



Parameter Selection For WYNER-ZIV Coding of Laplacian Sources

Debargha Mukherjee
Media Technologies Laboratory
HP Laboratories Palo Alto
HPL-2007-196
December 20, 2007*

WYNER-ZIV
Coding, distributed
coding, Laplacian
sources, arithmetic
coding, channel
coding

A large number of practical coding scenarios deal with sources such as transform coefficients that can be well modeled as Laplacians. For regular coding of such sources, samples are often quantized by a family of uniform quantizers possibly with a deadzone, and then entropy coded. For the Wyner-Ziv coding problem when correlated side-information is available at the decoder, the side-information can be modeled as obtained by additive Laplacian or Gaussian noise on the source. This paper deals with optimal choice of parameters for practical Wyner-Ziv coding in such scenarios, using the same quantizer family as in the regular codec to cover a range of rate-distortion trade-offs, given the variances of the source and additive noise. We propose and analyze a general encoding model that combines source coding and channel coding and show that at practical block lengths and code complexities, not pure channel coding but a hybrid combination of source coding and channel coding with right parameters provide optimal rate-distortion performance. Further, for the channel coded bit-planes we observe that only high-rate codes are useful. We also provide a framework for on-the-fly parameter choice based on non-parametric representation of a set of seed functions, for use in scenarios where variances are estimated during encoding. A good understanding of the optimal parameter selection mechanism is essential for building practical distributed codecs.

PARAMETER SELECTION FOR WYNER-ZIV CODING OF LAPLACIAN SOURCES

Debargha Mukherjee
Hewlett Packard Laboratories, Palo Alto, California, USA
Email: debargha@hpl.hp.com

ABSTRACT

A large number of practical coding scenarios deal with sources such as transform coefficients that can be well modeled as Laplacians. For regular coding of such sources, samples are often quantized by a family of uniform quantizers possibly with a deadzone, and then entropy coded. For the Wyner-Ziv coding problem when correlated side-information is available at the decoder, the side-information can be modeled as obtained by additive Laplacian or Gaussian noise on the source. This paper deals with optimal choice of parameters for practical Wyner-Ziv coding in such scenarios, using the same quantizer family as in the regular codec to cover a range of rate-distortion trade-offs, given the variances of the source and additive noise. We propose and analyze a general encoding model that combines source coding and channel coding and show that at practical block lengths and code complexities, not pure channel coding but a hybrid combination of source coding and channel coding with right parameters provide optimal rate-distortion performance. Further, for the channel coded bit-planes we observe that only high-rate codes are useful. We also provide a framework for on-the-fly parameter choice based on non-parametric representation of a set of seed functions, for use in scenarios where variances are estimated during encoding. A good understanding of the optimal parameter selection mechanism is essential for building practical distributed codecs.

1. INTRODUCTION

Inspired by the foundation laid by the Slepian-Wolfe [1] and Wyner-Ziv [2] theorems, in recent years immense attention has been devoted to practical source coding with side-information problems [3]-[10]. Most such work emphasizes using channel coding to convey a source, requiring the decoder to perform appropriate channel decoding based on correlated side-information. However, achieving virtually error-free transmission under practical code and complexity constraints always requires a non-trivial premium in the transmitted rate over the ideal channel coding rate given by the conditional entropy. Furthermore, in many realistic scenarios with finite block-lengths, the actual correlation statistics are non-stationary and can only be estimated with a large enough buffer for errors. This challenges the rationale for solely using channel codes for source coding with side-information problems. While in some prior work, a combination of source and channel coding has been used intuitively, this work addresses quantitatively the problem of finding the optimum balance between various feasible source and channel coding combinations under practical channel coding constraints, based on a realistic model for the source and correlation statistics. We assume a Laplacian source model since most transform coefficients are well-modeled as Laplacians. Further, the side-information available only at the decoder in a variety of applications can be well modeled by additive Generalized Gaussian noise on the source. In particular, if X denotes the source Laplacian random variable with variance σ_X^2 , and Y is the side-information available only at the decoder, then $Y = X + Z$, where Z is either *i.i.d.* Gaussian or *i.i.d.* Laplacian with variance σ_Z^2 in our models. Note that the results of this work trivially generalizes to a related $Y = \rho X + Z$ model with $0 \leq \rho \leq 1$, by replacing Y with Y/ρ , and σ_Z by σ_Z/ρ [15].

In any practical codec, X is quantized with a quantizer family ϕ to yield a quantization index random variable Q : $Q = \phi(X, QP)$, where QP parameterizes a family of quantizers that yield progressively coarse to fine quantization over a wide enough range. The simplest quantizer family is the uniform quantizer, given by $Q = \phi(X, QP) = \text{round}(X/QP)$. The uniform deadzone quantizer is actually more commonly used in practical codecs:

$$Q = \phi(X, QP) = \text{sign}(X) \times \lfloor |X|/QP \rfloor \quad (1)$$

In general QP may be continuous, but practically it takes values from a discrete set Ω_{QP} . While the results we present in this paper are for the uniform deadzone quantizer, the methodology applies to any family of quantizers yielding a range of rate-distortion trade-offs through a QP parameter. The quantized indices Q are typically truncated to a finite set $Q \in \Omega_Q = \{-q_{\max}, -q_{\max} + 1, \dots, -1, 0, 1, \dots, q_{\max} - 1, q_{\max}\}$, where q_{\max} is chosen large enough to make the overload probability negligible. Ideally q_{\max} depends on QP and σ_X , but for simplicity we simply refer to the set of all available quantization bins as Ω_Q .

In the source coding with side-information scenario under consideration in this work, we assume that the same quantizer family as in the regular codec, is used. The problem we address is then broadly stated as follows: Given a target upper-limit D_t on the overall expected distortion, and variances $\{\sigma_X^2, \sigma_Z^2\}$ for Laplacian X and Laplacian/Gaussian Z respectively, how should X be coded based on a given quantizer family. It may be convenient to specify D_t in terms of a target quantization parameter QP_t assuming regular coding (with no side-information) based on the same quantizer family.

2. GENERAL MODEL FOR SOURCE CODING WITH SIDE-INFORMATION

Once X has been quantized to Q , in a regular source coder the quantization bins are just entropy coded. In the source coding with side-information scenario, cosets are computed based on Q to yield a coset index random variable C : $C = \psi(Q, M)$, M being the coset

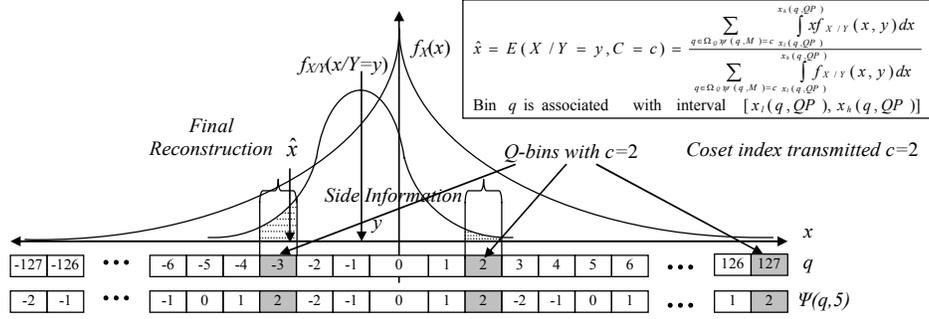


Figure 1. Optimum reconstruction based on coset index and side-information

modulus parameter, as follows:

$$C = \psi(Q, M) = \text{mod}_{c_2}(Q, M) = Q - M \{ \lfloor Q/M \rfloor + u(2Q - 2M \lfloor Q/M \rfloor - M) \} \quad (2)$$

$u(k)$ is the discrete unit step function: $u(k) = 0$ for $k < 0$, and $u(k) = 1$ for $k \geq 0$. Note that $C \in \Omega_C = \{ \lfloor -(M-1)/2 \rfloor, \dots, -1, 0, 1, \dots, \lfloor (M-1)/2 \rfloor \}$ is zero-centered as a matter of convention.

We next consider the case where C is further decomposed into S symbol planes $\{C_0, C_1, \dots, C_{S-1}\}$ where $C_i, i=0, 1, \dots, S-1$ is the $(i+1)$ th least significant symbol (i.e. C_0 is the least significant symbol, C_{S-1} is the most significant symbol) associated with a finite l_i -ary alphabet. The symbols C_i can be obtained from Q directly, given the alphabet-size vector $L = \{l_0, l_1, \dots, l_{S-1}\}$:

$$\text{Initialize: } Q_0 = Q; \quad \text{Compute: } C_i = \text{mod}_{c_i}(Q_i, l_i), \quad Q_{i+1} = \lfloor Q_i / l_i \rfloor \quad \text{for } i = 0, 1, \dots, S-1 \quad (3)$$

Thus $C_i \in \Omega_{C_i} = \{ \lfloor -(l_i-1)/2 \rfloor, \dots, -1, 0, 1, \dots, \lfloor (l_i-1)/2 \rfloor \}$ for $i = 0, 1, \dots, S-1$. The overall modulus M is given by $M = l_0 l_1 \dots l_{S-1}$. Note there is an one-to-one mapping between the overall coset index C and the symbol planes $\{C_0, C_1, \dots, C_{S-1}\}$, given by:

$$C + (l_0 l_1 \dots l_{S-1}) u(C) = \{C_0 + l_0 u(C_0)\} + l_0 \{C_1 + l_1 u(C_1)\} + l_0 l_1 \{C_2 + l_2 u(C_2)\} + \dots + (l_0 l_1 \dots l_{S-2}) \{C_{S-1} + l_{S-1} u(C_{S-1})\} \quad (4)$$

where $u(k)$ is the discrete unit step function with value 0 for negative k , and 1 for non-negative k . For simplicity, we use the notation $C_i = \xi_i^l(Q)$ to denote the mapping from Q to the i th symbol plane, given the alphabet size vector L .

