INFORMATION AND CONTROL 8, 36–55 (1965)

Indecomposable Channels with Side Information at the Transmitter*

FREDERICK JELINEK

School of Electrical Engineering, Cornell University, Ithaca, New York

In this paper the direct part and the strong converse of the coding theorem for two classes of Finite State Indecomposable Channels with Side Information at the Transmitter are proven. The question of membership in the first class can always be easily settled; to show that a channel belongs to the second class requires in general an infinite number of operations. A finite test is developed that is applicable if the given channel satisfies either of two additional restrictions. Fortunately, the second of these will be met by any "practical" indecomposable channel.

A PARTIAL LIST OF SYMBOLS AND ABBREVIATIONS

FSI channels	Finite state channels with side information at the transmitter
MISI channels	Finite state Markovian indecomposable channels with side information at the transmitter
SISI channels	Strongly indecomposable channels with side information at the transmitter
SIA matrix	Stochastic, indecomposable, aperiodic matrix
$\mathfrak{f}^m$	The set of all strategy letters f of order m
K, J	Channels with side information
K^m, J^m	Corresponding associated channels of order m
$\mathcal{C}$	A set of states in a Markov chain
$\mathcal{S}$	A set of subchannel output sequences
Σ	A subset of subchannel output sequences

I. INTRODUCTION

The main purpose of this paper is to prove the direct and the strong converse parts of the coding theorem for Finite State Markovian In-

* This work was supported, in part, by a Ford Foundation grant made through the Cornell College of Engineering.

decomposable Channels with Side Information at the Transmitter (MISI channels), and for Finite State Strongly Indecomposable Channels with Side Information at the Transmitter (SISI Channels). Since it is in general not possible to carry out the test for SISI character in a finite number of steps, we give in Section VI two simple sufficient conditions, the second of which any indecomposable channel will meet "in practice."

Both MISI and SISI channels form an interesting subclass of Finite State Channels with Side Information at the Transmitter (FSI channels) characterized by a set of transmission probability matrices $[p_s(y/x)]$ where $y \in \{0, \cdots, b-1\}$ are the possible output signals, $x \in \{0, \cdots, a-1\}$ are the possible input signals, and $s \in \{0, \cdots, h-1\}$ is an index specifying the channel state. The latter is allowed to change from one time interval to the next according to some statistical rule, and its identity is made known to the transmitter prior to the selection of the signal to be transmitted. Shannon (1958) gave a solution to the capacity problem for that subclass of FSI channels in which the states s_i were selected at time i $(= 1, 2, \cdots)$ with constant probability $r(s_i)$, independently of proceeding states $s_k (k < i)$ and of the channel inputs x_j and outputs y_j $(j = 1, 2, \cdots, i, i+1, \cdots)$.

II. DEFINITION OF FINITE STATE MARKOVIAN INDECOMPOSABLE CHANNELS WITH SIDE INFORMATION AT THE TRANSMITTER

The MISI channels we consider here are a generalization of the channels considered by Shannon (1958) in that their successive states are selected by an indecomposable aperiodic[1] Markov chain.

Consider the channel of Fig. 1 having inputs $x \in \{0, \cdots, a-1\}$, outputs $y \in \{0, \cdots, b-1\}$, and states $s \in \{0, \cdots, h-1\}$. Its operation at any time is specified by the probabilities $w(s_{i+1}, y_i/s_i, x_i)$ that the letter y_i will be received and that the next state will be s_{i+1}, given that the present state was s_i and that x_i was transmitted. The identity of the state s_i is to be made known to the transmitter which can then base the selection of the input x_i on this knowledge. The channel will be a MISI channel if its $w(\cdot, \cdot/\cdot, \cdot)$ satisfies the following restriction:

$$w(s_{i+1}/s_i, x_i) = \sum_{y_i=0}^{b-1} w(s_{i+1}, y_i/s_i, x_i) = r(s_{i+1}/s_i)$$

[1] A chain is indecomposable if in the terminology of Feller (1957), p. 355, it contains at most one closed state set other than the whole chain itself. For the definition of aperiodicity see Feller (1957), p. 353.

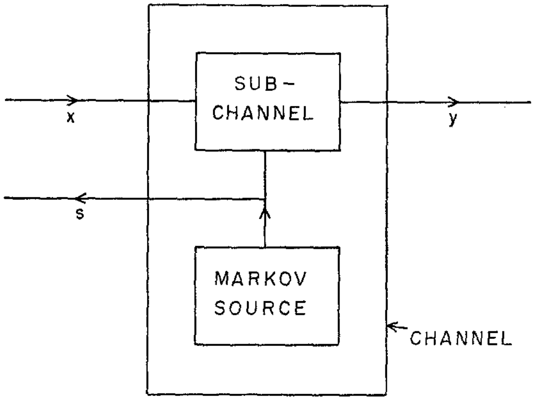

FIG. 1. Finite state Markovian indecomposable channel with side information at the transmitter.

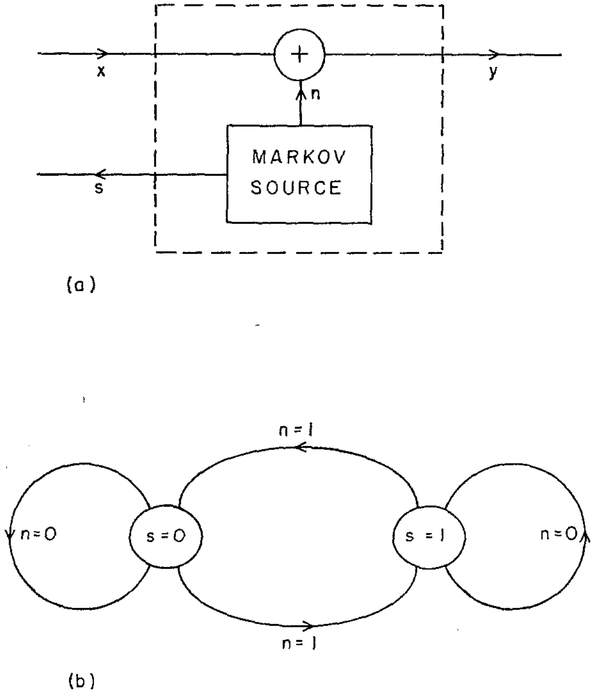

FIG. 2. Example of an MISI channel. In part (b),

$$P\{s_{i+1} = 0/s_i = 0\} = P\{s_{i+1} = 0/s_i = 1\} = \alpha$$

$$P\{s_{i+1} = 1/s_i = 0\} = P\{s_{i+1} = 1/s_i = 1\} = 1 - \alpha$$

for all

$$
\begin{aligned}
x_i &\in \{0, \cdots, a-1\} \\
s_i, s_{i+1} &\in \{0, \cdots, h-1\} \qquad (1) \\
i &= 1, 2, \cdots
\end{aligned}
$$

and $r(s_{i+1}/s_i)$ is the transition matrix of an indecomposable, aperiodic[1] Markov chain.

