ESD-TR-69-182

RUN LENGTH SYNCHRONIZATION TECHNIQUES

J. T. Barrows

JUNE 1969

Prepared for

AEROSPACE INSTRUMENTATION PROGRAM OFFICE
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

AD0691027

RUN LENGTH SYNCHRONIZATION TECHNIQUES

J. T. Barrows

JUNE 1969

Prepared for

AEROSPACE INSTRUMENTATION PROGRAM OFFICE
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

FOREWORD

This report was prepared by the Communications Techniques Department of The MITRE Corporation, Bedford, Massachusetts, under Contract AF 19(628)-5165. The work was directed by the Ground Instrumentation Engineering Division under the Aerospace Instrumentation Program Office, Air Force Electronic Systems Division, Laurence G. Hanscom Field, Bedford, Massachusetts. Robert E. Forney served as the Air Force Project Engineer for this program, identifiable as ESD (ESSIC) Project 5932, Range Data Transmission.

REVIEW AND APPROVAL

This technical Report has been reviewed and is approved.

GEORGE T. GALT, Colonel, USAF
Director of Aerospace Instrumentation
    Program Office

ABSTRACT

An important aspect of digital communications is the problem of determining efficient methods for acquiring block synchronization. In this paper we consider a sync technique based on the recognition of successive error-free digits from a known sequence.

The analysis of this technique draws from the theory of success runs. This theory is reviewed, and a simple recurrence relation is developed for computing the probability of the first occurrence of an error-free run of r digits in a binary sequence corrupted by noise. This relation is then applied to the analysis of the sync process, which utilizes an N-digit sync sequence as prefix to the data blocks. The results of this study show that this technique is a practical method for acquiring block synchronization.

ACKNOWLEDGMENT

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

SECTION I

INTRODUCTION

In digital communications an important problem is that of acquiring block synchronization at the receiver. A common block synchronization technique consists of prefixing a known sequence of length N digits to a D-digit data word. This scheme utilizes a correlation detector with an error threshold to recognize the sync prefix in the presence of channel noise. An in-sync indication is given after recognition of a fixed number of correctly spaced, consecutive, N-digit sequences.

In this paper we consider a different detection scheme. Instead of searching for N digits, allowing some errors, we search for a success run of length r, i.e., an r-digit error-free segment of the N-digit sync sequence $(r \leq N)$. The analysis of this scheme draws from the area in probability theory related to recurrent events and, more specifically, to success runs. We analyze what we term a "run-length" synchronization technique assuming, for lack of a priori knowledge, that the D-digit data words are randomly generated, and furthermore, that errors in the N-digit sync sequence occur randomly with probability p.

We begin by stating the criteria that will be used to evaluate the performance of this sync technique. Then the theory of success runs is reviewed. A new, simplified, recurrence relation is given for the probability of the first occurrence of a success run, which will facilitate the calculation of the various performance criteria.

The sync process is then introduced and analyzed. In addition, the results are presented in graphs for general applications.

1

SECTION II

ANALYSIS AND RESULTS

PERFORMANCE CRITERIA

The purpose of a synchronization technique is to determine the beginning of a data word. In a success-run technique the encoder prefixes an N-digit sequence to each data word of length D, and the decoder searches for a consecutive error-free run of r out of the N digits. When a run length of r digits is recognized, the system generates the remaining digits of the prefix and then declares word synchronization.

The performance of this particular synchronization scheme, when digit errors are present in the prefixed sync pattern, is evaluated by several performance criteria.

1. The probability P_N of acquiring sync is computed given that the decoder is examining the N-digit prefix.

2. The probability P_D of acquiring a (false) sync indication is computed given that the decoder is examining the D-digit (random) word.

3. The probability of acquiring a true sync for a $B = (D + N)$ digit block, designated by P_B, is lower bounded by assuming that the decoder examines first all D information digits and then the N sync digits. We upper bound this probability by assuming that the decoder examines first all N sync digits and then the D information digits. Hence

$$(1 - P_D) \, P_N < P_B < P_N \, .$$

3

4. The usual requirement imposed on a sync scheme is that, with probability greater than $1 - (10^{-k})$ (k a fixed integer), the system must be in true block sync within τ_0 seconds for a given transmission rate R. In general, the probability P_B of acquiring sync in one block will not satisfy this requirement. However, the probability of synchronizing correctly in T blocks is

$$P_T = 1 - (1 - P_B)^T \, ,$$

and the number of blocks that can be examined in τ_0 seconds at a data rate R is

$$T_0 = \frac{R}{B} \tau_0 \, .$$

Thus, the criterion to satisfy becomes $P_T > 1 - (10^{-k})$ with $T \leq T_0$, for a given k and channel error rate p.

THEORY OF SUCCESS RUNS

In this section we review the computation of the probability of success runs of length r in a sequence of Bernoulli trials. The first part of the discussion can be found in Feller[1] and is included here for completeness.

We are given a sequence of Bernoulli trials with p the probability that a trial results in a failure (a digit is in error) and $q = 1 - p$ the probability of success. (The words trial and digit are used interchangeably.) We say that a success run of length r, in a sequence of Bernoulli trials, occurs at the n th trial only if the n th trial results in the r th consecutive success

4

and also adds a new run to the sequence. Let u_n be the probability that a success run of length r is obtained at the n th digit, and f_n the probability that the first success run of length r occurs at the n th digit.

The probability that any r consecutive digits are correct is q^r. We look at the r digits numbered as $n - r + 1, n - r + 2, \ldots, n - 1, n$. Given that these r consecutive digits are correct implies that a success run must occur at one of these r digits. With probability u_{n-i}, a run occurs at the $(n - i)$ th digit, and therefore with probability $u_{n-i} q^i$ the n th trial also results in a success run of length r. The events, $q^i u_{n-i}$, are mutually exclusive for $i = 0, 1, 2, \ldots, r - 1$, by the definition of a success run (i.e., if an r-digit success occurs at the m th trial, then $u_{m+1} = u_{m+2} = \cdots = u_{m+r-1} = 0$). Adding these events, we obtain the recurrence relation

$$q^{r-1} u_{n-r+1} + q^{r-2} u_{n-r+2} + \cdots + q u_{n-1} + u_n = q^r \qquad (1)$$

for $n \geq r$. The boundary conditions are given as $u_0 = 1$, $u_1 = u_2 \ldots = u_{r-1} = 0$.

