On Constructions of Binary Locally Repairable Codes with Locality Two and Multiple Repair Alternatives via Autocorrelation Spectra of Boolean Functions

Deng Tang* Jian Liu† Sihem Mesnager ‡

Abstract

Distributed storage systems (in brief, DSSs) store data on several distributed nodes and are widely used in file system storage, large database storage, backup file, and cloud storage, etc. DSSs provide reliable access to data through redundancy spread over individually unreliable nodes, where the replication scheme and coding mechanism are two widespread techniques for ensuring reliability. In 2012, Gopalan et al. proposed locally repairable codes (LRCs for short) to minimize the number of nodes to be downloaded in repairing any node. A code over a finite alphabet is called LRC (with locality r) if every symbol in the encoding is a function of a small number (at most r) of other symbols of the codeword. In 2013 Pamies-Juarez et al. introduced LRCs with multiple repair alternatives, which allows repairing any node with different disjoint nodes. LRCs with multiple repair alternatives can increase the probability of being able to perform efficient repairs when there are multiple unavailable nodes (these nodes are failed or temporarily unavailable).

This paper proposes two large families of LRCs with multiple repair alternatives from Boolean functions. Each repair set has at most $r = 2$ symbols, which correspond to an interesting case in practice. We shall explore Boolean functions selected from the well-known Maiorana-McFarland class based on partial spreads, respectively. Moreover, we show that the number of the disjoint repair sets (denoted by t) of our LRCs can be determined entirely by the autocorrelation spectrum of the corresponding Boolean function. This achievement is obtained thanks to the relationship between the autocorrelation spectrum of the corresponding Boolean function and the number of disjoint repairs sets that we establish. Our results give rise LRCs with suitable parameters from special Boolean functions (such as bent functions) based on a construction method introduced by Ding in 2015 for designing linear codes based on the so-called "defining set" (involving mainly Boolean functions). The approach presented in this article introduces an interesting connection between LRCs (with multiple repair alternatives) and (the autocorrelation spectrum of) Boolean functions. Notably, it emphasizes a novel role of bent functions for designing LRCs. This connection has not been pointed out before to the best of our knowledge.

Keywords: Binary linear code, LRC, Boolean function, autocorrelation spectrum, Walsh transform, partial spread.

*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, P. R. China. Email: dtang@foxmail.com (D. Tang)

†School of Cybersecurity, College of Intelligence and Computing, Tianjin University, Tianjin, 300350, P. R. China. Email:jianliu.nk@gmail.com (J. Liu)

‡Department of Mathematics, University of Paris VIII, F-93526 Saint-Denis, University Sorbonne Paris Cité, LAGA, UMR 7539, CNRS, 93430 Villetaneuse and Telecom Paris, Polytechnic Institute of Paris, 91120 Palaiseau, France. Email: smesnager@univ-paris8.fr (S. Mesnager)

1 Introduction

The need for highly scalable and reliable extensive data storage systems is due to the fact that there is explosive growth in data. Distributed storage systems (DSSs) store data on several distributed nodes and are widely used in file system storage, large database storage, backup file, cloud storage, etc. The repair problem in DSSs addresses the recovery of the data encoded using erasure codes such as Reed-Solomon codes, *locally repairable codes* (LRCs) [9] etc. In recent years, the interest and attention on LRCs have proliferated. Several constructions and related results have been given (see. e.g. [10, 1, 22, 23]).

A binary linear code $\mathcal{C}$ of length n, dimension κ is a κ-dimensional subspace of $\mathbb{F}_2^n$. $\mathcal{C}$ is said to be an $[n, \kappa, d]$ linear code with minimum Hamming distance d. Linear codes are an important class of codes in coding theory. They have been extensively studied due to their significant applications in practical systems. In this paper, we shall focus on (binary) locally repairable codes (LRCs) which process four parameters by considering the locality r (in addition to those for usual linear codes). We briefly recall the terminology used in the literature in the context of LRC for distributed storage. Specifically, an LRC is said to have *locality* r if the value at any coordinate can be recovered by accessing at most r other coordinates, and to have *availability* t if every coordinate can be recovered from t disjoint repair sets of other coordinates. A code with multiple repair sets (called availability, see [15, 16] for instance) has the advantage of good parallel repairability, since for the target symbol, each repair set can be seen as a backup that can be accessed independently. Locally repairable codes with availability $t > 1$ have been extensively studied in recent years. An upper bound on the minimum distance of LRCs was derived in [19]. When such an upper bound is achieved with equality, the LRC code is optimal. In [20, 21], binary locally repairable codes were constructed by employing combinatorial structures. Constructions of locally repairable codes with availability $t > 1$ were proposed in the literature.

In this paper, we present two families of binary LRCs with multiple repair alternatives inspired by the design method of linear codes from the support set of a Boolean function presented by Ding in [6, 7]. We show that $[n, \kappa, d]$ LRCs with $r = 2$ can be constructed directly from the support set of a Boolean function, and the minimum distance d, as well as the availability t, depend only on the Walsh spectrum and the autocorrelation spectrum of the Boolean function. Using this connection, we firstly give a class of binary LRCs from Boolean functions with Maiorana-McFarland (M-M) constructions. By analyzing the related cryptographic criteria of the M-M functions, we obtain the explicit parameters of these LRCs, which leads to a large number of good LRCs with $r = 2$ and pre-defined t. Secondly, we provide another construction of LRCs with $r = 2$ from Boolean functions based on partial spreads, where the availability t relates to the dimension of the chosen spreads. The remainder of this extended abstract is organized as follows. In Section 2, we fix our notation and introduce some background and necessary preliminaries required for the subsequent sections. In Section 3, we start by recalling, namely, a construction method of binary linear codes from Boolean functions that we follow and present an important result (Theorem 2) on constructing binary LRCs from Boolean functions. We next explore some wide families of Boolean functions and investigate in Section 4 the design of LRCs from Boolean functions with M-M constructions (Subsection 4.1) and based on partial spreads (Subsection 4.2), respectively.

