THE CROSSCORRELATION DISTRIBUTION OF THE NIHO DECIMATION $d = 4(2^m - 1) + 1$: ONE UNKNOWN, AND ITS VALUE FOR $m \leq 15$

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. Let $n = 2m$, let α be a primitive element of $\mathbb{F}_{2^n}$, let $s_t = \mathrm{Tr}(\alpha^t)$ be the binary m-sequence it generates, and let $C_d(\tau)$ be the crosscorrelation between $\{s_t\}$ and its decimation by d. For the Niho decimation $d = 4(2^m - 1) + 1 = 2^{m+2} - 3$, Helleseth, Katz and Li proved that C_d takes at most five values when m is even and at most six when m is odd; the multiplicities are Open Problems 7 and 9 of the 2024 survey of Helleseth and Li, where both are stated to be open. We contribute three things. First, an elementary reduction: the four power moments available in the literature — three classical, the fourth a specialisation of a recent formula of Cui, Luo and Xiang — form a linear system that determines every multiplicity from N_4 alone when m is even, and from N_3 and N_4 when m is odd. We solve it explicitly, so each open problem becomes one missing integer (two for odd m). Second, we determine that integer exactly for every m with $2 \leq m \leq 15$, by an exhaustive computation of the whole distribution over $\mathbb{F}_{2^{2m}}$ for $2 \leq m \leq 11$ inside a proof assistant and for $12 \leq m \leq 15$ outside it. Third, we observe that the answer is governed at every one of those m by the same integer $T_m = 2^m \tau_m$, satisfying $T_1 = 1$, $T_2 = -7$ and $T_{k+2} = T_{k+1} - 4T_k$, that Xia, Li, Zeng and Helleseth found for the neighbouring decimation $3(2^m - 1) + 1$: for odd m the two decimations have the *identical* distribution, although they are inequivalent for $m \geq 5$, and for even m the distribution of $4(2^m - 1) + 1$ is that of $3(2^m - 1) + 1$ after the single substitution $T_m \mapsto 2^{m+1} + 2 - T_m$. We state this as a conjecture and are explicit that no proof for general m is offered. Apart from the conjecture itself, the four values computed outside the proof assistant, one elementary enumeration and one polynomial identity — each flagged where it occurs — every assertion below is machine-checked in Lean 4.

1. INTRODUCTION

Crosscorrelation of m-sequences. Fix $n \geq 1$ and a primitive element α of $\mathbb{F}_{2^n}$, and let $\mathrm{Tr}\colon \mathbb{F}_{2^n} \to \mathbb{F}_2$ be the absolute trace. The sequence

$$s_t = \mathrm{Tr}(\alpha^t), \qquad t = 0, 1, \ldots, 2^n - 2,$$

is a binary m-sequence of period $2^n - 1$. For d with $\gcd(d, 2^n - 1) = 1$ the decimated sequence $\{s_{dt}\}$ is again an m-sequence of the same period, and the *crosscorrelation* between the two is

$$C_d(\tau) = \sum_{t=0}^{2^n-2} (-1)^{s_{t+\tau} - s_{dt}}, \qquad \tau = 0, 1, \ldots, 2^n - 2, \tag{1}$$

the indices $t + \tau$ read modulo $2^n - 1$. The *crosscorrelation distribution* of d is the multiset $\{C_d(\tau) : 0 \leq \tau \leq 2^n - 2\}$, recorded as the multiplicity of each value; it does not depend on the choice of α. Writing $f(x) = \mathrm{Tr}(x^d)$ and

$$\mathrm{W}_f(a) = \sum_{x \in \mathbb{F}_{2^n}} (-1)^{f(x) + \mathrm{Tr}(ax)},$$

one has the standard identity [3, §2]

$$C_d(\tau) = -1 + \mathrm{W}_f(\alpha^\tau), \tag{2}$$

which turns the distribution of C_d into the distribution of the Walsh transform of a monomial trace function on the nonzero elements of $\mathbb{F}_{2^n}$.

Determining crosscorrelation distributions is a fifty-year programme; [3] is the current survey. This paper concerns one entry of it.

Niho decimations, and the two open problems. Let $n = 2m$ and write $q = 2^m$ and $Q = 2^n = q^2$ throughout. A decimation d is of *Niho type* when $d \equiv 1 \pmod{2^m - 1}$, that is $d = s(2^m - 1) + 1$ for an integer s. The parameter $s = 1$ gives $d = 2^m$, a power of two, for which C_d is just the autocorrelation of the m-sequence; $s = 2, 3, 4$ are the first nontrivial cases. This paper is about

$$d = 4(2^m - 1) + 1 = 2^{m+2} - 3.$$

Its value set is known. Resolving the last of Niho's conjectures, Helleseth, Katz and Li [2] proved that for this d

$$C_d(\tau) \in \{-1 + j \cdot 2^m\} \quad \text{with} \quad j \in \begin{cases} \{-1, 0, 1, 2, 4\}, & m \text{ even}, \\ \{-1, 0, 1, 2, 3, 4\}, & m \text{ odd}, \end{cases} \tag{3}$$

as quoted in [3, Thms. 11 and 17], so C_d is at most five-valued for even m and at most six-valued for odd m. What is not known is how often each value occurs. Helleseth and Li close the corresponding two sections of their survey with exactly that question [3, Open Problems 7 and 9]:

Problem 1.1 ([3, Open Problem 7])**.** Let $n = 2m$ with even m. Compute the crosscorrelation distribution of $C_d(\tau)$ for $d = 4(2^m - 1) + 1$.

Problem 1.2 ([3, Open Problem 9])**.** Let $n = 2m$ with odd m. Determine the distribution of $C_d(\tau)$ for $d = 4(2^m - 1) + 1$.