If C or equivalently the corresponding symbol planes $\{C_0, C_1, \dots, C_{S-1}\}$ were transmitted losslessly to the decoder, the decoder would perform an optimal reconstruction of a sample X based on the corresponding side-information $Y=y$, and the transmitted coset modulus $C=c$, as follows:

$$\hat{X}_{YC}(y, c) = E(X/Y = y, C = c) = E(X/Y = y, \psi(\phi(X, QP), M) = c) = \frac{\sum_{q \in \Omega_Q: \psi(q, M) = c} \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx}{\sum_{q \in \Omega_Q: \psi(q, M) = c} \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx} = \frac{\sum_{q \in \Omega_Q: \psi(q, M) = c} \mu(q, y, QP)}{\sum_{q \in \Omega_Q: \psi(q, M) = c} \pi(q, y, QP)} \quad (5)$$

where we have introduced the following definitions for convenience in the rest of the paper:

$$\pi(q, y, QP) = p_{Q|Y}(Q = q|Y = y) = \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx, \quad \mu(q, y, QP) = \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx \quad (6)$$

and the quantization bin q for quantization parameter QP is associated with interval $[x_i(q, QP), x_{i+1}(q, QP)]$. A decoding example is shown in Figure 1.

The coset C in general may be transmitted using a combination of source and channel coding. In particular, some of the symbol planes $\{C_0, C_1, \dots, C_{S-1}\}$ are source coded while the rest are channel coded. For channel coded symbol planes, we assume that the planes be only binary, i.e. $l_i=2$, because of the vast resources available in binary channel codes. However, extension to non-binary channel codes is straight-forward in principle. The symbol planes may be transmitted in any order, but the encoding parameters must be matched to that, and the decoding order must be the same as the encoding order [10]. Specifically, each channel coded bit-plane is be encoded at a rate equivalent to the conditional entropy of the bit-plane given Y and previously transmitted symbol-planes, and correspondingly decoded using a soft-input decoder driven by conditional probabilities of the bits given Y and previously decoded symbol planes. Likewise each source coded symbol plane needs to be encoded and decoded conditioned on the previously transmitted symbol-planes.

Now observe that since the lower significant symbols are less correlated with the side-information Y , there would be little gain in using channel coding for them. Also, if one or more channel coded symbol planes were transmitted prior to a source coded plane, then decoding errors could lead to a catastrophic loss of synchronization in the bit-stream during source decoding based on an incorrect context. Based on these considerations, we propose a model for coding where the least significant symbol C_0 is an m -ary source coded plane, while the remaining more significant symbol planes C_1, C_2, \dots, C_{S-1} are channel coded bit-planes. Thus, $M=2^{S-1}m$ and the alphabet-size vector $L = \{m, 2, 2, \dots, 2\}$, with $l_0=m$, and $l_i=2$ for $i=1, 2, \dots, S-1$. The S symbol planes are coded in order from C_0 to C_{S-1} .

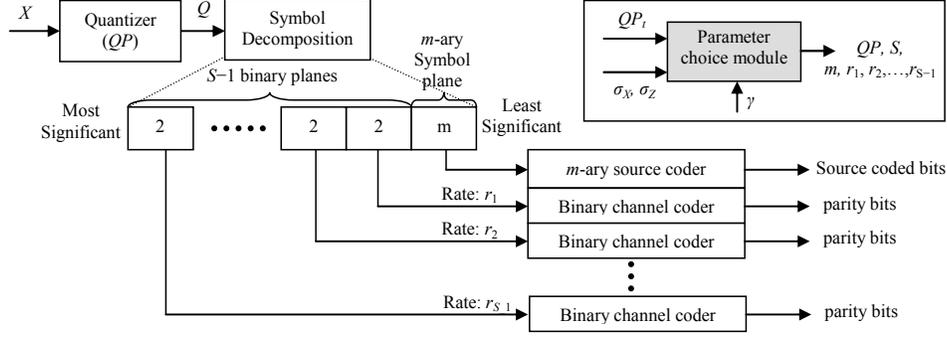


Figure 2. Model for Wyner-Ziv coding with source-channel combination codes

Figure 2 illustrates the model, the parameters to be selected and how they are used for encoding. Given a target quantization parameter for regular coding QP_b , and the source and correlation statistics σ_X and σ_Z , optimum parameter choice for Wyner-Ziv coding essentially involves obtaining: the optimum quantization parameter QP , the number of symbol planes $S \geq 1$, the alphabet-size of the source coded least significant plane $m \geq 1$, and the rates for the $S-1$ channel coded bit-planes r_1, r_2, \dots, r_{S-1} . There is an additional input parameter gamma to be explained in the next section.

Note that the above model is considerably generic, and covers all special cases of interest. For instance, $S=1$ corresponds to using only memoryless source coding (and no channel coding) used in [9], and may be preferred in many cases because of their simplicity. The case $m=1$, corresponds to not using any source coding bits since the least significant symbol plane is trivial. So $m=1$ and $S>1$ corresponds to only using channel coding. The case $m=1, S=1$ corresponds to the zero-rate coding case when the samples are not coded at all, but optimal reconstruction based on Y can still be conducted at the decoder. Finally, $m=\infty$ corresponds to the case when no coset computation is conducted on the quantized samples, and also S must be 1 in this case.

3. SEED FUNCTIONS AND NON-PARAMETRIC REPRESENTATION

In order to select the optimum parameters, we assume the availability of the following seed functions for a given distribution shape (Laplacian or Gaussian) of the noise Z in the $Y=X+Z$ model: (1) Unit-variance Coset Distortion function $d_{YC}(QP, M, \sigma_Z)$, providing the expected distortion incurred after optimum side-information based reconstruction of a unit-variance ($\sigma_X=1$) Laplacian source, coded with quantization parameter QP and coset modulus M , given the noise std. deviation σ_Z ; (2) Unit-variance Coset Conditional Entropy function $h_{YC}(QP, M, \sigma_Z)$, providing the conditional entropy of the coset indices given the side-information; (3) Unit-variance Coset Entropy function $h_C(QP, M)$, providing the entropy of the coset indices; and (4) Regular Distortion function $d_Q(QP)$, providing the expected distortion incurred by optimum reconstruction without side-information of a unit-variance Laplacian source quantized with parameter QP . The last one is needed only if distortion matching is required.

Note that corresponding functions for arbitrary source standard deviation σ_X – namely $D_{YC}(QP, M, \sigma_Z, \sigma_X)$, $H_{YC}(QP, M, \sigma_Z, \sigma_X)$, $H_C(QP, M, \sigma_X)$, and $D_Q(Q, \sigma_X)$ – can be derived from the unit-variance functions as follows:

$$\begin{aligned}
 D_{YC}(QP, M, \sigma_Z, \sigma_X) &= (\sigma_X^2) d_{YC}(QP / \sigma_X, M, \sigma_Z / \sigma_X) \\
 H_{YC}(QP, M, \sigma_Z, \sigma_X) &= h_{YC}(QP / \sigma_X, M, \sigma_Z / \sigma_X) \\
 H_C(QP, M, \sigma_X) &= h_C(QP / \sigma_X, M) \\
 D_Q(QP, \sigma_X) &= (\sigma_X^2) d_Q(QP / \sigma_X)
 \end{aligned} \tag{7}$$

Therefore it is enough to represent only the unit-variance forms helping to reduce the dimensionality by 1. In this section, we will derive expressions for these functions, based on which a non-parametric representation can be readily obtained. In the next section, we will use these function representations for optimal parameter selection.

