Some remarks on the nature of the cutoff rate parameter

Erdal Arıkan
Electrical-Electronics Engineering Department
Bilkent University, Ankara, TR-06800, Turkey
Email: arikan@ee.bilkent.edu.tr

***Abstract*— The cutoff rate is a multi-faceted parameter of communication with relations to the computational complexity of sequential decoding, the union bound on the probability of error in communication systems in general, and to certain types of guessing problems. We present a survey of selected results on the cutoff rate with the goal of bringing forth the fact that the cutoff rate is not a conservative quantity and hence needs to be interpreted carefully. We begin by reviewing Jacobs and Berlekamp's result on the computation problem in sequential decoding. Then, we review an example by Massey that shows that the cutoff rate can be "created" by splitting a quaternary erasure channel into two binary erasure channels. Finally, we show that such gains can be achieved for other channels if one combines multiple independent copies of a given channel and uses a coding scheme that is suitable for successive-cancellation decoding.**

I. Introduction

The cutoff rate of a discrete memoryless channel (DMC) W with an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, transition probabilities $W(y|x)$ is defined as

$$R_0(W) = -\log \sum_{y\in\mathcal{Y}} \left[\sum_{x\in\mathcal{X}} Q(x)\sqrt{W(y|x)} \right]^2$$

where Q is a probability distribution on $\mathcal{X}$, chosen to maximize the right hand side. The cutoff rate is a multi-faceted parameter of communication that owes its significance in part to its relation to sequential decoding. Sequential decoding is a decoding algorithm for tree codes invented by Wozencraft [1] with important later contributions by Fano [2]. The major drawback of sequential decoding is that the decoding complexity is a random variable.

In a seminal paper, Jacobs and Berlekamp [3] gave an elegant proof that the distribution of computation in sequential decoding follows a Paretian distribution with a mean that becomes unbounded as the coding rate approaches the channel cutoff rate. To derive this result, they formalized sequential decoding as a constrained search problem. They focused on the operations by the sequential decoder at an arbitrary but fixed depth N branches into the code tree from the root. For a tree code with rate R bits/branch, there are 2^{NR} nodes at level N. One of these nodes lies on the correct (actually transmitted) path, and the task of the sequential decoder is to find this correct node without looking ahead beyond level N. For a sequential decoder that eventually finds the correct node (possibly after several false attempts), the search task is in effect equivalent to a "guessing game" with the assistance of a helpful genie. The questions that the sequential decoder can ask are of the type "Is this node the correct node at level N?" to which the genie provides a truthful "yes" or "no" answer. Every false guess increases the complexity counter by one unit and the search continues until the sequential decoder finds the correct node and advances to the next level in the tree. The result in [3] states that (under certain technical conditions), the number of questions C_N that the decoder must ask to determine the correct node at level N has a Paretian distribution,

$$P(C_N > L) \gtrapprox L^{-\rho}$$

where the parameter $\rho \geq 0$ is determined by the rate R of the code. For $R < R_0$, $\rho > 1$ and the expectation of C_N is bounded for all $N \geq 1$. On the other hand, for $R > R_0$, $\rho < 1$ and the expectation of C_N grows without bound as N is increased.

In part due to its role as the cutoff rate of sequential decoding and in part due to its appearance in the union bound on the probability of error for communication systems in general, the parameter R_0 appeared prominently in the communications engineering literature in sixties and seventies. The classical book by Wozencraft and Jacobs [4] uses R_0 extensively as a practical measure of quality for signaling systems. Sequential decoding also was discussed in great detail by the information theory and coding books that were written in this period, notably, [4], [5], [6]. The argument in favor of R_0 as a practical parameter of communication was put forward lucidly in [7] and [8]. Communication at rates near R_0 appeared a reasonable and practically viable objective while achieving channel capacity appeared an unrealistic goal.

The notion of cutoff rate as a practical limit to achievable rates has been challenged from the very beginning. Pinsker [9] gave a method that combined sequential decoding and block codes to achieve channel capacity. Falconer [10] also devised a system using Reed-Solomon codes and sequential decoding that achieved channel capacity. Although these methods showed that the cutoff rate was not a fundamental barrier, they were too complicated at the time to have much practical impact.

