Statistical Span Property of Binary Run Sequences

Gangsan Kim, *Graduate Student Member, IEEE*, and Hong-Yeop Song, *Senior Member, IEEE*

***Abstract*— We define run sequences of period $2^n - 1$ as the binary sequences where the distribution of runs of 0's and runs of 1's is exactly same as that for the maximal length linear shift resister sequences of period $2^n - 1$. We first count the number of all the cyclically distinct run sequences of period $2^n - 1$. For each n-tuple, we consider the average number of occurrences over all the run sequences of period $2^n - 1$. We identify the n-tuples with average number 1 and, in particular, those that occur exactly once in every run sequence of period $2^n - 1$. We finally prove that, as n increases, the average number of every non-zero n-tuple approaches to 1.**

***Index Terms*— Binary sequences, pseudo-random sequences, run sequences, span sequences, de Bruijn sequences.**

I. Introduction

PSEUDO-RANDOM sequences are useful in various fields such as communication and cryptographic systems [5], [7], [16]. Many researchers have investigated various necessary and sufficient conditions for pseudo-random sequences. S. W. Golomb postulated three randomness properties: balance, run and ideal autocorrelation [5], [9]. In addition, there are span and multiplier properties [5], [6]. In this paper, we consider the binary sequences repeating periodically mostly with period $2^n - 1$ and we focus on the run property of binary sequences of period $2^n - 1$.

The above randomness properties can be defined by the corresponding properties of the m-sequences of period $2^n - 1$. An m-sequence is a maximal-length linear feedback shift register sequence satisfying a linear recursion whose characteristic polynomial is primitive over $\mathbb{F}_2$ [5], [7]. The number of cyclically distinct m-sequences of period $2^n - 1$ is shown in Table I for $n = 3, 4, 5$ and 6, where $\phi(\cdot)$ is the Euler's-phi-function [5], [7]. There have been a lot of studies about m-sequences as well as, in general, balanced binary sequences with ideal autocorrelation, multiplier and/or span properties [1], [2], [4], [5], [6], [7], [8], [9], [10], [14], [15], [16], [17], [18]. As far as both authors are aware of, there have not been much results on the run properties of binary sequences, or their relation with other randomness properties.

A binary sequence of period $2^n - 1$ is called a span sequence if all the non-zero n-tuples occur exactly once in its one period [5], [9]. We claim that all the span sequences of period $2^n - 1$ are in one-to-one correspondence with all the de Bruijn sequences of period 2^n. By inserting one 0 right after the longest run of 0's in the span sequence of period $2^n - 1$, we obtain a de Bruijn sequence of period 2^n. Conversely, by deleting one 0 in the longest run of 0's in the de Bruijn sequence of period 2^n, we obtain a span sequence of period $2^n - 1$. Therefore, the number s_n of cyclically distinct span sequences must be the same as that of de Bruijn sequences, which is given by [1] and [8]

$$s_n = 2^{2^{n-1}-n}$$

and is shown in Table I for $n = 3, 4, 5$ and 6. The (exhaustive) constructions and classification of s_n de Bruijn sequences have been studied a lot [8], [14], [15].

A run of 0's of length k in a binary sequence is a string of consecutive k 0s flanked by 1 and a run of 1's of length k is a string of consecutive k 1s flanked by 0 where the run is considered cyclically wrapped around since the sequence is considered periodically repeating [5], [6], [9], [11]. For example, when the ending of one period is a block $\ldots 011$ and the beginning is a block $1110\ldots$, we consider this a (final) run of 1's of length 5 of the sequence. A binary sequence of period $2^n - 1$ is said to have the run property and called a run sequence if its run distribution is the same as those of an m-sequence of the same period. In an m-sequence of period $2^n - 1$, there are 2^{n-2-i} runs of 0's of length i and 2^{n-2-i} runs of 1's of length i, for $i = 1, 2, \ldots, n-2$, with a single run of 0's of length $n-1$ and a single run of 1's of length n [5], [6], [9], [11]. Table II summarize the run distribution of an m-sequence of period $2^n - 1$. The number r_n of cyclically distinct run sequences is shown also in Table I for $n = 3, 4, 5$ and 6. This number was first determined in 2019 by the same authors and presented in a conference [11], but is revived here in full capacity by Theorem 1 in the beginning of Section II.

It is well-known that any span sequence must be a run sequence but not conversely [5], [7], [16]. Therefore, a span sequence is a special type of a run sequence. Any run sequence of period $2^n - 1$ must belong to one of the following types: a run sequence which is not a span sequence; a span sequence which is not an m-sequence; or an m-sequence. For example, of period 15, we have

1) a run sequence 000111100110101 which is not a span sequence (1011 is missing or 1010 appears twice);

Manuscript received 20 March 2022; revised 30 October 2022; accepted 13 November 2022. Date of publication 16 November 2022; date of current version 17 March 2023. This work was supported by the National Research Foundation of Korea (NRF) Grant through the Korea Government (MSIT) under Grant 2020R1A2C2011969. An earlier version of this paper was presented in part at the 9th International Workshop on Signal Design and its Applications in Communications (IWSDA 2019), Dongguan, China, October 2019 [DOI: 10.1109/IWSDA46143.2019.8966121]. *(Corresponding author: Hong-Yeop Song.)*

The authors are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, South Korea (e-mail: gs.kim@yonsei.ac.kr; hysong@yonsei.ac.kr).

Communicated by G. Kyureghyan, Associate Editor At Large for Sequences and Cryptography.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3222527.

Digital Object Identifier 10.1109/TIT.2022.3222527

TABLE I
THE NUMBER OF CYCLICALLY DISTINCT BINARY SEQUENCES WITH SOME RANDOMNESS PROPERTY

n	2^n-1	# m-sequences $= \frac{\phi(2^n-1)}{n}$	# Span sequences ($\triangleq s_n$)	# Run sequences ($\triangleq r_n$)
3	7	2	2	2
4	15	2	16	36
5	31	6	2,048	88,200
6	63	6	67,108,864	7,304,587,290,000

TABLE II
RUN DISTRIBUTION OF AN M-SEQUENCE OF PERIOD 2^n-1

run-length	# runs of 1's	# runs of 0's
n	1	0
$n-1$	0	1
$n-2$	2^0	2^0
$n-3$	2^1	2^1
$\vdots$	$\vdots$	$\vdots$
1	2^{n-3}	2^{n-3}
Total	2^{n-2}	2^{n-2}

2) a span sequence 000111100101101 which is not an m-sequence; or
3) an m-sequence 000111101011001 (a linear recursion $a_k = a_{k-1} + a_{k-4}$ for $k \geq 4$ is satisfied).