The solution of this recurrence relation (difference equation) is solved by introducing the generating function

$$\mathcal{U}(s) = \sum_{n=0}^{\infty} u_n s^n .$$

5

After some algebra we obtain

$$\mathcal{U}(s) = \frac{1 - s + pq^r s^{r+1}}{(1-s)(1-q^r s^r)} \ .$$

The value u_n represents the probability of a success run at the n th trial. We more precisely desire the probability of the first success run at the n th trial (f_n). The probability that a success occurs for the first time at trial number m and another success occurs at a later trial $n > m$ is, by definition $f_m u_{n-m}$. For example, the probability that a success occurs at the n th trial for the first time is $f_n = f_n u_0$. Recall that $u_0 = 1$. Since the u_{n-m} are mutually exclusive we have

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \dots f_n u_0 \tag{2}$$

for $n \geq 1$, with the boundary conditions on u_n given above. The right side of Equation (2) is the convolution of the sequences $\{f_i\} \{u_i\}$, which has the generating function $\mathcal{F}(s) \cdot \mathcal{U}(s)$, and the left side has the generating function $\mathcal{U}(s) - u_0$ with $u_0 = 1$. Thus the generating function for the recurrence times, $\mathcal{F}(s)$, is given by the relation

$$\mathcal{F}(s) = \frac{\mathcal{U}(s) - 1}{\mathcal{U}(s)} = \frac{q^r s^r (1-qs)}{1 - s(1 - pq^r s^r)} \ . \tag{3}$$

The general expression for f_n can be obtained by dividing through in the above equation. The results are given here as

6

$$f_n = q^r \sum_{k=0}^{} (-pq^r)^k \left[\binom{n-(k+1)r}{k} - q\binom{n-(k+1)r-1}{k} \right] \qquad (4)$$

where $\binom{n-(k+1)r}{k}$ is the binomial coefficient such that the summation terminates when $n-(k+1)r < 0$. The first few terms are:

$$f_1 = f_2 = \ldots = f_{r-1} = 0$$

$$f_r = q^r$$

$$f_{r+i} = pq^r \qquad \text{for } 1 \leq i \leq r$$

$$f_{2r+i} = pq^r [1 - q^r (1+(i-1)p)] \qquad 1 \leq i \leq r$$

$$\vdots$$

The mean and variance of f_n, the number of trials to the first success run of length r, are calculated from the first and second derivatives of the generating function $\mathcal{F}(s)$ evaluated at $s=1$ and are given respectively by

$$\mu = \frac{1-q^r}{pq^r} \quad , \qquad (5)$$

$$\sigma^2 = (pq^r)^{-2} - (2r+1)(pq^r)^{-1} - qp^{-2} . \qquad (6)$$

7

Table I lists the values of μ and σ for various p and r. The cumulative probability

$$F(n;r) \;=\; \sum_{i=1}^{n} f_i$$

is the probability that the first success run of length r occurs within n trials. For large n the computation of $F(n;r)$ becomes unwieldy using Equation (4). Feller[1] suggests that $F(n;r)$ can be approximated with a normal distribution. We show, however, that this approximation can be grossly inaccurate.

Specifically, let $A(n;r)$ be the normal approximation to $F(n;r)$; then

$$A(n;r) \;=\; \varphi(\alpha) - \varphi(\beta)$$

where

$$\alpha \;=\; \frac{n-\mu}{\sigma} \;, \qquad \beta \;=\; \frac{r-\mu}{\sigma} \;.$$

Using the expression of μ and σ given above, it can be shown that the lower limit, β, has value between -1 and 0 regardless of the values of r and p. Hence, as $n \to \infty$, we have $A(n;r) \to 1 - \varphi(\beta) \leq 0.8413$, but we know that $F(n;r) \to 1$ as $n \to \infty$.

We now examine f_n and $F(n;r)$ in detail to obtain an alternate technique for describing recurrence times in run-length problems. The algebraic details are given in the Appendix, hence only the results are presented here.

8

Table I

Values of μ and σ for the Probability of First Success Run of Length r

p	r	Mean μ	Standard Deviation σ
.5	10	2046.0	2037.47
.1	10	18.68	11.413
.01	10	10.573	2.074
.001	10	10.055	0.624
.0001	10	10.0055	0.196
.5	20	2,097,150.0	2,097,131.5
.1	20	72.253	57.473
.01	20	22.263	5.96
.001	20	20.212	1.712
.0001	20	20.021	0.536
.5	30	2.148×10^{9}	2.147×10^{9}
.1	30	225.9	202.9
.01	30	35.19	11.39
.001	30	30.47	3.123
.0001	30	30.046	0.974
.5	40	$\sim 2.2 \times 10^{12}$	$\sim 2.199 \times 10^{12}$
.1	40	666.550	634.69
.01	40	49.483	18.36
.001	40	40.832	4.803
.0001	40	40.082	1.491

In Equation (4), with $n = mr + i$ for $1 \le i \le r$, the upper limit on the summation is $m - 1$. Thus

$$f_n = q^r \sum_{k=0}^{m-1} (-pq^r)^k \left[\binom{n - (k+1)r}{k} - q \binom{n - (k+1)r - 1}{k} \right] \tag{7}$$

with the normal convention $\binom{i}{m-1} = 0$ for $i < m - 1$. We also have previously defined $F(N;r)$ as

$$F(N;r) = \sum_{n=0}^{N} f_n = \sum_{n=r}^{N} f_n \tag{8}$$

where we take advantage of the boundary conditions $f_0 = f_1 = \ldots f_{r-1} = 0$. Using these last two equations and the combinatorial relation

$$\binom{n}{m} = \sum_{k=m-1}^{n-1} \binom{k}{m-1} \tag{9}$$

we obtain

$$F(N;r) = q^r \sum_{k=0}^{m-1} (-pq^r)^k \left[\binom{N - (k+1)r + 1}{k+1} - q \binom{N - (k+1)r}{k+1} \right] \tag{10}$$

10

where $N = mr + i$ with $1 \leq i \leq r$. Looking at Equations (7) and (10) we see that they have a similar form, and with a little more manipulation we obtain

$$f_N = [1 - F(N-r-1;r)] \, pq^r \qquad (11)$$

for $N \geq r + 1$. For $N = r$ we have $f_r = q^r$. Equation (11) can be interpreted as the probability of the joint events (1) that a success run did not occur in the first $N-r-1$ digits, (2) the $(N-r)$ th digit is in error, and (3) the $(N-r+1)$ th to the N th digits are error-free. Thus f_N, the probability that the first r-digit success run is obtained at the N th digit, can be determined directly and in a form that can be verified intuitively.

Using Equation (11) we also obtain the recurrence relation on f_n as

$$f_{n+1} = f_n - pq^r \, f_{n-r} \qquad (12)$$

for $n \geq r + 1$ with initial conditions $f_1 = f_2 = \ldots = f_{r-1} = 0$, $f_r = q^r$ and $f_{r+1} = pq^r$. The derivation of Equation (12) is included in the Appendix. This formulation is advantageous for computation on a high-speed digital computer.

RUN LENGTH SYNCHRONIZATION PROCESS

The goal of any sync scheme is to optimize the probability of acquiring true word sync while minimizing the false sync probability. In this context we now discuss the properties of a sync detection process which utilizes the concept of run lengths.

11

The performance of the detection process of a run-length sync technique depends on the sync sequence used. Recall that we have an N-digit sync sequence, S(N), and are looking for the first occurrence of an error-free run of r digits, where $r \leq N$. In general, we need a sync sequence that has "good" correlation properties. Specifically, we require that each of the r-digit segments of S(N) be distinct (there are N-r+1 such segments). Then, once an r-digit run is found, the end-of-block can be uniquely ascertained. We also want the Hamming distance[2] between any two distinct r-digit segments of S(N) to be as large as possible. This requirement minimizes the probability of false sync when S(N) is being examined.

