Optimizing bit-by-bit power for minimal distortion*

Hiram Rodríguez-Díaz and Don H. Johnson†
Computer & Information Technology Institute
Department of Electrical & Computer Engineering, MS 366
Rice University
Houston, TX 77251–1892
hiramrd@rice.edu, dhj@rice.edu

March 28, 2003

Abstract

When transmitting a sampled signal digitally, an alternative to fixed power/bit is vary the power allocated to each bit to minimize signal distortion. We adjust the power here by allowing each bit interval's duration, whether a data bit or an error-correction bit, to vary and seek the set of bit durations that minimize the mean-squared reconstruction error. For smaller signal-to-noise ratios (SNR), the optimal solution amounted to only sending the most significant bits, with bit duration decreasing as the bit's significance to the sample decreases. As SNR increases, the optimal solution has more bits being transmitted, with durations becoming equal at high SNRs. The optimal solution yields significant gains in mean-squared error (several dB) over a wide range of channel signal-to-noise ratios when compared that provided by equal-duration bit intervals. When block error correction is performed, we derive the optimal decoder for variable bit-error probabilities and show that gains obtain here as well. For smaller signal-to-noise ratios, using no error correction was optimal as it expends power without sufficiently decreasing the mean-squared error.

Keywords: variable bit interval, distortion, power allocation, joint source-channel coding

*This work supported by the National Science Foundation under Grant CCR-0105558. This work was presented at ICASSP'03 [4].

†Contact person

1 Introduction

In any communication system, like telephone systems, computer networks, or cellular networks, error detection and correction has been an important issue. Many different strategies have been developed for achieving the best performance of the system while using the least amount of resources. Performance for digital systems is typically quantified by considering bit-error or block-error probabilities. When each bit is as important as the others, minimizing the average bit-error probability make sense. However, when some bits are more important than others to the fidelity of the received data, minimizing this error criterion may not yield the highest fidelity output. For example, in audio and video applications, the digital data represent samples from an A/D converter. In such cases, errors in the more significant bits are far more important than in others. When the digital data are K-bit quantized signal amplitudes, the transmitted amplitude s is related to the individual bits (to within a scaling and offset) as

$$s = \sum_{k=0}^{K-1} b_k 2^k \tag{1}$$

When communication errors occur, the mean-squared distortion between the transmitted and received amplitude is

$$\text{mse} = \sum_{k=0}^{K-1} P_e^{(k)} 2^{2k} \tag{2}$$

where $P_e^{(k)}$ is the probability the k^{th} bit is received in error. Thus, if the error probabilities were all equal, the most significant bit contributes $2^{2(K-1)}$ more to the mean-squared error than the least significant bit. A smaller mean-squared error could result if the error probability for the most significant bit were reduced and the error probability for the least significant bit were increased according to some kind of tradeoff.

When some data blocks are more important than others, such as packet headers, extra error protection in the form of stronger error correcting codes could be employed. It would seem that a similar approach could be applied to individual bits as well: could extra error protection be given bits on the basis of their importance to the ultimate received signal? Our approach is along this line; rather than a digital error correction code, we use what essentially is an *analog* repetition code. As shown in figure 1, we use variable-duration bit intervals to represent the various bits in the sample. Because we assume a white Gaussian noise channel and a matched filter receiver, each bit has an error probability different from that of other bits that have different transmission interval durations. Assuming a BPSK signal set, the error probability is $P_e = Q\left(\sqrt{2A^2T/N_0}\right)$ where A is the amplitude of the received sinusoidal signal. Longer bit intervals, which yield a smaller P_e, should be used for the most

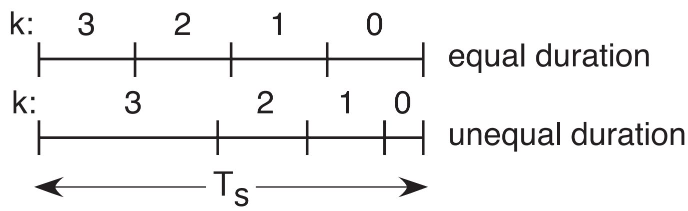

Figure 1: Normally, bit interval durations are equal. In our proposed scheme, they differ, with much more transmission time allocated to the most significant bit (here $k = 3$) than the least significant bit ($k = 0$). In this way, mean-squared error can be dramatically reduced. The total time to transmit each sampled signal is constrained to be no more than the sampling interval T_s.