Restriction (1) states that the channel inputs do not influence the selection of channel states, and justifies the diagramatic form of Fig. 1.

We may also define a transmission probability matrix $[p_s(y/x)]$ by the expression

$$p_{s_i}(y_i/x_i) = \sum_{s_{i+1}=0}^{h-1} w(s_{i+1}, y_i/s_i, x_i) \qquad (2)$$

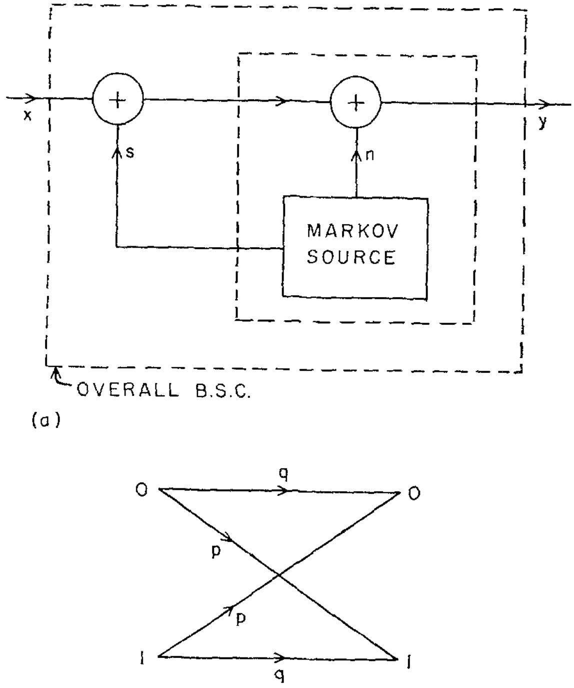

FIG. 3. Equivalent representation of the MISI channel of Fig. 2. If $s_2 = 0$ then $p = 1 - \alpha$, if $s_2 = 1$ then $p = \alpha$.

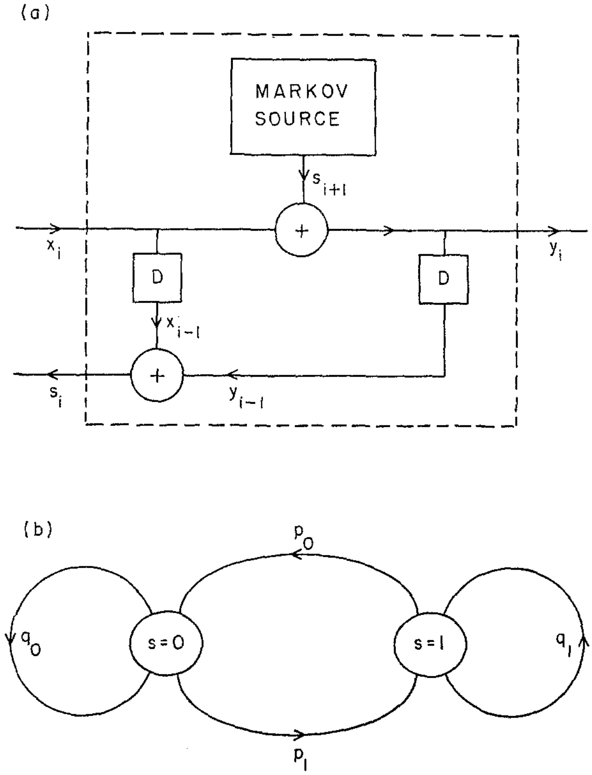

FIG. 4. A MISI channel constructed from a two-state binary symmetric channel with noiseless feedback. The boolean additive noise consists of the states of the Markov source whose transition probabilities are indicated in (b).

Figure 2 is an example of a channel satisfying restriction (1). The binary noise source (b) is actually memoryless, but the probability of noise is different in the two different states. The initial state can be made known to the receiver as follows: If $s_1 = 0$, the transmitter sends $x_1^* = 0$; if $s_1 = 1$, the transmitter sends $x_1^* = 1$. Then whenever $y_1 = 0$, $s_2 = 0$, and whenever $y_1 = 1$, $s_2 = 1$. Once the initial state is established, the channel is "hooked up" as in Fig. 3 (a), and this gives rise to the overall binary symmetric channel of Fig. 3 (b). In Fig. 3 (b) if $s_2 = 0$ (and the receiver knows whether this is so!!) then $p = 1 - \alpha$, while if $s_2 = 1$ then $p = \alpha$.

The second example, given in Fig. 4 (a), is a binary channel with feedback in which the Boolean additive noise is generated by the Markov source of Fig. 4 (b). Also here, restriction (1) is satisfied.

III. A SUMMARY OF RESULTS

We shall first prove the coding theorem and its strong converse for MISI channels. In Section IV we will construct out of MISI channels a sequence of mth order associated finite state channels without side information and find the capacity of the latter by showing that they are indecomposable (see Blackwell *et al.* (1958)). In Section V we will derive the capacity expression for the MISI channels by showing that any code for an associated channel can be translated into a code for the underlying MISI channel and vice versa.

The above manner of proof will lead naturally to the question of whether the coding theorem could be generalized to cover the entire class of Finite State Indecomposable Channels when side information about their current state is made available to the transmitter. This is possible only if all the channels in the sequence of mth order associated channels are indecomposable. Channels which satisfy this condition are termed Strongly Indecomposable (SISI channels), and for them we derive the appropriate capacity expression in Section VI.

It is then of interest to see whether perhaps all indecomposable channels are strongly indecomposable as well. In Section VII we provide a counterexample. We also specify two different and simple to check sufficient conditions under which an indecomposable channel is strongly indecomposable. It will be seen that the second of these will be naturally met by all "practical" indecomposable channels. In general, of course, to check for strong indecomposability is out of the question, since this would involve an infinite sequence of operations.

