# Computational Approaches to Open Problems in Combinatorics

Ilias S. Kotsireas



Wilfrid Laurier University Waterloo ON, Canada

ikotsire@wlu.ca

Joint work with:

Vijay Ganesh, Curtis, Bright, Albert Heinle, University of Waterloo

## **Summary and motivation**

- Combinatorics abounds with very hard problems featuring 100s/1000s of binary and ternary variables (ditto for Cryptography)
- Several important classes of such problems: defined via the concept of **autocorrelation**
- For many of these classes, traditional algorithmic methodologies exhibit saturation
- It is desirable to develop new algorithmic methodologies via **cross-fertilization** with other disciplines
- The development of **SAT solvers** has experienced tremendous growth in recent decades and continues to be a very active area of research
- The uncanny realization that autocorrelation-defined problems can be encoded as SAT problems, opens the door for a very fruitful interaction between the two areas



Hadamard matrices are  $n \times n$  matrices H with  $\pm 1$  elements such that  $H \cdot H^t = nI_n$ . trivial cases: n = 1 and n = 2.

well-known necessary condition:  $n \equiv 0 \pmod{4}$ 

the sufficiency of this condition is the celebrated Hadamard conjecture (1893)

"There exists a Hadamard matrix of order n, for every  $n \equiv 0 \pmod{4}$ "

smallest unresolved order until 1985: 268 smallest unresolved order until 2004: 428 smallest unresolved order until 2013: 668 3 unresolved cases < 1000: 668, 716, 892 10 unresolved cases < 2000: 1004, 1132, 1244, 1388, 1436, 1676, 1772, 1916, 1948, 1964 List of integers v < 500 for which no Hadamard matrices of order 4v are known consists of 13 integers:

167, 179, 223, 251, 283, 311, 347, 359, 419, 443, 479, 487, 491

all of them primes congruent to  $3 \pmod{4}$ .

**NEW RESULT**: Construction of a Hadamard matrix of order  $4 \cdot 251$  Djokovic, Golubitsky, Kotsireas, JCD, 2012.

## **Constructions for Hadamard matrices**

- 1. Kronecker product construction:  $HM(n), HM(m) \longrightarrow HM(nm)$
- 2. Gruner's theorem: if p and p+2 are twin primes, then there exist Hadamard matrices of order p(p+2) + 1.

3. Williamson method: uses the Williamson array:  $\begin{pmatrix} A & B & C & D \\ -B & A & -D & C \\ -C & D & A & -B \\ -D & -C & B & A \end{pmatrix}$ where A, B, C, D are circulant matrices whose first rows are complementary sequences.

4. Quadratic Residues of primes  $p \equiv 3 \pmod{4}$ , {seq(x^2 mod p,x=1..p)}

There are literally 100s of HM constructions ...

They all suffer from two kinds of disadvantages:

either they produce a sparse set of orders, or they fail for specific parameter values

### Williamson method: state-of-the-art

D. Z. Djokovic
Discr. Math. 115 (1993), no. 1-3, pp. 267-271

There is no Williamson matrix of order  $4 \cdot 35$ 

• W. H. Holzmann, H. Kharaghani, B. Tayfeh-Rezaie Des. Codes Cryptogr. 46 (2008), no. 3, pp. 343-352

There are no Williamson matrices of orders  $4 \cdot 47, 4 \cdot 53, 4 \cdot 59$ 

• It is not known whether a Williamson matrix of order  $4 \cdot 65$  exists.

The above non-existence results were obtained computationally.

An independent verification of the above non-existence results is highly desirable.

### **Representative Example**

Williamson solution for  $n = 29 \rightsquigarrow$  Hadamard matrix or order  $4 \cdot 29$ .

