Designs, Codes and Cryptography
https://doi.org/10.1007/s10623-024-01461-3

Investigation of the permutation and linear codes from the Welch APN function

Tor Helleseth[1] · **Chunlei Li**[1] · **Yongbo Xia**[2]

Received: 25 March 2024 / Revised: 25 June 2024 / Accepted: 9 July 2024

Abstract
Dobbertin in 1999 proved that the Welch power function x^{2^m+3} was almost perferct nonlinear (APN) over the finite field $\mathbb{F}_{2^{2m+1}}$, where m is a positive integer. In his proof, Dobbertin showed that the APNness of x^{2^m+3} essentially relied on the bijectivity of the polynomial $g(x) = x^{2^{m+1}+1} + x^3 + x$ over $\mathbb{F}_{2^{2m+1}}$. In this paper, we first determine the differential and Walsh spectra of the permutation polynomial $g(x)$, revealing its favourable cryptograhphic properties. We then explore four families of binary linear codes related to the Welch APN power functions. For two cyclic codes among them, we propose algebraic decoding algorithms that significantly outperform existing methods in terms of decoding complexity.

Keywords Permutation · Differential spectrum · Walsh spectrum · Linear codes · Cyclic codes · Algebraic decoding

Mathematics Subject Classification 94B05 · 94B35 · 11T06 · 11T71

1 Introduction

Let $\mathbb{F}_{2^n}$ denote the finite field of 2^n elements and $\mathbb{F}_{2^n}^*$ be its multiplicative group. Nonlinear functions over $\mathbb{F}_{2^n}$ have wide applications in cryptography and coding theory. In symmetric cryptography, block ciphers are designed by appropriate compositions of linear permutations and S-boxes that are the only nonlinear component. Hence the cryptographic properties of the nonlinear S-boxes are crucial to the security of the ciphers. Differential and linear attacks [5, 36] are two of the most powerful cryptographic attacks against block ciphers, and the link between these two approaches was investigated in [17]. To ensure good resistance to differential attacks, the differential uniformity of the nonlinear function used in an S-

✉ Yongbo Xia
xia@mail.scuec.edu.cn

Tor Helleseth
tor.helleseth@uib.no

Chunlei Li
chunlei.li@uib.no

[1] Department of Informatics, University of Bergen, 5008 Bergen, Norway

[2] Department of Mathematics and Statistics, South-Central Minzu University, Wuhan 430074, China

box should be low. The lowest possible differential uniformity is 2 and functions with this property are called almost perfect nonlinear (APN) functions. There has been much work and progress on APN functions; see, for example, [11, 13]. The nonlinearity quantifies the level of resistance of the function to the linear attack: the higher is the nonlinearity, the better is the resistance of the function against the linear attack. Besides the differential uniformity and the nonlinearity, there are also some other cryptographic criteria that measure the resistance of the nonlinear functions to various known attacks. For further details about this topic, the reader is referred to [13, 44] and references therein. The study on the cryptographically significant functions during the past decades shows that it is difficult to design a function attaining all good cryptographic criteria, and trade-offs must be considered.

Linear codes, particularly cyclic codes, have wide applications in reliable data storage and communication systems. In coding theory one of the most important topics is to construct linear codes with desirable properties and to explore efficient decoding for them. Constructing linear codes from nonlinear functions was extensively explored in the past decades [14, 22, 26, 32], and many optimal linear codes have been obtained from cryptographically significant functions [15, 24, 27, 28, 38], such as perfect nonlinear functions, almost perfect nonlinear functions, bent functions and plateaued functions. In those works, the minimum distances and weight distributions of the constructed codes and their duals were intensively studied (see for instance a recent survey by Li and Mesnager [32]). There are other parameters of linear codes, such as the covering radius [21] and coset weight distribution [18], that are of fundamental interest, particularly when evaluating the performance of linear codes in error correction. Nevertheless, due to their intractabilities, there has been limited research progress on such topics. It is well known that the problem of random syndrome decoding is NP-complete [4]. There do exist certain linear codes with efficient decoding. For instance, BCH codes, due to their special property, allow for efficient decoding with polynomial-time complexity [2]. However, efficiently decoding non-BCH cyclic codes remains a significant open problem, despite recent efforts to develop decoders for generic cyclic codes by investigating generalized error-locator polynomials [1, 16, 33].

In this paper, we first investigate important cryptographic properties, namely, the differential spectrum and Walsh spectrum, of the permutation polynomial $f(x) = x^{2^{m+1}+1} + x^3 + x$ over $\mathbb{F}_{2^{2m+1}}$, which we call the *Welch permutation* since it was used to prove the APNness of the Welch power function $F(x) = x^{2^m+3}$ [29]. In the second part, we explore two families of cyclic codes and two families of linear codes that are closely related to the Welch power function. For the two binary cyclic codes, we propose efficient algebraic decoders with complexity in the order of $O(N(\log N)^3)$, where $N = 2^{2m+1} - 1$ is the code length. For the second family of binary linear codes, it is shown to have at most five nonzero weights, which provides a partial resolution to the conjecture by Ding [24].

The remainder of this paper is organized as follows. Section 2 recalls basic definitions and auxiliary results. Section 3 determines the differential spectrum and Walsh spectrum of $g(x)$. Section 4 explores the properties and decoding of binary codes derived from the Welch APN power function, and Sect. 5 summarizes our contributions in this work.

2 Preliminaries

2.1 Cryptographic properties of vectorial Boolean functions

For a vectorial Boolean function $F(x)$ from $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^n}$, denote

$$N_F(a,b) = |\{x \in \mathbb{F}_{2^n} \mid F(x+a) + F(x) = b\}|. \tag{1}$$

The differential uniformity of $F(x)$ is defined by

$$\Delta_F = \max\left\{N_F(a,b) \mid a \in \mathbb{F}_{2^n}^*, b \in \mathbb{F}_{2^n}\right\}.$$

Nyberg defined a mapping $F(x)$ to be differentially δ-uniform if $\Delta_F = \delta$ [40]. It is clear that the equation $F(x+a) + F(x) = b$ have solutions in pairs. Thus, $\Delta_F = 2$ is the smallest possible value for the differential uniformity of $F(x)$. A function $F(x)$ is said to be almost perfect nonlinear (APN) if its differential uniformity is equal to 2. Equivalently, a function $F(x)$ is APN if its derivative function $D_aF(x) = F(x+a) + F(x)$, for any $a \in \mathbb{F}_{2^n}^*$, is a two-to-one function over $\mathbb{F}_{2^n}$.

Besides the differential uniformity, the differential spectrum of $F(x)$ is also an important notion for measuring its resistance against variants of differential cryptanalysis [6, 7, 9, 19]. Its definition is given as follows.

Definition 1 Let $F(x)$ be a function from $\mathbb{F}_{2^n}$ to itself and $N_F(a,b)$ be defined as in (1). Denote

$$\omega_i = |\left\{(a,b) \in \mathbb{F}_{2^n}^* \times \mathbb{F}_{2^n} \mid N_F(a,b) = i\right\}|.$$

The differential spectrum of $F(x)$ is defined as the multi-set of $N_F(a,b)$ for all $(a,b) \in \mathbb{F}_{2^n}^* \times \mathbb{F}_{2^n}$, which can be given by

$$\Omega_F = [\omega_0, \omega_1, \ldots, \omega_\delta], \tag{2}$$

where δ is the differential uniformity of $F(x)$.

It is easily seen that $\omega_i = 0$ in differential spectrum if i is odd. Moreover, we have the following properties

$$\sum_{i=0}^{\delta} \omega_i = 2^n(2^n-1) \text{ and } \sum_{i=0}^{\delta} (i \times \omega_i) = 2^n(2^n-1). \tag{3}$$

For any APN function over $\mathbb{F}_{2^n}$, there are only two possible values 0 and 2 in its differential spectrum. Thus, from the equalities in (3), the differential spectrum of an APN function can be uniquely determined.

Another important criterion of a vectorial Boolean function $F(x)$ is its nonlinearity, which can be given in terms of the (extended) Walsh transform of $F(x)$.

Definition 2 The extended Walsh transform of a vectorial Boolean function $F(x)$ at (a,b) is defined by

$$W_F(a,b) = \sum_{x \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(bF(x)+ax)},$$

where $a, b \in \mathbb{F}_{2^n}$. The extended Walsh spectrum of $F(x)$ is the multi-set

$$\Lambda_F = \{W_F(a,b) : a, b \in \mathbb{F}_{2^n}, b \neq 0\}. \tag{4}$$

The nonlinearity of F is given by

$$NL(F) = 2^{n-1} - \frac{1}{2} \max\{|W_F(a,b)| : a, b \in \mathbb{F}_{2^n}, b \neq 0\}.$$

Remark 1 Note that for a Boolean function $G(x)$ from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$, the extended Walsh transform reduces to the original Walsh-Hadamard transform

$$\widehat{G}(\lambda) = \sum_{x \in \mathbb{F}_{2^n}} (-1)^{G(x)+\mathrm{Tr}_1^n(\lambda x)}, \ \lambda \in \mathbb{F}_{2^n}.$$

Next we recall some results about the Walsh transforms of quadratic Boolean functions. Given a quadratic Boolean function $Q(x)$ from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$, the function $B(x,z) = Q(x+z) + Q(x) + Q(z)$ is a bilinear function in x and z. When x, z are expressed as vectors in $\mathbb{F}_2^n$, the bilinear function can be written as $B(x,z) = xBz^{\mathsf{T}}$, where $B = \left(b_{ij}\right)$ is the $n \times n$ symplectic matrix of $Q(x)$ satisfying that all diagonal elements of B are zero and $b_{ij} = 1$ for $1 \leq i, j \leq n$ if and only if the multivariate form of $Q(x)$ contains the term $x_i x_j$ [35]. The rank of $Q(x)$ is defined as the rank of its symplectic matrix B, which is always even. Let

$$V_Q = \{x \in \mathbb{F}_{2^n} \mid Q(x+z) + Q(x) + Q(z) = 0, \forall\ z \in \mathbb{F}_{2^n}\}.$$

By the rank-null theorem we have $\dim_{\mathbb{F}_2}\left(V_Q\right) + \mathrm{Rank}(Q) = n$. Note that

$$\left(\sum_{x \in \mathbb{F}_{2^n}} (-1)^{Q(x)}\right)^2 = \sum_{x \in \mathbb{F}_{2^n}} (-1)^{Q(x)} \sum_{z \in \mathbb{F}_{2^n}} (-1)^{Q(x+z)+Q(x)+Q(z)} = 2^n \sum_{x \in V_Q} (-1)^{Q(x)},$$

where V_Q is the $\mathbb{F}_2$-linear space defined as above. It is readily seen that $Q(x)$ is linear over V_Q. Hence one has

$$\sum_{x \in \mathbb{F}_{2^n}} (-1)^{Q(x)} = \begin{cases} \pm 2^{n - \mathrm{Rank}(Q)/2}, & \text{if } Q(x) = 0 \text{ for any } x \in V_Q, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, when λ runs through $\mathbb{F}_{2^n}$, the distribution of the Walsh transform $\widehat{Q}(x)$ can be given as follows.

