International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:9, 2008

On the Properties of Pseudo Noise Sequences with a Simple Proposal of Randomness Test

Abhijit Mitra

***Abstract*—**Maximal length sequences (m-sequences) are also known as pseudo random sequences or pseudo noise sequences for closely following Golomb's popular randomness properties: (P1) balance, (P2) run, and (P3) ideal autocorrelation. Apart from these, there also exist certain other less known properties of such sequences all of which are discussed in this tutorial paper. Comprehensive proofs to each of these properties are provided towards better understanding of such sequences. A simple test is also proposed at the end of the paper in order to distinguish pseudo noise sequences from truly random sequences such as Bernoulli sequences.

***Keywords*—**Maximal length sequence, pseudo noise sequence, punctured de Bruijn sequence, auto-correlation, Bernoulli sequence, randomness tests.

I. Introduction

MAXIMAL length sequences (which are also called m-sequences, pseudo random sequences or pseudo noise sequences) are certain binary sequences of length $N = 2^n - 1$ that satisfies a linear recurrence given by the corresponding primitive polynomial of degree n [1]. Although these sequences are not truly random because they can be predicted by a definite recurrence relation, nevertheless, they have many useful properties for which rapidly generated such m-sequences with 'fairly acceptable' randomness properties are essential components in a wide variety of modern applications including radar, spread spectrum, error correction, cryptographic systems, and Monte Carlo simulations. Acceptable m-sequences should exhibit no statistical bias in the occurrence of individual symbols or small blocks of symbols. With these goals in mind, in his classic book, S. Golomb [2] defined a pseudo noise (PN) sequence to be a periodic binary sequence that passes three well known statistical tests for randomness: balance, run, and, ideal autocorrelation; each of which we take up one by one in next section. It is also shown in [2] that such sequences can be rapidly generated using linear feedback shift register (LFSR) via primitive polynomials over a certain finite field (called, Galois field, or, $GF(q)$ where q denotes a finite prime number).

Although these PN sequences have many properties [3]-[5] apart from the three mentioned above, a simple and comprehensive account of all of their properties altogether is given in none of the existing literature including standard texts. We provide those in this tutorial paper along with the respective proofs towards better understanding of such sequences. A few applications of these properties in communication theory are also given in the form of *Lemmas*. Further, we propose a simple randomness test in order to distinguish the PN sequences from any fair coin tossing experiment, i.e., Bernoulli sequence. In the literature, there exist many randomness tests [6]-[7] for the same purpose yielding very good results. Our proposed test, however, is so simple that it can be easily taken up by the undergraduate students as an assignment in any 'advanced communication' course.

Manuscript received August 01, 2008.

A. Mitra is with the Department of Electronics and Communication Engineering, Indian Institute of Technology (IIT) Guwahati, Guwahati – 781039, India. E-mail: a.mitra@iitg.ernet.in.

II. Properties of Pseudo Noise Sequences

Before going into the properties of PN sequences, let us briefly describe the background materials first. We define a primitive polynomial $p(x)$ of degree n over a $GF(2^n)$ as

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0 \qquad (1)$$

where, all the coefficients a_i, $i = 0, 1, ..., n$ are members of $GF(2)$, i.e., integers $\{0, 1\}$ with $a_n = a_0 = 1$. We generate the recurring PN sequences with countably infinite list of values $(s_0, s_1, ...)$ via such primitive polynomials and define them as follows.

Definition 1: A PN sequence is a $(2^n - 1)$ length sequence that satisfies a linear recurrence, defined over $GF(2)$, given by any corresponding primitive polynomial $p(x)$ of degree n.

Remark 1: For any primitive polynomial $p(x)$, the reciprocal polynomial $p^*(x)$, defined as $p^*(x) = x^n p(x^{-1})$, is also a primitive polynomial, and the PN sequence generated by $p^*(x)$ is exactly the reverse of the PN sequence generated by $p(x)$.

Although three randomness properties (balance, run, and ideal autocorrelation) of these sequences are widely known, there exist some other properties also through which we usually characterize such sequences. Below, we provide all the properties of $(2^n - 1)$ length PN sequences along with a proof of each of them.

Property 1: **Recurrence:** In general, any PN sequence of order n satisfies the linear recurrence given by

$$s_{i+n} = \sum_{k=0}^{n-1} a_k s_{i+k}, \forall i \geq 0 \qquad (2)$$

where, as earlier, $a_k \in GF(2)$.