Problem 1.2 is introduced there with the sentence "Nevertheless, the distribution of $C_d(\tau)$ still remains open."

The neighbouring decimation $d = 3(2^m - 1) + 1$ *is* settled, by Xia, Li, Zeng and Helleseth [6], through a connection with the weight distribution of binary Zetterberg codes. Their answer is expressed in the parameter

$$\tau_m = \left(\tfrac{1+\sqrt{-15}}{4}\right)^m + \left(\tfrac{1-\sqrt{-15}}{4}\right)^m, \qquad \tau_1 = \tfrac{1}{2}, \quad \tau_2 = -\tfrac{7}{4}, \quad \tau_{k+2} = \tfrac{1}{2}\tau_{k+1} - \tau_k, \tag{4}$$

which we shall always use in the cleared form

$$T_m := 2^m \tau_m = \pi^m + \bar{\pi}^m, \qquad \pi = \tfrac{1+\sqrt{-15}}{2}, \quad \pi\bar{\pi} = 4, \tag{5}$$

an integer with $T_1 = 1$, $T_2 = -7$ and $T_{k+2} = T_{k+1} - 4T_k$; thus $T_3 = -11$, $T_4 = 17$, $T_5 = 61$, $T_6 = -7$, $T_7 = -251$, and so on. (In (4) we have restored the conjugation: [3] prints both summands as $\big((1 + \sqrt{-15})/4\big)^m$, but its own recurrence and its own alternative form $2\cos\big(m \arctan\sqrt{15}\big)$ both give the expression above.)

What is settled here. Let

$$N_j = \#\{\tau : 0 \le \tau \le 2^n - 2,\ C_d(\tau) = -1 + j \cdot 2^m\}, \qquad -1 \le j \le 4,$$

so that $\sum_j N_j = 2^n - 1$ by (3). Three contributions, in increasing order of what they leave unproved.

(i) The reduction (§3). Four power-moment identities for this crosscorrelation are available in print: the trivial count of $\mathbb{F}_{2^n}$ and the first two power moments of the Walsh transform [4, Cor. 7.4(ii),(iii)], in the form quoted as Lemma 3 of [3], and a fourth obtained by specialising a very recent third-moment formula of Cui, Luo and Xiang [1] to $p = 2$, $s = 4$. They are four linear equations in the five (even m) or six (odd m) unknowns N_j. Solving them (Theorems 3.1 and 3.2) leaves exactly one free parameter for even m and two for odd m; so Problem 1.1 is the single integer N_4, and Problem 1.2 is the pair (N_3, N_4). Cui, Luo and Xiang say as much for even m — "*only one additional equation is needed to fully determine the distribution*" [1, Remark 1(3)] — without printing the solved system, and they do not treat odd m; the solved system is written out here because we could not find it anywhere.

(ii) The numbers (§4). We compute the full distribution exhaustively for every m with $2 \le m \le 11$ (Theorems 4.1, 4.2 and 4.3), that is, over fields up to $\mathbb{F}_{2^{22}}$. Nothing in these computations is assumed: the field model is verified rather than postulated, $\gcd(d, 2^n - 1) = 1$ is checked, and the count of shifts τ whose crosscorrelation is *not* of the form $-1 + j \cdot 2^m$ with $-1 \le j \le 4$ comes out 0 every time, which

re-derives (3) at each m instead of assuming it. We could find no values of the distributions asked for in Problems 1.1 and 1.2 anywhere in the literature.

(iii) The pattern (§§5–6). The missing parameters are, at every m we can reach,

$$m \text{ even:} \quad N_3 = 0, \quad N_4 = \frac{q^2 + T_m - 2q - 1}{120}; \qquad m \text{ odd:} \quad N_3 = \frac{q-2}{3}, \quad N_4 = \frac{q^2 - T_m - 12q + 21}{120},$$

with T_m the integer (5) of the *neighbouring* decimation $s = 3$. We verify this identically for $2 \le m \le 11$ inside the formal development (Theorem 5.2) and for $12 \le m \le 15$ by an independent program (Remark 4.4), and we conjecture it in general (Conjecture 5.1); we prove nothing for general m, and say so again in §9. Two features of the pattern deserve to be stated on their own. For odd m the resulting distribution is *entry for entry* the published distribution of $3(2^m-1)+1$ (Theorem 6.1), even though for $m \ge 5$ the two decimations lie in different equivalence classes, so this is not an artefact of the symmetries of C_d. For even m the two distributions genuinely differ, and the closed form of one is obtained from the other by the substitution $T_m \mapsto 2^{m+1} + 2 - T_m$ (Proposition 6.2). We have no explanation for either phenomenon.

Provenance of the ingredients. The value set (3) is [2]; the low-order moments are classical, in the form of [4]; the third moment is [1]; the parameter T_m and the $s = 3$ distribution are [6]. What is offered here is the solved linear system, the computed distributions, the closed form, and the coincidence.

2. Preliminaries

Setting. Throughout, $m \ge 2$, $n = 2m$, $q = 2^m$, $Q = 2^n = q^2$, and $d = 4(2^m - 1) + 1 = 2^{m+2} - 3$ unless another decimation is named. For $2 \le m \le 15$ one checks directly that $\gcd(d, 2^n - 1) = 1$, so C_d is defined by (1); this is in contrast with $d = 3(2^m - 1) + 1$, for which $\gcd(d, 2^n - 1) = 5$ when $m \equiv 2 \pmod 4$, and it is why Problems 1.1 and 1.2 carry no congruence side condition while [6, Thm. 2(i)] does.