A natural choice which satisfies these requirements is a maximal length pseudo-random (PR) sequence[3]. Such a sequence, of length $2^k - 1$, is generated by the simple implementation of a k-stage shift register in which the feedback connections correspond to a k-degree primitive polynomial over the binary field[2]. The properties of this cyclic sequence are:

1. Given any k digits the next $2^k - 1 - k$ digits are uniquely determined. Thus any segment of the sequence with length greater than or equal to k is distinct.

2. The autocorrelation function is two-valued.

3. One half of the consecutive digits of the same kind are of length 1, 1/4 are of length 2, ... , $2^{-(k-1)}$ are of length k-1.

Using a portion of a PR sequence for S(N), and choosing N and r such that $k \leq r \leq N \leq 2^k - 1$, we not only satisfy the necessary requirements listed but can generate and detect the sync sequence quite easily.

The total probability of acquiring true sync in one block depends on the number of chances, (N-r+1), to find a run of r consecutive error-free

12

digits. This implies that r should be small. We must, however, choose r large to make the false sync probability within S(N) negligible. Thus the final choice of r depends on a trade-off analysis which is beyond the scope of this study.

The detection process is implemented by sequentially shifting the data stream through a k-stage shift register which corresponds to the maximal length sequence being used. At each clock time the contents of the register are mod 2 added according to the tap positions, and this sum digit is compared with the next data digit which is about to enter the register. A counter is increased by one if these two digits are the same, otherwise the counter is set to zero. Thus sync is effectively declared if and only if the count reaches r - k. This means that the r consecutive data digits match r digits of the pseudo-random sequence. This detection process is based on the property that any k consecutive digits of the pseudo-random sequence, in the absence of channel errors, uniquely generate the remaining r-k digits (and indeed, the remaining part of the PR sequence).

PROBABILITY OF OBTAINING TRUE SYNCHRONIZATION: P_N

A block consists of $B = D + N$ digits. In this section we consider the case in which the detector is examining the N-digit sync sequence which is some fixed portion of a $2^k - 1$ digit pseudo-random sequence. The probability at the n th trial of obtaining sync is the probability that the first count of r-k is attained at that trial. The detection process can be thought of as taking any k digits of S(N), generating the next r-k digits off-line, then comparing them with the corresponding r-k digits of the received sequence. Thus the first count of r-k is attained at the n th trial if and only if r consecutive digits of the received sync sequence correspond to some r-digit segment of S(N) . Hence, the probability of synchronizing at the n th trial

13

within the sync sequence is f_n, as given by Equation (7). And it follows that the probability of acquiring true word sync per block, P_N, from Equation (10) is

$$P_N = F(N;r) , \qquad (13)$$

which is the probability of obtaining the first run of r digits within N digits with channel error probability p.

This function $F(N;r)$ has been calculated for various values of p and r and illustrated on graphs in Figures 1, 2, and 3. Actually, the function $1 - F(N;r)$ is plotted versus N at values of $p = 10^{-1}$, 10^{-2}, and 10^{-3} for $r = 10$, 20, 30 and 40.

FALSE SYNCHRONIZATION PROBABILITY PER BLOCK: P_D

When the D-digit random data sequence is being examined a (false) sync is declared at the first digit in which a count of $r-k$ is reached in the detector. Therefore, the problem is still one of run lengths. The sort for relations cannot be obtained, however, simply by setting $p = q = 1/2$ in Equations (7) and (10), which were derived assuming the sync sequence was being examined.

In this case, the probability that an r-digit segment of the random sequence looks like a specific segment of the cyclic PR sequence is 2^{-r}; however, any of 2^k distinct r-digit segments of the PR sequence will yield an in-sync indication. (The all-zero segment, while not a part of the PR sequence, must be included as a possibility.) Thus, the probability that any r-digit segment of the randomly generated data word will give an in-sync indication is $2^{-(r-k)}$, not 2^{-r}. In view of this, we let f_n^* be the probability
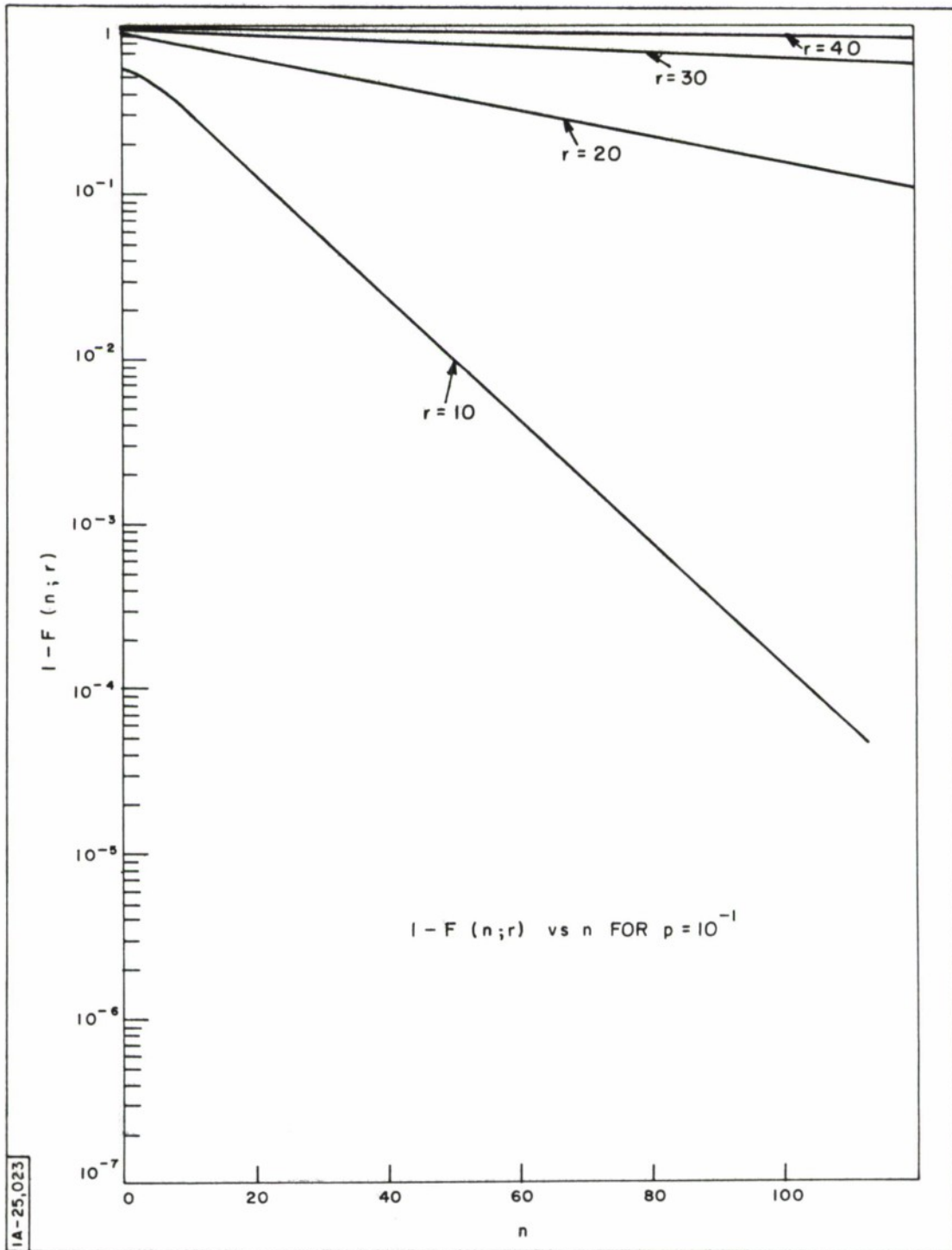
14

