INSTITUTE FOR SYSTEMS RESEARCH

TECHNICAL RESEARCH REPORT

A Communication Channel Modeled on Contagion

by F. Alajaji and T. Fuja

T.R. 93-78 r1

The Institute for Systems Research is supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

A Communication Channel Modeled on Contagion*

Fady Alajaji and Tom Fuja
Department of Electrical Engineering
Institute for Systems Research
University of Maryland
College Park, MD 20742

August 1993

Abstract

We introduce a binary additive communication channel with memory. The noise process of the channel is generated according to the contagion model of George Polya; our motivation is the empirical observation of Stapper *et. al.* that defects in semiconductor memories are well described by distributions derived from Polya's urn scheme. The resulting channel is stationary but not ergodic, and it has many interesting properties.

We first derive a maximum likelihood (ML) decoding algorithm for the channel; it turns out that ML decoding is equivalent to decoding a received vector onto *either* the closest codeword *or* the codeword that is farthest away, depending on whether an "apparent epidemic" has occurred. We next show that the Polya-contagion channel is an "averaged" channel in the sense of Ahlswede (and others) and that its capacity is zero. We then demonstrate that the Polya-contagion channel is a counter-example to the adage, "memory cannot decrease capacity"; the capacity of the Polya-contagion channel is actually *less* than that of the associated memoryless channel. Finally, we consider a finite-memory version of the Polya-contagion model; this channel is (unlike the original) ergodic with a non-zero capacity that increases with increasing memory.

Keywords: Channels with memory, additive noise, capacity.

[0*] Supported in part by NSF grant NCR-8957623; also by the NSF Engineering Research Centers Program, CDR-8803012. Parts of this paper were presented at the 1993 International Symposium on Information Theory, January 17-22, 1993, San Antonio, Texas.

1 Introduction: Communication via Contagion

We consider a discrete communication channel with memory in which errors spread in a fashion similar to the spread of a contagious disease through a population. The errors propagate through the channel in such a way that the occurrence of each "unfavorable" event (i.e., an error) increases the probability of future unfavorable events.

One motivation for the study of such channels is the "clustering" of defects in silicon; Stapper *et. al.* [1] have shown that the distribution of defects in semiconductor memories fits the Polya-Eggenberger (PE) distribution much better than the commonly used Poisson distribution. The PE distribution is one of the "contagious" distributions that can be generated by George Polya's urn model for the spread of contagion [2,3]. More generally, real-world communication channels often have memory; a contagion-based model offers an interesting alternative to the Gilbert model and others.

We begin by introducing a communication channel with additive noise modeled according to the Polya contagion urn scheme. The channel is stationary but not ergodic. We then present a maximum likelihood (ML) decoding algorithm for the channel; ML decoding for the Polya-contagion channel is shown to be equivalent to mapping the received vector onto either the codeword that is closest to the received vector *or* the codeword that is farthest away – depending on which possibility is more *extreme*. We then show that the Polya-contagion channel is in fact an "averaged" channel in the sense of Ahlswede and others [4,5]; – i.e., the block transition probability for the contagion channel is the average of the block transition probabilities of a class of binary symmetric channels, where the expectation is taken with respect to the beta distribution. Using De Finetti's results on exchangeability, we note that binary channels with additive exchangeable noise processes are averaged channels with binary symmetric channels as components.

Using the results of Ahlswede we show that the capacity of the Polya channel is zero; this result gives us a counter-example to the adage "memory can only increase capacity". We note that this adage applies only to stationary ergodic channels, and that for stationary non-ergodic channels, memory may *increase or decrease* capacity.

Finally, we consider a finite-memory version of the Polya-contagion model. The resulting channel is a stationary ergodic Markov channel with memory M; its capacity is positive and increases with M. As M increases, the finite-memory channel converges in distribution to the original Polya channel; however the capacity of the finite-memory channel *does not converge* to the capacity of the Polya channel.

2 Polya-Contagion Communication Channel

Consider a discrete binary additive communication channel – i.e., a channel for which the i^{th} output $Y_i \in \{0,1\}$ is the modulo-two sum of the i^{th} input $X_i \in \{0,1\}$ and the i^{th} noise symbol $Z_i \in \{0,1\}$; more succinctly, $Y_i = X_i \oplus Z_i$, for $i = 1, 2, 3, \ldots$

We assume that the input and noise sequences are independent of each other. The noise sequence $\{Z_i\}_{i=1}^{\infty}$ is drawn according to the Polya contagion urn scheme [6], as follows: An urn originally contains T balls, of which R are red and S are black ($T = R + S$); let $\rho = R/T$ and $\sigma = 1 - \rho = S/T$. We make successive draws from the urn; after each draw, we return to the urn $1 + \Delta$ balls of the same color as was just drawn. Note that if $\Delta = 0$, we get the classic case of independent drawings with replacement. In our problem we will assume that $\Delta > 0$ (contagion case) and that $\rho < \sigma$ – i.e. $\rho < 1/2$. Furthermore, we denote $\delta = \Delta/T$. Our sequence $\{Z_i\}$ corresponds to the outcomes of the draws from our Polya urn with parameters ρ and δ, where:

$$Z_i = \begin{cases} 1, & \text{if the } i^{th} \text{ ball drawn is red;} \\ 0, & \text{if the } i^{th} \text{ ball drawn is black.} \end{cases}$$

In Polya's model, a red ball in the urn represents a sick person in the population and a black ball in the urn represents a healthy person.

2.1 Block Transition Probability of the Channel

Definition 1 (Channel state) We define the state of the channel after the n^{th} transmission to be the total number of red balls drawn after n trials:

$$S_n \stackrel{\text{def}}{=} Z_1 + Z_2 + \cdots + Z_n = S_{n-1} + Z_n$$

$$S_0 = 0$$

The possible values of S_n are the elements of the set $\{0, 1, \ldots, n\}$. Therefore the channel at time n has $n + 1$ possible states. Furthermore, note that the sequence of states $\{S_n\}_{n=1}^{\infty}$ form a Markov chain, i.e.

$$P(S_n = s_n \mid S_{n-1} = s_{n-1}, S_{n-2} = s_{n-2}, \ldots, S_1 = s_1) = P(S_n = s_n \mid S_{n-1} = s_{n-1})$$

For a given input block $\underline{X} = [X_1, X_2, \ldots, X_n]$ and a given output block $\underline{Y} = [Y_1, Y_2, \ldots, Y_n]$, the block transition probability of the channel is given by

$$P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) = \prod_{i=1}^{n} P(Y_i = y_i \mid X_i = x_i, S_{i-1} = s_{i-1})$$

where

$$P(Y_i = y_i \mid X_i = x_i, S_{i-1} = s_{i-1}) = \left[\frac{\rho + s_{i-1}\delta}{1 + (i-1)\delta}\right]^{y_i \oplus x_i} \left[\frac{\sigma + (i - 1 - s_{i-1})\delta}{1 + (i-1)\delta}\right]^{1-(y_i \oplus x_i)}.$$

We thus obtain:

$$P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) = \frac{\rho(\rho+\delta)\cdots(\rho+(d-1)\delta)\sigma(\sigma+\delta)\cdots(\sigma+(n-d-1)\delta)}{(1+\delta)(1+2\delta)\cdots(1+(n-1)\delta)} \quad (1)$$

or

$$P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) = \frac{\Gamma(\frac{1}{\delta})\ \Gamma(\frac{\rho}{\delta} + d)\ \Gamma(\frac{\sigma}{\delta} + n - d)}{\Gamma(\frac{\rho}{\delta})\ \Gamma(\frac{\sigma}{\delta})\ \Gamma(\frac{1}{\delta} + n)} \quad (2)$$

where $d = d(\underline{y}, \underline{x}) = weight(\underline{z} = \underline{y} \oplus \underline{x}) = s_n$ and $\Gamma(\cdot)$ is the gamma function, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ for $x > 0$. To obtain equation (2) from equation (1), we used the fact that $\Gamma(x+1) = x\ \Gamma(x)$ which leads to the following identity:

$$\prod_{j=0}^{n-1}(\alpha + j\beta) = \beta^n\ \frac{\Gamma(\frac{\alpha}{\beta} + n)}{\Gamma(\frac{\alpha}{\beta})}$$

2.2 Properties of the Channel

Before analyzing the characteristics of the channel, we state from [7] the following definitions and lemma.

