INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
INFORMATION THEORETIC ASPECTS OF THE RELIABILITY OF BINARY SWITCHING NETWORKS

by

JOHN PARKAS

Submitted to the Department of Electrical Engineering in partial fulfilment of the requirements for the degree of Master of Science

Department of Electrical Engineering
Faculty of Pure and Applied Science,
The University of Ottawa,
Ottawa, Canada.

1966
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
Approved for the Department
of Electrical Engineering

Supervisor

Chairman of the examining committee

Chairman of the Department
ABSTRACT

The subject of this thesis is the analysis of binary switching networks.

Many schemes for improving the performance of an unreliable binary switching element (gate) have been advanced. Only those relying on the combination of many unreliable gates into a more reliable network — performing the original function — are here considered. Some of the well-known schemes are analysed and reduced to a common format.

This particular format was selected because it facilitated the application of information theoretic concepts to reliability improvement.

Information theoretic concepts, applied under certain restrictions, resulted in the development of a lower bound, LB, on the unreliability (probability of error) of binary switching networks.
ACKNOWLEDGEMENTS

The author wishes to thank Professor G.S. Glinski, Chairman of the Department of Electrical Engineering, who supervised this research project.
INFORMATION THEORETIC ASPECTS OF THE RELIABILITY OF BINARY SWITCHING NETWORKS

1. Introduction
2. Probabilistic Properties of Perfect Gates
3. Probabilistic Properties of Imperfect Gates
4. Networks of Perfect Gates
5. Networks of Imperfect Gates
6. Moore and Shannon Scheme for Increased Reliability
7. von Neumann Scheme for Increased Reliability
8. "Triangular Switching Network" Scheme for Increased Reliability
9. Entropy and Information
10. Information Theoretic Properties of Perfect Gates
11. Information Theoretic Properties of Noisy Gates
12. Information Theoretic Properties of Gate Networks
13. Application of Information Theory to Switching Networks
14. A Lower Bound on $H(Y/X)_{EQUIV}$
15. Conclusion

Appendix I - Convex Functions: Some Properties
Appendix II - Some Properties of $H(x)$
Bibliography

1
1
5
6
10
12
21
26
31
32
33
34
37
42
45
50
53
58
INFORMATION THEORETIC ASPECTS OF THE RELIABILITY OF
BINARY SWITCHING NETWORKS

1. Introduction

A theoretically perfect switching element is, in practice, invariably imperfect or "noisy". A standard method of reducing the probability of error, at this switching element's output, is to replace this single faulty element by a network of faulty elements. Moore and Shannon 12, von Neumann 15, et al. 14, have advanced different schemes using different mathematical approaches to achieve this reduction in error. By utilizing a common language to describe these schemes, using this language to compare a single element with an equivalent net, and then employing the concepts of entropy and information originating in information theory, it is hoped to gain further insight into the reliability of switching functions.

Although attention will be concentrated on the two-input, single-output, binary switching element, hereafter called a gate, much of the following will be equally applicable to the many-input, many-output, binary case. The extension to this more general case will be indicated as it occurs.

2. Probabilistic Properties of Perfect Gates

A perfect gate may be characterized by a function \( f \), such that

\[
f : A \times B \rightarrow C,
\]

where \( A = B = C = \{0, 1\} \).
and A, B are input spaces, 

C is the output space.

It may be noted that this definition of f is also the definition of a Boolean function. A common tabular representation of such a function is the "Canonical Sum of Fundamental Products" table. This table is constructed by listing the elements of the domain A x B, by considering each element as a binary number, and then arranging these numbers in ascending order. To each element of A x B, the function f then assigns a single element of the range C. An equivalent "transition diagram" is obtained by graphically displaying the elements of both domain and range and indicating relations by connecting lines (See Fig. 1, b) and c), for an illustration).

<table>
<thead>
<tr>
<th>A x B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>(0)</td>
</tr>
<tr>
<td>(0,1)</td>
<td>(0)</td>
</tr>
<tr>
<td>(1,0)</td>
<td>(0)</td>
</tr>
<tr>
<td>(1,1)</td>
<td>(1)</td>
</tr>
</tbody>
</table>

3 Equivalent Representations of the Boolean "AND" Function:
a) Gate Symbol, b) Tabular Form, c) Transition Diagram.

Fig. 1

Although, to this point, all definitions have been "deterministic" rather than "probabilistic", a more general case is attained, if the "transition diagram", Fig. 1c), is reconsidered in probabilistic terms as follows:
1) A conditional probability may be defined as the probability of occurrence of a specific output event, given a specific input event. A line between a specific input and output, therefore, indicates a conditional probability of 1. The absence of a line, accordingly, indicates a conditional probability of 0.

In other words, given the occurrence of an input event \((i, j)\), the presence of a line to an output \((k)\) indicates that 
\[
p \left\{ (k) / (i, j) \right\} = 1.
\]
The absence of a line to \(k\), consequently indicates 
\[
p \left\{ (k) / (i, j) \right\} = 0.
\]

2) To each input event \((i, j)\) assign a probability 
\[
p \left\{ (i, j) \right\},
\]
which, for convenience, will be called \(r_{ij}\).

These conventions lead logically to the representation of the transition diagram by the following matrix formula:

\[
Q = RP, \quad \text{(2.2)}
\]

where 
\[
R = \begin{bmatrix} r_{00} & r_{01} & r_{10} & r_{11} \end{bmatrix}
\]
is the input probability distribution vector,

\(P\) is a matrix whose elements are of the form 
\[
p \left\{ (k) / (i, j) \right\},
\]
where \((k)\) designates the column and \((i, j)\) the row. For convenient tabulation, both the row and column designators may be considered to be binary numbers, starting at zero. It can be seen that \(P\) characterizes the gate.
\[ Q = \begin{bmatrix} q_0 & q_1 \end{bmatrix} \] is the output probability distribution vector, with \( q_k \) the probability of occurrence of the output event (k).

Take, as an example of \( P \), the \( P \) of an "AND" gate.

It can be written as

\[
P = \begin{bmatrix}
  p \left\{ (0) \mid (0,0) \right\} & p \left\{ (1) \mid (0,0) \right\} \\
  p \left\{ (0) \mid (0,1) \right\} & p \left\{ (1) \mid (0,1) \right\} \\
  p \left\{ (0) \mid (1,0) \right\} & p \left\{ (1) \mid (1,0) \right\} \\
  p \left\{ (0) \mid (1,1) \right\} & p \left\{ (1) \mid (1,1) \right\}
\end{bmatrix}
\]

Note that the elements of \( P \), for the sake of conciseness, will henceforth be called \( P_{(ij)(k)} \) with \( (k) \) and \( (ij) \) as above. Note also, that for any row of \( P \), \( P_{(ij)(0)} + P_{(ij)(1)} = 1 \).

It should be noted that (2.2) is, by definition, applicable to the \( m \)-input, \( n \)-output switching network. In this case, \( R \) would be a row vector of \( 2^m \) elements, \( Q \) a row vector of \( 2^R \) elements, and \( P \) a \( 2^m \times 2^R \) matrix.

To conclude this section, the 16 possible gates will be listed for later use. The reason for the particular classification of Fig. 2) \(^4\), will also emerge later. For the moment, it is sufficient to note that the gates are arranged in rows with each gate in a row having the same number of lines terminating in an output of 1. For example, all gates in row 3 have 2 lines terminating in a 1.
3. **Probabilistic Properties of Imperfect Gates**

A perfect gate may be said to give the "correct" output when any input \((i,j)\) is applied. An imperfect or noisy gate, given input \((i,j)\) will be said to give rise to the "incorrect" output with a probability \(p_{ij}\).

(2.2), therefore, still applies, but the 1's and 0's of \(P\) are replaced, respectively, by the appropriate 1 - \(p_{ij}\)'s and \(p_{ij}\)'s (see fig. 3 for an illustration). In other words, \(P(ij)(k)\) is now either \(p_{ij}\) or 1\(-\)\(p_{ij}\). For example, in Fig. 3), \(P(01)(0)\) is 1\(-\)\(p_{01}\).
a) Transition Diagram and, b) Matrix Formula for a "Noisy" AND Gate.

Fig. 3.

It should be noted that the probability of error, \( P(e) \), of an imperfect element is given by

\[
P(e) = \frac{1}{2} \sum_{i=0}^{1} \sum_{j=0}^{1} t_{ij} p_{ij}
\]

\[
= R_i^t,
\]

where

\[
R = \begin{bmatrix}
P_{00} & P_{01} & P_{10} & P_{11}
\end{bmatrix}
\]

and superscript "t" indicates the transpose.

4. Networks of Perfect Gates

In all gate reliability improvement schemes that will be studied, a single faulty gate will be replaced by a network. For purposes of comparison with the original gate, this network must be reduced to an equivalent gate. In order to develop a standard method of reduction, attention will be concentrated on a simple type of network, the cascade. The cascade may be defined as a configuration of
gates where the last or output gate is designated the \( n \)th rank, the gate or gates directly feeding the \( n \)th rank are designated the \( n-1 \)th rank, and this process is continued till the source, which feeds the first rank, is reached. Since, by construction, there is no feedback, and any rank may only be fed by elements of the previous rank, time may be neglected, and combinatorial methods may be applied. The reduction method, however, must take cognizance of the fact that the output of a gate in the \( k \)th rank depends, in general, on the outputs of gates in all \( k-1 \) previous ranks.

To develop this reduction method, first consider a cascade of perfect gates. All gates will be designated by the superscript \( rs \), where \( r \) indicates the rank and \( s \) the row, starting from the top. The object, now, is to find a matrix characterization for the cascade. For this cascade to replace a single gate, the cascade's characteristic matrix must be identical with the single gate's matrix. In other words, if the cascade may be reduced to an equivalent gate characterized by \( P_{\text{EQUIV}} \), then (2.2) may be written as

\[
Q = RP_{\text{EQUIV}}. \tag{4.1}
\]

Since, in the two cases, \( R \) is unchanged, and \( Q \) must be unchanged in order for the cascade to be equivalent to the single gate, therefore, \( P_{\text{EQUIV}} \) must equal \( P \).

