A Communication Channel Modeled on Contagion

Fady Alajaji and Tom Fuja

Abstract—**We introduce a binary additive communication channel with memory. The noise process of the channel is generated according to the contagion model of G. 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. Finally, we consider a finite-memory version of the Polya-contagion model; this channel is (unlike the original) ergodic with a nonzero capacity that increases with increasing memory.**

Index Terms—**Channels with memory, additive noise, capacity, maximum likelihood decoding.**

I. 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 G. 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 [4].

We begin by introducing a communication channel with additive noise modeled by 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 carried out by 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 [5], [6], i.e., its block transition probability is the average of those 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.

Manuscript received September 8, 1993; revised September 8, 1994. This work was supported in part by the National Science Foundation under Grant NCR-8957623 and by the National Science Foundation Engineering Research Centers Program under Grant CDR-8803012. This paper was presented in part at the 1993 IEEE International Symposium on Information Theory, San Antonio, TX, Jan. 17–22, 1993.

The authors are with the Department of Electrical Engineering, Institute for Systems Research, University of Maryland, College Park, MD 20742.

IEEE Log Number 9406314.

We show that the capacity of the Polya channel is zero, and we also obtain the ϵ-capacity of the channel. We note that the zero capacity result provides a counter-example to the adage "memory can only increase 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 grows, the n-fold conditional distribution of the finite-memory channel converges to the n-fold conditional distribution of the original Polya channel; however the capacity of the finite-memory channel *does not converge* to the capacity of the Polya channel.

II. Polya-Contagion Communication Channel

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

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 [7], 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

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

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.

A. Block Transition Probability of the Channel

Definition 1 (Channel State): We define the state of the channel after the nth transmission to be the total number of red balls drawn after n trials:

$$S_n \triangleq Z_1 + Z_2 + \cdots + Z_n = S_{n-1} + Z_n, \qquad S_0 = 0.$$

The possible values of S_n are the elements of the set $\{0, 1, \cdots, n\}$. Furthermore, the sequence of states $\{S_n\}_{n=1}^{\infty}$ forms a Markov chain, i.e.,

$$P(S_n = s_n | S_{n-1} = s_{n-1}, S_{n-2} = s_{n-2}, \cdots, S_1 = s_1) = P(S_n = s_n | S_{n-1} = s_{n-1}).$$

For a given input block $\underline{X} = [X_1, X_2, \cdots, X_n]$ and a given output block $\underline{Y} = [Y_1, Y_2, \cdots, Y_n]$, the block (or n-fold) transition probability of the channel is given by

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

where

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

0018-9448/94$04.00

We thus obtain

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

or

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

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 (2) from (1), we used the fact that $\Gamma(x+1) = x\Gamma(x)$, which leads to the following identity:

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

B. Properties of the Channel

We first define a discrete channel to be stationary if for every *stationary* input process $\{X_i\}_{i=1}^\infty$, the joint input-output process $\{(X_i, Y_i)\}_{i=1}^\infty$ is stationary. Furthermore, a discrete channel is ergodic if for every *ergodic* input process $\{X_i\}_{i=1}^\infty$, the joint input–output process $\{(X_i, Y_i)\}_{i=1}^\infty$ is ergodic [10], [19].

Before analyzing the characteristics of the channel, we state from [8] 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 indexes $1, 2, \cdots, n$.

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 and identically distributed (i.i.d) random variables. Note that i.i.d. 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 $P(\underline{Y} = \underline{y} | \underline{X} = \underline{x})$ depends only on $\underline{x} \oplus \underline{y}$ since $P(\underline{Y} = \underline{y} | \underline{X} = \underline{x}) = P(\underline{Z} = \underline{y} \oplus \underline{x})$. 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 (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

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

and the correlation coefficient

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

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

3) Nonergodicity: It is shown in [7], [9] that $Z \triangleq \lim_{n\to\infty} S_n/n$ exists almost surely, where 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.

