THE MORE THE MERRIER:
EA FOR DESIGNING 5-VALUED SPECTRA BOOLEAN FUNCTIONS

CLAUDE CARLET, MARKO ĐURASEVIĆ, DOMAGOJ JAKOBOVIC, LUCA MARIOT, STJEPAN PICEK

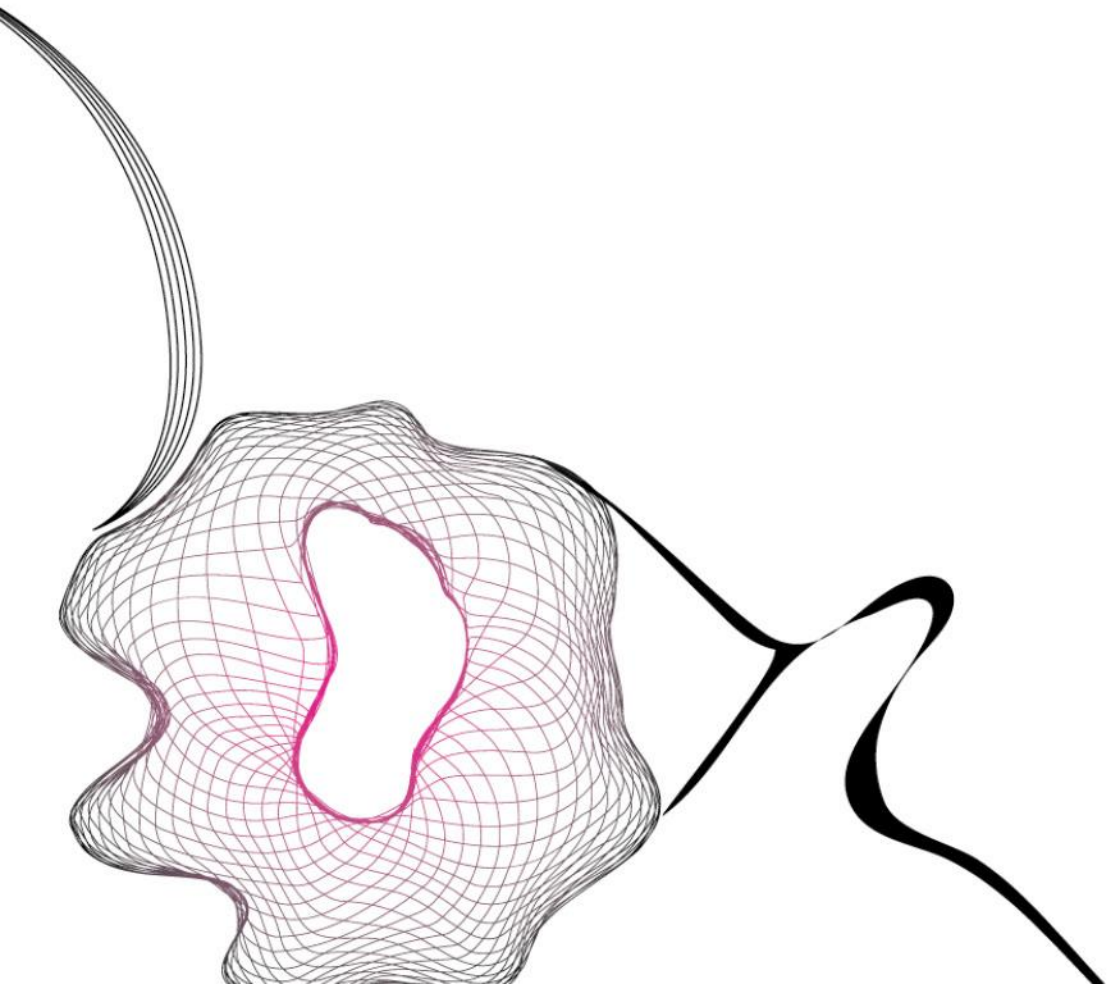
EVOAPPS 2025

TRIESTE, 24 APRIL 2025

UNIVERSITY OF TWENTE.