Figure 1. Cumulative Probability of Not Obtaining the First Success Within n Trials vs n for $p = 10^{-1}$
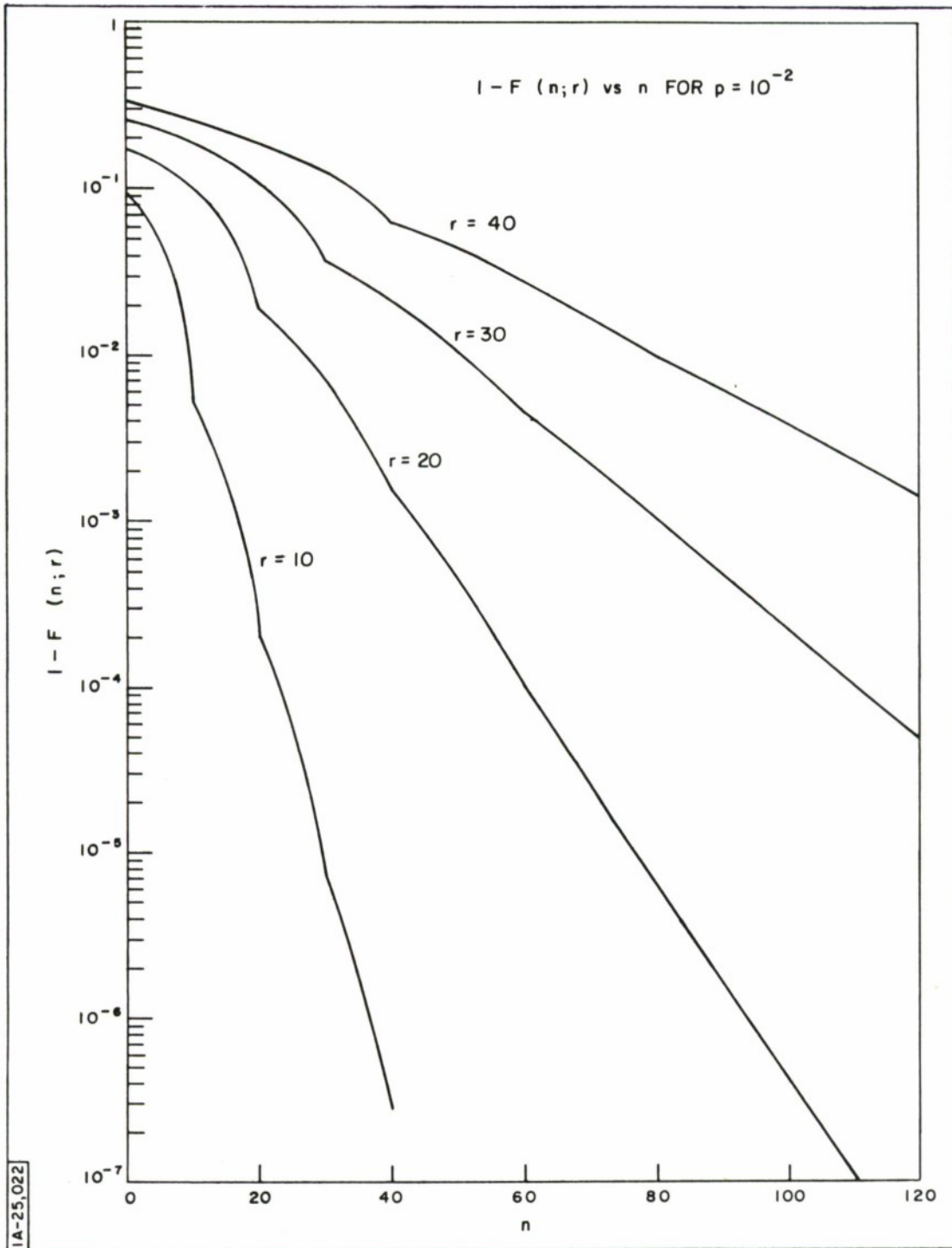
15

Figure 2. Cumulative Probability of Not Obtaining the First Success Within n Trials vs n for $p = 10^{-2}$
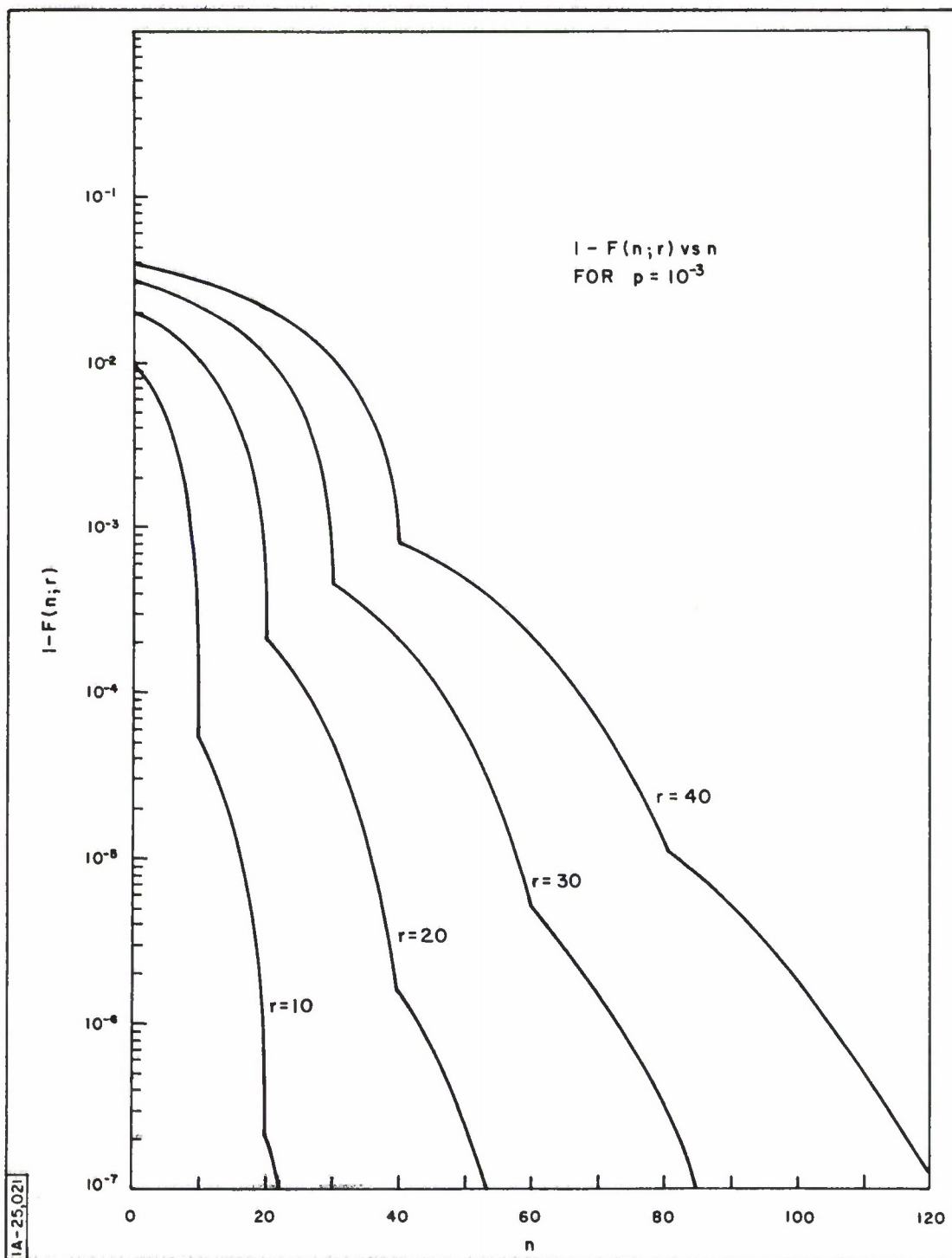
16