significant bits. By varying bit-interval duration, we are adjusting the transmitter power allocated to each bit. The bit-interval durations are constrained since the time taken to transmit each sample must be less than or equal to the sampling interval T_s. Conseqently, as shown in figure 1, the sum of the transmission intervals must be no more than T_s in either the equal or unequal bit interval case. From a power perspective, we are optimizing the bit-level allocation of a fixed amount of power available for each data block. Because of the nonlinear nature of $Q(\cdot)$, how the optimal bit-interval durations should be chosen subject to this constraint is not apparent.

When digital error correction is used, it could be advantageous to allow the transmission interval durations for *all* bits—data and error correction bits—to be optimized for minimum distortion. Because unequal bit error probabilities would result, code design and how to optimally decode need to be rethought. We take a detection theory approach here to derive an optimal physical layer transmission scheme and the accompanying decoding rule that applies whether coding is used or not. This approach creates an distortion function that we seek to optimize with respect to bit interval durations (constrained to sum to the sampling interval) and with respect to decoding rules. The resulting transmission scheme amounts to solving a joint source-channel coding problem by manipulating the power allocated to each bit so as to minimize the distortion of the reconstructed signal.

The use of power control is an old one [2]. However, power control typically applies to time durations much longer than that taken by individual bits. In cellular systems, power control is exercised over times measured in at least seconds. By varying the bit-interval duration, we are using a variable-rate transmission scheme. Again, variable datarate transmission has a long history [1], which amounts to adapting *all* bit interval durations to channel conditions. We want to explore adjusting the instantaneous datarate with the idea of minimizing the distortion of received sampled signals. Variable bit-interval transmission has been specified in an optical network application [3], with the idea of using different bit durations according to whether the bits occur in the preamble or in the data portion of the transmission. Furthermore, bit intervals are one, two, or three basic interval units long, and all data bits have the same duration. We vary bit durations on a continuum, seeking the minimal distortion durations for both data and error-correcting bits.

2 Results

Let the index j denote a value for the transmitted sampled amplitude, which we take here to be an integer in the range $[0, 2^{K-1}]$ (see equation 1). The decoded amplitude is indexed by i, and lies in the same interval. We take the distortion D between the received and transmitted amplitudes to be represented by

$$D = \sum_{i,j} \pi_j C_{i,j} \Pr[i|j] \ , \tag{3}$$

where $\Pr[i|j]$ is the probability that amplitude i is received given that amplitude j was transmitted and π_j is the *a priori* probability that amplitude j was sent. The key quantity is the so-called Bayes' cost $C_{i,j}$, the impact of receiving amplitude i when amplitude j was indeed sent. Many choices for the Bayes' cost can be made. For example, if $C_{j,j} = 0$ and $C_{i,j} = 1$, then the nature of the amplitude error doesn't matter and optimizing the distortion amounts to minimizing the average probability of error [5]. On the other hand, if $C_{i,j} = (i-j)^2$, then mean-squared error defines the distortion. Many distortion criteria, even non-symmetric ones, can be specified this way. We focus here on the mean-squared error choice.

We assume that N bits are used to transmit the K data bits ($N \geq K$). The additional $N - K$ bits provide some measure of error correction. We denote by $\mathbf{b}^{(j)}$ the transmitted N-bit sequence corresponding to amplitude j and by $\mathbf{b}^{(i)}$ the decoded bit sequence derived from the transmission of $\mathbf{b}^{(j)}$. The data bits constituting the sample amplitude are derived from these. The probability that the decoder yields the bit sequence $\mathbf{b}^{(i)}$ when $\mathbf{b}^{(j)}$ was transmitted equals

$$\Pr[i|j] = \sum_{\mathbf{b}^{(l)} \in \mathcal{R}_i} \prod_{n=0}^{N-1} [P_e^{(n)}]^{b_n^{(l)} \oplus b_n^{(j)}} [1 - P_e^{(n)}]^{1 - b_n^{(l)} \oplus b_n^{(j)}}$$

