Capacity of the Independent Increment Noise Channel

Michael R. Frey

Department of Statistics
University of North Carolina at Chapel Hill

LISS 44

30 October 1990

Abstract - A communications channel model with additive, independent increment, stochastically continuous noise is presented. The information capacity of this channel is obtained for a mean-square constraint imposed on the "generalized" intensity of the channel encoder. The capacity is derived by exploiting the Lévy decomposition of the noise to decompose the channel into a white Gaussian noise channel in parallel with a marked Poisson channel. The information capacity is then obtained using results for these latter channels. In particular, recent results for the Poisson channel with mean-square-constrained encoder intensity are exploited. Although only the channel without feedback is explicitly addressed, our results are also seen to be applicable in the presence of delayed causal feedback.

Supported by ARO Contract DAAL03-86-G-0022, ONR Contract N00014-86-K-0039 and NSF Contract NCR8713726.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) University of North Carolina Chapel Hill, North Carolina 27514	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0039
8c. ADDRESS (City, State and ZIP Code) Statistics & Probability Program Arlington, VA 22217	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.

11. TITLE (Include Security Classification)
Capacity of the Independent Increment Noise Channel

12. PERSONAL AUTHOR(S)
M.R. Frey

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM ______ TO ______	1990 October	26

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Independent-increment channels, Shannon theory, channel capacity

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A communications channel model with additive, independent increment, stochastically continuous noise is presented. The information capacity of this channel is obtained for a mean-square constraint imposed on the "generalized" intensity of the channel encoder. The capacity is derived by exploiting the Levy decomposition of the noise to decompose the channel into a white Gaussian noise channel in parallel with a marked Poisson channel. The information capacity is then obtained using results for these latter channels. In particular, recent results for the Poisson channel with mean-square-constrained encoder intensity are exploited. Although only the channel without feedback is explicitly addressed, our results are also seen to be applicable in the presence of delayed causal feedback.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION	
22a. NAME OF RESPONSIBLE INDIVIDUAL C.R. Baker	22b. TELEPHONE NUMBER (Include Area Code) (919) 962-2189	22c. OFFICE SYMBOL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

I. INTRODUCTION

The independent increment noise channel is an additive noise communications channel model incorporating causal feedback in which the noise is a process with independent increments. A diagram of the independent increment noise channel is given in Figure 1. In this model, a random message de-

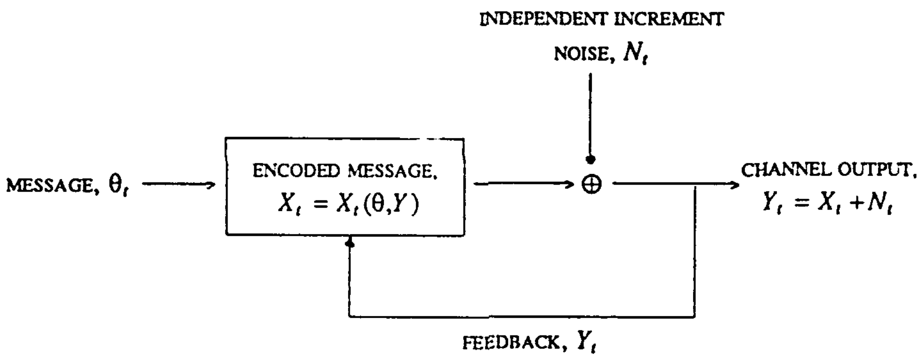

Figure 1. Independent increment noise channel model.

scribed by a stochastic process θ_t, $t \in [0,T]$ is encoded, perhaps using channel feedback, and the encoded message $X_t = X_t(\theta, Y)$ is transmitted. The encoded message is corrupted in transmission by the addition of an independent increment noise process N_t, $t \in [0,T]$ so that the signal observed by the receiver is $Y_t = X_t + N_t$. The channel model is presented in further detail in the next section.

Within the taxonomy of continuous-time channel models, the independent increment noise channel occupies a unique position. First, the independent increment noise channel is a natural point of departure for the study of more general continuous-time channels such as those with conditionally independent increment noise, Markov noise, or martingale noise. Independent increment noise is a special case of all three of these classes of noise processes. Second, the independent increment noise channel is the continuous-time analogue of the discrete memoryless channel [9]. Thus one finds properties of the discrete memoryless channel reflected in the independent increment noise channel. For instance, the capacity of the discrete memoryless channel is not increased by the availability of feedback [2], [17]. The same result holds under certain conditions for the independent increment noise channel [11]. Finally, both the white Gaussian noise channel and the Poisson channel with nonrandom noise intensity belong to the class of independent increment noise channels. Each of these two channel models is relevant to modern communication systems: for example, the Gaussian model for satellite systems and the Poisson model for laser/optical communications.

In this initial treatment of the independent increment noise channel, we derive its information capacity. We consider only independent increment noise channels with Lévy process noise; that is, stochastically continuous, independent increment noise. The Lévy process can be decomposed into a continuous-path Gaussian part and a purely discontinuous-path marked Poisson part [6]. The existence of this decomposition for Lévy noise processes allows us to split the independent increment noise channel into Gaussian and marked Poisson parts. Then the information capacity can be calculated using results for the capacity of the Gaussian and marked Poisson channels. Heretofore, the information capacity of the marked Poisson channel in any guise was unknown (although see Davis' treatment [4] of orthogonally polarized optical channels.) Therefore, to obtain the information capacity of the independent increment noise channel, some attention must be given to the marked Poisson channel.

A difficult part of formulating the capacity problem for the independent increment noise channel is the imposition of a suitable constraint on the channel encoding process. A suitable constraint should appeal to physical intuition, it should be capable of separation into two parts corresponding to the Lévy decomposition of the channel noise, and its two parts should admit finite solutions for the capacity of the corresponding Gaussian and marked Poisson subchannels. A mean-square constraint on a "generalized" channel intensity meets all these demands. While the capacity of the white Gaussian noise channel with mean-square constraint is well-known, such is not the case for the Poisson and marked Poisson channels. Previous work on the Poisson channel is limited to peak and average constraints [4], [8]. Thus we present here certain results for the capacities of the mean-square-constrained Poisson and marked Poisson channels. A fuller treatment of the mean-square-constrained Poisson and marked Poisson channels is planned for a separate paper.

The remainder of this paper is organized into three sections. The first section describes the channel model in detail. The second section states the capacity of the channel and outlines the steps in its derivation. Proofs of results used in this derivation are presented in the last section.

II. CHANNEL MODEL

The noise in the independent increment noise channel is modeled by a Lévy process. The properties of this process are given in [6], [7]. For convenience certain of these properties are restated here.

The Lévy Process

A stochastic process, ξ_t, $0 \le t \le T$, is an independent increment process if, for all $0 \le t_0 < t_1 < \ldots < t_n \le T$, the quantities $\xi_{t_0}, \xi_{t_1} - \xi_{t_0}, \ldots, \xi_{t_n} - \xi_{t_{n-1}}$ are mutually independent. We say a process ξ is stochastically continuous at time t_0 if

$$P\{|\xi_t - \xi_{t_0}| > \varepsilon\} \to 0$$

as $|t - t_0| \to 0$. If ξ is stochastically continuous over the whole interval $[0,T]$ (with the definition of stochastic continuity at the endpoints of the interval suitably modified), then ξ is said to be stochastically continuous. Stochastically continuous processes with independent increments have no discontinuities of the second kind w.p. 1 [6, p.168]. Every such process has a unique version which is càdlàg (right-continuous with extant left limits). This version of a stochastically continuous, independent increment process is called a Lévy process [7]. Every Lévy process can be represented in terms of independent Gaussian and Poisson processes as follows [6, Theorem 1, p.271]:

Let ξ_t, $0 \le t \le T$, be a Lévy process. Let J be the space of jumps which occur at the discontinuities in the paths of ξ. Zero is expressly excluded from belonging to J; jumps of size zero are disallowed. For $A \in \Sigma_J$, define $\psi_t(A)$ to be the number of points $s \in [0,t]$ for which $\xi_t - \xi_{t-0} \in A - \{0\}$. Also define $\bar{\psi}_t(A) = \psi_t(A) - \Pi_t(A)$ where $\Pi_t(A) = E[\psi_t(A)]$. Then

$$\xi_t = G_t + \int_{|x|>1} x\,\psi_t(dx) + \int_{|x|\le 1} x\,\bar{\psi}_t(dx) \qquad (1)$$

where G_t is an independent increment Gaussian process, $\psi_t(A)$ is an independent increment Poisson process for each $A \in \Sigma_J$, $\psi_t(A)$ and $\psi_t(B)$ are independent for $A \cap B = \emptyset$, and G_t is independent of $\psi_t(A)$ for every $A \in \Sigma_J$.