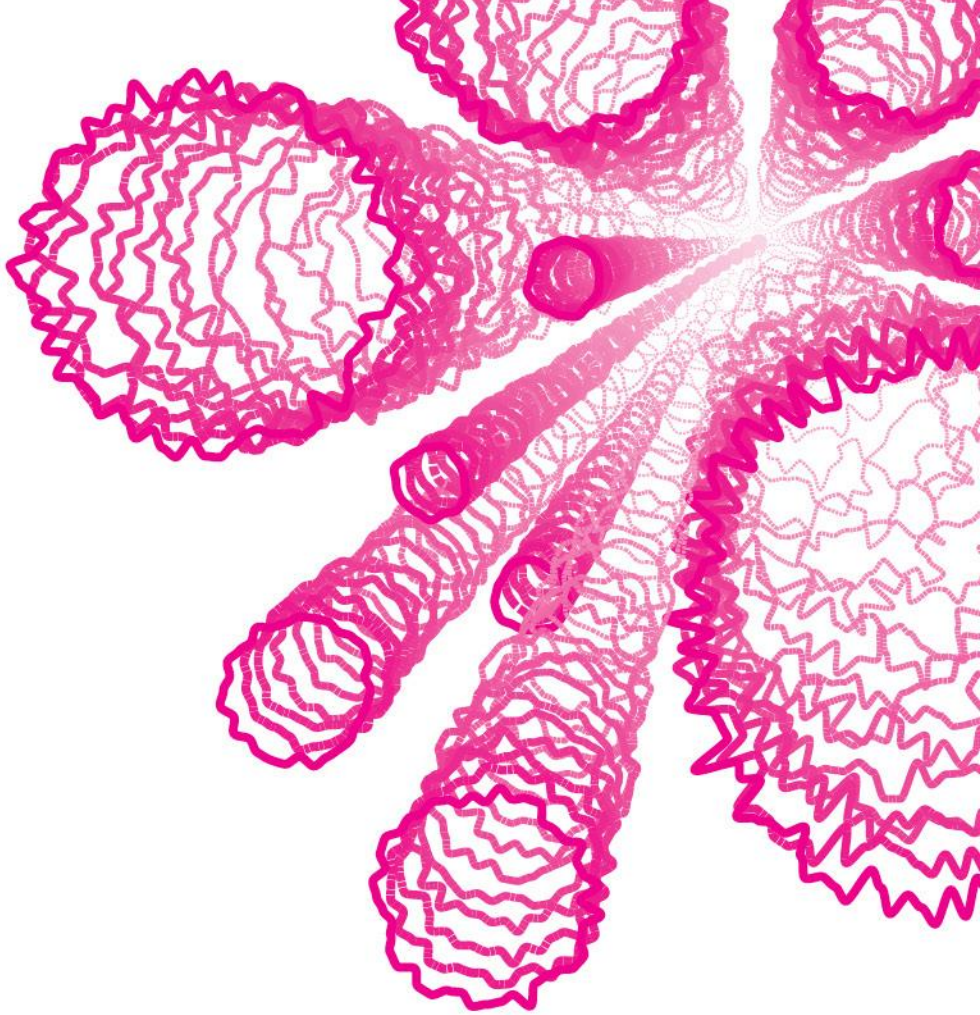
Radboud University

SUMMARY



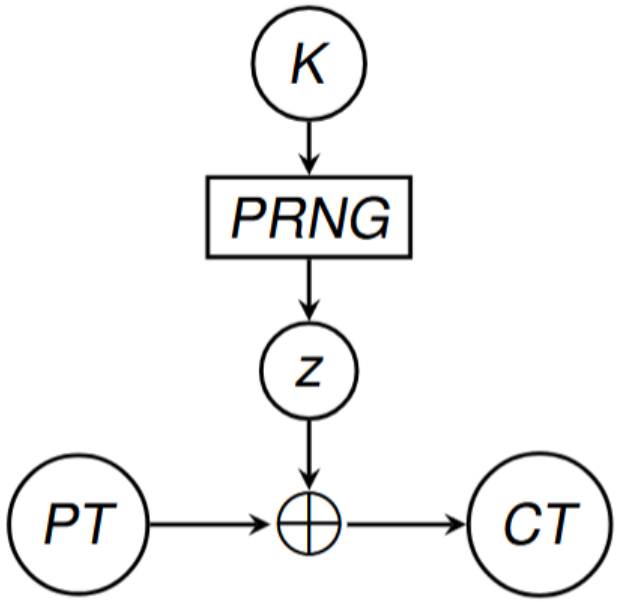
- 5-Valued Spectra Boolean Functions
- Genotype encodings & Fitness Functions
- Experimental Evaluation
- Conclusions and Future Directions

5-VALUED SPECTRA BOOLEAN FUNCTIONS



UNIVERSITY
OF TWENTE.

PRIMITIVES IN SYMMETRIC CRYPTOGRAPHY



(a) Stream cipher