Here, $b_n^{(l)}$ denotes the n^{th} bit in the transmitted sequence that corresponds to the l^{th} N-bit sequence and $b_n^{(j)}$ the same bit in the transmitted amplitude j. The notation $b_n^{(l)} \oplus b_n^{(j)}$ means the modulo-2 sum of the bits, which equals zero when the bits agree and one when they differ. $P_e^{(n)} = Q\left(\sqrt{2A^2 T_n / N_0}\right)$ denotes the probability the n^{th} bit is received in error. T_n denotes the duration assigned to the n^{th} bit. The product denotes the probability that a given bit sequence $\mathbf{b}^{(l)}$ is received. To find the amplitude, the decoder defines decoding regions $\mathcal{R}_i$ which includes all possible bit sequences that correspond to amplitude i.

To minimize the distortion, we want to chose the bit-interval durations T_n and the decoding regions $\mathcal{R}_i$ that jointly minimize (3) to yield the optimal distortion D^* subject to contraints on the word duration and the decision regions' properties.

$$D^* = \min_{\{\mathcal{R}_i\},\{T_n\}} \sum_{i,j} \pi_j C_{i,j} \Pr[i|j] \quad \text{subject to } \sum T_n = T_s, \bigcup \mathcal{R}_i = \{\mathbf{b}\}, \bigcap_{i_1 \neq i_2} \mathcal{R}_{i_1}\mathcal{R}_{i_2} = \emptyset \tag{4}$$

Here, $\{\mathbf{b}\}$ denotes the set of all N-bit sequences and $\emptyset$ means the empty set. The latter conditions mean that the decoding regions must include every possible received bit sequence and each received sequence can belong to only one decoding region. Normally, error correcting codes and the decoders are not designed with the ultimate interpretation of the bit sequences in mind. With such approaches, the probability $\Pr[i|j]$ would be optimized separately, then distortion considered. We take the approach here of determining how the entire physical layer should be structured so that distortion is minimized.

2.1 Uncoded case

Expression (2) for mean-squared error can be derived when no error correcting code is used ($N = K$) by explicitly using (1) that associates bits with amplitude values. Noting that $\pi_j \Pr[i|j] = \Pr[i,j]$, the sum in (3) is the expected value of $C_{i,j}$. Assuming the Bayes' cost function is $(i-j)^2$ and that the amplitudes are equally likely, the mean-squared error can be written as

$$\text{mse} = \mathcal{E}\left[\left(\sum_k \left(b_k^{(i)} - b_k^{(j)}\right) 2^k\right)^2\right]$$

Upon expanding the square, the only nonzero term occurs when the k^{th} bit in what is transmitted and received disagree. The cross-terms between differing bits disappear because we assume a white Gaussian noise channel, which means that the optimal receiver operates on each bit interval independently of the others, and produces statistically independent bit estimates. We are left with

$$\text{mse} = \sum_k \Pr\left[b_k^{(i)} \neq b_k^{(j)}\right] 2^{2k}$$

The probability in this expression is simply the probability the bit is received in error, and (2) results.

Analytically, a solution to minimizing the mean-squared error can be derived using Lagrange multipliers. The solution must satisfy the set of K equations found by differentiating expression (2) plus the Lagrange penalty term $\lambda\left(\sum_k T_k - T_s\right)$ with respect to the bit-interval durations and setting each of the derivatives to zero.

$$T_k e^{aT_k} = \frac{a\lambda^2 2^{4k}}{8\pi} \quad , k = 0, \ldots, K-1 \tag{5}$$

