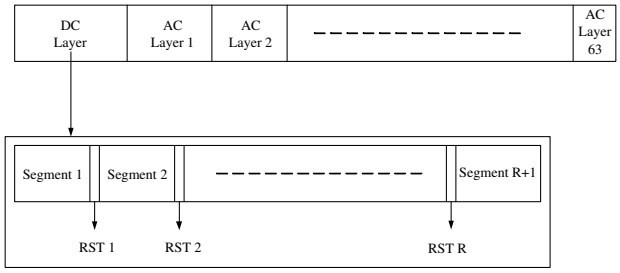


Fig. 1. JPEG layers and segments. Layer headers are not shown for simplicity.

is the addition of header information to the bitstream. An example of such a system can be a packet based video transmission system, where the output of the video encoder is divided into packets. These packets each contain a header and a payload field [34]. Each of the headers can have a header number, which is sequential. The decoder checks for these packet numbers, and in the case of a missing or an out of order packet, declares a packet error. Another well known method for performing error detection at the transport coder is to use error control coding. In this method, forward error correction codes (FEC) are used to add redundancy to the compressed bitstream in order to protect it from channel errors. A significant amount of research has been done in the field of error control coding for image and video transmission during the last two decades. A few examples can be found in [4], [14], [26], [29], [35].

Error detection can also be performed at the source decoder. There can be many different methods to accomplish this. For example, for detection of transmission errors in DPCM coded images, Rose and Heiman [36] compared the difference in pixel values between two neighboring lines to a threshold, and discarded the image segment if this difference is larger than the threshold. Another interesting scheme for error detection in the frequency domain was proposed by Lam and Reibman in [37].

In the distortion model presented in this paper, we do not use any packetization or error control coding to protect the entropy coded codewords corresponding to the DC and the AC layers. In the case of bit errors occurring in the transmitted bitstream, we assume that the decoder detects the first bit error (due to loss in synchronization of entropy decoding) and decodes all the coefficients in the rest of the segment as zero. All the coefficients before the codeword in error are decoded correctly, since a Huffman code is a prefix code. The assumption of bit error detection is based on the fact that in most cases the Huffman code being used is not a complete code, i.e. not all possible codewords are legitimate. Therefore, if a bit error occurs such that the resulting codeword is not in the decoding table, the decoder will declare an error [33]. It is possible that the codeword generated as a result of a bit error is legitimate, in which case the bit error will go undetected by the decoder, resulting in erroneous decoding of coefficients. However, it is highly likely that the decoder will detect the bit error since only a few codewords are in

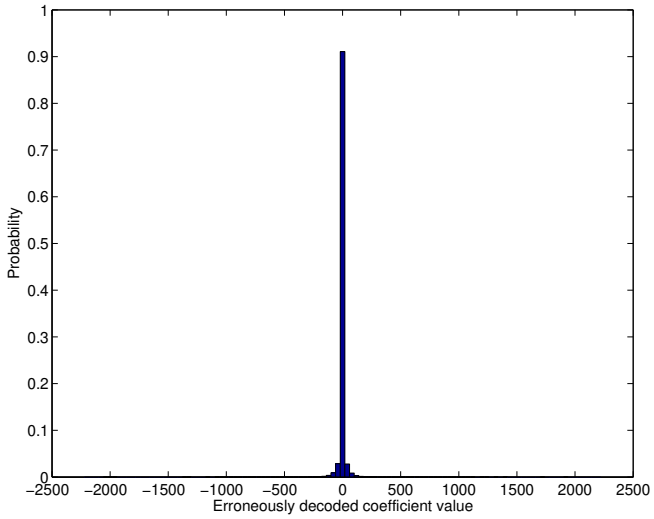


Fig. 2. Empirical distribution of decoded coefficient values in the case of bit errors.

the decoding table. The number of illegitimate codewords of different lengths depends on Huffman table, which is different for different images. In order to find out how often does the decoder detect bit errors, we introduced a random single bit error in different segments of 200 progressively coded JPEG images. Our decoder declared an error whenever it encountered a codeword that was not in the decoding table. The case when the decoder detects the error in exactly the same coefficient that was actually corrupted by the bit error in our simulations was termed as 'successful detection'. The cases when the bit error goes undetected or is detected at a later point were termed as 'unsuccessful detection'. 100000 iterations were performed over each image and the percentage of successful detection was found to be 80.4%. Note that these simulations were performed over a range of source coding rates (0.6 - 3 bpp), employing a variety of Huffman tables and codeword lengths, and 80.4% is the overall percentage of successful detection over this range of source coding rate. We also plotted the empirical distribution of decoded coefficient values that are affected by the bit error, as shown in Fig. 2. As the figure shows, the erroneous coefficients are zero with a probability of 0.91 in our simulations. We are getting this higher probability as compared to the successful detection percentage because it also includes the case when the decoder detects an error at a later point (and decodes the remaining coefficients as zeros). The coefficients decoded as non-zero represent the case when the bit error is not detected by the decoder. Since the percentage of successful error detection by the decoder is quite high (80.4%), we can fairly assume that the decoder detects the first error, to simplify our distortion model.

Error free transmission of headers and markers: Bit errors occurring during transmission can affect the headers, the markers, the DC layer, and the AC layers. If there is an error in the header, the entire

image will be damaged seriously, and it is very likely that the image cannot be recovered. In case of errors in RST markers, synchronization will be lost, corrupting the image until the next correctly received RST marker. In our paper, we only model the effects of quantization and transmission errors in the entropy coded DC and AC layers, and not in the headers and markers. We can assume that the headers and markers are transmitted error free, since they only constitute a small portion of the overall bitstream (appx. $4-5$ %), and hence powerful channel codes along with packetization and intelligent automatic repeat request (ARQ) techniques can be used to protect them. The approach of assuming error free transmission of headers, markers and a base layer (a basic representation of images and videos) is quite common in joint source-channel coding literature, and a few examples are given in [9], [29], [38], [39].

In a JPEG compressed bitstream, the $\mathrm{RST_m}$ markers incorporate a modulo-8 count, m. This count is started at zero for each scan and is incremented by one with the addition of each RST marker to the compressed bitstream. These modulo-8 RST markers are represented by two byte codes codes X 'FFD0' - X 'FFD7'. These codes are unique by default in the JPEG standard and do not occur anywhere else in the bitstream. Instead of transmitting these RST markers in the above mentioned format, we transmit segment lengths in bytes (similar to the scheme used by Fazel and Lhuillier in [9]) to specify the exact locations of RST markers to the decoder. These lengths are each represented by two bytes and are sent along with other headers and markers at the beginning of the bitstream before transmitting the entropy coded image data. Note that we transmit segment lengths instead of RST markers so that the decoder can reinsert the RST markers at their exact locations to achieve perfect synchronization. We call these headers, markers and segment lengths 'critical layer', as their error-free transmission is critical for image reconstruction. Since this critical layer only constitutes a small portion of the entire bitstream (appx. $4-5$ %), powerful channel codes can be used to guarantee its error free delivery at practical SNRs. In addition to channel coding, some other techniques such as packetization and ARQ can also be used to guarantee error free transmission of this critical layer. This layer is then parsed at the decoder, and all the headers and RST markers along with other markers are inserted at their original positions in the bitstream before initializing decoding. This way, the decoder knows the exact location of RST markers and hence perfect synchronization is achieved. Note that transmitting segment lengths is just one way to achieve perfect synchronization. There can be many other ways, depending on the particular application.

B. The Channel Model

We consider the channel to be a BSC, and derive our distortion model for a given bit error probability p_e. We assume that the bit errors are independent. Both the AWGN and the Rayleigh fading (with interleaving)

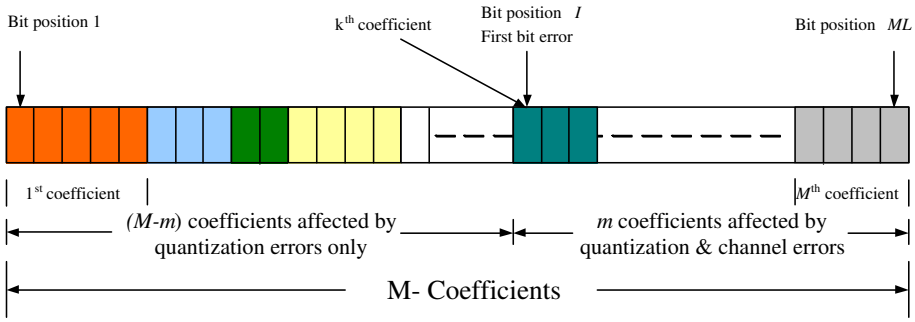


Fig. 3. Bit error in a segment. Different colors/shades represent different entropy coded coefficients.

channels can be represented as a BSC, given the bit error probabilities for these channels and the fact that the probability of making an error from 0 to 1 is the same as that of 1 to 0. Therefore, using this model we can find the distortion curves for any channel that can be represented as a BSC, if the source coding rate and the expression for bit error rate (BER) are known [40]. This makes our distortion model independent of modulation type and channel coding. We do not consider any packetization and encapsulation for the entropy coded raw bitstream, and it is assumed that the bitstream is directly transmitted over the channel without any overhead by streaming protocols. No error concealment is considered at the decoder.