2 Preliminaries

Given a finite set E, $\#E$ will denote its cardinality. Given a real number x, $|x|$ will denote its absolute value. Given a positive integer n, $[n]$ will denote the set $\{1, 2, \cdots, n\}$.

2.1 Boolean functions and related notions

For any positive integer m, we denote by $\mathbb{F}_2^m$ the vector space of m-tuples over the finite field $\mathbb{F}_2 = \{0, 1\}$, and by $\mathbb{F}_{2^m}$ the finite field of order 2^m. For simplicity, we denote by $\mathbb{F}_2^{m*}$ the set $\mathbb{F}_2^m \setminus \{\mathbf{0}\}$, and $\mathbb{F}_{2^m}^*$ denotes the set $\mathbb{F}_{2^m} \setminus \{0\}$, where $\mathbf{0}$ is the all-zero vector in $\mathbb{F}_2^m$. We use $+$ (resp. $\sum$) to denote the addition (resp. a multiple sum) in $\mathbb{Z}$ or in the finite field $\mathbb{F}_{2^m}$, and $\oplus$ (resp. $\bigoplus$) to denote the addition (resp. a multiple sum) in $\mathbb{F}_2$. For simplicity, when there is no ambiguity, we will use $+$ instead of $\oplus$. A Boolean function of m variables is a function from $\mathbb{F}_2^m$ into $\mathbb{F}_2$. If we identify $\mathbb{F}_2^m$ with $\mathbb{F}_{2^m}$, it is a mapping from $\mathbb{F}_{2^m}$ to $\mathbb{F}$. We shall denote by $\mathcal{B}_m$ the set of m-variable Boolean functions. The design of strong symmetric cipher systems requires that the underlying cryptographic Boolean function meet specific security requirements. Some of the required security criteria can be measured with the help of the autocorrelation function or using the Walsh transform as a tool. The *Walsh transform* of f in $\mathcal{B}_m$ at $\alpha \in \mathbb{F}_{2^m}$ is defined by $\widehat{\chi_f}(a) = \sum_{x \in \mathbb{F}_{2^m}} (-1)^{f(x)+\mathrm{Tr}_1^m(ax)}$, where $\mathrm{Tr}_1^m(x) = \sum_{i=0}^{m-1} x^{2^i}$ is the (absolute) trace function from $\mathbb{F}_{2^m}$ to $\mathbb{F}_2$. The multiset constituted by the values of the Walsh transform is called the *Walsh spectrum* of f. The *autocorrelation function* of a Boolean function f in $\mathcal{B}_m$ at a point $\alpha \in \mathbb{F}_{2^m}$ is defined by $A_f(\alpha) = \sum_{x \in \mathbb{F}_{2^m}} (-1)^{f(x)+f(x+\alpha)}$. The multiset constituted by the values of the autocorrelation function is called the *autocorrelation spectrum* of f. The well-known Wiener-Khintchine theorem connects the Walsh transform and the autocorrelation function (see e.g. [2]). A valuable reference on the theory of Boolean functions in cryptography and coding theory is [3].

2.2 Linear codes, LRC codes and related notions

An $[n, \kappa, d]_2$ linear code $\mathcal{C}$ over $\mathbb{F}_2$ is a κ-dimensional subspace of $\mathbb{F}_2^n$ with minimum Hamming distance d, where $d = \min_{a,b \in \mathcal{C}, a \neq b, d_H(a,b)}$ in which d_H denotes the Hamming distance between vectors (called codewords) $a = (a_1, a_2, \cdots, a_n) \in \mathcal{C}$ and $b = (b_1, b_2, \cdots, b_n) \in \mathcal{C}$, i.e., $d_H(a, b) = \#\{1 \leq i \leq n : a_i \neq b_i\}$. For a given codeword $a \in \mathcal{C}$, the Hamming weight $w_H(a)$ is defined as the number of nonzero coordinates. Usually, if the context is clear, we omit the subscript 2 by convention in the sequel (we shall write $[n, \kappa, d]$ instead of $[n, \kappa, d]_2$). We introduce the formal definition of locally repairable codes (LRCs) (see. [8] and also [18]).

Definition 1. *A linear code $\mathcal{C}$ is a LRC with locality r if for any $i \in [n]$, there exists a subset $\mathcal{R}_i \subset [n] \backslash \{i\}$ with $\#\mathcal{R}_i \leq r$ such that the i-th symbol c_i can be recovered by $\{c_j\}_{j \in \mathcal{R}_i}$. A set $\mathcal{R}_i$ is called a recovery or repair set for c_i. Furthermore, if for any $i \in [n]$, there are at least t disjoint repair sets with each set of size at most r symbols, we refer to such a code as an (r, t)-LRC.*

In order to maximize the reliability of storage systems it is desirable to obtain codes where lost data can be repaired by contacting a small number of nodes r where this number can be as small as $r = 2$.