Lemma 1 ([31, Theorem 6.2]) *Let $Q(x)$ be a quadratic form on $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$ with rank $2h$. Then its Walsh transform has the following distribution*

$$\widehat{Q}(\lambda) = \sum_{x \in \mathbb{F}_{2^n}} (-1)^{Q(x)+Tr_1^n(\lambda x)} = \begin{cases} \pm 2^{n-h}, & 2^{2h-1} \pm 2^{h-1} \text{ times}, \\ 0, & 2^n - 2^{2h} \text{ times}. \end{cases}$$

2.2 Linear codes from nonlinear functions

In this section we recall basics of linear codes and the two generic constructions for linear codes from nonlinear functions. Below we focus only on binary linear codes while the basics are valid for linear codes over finite fields in general [2, 35].

Basics of linear codes

An $[N, k, d]$ binary linear code $\mathcal{C}$ is a k-dimensional subspace of $\mathbb{F}_2^N$ with minimum (Hamming) weight d. The code $\mathcal{C}$ can be defined either by its generator matrix G as

$\mathcal{C} = \{xG \mid x \in \mathbb{F}_2^k\}$ or by its parity-check matrix H as $\mathcal{C} = \{c \in \mathbb{F}_2^n \mid cH^T = 0\}$. The *dual code* $\mathcal{C}^\perp$ is given by $\mathcal{C}^\perp = \left\{x \in \mathbb{F}_2^N \mid x_1c_1 + \cdots + x_Nc_N = 0,\ \forall (c_1, \ldots, c_N) \in \mathcal{C}\right\}$, which has the parity-check matrix H of $\mathcal{C}$ as its generator matrix. For a received vector $y = c+e$ with certain codeword $c \in \mathcal{C}$ and error vector $e \in \mathbb{F}_2^N$, the syndrome equation $s = yH^T = eH^T$ associates the error e with a coset of $\mathcal{C}$ in $\mathbb{F}_2^N$. The coset leader for each coset is defined as the element with minimum weight in the coset. A binary linear code $\mathcal{C}$ is called *cyclic* if for any $c = (c_1, \ldots, c_N) \in \mathcal{C}$, its cyclic shift $\sigma(c) = (c_N, c_1, \ldots, c_{N-1})$ is contained in $\mathcal{C}$. An $[N, k]$ binary cyclic code $\mathcal{C}$ can be equivalently seen as an ideal in $\mathbb{F}_2^N[x]/(x^N - 1)$. In this way, a binary cyclic code $\mathcal{C}$ can be uniquely defined by a binary monic polynomial $g(x)$ dividing $x^N - 1$, known as the *generator polynomial* of $\mathcal{C}$. Equivalently, the code $\mathcal{C} = \langle g(x) \rangle$, can be uniquely given by its *complete defining set* $S_\mathcal{C} = \{i : g(\alpha^i) = 0, 0 \le i < N\}$, where α is an N-th primitive root of unity. Since $g(\alpha^i) = 0$ iff $g(\alpha^{2i}) = 0$ for any $0 \le i < N$, the set $S_\mathcal{C}$ is usually partitioned into disjoint cyclotomic cosets modulo N. A subset of $S_\mathcal{C}$ that consists of coset leaders from each coset in $S_\mathcal{C}$ can uniquely define $\mathcal{C}$, and is therefore termed as the *primary defining set* of $\mathcal{C}$. When the (complete) defining set $S_\mathcal{C}$ contains $d - 1$ consecutive integers, the cyclic code $\mathcal{C}$ has minimum distance at least d according to the BCH bound [2].

Let $\mathcal{C}$ be a binary linear code of length N and minimum weight d. The space $\mathbb{F}_2^N$ can be then partitioned into cosets with respect to $\mathcal{C}$. For each coset, the coset leader is defined as one element with minimum weight in the coset. When the minimum weight of a coset is no greater than $\lfloor \frac{d-1}{2} \rfloor$, it has a unique coset leader; when its minimum weight is larger, a coset may have several elements with the minimum weight, indicating that the coset leader is not unique. For a received vector $y = c+e$ with certain codeword $c \in \mathcal{C}$ and error vector $e \in \mathbb{F}_2^N$, the syndrome $s = yH^\intercal = eH^\intercal$ associates the error e with a coset of $\mathcal{C}$ in $\mathbb{F}_2^N$. In particular, for the case of $s = 0$, it corresponds to the code $\mathcal{C}$, of which the coset leader is the zero vector. This indicates that when a codeword c is transmitted and the received vector $y = c + e$ is another codeword of $\mathcal{C}$, the process of error detection by the parity-check equation $s = yH^\intercal$ fails. The probability of the detection failure of the code $\mathcal{C}$ can be expressed in terms of its weight distribution, which is defined as $(A_0, A_1, \ldots, A_N)$, where A_i denote the number of codewords with Hamming weight i in the code $\mathcal{C}$ and it is obvious that $A_0 = 1$. Thanks to the MacWilliams identity, the weight distribution of $\mathcal{C}$ can be derived from the weight distribution $(1, B_1, \ldots, B_N)$ of its dual $\mathcal{C}^\perp$.

For a nonzero syndrome $s = yH^\intercal = eH^\intercal$, it belongs to a coset with a nonzero coset leader. The corresponding coset leader has the same syndrome as e, and it will be deemed as the error e added to the received vector y, since the coset leader has the minimum weight. The process can uniquely correct the error e when its weight is within the *packing radius* of $\mathcal{C}$ given by $t = \lfloor \frac{d(\mathcal{C})-1}{2} \rfloor$, for which the coset leader is unique; when an error e has weight beyond the packing radius t, it is likely that the corresponding coset doesn't have unique coset leader anymore. In this case, the error e cannot be uniquely decoded and the decoder may fail to return a correct codeword. The performance of the aforementioned error correction procedure can be evaluated in terms of weight distributions of cosets [2]. Unfortunately, a complete picture of weight distributions of all cosets is intractable. Instead, some attempts have been made in calculating the coset distribution $(1, K_1, K_2, \ldots, K_N)$, where K_i denotes the number of coset leaders with weight i, of the linear code $\mathcal{C}$ [18].

The largest weight of coset leaders of $\mathcal{C}$ is known as the *covering radius* of $\mathcal{C}$, which is defined by $\rho(\mathcal{C}) = \max\{\min\{d(y, c) : c \in \mathcal{C}\} : y \in \mathbb{F}_2^N\}$. The covering radius of $\mathcal{C}$ is a basic geometric parameter, which is a measure of the maximum distortion when $\mathcal{C}$ is used for data compression, and is the maximum weight of a correctable random error when $\mathcal{C}$ is used for error correction [21]. It is clear that the covering radius of a code is lower bounded by its packing radius t. The equality of such an inequality is achieved by *perfect codes*. In

addition, a linear code $\mathcal{C}$ is called a *quasi-perfect* if $\rho(\mathcal{C}) = t + 1$; and a quasi-perfect code is called *uniformly packed code* if $\rho(\mathcal{C})$ is the same as the external distance of $\mathcal{C}$, which is the number of non-zero weights in its dual $\mathcal{C}^{\perp}$.

Generic construction 1

Let F be a function from $\mathbb{F}_{2^n}$ to itself with $F(0) = 0$, and β be a primitive element of $\mathbb{F}_{2^n}$. A binary linear code $\mathcal{C}$ of length $2^n - 1$ can be constructed from F via the following parity-check matrix

$$H = \begin{bmatrix} 1 & \beta & \beta^2 & \dots & \beta^{2^n-2} \\ F(1) & F(\beta) & F(\beta^2) & \dots & F(\beta^{2^n-2}) \end{bmatrix}, \tag{5}$$

where each symbol stands for the column of its coordinate with respect to a basis of $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$. It is easy to verify that the dual code $\mathcal{C}^{\perp}$ is given by

$$\mathcal{C}^{\perp} = \left\{ \left(\mathrm{Tr}_1^n(ax + bF(x)) \right)_{x \in \mathbb{F}_{2^n}^*} : a, b \in \mathbb{F}_{2^n} \right\}.$$

For the nonlinear function F, the code $\mathcal{C}$ has dimension $2^n - 1 - 2n$. In particular, when $F(x)$ is a power function x^d, the code $\mathcal{C}$ is a cyclic code with primary defining set $\{1, d\}$. This generic construction has a long history and pertains to Delsarte's Theorem [22]. Note that for the dual code $\mathcal{C}^{\perp}$, the Hamming weight of a codeword $c_{a,b} \in \mathcal{C}^{\perp}$ is given by

$$\begin{aligned} \mathrm{wt}(c_{a,b}) &= 2^n - 1 - \#\left\{ x \in \mathbb{F}_{2^n}^* : \mathrm{Tr}_1^n(ax + bF(x)) = 0 \right\} \\ &= 2^{n-1} - \frac{1}{2} \sum_{x \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(ax+bF(x))} = 2^{n-1} - \frac{1}{2} W_F(a, b). \end{aligned} \tag{6}$$

Therefore, the weight distribution of $\mathcal{C}^{\perp}$ can be directly derived from the extended Walsh spectrum of $F(x)$ given by $\{W_F(a, b) : a, b \in \mathbb{F}_{2^n}\}$. This relation has led to a well-established coding-theory characterization of APN functions, almost bent (AB) functions [14].

Theorem 1 ([14]) *Let F be a function from $\mathbb{F}_{2^n}$ to itself with $F(0) = 0$ and n being odd. Let the code $\mathcal{C}$ be defined by a parity-check matrix H as in (5). Then $F(x)$ is an APN function if and only if the code $\mathcal{C}$ has minimum distance 5. Furthermore, $F(x)$ is an AB function if and only $\mathcal{C}^{\perp}$ is a $[2^n - 1, 2^n - 1 - 2n]$ uniformly packed code with minimum distance 5 and packing radius 3.*

Generic construction 2

Let $D = \{d_1, d_2, \dots, d_\ell\}$ be a subset of $\mathbb{F}_{2^n}$. A binary linear code having D as its defining set is given by

$$\mathcal{C}_D = \left\{ \mathbf{c}_a = \left(\mathrm{Tr}_1^n(ad_1), \mathrm{Tr}_1^n(ad_2), \dots, \mathrm{Tr}_1^n(ad_\ell) \right) : a \in \mathbb{F}_{2^n} \right\}.$$

It is clear that the code $\mathcal{C}_D$ has length ℓ and dimension at most n.

When the defining set D is properly chosen, the code $\mathcal{C}_D$ can have good or optimal parameters. The above construction is generic in the sense that all linear codes could be produced by selecting proper defining sets D. By considering defining sets D as the support or image of certain functions F over $\mathbb{F}_2^n$, researchers have proposed many families of few-weight linear codes with new code lengths, see e.g., [23–26, 37, 45]. Interested readers may refer to a recent survey by Li and Mesnager in [32] and references therein for more details about these two generic approaches.

3 Differential and Walsh spectra of the Welch permutation

For the Welch permutation $g(x) = x^{2^{m+1}+1} + x^3 + x$ over $\mathbb{F}_{2^n}$ with $n = 2m+1$, this section will determine its differential spectrum Ω_g as defined in (2) and its Walsh spectrum Λ_g as defined in (4).