Figure 3. Cumulative Probability of Not Obtaining the First Success Within n Trials vs n for $p = 10^{-3}$

17

that the n th trial within the random data sequence yields the first success run of r digits (with $p = 1/2$). We have $f_1^* = f_2^* = \ldots f_{r-1}^* = 0$ and $f_r^* = 2^{-(r-k)}$. Arguing in the manner used to interpret Equation (11) we obtain the general expression for f_n^* as

$$f_n^* = [1 - F^*(n-r-1; r)] \; (1/2) \; (1/2)^{r-k} \tag{14}$$

for $n > r$ where

$$F^*(n-r-1; r) = \sum_{j=r}^{n-r-1} f_j^* . \tag{15}$$

Summing Equation (14) over D digits we obtain the cumulative probability of false sync per block as

$$P_D = F^*(D; r) = \sum_{n=r}^{D} f_n^* \tag{16}$$

Equation (14) looks, at first glance, as if it describes an $r-k$ run-length process with $p = q = 1/2$. A closer analysis reveals that this is not true. However, if $2^{-(r-k)}$ is small enough, Equation (16) can be approximated by Equation (10) for a run length of $r-k$ and $q = 1/2$. That is, using the first two terms of Equation (10) as an approximation to $F(D;r)$ we have

$$F(D;r) \doteq q^r \; [1 + p \; (D - r)] \quad . \tag{17}$$

18

Similarly, for a run length of r-k we have

$$F(D; r-k) \doteq q^{rk}[1 + p \ (D - r + k)] \tag{18}$$

or, with $q = p = 1/2$,

$$F(D-k; r-k) \doteq 2^{-(r-k)}[1 + 1/2 \ (D-r)] \quad . \tag{19}$$

Now, $F^*(D;r)$ can be precisely written as

$$F^*(D;r) = 2^{-r+k} + 2^{-r+k-1} \sum_{n=r+1}^{D} [1 - F^*(n-r-1;r)] \quad , \tag{20}$$

or

$$F^*(D;r) = 2^{-r+k-1} \ (D-r+2) - 2^{-r+k-1} \sum_{m=r}^{D-r-1} F^*(m;r) . \tag{21}$$

Now, if 2^{-r+k-1} is small enough we can neglect the summation in Equation (21) to obtain

$$F^*(D;r) \doteq 2^{-r+k-1} \ (D-r+2) \quad , \tag{22}$$

which is the same as Equation (19). Thus we have

$$P_D \doteq F^*(D;r) = F(D-k; r-k) \quad , \tag{23}$$

19

which is a satisfactory approximation for most applications. The function $F(n; r-k)$ with $p = q = 1/2$ is plotted on the graphs in Figure 4 for values of $r - k = 10$, 20 and 30.

The false sync probability in the overlap regions of random data and sync sequence have been ignored in this analysis. Actually, the probability that the last few random digits look like an extension of the sync sequence is quite high. This "incoming" case could result in inadvertently finding a correct sync indication. In the "outgoing" case, where the first few random digits of the next data word could look like an extension of the sync sequence, the effects would be sufficiently detrimental to warrant the use of added protective measures. In practice, an end-of-sync-sequence counter could be used to count the digits between a detected run length and the end of the sync sequence. If the count reaches $N - r + 1$ or greater, the sync indication is ignored, thus eliminating this false sync possibility.

PROBABILITY OF SYNCHRONIZING WITHIN A GIVEN TIME

The usual requirement for any sync technique is that it must be able to attain an overall probability of synchronization greater than $1 - \left(10^{-k}\right)$ for k a fixed integer. One measure of performance for a sync technique is the time it takes to ensure this probability. Moreover, relative merits of various sync schemes can be determined by comparing their probability of synchronizing in a given time interval. This comparison obviously should be made using similar data rates and block lengths.