It is primarily for convenience that the Lévy process is emphasized in this presentation over the more general independent increment process which is not necessarily stochastically continuous. The general independent increment process has a unique decomposition [7] in terms of a deterministic part, a stochastically continuous (Lévy) part, and a stochastically discontinuous part. Nonrandom translations of the noise have no effect on channel information or capacity so we can assume the deterministic part is absent. The discontinuous part represents the process at the points at which it is not stochastically continuous. These points are no more than countable in number [7] and, since the location of these points are part of the specification of the process, can be avoided by channel users. Thus, we focus on the stochastically continuous part of the independent increment noise process and define the independent increment noise channel model with Lévy noise.

The Independent Increment Noise Channel Model

The independent increment noise channel is an additive noise model $Y_t = X_t + N_t$, $0 \le t \le T$ incorporating causal feedback (Figure 1). The process Y is the channel output, X is the encoded message, and N is the channel noise. $\mathbb{F}^{\theta}$ and $\mathbb{F}^{Y}$ denote the natural histories of the processes θ and Y, i.e., $\mathbb{F}_t^{\theta} = \sigma[\theta_s, 0 \le s \le t]$ and $\mathbb{F}_t^{Y} = \sigma[Y_s, 0 \le s \le t]$ for $t \in [0,T]$. We fix $\tau > 0$ and define $\mathbb{F}_t^{Y,\tau} = \sigma[Y_{s-\tau}, 0 \le s \le t]$, taking $Y_t \equiv 0$ for all $t < 0$. $\mathbb{F}^{Y,\tau}$ is the history of Y delayed by τ. The encoded message process X in the channel model is specified to be an $\mathbb{F}^{\theta} \vee \mathbb{F}^{Y,\tau}$-adapted functional of Y and the message process θ. Thus the channel model includes noiseless delayed causal feedback from the channel output Y and nonanticipative encoding of the message θ. We do not consider the case $\tau = 0$ of instantaneous feedback. This case seems impractical and raises presently unresolved technical questions. In the no-feedback version of the channel model, the encoded message is an $\mathbb{F}^{\theta}$-adapted functional $X_t = X_t(\theta)$ of only the message θ. The case $\tau > T$ is equivalent to the no-feedback case. Also as part of the channel model, X is required to be càdlàg with $X_0 \equiv 0$. The channel noise N is a Lévy process and is independent of θ. Following (1), we write

$$N_t = W_t + P_t \tag{2}$$

with $W_0 = P_0 = 0$. In (2), W_t is a zero-mean Wiener process [6], [12] with finite continuous variance measure β. By this it is meant that W_t has continuous paths and Gaussian independent increments $W_{t_2} - W_{t_1}$ such that, for all $0 \le t_1 < t_2 \le T$,

$$E[W_{t_2}-W_{t_1}]=0\,,\quad E[(W_{t_2}-W_{t_1})^2]=\beta([t_1,t_2])$$

where β is a finite continuous (w.r.t. Lebesgue measure) measure defined on $([0,T],\Sigma_{[0,T]})$. Also in (2), P_t is a purely discontinuous [15] (piecewise-constant) càdlàg process

$$P_t=\int_J m\,P_t(dm) \tag{3}$$

where, for each $F\in\Sigma_J$, the quantity $P_t(F)$ denotes the number of points $s\in[0,t]$ such that $P_s-P_{s-0}\in F$. $P_t(\cdot)$ is a marked Poisson point process with mark space $J\subset\mathbb{R}$. To distinguish the notations P_t and $P_t(\cdot)$, we remark that P_t is a jump process while, for instance, $P_t(J)$ is a counting process which gives the number of jumps (irrespective of size) of P_t in the interval $[0,t]$. It follows from the Lévy nature of N that W and P are independent processes. Also, for $F_1\cap F_2=\varnothing$, the counting processes $P(F_1)$ and $P(F_2)$ are independent. Because $P_t(F)$ is Poisson [6] for a given $F\in\Sigma_J$, it has a deterministic compensating measure B which here we assume takes the form

$$B([0,t]\times F)=\lambda b([0,t])\mu(F) \tag{4}$$

for some $\lambda\geq 0$ and all $F\in\Sigma_J$ and $t\in[0,T]$. The set functions b and μ in (4) are required to be finite measures on $([0,T],\Sigma_{[0,T]})$ and (J,Σ_J) respectively. The measure b is continuous w.r.t. Lebesgue measure. (Both b and β are continuous as a consequence of the stochastic continuity of N.) Finally, both μ and b are assumed to be standardized:

$$\mu(J)=1\,,\quad b([0,T])=T\,.$$

Looking back to (1), it is evident that deterministic terms have been arbitrarily added and subtracted to produce the decomposition of N in (2). We maintain that this does no harm because the noise decomposition in (2) is used only for calculating channel information and capacity - and those quantities are unaffected by nonrandom translations of the noise.

The paths of the Lévy noise process N in the channel model are càdlàg. Thus, without the comparable restriction (càdlàg paths) on the encoded message process X, the encoder could communicate without error at any desired rate by transmitting non-càdlàg paths. Hence, for finite information capacity, X must be càdlàg. Given that X is càdlàg, its paths can only have discontinuities of the first kind [1] and these are countable in number. Thus it is meaningful to define

$$X_{d,t}=\sum_{s\leq t}(X_s-X_{s-0})\,,\quad X_{c,t}=X_t-X_{d,t}\,.$$

Both X and N are càdlàg so $Y = X + N$ must be càdlàg as well and we define

$$Y_{d,t} = \sum_{s \le t} (Y_s - Y_{s-0}), \quad Y_{c,t} = Y_t - Y_{d,t}.$$

Analogous to (3), we can write X_d as

$$X_{d,t} = \int_J m \, X_{d,t}(dm)$$

where the marked Poisson-type point process $X_{d,t}(F)$ counts the number of jumps of X_d with size in $F \in \Sigma_J$ in the interval $[0,t]$. Likewise, we have

$$Y_{d,t} = \int_J m \, Y_{d,t}(dm).$$

Conversely, the marked point processes $X_d(\cdot)$ and $Y_d(\cdot)$ can be expressed in terms of their jump process counterparts as follows:

$$\begin{aligned} X_{d,t}(F) &= \sum_{s \le t} 1_F(X_{d,s} - X_{d,s-0}) \\ Y_{d,t}(F) &= \sum_{s \le t} 1_F(Y_{d,s} - Y_{d,s-0}) \end{aligned}, \quad F \in \Sigma_J.$$

Using the notation

$$\theta = \begin{bmatrix} \theta_c \\ \theta_d \end{bmatrix}$$

for the message process θ in the independent increment noise channel model, we have

$$Y_{c,t} = X_{c,t}(\theta_c, Y) + W_t, \tag{5}$$

$$Y_{d,t} = X_{d,t}(\theta_d, Y) + P_t. \tag{6}$$

The jump process channel expressed by (6) can be just as well viewed as a marked Poisson channel; that is, as a Poisson channel in the sense of [4], [8] in which marks (the jump size) are associated with each jump event in the channel. In terms of the marked point processes $X_{d,t}(\cdot)$, $P_t(\cdot)$, and $Y_{d,t}(\cdot)$, we write

$$Y_{d,t}(F) = X_{d,t}(\theta_d, Y)(F) + P_t(F), \quad F \in \Sigma_J. \tag{7}$$

The two formulations (6) and (7) are equivalent; however, (7) has the advantage of being applicable in other contexts so it is the formulation we use.