2.3 Binary linear codes from Boolean functions

Several constructions methods of linear codes from special functions (essentially from cryptographic Boolean functions which play a crucial role in symmetric cryptography) over finite fields have been presented in the recent literature (see [14]). Among many of his contributions to this topics, Ding proposed in [6] an efficient method to design a linear code from the support set $D = \{x \in \mathbb{F}_{2^m} : f(x) \neq 0\}$ of a Boolean function f in $\mathcal{B}_m$ using its univariate polynomial representation. We denote by n_f the size of D. Suppose that $D = \{d_1, d_2, \cdots, d_{n_f}\}$. Then Ding

defined a binary linear code $\mathcal{C}_D$ of length n_f as follows.

$$\mathcal{C}_D = \Big\{c_\alpha : \alpha \in \mathbb{F}_{2^m}\Big\}, c_\alpha = (\mathrm{Tr}_1^m(\alpha d_1), \mathrm{Tr}_1^m(\alpha d_2), \cdots, \mathrm{Tr}_1^m(\alpha d_{n_f})) \quad (1)$$

Theorem 1 ([7], Theorem 1)**.** *We keep the above notation. If $2n_f \neq -\widehat{\chi_f}(\alpha)$ for all $\alpha \in \mathbb{F}_{2^m}^*$, then $\mathcal{C}_D$ defined by (1) is a binary linear code with length n_f and dimension m, and its weight distribution is given by the following multiset: $\Big\{\Big\{\frac{2n_f+\widehat{\chi_f}(\alpha)}{4} : \alpha \in \mathbb{F}_{2^m}^*\Big\}\Big\} \cup \{\{0\}\}$.*

By choosing a basis of $\mathbb{F}_{2^m}$ over $\mathbb{F}_2$, $\mathbb{F}_{2^m}$ can then be viewed as an m-dimensional vector space over $\mathbb{F}_2$. Thus, each element of $\mathbb{F}_{2^m}$ can be identified with a binary row vector of length m. We now restate the above generic construction from the viewpoint of vector space. Let $D = \{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_n\} \subseteq \mathbb{F}_2^m$. The linear code $\mathcal{C}_D$ of length m over $\mathbb{F}_2$ defined from D is:

$$\mathcal{C}_D = \{(\mathbf{a}\cdot\mathbf{g}_1, \mathbf{a}\cdot\mathbf{g}_2, \cdots, \mathbf{a}\cdot\mathbf{g}_n) : \mathbf{a} \in \mathbb{F}_2^m\}, \quad (2)$$

where $\mathbf{a}\cdot\mathbf{g}_i$ is the dot product of $\mathbf{a}$ and $\mathbf{g}_i$. The set D is called the *defining set* of the resulting code $\mathcal{C}_D$.

3 A new construction method for designing LRC with locality 2 and multiple repair alternatives via autocorrelation spectra of Boolean functions

The following main result will play an important role in the rest of the paper.

Theorem 2. *Let f in $\mathcal{B}_m$ such that $f(0) = 0$ and $2n_f \neq -\widehat{\chi_f}(\alpha)$ for all $\alpha \in \mathbb{F}_{2^m}^*$. Then $\mathcal{C}_D$ is an $[n_f, m, d]$-LRC with $d = \min\{\frac{2n_f+\widehat{\chi_f}(\alpha)}{4} : \alpha \in \mathbb{F}_{2^m}^*\}$, $r = 2$, and $t = \min\Big\{\frac{4n_f+A_f(a)-2^m}{8} : a \in D\Big\}$, in which recall that n_f is the size of the support set of f and $A_f(a)$ is the autocorrelation function of f at point a.*

Proof. It follows from Theorem 1 that $\mathcal{C}_D$ has dimension m and minimum distance $\min\{\frac{2n_f+\widehat{\chi_f}(\alpha)}{4} : \alpha \in \mathbb{F}_{2^m}^*\}$. In the rest part of this proof. We will prove that $\mathcal{C}_D$ is a $(2,t)$-LRC with $t = \min\Big\{\frac{4n_f+A_f(a)-2^m}{8} : a \in D\Big\}$. Since $\mathrm{Tr}_1^m(\alpha x) = 0$ for all $\alpha \in \mathbb{F}_{2^m}^*$ if and only if $x = 0$, then for all $\alpha \in \mathbb{F}_{2^m}^*$, the i-th coordinate of the codeword c_α defined by (1) can be recovered from the subset $\{j_1, j_2\}$ if and only if $d_i + d_{j_1} + d_{j_2} = 0$, where $d_i, d_{j_1}, d_{j_2} \in D$. Hence, $\mathcal{C}_D$ is an $[n_f, m, d]$-LRC with $r = 2$, if for all $i \in \{1, \ldots, n_f\}$, there is at least one repair set of the i-th coordinate of $\mathcal{C}_D$ which has 2 elements. We now prove that $\mathcal{C}_D$ is an $(2,t)$-LRC with $t = \min\Big\{\frac{4n_f+A_f(a)-2^m}{8} : a \in D\Big\}$. From the above discussion, we know that for a given $i \in \{1, \ldots, n_f\}$, the i-th coordinate of $\mathcal{C}_D$ can recovered from the subset $\{j_1, j_2\}$ if and only if $d_i + d_{j_1} + d_{j_2} = 0$, where $d_i, d_{j_1}, d_{j_2} \in D$, which is equivalent to saying that, $d_{j_1} \in (d_i + D) \bigcap D$, where $d_i + D = \{d_i + d : d \in D\}$. Hence, in respect of the order of $\{j_1, j_2\}$, the number of disjoint repair sets of the i-th coordinate of $\mathcal{C}_D$, denoted by t_i, can be computed as

$$\begin{aligned} t_i &= \frac{1}{2}\#\left\{(j_1, j_2) : d_i + d_{j_1} + d_{j_2} = 0, d_i, d_{j_1}, d_{j_2} \in D\right\} \\ &= \frac{1}{2}\#\Big\{j_1 : d_{j_1} \in D \bigcap (d_i + D)\Big\} \\ &= \frac{1}{2}\#\Big(D \bigcap (d_i + D)\Big). \end{aligned}$$