III. 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 $\mathscr{C} = \{\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_M\}$, 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

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

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

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

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 are closed under addition and multiplication. 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 $P(\underline{Y} = \underline{y} | \underline{X} = \underline{x})$ 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, $P(\underline{Y} = \underline{y} | \underline{X} = \underline{x})$ is symmetric in $d(\underline{x}, \underline{y})$ about d_0.

Proof: 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 $d_i \triangleq d(\underline{y}, \underline{x}_i)$, for $i = 1, \cdots, M$. Compute also $d_{\max} \triangleq \max_{1 \le i \le M}\{d_i\}$ and $d_{\min} \triangleq \min_{1 \le i \le M}\{d_i\}$.

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 ⇔ 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 ⇔ maximum distance decoding.

Observations:

- Insight into the decoding rule: we can rewrite d_0 as $d_0 = n/2 + (1/\Delta)(T/2 - R)$. 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 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 farthest away from the equilibrium point d_0. In other words, the best decision is based on the following reasoning: either

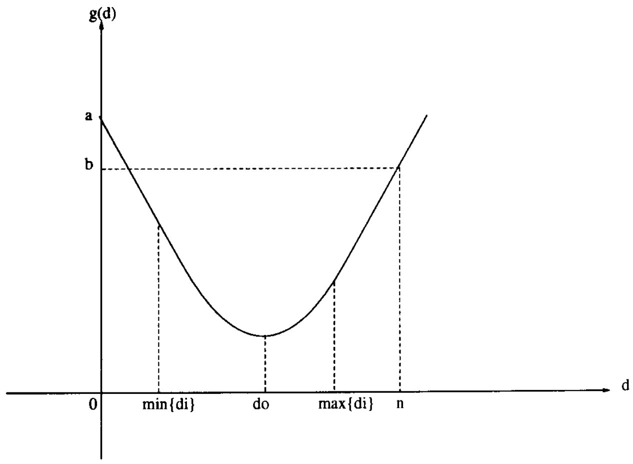

Fig. 1. Transition probability function versus Hamming distance, where $a = g(0)$ and $b = g(n)$.

many errors occurred during transmission—an *apparent epidemic*, to use the contagion interpretation—or very *few* errors occurred—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 $\delta < (1 - 2\rho)/(n - 1)$, 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.

IV. Averaged Communication Channels

Averaged channels with discrete memoryless components were first introduced by Jacobs [5] and then were analyzed by Ahlswede [6] and Kieffer [10]. 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 channels parameterized by θ:

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

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 n-fold transition probability of the channel specified by $\theta \in \Theta$.

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 channels parameterized by θ, i.e., if it is of the form

$$\begin{aligned} W_{\mathrm{avg}}^{(n)}(\underline{Y} = \underline{y} | \underline{X} = \underline{x}) &= \int_\Theta W_\theta^{(n)}(\underline{Y} = \underline{y} | \underline{X} = \underline{x})\, dG(\theta) \\ &= E_\theta\left[W_\theta^{(n)}(\underline{Y} = \underline{y} | \underline{X} = \underline{x})\right] \qquad (3) \end{aligned}$$

for some distribution $G(\cdot)$ on θ.

Note that if a channel is averaged with stationary components then it is stationary, and it may have memory. One way an averaged channel may be realized is as follows. From among the 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* additive channel—belongs to this class of channels. First we need to recall some results from [11], [12], [15].

Notation: Consider a discrete-time random process with alphabet D; let $\sigma(D^\infty)$ denote a σ-field consisting of subsets of D^∞, and let μ be a probability measure such that $(D^\infty, \sigma(D^\infty), \mu)$ forms a probability space. Finally, let $\mathbf{U}_n$: $D^\infty \to D$ denote a sampling function defined by $\mathbf{U}_n(u) = u_n$. Then the sequence of random variables $\{\mathbf{U}_n;\ n = 1, 2, \cdots\}$ is a discrete-time random process, to be denoted $[D, \mu, \mathbf{U}]$.