Definition 2 A finite sequence of random variables $\{Z_1, Z_2, \cdots, Z_n\}$ is said to be *exchangeable* if the joint distribution of $\{Z_1, Z_2, \cdots, Z_n\}$ is invariant with respect to permutations of the indices $1, 2, \cdots, n$.

Definition 3 An infinite sequence of random variables $\{Z_i\}_{i=1}^{\infty}$ is said to be *exchangeable* if for every finite n, the collection $\{Z_{i_1}, Z_{i_2}, \cdots, Z_{i_n}\}$ is exchangeable.

Lemma 1 Exchangeable random processes are *strictly stationary.*

Exchangeability was investigated by De Finetti (1931) who recognized its fundamental role for Bayesian statistics and modern probability. The main interest in adopting this concept is to use exchangeable random variables as an alternative to independent identically distributed (*iid*) random variables. Note that *iid* random variables are exchangeable. However, exchangeable random variables are *dependent* in general but symmetric in their dependence.

We now can study the properties of the channel:

1. **Symmetry:**

   The channel is *symmetric*. By this we mean that $P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x})$ depends only on $\underline{x} \oplus \underline{y}$ since $P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) = P(\underline{Z} = \underline{y} \oplus \underline{x})$. Due to the symmetry, if we want to maximize the mutual information $I(\underline{X}; \underline{Y})$ over all input distributions on $\underline{X}$, the result is maximized for equiprobable input n-tuples.

2. **Stationarity:**

   From equation (1) and the above definitions, we can conclude that the noise process $\{Z_i\}_{i=1}^{\infty}$ forms an *exchangeable* random process. The noise process is thus *strictly stationary* (by Lemma 1) and thus *identically distributed.* We get:

   $$P(Z_i = 1) = \rho = 1 - P(Z_i = 0) \quad \forall\, i = 1, 2, 3, \ldots$$

   and the correlation coefficient

   $$Cor(Z_i, Z_j) = \frac{Cov(Z_i, Z_j)}{\sqrt{Var(Z_i)\; Var(Z_j)}} = \frac{\delta}{1+\delta} > 0 \quad \forall\, i \neq j$$

   indicates the positive correlation among the random variables of the noise process.

3. **Non-Ergodicity :** $\{S_n/n\}$ is a martingale [8]; using the martingale convergence theorem, we obtain that $Z \stackrel{\text{def}}{=} \lim_{n\to\infty} S_n/n$ exists almost surely. It is shown in [6] that Z has the beta distribution with parameters ρ/δ and σ/δ. Thus the noise process $\{Z_i\}_{i=1}^{\infty}$ is *not ergodic* since its sample average does not converge to a constant.

3 Maximum Likelihood (ML) Decoding

Suppose M codewords are possible inputs to the channel with transition probability $P(\underline{Y} = \underline{y}|\underline{X} = \underline{x})$; the codebook is given by $\mathcal{C} = \{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M\}$, with each $\underline{x}_k \in \{0,1\}^n$. For a given received vector $\underline{y} \in \{0,1\}^n$ the maximum likelihood estimate of the transmitted codeword is

$$\underline{x} = \arg\max\{P(\underline{Y} = \underline{y}|\underline{X} = \underline{x}_k) : \underline{x}_k \in \mathcal{C}\}.$$

From equation (2), we can rewrite the transition probability of the channel as:

$$P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) = g(d(\underline{x}, \underline{y}))$$

where $g : [0,n] \to [0,1]$ is defined by

$$g(d) = A \cdot \Gamma\left(\frac{\rho}{\delta} + d\right) \cdot \Gamma\left(\frac{\sigma}{\delta} + n - d\right)$$

and A is a constant depending on n, ρ, and δ.

Recall that a positive-valued function $f(\cdot)$ is *log-convex* if $\log[f(\cdot)]$ is a convex function; log-convex functions are convex functions, and they're closed under addition and multiplication [9]. Furthermore, $\Gamma(\cdot)$ is strictly log-convex, meaning that $g(\cdot)$ defined above is strictly log-convex on the interval $[0,n]$. This observation leads to the following result.

Proposition 1 The transition probability function $P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x})$ of the Polya-contagion channel is strictly log-convex in $d(\underline{x}, \underline{y})$ and has a unique minimum at

$$d_0 = \frac{n}{2} + \frac{1 - 2\rho}{2\delta}.$$

Furthermore, $P(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x})$ is symmetric in $d(\underline{x}, \underline{y})$ about d_0.

Proof 1 As above, define $g(d) = P(\underline{Y} = \underline{y}|\underline{X} = \underline{x})$ for any $\underline{x}, \underline{y}$ such that $d(\underline{x}, \underline{y}) = d$; then $g(\cdot)$ is strictly log-convex. For $d_0 = (n/2) + ((1 - 2\rho)/2\delta)$, we obtain

$$g(d_0 + \epsilon) = g(d_0 - \epsilon) = A\ \Gamma\left(\frac{n}{2} + \frac{1}{2\delta} + \epsilon\right)\ \Gamma\left(\frac{n}{2} + \frac{1}{2\delta} - \epsilon\right)$$

for any ϵ; therefore $g(\cdot)$ is symmetric about d_0 and the strict convexity of $g(\cdot)$ means that a unique minimum occurs there. ■

Decoding Algorithm: From the results above, the ML decoding algorithm for the channel is as follows:

1. For a given n-tuple $\underline{y}$ received at the channel output, compute $d_i \stackrel{\text{def}}{=} d(\underline{y}, \underline{x}_i)$, for $i = 1, \ldots, M$. Compute also $d_{max} \stackrel{\text{def}}{=} \max_{1 \le i \le M}\{d_i\}$ and $d_{min} \stackrel{\text{def}}{=} \min_{1 \le i \le M}\{d_i\}$.

2. If $|d_{max} - d_0| \le |d_{min} - d_0|$, map $\underline{y}$ onto a codeword $\underline{x}_j$ for which $d_j = d_{min}$. In this case ML decoding $\Longleftrightarrow$ minimum distance decoding.

3. If $|d_{max} - d_0| > |d_{min} - d_0|$, map $\underline{y}$ onto a codeword $\underline{x}_j$ for which $d_j = d_{max}$. In this case ML decoding $\Longleftrightarrow$ maximum distance decoding.

