Cut-off Rate Channel Design

Prakash Narayan
University of Maryland,
College Park, MD 20742

Donald L. Snyder
Washington University,
St. Louis, MO 63130

Abstract

The eloquence with which Massey advocated the use of the cut-off rate parameter for the coordinated design of modulation and coding in communication systems caused many to redirect their thinking about how communication systems should be designed. Underlying his recommendation is the view that modulation and demodulation should be designed to realize a good discrete channel for encoding and decoding, rather than the prevailing view at the time, and still the view of many, that bit error-probability should be optimized. In this short paper, some of the research influenced by Massey's insightful suggestions on this subject is reviewed.

I Introduction

The use of the cut-off rate parameter R_0 in the study of single-user coded communication systems was first advocated by Wozencraft and Kennedy [11] in 1966. Unfortunately, their proposal to use R_0 as a criterion for the design of the modulation system remained largely unheeded. Then, in a remarkable paper in 1974, Massey [4] gave an eloquent argument in favor of the cut-off rate parameter as a criterion for the coordinated design of modulation and coding in a communication system. Calling to discard the heretofore popular "error probability" criterion on the grounds that it was apposite only for uncoded systems, he resurrected the earlier proposal of Wozencraft and Kennedy and showed that the R_0 criterion led to a rich "communication theory" of its own for coded communications. In particular, by interpreting R_0 as a function of the modulator and demodulator, Massey [4] demonstrated how it could be used to design the best discrete channel as seen by the encoder and decoder. He crowned his arguments by establishing that a simplex signal set maximized the cut-off rate of an infinite-bandwidth, additive, white Gaussian-noise channel for infinitely soft decisions. The "optimality" of such a signal set with respect to the error probability criterion has remained a conjecture for many years.

Massey's paper [4] opened the floodgates for a plethora of publications employing the cut-off rate criterion to assess the performance of coding and modulation schemes for a variety of applications ranging from optical communications to spread-spectrum systems to an extension to multiaccess channels. A fair citation of this field is beyond the scope of this paper; good sources of relevant publications are the *IEEE Transactions on Information Theory and Communications.*

A mention of the widespread use of the cut-off rate as a performance criterion must be qualified by the observation that it has not been accorded universal acceptance. A key reason is its seeming lack of fundamental significance, unlike that of the *capacity* of a communication channel. Another reason is the existence of channels with memory for which the cut-off rate, unlike capacity, exhibits anomalous behavior (see, e.g., [5]).

This paper focuses on the role of the cut-off rate criterion in the coordinated design of coding and modulation formats for the single-user, additive white Gaussian-noise channel [4]. We briefly review the work of Massey [4] when the modulated signals possess unlimited bandwidths. This problem is then considered in the presence of constraints on signal bandwidth; a complete solution remains elusive. Partial results by Narayan and Snyder [6] are then presented.

II The Cut-Off Rate and Its Properties

Let $\{W : \mathcal{X} \to \mathcal{Y}\}$ be a discrete memoryless channel (DMC) with finite input and output alphabets, $\mathcal{X}$ and $\mathcal{Y}$, respectively. Consider a (random) code of rate $R > 0$ with codewords of blocklength n, and a maximum-likelihood decoder. It is assumed that the codewords are independent and identically distributed (i.i.d.) and that the symbols in each codeword are i.i.d. with (an arbitrary) probability mass function P on $\mathcal{X}$. The average probability of a decoding error when this code is used on the DMC $\{W\}$ is bounded above [3, pp. 142-143] according to

$$P_e(P,W) \leq \exp[-n(E_0(\rho,P,W) - \rho R)], \quad 0 \leq \rho \leq 1 \tag{1}$$

where

$$E_0(\rho,P,W) = -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P(x)W^{1/(1+\rho)}(y|x)\right)^{1+\rho}\right]. \tag{2}$$

(All logarithms and exponentiations are with respect to the base 2.)

The upper bound in (2) is improved by choosing P and ρ so as to maximize $E_0(\rho,P,W) - \rho R$. To this end, consider

$$\max_P \max_{0\leq\rho\leq 1} E_0(\rho,P,W) - \rho R \tag{3}$$

and note from [3] that for

$$R \leq R_{crit}(P,W) = \frac{\partial E_0}{\partial \rho}(\rho,P,W)\,|_{\rho=1}\ , \tag{4}$$

$E_0(\rho,P,W)$ is maximized by $\rho = 1$.