The constant a equals $2A^2/N_0$. The somewhat curious appearance of 8π in the denominator can be traced to the Gaussian noise assumption and the evaluation of the derivative of $Q(\cdot)$. In our computations, we take the total time allocated to transmitting the K bits, T_s, to equal 1. Consequently, a is numerically equal to E_w/N_0, the ratio of the signal energy received during a *word* interval, the time taken to transmit the data, and the white noise spectral height. Because of the variable-duration bit intervals, the usual signal-to-noise ratio E_b/N_0 defined over a bit interval is not a constant; our SNR defined over the word interval is constant, and provides a way for comparing results. Because the left side of the equation is a monotonic function of the bit-interval durations, a unique solution exists. Furthermore, because the right side increases with k (i.e., increases as we go from least to most significant bits), the optimal bit-interval durations are strictly increasing: $T_0 < T_1 < \cdots < T_{K-1}$. However, the transcendental nature of this equation means no closed form solution exists. We did find that the relation between successive bit-interval durations satisfies

$$\frac{T_{k+1}}{T_k} = 2^4 e^{-a(T_{k+1}-T_k)}$$

We can approximate the solution when the signal-to-noise ratio parameter a is sufficiently small so that $e^{aT_k} \approx 1 + aT_k$. Using this approximation, we have that $T_k = 2^{-4}T_{k+1}$. Thus, the optimal bit intervals, particularly those corresponding to the least significant bits, decrease by a factor 16 as k decreases to zero.

Rather than solve (5) directly, we opted to solve the original optimization problem stated in (4). Because no error correction is considered as yet, optimization over the decision regions is not necessary as only data bits are sent. We numerically solved the optimization problem using Matlab's optimization program `fmincon`, which uses the Nelder-Mead direct search algorithm. Figure 2 shows the results for bit-interval durations and mean-squared error gain for 8-bit data. For small SNRs, transmitting only the most-significant bit minimizes the mean-squared error. In this SNR range, the mean-squared error decreased by a few decibels in comparison with that resulting from equal-duration bit intervals in this SNR range. As SNR increases, the next most significant bits are used, and the mean-squared error continues to decrease, becoming about 13 dB at its maximum. Maximum gain occurs when all bits are transmitted, but not when they alloted the same duration.