Proof: The relation $p(x) = 0$ is called the characteristic equation of polynomial (1). Putting this value and by replacing x^k with s_{i+k} in (1), we get (2) as, in $GF(2)$, minus signs can be changed to plus signs for modulo-2 arithmetic. ■

Remark 2: Specifying the initial values $(s_0, s_1, ..., s_{n-1})$ completely specifies the sequence. Indeed, specifying any n consecutive values specifies all remaining values. Since

there are only 2^n possible combinations of $s_0, ..., s_{n-1}$, any sequence satisfying (2) must repeat after the period of 2^n-1 as the all-zero state cannot occur unless the sequence is itself all zeros. This also leads to 2^n-1 different possible PN sequences depending upon the initial combination of $s_0, ..., s_{n-1}$. If we assume a set $\mathcal{S}$ containing all these PN sequences, it then follows from (2) that there are n linearly independent PN sequences in $\mathcal{S}$.

Property 2: **Closure:** The elements of any PN sequence $(s_0, s_1, ..., s_{2^n-2})$ of length N, formed with respect to the recursion (2), repeat the original sequence beyond s_{2^n-2}.

Proof: Following some elementary results in group theory, one can find that if $b \neq 0 \pmod q$, where $b \in GF(q)$, then $b^{q-1} = 1$. Multiplying both sides by b shows that $b^q - b = 0$ for $b \neq 0$. Since this is also true for $b = 0$, we see that every element in $GF(q)$ satisfies the following polynomial equation: $x^q - x = 0$ which has q roots with each root being exactly the elements in $GF(q)$. As $GF(q^n)$ is also a valid finite field, putting $q = q^n$ above yields the equation $x^{q^n} - x = 0$, or, $x^{q^n-1} = 1$. Mapping this result in (2) with $q = 2$, one obtains

$$s_{2^n-1+i} = s_i \quad \forall i \geq 0. \tag{3}$$

Following (3), any countably infinite list of values with recurrence (2) take the form $(s_0, s_1, ..., s_{2^n-2}, s_0, s_1, ...)$. This proves the property. ■

Property 3: **Shift:** If $\mathbf{s} = (s_i, s_{i+1}, ..., s_{i+2^n-2})$, $\forall i \geq 0$, is a PN sequence in $\mathcal{S}$, then any right or left cyclic shift on $\mathbf{s}$ is also in $\mathcal{S}$.

Proof: Consider any PN sequence $\mathbf{s}' = (s_{i+1}, s_{i+2}, ..., s_{i+2^n-1}) \in \mathcal{S}$ which comes out of a different initial state such that $\mathbf{s} \neq \mathbf{s}'$. Using *Property 2*, we get $s_{2^n-1+i} = s_i$. Hence $\mathbf{s}' = (s_{i+1}, s_{i+2}, ..., s_{i+2^n-2}, s_i)$, meaning it is nothing but a left circular shift on $\mathbf{s}$. Similarly, taking another $\mathbf{s}'' = (s_{i-1}, s_i, ..., s_{i+2^n-3}) \in \mathcal{S}$, it is easy to show that it is a right circular shift on $\mathbf{s}$. This holds true for all $i \geq 0$, which proves the stated property. ■

Remark 3: The above *Shift Equivalence Property* can also be proved in an alternate nice manner. Let us reformulate the above property as: any two different $\mathbf{s}$ and $\mathbf{s}'$, generated by the same primitive polynomial $p(x)$, must be in $\mathcal{S}$. We can then prove it as follows. For $n \geq 2$, it is well known that $2^{n-1} < 2^n - 1 < 2^n$, meaning construction of a length $2^n - 1$ PN sequence is not possible with less than n-tuples, where except the all zero state, we use all the other possible $2^n - 1$ states to generate such a sequence. Now, if any two $\mathbf{s}$ and $\mathbf{s}'$ are shift distinct, then the total number of different initial states needed to generate $\mathbf{s}$ and $\mathbf{s}'$, keeping $p(x)$ same, would be $2 \times (2^n - 1) > 2^n$ which is a contradiction since only $(2^n - 1)$ nonzero n-tuple states are available. Thus, any two $\mathbf{s}$ and $\mathbf{s}'$ must be shift equivalent.

Property 4: **Add:** The sum of any two PN sequences in $\mathcal{S}$ (formed componentwise, modulo 2, without carries) is another sequence in $\mathcal{S}$.