3.1. Coset Distortion Function

Assuming the minimum mean-squared-error reconstruction function in Eq. 5, the expected distortion D_{YC} given side information y and coset index c is given by:

$$E(D_{YC} / Y = y, C = c) = E([X - \hat{X}_{YC}(y, c)]^2 / Y = y, C = c) = E(X^2 / Y = y, C = c) - \hat{X}_{YC}(y, c)^2 \tag{8}$$

using $\hat{X}_{YC}(y, c) = E(X / Y = y, C = c)$. Marginalizing over y and c yields the overall expected distortion $E(D_{YC}) = D_{YC}(QP, M, \sigma_Z, \sigma_X)$:

$$\begin{aligned}
 D_{YC}(QP, M, \sigma_Z, \sigma_X) &= E(D_{YC}) = E(X^2) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{c \in \Omega_c} \hat{X}_{YC}(y, c)^2 p_{C/Y}(C = c / Y = y) f_Y(y) dy \\
 &= \sigma_X^2 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\sum_{q \in \Omega_q} \mu(q, y)}{\sum_{q \in \Omega_q} \pi(q, y)} \right)^2 p_{C/Y}(C = c / Y = y) f_Y(y) dy = \sigma_X^2 - \int_{-\infty}^{\infty} \left(\frac{\sum_{q \in \Omega_q} \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx}{\sum_{q \in \Omega_q} \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx} \right)^2 p_{C/Y}(C = c / Y = y) f_Y(y) dy
 \end{aligned} \tag{9}$$

where $p_{C/Y}(C = c / Y = y)$ is the conditional probability mass function of C given Y . Noting that:

$$p_{C/Y}(C = c / Y = y) = \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} p_{Q/Y}(Q = q / Y = y) = \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \pi(q, y, QP) = \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx \quad (10)$$

we can write:

$$D_{YC}(QP, M, \sigma_Z, \sigma_X) = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \frac{\left(\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \mu(q, y, QP) \right)^2}{\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \pi(q, y, QP)} \right\} f_Y(y) dy = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \frac{\left(\int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx \right)^2}{\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx} \right\} f_Y(y) dy = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \frac{\left(\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \mu(q, y, QP) \right)^2}{\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \pi(q, y, QP)} \right\} f_Y(y) dy \quad (11)$$

Defining:

$$m_{x/Y}^{(i)}(x, y) = \int_{-\infty}^x x' f_{X/Y}(x', y) dx' \quad (12)$$

we have:

$$\pi(q, y, QP) = p_{Q/Y}(Q = q / Y = y) = \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx = [m_{x/Y}^{(0)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(0)}(x_i(q, QP), y)] \quad (13)$$

$$\mu(q, y, QP) = \int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx = [m_{x/Y}^{(1)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(1)}(x_i(q, QP), y)]$$

Then we can rewrite Eq. 11 as:

$$D_{YC}(QP, M, \sigma_Z, \sigma_X) = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} \frac{\left(\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} [m_{x/Y}^{(1)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(1)}(x_i(q, QP), y)] \right)^2}{\sum_{q \in \Omega_Q \mathcal{W}(q, M) = c} [m_{x/Y}^{(0)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(0)}(x_i(q, QP), y)]} \right\} f_Y(y) dy \quad (14)$$

Certain boundary conditions are appropriate to consider. When $M \rightarrow \infty$, we have the case where the quantization index Q is transmitted noise free in entirety. In this case, the optimal reconstruction is conducted within the given quantization bin. This reconstruction function $\hat{X}_{YQ}(y, q)$, for $Y = y$ and $Q = q$, is given by:

$$\hat{X}_{YQ}(y, q) = E(X / Y = y, Q = q) = E(X / Y = y, \phi(X, QP) = q) \\ = \frac{\int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx}{\int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx} = \frac{\mu(q, y, QP)}{\pi(q, y, QP)} = \frac{m_{x/Y}^{(1)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(1)}(x_i(q, QP), y)}{m_{x/Y}^{(0)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(0)}(x_i(q, QP), y)} \quad (15)$$

Using this reconstruction, the expected Distortion with noise-free quantization bins (denoted D_{YQ}) is given by:

$$D_{YC}(QP, \infty, \sigma_Z, \sigma_X) = E(D_{YQ}) = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q} \frac{\left(\int_{x_i(q, QP)}^{x_{i+1}(q, QP)} x f_{X/Y}(x, y) dx \right)^2}{\int_{x_i(q, QP)}^{x_{i+1}(q, QP)} f_{X/Y}(x, y) dx} \right\} f_Y(y) dy = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q} \frac{\mu(q, y, QP)^2}{\pi(q, y, QP)} \right\} f_Y(y) dy \\ = \sigma_X^2 - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q} \frac{\left(m_{x/Y}^{(1)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(1)}(x_i(q, QP), y) \right)^2}{\left(m_{x/Y}^{(0)}(x_{i+1}(q, QP), y) - m_{x/Y}^{(0)}(x_i(q, QP), y) \right)} \right\} f_Y(y) dy$$

Note that this is also the distortion incurred when ideal Slepian Wolfe coding is used to transmit the quantization indices Q noise-free.

When $M = 1$ and/or $QP \rightarrow \infty$, it corresponds to the zero-rate case where no information is transmitted. The optimum reconstruction function $\hat{X}_Y(y)$, given side-information $Y=y$, is then given by:

$$\hat{X}_Y(y) = E(X / Y = y) = \int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx = m_{x/Y}^{(1)}(\infty, y) \quad (17)$$

The corresponding expected zero-rate distortion, denoted D_Y , is given by:

$$D_{YC}(\infty, M, \sigma_Z, \sigma_X) = D_{YC}(QP, 1, \sigma_Z, \sigma_X) = E(D_Y) = \sigma_X^2 - \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X/Y}(x, y) dx \right)^2 f_Y(y) dy = \sigma_X^2 - \int_{-\infty}^{\infty} m_{x/Y}^{(1)}(\infty, y)^2 f_Y(y) dy \quad (18)$$

Finally, when $QP \rightarrow 0$, no useful information is conveyed by the coset index C , and the distortion is the same as that for the zero-rate case in Eq. 18. The case $QP \rightarrow 0$ and $M \rightarrow \infty$, is a degenerate case.

3.2. Coset Conditional Entropy function

The coset conditional entropy function is the conditional entropy $H(C/Y)$ of the coset $C = \{C_0, C_1, \dots, C_{S-1}\}$ given side-information Y .

$$\begin{aligned}
H_{YC}(QP, M, \sigma_Z, \sigma_X) &= H(C/Y) = \int_{-\infty}^{\infty} (H(C/Y=y)) f_Y(y) dy \\
&= - \int_{-\infty}^{\infty} \left(\sum_{c \in \Omega_C} p_{C/Y}(C=c/Y=y) \log_2 p_{C/Y}(C=c/Y=y) \right) f_Y(y) dy = - \int_{-\infty}^{\infty} \left(\sum_{c \in \Omega_C} \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} p_{Q/Y}(Q=q/Y=y) \right\} \log_2 \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} p_{Q/Y}(Q=q/Y=y) \right\} \right) f_Y(y) dy \quad (19) \\
&= - \int_{-\infty}^{\infty} \left(\sum_{c \in \Omega_C} \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} \pi(q, y, QP) \right\} \log_2 \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} \pi(q, y, QP) \right\} \right) f_Y(y) dy \\
&= - \int_{-\infty}^{\infty} \left(\sum_{c \in \Omega_C} \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} [m_{X/Y}^{(0)}(x_h(q, QP), y) - m_{X/Y}^{(0)}(x_l(q, QP), y)] \right\} \log_2 \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} [m_{X/Y}^{(0)}(x_h(q, QP), y) - m_{X/Y}^{(0)}(x_l(q, QP), y)] \right\} \right) f_Y(y) dy
\end{aligned}$$

When $M \rightarrow \infty$, this function reduces to the conditional entropy $H(Q/Y)$, which is also the ideal Slepian-Wolfe rate:

$$\begin{aligned}
H_{YC}(QP, \infty, \sigma_Z, \sigma_X) &= H(Q/Y) = - \int_{-\infty}^{\infty} \left(\sum_{q \in \Omega_Q} p_{Q/Y}(Q=q/Y=y) \log_2 p_{Q/Y}(Q=q/Y=y) \right) f_Y(y) dy = - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q} \pi(q, y) \log_2 \pi(q, y) \right\} f_Y(y) dy \quad (20) \\
&= - \int_{-\infty}^{\infty} \left\{ \sum_{q \in \Omega_Q} [m_{X/Y}^{(0)}(x_h(q), y) - m_{X/Y}^{(0)}(x_l(q), y)] \log_2 [m_{X/Y}^{(0)}(x_h(q), y) - m_{X/Y}^{(0)}(x_l(q), y)] \right\} f_Y(y) dy
\end{aligned}$$

When $M=1$ and/or $QP \rightarrow \infty$, it is the zero-rate case, and so:

$$H_{YC}(\infty, M, \sigma_Z, \sigma_X) = H_{YC}(QP, 1, \sigma_Z, \sigma_X) = 0 \quad (21)$$

When $QP \rightarrow 0$, no useful information is conveyed by the coset index C , but because each coset index is equi-probable, we have:

$$H_{YC}(0, M, \sigma_Z, \sigma_X) = \log_2 M \quad (22)$$

The case $QP \rightarrow 0$ and $M \rightarrow \infty$, is a singularity.