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 nonergodic random processes are mixtures of stationary ergodic processes; by directly applying the ergodic decomposition theorem we get the following result.

Proposition 2: Any discrete channel with stationary (nonergodic) additive noise is an averaged channel whose components are channels with additive stationary ergodic noise.

Proof: Let $\{Z_i\}$ be the (nonergodic) noise sequence. Then the ergodic decomposition theorem states that $P(\underline{Z} = \underline{z}) = P(Z_1 = z_1, \cdots, Z_n = Z_n)$ 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 channels with stationary ergodic additive noise. □

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: We showed in Proposition 2 that the Polya channel is an averaged channel 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 $(\rho/\delta, \sigma/\delta)$ random variable:

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

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 $P(\underline{Y} = \underline{y} | \underline{X} = \underline{x})$ describes the Polya-contagion channel as in (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 below [9], [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 $Z_1, Z_2, \cdots$ is an infinite sequence of exchangeable random variables, then there exists a σ-field $\mathscr{F}$ and a distribution G such that, given $\mathscr{F}$, the random variables $Z_1, Z_2, \cdots$ 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

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

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

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.

V. Capacity of the Polya Channel

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

Definition 5: An (M, n, ϵ) code has M codewords, each with blocklength n, and average error probability not larger than ϵ. $R \geq 0$ is an *ϵ-achievable rate* if for every $\gamma > 0$ there exists, for sufficiently large n, (M, n, ϵ) codes with rate $(1/n)\log_2(M) > R - \gamma$. The maximum ϵ-achievable rate is called the *ϵ-capacity* C_ϵ. The *channel capacity* C is the maximum rate that is ϵ-achievable for all $0 < \epsilon < 1$. It follows immediately from the definition that $C = \lim_{\epsilon \to 0} C_\epsilon$.

In [21], Verdú and Han derived a formula for the capacity of arbitrary single-users channels (not necessarily stationary, ergodic, information stable, etc.).

Lemma 3: The channel capacity C is given by

$$C = \sup_{\underline{X}} \underline{I}(\underline{X}; \underline{Y}) \tag{4}$$

where the symbol $\underline{I}(\underline{X}; \underline{Y})$ is the *inf-information rate* between $\underline{X}$ and $\underline{Y}$ and is defined as the *liminf in probability*[1] the sequence of normalized information densities $(1/n) i_{\underline{X};\underline{Y}}(\underline{X}; \underline{Y})$, where

$$i_{\underline{X};\underline{Y}}(\underline{a}; \underline{b}) = \log_2 \frac{P^{(n)}_{\underline{Y}|\underline{X}}(\underline{b}|\underline{a})}{P^{(n)}_{\underline{Y}}(\underline{b})}. \tag{5}$$

Using the above lemma as well as the properties of the inf-information rate derived in [21], we obtain that the inf-information rate in (4) is maximized when the input process is equally likely Bernoulli (symmetry property), yielding the following expression for the capacity of the Polya channel:

$$C_{\text{Polya}} = 1 - \overline{H}(\underline{Z})$$

where $\overline{H}(\underline{Z})$ is the sup-entropy rate of the additive Polya noise process $\{Z_n\}$, defined as the limsup in probability of $(1/n)\log_2(1/P^{(n)}_{\underline{Z}}(\underline{Z}))$. Since the noise process is stationary, we obtain that the sup-entropy rate is equal to the supremum over the entropies of almost every ergodic component of the noise process [10], [21]:

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

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

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

and the essential supremum is defined by

$$\operatorname{ess}_\Theta \sup f(\theta) \triangleq \inf[r : dG(f(\theta) \leq r) = 1].$$

We know that the stationary ergodic components of the Polya channel are BSC's with crossover probability θ; 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 capacity of the channel:

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

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

ϵ-Capacity of the Polya Channel: Since the stationary Polya noise process $\{Z_n\}$ is a mixture of Bernoulli(θ) processes where the parameter θ is beta-distributed $(\rho/\delta, \sigma/\delta)$, it can be shown using the ergodic decomposition theorem [12], [19] that $(1/n)\log_2(1/P^{(n)}_{\underline{Z}}(\underline{Z}))$ converges in L^1 and almost surely (hence in distribution) to the random variable $V \triangleq h_b(U)$, where U is beta-distributed $(\rho/\delta, \sigma/\delta)$. The cumulative distribution function (cdf) of V is given by

$$F_V(a) \triangleq P(V \leq a) = F_U[h_b^{-1}(a)] + 1 - F_U[1 - h_b^{-1}(a)] \tag{7}$$

where $h_b^{-1}(a) \in [0, 1/2]$ is the smallest root of the equation $a = h_b(u)$, and $F_U(\cdot)$ is the cdf of U. Note that since U is beta distributed, $F_V(\cdot)$ is strictly increasing in the interval [0, 1]; it therefore admits an inverse $F_V^{-1}(\cdot)$. Now, applying the formula for ϵ-capacity in [21, Theorem 6] we obtain

$$C_\epsilon = 1 - F_V^{-1}(1 - \epsilon). \tag{8}$$

Note that $\lim_{\epsilon \to 0} C_\epsilon = 1 - F_V^{-1}(1) = 1 - 1 = 0$, as expected.

Observations:

- 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

[1] If A_n is a sequence of random variables, then its *liminf in probability* is the supremum of all reals α for which $P(A_n \leq \alpha) \to 0$ as $n \to \infty$. Similarly, its *limsup in probability* is the infimum of all reals β for which $P(A_n \geq \beta) \to 0$ as $n \to \infty$ [21].

channel with BSC's as components and the capacity of such a compound channel is equal to the infimum of the capacities of the BSC's.

- In [14], Pinsker and Dobrushin 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. By "a wide class of channels," they meant channels that are causal, historyless, and *information stable*. Information stable channels have the property that the input that maximizes mutual information and its corresponding output behave ergodically [20], [21]. In [16], Ahlswede showed that there are averaged channels for which the introduction of memory *decreases* capacity. The Polya-contagion channel is such a channel.
- The zero capacity result suggests that the Polya channel might not be a good model for a realistic channel. However, in Section VI we will consider a finite-memory channel that approximates the Polya channel as memory increases, but with a capacity that does *not* approach zero.

VI. 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 999th 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.

Assume once again that the channel output Y_i is the modulo-two sum of the input X_i and the noise Z_i; as for the Polya channel, assume that the input and noise sequences are independent. Then $\{Z_i\}_{i=1}^{\infty}$ is drawn according to the following urn scheme: an urn initially contains T balls—R red and S black ($T = R + S$). At the jth draw, $j = 1, 2, \cdots$, 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. Once again let $\rho = R/T < 1/2$, $\sigma = 1 - \rho = S/T$ and $\delta = \Delta/T$. Then the noise process $\{Z_i\}$ corresponds to the outcomes of the draws from the urn, where

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

Observation: With this modification of the original Polya urn scheme, the 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.

A. The Distribution of the Noise

During the initialization period ($n \le M$), 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 $n \ge M + 1$.

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 $n \ge M + 1$, and so

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

We now have that

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

where $e_i \in \{0, 1\}$. Thus the noise process $\{Z_i\}_{i=M+1}^{\infty}$ is a Markov process of order M. We shall refer to the resulting channel as the finite-memory contagion channel.

