d-Form Sequences: Families of Sequences with Low Correlation Values and Large Linear Spans

Andrew Klapper*
Department of Computer Science
915 Patterson Office Tower
University of Kentucky
Lexington, KY 40506-0027
E-mail: klapper@ms.uky.edu

September 11, 1995

Abstract

Large families of binary sequences with low correlation values and large linear span are critical for spread spectrum communication systems. In this paper we describe a method for constructing such families from families of homogeneous functions over finite fields, satisfying certain properties. We then use this general method to construct specific families of sequences with optimal correlations and exponentially better linear span than No sequences.

Index Terms–**Cross-correlation, autocorrelation, linear span, binary sequences, families of sequences, No sequences, spread spectrum, CDMA, finite fields, quadratic forms.**

I Introduction

The volume of communication traffic over a variety of media has been steadily increasing over the past few decades, and will continue to do so. This increase has led to a need for methods that allow many users to share communication channels. Among proposed methods for such sharing, Code Division Multiple Access holds great promise, in part due to its ability to resist interference from

*Project sponsored by the Natural Science and Engineering Research Council under Grant Number OGP0121648 and the National Security Agency under Grant Number MDA904-91-H-0012. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation hereon. Portions of this work have been presented at the 31st Annual Allerton Conference on Communications, Control, and Computing, Urbana-Champaign, IL, Sept., 1993.

hostile agents [17]. In order for this promise to be met, however, it is essential to find large families of easily generated binary sequences with high linear spans and low correlation function values. The smaller the pairwise cross-correlations and the larger the family, the higher the capacity of the system. Also, the higher the linear span, the harder it is for an adversary to jam or intercept messages.

Unfortunately, there are a limited number of known instances of such sequences. The correlation properties of sequences generated by various modified shift registers have been studied, including GMW sequences, [3], geometric sequences [8], cascaded GMW sequences [1, 9]. The results are often closely related to results from coding theory, as in the case of -1 decimations of m-sequences [18]. In a few cases families of sequences with good correlation properties have been found, such as Kasami sequences [4, 5], bent function sequences [10, 11, 15, 16], and No sequences [14]. Table (1) summarizes the properties of some of these families.

In this paper we present a general method for constructing families of sequences with low cross-correlation values from families of homogeneous functions (d-forms if the degree is d) with certain properties. The families of sequences that arise include No sequences [14]. We then exhibit a particular family of quadratic forms with the required property, and show that the linear spans of the resulting sequences, trace norm (or TN) sequences, are larger (in some cases, asymptotically exponentially larger) than the linear spans of No sequences. Moreover, we describe a method for implementing TN sequences that shows that they are no more difficult to generate than No sequences.

In his thesis No introduced a class of sequences called generalized No sequences [13]. These sequences are formed by iteratively applying functions of the form "trace of a power of x" to a No sequence with values in a finite field. He derived the balance and cross-correlation and autocorrelation values and their distributions for generalized No sequences. He described as open problems the determination of the linear span and the implementation of generalized No sequences.

A sequence is a TN sequence if and only if it is both a d-form sequence with $d = 2$ and a generalized No sequences. In this paper we resolve No's questions about linear span and implementation for this case of generalized No sequences. We also count the number of distinct families of TN sequences. No's results on the balance and correlation properties of generalized No sequences give the balance and correlation properties of TN sequences. We have, however, included sketches of the proofs of these results (Theorem III.2 and Proposition III.4) since they are examples of the application of the general result on cross-correlations of d-form sequences, they are not yet available in the archival literature, and they were discovered independently by the author.

In what follows, let e and m be positive integers, let $n = em$, and let Tr_m^n be the trace function from $GF(2^n)$ to $GF(2^m)$,

$$Tr_m^n(x) = \sum_{i=0}^{e-1} x^{2^{mi}}.$$

Also, let N_m^n be the norm function from $GF(2^n)$ to $GF(2^m)$,

$$N_m^n(x) = \prod_{i=0}^{e-1} x^{2^{mi}} = x^{(2^n-1)/(2^m-1)}.$$

Family	n	Size of Family	Maximum Correlation	Maximum Linear Span	Range of Imbalance
Gold	$2m+1$	2^n+1	$1+2^{\frac{n+1}{2}}$	$2n$	$[1, 2^{\frac{n+1}{2}}+1]$
Gold	$4m+2$	2^n-1	$1+2^{\frac{n+2}{2}}$	$2n$	$[1, 2^{\frac{n+2}{2}}+1]$
Kasami (Small Set)	$2m$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$\frac{3n}{2}$	$[1, 2^{\frac{n}{2}}+1]$
Kasami (Large Set)	$4m+2$	$2^{\frac{n}{2}}(2^n+1)$	$1+2^{\frac{n+2}{2}}$	$\leq \frac{5n}{2}$	$[1, 2^{\frac{n+2}{2}}+1]$
Bent	$4m$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$\binom{n/2}{n/4}2^{\frac{n}{4}}$	1
No	$2m$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$m(2^m-1)$	$[1, 2^{\frac{n}{2}}+1]$
TN	$2km$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$> 3mk(3k-1)^{m-2}$	$[1, 2^{\frac{n}{2}}+1]$

Table 1: Comparison of Properties of Families of Sequences of Period 2^n-1

For simplicity we write $q = 2^m$. By a d-form we mean a homogeneous function of degree d. That is, a function H is a d-form, if for any $x \in GF(q^e)$ and $y \in GF(q)$, we have

$$H(yx) = y^d H(x).$$

For example, any H which is a homogeneous polynomial of degree d (thinking of $GF(q^e)$ as an e-dimensional vector space over $GF(q)$) is a d-form. More generally, any sum of functions of the form

$$H(x) = Tr_m^{km}(N_{km}^n(x^r)),$$

where the sum of the coefficients in the base q expansion of $r(q^e-1)/(q^k-1)$ is congruent to d modulo $q-1$, is a d-form. Every d-form is a sum of functions of this form.

We generate d-form sequences in three steps. Our construction is based on the finite fields $GF(q)$ and $GF(q^e)$. We start with a sequence of powers of a primitive element in $GF(q^e)$. To this sequence we apply a d-form H mapping to $GF(q)$. We then raise the result to some power. Finally, we apply a trace function mapping to $GF(2)$. The precise definition of d-form sequences is as follows.

Definition I.1 *Let e and m be positive integers, and let $q = 2^m$. Let r and d be positive integers such that $\gcd(r, q-1) = \gcd(d, q-1) = 1$. Let α be a primitive element in $GF(q^e)$. Let $H(x)$ be a d-form on $GF(q^e)$ over $GF(q)$. Then the sequence $\mathbf{S}$ whose ith term is*

$$s_i = Tr_1^m((H(\alpha^i))^r) \tag{1}$$

*is called a d-*form sequence.

II Cross-Correlations of d-form Sequences

Recall that the cross-correlation with shift τ, of two sequences $\mathbf{S} = (s_1, s_2, \cdots)$ and $\mathbf{T} = (t_1, t_2, \cdots)$ of period N, is defined by

$$\Theta_{\mathbf{S},\mathbf{T}}(\tau) = \sum_{i=1}^{N}(-1)^{s_i + t_{i+\tau}}.$$

In our case $N = q^e - 1$. Our first result is a complete description of the cross-correlations of d-form sequences in terms of the zeros of d-forms.