The reduction method details will be developed in the next section which treats the more general case of cascades composed of noisy gates.
It is possible, however, to point out an interesting aspect of cascades of noiseless gates by consideration of Fig. 4. The single "AND" gate of Fig. 4a) is replaced by the simple, 3-gate cascade of Fig. 4b), which also performs the Boolean "AND" function. Although the cascade leaves the input-output relation of the simple gate unchanged, the input probability distribution to gate 21, an "AND" gate, is no longer R. In other words, if we assume that source R has 2 statistically independent outputs, A and B, such that
\[
\begin{align*}
    r_{00} &= p(A=0, B=0) = p(A=0) \cdot p(B=0) \\
    r_{01} &= p(A=0, B=1) = p(A=0) \cdot p(B=1) \\
    r_{10} &= p(A=1, B=0) = p(A=1) \cdot p(B=0) \\
    r_{11} &= p(A=1, B=1) = p(A=1) \cdot p(B=1)
\end{align*}
\]
then both the single gate and the cascade have an output
\[
Q = \left[\frac{r_{00}}{r_{00} + r_{01} + r_{10} + r_{11}}\right].
\]
Although the "AND" gate of Fig. 4a) has an input distribution of R, the "AND" gate (gate 21) of Fig. 4b) has an input distribution of \(R^{21} = \left[\frac{r_{00}}{r_{00} + r_{01} + r_{10} + r_{11}}\right]\). To sum up, the cascade of perfect gates in Fig. 4b), may be considered to be a means of changing the input probability distribution to an "AND" gate from statistical independence to a particular statistical dependence [7].

In passing, note that in the schematic of the cascade in Fig. 4b), a box with a cross product sign is placed between the first and second ranks. This symbolism is used since, from a set theory standpoint, the set of inputs to gate 21 is the cross product
a) AND Gate

b) An AND Gate Cascade Equivalent

Fig 4)
of the 2 previous output sets (the output sets of gates 11 and 12).

5. Networks of Imperfect Gates

A network reduction method, the Z-matrix method, will be developed in this section. Each gate of the network is characterized by its own \( P \) which corresponds to its own conditional probability of error structure. The Z-matrix method will relate, transform and reduce these \( P \)'s to the \( P_{\text{EQUIV}} \) of (4.1). The superscript \( rs \) on \( P_{\text{EQUIV}} \) will indicate that the network up to, and including gate \( rs \) has been reduced to a \( P_{\text{EQUIV}} \).

In order that the \( P_{\text{EQUIV}} \) may be obtained, a step-by-step reduction, starting at rank 1 and working to the right, must be carried out. As the reduction proceeds to the \( k \)th rank, a matrix characterizing the effect of the previous \( k-1 \) ranks, is required. This matrix, premultiplying the \( P \) matrix of a \( k \)th rank gate, will result in a \( P_{\text{EQUIV}} \). We will call this matrix the Z matrix.

To illustrate the method of obtaining this Z matrix and hence \( P_{\text{EQUIV}} \), consider a noisy network of the type shown in Fig. 4b):

1) Gate 11 is represented by \( P^{11} \) and \( Q^{11} = R^1 P^{11} \) \hspace{1cm} (5.2)
2) Gate 12 is represented by \( P^{12} \) and \( Q^{12} = R^1 P^{12} \) \hspace{1cm} (5.3)
3) Gate 21 is represented by \( P^{21} \) and \( Q^{21} = R^{21} P^{21} \) \hspace{1cm} (5.4)

The problem is to rewrite this last relation in the form

\[ Q^{21} = R Z^{21} P^{21} = R P^{21}_{\text{EQUIV}} \] \hspace{1cm} (5.5)

where \( Z^{21} P^{21} \) is, then the required \( P^{21}_{\text{EQUIV}} \)

and \( Z^{21} \) is the Z matrix sought.
It can be seen that $Z_{21}$ is a $4 \times 4$ matrix, since, by definition, any $R$ is a $1 \times 4$ matrix.

The elements of $Z_{21}$, and their significance, are derived as follows:

1) Since the element $P_{(ij)(k)}^{11}$ of $P_{11}$ indicates the probability of output $(k)$ occurring, source input $(i,j)$ having occurred;

2) Since the element $P_{(ij)(1)}^{12}$ of $P_{12}$ indicates the probability of output $(1)$ occurring, source input $(i,j)$ having occurred;

3) Then the element $z_{(ij)(1)}^{21}$ of $Z_{21}$ indicates the probability of $(k,1)$ occurring at the input to $21$, source input $(i,j)$ having occurred.

i.e. $z_{(ij)(kl)}^{21} = f(P_{ij}(k), P_{ij}(1))$ where the function is determined by the statistical relations between the elements of $P_{11}$ and $P_{12}$. These elements, of course, characterize the conditional probabilities of error of gates $11$ and $12$. As an example, consider the 'Noisy' case of Fig. 4b).

\[
P_{11} = \begin{bmatrix}
P_{00}^{11} & 1 - P_{00}^{11} \\
1 - P_{01}^{11} & P_{01}^{11} \\
1 - P_{10}^{11} & P_{10}^{11} \\
1 - P_{11}^{11} & P_{11}^{11}
\end{bmatrix}, \quad P_{12} = \begin{bmatrix}
P_{00}^{12} & 1 - P_{00}^{12} \\
1 - P_{01}^{12} & P_{01}^{12} \\
1 - P_{10}^{12} & P_{10}^{12} \\
1 - P_{11}^{12} & P_{11}^{12}
\end{bmatrix}
\]

In terms of previous definitions, choose two elements, say, $P_{(01)(0)}^{11} = P_{01}^{11}$, and $P_{(01)(1)}^{12} = P_{01}^{12}$, to be the elements of our
sample calculation. Assume that these errors occur with statistical independence. \( z^{21}(01)(01) \) is then a simple product of these elements, i.e.

\[
z^{21}(01)(01) = p_{01}^{11}(0) \cdot p_{01}^{12}(1) = p_{01}^{11} \cdot p_{01}^{12}
\]

and \( z^{21} = \)

\[
\begin{vmatrix}
(1 - p_{00}^{11})(1 - p_{00}^{12}) & (1 - p_{00}^{11}) & p_{00}^{12} & (1 - p_{00}^{11}) & p_{00}^{12} \\
(1 - p_{01}^{11}) & p_{01}^{12} & (1 - p_{01}^{11})(1 - p_{01}^{12}) & (1 - p_{01}^{11}) & p_{01}^{12} \\
(1 - p_{10}^{11}) & p_{10}^{12} & (1 - p_{10}^{11})(1 - p_{10}^{12}) & (1 - p_{10}^{11}) & p_{10}^{12} \\
(1 - p_{11}^{11}) & p_{11}^{12} & (1 - p_{11}^{11})(1 - p_{11}^{12}) & (1 - p_{11}^{11}) & p_{11}^{12}
\end{vmatrix}
\]

\( Z \) is a matrix representing the effect of all gates prior to the last one. Each component of the source, \( R \), is split into 4 parts by the row of the \( Z \) matrix with the same row index. Each column index designates the input of the last gate on which the source component acts.

6. **Moore and Shannon Scheme for Increased Reliability**

Moore and Shannon outlined a method for reliability improvement of networks in which certain general properties of relays were specified by certain parameters. Iteration techniques based on these parameters were developed and the resulting larger networks were made, accordingly, more reliable [5] [6].

The basic element, the relay, has the following properties:

1) Given an input signal of "0" (i.e. relay coil un-energized) the output will be "0" (i.e. contact open) with a probability
1-c and the output will be "1" (i.e. contact closed), with a probability c.

2) Given an input "1" (i.e. coil energized), the output will be "1" with probability a, and "0" with probability 1-a.

It can be seen that this characterization of a relay as a 1-input, 1-output noisy switching function is consistent with the definitions of section 3.

Consider now 2 perfect relays, the first with input space \( A = \{0,1\} \), and the second with input space \( B = \{0,1\} \), where the events of A and B are statistically independent, i.e.

if \( p(A=1) = r_A \) and therefore \( p(A=0) = 1-r_A \),

and if \( p(B=1) = r_B \) and therefore \( p(B=0) = 1-r_B \),

then the two relays together may be considered to have a source input space \( A \times B \), where

\[
\begin{align*}
p(0,0) &= (1-r_A)(1-r_B) = r_{00}, \\
p(0,1) &= (1-r_A)r_B = r_{01}, \\
p(1,0) &= r_A(1-r_B) = r_{10}, \\
p(1,1) &= r_Ar_B = r_{11}.
\end{align*}
\]

Using this source \( A \times B \), the relays may now be considered to be 2-input, 1-output switching functions. If in the 1-input characterization, the first relay was represented by \( P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \), in the 2-input characterization it is represented by \( P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \). Similarly, if the second relay was represented by \( P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \), it is now represented by \( P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \).
In the manner of section 3, the appropriate $p_{ij}$'s are inserted to give the noisy relay representation.

If the two relays are "ANDED" together, the "ANDING" function is a result of the geometry of the network. If, as in Moore and Shannon, errors are assumed to occur only in the relay contacts, the "AND" gate is, accordingly, noiseless (see Fig. 5).

Fig. 5.

It is apparent that no matter how complex the geometry of the network, if only one source (i.e. 2 inputs, or in functional terms, two arguments) is used, then the equivalent transition diagram will have noisy gates in the first rank only. This, in fact, is characteristic of the Moore and Shannon method.

As an illustration of the Moore and Shannon method, the following example will be worked out in detail. This is done in detail in order to fix the $Z$-matrix method firmly in mind, to provide a basis for comparison with other reliability increasing methods, and to provide a basis for later entropy applications.
a) Original AND    b) Equivalent Cascaded AND
A Moore and Shannon Iteration of an Original "AND" Net
Fig 6)
EXAMPLE

1) The network of Fig. 5a) is iterated by replacing each relay by 4 relays as indicated in Fig. 6b). As before, it is assumed that the error occurrences in the gates are statistically independent. It is also assumed that the relay coils are in parallel.