Theorem 2 *Let $n = 2m+1$ and $g(x) = x^{2^{m+1}+1} + x^3 + x$. Then the function $g(x)$ over $\mathbb{F}_{2^n}$ is differentially 4-uniform. Furthermore, its differential spectrum is given by*

$$\left[\omega_0 = 2^{2n-1} + 2^{2n-3} - 3 \cdot 2^{n-2}, \omega_2 = 2^{2n-2}, \omega_4 = 2^{2n-3} - 2^{n-2}\right].$$

Proof For $(a, b) \in \mathbb{F}_{2^n}^* \times \mathbb{F}_{2^n}$, let $N(a, b)$ be the number of solutions of the derivative equation $g(x+a) + g(x) = b$ in $\mathbb{F}_{2^n}$. Note that

$$\begin{aligned} & g(x+a) + g(x) + b \\ &= x^{2^{m+1}}a + xa^{2^{m+1}} + a^{2^{m+1}+1} + x^2a + xa^2 + a^3 + a + b \\ &= ax^{2^{m+1}} + ax^2 + (a^{2^{m+1}} + a^2)x + g(a) + b. \end{aligned}$$

Since $a \neq 0$, $g(x+a) + g(x) + b = 0$ is equivalent to that

$$x^{2^{m+1}} + x^2 + cx + d = 0, \tag{7}$$

where

$$c = a^{2^{m+1}-1} + a \text{ and } d = \frac{g(a) + b}{a}. \tag{8}$$

Note that $c = 0$ if and only if $a = 1$. Next we consider the following linearized equation

$$x^{2^{m+1}} + x^2 + cx = 0. \tag{9}$$

If $c = 0$, i.e., $a = 1$, then (9) have two solutions in $\mathbb{F}_{2^n}$, which are 0 and 1. If $c \neq 0$, i.e., $a \notin \mathbb{F}_2$, then by raising (9) to the power 2^m, we get

$$x + x^{2^{m+1}} + c^{2^m}x^{2^m} = 0. \tag{10}$$

Adding up (9) and (10), we get

$$c^{2^m}x^{2^m} + x^2 + (c+1)x = 0,$$

which implies

$$x^{2^m} = \frac{x^2}{c^{2^m}} + \frac{c+1}{c^{2^m}}x. \tag{11}$$

Substituting (11) into (10), we get

$$x^4 + (c^{2^{m+1}} + c^2 + 1)x^2 + c^{2^{m+1}+1}x = 0. \tag{12}$$

The above arguments show that when $c \neq 0$, the solutions of (9) must be those of (12). Note that the left hand side of (12) is a linearized polynomial over $\mathbb{F}_{2^n}$ with degree 4 and it may have 1, 2 or 4 roots in $\mathbb{F}_{2^n}$. Thus, the equation (9) also may have 1, 2 or 4 solutions in $\mathbb{F}_{2^n}$. Moreover, note that

$$c = a^{2^{m+1}-1} + a = \frac{a^{2^{m+1}} + a^2}{a}.$$

Besides $x = 0$, for any given $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$, it can be observed that $x = a$ must be a solution of (9). Thus, when $c \neq 0$, i.e., $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$, the number of solutions of (9) in $\mathbb{F}_{2^n}$ is 2 or 4.

Denote the number of $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$ such that (9) has two (resp. four) solutions in $\mathbb{F}_{2^n}$ by M_1 (resp. M_2). In what follows, we need to determine M_1 and M_2. We further investigate the equation (12). Since $x = 0$ and $x = a$ are its solutions, the polynomial on the left hand side of (12) has a factorization over $\mathbb{F}_{2^n}$ as follows

$$x^4 + (c^{2^{m+1}} + c^2 + 1)x^2 + c^{2^{m+1}+1}x = x(x + a)\left(x^2 + ax + \frac{c^{2^{m+1}+1}+1}{a}\right),$$

where $c = \frac{a^{2^{m+1}}+a^2}{a}$. By verifying that $a^2 + \frac{c^{2^{m+1}+1}+1}{a} = c^{2^{m+1}} + c^2 + 1$, we can check the validity of the above factorization. To determine the exact number of solutions to (9), we should investigate the solutions of the following quadratic equation

$$x^2 + ax + \frac{c^{2^{m+1}+1}+1}{a} = 0. \tag{13}$$

Note that

$$\begin{aligned}
&\mathrm{Tr}_1^n\left(\frac{c^{2^{m+1}+1}+1}{a^3}\right)\\
&= \mathrm{Tr}_1^n\left(\frac{a^2 + a^{2^{m+2}}}{a^{2^{m+1}}} \cdot \frac{a^{2^{m+1}} + a^2}{a^4}\right)\\
&= \mathrm{Tr}_1^n\left(\frac{a^4 + a^2 \cdot a^{2^{m+1}} + a^{2^{m+2}} \cdot a^{2^{m+1}} + a^2 \cdot a^{2^{m+2}}}{a^{2^{m+1}} \cdot a^4}\right)\\
&= \mathrm{Tr}_1^n\left(\frac{1}{a^{2^{m+1}}} + \frac{1}{a^2} + \frac{a^{2^{m+2}}}{a^4} + \frac{a^{2^{m+1}}}{a^2}\right)\\
&= \mathrm{Tr}_1^n\left(\frac{1}{a}\right) + \mathrm{Tr}_1^n\left(\frac{1}{a}\right) + \mathrm{Tr}_1^n\left(\frac{a^{2^{m+1}}}{a^2}\right) + \mathrm{Tr}_1^n\left(\frac{a^{2^{m+1}}}{a^2}\right)\\
&= 0.
\end{aligned}$$

Thus, (13) has two solutions in $\mathbb{F}_{2^n}$. This also shows that for any $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$, (12) always has four solutions in $\mathbb{F}_{2^n}$. By Theorem 1 in [20], one can get the solutions of (13), which can be represented as

$$x_1 = a\sum_{i=1}^{m}\left(\frac{c^{2^{m+1}+1}+1}{a^3}\right)^{2^{2i-1}}, \text{ and } x_2 = x_1 + a.$$

Note that $x_i \neq 0$ and $x_i \neq a$, $i = 1, 2$. Otherwise, by (13), it leads to $c = 0$, contradicting the fact that $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$. Next we should verify that whether x_1 is a solution of (9) or not. If x_1 is a solution of (9), so does x_2.

Let $y = \frac{x_1}{a}$, then (13) becomes into

$$y^2 + y + \frac{c^{2^{m+1}+1}+1}{a^3} = 0. \tag{14}$$

If x_1 is a solution of (9), we have

$$y^{2^{m+1}} + \frac{a^2}{a^{2^{m+1}}} y^2 + \frac{ca}{a^{2^{m+1}}} y = 0. \tag{15}$$

Substituting (14) into (15), we get

$$y^{2^{m+1}} + y + \left(\frac{c}{a}\right)^{2^{m+1}+1} = 0. \tag{16}$$

On the other hand, we can compute $y^{2^{m+1}} + y$ from (14) by

$$y^{2^{m+1}} + y = \sum_{i=0}^{m} \left(y^2 + y\right)^{2^i} = \sum_{i=0}^{m} \left(\frac{c^{2^{m+1}+1}}{a^3}\right)^{2^i}.$$

The computation details are given as follows:

$$\begin{aligned}
&y^{2^{m+1}} + y \\
&= \sum_{i=0}^{m} \left(\frac{c^{2^{m+1}+1}}{a^3}\right)^{2^i} \\
&= \sum_{i=0}^{m} \left(\frac{1}{a^{2^{m+1}}} + \frac{1}{a^2} + \frac{a^{2^{m+2}}}{a^4} + \frac{a^{2^{m+1}}}{a^2}\right)^{2^i} \\
&= \sum_{i=0}^{m} \left(\left(\frac{1}{a^2}\right)^{2^m} + \frac{1}{a^2} + \left(\frac{a^{2^{m+1}}}{a^2}\right)^2 + \frac{a^{2^{m+1}}}{a^2}\right)^{2^i} \\
&= \mathrm{Tr}_1^n\left(\frac{1}{a^2}\right) + \frac{1}{a^{2^{m+1}}} + \frac{a^{2^{m+1}}}{a^2} + \left(\frac{a^{2^{m+1}}}{a^2}\right)^{2^{m+1}} \\
&= \mathrm{Tr}_1^n\left(\frac{1}{a^2}\right) + \frac{1}{a^{2^{m+1}}} + \frac{a^{2^{m+1}}}{a^2} + \frac{a^2}{a^{2^{m+2}}} \\
&= \mathrm{Tr}_1^n\left(\frac{1}{a^2}\right) + 1 + \left(\frac{a^{2^{m+1}+a^2}}{a^2}\right)^{2^{m+1}} \cdot \frac{a^{2^{m+1}+a^2}}{a^2} \\
&= \mathrm{Tr}_1^n\left(\frac{1}{a^2}\right) + 1 + \left(\frac{c}{a}\right)^{2^{m+1}+1}.
\end{aligned} \tag{17}$$

By (17) and (16), we can conclude that for each $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$, the solution x_1 of (13) is also a solution of (9) if and only if $\mathrm{Tr}_1^n\left(\frac{1}{a^2}\right) = \mathrm{Tr}_1^n\left(\frac{1}{a}\right) = 1$. This implies that for each $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$, (9) has two (resp. four) solutions in $\mathbb{F}_{2^n}$ if and only if $\mathrm{Tr}_1^n\left(\frac{1}{a}\right) = 0$ (resp. $\mathrm{Tr}_1^n\left(\frac{1}{a}\right) = 1$). It is obvious that the number of $a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2$ such that $\mathrm{Tr}_1^n\left(\frac{1}{a}\right) = 0$ (resp. $\mathrm{Tr}_1^n\left(\frac{1}{a}\right) = 1$) is equal to $2^{n-1} - 1$. Thus, we obtain that $M_1 = M_2 = 2^{n-1} - 1$.

For each given $a \in \mathbb{F}_{2^n}^*$, denote the linearized polynomial on the left hand side of (9) by $L_a(x)$. Then, $L_a(x)$ is a linear transformation from the vector space $\mathbb{F}_{2^n}$ to itself. Let $A_i = \{a \in \mathbb{F}_{2^n} \setminus \mathbb{F}_2 \mid \mathrm{Tr}_1^n(\frac{1}{a}) = i\}$, where $i = 0, 1$. Then $\mathbb{F}_{2^n}^* = \{1\} \cup A_0 \cup A_1$. The above arguments have shown that the kernel of the linear transformation $L_a(x)$, denoted by $\ker L_a$, contains two elements of $\mathbb{F}_{2^n}$ if $a \in \{1\} \cup A_0$ and four elements if $a \in A_1$. Note that the

linear transformation $L_a(x)$ can also be regarded as a homomorphism from the additive group of $\mathbb{F}_{2^n}$ to itself. Thus, by the homomorphism theorem, the image of $L_a(x)$ has cardinality $\frac{2^n}{|\ker L_a|} = 2^{n-1}$ if $a \in \{1\} \cup A_0$ and has cardinality 2^{n-2} if $a \in A_1$. Moreover, for each element d in the image of $L_a(x)$, there exist exactly $|\ker L_a|$ elements x's in $\mathbb{F}_{2^n}$ such that $L_a(x) = d$.