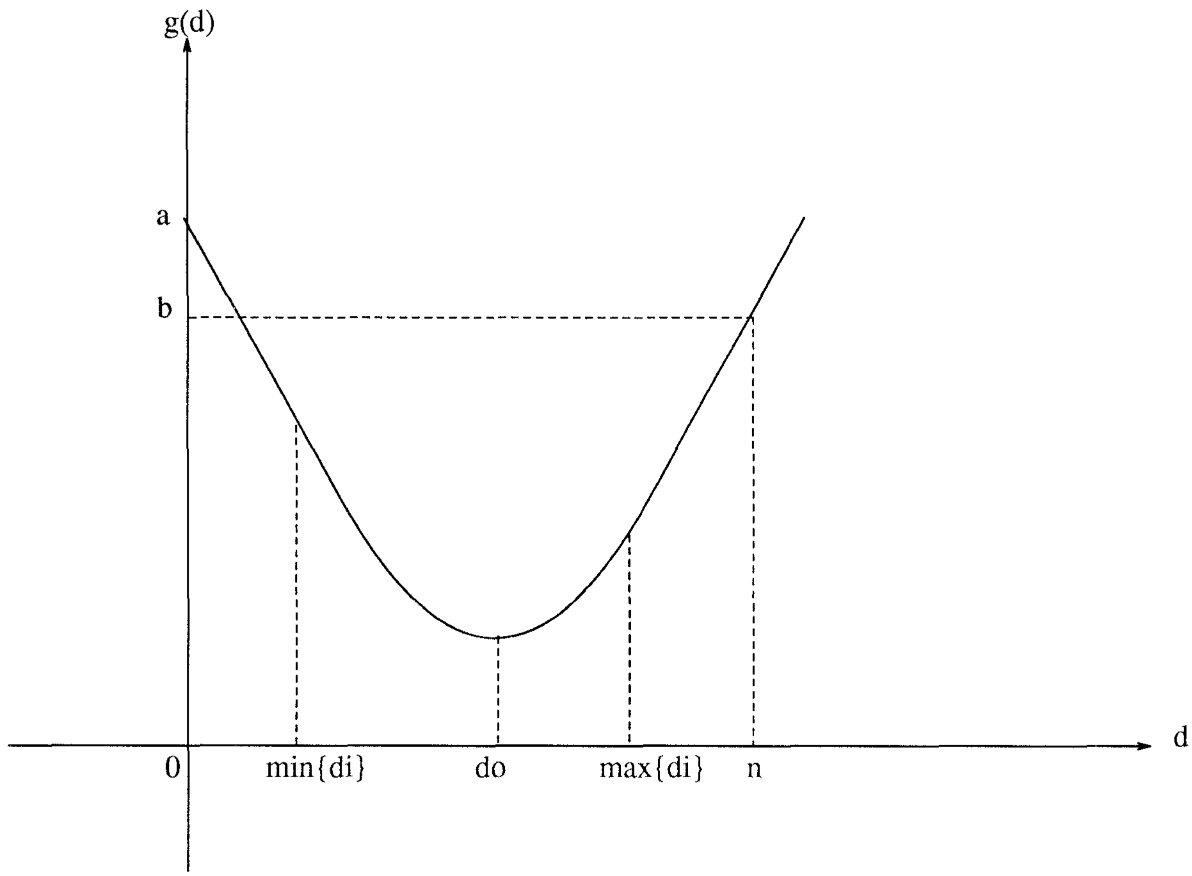

Figure 1: Transition probability function vs Hamming distance d

In Figure 1, we have that

$$a = g(0) = \frac{\Gamma(\frac{1}{\delta})\,\Gamma(\frac{\sigma}{\delta} + n)}{\Gamma(\frac{\sigma}{\delta})\,\Gamma(\frac{1}{\delta} + n)} \quad \text{and} \quad b = g(n) = \frac{\Gamma(\frac{1}{\delta})\,\Gamma(\frac{\rho}{\delta} + n)}{\Gamma(\frac{\rho}{\delta})\,\Gamma(\frac{1}{\delta} + n)}.$$

Observations:

- Insight into the decoding rule:

  – We can rewrite d_0 as:

  $$d_0 = \frac{n}{2} + \frac{1}{\Delta}\left(\frac{T}{2} - R\right).$$

  Note that $n/2$ is (of course) the distance the received n-tuple would be from the transmitted codeword if *half* of the bits get flipped; note also that $(T/2 - R)$ is the initial offset from having an equal number of red balls and black balls in the urn. Thus d_0 may be thought of as an equilibrium point.

  – The best estimate is then specified by the value of d_i that is furthest away from the equilibrium point d_0. In other words, the best decision is based on the following reasoning: either *many* errors occurred during transmission – an *apparent epidemic*, to use the contagion interpretation – or very *few* errors occured – an apparently healthy population.

- We note that if, $d_0 > n - 0.5$, then condition (2) in the above algorithm is always satisfied – meaning minimum distance decoding is optimal. The requirement $d_0 > n - 0.5$ is equivalent to the condition

  $$\delta < \frac{1-2\rho}{n-1},$$

  so if the parameter $\delta = \Delta/T$ is sufficiently small – i.e., there is sufficiently little memory in the system – minimum distance decoding is optimal. In particular, if $\delta = 0$, the draws from the urn are independent and the channel reduces to a binary symmetric channel with crossover probability ρ. Thus this observation is consistent with the fact that, for a BSC with crossover probability less than one-half, minimum-distance decoding is maximum likelihood decoding.

4 Averaged Communication Channels

Averaged channels with discrete memoryless components were first introduced by Jacobs [4] and then were analyzed by Ahlswede [5] and Kieffer [10], who investigated their capacity. We will show that the Polya-contagion channel is an averaged channel with components that are binary symmetric channels (BSC's).

Consider a family of stationary ergodic channels parameterized by θ:

$$\left\{W_\theta^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}), \theta \in \Theta\right\}_{n=1}^{\infty}$$

where $\underline{Y}$ and $\underline{X}$ are respectively the input and output blocks of the channel, each of length n. $W_\theta^{(n)}(\cdot)$ is the block transition probability of the stationary ergodic channel specified by the parameter $\theta \in \Theta$.

Definition 4 We say a channel is an "averaged" channel with stationary ergodic components if its block transition probability is the expected value of the transition probabilities of a class of stationary ergodic channels parameterized by θ – i.e., if it's of the form:

$$\begin{aligned} W_{\text{avg}}^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) &= \int_\Theta W_\theta^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) \; dG(\theta) \\ &= E_\theta[W_\theta^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x})] \qquad (3) \end{aligned}$$

where $(\Theta, \sigma(\Theta), G)$ is the probability space on which the random variable θ is defined.

Note that an averaged channel is stationary and may have memory. One way an averaged channel may be realized is as follows: From among the stationary ergodic components, nature selects one according to some probability distribution G. This component is then used for the entire transmission. However the selection is unknown to both the encoder and the decoder.