2) The original network's \( P_{\text{EQUIV}} = Z^{21}p^{21} \) is \( P_{\text{EQUIV}} = \)

\[
\begin{bmatrix}
(1-p_{10})(1-p_{00}) & (1-p_{10})p_{00} & (p_{10})(1-p_{00}) & (p_{10})p_{00} \\
(1-p_{01})(1-p_{12}) & (1-p_{01})p_{12} & (p_{01})(1-p_{12}) & (p_{01})p_{12} \\
(1-p_{10})(1-p_{12}) & (1-p_{10})p_{12} & (p_{10})(1-p_{12}) & (p_{10})p_{12} \\
(1-p_{11})(1-p_{12}) & (1-p_{11})p_{12} & (p_{11})(1-p_{12}) & (p_{11})p_{12}
\end{bmatrix}
\]

where in the \( Z^{21} \) matrix, the following identifications are made

\[
\begin{align*}
p_{10} &= p_{01} = c & p_{10} &= p_{11} = 1-a \\
p_{00} &= p_{12} = c & p_{01} &= p_{12} = 1-a
\end{align*}
\]

i.e.

\[
P_{\text{EQUIV}} = \begin{bmatrix}
1-c^2 & c^2 \\
1-ca & ca \\
1-ca & ca \\
1-a^2 & a^2
\end{bmatrix} \equiv \begin{bmatrix}
1-p_{00} & p_{00} \\
1-p_{01} & p_{01} \\
1-p_{10} & p_{10} \\
p_{11} & 1-p_{11}
\end{bmatrix}
\]

and, therefore, the equivalent matrix has the following conditional probabilities:
\[ p_{00} = c^2 \]
\[ p_{01} = ca \]
\[ p_{10} = ca \]
\[ p_{11} = 1-a^2 \]

3) In the equivalent cascaded "AND"

\[
\begin{align*}
p_{11} &= p_{12} = p_{13} = p_{14} = 1-a \\
p_{10} &= p_{10} = p_{10} = p_{10} = 1-a \\
p_{15} &= p_{16} = p_{17} = 18 \\
p_{15} &= p_{16} = p_{17} = 18 \\
\end{align*}
\]

By construction

\[
\begin{align*}
z_{21}p_{21} &= z_{22}p_{22} \\
z_{23}p_{23} &= z_{24}p_{24} \\
\end{align*}
\]

\[
\begin{array}{cccc}
(1-c)^2 & (1-c)c & c(1-c) & c^2 \\
(1-c)^2 & (1-c)c & c(1-c) & c^2 \\
(1-a)^2 & (1-a)a & a(1-a) & a^2 \\
(1-a)^2 & (1-a)a & a(1-a) & a^2 \\
\end{array}
\]

\[
\begin{array}{c}
1 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
\end{array}
\]
\[
\begin{bmatrix}
1-c^2 & c^2 \\
1-c^2 & c^2 \\
1-a^2 & a^2 \\
1-a^2 & a^2 \\
\end{bmatrix}
= \begin{bmatrix}
1-p_{00} & p_{00} \\
1-p_{01} & p_{01} \\
p_{10} & 1-p_{10} \\
p_{11} & 1-p_{11} \\
\end{bmatrix}
\equiv p_{21}^{\text{EQUIV}} = p_{22}^{\text{EQUIV}}.
\]

i.e.
\[
\begin{align*}
p_{00}^{21} &= p_{01}^{21} = c^2, & p_{10}^{21} &= p_{11}^{21} = 1-a^2, \\
p_{00}^{22} &= p_{01}^{22} = c^2, & p_{10}^{22} &= p_{11}^{22} = 1-a^2.
\end{align*}
\]

In a similar fashion it can be shown that

\[
p_{23}^{\text{EQUIV}} = p_{24}^{\text{EQUIV}} = \begin{bmatrix}
1-c^2 & c^2 \\
1-a^2 & a^2 \\
1-c^2 & c^2 \\
1-a^2 & a^2 \\
\end{bmatrix}
\equiv \begin{bmatrix}
1-p_{00} & p_{00} \\
1-p_{01} & p_{01} \\
p_{10} & 1-p_{10} \\
p_{11} & 1-p_{11} \\
\end{bmatrix}
\]

i.e.
\[
\begin{align*}
p_{00}^{23} &= p_{10}^{23} = c^2, & p_{10}^{23} &= p_{01}^{23} = 1-a^2, \\
p_{00}^{24} &= p_{10}^{24} = c^2, & p_{01}^{24} &= p_{11}^{24} = 1-a^2.
\end{align*}
\]

Repetition of this procedure for gates of rank 3, yields:

\[
2^{31}p_{31} = \begin{bmatrix}
(1-c^2)^2 & (1-c^2)^2 & c^2(1-c^2) & c^4 & 1 & 0 \\
(1-c^2)^2 & (1-c^2)c^2 & c^2(1-c^2) & c^4 & 0 & 1 \\
(1-a^2)^2 & (1-a^2)a^2 & a^2(1-a^2) & a^4 & 0 & 1 \\
(1-a^2)^2 & (1-a^2)a^2 & a^2(1-a^2) & a^4 & 0 & 1 \\
\end{bmatrix}
\]
\[
\begin{align*}
&= \begin{bmatrix}
1-(2c^2-c^4) & 2c^2-c^4 \\
1-(2c^2-c^4) & 2c^2-c^4 \\
(l-a^2)^2 & 1-(l-a^2)^2 \\
(l-a^2)^2 & 1-(l-a^2)^2 \\
\end{bmatrix} \\
\text{i.e.} & \quad p_{00}^{31} = p_{01}^{31} = 2c^2-c^4 \\
&\quad p_{10}^{31} = p_{11}^{31} = (l-a^2)^2 \\
&\quad z_{32}^{32} = \begin{bmatrix}
(l-c^2)^2 & (1-c^2)^2 & c^2(l-c^2) & c^4 \\
(l-a^2)^2 & (1-a^2)^2 & a^2(l-a^2) & a^4 \\
(l-c^2)^2 & (1-c^2)^2 & c^2(l-c^2) & c^4 \\
(l-a^2)^2 & (1-a^2)^2 & a^2(l-a^2) & a^4 \\
\end{bmatrix} \\
&\quad 1-(2c^2-c^4) \\
&\quad (l-a^2)^2 \\
&\quad 1-(l-a^2)^2 \\
&\quad 2c^2-c^4 \\
&\quad 1-(l-a^2)^2 \\
\end{align*}
\]

\[
\begin{align*}
&= \begin{bmatrix}
1-(2c^2-c^4) & 2c^2-c^4 \\
1-(2c^2-c^4) & 2c^2-c^4 \\
(l-a^2)^2 & 1-(l-a^2)^2 \\
(l-a^2)^2 & 1-(l-a^2)^2 \\
\end{bmatrix} \\
\text{i.e.} & \quad p_{00}^{32} = p_{10}^{32} = 2c^2-c^4 \\
&\quad p_{01}^{32} = p_{11}^{32} = (l-a^2)^2 \\
\text{Finally} & \quad z_{41}^{41} = \begin{bmatrix}
(l-c^2)^4 & (l-c^2)^2(2c^2-c^4) & (2c^2-c^4)(1-c^2)^2 & (2c^2-c^4)^2 \\
(l-c^2)^2(1-a^2) & (l-c^2)^2(2a^2-a^4) & (2c^2-c^4)(1-a^2)^2 & (2c^2-c^4)(2a^2-a^4) \\
(l-a^2)^2(1-c^2)^2 & (l-a^2)^2(2c^2-c^4) & (2a^2-a^4)(1-c^2)^2 & (2a^2-a^4)(2c^2-c^4) \\
(l-a^2)^4 & (l-a^2)^2(2a^2-a^4) & (2a^2-a^4)(1-a^2)^2 & (2a^2-a^4)^2 \\
\end{bmatrix} \begin{bmatrix}
1 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
\end{bmatrix}
\end{align*}
\]
\[
\begin{align*}
1-(2c^2-c^4)^2 & \quad (2c^2-c^4)^2 \\
1-(2c^2-c^4)(2a^2-a^4) & \quad (2c^2-c^4)(2a^2-a^4) \\
1-(2c^2-c^4)(2a^2-a^4) & \quad (2c^2-c^4)(2a^2-a^4) \\
1-(2a^2-a^4)^2 & \quad (2a^2-a^4)^2 \\
\end{align*}
\]

Identifying the \( p_{ij} \) of the equivalent cascaded "AND" leads to

\[
\begin{align*}
p_{00} &= (2c^2-c^4)^2 \\
p_{01} &= (2c^2-c^4)(2a^2-a^4) \\
p_{10} &= (2c^2-c^4)(2a^2-a^4) \\
p_{11} &= 1-(2a^2-a^4)^2 \\
\end{align*}
\]

4) The calculations are completed by comparing the \( p_{ij} \)'s of the original with the equivalent cascade.

It can be seen that the \( c \) of the original is replaced by \( 2c^2-c^4 \) in the equivalent cascade, and that the \( a \) of the original is replaced by \( 2a^2-a^4 \) in the equivalent cascade. This means that the relay network replacing the original relay network has \( p_{ij} \)'s less than the original relay's \( p_{ij} \)'s if \( c < .618 \) and \( a > .618 \). Under these conditions, therefore, the original A and B relays have been improved. If, in addition, \( ca \) of the original is more than the equivalent network's \( ca \), then

\[
P(e)_{\text{EQUIV}} < P(e)_{\text{ORIGINAL}} \text{ for all } R.
\]

While the approach presented here (in contrast with Moore and Shannon's original approach) is cumbersome and only indicates whether
probability of error has been improved, not how to improve this
reliability, it does provide a language. This language is easily
adaptable to an information theoretic approach developed in sub-
sequent sections. In general, other methods of increasing reliability
can easily be represented in terms of this language. In the next
two sections, where they will be treated more fully, we will observe
that:

1) Von Neumann's scheme can be shown to have
the $p_{ij}$'s $\neq 0$ in all ranks except the last.
2) Recursive triangular networks - although the
errors are defined in terms of probabilities
of the gates carrying out not the design
function but some other functions - can be
transformed into $p_{ij}$ representation, with
some or all $p_{ij}$'s $\neq 0$.

7. Von Neumann Scheme for Increased Reliability [15]

Using the format developed in sections 5 and 6, the
von Neumann scheme for increasing reliability of a Sheffer stroke
organ (gate 15 of Fig. 2) is represented graphically in Fig. 7.
Statistical independence of error occurrence is again assumed.

The first rank of gates represents the $N$ lines of
bundle $A$, $N$ of which are stimulated, and the $N$ lines of bundle $B$, $N$
of which are stimulated.

From section 5, it follows that,

$$P_{\text{EQUIV}} = 2^2 \cdot p^{21}$$

(5.5)
von Neumann Iteration of an Original Sheffer Stroke Organ

Fig. 7
\[ P_{\text{EQUIV}}^{21} = \begin{bmatrix}
\left(1-\xi\right)(1-\gamma) & \left(1-\xi\right)(\gamma) & \left(\xi\right)(1-\gamma) & \left(\xi\right)(\gamma) \\
\left(1-\xi\right)(1-\gamma) & \left(1-\xi\right)(\gamma) & \left(\xi\right)(1-\gamma) & \left(\xi\right)(\gamma) \\
\left(1-\xi\right)(1-\gamma) & \left(1-\xi\right)(\gamma) & \left(\xi\right)(1-\gamma) & \left(\xi\right)(\gamma) \\
\left(1-\xi\right)(1-\gamma) & \left(1-\xi\right)(\gamma) & \left(\xi\right)(1-\gamma) & \left(\xi\right)(\gamma) \\
\end{bmatrix} \in [1-\varepsilon, 1] \]

where \(i = 1, 2, \ldots, N\).

It should be noted that in von Neumann's exposition, \(0 \leq \xi, \gamma \leq 1\). The stipulation that \(p_{ij} \leq 0.5\), would seem to be inconsistent with the range of \(\xi\) and \(\gamma\). This apparent inconsistency is resolved if we associate \(p_{ij}^{1A}\) with either \(\xi\), or \(1-\xi\), whichever is less than 0.5.