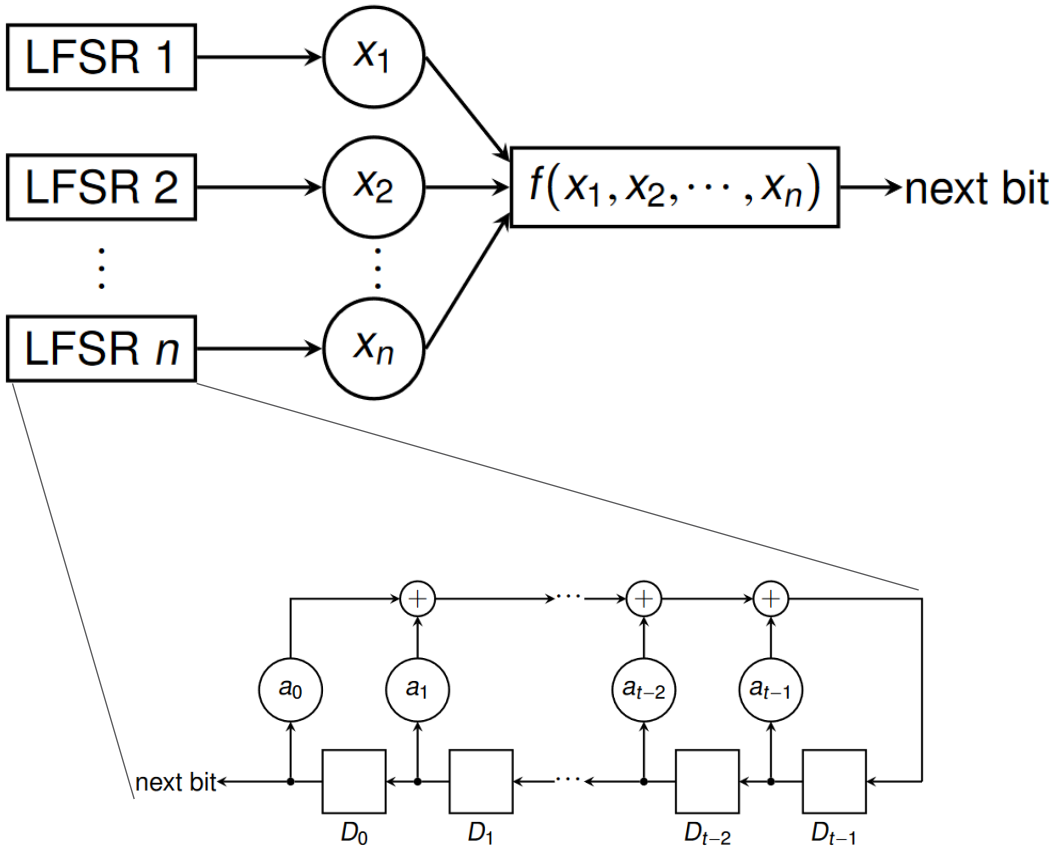
- Symmetric ciphers require several low-level **primitives** [5]:
  - Pseudorandom number generators (PRNG)
  - Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$
  - S-boxes $F : \{0,1\}^n \rightarrow \{0,1\}^n$
- Classic methods to build them: **algebraic constructions** [1, 2]
- Alternative: **(meta)heuristics** (GA, GP, …) [4, 10]