III. The Distortion Model

In this section we derive our model for distortion in the grayscale JPEG image due to quantization and channel errors. We use MSE as our distortion metric.

A. Assumptions and Notation

Our goal is to relate average image distortion, in terms of MSE, to source coding rate and bit error probability. This modelling is complicated due to the presence of Huffman coding, DPCM coding and run-length coding. A single bit error can cause the decoder to lose synchronization and corrupt the entire segment. Furthermore, it is difficult to precisely determine the coefficient position corresponding to a particular bit in error due to the different lengths of the entropy coded symbols.

Since our model consists of 64 layers, we will first outline our notation only for one layer in this section, and later generalize it for all the 64 layers. Therefore, in the following discussion, all variables meant are for one layer only.

In this paper, we make certain simplifying assumptions that allow us to derive the average MSE due to quantization and channel errors over a set of images. Specifically, we will model DCT coefficients as random processes that are wide sense stationary and ergodic. A RST marker is inserted in each layer after every M coefficients, which is constant for all layers. Since by assumption the first bit error in

a segment corrupts the entire segment from the bit in error to the next RST marker, we only need to consider the position I of the *first* bit in error, which we also model as a random variable. Thus multiple errors in the same codeword as well as all subsequent errors in the same segment, can be ignored. We index the coefficient to which the I^{th} bit belongs as k. All coefficients from the k^{th} coefficient to the end of the segment are corrupted, and assumed to be decoded as zero, while all previous coefficients from the first coefficient of the segment to the $(k-1)^{st}$ coefficient are assumed to be decoded correctly. We start by describing the distribution for the first bit error position I in a segment, denoted by $p_I(i)$. We assume that all the bits are corrupted independently with a bit error probability p_e, which is assumed to be constant for a layer. Then, $p_I(i)$ is given by

$$p_I(i) = p_e(1-p_e)^{i-1}. \tag{1}$$

Hence, an increase in the bit error rate will increase the probability of the first bit error occurring at an earlier location in a segment.

Consider a segment with M DCT coefficients indexed from 1 to M, as shown in Fig. 3. Note that this figure is a simplified representation of the actual bitstream. Each colored block is an entropy coded codeword and can represent one or more coefficients. Particularly in the case of AC layers, each block can represent a run-length coded codeword. We assume that the *average* length of a coded coefficient is a continuous random quantity denoted by L. We use the term average length because due to entropy coding it is difficult to model the exact number of bits used to represent individual coefficients. The distribution of L depends upon the source coding rate. On average, the coefficient index k corresponding to the first bit in error is given by $k \approx \lceil I/L \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function. Note that this relation holds only on *average*, and the distortion model based on it would only predict the *average* MSE over a set of images. The number of coefficients in a segment corrupted by a bit error is denoted by $m(L, I)$, or m for notational convenience. Thus $M - m$[1] coefficients are decoded correctly, while m coefficients are corrupted and decoded as zero. This is depicted in Fig. 3. We will first derive an expression for MSE conditioned on the knowledge of L and I, and later average over L and I. Since in JPEG encoding, the DC coefficients are DPCM coded, while the AC coefficients are coded directly, MSE needs to be modelled separately for DC and AC layers.

We will generalize our notation for each layer since our model consists of 64 layers. We will represent I, L, m and k for the n^{th} layer as I_n, L_n, m_n and k_n respectively, where the layer number n ranges from 0 to 63, with 0 for the DC layer and 63 for the highest frequency AC layer. Note that m_n and k_n

[1] M, m and k are obviously related: $m = M - k + 1,\ k = 1...M$

Notation	Description
M	The total number of coefficients in a segment. It is a constant for all layers in the image.
N	The total number of pixels in the image.
L_n	Average length of the coded coefficients in the n^{th} layer.
I_n	Position of the first bit error in a segment of the n^{th} layer.
k_n	Coefficient index to which the bit I_n belongs. It is a function of L_n and I_n.
m_n	The number of coefficients affected by quantization and channel errors in a segment in the n^{th} layer. It is a function of L_n and I_n.
$X^u_{k_n}$	k^{th} unquantized DCT coefficient in a segment of the n^{th} layer.
$X^q_{k_n}$	k^{th} quantized DCT coefficient in a segment of the n^{th} layer.
$\widehat{X}^q_{k_n}$	k^{th} quantized and erroneously decoded DCT coefficient in a segment of the n^{th} layer.
$\widehat{P}^q_{k_0}$	Erroneously decoded prediction value corresponding to the k^{th} quantized DC coefficient (DPCM coding).
ξ_{k_n}	Quantization error in the k^{th} DCT coefficient in a segment of the n^{th} layer.
$\sigma^2_{u_n}$	Variance of the unquantized DCT coefficients in the n^{th} layer.
$\sigma^2_{q_n}$	Variance of the quantized DCT coefficients in the n^{th} layer.
$\sigma^2_{\xi_n}$	Quantization error variance for the n^{th} layer.
MSE^{k_n}	MSE between the original and the received image, due to a bit error in the k^{th} coefficient in a segment of the n^{th} layer, without taking into account the effect of error propagation in the subsequent coefficients. Note that this is the value of MSE due to error in one coefficient (k^{th}) only.
MSE_{m_n}	MSE between the original and the received image, due to a bit error in the k^{th} coefficient in a segment of the n^{th} layer, taking into account the effect of error propagation in all the subsequent coefficients in the segment. Note that this is the value of MSE due to errors in m coefficients.
MSE_{M_n}	MSE between the original and the received image due to quantization errors in $M-m$ coefficients, and quantization and channel errors in m coefficients in a segment of the n^{th} layer.
MSE_{DC}	The total MSE between the original and the received image due to quantization and channel errors in all the segments of the DC layer.
MSE_{AC_n}	The total MSE between the original and the received image due to quantization and channel errors in all the segments of the n^{th} AC layer ($n = 1...63$).
MSE	The total MSE between the original and the received image due to quantization and channel errors in all the layers.

TABLE I

NOTATION.

are functions of L_n and I_n. To avoid confusion, we list our notation in Table I.

B. Distortion Model for the DC Layer ($n = 0$)

In this section, we will derive an expression for the expected value of MSE due to quantization and channel errors in the DC layer. Suppose the first bit error in a segment occurs at bit position I_0, corresponding to the coefficient indexed k_0. Due to orthonormality of DCT, we can analyze MSE in the DCT domain. Then, the average MSE (over an ensemble of images) between the original and the erroneous image due to error in the k_0^{th} coefficient, given L_0 and I_0, can be written as:

$$E\left(MSE^{k_0} \mid L_0, I_0\right) = \frac{1}{N} E\left[(X_{k_0}^u - \widehat{X}_{k_0}^q)^2 \mid L_0, I_0\right] \tag{2}$$

where the random variables $X_{k_0}^u$ and $\widehat{X}_{k_0}^q$ represent the unquantized and the erroneously decoded quantized DC coefficients respectively, and N is the total number of pixels in the entire image. We divide by N because we are considering the MSE in the entire image due to one coefficient in error. Equation (2) depends on L_0 and I_0, since k_0 is a function of L_0 and I_0. Note that we have not yet considered the effect of error propagation in a segment in (2).

In progressive JPEG, all DC coefficients, except the first coefficient, are DPCM coded, and only the DC *prediction* is represented in the bitstream. Thus, for all the coefficients other than the first DC coefficient:

$$\widehat{X}_{k_0}^q = X_{k_0-1}^q + \widehat{P}_{k_0}^q,\ k_0 = 2...M$$

where $X_{k_0-1}^q$ is the correctly decoded quantized coefficient at position k_0-1, while $\widehat{P}_{k_0}^q$ is the erroneously decoded prediction value. By assumption, the decoder detects the error and decodes $\widehat{P}_{k_0}^q$ as zero. Thus

$$\widehat{X}_{k_0}^q = X_{k_0-1}^q. \tag{3}$$

The quantized DC coefficient $X_{k_0}^q$ can be expressed as the sum of the unquantized coefficient $X_{k_0}^u$ and a quantization error ξ_{k_0}

$$X_{k_0}^q = X_{k_0}^u + \xi_{k_0}. \tag{4}$$

Since in the JPEG standard, the pixel values are level shifted to a signed representation by subtracting $2^{(Z-1)}$, where Z is the precision specified for the input data [7], before taking the DCT, we can fairly assume all the DCT coefficients to be zero mean. Now, representing the variances of the unquantized DC coefficients, the quantized DC coefficients, and the quantization error by $\sigma_{u_0}^2$, $\sigma_{q_0}^2$, and $\sigma_{\xi_0}^2$ respectively,

and expanding (2) using (3) and (4), the expected value of MSE for $k_0 = 2...M$ can be written as

$$\begin{aligned} E(MSE^{k_0}|L_0, I_0) &= \frac{1}{N}\left(E\left[(X^u_{k_0})^2 + (\widehat{X}^q_{k_0})^2 - 2X^u_{k_0}\widehat{X}^q_{k_0}|L_0, I_0\right]\right) \\ &= \frac{1}{N}\left(\sigma^2_{u_0} + \sigma^2_{q_0} - 2E\left[X^u_{k_0}X^u_{k_0-1} \mid L_0, I_0\right]\right. \\ &\quad \left. - 2E\left[X^u_{k_0}\xi_{k_0-1} \mid L_0, I_0\right]\right). \end{aligned} \tag{5}$$