Similar considerations apply to \(p_{ij}^{1B}\). This has the effect of making the function of the 1st rank gates dependent on the values of \(\xi\) and \(\gamma\).

The form of \(P_{\text{EQUIV}}^{21}\), however, is always

\[ P_{\text{EQUIV}}^{21} = \begin{bmatrix}
1-\mathcal{J} & \mathcal{J} \\
1-\mathcal{J} & \mathcal{J} \\
1-\mathcal{J} & \mathcal{J} \\
1-\mathcal{J} & \mathcal{J} \\
\end{bmatrix} \quad (7.2) \]

where \(\mathcal{J} = \left(1-\varepsilon\right)(1-\xi)(\gamma) + (\varepsilon)(\xi)(\gamma)\).

Each of the \(N\) lines emanating from the executive organs is split into two and the resulting \(2N\) lines are randomly connected to the restoring organs 1. With a sufficiently large \(N\), this has the effect of normalizing the error input to each executive organ around the mean \(\mathcal{J}\) with a very small dispersion from the mean. A large \(N\), by the central limit theorem, also makes the elements of the \(P_{\text{EQUIV}}^{21}\) statistically independent.

We can now, accordingly, write
\[
P_{\text{EQUIV}}^{34} = \begin{bmatrix}
1 - \omega & \omega \\
1 - \omega & \omega \\
1 - \omega & \omega \\
1 - \omega & \omega
\end{bmatrix}
\]

where
\[
\omega = (1 - \varepsilon)(1 - \varphi^2) + (\varphi^2).
\]

(7.5)

Repeating the process for the next rank of gates results in
\[
P_{\text{EQUIV}}^{44} = \begin{bmatrix}
1 - \psi & \psi \\
1 - \psi & \psi \\
1 - \psi & \psi \\
1 - \psi & \psi
\end{bmatrix}
\]

where
\[
\psi = (1 - \varepsilon)(1 - \omega^2) + (\varepsilon)(\omega^2).
\]

(7.6)

The final stage consists of setting a fiduciary level \( \Delta \), such that; if \( \Delta N \) or less of the \( N \) outputs of rank 4 fire (i.e. are 1 in our terminology) the system is said not to fire (i.e. to be 0); if \( (1 - \Delta)N \) or more of the lines fire, the system is said to fire; if neither of these conditions hold the system is said to be in error.

The final stage may be represented by a perfect N-line-input, 1-line-output, binary input, ternary output switching function, which in terms of the graphical model of section 5 would have the equation
\[
\begin{pmatrix}
(1-\psi)^N, (1-\psi)^{N-1}(\psi), \ldots, (1-\psi)(\psi)^{N-1}, \\
(1-\psi)^N, (1-\psi)^{N-1}(\psi), \ldots, (1-\psi)(\psi)^{N-1}, \\
(1-\psi)^N, (1-\psi)^{N-1}(\psi), \ldots, (1-\psi)(\psi)^{N-1}, \\
(1-\psi)^N, (1-\psi)^{N-1}(\psi), \ldots, (1-\psi)(\psi)^{N-1},
\end{pmatrix}
\begin{bmatrix}
100 \\
100 \\
\ldots \\
010 \\
\ldots \\
001 \\
001
\end{bmatrix}
\]

In explanation of (7.8): 1) The inputs to the last switching function may be considered to be binary numbers \(N\) digits long (i.e. there are \(2^N\) such inputs). 2) If there are \(\Delta N\) or less 1's in the input the output is a "0". 3) If there are \((1-\Delta)N\) or more 1's in this number the output is a 1. 4) If the number of 1's in the input number lies in between, the output is in error (F).

It can be seen that
\[
P_{\text{EQUIV}} = \sum_{j=0}^{\Delta N} \binom{N}{j}(\psi)^j(1-\omega)^{N-j} + \sum_{j=\Delta N+1}^{(1-\Delta)N-1} \binom{N}{j}(\psi)^j(1-\omega)^{N-j} + \sum_{j=(1-\Delta)N}^{N} \binom{N}{j}(\psi)^j(1-\omega)^{N-j}
\]

It can be seen that the system does behave as a Sheffer stroke,
1. If system operates perfectly \((\varepsilon = 0)\) then:
   a) If \(\xi = \eta = 1\) i.e. input to system is always 1, then output is 0
   b) If \(\xi = 0, \eta = 1\) then output is 0
   c) If \(\xi = 1, \eta = 0\) then output is 0
   d) If \(\xi = 0, \eta = 0\) then output is 0

2. If system operates with small Sheffer stroke organ error \((\varepsilon \approx 0)\) and argument error is kept within fiduciary levels, then:
   a) If \(\xi > (1-\Delta)\), \(\eta > (1-\Delta)\) then, \(\psi \approx 0\) and the output is most likely 0
   b) If \(\xi \geq (1-\Delta)\), \(\eta \geq (1-\Delta)\) then output is 0
   c) If \(\xi \leq (1-\Delta)\), \(\eta \leq (1-\Delta)\) then output is 0
   d) If \(\xi \leq \Delta\), \(\eta \leq \Delta\) then output is 0


In order to improve the reliability of a single gate, a method of iteration (or recursion) is developed and the resulting cascade is called a "Triangular Switching Network". The original gate is replaced by a homogeneous triangular network. A homogeneous network is one constructed of identical gates. Triangular means that the network is a cascade with 2 gates in the first rank and one in the second.

Each new gate is, in turn, replaced by a homogeneous triangular network. This process is repeated a requisite \(N\) times till the desired triangular switching network is achieved. It should be noted that the original single gate \((=3^0)\) is replaced by 3 gates \((=3^1)\) in 2 ranks \((2^1)\). The next, or second, iteration results in 9 gates \((=3^2)\) in 4 ranks \((=2^2)\). After \(N\) iterations there are a total of \(3^N\)
gates in $2^N$ ranks.

The noise, or error, in a gate is viewed from a different standpoint than the one characterized by conditional probabilities of error. It is assumed that a gate may perform switching functions other than its design function with stated probabilities. Certain physical considerations lead this analysis to the restriction of possible switching functions to the 6 positive functions (a positive function is a Boolean function which, in the sum of products form, contains no negated variable). These 6 functions are called $F$, $C$, $A$, $B$, $D$, and $T$. In terms of our classification of gates in Fig. 2, they are respectively gates 1, 5, 6, 7, 12, and 16. Any gate is assumed to perform these functions with the respective probabilities of $F_o$, $C_o$, $A_o$, $B_o$, $D_o$, and $T_o$. These probabilities are assumed statistically independent.

Although this error formulation is different, it has, obviously, an equivalent conditional probability of error representation. The relations between the two forms can be developed as follows:

1) $F_o + C_o + A_o + B_o + D_o + T_o = 1$

2) Assume an input distribution $R$, where
   as usual $\sum_{i,j} r_{ij} = 1$

3) Given an input $ij$, the gate, depending on which of the 6 functions it is performing, will give an output of either 0 or 1. Group the terms giving a zero
output in one category, those giving a one output in another.

i.e.

\[
\begin{align*}
P(\text{Output}=0) &= r_{00}^OF_0 + r_{01}^OF_0 + r_{10}^OF_0 + r_{11}^OF_0 \\
&+ r_{00}^C_0 + r_{01}^C_0 + r_{10}^C_0 \\
&+ r_{00}^A_0 + r_{01}^A_0 \\
&+ r_{00}^B_0 + r_{10}^B_0 \\
&+ r_{00}^D_0 \\
&= r_{00}(1-T_0) + r_{01}(F_0C_0 + A_0) + r_{10}(F_0 + C_0 + B_0) + r_{11}(F_0)
\end{align*}
\]

\[
\begin{align*}
P(\text{output}=1) &= \\
&+ r_{11}^C_0 \\
&+ r_{10}^A_0 + r_{11}^A_0 \\
&+ r_{01}^B_0 + r_{11}^B_0 \\
&+ r_{01}^D_0 + r_{10}^D_0 + r_{11}^D_0 \\
&= r_{00}(T_0) + r_{01}(T_0 + D_0 + B_0) + r_{10}(T_0 + D_0 + A_0) + r_{11}(1-1-F_0)
\end{align*}
\]

4) Putting the above in the matrix form of the conditional probability of error representation leads to

\[
P = \begin{bmatrix}
1-T_0 & T_0 \\
F_0 + C_0 + A_0 & T_0 + D_0 + B_0 \\
F_0 + C_0 + B_0 & T_0 + D_0 + A_0 \\
F_0 & 1-F_0
\end{bmatrix}
\]

5) The \( p_{ij} \)'s may, now, be identified when the design function is specified. If, for example, the gate is designed to be an AND gate (C), then,
\[
P = \begin{bmatrix}
1-P_{00} & P_{00} \\
1-P_{01} & P_{01} \\
1-P_{10} & P_{10} \\
P_{11} & 1-P_{11}
\end{bmatrix}, \quad \text{and}
\]
\[
P_{00}=T_0, \quad P_{01}=T_0+D_0+B_0, \quad P_{10}=T_0+D_0+A_0, \quad \text{and} \quad P_{11}=F_0.
\]

Having established the relations between the 2 approaches, let us revert to the analysis \[2\] under discussion. The iteration is developed in terms of the 6 functions, i.e. if the single gate probabilities are \(F_0, C_0, A_0, B_0, D_0,\) and \(T_0,\) then the first iteration yields, \(F_1, C_1, A_1, B_1, D_1,\) and \(T_1\) the respective probabilities of the network functioning as \(F, C, A, B, D\) and \(T.\) The network probabilities are, of course, in terms of the gate probabilities.