COMBINER PSEUDORANDOM GENERATOR

- **Idea:** use n LFSRs instead of one
- LFSRs outputs *combined* using a Boolean function:

$$f : \{0,1\}^n \rightarrow \{0,1\}$$

- Security of the PRG: **cryptographic properties of** $f : \{0,1\}^n \rightarrow \{0,1\}$ [2]



BOOLEAN FUNCTIONS

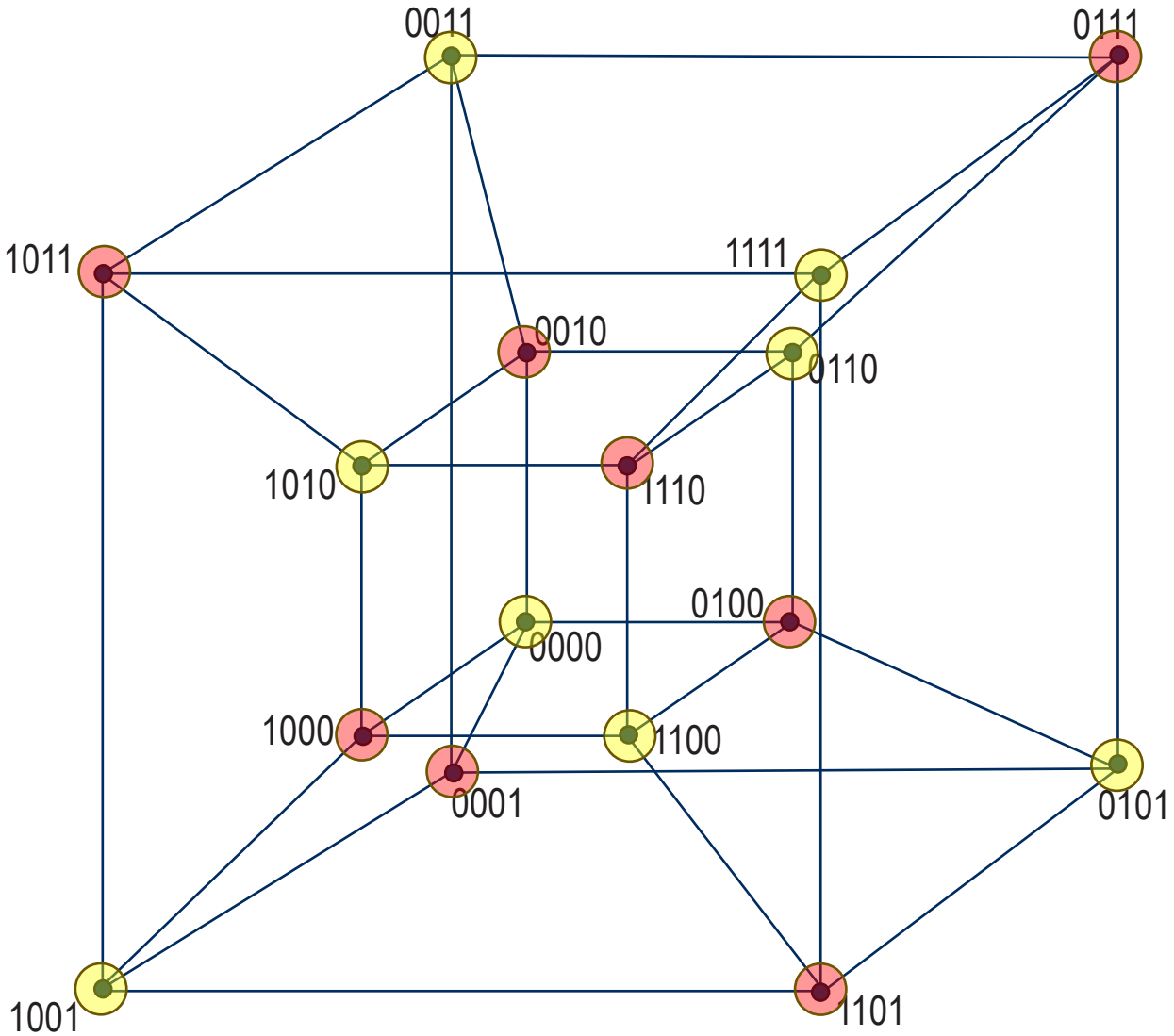
- A mapping $f : \{0,1\}^n \to \{0,1\}$ represented by a *truth table*