For the scheme presented in this paper, the probability P_B of acquiring true sync in a block of B digits can be bounded as
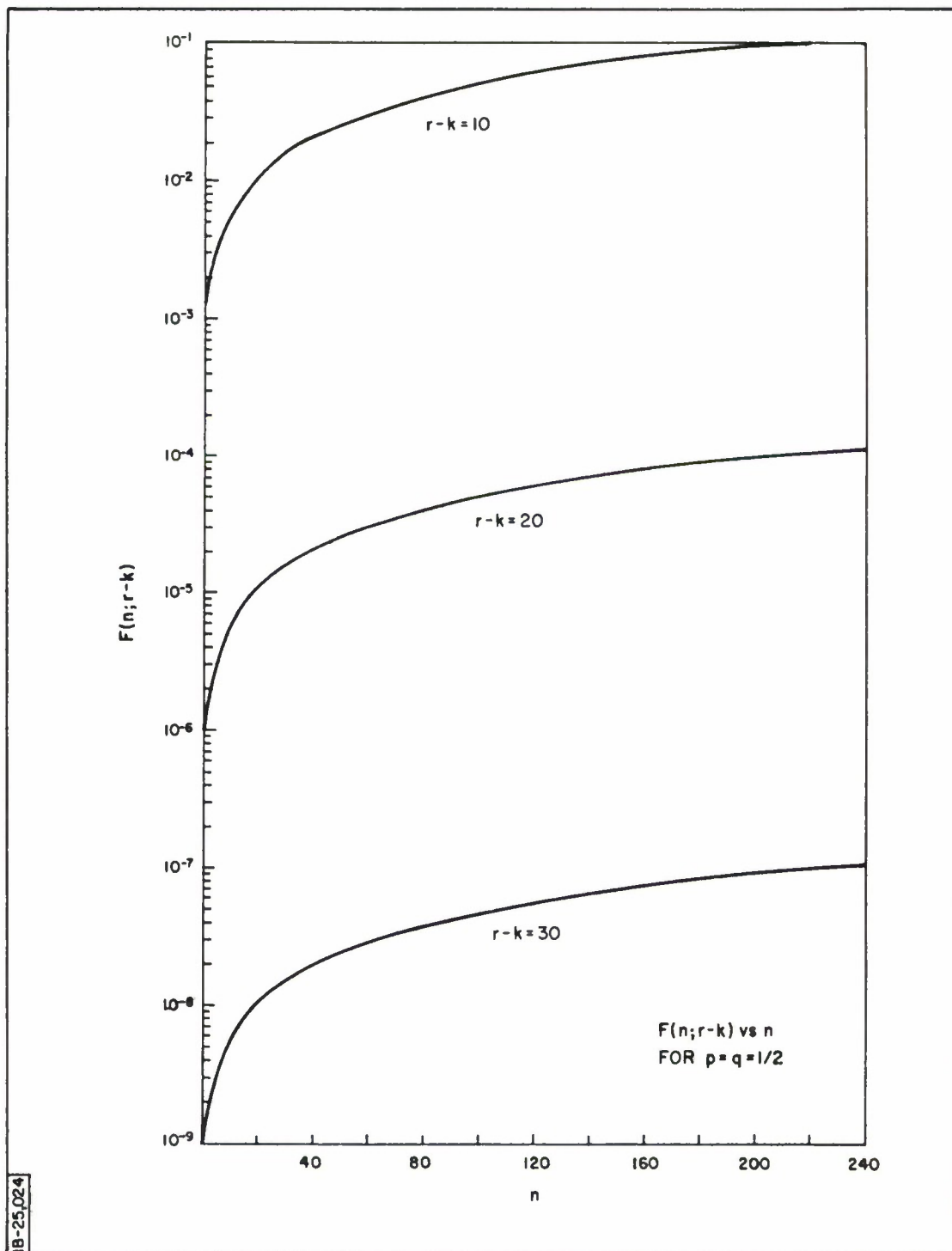
$$(1 - P_D) \, P_N < P_B < P_N \qquad (24)$$

20

Figure 4. Cumulative Probability of Obtaining the First Success
Within n Trials vs n for p = q = 1/2

21

where the expression $(1 - P_D)P_N$ indicates that the detector examines the total D-digit random word first and then the N-digit sync sequence.

For a given B, N, r, p, and k the probability P_B will not in general meet the requirement that $P_B > 1 - 10^{-k}$ and hence on the average more than one block will have to be examined. In T blocks the probability of correct sync is given by

$$P_T = 1 - (1 - P_B)^T . \tag{25}$$

The number of blocks that can be examined in τ seconds for a block of B digits and a data rate of R digits per second is $(R\tau/B$. Let T_0 be the minimum number of complete blocks needed for $P_{T_0} > 1 - 10^{-k}$ to be satisfied. In other words, T_0 is the smallest integer such that

$$1 - (1 - P_B)^{T_0} > 1 - 10^{-k} .$$

Therefore, T_0 must satisfy the inequality,

$$T_0 > \frac{k}{\log(\frac{1}{1 - P_B})} , \tag{26}$$

and is the smallest integer greater than or equal to the right side. The time it takes to examine this number of blocks is $T_0 B/R$. Thus if τ_0 is the maximum allowable time to achieve a probability greater than $1 - 10^{-k}$, the system must satisfy $T_0 B/R \leq \tau_0$.

22

Figure 5 is a plot of T_0 versus P_B for values of $1 - 10^{-k}$ of 0.9, 0.99 and 0.999 ($k = 1$, 2, and 3 respectively). Once the probability P_B of acquiring true sync per block is calculated, this graph allows the determination of the minimum number of blocks needed to be examined to ensure a total probability greater than $1 - (10^{-k})$.
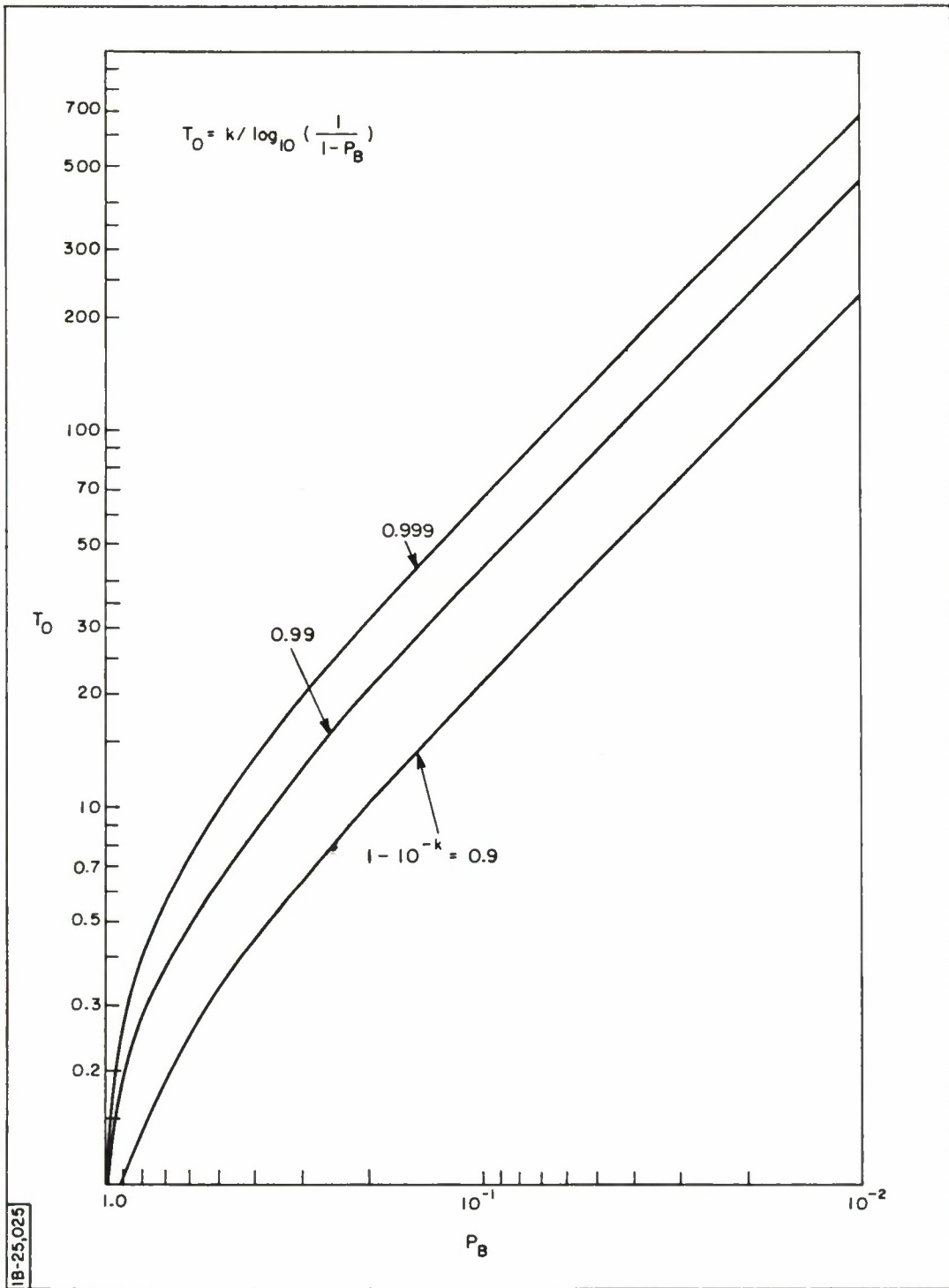
Figure 5. Plot of Minimum Number of Blocks to Guarantee a Probability of Acquiring Sync Greater than $1 - (10^{-k})$ vs the Probability of Synchronizing Per Block