IV. ASSOCIATED MISI CHANNELS AND THEIR CAPACITY

We will find the capacity of MISI channels by an approach similar to the one used by the author on two-way channels.[2]

Let $\{f\} = \mathfrak{f}^m$ be the complete set of order a^{h^m} of functions, called strategy letters, which map sequences $(s_{k-m+1}, \cdots, s_k)$, $s_i \in \{0, \cdots, h-1\}$, into the set $\{0, \cdots, a-1\}$ of channel inputs. Using the strategy letters, a transducer, and the given MISI channel denoted by K, one can construct an associated channel K^m as in Fig. 5. The inputs to K^m are letters $f \in \mathfrak{f}^m$, and the outputs are signals $y \in \{0, \cdots, b-1\}$. The

[2] See Jelinek (1963), section 3.

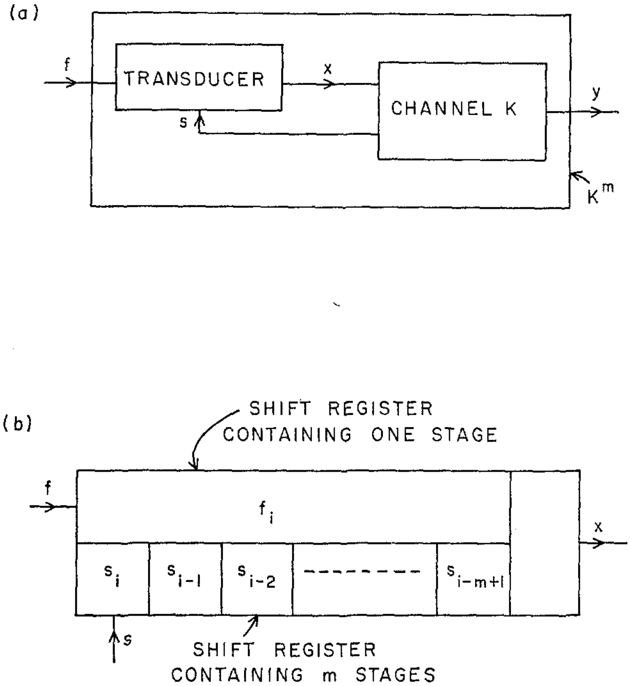

FIG. 5. (a) The associated MISI channel K^m. (b) The schematie diagram of the transducer.

transducer contains two shift registers: the first has one stage and contains one of the letters $f \in \mathfrak{f}^m$, the other has m stages, each containing indices $s \in \{0, \cdots, h-1\}$. The channel operates as follows: at time k, both transducer registers shift one step to the right, rejecting the contents of their rightmost stages. The first register is then filled with a particular K^m input f, and the first stage of the second register with the present state s_k of K. The transducer then puts out the K-input $f(s_k, \cdots, s_{k-m+1})$, which is transmitted through K under the probability law $p_{s_k}(\cdot/\cdot)$ and received as some signal $y \in \{0, \cdots, b-1\}$. The next time interval is then ready to start.

It is clear that the channel K^m has a well-defined transmission probability law

$$V(y_1, \cdots, y_n/f_1, \cdots, f_n, s_1, s_0, \cdots, s_{2-m}) \\ = p_{s_1}(y_1/f_1(s_1, s_0, \cdots, s_{2-m})). \tag{3}$$

$$\sum_{s_2,\cdots,s_n} \prod_{i=2}^{n} p_{s_i}(y_i/f_i(s_i, \cdots, s_{i-m+1}))\, r(s_i/s_{i-1})$$

specifying the output sequence given the input sequence and the initial K^m-channel state.

Let Q be a probability distribution over sequences of t letters $f \in \mathfrak{f}^m$ (sometimes, for clarity, we may write Q^t instead). Define the conditional probability measure P_Q over sequences of t signals $y \in \{0, \cdots, b-1\}$ given the initial state sequence $(s_1, s_0, \cdots, s_{2-m})$ by

$$\begin{aligned} &P_Q(y_1, \cdots, y_t/s_1, \cdots, s_{2-m}) \\ &= \sum_{f_1,\cdots,f_t} V(y_1, \cdots, y_t/f_1, \cdots, f_t, s_1, \cdots, s_{2-m})\, Q(f_1, \cdots, f_t) \\ &\qquad\qquad y_j \in \{0, \cdots, b-1\} \qquad j = 1, \cdots, t \\ &\qquad\qquad s_i \in \{0, \cdots, h-1\} \quad i = 2-m, \cdots, 1 \end{aligned} \tag{4}$$

In the following we will adopt the capital letter notation (5) for sequences of symbols of length t ending at time i:

$$Z_i^t = z_{i-t+1}, \cdots, z_i \tag{5}$$

where the letter z stands throughout (5) for a symbol taken from the same alphabet (e.g., $X_i^t = x_{i-t+1}, \cdots, x_i$). We will also write $Z^t = Z_t^t$ (e.g., $Y^t = y_1, \cdots, y_t$). Define the quantity

$$R(t, m/Q) = E_Q \frac{1}{t} \log \frac{V(Y^t/F^t)}{P_Q(Y^t)} \tag{6}$$

where

$$V(Y^t/F^t) = \sum_{S_1^m} r(S_1^m)\, V(Y^t/F^t, S_1^m)$$

and

$$P_Q(Y^t) = \sum_{S_1^m} r(S_1^m)\, P_Q(Y^t/S_1^m)$$

In (6) $r(S_1^m)$ stands for the stationary probability of the sequence S_1^m of states of the underlying Markov chain (1). We will always write E_Q for expectation with respect to the probability distribution Q given in the subscript.

We will now prove three lemmas. We will use them to prove Theorem 1 giving the capacity of the channel K.