(x_1, x_2, x_3)	000	001	010	011	100	101	110	111
$f(x_1, x_2, x_3)$	0	1	1	0	1	0	1	0

- The function must satisfy some properties to resist specific attacks:
  - **Balancedness**: equal number of 0s and 1s
  - High **Nonlinearity**: high Hamming distance from **linear** functions:

$$L_a(x) = a \cdot x = a_1 x_1 \oplus \ldots \oplus a_n x_n, \quad a_i, x_i \in \{0,1\}$$

NONLINEARITY – GEOMETRIC INTUITION



- **Example**: $n = 2$ variables (truth tables of $2^2 = 4$ bits, 16 functions in total)
- Linear functions and complements:

$a_1x_1 \oplus a_2x_2$	TT	TT $\oplus 1$
0	0000	1111
x_1	0011	1100
x_2	0101	1010
$x_1 \oplus x_2$	0110	1001

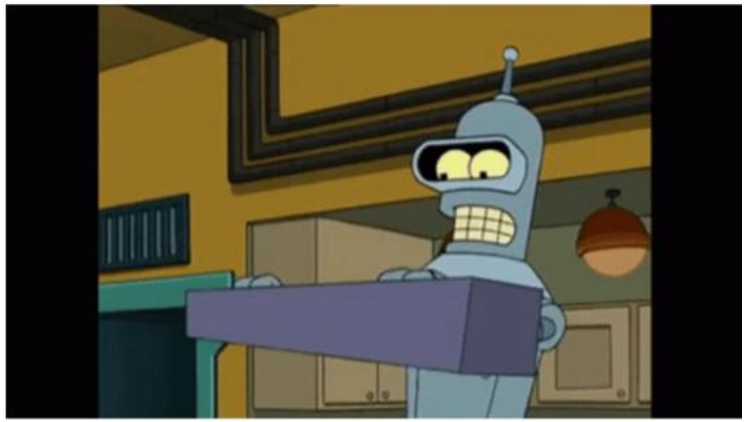
- Hamming distance from linear functions: **1**, e.g.: $f(x_1, x_2) = x_1x_2$