Proof: Let us take a simple recursion $s_{i+n} = s_i + s_{i+1}$. With this, adding $\mathbf{s}$ and $\mathbf{s}'$ yields the sequence $(s_{i+n}, ..., s_{i+2^n-2}, s_i, ..., s_{i+n-1})$. By *Property 3*, it is also in $\mathcal{S}$. The same holds true for any general recursion as given in (2) [the readers are encouraged to do it]. ■

Property 5: **Shift-and-Add:** The sum of a PN sequence in $\mathcal{S}$ and a cyclic shift of itself is another PN sequence in $\mathcal{S}$.

Proof: The proof is trivial which follows from *Properties 3* and *4*. ■

Remark 4: A PN sequence in $\mathcal{S}$ which obeys *Property 5*, will also obey "Shift-and-Subtract" rule as, in $GF(2)$, minus signs can be changed to plus signs.

Property 6: **Window:** If $2^n - 1$ non-overlapping windows of width n bits each are framed along any PN sequence in $\mathcal{S}$ (let us take n repetitions of the same PN sequence to place all these windows properly), each of the distinct $2^n - 1$ nonzero possible combinations of $s_0, ..., s_{n-1}$ is seen exactly once in each window.

Proof: The above property, in other words, can be written as: each possible n-tuple is seen only once in a full period of any PN sequence. To prove this, let us recall that the relation $p(x) = 0$ is called the characteristic equation of (1). Let u be a root of this equation, meaning $u^n = \sum_{k=0}^{n-1} a_k u^k$. Multiplying both sides by u^i, we get $u^{i+n} = \sum_{k=0}^{n-1} a_k u^{i+k}$ which shows that the sequence defined by $s_i = u^i$ satisfies the recurrence in (2). If $p(x)$ is an irreducible polynomial over $GF(2^n)$, this equation will have n distinct roots: $\{1, u^1, u^2, ..., u^{n-1}\}$. Then, any element of this field, i.e., $\{1, u^1, ..., u^{2^n-1}\}$, can be expressed in terms of linear combinations of $\{1, u^1, u^2, ..., u^{n-1}\}$ as given above, meaning, each element of $GF(2^n)$ will have a n-tuple vector representation. Let, G be an $n \times (2^n - 1)$ generator matrix with the rows to be the binary n-tuples which are these rectangular co-ordinates for all $2^n - 1$ elements $\{1, u^1, ..., u^{2^n-2}\}$ (it is easy to check from *Property 2* that $u^{2^n-1} = 1$). Reformulating the above result as $u^{i+n} = u^i \sum_{k=0}^{n-1} a_k u^k$, it then means any row of G is a linear combination of $\{1, u^1, u^2, ..., u^{n-1}\}$ multiplied by a unique element u^I where $I \equiv i \pmod{2^n - 1}$. If the combination $\sum_{k=0}^{n-1} a_k u^k$ is linearly independent, i.e., $\sum_{k=0}^{n-1} a_k u^k = 0$ iff all $a_k = 0$ where $a_k \in GF(2)$, only then each row will show a unique non-zero element. Conversely, any consecutive n bits in a PN sequence will represent this unique u^I iff the above combination is linearly independent. To show this, we only need to show that none of the elements $\{1, u^1, u^2, ..., u^{n-1}\}$ is zero (they must be distinct as stated above). Assume that any u^k, $k = 1, ..., n-1$, is zero ($u^0 = 1$ cannot be a zero) so that $\sum_{k=0}^{n-1} a_k u^k = 0$ with $a_k \neq 0$ for that k. As u^k is also a root of $p(x) = 0$, putting the value in the equation it yields $a_0 = 0$ which is a contradiction with the definition of $p(x)$. Hence, any $u^k \neq 0$, leading to the required result. ■

Remark 5: Property 6 is also called *Span* property since the n element subset $\{1, u^1, u^2, ..., u^{n-1}\}$ spans the entire $\mathcal{S}$, i.e., forms the basis of $GF(2^n)$. A PN sequence, which follows this property, is called a *Punctured de Bruijn* sequence of span n.

Remark 6: The n linearly independent PN sequences in $\mathcal{S}$ corresponds to n linearly independent solutions of $p(x) = 0$.

Property 7: **Balance:** Any PN sequence in $\mathcal{S}$ contains 2^{n-1} ones and $2^{n-1} - 1$ zeros.