We can fairly assume that the quantization error is uncorrelated with the unquantized DC coefficients; i.e. $E\left[X^u_{k_0}\xi_{k_0+i}\right] = 0\ \forall i$, and hence $\sigma^2_{q_0} = \sigma^2_{u_0} + \sigma^2_{\xi_0}$, where $X^u_{k_0}$ and ξ_{k_0} are zero mean. Hence the expected value of MSE (for $k_0 = 2...M$) becomes

$$E(MSE^{k_0}|L_0, I_0) = \frac{1}{N}\left(2\sigma^2_{u_0} + \sigma^2_{\xi_0} - 2r\left(1\right)\right), \tag{6}$$

where $r(1)$ is the autocorrelation function of the DC coefficients at lag 1. Using a similar methodology, MSE at a distance of $j-1$ coefficients from the k_0^{th} coefficient can be written as

$$E(MSE^{k_0+j-1}|L_0, I_0) = \frac{1}{N}(2\sigma^2_{u_0} + \sigma^2_{\xi_0} - 2r(j)),\ k_0 = 2...M. \tag{7}$$

As shown in Fig. 3, a bit error in the k_0^{th} coefficient will result in the loss of m_0 coefficients ($m_0 = M - k_0 + 1$). Now, assuming additivity, the expected value of MSE ($E(MSE_{m_0}|L_0, I_0)$) due to these m_0 coefficients for the case $m_0 = 1...M-1$ ($k_0 = 2...M$) can be written as

$$E(MSE_{m_0}|L_0, I_0) = \frac{1}{N}\left[m_0 \cdot (2\sigma^2_{u_0} + \sigma^2_{\xi_0}) - 2\sum_{j=1}^{m_0} r(j)\right],\ m_0 = 1...M-1.$$

Now, suppose a bit error occurs in the first coefficient ($k_0 = 1,\ m_0 = M$), then all the M coefficients will be corrupted. Hence, by our assumption, all the coefficients in the segment will be decoded as zero. Therefore, substituting $\widehat{X}^q_{k_0} = 0$ in (2), the expected value of MSE for $m_0 = M(k_0 = 1)$ becomes

$$E(MSE_{m_0}|L_0, I_0) = \frac{1}{N}\left(M\sigma^2_{u_0}\right),\ m_0 = M.$$

Hence, the expected value of the MSE due to m_0 coefficients in error can be written as:

$$E(MSE_{m_0}|L_0, I_0) = \begin{cases} \frac{1}{N}\left[m_0 \cdot (2\sigma^2_{u_0} + \sigma^2_{\xi_0}) - 2\sum\limits_{j=1}^{m_0} a^{|j|}\sigma^2_{u_0}\right], & m_0 = 1...M-1 \\ \frac{1}{N}\left(M\sigma^2_{u_0}\right), & m_0 = M \end{cases} \tag{8}$$

where we have used a first order auto-regressive model for the DC coefficients $\left(r(j) = a^{|j|}\sigma^2_{u_0}\right)$. Note that (8) only represents the effects of quantization and channel errors in the m_0 coefficients corrupted by channel errors. We only need to model quantization error for the remaining $M - m_0$ coefficients since they are not corrupted by channel errors. Adding this quantization distortion, the expected value of MSE becomes

$$E(MSE_M|L_0, I_0) = \frac{1}{N}\left[(M - m_0)\sigma^2_{\xi_0}\right] + E(MSE_{m_0}|L_0, I_0). \tag{9}$$

Now taking the expectation over I_0, we get

$$E(MSE_M|L_0=l_0) = \sum_{i_0=1}^{\lceil Ml_0\rceil}\left[\frac{M-m_0}{N}\sigma^2_{\xi_0}+E(MSE_{m_0}|L_0=l_0,I_0=i_0)\right]p_{I_0}(i_0), \quad (10)$$

where the ceiling appears in the summation because L_0 is the average length of the DC coefficients, which we modelled as a continuous random variable discussed in Section IV-A. We also need to consider the event when no bit error occurs in the entire segment and the distortion is solely due to quantization. The probability of such an event given L_0 is $p(\text{No error}|L_0=l_0)=1-\sum_{i_0=1}^{\lceil Ml_0\rceil}p_{I_0}(i_0)$, and the corresponding distortion is $\frac{1}{N}M\sigma^2_{\xi_0}$. Including this in our expression for MSE, taking expectation over L_0, and assuming the additivity of MSE due to all the $R+1$ segments in the DC layer, the expected value of MSE due to the DC layer becomes

$$E(MSE_{DC})=(R+1)\int_{\mathbb{R}^+}\left[\frac{1}{N}\left(1-\sum_{i_0=1}^{\lceil Ml_0\rceil}p_{I_0}(i_0)\right)M\sigma^2_{\xi_0}+E(MSE_M|L_0=l_0)\right]p_{L_0}(l_0)\mathrm{d}l_0, \quad (11)$$

where $p_{L_0}(l_0)$ is the probability density function (pdf) of L_0. We will present a model for $p_{L_n}(l_n)$ in Section IV-A.

This expression models the expected value of MSE between the original and the received image due to quantization and random bit errors in all the segments of the DC layer. If the DC layer is not transmitted, then it is obvious that $E(MSE_{DC})=\frac{M(R+1)}{N}\sigma^2_{u_0}$.

C. Distortion Model for the AC Layers ($n=1...63$)

The 63 AC subbands in the JPEG compressed image constitute the next 63 quality layers in our model. Similar to the case of DC layer, the expected value of MSE (over an ensemble of images) between the original and the erroneous image due to error in the k_n^{th} coefficient given L_n and I_n can be written as:

$$\begin{aligned} E\left(MSE^{k_n}\mid L_n,I_n\right) &= \frac{1}{N}E\left[(X^u_{k_n}-\widehat{X}^q_{k_n})^2\mid L_n,I_n\right] \\ &= \frac{1}{N}\left(E[(X^u_{k_n})^2\mid L_n,I_n]\right)=\frac{1}{N}\sigma^2_{u_n} \end{aligned} \quad (12)$$

where $X^u_{k_n}$ and $\widehat{X}^q_{k_n}$ are the k^{th} unquantized and erroneously decoded quantized coefficients of the n^{th} AC layer, and $\sigma^2_{u_n}$ is the variance of the zero-mean unquantized AC coefficients for the n^{th} AC layer. Equation (12) results from our assumption that the decoder decodes $\widehat{X}^q_{k_n}$ as zero, and from the fact that the AC coefficients are not DPCM coded.

Following similar steps as for the DC layer, the expected value of MSE due to the n^{th} AC layer can

be shown to be

$$E(MSE_{AC_n}) = \frac{R+1}{N}\int_{\mathbb{R}^+}\left[\sum_{i_n=1}^{\lceil Ml_n\rceil}\left((M-m_n)\sigma_{\xi_n}^2+m_n\sigma_{u_n}^2\right)p_{I_n}(i_n) + \left(1-\sum_{i_n=1}^{\lceil Ml_n\rceil}p_{I_n}(i_n)\right)M\sigma_{\xi_n}^2\right]p_{L_n}(l_n)dl_n. \quad (13)$$

where $\sigma_{\xi_n}^2$ is the variance of the quantization error for the n^{th} AC layer, and L_n models the average length of coded coefficients in the n^{th} AC layer. Note that there is no correlation term in the MSE expression for the AC layers. This is because the AC layers are not DPCM coded. Since we use a different distribution for the average coefficient length L_n for each layer, the effects of run-length and Huffman coding in the AC coefficients are incorporated automatically in our derivation.

This expression models the expected value of MSE between the original and the received image due to quantization and random bit errors in all the segments of the n^{th} AC layer. If the n^{th} AC layer is not transmitted, then $E(MSE_{AC_n}) = \frac{M(R+1)}{N}\sigma_{u_n}^2$.

D. Total Distortion

Assuming additivity of MSE due to individual layers (recall orthonormality of the DCT basis), the expected value of MSE can be written as

$$E(MSE) = E[MSE_{DC}] + \sum_{n=1}^{63}E[MSE_{AC_n}]. \quad (14)$$

This expression models the expected value of MSE (over an ensemble of images) between the original and the received image due to quantization and random bit errors in all the 64 coefficient layers.

IV. Simulations and Results

In this section we compare our model's prediction of average MSE against simulations. MSE is converted to PSNR assuming 8 bit unsigned representation for unquantized pixel values using the simple relation $\mathrm{PSNR} = 10\log_{10}\frac{255^2}{\mathrm{MSE}}$, since PSNR is commonly used for image quality assessment.

