DISTRIBUTION OF R-PATTERNS IN THE KERDOCK-CODE BINARY SEQUENCES AND THE HIGHEST LEVEL SEQUENCES OF PRIMITIVE SEQUENCES OVER Z_{2^l}

HONGGANG HU,DENGGUO FENG

ABSTRACT. The distribution of r-patterns is an important aspect of pseudo-randomness for periodic sequences over finite field.The aim of this work is to study the distribution of r-patterns in the Kerdock-code binary sequences and the highest level sequences of primitive sequences over Z_{2^l}.By combining the local Weil bound with spectral analysis,we derive the upper bound of the deviation to uniform distribution.As a consequence,the recent result on the quantity is improved.

1. INTRODUCTION

Pseudorandom sequences with a variety of statistical properties are important in many areas of communications and cryptography. Hence the development of a good pseudorandom sequence generator is a hot topic.Many sequences with nice pseudorandom properties have been found,such as Bent sequences,No sequences and interleaved sequences(see [5] and the references therein).There are also some interesting binary sequences derived from the rings Z_{2^l}. And in this paper we will study such two kinds of binary sequences:the Kerdock-code binary sequences and the highest level sequences of primitive sequences over Z_{2^l}.

In [8] some families of binary sequences of period $T = 2^m - 1$ were constructed.We call them the Kerdock-code binary sequences.The family size is larger than the known ones.And it is approximately $T^{l-1}/2^{l-1}$.The 0,1 distribution is asymptotically uniform.The crosscorrelations and nontrivial autocorrelations is upper bounded by $0.19l^2(2^{l-1} - 1)\sqrt{T+1}$.When $l = 3$,the nonlinearity of these sequences is upper bounded by $3\sqrt{2+\sqrt{2}}\sqrt{T+1}$.Furthermore,the linear complexity is of order $O(m^4)$ which is much larger than that of the so-called Z_4-linear sequences.Hence these sequences might be an attractive alternative in applications such as CDMA communication and cryptography.In this paper we will show that the distribution of r-patterns in the Kerdock-code binary sequences is asymptotically uniform.

The highest level sequences of primitive sequences over Z_{2^l} were introduced in the last century motivated by the potential cryptographic applications(see [1] and the references therein).In [1],it was proved that the period of the highest level sequence is $T = 2^{l-1}(2^m - 1)$ and the lower bound on the linear complexity is large.In [4] the authors revisted the highest level sequence.And they proved that the 0,1 distribution is asymptotically uniform and the absolute value of the autocorrelation function $C_T(\tau)$ is bounded by $2^{l-1}(2^{l-1} - 1)\sqrt{3(2^{2l} - 1)}2^{n/2} + 2^{l-1}$ for $\tau \neq 0$.By combining the local Weil bound with spectral analysis(I think the origin

Key words and phrases. r-pattern,exponential sum over Galois ring,Kerdock-code binary sequence,highest level sequence.

idea comes from [7, 10]),the two results were improved in [11].In [2, 3] the distribution of r-patterns in the highest level of p-adic sequences over Galois rings was investigated.By the similar method,we will give a new estimate on the deviation to uniform distribution in this paper.And the result improves the known ones in [2, 3].

The organization of this paper is as follows.In section 2 we will point out some basic knowledge of Galois ring $GR(2^l, m)$ and Fourier transformation on Z_{2^l} needed in the following sections.In section 3 the distribution of r-patterns in the Kerdock-code binary sequences will be investigated.And in section 4 the distribution of r-patterns in the highest level sequences of primitive sequences over Z_{2^l} will be investigated.Finally,section 5 concludes the paper.

2. Preliminaries

2.1. **Galois ring of characteristic 2^l.** The Galois ring $GR(2^l, m)$ is the unique Galois extension of degree m over Z_{2^l}.It is a ring of characteristic 2^l with 2^{lm} elements.And it is also a local ring with unique maximal ideal $2GR(2^l, m)$ and residue field F_{2^m}.Thus the set $GR(2^l, m)^*$ of units is $GR(2^l, m) \setminus 2GR(2^l, m)$.$GR(2^l, m)^*$ is a multiplicative group with the following group structure:

$$GR(2^l, m)^* \cong Z_{2^m-1} \times \underbrace{Z_{2^{l-1}} \times ... \times Z_{2^{l-1}}}_{m\ copies}$$

if $l = 2$;

$$GR(2^l, m)^* \cong Z_{2^m-1} \times Z_2 \times Z_{2^{l-2}} \times \underbrace{Z_{2^{l-1}} \times ... \times Z_{2^{l-1}}}_{m-1\ copies}$$

if $l \geq 3$.