Theorem II.1 *Let the integers m, e, and r, and primitive element $\alpha \in GF(q^e)$ be fixed, and let H_1 and H_2 be d-forms on $GF(q^e)$ over $GF(q)$ defining d-form sequences $\mathbf{S}^1$ and $\mathbf{S}^2$, as in equation (1). For any shift τ, let $z_\tau = |\{x \neq 0 \in GF(q^e) : H_1(x) + H_2(\alpha^\tau x) = 0\}|$. Then*

$$\Theta_{\mathbf{S}^1,\mathbf{S}^2}(\tau) = \frac{q z_\tau - (q^e - 1)}{q - 1}.$$

Proof: The proof of this theorem is an extension of the proof of the corresponding theorem on No sequences [14].

Let $L = (q^e - 1)/(q - 1)$. For any i, $0 \leq i < N$, we can write

$$i = i_1 L + i_2, \qquad 0 \leq i_1 < q - 1, \qquad 0 \leq i_2 < L.$$

Since for any $x \in GF(q^e)$, we have $x^L = N_m^{em}(x) \in GF(q)$, it follows that

$$H_j(x^i) = x^{d i_1 L} H_j(x^{i_2}).$$

Thus the terms of the sequence $\mathbf{S}^j$ can be written

$$s_i^j = Tr_1^m(\alpha^{d r i_1 L}(H_j(\alpha^{i_2}))^r).$$

Letting

$$f(i) = (H_1(\alpha^i))^r + (H_2(\alpha^{i+\tau}))^r,$$

we have

$$s_i^1 + s_{i+\tau}^2 = Tr_1^m(\alpha^{d r i_1 L} f(i_2)). \tag{2}$$

Whenever $f(i_2) \neq 0$, the sequence we get from equation (2) by letting i_1 vary is an m-sequence of period $q - 1$. It follows from the balance properties of m-sequences [2] that for a fixed i_2 the contribution of these terms to the cross-correlation is -1. On the other hand, when $f(i_2) = 0$, every term is zero, so the contribution of these terms for a fixed i_2 is $q - 1$. If we let z be the number of values of i_2, $0 \leq i_2 < L$, for which $f(i_2) = 0$, then

$$\Theta_{\mathbf{S}^1,\mathbf{S}^2}(\tau) = -(L - z) + (q - 1)z = qz - \frac{q^e - 1}{q - 1}.$$

To complete the proof, we observe that $f(i+L) = \alpha^{drL} f(i)$, so that

$$z = \frac{|\{i : f(i) = 0, 0 \le i < N\}|}{q-1}.$$

This is precisely $z_\tau/(q-1)$, since α^i ranges through all nonzero x as i ranges from 0 to $N-1$, and $f(i) = 0$ if and only if $H_1(\alpha^i) = H_2(\alpha^{i+\tau})$. □

In case $d = 2$ and H is a quadratic form (that is, H is a homogeneous polynomial of degree two), the number of zeros of H is well understood. Recall that the *rank* of a quadratic form H is the smallest integer t such that there is a set of coordinates in which H can be represented using only t variables. The number of solutions x to the equation $H(x) = 0$ (or, more generally, $H(x) = a$) is determined by the rank and, in the case of even rank, whether H is of one of two types. A nice treatment of this analysis can be found in Lidl and Neiderreiter's book [12].

Corollary II.2 *Let H_1 and H_2 be quadratic forms defining d-form sequences $\mathbf{S}^1$ and $\mathbf{S}^2$, and let τ be any integer. If the rank of $H_1(x) + H_2(\alpha^\tau x)$ is t, then*

$$\Theta_{\mathbf{S}^1,\mathbf{S}^2}(\tau) = -1$$

when t is odd, and

$$\Theta_{\mathbf{S}^1,\mathbf{S}^2}(\tau) = \pm q^{e-t/2} - 1$$

when t is even. In particular, if for every τ the rank is n or is odd, then the cross-correlations of $\mathbf{S}^1$ and $\mathbf{S}^2$ are three valued with values in $\{-q^{e/2} - 1, -1, q^{e/2} - 1\}$.

Proof: The corollary follows from the fact that a quadratic form with odd rank takes the value zero q^{e-1} times, while a quadratic form with even rank t takes the value zero $q^{e-1} \pm (q-1)q^{e-t/2-1}$ times [12]. □

Corollary II.3 *If $\mathcal{F}$ is a family of quadratic forms on $GF(q^e)$ over $GF(q)$ such that for any H_1 and H_2 in $\mathcal{F}$, and integer τ, the rank of the quadratic form $H_1(x) + H_2(\alpha^\tau x)$ is e or odd, then $\mathcal{F}$ defines a family of sequences with three valued cross-correlations with values in $\{-q^{e/2} - 1, -1, q^{e/2} - 1\}$.*

III Trace Norm (TN) Sequences

In this section we describe a class of families of quadratic form sequences that achieve the cross-correlations of Corollary II.3 and have large linear span. These are exactly the sequences that are simultaneously d-form sequences and generalized No sequences [13]. The results in this section were proved independently by No in his thesis. However, they do not appear in the archival literature and they are examples of the application of Theorem II.1, so we include the full details.

Definition III.1 *Let m and k be positive integers, let $q = 2^m$, and let $e = 2k$ and $n = 2mk$. Let r be a positive integer such that $\gcd(r, q-1) = 1$. Let α be a primitive element in $GF(q^{2k})$ and let γ be an element of $GF(q^k)$. Then the sequence $\mathbf{S}^\gamma$ whose ith term is*

$$s_i = Tr_1^m((Tr_m^{mk}(Tr_{mk}^{2mk}(\alpha^{2i}) + \gamma N_{mk}^{2mk}(\alpha^i)))^r)$$

is a Trace Norm *(or* TN*) sequence.*

Note that, if we let $T = q^k + 1$, then $N_{mk}^{2mk}(\alpha^i) = \alpha^{Ti}$. This sequence is a 2-form sequence based on

$$H(x) = Tr_m^{mk}(Tr_{mk}^{2mk}(x^2) + \gamma N_{mk}^{2mk}(x)).$$

We are interested in families of TN sequences with all parameters other than γ fixed.

Theorem III.2 *Let $\mathbf{S}^\gamma$ and $\mathbf{S}^\delta$ be two TN sequences, based on the same integers m, k, and r. Then the cross-correlations of $\mathbf{S}^\gamma$ and $\mathbf{S}^\delta$ are three valued, with values in $\{-q^k - 1, -1, q^k - 1\}$ unless $\mathbf{S}^\gamma = \mathbf{S}^\delta$ and $\tau = 0$.*

Proof: By Theorem II.1, it suffices to determine the number of zeros z of the quadratic form

$$\begin{aligned} H(x) &\stackrel{\text{def}}{=} Tr_m^{mk}(Tr_{mk}^{2mk}(x^2) + \gamma x^T) + Tr_m^{mk}(Tr_{mk}^{2mk}(\alpha^{2\tau}x^2) + \delta\alpha^{T\tau}x^T) \\ &= Tr_m^{mk}(Tr_{mk}^{2mk}((1+\alpha^{2\tau})x^2) + (\gamma + \delta\alpha^{T\tau})x^T). \end{aligned}$$

For simplicity, we write $A = (1 + \alpha^\tau)$ and $B = (\gamma + \delta\alpha^{T\tau})$.

Let $G(x) = Tr_{mk}^{2mk}(A^2x^2) + Bx^T$. Then G is a quadratic form on $GF(q^{2k})$ over $GF(q^k)$. Since $G(ax) = a^2G(x)$, and every element of a finite field of characteristic two has a unique square root, G takes on every nonzero value an equal number of times. If $w = |\{x : G(x) = 0\}|$, then G takes on any nonzero value precisely $(q^{2k} - w)/(q^k - 1)$ times. Moreover, Tr_m^{mk} is a balanced function. It follows that

$$z = \frac{wq^{k-1}(q-1) + q^{2k}(q^{k-1} - 1)}{q^k - 1} - 1$$

.