LEMMA 1. *For all positive integers m, K^m is an indecomposable channel.*[3]

PROOF: Consider the sequences $S_i^m = (s_{i-m+1}, \cdots, s_i)$ as h-ary numbers and represent them by their decimal equivalent

$$\sigma_i = h^0 s_i + h^1 s_{i-1} + \cdots + h^{m-1} s_{i-m+1}, \qquad s_j \in \{0, \cdots, h-1\} \quad (7)$$

Thus σ can take on values in the set $\{0, \cdots, h^m - 1\}$. Let $\{D(f)\}, f \in \mathfrak{f}^m$, be the set of $bh^m \times bh^m$ matrices whose entries, specified by the pair $(\bar{\sigma}, \bar{y})$, $\bar{y} \in \{0, \cdots, b-1\}$, are given in (8). The row (column) (σ', y') will precede the row (column) (σ'', y'') if either $\sigma' < \sigma''$ or if $\sigma' = \sigma''$ and $y' < y''$. The entries of the matrix $D(f)$ are

$$d((\sigma_i, y_{i-1}); (\sigma_{i+1}, y_i)/f) = \begin{cases} p_{s_i}(y_i/f(\sigma_i))\, r(s_{i+1}/s_i) \\ \quad \text{if } \sigma_{i+1} = h\sigma_i - h^m s_{i-m+1} + s_{i+1} \\ \quad \text{and } \sigma_i \text{ represents a possible state} \\ \quad \text{sequence with respect to the} \\ \quad \text{matrix } [r(s_{i+1}/s_i)] \\ 0 \quad \text{otherwise} \end{cases} \quad (8)$$

Thus if σ_i represents the sequence $(\alpha_{i-m+1}, s_{i-m-2}, \cdots, s_i)$ the element $d((\sigma_i, y_{i-1}); (\sigma_{i+1}, y_i)/f)$ can be nonzero only if σ_{i+1} represents the sequence $(s_{i-m-2}, \cdots, s_i, \alpha_{i+1})$; $\alpha_{i-m+1}, \ \alpha_{i+1}, s_j \in \{0, \cdots, h-1\}$, $j = i - m + 2, \cdots, i$.

It is clear from (8) that $\{D(f)\}, f \in \mathfrak{f}^m$ is a set of stochastic matrices which fully specifies K^m.

Blackwell *et al.* (1958) start with a rather complicated definition of indecomposable channels and in their Theorem 1 show that the following simple one is completely equivalent.

> A channel specified by a set of stochastic matrices $\{D(f)\}$ defines an indecomposable channel if and only if every finite product $D(f_1) \cdot D(f_2) \cdots D(f_k) = L$ is an indecomposable Markov matrix, $k = 1, 2, \cdots; f_i \in \mathfrak{f}^m$. (9)

We will show that our set, whose entries are given by (8), satisfies this condition.

For every $k = 1, 2, \cdots$ the matrix product L must be stochastic, indecomposable, and aperiodic (SIA) since, as pointed out by Thomasian (1963), if L is indecomposable and has period $\nu > 1$, L^ν is decompos-

[3] Indecomposable channels were first discussed by Blackwell *et al.* (1958). See also Wolfowitz (1961) and (1964).

able. A matrix L will be SIA if given any two states (σ', y') and (σ'', y'') there exists a third state (σ, y) which can be reached from either of the two given states in the same number of steps $t + 1$.

The matrix $r(s_{i+1}/s_1)$ defined by (1) is SIA. Hence so is any of its powers. Consider any two given sequences $(s_1', s_0', \cdots, s_{1-m}')$ and $(s_1'', s_0'', \cdots, s_{1-m}'')$ which have a nonzero probability of occurring (i.e. $s_j'(s_j'')$ can be reached from $s_{j-1}'(s_{j-1}'')$ in one step, $j = 2 - m, \cdots, 1$). There is an integer $\alpha \in \{1, 2, \cdots\}$ and a state s^* which can be reached from both s_1' and s_1'' in α steps. Therefore, for each integer $l = \alpha + m, \alpha + m + 1, \cdots$ there is a sequence $(s_1^*, s_0', \cdots, s_{1-m}')$ which can be reached from both given sequences in the same number of steps l.

It follows from (8) that if (σ', y') and (σ'', y'') are possible states of L (i.e., their respective columns have at least one nonzero entry) there will be a state (σ^*, y_1) $((\sigma^*, y_2))$ such that L^t will have a nonzero element $[(\sigma', y'); (\sigma^*, y_1)]$ $([(\sigma'', y''); (\sigma^*, y_2)])$ provided $tk \geqq l$. Since in any L the rows (σ, y), $y = 0, \cdots, h - 1$ are identical for any fixed σ (see (8)) then for some $(\bar{\sigma}, \bar{y})$ the matrix L^{t+1} will have nonzero entries in both positions $[(\sigma', y'); (\bar{\sigma}, \bar{y})]$ and $[(\sigma'', y''); (\bar{\sigma}, \bar{y})]$. QED.

Let M be a finite, say $D \times D$, indecomposable Markov matrix and let ϕ be a function from $\{0, \cdots, D - 1\}$ to $\mathfrak{f}^m$. We say that a source driving a channel K^m of Fig. 5 is governed by a pair (M, ϕ), if it operates as follows: at given time intervals an underlying Markov chain characterized by M changes from some state α to some state β $(\alpha, \beta \in \{0, \cdots, D - 1\})$ and the source then puts out a letter $\phi(\beta)$ and feeds it into the channel. Since M is indecomposable there exists a unique distribution $m(\cdot)$ over the set of states $\{0, \cdots, D - 1\}$ such that if the chain is started in state α with probability $m(\alpha)$, then the probability that the chain will be found at any time i in state β is $m(\beta)$ irrespective of i; $\alpha, \beta \in \{0, \cdots, D - 1\}$. If this is the way M operates, we say that the source (M, ϕ) is stationary. Given any integer t, a source (M, ϕ) has associated with it a probability distribution $Q^t(\cdot)$ over sequences $(f_1, f_2, \cdots, f_t), f_i \in \mathfrak{f}^m$, of source outputs. We say that a distribution $P^t(\cdot)$ over finite sequences $(f_1, \cdots, f_t), f_i \in \mathfrak{f}^m$ is *stationary* if and only if it is associated with some stationary source (M, ϕ).

DEFINITION: Let $\mathcal{V}_s^t$ be the set of all stationary distributions $Q^t(\cdot)$.

DEFINITION: $G(t, m) = \max_{Q \in \mathcal{V}_s^t} R(t, m/Q)$

LEMMA 2. *The strong capacity*[4] *of the channel* K^m *is given by*

$$C(m) = \lim_{t\to\infty} G(t, m) = \sup_t G(t, m) \tag{11}$$

That is, given any ϵ *and* λ, $\epsilon > 0$, $0 < \lambda < 1$, *for n sufficiently large there exists for* K^m *a code* $(n, 2^{n(C(m)-\epsilon)}, \lambda)$[4] *and there does not exist a code* $(n, 2^{n(C(m)+\epsilon)}, \lambda)$.[5]

PROOF: Since K^m is indecomposable the direct and strong converse coding theorems proven by Blackwell, Breiman, and Thomasian (1958) and by Wolfowitz,[6] respectively, apply. Actually, Wolfowitz states the strong converse only for that class of indecomposable channels where the outputs determine the channel state uniquely. The generalization to the full class of indecomposable channels considered by Blackwell *et al.* (1958) (where a not necessarily invertible function ψ is defined which when applied to the channel states $s \in \{0, \cdots, h-1\}$ produces the channel output) is obvious and easy, and we will not bother to spell it out.