For convenience, let B_a denote the image of the linear transformation $L_a(x) = x^{2^{m+1}} + x^2 + cx$, where $a \in \mathbb{F}_{2^n}^*$. We have obtained that $|B_a| = 2^{n-1}$ if $a \in \{1\} \cup A_0$ and $|B_a| = 2^{n-2}$ if $a \in A_1$. By (8), for a given element $a \in \mathbb{F}_{2^n}^*$, the correspondence between d and b is one-to-one. Recall that $N(a, b)$ denotes the number of solutions of (7) in $\mathbb{F}_{2^n}$. Thus, we can conclude that for each $a \in \{1\} \cup A_0$ (resp. $a \in A_1$), $N(a, b) = 2$ (resp. 4) iff $b \in aB_a + g(a) = \{ad + g(a) \mid d \in B_a\}$. In other cases, we all have $N(a, b) = 0$. Thus, the number of pairs $(a, b) \in \mathbb{F}_{2^n}^* \times \mathbb{F}_{2^n}$ such that $N(a, b) = 2$ (resp. 4) is equal to $2^{n-1} \cdot 2^{n-1}$ (resp. $\left(2^{n-1} - 1\right) \cdot 2^{n-2}$). This together with (3) gives the differential spectrum of $g(x)$. □

Note that $\mathrm{Tr}_1^n(ag(x)) = \mathrm{Tr}_1^n\left(a(x^{2^{m+1}+1} + x^3 + x)\right)$ is a quadratic Boolean function from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$. According to Lemma 1, the Walsh transform of $\mathrm{Tr}_1^n(ag(x))$ heavily depends on its rank. Below is an auxiliary result for the rank of $\mathrm{Tr}_1^n(ag(x))$.

Lemma 2 *Let s, n, l be positive integers satisfying $\gcd(s, n) = 1$ and let*

$$Q(x) = \sum_{i=1}^{l} Tr_1^n(c_i x^{2^{si}+1}),$$

where $c_i \in \mathbb{F}_{2^n}$ and at least one c_i is nonzero for $1 \le i \le l$. Then, the rank $2h$ of $Q(x)$ is in the range $n - 2l \le 2h \le n$.

Proof We consider the following equation

$$\begin{aligned}
&Q(x) + Q(z) + Q(x + z) \\
&= \mathrm{Tr}_1^n\left(\sum_{i=1}^{l}\left(c_i x^{2^{si}} z + c_i x z^{2^{si}}\right)\right) \\
&= \mathrm{Tr}_1^n\left(\sum_{i=1}^{l}\left(c_i x^{2^{si}} z + c_i^{2^{-is}} x^{2^{-is}} z\right)\right) \\
&= \mathrm{Tr}_1^n\left(z\sum_{i=1}^{l}\left(c_i x^{2^{si}} + c_i^{2^{-is}} x^{2^{-is}}\right)\right) \\
&= 0
\end{aligned}$$

for all $z \in \mathbb{F}_{2^n}$. The above equation holds if and only if

$$\sum_{i=1}^{l}\left(c_i x^{2^{si}} + c_i^{2^{-is}} x^{2^{-is}}\right) = 0,$$

Table 1 Walsh spectrum of $g(x)$

Value	Frequency
0	$9 \cdot 2^{2n-4} + 3 \cdot 2^{n-3} - 1$
$\pm 2^{m+1}$	$\frac{(5\cdot 2^{n-1}-2)}{3}\left(2^{n-2} \pm 2^{\frac{n-3}{2}}\right)$
$\pm 2^{m+2}$	$\frac{(2^{n-1}-1)}{3}\left(2^{n-4} \pm 2^{\frac{n-5}{2}}\right)$

which is equivalent to

$$\begin{aligned}\sum_{i=1}^{l}\left(c_i x^{2^{si}} + c_i^{2^{-is}} x^{2^{-is}}\right)^{2^{ls}} &= \sum_{i=1}^{l}\left(c_i^{2^{ls}} x^{2^{s(l+i)}} + c_i^{2^{s(l-i)}} x^{2^{s(l-i)}}\right)\\ &= \sum_{i=l+1}^{2l} c_{i-l}^{2^{ls}} x^{2^{si}} + \sum_{j=0}^{l-1} c_{l-j}^{2^{sj}} x^{2^{sj}}\\ &= 0.\end{aligned} \tag{18}$$

We can rewrite (18) in the following form

$$\sum_{i=0}^{2l} a_i x^{2^{si}} = 0, \tag{19}$$

where $a_i = c_{l-i}^{2^{si}}$ for $i = 0, 1, \ldots, l-1, a_l = 0$ and $a_i = c_{i-l}^{2^{ls}}$ for $i = l+1, l+2, \ldots, 2l$. Since $\gcd(s, n) = 1$, according to [10, Corollary 1], the equation (19) has at most 2^{2l} solutions in $\mathbb{F}_{2^n}$. The desired result then follows. □

With Theorem 2 and Lemma 2, we are ready to prove the following theorem.

Theorem 3 *Let $n = 2m + 1$ and $g(x) = x^{2^{m+1}+1} + x^3 + x$ be the Welch permutation of $\mathbb{F}_{2^n}$. Then the extended Walsh spectrum of $g(x)$ is given in Table 1.*

Proof It is easily seen that

$$W_g(0, b) = \sum_{x\in\mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(bx)} = \begin{cases} 2^n, & \text{if } b = 0,\\ 0, & \text{if } b \neq 0.\end{cases}$$

When $a \neq 0$,

$$W_g(a, b) = \sum_{x\in\mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n\left(ax^{2^{m+1}+1}+ax^3+(a+b)x\right)}.$$

Denote $\mathrm{Tr}_1^n\left(ax^{2^{m+1}+1} + ax^3\right)$ by $Q_a(x)$, which is a quadratic Boolean function from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$. Note that

$$Q_a(x) = \mathrm{Tr}_1^n\left(ax^{2^{m+1}+1} + ax^3\right) = \mathrm{Tr}_1^n\left(a^{2^m} x^{2^m+1} + a^{2^{2m}} x^{2^{2m}+1}\right).$$

Then, by Lemma 2 and taking $s = m$ and $l = 2$, we can conclude that the rank of $Q_a(x)$ is $n-3$ or $n-1$. When a runs through $\mathbb{F}_{2^n}^*$, assume that the number of $a \in \mathbb{F}_{2^n}^*$ such that $Q_a(x)$ has rank $n-(2i-1)$ is N_i, $i = 1, 2$. Then, by Lemma 1, when (a, b) runs through $\mathbb{F}_{2^n} \times \mathbb{F}_{2^n}$, the extended Walsh transform $W_g(a, b)$ of $g(x)$ has the following distribution

$$W_g(a,b) = \begin{cases} 0, & (2^n-1) + N_1(2^n - 2^{n-1}) + N_2(2^n - 2^{n-3}) \text{ times}, \\ \pm 2^{m+1}, & N_1(2^{n-2} \pm 2^{\frac{n-3}{2}}) \text{ times}, \\ \pm 2^{m+2}, & N_2(2^{n-4} \pm 2^{\frac{n-5}{2}}) \text{ times}. \end{cases}$$

Next we calculate the fourth power sum of $W_g(a, b)$. On one hand, we have

$$\sum_{a,b \in \mathbb{F}_{2^n}} \big(W_g(a,b)\big)^4 = 2^{4n} + 2^{4m+4} \cdot 2^{n-1} \cdot N_1 + 2^{4m+8} \cdot 2^{n-3} \cdot N_2. \tag{20}$$

On the other hand, we have

$$\begin{aligned} &\sum_{a,b \in \mathbb{F}_{2^n}} \big(W_g(a,b)\big)^4 \\ &= \sum_{x,y,u,v \in \mathbb{F}_{2^n}} \sum_{b \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(b(x+y+u+v))} \sum_{a \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(a(g(x)+g(y)+g(u)+g(v)))} \\ &= 2^{2n} T, \end{aligned} \tag{21}$$

where T denotes the number of $(x, y, u, v) \in (\mathbb{F}_{2^n})^4$ satisfying

$$\begin{cases} x + y + u + v = 0, \\ g(x) + g(y) + g(u) + g(v) = 0. \end{cases}$$

Let $N(a, b)$ be the number of solutions of $g(x + a) + g(x) = b$ in $\mathbb{F}_{2^n}$. Then, we have $T = \sum_{a,b \in \mathbb{F}_{2^n}} N(a,b)^2$. Using the notation and results in Theorem 2, we have

$$T = \sum_{a,b \in \mathbb{F}_{2^n}} N(a,b)^2 = 2^{2n} + 4\omega_2 + 16\omega_4 = 4 \cdot \left(2^{2n} - 2^n\right). \tag{22}$$

Combining (20), (21), (22) and the fact that $N_1 + N_2 = 2^n - 1$, we obtain N_1 and N_2. Thus, the value distribution of the extended Walsh transform of $g(x)$ can be derived as in Table 1. □

According to Theorem 3 and Definition 2, we get the following corollary.

Corollary 1 *Let $n = 2m + 1$ and $g(x) = x^{2^{m+1}+1} + x^3 + x$ be the Welch permutation over $\mathbb{F}_{2^n}$. Then, the nonlinearity $nl(g(x))$ of $g(x)$ is equal to $2^{n-1} - 2^{m+1}$.*

4 Binary codes related to the Welch APN function

4.1 Binary cyclic codes related to the Welch APN function

In this subsection we will discuss the properties of two families of binary cyclic codes, which are closely related to the Welch APN power function, and then present algebraic decoding for them.

Recall that $n = 2m + 1$ and β is a primitive element of $\mathbb{F}_{2^n}$. We start from a family of binary cyclic codes $\mathcal{C}_1$ with primary defining set $S_{\mathcal{C}_1} = \{1, d\}$, where $d = 2^m + 3$ is the Welch exponent. That is to say, the matrix

$$H = \begin{bmatrix} 1 & \beta & \beta^2 & \cdots & \beta^{2^n-2} \\ 1 & \beta^d & \beta^{2d} & \cdots & \beta^{(2^n-2)d} \end{bmatrix} \tag{23}$$

is a parity-check matrix of $\mathcal{C}_1$. Note that the Walsh spectrum of x^d was obtained by Canteaut et al. in [12]. Based on their result, it was known that x^d is an AB function. Therefore, it follows from Theorem 1 that $\mathcal{C}_1$ is a $[2^n - 1, 2^n - 1 - 2n]$ double-error-correcting uniformly packed code with packing radius 3. Below we discuss the algebraic decoding of this uniformly packed code.

Notice that for a cyclic code $\mathcal{C}$ with length N and a BCH bound $2t$, when an error has weight t, one can decode $\mathcal{C}$ by the well-known BCH decoder [2]:

- calculate the syndrome $s_j = y(\alpha^j) = \sum_{i=0}^{N-1} y_i \alpha^{ij}$ for $1 \le j \le 2t$ from the received vector y, where α is an N-th primitive root of unity;
- determine the *error-locator polynomial*

$$\sigma(x) = (1 - \alpha^{i_1} x) \cdots (1 - \alpha^{i_t} x) = 1 + \sigma_1 x + \cdots + \sigma_t x^t,$$

  where $i_1, \ldots, i_t$ are the t locations of the error, from the *key equation*

$$s_{i+t} + \sigma_1 s_{i+t-1} + \cdots + \sigma_t s_i = 0 \quad \text{for } 1 \le i \le t.$$

  by the Berlekamp–Massey algorithm;
- use the Chien algorithm to search roots $\alpha^{-i_1}, \ldots, \alpha^{-i_t}$ of $\sigma(x)$, thereby determining $i_1, \ldots, i_t$;
- use the Forney algorithm to determine the error values $e_{i_1}, \ldots, e_{i_t}$ (which is only needed for nonbinary codes).