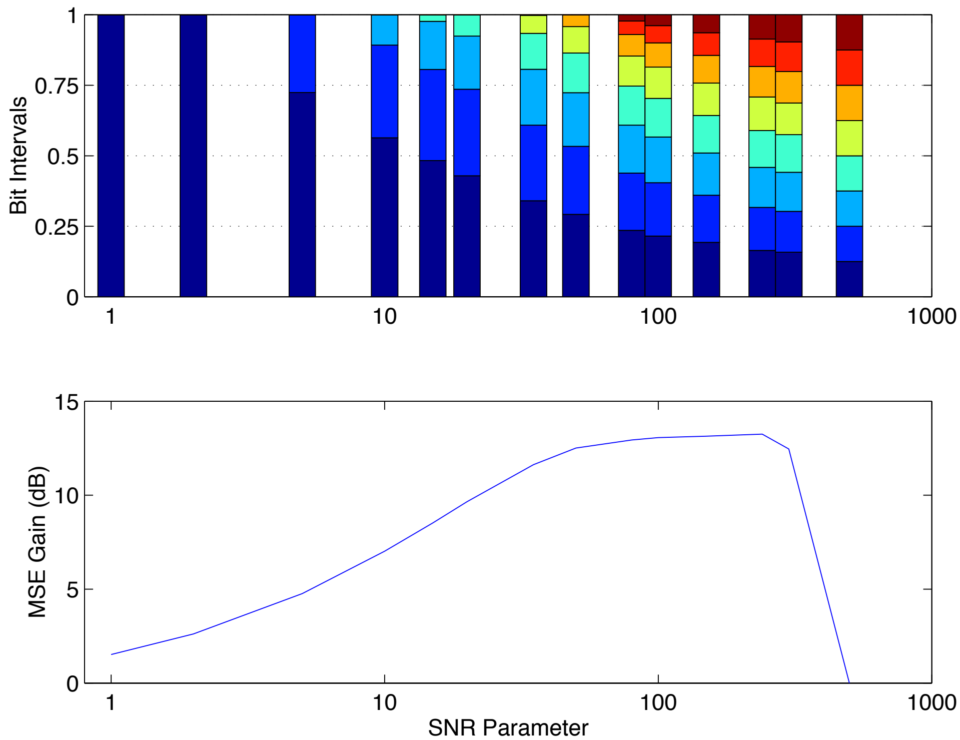

Figure 2: The top panel shows as a stacked bar histogram the bit intervals that optimize the mean-squared error of the reconstructed data. The most-significant bit occurs at the bottom of each stack and the least-significant at the top. The horizontal axis is the signal-to-noise parameter a found in equation (5), which is equivalent to E_w/N_0. The bottom panel shows the resulting gain (decrease) in mean-squared error relative to that of equal-duration bit intervals expressed in decibels. Here, positive gains correspond to smaller mean-squared errors.

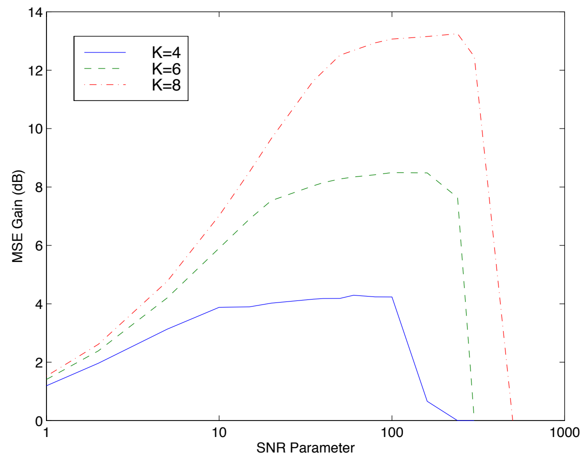

Figure 3: The mean-squared error gain is shown for three values of K, the number of bits used to represent the sample's value. The mean-squared error gains decrease as fewer data bits are used. The SNR at which equal-duration bit intervals occurs decreases as the number of data bits used in the sampling decreases.

Further SNR increases result in equal-duration bit intervals, which means that ultimately no gain (0 dB) occurs.

This variation of bit-interval durations and mean-squared error gain with SNR typified the behavior when a smaller number of bits comprised the sampled amplitude. Figure 3 shows the mean-squared error gains that resulted. The maximal gains decrease as the number of bits decreases, and the SNR at which equal-duration bit intervals are optimal decreases. We did not consider more than eight bits.

Because the optimal bit-interval durations change significantly with SNR, we studied how much gain in mean-squared error could be obtained with fixed, unequal duration bit intervals. Figure 4 shows that a gain over equal-duration bit intervals can be obtained for SNRs smaller than that used to derived the unequal duration bit intervals and for SNRs somewhat larger. The consequence of not varying the bit-interval durations with SNR can be several decibels, but significant gains remain. Beyond this SNR range, unequal-duration bit intervals yield a larger mean-squared error than do equal duration intervals.

2.2 Coded case

When error correction is incorporated into the communication scenario, the time taken for the coding bits can also be allowed to vary. We could not find a simple expression for the distortion when error