This gives a motivation of some investigation on the questions: which run sequences are span sequences or else why some run sequences fail to be a span sequence. This paper would like to provide some answers to these questions.

In this paper, we first count the number r_n of all the cyclically distinct binary run sequences of period 2^n-1. Then, we reports some results on the distribution of various n-tuples throughout all the cyclically distinct run sequences of period 2^n-1. We define the average number of occurrences of an n-tuple as the total number of occurrences in the set of all such run sequences divided by the number r_n. It is interesting to find that some n-tuples occur exactly once in all the run sequences so that not only the average number is one but the actual occurrence is exactly 1 "uniformly" over each and every run sequence. On the other hand, some n-tuples occur more than once in some sequences but the total number counts r_n so that the average number becomes 1. We may say in this case that the average number 1 is achieved "non-uniformly." For the remaining n-tuples, average number of occurrence turned out be different from 1.

The main result of this paper is summerized as two theorems: Theorem 1 counts the number of cyclically distinct run sequences of period 2^n-1 and Theorem 2 proves that the average number of occurrences of each and every n-tuple (determined in Lemma 1) approaches 1 as n increases, which we call *the statistical span property satisfied by the run sequences*.

II. MAIN RESULTS

First, we count the number of cyclically distinct binary run sequences of period 2^n-1. We assume that each of the cyclically equivalent classes is represented by the run sequence in which the unique run of 0's of length $n-1$ occurs in the beginning. Then, we set up a one-to-one correspondence between the set of all the representatives and the pairs of multiset permutations.

Definition 1 (Multiset Permutations [12]): Let $a_1, \ldots, a_k$ be some k distinct symbols and $n_1, n_2, , \ldots, n_k$ be some nonnegative integers. Define $\mathcal{R}(a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k})$ as the set of all the multiset permutations of length $N = n_1 + n_2 + \cdots + n_k$ that are formed by rearranging n_i copies of a_i for $i = 1, 2, \ldots, k$.

Example 1:

$$\mathcal{R}(0^2 1^1) = \{001, 010, 100\}.$$
$$\mathcal{R}(a^1 b^2 c^1 d^0) = \{abbc, abcb, acbb, babc, bacb, bbac, bbca, bcab, bcba, cabb, cbab, cbba\}.$$

We see that $|\mathcal{R}(0^2 1^1)| = 3$ and $|\mathcal{R}(a^1 b^2 c^1 d^0)| = 12$ in Example I. In general, the number of the multiset permutations in Definition I is given as a multinomial coefficient:

$$|\mathcal{R}(a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k})| = \binom{N}{n_1, n_2, \ldots, n_k} = \frac{N!}{n_1! n_2! \cdots n_k!}.$$

We now count the number of cyclically distinct run sequences of period 15, for example. The key observation is that the runs of 0's and the runs of 1's must alternate in any binary sequence. Each run sequence is appropriately cyclically-shifted so that the unique run 000 becomes the beginning. Since the number of runs of 1's is the same as that of runs of 0's in a run sequence of period 15, when the run sequence is cyclically shifted so that the beginning becomes 000, the ending must a run of 1's of some positive length. Then we can set up a one-to-one correspondence between the set of all such run sequences and the pairs of the multiset permutations, corresponding to the runs of 0's and the runs of 1's. For example, denoting by 0_k (or 1_k, resp.) a string of consecutive 0's (or 1's, resp.) of length k, we consider the following run sequence of length 15:

$$0_{x_1} 1_{y_1} 0_{x_2} 1_{y_2} 0_{x_3} 1_{y_3} 0_{x_4} 1_{y_4}.$$

By the assumption, we have $x_1 = 3$. By the definition, the integers x_2, x_3, x_4 must be some rearrangement of the run lengths $1, 1, 2$. Similarly, for the runs of 1's, the lengths

y_1, y_2, y_3, y_4 must be some rearrangement of the run lengths $1, 1, 2, 4$. For example, one may take such rearrangements as

$$(x_2, x_3, x_4) = (1, 2, 1) \quad \text{and} \quad (y_1, y_2, y_3, y_4) = (2, 1, 4, 1),$$

corresponding to one of the run sequences

$$0_3 1_2 0_1 1_1 0_2 1_4 0_1 1_1.$$

We now consider a pair of multiset permutations consisting of two distinct sets of symbols $\{a_1, a_2\}$ and $\{b_1, b_2, b_4\}$, corresponding to the lengths $\{1, 2\}$ of run of 0's and the lengths $\{1, 2, 4\}$ of run of 1's, respectively. Then the corresponding pair of the multiset permutations would be

$$(a_1 a_2 a_1, b_2 b_1 b_4 b_1) \in \mathcal{R}(a_1^2 a_2^1) \times \mathcal{R}(b_1^2 b_2^1 b_4^1).$$

It is not difficult to see that all the run sequences of period 15 (with the unique run 0_3 in the beginning) are in one-to-one correspondence with the members of $\mathcal{R}(a_1^2 a_2^1) \times \mathcal{R}(b_1^2 b_2^1 b_4^1)$. Therefore, the total number of cyclically distinct run sequences of period 15 is given by the size of the set $\mathcal{R}(a_1^2 a_2^1) \times \mathcal{R}(b_1^2 b_2^1 b_4^1)$ which is 36. This can be easily generalized to the period $2^n - 1$ and this proves the following theorem:

Theorem 1: The number of cyclically distinct binary run sequences of period $2^n - 1$, denoted by r_n, is given by

$$r_n = \frac{1}{2^{n-2}} \binom{2^{n-2}}{2^{n-3}, 2^{n-4}, \ldots, 2^0, 1}^2. \tag{1}$$

Proof: All the cyclically distinct binary run sequences of period $2^n - 1$ are in one-to-one correspondence with the members of

$$\mathcal{R}(a_1^{2^{n-3}} a_2^{2^{n-4}} \cdots a_{n-2}^{2^0}) \times \mathcal{R}(b_1^{2^{n-3}} b_2^{2^{n-4}} \cdots b_{n-2}^{2^0} b_n^1).$$

Note that the symbol b_{n-1} is missing in the second set of multiset permutations on the list of symbols $b_1, b_2, \ldots, b_{n-2}, b_n$ because a run of 1's of length $n-1$ is not in the run sequence as can be seen in Table II. ∎