Multiplicities, and the two normalisations. Two counts are in play and must be kept apart. On the sequence side, N_j counts *shifts*: $N_j = \#\{\tau : C_d(\tau) = -1 + jq\}$, with τ ranging over $0, \dots, 2^n - 2$. On the Walsh side, put

$$A_j = \#\{a \in \mathbb{F}_{2^n} : \mathrm{W}_f(a) = j \cdot 2^m\}.$$

By (2) the map $\tau \mapsto \alpha^\tau$ is a bijection from the shifts onto $\mathbb{F}_{2^n}^\times$, and $\mathrm{W}_f(0) = \sum_x (-1)^{\mathrm{Tr}(x^d)} = 0$ because $x \mapsto x^d$ is a permutation of $\mathbb{F}_{2^n}$ and Tr is balanced; this is recorded in [3, §2]. Hence

$$A_j = N_j + [\, j = 0 \,], \tag{6}$$

the extra element being $a = 0$. Power moments are naturally stated for the A_j and the open problems for the N_j; (6) is the only bookkeeping between them.

The four power moments.

Lemma 2.1 ([4, Cor. 7.4(ii),(iii)], as quoted in [3, Lemma 3]; [1, Cor. 1] specialised)**.** *For $d = 4(2^m - 1) + 1$ over $\mathbb{F}_{2^{2m}}$,*

$$\sum_j A_j = Q, \qquad \sum_j j A_j = q, \qquad \sum_j j^2 A_j = Q, \qquad \sum_j j^3 A_j = \begin{cases} Q, & m \text{ even}, \\ Q + 2q, & m \text{ odd}. \end{cases}$$

The zeroth identity is the count of $\mathbb{F}_{2^n}$. The next two are the first and second power moments of the Walsh transform of $\mathrm{Tr}(x^d)$ for an invertible exponent d, $\sum_{a \ne 0} \mathrm{W}_f(a) = Q$ and $\sum_{a \ne 0} \mathrm{W}_f(a)^2 = Q^2$; since $\mathrm{W}_f(0) = 0$ these are also the sums over all of $\mathbb{F}_{2^n}$, and dividing the l-th of them by q^l, with $\mathrm{W}_f(a) = j \cdot q$ and $Q = q^2$, gives $\sum_j j A_j = q$ and $\sum_j j^2 A_j = Q$.

The fourth is a specialisation. For a Niho exponent $d = s(p^m-1)+1$ over $\mathbb{F}_{p^n}$ with $n = 2m$, and $F(x) = x^d$, Cui, Luo and Xiang prove [1, Cor. 1]

$$\sum_{u,v\in\mathbb{F}_{p^n}} \mathrm{W}_F(u,v)^3 = p^{2n}\big(3p^n-2+(p^n-1)N_1'\big), \qquad N_1' = p^m+(s_1-1)(s_1-2)+(s_2-1)(s_2-2)-2, \tag{7}$$

with $s_1 = \gcd(s, p^m+1)$ and $s_2 = \gcd(s-1, p^m+1)$, so that the pair enters only through $(s_1-1)(s_1-2)+(s_2-1)(s_2-2)$. Splitting the left-hand side by u turns this into a statement about W_f: the term $u = 0$ contributes p^{3n}, and each of the p^n-1 terms with $u \neq 0$ contributes $\sum_a \mathrm{W}_f(a)^3$, since $x \mapsto x^d$ is a permutation and the substitution $x \mapsto u^{-1/d}x$ carries $\mathrm{W}_F(u,v)$ to $\mathrm{W}_f(vu^{-1/d})$. Hence

$$\sum_a \mathrm{W}_f(a)^3 = \frac{p^{2n}\big(3p^n-2+(p^n-1)N_1'\big)-p^{3n}}{p^n-1} = p^{2n}(N_1'+2).$$

At $p = 2$, $s = 4$ one has $s_1 = \gcd(4, 2^m+1) = 1$ always, so $(s_1-1)(s_1-2) = 0$, while $s_2 = \gcd(3, 2^m+1)$ is 1 for even m and 3 for odd m, so $(s_2-1)(s_2-2)$ is 0 respectively 2. Thus $N_1'+2 = q$ respectively $q+2$, and $\sum_a \mathrm{W}_f(a)^3$ is q^5 respectively $q^4(q+2)$; dividing by q^3 gives the stated fourth line. We record that the specialisation was checked against the computed censuses of §4: at $m = 4$, $\sum_j j^3A_j = 256 = q^2$, and at $m = 5$, $\sum_j j^3A_j = 1088 = q^2+2q$.

The computational model. All exhaustive computations below use one model of $\mathbb{F}_{2^n}$ and one transform, described here once.

Take $\mathbb{F}_{2^n} = \mathbb{F}_2[x]/(x^n + r_n(x))$ with r_n of degree $< n$, and identify a field element with the integer whose bit i is the coefficient of x^i. The polynomial is chosen so that the class β of x is primitive; *this is verified, not assumed*, by listing $1, \beta, \beta^2, \ldots, \beta^{2^n-2}$ and checking that the value 1 occurs exactly once in the list, which happens precisely when β has full order 2^n-1. Since Tr is $\mathbb{F}_2$-linear, $\mathrm{Tr}(y)$ is the parity of $y \wedge M$ for the mask M whose i-th bit is $\mathrm{Tr}(\beta^i)$; that each of those n values really lies in $\mathbb{F}_2$ is again checked rather than assumed.