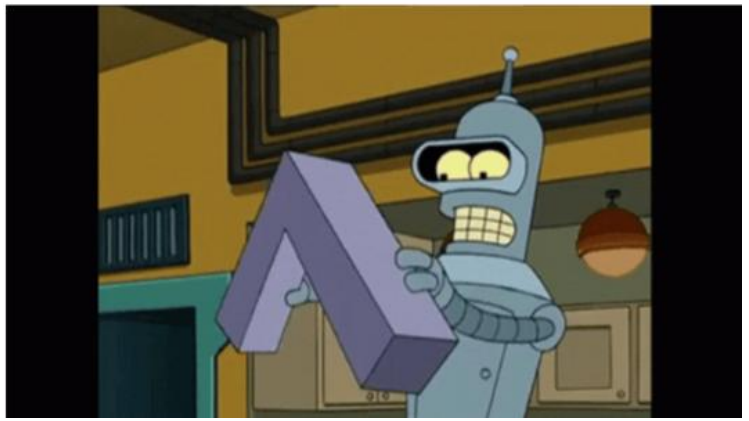
WALSH TRANSFORM

$$W_f(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus a \cdot x}$$

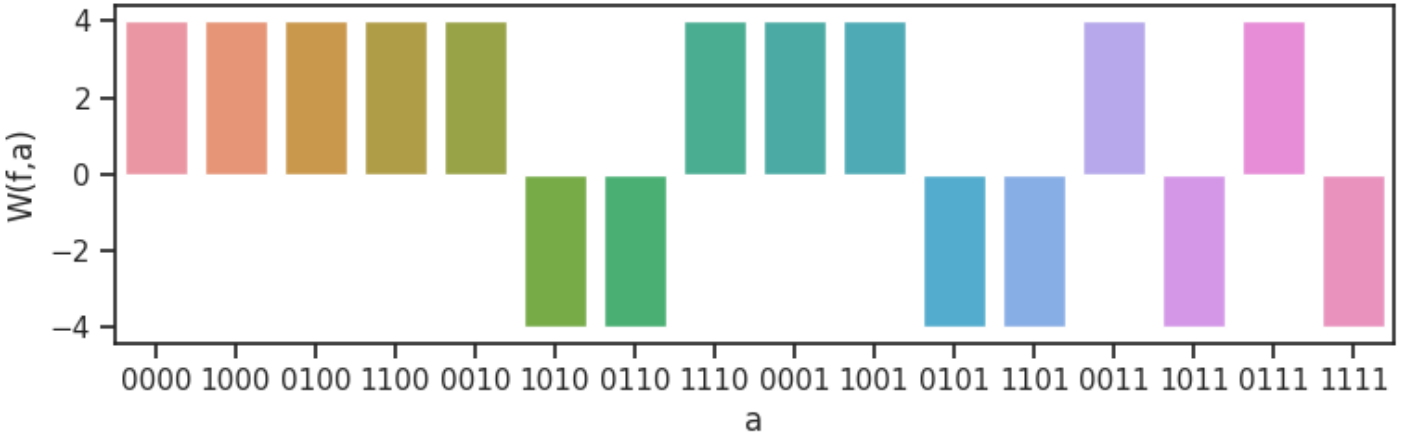
- Measures correlation with *linear* functions $L_a(x) = a \cdot x = a_1 x_1 \oplus \ldots \oplus a_n x_n,$
- Can be computed in $\mathcal{O}(n2^n)$ steps (FFT-like algorithm) [2]
- Nonlinearity is equal to: $nl_f = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |W_f(a)|$
- **Parseval's identity:** $\sum_{a \in \{0,1\}^n} [W_f(a)]^2 = 2^{2n}$
- **Covering radius bound:** $nl_f \leq 2^{n-1} - 2^{\frac{n}{2}-1}$

BENT FUNCTIONS





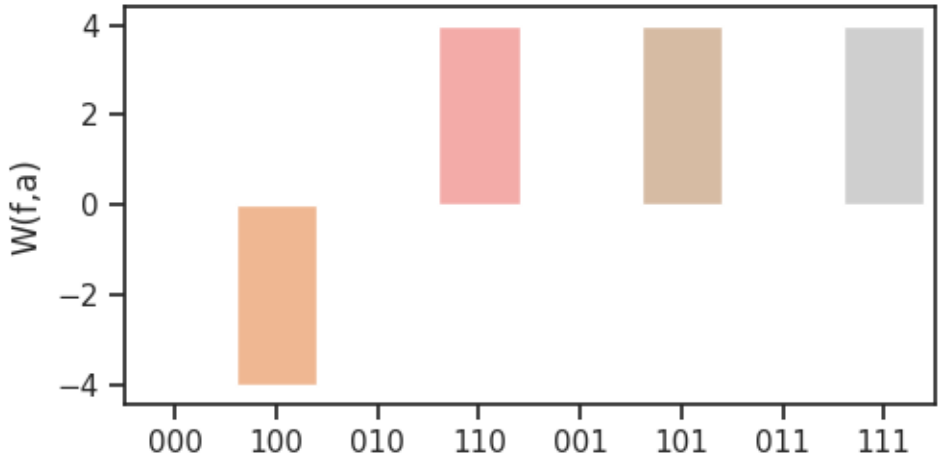
- Only two spectral values: $W_f(a) \in \{-2^{\frac{n}{2}}, +2^{\frac{n}{2}}\}$
- **PRO**: highest possible nonlinearity $nl_f = 2^{n-1} - 2^{\frac{n}{2}-1}$
- **CONS**: unbalanced, exist only for n even [2]