With the help of an additional theorem by Wolfowitz (1963b) the capacity expression would normally be given by

$$C^*(m) = \lim_{t\to\infty} \max_{Q^t} \min_{S_1^m} \frac{1}{t} E_{Q^t} \log \frac{V(Y^t/F_1^t, S_1^m)}{P_{Q^t}(Y^t/S_1^m)} \tag{12}$$

where the expectation is, unlike in (6), taken with a fixed S_1^m. Now Theorem 2 of Blackwell *et al.* asserts that for a fixed Q

$$\left| \min_{S_1^m} \frac{1}{t} \log \frac{V(Y^t/F_1^t S_1^m)}{P_Q(Y^t/S_1^m)} - \frac{1}{t} \log \frac{V(Y^t/F^t)}{P_Q(Y^t)} \right| \to 0 \qquad \text{a.e. and } L_1(Q)$$

[4] We will say throughout this paper that a channel has strong (weak) capacity if and only if both the direct and the strong (weak) converse parts of the coding theorem can be proven for it. For discussion of the strong and weak converses see Wolfowitz (1961), section 7.6. The strong converse implies the weak one but not vice versa.

[5] We are using here the notation of Wolfowitz (1961), p. 15. Thus (n, N, λ) is a block code using channel input sequences of length n and accommodating N messages which when sent are each correctly decoded with probability exceeding $1 - \lambda$.

[6] The original version of the proof is in Wolfowitz (1963b), however there is a slight gap in the argument which is fully corrected in the indecomposable channel section of Wolfowitz (1964).

(c.f. (6)). Hence we may replace (11) by[7]

$$\lim_{t\to\infty} \max_{Q^t} \frac{1}{t} E_{Q^t} \log \frac{V(Y^t/F^t)}{P_{Q^t}(Y^t)} \tag{13}$$

Blackwell *et al.* in their Theorem 3 prove this capacity expression with the help of sources (M, ϕ) defined above. They show that the limit $C^*(m)$ is independent of the initial state distribution of the chain M. We are, therefore, free to insist on the stationary distribution. This proves (11). Q. E. D.

V. THE CAPACITY OF MISI CHANNELS

We are now ready to use the results of the preceding section to derive the capacity expression for the MISI channels.

LEMMA 3. *Any code for the channel K^m is a code for the channel K (see Fig. 5); the transmission rate and the probability of error remain the same. Conversely, any code for the channel K is a code for a channel K^μ for some sufficiently large integer μ.*

PROOF: A code for the channel K^m maps integers $\{1, \cdots, N\}$ into sequences of the form $(f_1, f_2, \cdots, f_n)$, $f_i \in f^m$, and maps sequences $(y_1, y_2, \cdots, y_n)$, $y_i \in \{0, \cdots, b-1\}$ into integers $\{1, \cdots, N\}$. Hence it maps integers $\{1, \cdots, N\}$ into sequences

$$(f_1(S_1^m), f_2(S_2^m), \cdots, f_n(S_n^m)) = (x_1, x_2, \cdots, x_n)$$

of input signals depending on the identity of the sequence

$$(S_1^m, S_2^m, \cdots, S_n^m) = (s_{2-m}, \cdots, s_1, \cdots, s_n)$$

of successive states of the channel K. This proves the first assertion, provided that it is assumed that the transmitter, before starting its operation at time *1* has knowledge of the preceding m or more K-channel states. If $m < n$, this is certainly satisfied after the transmission of the first message. Any necessary adjustments in our argument if $m \geqq n$ are trivial.

Now consider any block code for the channel K (other kinds of codes

[7] It is clear that there was nothing to prevent Wolfowitz (1963b) to give the capacity in the form of (13). However, the form of (12) was preferable since in the general case the distribution of the state S_1^m depends on Q. Condition (1) assumes that this is not so in our case, and if it holds, form (12) and especially (11) makes the evaluation of $C(m)$ easier, since the convergence to the limit is faster.

transmitting at constant rate can always be regarded as special cases of block codes). If possible decision errors at the receiver are not to carry over from one message to another, then the signals transmitted during a given block interval may depend, if need be, on some channel states of a preceding block interval, but not on signals transmitted during a preceding block interval. Moreover, even the past channel state dependence, if any, must be finite. Thus there will exist an integer ν such that the code will map a message $i \in \{1, \cdots, N\}$ into a function ψ_i which itself maps state sequences

$$(s_{2-\nu}, \cdots, s_1, \cdots, s_n), s_i \in \{0, \cdots, h-1\}$$

into signal sequences $(x_1, \cdots, x_n), x_i \in \{0, \cdots, a-1\}$. Moreover, the mapping ψ_i must be nonanticipatory, that is, the selection of the signal x_j cannot depend on the identity of any state s_l, $l > j$. Thus it must be possible to write ψ_i as a sequence of functions $g_1^i, g_2^i, \cdots, g_n^i$ where g_k^i maps state sequences $(s_{2-\nu}, \cdots, s_1, \cdots s_k)$ into channel inputs $x_k \in \{0, \cdots, a-1\}$, $k = 1, 2, \cdots n$. It follows, therefore, that such a code for K is also a code for the channel $K^{\nu+n}$, which proves the lemma. Q. E. D.

THEOREM 1. *The strong capacity*[5] *of a Finite Markovian Indecomposable Channel with Side Information at the Transmitter whose operation is defined in (1) and (2) is given by the expression*

$$C = \lim_{m\to\infty} C(m) = \sup_m C(m) \tag{14}$$

That is, given any ϵ *and* λ, $\epsilon > 0$, $0 < \lambda < 1$, *for n sufficiently large there exists a code* $(n, 2^{n(C-\epsilon)}, \lambda)$[4] *and there does not exist a code* $(n, 2^{n(C+\epsilon)}, \lambda)$.[4]

PROOF: It follows directly from Lemmas 1, 2, and 3 that the capacity is given by the expression $\sup_m C(m)$. However, for any positive integers $\mu < \nu$, a code for the channel K^μ can be directly translated into a code for the channel K^ν which would have the same probability of error. Hence $C(\mu) \leqq C(\nu)$, and (14) holds. Q. E. D.