3.3. Coset Entropy function

The coset entropy function is simply the entropy $H(C)$ of the coset $C = \{C_0, C_1, \dots, C_{S-1}\}$:

$$\begin{aligned}
H_C(QP, M, \sigma_X) &= H(C) = - \sum_{c \in \Omega_C} p_C(c) \log_2 p_C(c) = - \sum_{c \in \Omega_C} \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} p_Q(q) \right\} \log_2 \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} p_Q(q) \right\} \\
&= - \sum_{c \in \Omega_C} \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} \int_{x_l(q, QP)}^{x_h(q, QP)} f_X(x) dx \right\} \log_2 \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} \int_{x_l(q, QP)}^{x_h(q, QP)} f_X(x) dx \right\} \quad (23) \\
&= - \sum_{c \in \Omega_C} \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} [m_X^{(0)}(x_h(q, QP)) - m_X^{(0)}(x_l(q, QP))] \right\} \log_2 \left\{ \sum_{q \in \Omega_Q, \psi(q, M)=c} [m_X^{(0)}(x_h(q, QP)) - m_X^{(0)}(x_l(q, QP))] \right\}
\end{aligned}$$

where we defined: $m_X^{(i)}(x) = \int_{-\infty}^x x'^i f_X(x') dx'$. When $M \rightarrow \infty$, the entropy function converges to the entropy $H(Q)$:

$$H_C(QP, \infty, \sigma_X) = H(Q) = - \sum_{q \in \Omega_Q} p_Q(q) \log_2 p_Q(q) = - \sum_{q \in \Omega_Q} [m_X^{(0)}(x_h(q, QP)) - m_X^{(0)}(x_l(q, QP))] \log_2 [m_X^{(0)}(x_h(q, QP)) - m_X^{(0)}(x_l(q, QP))] \quad (24)$$

$QP \rightarrow 0$ is a singularity.

3.4. Regular Distortion function

When regular coding is used, the quantization indices Q are just source coded, and no side-information Y is assumed to be available at the decoder. This reconstruction function $\hat{X}_Q(q)$ is then given by:

$$\hat{X}_Q(q) = E(X/Q=q) = E(X/\phi(X, QP)=q) = \frac{\int_{x_l(q)}^{x_h(q)} x f_X(x) dx}{\int_{x_l(q)}^{x_h(q)} f_X(x) dx} = \frac{m_X^{(1)}(x_h(q)) - m_X^{(1)}(x_l(q))}{m_X^{(0)}(x_h(q)) - m_X^{(0)}(x_l(q))} \quad (25)$$

while the expected distortion is given by:

$$D_Q(QP, \sigma_X) = E(D_Q) = \sigma_X^2 - \sum_{q \in \Omega_Q} \frac{\left(\int_{x_l(q)}^{x_h(q)} x f_X(x) dx \right)^2}{\int_{x_l(q)}^{x_h(q)} f_X(x) dx} = \sigma_X^2 - \sum_{q \in \Omega_Q} \frac{\left(m_X^{(1)}(x_h(q)) - m_X^{(1)}(x_l(q)) \right)^2}{\left(m_X^{(0)}(x_h(q)) - m_X^{(0)}(x_l(q)) \right)^2} \quad (26)$$

3.5. Non-parametric representation

When X is Laplacian, and Z is Laplacian or Gaussian, closed-form expressions for the functions: $m_{X/Y}^{(0)}(x, y)$, $m_{X/Y}^{(1)}(x, y)$, $m_X^{(1)}(x)$, $m_X^{(0)}(x)$ and $f_Y(y)$, can be readily obtained. For the Gaussian case, the $erf()$ function is required, but a polynomial approximation provided in *Numerical Recipes* [11] may be used to evaluate it in closed form. The values of the seed functions at a set of given points can be readily computed by evaluating the expressions and boundary conditions above in conjunction with 1-dimensional numerical integration and summation.

We adopted a non-parametric function representation approach, where the above functions are computed over an appropriate sampling grid and stored. In order to reduce the dimensionality of the function representation, unit variance samples ($\sigma_X=1$) are computed. Evaluation of the functions at an arbitrary point involves interpolating from the unit-variance grid and normalizing appropriately. For the sampled points, the values are evaluated based on the chosen Laplacian-Laplacian or Laplacian-Gaussian model.

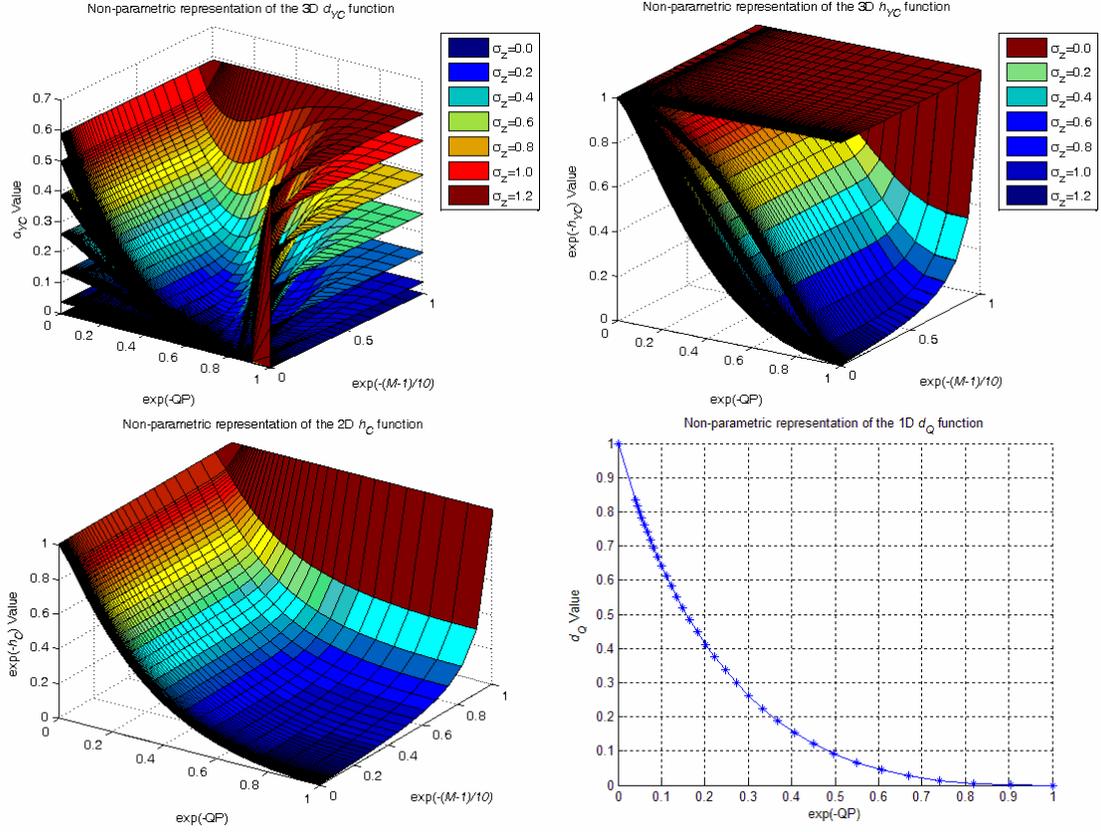


Figure 3. Non-parametric representation of unit-variance seed functions for the Laplacian-Laplacian model

Note that these samples need to be evaluated only once, but the resultant non-parametric representation can be used for any coding application with the same statistical model and the quantizer family. Figure 3 shows the functions that are represented for the uniform deadzone quantizer family for the Laplacian-Laplacian model. For more convenient representation of boundary conditions and better interpolation performance, the axes and the functions are modified by exponentiation as shown.

4. OPTIMUM PARAMETER CHOICE IN A RATE-DISTORTION SENSE

For a given set of parameters $\{QP, S, m\}$ in our coding model, the expected distortion D is given by: $D_{C_1}(QP, 2^{S-1}m, \sigma_Z, \sigma_X) = (\sigma_X)^2 d_{C_1}(QP/\sigma_X, 2^{S-1}m, \sigma_Z/\sigma_X)$, assuming noise-free decoding of the channel coded bit-planes. In practice however, since there is always a small probability of decoding error, this is really a lower bound for the expected distortion.