# of binary variables: 
$$4 \cdot \left(\frac{29-1}{2}\right) = 56$$

Diophantine equation satisfied:

$$\left(\sum_{i=1}^{29} a_i\right)^2 + \left(\sum_{i=1}^{29} b_i\right)^2 + \left(\sum_{i=1}^{29} c_i\right)^2 + \left(\sum_{i=1}^{29} d_i\right)^2 = 9^2 + 1^2 + 5^2 + (-3)^2 = 116 = 4 \cdot 29$$

Note that the Diophantine equation  $x^2 + y^2 + z^2 + t^2 = 116$  has 4 solutions in general (up to sign/permutation of vars)

[0, 0, 4, 10], [0, 4, 6, 8], [1, 3, 5, 9], [3, 3, 7, 7]

PowersRepresentations[116, 4, 2]

### **Representative Example cont'd**

these 4 sequences also have PSD equal to 116:

1,	8.510326027,	60.18928195,	25.99396451,	21.30642745,	115.9999999
2,	44.95677929,	1.552692648,	29.80302391,	39.68750409,	115.9999999
3,	77.29326395,	21.80423241,	15.44094978,	1.461553876,	116.0000000
4,	1.884068098,	60.92130947,	.2803757623,	52.91424662,	116.0000000
5,	52.12681974,	.4087514912,	44.29945606,	19.16497262,	115.9999999
6,	42.87772654,	44.68485046,	7.517749906,	20.91967303,	115.9999999
7,	2.768255979,	20.28734866,	1.002095058,	91.94230028,	116.0000000
8,	1.266406791,	75.69082274,	12.47835551,	26.56441492,	116.0000000
9,	13.36443427,	13.27812572,	71.05938529,	18.29805467,	116.0000000
10,	36.62394854	, 5.727979124	, 27.04815073	, 46.59992153,	, 115.9999999
11,	.8255022273	, 59.96354138	, 17.38780621	, 37.82315007,	, 115.9999999
12,	48.45774811	, 6.284747246	, 55.89279515	, 5.364709424,	, 115.9999999
13,	14.61672118	, 42.54132097	, 32.30631413	, 26.53564366	, 115.9999999
14,	34.42799914	, 6.664995510	, 67.48957786	, 7.417427483	, 116.0000000

## **Combination of SAT and CAS**

MathCheck: A Math Assistant based on a Combination of Computer
Algebra Systems and SAT Solvers
Ed Zulkoski, Vijay Ganesh, and Krzysztof Czarnecki
International Conference on Automated Deduction (CADE 2015), Berlin,
Germany, August 1-7, 2015
Main idea: SAT solvers have finely-tuned search procedures not available in CAS,
but lack the expressiveness and domain-specific knowledge of a CAS. For this reason

we use CAS code and domain-specific knowledge to considerably cut down the search space before searching with a SAT solver.

MathCheck embeds the functionality of a computer algebra system (CAS) within the inner loop of a conflict-driven clause-learning SAT solver. SAT+CAS systems, a la MathCheck, can be used as an assistant by mathematicians to either counterexample or finitely verify open universal conjectures on any mathematical topic (e.g., graph and number theory, algebra, geometry, etc.) supported by the underlying CAS system. Such a SAT+CAS system combines the efficient search routines of modern SAT solvers, with the expressive power of CAS, thus complementing both.

The key insight behind the power of the SAT+CAS combination is that the CAS system can help cut down the search-space of the SAT solver, by providing learned clauses that encode theory-specific lemmas, as it searches for a counterexample to the input conjecture (just like the T in DPLL(T)).

In addition, the combination enables a more efficient encoding of problems than a pure Boolean representation.

The paper leverages the graph-theoretic capabilities of an open-source CAS, called SAGE, on two case studies, to make progress on two long-standing open mathematical conjectures from graph theory regarding properties of hypercubes.

## The MapleSAT solver

Experiments performed with the MapleSAT solver:

Understanding VSIDS Branching Heuristics in Conflict-Driven Clause-Learning SAT Solvers

Liang, Jia Hui and Ganesh, Vijay and Zulkoski, Ed and Zaman, Atulan and Czarnecki, Krzysztof, Hardware and Software: Verification and Testing: 11th International Haifa Verification Conference, 2015

- Empirically, MapleSAT is very effective at solving cryptographic and combinatorial problems from the annual SAT competition.
- The main innovation of the solver is a new branching heuristic based-off of a well-known algorithm in reinforcement learning literature.
- Details available in the upcoming:

Exponential Recency Weighted Average Branching Heuristic for SAT Solvers Liang, Jia Hui and Ganesh, Vijay and Poupart, Pascal and Czarnecki, Krzysztof Proceedings of AAAI-16, 2016

### **Encoding Hadamard as SAT**

- let *n* be odd and set  $m = \frac{n-1}{2}$
- use 4 sets of boolean variables  $a_1, \ldots, a_m, b_1, \ldots, b_m, c_1, \ldots, c_m, d_1, \ldots, d_m$  to encode the first rows of the symmetric circulant matrices A, B, C, D
- use 2-bit and 3-bit adders to verify the autocorrelation equations
- use ordering conditions and the Espresso logic minimizer, and up to 7-bit adders (resulted in using fewer variables but more clauses)

EXAMPLE:  $4 \cdot 35$  can be written as the sum of 4 odd squares in 3 ways:

 $1^2 + 3^2 + 3^2 + 11^2 = 1^2 + 3^2 + 7^2 + 9^2 = 3^2 + 5^2 + 5^2 + 9^2 = 4 \cdot 35$ 

we can deduce the number of  $\pm 1$ 's in each of the first rows of A, B, C, D, e.g.  $a_1 + \cdots + a_{35} = 3$ . implies 19 +1's and 16 -1's.

**RELATED WORK:** 

Handbook of Satisfiability, IOS Press, 2009

- Armin Biere, Marijn Heule, Hans van Maaren, Toby Walsh
- Chapter 17. Combinatorial Designs by SAT Solvers, by Hantao Zhang

### Autocorrelation of finite sequences

• The **periodic autocorrelation function** associated to a finite sequence  $A = [a_0, \ldots, a_{n-1}]$  of length n is defined as

$$P_A(s) = \sum_{k=0}^{n-1} a_k a_{k+s}, \ s = 0, \dots, n-1,$$

where k + s is taken modulo n, when k + s > n.

• The aperiodic autocorrelation function associated to a finite sequence  $A = [a_0, \ldots, a_{n-1}]$  of length n is defined as

$$N_A(s) = \sum_{k=0}^{n-1-s} a_k a_{k+s}, \ s = 0, \dots, n-1,$$

We are mostly concerned with binary  $\{-1, +1\}$ , ternary  $\{-1, 0, +1\}$  and 4th roots of unity  $\{\pm 1, \pm i\}$  sequences.

Note that for sequences with complex number elements,  $a_{k+s}$  is replaced by  $\overline{a_{k+s}}$ .

Example:  $n = 7, A = [a_1, ..., a_7]$ 

$P_A(0)$	=	$a_{1}^{2} + a_{2}^{2} + a_{3}^{2} + a_{4}^{2} + a_{5}^{2} + a_{6}^{2} + a_{7}^{2}$
$P_A(1)$	=	$a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_6 + a_6a_7 + a_7a_1$
$P_A(2)$	=	$a_1a_3 + a_2a_4 + a_3a_5 + a_4a_6 + a_5a_7 + a_6a_1 + a_7a_2$
$P_A(3)$	=	$a_1a_4 + a_2a_5 + a_3a_6 + a_4a_7 + a_5a_1 + a_6a_2 + a_7a_3$
$P_A(4)$	=	$a_1a_4 + a_2a_5 + a_3a_6 + a_4a_7 + a_5a_1 + a_6a_2 + a_7a_3$
$P_A(5)$	=	$a_1a_3 + a_2a_4 + a_3a_5 + a_4a_6 + a_5a_7 + a_6a_1 + a_7a_2$
$P_A(6)$	=	$a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_6 + a_6a_7 + a_7a_1$

$$\begin{array}{rclrcl} N_A(0) &=& a_1{}^2 + a_2{}^2 + a_3{}^2 + a_4{}^2 + a_5{}^2 + a_6{}^2 + a_7{}^2 \\ N_A(1) &=& a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_6 + a_6a_7 \\ N_A(2) &=& a_1a_3 + a_2a_4 + a_3a_5 + a_4a_6 + a_5a_7 \\ N_A(3) &=& a_1a_4 + a_2a_5 + a_3a_6 + a_4a_7 \\ N_A(4) &=& a_1a_5 + a_2a_6 + a_3a_7 \\ N_A(5) &=& a_1a_6 + a_2a_7 \\ N_A(6) &=& a_1a_7 \end{array}$$

$$P_A(s) = N_A(s) + N_A(n-s), s = 1, ..., n-1$$

### **Circulant matrices**

A  $n \times n$  matrix C(A) is called **circulant** if every row (except the first) is obtained by the previous row by a right cyclic shift by one.