At this point we should like to remark that it is really questionable whether one ought to speak about the capacity of channels with underlying *indecomposable* Markov chains (1). Imbedded in the set of states of such a chain is a smallest closed subset of states.[8] Such a chain then contains one irreducible subchain, and it is the latter only which is used to compute $R(t, m/Q)$ (see (6)) and hence the capacity C. We are

[8] See Feller (1957), p. 349.

dealing here with an asymptotic theory and thus all our results pertain to block lengths large enough to assume essentially stationary operation of the underlying chain for an overwhelming fraction of the transmission process.

Consider two channels K and K', with underlying chains M and M' such that M' is the unique irreducible subchain of the indecomposable chain M. Let the collection of states of M' be denoted by $\mathcal{C}'$ and let the corresponding subset of states of M be denoted by $\mathcal{C}$. If K has the transmission matrices $[p_s(y/x)]$, and K' the matrices $[p_s'(y/x)]$, and if the corresponding states in $\mathcal{C}$ and $\mathcal{C}'$ are given the same label, then we will call K' *the irreducible channel imbedded in* K, provided

$$\begin{aligned} p_s'(y/x) = p_s(y/x) \quad \text{for all} \quad & s \in \mathcal{C} \\ & x \in \{0, \cdots, a-1\} \\ & y \in \{0, \cdots, b-1\} \end{aligned} \tag{15}$$

We then have the

COROLLARY. *The capacities of a Finite Markovian Indecomposable Channel with Side Information at the Transmitter, and of the irreducible channel imbedded in it are identical.*

VI. THE CAPACITY OF FINITE STATE STRONGLY INDECOMPOSABLE CHANNELS WITH SIDE INFORMATION AT THE TRANSMITTER

We found the capacity expression of MISI channels by showing that the associated channels are indecomposable. This brings up the question of whether it would be possible to handle the general Indecomposable Channels (Blackwell *et al.* (1958)) when side information about their state would be available at the transmitter. Such channels are shown schematically in Fig. 6, and their operation is as follows:

The channel consists of inputs $x \in \{0, \cdots, a-1\}$, states $s \in \{0, \cdots, h-1\}$, and of a function ψ which maps the states s onto the channel outputs $y \in \{0, \cdots, b-1\}$, $b \leqq h$. At discrete time intervals $i = 1, 2, \cdots$ the transmitter selects, with knowledge of the present subchannel state s_i, the input x_i. The subchannel then puts out, with probability $w(s_{i+1}/x_i, s_i)$, its next state s_{i+1} which is simultaneously made known to the transmitter and fed into the output transducer. The latter finally puts out the channel output $y_i = \psi(s_{i+1})$ and the next time interval is ready to start.

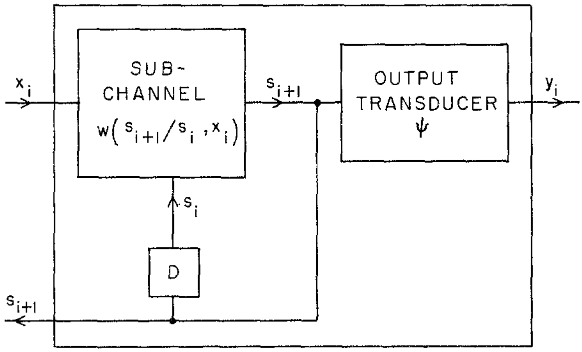

FIG. 6. The general indecomposable channel with side information at the transmitter.

It is clear that the MISI channels can be formulated so that they would appear to be essentially a subclass of the channels of Fig. 6. All one must do is to let states of the present channel be designated by pairs (s_{i+1}, y_i), where s_{i+1} are the states and y_i are the outputs of the MISI channel, and let the function ψ be defined by $\psi(s, y) = y$. Then the present and MISI channels would be completely equivalent except that in the latter the transmitter is told only s and not the pair (s, y). But this is an insignificant detail.

Now a general code for the channel of Fig. 6 would be a mapping of messages into sequences of functions $f \in \mathfrak{f}^m$ for some m, and a mapping of sequences of outputs y onto the set of messages, exactly as described in the proof of Lemma 3. One would again construct an associated channel J^m from the channel of Fig. 6, designated by J, just as K^m of Fig. 5 was constructed from K of Fig. 1. The crucial step then would be to prove an analogue of Lemma 1—with this done the capacity expression would be at hand.

Let $\{G_m(f)\}$, $f \in \mathfrak{f}^m$ be a set of $h^m \times h^m$ matrices whose entries $g_m(\sigma_i ; \sigma_{i+1}/f)$ are given in (16). The numbers σ are defined as in (7).

$$g_m(\sigma_i ; \sigma_{i+1}/f) = \begin{cases} w(s_{i+1}/s_i , f(\sigma_i)) \\ \quad \text{if} \quad \sigma_{i+1} = h\sigma_i - h^m s_{i-m+1} \\ \quad + s_{i+1} \\ 0 \text{ otherwise} \end{cases} \tag{16}$$

The associated channel J^m will be indecomposable if condition (9) is met when the letter D is replaced by G_m. (17)

DEFINITION. The class of Strongly Indecomposable Channels with Side Information (SISI channels) consists of those channels J which meet (17) for all positive integers $m = 1, 2, \cdots$. (18)

The suggested method of finding channel capacity works only for SISI channels and it will be shown in Section VI that not all indecomposable channels J of Fig. 6 are SISI.

The associated channel J^m has a well-defined transmission probability law

$$V'(Y^n/F^n, S_1^m) = \sum_{\substack{S^n \ni \\ [\psi(s_2), \cdots, \psi(s_{n+1})] = Y^n}} \prod_{i=1}^{n} w(s_{i+1}/s_i, f(S_i^m)), \quad (19)$$

and given an input probability distribution Q over sequences $F^n, f \in \mathfrak{f}^m$, there is a conditioned output probability law

$$P_Q'(Y^n/S_1^m) = \sum_{F^n} V'(Y^n/F^n, S_1^m)\, Q(F^n) \quad (20)$$

It then follows from the results of Sections III and IV that Theorem 2 holds.