Corollary 1: Let r_n be as given in Theorem 1. Then, as n increases indefinitely,

$$r_n \quad \longrightarrow \quad c 2^{2^n - \frac{1}{2}n^2 + \left(\frac{3}{2} - \log_2 \pi\right)n}, \tag{2}$$

for some constant c in the range $e^{-1/6} \leq 8(\frac{e}{\pi})^4 c < 1$ or $0.189 \leq c < 0.223$, and hence,

$$\frac{r_n}{r_{n-1}} \quad \longrightarrow \quad 2^{2^{n-1} - n + 2 - \log_2 \pi}.$$

Proof: It is easy and straightforward using the inequality [13]

$$\sqrt{2\pi m}\left(\frac{m}{e}\right)^m < m! < \sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{\frac{1}{12m}}$$

for any positive integer m, and the stirling's approximation [3],

$$\lim_{m \to \infty} \frac{m!}{\sqrt{2\pi m}\left(\frac{m}{e}\right)^m} = 1.$$

∎

Corollary 2: Let r_n and s_n be the number of cyclically distinct run and span sequences of period $2^n - 1$, respectively. Then, as n increases indefinitely,

$$\frac{r_n}{s_n} \quad \longrightarrow \quad c 2^{2^{n-1} - \frac{1}{2}n^2 + \left(\frac{5}{2} - \log_2 \pi\right)n}, \tag{3}$$

where the same constant c is given in (2).

Definition 2: Consider the set of all the cyclically distinct r_n binary run sequences of period $2^n - 1$. For any binary n-tuple, we count the number N of occurrences of this n-tuple throughout all these run sequences. We define the average number of occurrences (or, the average number, for short) of this n-tuple as N/r_n.

Now, we investigate the n-tuple distribution property over all the cyclically distinct r_n binary run sequences of period $2^n - 1$, where r_n is given in Theorem 1. Figure 1 (a) and (b) show the average number of occurrences of each n-tuple in all the cyclically distinct run sequences of period $2^n - 1$, for $n = 5$ in (a) and $n = 6$ in (b). The average number of an n-tuple is given in Definition II. It is surprising that there exist some not-all-zero n-tuples such that their average numbers are equal to 1. It is interesting to observe that there are only three values as an average number, which are $6/7$, $7/7 = 1$ and $8/7$ for $n = 5$ in (a). On the other hand, there are six values for $n = 6$ in (b) which are $196/225, 210/225, 224/225$, $225/225 = 1$, $240/225$ and $256/225$. We first observe an obvious lower bound on the average number of any not-all-zero n-tuple:

Proposition 1: Let r_n and s_n be the number of cyclically distinct binary run and span sequences of period $2^n - 1$, respectively. Then, the average number of any not-all-zero n-tuple is lower bounded by s_n/r_n.

The above lower bound is obtained since any not-all-zero n-tuple appears exactly once at least on the s_n span sequences. One may wonder how many times an n-tuple appears throughout all the run sequences including span sequences. In the remaining of this paper, we prove that this number approaches r_n so that the average number approaches 1 as n increases indefinitely. Before this journey, we would like to mention one more property of these average numbers of all the not-all-zero n-tuples. From Fig. 1(a), we observe that the sum of all the average numbers of all the not-all-zero 5-tuples becomes $2^5 - 1$ so that the average (over all the 5-tuples) of all these average numbers becomes 1. This can be proved in general for any positive integer n.

Proposition 2: For $k = 1, 2, \ldots, 2^n - 1$, denote by A_k the average number of the binary n-tuple whose decimal representation is k. Here, the average is over all the cyclically distinct binary run sequences. Then

$$\sum_{k=1}^{2^n - 1} A_k = 2^n - 1.$$

Proof: Let all the cyclically distinct run sequences be ordered somehow such that we may call an i-th run sequence for $i = 1, 2, \ldots, r_n$. Denote by $B_{k,i}$ the number of occurrences of the n-tuple k in the i-th run sequence. Then,

$$A_k = \frac{1}{r_n} \sum_{i=1}^{r_n} B_{k,i}.$$

Therefore,

$$\sum_{k=1}^{2^n - 1} A_k = \sum_{k=1}^{2^n - 1} \frac{1}{r_n} \sum_{i=1}^{r_n} B_{k,i} = \frac{1}{r_n} \sum_{i=1}^{r_n} \sum_{k=1}^{2^n - 1} B_{k,i} = 2^n - 1.$$

∎

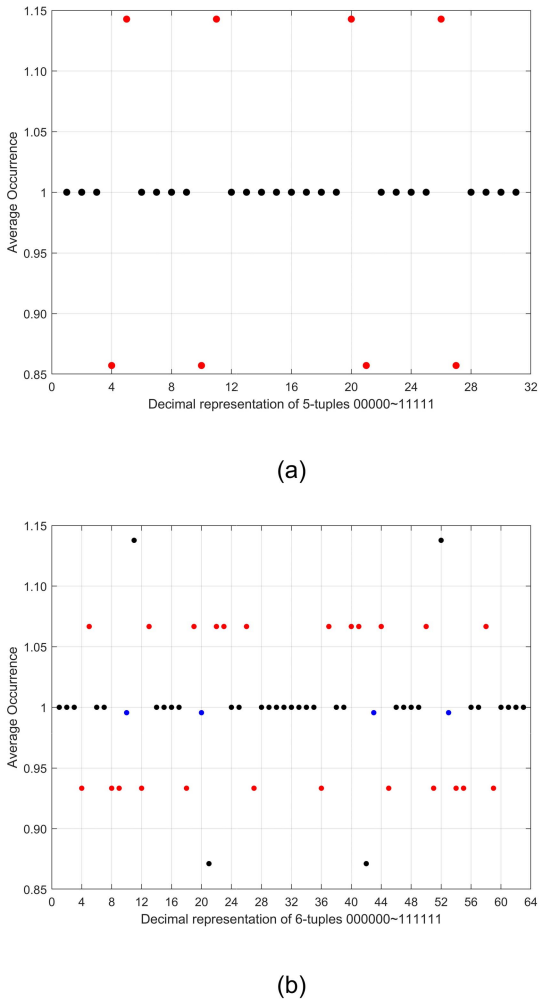


Fig. 1. The average number of occurrences of (a) 5-tuples and (b) 6-tuples.

The average number 1 of an n-tuple can be obtained if every run sequence contains the n-tuple exactly once (uniform case) or if some contains more than once but the total number throughout all the run sequences is r_n (non-uniform case). It would be interesting to identify which n-tuples have the average number 1 in uniform way. Now, Proposition 3 identifies those n-tuples with the average number 1 uniformly. Recall that we denote by 0_k (or 1_k, resp.) a set of consecutive 0's (or 1's, resp.) of length k. We use a run 0_k or a run 1_k if it is flanked by the other symbol. That is, a '0_k' can be different from a 'run 0_k.'