The continuous-path part X_c of the encoded message is assumed to take the form

$$X_{c,t} = \int_0^t \phi_s b(ds) \tag{8}$$

where $\phi_t = \phi_t(\theta_c, Y)$ is an $\mathbb{F}^{\theta_c} \vee \mathbb{F}^{Y,\tau}$-predictable functional. Also, the measurability, integrability, and identifiability conditions laid out in [10] for (5) and (8) are assumed to be satisfied. If the variance measure β of W in (2) is absolutely continuous w.r.t. b and if a mean-square constraint

$$\frac{1}{T} E\left[\int_0^T \phi_t^2 b(dt)\right] \le P^2 \tag{9}$$

is imposed on the encoding kernel ϕ, then (5) and (8) describe a form of the additive white Gaussian noise channel treated by Kadota, Zakai, and Ziv [10]. Therefore, to use their results for information capacity, we assume (9) to be in force and take $\beta \ll b$. In fact, it is assumed that β and b are scaled versions of one another $\beta = \eta b$. The scalar $\eta \ge 0$ is termed the Wiener noise intensity of N_t. η is the counterpart to the Poisson noise intensity λ introduced in (4).

The compensator A of the discontinuous part X_d in (6) is assumed to have the form

$$A([0,t] \times F) = \int_0^t \int_F \chi_s(m) \mu(dm) b(ds) \tag{10}$$

where the (nonnegative) jump intensity $\chi_t(m)$ is a $\mathbb{F}^{\theta_d} \vee \mathbb{F}^{Y,\tau}$-predictable functional, μ is the finite measure on (J, Σ_J) introduced in (4), and b is the measure already appearing in (4) and (8). The measures b in (4), (8) and (10) need not have been chosen to be identical; however, this is a natural choice. b is called the base measure of the independent increment noise channel. In usual applications, b is Lebesgue and the reader uninterested in more abstract situations is invited to view b as Lebesgue for the remainder of this paper.

Suppose a mean-square constraint similar to (9) is imposed on the jump intensity $\chi_t(m)$:

$$\frac{1}{T} E\left[\int_0^T \int_J \chi_t^2(m) \mu(dm) b(dt)\right] \le Q^2 . \tag{11}$$

With χ in (10) satisfying (11), the subchannel model in (7) is recognized to be a straightforward generalization of the Kabanov Poisson-type point process channel model [8] with a mean-square constraint in place of the peak constraint considered there

The mean-square constraints (9) and (11) suggest a way to impose a single mean-square constraint on the encoder output X by a formal unification of ϕ and χ. For all $t \geq 0$, define

$$\Psi_t(m) = \begin{cases} \chi_t(m), & m \in J \\ \phi_t, & m = 0 \end{cases}$$

to be the formal intensity of the càdlàg process $X = X_c + X_d$ with the parts X_c and X_d as in (8) and (10). We introduce the formal mark measure $\nu = \mu + \delta_0$ defined in the obvious way on the whole of the space $M = J \cup \{0\}$. Here, δ_0 denotes the atomic measure which assigns all of its unit mass to the point 0. Since $\mu(J) = 1$, ν gives equal weight to the continuous and discontinuous parts of the channel. Now the two constraints (9) and (11) are combined in the single mean-square constraint

$$E\left[\frac{1}{T}\int_0^T\int_M \Psi_t^2(m)\nu(dm)b(dt)\right] \leq \Pi^2 \tag{12}$$

on the formal intensity Ψ of the encoded message process X. In this context, we call the LHS of (12) the formal power of the encoder output and Π^2 the formal power available to the encoder.

To summarize, the independent increment noise channel model is given by (5) and (7) with Lévy noise N in (2). The noise variance β measure is related to the channel base measure by $\beta = \eta b$ for some $\eta \geq 0$. The noise compensating measure B is correspondingly assumed to take the form $\lambda b \mu$ for some $\lambda \geq 0$. The encoded message X is assumed to be càdlàg with $\mathbb{F}^\theta \vee \mathbb{F}^{Y,\tau}$-predictable ($\tau > 0$) formal intensity Ψ satisfying the mean-square constraint (12). The channel base measure b is assumed to be continuous with $b([0,T]) = T$ while the formal mark measure ν has an atom of unit mass at zero and unit mass (without atoms) elsewhere:

$$\nu(\{0\}) = 1, \quad \nu(J) = 1.$$

III. CHANNEL CAPACITY

In this section the information capacity of the independent increment noise channel is given and its derivation is outlined. Proofs of results used in the derivation are given in the next section. We begin by recalling the definitions of channel information and channel capacity.

Channel information is defined to be the average mutual information [14] in the message θ and the channel output Y over the interval $[0,T]$:

$$I^T[\theta,Y] = E\left[\ln \frac{d\mu_{\theta Y}}{d\mu_{\theta\times Y}}\right]$$

provided $\mu_{\theta Y} \ll \mu_{\theta\times Y}$; otherwise $I^T[\theta,Y]=\infty$. Here, $\mu_{\theta Y}$, μ_θ, and μ_Y are the joint and marginal measures induced by the message and output processes, θ and Y, on the spaces S_θ, S_Y, and $S_\theta \times S_Y$ where S_θ and S_Y are the spaces of trajectories of θ and Y over the interval $[0,T]$. Also, $\mu_{\theta\times Y}$ is the product measure induced by μ_θ and μ_Y. The information capacity of the independent increment noise channel is the supremum of the channel information

$$\mathbb{C} = \sup_\theta \sup_X \frac{1}{T} I^T[\theta,Y]$$

where θ is any jointly measurable process defined over the interval $[0,T]$ and $X = X(\theta,Y)$ is any càdlàg $\mathbb{F}^\theta \vee \mathbb{F}^{Y,\tau}$-adapted message encoding satisfying the mean-square constraint (12) on its formal intensity.

The independent increment noise channel with delayed feedback considered here is memoryless in the sense of Kadota in [9] and Kadota, Zakai, and Ziv in [11]. In [11] it is shown that the information capacity of a continuous-time memoryless channel can not be increased by feedback. Thus, for the present purpose of obtaining the information capacity of the independent increment noise channel with $\tau > 0$, we can and do restrict our attention to the no-feedback case. This significantly simplifies the problem. In particular, we exploit the facts that in the absence of feedback the processes X and N are independent and that Y_c and Y_d are independent if and only if the components θ_d and θ_c of the message are independent.

The parallel Gaussian and marked Poisson subchannels of the independent increment noise channel dominate its structure. In the absence of feedback, these two subchannels have two points of interaction: the transmitted message components θ_c and θ_d may be dependent and the encoders of the two subchannels compete for the available formal power Π^2. Thus, the independent increment noise channel is distinguished by the possibility of redundantly encoding some or all of the message to be transmitted into both X_d and X_c. It is not immediately clear what effect such redundancy has on the combined capacity of the two subchannels. This question is settled with the aid of some definitions.

Let the combined message and output of n additive noise channels be expressed by $\theta=(\theta_1,\theta_2,\ldots,\theta_n)$ and $Y=(Y_1,Y_2,\ldots,Y_n)$, respectively. Also, define $N=(N_1,N_2,\ldots,N_n)$ and $X=(X_1,X_2,\ldots,X_n)$. The parallel combination channel (θ,X,N,Y) with n subchannels is defined to have the following properties:

1. $Y_{i,t} = X_{i,t} + N_{i,t}, \quad t \in [0,T],\ i = 1,2,\ldots,n$,

2. $X_{i,t} = X_{i,t}(\theta_i)$ is $\mathbb{F}^{\theta_i}$-adapted , $\quad i = 1,2,\ldots,n$,

3. θ , N are independent,

4. $N_1, N_2, \ldots, N_n$ are mutually independent.