24

SECTION III

CONCLUSIONS

We have analyzed a synchronization technique that searches for an r-digit error-free run within an N-digit sync sequence which is prefixed to the data words. The N-digit sequence is chosen to be a portion of a cyclic pseudo-random sequence of length 2^k-1 digits. A simple implementation has been described which employs a k-stage shift register as detector $(k < r)$. This detector examines the data stream digit by digit, while indexing a counter, enabling a search for a pattern much longer than the register length. Most commonly used sync techniques require examination of entire sync sequence lengths at one time.

In order to evaluate this run-length technique, expressions for several performance criteria have been derived. These criteria include the probability of acquiring true sync when the detector is examining the N-digit sync sequence, and also the probability of a false sync given that the detector is examining the data word. Graphs are provided which illustrate these expressions for parameter values that would be chosen for practical schemes. An interesting property of the analyzed sync scheme is that the probability of false sync increases linearly with the length of the data word. In sync schemes that allow up to e errors in the sync prefix, the probability of falsely synchronizing within the data word increases as $\binom{n}{e}$ where n is the length of the data word.

A more general result of this research which is applicable to problems in other areas is the derivation of a new recurrence relation for the probability of the first success run in a sequence of Bernoulli trials. The formulation presented here is advantageous for computation on a high-speed digital computer.

25

REFERENCES

1. W. Feller, An Introduction to Probability Theory and Its Application. New York: John Wiley and Sons, Inc., 1961.

2. S. W. Golomb, Sequences with Randomness Properties, Glen L. Martin Co., internal report, Baltimore, Maryland, June 14, 1965.

3. W. W. Peterson, Error-Correcting Codes, Cambridge, Mass.: The MIT Press, 1961.

APPENDIX

Given the generating function

$$\mathcal{F}(s) = \frac{q^r s^r (1-qs)}{1 - s(1-pq^r s^r)} = \sum_{n=0}^{\infty} f_n s^n$$

we derive the general expression for f_n. Let

$$\mathcal{F}(s) = q^r s^r (1-qs) \, G(s)$$

where

$$G(s) = \frac{1}{1 - s(1-q^r s^r p)} = \sum_{n=0}^{\infty} (1 - pq^r s^r)^n s^n$$

is expressed as a geometric series. Now $(1 - pq^r s^r)^n$ is given as

$$(1-pq^r s^r)^n = \sum_{i=0}^{n} \binom{n}{i} (-pq^r s^r)^i \quad .$$

Therefore

$$G(s) = \sum_{n=0}^{\infty} s^n \sum_{i=0}^{n} \binom{n}{i} (-pq^r s^r)^i$$

29

$$= \sum_{n=0}^{\infty} s^n \left[1 - \binom{n}{1} (pq^r) s^r + \binom{n}{2} (pq^r)^2 s^{2r} \ldots + (-1)^n (pq^r)^n s^{nr} \right]$$

$$= \sum_{n=0}^{\infty} g_n s^n \ .$$

The coefficient g_n is obtained by collecting terms of like power in the above expression. That is,

$$g_n s^n = s^n \left[1 - \binom{n-r}{1} pq^r + \binom{n-2r}{2} (pq^r)^2 \ldots + \binom{n-mr}{m} (-pq^r)^m \right]$$

where m is the greatest integer less than or equal to $n/(r+1)$. Using this summation for $G(s)$ in the expression for $\mathcal{F}(s)$, we obtain

$$\mathcal{F}(s) = q^r s^r \sum_{n=0}^{\infty} g_n (1-qs) s^n$$

$$= q^r s^r \left[g_0 + s(g_1 - qg_0) + s^2 (g_2 - qg_1) + \ldots \ldots \right]$$

with $g_0 = 1$. It is apparent that $f_0 = f_1 = \ldots f_{r-1} = 0$, and if we define $g_{-1} = 0$ we can write

30

$$\mathcal{F}(s) = \sum_{n=r}^{\infty} q^r (g_{n-r} - qg_{n-r-1}) s^n .$$

Thus

$$f_n = q^r \left[g_{n-r} - qg_{n-r-1} \right]$$

for $n \geq r$. Using the above expression for g_n we have

$$f_n = q^r \sum_{k=0}^{m} (-pq^r)^k \left[\binom{n-(k+1)r}{k} - q \binom{n-(k+1)r-1}{k} \right]$$

where the upper limit on the summation m is the greatest integer less than or equal to $\dfrac{n-r}{r+1}$.

*   *   *   *   *

Now Equation (10) and the simple expression

$$f_N = pq^r \left[1 - F(n-r-1; r) \right]$$

for $N > r$ are derived from the previous results.

31

$$F(N;r) = \sum_{n=r}^{N} f_n$$

$$= q^r \sum_{n=r}^{N} \sum_{k=0}^{k} (-pq^r)^k \left[\binom{n - (k+1)r}{k} - q \binom{n - (k+1)r - 1}{k} \right]$$

We interchange the order of summation and look at the terms

$$\sum_{n=r}^{N} \binom{n - (k+1)r}{k} \quad \text{and} \quad \sum_{n=r}^{N} \binom{n - (k+1)r - 1}{k}$$

Starting with the well-known combinatorial relation

$$\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-1}{m}$$

we repeatedly use this expansion on the second term of the previous iteration as

$$\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-2}{m-1} + \binom{n-2}{m}$$

$$= \binom{n-1}{m-1} + \binom{n-2}{m-1} + \binom{n-3}{m-1} + \binom{n-3}{m}$$

32

and so on, to obtain the relation

$$\binom{n}{m} = \sum_{j=m-1}^{n-1} \binom{j}{m-1}$$

Using this relation with the expression

$$\sum_{n=r}^{N} \binom{n-(k+1)r}{k} = \sum_{n=(k+1)r+k}^{N} \binom{n-(k+1)r}{k} \quad ,$$

we make a change of variable to obtain

$$\sum_{n=r}^{N} \binom{n-(k+1)r}{k} = \sum_{m=k}^{N-(k+1)r} \binom{m}{k} = \binom{N-(k+1)r+1}{k+1} \quad .$$

Similarly, we have

$$\sum_{n=r}^{N} \binom{n-(k+1)r-1}{k} = \binom{N-(k+1)r}{k+1}$$

Finally, we use these last results in the expression for $F(N;r)$ above to obtain Equation (10) as

$$F(N;r) = q^r \sum_{k=0}^{r} (-pq^r)^k \left[\binom{N-(k+1)r+1}{k+1} - q\binom{N-(k+1)r}{k+1} \right]$$

33

for $N \geq r$. Letting $N = mr + i$ with $m \geq 0$ and $1 \leq i \leq r$, the upper limit on this summation is $m - 1$.