A. Model Parameters

In order to predict the MSE using (11), (13) and (14), certain model parameters are needed. These parameters consist of $\sigma_{u_n}^2$, $\sigma_{\xi_n}^2$, a, and distributions $p_{L_n}(l_n)$ for L_n ($n = 0...63$). We used a training database of randomly selected 200 grayscale photographic images, compressed at different source coding rates ranging from 0.6 bits per pixel (bpp) to 3 bpp, to derive these model parameters. The value of

a was found to be 0.94. We used a constant segment size of 64 coefficients ($M = 64$). We model L_n with a truncated Gaussian distribution (see appendix) as shown in Fig. 4, and fit the dependence of the parameters on the source coding rate. The parameters of the truncated Gaussian distribution, namely ρ_{L_n} and $\sigma^2_{L_n}$ were computed numerically using estimated mean and variance of L_n (see appendix). We used four-parameter logistic curve fitting to express ρ_{L_n} and $\sigma^2_{L_n}$ as functions of source coding rate for each of the 64 coefficient layers. The mapping function used is given in (15), while the fitting is done using MATLAB's *fminsearch*.

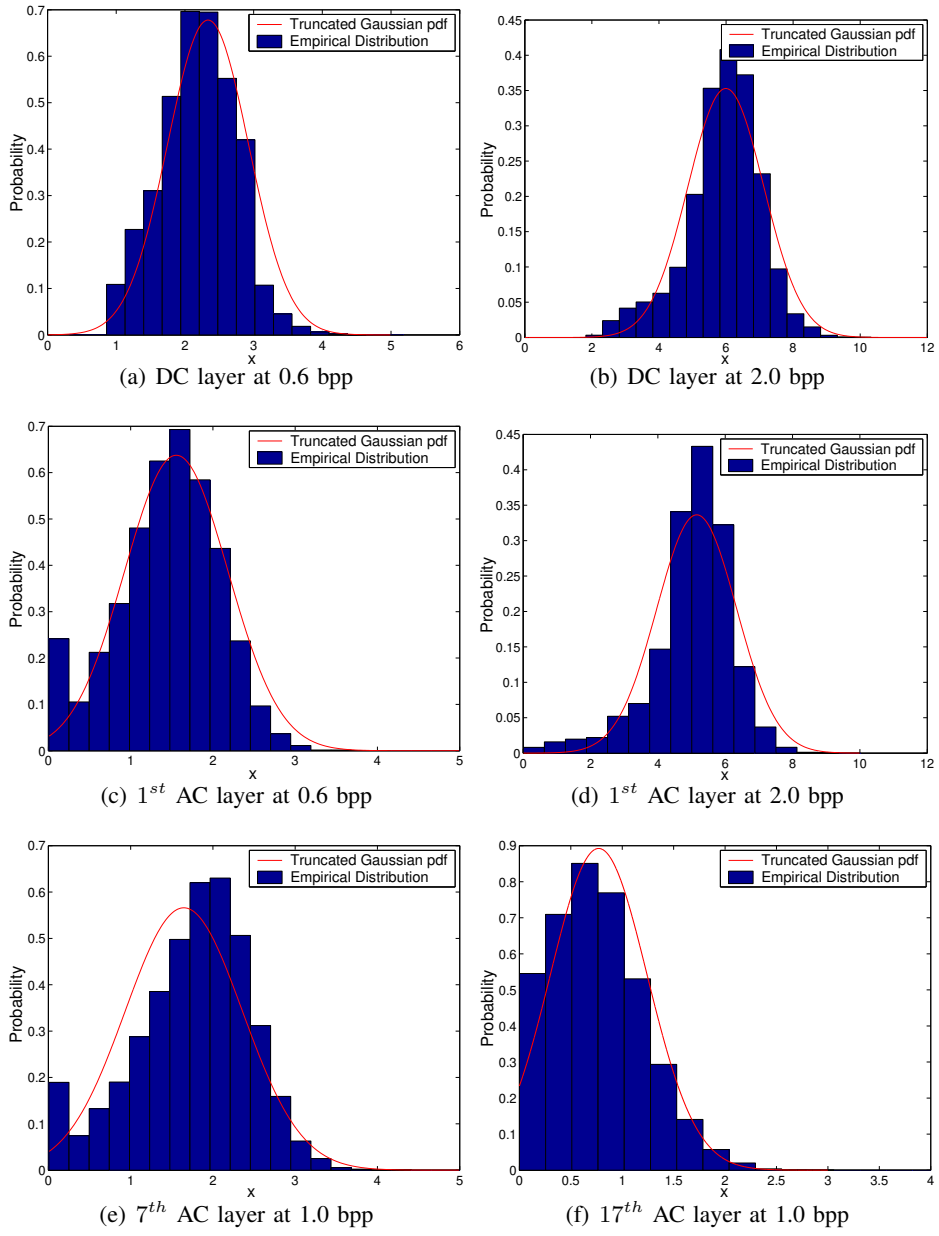
$$\Phi(x) = \beta_1 \text{logistic}\left(\beta_2, (x - \beta_3)\right) + \beta_4 \tag{15}$$

$$\text{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(\tau x)} \tag{16}$$

where x is the source coding rate in bits per pixel (bpp), and β_1, β_2, β_3, and β_4 are fitting parameters. Φ can be either ρ_{L_n}, or for the case of $\sigma^2_{L_n}$, we take the 'max' of 0 and Φ ($\sigma^2_{L_n} = \max(0, \Phi)$), since the variance cannot be negative. We used five-parameter logistic curve fitting given by (17) to model the natural $\log$ of quantization error variance $\sigma^2_{\xi_n}$ as a function of the source coding rate for each layer.

$$\ln\left(\sigma^2_{\xi_n}(x)\right) = \beta_1 \text{logistic}\left(\beta_2, (x - \beta_3)\right) + \beta_4 + \beta_5 x. \tag{17}$$

By using this curve fitting, the parameters of our model can be computed for any bit rate. Furthermore, since L_n is now a continuous function of the source coding rate, our expressions for MSE are also continuous functions of the source coding rate. Fig. 4 shows the empirical distributions of the coefficient lengths along with the approximated truncated Gaussian distributions. We noted that the truncated Gaussian approximation is a good fit for the empirical distribution for a variety of coefficient layers and source coding rates. At low bit rates, there is a peak in the first bin of AC layers. This is because many quantized AC coefficients are zero at low bit rates, and hence are coded very efficiently by run-length coding. Sometimes the entire coefficient layer is zero and hence represented by just one symbol, resulting in a peak near zero. However, it is still appropriate to use truncated Gaussian because we found it to be the best fit over a variety of distributions. The purpose of using a distribution fit instead of the actual empirical distribution was to reduce the parameters of the model, which we accomplished by using the truncated Gaussian distribution for all layers. Fig. 5 (a) - (d) show the curves for the parameters of the truncated Gaussian distributions (ρ_{L_n} and $\sigma^2_{L_n}$) as functions of the source coding rate along with the four-parameter logistic curve fitting. Fig. 5 (e) - (f) show the quantization error variance $\sigma^2_{\xi_n}$ as a function of the source coding rate along with the five-parameter logistic curve fitting. Note that these curves approximate the actual values quite accurately.



(a) DC layer at 0.6 bpp (b) DC layer at 2.0 bpp

(c) 1^{st} AC layer at 0.6 bpp (d) 1^{st} AC layer at 2.0 bpp

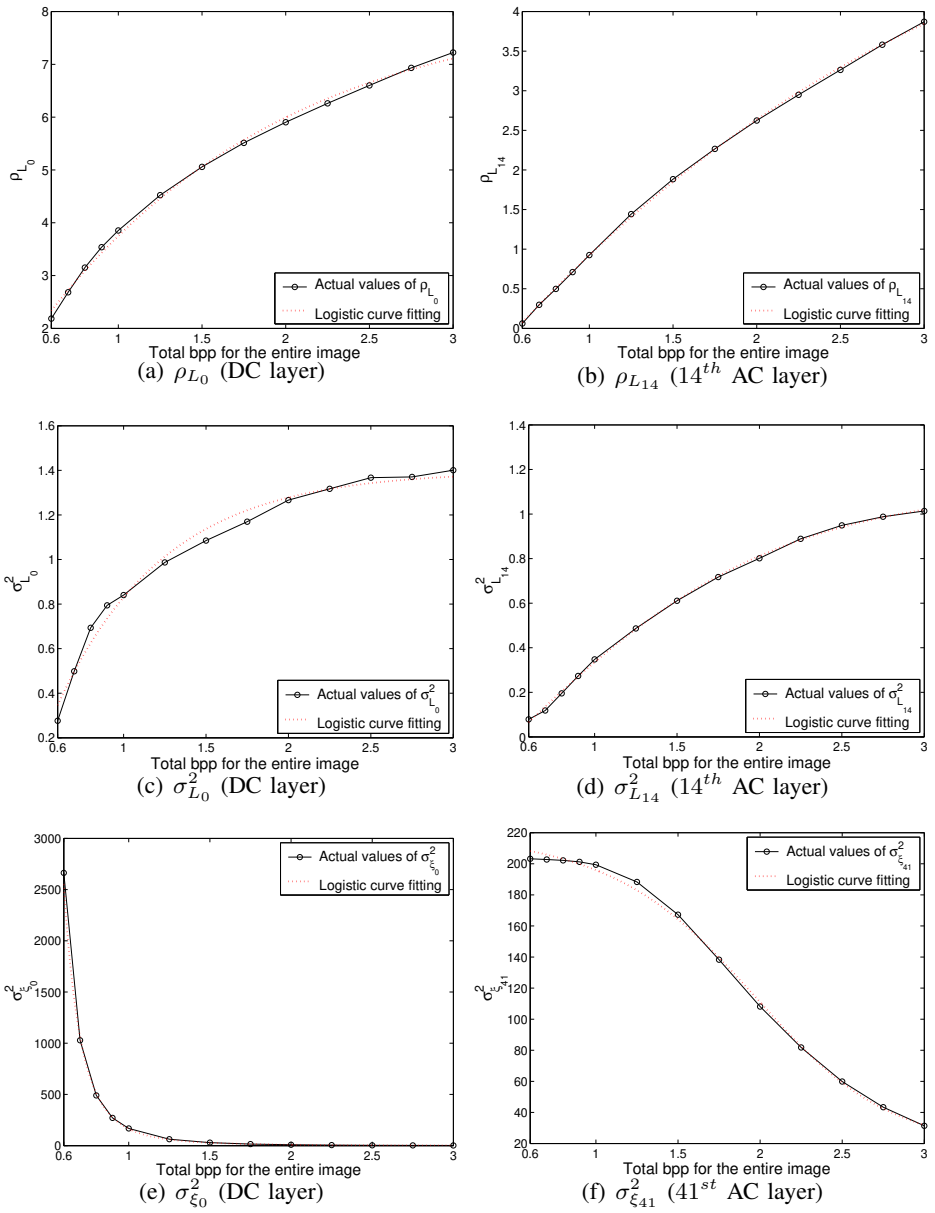
(e) 7^{th} AC layer at 1.0 bpp (f) 17^{th} AC layer at 1.0 bpp