The quantity $\max_P E_0(1,P,W)$ is called the *cut-off rate* of the DMC $\{W\}$, denoted $R_0(W)$, and is given by

$$R_0(W) = \max_P\left[-\log\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P(x)W^{1/2}(y|x)\right)^2\right]. \tag{5}$$

Observe that $R_0(W)$ depends on $\{W\}$ but not on P. If P^* achieves the maximum in (5), we obtain from (2) that

$$P_e(P^*, W) \leq \exp\left[-n(R_0(W) - R)\right]. \tag{6}$$

This simpler bound on average error probability is useful if $R < R_0(W)$, and is quite accurate for $R \cong R_{crit}(P^*, W)$. Massey [4] concluded from (6) that for block codes with maximum-likelihood decoding, $R_0(W)$ determines a range of code rates at which reliable communication can be assured, as well as the coding complexity, reflected by n, to achieve a specified level of reliability. (For a different interpretation of cut-off rate for channels with memory, see [5].)

The cut-off rate of a DMC $\{W\}$ also affords other interpretations. For instance, as pointed out by Csiszár [2], $R_0(W)$ equals Rényi capacity [7] or "information radius" of order $\alpha = 1/2$, and the β-cut-off rate [2] for $\beta = -1$.

The cut-off rate parameter plays an important role in assessing the performance of codes different from that considered above. For instance, Viterbi [10] has shown for convolutional coding with maximum-likelihood decoding that the average probability of decoding error on a DMC $\{W\}$ is bounded above according to

$$P_e(P^*, W) \leq k_R \exp[-nR_0(W)] \tag{7}$$

where n is the constraint length, and k_R varies slowly with R. Also, the cut-off rate is a key parameter in sequential decoding, wherein the receiver decodes a code with a tree structure by computing the metrics of, and making tentative hypotheses on, successive branches of the tree and by changing these hypotheses when subsequent choices indicate an earlier incorrect hypothesis. The cut-off rate $R_0(W)$ of a DMC $\{W\}$ is the $R_{comp}(W)$ for sequential decoding (cf. Arikan [1]), which is the rate above which the average computational complexity of the decoding algorithm becomes unbounded.

III The Additive White Gaussian Noise Channel

Let $\mathcal{X} = \{1, \cdots, a\}$ be the (finite) encoder alphabet. The corresponding (modulated) signal set is $\mathcal{S} = \{s_i(t), 0 \leq t \leq T;\ i = 1, \cdots, a\}$, where $s_i(t)$ is the signal transmitted by the sender over the waveform channel when the encoder produces the symbol i. Signals are transmitted and received on the interval $[0, T]$. The (random) signal $Z(t)$ received at the output of the additive, white, Gaussian-noise (AWGN) waveform-channel is

$$Z(t) = s_i(t) + N(t), \tag{8}$$

where $N(t)$ is white Gaussian-noise with power-spectral density $N_0/2$ W/hz. The demodulator then produces an output from the alphabet $\mathcal{Y} = \{1, \cdots, b\}$.

From a coding viewpoint, the modulator, waveform channel, and demodulator together constitute a discrete channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. By virtue of the stationarity and independent increments property of $\{N(t), 0 \leq t \leq T\}$, this channel is also memoryless, and hence a DMC; we denote this DMC by $\{W\}$. It is important to note that $\{W\}$ depends on the choice of the signal set $\mathcal{S}$, although this dependence will not be displayed for notational convenience.

The cut-off rate of the DMC $\{W\}$ is not decreased by using a finer output quantization at the demodulator. In this treatment, we shall restrict ourselves to the limiting situation when the output quantization is arbitrarily fine, i.e., $b = \infty$. For the effects of finite quantization, see [4]. It then follows from [4] and (2) that

$$E_0(1, P, W) = -\log \sum_{i,j=1}^{a} P(i)P(j)\exp[-s_{ij}/4N_0] \tag{9}$$

where

$$s_{ij} = \int_0^T [s_i(t) - s_j(t)]^2\, dt. \tag{10}$$

The problem of coordinated design of the encoder and modulator, using the cut-off rate criterion, can now be stated as follows.

Problem 1: Determine

$$\max_{S} R_0(W)$$

or, equivalently,

$$\max_{S} \max_{P} E_0(1, P, W) \tag{11}$$

subject to

$$P(i) \geq 0,\ i = 1, \cdots, a; \quad \sum_{i=1}^{a} P(i) = 1; \tag{12}$$

and the "average energy" constraint

$$\sum_{i=1}^{a} P(i) \left(\int_0^T s_i^2(t)dt \right) = E. \tag{13}$$

Theorem 1 (Massey [4]): The maximum in Problem 1 is attained by the *simplex* signal set S^* characterized by

$$\frac{1}{a} \sum_{i=1}^{a} s_i^*(t) = 0, \quad 0 \leq t \leq T; \tag{14}$$

$$s_{ij}^* = s, \quad i \neq j; \tag{15}$$

$$\int_0^T s_i^{*2}(t)dt = E, \quad i = 1, \cdots, a; \tag{16}$$

where s is a constant for distinct signals and with the code symbols being chosen according to the uniform probability distribution

$$P^*(i) = 1/a, \quad i = 1, \cdots, a. \tag{17}$$

The corresponding maximal cut-off rate is

$$R_0^* = \log a - \log\left[1 + (a-1)\exp\left(-\frac{aE}{2(a-1)N_0}\right)\right]. \tag{18}$$

Proof: See [4].

