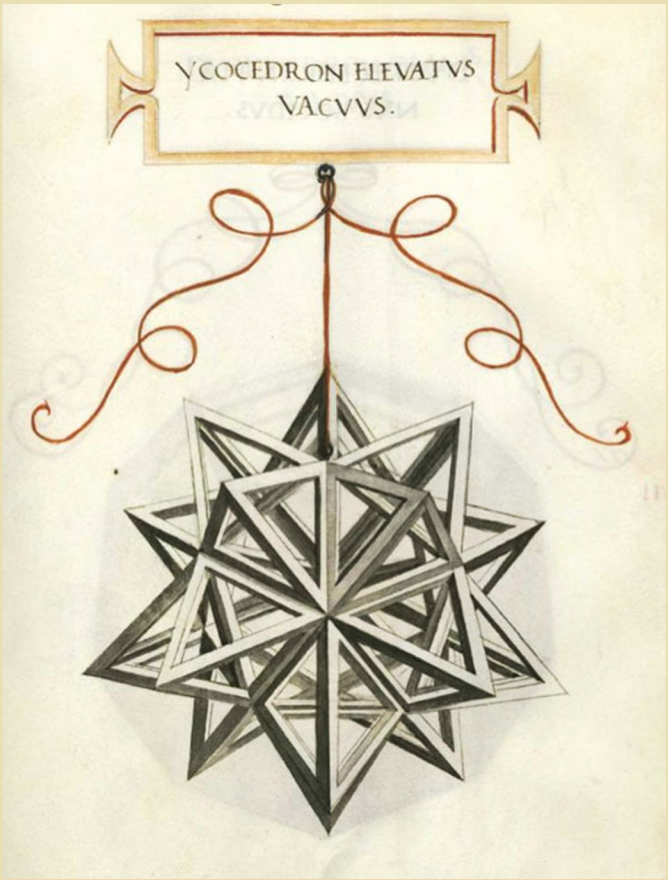
Colloquium on Combinatorial Designs

2020.9.26 8:30-12:00

https://meeting.tencent.com/s/3oOImSk5NwWR

ID: 572 282 353



Invited Speakers

Shuxing Li (Simon Fraser University)

Han Mao Kiah (Nanyang Technological University)

Xin Wang (Soochow University)

Organisers: Tao Feng, Xiande Zhang, Yue Zhou

Meeting Information

2020.9.26, Saturday

Tencent meeting ID：572 282 353

https://meeting.tencent.com/s/3oOImSk5NwWR

Program

8:30-9:30	**Shuxing Li**
	Intersection Distribution and Its Application
9:30-10:30	**Han Mao Kiah**
	Balancing a la Knuth for DNA-based Data Storage
10:30-11:30	**Xin Wang**
	New Upper Bounds for Wide-Sense Frameproof Codes

Intersection Distribution and Its Application

Shuxing Li
Simon Fraser University

Abstract

Given a polynomial f over finite field $\mathbb{F}_q$, its intersection distribution concerns the collective behaviour of a series of polynomials $\{f(x) + cx | c \in \mathbb{F}_q\}$. Each polynomial f canonically induces a $(q+1)$-set S_f in the classical projective plane $PG(2, q)$ and the intersection distribution of f reflects how the point set S_f interacts with the lines in $PG(2, q)$.

Motivated by the long-standing open problem of classifying oval monomials, which are monomials over $\mathbb{F}_{2^m}$ having the same intersection distribution as x^2, we consider the next simplest case: classifying monomials over $\mathbb{F}_q$ having the same intersection distribution as x^3. Some characterizations of such monomials are derived and consequently a conjectured complete list is proposed.

Among the conjectured list, we identify two exceptional families of monomials over $\mathbb{F}_{3^m}$. Interestingly, new examples of Steiner triple systems follow from them, which are nonisomorphic to the classical ones.

This is joint work with Gohar Kyureghyan and Alexander Pott.

References

[1] G. Kyureghyan, S. Li, and A. Pott. On the intersection distribution of degree three polynomials and related topics, arXiv:2003.10040, submitted.

[2] S. Li, and A. Pott. Intersection distribution, non-hitting index and Kakeya sets in affine planes, *Finite Fields and Their Applications*, 2020.

Balancing a la Knuth for DNA-based Data Storage

Han Mao Kiah
Nanyang Technological University, Singapore

The imbalance of a binary word refers to the absolute difference between the number of ones and the number of zeros in it. A word is balanced if its imbalance is at most one and a code is balanced if all its codewords are balanced. In 1986, motivated by applications in optical disks, Knuth [1] proposed a simple and efficient scheme that transforms binary messages into balanced codes.

In recent years, advances in synthesis and sequencing technologies have made DNA macromolecules an attractive medium for digital information storage. Besides being biochemically robust, DNA strands offer ultrahigh storage densities of 10^15 - 10^20 bytes per gram of DNA, as demonstrated in recent experiments (see [2, Table 1]). However, this non-traditional media presents new challenges to the coding community (see [3] for a survey). One particular challenge has rekindled interest in balanced codes. Specifically, a DNA string comprises four bases or letters: A, C, T and G, and a string is GC-rich (or GC-poor) if a high (or low) proportion of the bases corresponds to either G or C. Since GC-rich or GC-poor DNA strings are prone to both synthesis and sequencing errors, we aim to reduce the difference with the number of G and C and the number of A and T on every DNA codeword. This requirement is equivalent to reducing the imbalance of a related binary word.