Proposition 3: An n-tuple has the average number 1 uniformly if and only if it is one of the following seven cases:

$$a0_{n-2}b \quad \text{and} \quad a1_{n-2}b,$$

where $a, b \in \{0, 1\}$, except for the all-zero n-tuple.

Proof: For sufficiency, consider any one run sequence of period $2^n - 1$. Then, we observe the following unique occurrence of each of the above seven n-tuples:

- An n-tuple $10_{n-2}1$ occurs once since there exists a unique run 0_{n-2}.
- Each of two n-tuples 10_{n-1} and $0_{n-1}1$ occurs once since there exists a unique run 0_{n-1} and no run 0_k for $k \geq n$.
- An n-tuple $01_{n-2}0$ occurs once since there exists a unique run 1_{n-2}.
- Each of three n-tuples 01_{n-1}, $1_{n-1}0$, and 1_n occurs once since there exists a unique run 1_n and no run 1_k for $k = n-1$ or $k > n$.

We omit the proof of necessity since it is a straightforward but complicated application of the run property again. ■

We now determine the average number of each of the n-tuples. Any n-tuple may begin by either 0 or 1 and end by 0 or 1, and hence, we may consider its four different forms in the following:

	generic form of an n-tuple
Case 1	$0_x 1_{\alpha_1} 0_{\beta_1} 1_{\alpha_2} 0_{\beta_2} \cdots 1_{\alpha_m} 0_{\beta_m} 1_y$
Case 2	$1_y 0_{\beta_1} 1_{\alpha_1} 0_{\beta_2} 1_{\alpha_2} \cdots 0_{\beta_m} 1_{\alpha_m} 0_x$
Case 3	$0_x 1_{\alpha_1} 0_{\beta_1} 1_{\alpha_2} 0_{\beta_2} \cdots 1_{\alpha_m} 0_y$
Case 4	$1_x 0_{\beta_1} 1_{\alpha_1} 0_{\beta_2} 1_{\alpha_2} \cdots 0_{\beta_m} 1_y$

(4)

where x, y are some positive integers and where α_k, β_k are some positive run-lengths inside the n-tuple except that $\beta_m = 0$ in Case 3 and $\alpha_m = 0$ in Case 4. With this assumption, we must have

$$x + y + \sum_{k=1}^{m} (\alpha_k + \beta_k) = n. \tag{5}$$

Note that the initial and final consecutive symbols can be a part of some longer runs according to the symbols flanking the n-tuple. For example, consider an n-tuple of Case 1. If 1 comes just before this n-tuple, then 0_x is a run 0_x. However, if 0 comes, then 0_x is a part of a run of 0's of length $x + 1$ or more.

Lemma 1: Consider a binary not-all-zero n-tuple with the notation and assumption leading to (4) and (5). Let p_k be the number of runs 1_k's and z_k be the number of runs 0_k's in the n-tuple except for $0_x, 0_y, 1_x, 1_y$ in either the beginning or the ending. Then, the average number of the n-tuple is given by the following:

Case 1 and Case 2:

$$M \times \frac{2^{n-x-1} - \sum_{k=x}^{n-2} z_k}{2^{n-2} - \sum_{k=1}^{n-2} z_k} \times \frac{2^{n-y-1} - \sum_{k=y}^{n-2} p_k}{2^{n-2} - \sum_{k=1}^{n-2} p_k},$$

Case 3:

$$M \times \frac{2^{n-\max(x,y)-1} - \sum_{k=\max(x,y)}^{n-2} z_k}{2^{n-2} - \sum_{k=1}^{n-2} z_k} \times \frac{2^{n-\min(x,y)-1} - 1 - \sum_{k=\min(x,y)}^{n-2} z_k}{2^{n-2} - 1 - \sum_{k=1}^{n-2} z_k},$$

Case 4:

$$M \times \frac{2^{n-\max(x,y)-1} - \sum_{k=\max(x,y)}^{n-2} p_k}{2^{n-2} - \sum_{k=1}^{n-2} p_k} \times \frac{2^{n-\min(x,y)-1} - 1 - \sum_{k=\min(x,y)}^{n-2} p_k}{2^{n-2} - 1 - \sum_{k=1}^{n-2} p_k},$$

where

$$M \triangleq \frac{1}{r_n} \times \frac{\left(2^{n-2} - \sum_{k=1}^{n-2} z_k\right)!}{\prod_{k=1}^{n-2} \{(2^{n-2-k} - z_k)!\}} \times \frac{\left(2^{n-2} - \sum_{k=1}^{n-2} p_k\right)!}{\prod_{k=1}^{n-2} \{(2^{n-2-k} - p_k)!\}}. \quad (6)$$

Proof: We prove the theorem for Case 1 in (4), where the n-tuple is of the form $0_x 1_{\alpha_1} 0_{\beta_1} 1_{\alpha_2} 0_{\beta_2} \cdots 1_{\alpha_m} 0_{\beta_m} 1_y$. All other cases can be treated similarly. Consider any n-tuple of the form

$$0_x \underline{1_{\alpha_1} 0_{\beta_1} 1_{\alpha_2} 0_{\beta_2} \cdots 1_{\alpha_m} 0_{\beta_m}} 1_y. \quad (7)$$

We count the number of occurrences of this n-tuple in all the cyclically distinct binary run sequences of period $2^n - 1$. It can be determined by counting the number of run sequences (not necessarily cyclically distinct) which contain this n-tuple in the beginning. We count this number in two steps. The first counts the run sequences in which the underlined part of the pattern in (7) together with a single preceding 0 and succeeding 1 appears as $n - (x + y) + 2$ consecutive bits in the same positions as (7). In the second step, we obtain the final answer by considering 0_x in the beginning and 1_y in the ending of the n-tuple.