Proof: As all zero initial combination of $s_0, ..., s_{n-1}$ is not permitted, all the other combinations take equivalent decimal values between 1 and $2^n - 1$. Since there are 2^{n-1} odd numbers and $2^{n-1} - 1$ even numbers between this range, with

binary representations finishing with 1's and 0's, respectively, we get the above result. ■

Property 8: **Run:** In any PN sequence, $1/2$ of the runs have length 1, $1/4$ have length 2, $1/8$ have length 3, $1/16$ have length 4, and so on, as long as these fractions give integral numbers of runs. Also, in each case, the number of runs of 0's is equal to the number of run's of 1's.

Proof: A run of length k is a basically a block of k consecutive identical digits that is not contained in a longer block of consecutive digits. In other words, it is a block $(s_i, s_{i+1}, ..., s_{i+k-1})$ in any PN sequence such that $s_{i-1} \neq s_i = s_{i+1} = ... = s_{i+k-1} \neq s_{i+k}$. With this definition of run, if we now want to prove the run property, we have to take the help of *Property 6* which states that every non-zero n-tuple occurs exactly once in any PN sequence. Since every possible n-tuples occur for a single time, the n-tuple $\underbrace{11...1}_{n}$ must occur only once and it must be preceded and followed by a 0 each, indicating a run of length n of ones. As this $(n+2)$-tuple with zeros at both the extreme ends can alternately be visualized as

$$\overbrace{0\underbrace{11...1}_{n}0}^{n+2} \rightarrow 01\overbrace{\underbrace{11...1}_{n-1}0}^{n} \quad \text{or} \quad \overbrace{0\underbrace{11...1}_{n-1}10}^{n}$$

it is evident that there can be no run of length $(n-1)$ ones in the sequence since the n-tuples 11...10 or 01...1 occur only once in the sequence. However, there will be exactly one run of length $(n-1)$ zeros, represented as $1\underbrace{00...0}_{n-1}1$, since all-zero n-tuple cannot occur. Now let us consider the run of ones of length r where $0 < r \leq n-2$. Each such run can be represented in the form of n-tuples as:

$$\overbrace{0\underbrace{11...1}_{r}0\underbrace{xx...x}_{n-r-2}}^{n}$$

where x are any arbitrary digits (either 0 or 1). Such arbitrary digits can be arranged in 2^{n-r-2} ways. This means the number of such n-tuples of run length r is 2^{n-r-2}. A similar argument gives the same number of runs of zeros of length r. Hence, any run of length r, $0 < r \leq n-2$, occurs $2^{n-1}2^{-r}$ times including runs of both ones and zeros. Runs of length n and $n-1$ occur only once for zeros and ones, respectively. This completely determines the run structure of PN sequences as given above. ■

Remark 7: The maximum length of a run cannot be greater than n which comes from *Property 1* that is governed by the primitive polynomial of (1). As shown above, one run of this length and one run of length $n-1$ will take place in the total sequence. Putting the value of r for all other runs, the total number of runs comes out to be $K = (1+1+2...+2^{n-2}) = 2^{n-1}$. Any PN sequence always follows this total number of runs in a single period.

Property 9: **Ideal Autocorrelation:** The autocorrelation function $r(i)$ of any PN sequence of length N is given by

$$r(i) = \begin{cases} 1 & \text{for } i = 0 \\ -\frac{1}{N} & \text{for } 1 \leq |i| \leq N-1. \end{cases} \tag{4}$$

Proof: We usually define the autocorrelation function (ACF) of any real sequence as $r(i) = \frac{1}{N}\sum_{j=0}^{N-1} s_j s_{j+i}$ for $|i| \geq 0$. For a binary sequence where any $s_i \in \{0,1\}$, let us replace 1's by -1's and 0's by 1's so that any sequence $(s_0, s_1, ..., s_{2^n-2})$ can be represented as $\{(-1)^{s_0}, (-1)^{s_1}, ..., (-1)^{s_{2^n-2}}\}$. In that case, the ACF takes the form $\frac{1}{N}\sum_{j=0}^{N-1}(-1)^{s_j+s_{j+i}}$. Now, if A is the number of places where any sequence $(s_0, s_1, ..., s_{2^n-2})$ and its i cyclic shift $(s_i, s_{i+1}, ..., s_{i-1})$ agree and D is the number of places where they disagree so that $A+D=N$, then, from the above definition of ACF,

$$r(i) = \frac{A-D}{N}. \tag{5}$$

From *Property 5*, the elementwise sum of the above two sequences must be in $\mathcal{S}$. Then D will denote the number of 1's in the resultant sequence and A would represent the number of 0's (mod 2). By *Property 7*, $A = 2^{n-1}-1$ and $D = 2^{n-1}$ for any $i \neq 0$. For $i = 0$, the resultant sequence will be a all zero sequence, leading to a value of ACF$= 1$. We thus get (4). ■

Remark 8: Every *punctured de Bruijn* sequence of span n is balanced and has shift-and-add, run and ideal autocorrelation properties.