The value of w can be computed by cases, depending on τ, γ, and δ. For example, suppose $\gamma \neq \delta\alpha^{T\tau}$ and $\alpha^\tau \neq 1$, i.e., $B \neq 0$ and $A \neq 0$. Then $G(x) = 0$ if and only if there is a $\mu \in GF(q^k)$ such that

$$Tr_{mk}^{2mk}(Ax) = \mu \qquad \text{and} \qquad Bx^T = \mu^2.$$

Thus we must count the common zeros of a linear function and a quadratic form in two variables over $GF(q^k)$. These have been completely analyzed [7]. Observe that the quadratic form Bx^T is zero only for $x = 0$, hence in the terminology of [7], is a Type III quadratic form, and has rank two. This allows us to apply Proposition 3.4 of [7] to show that $w = 1$ or $2q^k - 1$. It follows that $\Theta_{\mathbf{S}^\gamma,\mathbf{S}^\delta}(\tau) = -q^k - 1$ or $q^k - 1$. □

The imbalance $I(\mathbf{S})$ of a binary sequence $\mathbf{S}$ is the number of zeros in $\mathbf{S}$ minus the number of ones. The proof of the Theorem III.2 can be used to find the imbalance of a TN sequence.

Proposition III.3 *The imbalance of a TN sequence is -1, $-q^k-1$, or q^k-1.*

Proof: If $\gamma = 0$, then we have a GMW sequence, which is well known to have imbalance -1. If $\gamma \neq 0$, then we can count as in Case 3 of the proof of Theorem III.2 with $A = 1$ and $B = \gamma$. The proposition follows. □

There are six possible distributions of values of the cross-correlation of two TN sequences.

Proposition III.4 *Let $\mathbf{S}^\gamma$ and $\mathbf{S}^\delta$ be two TN sequences based on the same field parameters m and k and exponent r. The distribution of values of the cross-correlation of $\mathbf{S}^\gamma$ and $\mathbf{S}^\delta$ is given by one of the following cases. The first three cases correspond to autocorrelations.*

1. *$q^{2k}-1$ occurs once and -1 occurs $q^{2k}-2$ times.*
2. *$q^{2k}-1$ occurs once, -1 occurs q^k times, q^k-1 occurs $q^{2k}/2$ times, and $-q^k-1$ occurs $q^{2k}/2-q^k-2$ times.*
3. *$q^{2k}-1$ occurs once, -1 occurs q^k times, q^k-1 occurs $q^{2k}/2-2$ times, and $-q^k-1$ occurs $q^{2k}/2-q^k$ times.*
4. *q^k-1 occurs $(q^{2k}+q^k)/2-1$ times and $-q^k-1$ occurs $(q^{2k}-q^k)/2$ times.*
5. *q^k-1 occurs $(q^{2k}+q^k)/2$ times and $-q^k-1$ occurs $(q^{2k}-q^k)/2-1$ times.*
6. *-1 occurs q^k+1 times, q^k-1 occurs $(q^{2k}+q^k)/2$ times, and $-q^k-1$ occurs $(q^{2k}-3q^k)/2-2$ times.*
7. *-1 occurs q^k+1 times, q^k-1 occurs $(q^{2k}+q^k)/2-2$ times, and $-q^k-1$ occurs $(q^{2k}-3q^k)/2$ times.*
8. *-1 occurs q^k+1 times, q^k-1 occurs $(q^{2k}-q^k)/2-1$ times, and $-q^k-1$ occurs $(q^{2k}-q^k)/2-1$ times.*

Proof: If $\gamma = \delta = 0$, then we have the autocorrelation of a GMW sequence. Otherwise, let C (respectively, D; respectively, E) be the number of occurences of $\Theta_{\mathbf{S}^\gamma,\mathbf{S}^\delta}(\tau) = -1$ (respectively, $\Theta_{\mathbf{S}^\gamma,\mathbf{S}^\delta}(\tau) = -q^k-1$; respectively, $\Theta_{\mathbf{S}^\gamma,\mathbf{S}^\delta}(\tau) = q^k-1$). If $\gamma \neq \delta = 0$, then in the proof of Theorem III.2 we have $B \neq 0$, and the only values that appear are $-q^k-1$ and q^k-1. That is, $C = 0$. Otherwise the cross-correlation is -1 for q^k+1 shifts, so $C = q^k+1$. In general, we have

$$\sum_\tau \Theta_{\mathbf{S}^\gamma,\mathbf{S}^\delta}(\tau) = I(\mathbf{S}^\gamma)I(\mathbf{S}^\delta),$$

where $I(\mathbf{S}^\gamma)$ and $I(\mathbf{S}^\delta)$ are the imbalances of $\mathbf{S}^\gamma$ and $\mathbf{S}^\delta$, respectively.

Thus we have an equation

$$-C + (-q^k-1)D + (q^k-1)E = I(\mathbf{S}^\gamma)I(\mathbf{S}^\delta).$$

Moreover,

$$C + D + E = q^{2k}-1.$$

Considering the various values of τ, γ, δ, and C, we can solve these equations to get the various possibilities in the statement of the proposition. □

IV Linear Span of TN Sequences

In this section we show that the linear span of a TN sequence is at least that of a GMW sequence with the same period, and that for q large enough and $k = 2$, it exceeds the linear span of a No sequence with the same period. The development is similar to that in the case of No sequences, with some additional complication due to the additional trace function.

Key [6] showed that if we express the ith term of a sequence $\mathbf{S}$ as a polynomial in α^i, then the linear span $\lambda_{\mathbf{S}}$ is the number of monomials in the polynomial. That is, we must count the monomials in the polynomial

$$\begin{aligned} s(x) &= Tr_1^m((Tr_m^{mk}(Tr_{mk}^{2mk}(x^2) + \gamma x^T))^r) \\ &= Tr_1^m((Tr_m^{mk}(x^2 + x^{2q^k} + \gamma x^T))^r) \\ &= Tr_1^m((Tr_m^{mk}(x^2(1 + \gamma y + y^2)))^r), \end{aligned}$$

where $y = x^{q^k - 1}$. Expanding the trace functions, we see that

$$s(x) = \sum_{j=0}^{m-1} (\sum_{i=0}^{k-1} x^{2^{j+1}q^i}(1 + \gamma y + y^2)^{2^j q^i})^r.$$

Lemma IV.1 *Distinct terms from the outer sum have distinct degree monomials.*

Proof: First reduce all exponents mod $q^k - 1$ (which divides $q^{2k} - 1$, so this reduction is compatible with $x^{q^{2k}} = x$). Then y becomes 1, so the degrees of the monomials resulting from the jth term are the degrees of the monomials in

$$(\sum_{i=0}^{k-1} x^{2^{j+1}q^i})^r = (\sum_{i=0}^{k-1} x^{q^i})^{r2^{j+1}}, \quad j = 0, \cdots, m-1.$$

Thus it suffices to compare the degrees of the monomials in the terms

$$(\sum_{i=0}^{k-1} x^{q^i})^r \qquad \text{and} \qquad (\sum_{i=0}^{k-1} x^{q^i})^{r2^j}.$$

These polynomials have monomials with degrees

$$A = \sum_{i=0}^{k-1} a_i q^i, \qquad \text{with} \qquad \sum_{i=0}^{k-1} a_i = r < q \tag{3}$$

and

$$B = \sum_{i=0}^{k-1} b_i 2^j q^i, \qquad \text{with} \qquad \sum_{i=0}^{k-1} b_i = r < q, \tag{4}$$

respectively. In particular, $0 \leq a_i, b_i < q$, so the representations in equations (3) and (4) are the (unique) base q representations.