We will show that the Polya channel – and indeed *any* non-ergodic additive channel – belongs to this class of channels. But first we need to recall some results from [11,12]:

Notation: Consider a discrete time random process with alphabet D, σ-field $\sigma(D^\infty)$ consisting of subsets of the space D^∞ of sequences $u = (u_1, u_2, \ldots)$, $u_i \in D$, a probability measure μ on the space $(D^\infty, \sigma(D^\infty))$ forming a probability space $(D^\infty, \sigma(D^\infty), \mu)$ and a coordinate or sampling function $\mathbf{U}_n : D^\infty \to D$ defined by $\mathbf{U}_n(u) = u_n$. The sequence of random variables $\{\mathbf{U}_n; n = 1, 2, \ldots\}$ is a discrete time random process. As convenient, random processes will be denoted by either $\{\mathbf{U}_n\}$ or by $[D, \mu, \mathbf{U}]$.

Lemma 2 (Ergodic Decomposition) Let $[D, \mu, \mathbf{U}]$ be a stationary, discrete time random process. There exists a class of stationary ergodic measures $\{\mu_\theta; \theta \in \Theta\}$ and a probability measure G on an event space of Θ such that for every event $F \subset \sigma(D^\infty)$ we can write:

$$\mu(F) = \int_\Theta \mu_\theta(F) \; dG(\theta)$$

Remark: The ergodic decomposition theorem states that, in an appropriate sense, all stationary non-ergodic random processes are a mixture of stationary ergodic processes; if we are viewing a stationary non-ergodic process, we are viewing a stationary ergodic process selected by nature according to some probability measure G. Therefore, by directly applying the ergodic decomposition theorem we get the following result:

Proposition 2 Any discrete channel with stationary (non-ergodic) additive noise is an averaged channel with channels with additive stationary ergodic noise as components.

Proof 2 Let $\{Z_i\}$ be the (non-ergodic) noise sequence. Then the ergodic decomposition theorem states that $P(\underline{Z} = \underline{z}) = P(Z_1 = z_1, \ldots, Z_n = z_n)$ may be written as the expected value of the distribution of a class of stationary ergodic processes; since the noise and input sequences are independent, we have $W^{(n)}(\underline{Y} = \underline{y}|\underline{X} = \underline{x}) = P(\underline{Z} = \underline{y} - \underline{x})$ and so $W^{(n)}(\underline{Y} = \underline{y}|\underline{X} = \underline{x})$ may likewise be expressed as the expected value of the transition probabilities of a class of stationary ergodic additive noise channels. ■

Proposition 3 The binary Polya-contagion channel is an averaged channel; its components are BSC's with crossover probability θ, where θ is a beta-distributed random variable with parameters ρ/δ and σ/δ.

Proof 3 We showed in Proposition 2 that the Polya channel is an averaged whose components are channels with additive stationary ergodic noise. To prove the rest of the proposition we just note that, if we let $f_\Theta(\theta)$ be the pdf of a beta-distributed random variable with parameters ρ/δ and σ/δ – i.e.,

$$f_\Theta(\theta) = \begin{cases} \dfrac{\Gamma(1/\delta)}{\Gamma(\rho/\delta)\Gamma(\sigma/\delta)} \theta^{\rho/\delta - 1}(1-\theta)^{\sigma/\delta - 1}, & \text{if } 0 < \theta < 1; \\ 0, & \text{otherwise,} \end{cases}$$

then

$$\int_0^1 \theta^{d(\underline{x},\underline{y})}(1-\theta)^{n-d(\underline{x},\underline{y})} f_\Theta(\theta) d\theta = P(\underline{Y} = \underline{y}|\underline{X} = \underline{x}),$$

where $P(\underline{Y} = \underline{y}|\underline{X} = \underline{x})$ is the transition probability of the Polya-contagion channel from (2). ■

Observation: We could have proved part of Proposition 3 by using De Finetti's results on exchangeability, since the additive noise process of the Polya channel is a binary exchangeable random process. De Finetti's results are summarized in the following theorem and corollary [8,13]:

Theorem 1 (De Finetti) For an infinite sequence of random variables, the concept of exchangeability is equivalent to that of conditional independence with a common marginal distribution; i.e. if $Z_1, Z_2, \ldots$ is an infinite sequence of exchangeable random variables, then there exists a σ-field $\mathcal{F}$ and a distribution G such that, given $\mathcal{F}$, the random variables $Z_1, Z_2, \ldots$ are conditionally independent with distribution function G.

Corollary 1 For every infinite sequence of exchangeable random variables $\{Z_i\}$ such that $Z_i \in \{0,1\}$, there corresponds a probability distribution G concentrated on the interval $(0,1)$ such that:

$$P(Z_1 = e_1, Z_2 = e_2, \ldots, Z_n = e_n) = \int_0^1 \theta^k \ (1-\theta)^{n-k} \ dG(\theta)$$

where $k = e_1 + e_2 + \cdots + e_n$ and $e_i \in \{0,1\}$ for $i = 1, 2, \ldots, n$.

This brings us to the following more general result:

Proposition 4 Any binary channel with an exchangeable additive noise process is an averaged channel with binary symmetric channels (BSC's) as its components.

5 Capacity of Averaged Communication Channels

Strong vs Weak Capacity: We briefly describe what we mean by the weak (or *operational*) capacity and the strong capacity of a communication channel.

Consider a discrete (not-necessarily memoryless) channel with common input and output alphabet A; let $W^{(n)}(\underline{Y} = \underline{y} | \underline{X} = \underline{x})$ be the block transition probability describing the channel.

Definition 5 An (M, n, ϵ) code for this channel is a collection of M pairs

$$\{(\underline{x}_1, B_1), (\underline{x}_2, B_2), \ldots, (\underline{x}_M, B_M)\}$$

where

- $\underline{x}_i \in A^n$ for $i = 1, 2, \ldots, M$; these are the *codewords.*

- $B_i \subset A^n$ such that $B_i \bigcap B_j = \emptyset$ for $i \neq j$; these are the decoding sets for the code – i.e., if an element of B_i is received then it will be assumed that $\underline{x}_i$ was sent.

- The maximum decoding error probability is given by ϵ; that is, if we let $\underline{Y}_i$ be the random n-tuple appearing at the channel output when the codeword $\underline{x}_i$ is transmitted – and so $P(\underline{Y}_i = \underline{y}) = W^{(n)}(\underline{y}|\underline{x}_i)$ – then

$$\max\{P(\underline{Y}_i \notin B_i) : i = 1, 2, \ldots, M\} \leq \epsilon.$$

The *rate* of an (M, n, ϵ) code is $R = (1/n)\log_2(M)$.

A rate R is *admissable* if, for any $\epsilon > 0$, there exists (for sufficiently large n) an (M, n, ϵ) code with $M \geq 2^{nR}$. We define the weak capacity of a channel as the supremum of all admissable rates. We establish that a particular nonnegative number C_w (resp., C_s) is the weak (strong) capacity of a given channel by proving a *coding theorem* and a *weak (strong) converse.*

Coding Theorem For any $R < C$, there exists a sequence of $(2^{nR}, n, \epsilon_n)$ codes such that $\lim_{n\to\infty} \epsilon_n = 0$. Here C represents either C_w or C_s, according to the capacity sought.

Weak Converse Given any sequence of $(2^{nR}, n, \epsilon_n)$ codes with $R > C_w$, there exists an $\epsilon > 0$ such that for sufficiently large n, $\epsilon_n > \epsilon$.

Strong Converse Given any sequence of $(2^{nR}, n, \epsilon_n)$ codes with $R > C_s$, $\lim_{n\to\infty} \epsilon_n = 1$.