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

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

$$\begin{aligned} P_M(\underline{Y} = \underline{y} | \underline{X} = \underline{x}) &= P(\underline{Z} = \underline{e}) \\ &= \prod_{i=1}^{n} P(Z_i = e_i | Z_{i-1} = e_{i-1}, \cdots, Z_{i-M} = e_{i-M}) \\ &= L \prod_{i=M+1}^{n} \left[\frac{\rho + s_{i-1}\delta}{1 + M\delta}\right]^{e_i} \left[\frac{\sigma + (M - s_{i-1})\delta}{1 + M\delta}\right]^{1-e_i} \end{aligned} \tag{9}$$

where

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

$e_i = x_i \oplus y_i$, $k = e_1 + \cdots + e_M$ and $s_{i-1} = e_{i-1} + \cdots + e_{i-M}$.

We thus see that the noise process (and so the channel) is stationary.

We now consider the properties of the noise process $\{Z_i\}$. Define $\{W_n\}$ to be the process obtained by M-step blocking $\{Z_n\}$, i.e., $W_n = (Z_n, Z_{n+1}, Z_{n+2}, \cdots, Z_{n+M-1})$. 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 $(0 \cdots 01), \cdots$, 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 $\Pi = [\pi_0, \pi_1, \cdots, \pi_{2^M-1}]$, 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

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

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

$$p_{ij} = \begin{cases} \dfrac{\sigma + [M - w(i)]\delta}{1 + M\delta}, & \text{if } j = 2i \text{ (modulo } 2^M) \\ \dfrac{\rho + w(i)\delta}{1 + M\delta}, & \text{if } j = (2i + 1) \text{ (modulo } 2^M) \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

We have thus shown that the Markov process $\{W_n\}$ is irreducible and aperiodic; therefore it is *strongly mixing* and hence ergodic. Since the additive noise process is stationary and mixing, the resulting additive noise channel is *ergodic* [19], [20].

Observation: For $M = 1$ the one-step transition probability matrix of $\{Z_n\}$ is

$$\Pi = \frac{1}{1+\delta}\begin{bmatrix} \sigma+\delta & 1-\sigma \\ \sigma & 1-\sigma+\delta \end{bmatrix}. \tag{11}$$

Clearly one can choose δ and σ to "match" the transition probabilities of an arbitrary irreducible two-state Markov chain.

B. 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 nondecreasing 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

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

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

Proof: The capacity is given by

$$\begin{aligned} C_M &= 1 - H(Z_{M+1}|Z_M, Z_{M-1}, \cdots, Z_1) \\ &= 1 + \sum_{i,j=0}^{2^M-1} \pi_i p_{ij} \log_2 p_{ij} \\ &= 1 - \sum_{k=0}^{M}\binom{M}{k} L_k h_b\left(\frac{\rho + k\delta}{1 + M\delta}\right). \end{aligned} \tag{13}$$

The monotonicity of C_M in M follows from (13) because the Markov noise process is stationary and conditioning can only decrease entropy. □

Proposition 6: The following equality holds:

$$\lim_{M\to\infty} C_M = 1 - \int_0^1 h_b(z) f_Z(z)\, dz \tag{14}$$

where $f_Z(z)$ is the beta$(\rho/\delta, \sigma/\delta)$ pdf and $h_b(\cdot)$ is the binary entropy function.

Proof: If we examine the quantity $\binom{M}{k} L_k$ 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 Mth draw, as defined in Section II-A. Thus we have

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

where $T_M = S_M/M$. We know by the nonergodicity property 3) in Section II-B, that $T_M = S_M/M$ converges almost surely to a beta-distributed random variable Z with parameters ρ/δ and σ/δ. Furthermore, since $h_b(\cdot)$ is bounded and continuous, the "weak equivalence" theorem [18] implies that

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

Observation: As the memory M grows, $P_M^{(n)}(\cdot) \to P_{\text{Polya}}^{(n)}(\cdot)$, 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 (nonergodic) Polya channel, then

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\underline{X}} I(\underline{X}; \underline{Y}) = 1 - \int_0^1 h_b(z) f_Z(z)\, dz. \tag{15}$$

VII. Summary

In this correspondence we considered a discrete channel with memory in which errors "spread" like the spread of a contagious disease through a population; the channel is based on Polya's model for contagion. The channel is stationary and nonergodic. 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 n-fold transition distribution of the finite-memory contagion channel converges to the n-fold transition distribution of the original Polya-contagion channel, but its capacity does not converge to the capacity of the Polya channel.