Remark 9: The value of $D = 2^{n-1}$ signifies the minimum distance between any two sequences in $\mathcal{S}$, i.e., in how many digits they differ from each other. This is treated as an important parameter in communication theory. Also, this denotes the weight of a PN sequence which is defined as the number of 1's in it.

Remark 10: Another important measure in communication theory is *Figure of Merit* (FoM), which is defined as

$$F_M = \frac{r^2(0)}{\sum_{i\neq 0} r^2(i)} = \frac{r^2(0)}{2\sum_{i=1}^{N-1} r^2(i)} \tag{6}$$

since ACF is a symmetric function. FoM usually gives a quantitative spectral information of the sequences such as, for low FoM, the spectrum is narrow or irregular whereas for high FoM, it is the other way round. From (4), we see that PN sequences have an FoM $= \frac{N^2}{2(N-1)}$, which yields quite a high value for high N. It then follows that such sequences must have applications in wideband communications.

Property 10: **Construction of Hadamard Matrices:** If an $(N+1)\times(N+1)$ array is formed whose rows are each of the PN sequences in $\mathcal{S}$, by replacing 1's with -1's and 0's with 1's of each sequence, along with adding an initial row of length N and an initial column of length $(N+1)$ with all 1's, the resultant array is a $2^n \times 2^n$ Hadamard matrix.

Proof: Any $n \times n$ real matrix H_n with all its entries as ± 1 is called a Hadamard matrix if it satisfies the relation: $H_n H_n^T = nI$, where the superscript T denotes the transposition operation and I is an $n \times n$ unit matrix. Following the construction of H_{2^n} as given above, it is easy to check that $H_{2^n} H_{2^n}^T = 2^n I$ which follows from a trivial modification of *Property 9* in the expression (4). ■

We show some of these properties of PN sequences via Figs. 1 to 4. In Fig. 1, the LFSR structure is shown corresponding to the recurrence $s_{i+4} = s_{i+1} + s_i$. Fig. 2 shows all the possible 15 PN sequences obtained from this recurrence simply by left

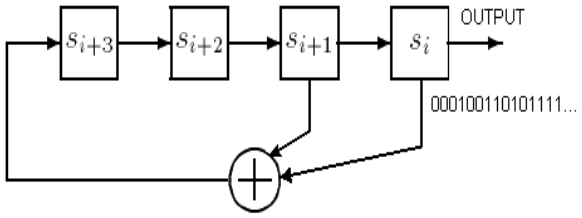

Fig. 1. The LFSR structure corresponding to the recurrence relation $s_{i+4} = s_{i+1} + s_i$.

shifting the first sequence 15 times. Fig. 3 demonstrates the window property of this sequence. Finally, the typical ACF nature of this sequence is depicted in Fig. 4.

III. Application of These Properties in Communication Theory

If we consider a set $\mathcal{T}$ that contains all the possible 2^{2^n-1} combinations of binary (2^n-1)-tuple, then the set $\mathcal{S}$ containing all PN sequences is definitely a proper subset of $\mathcal{T}$. $\mathcal{T}$ is called an N digit permutation alphabet while $\mathcal{S}$ is called an N digit PN alphabet. *Properties 1* to *10* of $\mathcal{S}$, as stated and proved above, have certain importance in the theory of communications [8]. Firstly, the PN sequences in $\mathcal{S}$ form 'block codes' or 'forward error correction codes', with the properties: systematic, cyclic, and linearity, that enable a limited number of errors to be detected and corrected without retransmission. Such a code is referred to as (N, n) code, meaning, n information bits are encoded in $N = 2^n - 1$ code bits with a code rate n/N. Secondly, such codes always maintain a minimum Hamming distance among them. Next, the dual to such a block code is called $(N, N-n)$ Hamming code, which is an important code in digital communications. Below, we state a few utilities of the stated properties of $\mathcal{S}$ in the form of *Lemmas*. The proofs to these *Lemmas* are not provided as it is easy to prove them using some of the properties given in the last section.