Remark: It is interesting to note that the conjectured optimality of the simplex set with respect to the probability of error criterion, subject to (13) with $P(i) = 1/a$, $1 \leq i \leq a$ (the "strong" simplex conjecture) was recently disproved [9]. The "weak" simplex conjecture is still unresolved when (13) is replaced by the constraint $\int_0^T s_i^2(t)dt = E$, $1 \leq i \leq a$.

Either by virtue of law or nature, it is generally necessary to impose constraints on the portion of the frequency spectrum that the signals transmitted by a sender can occupy. There is no universal measure of bandwidth for a signal of finite duration; several measures have been proposed, and we consider two of these below.

Consider the signal set $\mathcal{S} = \{s_i(t), 0 \leq t \leq T,\ i = 1, \cdots, a\}$ used according to the probability distribution $\{P(i), i = 1, \cdots, a\}$. The *squared second moment bandwidth* of $\mathcal{S}$ is defined by

$$B_{SM}(\mathcal{S}; P) = \frac{\sum_{i=1}^{a} P(i) \int_0^T [ds_i(t)/dt]^2\, dt}{\sum_{i=1}^{a} P(i) \int_0^T [s_i(t)]^2\, dt}. \tag{19}$$

If $S_i(f), -\infty < f < \infty$, denotes the Fourier transform of $s_i(t), 0 \leq t \leq T$, the *fractional out-of-band energy* of $\mathcal{S}$ is the fraction of the total average energy of $\mathcal{S}$ lying outside a prespecified frequency band $[-F, F]$ and is defined by

$$B_{OB(F)}(\mathcal{S}; P) = \frac{\sum_{i=1}^{a} P(i) \left[\int_{-\infty}^{\infty} |S_i(f)|^2 df - \int_{-F}^{F} |S_i(f)|^2 df\right]}{\sum_{i=1}^{a} P(i) \int_{-\infty}^{\infty} |S_i(f)|^2 df}. \tag{20}$$

We now state two problems of coordinated encoder and signal design subject to constraints on the bandwidth of the signal set, using the cut-off rate criterion. These problems are obvious extensions of Problem 1 above, and areas yet unresolved in full generality.

Problem 2A: Same as Problem 1 above, with the additional squared second-moment bandwidth constraint

$$B_{SM}(\mathcal{S}; P) \leq \beta^2. \tag{21}$$

Problem 2B: Same as Problem 1 above, with the additional fractional out-of-band energy constraint

$$B_{OB(F)}(\mathcal{S}; P) \leq \epsilon. \tag{22}$$

Remark: Problems 2A and 2B reduce to Problem 1 upon setting $\beta^2 = \infty$ in (21) and $\epsilon = 1$ in (22), respectively.

Modified versions of Problems 2A and 2B above have been, in effect, solved in [6] albeit in the context of a multiaccess AWGN channel. Rather than determining signal sets that maximize $R_0(W)$ under constraints on average energy and bandwidths, we instead seek to identify signal sets with minimal bandwidths, in the sense of (19) and (20), from among those that achieve the maximal cut-off rate R_0^* in (18).

Problem 3A: Consider the family $\sum$ of all simplex signal sets $\mathcal{S}$ satisfying (14)-(16), with the signals being used equiprobably in accordance with (17), that is, $P(i) = 1/a, 1 \leq i \leq a$. Determine

$$\min_{\mathcal{S}\in\sum} B_{SM}(\mathcal{S}; P). \tag{23}$$

Problem 3B: For the same setup as in Problem 3A above, determine

$$\min_{\mathcal{S}\in\sum} B_{OB(F)}(\mathcal{S}; P). \tag{24}$$

Remark: Clearly, if the minimal bandwidth in Problem 3A (resp. Problem 3B) satisfies constraint (19) (resp. (20)), then the corresponding (optimal) signal set is optimal for Problem 2A (resp. Problem 2B) too.