Disregarding the mean-square constraint on the formal intensity, we observe that the independent increment noise channel without feedback is a parallel combination channel with two subchannels.

The capacity of the parallel combination channel is

$$\mathbb{C} = \sup_{\theta\in B_1} \sup_{\substack{X_i\in A_i \\ X\in C}} \frac{1}{T} I^T[\theta,Y]$$

where B_1 is the class of n-tuples $(\theta_1,\ldots,\theta_n)$ of jointly measurable processes and A_i, $i=1,2,\ldots,n$ is the class of $\mathbb{F}^{\theta_i}$-adapted encodings. Also, C is a class of encodings satisfying a fixed collection of constraints; for the independent increment noise channel, C is the class of càdlàg processes with formal intensity satisfying (12). It is clear that the capacity of the parallel combination channel is at least as great as the sum of the capacities of the individual subchannels. The following theorem states that nothing is gained by redundantly encoding some part of the message; allowing components of θ to be dependent does not increase capacity. The theorem's proof is given in the next section.

Theorem 1: Let (θ,X,N,Y) be a parallel combination channel with 2 subchannels (θ_1,X_1,N_1,Y_1) and (θ_2,X_2,N_2,Y_2). The capacities of the two subchannels are

$$\mathbb{C}_1(C_{X_2}) = \sup_{\theta_1\in B} \sup_{X_1\in C_{X_2}\cap A_1} \frac{1}{T} I^T[\theta_1,Y_1],$$

$$\mathbb{C}_2(C_{X_1}) = \sup_{\theta_2 \in B} \sup_{X_2 \in C_{X_1} \cap A_2} \frac{1}{T} I^T[\theta_2, Y_2]$$

where B is the class of jointly measurable processes, A_i, $i=1,2$ is the class of $\mathbb{F}^{\theta_i}$-adapted encodings, C_{X_1} is the set of all encodings $X \in C$ with first component X_1 fixed, and C_{X_2} is the set of all encodings $X \in C$ with second component X_2 fixed. Then

$$\mathbb{C} = \sup_{(X_1, X_2) \in C} \left\{ \mathbb{C}_1(C_{X_2}) + \mathbb{C}_2(C_{X_1}) \right\}.$$

Similar statements can be made for parallel combination channels with more than 2 subchannels. In terms of our channel model, Theorem 1 states that the information capacity of the independent increment noise channel is

$$\mathbb{C} = \sup_{R^2 + Q^2 \leq \Pi^2} \left\{ \mathbb{C}_W(R) + \mathbb{C}_P(Q) \right\} \tag{13}$$

where $\mathbb{C}_W(R)$ is the capacity of the Gaussian channel (5), (8) with Wiener noise W and mean-square constraint parameter R in (9) and where $\mathbb{C}_P(Q)$ is the capacity of the marked Poisson channel (7), (10) with marked Poisson noise P and mean-square constraint parameter Q in (11). The capacity $\mathbb{C}_W(R)$ is known [10]; with or without feedback it is

$$\mathbb{C}_W(R) = \frac{R^2}{2\eta}. \tag{14}$$

Therefore, to obtain the capacity of the independent increment noise channel, it only remains to determine the capacity of the marked Poisson channel with mean-square constraint (11) and to then find the supremum in (13).

The following theorem gives the capacity of the marked Poisson channel. The proof is given in the next section.

Theorem 2: Consider the marked Poisson channel model (7) with signal compensator (10) and noise compensator (4). Suppose that the channel base and mark measures b and μ are both finite and continuous and let the encoder intensity $\chi_t(m)$ be $\mathbb{F}^Y \vee \mathbb{F}^{\theta_d}$-predictable and mean-square-constrained as in (11) for some $Q \geq 0$. Then $\mathbb{C}_P(Q) = D(\lambda, Q)$ where $D(0, Q) = 2Q/e$ and, for $\lambda > 0$,

$$D(\lambda, Q) = \frac{Q^2}{\lambda\alpha^2}\left[1+\alpha\right]\ln(1+\alpha) - \lambda\left[\frac{P^2}{\alpha\lambda^2}+1\right]\ln\left[\frac{P^2}{\alpha\lambda^2}+1\right] \tag{15}$$

where α is the unique solution of the equation

$$\ln\left[\frac{P^2}{\alpha\lambda^2}+1\right] + 2 = \left[\frac{2}{\alpha}+1\right]\ln(1+\alpha).$$

The information capacity given in Theorem 2 is the instantaneous feedback capacity. When the proof of this result is examined, however, it is seen that feedback is not used in its derivation. Thus (15) is also the capacity for delayed feedback ($\tau > 0$) and, in particular, for the case in which the encoder intensity $\chi_t(m)$ is restricted to be $\mathbb{F}^{\theta_d}$-predictable - the no-feedback case. Causal feedback with or without delay cannot increase capacity in the marked Poisson channel considered here.

With regard to Theorem 2, we note that although the theorem is presented for a mark space $J \subset \mathbb{R}$ of jumps, one obtains the same result for more general mark spaces - provided the associated mark measure is finite.

The following theorem follows from (13) using Theorem 2 and (14).

Theorem 3: Fix $\tau > 0$. The information capacity of the independent increment noise channel is $\mathbf{C} = \Gamma(\lambda, \eta, \Pi)$ where

$$\Gamma(\lambda, \eta, \Pi) = \sup_{0 \le Q \le \Pi}\left\{\frac{\Pi^2 - Q^2}{2\eta} + D(\lambda, Q)\right\}. \tag{16}$$

The supremum in (16) does not in general admit analytic evaluation for $\lambda > 0$. However, for $\lambda = 0$, Γ can be more simply expressed. Using $D(0, Q) = 2Q/e$, the supremum in (16) is found to be

$$\Gamma(0, \eta, \Pi) = \begin{cases} \dfrac{\Pi^2}{2\eta} + \dfrac{2\eta}{e^2}, & \Pi > \dfrac{2\eta}{e} \\ \dfrac{2}{e}\Pi, & \Pi \le \dfrac{2\eta}{e} \end{cases}. \tag{17}$$

$\Gamma(0, \eta, \Pi)$ is presented as a function of Π in Figure 2. For comparison, the corresponding Gaussian and Poisson subchannel capacities $\Pi^2/(2\eta)$ and $2\Pi/e$ are also shown.