Let

$$b_i = c_i + 2^{m-j} d_i, \qquad 0 \le c_i < 2^{m-j}, \qquad 0 \le d_i < 2^j.$$

Then

$$\begin{aligned} B &= \sum_{i=0}^{k-1} c_i 2^j q^i + d_i q^{i+1} \\ &\equiv \sum_{i=0}^{k-1} (d_{i-1} + c_i 2^j) q^i \pmod{q^k - 1} \end{aligned}$$

is the (unique) base q representation modulo $q^k - 1$ (where arithmetic on the subscripts is modulo k).

Suppose $A = B$. Then for every i, $a_i = d_{i-1} + c_i 2^j$. Consequently,

$$r = \sum_{i=0}^{k-1} (d_{i-1} + c_i 2^j) = \sum_{i=0}^{k-1} (c_i + d_i 2^{m-j}).$$

Letting $C = \sum_i c_i$ and $D = \sum_i d_i$, this reduces to

$$r = D + 2^j C = C + 2^{m-j} D,$$

so

$$(2^j - 1)C = (2^{m-j} - 1)D.$$

Let t be the greatest common divisor of m and j. Then the greatest common divisor of $2^j - 1$ and $2^{m-j} - 1$ is $2^t - 1$, so there is an integer E such that

$$C = \frac{2^{m-j} - 1}{2^t - 1} E \qquad \text{and} \qquad D = \frac{2^j - 1}{2^t - 1} E.$$

It follows that

$$\begin{aligned} r &= E \left(\frac{2^j - 1}{2^t - 1} + 2^j \frac{2^{m-j} - 1}{2^t - 1} \right) \\ &= E \left(\frac{2^m - 1}{2^t - 1} \right). \end{aligned}$$

However, the greatest common divisor of r and $2^m - 1$ is one, so we must have $t = m$, i.e., m divides j, which is impossible. $\square$

This lemma implies that

$$\lambda_{\mathbf{S}} = m \cdot \left| \left\{ \text{ monomials in } \left(\sum_{i=0}^{k-1} x^{2q^i} (1 + \gamma y + y^2)^{q^i} \right)^r \right\} \right|. \tag{5}$$

Let

$$r = \sum_{j=0}^{m-1} r_j 2^{\cdot}$$

Then for any z,

$$\begin{aligned}\left(\sum_{i=0}^{k-1} z^{q^i}\right)^r &= \prod_{r_j \neq 0} \left(\sum_{i=0}^{k-1} z^{2^j q^i}\right) \\ &= \sum_{(i_j)} \prod_{r_j \neq 0} z^{2^j q^{i_j}} \\ &= \sum_{(i_j)} z^{\sum_{r_j \neq 0} 2^j q^{i_j}},\end{aligned}$$

where the outer sums are over all vectors of i_j, indexed by those j such that $r_j \neq 0$, and with $0 \leq i_j < k$. Consider a term with exponent $\sum_{r_j \neq 0} 2^j q^{i_j}$. For each $i = 0, \cdots, k-1$, let a_i be the number whose binary representation has a 1 in the jth bit if and only if $i_j = i$. Then

$$\sum_{r_j \neq 0} 2^j q^{i_j} = \sum_{i=0}^{k-1} a_i q^i,$$

and this last is the base q expansion of this exponent. For a given bit j, at most one a_i has a 1 in bit j, so the bitwise "AND" of any two different a_is is zero. If r has a 1 in bit j, then some a_i has a 1 in bit j, so the bitwise "OR" of all the a_i equals r. That is, the bits of r are distributed among the a_i, maintaining their relative bit positions. In particular, $a_i \leq r < q$. For example, writing r and the a_is in base 2, if $k = 2$ and $r = 11110$, then we can have $a_0 = 11110$ and $a_1 = 00000$, or $a_0 = 11010$ and $a_1 = 00100$, or $a_0 = 01010$ and $a_1 = 10100$, etc.

It follows that the rth power in equation (5) can be expanded as

$$\sum_{\bar{a}} \prod_{i=1}^{k} (x^{2q^i}(1 + \gamma y + y^2)^{q^i})^{a_i} = \sum_{\bar{a}} x^{2\sum a_i q^i}(1 + \gamma y + y^2)^{\sum a_i q^i},$$

where the sum is over all $\bar{a} = (a_0, \cdots, a_{k-1})$ such that $0 \leq a_i$, the bitwise "AND" of any two different a_is is zero, and the bitwise "OR" of all the a_i equals r. Reducing exponents modulo $q^k - 1$, we get monomials of degree $2\sum_i a_i q^i$. Each such expression is twice the unique base q representation of an integer, hence these monomials are pairwise distinct. Thus we have the following proposition.

Proposition IV.2 *The linear span of the TN sequence* $\mathbf{S}$ *is given by*

$$\lambda_{\mathbf{S}} = m \cdot \sum_{\bar{a}} |\{ \text{ monomials in } (1 + \gamma y + y^2)^{\sum a_i q^i} \}|,$$

where the sum is over all $\bar{a} = (a_0, \cdots, a_{k-1})$ *such that* $0 \leq a_i$*, the bitwise "AND" of any two* a_i*s is zero, and the bitwise "OR" of all the* a_i *equals* r*.*

The cardinalities in the sums in this proposition have been evaluated by No and Kumar [14] in describing the linear span of No sequences as follows. Suppose $t < q^k/2$ is a positive integer with R runs of ones, of lengths $L_1, \cdots, L_R$ and $\gamma \neq 0$ in its base two expansion[1]. Let $\epsilon = -1$ if the quadratic $y^2+\gamma y+1$ is reducible over $GF(q^k)$ (that is, if $Tr_1^{mk}(1/\gamma) = 0$) and $\epsilon = 1$ otherwise. For a fixed primitive element $\alpha \in GF(q^{2k})$, there is a unique (principal) root $\delta = \alpha^b$ of the quadratic $y^2+\gamma y+1$ such that $0 \le b < (q^{2k}-1)/2$. When $\epsilon = -1$, we have $b = c(q^k+1)$, and when $\epsilon = 1$, we have $b = c(q^k-1)$. Thus $0 \le c \le q^k/2$. Let $g = \gcd(c, q^k+\epsilon)$. Then the number of nonzero terms in $(y^2+\gamma y+1)^t$ is

$$\prod_{j=1}^{R} \left(2^{L_j+1} - 1 - 2 \left\lfloor \frac{(2^{L_j}-1)g}{q^k+\epsilon} \right\rfloor \right).$$

When $\gamma = 0$, the number of nonzero terms is $2^{\mathrm{wt}(t)}$, where $\mathrm{wt}(t) = \sum_j L_j$ is the number of ones in the binary expansion of t.