Note that the underlined part of (7) contains $\sum_{k=1}^{n-2} p_k$ runs of 1's and $\sum_{k=1}^{n-2} z_k$ runs of 0's from the assumption. Now the run distribution of the remaining part of the sequences must contain the runs and their frequencies as given in Table III. Similar to the counting in Theorem 1, the number of run sequences containing the underlined part of the pattern (7) from $(x+1)^{th}$ to $(n-y)^{th}$ bit positions in the beginning is given as the size of the set of multiset permutation pairs:

$$\begin{aligned} &\left|\mathcal{R}\left(a_1^{2^{n-3}-z_1} a_2^{2^{n-4}-z_2} \cdots a_{n-2}^{2^0-z_{n-2}} a_{n-1}^1\right)\right| \\ &\times \left|\mathcal{R}\left(b_1^{2^{n-3}-p_1} b_2^{2^{n-4}-p_2} \cdots b_{n-2}^{2^0-p_{n-2}} b_n^1\right)\right| \\ &= \frac{\left(2^{n-2} - \sum_{k=1}^{n-2} z_k\right)!}{\prod_{k=1}^{n-2} (2^{n-2-k} - z_k)!} \times \frac{\left(2^{n-2} - \sum_{k=1}^{n-2} p_k\right)!}{\prod_{k=1}^{n-2} (2^{n-2-k} - p_k)!} = r_n M. \end{aligned} \quad (8)$$

Note that this number counts those run sequences in which the first x terms may contain some 1's or the last y terms may contain some 0's inside the n-tuple in the beginning. Therefore, we have to consider only those portions by multiplying the ratio of those containing 0_x in the beginning and 1_y in the ending of the first n positions. This ratio turns out to be given as the probability of having 0_x as a part of a run of 0's of length x or more and having 1_y as a part of a run of 1's of length y or more, preceding and succeeding the underlined part of the pattern (7), respectively. It is given as

$$\frac{1 + \sum_{k=x}^{n-2} \left(2^{n-2-k} - z_k\right)}{2^{n-2} - \sum_{k=1}^{n-2} z_k} \times \frac{1 + \sum_{k=y}^{n-2} \left(2^{n-2-k} - p_k\right)}{2^{n-2} - \sum_{k=1}^{n-2} p_k}. \quad (9)$$

Multiplying the above two numbers in (8) and (9) and then dividing the result by r_n gives the final expression in the theorem for Case 1. ∎

TABLE III
RUN DISTRIBUTION EXCEPT FOR THE RUNS IN $1_{\alpha_1} 0_{\beta_1} 1_{\alpha_2} 0_{\beta_2} \cdots 1_{\alpha_m} 0_{\beta_m}$

run-length	# runs of 1's	# runs of 0's
n	1	0
$n-1$	0	1
$n-2$	$2^0 - p_{n-2}$	$2^0 - z_{n-2}$
$n-3$	$2^1 - p_{n-3}$	$2^1 - z_{n-3}$
$\vdots$	$\vdots$	$\vdots$
1	$2^{n-3} - p_1$	$2^{n-3} - z_1$
Total	$2^{n-2} - \sum_{k=1}^{n-2} p_k$	$2^{n-2} - \sum_{k=1}^{n-2} z_k$

Example 2: We calculate the average number of the following two 5-tuples (all in Case 1) using the formula in Lemma 1. First, the number r_5 is given by

$$r_5 = \frac{1}{2^3}\left(\frac{2^3!}{2^2!2^1!2^0!}\right)^2 = 88200.$$

1) 5-tuple 00011: $p_k = z_k = 0$ for all k and $x = 3, y = 2$. Therefore, the average number becomes $\frac{1}{8}M$ which is given by

$$\frac{1}{8r_5}\left(\frac{2^3!}{2^2!2^1!2^0!}\right)^2 = 1.$$

2) 5-tuple 01011: $p_1 = z_1 = 1$ and $x = 1, y = 2$. Therefore, the average number becomes $\frac{4}{7}M$ which is given by

$$\frac{4}{7r_5}\left(\frac{(2^3-1)!}{(2^2-1)!2^1!2^0!}\right)^2 = \frac{8}{7}.$$

Proposition 4 (Corollary to Lemma 1): The average number is 1 only for the following n-tuples:

$$a0_k 1_{n-2-k} b \text{ and } a1_{n-2-k} 0_k b,$$

where $a, b \in \{0, 1\}$ and $k = 0, 1, \ldots, n-2$, except for the all-zero n-tuple. Of these, only those seven n-tuples in Proposition 3 achieve the average number 1 uniformly, and all others achieve non-uniformly.

Table IV shows all the twenty-three 5-tuples described in Prop. II having the average number 1. These 5-tuples appear also in Fig. 1 (a) as 23 black dots. Of these, only those seven 5-tuples in Prop. 3 achieve the average number 1 uniformly.

Proposition II describes that only a few n-tuples have the average number 1 for any given finite value of n. However, in Theorem 2, we show that all the n-tuples except for the all-zero have the average number 1 as n increases indefinitely. That is, Theorem 2 shows that both supremum and infimum of the set of all the possible average numbers for all the non-zero n-tuples approach to 1. For this, we proceed by separating all the non-zero n-tuples into 4 cases defined in (4).

Theorem 2: Let $\mathcal{E}_n, \mathcal{F}_n, \mathcal{G}_n$ and $\mathcal{H}_n$ be the sets of all the average numbers for all the non-zero n-tuples which have the

TABLE IV
ALL THE 5-TUPLES WITH THE AVERAGE NUMBER 1 IN PROPOSITIONS 3 AND II

n-tuples	comment
00011, 01001, 00111, 01101, 11100, 10110, 11000, 10010, 00010, 01000, 00110, 01100, 11101, 10111, 11001, 10011	non-uniform, Proposition 4
01110, 00001, 01111, 10000, 11110, 10001, 11111	uniform, Propositions 3 and 4

form of Case 1,2,3 and 4 in (4), respectively, throughout all the cyclically distinct run sequences of period 2^n-1. Then

$$\lim_{n\to\infty}(\sup\mathcal{E}_n)=\lim_{n\to\infty}(\inf\mathcal{E}_n)=1.$$

Similarly for $\mathcal{F}_n,\mathcal{G}_n$ and $\mathcal{H}_n$.

Proof: In this proof, we only show that both $\sup\mathcal{E}_n$ and $\inf\mathcal{E}_n$ approach to 1 as n increases. All other cases can be treated similarly. We choose any binary n-tuple in Case 1 of (4). Then the parameters x,y,α_i,β_i and $i=1,2,\ldots,m$ are determined from this n-tuple. They satisfy the relation (5). We use the same notation z_k,p_k as in the proof of Lemma 1. Then we have

$$\sum_{j=1}^{m}\beta_j=\sum_{k=1}^{n-2}kz_k \quad\text{and}\quad \sum_{j=1}^{m}\alpha_j=\sum_{k=1}^{n-2}kp_k.$$

Therefore, (5) becomes

$$x+y+\sum_{k=1}^{n-2}k(z_k+p_k)=n, \tag{10}$$

since the underlined part of the n-tuple in (7) does not have any run of length $n-1$ or more.