Weak Capacity of Averaged Channels: The strong capacity of an averaged channel does not in general exist [5], since the strong converse to the channel coding theorem may not hold. However it was shown by Ahlswede [5] that the weak converse holds for these channels. We now give the formula of the weak capacity of an averaged channel [10]:

Lemma 3 Consider the averaged channel with stationary ergodic components described by (3); assume common input and output alphabets A and "averaging" distribution $G(\cdot)$, which may be either discrete or continuous.

Then the weak capacity of the averaged channel is given by

$$C_{\text{avg}} = \lim_{\alpha\to 0} C(\alpha) \tag{4}$$

where

$$C(\alpha) = \max_{Q} \sup_{\{E\in\sigma(\Theta):\, G(E)\geq 1-\alpha\}} \inf_{\theta\in E} i(Q; W_\theta), \tag{5}$$

$$i(Q; W_\theta) = \lim_{n\to\infty} \frac{1}{n} I(Q^{(n)}; W_\theta^{(n)})$$

and

$$I(Q^{(n)}; W_\theta^{(n)}) = \sum_{\underline{x},\underline{y}\in A^n} W_\theta^{(n)}(\underline{y} \mid \underline{x})\ Q^{(n)}(\underline{x})\ \log_2 \frac{W_\theta^{(n)}(\underline{y} \mid \underline{x})}{\sum_{\underline{\tilde{x}}\in A^n} W_\theta^{(n)}(\underline{y} \mid \underline{\tilde{x}}) Q^{(n)}(\underline{\tilde{x}})}$$

Capacity of the Polya Channel: We use the above lemma to compute the capacity of the Polya channel. Since the additive noise is independent of the input, the maximization over the input distribution Q in equation (5) is realized by the uniform input distribution. We can therefore interchange the inf and max in (5) and get:

$$\max_Q\ i(Q; W_\theta) = 1 - h(W_\theta)$$

The resulting capacity of the channel is:

$$C_{\text{Polya}} = 1 - \text{ess}_\Theta \sup\ h(W_\theta) \tag{6}$$

where

- the noise entropy rate $h(W_\theta)$ is given by

$$h(W_\theta) = -\lim_{n\to\infty} \frac{1}{n} \sum_{\underline{x},\underline{y}\in A^n} W_\theta^{(n)}(\underline{y} \mid \underline{x})\ Q^{(n)}(\underline{x})\ \log_2\ W_\theta^{(n)}(\underline{y} \mid \underline{x})$$

- and the essential supremum is defined by

$$\text{ess}_\Theta \sup\ f(\theta) \stackrel{\text{def}}{=} \inf\ [r : dG(f(\theta) \le r) = 1]$$

We know that the stationary ergodic components of the Polya channel are BSC's; therefore the noise entropy rate is given by $H(W_\theta) = h_b(\theta)$, where $h_b(x) = -x\log_2(x) - (1-x)\log_2(1-x)$. Equation (6) then yields the weak capacity of the channel:

$$C_{\text{Polya}} = 1 - \text{ess}_\Theta \sup\ h_b(\theta)$$

Since θ has the beta distribution on [0,1], we obtain $\text{ess}_\Theta \sup\ h_b(\theta) = 1$ which, in turn, implies $C_{\text{Polya}} = 0$.

Comments: The zero-capacity of the Polya channel is due to the fact that θ can occur in any neighborhood of the point 1/2 with positive probability. This channel behaves like a compound channel and the capacity of compound channels is defined from a pessimistic point of view as the worst case achievable rate.

The zero capacity result suggests that the Polya channel might not be a good model for a realistic channel. However in Section 7 we will consider a finite-memory channel that approximates the Polya channel as memory increases, but with a capacity that does *not* approach zero. Before we do so, however, we first point out that the Polya channel provides a counterexample to the adage "memory increases capacity"; this is the subject of the next section.

6 Effect of Memory on the Capacity of the Polya Channel

Pinsker and Dobrushin [14] showed that "for a wide class" of channels, the capacity of a channel with memory is *not less* than the capacity of the "equivalent" memoryless channel. They considered a channel with input alphabet A, output alphabet B, and n-fold transition probability $W^{(n)}(y_1, \ldots, y_n \mid x_1, \ldots, x_n)$, $x_i \in A$, $y_i \in B$, such that:

$$W^{(n-1)}(y_1, \ldots, y_{n-1} \mid x_1, \ldots, x_{n-1}) = \sum_{y_n \in B} W^{(n)}(y_1, \ldots, y_n \mid x_1, \ldots, x_n). \tag{7}$$

Specifically, they considered such channels with an operational capacity given by

$$C = \lim_{n \to \infty} \frac{1}{n} C_n \tag{8}$$

where

$$C_n = \sup_{(X_1, \ldots, X_n)} I\left((X_1, \ldots, X_n); (Y_1, \ldots, Y_n)\right)$$

They then defined the memoryless channel associated with this channel to have n-fold transition probability

$$\check{W}^{(n)}(y_1, \ldots, y_n | x_1, \ldots, x_n) = \prod_{i=1}^{n} \check{W}_i(y_i | x_i),$$

where

$$\check{W}_i(y_i \mid x_i) = \sum_{y_1, \ldots, y_{i-1} \in B} W^{(i)}(y_1, \ldots, y_i \mid x_1, \ldots, x_i) \tag{9}$$

and the left-hand side of (9) is assumed to be independent of $(x_1, \ldots, x_{n-1})$. Thus the one-step transition probabilities of the memoryless channel are equal to the per-letter marginals of the channel with memory. The capacity of the associated memoryless channel is denoted by $\tilde{C}_n$, and they showed that:

$$C_n \geq \tilde{C}_1 + \tilde{C}_2 + \cdots + \tilde{C}_n.$$

By "a wide class of channels", they made some implicit assumtions:

- The channels are non-anticipatory and historyless. The non-anticipatory property can be seen from equation (7), where it is implicitely assumed that

$$W^{(n-1)}(y_1, \ldots, y_{n-1}|x_1, \ldots, x_{n-1}) = W^{(n-1)}(y_1, \ldots, y_{n-1}|x_1, \ldots, x_n).$$

  Furthermore, equation (9) assumes the distribution on the n^{th} channel output given the first n channel inputs depends only on the n^{th} input – i.e., the original channel has no input memory; it is historyless.

- The channels are asymptotically mean stationary [15] and ergodic, since otherwise equation (8) may not represent the operational capacity.

If we restrict ourselves to stationary channels, then $\check{W}_n(y_n \mid x_n) = \check{W}(y_n \mid x_n)$ for all n and so $C_n \geq n\tilde{C}$, where $\tilde{C}_1 = \tilde{C}_2 = \ldots = \tilde{C}_n = \tilde{C}$. Thus we get $C \geq \tilde{C}$

In [16] Ahlswede showed that there are averaged channels for which the introduction of memory *decreases* capacity. We briefly show that the Polya-contagion channel is such a channel.