It is possible, by the methods of section 5, to arrive at the same recursion formulas obtained in this analysis. First assuming that because of physical symmetry, \(A_0=B_0,\) it can be seen that

\[
P_{\text{EQUIV}} = \begin{bmatrix}
(1-T_0)^2 & (1-T_0)T_0 & (T_0(1-T_0)) & T_0^2 \\
(F_0+C_0+A_0)^2 & (F_0+C_0+A_0)(T_0+D_0+A_0) & (T_0+D_0+A_0)(F_0+C_0+A_0) & (T_0+D_0+A_0) \\
F_0^2 & F_0(1-F_0) & (1-F_0)F_0 & (1-F_0)^2
\end{bmatrix}
\]
\[
\begin{bmatrix}
1-T_1 & T_1 \\
F_1 + C_1 + A_1 & T_1 + D_1 + A_1 \\
F_1 & T_1
\end{bmatrix},
\]

where
\[
T_1 = T_0 \left\{ (1-T_0)^2 + 2(1-T_0)(T_0+D_0+A_0) + T_0(1-F_0) \right\}
\]
\[
= T_0 \left\{ 1+2D_0+2D_0+2A_0+T_0(-T_0+1-F_0-2D_0-2A_0) \right\}
\]
\[
= T_0 \left\{ 1+2D_0+2A_0+T_0(C_0-D_0) \right\},
\]

and similarly,
\[
F_1 = F_0 \left\{ 1+2C_0+2A_0+F_0(D_0-C_0) \right\}
\]
\[
A_1 = A_0 \left\{ 2A_0+A_0(C_0+D_0)+4C_0 D_0+2C_0 T_0+2D_0F_0 \right\}
\]
\[
D_1 = D_0 \left\{ 2A_0+2A_0+2C_0 T_0+D_0(2+C_0-2T_0) - D_0^2 \right\}
\]
\[
C_1 = C_0 \left\{ 2A_0+2A_0+2D_0F_0+C_0(2+D_0-2F_0) - C_0^2 \right\}.
\]

It can be seen that for \( N \) iterations the recursion formulas above remain of the same form but the subscripts are all enlarged by \( N \). As an example
\[
F_{N+1} = F_N \left\{ 1+2C_N+2A_N+F_N(D_N-C_N) \right\}.
\]

This analysis yielded several interesting results, among which are the following:

1) If \( F_0+T_0 \neq 0 \), if \( (C_0+D_0) \neq 0 \), then \( \lim_{N \to \infty} (C_N+D_N) = 1 \)

2) Given \( F_0=T_0=0 \), if \( (C_0+D_0) \neq 0 \), then \( \lim_{N \to \infty} (C_N+D_N) = 1 \).
In particular, if \( C_0 \approx D_0 \), then the network converges to \( C \),

\[
D_0 \approx C_0 \approx D_0
\]

\[
C_0 = D_0, \text{ then } \lim_{N \to \infty} C_N = \lim_{N \to \infty} D_N = 1/2.
\]

3) If \( C_0 + A_0 + B_0 \) (= \( C_0 + 2A_0 \)) = 1, then if the desired network function is \( C \), an iteration improves the network.

4) If \( C_0 + F_0 = 1 \), then if the desired network function is \( C \), an iteration degrades the network.

Having shown that various reliability improvement schemes can, without great difficulty, be represented by the \( Z \) matrix format, we now consider the application of entropy to this format.

9. Entropy and Information:

The mathematical concept of entropy is a useful way of showing the relation of an input set \( X = \{x_1, x_2, \ldots, x_n\} \), where each event \( x_k \) has a stationary probability of occurrence \( p(x_k) \), to an output set \( Y = \{y_1, y_2, y_m\} \), each of whose events \( y_i \) has a stationary probability of occurrence \( p(y_i) \). The two sets are related by a set of probabilities of joint occurrences \( p(x_k, y_i) \). In all cases to be considered, the statistical relations \( p(x_k, y_i) = p(x_k/y_i)p(y_i) = p(y_i/x_k)p(x_k) \), will, of course, hold.

The entropy function is a convex function (see appendix I) of the type \(-\sum \log Z_j \) (where \( \log Z_j \) will be understood to designate \( \log_2 Z_j \) throughout).
There are many excellent texts, which may be consulted for a detailed analysis of entropy. To relate entropy to gates, however, it is sufficient, for our purposes, to state only the basic equation, i.e.

\[ I(X;Y) = H(X) - H(X/Y) = H(Y) - H(Y/X) \]  \hfill (9.1)

where

\[ I(X;Y), \text{ the transinformation, or mutual information, is} \]

defined as

\[ I(X;Y) = \sum_{X} \sum_{Y} p(x_k, y_1) \log \frac{p(x_k)}{p(y_1)} , \]

\[ H(X), \text{ the entropy, or average self information of} X, \]

is defined as

\[ H(X) = -\sum_{X} p(x_k) \log p(x_k), \]

\[ H(Y), \text{ the entropy of} Y, \text{ is defined as} \]

\[ H(Y) = -\sum_{Y} p(y_1) \log p(y_1), \]

\[ H(X/Y), \text{ is, therefore, defined as} \]

\[ H(X/Y) = -\sum_{X} \sum_{Y} p(x_k, y_1) \log p(x_k/y_1), \]

\[ H(Y/X), \text{ is, accordingly, defined as} \]

\[ H(Y/X) = -\sum_{X} \sum_{Y} p(x_k, y_1) \log p(y_1/x_k). \]

10. **Information Theoretic Properties of Perfect Gates**

To apply the foregoing entropy definitions to the gate characterizations of section 2 requires the terminology of section 2 be given its equivalent name. A x B becomes X, C becomes Y, and (2.1) becomes

\[ f: X \rightarrow Y \]  \hfill (10.1)

For a perfect gate, the function performed is a mapping (i.e. for every \( x_k \) there is one and only one \( y_1 \)). The conditional proba-
bilities $p(y_i/x_k)$ are, therefore, equal to either 0 or 1. This means, by
definition, that $H(Y/X) = 0$. This leads to equation (9.1) having the form

$$I(X;Y) = H(X) - H(X/Y) = H(Y) - 0. \quad (10.2)$$

(10.2) may be interpreted as follows: The average
amount of information at the input $H(X)$ is not all available at the output since the average amount of information there is $H(Y)$, i.e. an average amount of information $H(X/Y)$ called the equivocation (measuring the average confusion about $X$, knowing $Y$) is lost in the process.

The terms of (10.2) are computed as follows:

$$H(X) = -R \log R^t, \quad (10.3)$$
$$H(Y) = -Q \log Q^t, \quad (10.4)$$

Where superscript "t" indicates the matrix transpose,

and $R$ and $Q$ are defined before in section 2.

11. Information Theoretic Properties of Noisy Gates

The introduction of noise into the gate leads to

$H(Y/X) \neq 0$, by definition. This average amount of information $H(Y/X)$ is called the noise or error entropy and measures the average confusion about $Y$, knowing $X$. i.e. knowing $X$, $Y$ can no longer be determined with certainty.

(9.1) may be rewritten as:

$$H(X) + H(Y/X) = H(Y) + H(X/Y) = H(X,Y), \quad (11.1)$$

where the terms are computed as follows:

1) $H(X)$ is unchanged from the perfect case of (10.3),

2) $H(Y/X) = RK^t, \quad (11.2)$
where

\[ K = \left[ H(p_{00}), H(p_{01}), H(p_{10}), H(p_{11}) \right] \tag{11.3} \]

and \( H(p_{ij}) = -\{ p_{ij} \log p_{ij} + (1-p_{ij}) \log(1-p_{ij}) \}. \tag{11.4} \]

Note that a probability, such as \( p_{ij} \), appearing in brackets after \( H \) will indicate an entropy of the form of (11.3).

3) \( H(Y) \) retains the form of (10.4), but, of course, the value of \( Q \) will be changed.

12. **Information Theoretic Properties of Gate Networks**

When a noisy gate is replaced by an equivalent gate cascade, then the input source, \( R \), is unchanged, hence \( H(X) \) is unchanged, but the reliability is improved if

\[ P(e)_{\text{ORIGINAL}} = \sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} P(i)(\text{ORIGINAL}) \] \[ P(e)_{\text{EQUIV}} = \sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} P(i)(\text{EQUIV}). \tag{12.1} \]

It could be profitable to relate \( P(e) \) to an entropy term, \( H(Y/X) \), under certain restrictions, is such an entropy. A justification, for the previous statement, is developed on the following 6 steps:

1) Attention will be restricted to sources whose 2 outputs, \( A \) and \( B \), are statistically independent,

\[ i.e. \ R = \left[ (1-r_A)(1-r_B), (1-r_A)r_B, r_A(1-r_B), r_Ar_B \right]. \]

This restriction is much less a constraint on the general case than it would appear to be on first sight. Many cases of sources with statistically dependent outputs may, as in section 4, be represented by a statistically independent source followed by a rank, or ranks, of perfect gates. These ranks of perfect gates may be considered as additions to the front end of the existing cascade. Analytic methods,
Graphical Representation of $H(Y|X)$ and $P(e)$

Fig. 8
developed in subsequent sections, will be equally applicable to the new cascade.

2) \( P(e) = \sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} p_{ij} \), and therefore, from above,

\[
P(e) = (1-r_A)(1-r_B)p_{00} + (1-r_A)r_Bp_{01} + r_A(1-r_B)p_{10} + r_Ar_Bp_{11}.
\]  

(12.2)

3) Consider \( r_A \) and \( r_B \) to be the \( x \) and \( y \) axes of a 3-dimensional cartesian coordinate system. Along any fixed \( r_A, Z = P(e) \) is a linear function of \( r_B \) (i.e. a line). Along any fixed \( r_B, Z = P(e) \) is a linear function of \( r_A \) (i.e. a line). \( Z = P(e) \) is a surface which at the four corners under consideration \((0,0), (0,1), (1,0), (1,1)\) is \( p_{00}, p_{01}, p_{10}, \) and \( p_{11} \) respectively (see Fig. 8).

4) From Fig. 9, the relation between \( p_{ij} \) and \( H(p_{ij}) \), it is apparent that

\[
\sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} H(p_{ij}) = H(Y/X) \geq \sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} p_{ij} = P(e)
\]

(12.3)

when \( p_{ij} < .773 \)

Relation between \( H(p_{ij}) \) and \( p_{ij} \)

Fig. 9
5) Restricting $p_{ij}$ to the domain $0 \leq p_{ij} \leq 0.5$, both $p_{ij}$ and $H(p_{ij})$ are monotonically increasing with $p_{ij}$, and therefore, the surface $z = H(Y/X)$ has its minimum at the same corner as the surface $z = F(e)$.

6) With restriction $0 \leq p_{ij} \leq 0.5$, $F(e)_{\text{ORIG.}} \geq F(e)_{\text{EQUIV}}$ for all $R$, if and only if

$$H(Y/X)_{\text{ORIG.}} \geq H(Y/X)_{\text{EQUIV}}.$$  \hfill (12.4)

The stipulation that $p_{ij}$ be less than, or equal to, 0.5 is not unreasonable. A gate is customarily designed to perform its design function over 50% of the time, regardless of input. If the hardware doesn't conform to this specification, the design function may be renamed. As an example, consider an "OR" gate with $p_{11} = 0.8$ and all other $p_{ij} \leq 0.5$. This gate performs as an "EXCLUSIVE OR" gate within specifications.

Since $F(e)$ has been related to $H(Y/X)$, the following final sections will be devoted to obtaining the $H(Y/X)_{\text{EQUIV}}$ of a network by considering the entropies of the component gates, and to obtaining a lower bound for $H(Y/X)_{\text{EQUIV}}$.

13. Application of Information Theory to Switching Networks

To recapitulate, a format for the representation of gate networks and some mathematical relations from information theory applicable to gates have thus far been developed. How can these, most usefully, be applied?