Combining this with Proposition IV.2, we have

Theorem IV.3 *Let* $\mathbf{S}$ *be a TN sequence based on fields* $GF(q)$ *and* $GF(q^k)$*, exponent* r*, and coefficient* $\gamma \in GF(q^k)$*. When* $\gamma \neq 0$*, let* $\epsilon = -1$ *if the quadratic* $y^2+\gamma y+1$ *is reducible over* $GF(q^k)$ *and* $\epsilon = 1$ *otherwise. Let* c *be such that*

$$\delta = \alpha^{c(q^k+1)} \qquad \textit{when } \epsilon = -1$$

and

$$\delta = \alpha^{c(q^k-1)} \qquad \textit{when } \epsilon = 1$$

is a root of $y^2+\gamma y+1$*, with* $0 \le c \le q^k/2$*. Let* $g = \gcd(c, q^k+\epsilon)$*.*

For each $\bar{a} = (a_0, \cdots, a_{k-1})$ *such that* $0 \le a_i$*, the bitwise "AND" of any two different* a_i*s is zero, and the bitwise "OR" of all the* a_i *equals* r*, let* $R_{\bar{a}}$ *be the number of runs of ones in* $\sum a_i q^i$*, and let* $L_{\bar{a},1}, \cdots, L_{\bar{a},R_{\bar{a}}}$ *be the lengths of the runs. Then the linear span of* $\mathbf{S}$ *is given by*

$$\lambda_{\mathbf{S}} = m \cdot \sum_{\bar{a}} \prod_{j=1}^{R_{\bar{a}}} \left(2^{L_{\bar{a},j}+1} - 1 - 2 \left\lfloor \frac{(2^{L_{\bar{a},j}}-1)g}{q^k+\epsilon} \right\rfloor \right).$$

If $\gamma = 0$*, then the linear span of* $\mathbf{S}$ *is given by*

$$\lambda_{\mathbf{S}} = m \cdot \sum_{\bar{a}} 2^{\mathrm{wt}(r)}.$$

[1] It should be noted that the necessary assumption that $t < q^k/2$ was not explicitly stated by No and Kumar. It poses no difficulty for their analysis, however, since any exponent over $GF(q^k)$ which is relatively prime to q^k-1 can be assumed to be less than q^k-1. Such an exponent thus has at least one zero in the first mk bits of its base two expansion. Since No sequences are invariant under cyclic shift of the exponent r, we can assume the exponent has a zero as its high bit. In the more general case of TN sequences, we can also cyclically shift r so its high bit is zero, and therefore the high bit of each $\sum_i a_i q^i$ is zero.

We next want to determine how the linear span can be maximized. First observe that if we choose γ so that

$$\gcd(c, q^k + \epsilon) < \frac{q^k + \epsilon}{2^{m-1} - 1} \leq \frac{q^k + \epsilon}{2^{L_{\bar{a},j} - 1}},$$

then all terms involving the floor will disappear, and the linear span will be

$$\lambda_{\mathbf{S}} = m \cdot \sum_{\bar{a}} \prod_{j=1}^{R_{\bar{a}}} \left(2^{L_{\bar{a},j}+1} - 1\right).$$

This can be done independently of the choice of r (say by choosing c relatively prime to $q^k + \epsilon$), so the maximum value of the linear span occurs with such a choice.

Next observe that the effect on an $\bar{a}$ of increasing the number of ones in the binary expansion of r is either to increase the length of a run of ones in $\sum_i a_i q^i$ by one, or to merge a run of length L and a run of length K into a run of length $L + K + 1$. In either case, the contribution to the linear span is increased, since

$$(2^{L+1} - 1) < (2^{L+2} - 1) \qquad \text{and} \qquad (2^{L+1} - 1)(2^{K+1} - 1) < (2^{K+L+2} - 1).$$

Thus the linear span is maximized by maximizing $\mathrm{wt}(r)$. We have $r < q - 1$, so $\mathrm{wt}(r)$ is maximized at $m - 1$, that is, when r has one zero and $m - 1$ ones in its binary expansion. Since the sequence, and hence the linear span, is independent of cyclic shifts of r, the maximal linear span occurs when r is a string of ones followed by a single zero. That is $r = 2^{m-1} - 1$.

We next estimate the linear span for these r by recursively estimating the linear span for the exponent r_i which consists of i ones followed by $m - i$ zeros. If $i > 0$, we can produce r_i from r_{i-1} by replacing the first zero with a one. Each $\bar{a}'$ for r_{i-1} gives rise to k $\bar{a}$'s for r_i, depending on which a_i receives the new bit. For $k - 1$ of them, we are adding a run of length 1, hence multiplying the contribution to the linear span by 3. For the remaining $\bar{a}$, we are increasing the length of a run by 1, and this run has length at most $i - 1$. Thus the contribution of this term is

$$\eta = \frac{2^{t+1} - 1}{2^t - 1}$$

times the contribution of the original term, for some t such that $2 \leq t < i$. Therefore $2 < \eta \leq 7/3$. Also, when r has a single one in its binary expansion, the linear span is exactly $3k$. It follows that the linear span satisfies

$$3mk(3k - 1)^{m-2} < \lambda_{\mathbf{S}} < m(3k)^2(3k - \frac{2}{3})^{m-3}.$$

For large enough $n = 2mk$, the linear span is maximized by taking $k = 2$, and can be made at least

$$\frac{3n}{2}5^{n/4-2}.$$

We have shown experimentally that the base of 5 in this expression in fact gets close to 5.24 as n increases, and we can come closer to this value theoretically by improving the estimates for r_i for small i. By comparing the values achieved, we find that for $n \geq 40$ the linear span is maximized by taking $k = 2$. For smaller n, experimental data show that the linear span is maximized by taking $k = 1$, i.e., by No sequences. The linear span in this case is at most

$$\frac{n}{2}(2^{n/2} - 1) < \frac{n}{2}4^{n/4}.$$

The maximum linear span of TN sequences with $k = 2$ and period $2^n - 1$ grows with n at a rate of $O(n \cdot 5^{n/4})$. This is exponentially larger than the rate of growth of the maximum linear span of No sequences with the same period, that growth rate being $O(n \cdot 4^{n/4})$.

Theorem IV.4 *The maximum possible linear span for a TN sequence is achieved by taking $k = 2$ and $r = 2^{n/4-1} - 1$ when $n \geq 40$, and by taking $k = 1$ and $r = 2^{n/2-1} - 1$ when $n < 40$.*

In general, we see that the linear span of TN sequences can be the same as or greater than that of No sequences while performing the exponentiation in a smaller field, with a smaller exponent r. This leads to more efficient implementation of generators of the sequences, as discussed in the next section.

We would also like to know how many sequences in a family have this maximum linear span. In fact, this happens for most sequences when $r = 2^{m-1} - 1$. As has been shown above, for the linear span to be maximal we must have

$$\gcd(c, 2^{mk} + \epsilon) < \frac{2^{mk} + \epsilon}{2^{m-1} + \epsilon} \quad \text{and} \quad 0 \leq c \leq 2^{km-1}. \tag{6}$$

Thus we need to know how many such choices of c and ϵ arise from parameters $\gamma \in GF(q^k)$.

Lemma IV.5 *For any $\epsilon \in \{\pm 1\}$ and c such that $0 \leq c \leq q^k/2$, there is a (unique) $\gamma \in GF(q^k)$ which gives rise to ϵ, c.*

Proof: First let $\epsilon = -1$, so

$$\delta = \alpha^{c(q^k+1)}.$$

Then $\delta = N^{2km}_{km}(\alpha^c) \in GF(q^k)$. Therefore, $\gamma = (\delta^2 + 1)/\delta \in GF(q^k)$ and $y^2 + \gamma y + 1$ is a reducible polynomial with root δ.