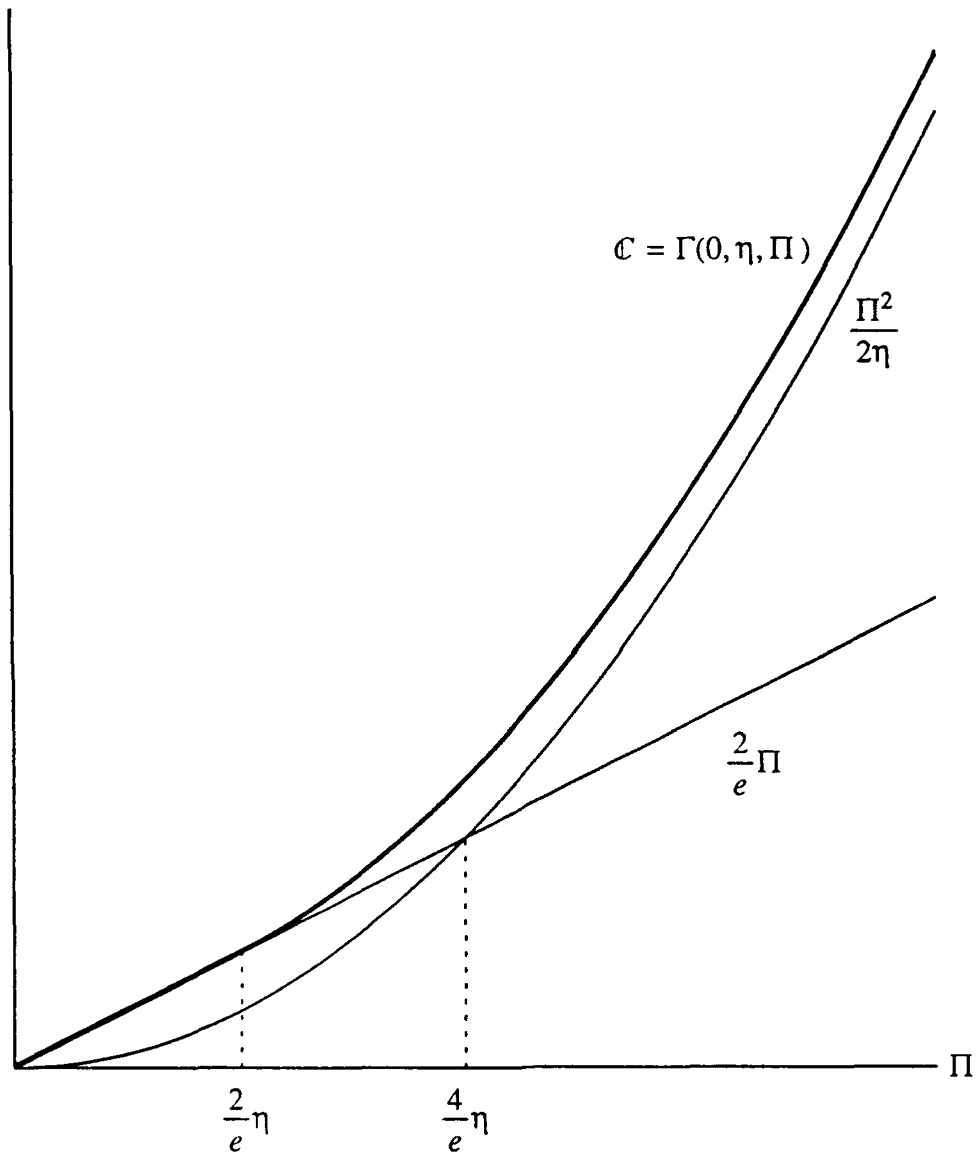

Figure 2. Information capacity ℂ of the independent increment noise channel with $\lambda = 0$.

The expression for $\Gamma(0,\eta,\Pi)$ in (17) exhibits a transition point at $2\eta/e$. This point is obtained by solving for the zero of the derivative of

$$\frac{\Pi^2-Q^2}{2\eta}+\frac{2}{e}Q$$

with respect to Q. For $\Pi<2\eta/e$, to maximize channel capacity, one does best to allocate all available formal power to the marked Poisson subchannel. For $\Pi>2\eta/e$, the formal power should be apportioned between the two subchannels with formal power equal to $(2\eta/e)^2$ assigned to the marked Poisson subchannel and the remainder to the Gaussian subchannel. For maximum information rate in the channel, at least some formal power must be assigned to the Poisson subchannel in all cases where $\lambda=0$.

The optimal distribution of the formal power Π^2 among the two subchannels of the independent increment noise channel for the case $\lambda>0$ can be addressed numerically, affording further insight to the nature of the channel. Numerical calculations involving Γ are somewhat simplified by the fact that Γ is first-order homogeneous; i.e. $\Gamma(wx,wy,wz)=w\Gamma(x,y,z)$ for all $w>0$. The results of calculations made for the optimal distribution of formal power are displayed in Figure 3 for a range of λ/Π and η/Π. There are three possible forms the optimal distribution might take: assignment of all of Π^2 to the Poisson subchannel, assignment of all of Π^2 to the Gaussian subchannel, or a nontrivial apportionment of Π^2 to both subchannels. According to Figure 3, all three possibilities do occur for the independent increment noise channel.

Figure 3 shows, in particular that for $\lambda\geq\eta$, the optimal distribution of formal power is to assign all of Π^2 to the Gaussian subchannel. Thus, in this case, the capacity of the independent increment noise channel is simply

$$\mathbf{C}=\frac{\Pi^2}{2\eta},\quad \lambda\geq\eta. \tag{18}$$

A proof of (18) is given in outline as follows: Hold λ and η fixed with $\lambda\geq\eta>0$. Let $g(Q)$ and $d(Q)$ denote the first partial derivatives with respect to Q of $Q^2/(2\eta)$ and $D(\lambda,Q)$, respectively. The second derivative of $D(1,Q)$ is found, upon evaluation at $Q=0$, to be equal to 1. Because $D(\lambda,Q)$ is first-order homogeneous, it then follows that $d'(0)=1/\lambda$. Also, $d'(Q)$ is a decreasing function of Q. $g'(Q)=1/\eta$ so, for $\lambda\geq\eta$ and all $Q\geq0$, $g'(Q)\geq d'(Q)$. We have $g(0)=d(0)=0$ so $g(Q)\geq d(Q)$ for all $Q\geq0$. Then the supand in (16) is nonincreasing in Q and the optimal choice of Q is $Q=0$. (18) follows.