The expected rate $R = R_{SC} + R_{CC}$ where R_{SC} is the rate for the source coded least significant symbol plane and R_{CC} is the sum of rates for all channel coded bit-planes. The expected source coding rate R_{SC} is given by:

$$R_{SC} = H_C(QP, m, \sigma_X) = h_C(QP/\sigma_X, m) \quad (27)$$

The ideal channel coding rate h_i for the i th bit-plane C_i , $i=1, 2, \dots, S-1$, is the conditional entropy: $H(C_i | C_0, C_1, \dots, C_{i-1}, Y)$, when coded and decoded in the least to most significant order. Since $H(C_i | C_0, C_1, \dots, C_{i-1}, Y) = H(C_0, C_1, \dots, C_{i-1}, C_i | Y) - H(C_0, C_1, \dots, C_{i-1} | Y)$, the ideal rate $h_i = h_{YC}(QP/\sigma_X, 2^i m, \sigma_Z/\sigma_X) - h_{YC}(QP/\sigma_X, 2^{i-1} m, \sigma_Z/\sigma_X)$. Next, we define a function $\lambda(h) > h$ which provides the practical channel coding rate after considering any premium in the channel coding rate one needs to pay over the ideal in order to achieve virtually error-free transmission in practice. The practical channel coding rate r_i for C_i is then $r_i = \lambda(h_i)$. The total channel coding rate R_{CC} is the sum of the rates r_i for all channel coded bit-planes:

$$R_{CC} = \sum_{i=1}^{S-1} \lambda(h_{YC}(QP/\sigma_X, 2^i m, \sigma_Z/\sigma_X) - h_{YC}(QP/\sigma_X, 2^{i-1} m, \sigma_Z/\sigma_X)) \quad (28)$$

We assume a rate-adaptive model for the channel coder, where the bit-planes can be coded at any arbitrary rate as determined by the parameter choice mechanism. Sometimes, the codes for a range of rates may be derived from a common code family. Examples include variable punctured codes derived from a mother lower rate code, the recently proposed LDPC accumulate code, and so on. Alternatively, there can be multiple codes each covering a non-overlapping range of rates. The $\lambda(h)$ function then depends on the particular code used at a particular rate range, its efficiency at the operating rate, as well as factors such as block length. Because it can

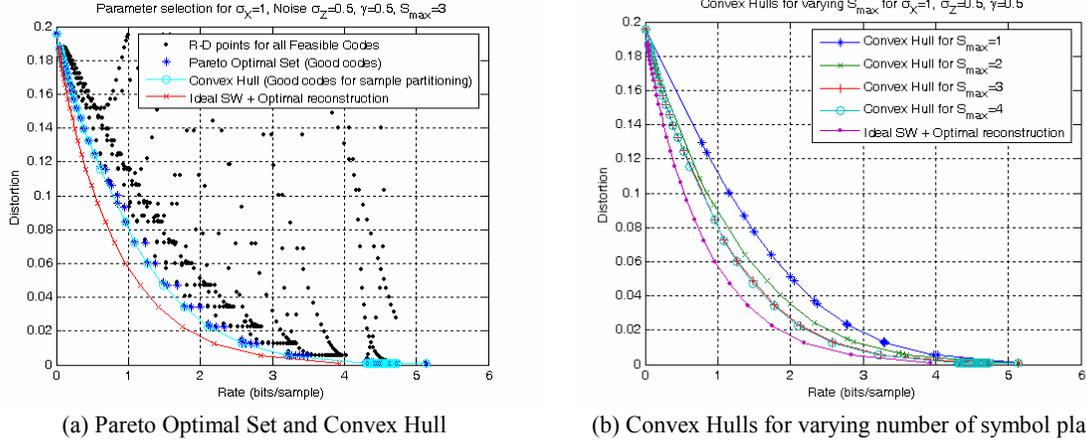


Figure 4. R-D Characteristics for Laplacian $\sigma_x=1$, Laplacian $\sigma_z=0.5$, $\gamma=0.5$, $\Omega_{QP}=\{0.1,0.2,\dots,2.9,3.0\}$

be quite involved to characterize this function accurately, we assume $\lambda(h) = (1+\gamma)h$ in this work, where γ is a constant. However, this simplifying assumption can still be very useful and provide a good understanding of the parameter selection mechanism.

Now that we know how to obtain the distortion and rate for each $\{QP, S, m\}$ combination, we can describe how the parameter selection mechanism for the general Wyner-Ziv problem would ideally work to optimize the triple $\{QP, S, m\}$. Typically, QP is restricted to lie within a finite set Ω_{QP} . If QP is allowed to be continuous then Ω_{QP} could have arbitrarily fine granularity but for the purpose of this work, we assume Ω_{QP} to be discrete, S can have values in $\{1, 2, \dots, S_{\max}\}$ from practicality considerations, and m can take values in $\{1, 2, \dots, m_{\max}, \infty\}$. At higher values of m there is typically little or no difference in rate-distortion performance, and as such it is better to shut off coset computation and channel coding entirely after m_{\max} . For each allowable $\{QP, S, m\}$ combination, the R-D point is obtained from the seed functions. From these points, only the Pareto Optimal subset of points can be regarded as ‘good’ points for coding. For these set of points, there is no other feasible point that yields a lower rate for a lower distortion. For a given target QP_b , the target distortion D_t is obtained by evaluating $D_t = D_Q(QP_b, \sigma_x)$, and the point from the Pareto Optimal set that yields the least rate with distortion not larger than D_t is chosen. Once the optimal $\{QP, S, m\}$ has been obtained, the channel coding rates can also be readily read off.

A somewhat better strategy to use is to seek the convex hull, which is a subset of the Pareto Optimal set. The mechanism to generate the convex hull is outlined in [9][10]. Since the convex hull has fewer points joined by linear segments, the coding strategy for an intermediate point at a given D_t is to pseudo-randomly select between two nearest codes (convex hull vertices) with appropriate probabilities. In other words, the samples are partitioned into two separate smaller sequences at appropriate proportion, and each is coded with a different set of parameters. This strategy would yield a Rate-distortion point that is exactly on the convex hull.

Figure 4(a) illustrates the parameter choice mechanism in terms of the RD characteristics for the Laplacian-Laplacian model. Figure 4(b) shows how the convex hull changes with S_{\max} . As seen, there is little or no gain in increasing the number of symbols beyond a few. Table 1 presents a table generated showing the mapping from QP_t to the optimal parameters assuming partitioned codes on the convex hull. Note that codes with $QP = \infty$, $S = 0$, $m = 1$ correspond to the zero-rate code case. Thus at lower qualities, we quickly reach the situation where the optimal coding is simply not to send anything at all (zero-rate coding), but just reconstruct using the side-information. The rates r_{ij} in the table correspond to additional bits that need to be sent besides the side-information Y . In conventional channel coding terms, the code rate is then $1/(1+r_{ij})$. If a systematic channel code with this rate $1/(1+r_{ij})$ is used, only the parity bits are transmitted. As seen from the table, we see that only high rate codes (small r_{ij}) are useful, especially for the higher bit-planes. Since high rate codes are also more likely to be numerically unstable, in many practical scenarios it may be sufficient to use at most a single channel coded bit-plane (*i.e.* $S_{\max}=2$).

Table 1. Mapping from QP_t to optimal WZ parameters for the Laplacian-Laplacian model with $\sigma_x=1$, $\sigma_z=0.5$, $\gamma=0.5$, $S_{\max}=3$, $\Omega_{QP}=\{0.1,0.2,\dots,2.9,3.0\}$, $m_{\max}=32$. The first code $\{QP_1, S_1, m_1, r_{11}, r_{21}\}$ is chosen with probability α , while the second code $\{QP_2, S_2, m_2, r_{12}, r_{22}\}$ is chosen with probability $(1-\alpha)$. r_{ij} is the channel coding rate for the symbol plane C_i in the j th code.

QP_t	QP_1	S_1	m_1	r_{11}	r_{21}	QP_2	S_2	m_2	r_{12}	r_{22}	α
0.1	0.1	3	22	0.0898	0.0014	0.1	3	23	0.0750	0.0009	0.1805
0.2	0.3	3	5	0.2503	0.0165	0.2	3	8	0.2285	0.0125	0.0151
0.3	0.4	3	4	0.2002	0.0104	0.3	3	5	0.2503	0.0165	0.1248
0.4	0.5	3	3	0.2274	0.0143	0.4	3	4	0.2002	0.0104	0.3854

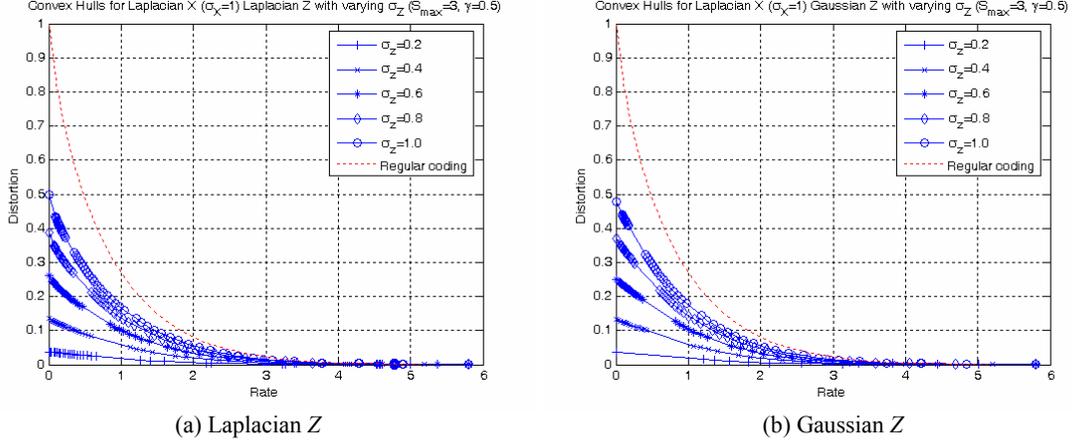


Figure 5. R-D Characteristics for unit variance Laplacian X with Laplacian or Gaussian Z with varying variance σ_Z , compared to regular coding