Lemma 1: The PN alphabet $\mathcal{S}$ forms a block code, closed under addition and multiplication over $GF(2)$, of length 2^n-1, span n, weight 2^{n-1} and minimum distance 2^{n-1}.

Lemma 2: The minimum distance between any two codes in N digit PN alphabet $\mathcal{S}$ is always greater than or equal to the minimum distance for any other N digit permutation alphabet $\mathcal{T}$.

Lemma 3: The probability that a transmitted sequence of an N digit PN alphabet $\mathcal{S}$ is incorrectly received as another sequence in $\mathcal{S}$ is always less than or equal to the corresponding probability for any N digit permutation alphabet $\mathcal{T}$.

IV. A Proposal to Distinguish PN Sequences from Bernoulli Sequences

Mainly, window, balance, run and ideal autocorrelation are the four properties which justify the name pseudo-random sequence as these are the properties which one would expect from the resultant sequence formed by tossing a fair coin $2^n - 1$ times (say, 'head' is assigned the value 1 and 'tail' is 0). In fact, these are the properties for which PN sequences are fairly useful in secured communications. Nevertheless, PN sequences are not truly random as all the stated 10 properties hold true for *all* PN sequences whereas in a fair coin tossing experiment (Bernoulli experiment), there must be some variations from sequence to sequence. This is why PN sequences are not suitable for serious encryption purposes. Many tests thus have been proposed, including the popular chi-square test that examines the uniformity of distribution, to distinguish PN sequences from truly random sequences such as Bernoulli sequence. Below, we provide a brief account of two such standard tests along with a simple proposal of a new randomness test which is quite easy for checking purpose. However, before dealing with these tests, we briefly discuss certain probabilistic values of these sequences some of which would be useful in the discussion of the randomness tests in the sequel.

A Bernoulli sequence is a discrete-time stochastic sequence consisting of a finite or infinite number of independent random variables $X_1, X_2, X_3, ...$, such that (a) for each i, the value of X_i is either 0 or 1; and (b) $\forall i$, the probability that $X_i = 1$ is the same number p_1. For sufficiently large number of trials,

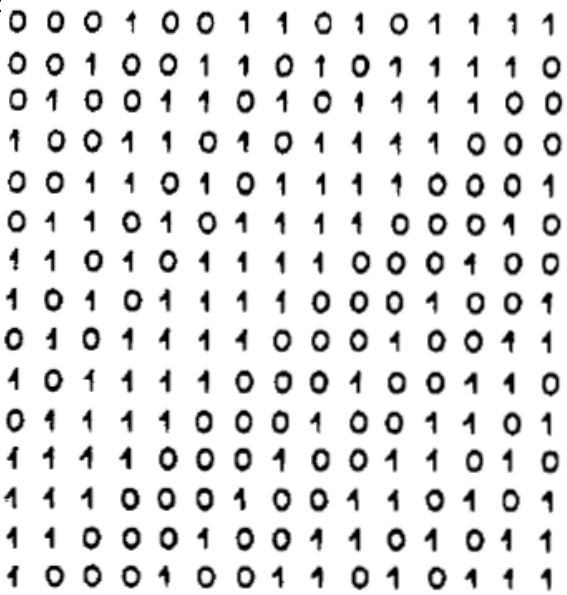

Fig. 2. The 15 possible PN sequences corresponding to the recurrence of Fig. 1.

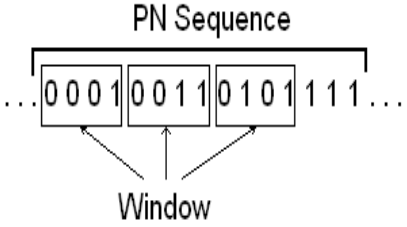

Fig. 3. The window or span property corresponding to the PN sequence obtained from Fig. 1.