THEOREM 2. *The capacity* [5]*C of a SISI channel is given by*

$$C = \lim_{m\to\infty} \lim_{n\to\infty} \max_{Q \in \mathfrak{V}_s^n} \min_{S_1^m} E_Q \frac{1}{n} \log \frac{V'(Y^n/F^n, S_1^m)}{P_Q'(Y^n/S_1^m)} \quad (21)$$

VII. COUNTEREXAMPLES AND SUFFICIENT CONDITIONS FOR SISI CHANNEL CHARACTER

In this section we will state two alternate sufficient conditions for SISI character, the first of which is quite an obvious one, while the second, it is reasonable to assert, will in "practice" be met by all SISI channels. We will then introduce an example showing that the second sufficient condition cannot be directly weakened.

Consider the set of a $h^m \times h^m$ matrices $\{\Gamma_m(x)\}$ whose entries $\gamma_m(\sigma_i ; \sigma_{i+1}/x)$ are given by

$$\gamma_m(\sigma_i ; \sigma_{i+1}/x) = \begin{cases} w(s_{i+1}/s_i, x) \\ \quad \text{if } \sigma_{i+1} = h\sigma_i \\ \quad - h^m s_{i-m+1} + s_{i+1} \\ 0 \text{ otherwise} \end{cases} \quad (22)$$

where the numbers σ are defined as in (7). The channel is indecomposable if the set $\{\Gamma_1(x)\}$ satisfies the condition (9), and it is obvious that in that case the set $\{\Gamma_m(x)\}$ will satisfy (9) for all positive integers m.

Using the set $\{\Gamma_m(x)\}$ we can generate a new set of $h^m \times h^m$ matrices $\{\Lambda_m\}$ of size $\exp_a h^m$ as follows:

$$\begin{aligned}&\Lambda_m \in \{\Lambda_m\} \quad \text{if and only if for all} \quad i = 0, 1, \cdots, h^m - 1\\ &\text{the } i\text{th row of } \Lambda_m \text{ is identical with the } i\text{th row of some matrix}\\ &\Gamma_m(x), x \in 0, 1, \cdots, a - 1.\end{aligned} \tag{23}$$

Comparing (16), (22), (23) and the definition of the set $\mathfrak{f}^m$ it follows that the set $\{G_m(f)\}$ is identical with the set $\{\Lambda_m\}$.

To test whether any set of matrices $\{A\}$ guarantees indecomposability of its corresponding channel[3] one proceeds, after Thomasian (1963), to construct the matrix set $\{A'\}$ by taking each matrix of $\{A\}$ and replacing its positive entries by ones, and then retaining only those matrices which are different. The channel is indecomposable if the matrices $\{A'\}$ satisfy condition (9), and Thomasian's test enables us to determine in a finite number of steps whether this is so. We state the following lemma without proof.

LEMMA 4. *The indecomposable channel J is SISI if whenever the corresponding rows of any two matrices $\alpha, \beta \in \{\Gamma_1'(x)\}$ are different, one of them contains no zeros.*

LEMMA 5. *Let the transmission probability distribution of the channel J be such that for every pair $x', x'' \in \{0, \cdots, a - 1\}$ and every state s there exists a state $s^*(s, s^* \in \{0, \cdots, h - 1\})$ such that simultaneously*

$$w(s^*/s, x') \neq 0, \qquad w(s^*/s, x'') \neq 0 \tag{24}$$

Then the necessary and sufficient condition for J to be SISI is that J^1 be indecomposable (i.e. J is indecomposable whenever $\{\Lambda_1\}$ satisfies (9)).

PROOF: Now J is SISI if for all m, J^m is indecomposable. In turn, J^m is indecomposable if the matrix $L = G_m(f_0) \cdots G_m(f_{k-1})$ is indecomposable (see (9)), for any sequence $(f_0, f_1, \cdots, f_{k-1})$, where $f_i \in \mathfrak{f}^m$ and $k = 1, 2, \cdots$, i.e. if L is a matrix of a Markov chain whose any two J^m-states $S^{m'}$ and $S^{m''}$ have a possible common successor state S^{m*}. Suppose that from the J^m-state $S^{m'}$ one can reach a state $(t'_{j-m+1}, \cdots, t_{j-i} = \alpha, \cdots, t_j') = T_j^{m'}$ and from $S^{m''}$ one can reach a state $(t''_{l-m+1}, \cdots, t''_{l-i} = \alpha, \cdots, t_l'') = T_l^{m''}$. Then it follows from the definition of the matrices $\{G_m(f)\}$ and of L that if (24) holds, $S^{m'}$ and $S^{m''}$ will have a possible successor state.

On the basis of the above observation we will prove the lemma by showing that whenever J^m is decomposable so is J^1. Let $S_0^{m'} = (s'_{-m+1}, \cdots, s_0' = \eta)$ and $S_0^{m''} = (s''_{-m+1}, \cdots, s_0'' = \xi)$ be two J^m-states which, relative to L (where the sequence $(f_0, f_1, \cdots, f_{k-1})$ is fixed), have no possible common successor state, and let the periodic sequence $(f_0^*, f_1^*, \cdots)$, $f_{ik+j}^* = f_j$, $f_j \in \mathfrak{f}^m$, $i = 0, 1, \cdots; j = 0, \cdots, k-1$, be an input to J^m. Designate by $\mathcal{S}'$ the set of possible subchannel output sequences $\{(s_1', s_2', \cdots)\}$ (see Fig. 6) when the transducer (see the paragraph preceding (16)) was started in state $S_0^{m'}$, and by $\mathcal{S}''$ the set $(s_1'', s_2'', \cdots)$ when it was started in state $S_0^{m''}$. Designate further by $\mathcal{S}_j'(j \in 0, \cdots, k-1)$ the set of different states s'_{ik+j} $i = 0, 1, \cdots$ occurring in the sequences $\mathcal{S}'$, and by $\mathcal{S}_j''$ the set of different states s''_{ik+j} occurring in the sequences $\mathcal{S}''$. By the remark of the preceding paragraph, and by the assumption about S_0' and S_0'', $\mathcal{S}_j' \cap \mathcal{S}_j''$ must be empty for all $j = 0, \cdots, k-1$. This means that for every pair of sequences $(s_1', \cdots s'_{ik+j}) \in \mathcal{S}'$ and $(s_1'', \cdots, s''_{lk+j}) \in \mathcal{S}''$ the subchannel inputs $x' = f_j(s'_{ik+j}, \cdots, s'_{ik+j-m+1})$ and $x'' = f_j(s''_{lk+j}, \cdots, s''_{lk+j-m+1})$ must be such that

$$\text{if} \quad w(\alpha/s'_{ik+j}, x') > 0 \quad \text{then} \quad w(\alpha/s''_{lk+j}, x'') = 0$$

and

$$\text{if} \quad w(\beta/s''_{lk+j}, x'') > 0 \quad \text{then} \quad w(\beta/s_{ik+j}, x') = 0 \tag{25}$$

where $\alpha, \beta \in \{0, \cdots, h-1\}$. Using the functions $f_0, \cdots, f_{k-1} \in \mathfrak{f}^m$ we will now construct functions $\phi_0, \cdots, \phi_{k-1} \in \mathfrak{f}^1$ such that $L' = G(\phi_0) \cdots G(\phi_{k-1})$ will be decomposable:

Taking all states $\nu \in \mathcal{S}_j'$ in turn select some sequence $(s'_{ik+j-m+1}, \cdots, s'_{ik+j} = \nu) \in \mathcal{S}'$ and let $\phi_j(\nu) = f_j(s'_{ik+j-m+1}, \cdots, s'_{ik+j} = \nu)$.

Also, taking all states $\mu \in \mathcal{S}_j''$ in turn select some sequence

$$(s''_{lk+j-m+1}, \cdots, s''_{lk+j} = \mu) \in \mathcal{S}'' \quad \text{and let}$$
$$\phi_j(\mu) = f_j(s''_{lk+j-m+1}, \cdots, s''_{lk+j} = \mu) \tag{26}$$

Finally for all states $\tau \notin \mathcal{S}_j' \cup \mathcal{S}_j''$, let arbitrarily $\phi_j(\tau) = 0$. The procedure (26) is to be carried out for all $j = 0, \cdots, k-1$, and thus the sequence $\phi_0, \cdots, \phi_{k-1}$ is fully defined. Now consider the last states η and ξ of the two starting J^m-states $S_0^{m'}$ and $S_0^{m''}$, and let the periodic sequence $(\phi_0^*, \phi_1^*, \cdots)$, $\phi_{ik+j}^* = \phi_j$, $i = 0, 1, \cdots, j = 0, \cdots, k-1$, be an input to J^1. Then, designating by Σ' and Σ'' the possible subchannel output sequences when η and ξ, respectively, are the starting states, and defining Σ_j' and Σ_j'' analogously to S_j' and S_j'', it follows from (25) and (26) that $\Sigma_j \subseteq S_j'$ and $\Sigma_j'' \subseteq S_j''$. Hence $\Sigma_j' \cap \Sigma_j''$ is empty for all $j = 0, \cdots, k-1$ and thus J^1 is decomposable, which proves the sufficiency of the condition. The necessity follows directly from (18). Q.E.D.

We will conclude this section by an example of a non-SISI channel whose associated channel J^1 is indecomposable. Thus it will be shown that the condition (24) cannot be dispensed with. Nevertheless, it should be remarked again that (24) will in "practice" always be met.

Consider the binary, three state subchannel of Fig. 6 characterized by the matrices

$$\Gamma'(0) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \qquad \Gamma''(1) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Then the matrices $\{\Lambda_1'\}$ are identical with the matrices $\{\Gamma'(x)\}$, $x = 0, 1$. Using the terminology of Wolfowitz (1963a) the products $\Gamma'(0)\Gamma'(0)$, $\Gamma'(0)\,\Gamma'(1)$, $\Gamma'(1)\Gamma'(0)$, and $\Gamma'(1)\Gamma'(1)$ are all "scrambling" matrices and thus J^1 is indecomposable. However, the set $\{\Lambda_2'\}$ contains the matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

where the row and column numbering scheme (7) was used, and it can be seen that the sets $\{(0, 0), (0, 1), (1, 0)\}$ and $\{(1, 2), (2, 1), (2, 2)\}$ of states (s_{i-1}, s_i) are closed.[8] Thus J^2 is not indecomposable, and therefore J is not SISI.

VIII. CONCLUSION

It is perhaps not necessary to point out that the obtained capacity expressions are not computable as they stand, and that for the MISI channel the trouble is not the limiting procedure (12), but the procedure (14). Equally, for the SISI channel it is the limit with respect to m which causes the difficulty in (21).

One would be tempted to conjecture that for SISI channels which satisfy the condition of Lemma 5 the capacity is obtained with $m = 1$, and, even more strongly, that for all MISI channels $C = C(1)$ (see (14)) and that the optimizing distribution makes the symbols in the sequences F^t independent (see (12)). A forthcoming paper by the author will explore the question of capacity computability for MISI channels. It will be shown that our conjecture is provable if an additional condition is imposed.

Acknowledgment

The author gratefully acknowledges his debt to Professor J. Wolfowitz of Cornell Iniversity who called the author's attention to and made available the as yet unpublished results Wolfowitz (1963a), (1963b), and (1964).

Received: October 4, 1963.

References

Blackwell, D., Breiman, L., and Thomasian, A. J. (1958), Proof of Shannon's transmission theorem for finite-state indecomposable channels. *Ann. Math. Statist.* **29**, No. 4.

Feller, W. (1957), "An Introduction to Probability Theory and Its Applications," Vol. 1, 2nd ed. Wiley, New York.

Jelinek, F. (1963), Loss in information transmission through two-way channels. *Inform. Control* **6**, No. 4.

Shannon, C. E. (1958), Channels with side information at the transmitter. *IBM J* **2**, No. 4, 289–293.

Thomasian, A. J. (1963), A finite criterion for indecomposable channels. *Ann. Math. Statist.* **34**, 337–338.

Wolfowitz, J. (1961), "Coding Theorems of Information Theory." Springer-Verlag, Berlin, and Prentice Hall, Englewood Cliffs, N. J.

Wolfowitz, J. (1963a), Products of indecomposable, aperiodic, stochastic matrices. *Proc. Am. Math. Soc.* **14**, No. 5, 733–737.

Wolfowitz, J. (1963b), The capacity of an indecomposable channel. *Sankhya: Indian J. Statist. Ser. A* **25**, Part I, 101–108.

Wolfowitz, J. (1964), "Coding Theorems of Information Theory," 2nd ed. Springer-Verlag, Berlin and Prentice Hall, Englewood Cliffs, N. J.