Next let $\epsilon = 1$, so

$$\delta = \alpha^{c(q^k-1)}.$$

Then $\delta \in GF(q^{2k})$, so it satisfies a quadratic equation $y^2 + \gamma y + \phi$ for some $\gamma, \phi \in GF(q^k)$. By Galois theory we have $\phi = N^{2mk}_{mk}(\delta) = \delta^{q^k+1} = 1$. Furthermore, if this equation were reducible, we would have $\delta \in GF(q^k)$, so $\delta^{q^k} = \delta$. That is, $\alpha^{c(q^k-1)^2} = 1$. This implies that $q^k + 1$ divides c, which is false. $\square$

Thus it suffices to count the number of c and ϵ satisfying equation (6). We first observe that $c = 0$ corresponds to $\gamma = 0$. Thus if we let c range from 1 to $q^k + \epsilon - 1$, we will have counted each γ exactly twice ($y^2 + 1$ is the only quadratic equation with constant term 1 that has a repeated root). It follows that for each ϵ, the number of γ is

$$\frac{1}{2}|\{c : \gcd(c, 2^{mk} + \epsilon) \leq \frac{2^{mk} + \epsilon}{2^{m-1} - 1}\}| = \frac{1}{2} \sum_{\substack{t | 2^{mk}+\epsilon \\ 2^{m-1} < t}} \phi(t),$$

where $\phi(t)$ is Euler's function. The total,

$$\frac{1}{2}\left(\sum_{\substack{t|2^{mk}+1 \\ 2^{m-1}<t}} \phi(t) + \sum_{\substack{t|2^{mk}-1 \\ 2^{m-1}<t}} \phi(t) \right),$$

is greater than $(\phi(2^{mk} - 1) + \phi(2^{mk} + 1))/2$, and is very nearly $2^{mk} - 1$. That is, the great majority of sequences in this family have the maximum linear span.

V Implementation of TN Sequence Generators

Consider the TN sequence $\mathbf{S}^\gamma$ with

$$\begin{aligned} s_i^\gamma &= Tr_1^m((Tr_m^{mk}(Tr_{mk}^{2mk}(\alpha^{2i}) + \gamma\alpha^{iT}))^r) \\ &= Tr_1^m((Tr_m^{2mk}(\alpha^{2i}) + Tr_m^{mk}(\gamma\alpha^{iT}))^r) \end{aligned}$$

and let $n = 2mk$. Since α is a primitive element in $GF(q^{2k})$, α^2 is also a primitive element in $GF(q^{2k})$ and $\beta \stackrel{def}{=} \alpha^T$ is a primitive element in $GF(q^k)$. It follows that the sequence

$$Tr_m^{2mk}(\alpha^{2i})$$

is an m-sequence of elements in $GF(q)$, and can be generated by a linear feedback shift register (LFSR) of length n/m over $GF(q)$. That is, the elements of the register are elements of $GF(q)$, and the feedback function is a linear function in n/m variables. Such a register requires only n bits of storage. The arithmetic required is at most n/m multiplications by constants in $GF(q)$ (the coefficients of the minimum polynomial of α^2, some of which may be zero), and at most $n/m - 1$ additions in $GF(q)$. The arithmetic can be minimized if α is chosen to minimize the number of nonzero coefficients in its minimal polynomial.

Similarly, the sequence

$$Tr_m^{mk}(\gamma\alpha^{iT}) = Tr_m^{mk}(\gamma\beta^i)$$

is an m-sequence over $GF(q)$ which can be generated by a LFSR of length $n/(2m)$ over $GF(q)$. This requires $n/2$ bits, and at most $n/(2m)$ multiplications and $n/(2m) - 1$ additions in $GF(q)$, which can be minimized by choosing α to minimize the number of nonzero coefficients in the minimal polynomial of β. Thus the total amount of $GF(q)$ arithmetic required to implement a TN sequence is minimized by choosing α so the total number of nonzero coefficients in the minimal polynomials of α and β is minimized.

One extra addition is required to combine the outputs of the two LFSRs. The result is then raised to the rth power, and the trace to $GF(2)$ computed. However, in representing elements of $GF(q)$ as m-bit vectors, we can choose a basis so that the trace of an element is always given by projection onto a fixed component, say the first. Thus we only need to compute a single bit of the rth power.

The different choices of γ correspond to different initial loadings of the second LFSR. Thus an entire family of TN sequences can be implemented by a single hardware circuit. Changing to a new sequence is possible by simply resetting the initial loading of the second LFSR.

VI The Number of Distinct Families of TN Sequences

It is useful to know how many distinct families of TN sequences (not necessarily with maximum linear span) there are with the parameters $n = 2\ell$, and thus the period, fixed. In this section we keep n fixed and let the factorization $\ell = mk$ vary. We show that each choice of parameters m, k, r (up to multiplication by a power of 2), and α (up to raising to an exponent which is a power of two) gives rise to a distinct family of sequences, in the sense that no sequence in one family is a cyclic shift of a sequence in another family. For any fixed even integer n we write

$$S(m,\alpha,r) = \{\mathbf{S}^{\gamma} : s_i = Tr_1^m((Tr_m^{mk}(Tr_{mk}^{2mk}(\alpha^{2i}) + \gamma\alpha^{iT}))^r) \text{ and } \gamma \in GF(2^k)\},$$

where m divides $n/2$, $k = n/(2m)$, α is a primitive element of $GF(2^n)$, and r is relatively prime to $2^m - 1$.

Proposition VI.1 *Let $n = 2\ell$, $N = 2^n - 1$, m_1, and m_2 be divisors of ℓ. Let r_1 and r_2 be integers such that $1 \leq r_i < 2^{m_i} - 1$ and r_i is relatively prime to $2^{m_i} - 1$. Let α_1 and α_2 be primitive elements in $GF(2^n)$. Then $S(m_1, \alpha_1, r_1)$ and $S(m_2, \alpha_2, r_2)$ are distinct unless either*

1. *$m_1 = m_2$, and for some integers u and v, $0 \leq u < n$, and $0 \leq v < m_1$, $\alpha_2 = \alpha_1^{2^u}$, and $r_1 = 2^v \cdot r_2$, or*
2. *r_1 and r_2 are powers of 2 and for some integer u, $0 \leq u < n$ and $\alpha_2 = \alpha_1^{2^u}$.*

In each of these cases

$$S(m_1, \alpha_1, r_1) = S(m_2, \alpha_2, r_2).$$

Proof: Suppose that we have a pair of sequences in $S(m_1, \alpha_1, r_1)$ and $S(m_2, \alpha_2, r_2)$ respectively, such that one is a cyclic shift of the other. Since n (and hence ℓ) is fixed, there are integers a, b, c, and d with $abcd = \ell$ and b relatively prime to c, such that $m_1 = ab$ and $m_2 = ac$. If r_1 and r_2 are powers of two, they can be factored out of the trace functions. This is the second case of the proposition, and the sequences can be written with $m_1 = m_2$. The following lemma shows that when r_1 and r_2 are not both powers of two, the degrees m_1 and m_2 must be equal.

Lemma VI.2 *Assume r_1 and r_2 are not both powers of two. Then $m_1 = m_2$.*

Proof of lemma: For some $\gamma_1, \gamma_2 \in GF(2^\ell)$, and $\delta \in GF(2^n)$ we have

$$Tr_1^{ab}((Tr_{ab}^{\ell}(Tr_\ell^n(\alpha_1^{2i}) + \gamma_1\alpha_1^{iT}))^{r_1}) = Tr_1^{ac}((Tr_{ac}^{\ell}(Tr_\ell^n(\delta\alpha_2^{2i}) + \gamma_2\alpha_2^{iT}))^{r_2}) \tag{7}$$

for every i. For some e we have $\alpha_2 = \alpha_1^e$, so equation (7) holds if and only if for every $x \in GF(2^n)$,

$$Tr_1^{ab}((Tr_{ab}^{\ell}(Tr_\ell^n(x^2) + \gamma_1 x^T))^{r_1}) = Tr_1^{ac}((Tr_{ac}^{\ell}(Tr_\ell^n(\delta x^{2e}) + \gamma_2 x^{Te}))^{r_2}). \tag{8}$$