For the code $\mathcal{C}_1$ defined by H in (23), although it has minimum distance 5, we cannot apply BCH decoder when there are double errors in the received vector. Under such a circumstance, one can consider directly the following system of syndrome equations

$$\begin{cases} x_1 + x_2 = s_1, \\ x_1^d + x_2^d = s_2, \end{cases} \tag{24}$$

where $x_t = \beta^{i_t}$ for $t = 1, 2$ and β is a primitive element in $\mathbb{F}_{2^n}$. The task is to efficiently find x_1, x_2 for a given syndrome $s = (s_1, s_2) = yH^{\mathsf{T}}$.

Letting $y_t = x_t / s_1$ for $t = 1, 2$, Eq. (24) is equivalent to $y_1 = y_2 + 1$ and $y_1^d + y_2^d = \frac{s_2}{s_1^d}$. That is to say, it suffices to focus on finding the solution to the equation $(y + 1)^d + y^d = b$, where $b = \frac{s_2}{s_1^d}$. Dobbertin [29] showed that for the Welch exponent $d = 2^m + 3$, the derivative equation of x^d can be written as

$$(x + 1)^d + x^d = \left(x + x^{2^m}\right)\left(x^2 + x + 1\right) + 1 = g\left(x + x^{2^m}\right) + 1,$$

where $g(x) = x^{2^{m+1}+1} + x^3 + x$ is the corresponding Welch permutation. Let $z = x + x^{2^m}$. The task of correcting double errors for $\mathcal{C}_1$ therefore can be rearranged as follows:

Step 1: solve the equation $g(z) = c = 1 + s_2/s_1^d$;
Step 2: solve the equation $y + y^{2^m} = \eta$, where η is the solution obtained in Step 1;
Step 3: determine error positions i_1, i_2 from $x_t = s_1 y_t$ for $t = 1, 2$.

For the first step, one can find the preimage of c with the help of the compositional inverse $g^{-1}(x)$ of the permutation $g(x)$. Nevertheless, we don't have an explicit expression of the compositional inverse $g^{-1}(x)$ yet. A straightforward way is to exhaust possible $z \in \mathbb{F}_{2^n}$ for the equation $g(z) + c = 0$. For each evaluation $g(z)$, the Chien search method can reduce the computational complexity from $O(t^2)$ to $O(t)$. The optimization in this part is negligible for $t = 2$. Another way is to calculate $\gcd(z^{2^n-1} - 1, g(z) + c)$ over the polynomial ring $\mathbb{F}_{2^n}[x]$, which gives a linear term $z + z_0$. This method can be further optimized based on the form of $g(z)$. As observed in [29], the equation $g(z) = c$ for $c \neq 0$ can be rewritten as

$$z^{2^{m+1}} = z^2 + 1 + \frac{c}{z}.$$

Raising this equation to the power of 2^{m+1} gives

$$z^2 = z^{2^{m+1}+1} + 1 + \frac{c^{2^{m+1}}}{z^{2^{m+1}}} = \left(z^2 + 1 + \frac{c}{z}\right)^2 + 1 + \frac{c^{2^{m+1}}}{z^2 + 1 + \frac{c}{z}}.$$

Rearranging the above equation gives

$$g_0(z) = z^9 + cz^6 + z^5 + cz^4 + (c^{2^m} + c)^2 z^3 + c^2 z + c^3.$$

Dobbertin showed that $g_0(z)$ can only have one solution in $\mathbb{F}_{2^n}$. Hence, an alternative way to solve $g(z) = c$ is to calculate $\gcd(g(z) + c, g_0(z))$. To compare this calculation with the typical root searching and the calculation of $\gcd(z^{2^n-1} - 1, g(z) + c)$, we recall the result from [30].

Theorem 4 ([30, Theorem 5.4]) *Let $\mathbb{F}_q$ be the finite field of q elements and $\mathbb{F}_q[x]_t$ be the polynomials in $\mathbb{F}_q[x]$ of degree t. Let e, d be positive integers such that $q > d(2e - d + 1)/2$ and $e > d$. Let $t_g^{\mathrm{div}}, t_g^{\div}, t_g^{-,\times}$ be the polynomial divisions, divisions, addition/multiplications in $\mathbb{F}_q$. Given a polynomial $g \in \mathbb{F}_q[x]_e$, the average number $\mathbb{E}\left[t_g^{\mathrm{w}}\right]$ of operations $\mathrm{w} \in \{\mathrm{div}, \div, -, \times\}$ performed on (uniform distributed) inputs from $\mathbb{F}_q[x]_d$ is bounded in the following way:*

$$\left|\frac{\mathbb{E}\left[t_g^{\mathrm{div}}\right]}{d+1} - 1\right| \leq \frac{de}{q}, \quad \left|\frac{\mathbb{E}\left[t_g^{\div}\right]}{e+d+1} - 1\right| \leq \frac{de}{q}, \quad \left|\frac{\mathbb{E}\left[t_g^{-,\times}\right]}{de} - 1\right| \leq \frac{de}{q}.$$

Note that $\gcd(z^{2^n-1} - 1, g(z) + c) = \gcd(g(z) + c, g_1(z))$, where g_1 is the remainder polynomial with degree less than $\deg(g)$. Hence $\gcd(z^{2^n-1} - 1, g(z) + c)$ has more operations than $\gcd(g(z) + c, g_0(z))$, where $\deg(g_0) = 9$. According to the above theorem, for the polynomial $g(z) + c \in \mathbb{F}_{2^n}[x]$ of degree $e = 2^{m+1} + 1$, calculating $\gcd(g(z) + c, g_0(z))$ with $d = \deg(g_0) = 9$ on average takes $d + 1 = 10$ polynomial divisions, $e + d + 1 = 2^{m+1} + 10 \approx \sqrt{q}$ divisions and $ed = 9(2^{m+1} + 1) \approx 9\sqrt{q}$ addition and multiplications in $\mathbb{F}_q$ for $q = 2^n = 2^{2m+1}$. On the other hand, finding roots of $g(z) = c$ with Chien search method for $t = 2$ takes on average $\frac{tq}{2} = q$ operations in $\mathbb{F}_q$. In this sense, it is better to calculate $\gcd(g(z) + c, g_0(z))$ in solving the equation $g(z) = c$ for Step 1.

Suppose η is the root of $g(z) = c$ in Step 1. From the equality $y^{2^m} + y = \eta$, we can obtain the quadratic equation $y^2 + y = \eta^2 + \eta^{2^{m+1}}$. Let $\theta = \beta^2 + \beta^{2^{m+1}}$. Then, it satisfies $\mathrm{Tr}_1^n(\theta) = 0$. Suppose for a normal basis $\beta^{2^i}, i = 0, 1, \ldots, n-1$, the element $\theta = \sum_{i=0}^{n-1} \theta_i \beta^{2^i}$

and $y = \sum_{i=0}^{n-1} y_i \beta^{2^i-1}$. Then we obtain the following system of n linear equations with rank $n-1$ in n variables in $\mathbb{F}_2$:

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 1 \\ 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n-1} \end{bmatrix},$$

which has solutions $y_i = y_{n-1} + \sum_{j=0}^{i} \theta_j$ for $i = 0, 1, \dots, n-2$ and $y_{n-1} \in \mathbb{F}_2$. Here we provide this elementary process to show its complexity is in order of $O(n)$ instead of the typical complexity $O(n^3)$ for solving linearized equations over $\mathbb{F}_{2^n}$.

With the two solutions y_1, y_2 in Step 2, the two corresponding error positions i_1, i_2 can be immediately obtained from $\beta^{i_1} = x_1 = y_1 s_1$, $\beta^{i_2} = x_2 = y_2 s_1$.

We now discuss another family of binary cyclic codes closely related to the Welch permutation. For the Welch permutation $g(x) = x^{2^{m+1}+1} + x^3 + x$, we define a cyclic code $\mathcal{C}_2$ with the primary defining set $\{1, 3, 2^{m+1}+1\}$. We see that the code $\mathcal{C}_2$ is a subcode of the trivial double-error-correcting BCH code, namely, its primary defining set is $\{1, 3\}$. Interestingly, the code $\mathcal{C}_2$ actually has properties rather similar to that of the triple-error-correcting BCH code with primary defining set $\{1, 3, 5\}$. With the fact that $(2^{2(m+1)} + 1) \equiv 3 \pmod{2^{2m+1} - 1}$, we see that the defining set of $\mathcal{C}_2$ can be written as $\{1, 2^k + 1, 2^{2k} + 1\}$, where $k = m + 1$. The first author and Bracken [8] showed that $\mathcal{C}_2$ has minimum distance 7. Note that the dual code $\mathcal{C}_2^{\perp}$ is given by

$$\mathcal{C}_2^{\perp} = \left\{ \left(\mathrm{Tr}_1^n(ax^{2^{m+1}+1} + bx^3 + cx) \right)_{x \in \mathbb{F}_{2^n}^*} : a, b, c \in \mathbb{F}_{2^n} \right\}.$$

Luo [34] determined the weight distribution of binary codes given by

$$\mathcal{D}_k = \left\{ \left(\mathrm{Tr}_1^n(ax^{2^{2k}+1} + bx^{2^k+1} + cx) \right)_{x \in \mathbb{F}_{2^n}^*} : a, b, c \in \mathbb{F}_{2^n} \right\},$$

when $n/\gcd(n, k)$ is odd. From [34, Theorem 1] one can readily see that the codes $\mathcal{D}_1$ and $\mathcal{D}_{m+1}$ have exactly the same weight distribution with a 5-weight spectrum

$$\left\{ 2^{n-1}, 2^{n-1} \pm 2^{(n-1)/2}, 2^{n-1} \pm 2^{(n+1)/2} \right\}.$$

This implies that $\mathcal{C}_2$ with defining set $\{1, 3, 2^{m+1}\}$ and the triple-error-correcting BCH code with defining set $\{1, 3, 5\}$ have exactly the same weight distribution. Since the terms $x^{2^{m+1}+1}$ and x^3 have algebraic degree 2, the code $\mathcal{C}_2^{\perp}$ is a super-code of first-order binary Reed-Muller codes in the second-order Reed–Muller code. It is worth noting that in this context, Kai–Uwe Schmidt has made significant contributions, including first-order generalized Reed–Muller codes [41] and complementary sets in the context of sequence design [42].