The most general approach is that of Cowan and Winograd [3].
Their argument, in simplified form, is as follows:

1) In developing information theory, Shannon considered the case of communication through a simple noisy channel. Given a message input set $X$ and a simple noisy channel, fully specified by the $p(y_j/x_k)$'s, he defined a channel capacity $C$. This $C = \max_{X} I(X;Y)$, is obtained by varying the symbol transmission probabilities of the input set. $C$ is an upper bound on the average amount of information that can be provided by each received symbol about the corresponding transmitted signal. Shannon showed that if $H(X) \leq C$, then there exists at least one code such that a suitably encoded message may be transmitted over the channel and recovered with an arbitrarily small frequency of identification error. Coding may be considered to be the replacement of a set of symbols by another, usually larger, set according to a fixed scheme. In other words, given a fixed channel with capacity $C$, and a message ensemble $X$, such that $H(X) \leq C$, a coder may be found to match the message sequences to the noisy channel so that a decoder at the other end of the channel will decode the received messages with an arbitrarily small probability of error.

Smallness of error is obtained at the expense of complexity at the coder and decoder and time delay in decoding.

2) Cowan and Winograd, in an analogous manner, define a computation capacity $C^*$ for a module (a complex switching network) and show that, under certain restrictions, an arbitrarily reliable network, composed of modules all having computation capacity $C^*$, may be designed. These results are obtained by considering the coder-channel-
decoder trilogy to be incorporated in every module. Every increase in
network reliability entails more modules and more complex modules.

While this approach is of prime importance in formu-
la
ting a general theory of reliable computation, it is possible that
information theory concepts may also be used in smaller and more specific
areas of reliable computation, such as obtaining $H(Y/X)_{\text{EQUIV}}$, or a
lower bound on it.

Why a lower bound on $H(Y/X)_{\text{EQUIV}}$? In section 12 it
was noted that $H(Y/X)_{\text{EQUIV}}$ directly reflects $P(e)$. By finding such a
lower bound, therefore, a lower limit is set on $P(e)$.

In order to attempt a reduction in the arbitrariness
of the choice, the gate network must be re-examined. A network of gates
may be considered to be a computation channel. Unlike the simple com-
munication channel, the computation channel may be considered to be
not fixed, but variable, depending on the exact configuration of the
constituent gates. Synthesis of a more reliable gate network, accord-
ingly, may be considered to be the synthesis of a better computation
channel.

Whereas in the simple communication case, the fixed
characteristics of a channel result in a single statistical figure of
merit $C$, which is then used in obtaining an optimal code, in a compu-
tation channel we may reasonably consider that only the constituent gates
have fixed characteristics. The figure of merit of the network synthe-
sized from these gates would depend on the configuration of the gates
and vary with the varying arrangement of these gates. By analogy with
the communication case, then, a likely approach is to define some figure(s) of merit for the individual gates; obtain an optimal figure for the equivalent gate (the synthesized computation channel); finally, it is to be hoped that in the process of obtaining this optimal figure there will be indication of how to arrange (code) the gates in order to achieve this optimum.

In section 5, it was seen that a gate network may be explicitly reduced to an equivalent gate by what was designated the Z-matrix approach. It can, therefore, be seen that once a gate network is designed, its theoretical performance can be directly verified by use of Z-matrix techniques. No statistical (entropy) techniques need be used. Statistical techniques, however, may be useful, given particular gates, in giving lower bounds on network error towards which coding may aim. It may also conceivably be used to gain better insight into all types of switching networks, sequential (with feedback) as well as combinatorial.

Starting, then, with an individual gate, it can be seen that the matrix K (its characterization in the entropy field) is not a substantial economy over the probability matrix P. Any attempt to develop an exact analogue of the Z-matrix method in the entropy field would lead to a method nearly as laborious as the Z-matrix. Even if the attempt were successful, the result would be a mere duplication, giving no additional insights into networks.

If the figure(s) of merit characterizing a gate are to be simpler (or at least different) than the K matrix, a re-examination of
the Z-matrix method is called for. These new figure(s) must, after all, be at least based on the Z-matrix.

The crucial fact to note is that the sum of the 4 elements of any row of the Z matrix is 1. Taking as a point of departure the equation

$$Z^{rs}P^{rs} = P^{rs}_{EQUIV}$$

which characterizes the Z-matrix method, this reduction method may be reinterpreted in the following terms:

1. Each gate $rs$, of a network takes 4 input distributions (the $Z^{rs}$ matrix) and transforms them into 4 output distributions (the $P^{rs}_{EQUIV}$ matrix).

2. The output of gate rs and another gate, say rt, then feed into a gate of the next rank, say $(r+1)u$. In other words, $P^{rs}_{EQUIV}$ and $P^{rt}_{EQUIV}$ are related to each other and together form the new Z matrix, $Z^{(r+1)u}$, which feeds gate $(r+1)u$.

3. The process is repeated till the last gate is reached. The last gate then transforms its Z into the required $P_{EQUIV}$ of the network.

To sum up, the Z-matrix method, instead of being regarded as the repeated multiplication of conditional probability matrices, may be regarded as the repeated transformation of input distributions into output distributions by a cascade of gates.

It must be noted that at every step of the Z-matrix procedure the output distribution may, in terms of entropy, be represented by the appropriate $K$ of (11.3). Since $H(Y/X) = RK^T$ (11.2), it can
immediately be seen that in order to reduce \( H(Y/X) \) it is sufficient to reduce all 4 elements of \( K^t \). If, in turn, a lower bound is found on each element of \( K^t \); the least of these then chosen and transmitted through the net, a lower bound will have been established on the smallest element of \( K^t_{EQUIV} \), hence on \( K^t_{EQUIV} \) and hence on \( H(Y/X)_{EQUIV} \).

Such a lower bound will be developed in the next section.

14. A Lower Bound on \( H(Y/X)_{EQUIV} \)

There is no loss in generality if in developing the required lower bound, the terminology of section 2 is used. i.e.,

\[
\begin{bmatrix}
  r_{00} & r_{01} & r_{10} & r_{11} \\
  p(00)(0) & p(00)(1) \\
  p(01)(0) & p(01)(1) \\
  p(10)(0) & p(10)(1) \\
  p(11)(0) & p(11)(1)
\end{bmatrix}
= \begin{bmatrix}
  q_0 & q_1
\end{bmatrix}
\]

(2.2)

or \( RP = Q \), where \( R \) now represents a row \( kl \) of \( Z^{RS} \), \( P \) represents the gate, and \( Q \) represents a row \( kl \) of \( \text{PR}^{EQUIV} \).

By Theorem 2 of Appendix I, it can be seen that

\[
\sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} p_{ij}(0) \log p_{ij}(0) \geq q_0 \log q_0
\]

(14.1)

and that

\[
\sum_{i=0}^{1} \sum_{j=0}^{1} r_{ij} p_{ij}(1) \log p_{ij}(1) \geq q_1 \log q_1
\]

(14.2)
Adding (14.1) and (14.2) leads to

\[
\sum_{i=0}^{l} \sum_{j=0}^{l} r_{ij} H(p_{ij}) \leq H(q_1) \tag{14.3}
\]

If the left-hand side of (14.3) is defined as \( \alpha \), it can be seen that \( \alpha \) is a lower bound on \( H(q_1) \). It is, however, unsatisfactory for at least two reasons:

1. It is composed of a mixture of entropy and probability terms.

2. If \( P \) is noiseless, i.e. all \( p_{ij} = 0 \), then \( \alpha = 0 \).

Stated another way, \( \alpha \) is far from being a greatest lower bound which would, of course, be most desirable. In the Moore and Shannon case, for example, \( \alpha \) would equal zero after the first rank. It would accordingly be of little use.

In order to develop a lower bound based on an approach such as the one embodied in (14.3) we will make the 3 following simplifying assumptions:

1. Let \( q_1 \leq 0.5 \). This is done purely for convenience and consistency. Whether \( q_1 \leq 0.5 \) or \( q_0 \leq 0.5 \) is irrelevant in the same sense that this is a problem in coding and not of lower bounds on entropies (cf. step 6 of section 12, for an analogous case).

2. In any gate there is both a maximum and a minimum \( p_{ij} \) (though they may be equal). Call these respectively \( p_{\text{max}} \) and \( p_{\text{min}} \).

3. For a perfect gate (all \( p_{ij} = 0 \)), let \( q_{1p} \) designate \( q_1 \) and let \( \Delta q_1 \) be the difference between the actual \( q_1 \) and the perfect \( q_1 \) (i.e. \( q_{1p} \)). In other words \( q_1 = q_{1p} + \Delta q_1 \). \( \tag{14.4} \)
On the basis of assumption 2, we can define a lower bound on \( \Delta q \), as
\[
\Delta q_{\text{min}}^\Delta = (1-q_{1p})p_{\text{min}}q_{1p}p_{\text{max}}^* \tag{14.5}
\]
It now follows that if we define \( q_{\text{1min}} \) as
\[
q_{\text{1min}}^\Delta = q_{1p} + q_{\text{1min}}
\]
then, by definition, \( q_{\text{1min}} \leq q_{1} \tag{14.7} \)
and hence \( H(q_{\text{1min}}) \leq H(q_{1}) \) \tag{14.8}.

\( H(q_{\text{1min}}) \) is, accordingly, another lower bound on \( H(q_{1}) \).

As a final step, a lower bound on \( H(q_{1}) \), call it \( H(LB) \), must be found. \( H(LB) \) should satisfy the following properties:

1. \( H(LB) \) should be less than \( H(q_{\text{1min}}) \). This is still a lower bound on \( H(q_{1}) \) and is done so that the \( p_{ij} \)'s on which \( H(LB) \) will be based will be reduced from 4 to 2 \( (p_{\text{min}} \) and \( p_{\text{max}} \)) for each gate.

2. \( H(LB) \) must be greater than \( H(p_{\text{min}}) \). This is evident from (14.3). It follows that because
\[
\sum_{i=0}^{1} \sum_{j=0}^{I} r_{ij} H(p_{\text{min}}) = H(p_{\text{min}})
\]
and because
\[
\sum_{i=0}^{1} \sum_{j=0}^{I} r_{ij} H(p_{ij}) = \alpha, \text{ therefore, } H(p_{\text{min}}) \leq \alpha. \tag{14.9}
\]
This is, in entropy terms, the well-known fact that the error out of a gate is at least as great as the minimum error in the gate.

3. \( H(LB) \) should increase with increasing \( p_{\text{min}} \).

4. When the \( p_{ij} \)'s all equal zero, \( H(LB) \) should equal \( H(q_{1p}) \).

\[
H(LB) = H(p_{\text{min}}) + H(q_{1p}) \frac{H(p_{\text{min}})}{H(p_{\text{min}} + p_{\text{max}})} \tag{14.10}
\]
satisfies the above 4 properties under the following qualifications.