Fig. 4. Empirical distributions of L_n and their corresponding truncated Gaussian approximations at different total source coding rates for the DC and different AC layers.

B. Simulation Details

Two different sets of 200 512×512 randomly selected grayscale photographic images from a database of over 2000 images[2] were used in the simulations, one for training and the other for testing. As discussed in the previous section, the training database was used to obtain the model parameters, which in turn

[2]We randomly selected our images from the two-CD set of 'Austin & Vicinity - The world of nature' and 'Austin and Vicinity - The human world'.



(a) ρ_{L_0} (DC layer) (b) $\rho_{L_{14}}$ (14^{th} AC layer)

(c) $\sigma^2_{L_0}$ (DC layer) (d) $\sigma^2_{L_{14}}$ (14^{th} AC layer)

(e) $\sigma^2_{\xi_0}$ (DC layer) (f) $\sigma^2_{\xi_{41}}$ (41^{st} AC layer)

Fig. 5. Model parameters as functions of source coding rate for different layers together with logistic curve fitting. ρ_{L_n} and $\sigma^2_{L_n}$ are truncated Gaussian parameters, and $\sigma^2_{\xi_n}$ is the quantization error variance.

were used to predict the average MSE as a function of the source coding rate (bpp) and channel BER. The test database was used to compute the actual average distortion introduced in the 'test' set of images by comparing the original images with the quantized and erroneous images at different source coding rates and bit error rates (BERs). Random bit errors were introduced in the compressed bitstream at the given BERs in order to simulate the effects of channel errors. RST markers were introduced after every 64 coefficients ($M = 64$). We used source coding rates from 0.6 bpp to 3 bpp, and BERs from 10^{-2} to

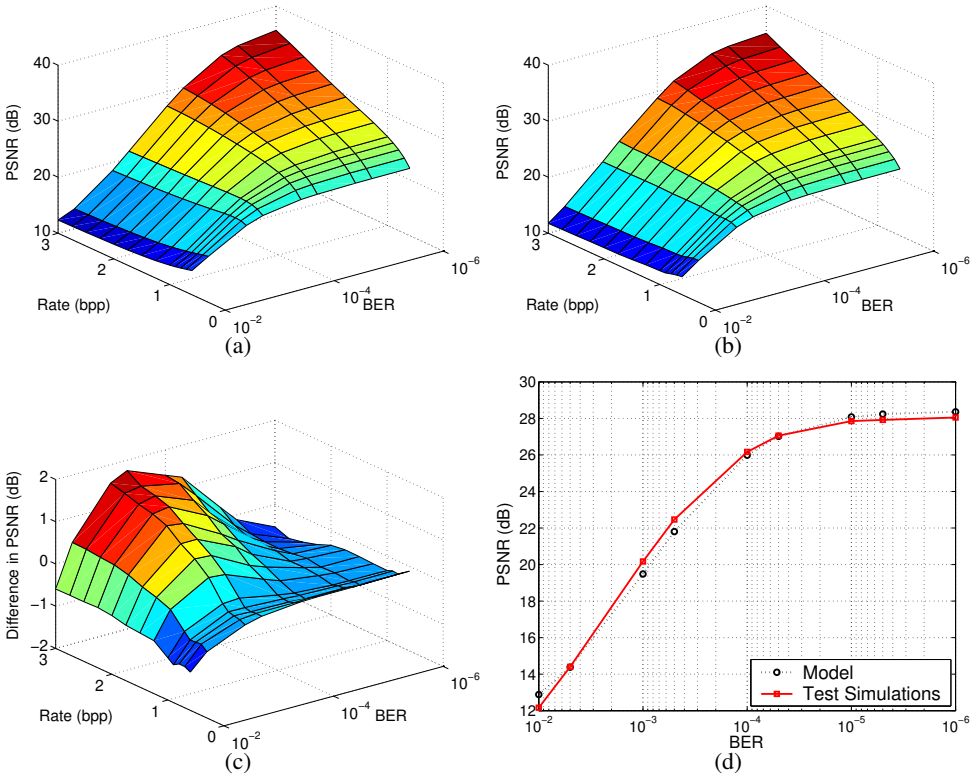


Fig. 6. PSNR vs. BER and bpp for quantization and bit errors in all the layers, obtained using the model and the test simulations. (a) Model (Eq. 14). (b) Simulations over the test database. (c) Difference between (a) and (b). (d) Slices of (a) and (b) at 1.25 bpp.

10^{-6} for our simulations. 20 channel instantiations were used for each image at each source coding rate and BER in order to compute the average MSE and PSNR. The average PSNR values obtained using simulations were then compared to those obtained using our model. During the simulations, the headers and the markers were separated and it was assumed that they were transmitted without errors. This is a valid assumption since powerful channel codes could be used to transmit the headers and markers, which constitute a very small portion of the overall bitstream, without any bit errors. We did not use any error concealment at the decoder.

C. Results and Discussion

Fig. 6, Fig. 8 and Fig. 7 show the PSNR curves obtained using the model and the test simulations, and their differences. The curves in Fig. 6 represent the PSNR obtained for the case of quantization and channel errors in all the 64 layers. The curve in Fig. 6 (a) is obtained using the MSE given by the model (14), whereas the curve in Fig. 6 (b) is obtained using simulations over the test database. Fig. 6 (c) shows the difference in PSNR obtained using the model and the test simulations, whereas Fig. 6 (d)

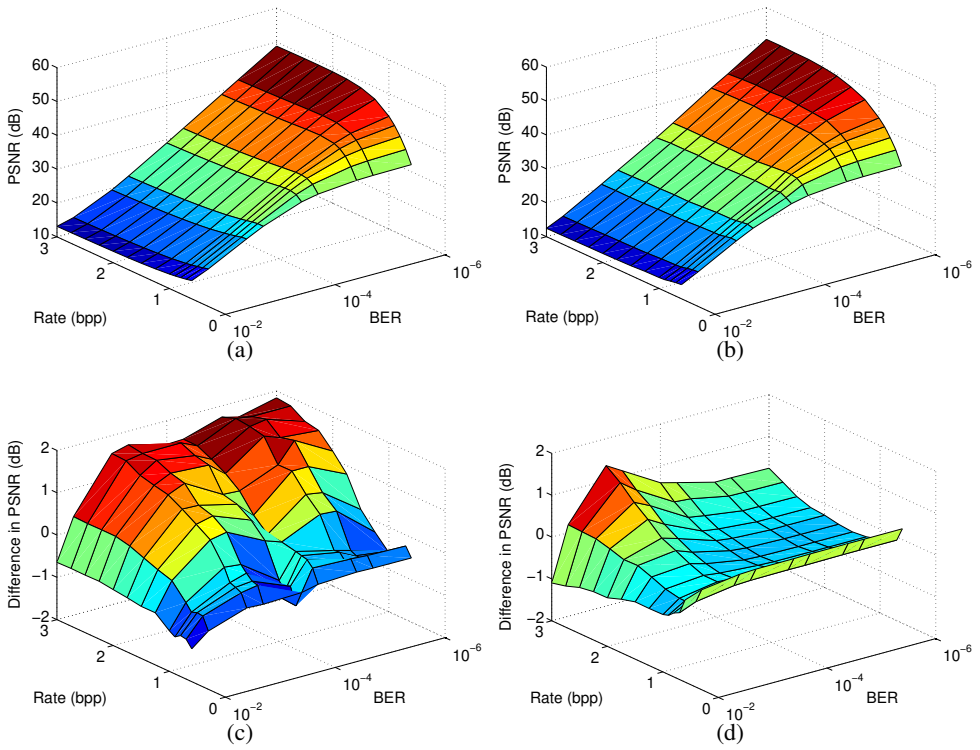


Fig. 7. PSNR vs. BER and bpp for quantization and bit errors in individual layers. (a) Model for the DC Layer (Eq. 11). (b) Test simulations for the DC Layer. (c) PSNR difference between the model (a) and the simulations (b) for the DC layer. (d) PSNR difference between the model and the simulations for the first AC layer.

shows a slice of Fig. 6 (a) and (b) at 1.25 bpp for comparison. Fig. 8 shows the difference in PSNR obtained using two different test simulations with 20 channel instantiations each. The PSNR curves in Fig. 7 are for the case of quantization and channel errors in individual layers. For the DC layer (a)-(c), we plot the results obtained using the model, the test simulations, and their difference; whereas for the first AC layer, we plot the difference in PSNR. Note that the PSNR values shown in Fig. 7 are for the entire image, however, the quantization and channel errors exist only in one layer (DC or AC), and not in any other layer. The source coding rate shown in Fig. 7 is for the entire image. These curves are useful in quantifying the deviation in the model from the actual test simulations for *individual* layers. Figs. 9 and 10 show original test images and their compressed and erroneous versions at different BERs. The long streaks in these images are due to the error propagation effects in the DPCM and entropy coded coefficients. These streaks show that our assumption that the decoder detects the first bit error and decodes all the subsequent coefficients in the segment as zeros, is quite valid.