Charpin et al. [18] showed that the coset weight distribution of triple-error-correcting BCH code of length $N = 2^n - 1$ is given by

$$K_0 = 1, K_1 = \binom{N}{1}, K_2 = \binom{N}{2}, K_3 = \binom{N}{3},$$

$$K_4 = \frac{N(5N^2 + 10N - 3)}{6}, K_5 = \frac{4N(N+2)}{3},$$

where K_i is the number of coset leaders with weight i. Moreno and Castro in [39] showed that binary cyclic codes with primary defining set $\{1, 2^k+1, 2^{2k}+1\}$, where $\gcd(k, n) = 1$, have covering radius 5. This implies that the cyclic code $\mathcal{C}_2$ has covering radius 5. Experimental results show that for $m \geq 3$, the cyclic code $\mathcal{C}_2$ has the same coset distribution as above. In our view, this is an interesting connection. Nevertheless, we are not able to provide a theoretical proof for this fact.

Below we discuss the decoding of this triple-error-correcting code. Suppose a received vector y contains an error e of weight 2. Since $\mathcal{C}_2$ has defining set $\{1, 3, 2^{m+1}+1\}$, the error e can be corrected with a BCH decoder. On the other hand, since the code length is $2^n - 1$, finding the roots of the error-locator polynomial directly would be costly when n increases. Instead, the following process works more efficiently. Let $(s_1, s_2, s_3) = yH^{\perp}$. We obtain the following system of equations as in (24):

$$\begin{cases} x_1 + x_2 = s_1, \\ x_1^3 + x_2^3 = s_2, \\ x_1^{2^{m+1}+1} + x_2^{2^{m+1}+1} = s_3, \end{cases}$$

where $x_t = \beta^{i_t}$ for $t = 1, 2$ and β is a primitive element in $\mathbb{F}_{2^n}$. The first two equations immediately leads to the quadratic equation $s_1x^2 + s_1^2x = s_1^3 + s_2$, where $x = x_1$ or x_2. This quadratic equation can be further transformed to $y^2 + y = 1 + \frac{s_2}{s_1^3}$ by letting $y = \frac{x}{s_3}$. As discussed earlier, this equation can be solved in $O(n)$ operations in $\mathbb{F}_2$.

Now we consider the decoding of triple errors in a vector $y = c + e$. Similarly, we need to solve the following system of equations

$$\begin{cases} x_1 + x_2 + x_3 = s_1, \\ x_1^3 + x_2^3 + x_3^3 = s_2, \\ x_1^{2^{m+1}+1} + x_2^{2^{m+1}+1} + x_3^{2^{m+1}+1} = s_3, \end{cases} \tag{25}$$

where $x_t = \beta^{i_t}$ for $t = 1, 2, 3$. We first transform the above equations into simplified ones in two variables.

Assume $s_1 = 0$. Substituting $x_3 = x_1 + x_2$ in the second and third equations in (25) gives

$$\begin{cases} x_1x_2^2 + x_1^2x_2 = s_2, \\ x_1x_2^{2^k} + x_1^{2^k}x_2 = s_3, \end{cases}$$

where $k = m + 1$. Assume $s_1 \neq 0$. Letting $x_t = s_1(y_t + 1)$ for $t = 1, 2, 3$. Then (25) becomes

$$\begin{cases} (y_1+1)^3 + (y_2+1)^3 + (y_1+y_2+1)^3 = s_2/s_1^3, \\ (y_1+1)^{2^k+1} + (y_2+1)^{2^k+1} + (y_1+y_2+1)^{2^k+1} = s_3/s_1^{2^k+1}, \end{cases}$$

implying

$$\begin{cases} y_1y_2^2 + y_1^2y_2 = 1 + s_2/s_1^3, \\ y_1y_2^{2^k} + y_1^{2^k}y_2 = 1 + s_3/s_1^{2^k+1}. \end{cases}$$

Therefore, for any $s_1 \in \mathbb{F}_{2^n}$, it suffices to focus on only the following equations in y_1, y_2:

$$\begin{cases} y_1y_2^2 + y_1^2y_2 = \delta, \\ y_1y_2^{2^{m+1}} + y_1^{2^{m+1}}y_2 = \tau, \\ y_1y_2^{2^m} + y_1^{2^m}y_2 = \tau^{2^m}, \end{cases} \tag{26}$$

where $(\delta, \tau) = (s_2, s_3)$ for $s_1 = 0$, and $(\delta, \tau) = (1 + s_2/s_1^3, 1 + s_3/s_1^{2^d+1})$ for $s_1 \neq 0$. Note that if $\delta\tau = 0$, the equation implies that $y_1 = y_2$, which is invalid here.

Furthermore, taking $z = y_1/y_2$, we derive the following equations from the above system

$$\begin{cases} y_2^3 = \frac{\delta}{z^2+z}, \\ y_2^{2^{m+1}+1} = \frac{\tau}{z^{2^{m+1}}+z}, \\ y_2^{2^m+1} = \frac{\tau^{2^m}}{z^{2^m}+z}, \end{cases} \implies \begin{cases} y_2^{2^{m+1}-2} = \frac{\tau}{z^{2^{m+1}}+z} \cdot \frac{z^2+z}{\delta}, \\ y_2^{2^m} = \frac{\tau}{z^{2^{m+1}}+z} \cdot \frac{z^{2^m}+z}{\tau^{2^m}}. \end{cases}$$

Assume $w = z^{2^{m+1}} + z$. It is readily seen that $z^2 + z = w^{2^{m+1}} + w$ and $z^{2^m} + z = w^{2^m}$. Then we have

$$\begin{cases} y_2^{2^{m+1}-2} = \frac{\tau}{\delta} \cdot \frac{w^{2^{m+1}}+w}{w}, \\ y_2^{2^{m+1}} = (y_2^{2^m})^2 = \left(\frac{\tau}{\tau^{2^m}} \cdot \frac{w^{2^m}}{w}\right)^2, \\ y_2^2 = (y_2^{2^m})^{2^{m+2}} = \left(\frac{\tau}{\tau^{2^m}} \cdot \frac{w^{2^m}}{w}\right)^{2^{m+2}}. \end{cases} \tag{27}$$

From the above equations, the fact $y_2^{2^{m+1}-2} y_2^2 = y^{2^{m+1}}$ gives

$$\frac{\tau}{\delta} \cdot \frac{w^{2^{m+1}} + w}{w} \left(\frac{\tau}{\tau^{2^m}} \cdot \frac{w^{2^m}}{w}\right)^{2^{m+2}} = \left(\frac{\tau}{\tau^{2^m}} \cdot \frac{w^{2^m}}{w}\right)^2,$$

i.e.,

$$\frac{\tau^{1+2^{m+2}}}{\delta\tau^2} \frac{(w^{2^{m+1}} + w)w^2}{w^{2^{m+2}+1}} = \frac{\tau^2}{\tau^{2^{m+1}}} \cdot \frac{w^{2^{m+1}}}{w^2}.$$

Rearranging this above equation yields

$$w^{3\cdot(2^{m+1})} = \gamma(w^{2^{m+1}+3} + w^4),$$

where $\gamma = \frac{\tau^{3(2^{m+1}-1)}}{\delta}$. Following Dobbertin's method in [29], we denote $\overline{w} = w^{2^{m+1}}$ and $\overline{\gamma} = \gamma^{2^{m+1}}$. Then we have $\overline{w}^{2^{m+1}} = w^2$. The above equation and its 2^{m+1}-th power give two equations

$$\begin{cases} \overline{w}^3 + \gamma(\overline{w}w^3 + w^4) = 0, \\ w^6 + \overline{\gamma}(w^2\overline{w}^3 + \overline{w}^4) = 0. \end{cases} \tag{28}$$

Substituting the first equation to the second one in (28) gives $\overline{\gamma}\gamma(\overline{w}+w)(\overline{w}+w^2)+w^3 = 0$. By the first equation and the new equation, we denote

$$\begin{cases} h_1 = \overline{w}^3 + \gamma w^3 \cdot \overline{w} + \gamma w^4 = 0, \\ h_2 = \overline{w}^2 + (w + w^2)\overline{w} + \gamma_1 w^3 = 0. \end{cases} \tag{29}$$

where $\gamma_1 = 1 + (\overline{\gamma}\gamma)^{-1}$.

Below we will eliminate $\overline{w}$ from h_1, h_2. For reader's convenience, we include the process despite its simplicity. Viewing h_1, h_2 as polynomials in variable $\overline{w}$, by the Euclidean method, we have

$$h_1 = \left(\overline{w} + (w + w^2)\right) \cdot h_2 + h_3,$$

where $h_3 = w^2(w^2 + (\gamma + \gamma_1)w + 1)\overline{w} + w^4(\gamma_1 w + (\gamma + \gamma_1))$. Furthermore, rewrite h_2 and h_3 as $h_2 = \overline{w}^2 + \phi_1\overline{w} + \phi_2$ and $h_3 = \phi_3\overline{w} + \phi_4$ for simplicity. The equation system (29) is equivalent to the following system

$$\begin{cases} h_2 = \overline{w}^2 + \phi_1\overline{w} + \phi_2 = 0, \\ h_3 = \phi_3\overline{w} + \phi_4 = 0. \end{cases} \tag{30}$$

From (30), we can eliminate $\overline{w}$ by calculating

$$\begin{aligned} h_4(w) &= \phi_3^2 h_2 + (\phi_3\overline{w} + \phi_1\phi_3 + \phi_4)h_3 \\ &= \phi_3^2(\overline{w}^2 + \phi_1\overline{w} + \phi_2) + (\phi_3\overline{w} + \phi_1\phi_3 + \phi_4)(\phi_3\overline{w} + \phi_4) \\ &= \phi_2\phi_3^2 + \phi_1\phi_3\phi_4 + \phi_4^2. \end{aligned}$$

More explicitly,

$$\begin{aligned} h_4(w) &= \gamma_1 w^3 w^4(w^2 + (\gamma + \gamma_1)w + 1)^2 \\ &\quad +(w + w^2)w^2(w^2 + (\gamma + \gamma_1)w + 1)w^4((\gamma_1)w + (\gamma + \gamma_1) \\ &\quad +w^8(\gamma_1 w + (\gamma + \gamma_1)^2 \\ &= w^7\left(\gamma(1 + \gamma_1)w^3 + (\gamma(\gamma + 1)(\gamma_1 + 1) + \gamma_1^3)w^2 + \gamma w + \gamma\right) \\ &= \gamma(1 + \gamma_1)w^7\left(w^3 + \sigma_1 w^2 + \sigma_2\gamma w + \sigma_3\right) \\ &= \gamma(1 + \gamma_1)w^7 \cdot h(w), \end{aligned} \tag{31}$$

where $\sigma_1, \sigma_2, \sigma_3$ are coefficients derived from the cubic polynomial in the last second equation and $h(w) = w^3 + \sigma_1 w^2 + \sigma_2 w + \sigma_3$. Therefore, the system (28) is reduced to the cubic equation $h(w) = 0$.