0.5	0.6	3	2	0.3602	0.0431	0.5	3	3	0.2274	0.0143	0.7443
0.6	0.8	3	2	0.1818	0.0087	0.7	3	2	0.2578	0.0194	0.5528
0.7	1.3	3	1	0.3224	0.0296	0.9	3	2	0.1271	0.0039	0.2074
0.8	1.3	3	1	0.3224	0.0296	0.9	3	2	0.1271	0.0039	0.9260
0.9	1.8	3	1	0.1591	0.0048	1.7	3	1	0.1835	0.0069	0.4643
1.0	∞	0	1	0	0	3.0	3	1	0.0287	0.0001	0.0874
1.1	∞	0	1	0	0	∞	0	1	0	0	1.0000
...	∞	0	1	0	0	∞	0	1	0	0	1.0000

In a practical codec, normalized tables (i.e. $\sigma_X=1$) of the above form may be pre-computed and stored for a range of values of σ_Z in small increments. Furthermore the granularity of the normalized QP and QP_t values must be sufficiently fine. Then for a given QP_t , σ_X and σ_Z , one can look up the closest entry corresponding to QP_t/σ_X in the normalized table that is closest to σ_Z/σ_X , and then scale all the QP values read from the entry by σ_X . The closest allowable QP to this scaled value may then be used for actual coding.

Figure 5 shows how the R-D characteristics change with variance of Z for both the Laplacian and Gaussian models for Z . Note that even for very high values of σ_Z , the Wyner-Ziv coder does significantly better than regular coding.

The seed functions can also be used to obtain the optimal parameters for the simpler Slepian-Wolfe problem, where QP is fixed and only $\{S, m\}$ are to be selected. The following strategy is used for this search: For each S in $\{1, 2, \dots, S_{\max}\}$, search m in increasing order until the conditional entropy of the coded part $H_{YC}(QP, 2^{S-1}m, \sigma_Z, \sigma_X)$ comes very close to the ideal Slepian-Wolfe rate $H_{YC}(QP, \infty, \sigma_Z, \sigma_X)$ for the given QP . Specifically, for each S , find the smallest m such that:

$$H_{YC}(QP, \infty, \sigma_Z, \sigma_X) - H_{YC}(QP, 2^{S-1}m, \sigma_Z, \sigma_X) < \varepsilon \quad (29)$$

for a sufficiently small ε . Compute the total rate $R_{SC} + R_{CC}$ for this $\{S, m\}$ combination. Finally pick the combination that yields the least rate over the set of S values. Note that while there is some redundancy in the search for m for the different S values, eliminating such redundancies is a trivial extension of the above algorithm.

A fast parameter choice algorithm for the Wyner-Ziv problem can now be conceived of based on decomposing the search into two smaller steps, the first for finding the right QP based on a given QP_t , and the second for finding the best m and S combination using the Slepian-Wolfe search above. This algorithm is more appropriate when fast parameter selection is desired during encoding based on on-the-fly estimates for σ_X and σ_Z . For a given QP_t , the distortion is computed using the regular distortion function to yield the target distortion D_t . Thereafter, we find QP such that the coset distortion function computed for $M=\infty$ matches D_t . In other words we solve for QP such that:

$$D_{YC}(QP, \infty, \sigma_Z, \sigma_X) = D_Q(QP, \sigma_X) = D_t \quad (30)$$

The solution can be readily obtained by a search over QP . Then we find the best M for that QP using the mechanism previously outlined. Figure 6 provides the R-D characteristics for fast vs. optimal search. As we see, in most cases, the fast method yields points close to the convex hulls.

5. PRACTICAL CODING RESULTS

The proposed framework was implemented in the following practical setting. An adaptive arithmetic coder was used as the source coder for the least significant plane. For the channel coded more significant bit-planes, a set of systematic repeat-accumulate (RA) codes [12][13] was used, where only the parity bits are transmitted. The advantage of using these codes is primarily in the flexibility of selecting arbitrary rates between 0 and 1, while still maintaining performance competitive with Turbo or LDPC codes. Furthermore,

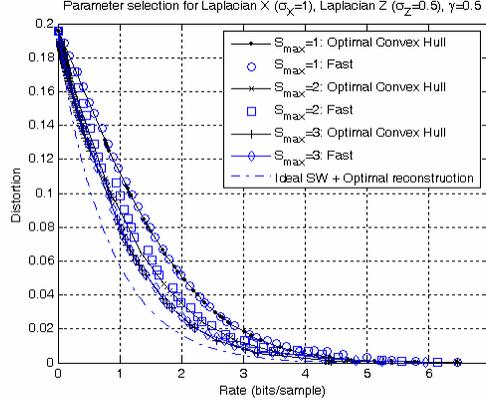


Figure 6. Fast versus Optimal parameter selection

encoding is extremely simple, which makes it an attractive choice in many low encoding complexity applications. In particular, we use Irregular Repeat Accumulate codes (IRA) where the grouping factor and the repetition profiles were optimized and stored based on the technique in [14] for the BIAWGNC case for a variety of rates. Even though the channel here is different we hope that by matching the conditional entropy to the code rate, we would get performance close to the optimal. For a given desired rate obtained by the parameter selection process, the degree distributions corresponding to the closest rate in the pre-computed tables is used, but the grouping factor is tweaked to yield the exact desired rate. A soft decoder is used, where decoding for the channel coded bit-planes is conducted progressively from C_1 through C_{S-1} . For decoding the i th symbol plane C_i – the soft probabilities: $P(C_i/Y, C_0, C_1, \dots, C_{i-1})$ is computed for the systematic bits, and provided as input to a soft decoder. There is no uncertainty in the parity bits since they are transmitted noise-free.

$$p(C_i = c_i / \{C_k = c_k : k \in \{0, 1, \dots, i-1\}, Y = y\}) = \frac{p(\{C_k = c_k : k \in \{0, 1, \dots, i\} / Y = y\})}{p(\{C_k = c_k : k \in \{0, 1, \dots, i-1\} / Y = y\})} = \frac{\sum_{q \in \Omega_Q} \pi(q, y, QP) \sum_{q \in \Omega_Q} [m_{X/Y}^{(0)}(x_i(q), y) - m_{X/Y}^{(0)}(x_i(q), y)]}{\sum_{q \in \Omega_Q} \pi(q, y, QP) \sum_{q \in \Omega_Q} [m_{X/Y}^{(0)}(x_i(q), y) - m_{X/Y}^{(0)}(x_i(q), y)]} = \frac{\sum_{q \in \Omega_Q} \xi_k^i(q) = c_i \forall k \in \{0, 1, \dots, i\}}{\sum_{q \in \Omega_Q} \xi_k^i(q) = c_i \forall k \in \{0, 1, \dots, i-1\}} \quad (31)$$

We found that the difference between the theoretical source coding rate and the practical source coding rate with the arithmetic coder was minimal. On the other hand, $\gamma = 0.5$ or more was required on an average to get good channel coding performance (low bit-error rate) with our IRA based channel coder. As a result, we used this value in the parameter selection process using our simplified $\lambda(h) = (1 + \gamma)h$ model. However, we also observed that the performance degrades for parity rates < 0.1 (i.e. high rate codes). So ideally, the optimization process should have used a $\lambda(h)$ function where $\gamma(h) = \lambda(h)/h - 1$ is not constant, but increases at lower rates. While we expect optimized irregular LDPC codes to perform better, it is unlikely that for a practical block length of up to a few thousand bits, and very low parity bit rates, the parameter γ can be reduced to lower than 0.25 even for the best of channel codes.

To test the performance, a Laplacian source X with a given variance σ_X was generated; followed by adding a Laplacian noise Z with a given variance σ_Z to it, in order to obtain the side information Y . The parameter selection process was then invoked to obtain the optimal coding parameters, and the source X is then coded using them. The coded data is decoded based on the side-information Y . In the results we present, we assumed $S_{max} = 2$ in order to allow a maximum of one channel coded bit-plane. The block size was chosen as 6336 bits.

Figure 7 presents the practical coding results with arithmetic and IRA coders compared against both the theoretical Pareto Optimal set of points optimized for Laplacian X ($\sigma_X = 1$), Gaussian Z ($\sigma_Z = 1$), and $\gamma = 0.5$. The parameter choice mechanism used is one where the nearest Pareto Optimal point is chosen for a given target QP_r . Also shown is the RD characteristics for ideal Slepian Wolfe coding followed by optimal reconstruction. Figure 7(a) shows the results for $S_{max} = 1$, where a single source coded symbol plane is used. Figure 7(b) shows the results for $S_{max} = 2$, where one channel coded symbol plane is used. We observe that the distortions in the practical coding points in Figure 7(b) are slightly worse than that of the corresponding Pareto Optimal points because of a few uncorrected errors. But generally, this case yields better RD performance than pure source coding. In both cases we also show the results when there is a channel mismatch, so that the actual noise Z has variance 0.6 rather than the designed 0.5. This is where the pure source coding strategy is a clear winner. The channel coding strategy yields a large number of channel decoding errors, degrading the performance substantially.