We showed in the previous section that the capacity of the Polya channel is zero. Now, let us compute $\tilde{C}$, the capacity of the associated memoryless channel. The transition probability of the associated memoryless channel $\check{W}(\cdot)$ is

$$\begin{aligned}
\check{W}(Y_n = y_n|X_n = x_n) &= \sum_{y_1,\ldots,y_{n-1}\in\{0,1\}} \int_\Theta W_\theta^{(n)}(\underline{Y} = \underline{y}|\underline{X} = \underline{x})\ dG(\theta) \\
&= \int_\Theta \sum_{y_1,\ldots,y_{n-1}\in\{0,1\}} W_\theta^{(n)}(\underline{Y} = \underline{y}|\underline{X} = \underline{x})\ dG(\theta) \\
&= \int_\Theta W_\theta^{(1)}(Y_n = y_n|X_n = x_n)\ dG(\theta)
\end{aligned}$$

$$= \int_0^1 \theta^{y_n \oplus x_n} (1-\theta)^{1-(y_n \oplus x_n)} f_\Theta(\theta)\, d\theta$$

$$= \rho^{y_n \oplus x_n} (1-\rho)^{1-(y_n \oplus x_n)}$$

where $\Theta = [0,1]$, and $dG(\theta) = f_\Theta(\theta)\, d\theta$ is the beta distribution with parameters ρ/δ and σ/δ given in Proposition 3.

Thus we observe that the memoryless channel equivalent to the Polya channel is a BSC with crossover probability ρ; this leads us to the conclusion that, for $\rho \neq 1/2$, the memory in the Polya channel *decreases* capacity.

Finally, we can find examples of stationary non-ergodic channels for which memory increases capacity [17]; this leads us to conclude that for stationary *non-ergodic* channels, memory may *increase* or *decrease* capacity.

7 Finite-Memory Contagion Channel

An unrealistic aspect of the Polya channel is its infinite memory. Consider, for instance, the millionth ball drawn from Polya's urn; the very *first* ball drawn from the urn and the 999,999'th ball drawn from the urn have an identical effect on the outcome of the millionth draw. In the context of a communication channel, this is not reasonable; we would assume that the effects of the "disease" fade in time. We now consider a more realistic model for a contagion channel with finite memory, where the noise in the additive channel is generated according to a modified version of the Polya urn scheme.

Consider a discrete binary additive non-anticipatory communication channel described by the following equation: $Y_i = X_i \oplus Z_i$, for $i = 1, 2, 3, \ldots$ where $\oplus$ is modulo-two addition, and X_i, Z_i and Y_i are respectively the i^{th} input, noise and output symbols of the channel.

We assume, as for the Polya channel, that the input and noise sequences are independent of each other. The sequence of random variables $\{Z_i\}_{i=1}^{\infty}$ of the noise process is modeled according to the following urn scheme: An urn contains originally T balls, of which R are red and S are black ($T = R + S$). At the j'th draw, $j = 1, 2, \ldots$, we select a ball from the urn and replace it with $1 + \Delta$ balls of the same color ($\Delta > 0$); then, M draws later - after the $(j + M)$'th draw - we retrieve from the urn Δ balls of the color picked at time j. Here also, we let $\rho = R/T$, $\sigma = 1 - \rho = S/T$ and $\delta = \Delta/T$. Furthermore, we assume that $\rho < 1/2$. The noise process

$\{Z_i\}$ corresponds to the outcomes of the draws from the urn, where:

$$Z_i = \begin{cases} 1, & \text{if the } i^{th} \text{ ball drawn is red;} \\ 0, & \text{if the } i^{th} \text{ ball drawn is black.} \end{cases}$$

Observation: With this modification of the original Polya urn scheme, the total number of balls in the urn is constant ($T + M\Delta$ balls) after an initialization period of M draws. It also limits the effect of any draw to M draws in the future.

7.1 The Distribution of the Noise

During the initialization period ($n \leq M$), the process $\{Z_i\}$ of the finite-state channel is identical to the Polya noise process discussed earlier. We now study the noise process for $n \geq M + 1$.

Let R_n be the number of red balls in the urn after n draws, T_n be the total number of balls in the urn after n draws, and $r_n = R_n/T_n$. Then $T_n = T + M\Delta$ for $n \geq M + 1$, and so

$$\begin{aligned} r_n &= \frac{R + (Z_n + Z_{n-1} + \cdots + Z_{n-M+1})\Delta}{T + M\Delta} \\ &= \frac{\rho + (Z_n + Z_{n-1} + \cdots + Z_{n-M+1})\delta}{1 + M\delta} \end{aligned}$$

We now have that:

$$\begin{aligned} P(Z_n = 1 | Z_1 = e_1, \ldots, Z_{n-1} = e_{n-1}) &= \frac{\rho + (e_{n-1} + e_{n-2} + \cdots + e_{n-M})\delta}{1 + M\delta} \\ &= r_{n-1} \\ &= P(Z_n = 1 | Z_{n-M} = e_{n-M}, \ldots, Z_{n-1} = e_{n-1}) \end{aligned}$$

where $e_i = 0$ or 1, for $i = 1, 2, \ldots, n-1$ and where $n \geq M + 1$. Thus the noise process $\{Z_i\}_{i=M+1}^{\infty}$ is a Markov process of order M. The resulting channel is thus a Markov channel with memory M; we shall refer to it as the finite-memory contagion channel.

For an input block $\underline{X} = [X_1, X_2, \ldots, X_n]$ and an output block $\underline{Y} = [Y_1, Y_2, \ldots, Y_n]$, the block transition probability of the resulting binary channel is as follows:

- For blocklength $n \leq M$, the block transition probability of this channel is identical to that of the Polya-contagion channel given by equations (1) and (2).

- For $n \geq M+1$, we obtain:

$$\begin{aligned} P_M(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) &= P(\underline{Z} = \underline{e}) \\ &= \prod_{i=1}^{n} P(Z_i = e_i \mid Z_{i-1} = e_{i-1}, \ldots, Z_{i-M} = e_{i-M}) \\ &= L \prod_{i=M+1}^{n} \left[\frac{\rho + s_{i-1}\delta}{1+M\delta}\right]^{e_i} \left[\frac{\sigma + (M - s_{i-1})\delta}{1+M\delta}\right]^{1-e_i} \end{aligned} \tag{10}$$

  where

$$e_i = x_i \oplus y_i,$$

$$L = \frac{\prod_{i=0}^{k-1}(\rho + i\delta)\prod_{j=0}^{M-1-k}(\sigma + j\delta)}{\prod_{\ell=1}^{M-1}(1+\ell\delta)},$$

$$k = e_1 + \cdots + e_M,$$

  and

$$s_{i-1} = e_{i-1} + \cdots + e_{i-M}.$$

By examining the above equation we see that the noise process (and thus the channel) is *stationary*.

Observation: Obviously, as M grows, the finite-memory contagion channel converges in *distribution* to the Polya-contagion channel, i.e. $P_M(\cdot) \to P_{Polya}(\cdot)$.

We now consider the properties of the M'th order stationary Markov noise process. Define $\{W_n\}$ to be the process obtained by M-step blocking the process $\{Z_n\}$ – i.e. $W_n = (Z_n, Z_{n+1}, Z_{n+2}, \ldots, Z_{n+M-1})$. Then $\{W_n\}$ is a one-step Markov process with 2^M states; we denote each state by its decimal representation; i.e. state 0 corresponds to state $(0\cdots 00)$, state 1 corresponds to state $(0\cdots 01)$, ..., and state $(2^M - 1)$ corresponds to state $(1\cdots 11)$.