For the distribution itself, $a \mapsto \mathrm{Tr}(a\,\cdot)$ is an $\mathbb{F}_2$-linear bijection of $\mathbb{F}_{2^n}$ onto its dual, so with $g[x] = (-1)^{\mathrm{Tr}(x^d)}$ indexed by the bit vector of x and $G[u] = \sum_x (-1)^{u\cdot x} g[x]$ the ordinary Walsh–Hadamard transform, the multiset $\{G[u] : u \neq 0\}$ equals $\{\mathrm{W}_f(a) : a \neq 0\}$ and $G[0] = \mathrm{W}_f(0)$. One transform of $n \cdot 2^n$ butterflies therefore yields the whole distribution. Because we only ever want the multiset, the permutation of indices costs nothing.

Each computation reported below is one such transform, from which five things are read off: that the field model checks out; that $\gcd(d, 2^n-1) = 1$; that $G[0] = 0$; the seven counts

$$[\,N_{-1},\ N_0,\ N_1,\ N_2,\ N_3,\ N_4,\ \mathrm{off}\,],$$

where off is the number of shifts whose crosscorrelation is *not* of the form $-1+j\cdot 2^m$ with $-1 \le j \le 4$; and that the four identities of Lemma 2.1 hold on those counts. A value off $= 0$ is (3) re-derived at that m.

3. The moment reduction: one unknown, or two

The following two theorems are pure integer linear algebra and involve no computation over a field. They are stated for indeterminate Q and q; the relation $Q = q^2$ is never used.

Theorem 3.1. *Let $Q, q, N_{-1}, N_0, N_1, N_2, N_4$ be integers satisfying the four identities of Lemma 2.1 in the five-valued shape, that is*

$$N_{-1}+(N_0+1)+N_1+N_2+N_4 = Q, \qquad -N_{-1}+N_1+2N_2+4N_4 = q,$$
$$N_{-1}+N_1+4N_2+16N_4 = Q, \qquad -N_{-1}+N_1+8N_2+64N_4 = Q.$$

Then

$$3N_{-1} = Q-q+12N_4, \quad 2N_0 = Q-q-30N_4-2, \quad N_1 = q+20N_4, \quad 6N_2 = Q-q-60N_4.$$

Theorem 3.2. *Let $Q, q, N_{-1}, N_0, N_1, N_2, N_3, N_4$ be integers satisfying the four identities in the six-valued shape, that is*

$$N_{-1} + (N_0 + 1) + N_1 + N_2 + N_3 + N_4 = Q, \qquad -N_{-1} + N_1 + 2N_2 + 3N_3 + 4N_4 = q,$$
$$N_{-1} + N_1 + 4N_2 + 9N_3 + 16N_4 = Q, \qquad -N_{-1} + N_1 + 8N_2 + 27N_3 + 64N_4 = Q + 2q.$$

Then

$$3N_{-1} = Q - 2q + 3N_3 + 12N_4, \qquad 2N_0 = Q + q - 8N_3 - 30N_4 - 2,$$
$$N_1 = 6N_3 + 20N_4, \qquad 6N_2 = Q + q - 24N_3 - 60N_4.$$

Proof of Theorems 3.1 and 3.2. Elimination in $\mathbb{Z}$, in four steps. Subtracting the second identity from the fourth cancels N_{-1} and N_1 and yields the stated relation for $6N_2$; adding the second to the third and substituting that relation yields N_1; subtracting the second from the third and substituting again yields $3N_{-1}$; and the first identity then yields $2N_0$. The odd case is the same elimination with N_3 carried along as a second free parameter. Both are verified in the formal development by a decision procedure for linear integer arithmetic, with no appeal to computation over $\mathbb{F}_{2^n}$. □

Two remarks keep these from being vacuous or weaker than they look.

Proposition 3.3. *The reductions are determinations, not restatements. If two quadruples of integers both satisfy the hypotheses of Theorem 3.1 with the* same *value of N_4, then they agree entry by entry; likewise in Theorem 3.2 with the same N_3 and N_4. Moreover neither hypothesis is empty: the tuple $(88, 89, 56, 20, 2)$ satisfies the even system at $q = 16$, $Q = 256$, and the tuple $(350, 412, 160, 86, 10, 5)$ satisfies the odd system at $q = 32$, $Q = 1024$; both are true censuses, from §4.*

Proposition 3.4. *The fourth moment is needed. At $q = 16$, $Q = 256$ and $N_4 = 2$, the tuple $(N_{-1}, N_0, N_1, N_2) = (89, 86, 59, 19)$ satisfies the first three identities of Lemma 2.1 exactly as the true census $(88, 89, 56, 20)$ does, and differs from it; it fails the third-moment identity. So a reduction resting on the first three identities alone would not determine the distribution from N_4, and Theorem 3.1 genuinely uses* [1].

Theorems 3.1 and 3.2 are what turn the two open problems into finite questions about one, respectively two, integers. We emphasise that the input is entirely other people's: the achievement is only that the system is solved and the solution recorded.

4. The distribution, computed

We now report the exhaustive computations. Each theorem below is a single evaluation of the transform of §2 at the stated m, checked in full: the field model, the coprimality, $\mathrm{W}_f(0) = 0$, the seven counts, and the four moment identities of Lemma 2.1.

Theorem 4.1. *For $d = 4(2^m - 1) + 1$ over $\mathbb{F}_{2^{2m}}$ and $m = 2, 4, 6, 8, 10$, the crosscorrelation distribution is*

m	N_{-1}	N_0	N_1	N_2	N_3	N_4
2	4	5	4	2	0	0
4	88	89	56	20	0	2
6	1476	1520	724	342	0	33
8	23920	24539	11056	5480	0	540
10	384124	392750	175724	87242	0	8735

In each case no shift has a crosscorrelation outside $\{-1 + j \cdot 2^m : -1 \le j \le 4\}$, and $N_3 = 0$, so C_d is five-valued (four-valued at $m = 2$, where $N_4 = 0$ as well).