Let $\mu \in GF(2^n)$ satisfy $Tr_\ell^n(\mu^2) + \gamma_1\mu^T \neq 0$ (such a μ always exists) and restrict equation (8) to x of the form $x = \mu y$, $y \in GF(2^\ell)$. Then

$$Tr_1^{ab}((Tr_{ab}^{\ell}(\sigma_1 y^2))^{r_1}) = Tr_1^{ac}((Tr_{ac}^{\ell}(\sigma_2 y^{2e}))^{r_2}), \tag{9}$$

where $\sigma_1 = Tr_\ell^n(\mu^2) + \gamma_1\mu^T$ and $\sigma_2 = Tr_\ell^n(\delta\mu^{2e}) + \gamma_2\mu^{Te}$. Note that $\sigma_2 \neq 0$. As in the derivation of the linear span of a GMW sequence, it can be shown that the number of nonzero terms in the expansion of the left hand side is

$$ab \cdot (cd)^{\mathrm{wt}(r_1)},$$

while the number of nonzero terms on the right hand side is

$$ac \cdot (bd)^{\mathrm{wt}(r_2)}.$$

It follows that

$$c^{\mathrm{wt}(r_1)-1} d^{\mathrm{wt}(r_1)} = b^{\mathrm{wt}(r_2)-1} d^{\mathrm{wt}(r_2)}.$$

Without loss of generality, we may assume that $\mathrm{wt}(r1) \le \mathrm{wt}(r_2)$. It follows from the fact that b and c are relatively prime that either $b = 1$, or $\mathrm{wt}(r_2) = 1$. In the latter case we also must have $\mathrm{wt}(r_1) = 1$, that is, r_1 and r_2 are powers of two, which we have assumed is not the case. Therefore $b = 1$ and

$$c^{\mathrm{wt}(r_1)-1} = d^{\mathrm{wt}(r_2)-\mathrm{wt}(r_1)}. \tag{10}$$

Now if we further restrict y to be in $GF(2^{ac})$ in equation (9), we get

$$Tr_1^{a}((Tr_a^{ac}(\pi_1 y^2))^{r_1}) = Tr_1^{ac}(\pi_2 y^{2er_2}),$$

for nonzero π_1 and π_2. The number of nonzero terms on the left hand side is

$$a \cdot c^{\mathrm{wt}(r_1)}$$

while the number of nonzero terms on the right hand side is ac. It follows that either $c = 1$, in which case $m_1 = m_2$, or $\mathrm{wt}(r_1) = 1$. In the latter case, from equation (10) and the fact that r_1 and r_2 are not both powers of two, we see that $d = 1$. This implies that

$$Tr_1^{\ell}(Tr_\ell^n(x^2) + \gamma_1 x^T) = Tr_1^{\ell}((Tr_\ell^n(\delta x^{2e}) + \gamma_2 x^{Te})^{r_2}).$$

By considering the linear spans of the sequences corresponding to these two functions, we see that this is possible only if $\mathrm{wt}(r_2) = 1$. This proves the lemma. □

The completion of the proof of the proposition is essentially the same as the proof of the corresponding result of No and Kumar (Lemma 2 of [14]). The details are left to the reader. □

n	Period	$\mathbf{N}_{\mathrm{TN}}$
6	63	12
8	255	32
10	1 023	360
12	4 095	1 008
14	16 383	13 608
16	65 535	34 816
18	262 143	381 024
20	1 048 575	1 560 000
22	4 194 303	21 125 632
24	16 777 215	41 748 480

Table 2: Number of Distinct Families of TN Sequences of Period $2^n - 1$

Theorem VI.3 *Let $n = 2\ell$. The number $\mathbf{N}_{\mathrm{TN}}$ of distinct families of TN sequences of period $2^n - 1$ is given by*

$$\mathbf{N}_{\mathrm{TN}} = \frac{\phi(2^n-1)}{n} \cdot \left(\sum_{m|\ell} \left(\frac{\phi(2^m-1)}{m} - 1 \right) + 1 \right),$$

where $\phi(\cdot)$ is Euler's phi function.

Proof: If α_1 and α_2 are primitive elements in $GF(2^n)$, then they are equivalent for purposes of generating families of TN sequences if they are in the same Galois coset, i.e., if $\alpha_2 = \alpha_1^{2^j}$ for some j. If $m|\ell$ is chosen, and $0 \le r_1, r_2 < 2^m - 1$, then we say r_1 is equivalent to r_2, written $r_1 \sim_m r_2$, if $r_2 \equiv 2^j r_1 \bmod 2^m - 1$ for some j. By Proposition VI.1, a family of TN sequences is uniquely determined by the following: a choice of Galois coset of primitive elements of $GF(2^n)$; a choice of divisor m of ℓ; and a choice of $\sim_m$ equivalence class r, with $r \not\sim_m 1$. In addition, there is a family of TN sequences for each choice of Galois equivalence class of primitive elements and $r = 1$.

The number of Galois equivalence classes of primitive elements is $\phi(2^n - 1)/n$. For a given m, the number of $\sim_m$ equivalence classes is $\phi(2^m - 1)/m$, which proves the theorem. □

The values of $\mathbf{N}_{\mathrm{TN}}$ for the first few n are summarized in Table 2. Note that the No sequences correspond to the choice $m = \ell$. Thus, if we restrict the sum in Theorem VI.3 to $m = \ell$, we obtain the number of No sequences, as computed previously [14].

VII Complete Families of Degree d d-Form Sequences

In this section we consider families of d-form sequences in which the fields $GF(q)$ and $GF(q^e)$, exponent r, the degree d, and the primitive element α are fixed, but the d-form $H(x)$ is free to vary through all nonzero polynomials that are homogeneous of degree d on $GF(q^e)$ over $GF(q)$. We call such a family a *complete family of degree d d-form sequences*. Here we restrict our attention to the case when $e = 2$. Note that for $d = 2$, this gives a family of No sequences. In general the size N of such a family (identifying sequences that are cyclic shifts of each other) is

$$\begin{aligned} N &= \begin{cases} \frac{q^{d+1}-1}{q^2-1} & \text{if } d \text{ is odd,} \\ \frac{q^{d+1}-q}{q^2-1}+1 & \text{if } d \text{ is even,} \end{cases} \\ &\sim q^{d-1}. \end{aligned}$$

Furthermore, cross-correlations in a complete family of degree d d-form sequences are bounded as follows.

Theorem VII.1 *If $\mathcal{F}$ is a complete family of degree d d-form sequences, then every cross-correlation of any two sequences from $\mathcal{F}$ is contained in the set $\{-q-1, -1, q-1, 2q-1, \cdots, (d-1)q-1\}$. In particular, the cross-correlations are at most $d+1$ valued.*

Proof: As we have shown, it suffices to compute the number z of nontrivial zeros of any d-form $H(x, y)$ in two variables over $GF(q)$. We show by induction that z is in the set $\{i(q-1) : 0 \leq i \leq d\}$, and the theorem follows from Theorem II.1. We can take $d = 2$ as base case, since any quadratic polynomial in two variables over $GF(q)$ has 0, 1, or 2 roots.

For the induction case, observe that if we apply a change of coordinates

$$x \mapsto ax + by, \qquad y \mapsto cx + dy,$$

then $H(x, y)$ will be put in the form

$$H(x, y) = xG(x, y) + H(b, d)y^d$$

where $G(x, y)$ is a $d-1$-form. If H has no nontrivial roots, we are done. Otherwise, we can find such a change of coordinates for which $H(b, d) = 0$. In other words, we can assume

$$H(x, y) = xG(x, y).$$

It follows that $H(x, y) = 0$ if either $x = 0$ or $G(x, y) = 0$. There are exactly $q - 1$ nontrivial pairs (x, y) with $x = 0$. By the homogeneity of G, they are either all roots of G, or none are. Thus the number of roots of H is either equal to the number of roots of G, or is $q - 1$ greater than the number of roots of G. The theorem follows by induction on G. $\square$

If $d = 3$, the size of such a family is about the same as a family of Gold sequences. The maximum correlations are the same as for Gold sequences with even n. The linear spans of d-form sequences, however, are generally much larger.