References

[1] C. H. Stapper, A. N. McLarent, 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] ——, "Sur l'Interprétation de Certaines Courbes de Fréquences," *Comptes Rendus C. R.*, Vol. 187, pp. 870–872, 1928.

[4] L. Kanal and A. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, pp. 724–744, July 1978.

[5] K. Jacobs, "Almost periodic channel," *Colloq. Combinator. Methods Probability Theory*, Aarhus, Denmark, 1962, pp. 118–126.

[6] R. Ahlswede, "The weak capacity of averaged channels," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, vol. 11, pp. 61–73, 1968.

[7] G. Polya, "Sur Quelques Points de la Théorie des Probabilités," *Ann. Inst. H. Poincarré*, vol. 1, pp. 117–161, 1931.

[8] R. Taylor, P. Daffer, and R. Patterson, *Limit Theorems for Sums of Exchangeable Random Variables.* Rowman & Allanheld, Totowa, NJ, 1985.

[9] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: Wiley, vol. 2, 1971.

[10] J. C. Kieffer, "A general formula for the capacity of stationary nonanticipatory channels," *Inform. Control*, vol. 26, pp. 381–391, 1974.

[11] R. Gray and L. D. Davisson, "The ergodic decomposition of sationary discrete random processes," *IEEE Trans. Inform. Theory*, vol. 20, pp. 625–636, Sept. 1974.

[12] ——, *Ergodic and Information Theory.* Dowden, Hutchinson & Ross, Inc., Stroudsburg, PA, 1977.

[13] Y. C. Tong, *Probability Inequalities in Multivariate Distributions*. New York: Academic, 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*. New York: SpringeI-Verlag, 1988.
[16] R. Ahlswede, "Certain results in coding theory for compound channels I," in *Proc. Bolyai Colloq. Inform. Theory*, Debrecen, Hungary, 1967, pp. 35–60.
[17] F. Alajaji, "New results on the analysis of discrete communication channels with memory," Ph.D. dissertation, Dep. Elect. Eng., Univ. Maryland, August 1994.
[18] P. Billingsley, *Probability and Measure*. New York: Wiley, 1979.
[19] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
[20] M. S. Pinsker, *Information and Information Stability of Random Variables and Processess*. San Francisco: Holden Day, 1964.
[21] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

Coding Theorems for a Continuous-Time Gaussian Channel with Feedback

Shunsuke Ihara

Abstract—**The main aim of the present paper is to prove a coding theorem for a continuous-time Gaussian channel with feedback, under an average power constraint. In the case of discrete-time, the coding theorem for the feedback Gaussian channel has been shown by Cover and Pombra.**

Index Terms—**Channel coding theorem, Feedback capacity, Gaussian channel with feedback.**

I. Introduction

The main aim of the present paper is to prove a coding theorem for a continuous-time Gaussian channel (GC) with feedback. The model of the GC is given by

$$Y(t) = X(t) + Z(t), \qquad 0 \le t \le T \tag{1}$$

where the noise $Z = \{Z(t); 0 \le t \le T\}$ is a zero-mean Gaussian process, $X = \{X(t)\}$ is a channel input and $Y = \{Y(t)\}$ is the corresponding output. We assume that an average power constraint (in a generalized sense) is imposed on the channel inputs. The capacity C_f of the GC with feedback is defined as the supremum of mutual information one can transmit over the channel.

The channel coding theorem to be proved is stated as follows: i) an information transmission rate R less than C_f is achievable; and conversely, ii) a rate R greater than C_f is not achievable. The capacity C_f is sometimes called the *information capacity*. On the other hand, the supremum of achievable rates is said to be the *coding capacity*. Thus the coding theorem means that the coding capacity is equal to the information capacity.