$$C(A) = \begin{bmatrix} a_0 & a_1 & \dots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_0 & \dots & a_{n-3} & a_{n-2} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_2 & a_3 & \dots & a_0 & a_1 \\ a_1 & a_2 & \dots & a_{n-1} & a_0 \end{bmatrix}$$

- Consider a finite sequence  $A = [a_0, \ldots, a_{n-1}]$  of length n and the circulant matrix C(A) whose first row is equal to A. Then  $P_A(i)$  is the inner product of the first row of C(A) and the i + 1 row of C(A).
- symmetry property  $\rightsquigarrow P_A(s) = P_A(n-s), s = 1, \dots, n-1.$ • 2<sup>nd</sup> ESF property  $\rightsquigarrow P_A(1) + P_A(2) + \dots + P_A(n-1) = 2e_2(a_0, \dots, a_{n-1})$
- $\rightsquigarrow$   $N_A(s) + N_A(n-s) = P_A(s), s = 1, \dots, n-1.$

### **Complementary Sequences** Definition:

Let  $\{A_i\}_{i=1,...,t}$  be t sequences of length v with complex elements. The sequences  $\{A_i\}_{i=1,...,t}$  are called complementary, if

$$\sum_{i=1}^{t} PAF_{A_i} = [\alpha_0, \underbrace{\alpha, \dots, \alpha}_{v-1 \text{ terms}}]$$

with the convention:

$$PAF_{A_i} = [PAF_{A_i}(0), PAF_{A_i}(1), \dots, PAF_{A_i}(v-1)].$$

Algorithms and Metaheuristics for Combinatorial Matrices, Ilias S. Kotsireas, in Handbook of Combinatorial Optimization, 2nd edition, Pardalos, P. M., Du, D.-Z., Graham, R. L. (eds) pp. 283-309, Springer 2013

## **Unified description of combinatorial objects**

number/type of sequences	defining property	name	
1 binary	aper. autoc. $0, \pm 1$	Barker sequences	
1 ternary	per. autoc. 0	circulant weighing matrices	
2 binary	aper. autoc. 0	Golay sequences	
2 binary	per. autoc. 0	Hadamard matrices	
2 binary	per. autoc. 2	D-optimal matrices	
2 binary	per. autoc. $-2$	Hadamard matrices	
2 ternary	aper. autoc. 0	TCP	
2 ternary	per. autoc. 0	Weighing matrices	
3 binary	aper. autoc. const.	Normal sequences	
4 binary	aper. autoc. 0	Base sequences	
4 binary	aper. autoc. 0	Turyn type sequences	
4 ternary	aper. autoc. 0	T-sequences	
4 binary	per. autoc. 0	Williamson Hadamard	
$2 \dots 12$ binary	per. autoc. zero	PCS	

## **Power Spectral Density, PSD**

**Seberry & Gysin** first introduced the PSD concept in the search for complementary sequences of various kinds.

#### **Definition:**

 $PSD([a_1, \ldots, a_n], k)$  denotes the k-th element of the power spectral density sequence, i.e. the square magnitude of the k-th element of the discrete Fourier transform (DFT) sequence associated to  $[a_1, \ldots, a_n]$ .

The DFT sequence associated to  $[a_1, \ldots, a_n]$  is defined as

$$DFT_{[a_1,\ldots,a_n]} = [\mu_0,\ldots,\mu_{n-1}], \text{ with } \mu_k = \sum_{i=0}^{n-1} a_{i+1} \omega^{ik}, k = 0,\ldots,n-1$$

where  $\omega = e^{\frac{2\pi i}{n}} = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right)$  is a primitive *n*-th root of unity.