As can be seen from these figures, our model predicts the distortion introduced in images due to quantization and channel errors with reasonable accuracy. The difference in the average PSNR obtained

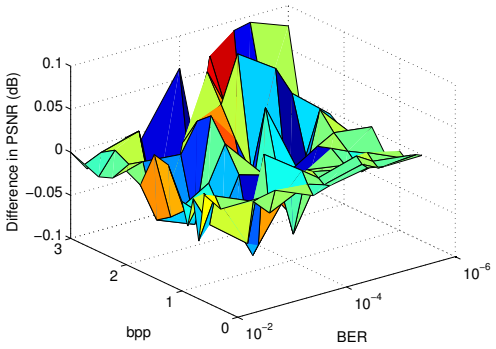
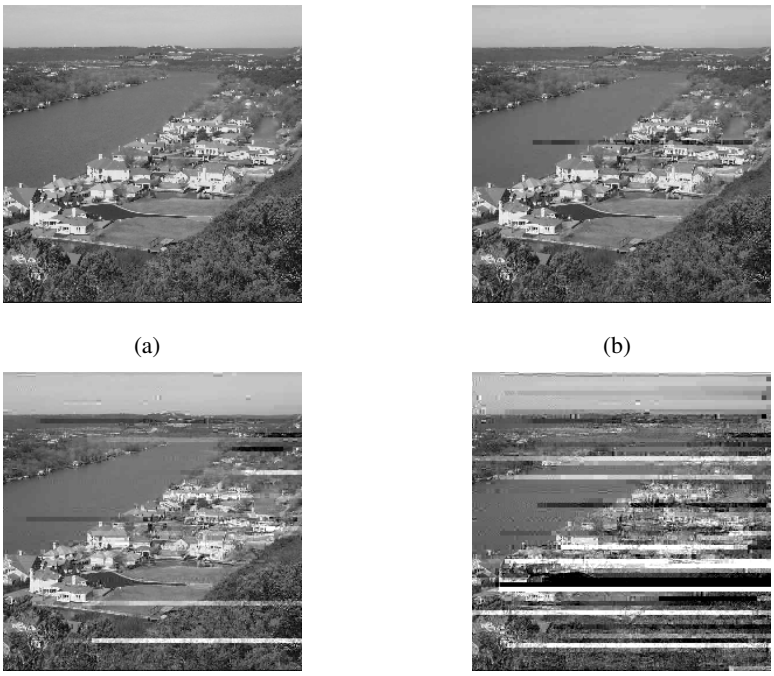


Fig. 8. Difference in PSNR obtained using two different test simulations with 20 channel instantiations each.

using the model and the test simulations is within 2 dB at all points as shown in Fig. 6 (c). For the case of individual layers, this difference is within 2 dB for the DC layer (Fig. 7 (c)), and 1.5 dB for the first AC layer (Fig. 7 (d)). This shows that our model predicts the average distortion very closely to that obtained using the simulations. This is also evident from the PSNR values shown in Figs. 9 and 10 for two test images along with those predicted by the model at different BERs. Also, Fig. 8 shows that the difference in PSNR between two different test simulations is within 0.1 dB at all points. This shows that the variance in our results is low.

V. Applications

The model presented in Sec. III can be used for a variety of applications. On one hand it can predict the quality (in terms of MSE and PSNR) of a JPEG compressed image when transmitted over a memoryless channel, and on the other hand it can be used to devise efficient JSCC techniques for image communication. The distortion introduced due to quantization and channel errors in individual coefficient layers can be predicted with reasonable accuracy, and the contribution of different layers to overall distortion can be used to provide different levels of error protection to these layers with the goal of minimizing the expected value of total distortion. A tradeoff between the source and the channel coding rates can also be performed using this model given a total bit-rate, with the expected value of MSE as the cost function. JSCC techniques could be designed off-line for a database of images and various channel conditions, with average distortion as the minimization criterion, and need not be re-optimized for each image. The parameters of these JSCC schemes could then be changed in real-time, suited for a particular channel condition without going through the optimization procedure again. This enables a significant reduction in the computational complexity as compared to existing 'per-image' based JSCC techniques. The per-image based JSCC techniques would perform better in terms of lower



(a) (b)

(c) (d)

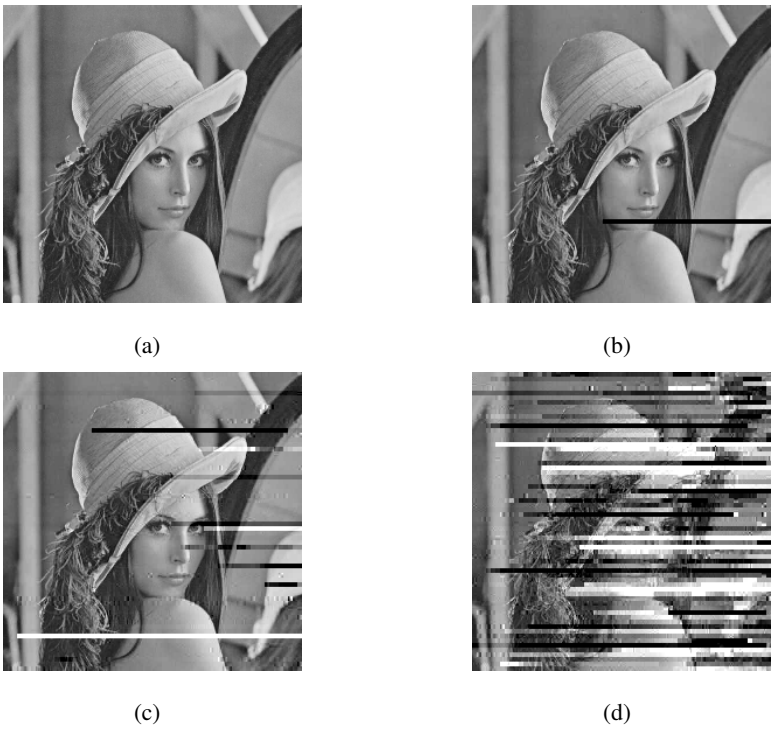
Fig. 9. Effects of quantization and channel errors in a test image at 1.25 bpp and different BERs.(a) Error free test image at 8 bits per pixel (bpp). (b) 1.25 bpp and 10^{-4} BER. PSNR=25.96 dB (model prediction: 25.98 dB). (c) 1.25 bpp and 10^{-3} BER. PSNR=20.14 dB (model prediction: 19.47 dB). (d) 1.25 bpp and 10^{-2} BER. PSNR=12.81 dB (model prediction: 12.94 dB).

distortion, however, computational complexity of such schemes prohibits their use for practical real-time image communication. In order to illustrate the use of our distortion model, we present a very simple application in Sec. V-A.

A. Unequal Power Allocation

As a simple application of the distortion model described in this paper, we transmit different JPEG layers using an unequal power allocation (UPA) scheme. All the images are coded at 1 bpp in 64 layers using the spectral selection mode of operation and Huffman coding. Again, headers and markers are assumed to be transmitted error free. The DC layer is kept separate from the AC layers, and the 63 AC layers are then grouped into 3 subgroups to keep the computational complexity low. The layers are grouped such that each subsequent group of AC layers has approximately $2/3$ of the remaining energy of the quantized coefficients (see Table II).