Theorem 4.2. *For $d = 4(2^m - 1) + 1$ over $\mathbb{F}_{2^{2m}}$ and $m = 3, 5, 7, 9$, the crosscorrelation distribution is*

m	N_{-1}	N_0	N_1	N_2	N_3	N_4
3	18	27	12	4	2	0
5	350	412	160	86	10	5
7	5922	6197	2772	1324	42	126
9	95718	98742	43560	21826	170	2127

In each case no shift falls outside the six-element value set, and $N_3 = (2^m - 2)/3$; at $m = 3$ the value $-1 + 4 \cdot 2^m$ happens not to occur, so C_d is there five-valued.

Theorem 4.3. *For $m = 11$, $n = 22$, $d = 8189$, the crosscorrelation distribution of $d = 4(2^m - 1) + 1$ over $\mathbb{F}_{2^{22}}$ is*

$$(N_{-1}, N_0, N_1, N_2, N_3, N_4) = (1536458,\ 1574047,\ 699292,\ 349064,\ 682,\ 34760),$$

with no shift outside the six-element value set.

Proof of Theorems 4.1, 4.2 and 4.3. Exhaustive computation. For each m, one Walsh–Hadamard transform of the ± 1 truth table of $\mathrm{Tr}(x^d)$ over all 2^{2m} elements of $\mathbb{F}_{2^{2m}}$, followed by a linear scan of the $2^{2m} - 1$ nonzero spectral values into the seven bins described in §2. Nothing is sampled and no symmetry is used to shorten the search: the search space is the whole field, of size 16, 64, ..., 4 194 304 elements for $m = 2, \ldots, 11$. The primitivity of the reduction polynomial is established inside the same computation by exhibiting the full multiplicative order of β, and the coprimality of d with $2^n - 1$ by a Euclidean algorithm. These claims rest on exhaustive computation and on nothing else. □

Remark 4.4. Beyond the reach of the formal development, the same distribution was computed by an independent program in C for $m = 12, 13, 14, 15$, that is up to $\mathbb{F}_{2^{30}}$, and agrees with the closed form of §5 at each:

m	N_{-1}	N_0	N_1	N_2	N_3	N_4
12	6149736	6291449	2797576	1398780	0	139674
13	24600654	25170992	11185200	5590846	2730	558441
14	98420788	100655420	44755204	22367102	0	2236941
15	393679986	402679977	178950372	89476324	10922	8944242

These four are *measurements*, outside the machine-checked development, and are quoted here only as further evidence for Conjecture 5.1. The cost is the transform array, $4 \cdot 2^n$ bytes: 4 GiB at $m = 15$, where one run took about 160 seconds of CPU time on the machine used here, and 16 GiB at $m = 16$, which we did not attempt.

Combining the computed values with §3 gives the following, which is the precise sense in which the two open problems are reduced.

Corollary 4.5. *For each m in Theorems 4.1, 4.2 and 4.3 the computed census satisfies the four identities of Lemma 2.1, so by Proposition 3.3 it is the only census compatible with those identities and with its own value of N_4 (even m), respectively of N_3 and N_4 (odd m). Thus for those m Problems 1.1 and 1.2 are answered, and answered twice over: the tables are the answer, and the moment system independently re-derives all but one entry (two, for odd m) of each.*

5. A closed form for the missing parameter

The data of §4 are governed by the integer T_m of (5), which belongs to a *different* decimation. Written out in full, with $q = 2^m$:

Conjecture 5.1. Let $m \geq 2$, $n = 2m$, $q = 2^m$ and $d = 4(2^m - 1) + 1$, and let T_m be as in (5). Then the crosscorrelation distribution of C_d is

$$\begin{aligned}
& & N_{-1} = \tfrac{1}{30}(11q^2 + T_m - 12q - 1), \quad & N_0 = \tfrac{1}{8}(3q^2 - T_m - 2q - 7),\\
& m \text{ even:} & N_1 = \tfrac{1}{6}(q^2 + T_m + 4q - 1), \quad & N_2 = \tfrac{1}{12}(q^2 - T_m + 1),\\
& & N_3 = 0, \quad & N_4 = \tfrac{1}{120}(q^2 + T_m - 2q - 1);\\
& & N_{-1} = \tfrac{1}{30}(11q^2 - T_m - 22q + 1), \quad & N_0 = \tfrac{1}{24}(9q^2 + 3T_m + 16q - 23),\\
& m \text{ odd:} & N_1 = \tfrac{1}{6}(q^2 - T_m - 3), \quad & N_2 = \tfrac{1}{12}(q^2 + T_m - 2q + 11),\\
& & N_3 = \tfrac{1}{3}(q - 2), \quad & N_4 = \tfrac{1}{120}(q^2 - T_m - 12q + 21).
\end{aligned}$$

By Theorems 3.1 and 3.2 the whole of each display follows from its last line together with N_3, so the content of the conjecture is exactly the closed form for N_4 (and, for odd m, for N_3). What we can prove is that it is correct wherever we can check it.