The Teichmüller set Γ of $GR(2^l, m)$ is $\{0, 1, \xi, \xi^2, ..., \xi^{2^m-2}\}$,where ξ is a primitive $(2^m - 1)$th root of unity.Each element $x \in GR(2^l, m)$ has a unique 2-adic representation

$$x = x_0 + 2x_1 + ... + 2^{l-1}x_{l-1}$$

where $x_0, x_1, ..., x_{l-1} \in \Gamma$.

The Frobenius automorphism from $GR(2^l, m)$ to $GR(2^l, m)$ acts as follows:

$$F(x) = x_0^2 + 2x_1^2 + ... + 2^{l-1}x_{l-1}^2.$$

F fixes only elements of Z_{2^l},and generates the Galois group of $GR(2^l, m)/Z_{2^l}$,which is a cyclic group of order m.The trace map Tr from $GR(2^l, m)$ to Z_{2^l} is defined by

$$Tr(x) = x + F(x) + ... + F^{m-1}(x).$$

Borrowing the notation in [11],let MSB denotes the most significant bit map,i.e.,

$$MSB(x_0 + 2x_1 + ... + 2^{l-1}x_{l-1}) = x_{l-1}.$$

2.2. **Exponential sum over Galois ring.** The canonical additive character ψ over Z_{2^l} is defined by $\psi(x) = e^{2\pi ix/2^l}, \forall x \in Z_{2^l}$.For any $\beta \in GR(2^l, m)$,the additive character Ψ_β over $GR(2^l, m)$ is defined by $\Psi_\beta(x) = (\psi \circ Tr)(\beta x) = e^{2\pi iTr(\beta x)/2^l}$.When $\beta = 1, \Psi_\beta$ is the canonical additive character ψ over $GR(2^l, m)$.

The following lemma is contained in [11].And it follows easily from the Weil exponential sum over Galois ring(see [6, 9]).

Lemma 2.1. *For any $\lambda \in GR(2^l, m), \lambda \neq 0$, we have*

$$|\sum_{x\in\Gamma} \Psi_\lambda(x)| \leq (2^{l-1} - 1)\sqrt{2^m}.$$

2.3. **Fourier transformation on Z_{2^l}.** For any $k \in Z_{2^l}$,we denote by ψ_k the additive character over Z_{2^l}

$$\psi_k(x) = e^{2\pi i k x/2^l}, \forall x \in Z_{2^l}.$$

Let μ be the map from Z_{2^l} to $\{\pm 1\}$ such that $\mu(x) = (-1)^c$ where c is the most significant bit of $t \in Z_{2^l}$,i.e.,it maps $0, 1, ..., 2^{l-1}-1$ to $+1$ and $2^{l-1}, 2^{l-1}+1, ..., 2^l-1$ to -1.We can express the map μ as a linear combination of characters:

$$\mu = \Sigma_{j=0}^{2^l-1} \mu_j \psi_j,$$

where $\mu_j = \frac{1}{2^l}\sum_{x=0}^{2^l-1} \mu(x)\psi_j(-x)$.

The next lemma is the corollary 14 of [8].

Lemma 2.2. *With the notations as above,for any $l \geq 4$,the following estimate holds:*

$$\Sigma_{j=0}^{2^l-1}|\mu_j| < \frac{2l\ln(2)}{\pi} + 1.$$

Finally,we will give the definition of r-pattern.

Definition 2.3. Suppose that $(c_t)_{t=0}^{\infty}$ is a binary sequence with period T.$\forall(\overline{z}, \overline{s}) \in F_2^r \times [0, T)^r$ is called a r-pattern of $(c_t)_{t=0}^{\infty}$, where $\overline{s} = (s_1, s_2, ..., s_r), 0 \leq s_1 < s_2 < ... < s_r < T$.

Throughout the following sections,let ξ be the generator of the multiplicative group $\Gamma^* = \Gamma \setminus \{0\}$.

3. r-patterns in the Kerdock-code binary sequences

First,we define a cyclic code as:

$$S_m = \{(Tr(\lambda\xi^t))_{t=0}^{2^m-2} | \lambda \in GR(2^l, m)\}.$$

From which we can get a binary code s_m as $s_m = MSB(S_m)$.

Any Kerdock-code binary sequence $(c_t)_{t=0}^{\infty}$ with period $T = 2^m-1$ can be defined as $c_t = MSB(Tr(\lambda\xi^t))$,where $\lambda \in GR(2^l, m)^*$.