For correctable syndromes $s = (s_1, s_2, s_3)$ derived from an error e of weight 3, it can be verified, according to the criteria in [43], that $h(w)$ has three roots. In order to obtain the roots of $h(w)$, we follow the method by Berlekamp and Solomon [3]. Multiplying $h(w)$ by $w + \sigma_1$, we obtain a linearized polynomial

$$L(w) = (w + \sigma_1)h(w) = w^4 + (\sigma_2 + \sigma_1^2)w^2 + (\sigma_3 + \sigma_1\sigma_2) + \sigma_1\sigma_3.$$

With a basis $\beta_1, \ldots, \beta_n$ of $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$, we can express $L(w) = 0$ as a system of m linear equations in n variables $w_1, \ldots, w_n \in \mathbb{F}_2$. From the possible 4, 2, 1 solutions to $L(w) = (w+\sigma_1)h(w)$, we can get 3, 1 roots of the cubic polynomial $h(w)$ in general. However, since $\sigma_1, \sigma_2, \sigma_3$ were derived from γ, the cubic polynomial $h(w)$ actually has 3 roots. This process has complexity in the order of $O(n^3)$ operations in $\mathbb{F}_2$. Given a root w of $h(w)$, from the equation $z^{2^{m+1}} + z = w$, one can obtain the equation $z^2 + z = w^{2^{m+1}} + w$ and can find the roots z in $O(n)$ operations in $\mathbb{F}_2$. By (27) we can get the unique root y_2 from δ, τ, w; and by $z = y_1/y_2$, one can get solutions (y_1, y_2) from $\{(y_2 z, y_2), (y_2 z + y_2, y_2)\}$ for the system (26) of equations. Furthermore, for either $s_1 = 0$ or $s_1 \neq 0$, we can get two solutions (x_1, x_2, x_3) from $(y_1, y_2) \in \{(y_2 z, y_2), (y_2 z + y_2, y_2)\}$. Here it is to be noted that three roots w of $h(w)$ leads to six solutions (x_1, x_2, x_3) for the syndrome equations. These six solutions correspond to the 6 permutations of one error e with support $\{i_1, i_2, i_3\}$.

To summarize, the decoding of the cyclic code $\mathcal{C}_2$ for three errors can proceed as follows:

- given a syndrome $(s_1, s_2, s_3) = yH^T$, calculate the corresponding δ, τ from (s_1, s_2, s_3) and $\gamma = \frac{\tau^{3(2^{m+1}-1)}}{\delta}$;
- calculate the polynomial $h(w)$ as in (31);
- construct the linearized polynomial $L(w)$ from $h(w)$ and find its solutions in $\mathbb{F}_{2^n}$;
- for the solution w, calculate the intermediate parameters y_1, y_2, and then use them to calculate $x_t = \beta^{i_t}$ for $t = 1, 2, 3$

- recover the codeword $c = y + e$ with the support of e being $\{i_1, i_2, i_3\}$.

Denote by $N = 2^n - 1$ the length of the code $\mathcal{C}_2$. In the above decoding procedure, the calculation of syndrome takes $O(N \log N)$ operations in $\mathbb{F}_2$; the calculation of $h_4 = \gcd(h_1, h_2)$ is independent of the code length N and solving $L(w) = 0$ takes $O((\log N)^3)$ operations in $\mathbb{F}_2$. This decoder significantly outperforms the syndrome decoder with complexity $O(N^3)$ and recent decoders for cyclic codes in [1, 16], which have complexity at least $O(N^2)$ for triple-error-correcting cyclic codes.

4.2 Binary linear codes related to the Welch permutation

In this section we will discuss two families of binary linear codes that are relevant to the Welch permutation.

For the Welch permutation $g(x)$, the first family of binary linear code $\mathcal{C}_3$ is given by a parity-check matrix similar to (23), where β^{di} is replaced by $g(\beta^i)$. It is well known [13, 14] that $\mathcal{C}_3$ defined in this way has minimum distance at most 5, and dimension $2^n - 1 - 2n$. According to the proof of Theorem 2, there exist $(x, y, z) \in \left(\mathbb{F}_{3^n}^*\right)^3$ such that

$$\begin{cases} x + y + z = 0, \\ g(x) + g(y) + g(z) = 0. \end{cases}$$

Thus, the code $\mathcal{C}_3$ has minimum distance 3. In addition, for its dual code,

$$\mathcal{C}_3^{\perp} = \left\{ \left(\mathrm{Tr}_1^n(ag(x) + bx) \right)_{x \in \mathbb{F}_{2^n}^*} : a, b \in \mathbb{F}_{2^n} \right\}.$$

According to (6), the weight of nonzero codewords in $\mathcal{C}_3^{\perp}$ can be expressed in terms of the extended Walsh transform of $g(x)$. From the Walsh spectrum obtained in Sect. 3, we see that the weight distribution of $\mathcal{C}_3^{\perp}$ is obtained accordingly, which is a 5-weight spectrum

$$\{2^{n-1}, 2^{n-1} \pm 2^{m+1}, 2^{n-1} \pm 2^m\}.$$

Now let's consider another binary linear code related to the Welch permutation. Ding et al. in [24, 26] introduced a generic construction of binary linear codes from a subset $D = \{d_1, d_2, \ldots, d_\ell\}$ of $\mathbb{F}_{2^n}$ and the absolute trace function $\mathrm{Tr}_1^n(\cdot)$ from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$ as

$$\mathcal{C}_D = \left\{ \mathbf{c}_a = \left(\mathrm{Tr}_1^n(ad_1), \mathrm{Tr}_1^n(ad_2), \ldots, \mathrm{Tr}_1^n(ad_\ell) \right) : a \in \mathbb{F}_{2^n} \right\}.$$

When the defining set D is properly chosen, the code $\mathcal{C}_D$ can have a few nonzero weights. Particularly, when the defining set is given as $D(F) = \{F(x) : x \in \mathbb{F}_2^n\}$ with a two-to-one function F on $\mathbb{F}_{2^n}$, the Hamming weight of a codeword $\mathbf{c}_a$ in $\mathcal{C}_{D(F)}$ is given by

$$\begin{aligned} \mathrm{wt}(\mathbf{c}_a) &= \left| \{1 \le i \le 2^{n-1} : \mathrm{Tr}_1^n(ad_i) = 1\} \right| \\ &= \frac{1}{2} \left(2^{n-1} - \sum_{d \in D(F)} (-1)^{\mathrm{Tr}_1^n(ad)} \right) \\ &= \frac{1}{2} \left(2^{n-1} - \frac{1}{2} \sum_{x \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(aF(x))} \right) \\ &= 2^{n-2} - \frac{1}{4} \sum_{x \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(aF(x))}. \end{aligned} \tag{32}$$

That is to say, for studying the Hamming weight properties of the code $\mathcal{C}_{D(F)}$, it is critical to study the possible values of the exponential sum $W_F(a, 0)$.

In [24], Ding investigated the properties of binary linear codes from the images of certain functions on $\mathbb{F}_{2^n}$ and proposed several conjectures on properties of the constructed codes, including the following one from Welch APN power function.

Conjecture 1 ([24, Conjecture 33]) *Let $n = 2m + 1$ and $F(x) = x^{2^m+3}$. Let $f(x) = F(x) + F(x + 1) + 1$ and $D(f) = \{d_1, d_2, \ldots, d_\ell\} = \{f(x) \,|\, x \in \mathbb{F}_{2^n}\}$. Define the binary code $\mathcal{C}_{D(f)}$ as*

$$\mathcal{C}_{D(f)} = \left\{ \boldsymbol{c}_a = \left(Tr_1^n(ad_1), Tr_1^n(ad_2), \ldots, Tr_1^n(ad_\ell) \right) : a \in \mathbb{F}_{2^n} \right\}.$$

If $n \in \{5, 7\}$, then $\mathcal{C}_{D(f)}$ is a three-weight code with length 2^{n-1} and dimension n. If $n \geq 9$, then $\mathcal{C}_{D(f)}$ is a five-weight code with length 2^{n-1} and dimension n.

For the Welch APN power function $F(x) = x^{2^m+3}$ and $f(x) = F(x + 1) + F(x) + 1$, it is easy to verify that

$$f(x) = F(x + 1) + F(x) + 1 = (x + x^{2^m})(x^2 + x + 1) = g(x + x^{2^m}),$$

where $g(x)$ is the Welch permutation of $\mathbb{F}_{2^n}$. With the properties of $g(x)$ discussed in Sect. 3, we present the following result on the code $\mathcal{C}_{D(f)}$.

Theorem 5 *Let $n = 2m + 1$ for a positive integer $m \geq 2$. The binary linear code $\mathcal{C}_{D(f)}$ defined in Conjecture 1 has length 2^{n-1}, dimension n and its nonzero weights are contained in the following set:*

$$\left\{ 2^{n-2}, 2^{n-2} \pm 2^{\frac{n-3}{2}}, 2^{2m-1} \pm 2^{\frac{n-1}{2}} \right\}.$$

Proof It is clear that the length of $\mathcal{C}_{D(f)}$ is 2^{n-1} since $f(x) = g(x + x^{2^m})$ is a two-to-one function. As for the dimension, since $\mathcal{C}_{D(f)}$ is linear, we need to consider the number of $a \in \mathbb{F}_{2^n}$ such that $\mathrm{Tr}_1^n(af(x)) = 0$ for any $x \in \mathbb{F}_{2^n}$, equivalently, $\sum_{x \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(af(x))} = 2^n$.

Define $T_0 = \{x + x^{2^m} \,|\, x \in \mathbb{F}_{2^n}\}$ and $T_1 = \{x + 1 \,|\, x \in T_0\}$. Note that $x + x^{2^m}$ is a two-to-one function over $\mathbb{F}_{2^n}$. Thus $T_0 \cup T_1 = \mathbb{F}_{2^n}$. Moreover, we have $\mathrm{Tr}_1^n(1) = 1$ since n is odd, $\mathrm{Tr}_1^n(x) = 0$ for any $x \in T_0$ and $\mathrm{Tr}_1^n(x) = 1$ for any $x \in T_1$. Since $g(x)$ is a permutation of $\mathbb{F}_{2^n}$, one has

$$\sum_{z \in T_0} (-1)^{\mathrm{Tr}_1^n(bg(z))} + \sum_{z \in T_1} (-1)^{\mathrm{Tr}_1^n(bg(z))} = \sum_{z \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}_1^n(bg(z))} = 0,$$

which implies that

$$\begin{aligned} \sum_{z \in T_0} (-1)^{\mathrm{Tr}_1^n(bg(z))} &= \sum_{z \in T_1} (-1)^{\mathrm{Tr}_1^n(bg(z)+1)} \\ &= \sum_{z \in T_0} (-1)^{\mathrm{Tr}_1^n(bg(z+1)+1)}. \end{aligned}$$

Table 2 Some numerical results

Values of	Weight enumerator of $\mathcal{C}_{D(f)}$
5	$1+6x^{10}+15x^8+10x^6$
7	$1+63x^{32}+36x^{28}+28x^{36}$
9	$1+x^{144}+108x^{120}+285x^{128}+108x^{136}+9x^{112}$
11	$1+440x^{496}+408x^{528}+22x^{480}+1155x^{512}+22x^{544}$

Therefore, for any $a \in \mathbb{F}_{2^n}^*$, we have

$$\begin{aligned}\sum_{x\in\mathbb{F}_{2^n}}(-1)^{\mathrm{Tr}_1^n(af(x))} &= 2\sum_{z\in T_0}(-1)^{\mathrm{Tr}_1^n(ag(z))}\\ &= \sum_{z\in T_0}(-1)^{\mathrm{Tr}_1^n(ag(z))}+\sum_{z\in T_0}(-1)^{\mathrm{Tr}_1^n(ag(z+1)+1)}\\ &= \sum_{z\in T_0}(-1)^{\mathrm{Tr}_1^n(ag(z)+z)}+\sum_{z\in T_0}(-1)^{\mathrm{Tr}_1^n(ag(z+1)+z+1)}\\ &= \sum_{z\in T_0}(-1)^{\mathrm{Tr}_1^n(ag(z)+z)}+\sum_{z\in T_1}(-1)^{\mathrm{Tr}_1^n(ag(z)+z)}\\ &= \sum_{x\in\mathbb{F}_{2^n}}(-1)^{\mathrm{Tr}_1^n(ag(x)+x)}.\end{aligned}$$

By the extended Walsh spectrum of $g(x)$ in Theorem 3, it is clear that $W_f(a,0) = \sum_{x\in\mathbb{F}_{2^n}}(-1)^{\mathrm{Tr}_1^n(af(x))} = W_g(a,1) \neq 2^n$ for any nonzero $a \in \mathbb{F}_{2^n}$. This means that for different $a \in \mathbb{F}_{2^n}^*$, the codewords $\mathbf{c}_a$ are different. Thus, $\mathcal{C}_{D(f)}$ has dimension n.