Theorem 5.2. *Conjecture 5.1 holds for every m with $2 \leq m \leq 11$.*

Proof. For each such m, compare the exhaustively computed census of Theorems 4.1, 4.2 and 4.3 with the closed form evaluated at that m from the recurrence $T_1 = 1$, $T_2 = -7$, $T_{k+2} = T_{k+1} - 4T_k$. All ten comparisons are equalities of explicit integers. This is a finite verification and carries no information about any $m > 11$. □

We do not know why T_m appears. For $s = 3$ it appears because [6] identify the relevant count with the weight distribution of a binary Zetterberg code, whose Frobenius data lives in $\mathbb{Q}(\sqrt{-15})$; the number $\pi = (1 + \sqrt{-15})/2$ of (5) is the same one. For $s = 4$ the object to be counted is, as recorded in [3, §5], the number of roots in the unit circle $U = \{x \in \mathbb{F}_{2^n} : x^{2^m+1} = 1\}$ of

$$g_a(x) = x^7 + ax^4 + \bar{a}x^3 + 1, \qquad \bar{a} := a^{2^m},$$

which is $j + 1$ when $C_d(\tau) = -1 + j \cdot 2^m$ and $a = \alpha^\tau$; the $s = 3$ case replaces g_a by $x^5 + ax^3 + \bar{a}x^2 + 1$. Identifying an object whose point counts reproduce the $s = 4$ root distribution and whose Weil numbers are the $\pi, \bar{\pi}$ of (5) would turn Conjecture 5.1 into a theorem. We have not found one.

6. The two decimations $s = 3$ and $s = 4$

Xia, Li, Zeng and Helleseth [6] determined the distribution for $d_3 = 3(2^m - 1) + 1$; in the normalisation of this paper, and as quoted in [3, Thms. 10 and 16], it is

$$\begin{aligned}
& m \text{ even, } m \not\equiv 2 \ (4): & N_{-1} = \tfrac{1}{30}(11q^2 - T_m - 10q + 1), \quad & N_0 = \tfrac{1}{8}(3q^2 + T_m - 4q - 9),\\
& & N_1 = \tfrac{1}{6}(q^2 - T_m + 6q + 1), \quad & N_2 = \tfrac{1}{12}(q^2 + T_m - 2q - 1),\\
& & N_3 = 0, \quad & N_4 = \tfrac{1}{120}(q^2 - T_m + 1);
\end{aligned} \tag{8}$$

$$\begin{aligned}
& m \text{ odd:} & N_{-1} = \tfrac{1}{30}(11q^2 - T_m - 22q + 1), \quad & N_0 = \tfrac{1}{24}(9q^2 + 3T_m + 16q - 23),\\
& & N_1 = \tfrac{1}{6}(q^2 - T_m - 3), \quad & N_2 = \tfrac{1}{12}(q^2 + T_m - 2q + 11),\\
& & N_3 = \tfrac{1}{3}(q - 2), \quad & N_4 = \tfrac{1}{120}(q^2 - T_m - 12q + 21).
\end{aligned} \tag{9}$$

The odd display (9) is, term for term, the odd display of Conjecture 5.1. That is not a coincidence we can explain, and it is not forced by the symmetries of the crosscorrelation, as the next theorem records.

Theorem 6.1. *For $m = 3, 5, 7, 9, 11$ the crosscorrelation distribution of $d_4 = 4(2^m - 1) + 1$ coincides, value by value and multiplicity by multiplicity, with the distribution* (9) *of $d_3 = 3(2^m - 1) + 1$ established in* [6]*. For even m this fails: at $m = 8$ the two distributions differ, and so do the closed forms* (8) *and Conjecture 5.1 at $m = 12$ (whereas at $m = 4$ they agree). Moreover the agreement for odd $m \geq 5$ is not an equivalence: for each odd m with $5 \leq m \leq 15$ one has $d_3 \not\equiv 2^j d_4^{\pm 1} \pmod{2^{2m} - 1}$ for every j.*

Proof. The first assertion is the comparison of the computed censuses of Theorems 4.2 and 4.3 with (9) evaluated at those m; at $m = 3, 5, 7, 9$ the distribution of d_3 was in addition recomputed here from scratch (Theorem 7.1 below) and agrees with (9), so at those four m the comparison is between two computed censuses, while at $m = 11$ it is between the computed census of d_4 and the published formula (9). The second assertion at $m = 8$ compares two computed censuses, which differ already in the first cell (23920 against 23952); at $m = 4$ and $m = 12$ it compares the closed forms, which agree at $m = 4$ — as they must, the two decimations being equivalent there — and differ at $m = 12$. For the third, recall from [3, Lemma 2] that the distribution of C_d is invariant under $d \mapsto 2^j d$ and $d \mapsto d^{-1}$ modulo $2^n - 1$; the equivalence class of d_4 therefore consists of at most $2n$ residues, and enumerating them for each odd m with $5 \le m \le 15$ and intersecting with the Niho residues $\{t(2^m - 1) + 1 : 1 \le t \le 2^m\}$ gives the parameter sets

$$\{4, 10, 24, 30\}\ (m = 5), \quad \{4, 19, 111, 126\}\ (m = 7), \quad \{4, 147, 367, 510\}\ (m = 9),$$

$$\{4, 586, 1464, 2046\}\ (m = 11), \quad \{4, 1171, 7023, 8190\}\ (m = 13), \quad \{4, 9363, 23407, 32766\}\ (m = 15),$$

none of which contains 3. (At $m = 3$ and $m = 4$ the classes are $\{3, 4, 6, 7\}$ and $\{3, 4, 14, 15\}$, so there the two decimations *are* equivalent and their agreement carries no information.) This last enumeration is elementary — at most $2n$ residues to list and 2^m Niho parameters to test against them — but, unlike the rest of this theorem, it is not part of the machine-checked development; it was carried out twice, by independent programs, which agree. □

The even case is different, and the difference is itself a substitution.

Proposition 6.2. *For even m, the closed form of Conjecture 5.1 is obtained from the published $s = 3$ closed form* (8) *by the single substitution*

$$T_m \ \longmapsto\ 2^{m+1} + 2 - T_m,$$

applied to all six entries simultaneously. Equivalently, writing $X_m = q^2 + 1 - 120N_4$ for the $s = 3$ normalisation of the rarest multiplicity, the $s = 4$ value of X_m is $2^{m+1} + 2 - T_m$ where the $s = 3$ value is T_m.