1. \( H(LB) \leq H(q_{\text{1min}}) \) only under the empirically derived
restriction that $p_{\text{min}} < 0.02$. This area of small error is, however, the
most commonly occurring and hence most interesting, as has been noted by
Zemanek.\[1\]

A justification may be attempted in the following steps:

a. With a given $p_{\text{min}}$, it may be noted that

$$\frac{H(\text{LB})}{H(p_{\text{min}})} \rightarrow \text{Maximum, while } H(q_{\text{1min}}) \text{ increases only slightly.}$$

b. We are, therefore, interested in the limits on $p_{\text{min}}$,
such that $H(q_{\text{1p}})\frac{H(p_{\text{min}})}{H(2p_{\text{min}})} \leq H(q_{\text{1min}}) - H(p_{\text{min}}),$
or, by a. above $H(q_{\text{1p}})\frac{H(p_{\text{min}})}{H(2p_{\text{min}})} = H(q_{\text{1p}}[1-2p_{\text{min}}]+p_{\text{min}}) - H(p_{\text{min}})$ \hspace{1cm} (14.11).

c. Since $q_{\text{1p}} \leq 0.5$ by construction, therefore, $2q_{\text{1p}}p_{\text{min}} < p_{\text{min}}$, and

hence $H(q_{\text{1p}}) \leq H(q_{\text{1min}})$ \hspace{1cm} (14.12).

d. This satisfies the prerequisites of Theorem 2, Appendix 2, and
we may accordingly impose a relaxed upper bound on $\frac{H(p_{\text{min}})}{H(2p_{\text{min}})}$ of

$$\frac{H(p_{\text{min}})}{H(2p_{\text{min}})} \leq (1-2p_{\text{min}})q_{\text{1p}}$$

or

$$q_{\text{1p}} \leq H(2p_{\text{min}})$$

This inequality is easily solved graphically ($p_{\text{min}} \leq 0.1$)
and indicates an upper bound on $p_{\text{min}}$, which is determined to be about 0.02
(See Table 1).

15 Conclusion

Converting $H(\text{LB})$ into $\text{LB}$ by means of entropy tables results
in a lower bound on the probability of error out of any specific gate.
A problem remains in compounding two $H(\text{LB})$'s to obtain the $H(q_{\text{1p}})$
of the next gate. In the case of Moore and Shannon, and von Neumann where errors are made statistically independent, a solution in the entropy field is to approximate $H(LB_1 \cdot LB_2)$, the minimum $H(q_{1p})$, by $H(LB_1)H(LB_2)$, (cf. Theorem 1, Appendix 2).

This problem of compounding $q_{1p}$ in the case of statistically dependent errors and, ultimately, in multi-input, multi-output sequential circuits is a matter for further inquiry. So, unfortunately, is the answer to the overriding problem, i.e., are there coding methods inherent in some form of $H(LB)$?

In the meantime, $H(LB)$ may be used to indicate a lower bound on error in a network. This may be done directly in some simple cases and with conversion from entropy to probability between gates in the case of more complex cascades.

To conclude:

1. Some well-known gate-reliability-improvement methods have been presented in a single format (the $Z$ matrix method) which facilitates the application of entropy concepts.

2. A lower bound, $H(LB)$, was synthesized. When converted back into probability, it gave a lower bound on the conditional probabilities of error of the gate equivalent to a network composed of specified gates.

3. This lower bound, while not a greatest lower bound, does, at least, make the derivation of a greatest lower bound a possibility.
4. In the meanwhile, it is to be hoped that even $H(LB)$ may be of use in further analysis of the more complex networks mentioned - i.e. sequential circuits with feedback, having component gates with non-statistically independent probabilities, and possibly many non-binary inputs and outputs.
### TABLE 1

Some Comparisons of $H(\text{LB})$ and $H(\text{q}_{1\text{min}})$

<table>
<thead>
<tr>
<th>$p_{\text{min}}$</th>
<th>$H(p_{\text{min}})$</th>
<th>$q_{1\text{p}}$</th>
<th>$H(q_{1\text{p}})$</th>
<th>$H(q_{1\text{min}})$</th>
<th>$H(\text{LB})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{max}}$</td>
<td>$p_{\text{min}} + p_{\text{max}}$</td>
<td>$\frac{H(p_{\text{min}} + p_{\text{max}})}{H(p_{\text{min}})}$</td>
<td>.03</td>
<td>.081</td>
<td>.2395</td>
</tr>
<tr>
<td>.06</td>
<td>.05</td>
<td>.291</td>
<td>.395</td>
<td>.367</td>
<td></td>
</tr>
<tr>
<td>.327</td>
<td>.594</td>
<td>.971</td>
<td>.9785</td>
<td>.771</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p_{\text{min}}$</th>
<th>$H(p_{\text{min}})$</th>
<th>$q_{1\text{p}}$</th>
<th>$H(q_{1\text{p}})$</th>
<th>$H(q_{1\text{min}})$</th>
<th>$H(\text{LB})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{max}}$</td>
<td>$p_{\text{min}} + p_{\text{max}}$</td>
<td>$\frac{H(p_{\text{min}} + p_{\text{max}})}{H(p_{\text{min}})}$</td>
<td>.02</td>
<td>.081</td>
<td>.192</td>
</tr>
<tr>
<td>.04</td>
<td>.05</td>
<td>.291</td>
<td>.358</td>
<td>.311</td>
<td></td>
</tr>
<tr>
<td>.242</td>
<td>.583</td>
<td>.971</td>
<td>.961</td>
<td>.711</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p_{\text{min}}$</th>
<th>$H(p_{\text{min}})$</th>
<th>$q_{1\text{p}}$</th>
<th>$H(q_{1\text{p}})$</th>
<th>$H(q_{1\text{min}})$</th>
<th>$H(\text{LB})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{max}}$</td>
<td>$p_{\text{min}} + p_{\text{max}}$</td>
<td>$\frac{H(p_{\text{min}} + p_{\text{max}})}{H(p_{\text{min}})}$</td>
<td>.08</td>
<td>.081</td>
<td>.189</td>
</tr>
<tr>
<td>.100</td>
<td>.05</td>
<td>.291</td>
<td>.347</td>
<td>.225</td>
<td></td>
</tr>
<tr>
<td>.469</td>
<td>.301</td>
<td>.971</td>
<td>.958</td>
<td>.433</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p_{\text{min}}$</th>
<th>$H(p_{\text{min}})$</th>
<th>$q_{1\text{p}}$</th>
<th>$H(q_{1\text{p}})$</th>
<th>$H(q_{1\text{min}})$</th>
<th>$H(\text{LB})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{max}}$</td>
<td>$p_{\text{min}} + p_{\text{max}}$</td>
<td>$\frac{H(p_{\text{min}} + p_{\text{max}})}{H(p_{\text{min}})}$</td>
<td>.02</td>
<td>.081</td>
<td>.140</td>
</tr>
<tr>
<td>.02</td>
<td>.05</td>
<td>.291</td>
<td>.323</td>
<td>.249</td>
<td></td>
</tr>
<tr>
<td>.141</td>
<td>.575</td>
<td>.971</td>
<td>.972</td>
<td>.640</td>
<td></td>
</tr>
</tbody>
</table>
\[
\begin{array}{|c|c|c|c|}
\hline
p_{\text{min}} = .001 & H(p_{\text{min}}) = .011 & q_{1p} & H(q_{1p}) & H(q_{1\text{min}}) & H(LB) \\
p_{\text{max}} = .001 & p_{\text{min}} + p_{\text{max}} = .002 & .001 & .011 & .021 & .016 \\
H(p_{\text{min}} + p_{\text{max}}) = .021 & .01 & .081 & .087 & .066 \\
H(p_{\text{min}}) & H(p_{\text{min}} + p_{\text{max}}) = .549 & \\
\hline
\end{array}
\]
Appendix I

Convex Functions: Some Properties [8] [9] [13]

Definition: \( f(x) \) is convex in \((a,b)\) if \( \frac{d^2 f(x)}{dx^2} \geq 0 \) in \((a,b)\).

Theorem 1: If \( f(x) \) is convex in \((a,b)\), then for every \( x_1, x_2 \) in \((a,b)\) such that \( x_1 < x_2 \), \( \alpha f(x_1) + (1-\alpha)f(x_2) \geq f(\alpha x_1 + (1-\alpha)x_2) \) for \( 0 \leq \alpha \leq 1 \).

Proof: Define an \( x_\alpha \) such that \( x_1 < x_\alpha < x_2 \).

It follows that \( 0 \leq \frac{x_\alpha - x_1}{x_2 - x_1} \leq 1 \), accordingly, define \( 1-\alpha \), and \( \alpha \) as follows:

\[
\frac{x_\alpha - x_1}{x_2 - x_1} = 1-\alpha, \quad \text{and} \quad \frac{x_2 - x_\alpha}{x_2 - x_1} = \alpha.
\]

Now define \( q(x_\alpha) \) as the straight line joining \( f(x_1) \) to \( f(x_2) \)

i.e. \( q(x_\alpha) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot (x_\alpha - x_1). \) \( (I1) \)

By definition of convexity, \( \frac{df(x)}{dx} \) increases monotonically with increasing \( x \), and hence \( f(x) \) may be represented by an upward opening curve such as that of Fig. A. It may, hence, be added that \( q(x_\alpha) \geq f(x_\alpha) \). \( (I2) \)
But, from (I1),
\[ q(x_a) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} (x_a - x_1), \]
and rearranging the terms
\[ q(x_a) = \frac{x_2 - x_a}{x_2 - x_1} f(x_1) + \frac{x_a - x_1}{x_2 - x_1} f(x_2). \]  
(13)

Substituting the definitions of \( a \) and \( 1 - a \) leads to
\[ q(x_a) = af(x_1) + (1-a)f(x_2). \]  
(14)

Note, that by decomposition,
\[ x_a = \frac{x_2 - x_a}{x_2 - x_1} x_1 + \frac{x_a - x_1}{x_2 - x_1} x_2, \]  
(15)

and, again, by definition of \( a \) and \( 1 - a \),
\[ x_a = ax_1 + (1-a) x_2, \]  
(16)

de yields
\[ f(x_a) = f(ax_1 + (1-a) x_2). \]  
(17)

Substituting (14) and (17) in (I2), obtain
\[ af(x_1) + (1-a)f(x_2) \geq f(ax_1 + (1-a) x_2). \]  
(18)

Theorem 2: If \( \sum_{i=1}^{n} q_i = 1 \) and \( q_i \geq 0 \), and if \( Q(x) \) is convex, then
\[ \sum_{i=1}^{n} q_i Q(x_i) \geq Q \left( \sum_{i=1}^{n} q_i x_i \right). \]  

Proof: Assume that the theorem is true for \( \sum_{i=1}^{n-1} \), and prove it true for
\[ \sum_{i=1}^{n} \] by induction.
\[Q\left(\sum_{i=1}^{n} q_i x_i\right) = Q\left(q_1 x_1 + \left[\frac{q_2 x_2 + \cdots + q_n x_n}{q_2 + \cdots + q_n}\right]\right)\]

\[\leq q_1 Q(x_1) + \left[\frac{q_1 + \cdots + q_n}{q_2 + \cdots + q_n}\right] Q\left(\frac{q_2 x_2 + \cdots + q_n x_n}{q_2 + \cdots + q_n}\right), \quad (I9)\]

by Theorem 1.