UNIVERSITY OF TWENTE.

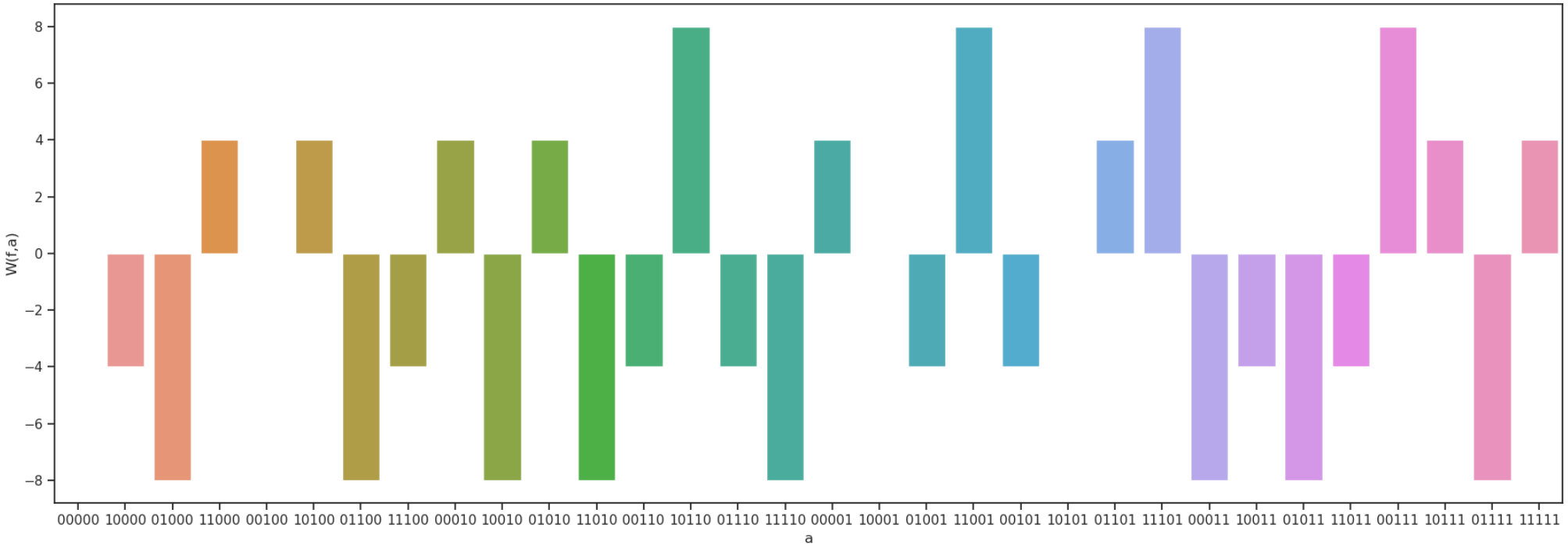
PLATEAUED FUNCTIONS

- Three spectral values: $W_f(a) \in \{-2^\lambda, 0 + 2^\lambda\}$
- **PRO**: exist for any n even or odd, can be balanced, achieve good nonlinearity [2]



5-VALUED SPECTRA FUNCTIONS

- Five spectral values: $W_f(a) \in \{0, \pm 2^{\lambda_1}, \pm 2^{\lambda_2}\}$ [8]
- **PRO**: more combinations available to get good crypto properties



UNIVERSITY OF TWENTE.