These 4 groups of layers are then transmitted using 4-QAM modulation over a Rayleigh flat fading channel using unequal power for different layers, bounded by a total power constraint. We assume that the channel is known at the transmitter, and that it stays constant for a group of layers but varies between



(a) (b)

(c) (d)

Fig. 10. Effects of quantization and channel errors in Lena image at 1 bpp and different BERs. (a) Error free Lena image at 8 bpp. (b) 1 bpp and 10^{-4} BER. PSNR=24.95 dB (model prediction: 24.79 dB). (c) 1 bpp and 10^{-3} BER. PSNR=18.57 dB (model prediction: 19.08 dB). (d) 1 bpp and 10^{-2} BER. PSNR=12.44 dB (model prediction: 12.76 dB).

different groups of layers. We further assume that each symbol is transmitted in unit time. Then, if E_s is the symbol power (energy), and N_0 is the noise variance, then E_s/N_0 is the SNR per symbol. Since we are using 4-QAM modulation, the average SNR per bit (E_b/N_0) is related to E_s/N_0 as $E_s/N_0 = 2E_b/N_0$. If T is the total number of symbols corresponding to the length of the bitstream, then the total power (P_{TOT}) transmitted over the length of the bitstream is E_sT. Our goal is to minimize the expected value of MSE by varying the transmit power for different groups of layers, while keeping the total power constant over the length of the symbol stream. Individual symbols within a group of layers are transmitted with equal power. We represent MSE in terms of E_b/N_0, and formulate and solve the UPA problem numerically as a constrained minimization problem. The relation between the instantaneous BER and E_b/N_0 for a Rayleigh fading channel for 4-QAM modulation [40] is given by:

$$BER = Q\left(\sqrt{2\left|H\right|^2 \frac{E_b}{N}_0}\right), \tag{18}$$

where the channel H is a circularly symmetric complex Gaussian random variable with mean 0 and variance 1.

Let $x_i = E_{b_i}/N_0$ ($i = 1..4$) be the SNR per bit for the i^{th} group of layers, $\vec{x} = [x_1\ x_2\ x_3\ x_4]^T$, and

$MSE(\vec{x})$ be the corresponding MSE as a function of SNR per bit for individual groups of layers. Then, in accordance with our grouping of layers in Table II, the expected value of MSE can be written as

$$E\left(MSE(\vec{x})\right) = E(MSE_{DC}(x_1)) + \sum_{n=1}^{8} E\left(MSE_{ACn}(x_2)\right) + \sum_{n=9}^{19} E\left(MSE_{ACn}(x_3)\right) + \sum_{n=20}^{63} E\left(MSE_{ACn}(x_4)\right). \quad (19)$$

Our objective is to minimize $E\left(MSE(\vec{x})\right)$

$$\min_{\vec{x}} E\left(MSE(\vec{x})\right),$$

with the equality constraint

$$g(\vec{x}) = \sum_{i=1}^{4} l_i x_i = P_{TOT}, \quad (20)$$

where l_i is number of bits in the i^{th} group of layers. In (20), we have assumed $N_0 = 1$, without loss of generality.

We minimize MSE using our model for 500 channel realizations at each *average* SNR per bit, given as $(\frac{E_b}{N_0})_{avg} = \frac{\sum_{i=1}^{4} l_i x_i}{\sum_{i=1}^{4} l_i}$, and then take the average of these MSE values. These results are shown in Fig. 11 with PSNR plotted against $(\frac{E_b}{N_0})_{avg}$. Note that this SNR is the average SNR per bit for the entire bitstream, whereas the actual SNRs per bit associated with individual groups of layers can be different based on the group's contribution to the total MSE. For comparison, the PSNR curve for an equal power allocation (EPA) scheme is also shown. To observe a fair comparison, the bitstream used for EPA is derived from 'baseline' JPEG coded images. Using progressive DCT with EPA would not be a fair comparison since for progressive streams the importance of bits decreases with distance from the start of the bitstream. We use baseline coding for EPA because in baseline mode there is no layering as opposed to the progressive DCT based mode, and hence all the parts of the bitstream have roughly equal importance. The source coding rate for the baseline JPEG is kept the same as that of the progressive DCT coded images. The total number of RST markers is also the same for both the cases. In both the cases, a RST marker is inserted after every 64 coefficients. For the progressive mode, all the 64 coefficients are from the same layer, whereas for the baseline mode these 64 coefficients correspond to the 64 sub-bands.

We first obtain the UPA strategy using our distortion model, and then use this strategy to transmit images from the test database. Both these curves along with the EPA curve are shown in Fig. 11. Note that the curve for EPA is obtained using simulations, since we did not derive a model for the baseline case.

In the case of UPA, at lower SNRs, it is sometimes optimum to transmit only a few groups of layers.

Group No.	Layers
1	DC Layer
2	AC Layer 1-8
3	AC Layer 9-19
4	AC Layer 20-63

TABLE II

GROUPS OF DCT LAYERS TRANSMITTED USING UNEQUAL POWER.

Hence in such cases, all the power is allocated to these groups of layers and the remaining groups of layers are not transmitted. Hence nothing is transmitted for the remaining length of the symbol stream in these scenarios and P_{TOT} is kept constant. As can be seen from Fig. 11, the UPA test simulations perform very closely to the UPA curve obtained using the model (within 0.9 dB at all points). Hence this model can be used to design UPA and JSCC schemes with reasonable accuracy. This will also reduce the system complexity significantly as compared to 'per-image' based optimizations in real-time. Also, as shown in Fig. 11, by using the UPA scheme, we get a PSNR gain of around 6.5 dB at 5 dB $(E_b/N_0)_{avg}$ as compared to EPA. This gain reduces as we move towards higher SNR. For example, at 20 dB $(E_b/N_0)_{avg}$, the PSNR gain is around 2.5 dB. This is because at high SNR there are almost no channel errors and the distortion is mostly due to quantization errors. For the error free case, PSNR for the baseline and the progressive modes are 25.95 dB and 25.53 dB respectively. The baseline does not clearly outperform the progressive mode when the channel is good, because in baseline mode, there is no DPCM coding in the DC layer since there is a RST marker after each block of subbands, resulting in a performance loss.

VI. CONCLUSION

In this paper, we presented a model for estimating the distortion introduced in an image as a result of quantization and random bit errors when compressed by a JPEG encoder, and transmitted over a noisy/fading channel. This model is unique since it incorporates modelling of DPCM and entropy coding for estimating distortion. Furthermore, since this model can predict the average distortion over a large set of images, it can be used for designing efficient JSCC techniques for transmission of a large set of images over noisy/fading channels, with no need of carrying out the associated optimization procedures again for each image and varying channel conditions. This eliminates the image dependence of the designed JSCC schemes, hence significantly reducing the computational complexity for real-time applications. To our knowledge, such a model for any practical coding standard has not been presented previously. Simulation results show that the distortion predicted by the model estimates the value obtained via simulations quite

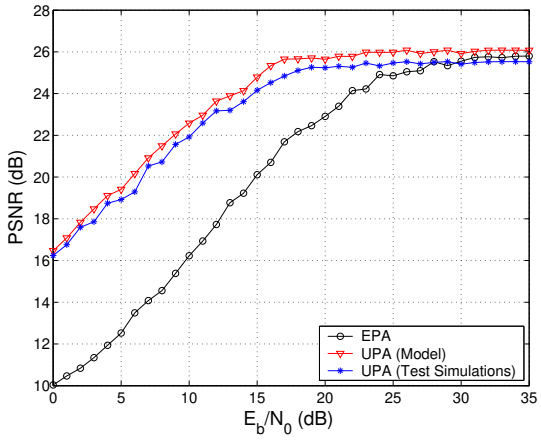


Fig. 11. PSNR comparison for unequal power and equal power allocation schemes at 1 bpp source coding rate.

closely. As a simple application of this model, we presented an UPA scheme for JPEG compressed images for transmission over Rayleigh flat fading channels. Different layers in the JPEG compressed images were transmitted using unequal power while keeping the total transmitted power over the length of the symbol stream constant. Results show a 6.5 dB PSNR gain over an EPA scheme at 5 dB SNR. In future, we plan to formulate similar distortion models for MPEG-4 and H.263 video coding standards. In addition, we plan to devise joint source-channel coding and unequal error protection schemes based on these models for JPEG, MPEG-4 and H.263 standards.

APPENDIX

We define the truncated Gaussian distribution as

$$p(x|\rho,\sigma^2) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{\exp\ -\frac{(x-\rho)^2}{2\sigma^2}}{\sigma \mathrm{erfc}\left(\frac{\epsilon-\rho}{\sqrt{2}\sigma}\right)}, & x \geq \epsilon \\ 0, & x < \epsilon \end{cases} \tag{21}$$

where ρ and σ^2 are the distribution parameters, ϵ is a small positive constant, and $\mathrm{erfc}(\cdot)$ is the complementary error function. We choose ϵ to be greater than 0 because when a layer is encoded, the average length is always greater than 0. It is possible to show that

$$E(X|\rho,\sigma^2) = \rho + \sqrt{\frac{2}{\pi}}\sigma \frac{\exp\left(-\frac{(\epsilon-\rho)^2}{2\sigma^2}\right)}{\mathrm{erfc}\left(\frac{\epsilon-\rho}{\sqrt{2}\sigma}\right)}, \tag{22}$$

$$E\left(X^2|\rho,\sigma^2\right) = \rho^2 + \sigma^2 + \sqrt{\frac{2}{\pi}}\sigma \frac{\exp\left(-\frac{(\epsilon-\rho)^2}{2\sigma^2}\right)(\epsilon+\rho)}{\mathrm{erfc}\left(\frac{\epsilon-\rho}{\sqrt{2}\sigma}\right)}. \tag{23}$$

We used the mean and the variance of the average coefficient lengths obtained from the empirical distributions to find ρ and σ^2 using numerical methods. Note that this numerical solution is only required during training the model and not during its testing.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 & 623–656, Jul 1948.