A paper by Massey [11] revealed a truly interesting aspect of the cutoff rate by showing that it could be "created" by simply splitting a given channel into componential subchannels. Massey noticed that the cutoff rate of an M'ary erasure channel could be improved by splitting it into $n = \log_2(M)$ binary erasure channels (BEC). Massey's example did not involve any sophisticated coding schemes (unlike Pinsker's and Falconer's) to improve the cutoff rate; simply splitting a given channel lead to an improved sum cutoff rate. If a channel W is split into two subchannels W_1, W_2, the channel capacity function always satisfies $C(W_1) + C(W_2) \leq C(W)$. Massey's example demonstrated the possibility that $R_0(W_1) + R_0(W_2) > R_0(W)$. In light of Massey's example, Gallager [5] concluded that "the cutoff rate is not really a fundamental parameter of communication."

In the rest of this presentation, we discuss Massey's example and suggest a method for achieving cutoff rate gains for arbitrary DMC's. The emphasis is on simple schemes that can be implemented in practice. The methods presented here have common elements with well-known coded-modulation techniques, namely, Imai and Hirakawa's [13] multi-level coding scheme and Ungerboeck's [14], [15] set-partitioning idea. These connections are discussed in [16].

II. Massey's example

Consider a quaternary erasure channel (QEC), $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_1 \times \mathcal{Y}_2$ where $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, ?\}$, and

$$W(y_1 y_2 | x_1 x_2) = \begin{cases} 1 - \epsilon, & y_1 y_2 = x_1 x_2 \\ \epsilon, & y_1 y_2 = ?? \end{cases}$$

where $0 \leq \epsilon \leq 1$ is the erasure probability. Such a QEC W can be decomposed into two BEC's (binary erasure channels): $W_i : \mathcal{X}_i \to \mathcal{Y}_i$, $i = 1, 2$. In this decomposition, a transition $(x_1, x_2) \to (y_1, y_2)$ over the QEC is viewed as two transitions, $x_1 \to y_1$ and $x_2 \to y_2$, taking place on the respective component channels, with

$$W_i(y_i | x_i) = \begin{cases} 1 - \epsilon, & y_i = x_i \\ \epsilon, & y_i = ? \end{cases}$$

These BEC's are fully correlated in the sense that an erasure occurs either in both or in none.

Instead of direct coding of the QEC W, Massey suggested applying independent encoding of the component BEC's W_1 and W_2, ignoring the correlation between the two channels. The second alternative presents significant advantages with respect to the cutoff rate criterion. A simple computation shows that the sum of the cutoff rates of the two component BEC's obtained by splitting the given QEC is $2R_0(\text{BEC}) = 2[1 - \log(1 + \epsilon)]$; this exceeds the cutoff rate of the QEC, $R_0(\text{QEC})$, as shown in Fig. 1. It is remarkable that the gap between the cutoff rate and the capacity of the QEC is bridged significantly by simply splitting the channel. This improvement is obtained at no extra system complexity.

III. Cutoff rate improvement for BEC and BSC

An essential element of achieving gains in the sum cutoff rate is to split a given channel into (positively) correlated subchannels. If the subchannels are independent, no gain can be expected due to Gallager's parallel channels theorem [5, p.149]. On the other hand, it is not clear if an arbitrary channel can be readily split into subchannels, as the QEC was split into two BEC's. To facilitate channel splitting, we propose channel combining as a preliminary step before splitting is applied. We illustrate this idea by two examples in this section.

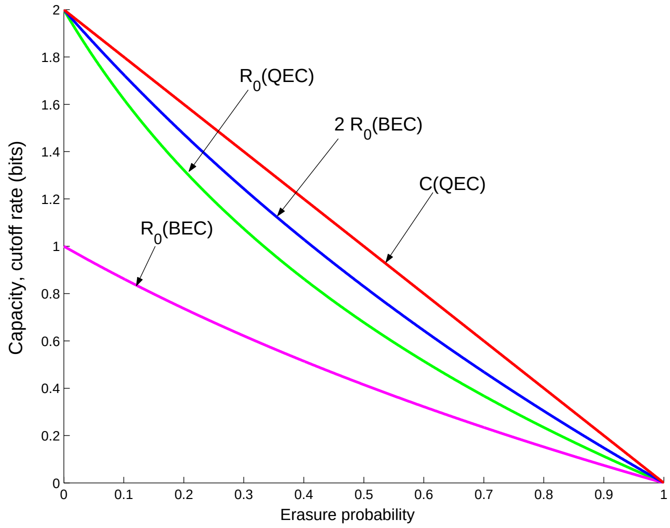

Fig. 1. Capacity and cutoff rate for the splitting of a QEC.

Example 1 (BEC): Let $V : \mathcal{X} \to \mathcal{Y}$ be a BEC with alphabets $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, ?\}$, and erasure probability ϵ. Consider combining two copies of V as shown in Fig. 2. In the figure, $\oplus$ denotes addition modulo-2. The combined channel is a DMC $W : \mathcal{X}^2 \to \mathcal{Y}^2$ with transition probabilities $W(y_1 y_2 | u_1 u_2) = V(y_1 | u_1 \oplus u_2) V(y_2 | u_2)$. For cutoff rate calculations we need to assign a probability distribution on the inputs (u_1, u_2) of the channel W. We use the assignment $(U_1, U_2) \sim Q_1(u_1) Q_2(u_2)$ where Q_1, Q_2 are uniform on $\{0, 1\}$. This assignment is consistent with a coding scheme where the inputs u_1 and u_2 of W are encoded independently, as in Massey's scheme. In Massey's scheme the decoder is also split into two non-communicating decoders; it turns out this is too strong a condition for achieving cutoff rate gains in general. We also split the decoding task, but allow the decoders communicate with each other in a decision feed-

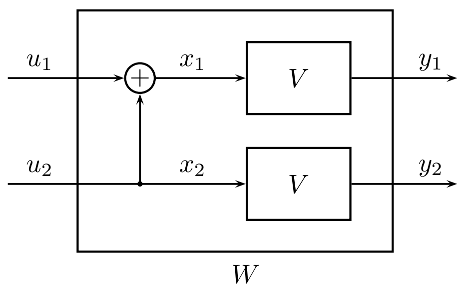

Fig. 2. Combining two binary-input channels.

forward fashion. The resulting decoder structure is known generally as a *successive cancellation* decoder.

Under successive cancellation, the achievable cutoff rates by the first and second stage decoders are given by

$$R_0(U_1;Y_1Y_2) = 1 - \log(1 + 2\epsilon - \epsilon^2)$$
$$R_0(U_2;Y_1Y_2U_1) = 1 - \log(1 + \epsilon^2)$$

where in the second line U_1 (the input of the first channel) is regarded as side-information available perfectly to the second decoder. A successive cancellation decoder with reliable estimates at each stage comes close to achieving this perfect side-information assumption.

Some insight can be gained into this example by noting that the subchannels $U_1 \to Y_1Y_2$ and $U_2 \to Y_1Y_2U_1$ created by splitting the combined channel are themselves BEC's. To see this observe that given Y_1Y_2 user 1 can determine U_1 with certainty only if neither Y_1 nor Y_2 is an erasure. If either Y_1 or Y_2 is an erasure, due to randomization by U_2, U_1 is equally likely to be 0 or 1, i.e., it is erased. So, this is an erasure channel with erasure probability $\epsilon_1 \stackrel{\Delta}{=} 1 - (1-\epsilon)^2 = 2\epsilon - \epsilon^2$. On the other hand, decoder 2 observes $Y_1Y_2U_1$ and fails to decode U_2 with certainty only if both Y_1 and Y_2 are erasures, which occurs with probability $\epsilon_2 \stackrel{\Delta}{=} \epsilon^2$.