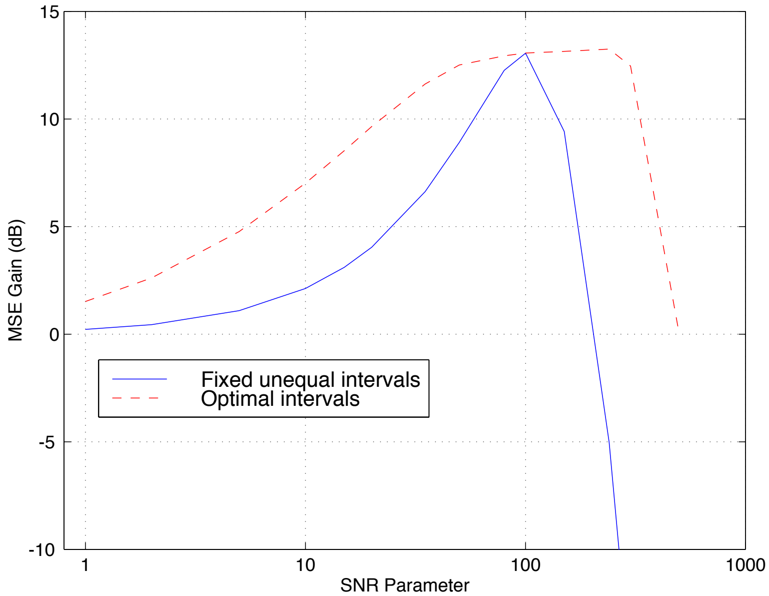

Figure 4: The gain in mean-squared error relative to equal-duration bit intervals is plotted (solid line) as a function of SNR when the optimal intervals for SNR=100 are used. Negative gain corresponds to larger mean-squared error than that obtained with equal-duration bit intervals. The dashed line indicates the gains obtained with optimal bit-interval durations. The number of bits $K = 8$.

---

correction is employed. We solved the optimization problem stated in (4) for a (7,4) Hamming code. The bit-interval durations of both data and error-correction bits were allowed to vary independently while imposing the constraint that the sum of the durations must equal the sampling interval T_s. With this constraint, the data rate equals that when no coding is used. Note that this approach means that the when error correction is used, less time is available to transmit the data bits, which leads to larger error probabilities. This effect is countered by the ability to correct errors (single-bit error correction in the case of the (7,4) code). However, it could well be that when the quality of the communications system is judged on the basis of signal distortion (mean-squared error), error correction may not be the best solution.

Figure 5 shows the result of our optimization. As expected, the durations of the most significant bits are longer than those of the least significant bits until large SNRs are reached. Interestingly, *no* error correction is optimal for SNRs smaller than a threshold (here about SNR $= 60$). Once this threshold is exceeded, all error-correction bits appear and their durations are comparable to that of the least-significant bits. Well above this threshold, equi-duration intervals are optimal. Below threshold, the gain in mean-squared error relative to the equal-duration, no-error-correction case is a few decibels larger than the gain when no error correction is used (figure 3, $K = 4$). The dashed