Proof. Substitute and compare coefficients; each of the six entries matches. For instance $\frac{1}{120}(q^2 - (2q + 2 - T_m) + 1) = \frac{1}{120}(q^2 + T_m - 2q - 1)$. For $m \equiv 2 \pmod 4$ the right-hand side of (8) is a formal expression only, since $3(2^m - 1) + 1$ is not a decimation there; the substitution rule nevertheless produces the correct $s = 4$ answer at $m = 2, 6, 10, 14$, as the data of §4 and Remark 4.4 show. □

Remark 6.3. Proposition 6.2 is arithmetic bookkeeping on Conjecture 5.1 and (8), not an independent claim. We state it because it is the shortest description of the whole conjecture: *for odd m, $s = 4$ has the $s = 3$ distribution; for even m, it has the $s = 3$ distribution with T_m replaced by $2^{m+1} + 2 - T_m$.*

7. Controls

A distribution is a tuple of large integers, and a routine that quietly returned the wrong tuple would look exactly like one that returned the right one. We therefore record what was done to separate the two. None of this section is a new claim.

Theorem 7.1. *The published distribution of $d_3 = 3(2^m - 1) + 1$* [6] *is reproduced by the computation of §2 at $m = 3, 4, 5, 7, 8, 9$, in all five respectively six multiplicities:*

m	N_{-1}	N_0	N_1	N_2	N_3	N_4
3	18	27	12	4	2	0
4	88	89	56	20	0	2
5	350	412	160	86	10	5
7	5922	6197	2772	1324	42	126
8	23952	24419	11216	5400	0	548
9	95718	98742	43560	21826	170	2127

In particular the values $T_3 = -11$ *and* $T_5 = 61$ *of* (5) *come out of the computation as they must. The residues* $m \equiv 2 \pmod 4$ *are absent because* $\gcd(d_3, 2^{2m} - 1) = 5$ *there, which the computation reports rather than assumes.*

This is somebody else's theorem recomputed with our arithmetic, and it was run before any value for $s = 4$ was looked at; it is the gate the pipeline had to pass.

Proposition 7.2. *The transform computes what* (1) *defines. Evaluating* (1) *literally — building* $\{s_t\}$ *and* $\{s_{dt}\}$ *as bit strings and summing over one full period for each* τ*, at a cost of* $(2^n - 1)^2$ *operations — reproduces the census obtained from the Walsh–Hadamard route at* $m = 2, 3, 4, 5$ *for* $s = 4$ *and at* $m = 3, 5$ *for* $s = 3$*; at* $m = 4, 5$ *for* $s = 4$ *it does so in a second field model as well.*

Proposition 7.3. *The following negative and independence controls hold.*

(1) Independence of the field model. *A second primitive reduction polynomial yields the same census for* $s = 4$ *at every* m *with* $3 \le m \le 8$*, and likewise for the literal evaluation of* (1) *at* $m = 4, 5$*. This must be so, the distribution being an invariant of* (n, d)*.*

(2) The trivial decimation is trivial. *For* $s = 1$*,* $d = 2^m$ *is a power of two, so* C_d *is the autocorrelation of the* m*-sequence, and the computation returns Golomb's ideal two-level property at* $m = 5, 6, 7$*: the value* -1 *at all* $2^n - 2$ *nonzero shifts and one shift with the off-lattice value* $2^n - 1$*. This is the only place in this work where the off-lattice counter is nonzero.*

(3) The hypotheses bite. *The primitivity check fails on* $x^{10} + x + 1$ *and on* $x^{10} + x^2 + 1$ *while passing on* $x^{10} + x^3 + 1$*, the polynomial actually used at* $n = 10$*; the coprimality check fails for* $s = 2$ *at* $m = 5$ *and for* $s = 3$ *at* $m = 6$ *and* $m = 10$*, and passes for* $s = 4$ *throughout.*

(4) The answer depends on d. *At* $m = 8$ *the censuses for* $s = 3$ *and* $s = 4$ *differ.*

(5) Too large and too small are refuted. *Moving* N_4 *by one in either direction contradicts the computed census at* $m = 8$ *and at* $m = 9$*. Likewise the value set is tight in both directions:* $N_3 \neq 0$ *at* $m = 7, 9$*, so the five-valued shape is refuted for odd* m*;* $N_3 = 0$ *at* $m = 8, 10$*, so no sixth value occurs for even* m*; and the off-lattice count is* 0 *at* $m = 9, 10$*.*

8. Verification

Every theorem, proposition and corollary stated above — with the three exceptions named below — is formally verified in Lean 4 [5]; the development contains no `sorry` and no hand-written axiom. It consists of five modules and 41 theorems, written against Lean's built-in prelude alone and without *Mathlib*, so that every statement is an equation between integers, arrays of natural numbers and booleans. The reduction of §3 (Theorems 3.1 and 3.2, Propositions 3.3 and 3.4) and the tightness of the value set asserted in the second half of Proposition 7.3(5) are kernel reasoning, discharged by a decision procedure for linear integer arithmetic and by evaluation of closed forms; they carry no evaluation axiom. The exhaustive field computations — Theorems 4.1, 4.2, 4.3, 7.1, Propositions 7.2, 7.3(1)–(4) and the refutations in the first half of 7.3(5) — are delegated to compiled evaluation, and each such statement carries the standard axioms of the ambient library together with exactly one evaluation axiom generated for it. The comparisons of computed censuses with closed forms, including Theorem 5.2 and the first two assertions of Theorem 6.1, are again kernel evaluation. The whole development re-checks in about 139 seconds at a peak resident set of about 1.7 GB, the $m = 11$ computation of Theorem 4.3 accounting for roughly two thirds of that. The three exceptions are the third assertion of Theorem 6.1 (the equivalence-class enumeration) and the four values of Remark 4.4, which are computations outside the formal development, and Proposition 6.2, which is a polynomial identity in q and T_m between two closed forms and is verified here by substituting and comparing coefficients. Each is labelled where it appears. Every number reported anywhere in this paper was produced first by an independent implementation in C and only then stated and checked formally, so the formal run re-checks a known answer rather than searching for one. The development lives in a repository that is private at the time of writing; repository reference to be added at submission.