[2] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. com-27, pp. 1644–1659, Nov. 1979.

[3] J. W. Modestino, D. G. Daut, and A. L. Vickers, "Combined source-channel coding of images using the block cosine transmform," *IEEE Trans. Commun.*, vol. com-29, pp. 1261–1274, Sept. 1981.

[4] A. A. Alatan, M. Zhao, and A. N. Akansu, "Unequal error protection of SPIHT encoded image bit streams," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 814–818, June 2000.

[5] A. Alavi, R. Link, and S. Kallel, "Adaptive unequal error protection for subband image coding," *IEEE Trans. Broadcast.*, vol. 46, pp. 197–205, Sep 2000.

[6] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 827–838, Nov. 1987.

[7] W. B. Pennebaker and J. L. Mitchell, "JPEG still image data compression standard." Van Nostrand Reinhold, 1993.

[8] M. J. Ruf and J. W. Modestino, "Operational rate-distortion performance for joint source and channel coding of images," *IEEE Trans. Image Processing*, vol. 8, pp. 305–320, Mar 1999.

[9] K. Fazel and J. J. Lhuillier, "Application of unequal error protection codes on combined source-channel coding of images," *in Proc. IEEE Int. Conf. Comm.*, vol. 3, pp. 898–903, Apr 1990.

[10] D. Comstock and J. Gibson, "Hamming coding of DCT-compressed images over noisy channels," *IEEE Trans. Commun.*, vol. COM-32, pp. 856–861, Jul 1984.

[11] N. Tanabe and N. Farvardin, "Suband image coding using entropy-coded quantization over noisy channels," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 926–943, Jun 1992.

[12] V. Chande and N. Farvardin, "Progressive transmission of images over memoryless noisy channels," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 850 –860, June 2000.

[13] P. G. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Lett.*, vol. 4, pp. 189–191, Jul 1997.

[14] P. Sherwood and K. Zeger, "Error protection for progressive image transmission over memoryless and fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1555–1559, Dec 1998.

[15] P. Cosman, J. Rogers, P. Sherwood, and K. Zeger, "Combined forward error control and packetized zerotree wavelet encoding for transmission of images over varying channels," *IEEE Trans. Image Processing*, vol. 9, pp. 982–993, Jun 1998.

[16] J. Rogers and P. Cosman, "Wavelet zerotree image compression with packetization," *IEEE Signal Processing Lett.*, vol. 5, pp. 105–107, May 1998.

[17] A. J. Goldsmith and M. Effros, "Joint design of fixed-rate source codes and multiresolution channel codes," *IEEE Trans. Commun.*, vol. 46, pp. 1301–1312, Oct 1998.

[18] J. Hagenauer, "Rate compatible punctured convolutional codes and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.

[19] I. Kozintsev and K. Ramchandran, "Robust image transmission over energy-constrained time-varying channels using multiresolution joint source-channel coding," *IEEE Trans. Signal Processing*, vol. 46, pp. 1012–26, Apr. 1998.

[20] J. Garcia-Frias and J. D. Villasenor, "An analytical treatment of channel-induced distortion in run length coded subbands," *Proc. Data Compression Conference*, pp. 52–61, Mar 1997.

[21] M. Bystrom and J. W. Modestino, "Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 880–890, Jun 2000.

[22] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Trans. Image Processing*, vol. 9, pp. 340–356, Mar. 2000.

[23] Y. Eisenberg, C. Luna, T. Pappas, R. Berry, and A. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 411–424, Jun 2002.

[24] A. Nosratinia, J. Lu, and B. Aazhang, "Source-channel rate allocation for progressive transmission of images," *IEEE Trans. Commun.*, vol. 51, pp. 186–196, Feb 2003.

[25] J. Cai and C. W. Chen, "Robust joint source-channel coding for image transmission over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 962–966, Sep 2000.

[26] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the internet using unequal error protection," *IEEE Trans. Image Processing*, vol. 12, pp. 121–131, Feb 2003.

[27] P. Burlina and F. Alajaji, "An error resilient scheme for image transmission over noisy channels with memory," *IEEE Trans. Image Processing*, vol. 7, pp. 593–600, Apr 1998.

[28] H. Man, F. Kossentini, and M. J. T. Smith, "A family of efficient and channel error resilient wavelet/subband image coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 95–108, Feb. 1999.

[29] W. Xiang, S. A. Barbulescu, and S. S. Pietrobon, "Unequal error protection applied to jpeg image transmission using turbo codes," *IEEE Proceddings, Information Theory Workshop, 2001*, pp. 64–66, 2001.

[30] H. Gharavi and S. M. Alamouti, "Multipriority video transmission for third-generation wireless communication systems," *Proceedings of the IEEE*, vol. 87, pp. 1751–1763, Oct. 1999.

[31] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," *Proceedings of the IEEE*, vol. 87, pp. 1164–1777, Oct. 1999.

[32] J. Song and K. J. R. Liu, "Robust progressive image transmission over OFDM systems using space-time block code," *IEEE Trans. Multimedia*, vol. 4, pp. 394–406, Sep 2002.

[33] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, pp. 974–997, May 1998.

[34] N. Ohta, *Packet Video: Modeling and Signal Processing.* Norwood, MA: Artech House, 1994.

[35] A. Alavi, R. Link, and S. Kallel, "Adaptive unequal error protection for subband image coding," *IEEE Trans. Broadcast.*, vol. 46, pp. 197–205, Sept. 2000.

[36] K. M. Rose and A. Heiman, "Enhancement of one-dimensional variable-length DPCM images corrupted by transmission errors," *IEEE Trans. Commun.*, vol. 37, pp. 373–379, Apr 1989.

[37] W.-M. Lam and A. Reibman, "An error concealment algorithm for images subject to channel errors.," *IEEE Trans. Image Processing*, vol. 4, pp. 533–542, May 1995.

[38] K. Sayood, H. H. Otu, and N. Demir, "Joint source/channel coding for variable length codes," *IEEE Trans. Commun.*, vol. 48, pp. 787–794, May 2000.

[39] R. Zhang, L. Regunathan, and K. Rose, "Optimal estimation for error concealment in scalable video coding," *in Proc. Asilomar Conf. Signals, Systems, Computers*, vol. 2, pp. 1374–1378, Nov 2000.

[40] M. K. Simon and M. S. Alouni, "Digital communication over fading channels." John Wiley & Sons, 2000.

Muhammad F. Sabir received his B.S. degree in Electronics Engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan in 1999 and his M.S. in Electrical Engineering from The University of Texas at Austin, Austin, Texas, USA, in 2002. He is currently a Ph.D. student at The University of Texas at Austin. His research interests include joint source-channel coding, real-time image and video communication, multiple-input multiple-output systems, and space-time coding. He is a member of the Laboratory for Image and Video Engineering, and the Wireless Networking and Communications Group at The University of Texas at Austin.

Hamid R. Sheikh completed his Bachelor's degree in Electrical Engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1998 and his MS and Ph.D. from The University of Texas, Austin, Texas, USA, in 2001 and 2004 respectively. His research interests include full-reference and no-reference quality assessment, application of natural scene statistics models and human visual system models for solving image and video processing problems, and image and video codecs and their embedded implementation.

Robert W. Heath Jr. (S'96 - M'01) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, in 1996 and 1997 respectively, and the Ph.D. from Stanford University, Stanford, CA, in 2002, all in electrical engineering.

From 1998 to 1999, he was a Senior Member of the Technical Staff at Iospan Wireless Inc, San Jose, CA where he played a key role in the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. From 1999 to 2001 he served as a Senior Consultant for Iospan Wireless Inc. In 2003 he founded MIMO Wireless Inc. Since January 2002, he has been with the Department of Electrical and Computer Engineering at The University of Texas at Austin where he serves as an Assistant Professor as part of the Wireless Networking and Communications Group. His research interests include interference management in wireless networks, sequence design, and all aspects of MIMO communication including antenna design, practical receiver architectures, limited feedback techniques, and scheduling algorithms.

Dr. Heath serves as an Associate Editor for the IEEE Transactions on Communication and the IEEE Transactions on Vehicular Technology.

Alan C. Bovik is currently the Curry/Cullen Trust Endowed Chair at The University of Texas at Austin and is the Director of the Laboratory for Image and Video Engineering (LIVE) in the Center for Perceptual Systems. His current research interests include digital video, image processing, and computational aspects of biological visual perception. He has published over 400 technical articles in these areas and holds two U.S. patents. He is also the editor/author of the Handbook of Image and Video Processing, published by Academic Press in April of 2000, with Second Edition forthcoming in 2005.

Dr. Bovik was named Distinguished Lecturer of the IEEE Signal Processing Society in 2000, received the IEEE Signal Processing Society Meritorious Service Award in 1998, the IEEE Third Millennium Medal in 2000, and is a two-time Honorable Mention winner of the international Pattern Recognition Society Award for Outstanding Contribution (1988 and 1993). He is a Fellow of the IEEE and has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996-1998; Editor-in-Chief, IEEE Transactions on Image Processing, 1996-2002; Editorial Board, The Proceedings of the IEEE, 1998-2004; Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003-present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, Texas, in November, 1994.

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions.