

# Computation of Channel Capacity

Bernard Lampe, *Member, IEEE*

**Abstract**—The maximum capacity of a discrete memoryless channel can be computed as a solution to a convex optimization problem. An efficient algorithm for numerically solving for the capacity of a channel was developed by Blahut. Blahut's algorithm exploits the convex structure of the problem by formulating an iterative set of equations which converges on the optimal solution. In this paper we demonstrate Blahut's algorithm by computing the channel capacities of discrete  $n^{th}$  binary channel of varying sizes. Then we observe the changes in channel capacity as the number of input and output symbols varies.

**Index Terms**—Blahut, Channel Capacity, Binary Channel

## I. INTRODUCTION

CHANNELS are fundamental abstract representations of a communication process. They are useful for mathematically modeling the characteristics of a particular communication. If an entity  $A$  communicates with entity  $B$ ,  $A$  will physically act to transfer information via an available transport medium. A successful transfer of information occurs when  $A$  and  $B$  agree on what has been sent. The entity  $A$  is able to send any arbitrary set and sequence of symbols desired. However, if  $A$  does send information arbitrarily in a noisy channel, then a particular observed output may have resulted from multiple inputs. In effect the receiving entity,  $B$ , is unable to determine what was sent and information is lost. If the sending party reduces the symbols being sent to a subset of those available, a scheme can be devised which will prevent ambiguity on the output of the channel. By imposing such a scheme the rate at which  $A$  is sending signals is reduced. However, the confusion of the receiving party is also reduced. The act of designing such a scheme is known as coding. Claude Shannon was the first to discover that for any channel, there is a maximum rate at which you can unambiguously send information. The maximum rate is known as the channel capacity and is defined as the maximum of the mutual information between the input and output over all possible distributions of the input. [2, p. 183-184].

This analysis will discuss discrete memoryless channels. A channel is memoryless if any particular received output symbols is only dependent upon the input which was induced. A formal description of a discrete channel consists of a set of input symbols  $X = \{x_j\}, j = 0, 1, \dots, m$ , and set of output symbols  $Y = \{y_k\}, k = 0, 1, \dots, n$  and a probability transition matrix  $Q_{k|j}$  as shown in equation 1. The transition matrix specifies the probability of receiving any particular output symbol when given an input symbol. Throughout the remainder of the paper, we may refer to individual transition probabilities as  $p(k|j) = p(y_k|x_j)$ . This means the probability of receiving symbol  $y_k$  when symbol  $x_j$  was sent.

$$Q_{k|j} = \begin{pmatrix} p(y_0|x_0) & p(y_1|x_0) & \dots & p(y_n|x_0) \\ p(y_0|x_1) & p(y_1|x_1) & \dots & p(y_n|x_1) \\ \vdots & \vdots & \ddots & \vdots \\ p(y_0|x_m) & p(y_1|x_m) & \dots & p(y_n|x_m) \end{pmatrix} \quad (1)$$

For a discrete memoryless channel, Claude Shannon defined the channel capacity as the maximum of the mutual information between the input and output as in equation 2. This maximization is taken over all input probability distributions. Shannon also showed that this particular function is concave over a closed convex set [3]. Therefore, it could be solved using a simple gradient search or Blahut's iterative algorithm which we are exhibiting here [2, p. 191].

$$C = \max_{p(x_j)} I(X; Y) \quad (2)$$

$$p(x_j) \geq 0, \forall j \text{ and } \sum_{j=0}^m p(x_j) = 1 \quad (3)$$

The mutual information between the input and output, expressed in bits, is defined by Shannon as in equation 4. Shannon proved the existence of an optimal distribution of the input symbols  $p(x_j)$  which could achieve the channel capacity. The proof is non-constructive, but Blahut composed an algorithm that computes the optimal distribution by taking advantage of the convex nature of the problem. His algorithm begins with a guess of the input distribution  $p(x_j)$  and numerically iterates until the distribution represent the maximum mutual information  $I(X; Y)$  [1].

$$I(X; Y) = \sum_{k=0}^n \sum_{j=0}^m p(y_k|x_j) p(x_j) \log_2 \left( \frac{p(y_k|x_j)}{\sum_{j=0}^m p(y_k|x_j) p(x_j)} \right) \quad (4)$$

### A. Blahut's Algorithm

Blahut's algorithm is an iterative solution to the equation 2. The algorithm takes advantage of the concave nature of the definition of mutual information. Mutual information is a concave function over any choice of  $p(x_j)$  for the input distribution. Therefore, the equation can be casted as a convex optimization. Blahut constructed an iterative mapping from one probability distribution to another which will monotonically converge on the maximum of the concave mutual information function. Given a single input and output entity, this maximum is the global maximum mutual information and subsequently the channel capacity [1].

Blahut's algorithm begins by choosing a starting probability distribution at iteration  $r = 0$  as in equation 5. In practice,

choosing a uniform distribution at iteration  $r = 0$  works well. The mapping formulated by Blahut computes the next iteration  $r + 1$  of the probability distribution via equations 6 and 7 [1].

$$p(x_j)^r \geq 0, \forall j \text{ and } \sum_{j=0}^m p(x_j)^r = 1 \quad (5)$$

$$c(x_j)^r = \exp \left( \sum_{k=0}^n Q_{k|j} \log_2 \left( \frac{Q_{k|j}}{\sum_{j=0}^m p(x_j)^r Q_{k|j}} \right) \right) \quad (6)$$

$$p(x_j)^{r+1} = p(x_j)^r \frac{c(x_j)^r}{\sum_{j=0}^m p(x_j)^r c(x_j)^r} \quad (7)$$

Blahut also formulated stopping criteria by specifying an equation which determines the amount of convergence and channel capacity at each iteration  $r$ . The formulation relies on the inequality in equation 8. The stopping criterion for the algorithm occurs when the difference between the upper and low bound of  $C^r$  becomes smaller than a predefined epsilon  $\epsilon$  as in equation 9. Upon convergence the channel capacity is computed as the lower bound of equation 8 as in equation 10 [1].

$$\log_2 \left( \sum_{j=0}^m p(x_j)^r c(x_j)^r \right) \leq C^r \leq \log_2 (\max_j c(x_j)^r) \quad (8)$$

$$\log_2 (\max_j c(x_j)^r) - \log_2 \left( \sum_{j=0}^m p(x_j)^r c(x_j)^r \right) < \epsilon \quad (9)$$

$$C = \max_{p(x_j)} I(X; Y) = \log_2 \left( \sum_{j=0}^m p(x_j)^r c(x_j)^r \right) \quad (10)$$

### B. Binary Erasure Channel Capacity

We implemented Blahut's algorithm using matlab. Also, we validated the algorithm performance by computing the channel capacity of the simple binary erasure channel. The channel capacity of such a simple channel can be computed analytically and therefore makes a good candidate for validation of the algorithm. A depiction of the simple channel is in figure 1. This simple channel models a communication channel in which bits could be lost (erased) during transmission with probability  $\alpha$ . It has been analytically shown that the capacity of this channel is  $C = 1 - \alpha$  and the optimal distribution of the input is achieved with  $p(x_0) = p(x_1) = \frac{1}{2}$  [2, p. 88-189].

The transition matrix for the binary erasure channel is in equation 11. There are two rows representing the input symbols (bits) and three columns representing the output symbols (bits or erased). For verification we defined  $\alpha = 0.2$ , and the matlab implementation computed the  $C = 1 - \alpha = 0.8$  and the input distribution which maximized the channel capacity as  $p(x_0) = p(x_1) = 0.5$ . This result gives evidence of the correctness of implementation. Other parameters and transition matrices were tested, and all validated the correctness of the code.

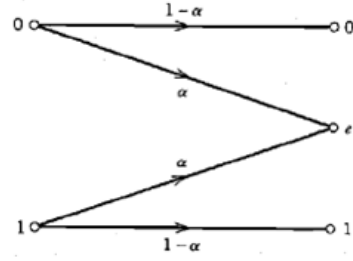


Fig. 1: Binary Erasure Channel

$$Q_{k|j} = \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{pmatrix} \quad (11)$$

## II. EXPERIMENTS

### A. $N^{th}$ Binary Channel Capacity

With the validated matlab implementation of Blahut's algorithm, we explore the channel capacity of a  $n^{th}$  binary channel of varying sizes. The channel is defined by two positive integers,  $m$  and  $n$ . The channel transition matrix is specified by  $Q_{k|j} = \{p(k|j)\}_{j=0, \dots, m}^{m, n}$ . The values of  $p(k|j)$  are given by equation 12 where  $j = 0, \dots, m$  and  $k = 0, \dots, n$ .

$$p(k|j) = \binom{n}{k} \left( \frac{j}{m} \right)^k \left( 1 - \frac{j}{m} \right)^{n-k} \quad (12)$$

The channel defined by 12 is called a generalized  $n^{th}$  binary channel because the channel matrix is characterized by a binomial distribution with  $n$  coin flipping trials. Note that the described channel is of dimensions  $m+1$  inputs and  $n+1$  outputs. We will be computing the channel capacities  $C_{m+1, n+1}$  for channels of dimensions where  $m = 10, 20, 30, 40, 50$  and  $n = 10, 20, 30, 40, 50$ . If we graph the transition matrix for a channel of size  $(m, n) = (50, 50)$ , as in figure 2, we see that most of the transition probability is concentrated in the matrix diagonal. In addition, the transition probability at the diagonal extremes is one. Therefore, in this channel, it is possible to send the symbol  $x_0$  and  $x_{50}$  without error. However, other symbols may be subject to error during transmission.