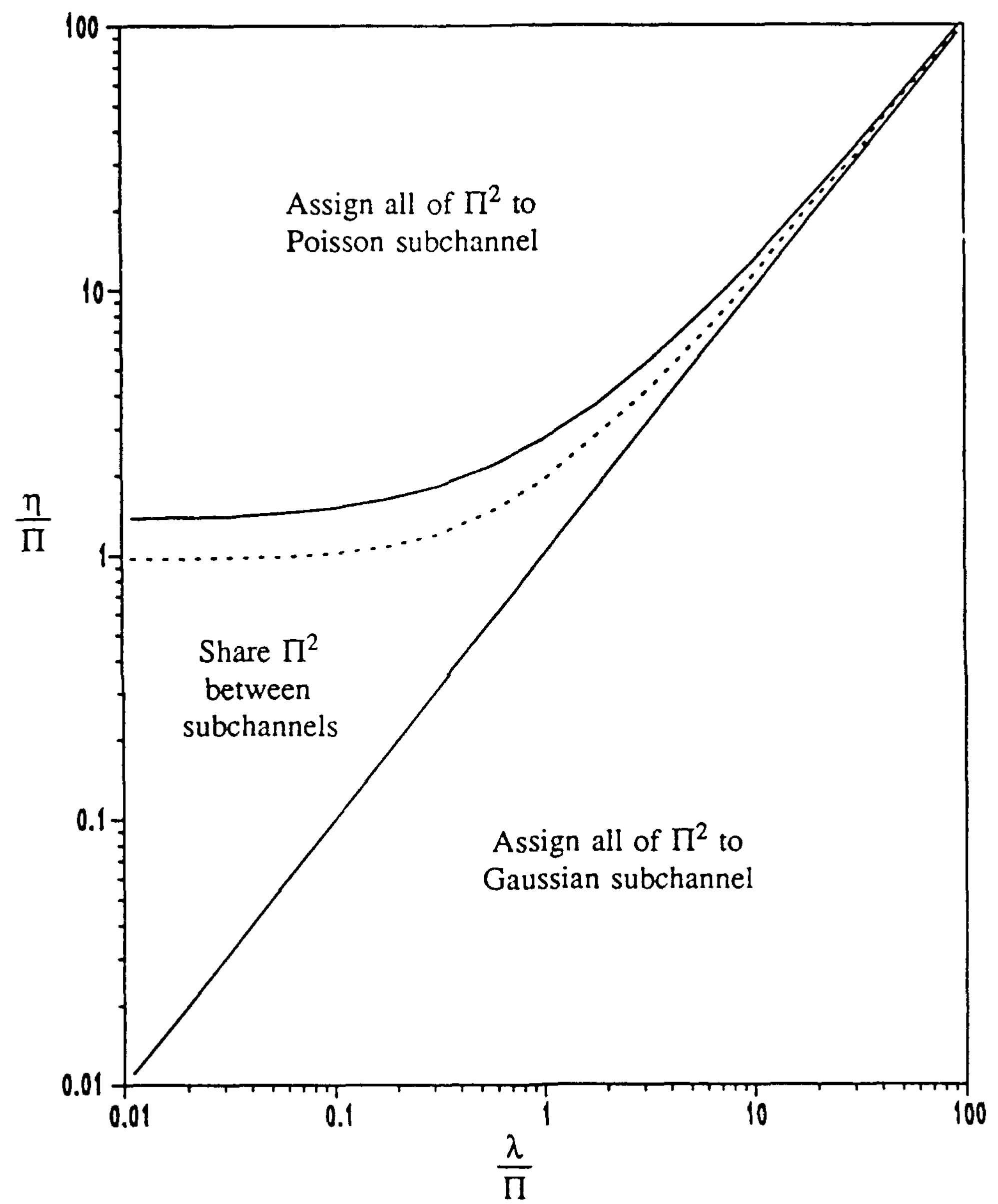

Figure 3. Optimal assignment of formal power, Π^2. Three regions shown as functions of the channel noise parameters, $\{\lambda, \eta\}$ scaled by Π. The dotted line is the contour along which each subchannel is optimally assigned half the available formal power.

For $\lambda/\eta < 1$, Figure 3 shows that the optimal distribution of Π^2 follows the general scheme observed for the special case $\lambda = 0$; for Π/η below some transition point (which varies as a function of λ/η) one optimally uses only the Poisson subchannel and above this transition point one optimally uses both subchannels. This point is illustrated in Figure 4 where the channel capacity $\mathbb{C} = \Gamma(\lambda, \eta, \Pi)$ is plotted for the representative case $\lambda/\eta = 0.1$. The similarity with the plot of $\Gamma(0, \eta, \Pi)$ in Figure 2 is evident.

The capacity formula (17) is given for the independent increment noise channel with delayed causal feedback even though (17) was only derived for the case of message encoding without feedback. As has already been observed, it follows from the result of Kadota, Zakai, and Ziv in [11] that the information capacity is the same in the two cases of delayed feedback and no feedback.

A quantity of significant practical interest for a channel is the coding capacity [2], [17]. Briefly, it is the supremum of information rates - the logarithm of the number of messages of length T divided by T - for which the error probability can be made arbitrarily small for sufficiently large T. The coding capacity $\mathbb{C}_{CODING}$ of the independent increment noise channel follows in straightforward fashion from Theorem 3. For the present channel model with $\tau > 0$, $\mathbb{C}_{CODING} = \Gamma(\lambda, \eta, \Pi)$ also. Skipping the details, this is proved first by showing that the coding capacity of the marked Poisson channel in Theorem 2 has the same expression as that given for the information capacity. An argument similar to Wyner's [18] for the peak-constrained Poisson channel proves this. Combined with the known coding capacity [2] of the additive Gaussian white noise channel, this gives that $\mathbb{C}_{CODING} \geq \Gamma(\lambda, \eta, \Pi)$. Fano's inequality completes the proof.

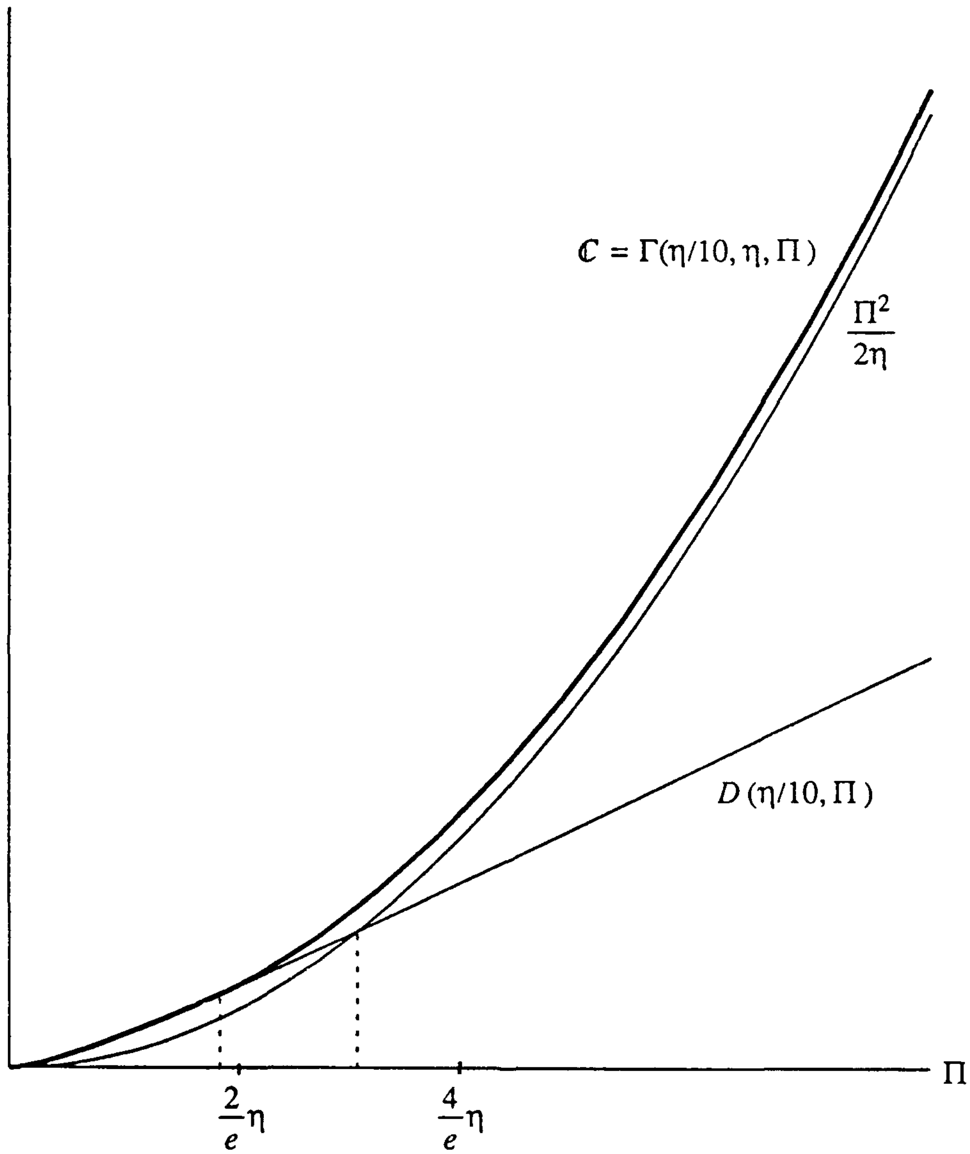

Figure 4. Information capacity $\mathbb{C}$ of the independent increment noise channel with $\lambda = \eta/10$. Compare with Figure 2.

IV. PROOFS

This section is divided into two subsections. In the first, a result for the information in the parallel combination channel is obtained, leading to a proof of Theorem 1. The second subsection contains a proof of Theorem 2.

Capacity of the Parallel Combination Channel

In this subsection we show that the capacity of the parallel combination channel is the sum of the capacities of its n subchannels. Repeated use is made of Kolmogorov's formula [13]. Let Z_1, Z_2, Z_3, and Z_4 be four random elements defined on a common probability space. Kolmogorov's formula is

$$I[(Z_1,Z_1),Z_3] = I[Z_1,Z_3] + E[I[Z_2,Z_3|Z_1]]. \tag{19}$$

When generalized to average conditional informations, Kolmogorov's formula takes the form [13]

$$E[I[(Z_1,Z_2),Z_3|Z_4]] = E[I[Z_1,Z_3|Z_4]] + E[I[Z_2,Z_3|(Z_1,Z_4)]]. \tag{20}$$

Multiinformation [15] appears naturally in the following discussion. The multiinformation $M[Z_1,\dots,Z_n]$ in n random elements $Z_1,\dots,Z_n$ is defined as a relative entropy in [15]. An equivalent definition is

$$M[Z_1,\dots,Z_n] = \sum_{k=2}^{n} I[(Z_1,\dots,Z_{k-1}),Z_k]. \tag{21}$$

Multiinformation is an extension of average mutual information; for two random elements $M[Z_1,Z_2]=I[Z_1,Z_2]$. Like average mutual information, multiinformation is nonnegative and zero if and only if the corresponding random elements are mutually independent.