Now, in the preceding equation let $x = k + 1$. Then

$$F(N;r) = q^r \sum_{x=1}^{m} (-pq^r)^{x-1} \left[\binom{N - xr + 1}{x} - q \binom{N - xr}{x} \right] .$$

Multiplying both sides by $(-pq^r)$ we have

$$-pq^r \; F(N;r) = q^r \sum_{x=1}^{m} (-pq^r)^{x} \left[\binom{N - xr + 1}{x} - q \binom{N - xr}{x} \right]$$

$$= q^r \sum_{x=1}^{m} (-pq^r)^{x} \left[\binom{N + r - (x+1)r + 1}{x} - q \binom{N + r - (x+1)r}{x} \right]$$

This summation is the expression for f_{N+r+1} with the $x = 0$ term missing. Thus,

$$-pq^r \; F(N;r) = \left[f_{N+r+1} \quad -q^r p \right]$$

where pq^r is the zero order term. Rearranging, we have

$$f_{N+r+1} = pq^r \left[1 - F(N;r) \right]$$

34

or

$$f_N = pq^r \left[1 - F(N-r-1; r) \right] \quad .$$

The recurrence relation

$$f_{n+1} = f_n - pq^r f_{n-r} \quad ,$$

which can be taken as the defining equation for the probability of obtaining the first run of r successes at the n th trial, is derived from the previous result. That is,
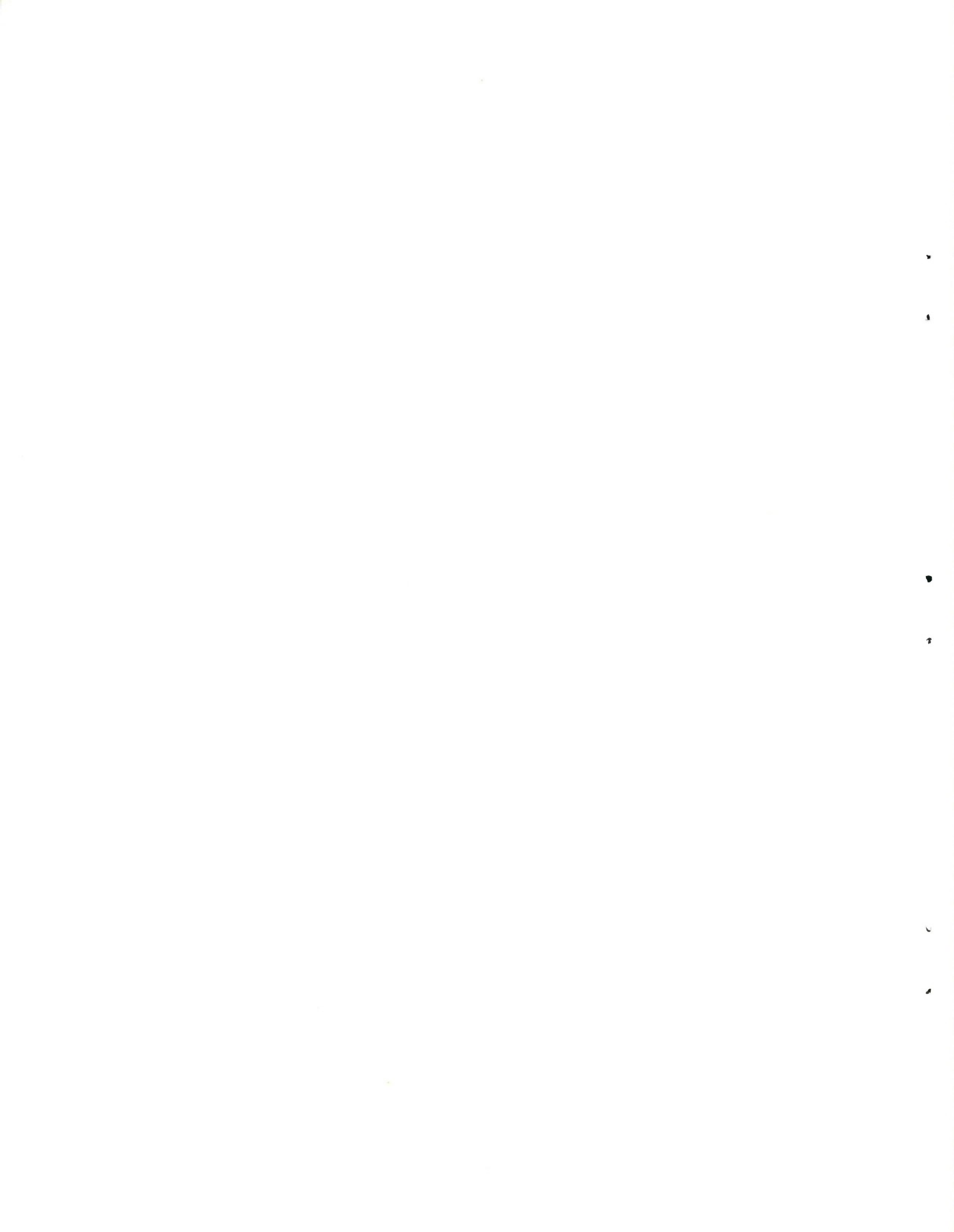
$$f_{n+1} = pq^r \left[1 - F(n-r; r) \right]$$

and

$$f_n = pq^r \left[1 - F(n-r-1; r) \right] .$$

Subtracting the second equation from the first we have

$$f_{n+1} = f_n - \left[F(n-r; r) - F(n-r-1; r) \right] pq^r \quad .$$

But the term in brackets is just f_{n-r}. Hence,

$$f_{n+1} = f_n - pq^r f_{n-r} \quad .$$

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
The MITRE Corporation Bedford, Massachusetts	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

RUN LENGTH SYNCHRONIZATION TECHNIQUES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

N/A

5. AUTHOR(S) *(First name, middle initial, last name)*

J. T. Barrows

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
JUNE 1969	38	--

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF 19(628)-5165	ESD-TR-69-182
b. PROJECT NO.	
705B	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	MTP-73

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Aerospace Instrumentation Program Office, Electronic Systems Division, Air Force Systems Command, L. G. Hanscom Field, Bedford, Massachusetts
N/A	

13. ABSTRACT

An important aspect of digital communications is the problem of determining efficient methods for acquiring block synchronization. In this paper we consider a sync technique based on the recognition of successive error-free digits from a known sequence.

The analysis of this technique draws from the theory of success runs. This theory is reviewed, and a simple recurrence relation is developed for computing the probability of the first occurrence of an error-free run of r digits in a binary sequence corrupted by noise. This relation is then applied to the analysis of the sync process, which utilizes an N-digit sync sequence as prefix to the data blocks. The results of this study show that this technique is a practical method for acquiring block synchronization.

DD FORM 1 NOV 65 **1473**

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
DIGITAL COMMUNICATIONS BLOCK SYNCHRONIZATION						