### **PSD** criterion

Williamson Hadamard matrices: 4 complementary sequences of length n, (odd) PAF constant: 0, PSD constant: 4n.

$$PAF(A, s) + PAF(B, s) + PAF(C, s) + PAF(D, s) = 0, \quad s = 1, \dots, \frac{n-1}{2}$$
$$PSD(A, s) + PSD(B, s) + PSD(C, s) + PSD(D, s) = 4n, \quad s = 1, \dots, \frac{n-1}{2}$$

if for a certain sequence  $A = [a_1, \ldots, a_n]$  there exists  $s \in \{1, \ldots, n-1\}$  with the property that PSD(A, s) > 4n, then this sequence cannot be used to construct 4 such complementary sequences

**Important Consequence:** we can now **decouple** the PAF equations, roughly corresponding to cutting down the complexity by four.





## http://top500.org/



TIANHE-2	(MILKYWAY-2)	
Site:	National Super Computer Center in Guangzhou	
Cores:	$3,\!120,\!000$	
Linpack Performance (Rmax)	33,862.7 TFlop/s	
Theoretical Peak (Rpeak)	54,902.4 TFlop/s	
Memory:	$1,024,000  \mathrm{GB}$	
Processor:	Intel Xeon E5-2692v2 12C 2.2GHz	
Compiler:	icc	

2007: open problem,  $2^{50}$  ops  $\rightsquigarrow$  2015: ex. search in 10 minutes

# Compression of complementary sequences

#### **Definition:**

Let  $A = [a_0, a_1, \dots, a_{v-1}]$  be a complex sequence of length v = dm. Set  $a_j^{(d)} = a_j + a_{j+d} + \dots + a_{j+(m-1)d}$ , for  $j = 0, \dots, d-1$ . Then we say that the sequence  $A^{(d)} = [a_0^{(d)}, a_1^{(d)}, \dots, a_{d-1}^{(d)}]$  of length d is the m-compression of A.

PhD thesis of Yoseph Strassler, (1997), Bar Ilan University, Israel.

#### **Example:**

$$\begin{split} A &= CW(24,9) = [0,0,0,-1,-1,0,0,0,0,0,1,-1,0,0,0,-1,1,0,0,1,0,0,-1,-1] \\ m &= 2, \quad d = 12, \quad \rightsquigarrow \quad A^{(12)} = [0,0,0,-2,0,0,0,1,0,0,0,-2] \\ m &= 3, \quad d = 8, \quad \rightsquigarrow \quad A^{(8)} = [1,0,1,-1,-1,0,-1,-2] \end{split}$$

#### Theorem: Djokovic-Kotsireas (2012)

Let  $\{A_i\}_{i=1,...,t}$  be t complementary sequences, of length v each, with complex elements  $A_i = [a_{i0}, a_{i1}, \ldots, a_{i,v-1}]$ , for  $i = 1, \ldots, t$  and  $\sum_{i=1}^t PAF_{A_i} = [\alpha_0, \underbrace{\alpha, \ldots, \alpha}_{v-1 \text{ terms}}]$ . Assume that v = dm and set  $a_{ij}^{(d)} = a_{i,j} + a_{i,j+d} + \cdots + a_{i,j+(m-1)d}$  for  $i = 1, \ldots, t$ and  $j = 0, \ldots, d-1$ . Let  $A_i^{(d)}$  be the t sequences  $A_i^{(d)} = [a_{i0}^{(d)}, \ldots, a_{i,d-1}^{(d)}]$ , for  $i = 1, \ldots, t$ . Then the t sequences  $\{A_i^{(d)}\}_{i=1,...,t}$ , of length d each, are also complementary and we

have:

$$\sum_{i=1}^{t} PAF_{A_{i}^{(d)}} = [\alpha_{0} + (m-1)\alpha, \underbrace{m\alpha, \dots, m\alpha}_{d-1 \text{ terms}}]$$
(1)

$$\sum_{i=1}^{\circ} PSD_{A_i^{(d)}} = [\beta_0, \underbrace{\beta, \dots, \beta}_{d-1 \text{ terms}}]$$
(2)