By slightly modifying the formula in Lemma 1 for Case 1, the average number of the n-tuple can be expressed as follows:

$$\frac{M}{2^{x+y-2}}\times\left(2^{x-1}\frac{2^{n-x-1}-\sum_{k=x}^{n-2}z_k}{2^{n-2}-\sum_{k=1}^{n-2}z_k}\right)\times\left(2^{y-1}\frac{2^{n-y-1}-\sum_{k=y}^{n-2}p_k}{2^{n-2}-\sum_{k=1}^{n-2}p_k}\right). \tag{11}$$

For convenience, denote each of the above three terms by M_n, Z_n and P_n, respectively. In the remaining of this proof, we show that some upper bounds and lower bounds for each of these M_n, Z_n and P_n approach to 1 as n increases.

First, we consider the expressions Z_n and P_n. We take care of the value Z_n and then the expression P_n can be treated similarly. Claim that

$$\frac{2^{n-2}-n2^{\frac{n}{2}-2}}{2^{n-2}}<Z_n<\frac{2^{n-2}}{2^{n-2}-\frac{n}{2}}.$$

Then, both upper and lower bounds above approach to 1 as n increases. To prove this claim, we distinguish two cases: $x>n/2$ or $x\le n/2$ from (11).

If $x>n/2$, then $z_x=z_{x+1}=\cdots=z_{n-2}=0$. Therefore,

$$Z_n=2^{x-1}\frac{2^{n-x-1}}{2^{n-2}-\sum_{k=1}^{n-2}z_k}=\frac{2^{n-2}}{2^{n-2}-\sum_{k=1}^{n-2}z_k}.$$

Using the inequalities $0\le\sum_{k=1}^{n-2}z_k<\frac{n}{2}$, we have

$$1\le Z_n<\frac{2^{n-2}}{2^{n-2}-\frac{n}{2}}.$$

Note that the following inequalities hold in general:

$$0\le\sum_{k=x}^{n-2}z_k\le\sum_{k=1}^{n-2}z_k<\frac{n}{2}$$

from the definition of z_k. We now consider the case $x\le n/2$. Using the above inequalities, we have

$$2^{x-1}\frac{2^{n-x-1}-\frac{n}{2}}{2^{n-2}}<2^{x-1}\frac{2^{n-x-1}-\sum_{k=x}^{n-2}z_k}{2^{n-2}-\sum_{k=1}^{n-2}z_k}<2^{x-1}\frac{2^{n-x-1}}{2^{n-2}-\frac{n}{2}}.$$

Using $x\le\frac{n}{2}$, this results in the following.

$$\frac{2^{n-2}-n2^{\frac{n}{2}-2}}{2^{n-2}}<Z_n<\frac{2^{n-2}}{2^{n-2}-\frac{n}{2}}.$$

Finally, we claim that

$$\left(1-\frac{n/2}{2^{\frac{n}{2}-2}}\right)^n<M_n<\left(\frac{2^{n-2}}{2^{n-2}-\frac{n}{2}}\right)^n.$$

Then, both upper and lower bounds above approach to 1 as n increases. To prove this claim, we first manipulate the expression of M in (6) and obtain a new expression for M as follows:

$$M_n=\frac{M}{2^{(x+y)-2}}=\frac{2^{n-(x+y)}\prod_{k=1}^{n-2}(P(2^{n-k-2},z_k)P(2^{n-k-2},p_k))}{P(2^{n-2},A)P(2^{n-2},B)}, \tag{12}$$

where we use the permutation notation

$$P(N,K)\triangleq\frac{N!}{(N-K)!}=N(N-1)\cdots(N-K+1),$$

and the notation for convenience

$$A\triangleq\sum_{k=1}^{n-2}z_k<\frac{n}{2} \quad\text{and}\quad B\triangleq\sum_{k=1}^{n-2}p_k<\frac{n}{2}, \tag{13}$$

where the upper bound $n/2$ comes from the definition of z_k and p_k. The number $P(N,K)$ can be bounded by

$$(N-K)^K\le P(N,K)\le N^K. \tag{14}$$

We take care of the upper bound on M_n using the above inequalities. Therefore,

$$M_n\le\frac{2^{n-(x+y)}\prod_{k=1}^{n-2}\left(\left(2^{n-k-2}\right)^{z_k}\left(2^{n-k-2}\right)^{p_k}\right)}{(2^{n-2}-A)^A(2^{n-2}-B)^B}=\frac{2^{(n-2)A}}{(2^{n-2}-A)^A}\times\frac{2^{(n-2)B}}{(2^{n-2}-B)^B},$$

where the numerator in the final form was obtained by using the relation (10) and the notation (13). Since both A and B are upper bounded by $n/2$, the above expression is upper bounded by

$$\left(\frac{2^{n-2}}{2^{n-2}-\frac{n}{2}}\right)^n.$$

For the lower bound on M_n in (12), we first observe that the values of z_k and p_k for $k > n/2$. For example, $z_{n/2}$ is the number of runs $0_{n/2}$ inside the underlined part of the n-tuple in (7). Observe that it is at most 1. If it is 1 then all the values of z_k for $k > n/2$ must be 0. In fact, z_k could be 1 only for at most one value of $k \geq n/2$. Similarly, for the values of p_k for $k \geq n/2$. The conclusion is that the value z_k or p_k for $k \geq n/2$ must be 0 or 1.

We use this to split the product in the numerator of M_n in (12) into two terms: one product for k from 1 to $n/2-1$ which can be lower bounded by (14) and the other product for k from $n/2$ to $n-2$ which can be simply changed using the relation $P(N,1)=N$. The denominator can be replaced by its upper bound using (14). This gives a lower bound on M_n as follows:

$$\begin{aligned}
M_n =& \frac{2^{n-(x+y)}}{P(2^{n-2},A)P(2^{n-2},B)} \times \prod_{n/2\leq k\leq n-2} (2^{n-k-2})^{(z_k+p_k)} \\
&\times \prod_{1\leq k<n/2} P(2^{n-k-2},z_k)P(2^{n-k-2},p_k) \\
\geq& \frac{2^{n-(x+y)}}{(2^{n-2})^A(2^{n-2})^B} \times \prod_{n/2\leq k\leq n-2} (2^{n-k-2})^{(z_k+p_k)} \\
&\times \prod_{1\leq k<n/2} (2^{n-k-2}-z_k)^{z_k}(2^{n-k-2}-p_k)^{p_k} \\
=& \prod_{1\leq k<n/2} \left(1-\frac{z_k}{2^{n-k-2}}\right)^{z_k} \left(1-\frac{p_k}{2^{n-k-2}}\right)^{p_k},
\end{aligned}$$