The following theorem is the main result on the distribution of r-patterns in the Kerdock-code binary sequences.

Theorem 3.1. *Suppose that $(c_t)_{t=0}^{\infty}$ is a Kerdock-code binary sequence with period $T = 2^m - 1$.And it is defined as above.$(\overline{z}, \overline{s})$ is a r-pattern of $(c_t)_{t=0}^{\infty}$,where $\overline{z} = (z_1, z_2, ..., z_r) \in F_2^r, \overline{s} = (s_1, s_2, ..., s_r), 0 \leq s_1 < s_2 < ... < s_r < T$.Let $N_{(\overline{z},\overline{s})}$ denotes the number of $(\overline{z}, \overline{s})$ in one cycle of $(c_t)_{t=0}^{\infty}$.If $\xi^{s_1}, \xi^{s_2}, ..., \xi^{s_r}$ is linear independent,we have*

$$|N_{(\overline{z},\overline{s})} - \frac{2^m-1}{2^r}| < [(\frac{l\ln(2)}{\pi} + 1)^r - \frac{1}{2^r}][(2^{l-1} - 1)\sqrt{2^m} + 1].$$

Proof. By the similar method in the proof of theorem 2 in [4],we know

$$N_{(\overline{z},\overline{s})} = \sum_{t=0}^{2^m-2} \frac{1}{2} \sum_{d_1=0,1} (-1)^{d_1(c_{t+s_1}+z_1)} ... \frac{1}{2} \sum_{d_r=0,1} (-1)^{d_r(c_{t+sr}+z_r)}$$

$$= \frac{1}{2^r} \sum_{d_1,d_2,...,d_r=0,1} (-1)^{d_1 z_1 + ... + d_r z_r} \sum_{t=0}^{2^m-2} (-1)^{d_1 c_{t+s_1} + ... + d_r c_{t+sr}}.$$

Thus

$$N_{(\overline{z},\overline{s})} - \frac{2^m-1}{2^r} = \frac{1}{2^r} \sum_{(d_1,...,d_r) \neq (0,...,0)} (-1)^{d_1 z_1 + ... + d_r z_r} \sum_{t=0}^{2^m-2} (-1)^{d_1 c_{t+s_1} + ... + d_r c_{t+sr}}.$$

$$|N_{(\overline{z},\overline{s})} - \frac{2^m-1}{2^r}| \leq \frac{1}{2^r} \sum_{(d_1,...,d_r) \neq (0,...,0)} |\sum_{t=0}^{2^m-2} (-1)^{d_1 c_{t+s_1} + ... + d_r c_{t+sr}}|.$$

Now we will give a estimate on $|\sum_{t=0}^{2^m-2}(-1)^{d_1 c_{t+s_1}+...+d_r c_{t+sr}}|$.Suppose there are exactly k elements among $d_1, ..., d_r$ are 1,and the others are $0,1 \leq k \leq r$.Without loss of generality,let $d_1 = 1, ..., d_k = 1, d_{k+1} = 0, ..., d_r = 0$.Since $\mu = \Sigma_{j=0}^{2^l-1} \mu_j \psi_j$,we have

$$|\sum_{t=0}^{2^m-2} (-1)^{d_1 c_{t+s_1} + ... + d_r c_{t+sr}}| = |\sum_{t=0}^{2^m-2} (-1)^{c_{t+s_1} + ... + c_{t+s_k}}|$$
$$= |\sum_{j_1=0}^{2^l-1} ... \sum_{j_k=0}^{2^l-1} \mu_{j_1} ... \mu_{j_k} \sum_{t=0}^{2^m-2} \Psi_\beta(\xi^t)|.$$

Here $\beta = \lambda(j_1 \xi^{s_1} + ... + j_k \xi^{s_k})$.$\beta \neq 0$ as $\lambda \in GR(2^l, m)^*$,and $\xi^{s_1}, \xi^{s_2}, ..., \xi^{s_r}$ is linear independent.Thus

$$|\sum_{t=0}^{2^m-2} (-1)^{d_1 c_{t+s_1} + ... + d_r c_{t+sr}}| \leq (\sum_{j=0}^{2^l-1} |\mu_j|)^k |\sum_{t=0}^{2^m-2} \Psi_\beta(\xi^t)|$$
$$< (\frac{2l \ln(2)}{\pi} + 1)^k [(2^{l-1} - 1)\sqrt{2^m} + 1].$$

Finally,we have

$$|N_{(\overline{z},\overline{s})} - \frac{2^m-1}{2^r}| < \frac{1}{2^r} \sum_{k=1}^{r} \binom{r}{k} (\frac{2l \ln(2)}{\pi} + 1)^k [(2^{l-1} - 1)\sqrt{2^m} + 1]$$
$$= [(\frac{l \ln(2)}{\pi} + 1)^r - \frac{1}{2^r}][(2^{l-1} - 1)\sqrt{2^m} + 1].$$

This completes the proof. □

Remark 3.2. When r=1,this theorem is almost the same as Theorem 5 in [8].