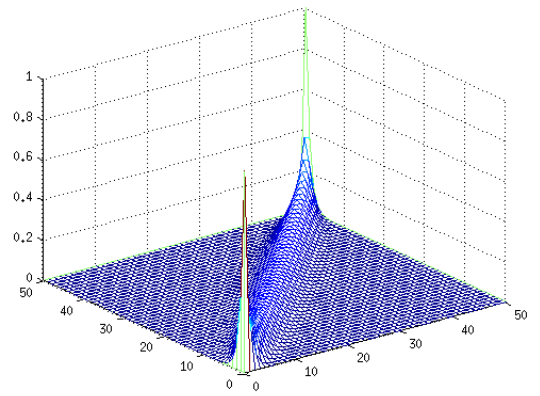


Fig. 2: Transition Matrix of  $n^{th}$  Binary Channel

### B. Analysis of $N^{th}$ Binary Channel Capacities

The channel capacities were computed, in bits, for varying sizes of the  $n^{th}$  binary channel. The results are below in table I. By analyzing the table of capacities, two obvious trends appear.

TABLE I:  $N^{th}$  Binary Channel Capacities

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$m = 10$	1.7771	2.1397	2.3636	2.5054	2.6051
$m = 20$	1.7774	2.1401	2.3672	2.5371	2.6719
$m = 30$	1.7776	2.1405	2.3689	2.5372	2.6701
$m = 40$	1.7777	2.1407	2.3695	2.5375	2.6720
$m = 50$	1.7777	2.1408	2.3695	2.5382	2.6715

The first is seen when examining the columns of of the data. When the number of output symbols remains constant, the capacity of the channel does not increase significantly when the number of input symbols is increased. For example, the difference between the channel capacities when the number of inputs is  $m = 10$  and  $m = 50$ , but the number of outputs remains the same  $n = 50$  is only 0.0664 bits.

The second trend is observed when examining the rows of the data. When the number of input symbols is constant, and the number of output symbols increases, the channel capacity increase considerably. If we plot the rows of the channel capacity matrix in table I, we see a logarithmic increase in the channel capacities as the number of output symbols increases. The plot is in figure 3.

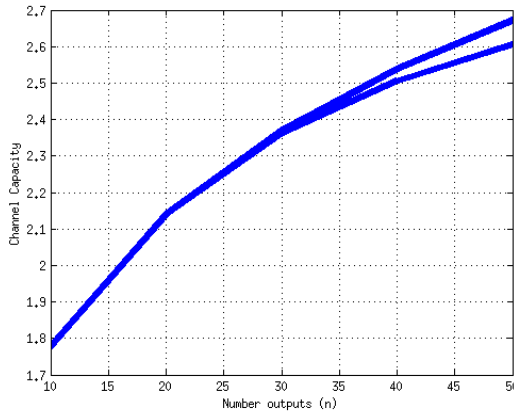


Fig. 3: Channel Capacities as  $n$  Increases

Given the observations made above, it is intuitively clear that the number of output symbols is the limiting factor in this channel. A channel characterized by equation 12 appears to be “output limited” in that the number of output symbols is the dominating factor in the channel capacity. If you examine the transition matrix in figure 2 you will notice that one particular input symbol can map probabilistically to a narrow range of adjacent output symbols with the maximum probability density for a particular symbol being on the correct output symbols. Therefore, increasing the number of output symbols results in an input symbol being mapped unambiguously to a unique

range of output symbols (i.e. the output symbols ranges do not overlap). Therefore, to increase the channel capacity, the number of output symbols must be increased such that every input symbols has a range of output symbols in which to map. Thereby giving each input symbols it’s own unique range of adjacent output symbols so that the receiving entity can determine what has been sent. However, there are diminishing returns in that increasing the number of output symbols results in a logarithm increase of the channel capacity.

### III. CONCLUSION

In this paper we have briefly explained the general definition of a channel as characterized by Claude Shannon. In addition, we have detailed Blahut’s iterative algorithm for computing the channel capacity and optimizing input probability distribution for a particular channel. We implemented the algorithm in matlab and verified the code using simple channels which have analytically tractable solutions. Finally, we described a  $N^{th}$  binary channel and computed channel capacities for a range of channels sizes. Through analysis of these capacities, it was observed that the capacity of the  $n^{th}$  binary channel is increasing when the number of output symbols increases. It was also observed that the capacity is not significantly increased with an increasing number of input symbols. This leads to the conclusion that this particular type of channel is “output limited”.

### REFERENCES

- [1] Richard E. Blahut. Computations of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, pages 460–473, July 1972.
- [2] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley and Sons, Inc., 1991.
- [3] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October 1948.

**Bernard Lampe** (M’09) became an IEEE Member (M) in 2009 and received his bachelors of science degree from The University of Michigan in Ann Arbor, Michigan, USA in 2009.

Mr. Lampe is also a member of the American Society for Computing Machines (ACM) since 2009.