In this talk, we revisit Knuth's balancing method and look at techniques that adapt the method to address coding problems in DNA-based data storage.

In the first part, we look at constructions of address sequences that are critical in DNA-based data storage with random-access capabilities. Specifically, Yazdi et al. [2] developed an architecture that allows selective access to encoded DNA strands through the process of hybridization. The technique involves prepending information-carrying DNA strands with specially chosen address sequences called primers. By combining Knuth's balancing method with cyclic codes, we provide efficient and explicit constructions of such primer sets [4].

In the second part, in addition to the GC-balanced constraints, we look at codes that correct a single insertion or deletion or substitution and whose codewords obey the homopolymer runlength constraints. Besides the code constructions, we also propose linear-time encoders for our codebooks [5].

In the third part, we apply Knuth's balancing method to a beautiful and important class of codes, the polar codes. Invented by Arıkan [6], polar codes achieve capacity for many channels with low encoding and decoding complexities. We study a generalization of Knuth's balancing method, specifically, a technique of Mazumdar, Roth, and Vontobel [7], and provide means of transforming messages into balanced polar codewords while retaining the low complexities of the polar encoding and decoding algorithms.

References

[1] D. Knuth, "Efficient balanced codes," IEEE Transactions on Information Theory, vol. 32, no. 1, pp. 51–53, 1986.

[2] S. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free DNA-based data storage," Scientific reports, vol. 7, no. 1, p. 5011, 2017.

[3] S. Yazdi, H. M. Kiah, E. R. Garcia, J. Ma, H. Zhao, and O. Milenkovic, "DNA-based storage: Trends and methods," IEEE Trans. Molecular, Biological, Multi-Scale Communications, vol. 1, no. 3, pp. 230–248, 2015.

[4] Y. M. Chee, H. M. Kiah, and H. Wei, "Efficient and explicit balanced primer codes," IEEE Transactions on Information Theory, vol. 66, no. 9, pp. 5344--5357, doi:10.1109/TIT.2020.2977915, 2020.

[5] K. Cai, Y. M. Chee, R. Gabrys, H. M. Kiah, T. T. Nguyen, "Optimal Codes Correcting a Single Indel / Edit for DNA-Based Data Storage", preprint arXiv:1910.06501 (see also doi:10.1109/ISIT.2019.8849643, doi:10.1109/ICASSP40776.2020.9053256), 2019.

[6] E. Arıkan. "Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels", IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3051–3073, 2009.

[7] A. Mazumdar, R. M. Roth, and P. O. Vontobel, "On linear balancing sets," Advances in Mathematics of Communications, vol. 4, no. 3,pp. 345–361, 2010.

[8] U. Gupta, H. M. Kiah, A. Vardy, H. Yao, “Polar Codes with Balanced Codewords,” in 2020 IEEE International Symposium on Information Theory (ISIT), doi:10.1109/ISIT44484.2020.9174042, 2020.

New Upper Bounds for Wide-Sense Frameproof Codes

Xin Wang

Soochow University

xinw@suda.edu.cn

Abstract

Let Q be a finite alphabet of size q. Given a t-subset $X = \{x^1, x^2, \ldots, x^t\} \subset Q^n$. Denote x_i^j as the i-th component of x^j for $1 \leq i \leq n$ and $1 \leq j \leq t$, a position i is called undetectable for X if the values of the words of X match in their i-th position: $x_i^1 = x_i^2 = \cdots = x_i^t$. The wide-sense descendant set of X is defined by

$$wdesc(X) = \{y \in Q^n | y_i = x_i^1, i \in U(X)\},$$

where $U(X)$ is the set of undetectable positions for X.

A code $\mathcal{C} \subset Q^n$ is an (n, q) wide-sense t-frameproof code, if

$$wdesc(X) \cap \mathcal{C} = X$$

for all $X \subset \mathcal{C}$ with $|X| \leq t$.

The upper bounds for Wide-Sense Frameproof Codes are closely related to Sperner families and intersecting families in extremal set theory [2, 4, 5]. In this talk, we give new upper bounds based on entropy method to improve the previous results asymptotically.

Open problem

1. The explicit construction of size exponential in n.
2. To improve the probabilistic lower bound of wide-sense frameproof code.
3. The gaps between the lower bounds and upper bounds are quite large even for the case of binary (wide-sense) 2-frameproof codes.
4. The extremal structure of wide-sense frameproof codes for small length.

References

[1] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, IEEE Trans. Inf. Theory, vol. 44, no. 5, pp. 1897-1905, 1998.

[2] D. R. Stinson and R. Wei, Combinatorial properties and constructions of traceability schemes and frameproof codes, SIAM Journal on Discrete Mathematics, no. 11, pp. 41-53, 1998.

[3] S. R. Blackburn, Combinatorial schemes for protecting digital content, Surveys in combinatorics, 2003.

[4] A. Panoui, Wide-sense fingerprinting codes and honeycomb arrays, PHD thesis, 2012.

[5] J. Zhou and W. Zhou, Wide-sense 2-frameproof codes, Des. Codes Cryptogr., online.