Furthermore, it follows from (32) that

$$\mathrm{wt}(\mathbf{c}_a) = 2^{n-2} - \frac{1}{4}\sum_{x\in\mathbb{F}_{2^n}}(-1)^{\mathrm{Tr}_1^n(ag(x)+x)}.$$

From the extended Walsh spectrum of $g(x)$ in Table 1, the possible nonzero weights of the code $\mathcal{C}_{D(f)}$ can be directly determined. □

Theorem 5 provides a partial resolution to the conjecture by Ding [24]. It appears that new technique is required to completely settle the conjecture and determine the weight distribution of the code $\mathcal{C}_{D(f)}$. With the help of Magma, we list some numerical results in Table 2, which are in accordance with Theorem 5.

5 Conclusion

The contributions in this paper are twofold. First, we completely determined the differential spectrum and the Walsh spectrum of the permutation polynomial $g(x)$ from the Welch APN power function x^{2^m+3} over $\mathbb{F}_{2^{2m+1}}$. Second, we explore two families of cyclic codes and two families of linear codes derived from the Welch APN power function. For the two cyclic codes, their properties have been well studied, and we present efficient algebraic decoders for them; for the two linear codes, the weight distribution of the first family can be easily

obtained from the Walsh spectrum of $g(x)$, and the weight spectrum of the second one was investigated, which partially solved a conjecture by Ding in [24]. The Welch permutation $g(x)$ appears to have good cryptographic properties and some other cryptographic criteria may deserve further investigation.

Acknowledgements Y. Xia was supported in part by the National Natural Science Foundation of China under Grant 62171479 and in part by the Fundamental Research Funds for the Central Universities, South-Central Minzu University under Grant CZZ23004. C. Li and T. Helleseth were supported by the Research Council of Norway under Grant 311646/O70.

Author Contributions The authors contributed equally to this work. Helleseth Tor led and supervised this research. Chunlei Li and Yongbo Xia wrote the main manuscript text. All authors reviewed the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declared that they have no Conflict of interest in connection with the work submitted.

References

1. Augot D., Bardet M., Faugere J.-C.: On the decoding of binary cyclic codes with the Newton identities. J. Symb. Comput. **44**(12), 1608–1625 (2009).
2. Berlekamp E.: Algebraic Coding Theory, Revised World Scientific Publishing, Singapore (2015).
3. Berlekamp E.R., Rumsey H., Solomon G.: On the solution of algebraic equations over finite fields. Inf. Control **10**(6), 553–564 (1967).
4. Berlekamp E., McEliece R., Van Tilborg H.: On the inherent intractability of certain coding problems (Corresp.). IEEE Trans. Inf. Theory **24**(3), 384–386 (1978).
5. Biham E., Shamir A.: Differential cryptanalysis of DES-like cryptosystem. J. Cryptol. **4**(1), 3–72 (1991).
6. Blondeau C., Canteaut A., Charpin P.: Differential properties of power functions. Int. J. Inf. Coding Theory. **1**(2), 149–170 (2010).
7. Blondeau C., Canteaut A., Charpin P.: Differential properties of $x \mapsto x^{2^t-1}$. IEEE Trans. Inf. Theory **57**(12), 8127–8137 (2011).
8. Bracken C., Helleseth T.: Triple-error-correcting BCH-like codes. In: 2009 IEEE International Symposium on Information Theory, pp. 1723–1725. IEEE Xplore, Seoul, Korea (South) (2009).
9. Bracken C., Leander G.: A highly nonlinear differentially 4-uniform power mapping that permutes fields of even degree. Finite Fields Appl. **16**(4), 231–242 (2010).
10. Bracken C., Byrne E., Markin N., McGuire G.: Determining the nonlinearity of a new family of APN functions. In: Boztas S., Lu H.-F. (eds.) Applied Algebra, Algebraic Algorithms and Error-Correcting Codes. AAECC 2007. Lecture Notes in Computer Science, vol. 765, pp. 72–79. Springer, Berlin (2007).
11. Budaghyan L.: Construction and Analysis of Cryptographic Functions. Springer, Cham (2014).
12. Canteaut A., Charpin P., Dobbertin H.: Binary m-sequences with three-valued crosscorrelation: a proof of Welch's conjecture. IEEE Trans. Inf. Theory **46**(1), 4–8 (2000).
13. Carlet C.: Boolean Functions for Cryptography and Coding Theory. Cambridge University Press, Cambridge (2021).
14. Carlet C., Charpin P., Zinoviev V.: Codes, bent functions and permutations suitable for DES-like cryptosystems. Des. Codes Cryptogr. **15**(2), 125–156 (1998).
15. Carlet C., Ding C., Yuan J.: Linear codes from perfect nonlinear mappings and their secret sharing schemes. IEEE Trans. Inf. Theory **51**(6), 2089–2102 (2005).
16. Caruso F., Orsini E., Sala M., Tinnirello C.: On the shape of the general error locator polynomial for cyclic codes. IEEE Trans. Inf. Theory **63**(6), 3641–3657 (2017).
17. Chabaud F., Vaudenay S.: Links between differential and linear cryptanalysis. In: De Santis A. (ed.) Advances in Cryptology - EUROCRYPT'94. Lecture Notes in Computer Science, vol. 950, pp. 356–365. Springer, Berlin (1995).
18. Charpin P., Helleseth T., Zinoviev V.A.: The coset distribution of triple-error-correcting binary primitive BCH codes. IEEE Trans. Inf. Theory **52**(4), 1727–1732 (2006).

19. Charpin P., Kyureghyan G.M., Suder V.: Sparse permutations with low differential uniformity. Finite Fields Appl. **28**, 214–243 (2014).
20. Chen C.L.: Formulas for the solutions of quadratic equations over GF(2^m). IEEE Trans. Inf. Theory **28**(5), 792–794 (1982).
21. Cohen G., Karpovsky M., Mattson H., Schatz J.: Covering radius-survey and recent results. IEEE Trans. Inf. Theory **31**(3), 328–343 (1985).
22. Delsarte P.: On subfield subcodes of modified Reed–Solomon codes (Corresp.). IEEE Trans. Inf. Theory **21**(5), 575–576 (1975).
23. Ding C.: Linear codes from some 2-designs. IEEE Trans. Inf. Theory **61**(6), 3265–3275 (2015).
24. Ding C.: A construction of binary linear codes from Boolean functions. Discret. Math. **339**(9), 2288–2303 (2016).
25. Ding K., Ding C.: A class of two-weight and three-weight codes and their applications in secret sharing. IEEE Trans. Inf. Theory **61**(11), 5835–5842 (2015).
26. Ding C., Niederreiter H.: Cyclotomic linear codes of order 3. IEEE Trans. Inf. Theory **53**(6), 2274–2277 (2007).
27. Ding C., Wang X.: A coding theory construction of new systematic authentication codes. Theor. Comput. Sci. **330**(1), 81–99 (2005).
28. Ding C., Li C., Li N., Zhou Z.: Three-weight cyclic codes and their weight distributions. Discret. Math. **339**(2), 415–427 (2016).
29. Dobbertin H.: Almost perfect nonlinear power functions on GF(2^n): the Welch case. IEEE Trans. Inf. Theory **45**(4), 1271–1275 (1999).
30. Giménez N., Matera G., Pérez M., Privitelli M.: Average-case complexity of the Euclidean algorithm with a fixed polynomial over a finite field. Comb. Probab. Comput. **31**(1), 166–183 (2022).
31. Helleseth T., Kumar P.V.: Sequences with low correlation. In: Pless V.S., Huffman W.C. (eds.) Handbook of Coding Theory, vol. II, pp. 1765–1853. North-Holland, Amsterdam (1998).
32. Li N., Mesnager S.: Recent results and problems on constructions of linear codes from cryptographic functions. Cryptogr. Commun. **12**(5), 965–986 (2020).
33. Lin T.-C., Lee C.-D., Chen Y.-H., Truong T.-K.: Algebraic decoding of cyclic codes without error-locator polynomials. IEEE Trans. Commun. **64**(7), 2719–2731 (2016).
34. Luo J.: On binary cyclic codes with five nonzero weights. https://doi.org/10.48550/arXiv.0904.2237 (2009)
35. MacWilliams F.J., Sloane N.J.A.: The Theory of Error-Correcting Codes. North-Holland, Amsterdam (1977).
36. Matsui M.: Linear cryptanalysis method for DES cipher. In: Helleseth T. (ed.) Advances in Cryptology - Eurocrypt'93. Lecture Notes in Computer Sceince, vol. 765, pp. 386–397. Springer, Berlin (1994).
37. Mesnager S.: Linear codes with few weights from weakly regular bent functions based on a generic construction. Cryptogr. Commun. **9**(1), 71–84 (2017).
38. Mesnager S., Qu L.: On two-to-one mappings over finite fields. IEEE Trans. Inf. Theory **65**(12), 7884–7895 (2019).
39. Moreno O., Castro F.N.: Divisibility properties for covering radius of certain cyclic codes. IEEE Trans. Inf. Theory **49**(12), 3299–3303 (2003).
40. Nyberg K.: Differentially uniform mappings for cryptography. In: Helleseth, T. (ed.) Advances in Cryptology – Eurocrypt'93. Lecture Notes in Computer Sceince, vol. 765, pp. 55–64. Springer, Berlin (1994)
41. Schmidt K.-U.: On cosets of the generalized first-order Reed-Muller code with low OFDM. IEEE Trans. Inf. Theory **52**(7), 3220–3232 (2006).
42. Schmidt K.-U.: Complementary sets, generalized Reed–Muller codes, and power control for OFDM. IEEE Trans. Inf. Theory **53**(2), 808–814 (2007).
43. Williams K.S.: Note on cubics over GF(2^n) and GF(3^n). J. Number Theory **7**(4), 361–365 (1975).
44. Wu C.-K., Feng D.: Boolean Functions and Their Applications in Cryptography. Springer, Berlin (2016).
45. Zhou Z., Li N., Fan C., Helleseth T.: Linear codes with two or three weights from quadratic bent functions. Des. Codes Cryptogr. **81**(2), 283–295 (2016).