It is not difficult to see that $\#\Big(D\bigcap(d_i+D)\Big)$ must be even, since $d_{j_1}\in D\bigcap(d_i+D)$ if and only if $d_{j_1}+d_i\in D\bigcap(d_i+D)$, and $d_{j_1}\neq d_{j_1}+d_i$ (note that $f(0)=0$ implies $d_i\neq 0$ for all $d_i\in D$). Since $\#D=\#d_i+D=n_f$, the autocorrelation function $A_f(d_i)$ can be written as

$$\begin{aligned}A_f(d_i)&=\sum_{x\in\mathbb{F}_{2^m}}(-1)^{f(x)+f(x+d_i)}\\&=\sum_{x\in D\bigcap(d_i+D)}(-1)^0+\sum_{x\in\Big(D\setminus(d_i+D)\Big)\bigcup\Big((d_i+D)\setminus D\Big)}(-1)^1+\sum_{x\in\mathbb{F}_{2^m}\setminus(D\bigcup(d_i+D))}(-1)^0\\&=2t_i+2(-1)(\#D-2t_i)+2^m-(2\#D-2t_i)\\&=8t_i-4n_f+2^m.\end{aligned}$$

Therefore, we have $t_i=\frac{4n_f+A_f(d_i)-2^m}{8}$, and thus $t=\min\Big\{\frac{4n_f+A_f(a)-2^m}{8}:a\in D\Big\}$. □

4 Binary LRCs with locality 2 and multiple repair alternatives from specific wide families of Boolean functions

4.1 LRCs from (bent) Boolean functions through the M-M constructions

First, recall that the *Hamming distance* between two Boolean functions f_1 and f_2 in $\mathcal{B}_m$ is equal to the weight of $f_1\oplus f_2$. The minimum distance between f in $\mathcal{B}_m$ and the set of all affine functions $l_b\oplus\epsilon$ ($b\in\mathbb{F}_2^m,\epsilon\in\mathbb{F}_2$), called the *nonlinearity* of f, is denoted by $\mathrm{nl}(f)$ and satisfies the relation $\mathrm{nl}(f)=2^{n-1}-\frac{1}{2}\max_{b\in\mathbb{F}_2^m}|\widehat{\chi_f}(b)|$. Because of Parseval's relation ([11]), it is upper bounded by $2^{m-1}-2^{m/2-1}$. This bound is tight for m even. Functions achieving the equality are called *bent* ([5, 17]). Bent functions are interesting combinatorial objects with many connections in many domains.(see [4],[3],[13]).

Lemma 1 ([17]). *A Boolean function f in $\mathcal{B}_m$ is bent if and only if $A_f(\omega)=0$ for any $\omega\in\mathbb{F}_2^{m*}$.*

As a first consequence of Theorem 2 using Lemma 1, we derive the following result highlighting that the bent functions allow the constructions of binary LRCs with locality 2 and multiple repair alternatives.

Corollary 1. *Let $f\in\mathcal{B}_m$ (where $m\geq 4$) be a bent function such that $f(0)=0$. Then $\mathcal{C}_D$ is an $[n_f,m,\frac{n_f}{2}-2^{\frac{m}{2}-2}]$-LRC with $r=2$ and $t=\frac{4n_f-2^m}{8}$.*

It is known that any bent function in m variables has Hamming weight $2^{m-1}-2^{\frac{m}{2}-1}$ or $2^{m-1}+2^{\frac{m}{2}-1}$. Then by Corollary 1 we can get $[2^{m-1}-2^{\frac{m}{2}-1},m,2^{m-2}-2^{\frac{m}{2}-1}]$-LRC with $(r,t)=(2,2^{m-2}-2^{\frac{m}{2}-2}-2^{m-3})$ and $[2^{m-1}+2^{\frac{m}{2}-1},m,2^{m-2}]$-LRC with $(r,t)=(2,2^{m-2}+2^{\frac{m}{2}-2}-2^{m-3})$.

The well-known class of Maiorana-McFarland (M-M) bent functions was discovered independently by Maiorana and McFarland (see [5, 12]), which includes a huge number of bent functions. The M-M construction produce bent functions indeed in $\mathcal{B}_m$ where $m=2k$ but it was generalized into a more general case as follows (see, e.g., [3]).

Construction 1. *Let m be a positive integer and s_1,s_2 be two positive integers such that $s_1+s_2=m$. Define a Boolean function $f\in\mathcal{B}_m$ as follows*

$$f(x,y)=\phi(x)\cdot y+g(x),\tag{3}$$

where $x\in\mathbb{F}_2^{s_1},y\in\mathbb{F}_2^{s_2}$, ϕ be an arbitrary mapping from $\mathbb{F}_2^{s_1}$ to $\mathbb{F}_2^{s_2}$, and g is an arbitrary Boolean function in s variables.

For a mapping ϕ from $\mathbb{F}_2^{s_1}$ to $\mathbb{F}_2^{s_2}$, we denote $\mathrm{Ker}(\phi) = \{x \in \mathbb{F}_2^{s_1} \mid \phi(x) = 0\}$, $\mathrm{Im}\phi = \{\phi(x) \mid x \in \mathbb{F}_2^{s_1}\}$, and $\phi(U) = \{\phi(x) \mid x \in U\}$ for a subset $U \subseteq \mathbb{F}_2^{s_1}$. Note that for $(a,b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$, the Walsh transform of f in (3) can be written as

$$\widehat{\chi_f}(a,b) = \sum_{x \in \mathbb{F}_2^{s_1}, y \in \mathbb{F}_2^{s_2}} (-1)^{\phi(x)\cdot y + g(x) + a\cdot x + b\cdot y} = \sum_{x \in \mathbb{F}_2^{s_1}} (-1)^{g(x)+a\cdot x} \sum_{y \in \mathbb{F}_2^{s_2}} (-1)^{(\phi(x)+b)\cdot y}.$$

Then, the following corollary is a direct consequence.