Lemma 1: Let (θ,X,N,Y) be a parallel combination channel with n subchannels. Then, for all $i \neq j$,

$$I^T[(\theta_i,\theta_j),Y_j] = I^T[\theta_j,Y_j],$$

$$I^T[(\theta_i,Y_i),Y_j] = I^T[\theta_i,Y_j],$$

$$I^T[(\theta_i,(Y_i,Y_j)] = I^T[\theta_i,Y_i].$$

Proof: We prove the first equality; the others are proved similarly. θ_i and N_j are independent so

$$I^T[\theta_i,Y_j|\theta_j] = I^T[\theta_i,X_j(\theta_j)+N_j|\theta_j] = 0.$$

Then, by Kolmogorov's formula (19),

$$I^T[(\theta_i,\theta_j),Y_j] = I^T[\theta_j,Y_j] + E[I^T[\theta_i,Y_j|\theta_j]]$$

$$= I^T[\theta_j,Y_j].$$

Proposition 1: Let (θ,X,N,Y) be a parallel combination channel with n subchannels. Then,

$$I^T[\theta,Y] = \sum_{i=1}^{n} I^T[\theta_i,Y_i] - M^T[Y_1,Y_2,...,Y_n]$$

where $M^T[Y_1,Y_2,...,Y_n]$ is the multiinformation in the n subchannel outputs over the interval $[0,T]$.

Proof: Consider $n=2$. From (19),

$$I^T[(\theta_1,\theta_2),(Y_1,Y_2)] = I^T[\theta_1,(Y_1,Y_2)] + E[I^T[\theta_2,(Y_1,Y_2)|\theta_1]].$$

By Lemma 1 above, $I^T[\theta_1,(Y_1,Y_2)]=I^T[\theta_1,Y_1]$. Then, using (20),

$$I^T[(\theta_1,\theta_2),(Y_1,Y_2)] = I^T[\theta_1,Y_1]$$

$$+ E[I^T[\theta_2,Y_1|\theta_1]]$$

$$+ E[I^T[\theta_2,Y_2|(\theta_1,Y_1)]].$$

Also

$$I^T[\theta_2,Y_2|(\theta_1,Y_1)] = I^T[\theta_2,Y_2|(\theta_1,N_1)] = I^T[\theta_2,Y_2|\theta_1]$$

since Y_2 is conditionally independent of N_1 given (θ_1,θ_2). So

$$I^T[(\theta_1,X_2),(\theta_1,Y_2)] = I^T[\theta_1,Y_1]$$

$$+ E[I^T[\theta_1,Y_2|Y_1]]$$

$$+ E[I^T[\theta_2,Y_2|\theta_1]].$$

Kolmogorov's formula is used twice more to obtain

$$E[I^T[\theta_2, Y_2|\theta_1]] = I^T[(\theta_1,\theta_2), Y_2] - I^T[\theta_1, Y_2],$$

$$E[I^T[\theta_1, Y_2|Y_1]] = I^T[(\theta_1, Y_1), Y_2] - I^T[Y_1, Y_2].$$

Then, using the preceding lemma, we get

$$\begin{aligned} I^T[(\theta_1,\theta_2),(Y_1,Y_2)] = I^T[\theta_1, Y_1] \\ + I^T[\theta_2, Y_2] - I^T[\theta_1, Y_2] \\ + I^T[\theta_1, Y_2] - I^T[Y_1, Y_2] \end{aligned}$$

which is the desired result for $n=2$. Similarly, for $n=k$ one can show that

$$\begin{aligned} I^T[(\theta_1,\ldots,\theta_k),(Y_1,\ldots,Y_k)] = I^T[(\theta_1,\ldots,\theta_{k-1}),(Y_1,\ldots,Y_{k-1})] \\ + I^T[\theta_k, Y_k] - I^T[(Y_1,\ldots,Y_{k-1}), Y_k]. \end{aligned}$$

Then, by recursion and the relation (21) the result is obtained for all n. Therefore, the proof is complete.

Since multiinformation is always nonnegative, Proposition 1 implies that for parallel channels without feedback, the channel information in the composite channel is dominated by the sum of the informations in the component subchannels. Proposition 1 quantifies the degree to which the composite mutual information is dominated by the sum of its component informations - according to Proposition 1, the sum of the component informations exceeds the composite information by exactly the multiinformation $M^T[Y_1,\ldots,Y_n]$ in the channel output. Since the outputs $Y_1,\ldots,Y_n$ of the parallel combination channel are mutually independent iff $\theta_1,\ldots,\theta_n$ are mutually independent, the proposition further implies that the composite information equals the sum of the component informations if and only if the components of the input are mutually independent. We state this formally as a corollary from which Theorem 1 follows directly.

Corollary: Let (θ, X, N, Y) be a parallel combination channel with n subchannels. Then,

$$I^T[\theta, Y] \le \sum_{i=1}^{n} I^T[\theta_i, Y_i]$$

with equality iff $\theta_1, \ldots, \theta_n$ are mutually independent. Also,

$$\mathbb{C} = \sum_{i=1}^{n} \mathbb{C}_i$$

where $\mathbb{C}$ is the capacity of the parallel combination channel and $\mathbb{C}_i$ is the capacity of its ith subchannel.

The corollary implies that, in seeking the capacity of the parallel combination channel, one need only consider messages θ such that $\theta_1, \theta_2, \ldots, \theta_n$ are mutually independent.

The foregoing calculations are similar to some performed by Kadota [8] for time segments of the continuous-time memoryless channel.

Capacity of the Marked Poisson Channel

The proof of Theorem 2 relies on results given in [5] for the capacity of the Poisson channel (unmarked) with a mean-square constraint. These results are repeated here as Lemmas 2 and 3.

Lemma 2: Fix $Q \geq 0$, $\lambda \geq 0$ and let A_Q denote the set of all nonnegative random variables X with constrained second moment $E[X^2] \leq Q^2$. Also, $I[X] \equiv E[X \ln X] - E[X] \ln E[X]$. Then

$$\sup_{X \in A_Q} I[X + \lambda] = D(\lambda, Q).$$

Lemma 3: Consider the Poisson channel with Poisson-type point process channel output Y_t, $t \in [0,T]$ with compensator

$$\int_0^t (\chi_s + \lambda) b(ds)$$

where $\lambda \geq 0$ is the noise intensity, b is the channel base measure ($b[0,T] = T$), and $\chi_t = \chi_t(\theta, Y)$ is a predictable encoding of the message process θ_t, $t \in [0,T]$. Let the encoder intensity χ_t be mean-square-constrained, $E[\chi_t^2] \leq Q^2$ for all $t \in [0,T]$. The information capacity of this channel is $D(\lambda, Q)$.

Lemma 4: Let D be a Borel subset of $\mathbb{R}$ and define A_D to be the class of random variables with range in D. Let f be a real function and suppose f^{-1}, the inverse of f, exists on D. For $P \in D$, define $A = \{X \in A_D : E[f(X)] \le f(P)\}$. Let g be a real function such that $g \circ f^{-1}$ is concave and nondecreasing. Then

$$\max_{X \in A} E[g(X)] = g(P).$$

Proof: Define $h = g \circ f^{-1}$. Then

$$E[g(X)] = E[h(f(X))] \le h(E[f(X)]) \le h(f(P)) = g(P).$$

Let $X = P$. Then $X \in A$ and $E[g(X)] = g(P)$. The result follows.