## Periodic Golay pairs of length 68

Consider the following two sequences of length 34 each, with  $\{-2, 0, +2\}$  elements:  $A^{(34)} = [0, 0, 0, 2, 0, 0, -2, 0, 0, 0, 2, -2, 0, 0, -2, 0, 0, 2, 0, 0, 0, 2, 2, -2, 0, 0, -2, 0, 0, 2, 0, 2, 0, 2]$  $B^{(34)} = [0, 0, -2, 2, 0, 2, 0, -2, -2, 0, 2, 2, 0, 2, -2, 0, 2, 0, 2, 2, 0, 2, 2, 0, 2, 2, 0, -2, 2, 0, -2, -2]$ 

These two sequences satisfy the following properties:

1. 
$$PAF(A^{(34)}, s) + PAF(B^{(34)}, s) = 0, s = 0, 1, ..., 33;$$

2. 
$$PSD(A^{(34)}, s) + PSD(B^{(34)}, s) = 2 \cdot 68 = 136, \ s = 0, 1, \dots, 33;$$

3. 
$$PSD(A^{(34)}, 17) = 100$$
 and  $PSD(B^{(34)}, 17) = 36;$ 

4. 
$$\sum_{i=1}^{34} A_i^{(34)} = 6$$
 and  $\sum_{i=1}^{34} B_i^{(34)} = 10;$ 

5. The total number of 0 elements in  $A^{(34)}$  and  $B^{(34)}$  is equal to 34;

- 6. The total number of  $\pm 2$  elements in  $A^{(34)}$  and  $B^{(34)}$  is equal to 34;
- 7.  $A^{(34)}$  contains 21 zeros and  $B^{(34)}$  contains 13 zeros.

 $A^{(34)}$  and  $B^{(34)}$  are the 2-compressed sequences of two  $\{-1, +1\}$  sequences of length 68 each, that form a particular **periodic Golay pair of length** 68:



 $\rightsquigarrow$  Hadamard matrices of order  $2 \cdot 68$ 

Djokovic, Dragomir; Kotsireas, Ilias; Recoskie, Daniel; Sawada, Joe Charm bracelets and their application to the construction of periodic Golay pairs. Discrete Appl. Math. 188 (2015), 32-40.

## **Periodic Golay pairs of length** 72

Using the same machinery, we also found **periodic Golay pairs of length** 72

Dragomir Djokovic and Ilias Kotsireas, Periodic Golay pairs of length 72 in:

Springer Proceedings in Mathematics & Statistics, Vol. 133 Algebraic Design Theory and Hadamard Matrices ADTHM, Lethbridge, Alberta, Canada, July 2014 Colbourn, Charles J. (Ed.) 2015

Only known example of a length of a periodic Golay pair that is divisible by 3

SDS(72; 36, 30; 30)

10 million lines of C code, generated with Maple meta-programming next open case: order 90

## **Interactions with Coding Theory**

• Gröbner Bases, Coding, and Cryptography M. Sala, T. Mora, L. Perret, S. Sakata, C. Traverso (Editors)

**Open problem:** Does there exists a binary linear [72, 36, 16] code? The answer lies in being able to construct an ample supply of skew-Hadamard matrices of order 72.

• Information security, coding theory and related combinatorics. Information coding and combinatorics

Dean Crnkovic and Vladimir Tonchev (Editors)NATO Science for Peace and Security Series D:Information and Communication Security, 29.IOS Press, Amsterdam, 2011.

### **Interactions with Quantum Computing**

Weighing matrices are generalizations of Hadamard matrices.

 $W \cdot W^t = kI_n$ 

- "Weighing matrices and optical quantum computing" S. Flammia and S. Severini, J. Phys. A: Math. Theor. 42 (2009) 065302
- "Quantum Algorithms for Weighing Matrices and Quadratic Residues" W. van Dam, Algorithmica 34, (2002) pp. 413428.

## Acknowledgement

This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network, SHARCNET, www.sharcnet.ca and Compute/Calcul Canada.



## **Future work roadmap**

- integrate symmetries/compression, to our SAT encoding of autocorrelation problems
- improve the SAT encoding of autocorrelation problems
- improve and further optimize algorithmic implementations
- explore the applicability of new HPC paradigms: Intel MIC architecture
- develop a custom-tailored SAT solver for autocorrelation problems