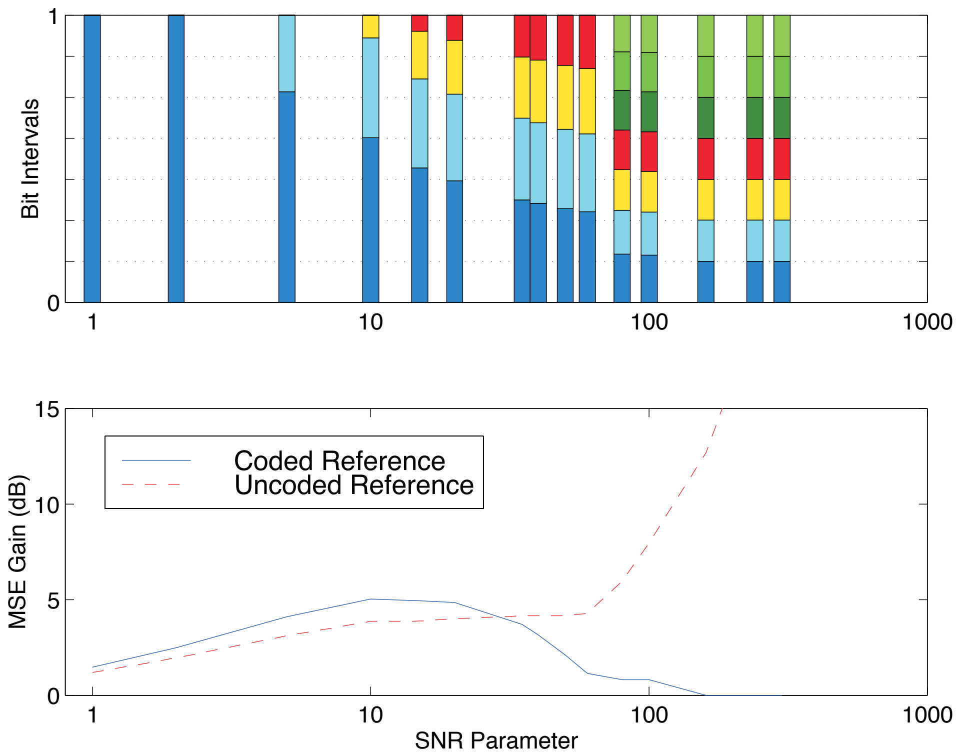

Figure 5: The top plot shows the optimal (minimum mean-squared errror) bit-interval durations as a function of E_w/N_0 for a (7,4) Hamming code. The most significant bit lies at the botttom of each column and the least significant bit the fourth interval from the bottom. The three error-correction bit intervals are shown at the top of the columns, and only occur in the last five columns. The bottom plot shows the gains in mean-squared error that accrue in comparison to two reference values. The solid line shows the gain that results when the reference is the mean-squared error incurred by assigning equal-duration bit intervals to data and error correction bits (equal to $T_s/7$ in this case). The dashed line shows the gain when the reference is equal duration bit-intervals for the *no* error correction case.

line in figure 5 shows the mean-squared error achieved in comparison to the case when no error-correction code is used. When this gain is less than the gain over the coded case, it indicates that imposing the (7,4) code actually worsens mean-squared error when bit-interval durations are equal. When the gain is higher, error-correction improves mean-squared error. Error-correction improves the mean-squared error only for SNRs greater than about 40 in our example. When optimal-duration error correction is present at higher SNRs, large gains become apparent.

3 Conclusions

The optimality of unequal bit-interval durations for minimizing mean-squared error is not surprising. What is surprising is that the improvement in mean-squared error is so large (figure 3). Maximal gain (reduction) of the mean-squared error occurs in a signal-to-noise ratio range typical of wireline communication systems (greater than 10 dB SNR for each bit). In this range, unequal bit-interval durations for all the data bits provides the smallest mean-squared error. The SNR would need to be greater than 20 dB for equal-duration bit intervals to be optimal. For small SNRs more typical of wireless systems (less than 10 dB), unequal bit-interval durations *and* not transmitting some of the less significant bits is optimal. Thus, it may well be the case that it is not worthwhile using more than a few bits in A/D converters when the data will be transmitted over very noisy channels.

The unequal bit-interval durations may be difficult to coordinate between transmitter and receiver, especially in a wireless multicast situation. That said, our results can be interpreted as how to allocate power on a bit-by-bit basis and how many bits to use to represent data. Thus, in addition to power control, additional gains can be achieved by carefully allocating transmitter power at the level of single bits and using this fixed allocation regardless of the actual SNR.

The single-bit block error correction considered here was only effective at large signal-to-noise ratios. At smaller ratios, the power consumed by transmitting error correction bits is better spent on the data when signal distortion is the performance metric. The error-correction afforded by coding did not compensate for the increased error probability. The block code we used was primitive in the sense that all data bits must be transmitted in order for the coding to be effective. Thus, it is not surprising that coding only becomes effective at high SNRs. Future work should develop more sophisticated coding schemes which can accommodate variable-length data words while still minimizing distortion.

In summary, our results suggest that considering signal distortion at the physical as well as higher layers can be fruitful. Our results are not tied to mean-squared error; any sample-to-sample distortion measure can be used in equation (4).

References

[1] J.K. Cavers. Variable-rate transmission for Rayleigh fading channels. *IEEE Trans. Comm.*, COM-20:15–22, 1972.

[2] J.F. Hayes. Adaptive feedback communications. *IEEE Trans. Communication Technology*, COM-16:29–34, 1968.

[3] MOST Specification of Physical Layer, 2001. Rev. 1, http://www.mostcooperation.com.

[4] H. Rodríguez-Díaz and D.H. Johnson. Optimizing physical layer data transmission for minimal signal distortion. In *ICASSP '03*, Hong Kong, 2003.

[5] H.L. van Trees. *Detection, Estimation, and Modulation Theory, Part I.* Wiley, New York, 1968.