Based on the above results we conclude that in applications where there is a large margin of error in the channel correlation estimation, the pure source coding strategy is actually quite reasonable. Alternatively, we could design for a value of γ larger than what just the inefficiency in channel coding performance would need, but that would automatically put more bits on the source coding part.

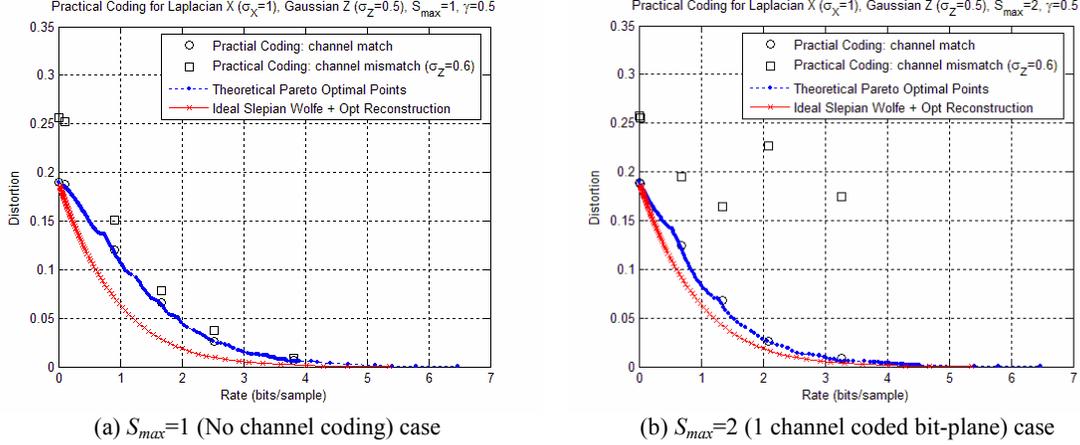


Figure 7. Practical coding results for Laplacian X ($\sigma_X=1$), Gaussian Z ($\sigma_Z=1$) compared against theoretical and ideal, for the cases: (a) when no channel coding is used (b) a single channel coded bit-plane is used along with source coding. Also shown are coding results when there is a channel estimation mismatch.

We have actually used the pure source coding strategy in applications [9][15], where the inaccuracy of correlation estimation is deemed too large.

6. APPENDIX

We provide expressions for Laplacian X and for two particular cases for Z , Gaussian and Laplacian, for use in generation of the non-parametric representation of the seed functions, computation of the optimal reconstruction functions, as well as computation of the channel soft conditional probabilities for channel decoding.

6.1. Expressions for Laplacian X Gaussian noise Z

We first specialize for the case of Laplacian X and Gaussian Z , *i.e.*:

$$f_X(x) = \frac{1}{\sqrt{2}\sigma_X} e^{-\frac{|\sqrt{2}x|}{\sigma_X}}, \quad f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma_Z} e^{-\frac{1}{2}\left(\frac{z}{\sigma_Z}\right)^2} \quad (32)$$

In the following, we assume:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (33)$$

Then, defining $\beta(x) = e^{-\frac{\sqrt{2}x}{\sigma_X}}$, we have

$$m_X^{(0)}(x) = \begin{cases} \frac{\beta(x)}{2}, & x \leq 0 \\ 1 - \frac{1}{2\beta(x)}, & x > 0 \end{cases} \quad m_X^{(1)}(x) = \begin{cases} \frac{\beta(x)}{2\sqrt{2}}(\sqrt{2}x - \sigma_X), & x \leq 0 \\ -\frac{1}{2\sqrt{2}\beta(x)}(\sqrt{2}x + \sigma_X), & x > 0 \end{cases} \quad (34)$$

Further defining:

$$\gamma_1(x) = \text{erf}\left(\frac{\sigma_X x - \sqrt{2}\sigma_Z^2}{\sqrt{2}\sigma_X\sigma_Z}\right), \quad \gamma_2(x) = \text{erf}\left(\frac{\sigma_X x + \sqrt{2}\sigma_Z^2}{\sqrt{2}\sigma_X\sigma_Z}\right) \quad (35)$$

and using $Y=X+Z$, we have:

$$f_{XY}(x, y) = \frac{1}{2\sqrt{\pi}\sigma_X\sigma_Z} e^{-\frac{|x\sqrt{2}|}{\sigma_X}} e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_Z}\right)^2} \quad (36)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \frac{1}{2\sqrt{2}\beta(y)\sigma_X} e^{-\sigma_Z^2/\sigma_X^2} [\gamma_1(y)+1.0 - \beta(y)^2(\gamma_2(y)-1.0)]$$

$$f_{X|Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{\sqrt{2}\beta(y)}{\sqrt{\pi}\sigma_Z} \frac{e^{-\frac{|x\sqrt{2}|}{\sigma_X}} e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_Z}\right)^2}}{[\gamma_1(y)+1.0 - \beta(y)^2(\gamma_2(y)-1.0)]}$$

Given $f_{X|Y}(x, y)$, the moments can now be computed:

$$m_{X/Y}^{(0)}(x, y) = \begin{cases} \frac{1}{[\gamma_1(y) + 1.0 - \beta(y)^2(\gamma_2(y) - 1.0)]} \beta(y)^2 [1 - \operatorname{erf}(\frac{\sigma_X(y-x) + \sqrt{2}\sigma_Z^2}{\sqrt{2}\sigma_X\sigma_Z})], & x \leq 0 \\ 1 - \frac{1}{[\gamma_1(y) + 1.0 - \beta(y)^2(\gamma_2(y) - 1.0)]} [1 + \operatorname{erf}(\frac{\sigma_X(y-x) - \sqrt{2}\sigma_Z^2}{\sqrt{2}\sigma_X\sigma_Z})], & x > 0 \end{cases} \quad (37)$$

$$m_{X/Y}^{(1)}(x, y) = \begin{cases} \frac{\beta(y)^2 [y + \sqrt{2} \frac{\sigma_Z^2}{\sigma_X}] [1 - \operatorname{erf}(\frac{\sigma_X(y-x) + \sqrt{2}\sigma_Z^2}{\sqrt{2}\sigma_X\sigma_Z})] - \frac{\sqrt{2}}{\sqrt{\pi}} \sigma_Z \beta(x)^2 e^{-\frac{1}{2}(\frac{\sigma_X(y-x) - \sqrt{2}\sigma_Z^2}{\sigma_X^2})^2}}{[\gamma_1(y) + 1 - \beta(y)^2(\gamma_2(y) - 1)]}, & x \leq 0 \\ -\beta(y)^2 [y + \sqrt{2} \frac{\sigma_Z^2}{\sigma_X}] (\gamma_2(y) - 1) + [y - \sqrt{2} \frac{\sigma_Z^2}{\sigma_X}] [\gamma_1(y) - \operatorname{erf}(\frac{\sigma_X(y-x) - \sqrt{2}\sigma_Z^2}{\sqrt{2}\sigma_X\sigma_Z})] - \frac{\sqrt{2}}{\sqrt{\pi}} \sigma_Z e^{-\frac{1}{2}(\frac{\sigma_X(y-x) - \sqrt{2}\sigma_Z^2}{\sigma_X^2})^2}}{[\gamma_1(y) + 1 - \beta(y)^2(\gamma_2(y) - 1)]}, & x > 0 \end{cases}$$

A special case used for the optimal reconstruction and distortion functions in the zero-rate case is when $x \rightarrow \infty$. In this case,

$$m_{X/Y}^{(1)}(\infty, y) = \frac{-\beta(y)^2 [y + \sqrt{2} \frac{\sigma_Z^2}{\sigma_X}] (\gamma_2(y) - 1) + [y - \sqrt{2} \frac{\sigma_Z^2}{\sigma_X}] (\gamma_1(y) + 1)}{[\gamma_1(y) + 1 - \beta(y)^2(\gamma_2(y) - 1)]} = y - \sqrt{2} \frac{\sigma_Z^2}{\sigma_X} \frac{[\gamma_1(y) + 1 + \beta(y)^2(\gamma_2(y) - 1)]}{[\gamma_1(y) + 1 - \beta(y)^2(\gamma_2(y) - 1)]} \quad (38)$$

6.2. Expressions for Laplacian X Laplacian noise Z

We next specialize for the case of Laplacian X and Laplacian Z , *i.e.*:

$$f_X(x) = \frac{1}{\sqrt{2}\sigma_X} e^{-\frac{|x|}{\sigma_X}}, \quad f_Z(z) = \frac{1}{\sqrt{2}\sigma_Z} e^{-\frac{|z|}{\sigma_Z}} \quad (39)$$