Let $f_{(\overline{z},\overline{s})}$ denotes the proportion of $(\overline{z},\overline{s})$ in one cycle of $(c_t)_{t=0}^{\infty}$,we have the following corollary.

Corollary 3.3. With the notations as above,we have

$$|f_{(\overline{z},\overline{s})} - \frac{1}{2^r}| < [(\frac{l \ln(2)}{\pi} + 1)^r - \frac{1}{2^r}][(2^{l-1} - 1)\sqrt{2^m} + 1]/(2^m - 1)$$
$$\approx C_l / \sqrt{2^m},$$

where C_l is a constant in l of order $l^r 2^l$.

Hence when $m \to +\infty, f_{(\overline{z},\overline{s})} \to \frac{1}{2^r}$,for any r-pattern $(\overline{z},\overline{s})$ which satisfies the condition in Theorem 3.1.

4. R-PATTERNS IN THE HIGHEST LEVEL SEQUENCES OF PRIMITIVE SEQUENCES OVER Z_{2^l}

Let $T = 2^{l-1}(2^m - 1)$.Any primitive sequences $(a_t)_{t=0}^{\infty}$ over Z_{2^l} has the well known trace deccription:

$$a_t = Tr(\alpha\gamma^t),$$

where $\alpha \in GR(2^l, m)^*, \gamma = \xi(1 + 2\xi_1), \xi_1 \in GR(2^l, m)^*$. Therefore the highest level sequences $(c_t)_{t=0}^{\infty}$ of primitive sequences over Z_{2^l} has the following description:

$$c_t = MSB(Tr(\alpha\gamma^t)).$$

The following theorem is the main result on the distribution of r-patterns in highest level sequences of primitive sequences over Z_{2^l}.

Theorem 4.1. *Suppose that $(c_t)_{t=0}^{\infty}$ is a highest level sequence of primitive sequences over Z_{2^l} with period $T = 2^{l-1}(2^m - 1)$.$(\overline{z}, \overline{s})$ is a r-pattern of $(c_t)_{t=0}^{\infty}$,where $\overline{z} = (z_1, z_2, ..., z_r) \in F_2^r, \overline{s} = (s_1, s_2, ..., s_r), 0 \le s_1 < s_2 < ... < s_r < T$.Let $N_{(\overline{z},\overline{s})}$ denotes the number of $(\overline{z}, \overline{s})$ in one cycle of $(c_t)_{t=0}^{\infty}$.If $\gamma^{s_1}, \gamma^{s_2}, ..., \gamma^{s_r}$ is linear independent,we have*

$$|N_{(\overline{z},\overline{s})} - \frac{2^{l-1}(2^m - 1)}{2^r}| < 2^{l-1}[(\frac{l\ln(2)}{\pi} + 1)^r - \frac{1}{2^r}][(2^{l-1} - 1)\sqrt{2^m} + 1].$$

Proof. (sketch)Note that for any $\beta \neq 0$

$$\begin{aligned}|\Sigma_{t=0}^{T-1}\Psi_\beta(\gamma^j)| &= |\Sigma_{t=0}^{2^{l-1}-1}\Sigma_{x\in\Gamma^*}\Psi_{\beta(1+2\lambda)^t}(x)| \\ &\le 2^{l-1}[(2^{l-1} - 1)\sqrt{2^m} + 1].\end{aligned}$$

Then by the same method in the proof of Theorem 3.1,we will get the upper bound on $|N_{(\overline{z},\overline{s})} - \frac{2^{l-1}(2^m-1)}{2^r}|$. □

Remark 4.2. When r=1,this theorem is the same as Theorem 3.3 in [11].

Let $f_{(\overline{z},\overline{s})}$ denotes the proportion of $(\overline{z}, \overline{s})$ in one cycle of $(c_t)_{t=0}^{\infty}$.It is interesting that the following corollary is the same as corollary 3.3.