The coding theorem has been studied for various communication channels, mostly for channels without feedback. Recently, Cover and Pombra [3] proved the coding theorem for a discrete-time GC with feedback. For the continuous-time GC with feedback, however, no proof of the coding theorem in full generality has been known. Indeed, the information capacity and the coding capacity have been studied rather separately. Some previous work [1], [6], [8], [9], [12] examined the information capacity, while the coding capacity was studied in [2], [15].

The statement of our coding theorem will be given in Section II (Theorem 1). It is known that the capacity of the GC is attained by sending a Gaussian message and using linear feedback [8] (see Theorem 2). This fact enables us to prove the coding theorem for the continuous-time GC with feedback.

Manuscript received April 2, 1991; revised September 8, 1994.
The author is with the School of Informatics and Sciences, Nagoya University, Nagoya 464-01, Japan.
IEEE Log Number 9406220.

II. Coding Theorem

We consider the GC (1) with feedback and assume that the terminal time T is finite. A message θ, to be transmitted, is a random variable taking values in an arbitrary measurable space $(\Omega_\theta, \mathscr{B}_\theta)$. We assume that the feedback channel is noiseless and without time delay, so that the channel input X is a nonanticipative process with respect to Y, i.e., X is of the form

$$X(t) = x(t, \theta, Y_0^t) \tag{2}$$

where Y_0^t stands for the path $Y(s)$, $0 \le s < t$, and $x(t, \theta, y_0^t)$ is a measurable functional of $t \in [0, T]$, $\theta \in \Omega_\theta$, and $y = \{y(s); 0 \le s \le T\}$. More precisely, we assume the following conditions (A.1)–(A.3) (cf. [12], [14]):

(A.1) the message θ is independent of the noise Z;
(A.2) for almost all $\theta \in \Omega_\theta$, $X(t) = x(t, \theta, Y_0^t)$ is $\sigma_t(Y)$-measurable at any moment t, where $\sigma_t(Y)$ is the σ-field generated by $\{Y(s);\ s < t\}$;
(A.3) the stochastic equation (1) has a unique strong solution $Y = \{Y(t)\}$.

Let us impose on the channel inputs a constraint that is given in terms of the covariance function. Let $\mathscr{P}$ be a class of covariance functions on the time interval $[0, T]$. A stochastic process X with covariance function $\Gamma_X(\cdot, \cdot)$ is possible to be input if

$$\Gamma_X(\cdot, \cdot) \in \mathscr{P}. \tag{3}$$

Define a class $\mathscr{A}(\mathscr{P})$ of all admissible pairs (θ, X) by

$$\mathscr{A}(\mathscr{P}) = \{(\theta, X); (\theta, X) \text{ satisfies (A.1)–(A.3) and (3)}\}.$$

Then, under the constraint (3), the capacity $C_f = C_f(\mathscr{P})$ of the GC (1) with feedback is defined by

$$C_f(\mathscr{P}) = \sup\{I(\theta, Y_0^T); (\theta, X) \in \mathscr{A}(\mathscr{P})\} \tag{4}$$

where $I(\theta, Y_0^T)$ denotes the mutual information between the message θ and the output Y_0^T. Note that the mutual information can be written in the form

$$I(\theta, Y) = \int \log \varphi(x, y)\, d\mu_{\theta Y}(x, y)$$

where

$$\varphi(x, y) = \frac{d\mu_{\theta Y}}{d\mu_\theta \times \mu_Y}(x, y) \tag{5}$$

is the Radon–Nikodym derivative of $\mu_{\theta Y}$ with respect to $\mu_\theta \times \mu_Y$. Here μ_θ, μ_Y, and $\mu_{\theta Y}$ denote the probability distributions of θ, Y, and (θ, Y), respectively. The capacity $C_0 = C_0(\mathscr{P})$ of the same GC without feedback is defined in the same way. We