Tedious calculations [17] reveal the following properties about the process $\{W_n\}$.

- $\{W_n\}$ is a homogeneous stationary Markov process with stationary distribution $\Pi = [\pi_0, \pi_1, \ldots, \pi_{2^M-1}]$, where π_i is computed as follows. Let $w(i)$ denote the number of 1's in the binary representation of the decimal integer i. Then

$$\pi_i = \frac{\prod_{j=0}^{w(i)-1}(\rho + j\delta) \prod_{k=0}^{M-1-w(i)}(\sigma + k\delta)}{\prod_{\ell=1}^{M-1}(1+\ell\delta)}.$$

- If we let $\{p_{ij}\}$ be the one-step transition probabilities, then

$$p_{ij} = \begin{cases} \dfrac{\sigma + (M - w(i))\delta}{1 + M\delta}, & \text{if } j = 2i \ (modulo\ 2^M); \\ \dfrac{\rho + w(i)\delta}{1 + M\delta}, & \text{if } j = (2i+1) \ (modulo\ 2^M); \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Finally, from these transition probabilities, we clearly see that any state can reach any other state with positive probability in a finite number of steps. Therefore the Markov process W_n is *irreducible*. Furthermore it is *aperiodic*; thus it is *strongly mixing* (and hence ergodic) [19].

Since the additive noise process is stationary and mixing, the resulting additive noise channel is therefore *ergodic* [20, p. 205].

7.2 Capacity of the Finite-Memory Contagion Channel

Using the results in the previous section, we arrive at the following proposition.

Proposition 5 The capacity C_M of M-memory contagion channel is non-decreasing in M. It is given by:

$$C_M = 1 - \sum_{k=0}^{M} \binom{M}{k} L_k \ h_b\left(\frac{\rho + k\delta}{1 + M\delta}\right) \tag{12}$$

where

$$L_k = \frac{\prod_{j=0}^{k-1} (\rho + j\delta) \ \prod_{\ell=0}^{M-k-1} (\sigma + \ell\delta)}{\prod_{m=1}^{M-1} (1 + m\delta)}$$

and $h_b(\cdot)$ is the binary entropy function.

Proof 5 The capacity is given by

$$\begin{aligned} C_M &= 1 - H(Z_{M+1} \mid Z_M, Z_{M-1}, \ldots, Z_1) \\ &= 1 + \sum_{i,j=0}^{2^M-1} \pi_i \ p_{ij} \ \log_2 p_{ij} \\ &= 1 - \sum_{k=0}^{M} \binom{M}{k} L_k \ h_b\left(\frac{\rho + k\delta}{1 + M\delta}\right) \end{aligned}$$

The monotonicity of C_M in M follows from the fact that conditioning can only decrease entropy. ■

Recalling that, if we let M grow, our finite-memory contagion channel converges in distribution to the original Polya-contagion channel, we obtain the following result.

Proposition 6 The following equality holds:

$$\lim_{M\to\infty} C_M = 1 - \int_0^1 h_b(z)\ f_Z(z)dz \tag{13}$$

where $f_Z(z)$ is the beta(ρ/δ, σ/δ) pdf given in Proposition 3 and $h_b(\cdot)$ is the binary entropy function.

Proof 6 If we examine the quantity $\begin{pmatrix} M \\ k \end{pmatrix} L_k$ in the formula of C_M, we note that it is equal to the probability that $S_M = k$, where S_M is the state of the original Polya-contagion channel after the M'th draw, as defined in Section 2.1. We thus have:

$$\begin{aligned} C_M &= 1 - \sum_{k=0}^{M} h_b\left(\frac{\rho + k\delta}{1 + M\delta}\right) P(S_M = k) \\ &= 1 - \sum_{\tau\in\{k/M:k=0,1,\ldots,M\}} h_b\left(\frac{\frac{\rho}{M} + \tau\delta}{\frac{1}{M} + \delta}\right) P\left(\frac{S_M}{M} = \tau\right) \\ &= 1 - E_{T_M}\left[h_b\left(\frac{\frac{\rho}{M} + T_M\delta}{\frac{1}{M} + \delta}\right)\right] \end{aligned}$$

where $T_M = S_M/M$. We know by Property 3 in Section 2.2, that $T_M = S_M/M$ converges almost surely to a beta-distributed random variable Z with parameters ρ/δ and σ/δ. This almost surely convergence implies convergence in distribution; furthermore, since $h_b(\cdot)$ is a bounded and continuous function, the "weak equivalence" theorem [18] implies that

$$\begin{aligned} \lim_{M\to\infty} E_{T_M}\left[h_b\left(\frac{\frac{\rho}{M} + \frac{S_M}{M}\delta}{\frac{1}{M} + \delta}\right)\right] &= E_Z[h_b(Z)] \\ &= \int_0^1 h_b(z)\ f_Z(z)dz, \end{aligned}$$

which proves the proposition.

■

Observation: As the memory grows, the (ergodic) finite-memory contagion channel converges in distribution to the (non-ergodic) Polya-contagion channel, but the capacity C_M of the finite-memory channel *does not converge* to the capacity of the Polya-contagion channel (which is zero). On the contrary, C_M increases in M and converges to $1 - \int_0^1 h_b(z)\ f_Z(z)dz$. In addition, it can be shown [17] that, if we let $I(\underline{X};\underline{Y})$ denote the mutual information between the input vector $\underline{X}$ and output vector $\underline{Y}$ connected over the original (non-ergodic) Polya channel, then

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\underline{X}} I(\underline{X};\underline{Y}) = 1 - \int_0^1 h_b(z)\ f_Z(z)dz. \tag{14}$$

The left side of equation (14) is called the *information rate capacity* of the Polya channel; we have thus demonstrated that, as we let the memory in the finite-memory contagion channel increase, not only does the channel converge in distribution to the Polya channel, but the information rate capacities also converge to that of the Polya channel. However, there is no convergence in the weak capacity – the *operational* capacity. It seems reasonable to assume that this is due to the non-ergodic nature of the Polya channel. In the following proposition we examine this question.

Proposition 7 Consider a sequence of non-anticipatory stationary ergodic channels; let the n-fold transition probability of the M^{th} channel be denoted $W_M^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x})$. Let C_M denote the (weak) capacity of the M^{th} channel. Finally, suppose this sequence of channels satisfies the following conditions.

1. As M grows, they converge in distribution to a non-anticipatory stationary channel – i.e., if we let $W_*^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x})$ denote the n-fold transition probability of the limiting channel, then for any real n-tuples $\underline{x}$ and $\underline{y}$,

$$\lim_{M\to\infty} W_M^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}) = W_*^{(n)}(\underline{Y} = \underline{y} \mid \underline{X} = \underline{x}).$$

2. The "information rate capacities" of the channels converge to that of the limiting channel – i.e., if $I_M(\underline{X};\underline{Y})$ denotes the n-fold mutual information between the inputs and outputs of the M^{th} channel, and $I_*(\underline{X};\underline{Y})$ denotes the same for the limiting channel, then

$$\lim_{M\to\infty} \lim_{n\to\infty} \frac{1}{n} \sup_{\underline{X}} I_M(\underline{X};\underline{Y}) = \lim_{n\to\infty} \frac{1}{n} \sup_{\underline{X}} I_*(\underline{X};\underline{Y}).$$

Let C_* denote the (weak) capacity of the limiting channel; then a *sufficient but not necessary* condition that

$$\lim_{M\to\infty} C_M = C_*$$

is that the limiting channel be *ergodic.*

Proof 7 The proof that ergodicity is sufficient is trivial. All the channels are ergodic, so the information rate capacities are equal to the corresponding operational capacities; condition (2.) says the information rate capacities converge, and so the operational capacities must too.