Corollary 2. *Let f be the function generated by Construction 1, then for any $(a,b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$ we have*

$$\widehat{\chi_f}(a,b) = \begin{cases} 2^{s_2} \sum\limits_{x \in \phi^{-1}(b)} (-1)^{g(x)+a\cdot x}, & b \in \mathrm{Im}\phi, \\ 0, & b \notin \mathrm{Im}\phi. \end{cases} \tag{4}$$

Furthermore, let U be a subset of $\mathbb{F}_2^{s_1}$ and V be a subset of $\mathbb{F}_2^{s_2}$. If ϕ is an injection from $\mathbb{F}_2^{s_1} \setminus U$ to $\mathbb{F}_2^{s_2} \setminus V$, then for $(a,b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$, we have

$$\widehat{\chi_f}(a,b) = \begin{cases} 2^{s_2} \sum\limits_{x \in \phi^{-1}(b)} (-1)^{g(x)+a\cdot x}, & b \in \phi(U), \\ 2^{s_2}(-1)^{g(\phi^{-1}(b))+a\cdot\phi^{-1}(b)}, & b \in \mathbb{F}_2^{s_2} \setminus V, \\ 0, & b \notin \mathrm{Im}\phi. \end{cases} \tag{5}$$

Theorem 3. *Let f be the function generated by Construction 1, where $g \equiv 0$, U is a subspace of $\mathbb{F}_2^{s_1}$. Define ϕ as a mapping from $\mathbb{F}_2^{s_1}$ to $\mathbb{F}_2^{s_2}$ satisfying ϕ is additive homomorphic from U to V, and ϕ is injective from $\mathbb{F}_2^{s_1} \setminus U$ to $\mathbb{F}_2^{s_2} \setminus V$. Then, for any $(a,b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$,*

$$\widehat{\chi_f}(a,b) = \begin{cases} (-1)^{a\cdot\phi^{-1}(b)} 2^{s_2} \#\mathrm{Ker}(\phi), & \text{if } a \in \mathrm{Ker}(\phi)^\perp, b \in \phi(U), \\ (-1)^{a\cdot\phi^{-1}(b)} 2^{s_2}, & \text{if } b \in \mathbb{F}_2^{s_2} \setminus V, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Proof. Since ϕ is additive homomorphic from U to V, we have that for any $b \in \phi(U)$, $\phi^{-1}(b) = u + \mathrm{Ker}(\phi)$, $u \in U$. For convenience, we denote by u_b the coset representative of $\phi^{-1}(b)$.

According to (5), we only need to consider the case for $b \in \phi(U)$. Suppose that $\mathrm{Ker}(\phi)$ has dimension w, then $\mathrm{Ker}(\phi) = \{\sum_{i=1}^{w} c_i\kappa_i \mid c_i \in \mathbb{F}_2, i = 1, \ldots, w\}$ for a basis $\{\kappa_1, \ldots, \kappa_w\}$ on $\mathbb{F}_2^{s_1}$. Let $\phi^{-1}(b) = u_b + \mathrm{Ker}(\phi)$, then for $a \in \mathbb{F}_2^{s_1}$ and $b \in \phi(U)$, we have

$$\begin{aligned} \widehat{\chi_f}(a,b) &= 2^{s_2} \sum_{x \in u_b + \mathrm{Ker}(\phi)} (-1)^{a\cdot x} \\ &= (-1)^{a\cdot u_b} 2^{s_2} \sum_{y \in \mathrm{Ker}(\phi)} (-1)^{a\cdot y} \\ &= (-1)^{a\cdot u_b} 2^{s_2} \sum_{c \in \mathbb{F}_2^w} (-1)^{a\cdot(\sum_{i=1}^{w} c_i\kappa_i)} \\ &= (-1)^{a\cdot u_b} 2^{s_2} \sum_{c \in \mathbb{F}_2^w} (-1)^{\sum_{i=1}^{w} c_i(a\cdot\kappa_i)} \\ &= (-1)^{a\cdot u_b} 2^{s_2} \sum_{c \in \mathbb{F}_2^w} (-1)^{c\cdot d} \\ &= \begin{cases} (-1)^{a\cdot u_b} 2^{s_2+w}, & \text{if } a \in \mathrm{Ker}(\phi)^\perp, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

where $d = (a \cdot \kappa_1, \ldots, a \cdot \kappa_w) \in \mathbb{F}_2^w$. Note that when $a \in \mathrm{Ker}(\phi)^\perp$, then $(-1)^{a \cdot u_b}$ is independent with the choice of u_b in the coset $\phi^{-1}(b)$, so we denote $u_b = \phi^{-1}(b)$ for convenience. The desired result follows. □

Recall that the defining set D of f is defined as the support set of f. Theorem 4 gives the autocorrelation values of functions from Construction 1. Due to the limit in space, the (long) proof of Theorem 4 was removed. It will be included in the full version.

Theorem 4. *Let f be the function generated by Construction 1, where $g \equiv 0$, U is a k-dimensional subspace of $\mathbb{F}_2^{s_1}$. Define ϕ as a mapping from $\mathbb{F}_2^{s_1}$ to $\mathbb{F}_2^{s_2}$ satisfying ϕ is additive homomorphic from U to V, and ϕ is injective from $\mathbb{F}_2^{s_1} \setminus U$ to $\mathbb{F}_2^{s_2} \setminus V$. Then, for any $(a, b) \in D$,*

$$A_f(a,b) \quad = \quad \begin{cases} 2^{s_2+k}, & \text{if } a \in \mathrm{Ker}(\phi) \setminus \{0\}, b \in \phi(U)^\perp, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The following result is obtained directly by combining Theorem 2, Theorem 3, and Theorem 4. It can be easily checked that Corollary 1 is a particular case of Theorem 5 below with $U = \{0\}$.