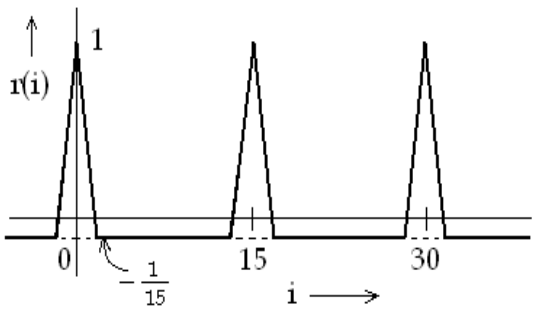

Fig. 4. The ACF of the PN sequence obtained from Fig. 1.

we assign p_1 and p_0 (the probability of $X_i = 0$) as $p_1 = p_0 = p = 1/2$ which is a fairly reasonable assumption. We assume this condition in the rest of this paper instead of taking $p_0 = (1 - p_1)$. Using this condition, we get the mean and variance of a Bernoulli sequence as: $E[X] = p_1 = 1/2$ and $Var[X] = p_1 p_0 = 1/4$, respectively. However, from *Property 7* of a PN sequence, we can write

$$p_1 = \frac{2^{n-1}}{2^n - 1}, \quad p_0 = \frac{2^{n-1} - 1}{2^n - 1} \tag{7}$$

where p_1 and p_0 denote the probability of getting a 1 and 0 respectively for the said PN sequence. From (7), we get

$$E[S] = \frac{2^{n-1}}{2^n - 1}, \quad Var[S] = \frac{2^{n-1}(2^{n-1} - 1)}{(2^n - 1)^2}. \tag{8}$$

A. Two Standard Tests

Among the many standard randomness tests available in literature, we describe here one time domain test, i.e., matrix rank test [4], and another frequency domain test, i.e., spectral test [7].

1) Matrix rank test: Let $s_0, s_1, ..., s_{m-1}$ be m consecutive binary digits from a PN sequence of length $N = 2^n - 1$, where $m < N$, which form the matrix

$$M = \begin{bmatrix} s_0 & s_1 & ... & s_{m-y} \\ s_1 & s_2 & ... & s_{m-y+1} \\ ... & & ... & \\ s_{y-1} & s_y & ... & s_{m-1} \end{bmatrix} \tag{9}$$

with $n < y < m/2$. Then $rank(M)$ over $GF(2)$ must always be less than y as, from *Property 1*, there exists only n linearly independent PN sequences in $\mathcal{S}$. If, on the other hand, $s_0, s_1, ..., s_{m-1}$ is a segment of a coin tossing sequence, then, $Pr\{rank(M) < y\} \leq 2^{2y-m-1}$ where $Pr\{rank(M) < y\}$ is the probability of $rank(M) < y$. This value becomes very small if $y \ll m/2$. The question "is $rank(M) = y$?" is thus a test on certain number of digits from any PN sequence to show a departure from true randomness. Here, as said above, we have taken $p_0 = p_1 = 1/2$.

2) Spectral test: This test is a sequential test which ultimately outputs whether a sequence is 'not random' or it 'may be random'. First, representing the binary string $s_0, s_1, ..., s_{N-1}$ with values ± 1 as in ACF, the fast Fourier transform (FFT) of the string is computed. The FFT consists of $\frac{N}{2} + 1$ independent values of the cosine transform while $\frac{N}{2} - 1$ independent values of the sine transform. These are represented as the indexed variables v_m and w_m. Next, $\frac{N}{2} - 1$ estimates of the periodogram is computed as $I_m = (1/N)[v_m^2 + w_m^2]$ and the expected value of r-th power of I_m, i.e., $E[I_m^r]$ for $r \geq 2$, is deduced within the range $m \leq (N/2) - 1$. It can be shown that excluding the boundary expectation values, i.e., $E[I_0^r]$ and $E[I_{N/2}^r]$, all the other mean values $E[I_m^r]$ are almost same if the input sequence is a true random sequence. On the other hand, for any PN sequence, $E[I_m^r]$ will have divergent values for different order moments and different m. Readers are suggested to test it on $Var[I_m]$.

B. The Proposal: One Simple Test

Although for less number of digits, one can simply calculate $E[S]$ and $Var[S]$ as given in (8) that would yield same value for all PN sequences all the time whereas for truly random sequences these will go on fluctuating; one cannot rely on these measurements as they almost converge to $E[X]$ and $Var[X]$ for higher values of n.