If $d = 4$, the size of such a family is about the same as a large set of Kasami sequences. The maximum correlations are about one and a half times those of a large set of Kasami sequences. Again, the linear spans of d-form sequences are generally much larger.

For $d \geq 5$, the maximum correlations increase linearly in d, while the size of the family increases exponentially in d.

VIII Acknowledgements

The author thanks the referees of this paper for pointing out the previous results on generalized No sequences in No's thesis and for making a number of suggestions that improved the manuscript.

References

[1] M. Antweiller and L. Bömer, "Complex sequences over $GF(p^m)$ with a two-level autocorrelation function and a large linear span," *IEEE Trans. on Info. Theory*, vol. 38, pp. 120-130, Jan. 1992.

[2] S. Golomb, *Shift Register Sequences,* Aegean Park Press: Laguna Hills, CA, 1982.

[3] B. Gordon, W. H. Mills, and L. R. Welch, "Some new difference sets," *Canad. J. Math.* vol. 14 pp. 614-625, 1962.

[4] T. Kasami, "Weight distribution formula for some classes of cyclic codes," Coordinated Science Laboratory, University of Illinois, Urbana, Tech. Rep. R-285 (AD632574), 1966.

[5] T. Kasami, "Weight distribution of Bose-Chaudhuri-Hocquenghem codes," in *Combinatorial Mathematics and its Applications.* Chapel Hill, NC: University of North Carolina Press, 1969.

[6] E. L. Key, "An Analysis of the structure and complexity of nonlinear binary sequence generators," *IEEE Trans. Info. Theory,* vol. IT-22 no. 6, pp. 732-736, Nov. 1976.

[7] A. Klapper, "Cross-correlations of geometric sequences in characteristic two," *Designs, Codes, and Cryptography*, vol. 3, pp. 347-377, 1993.

[8] A. Klapper, A.H. Chan, and M. Goresky, "Cross-correlations of linearly and quadratically related geometric sequences and GMW Sequences," *Discrete Applied Mathematics,* vol. 46, pp. 1-20, 1993.

[9] A. Klapper, A.H. Chan, and M. Goresky, "Cascaded GMW Sequences," *IEEE Trans. on Info. Theory,* vol. IT-39, pp. 177-183, 1993.

[10] P. V. Kumar and R. A. Scholtz, "Bounds on the linear span of bent sequences," *IEEE Trans. Info. Theory,* vol. IT-29, pp. 854-862, Nov. 1983.

[11] A. Lempl and M. Cohn, "Maximal families of bent sequences," *IEEE Trans. Info. Theory,* vol. IT-28, pp. 865-868, Nov. 1982.

[12] R. Lidl and H. Niederreiter *Finite Fields* in *Encyclopedia of Mathematics Vol. 20,* Cambridge University Press: Cambridge, 1983.

[13] J. No, *A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span,* Doctoral Dissertation, University of Southern California, 1988.

[14] J. No and P. V. Kumar, "A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span," *IEEE Trans. on Info. Theory*, vol. 35, pp. 371-379, 1989.

[15] J. D. Olsen, R. A. Scholtz, and L. R. Welch, "Bent-function sequences," *IEEE Trans. Inform. Theory,* vol. IT-28, pp. 858-864, Nov. 1982.

[16] O. Rothaus, "On bent functions," *Journal of Combinatorial Theory Series A*, vol. 20, pp. 300-305, 1976.

[17] M. Simon, J. Omura, R. Scholtz, and B. Levitt, *Spread-Spectrum Communications Vol. 1*, Computer Science Press: 1985.

[18] J. Wolfmann, "New bounds on cyclic codes from algebraic curves," in *Proc. 1988 Conference on Coding Theory and Its Applications,* G. Cohen, J. Wolfmann, Eds., *Lecture Notes in Computer Science Vol. 388*, Springer-Verlag: Berlin, pp. 47-62, 1989.

Table 1:

Family	n	Size of Family	Maximum Correlation	Maximum Linear Span	Range of Imbalance
Gold	$2m+1$	2^n+1	$1+2^{\frac{n+1}{2}}$	$2n$	$[1, 2^{\frac{n+1}{2}}+1]$
Gold	$4m+2$	2^n-1	$1+2^{\frac{n+2}{2}}$	$2n$	$[1, 2^{\frac{n+2}{2}}+1]$
Kasami (Small Set)	$2m$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$\frac{3n}{2}$	$[1, 2^{\frac{n}{2}}+1]$
Kasami (Large Set)	$4m+2$	$2^{\frac{n}{2}}(2^n+1)$	$1+2^{\frac{n+2}{2}}$	$\leq \frac{5n}{2}$	$[1, 2^{\frac{n+2}{2}}+1]$
Bent	$4m$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$\binom{n/2}{n/4}2^{\frac{n}{4}}$	1
No	$2m$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$m(2^m-1)$	$[1, 2^{\frac{n}{2}}+1]$
TN	$2km$	$2^{\frac{n}{2}}$	$1+2^{\frac{n}{2}}$	$> 3mk(3k-1)^{m-2}$	$[1, 2^{\frac{n}{2}}+1]$

Table 1: Comparison of Properties of Families of Sequences of Period $2^n - 1$

Table 2:

n	Period	$\mathbf{N}_{\mathrm{TN}}$
6	63	12
8	255	32
10	1 023	360
12	4 095	1 008
14	16 383	13 608
16	65 535	34 816
18	262 143	381 024
20	1 048 575	1 560 000
22	4 194 303	21 125 632
24	16 777 215	41 748 480

Table 2: Number of Distinct Families of TN Sequences of Period $2^n - 1$

Captions of Tables:

1. Comparison of Properties of Families of Sequences of Period $2^n - 1$
2. Number of Distinct Families of TN Sequences of Period $2^n - 1$

Footnote 1:

It should be noted that the necessary assumption that $t < q^k/2$ was not explicitly stated by No and Kumar. It poses no difficulty for their analysis, however, since any exponent over $GF(q^k)$ which is relatively prime to $q^k - 1$ can be assumed to be less than $q^k - 1$. Such an exponent thus has at least one zero in the first mk bits of its base two expansion. Since No sequences are invariant under cyclic shift of the exponent r, we can assume the exponent has a zero as its high bit. In the more general case of TN sequences, we can also cyclically shift r so its high bit is zero, and therefore the high bit of each $\sum_i a_i q^i$ is zero.

Biography

Andrew M. Klapper was born in White Plains, New York, in 1952. He received the A.B. degree in mathematics from New York University, New York, NY, in 1974, the M.S. degree in applied mathematics from SUNY at Binghamton, Binghamton, NY, in 1975, the M.S. degree in mathematics from Stanford University, Stanford, CA, in 1976, and the Ph.D. degree in mathematics from Brown University, Providence, RI, in 1982. His thesis, in the area of arithmetic geometry, concerned the existence of canonical subgroups in formal grouplaws.

From 1981 to 1984 he was a post doctoral fellow and visiting professor in the Department of Mathematics and Computer Science at Clark University, Worcester, MA. From 1984 to 1991 he was an assistant professor in the College of Computer Science at Northeastern University, Boston, MA. From 1991 to 1993 he was an assistant professor in the Computer Science Department at the University of Manitoba in Winnipeg, Manitoba, Canada. Currently he is an assistant professor in the Computer Science Department at the University of Kentucky. His past research has included work on algebraic geometry over p-adic integer rings, computational geometry, modelling distributed systems, structural complexity theory, and cryptography. His current interests include statistical properties of pseudo-random sequences, stream ciphersm, public key cryptosystems, and morris dancing.