Theorem 5. *Let f be an m-variable ($m \geq 4$) M-M function generated by Construction 1, where $m = s_1 + s_2$, $g \equiv 0$, and U is a k-dimensional subspace of $\mathbb{F}_2^{s_1}$. Define ϕ as a mapping from $\mathbb{F}_2^{s_1}$ to $\mathbb{F}_2^{s_2}$ satisfying ϕ is additive homomorphic from U to V, $\#U \neq \#\mathrm{Ker}(\phi) < 2^{s_1-1}$ if $k \geqslant 1$, and ϕ is injective from $\mathbb{F}_2^{s_1} \setminus U$ to $\mathbb{F}_2^{s_2} \setminus V$. Then, $\mathcal{C}_D$ defined in (2) is an $[n_f, m, \frac{n_f}{2} - 2^{s_2-2}\#\mathrm{Ker}(\phi)]$-LRC with $r = 2$ and*

$$t = \begin{cases} \frac{4n_f - 2^{s_2+k} - 2^m}{8}, & \text{if } \mathrm{Ker}(\phi) \setminus \{0\} \neq \emptyset,\ \phi(U)^\perp \neq \emptyset, \\ \frac{4n_f - 2^m}{8}, & \text{otherwise.} \end{cases}$$

Proof. Since $g \equiv 0$ and ϕ is additive homomorphic from U to V, then $f(0,0) = \phi(0) \cdot 0 + 0 = 0$. It can be easily checked that since $\#\mathrm{Ker}(\phi) < 2^{s_1-1}$, then for any $(a, b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$,

$$\begin{aligned} 2n_f + \widehat{\chi_f}(a,b) &= 2^{s_1+s_2} - \widehat{\chi_f}(0,0) + \widehat{\chi_f}(a,b) \\ &= 2^{s_1+s_2} - 2^{s_2}\#\mathrm{Ker}(\phi) + \widehat{\chi_f}(a,b) \\ &\geqslant 2^{s_1+s_2} - 2^{s_2+1}\#\mathrm{Ker}(\phi) > 0. \end{aligned}$$

Hence, according to Theorem 2, we know that $\mathcal{C}_D$ is an $[n_f, m, d]$-LRC with $d = \min\{\frac{2n_f + \widehat{\chi_f}(a,b)}{4} : (a,b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}, (a,b) \neq (0,0)\}$, $r = 2$, and $t = \min\{\frac{4n_f + A_f(a,b) - 2^m}{8} : (a,b) \in D\}$. From Theorem 3 and Theorem 4, we only need to prove that there exists $(a, b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$ such that $\widehat{\chi_f}(a,b) = -2^{s_2}\#\mathrm{Ker}(\phi)$. If $U = \{0\}$, then we have $\#\mathrm{Ker}(\phi) = 1$, and thus $\widehat{\chi_f}(a,b) = -2^{s_2}$ for some $(a, b) \in \mathbb{F}_2^{s_1} \times \mathbb{F}_2^{s_2}$. If $\dim(U) \geqslant 1$, then we choose $a \in \mathrm{Ker}(\phi)^\perp \setminus U^\perp$, and thus for any $b \in \phi(U)$, $a \cdot \phi^{-1}(b) \neq 0$. The desired result is therefore deduced thanks to Theorem 3. □

4.2 LRCs from Boolean functions based on partial spreads

Spreads and partial spreads are fundamental objects in several fields, including the theory of (bent) Boolean functions. We emphasize below that they also play a role in constructing LRCs. Recall that a partial k-spread of the vector space $\mathbb{F}_2^m$ is a collection μ of k-dimensional subspaces $V_1, V_2, \cdots, V_s$ of $\mathbb{F}_2^m$ such that $V_i \cap V_j = \{\mathbf{0}\}$ for $1 \leq i \neq j \leq s$. Such a collection is called a spread if, in addition, $\bigcup_{i=1}^s V_i = \mathbb{F}_2^m$. Particularly, for $m = 2k$, a partial k-spread of $\mathbb{F}_2^m$ with $m = 2k$ is a set of pairwise supplementary of k-dimensional subspaces of $\mathbb{F}_2^m$. In this case, a k-spread of $\mathbb{F}_2^m$ can be easily obtained from the finite field $\mathbb{F}_{2^m}$, in which a k-dimensional subspace of $\mathbb{F}_2^m$ can be viewed as

an additive group of $\mathbb{F}_{2^m}$. Indeed, let α be a primitive element of $\mathbb{F}_{2^m}$ and $\gamma = \alpha^{2^k+1}$, then we can easily verify that $V_1, V_2, \cdots, V_{2^k+1}$ defined by $V_i = \{\alpha^{i-1}, \alpha^{i-1}\gamma, \alpha^{i-1}\gamma^2, \cdots, \alpha^{i-1}\gamma^{2^k-2}\} \cup \{0\}$ for any $1 \leq i \leq 2^k+1$ form a spread of $\mathbb{F}_{2^m}$. In the sequel, we consider Boolean functions with support constituted by partial spreads and the parameters of results in locally repairable codes.