Notice that the average of ϵ_1 and ϵ_2 equals ϵ. This conservation of erasure rates is explained by noting that the mutual information terms for the channels under consideration satisfy $I(U_1;Y_1Y_2) = 1 - \epsilon_1$, $I(U_2;Y_1Y_2U_1) = 1 - \epsilon_2$, $I(U_1U_2;Y_1Y_2) = 2(1-\epsilon)$. But by the chain rule for mutual information, $I(U_1U_2;Y_1Y_2) = I(U_1;Y_1Y_2) + I(U_2;Y_1Y_2|U_2)$. (Here, $I(U_2;Y_1Y_2|U_2) = I(U_2;Y_1Y_2U_2)$ since U_1 and U_2 are independent.)

The sum cutoff rate under this scheme is given by

$$R_{0,S}(U_1U_2;Y_1Y_2) = 2 - \big[\log(1+\epsilon_1) + \log(1+\epsilon_2)\big]$$

which is to be be compared with $2R_0(V) = 2(1 - \log(1+\epsilon))$. These cutoff rates are shown in Fig. 3. The figure shows that the sum cutoff rate is improved for all $0 < \epsilon < 1$. This is to be expected since the function $1 - \log(1+\epsilon)$ is a convex function of $0 \le \epsilon \le 1$ and by the conservation of erasures $\epsilon = \frac{1}{2}(\epsilon_1 + \epsilon_2)$.

Example 2 (BSC): Let $V : \mathcal{X} \to \mathcal{Y}$ be a BSC with $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and crossover probability $0 \le \epsilon \le 1/2$. The cutoff rate of the BSC is given by

$$R_0(V) = 1 - \log(1 + \gamma(\epsilon))$$

where $\gamma(\delta) := \sqrt{4\delta(1-\delta)}$ for $0 \le \delta \le 1$.

We combine two copies of the BSC using the same method as in Fig. 2 and use the same input probability assignment, $(U_1,U_2) \sim Q_1(x_1)Q_2(x_2)$ where Q_1, Q_2 are uniform on $\{0,1\}$. The cutoff rates $R_0(U_1;Y_1Y_2)$ and $R_0(U_2;Y_1Y_2U_1)$ can be obtained by direct calculation; however, it is instructive to obtain them by the following argument. The input and output variables of the channel W are related by $Y_1 = U_1 \oplus U_2 \oplus E_1$ and $Y_2 = Y_2 \oplus E_2$ where E_1 and E_2 are independent noise terms, each taking the values 0 and 1 with probabilities $1-\epsilon$ and ϵ, respectively. Decoder 1 sees effectively the channel $U_1 \to U_1 \oplus E_1 \oplus E_2$, which is a BSC with crossover probability $\epsilon_2 = 2\epsilon(1-\epsilon)$ and has cutoff rate

$$R_0(U_1;Y_1Y_2) = 1 - \log(1 + \gamma(\epsilon_2))$$

Decoder 2 sees the channel $U_2 \to Y_1Y_2U_1$, which is equivalent to the channel $U_2 \to (Y_1 \oplus U_1, Y_2) = (U_2 \oplus E_1, U_2 \oplus E_2)$, which in turn is a BSC with diversity order 2 and has cutoff rate

$$R_0(U_2;Y_1Y_2U_1) = 1 - \log(1 + \gamma(\epsilon)^2)$$

Thus, the sum cutoff rate with this splitting scheme is given by

$$R_{0,S}(U_1U_2;Y_1Y_2) = 2 - \big[\log(1 + \gamma(\epsilon_2)) + \log(1 + \gamma(\epsilon)^2)\big]$$

which is larger than $2R_0(V)$ for all $0 < \epsilon < 1/2$, as shown in Fig. 4.

Extensions of these channel combining methods can be found in [17] and [16]. The above examples illustrate the main point that significant gains in cutoff rate can be achieved with negligible extra system complexity.