where the last equality is obtained by the relation (10) and some straightforward simplification on the various powers of 2. This can be lower bounded by taking the minimum of $(1-z_k/2^{n-k-2})$ over all $1 \leq k < n/2$. We denote this minimum by $(1-z_\gamma/2^{n-\gamma-2})$ for some $k=\gamma$. Similarly, we write the minimum of the second factors as $(1-p_\delta/2^{n-\delta-2})$ for some $k=\delta$. Then, we have

$$\begin{aligned}
M_n &\geq \left(1-\frac{z_\gamma}{2^{n-\gamma-2}}\right)^{\sum_{1\leq k<n/2} z_k} \left(1-\frac{p_\delta}{2^{n-\delta-2}}\right)^{\sum_{1\leq k<n/2} p_k} \\
&> \left(1-\frac{n/2}{2^{\frac{n}{2}-2}}\right)^n,
\end{aligned}$$

since the exponents are less than A and B, respectively, (where both A and B are less than $n/2$) and also z_γ and p_δ are also less than $n/2$, and $1 \leq \gamma, \delta < n/2$. ■

Figure 2 shows the maximum and minimum values of $\mathcal{E}_n \cup \mathcal{F}_n \cup \mathcal{G}_n \cup \mathcal{H}_n$ for $n = 3, 4, \ldots, 25$. The maximum and minimum values are farthest away from each other when $n = 5$, and then gradually converge to 1 as n increases. It would be enough to say that the average numbers of all the not-all-zero n-tuples are essentially 1 when $n \geq 20$. For

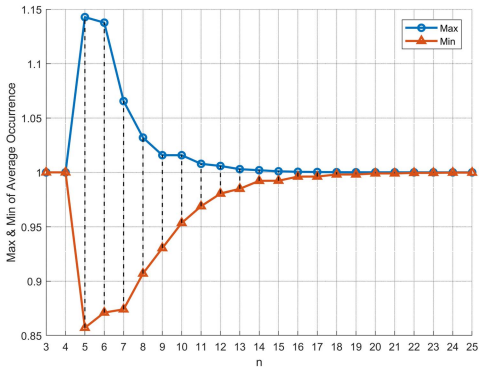
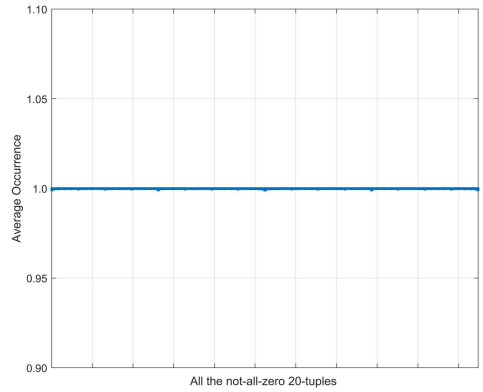
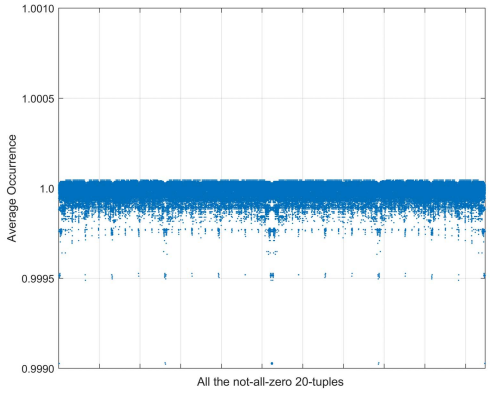


Fig. 2. The maximum and minimum values of the average numbers of all the n-tuples.



(a)



(b)

Fig. 3. The average number of occurrences of each 20-tuples.

$n = 20$, the average numbers are between 0.9990 and 1.0001. Figure 3(a) and (b) obtained from Lemma 1 show these values.

III. Concluding Remarks

It is well-known that a span sequence is a run sequence but not conversely. The total number s_n of span sequences of period 2^n-1 is well-known to be $s_n = 2^{2^{n-1}-n}$. In this paper, we investigated various relations of run properties and span properties of binary sequences of period 2^n-1.

We first count the number r_n of all the cyclically distinct binary run sequences of period 2^n-1 (Thm. 1). We check

TABLE V
DISTRIBUTION OF OCCURRENCES IN A RUN SEQUENCE OF PERIOD 31

	# occurrences							# sequences
	0	1	2	3	4	5	⋯	(total = 88200)
Distribution 0	0	31	0	0	0	0	⋯	2048 (span)
Distribution 1	2	27	2	0	0	0	⋯	6144
Distribution 2	4	23	4	0	0	0	⋯	26112
Distribution 3	6	19	6	0	0	0	⋯	33920
Distribution 4	7	18	5	1	0	0	⋯	6912
Distribution 5	8	17	4	2	0	0	⋯	1152
Distribution 6	8	15	8	0	0	0	⋯	7680
Distribution 7	10	13	6	2	0	0	⋯	2304
Distribution 8	10	11	10	0	0	0	⋯	1536
Distribution 9	12	7	12	0	0	0	⋯	320
Distribution 10	14	7	8	0	2	0	⋯	72
Average distribution	5.47	20.22	5.15	0.157	0.00163	0	⋯	sum = 31.00
Normalized average	0.176	0.652	0.166	0.005	0.00005	0	⋯	sum = 1.00

the rate of increase of the sequence r_n as n goes to infinity (Cor. 1) and find that only a tiny small portion of these are span sequences (Cor. 2). We define the average number of occurrences of an n-tuple throughout all the r_n run sequences (Def. II). We see that this average number of any binary n-tuple is lower bounded by s_n/r_n (Prop. 1) and that the sum (over all n-tuples) of all these average numbers is 2^n-1 (Prop. 2). We are able to identify those binary n-tuples with the average number 1 (Prop. II) and especially those with the average number 1 uniformly (Prop. 3). As a main result, we prove that the average number of any not-all-zero n-tuple approaches 1 as n increase indefinitely (Thm. 2). We may call this a statistical span property of run sequences.

One of future research topic would be to consider the set of run sequences over some non-binary alphabet. Similar definition would be possible from the q-ary m-sequences of period q^n-1. Here, q different types of runs exist and the order of these runs matters, unlike the binary cases in which only two different types of run alternate. Calculating the number of cyclically distinct q-ary run sequences would be an interesting and much more difficult problem.

Another interesting future research would be to consider the set of binary run sequences under the equivalence of both *rotation* and *reversal*. In this paper, we only consider the equivalence class of rotation or cyclic shift. The problem would become quite different and much more difficult if we consider both operations for the equivalence class.