Lemma 2. *Let $m = 2k \geq 4$ be an integer and $\Omega_s = \{V_1, V_2, \cdots, V_s\}$ be a partial k-spread of $\mathbb{F}_2^m$, where $2 \leq s \leq 2^k+1$. Let $f_s \in \mathcal{B}_m$ be the Boolean function with support $\Omega_s \setminus \{\boldsymbol{0}\}$, i.e., $\mathrm{supp}(f_s) = \Omega_s \setminus \{\boldsymbol{0}\}$. Then we have*

$$A_{f_s}(a) = \begin{cases} 2^m, & \text{if } a = \boldsymbol{0} \\ 2^m + 4s^2 - 2^{k+2}s - 8s + 2^{k+2}, & \text{if } a \in \mathrm{supp}(f_s) \\ 2^m + 4s^2 - 2^{k+2}s, & \text{if } a \in \mathbb{F}_2^{m*} \setminus \mathrm{supp}(f_s) \end{cases}$$

Proof. The well-known Wiener-Khintchine Theorem (see e.g., [2]) shows that for any m-variable Boolean function h and an arbitrary vector $a \in \mathbb{F}_2^m$, we have (where $\cdot$ denotes a scalar product in $\mathbb{F}_2^m$):

$$A_h(a) = 2^{-m} \sum_{u \in \mathbb{F}_2^m} \widehat{\chi_h}^2(u)(-1)^{u \cdot a}. \tag{8}$$

Therefore, it is sufficient to determine the values of $\widehat{\chi_{f_s}}^2(u)$ for all $u \in \mathbb{F}_2^m$. We now consider the values of $\widehat{\chi_{f_s}}(u)$, where $u \in \mathbb{F}_2^{m*}$, by considering $\widehat{\chi_{f_s}}(u) = -2\sum_{x \in \mathrm{supp}(f_s)}(-1)^{u \cdot x}$. Basically, our discussion is based on the fact that, for any $1 \leq i \leq s$, we have $\sum_{x \in V_i \setminus \{\boldsymbol{0}\}}(-1)^{u \cdot x} = 2^k - 1$ if $u \perp V_i$ and $\sum_{x \in V_i \setminus \{\boldsymbol{0}\}}(-1)^{u \cdot x} = -1$ otherwise. Note also that for any $1 \leq i \neq j \leq s$ if we have both $u \perp V_i$ and $u \perp V_j$ then u must be the all-zero vector, i.e., $u = \boldsymbol{0}$. Then we can straightforwardly obtain that

$$\widehat{\chi_{f_s}}(u) = \begin{cases} 2^m - 2s(2^k - 1), & \text{if } u = \boldsymbol{0} \\ -2^{k+1} + 2s, & \text{if } u \in \Omega_s' \\ 2s, & \text{if } u \notin \Omega_s' \end{cases}, \tag{9}$$

where $\Omega_s' = \bigcup_{i=1}^s V_i^\perp \setminus \{\boldsymbol{0}\}$, in which $V_i^\perp$ denotes the orthogonal subspace of V_i.

We are ready now to give the values of $A_{f_s}(a)$ for all $a \in \mathbb{F}_2^m$. Clearly, by the definition of autocorrelation function we can directly get $A_{f_s}(\boldsymbol{0}) = 2^m$. For any $a \in \mathbb{F}_2^m \setminus \{0\}$, two cases can occur.

Case A. $a \in \mathrm{supp}(f_s)$. Note that $\Omega_s' \cup \{0\}$ is also a partial k-spread of $\mathbb{F}_2^m$, constituted by s subspaces of $\mathbb{F}_2^m$. Since $a \in \mathrm{supp}(f_s) = \bigcup_{i=1}^s V_i \setminus \{\boldsymbol{0}\}$ and $\Omega_s' = \bigcup_{i=1}^s V_i^\perp \setminus \{\boldsymbol{0}\}$, there exists exact one V_j such that $a \perp V_j$, where $1 \leq j \leq s$. This implies that $\sum_{u \in \Omega_s'}(-1)^{a \cdot u} = (2^k - 1) + (-1) \cdot (s-1) = 2^k - s$. In addition, we have $\sum_{u \in \mathbb{F}_2^{m*} \setminus \Omega_s'}(-1)^{a \cdot u} = s - 2^k - 1$ since $\sum_{u \in \mathbb{F}_2^{m*}}(-1)^{a \cdot u} = -1$ for $a \neq \boldsymbol{0}$. Then by (8), we have

$$\begin{aligned} A_f(a) &= 2^{-m}\left(\left(2^m - 2s(2^k-1)\right)^2 + (-2^{k+1}+2s)^2 \sum_{u \in \Omega_s'}(-1)^{a \cdot u} + (2s)^2 \sum_{u \in \mathbb{F}_2^{m*} \setminus \Omega_s'}(-1)^{a \cdot u}\right) \\ &= 2^{-m}\left(\left(2^m - 2s(2^k-1)\right)^2 + (-2^{k+1}+2s)^2(2^k - s) + (2s)^2(s - 2^k - 1)\right) \\ &= 2^{-m}\left(2^{2m} + s^2 2^{m+2} - s2^{m+k+2} + 2^{m+k+2} - s2^{m+3}\right) \\ &= 2^m + 4s^2 - 2^{k+2}s - 8s + 2^{k+2}. \end{aligned}$$

Case B. $a \in \mathbb{F}_2^{m*} \setminus \mathrm{supp}(f_s)$. Similar to Case A above, we have $\sum_{u \in \Omega_s'}(-1)^{a \cdot u} = -s$ and

$\sum_{u\in\mathbb{F}_2^{m*}\setminus\Omega'_s}(-1)^{a\cdot u}=s-1$. Then by (8), we have

$$\begin{aligned} A_f(a) &= 2^{-m}\Big(\big(2^m-2s(2^k-1)\big)^2+(-2^{k+1}+2s)^2(-s)+(2s)^2(s-1)\Big)\\ &= 2^m+4s^2-2^{k+2}s. \end{aligned}$$

The assertion of theorem follows from the two cases above. This completes the proof. □

We are ready now to present the parameters of LRCs derived from the Boolean functions based on partial spreads.