But, by the assumption above,

\[Q\left(\frac{q_2 x_2 + \cdots + q_n x_n}{q_2 + \cdots + q_n}\right) \leq \frac{q_2}{q_2 + \cdots + q_n} Q(x_2) + \cdots + \frac{q_n}{q_2 + \cdots + q_n} Q(x_n), \quad (I10)\]

i.e., substituting (I10) in (I9) yields

\[Q\left(\sum_{i=1}^{n} q_i x_i\right) \leq \sum_{i=1}^{n} q_i Q(x_i). \quad (I11)\]

But the assumption is true for \(n-1=2\) (a restatement of Theorem 1), hence the theorem is true for all \(n\).

Note: \(\ln x \leq x-1, \quad (I12)\)

and, hence, \(\log x (= \ln x \log e) \leq (x-1) \log e. \quad (I13)\)

This relation (I12) may be derived as follows:

1. Call \(\ln x\), \(f(x)\).
2. Define a tangent to \(f(x)\) at \(x=x_0\) as \(q(x)\).
3. Since \(-f(x)\) is convex, \(q(x) \geq f(x)\), for any \(x). \quad (I14)\)
4. Let \(x_0=1\), then \(q(x)=x-1\), and substituting for \(f(x)\) and \(q(x)\) in (I14),

\[\ln x \leq x-1 \quad \text{ (See Fig. B)}\]

Fig. B
Appendix II

Some Properties of $H(x)$

Given, in this appendix, that $0 \leq b \leq a \leq 0.5$, it follows:

Lemma 1. (i) $a \leq -(1-a) \log(1-a)$

(ii) $-(1-a) \log(1-a) \leq -a \log a$

(iii) $-a \log a \leq H(a)$

Proof:

(1) The function $x \log x$ is convex; hence, by Theorem 1 of Appendix 1, for any $x$ between $x = .5$ and $x = 1$, the function is less than the line joining these two points. The equation of this line is $x - 1$.

In other words, for $0.5 \leq x \leq 1$, $x \log x \leq x - 1$. (III)

Let $x = 1 - a$, and substitute in (III),

i.e. $a \leq -(1-a) \log(1-a)$.

(II) By reference to Figure C, note the following:

a. In the range 0 to 1, the convex function $-\log x$ has a slope of minimum magnitude at $x = 1$, hence so has the function $\log x$ in this range.

b. $\log x = \frac{\ln x}{\ln e} = 1.44 \ln x$, (II2)

hence, the slope at $x = 1$ is $1.44 \frac{d \ln x}{dx} \bigg|_{x = 1} = 1.44$. (II3)

Through $\log 1$ and $\log(1-a)$ draw a straight line, and call it $y_2$. The slope of $y_2$ is, by definition, greater than $y_1$.

Observe that

$$\frac{a}{1-a} = \frac{-y_2 \text{ at } x = 1-a}{-y_2 \text{ at } x = a}.$$ (II4)
Now, by convexity of $-\log x$,

$$-y_2 \text{ at } x = a \leq -\log a,$$  \hspace{1cm} (II5)

and, by construction,

$$-y_2 \text{ at } x = 1-a = -\log(1-a),$$  \hspace{1cm} (II6)

hence, substituting (II5) and (II6) in (II4) yields

$$\frac{a}{1-a} \geq \frac{-\log(1-a)}{-\log a},$$  \hspace{1cm} (II7)

or

$$-(1-a)\log(1-a) \leq -a\log a.$$  

(III) Since $H(a)$ is by definition

$$H(a) = -[a\log a + (1-a)\log(1-a)],$$

and both right-hand terms are greater than zero, it follows that

$$-a\log a \leq H(a).$$

Theorem 1: (1) $ab \leq bH(a)$

(II) $bH(a) \leq aH(b)$

(III) $aH(b) \leq H(ab)$

(IV) $H(ab) \leq H(a)H(b)$.

Proof: (1) By lemma 1, $a \leq H(a)$, hence, $ab \leq bH(a)$.

(II) An equivalent statement is $b \leq \frac{H(b)}{H(a)}$.  \hspace{1cm} (II8)
It can be seen (by reference to Figure D) that if a line joining the
origin and \( H(x) \) at \( x = b \), call it the function \( y_1 \), is constructed, then,
\[
\frac{b}{a} = \frac{y_1 \text{ at } b}{y_1 \text{ at } a}.
\]  \hspace{1cm} (II9)

But, by construction,
\[
y_1 \text{ at } b = H(b)
\]  \hspace{1cm} (II10)

and, by convexity,
\[
y_1 \text{ at } a > H(a).
\]  \hspace{1cm} (II11)

Substituting (II10) and (II11) in (II9) yields
\[
\frac{b}{a} \leq \frac{H(b)}{H(a)}.
\]

(II11) By (II) above, \( \frac{H(ab)}{H(b)} \geq \frac{ab}{b} = a \),
\[
\frac{b}{a} \leq \frac{H(b)}{H(a)}.
\]  \hspace{1cm} (II12)

hence, \( aH(b) \leq H(ab) \).

(IV) Make the following definitions:
\[
-(1-x) \log(1-x) \equiv x + \Delta x',
\]  \hspace{1cm} (II13)

and 
\[
-x \log x \equiv x + \Delta x.
\]  \hspace{1cm} (II14)

By lemma 1, \( \Delta x \geq \Delta x' > 0 \).
\[
(II15)
\]

\( H(a)H(b) \) may now be rewritten as
\[
H(a)H(b) = \frac{(a+\Delta a) + (a+\Delta a')}{(b+\Delta b) + (b+\Delta b')} \cdot \frac{(b+\Delta b) + (b+\Delta b')}{2a+\Delta a+\Delta a'} \cdot \frac{2b+\Delta b+\Delta b'}{4ab+2b\Delta a+2a\Delta b+\Sigma},
\]  \hspace{1cm} (II16)

where \( \Sigma = 2a\Delta b' + 2b\Delta a' + \Delta a\Delta b + \Delta a'\Delta b + \Delta a\Delta b' + \Delta a'\Delta b' \), and therefore \( \Sigma \geq 0 \).

Now, \(-ab \log a + ab \log b\)
\[
= b(a+\Delta a) + a(b+\Delta b)
\]  \hspace{1cm} (II17)

\[
= 2ab + a\Delta b + b\Delta a.
\]  \hspace{1cm} (II18)

Note that \(-ab \log a > -(1-ab) \log(1-ab) \) by lemma 1.
It follows that 

\[ H(ab) = -\left[ a \log ab + (1-ab) \log(1-ab) \right] \]

\[ \leq -2ab \log ab \]

\[ = 4ab + 2b\Delta a + 2a\Delta b. \]  

(II19)

Comparing (II19) with (II16), it can be seen that even an upper bound on \( H(ab) \) is less than \( H(a)H(b) \), i.e.,

\[ H(ab) \leq H(a)H(b). \]

Theorem 2:

Given \( 0 \leq d \leq c \leq b \leq a \leq 0.5 \)

\[ \frac{H(a)-H(c)}{H(b)-H(d)} \leq \frac{a-c}{b-d} \]

Proof: Since \(-H(x)\) is convex, the slope of \( H(x) \) decreases with \( x \).

By the Mean Value Theorem the slope of the line joining \( H(a) \) and \( H(b) \), has a slope less than that of the tangent at \( b \). Similarly, the slope of the line joining \( H(c) \) and \( H(d) \) is greater than that of the tangent at \( c \).

But, by convexity, \( H'(x) \) at \( c \geq H'(x) \) at \( b \), hence,

\[ \frac{H(a)-H(b)}{a-b} \leq \frac{H(c)-H(d)}{c-d} \]  

(See Fig. E)  

(II20)

Fig. E
Consider now, \( \frac{H(x) - H(d)}{x-d} \), where \( d \leq x \leq a \),  \hspace{1cm} (\text{II}21) \\
and \( \frac{H(a) - H(x)}{a-x} \).  \hspace{1cm} (\text{II}22)

Observe that  
\( (\text{II}21) \gg \frac{H(a) - H(d)}{a-d} \), and  
\( (\text{II}22) \ll \frac{H(a) - H(d)}{a-d} \).

Hence, substituting \( x = c \) in (\text{II}21) and \( x = b \) in (\text{II}22) leads to  
\( \frac{H(a) - H(c)}{H(b) - H(d)} \ll \frac{a-c}{b-d} \).
References

1. S. AMAREL and J. A. BRZOZOWSKI
   "Reliability of Triangular Switching Networks"
   Redundancy Techniques for Computing Systems
   R. H. Wilcox and W. C. Mann Eds.

2. J. A. BRZOZOWSKI
   "Reliability of "Triangular" Switching Networks with Intermittent Failures"

3. J. D. COWAN and S. WINOGRAD
   "Reliable Computation in the Presence of Noise"

4. P. ELIAS
   "Computation in the Presence of Noise"

5. ESSARY and PROSCHAN
   "Reliability of Coherent Systems"
   Redundancy Techniques for Computing Systems
   R. H. Wilcox and W. C. Mann Eds.

6. R. M. FANO
   "Transmission of Information"

7. A. GILL
   "Synthesis of Probability Transformers"

8. HARDY, LITTLEWOOD, and POLYA
   "Inequalities"
   Cambridge University Press, 1934.

9. A. I. KHINCHIN
   "Mathematical Foundations of Information Theory"
   Dover, New York, 1957.

10. S. KULLBACK
    "Information Theory and Statistics"
11. R. S. LEDLEY
"Digital Computer and Control Engineering"

12. E. F. MOORE and C. E. SHANNON
"Reliable Circuits Using Less Reliable Relays"

13. F. M. REZA
"An Introduction to Information Theory"

14. J. TOOLEY
"Network Coding for Reliability"

15. J. VON NEUMANN
"Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components"
Automata Studies
C. E. Shannon and J. McCarthy Eds.

16. H. ZEMANEK
"The Logics and Information Theory of Sequential Circuits"
Automatic and Remote Control
J.F. Coates Ed.
VITA

Name: John Parkas
Date of birth: 4 May 1936
Place of birth: Milan, Italy

Education:
High School
  Saskatoon City Park
  Collegiate Institute

B. Eng. (El.) McGill University