Remark 8.1. The ceiling at $m = 11$ is an artefact of evaluation speed, not of the mathematics. Under the proof assistant's bytecode interpreter the case $m = 11$ takes about 103 seconds, while $m = 12$

had not finished after 25 minutes and was abandoned; the same computation in C takes about 3.2 seconds. That gap, and not the size of $\mathbb{F}_{2^{24}}$, is why $m = 12, \ldots, 15$ appear in Remark 4.4 rather than as theorems.

9. What is open

Problems 1.1 and 1.2 themselves. They are open. What is settled is that each is a single integer, explicitly (§3); that the integer is known at every m with $2 \le m \le 15$ (§4, Remark 4.4); and that a closed form fits all fourteen of those values (§5). Nothing above proves the closed form for a single $m > 15$. We note for the record that [1] reduce the even case to "one additional equation" and call determining this distribution a natural next step.

The odd-m coincidence. The cleanest formulation of Problem 1.2's conjectural answer is Theorem 6.1: for odd m, $s = 3$ and $s = 4$ have the same distribution. Both distributions are controlled by root counts in the unit circle $U = \{x : x^{2^m+1} = 1\}$, for $x^5 + ax^3 + \bar{a}x^2 + 1$ and $x^7 + ax^4 + \bar{a}x^3 + 1$ respectively. A correspondence between the two families of curves that preserves the number of roots in U, valid for odd m only, would prove it outright and would be the most valuable single step here.

The one easy multiplicity. For odd m the conjecture asserts $N_3 = (2^m - 2)/3$, that is, the number of a for which g_a has exactly four roots in U. The relevant gcd there is $\gcd(3, 2^m + 1) = 3$ — the same $s_2 = 3$ that makes the third moment of [1] differ between the parities — and a direct count looks within reach. It would halve Problem 1.2.

Identifying T_m intrinsically. For $s = 3$ the appearance of $\mathbb{Q}(\sqrt{-15})$ is explained by Zetterberg codes. For $s = 4$ the same $\pi = (1+\sqrt{-15})/2$ appears, and Proposition 6.2 says the even-m parameter is exactly $2^{m+1} + 2 - T_m$. Whatever counts the roots of g_a ought to carry the same Weil numbers; naming it is the route from Conjecture 5.1 to a theorem.

More data. Not the bottleneck. With the present algorithm $m = 16$ needs 16 GiB, and the fourteen values already available fix the closed form far beyond the number of parameters in it. What is missing is a proof.

References

[1] Y. Cui, J. Luo and C. Xiang, *Walsh Spectrum and Boomerang Properties of Locally-APN Niho Functions*, arXiv:2605.01786v1 (3 May 2026). All numbering cited above — Corollary 1 and Remark 1 — is that of v1, the only version at the time of writing; we found no journal version in Crossref, dblp, OpenAlex or Semantic Scholar.

[2] T. Helleseth, D. J. Katz and C. Li, *The resolution of Niho's last conjecture concerning sequences, codes, and Boolean functions*, IEEE Trans. Inform. Theory **67** (2021), no. 10, 6952–6962, doi:10.1109/TIT.2021.3098342; preprint arXiv:2006.12239.

[3] T. Helleseth and C. Li, *An updated review on crosscorrelation of m-sequences*, arXiv:2407.16072v1 (22 July 2024); announced there as an invited chapter to the Series on Coding Theory and Cryptology, World Scientific. All numbering cited above — Open Problems 7 and 9, Theorems 10, 11, 16 and 17, Lemmas 2 and 3, and the section numbers — is that of v1, the only version at the time of writing; we found no record of the printed chapter in Crossref, dblp, OpenAlex, Semantic Scholar or zbMATH, and so cite the preprint.

[4] D. J. Katz, *Weil sums of binomials: properties, applications and open problems*, in: Combinatorics and Finite Fields: Difference Sets, Polynomials, Pseudorandomness and Applications (K.-U. Schmidt and A. Winterhof, eds.), De Gruyter, Berlin, 2019, 109–134, doi:10.1515/9783110642094-007; preprint arXiv:1805.10452. We cite the numbering of the preprint (v3), whose Corollary 7.4(ii),(iii) states the two moments used here; the book chapter renumbers them. [3, Lemma 3] reproduces that corollary, but keys it to a different paper of Katz. In the numbering of the preprint arXiv:1407.7923v3, the same two moments are Corollary 2.6(i),(ii) of D. J. Katz, *Divisibility of Weil sums of binomials*, Proc. Amer. Math. Soc. **143** (2015), no. 11, 4623–4632, doi:10.1090/proc/12687.

[5] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science 12699, Springer, 2021, 625–635, doi:10.1007/978-3-030-79876-5_37.

[6] Y. Xia, N. Li, X. Zeng and T. Helleseth, *An open problem on the distribution of a Niho-type cross-correlation function*, IEEE Trans. Inform. Theory **62** (2016), no. 12, 7546–7554, doi:10.1109/TIT.2016.2614333.