We finally propose a question from the different perspective. Each run sequence has 2^n-1 binary n-tuples which are not necessarily all distinct. If it is a span sequence then all these n-tuples are distinct. Otherwise, some appear multiple times and some others do not appear at all. Therefore, we may ask the following: what is the (average) distribution of n-tuples in a random run sequence of period 2^n-1?

For $i=0,1,2,\ldots$, let a_i be the number of n-tuples that occur exactly i many times in a run sequence of period 2^n-1. Then, we have at least the following two conditions:

$$a_0+a_1+a_2+\cdots=2^n-1,$$

since the total number of n-tuples is 2^n-1, and

$$1a_1+2a_2+3a_3+\cdots=2^n-1,$$

since the period of the sequence is 2^n-1.

For examples, any span sequence of period 2^5-1 has a distribution $(a_0,a_1,a_2,a_3,\ldots)=(0,31,0,0,\ldots)$, which is denoted as Distribution 0 in Table V. There are 2048 such run sequences (all of which are in fact span sequences). All the distributions of 5-tuples in the 88200 run sequences of period 2^5-1 are classified as eleven different distributions in Table V. Distribution 1 indicates twenty-seven 5-tuples appear 1 time, two 5-tuples appear 2 times, and remaining two 5-tuples do not appear at all. There are 6144 run sequences having this type of distribution. Here, the average of the distributions indicates that, in a random run sequence of period 31, 5.47 5-tuples do not appear at all, 20.22 of them appear exactly 1 time, 5.15 of them appear exactly 2 times, etc. If this distribution is normalized by the total number $2^n-1=31$, it becomes a probability distribution. It is wide open what this distribution would become as n increases indefinitely.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for careful review and many helpful suggestions, especially those leading to the open questions in the last section.

REFERENCES

[1] N. G. de Bruijn, "A combinatorial problem," *Koninklijke Nederlandse Akademie Wetenschappen*, vol. 49, no. 7, pp. 758–764, Jun. 1946.
[2] U. Cheng and S. W. Golomb, "On the characteristics of PN sequences (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 4, p. 600, Jul. 1983.
[3] J. Dutka, "The early history of the factorial function," *Archive Hist. Exact Sci.*, vol. 43, no. 3, pp. 225–249, 1991.
[4] P. Gaal and S. Golomb, "Exhaustive determination of (1023, 511, 255)-cyclic difference sets," *Math. Comput.*, vol. 70, no. 233, pp. 357–366, Jan. 2001.

[5] S. W. Golomb, *Shift Register Sequences*, 3rd ed. Hackensack, NJ, USA: World Scientific, 2017.
[6] S. W. Golomb, "On the classification of balanced binary sequences of period $2^n - 1$," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 730–732, Nov. 1980.
[7] S. W. Golomb and G. Gong, *Signal Design for Good Correlation: for Wireless Communication, Cryptography and Radar*. New York, NY, USA: Cambridge Univ. Press, 2005.
[8] M. Hall, Jr., *Combinatorial Theory*, 2nd ed. New York, NY, USA: Wiley, 1986.
[9] T. Helleseth, "Golomb's randomness postulates," in *Encyclopedia of Cryptography and Security*. Boston, MA, USA: Springer, 2011, pp. 516–517, doi: 10.1007/0-387-23483-7_178.
[10] J.-H. Kim and H.-Y. Song, "Existence of cyclic Hadamard difference sets and its relation to binary sequences with ideal autocorrelation," *J. Commun. Netw.*, vol. 1, no. 1, pp. 14–18, Mar. 1999.
[11] G. Kim, M. H. Lee, and H.-Y. Song, "Some notes on the binary sequences of length $2^n - 1$ with the run property," in *Proc. 9th Int. Workshop Signal Design Appl. Commun. (IWSDA)*, Dongguan, China, Oct. 2019.
[12] N. Loehr, *Bijective Combinatorics*. Boca Raton, FL, USA: CRC Press, 2011.
[13] H. Robbins, "A remark on Stirling's formula," *Amer. Math. Monthly*, vol. 62, no. 1, pp. 26–29, 1955.
[14] J. Sawada, A. Williams, and D. Wong, "A surprisingly simple de Bruijn sequence construction," *Discrete Math.*, vol. 339, no. 1, pp. 127–131, Jan. 2016.
[15] J. Sawada, A. Williams, and D. Wong, "A simple shift rule for k-ary de Bruijn sequences," *Discrete Math.*, vol. 340, no. 3, pp. 524–531, Mar. 2017.
[16] H.-Y. Song, "Feedback shift register sequences," in *Wiley Encyclopedia of Telecommunications*. Hoboken, NJ, USA: Wiley, 2003.
[17] H. Y. Song and S. W. Golomb, "On the existence of cyclic Hadamard difference sets," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1266–1268, Jul. 1994.
[18] S. W. Golomb and H.-Y. Song, "A conjecture on the existence of cyclic Hadamard difference sets," *J. Stat. Planning Inference*, vol. 62, no. 1, pp. 39–41, Jul. 1997.

Gangsan Kim (Graduate Student Member, IEEE) received the B.S. degree in electronic engineering from Yonsei University in 2016. He is currently pursuing the Ph.D. degree with Yonsei University, Seoul, South Korea, under the supervision of Prof. Hong-Yeop Song. His area of research interests include coding theory, PN sequences, related discrete mathematics, and their applications to digital communication systems.

Hong-Yeop Song (Senior Member, IEEE) received the B.S. degree in electronic engineering from Yonsei University, Seoul, South Korea, in 1984, and the M.S.E.E. and Ph.D. degrees from the University of Southern California, Los Angeles, CA, USA, in 1986 and 1991, respectively. He spent two years as a Research Associate at USC and then two years as a Senior Engineer in standard team of Qualcomm Inc., San Diego, CA, USA. Since September 1995, he has been with the Department of Electrical and Electronic Engineering, Yonsei University. His research interests include digital communications and channel coding, design and analysis of various pseudo-random sequences for communications, and cryptography. He is a member of the National Academy of Engineering of Korea (NAEK), Mathematical Association of America (MAA), Korean Mathematical Society (KMS), KICS, IEIE, and KIISC. He received the 2017 Special Contribution Award from Korean Mathematical Society and the 2021 S.J.Choi Award from the Korean Government, both for his contribution to the global wide-spread of the fact that S. J. Choi (1646–1715) from South Korea had discovered a pair of orthogonal Latin squares of order nine much earlier than Euler. He was the Chair of IEEE IT Society Seoul Chapter from 2009 to 2016. He served as the General Co-Chair for IEEE ITW 2015, Jeju, South Korea.