Theorem 6. *Let $m=2k\geq 4$ be an integer and $\Omega_s=\{V_1,V_2,\cdots,V_s\}$ be a partial k-spread of $\mathbb{F}_2^m$, where $2\leq s\leq 2^k+1$. Let $f_s\in\mathcal{B}_m$ be the Boolean function with support $\Omega_s\setminus\{\boldsymbol{0}\}$. Then the binary linear code $\mathcal{C}_{D_s}$ defined by (2) is an $[(2^k-1)s,m,2^{k-1}(s-1)]$-LRC with $r=2$ and $t=\frac{4s^2-12s+2^{k+2}}{8}$.*

Proof. It can be easily seen that the linear code $\mathcal{C}_{D_s}$ has length $(2^k-1)s$. Then by Theorem 1 and (9) we obtain that $\mathcal{C}_{D_s}$ has dimension m. Finally, by Theorem 2 and Lemma 2 we immediately get that $r=2$ and $t=\frac{4s^2-12s+2^{k+2}}{8}$, which completes the proof. □

5 Conclusions

We have presented a new approach to design locally repairable codes (LRCs) with locality two and multiple repair alternatives by employing Boolean functions. Specifically, we focused on codes achieving minimum repair locality and maximum rate. We analyzed those codes using new tools and proposed an explicit construction method for designing them. A connection between LRCs (with multiple repair alternatives) and (the autocorrelation spectrum of) Boolean functions has been pointed out for the first time in this context (to the best of our knowledge), emphasizing notably a novel role of bent functions for designing LRCs.

References

[1] H. Cai, Y. Miao, M. Schwartz, and X. Tang. On optimal locally repairable codes with multiple disjoint repair sets. *IEEE Trans. Inf. Theory*, 66(4), pp. 2402–2416, 2019.

[2] C. Carlet. Partially-bent functions. *Designs, Codes and Cryptography*, 3(2):135–145, 1993.

[3] C. Carlet. Boolean Functions for Cryptography and Coding Theory. *Cambridge University Press*, Cambridge, U.K., 2021.

[4] C. Carlet and S. Mesnager. Four decades of research on bent functions. *Des. Codes Cryptogr.*, vol. 78, pp. 5-50, 2016.

[5] J.F. Dillon. Elementary Hadamard difference sets. *PhD thesis, Univ. of Maryland*, 1974.

[6] C. Ding. Linear codes from some 2-designs. *IEEE Trans. Inf. Theory*, 61(6), pp. 3265–3275, 2015.

[7] C. Ding. A construction of binary linear codes from Boolean functions. *Discrete Mathematics*, 339(9), pp. 2288–2303, 2016.

[8] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925–6934, 2012.

[9] C. Huang, H. Simitci, Y. Xu, A.Ogus, B.Calder, Parikshit Gopalan, J. Li, and S. Yekhanin. Erasure coding in windows azure storage. In *2012-USENIX Annual Technical Conference USENIX-ATC 12*, pp. 15–26, 2012.

[10] L. Jin, H. Kan, and Y. Zhang. Constructions of locally repairable codes with multiple recovering sets via rational function fields. *IEEE Trans. Inf. Theory*, 66(1), pp. 202–209, 2019.

[11] F.J. MacWilliams and N. J. A. Sloane. The theory of error-correcting codes, Vol 16. *Elsevier*, 1977.

[12] R. L McFarland. A family of difference sets in non-cyclic groups. *Journal of Combinatorial Theory, Series A*, 15(1), pp. 1–10, 1973.

[13] S. Mesnager. Bent Functions - Fundamentals and Results. *Springer*, Switzerland, 2016.

[14] S. Mesnager. Linear codes from functions. *A Concise Encyclopedia of Coding Theory* CRC Press/Taylor and Francis Group (Publisher) London, New York, 2021 (Chapter 20, 94 pages).

[15] L. Pamies-Juarez, H. D. L. Hollmann, and F. Oggier. Locally repairable codes with multiple repair alternatives. In *2013 IEEE international symposium on information theory*, pages 892–896, 2013.

[16] A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath. Locality and availability in distributed storage. *IEEE Trans. Inf. Theory*, 62(8), pp. 4481–4493, 2016.

[17] O. S Rothaus. On "bent" functions. *Journal of Combinatorial Theory, Series A*, 20(3) pp. 300–305, 1976.

[18] I. Tamo and A. Barg. A family of optimal locally recoverable codes. *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, 2014.

[19] I. Tamo, A. Barg, and A. Frolov. Bounds on the parameters of locally recoverable codes. *IEEE Trans. Inf. Theory*, 62(6), pp. 3070–3083, 2015.

[20] A. Wang and Z. Zhang. Repair locality with multiple erasure tolerance. *IEEE Trans. Inf. Theory*, 60(11), pp. 6979–6987, 2013.

[21] A. Wang and Z. Zhang. An integer programming based bound for locally repairable codes. *IEEE Trans. Inf. Theory*, 61(10), pp. 5280–5294, 2015.

[22] J. Wang, K. Shen, X. Liu, and C. Yu. Construction of binary locally repairable codes with optimal distance and code rate. *IEEE Communications Letters*, vol. 25, no. 7, pp. 2109 - 2113, 2021.

[23] Y. Zhang and H. Kan. Locally repairable codes from combinatorial designs. *Science China Information Sciences*, 63(2), pp. 1–15, 2020.