To see that ergodicity is not necessary, we briefly sketch a counter-example. Let $\{U_i^M\}_{i=0}^{\infty}$ be a stationary, mixing binary Markov process indexed by the parameter M; assume that $P(U_0^M = 0) = P(U_0^M = 1) = 1/2$ and that U_i^M has one-step transition matrix

$$Q_{U^M} = \begin{pmatrix} 2^{-M} & 1-2^{-M} \\ 1-2^{-M} & 2^{-M} \end{pmatrix}.$$

We create a noise process $\{Z_i^M\}_{i=0}^{\infty}$ by two-blocking the process $\{U_i^M\}$ – i.e., $Z_i^M = (Z_{i1}^M, Z_{i2}^M) = (U_{2i}^M, U_{2i+1}^M)$ for $i = 0, 1, 2, \ldots$. Then $\{Z_i^M\}$ is a one-step Markov chain with four states and transition matrix

$$Q_{Z^M} = \begin{pmatrix} 2^{-2M} & 2^{-M}(1-2^{-M}) & (1-2^{-M})^2 & 2^{-M}(1-2^{-M}) \\ 2^{-M}(1-2^{-M}) & (1-2^{-M})^2 & 2^{-M}(1-2^{-M}) & 2^{-2M} \\ 2^{-2M} & 2^{-M}(1-2^{-M}) & (1-2^{-M})^2 & 2^{-M}(1-2^{-M}) \\ 2^{-M}(1-2^{-M}) & (1-2^{-M})^2 & 2^{-M}(1-2^{-M}) & 2^{-2M} \end{pmatrix}.$$

Now consider the channel with input/output alphabet $\{00, 01, 10, 11\}$, where the i^{th} input $X_i = (X_{i1}, X_{i2})$ is related to the i^{th} output $Y_i = (Y_{i1}, Y_{i2})$ by $Y_i = (X_{i1} \oplus Z_{i1}^M, X_{i2} \oplus Z_{i2}^M)$. $\{Z_i^M\}$ is a stationary mixing process; thus the channel is stationary ergodic [20]. For finite M, the capacity – both operational and information rate – is given by $C_M = 2 - H(Z_2^M|Z_1^M)$ bits/channel use. From Q_{Z^M} we observe that $\lim_{M\to\infty} H(Z_2^M|Z_1^M) = 0$; thus, $\lim_{M\to\infty} C_M = 2$ bits/channel use.

As M increases, the process $\{Z_i^M\}$ converges in distribution to a stationary non-ergodic process $\{Z_i^*\}$ with two equiprobable components – $\{01, 01, 01, 01, \ldots\}$ and $\{10, 10, 10, 10, \ldots\}$. The information rate capacity of this – that is, $\lim_{n\to\infty} \max_{\underline{X}} (1/n) I_*(\underline{X}; \underline{Y})$ – is two bits/channel use. Thus both of the conditions above are met. However, this limiting channel is a mixture of two deterministic channels, and its operational capacity is also two bits per channel use. Thus the ergodicity of the limiting channel is not a necessary condition.

■

8 Summary

In this paper we considered a discrete channel with memory in which errors spread like the spread of a contagious disease through a population. We analyzed a communication channel with additive noise modeled by Polya's model for the spread of contagion. The channel is stationary and non-ergodic. We first presented a maximum likelihood (ML) decoding algorithm for the channel, and then showed that this channel is in fact an "averaged" channel, and its capacity is zero. Using De Finetti's results on exchangeability, we noted that binary channels with additive exchangeable noise processes are averaged channels with binary symmetric channels as components. The zero capacity result illustrates a counter-example to the adage "memory can only increase capacity".

Finally, we considered a finite-memory version of the Polya-contagion model. The resulting channel is a stationary ergodic Markov channel with memory M; its capacity is positive and increases with M. As M increases, the finite-memory contagion channel converges in distribution to the original Polya-contagion channel, but its capacity *does not converge* to the capacity of the Polya channel.

9 References

[1] C. H. Stapper, A. N. McLaren and M. Dreckmann, "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product", *IBM J. Res. Develop.*, Vol. 24, No. 3, pp. 398-409, May 1980.

[2] G. Polya and F. Eggenberger, "Über die Statistik Verketteter Vorgänge", *Z. Angew. Math. Mech.*, Vol. 3, pp. 279-289, 1923.

[3] G. Polya and F. Eggenberger, "Sur l'Interpretation de Certaines Courbes de Fréquences", *Comptes Rendus C. R.* , Vol. 187, pp. 870-872, 1928.

[4] K. Jacobs, "Almost Periodic Channels", *Colloquium on Combinatorial Methods in Probability Theory, Aarhus*", pp. 118-126, 1962.

[5] R. Ahlswede, "The Weak Capacity of Averaged Channels", *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 11, pp. 61-73, 1968.

[6] G. Polya, "Sur Quelques Points de la Théorie des Probabilités", *Ann. Inst. H. Poincaré*, Vol. 1, pp. 117-161, 1931.

[7] R. Taylor, P. Daffer, R. Patterson, *Limit Theorems for Sums of Exchangeable Random*

Variables, Rowman & Allanheld Inc., 1985.

[8] W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley & Sons Inc., Second Edition, Vol. 2, 1971.

[9] B. C. Carlson, *Special Functions of Applied Mathematics*, Academic Press, 1977.

[10] J. C. Kieffer, "A General Formula for the Capacity of Stationary Nonanticipatory Channels", *Information and Control*, Vol. 26, pp. 381-391, 1974.

[11] R. Gray and L. D. Davisson, "The Ergodic Decomposition of Stationary Discrete Random Processes", *IEEE Transactions on Information Theory*, Vol. 20, No. 5, pp. 625-636, 1974.

[12] R. Gray and L. D. Davisson, *Ergodic and Information Theory*, Dowden, Hutchinson & Ross, Inc., 1977.

[13] Y. C. Tong, *Probability Inequalities in Multivariate Distributions*, Academic Press, 1980.

[14] R. L. Dobrushin and M. S. Pinsker, "Memory Increases Transmission Capacity", *Problemy Peredachi Informatsii*, Vol. 5, No. 1, pp. 94-95, 1969.

[15] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag New York Inc., 1988.

[16] R. Ahlswede, "Certain Results in Coding Theory for Compound Channels I," *Proceedings Bolyai Colloquium on Information Theory*, Debrecen, Hungary, pp. 35-60, 1967.

[17] F. Alajaji, "New Results on the Analysis of Discrete Communication Channels with Memory", Ph.D. Dissertation, Department of Electrical Engineering, University of Maryland, College Park, MD 20742, in preparation.

[18] P. Billingsley, *Probability and Measure*, Wiley, NY, 1979.

[19] K. Petersen, *Ergodic Theory*, Cambridge University Press, 1983.

[20] R. M. Gray, *Entropy and Information Theory*, Springer-Verlag New York Inc., 1990.