Lemma 5: Consider the Poisson channel of Lemma 3 except, instead of the mean-square constraint imposed there on the encoder intensity, let the encoder constraint be

$$\frac{1}{T}\int_0^T E[\chi_t^2]b(dt) \le Q^2. \tag{22}$$

The capacity is still $C = D(\lambda, Q)$.

Proof: The channel information is [12]

$$I^T[\theta, Y] = \int_0^T E[\eta_t \ln \eta_t - \hat{\eta}_t \ln \hat{\eta}_t] b(dt).$$

Here $\eta_t = \chi_t + \lambda$ and $\hat{\eta}_t$ is the predictable version of $E[\eta_t | \mathbb{F}_t^Y]$. By Jensen's inequality,

$$I^T[\theta, Y] \le \int_0^T I[\chi_t + \lambda] b(dt).$$

By Lemma 2, $I[X + \lambda]$ has the upper bound $D(\lambda, E^{1/2}[X^2])$. Therefore

$$I^T[\theta, Y] \le \int_0^T D(\lambda, \sqrt{M(t)}) b(dt)$$

where $M(t) = E[\chi_t^2]$. Thus

$$C \le \frac{1}{T} \sup_{M(t) \in \Phi} \int_0^T D(\lambda, \sqrt{M(t)}) b(dt) \tag{23}$$

where Φ is the class of nonnegative functions

$$\Phi = \{M(t): M(t) \geq 0,\ \frac{1}{T}\int_0^T M(t) b(dt) \leq Q^2\}.$$

The function $D(\lambda, \sqrt{x})$ is nonnegative, increasing, and concave as a function of x. Also, b is finite so, by Lemma 4, the supremum in (23) is realized for $M(t)$ equal to a constant. Thus we have

$$\mathbf{C} \leq \frac{1}{T}\int_0^T D(\lambda, \sqrt{Q^2}) b(dt) = D(\lambda, Q).$$

The constraint $E[\chi_t^2] = Q^2$ for all $t \in [0,T]$ is stronger than the constraint (22) so $\mathbf{C} \geq D(\lambda, Q)$, also. This completes the proof.

Proof of Theorem 2: We begin just as in the proof of Lemma 5. The channel information in the marked Poisson channel is

$$\int_J \int_0^T E[\eta_t(m)\ln\eta_t(m) - \hat{\eta}_t(m)\ln\hat{\eta}_t(m)] b(dt)\mu(dm);$$

this expression is easily inferred from that given in [12] for the Poisson channel. Here $\eta_t(m) = \chi_t(m) + \lambda$ and $\hat{\eta}_t(m)$ is the predictable version of $E[\eta_t(m)|\mathbb{F}_t^Y]$. By Jensen's inequality,

$$I^T[\theta, Y] \leq \int_J \int_0^T I[\chi_t(m) + \lambda] b(dt)\mu(dm).$$

By Lemma 2, $I[X + \lambda]$ has the upper bound $D(\lambda, E^{1/2}[X^2])$. Therefore

$$I^T[\theta, Y] \leq \int_J \int_0^T D(\lambda, \sqrt{M(t, m)}) b(dt)\mu(dm)$$

where $M(t, m) = E[\chi_t^2(m)]$. Thus

$$\mathbf{C} \leq \frac{1}{T} \sup_{M(t,m) \in \Phi} \int_J \int_0^T D(\lambda, \sqrt{M(t, m)}) b(dt)\mu(dm) \tag{24}$$

where Φ is the class of nonnegative functions

$$\Phi = \{M(t, m): M(t, m) \geq 0,\ \frac{1}{T}\int_J \int_0^T M(t, m) b(dt)\mu(dm) \leq Q^2\}.$$

The function $D(\lambda, \sqrt{x})$ is nonnegative, increasing, and concave as a function of x. Also, b and μ are finite so, by Lemma 4, the supremum in (24) is realized for $M(t,m)$ equal to a constant. Thus we have

$$\mathbb{C} \leq \frac{1}{T}\int_J\int_0^T D(\lambda, \sqrt{Q^2})b(dt)\mu(dm) = D(\lambda, Q). \tag{25}$$

We now show that the RHS of (25) is also a lower bound on $\mathbb{C}$. Let the marks on the channel output be ignored; i.e., process the output Y of the marked Poisson channel being considered here as follows:

$$\{Y_t(F), F \in \Sigma_J\} \to Y_t(J).$$

In other words, only the path $Y(J)$ of the channel output is retained. In such cases $I^T[\theta, Y] \geq I^T[\theta, Y(J)]$ so $\mathbb{C}$ is lower-bounded by the capacity of the Poisson channel (unmarked) with channel output $Y(J)$. This second channel has noise and encoder intensities

$$\int_J \lambda\mu(dm) = \lambda, \quad \int_J \chi_t(m)\mu(dm).$$

Suppose the encoder intensity $\chi_t(m)$ is chosen to be a function $\chi_t \equiv \chi_t(m)$ of only t - not of m. Then we have a Poisson channel with noise intensity λ and encoder intensity χ_t with constraint

$$\frac{1}{T}\int_0^T E[\chi_t^2]b(dt) \leq Q^2.$$

Therefore, by Lemma 5, $\mathbb{C} \geq D(\lambda, Q)$ and the theorem is proved.

ACKNOWLEDGEMENT

The author considers it his pleasant duty to thank Prof. Charles R. Baker for proposing the present study.

REFERENCES

[1] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[2] R. Blahut, *Principles and Practice of Informat:ɔn Theory*, Addison-Wesley Publ. Co., Reading, MA., 1987.

[3] P. Brémaud, "An Averaging Principle for Filtering a Jump Process with Point Process Observations," *IEEE Trans. Info. Theory*, Vol. 34, No. 3, pp. 582-585, 1988.

[4] M. Davis, "Capacity and cut-off rate for Poisson-type channels," *IEEE Trans. on Information Theory*, Vol. 26, No. 6, pp. 710-715, 1980.

[5] M.R. Frey, "Information Capacity of the Mean-Square-Constrained Poisson Channel," *Annals of Appl. Prob.*, in review.

[6] I. Gikhman and A. Skorohod, *Introduction to the Theory of Random Processes*, Saunders Co., 1969.

[7] K. Ito, *Stochastic Processes*, Matematisk Institut, Aarhus Universitet, Lecture Notes Series, No. 16, August 1969.

[8] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Prob. and Appl.*, Vol. 23, 1978.

[9] T.T. Kadota, "On the Mutual Information of Memoryless Channels," *IEEE Trans. Info. Theory*, Vol. 17, No. 2, pp. 140-143, March 1971.

[10] T.T. Kadota, M. Zakai, and J. Ziv, "Mutual Information of the White Gaussian Channel With and Without Feedback," *IEEE Trans. Info. Theory*, Vol. 17, No. 4, July 1971.

[11] T.T. Kadota, M. Zakai, and J. Ziv, "Capacity of a Continuous Memoryless Channel with Feedback," *IEEE Trans. Info. Theory*, Vol. 17, No. 4, July 1971.

[12] G. Kallianpur, *Stochastic Filtering Theory*, Springer-Verlag, Inc., New York, 1980.

[13] R. Liptser and A. Shiryayev, *Statistics of Random Processes, Vol. II, Applications*, Springer-Verlag, New York, 1977.

[14] M.S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.

[15] A. Segall and T. Kailath, "Radon-Nikodym Derivatives with Respect to Measures Induced by Discontinuous Independent Increment Processes," *Annals of Prob.*, Vol. 3, No. 3, pp. 449-464, 1975.

[16] M. Studený, "Multiinformation and the Problem of Characterization of Conditional Independence Relations," *Prob. Contr. and Info. Theory*, Vol. 18, No. 1, pp. 3-16, 1989.

[17] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer-Verlag, Berlin, Third Edition, 1978.

[18] A.D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel - Part I, II," *IEEE Trans. Info. Theory*, Vol. 34, No. 6, November 1988.