Corollary 4.3. With the notations as above,we have

$$\begin{aligned}|f_{(\overline{z},\overline{s})} - \frac{1}{2^r}| &< [(\frac{l\ln(2)}{\pi} + 1)^r - \frac{1}{2^r}][(2^{l-1} - 1)\sqrt{2^m} + 1]/(2^m - 1) \\ &\approx C_l/\sqrt{2^m},\end{aligned}$$

where C_l is a constant in l of order $l^r 2^l$.

The estimate in [2, 3] is of the same shape with respect to m.But the constant C_l is of order $2^{(r+1)l}$.So our estimate is more sharp with respect to l.

5. CONCLUDING REMARKS

The distribution of r-patterns in the Kerdock-code binary sequences and the highest level sequences of primitive sequences over Z_{2^l} is studied in this paper.By combining the local Weil bound with spectral analysis,we derive the upper bound of the deviation to uniform distribution.And the results show that they are asymptotically uniform.As a consequence,the recent result on the the highest level sequences of primitive sequences over Z_{2^l} is improved.

Moreover,suppose a binary sequence $(c_t)_{t=0}^{\infty}$ is defined by

$$c_t = MSB(Tr(\alpha\gamma^t))$$

where $\alpha \in GR(2^l, m)^*, \gamma = \xi(1 + 2^i\xi_1), \xi_1 \in GR(2^l, m)^*, 1 < i < l$.Then the period of $(c_t)_{t=0}^{\infty}$ is $2^{l-i}(2^m - 1)$.And it is easy to check that the distribution of r-patterns in $(c_t)_{t=0}^{\infty}$ is also asymptotically uniform.

References

[1] Z.D.Dai,Binary sequences derived from ML-sequences over rings I: Periods and minimal polynomials,J.Cryptology,vol.5, pp. 193-207,1990.

[2] Z.D.Dai,Y.Dingfeng,W.Ping,and F.Genxi,Distribution of r-patterns in the highest level of p-adic sequences over Galois rings,in Proc. Golomb Symp., Univ. Southern California, Los Angeles, CA, 2002.

[3] Z.D.Dai,Y.Dingfeng,W.Ping,and F.Genxi,Distribution of r-grams in the highest level of p-adic expabsion of MLS over Galois rings and the asymptotical uniformity,in Proc. ChinaCrypt'2002,Beijing:Publishing House of Electronics Industry, 2002.

[4] S.Fan and W.Han,Random properties of the highest level sequences of primitive sequences over,IEEE Trans. Inform. Theory, vol. 49, pp. 1553-1557, June 2003.

[5] T.Helleseth and P.V.Kumar,Sequences with low correlation,in Handbook of Coding Theory, V.S.Pless and W.C.Huffman, Eds. Amsterdam, The Netherlands: North-Holland, 1998, vol.II, pp. 1765-1853.

[6] T.Helleseth,P.V.Kumar,and A.G.Shanbhag,Exponential sums over Galois rings and their applications,in Finite Fields and Applications. ser. Lecture Notes Series 233, S.D.Cohen and H. Nierreiter, Eds. Cambridge, U.K.: Cambridge Univ.Press,1996, pp. 109-128.

[7] J.Lahtonen,On the odd and the aperiodic correlation properties of the binary Kasami sequences,IEEE Trans.Inform.Theory, vol.41, pp.1506-1508,Sept.1995.

[8] Jyrki Lahtonen, San Ling, Patrick Sol and Dmitrii Zinoviev,Z8-Kerdock codes and pseudorandom binary sequences, Journal of Complexity,Volume 20,Issues 2-3(April-June 2004),Pages 318-330

[9] P.V.Kumar, .Helleseth,and A.R.Calderbank,An upper bound for Weil exponential sums over Galois rings and applications,IEEE Trans. Inform. Theory,vol.41, pp.456-468, Mar.1995.

[10] D.V.Sarwate,An upper bound on the aperiodic autocorrelation function for a maximal-length sequence,IEEE Trans. Znform Theory, vol.IT-30,pp.685-687,1984.

[11] P.Sole,D.Zinoviev,The Most Significant Bit of Maximum-Length Sequences Over Z_{2^l}: Autocorrelation and Imbalance,IEEE Transactions on Information Theory,Volume 50,Issue 8,Aug.2004,Page(s): 1844-1846

State Key Laboratory of Information Security(Graduate School of Chinese Academy of Sciences),Beijing,100039

Institute of Electronics,Chinese Academy of Sciences,Beijing,100080
E-mail address: hghu@ustc.edu