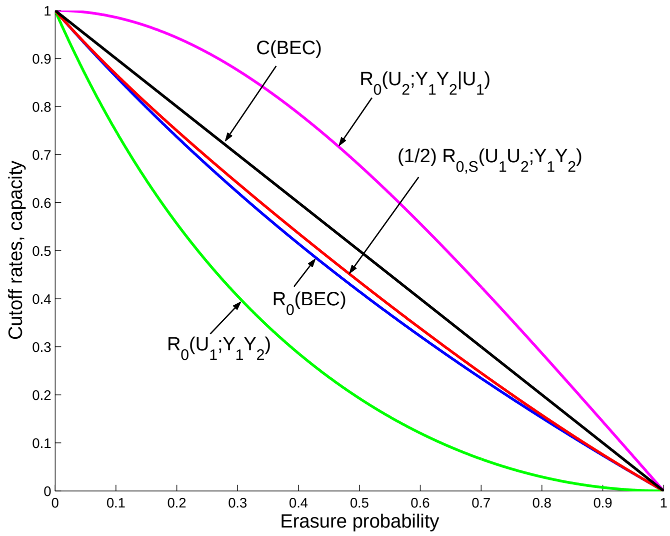

Fig. 3. Cutoff rates for the splitting of BEC.

IV. Summary

We have given a brief survey of the notion of cutoff rate as regards its fundamental nature for communication systems. We pointed out the fundamental nature of the cutoff rate in relation to the computation problem in sequential decoding by citing the main result of [3]. Next, we reviewed Massey's example to show that the cutoff rate can be exceeded without any difficulty even by sequential decoding itself if one uses more than one sequential decoder. Finally, we have given a brief description of a general method for improving the cutoff rates of arbitrary DMC's.

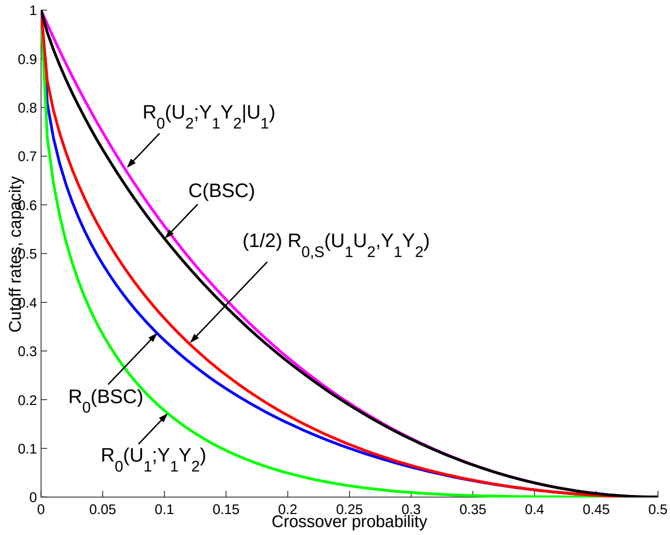

Fig. 4. Cutoff rates for the splitting of BSC.

References

[1] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*. M.I.T. Press: Cambridge, Mass., 1961.

[2] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 64–74, April 1963.

[3] I. M. Jacobs and E. R. Berlekamp, "A lower bound on to the distribution of computation for sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 167–174, April 1967.

[4] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. Wiley: New York, 1965.

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

[6] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill: New York, 1979.

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

[8] J. L. Massey, "Coding and modulation in digital communication," in *Proc. Int. Zurich Seminar on Digital Communication*, (Zurich, Switzerland), pp. E2(1)–E2(24), 1974.

[9] M. S. Pinsker, "On the complexity of decoding," *Problemy Peredachi Informatsii*, vol. 1, no. 1, pp. 113–116, 1965.

[10] D. D. Falconer, "A hybrid coding scheme for discrete memoryless channels," *Bell Syst. Tech. J.*, pp. 691–728, March 1969.

[11] J. L. Massey, "Capacity, cutoff rate, and coding for a direct-detection optical channel," *IEEE Trans. Comm.*, vol. COM-29, pp. 1615–1621, Nov. 1981.

[12] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, March 1985.

[13] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, May 1977.

[14] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets, Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, pp. 5–11, February 1987.

[15] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets, Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, pp. 12–21, February 1987.

[16] E. Arıkan, "Channel combining and splitting for cutoff rate improvement," *IEEE Trans. Inform. Theory*, vol. IT-52, February 2006.

[17] E. Arıkan, "Channel combining and splitting for cutoff rate improvement," in *Proc. 2005 IEEE Int. Symp. Inform. Theory*, (Adelaide, Australia), 4-9 September 2005.