Our proposed test is a direct consequence of *Property 7* (balance) which doesn't require any of these measurements. It is also simple in the sense that it neither needs to compute the rank of any matrix nor to compute the periodogram or Fourier transforms. *Property 7*, in other words, can be stated as 'the difference of number of 1's and number of 0's in a single period of any PN sequence is always 1'. As the all valid PN sequences are shifted versions of any basic sequence (by *Property 3*), following the proof of *Property 7* it also then means 'number of 0's cannot be more than the number of 1's in a single period of any PN sequence'. Thus, if we want to find the probability of "number of 2^{n-1} zeros in a single period of $2^n - 1$ length PN sequence", the outcome should always be zero as this is an absurd event. However, for any truly random sequence like Bernoulli sequence, the probability of getting 2^{n-1} zeros in $2^n - 1$ independent binary number outcomes is given as $\binom{2^n-1}{2^{n-1}} p^{2^{n-1}} (1-p)^{2^n - 2^{n-1} - 1}$, where it is assumed that p is probability of getting a single zero. As earlier, putting $p = 1 - p = \frac{1}{2}$, we get the value of this expression as

$$Pr\{0^{2^{n-1}} 1^{2^{n-1}-1}\} = \frac{1}{2^{2^n+n-2}} \frac{(2^n - 1)!}{(2^{n-1} - 1)!^2} \tag{10}$$

where the L.H.S. means the probability of getting 2^{n-1} zeros and $2^{n-1}-1$ ones. Calculations show that for $N = 3$, $Pr\{.\} \approx 0.4$; for $N = 7$, $Pr\{.\} \approx 0.3$; for $N = 15$, $Pr\{.\} \approx 0.2$; for $N = 31$, $Pr\{.\} \approx 0.1$, and so on. This in turn means that after certain iterations, for any PN sequence the required probability will be zero but for a truly random sequence, the value will go on fluctuating with a value greater than zero and ultimately will converge to the theoretical value for a large number of digits. It is easy to test the proposed approach in any standard software.

V. Concluding Remarks

In this paper, we have provided a simple and comprehensive discussion of all of the properties of PN sequences, including the less known properties like window, shift-and-add or closure, along with the respective proofs towards the better understanding of such sequences. A few applications of these properties in communication theory have also been given in the form of *Lemmas* without proving them as the the task becomes trivial with the knowledge of the discussed properties. Lastly, a simple randomness test is proposed in order to distinguish the said PN sequences from any fair coin tossing experiment, i.e., Bernoulli sequence. The proposed test is effective yet simple in comparison with the standard tests existing in the literature and thus can be easily taken up by the undergraduate students as an assignment.

References

[1] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*. Boston, MA: Kluwer Academic, 1987.
[2] S. Golomb, *Shift Register Sequences, Revised edition*. Laguna Hills, CA: Aegean Park Press, 1982.
[3] S. Blackburn, "A Note on Sequences with the Shift and Add Property," *Designs, Codes, and Crypt.*, vol. 9, pp. 251-256, 1996.
[4] F. J. MacWilliams and J. A. Sloane, "Pseudo-Random Sequences and Arrays," *Proc. IEEE*, vol. 64, no. 12, pp. 1715-1729, Dec. 1976.
[5] S. A. Fredricsson, "Pseudo-Randomness Properties of Binary Shift Register Sequences," *IEEE Trans. Inform. Theory*, vol. 21, pp. 115-120, Jan. 1975.
[6] D. E. Knuth, *The Art of Computer Progamming*. Reading, MA: Addison-Wesley, 1968.
[7] F. A. Feldman, "A New Spectral Test for Nonrandomness and the DES," *IEEE Trans. Soft. Engg.*, vol. 16, no. 3, pp. 261-267, March 1990.
[8] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

Abhijit Mitra was born in Serampore, India, in 1975. He received the B.E.(Honors) degree from the Regional Engineering College, Durgapur, India, in 1997, the M.E.Tel.E. degree from Jadavpur University, India, in 1999 and the Ph.D. degree from the Indian Institute of Technology, Kharagpur, India, in 2004, all in electronics and communication engineering.

Since 2004, he has been with the Department of Electronics and Communication Engineering at the Indian Institute of Technology, Guwahati, India, as an Assistant Professor. He visited Indian Statistical Institute (ISI), Kolkata, as a Visiting Scientist during June-July and December 2007. He is the recipient of URSI Young Scientist Award and INAE Summer Fellowship for Young Engineering Teachers, both for 2008. He has also been elected as an Associate of Indian Academy of Sciences (IAS) for 2008-2011. His research interests include adaptive signal processing and signal processing applications in wireless communications with the primary emphasis on low complexity realizations.

Dr. Mitra has been a member of IEEE since 2003 and presently serves as a reviewer of many IEEE Transactions. He is also a member of Indian Science Congress Association and Institution of Electronics and Telecommunication Engineers,