Defining:

$$\alpha(x) = e^{-\frac{x}{\sigma_X}}, \quad \beta(x) = e^{-\frac{x}{\sigma_Z}}, \quad \gamma(x) = (\sigma_X \alpha(x) - \sigma_Z \beta(x)) \quad (40)$$

Eq. 34 still applies for $m_X^{(0)}(x)$ and $m_X^{(1)}(x)$. Further, using $Y=X+Z$, we have

$$f_{X/Y}(x, y) = \frac{1}{2\sigma_X\sigma_Z} e^{-\sqrt{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Z^2}} |x-y|}$$

$$f_Y(y) = \begin{cases} \frac{\gamma(|y|) e^{-\sqrt{2}|y|(\frac{1}{\sigma_X} + \frac{1}{\sigma_Z})}}{\sqrt{2}(\sigma_X^2 - \sigma_Z^2)}, & \sigma_Z \neq \sigma_X \\ \frac{(\sigma_X + \sqrt{2}|y|) e^{-\frac{|y|}{\sigma_X}}}{2\sqrt{2}\sigma_X^2}, & \sigma_Z = \sigma_X \end{cases} \quad f_{X/Y}(x, y) = \begin{cases} \frac{e^{-\sqrt{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Z^2}} |x-y|} (\sigma_X^2 - \sigma_Z^2)}{\sqrt{2}\sigma_X\sigma_Z\gamma(|y|)}, & \sigma_Z \neq \sigma_X \\ \frac{\sqrt{2} e^{-\sqrt{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Z^2}} |x-y|}}{(\sigma_X + \sqrt{2}|y|)}, & \sigma_Z = \sigma_X \end{cases} \quad (41)$$

The partial moments can now be calculated as follows:

$$m_{X/Y}^{(0)}(x, y) = \begin{cases} 1 - m_{X/Y}^{(0)}(-x, y) & y \leq 0 \\ m_{X/Y}^{(0)}(x, y) & y > 0 \end{cases} \quad \text{where } m_{X/Y}^{(0)}(x, y) = \begin{cases} \frac{(\sigma_X - \sigma_Z)\beta(x+y)\alpha(x)}{2\gamma(|y|)} & x \leq 0 \\ \frac{\beta(|y|)(\sigma_X + \sigma_Z)\alpha(x)/\beta(x) - 2\sigma_Z}{2\gamma(|y|)} & x \leq |y| \\ 1 - \frac{(\sigma_X - \sigma_Z)\alpha^2(|y|)\beta(|y|)}{2\gamma(|y|)\alpha(x)\beta(x)} & x > |y| \end{cases} \quad \left. \begin{matrix} \\ \\ \\ \end{matrix} \right\} \begin{matrix} \frac{\sigma_X\beta^2(x)}{2(\sigma_X + \sqrt{2}|y|)} & x \leq 0 \\ \frac{\sigma_X + 2\sqrt{2}x}{2(\sigma_X + \sqrt{2}|y|)} & x \leq |y| \\ 1 - \frac{\sigma_X\beta^2(|y|)}{2(\sigma_X + \sqrt{2}|y|)\beta^2(x)} & x > |y| \end{matrix} \right\} \text{for } \sigma_X = \sigma_Z$$

$$m_{X/Y}^{(1)}(x, y) = \begin{cases} m_{X/Y}^{(1)}(-x, y) - m_{X/Y}^{(1)}(\infty, y) & y \leq 0 \\ m_{X/Y}^{(1)}(x, y) & y > 0 \end{cases} \quad \text{where:}$$

$$m_{X/Y}^{(1)}(x, y) = \begin{cases} \frac{(\sqrt{2}x(\sigma_X + \sigma_Z) - \sigma_X\sigma_Z)(\sigma_X - \sigma_Z)\alpha(x)\beta(x)\beta(|y|)}{2\sqrt{2}\gamma(|y|)(\sigma_X + \sigma_Z)} & x \leq 0 \\ \frac{\beta(|y|)\{(\sigma_X + \sigma_Z)^2(\sqrt{2}x(\sigma_X - \sigma_Z) - \sigma_X\sigma_Z)\alpha(x)/\beta(x) + 4\sigma_X^2\sigma_Z^2\}}{2\sqrt{2}\gamma(|y|)(\sigma_X^2 - \sigma_Z^2)} & x \leq |y| \\ \frac{2\sqrt{2}\gamma\beta(|y|)\sigma_X(\sigma_X^2 - \sigma_Z^2) + 4\sigma_X^2\sigma_Z^2(\beta(|y|) - \alpha(|y|)) - (\sigma_X - \sigma_Z)^2(\sigma_X\sigma_Z + \sqrt{2}x(\sigma_X + \sigma_Z))}{2\gamma(|y|)} \frac{\alpha^2(|y|)\beta(|y|)}{\alpha(x)\beta(x)} & x > |y| \end{cases} \quad (42)$$

$$\left. \begin{matrix} \\ \\ \\ \end{matrix} \right\} \begin{matrix} \frac{\sqrt{2}\sigma_X\beta^2(x)(2\sqrt{2}x - \sigma_X)}{8(\sigma_X + \sqrt{2}|y|)} & x \leq 0 \\ \frac{\sqrt{2}(4x^2 - \sigma_X^2)}{8(\sigma_X + \sqrt{2}|y|)} & x \leq |y| \\ \frac{1}{2}|y| - \frac{\sigma_X\beta^2(|y|)(\sqrt{2}\sigma_X + 4x)}{8(\sigma_X + \sqrt{2}|y|)\beta^2(x)} & x > |y| \end{matrix} \right\} \text{for } \sigma_X = \sigma_Z$$

Also note:

$$m_{X/Y}^{(1)}(\infty, y) = \text{sgn}(y)m_{X/Y+}^{(1)}(\infty, y), \text{ where:} \quad (43)$$

$$m_{X/Y+}^{(1)}(\infty, y) = \begin{cases} \frac{|y|\alpha(|y|)\sigma_X(\sigma_X^2 - \sigma_Z^2) + \sqrt{2}\sigma_X^2\sigma_Z^2(\beta(|y|) - \alpha(|y|))}{\gamma(|y|)(\sigma_X^2 - \sigma_Z^2)}, & \sigma_X \neq \sigma_Z \\ |y|/2, & \sigma_X = \sigma_Z \end{cases}$$

7. REFERENCES

- [1] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inf. Theory, vol. IT-19, pp. 471–480, July 1973.
- [2] A. D. Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inf. Theory, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [3] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," in Proc. IEEE Data Compression Conf., 1999, pp. 158–167.
- [4] R. Puri and K. Ramchandran, "PRISM: A 'reversed' multimedia coding paradigm," Proc. IEEE Int. Conf. Image Processing, Barcelona, Spain, 2003.
- [5] A. Aaron, R. Zhang, B. Girod, "Transform-domain Wyner-Ziv coding for video," Proc. Visual Communications and Image Processing, San Jose, California, SPIE vol. 5308, pp. 520-528, Jan. 2004.
- [6] Q. Xu, Z. Xiong, "Layered Wyner-Ziv video coding," Proc. Visual Comm. and Image Proc., San Jose, California, SPIE vol. 5308, pp. 83-91, 2004.
- [7] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding," Proceedings of the IEEE, Special Issue on Video Coding and Delivery, vol. 93, no. 1, pp. 71-83, January 2005.
- [8] Z. Xiong, A. D. Liveris, S. Cheng, "Distributed source coding for sensor networks," IEEE Signal Proc. Magazine, pp. 80-94, Sept 2004.
- [9] D. Mukherjee, B. Macchiavello, R. L. de Queiroz, "A simple reversed complexity Wyner-Ziv video coding mode based on a spatial reduction framework," Proc. Visual Comm. and Image Proc., San Jose, California, Jan 2007.
- [10] D. Mukherjee, "Optimal parameter choice for Wyner-Ziv coding of Laplacian sources with decoder side-information," HP Labs Tech. Report, HPL-2007-34.
- [11] William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling, Numerical Recipes in C, Second Edition, Cambridge University Press, 1992.
- [12] D. Divsalar, H. Jin, R. J. McEliece, "Coding theorems for Turbo-like codes," Proc. 36th Allerton Conf. on Communications, Control and Computing, Allerton, Illinois, Sept. 1998, pp. 201-210.
- [13] H. Lin, A. Khandekar, R. McEliece, "Irregular repeat-accumulate codes," Proc. 2nd Int. Symposium Turbo codes and Related Topics, Brest, France, Sept. 2000, pp. 1-8.
- [14] A. Roumy, S. Guemghar, G. Caire, S. Verdu, "Design Methods for Irregular Repeat-Accumulate codes," IEEE Trans. Information Theory, vol. 50, no. 8, Aug 2004.
- [15] R. Samadani, D. Mukherjee, "Photoplus: Auxiliary information for printed images based on distributed source coding," Proc. Visual Comm. and Image Proc., San Jose, California, Jan 2008.