The solutions to Problems 3A and 3B are provided by the following

Theorem 2 (Narayan-Snyder [6]):

1. The minimum in Problem 3A is attained by the "raised-cosine" simplex set $\mathcal{S}^{SM} = \{s_i^{SM}(t), 0 \leq t \leq T; 1 \leq i \leq a\}$ given by

$$s_i^{SM}(t) = \sqrt{\frac{aE}{a-1}}\left[\left(\frac{a-1}{a}\right)\phi_i(t) - \frac{1}{a}\sum_{j=1,j\neq i}^{a}\phi_j(t)\right]. \tag{25}$$

where

$$\phi_i(t) = \left(\frac{2}{T}\right)^{1/2}\sin\frac{i\pi t}{T}, \qquad 1 \leq i \leq a. \tag{26}$$

The corresponding minimal squared second-moment bandwidth equals

$$B_{SM}(\mathcal{S}^{SM}) = \frac{\pi^2}{6T^2}(a+1)(2a+1). \tag{27}$$

2. The minimum in Problem 3B is attained by the "prolate spheroidal wave" simplex set $\mathcal{S}^{OB(F)} = \left\{s_i^{OB(F)}(t),\ 0 \leq t \leq T;\ 1 \leq i \leq a\right\}$ given by

$$s_i^{OB(F)}(t) = \sqrt{\frac{aE}{a-1}}\left[(a-1a)\,\Psi_i(t) - \frac{1}{a}\sum_{j=1,j\neq i}^{a}\Psi_j(t)\right], \tag{28}$$

where

$$\Psi_i(t) = \sqrt{\frac{1}{\lambda_{i-1}(2\pi FT)}}\zeta_{i-1}(t), \tag{29}$$

with $\zeta_{i-1}(t)$ being a prolate spheroidal wave function with eigenvalue $\lambda_{i-1}(2\pi FT)$. The corresponding minimal fractional out-of-band energy equals

$$B_{OB(F)}\left(\mathcal{S}^{OB(F)}\right) = a - \sum_{i=0}^{a-1} \lambda_i(2\pi FT). \tag{30}$$

Proof: The proof follows from [6, Section III].

IV Conclusion

In addition to the results on the AWGN channel reviewed in this paper, several authors have gainfully used the cut-off rate parameter as a criterion for signal design. For instance, Snyder and Rhodes [8] have identified modulation formats that maximize the cut-off rate parameter of a single-user, shot-noise limited optical channel with infinite bandwidth under simultaneous constraints on average energy and peak amplitude (see also Wyner [12]). In [6], some of these results are extended to a two-sender multiaccess channel by maximizing the "cut-off rate region" and identifying the optimal signal sets. Also, conditions are established under which this optimality is preserved when constraints are imposed on signal bandwidth. This cumulative body of work bears out Massey's thesis [4] that the cut-off rate parameter of a DMC leads to a rich communication theory of its own, offering useful insights into the coordinated design of efficient coding and modulation systems. At the same time, we should be careful not to overstate its importance as it lacks the fundamental significance of channel capacity.

V Acknowledgments

Prakash Narayan considers it an honor to have been the beneficiary of Jim Massey's expertise as well as sustained help and encouragement, and expresses with pleasure his deep sense of gratitude to Jim.

References

[1] E. Arikan, "An upper bound on the cut-off rate of sequential decoding," *IEEE Trans. Inform. Theory*, vol. 34, no. 1, pp. 55-63, 1988.

[2] I. Csiszár, "Generalized cut-off rates and Rényi's information measures," *Proc. IEEE Int. Symp. Inform. Th.*, Budapest, Hungary, p. 73, June 1991.

[3] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley, New York, 1968.

[4] J. L. Massey, "Coding and modulation in digital communications," *Proc. Int. Zurich Seminar on Dig. Comm.*, Switzerland, pp. E2(1)-E2(4), March 1974.

[5] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 1, pp. 44-53, 1984.

[6] P. Narayan and D. L. Snyder, "Signal set design for band-limited multiple-access channels with soft decision demodulation," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 4, pp. 539-556, 1987.

[7] A. Rényi, "On measures of entropy and information," *Proc. 4th Berkeley Symp. Math. Statist. Probability*, vol. 1, Univ. Calif. Press, Berkeley, pp. 547-561, 1961.

[8] D. L. Snyder and I. B. Rhodes, "Some implications of the cut-off rate criterion for coded direct detection optical communication systems," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 327-338, 1980.

[9] M. Steiner, "New results in signal design for the AWGN channel," *Proc. IEEE Int. Symp. Inform. Th.*, Budapest, Hungary, p. 204, June 1991.

[10] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 327-338, 1980.

[11] J. M. Wozencraft and R. S. Kennedy, "Modulation and demodulation for probabilistic coding," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 291-297, 1966.

[12] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel-part I," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449-1461, 1988.