OPTIMIZATION PROBLEM

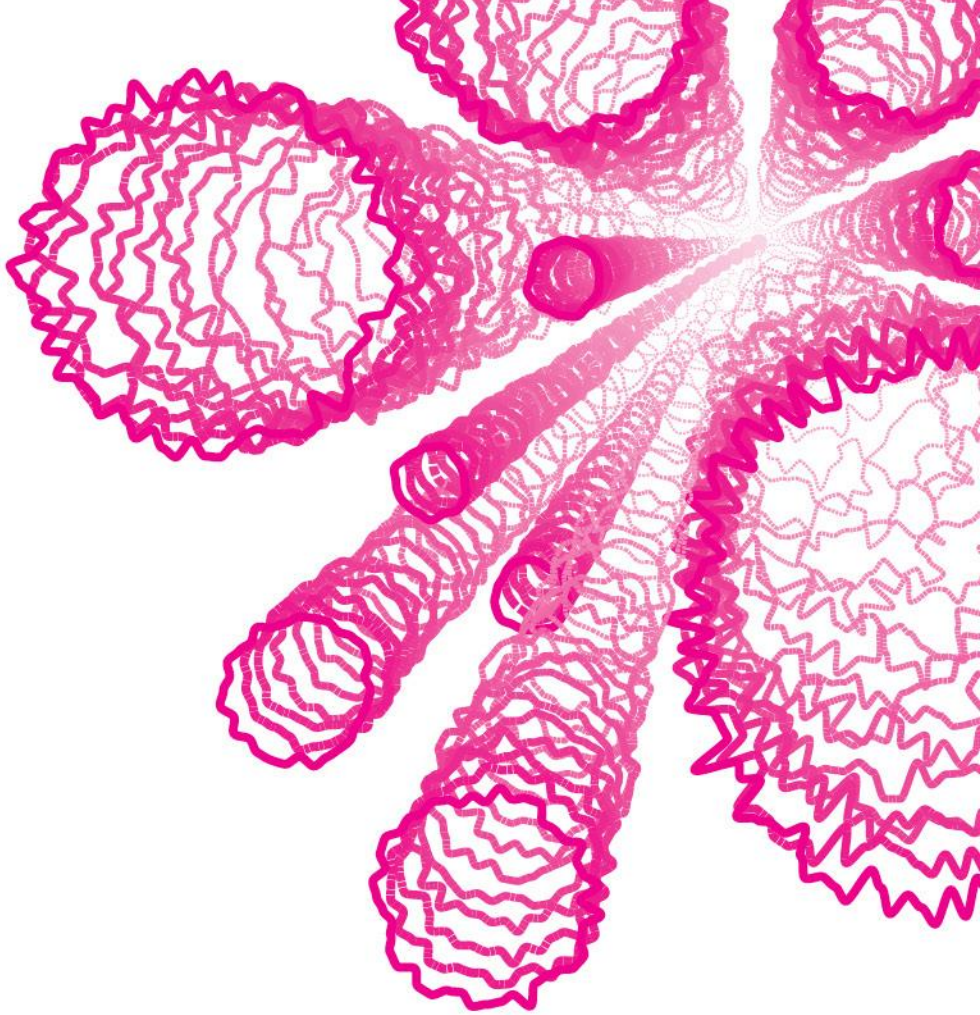
- Given $n \in \mathbb{N}$, how do we fill the truth table so that f :
  - Has a **5-valued Walsh spectrum**
  - Is **highly nonlinear** and **balanced**

(x_1, x_2, x_3)	000	001	010	011	100	101	110	111
$f(x_1, x_2, x_3)$	?	?	?	?	?	?	?	?

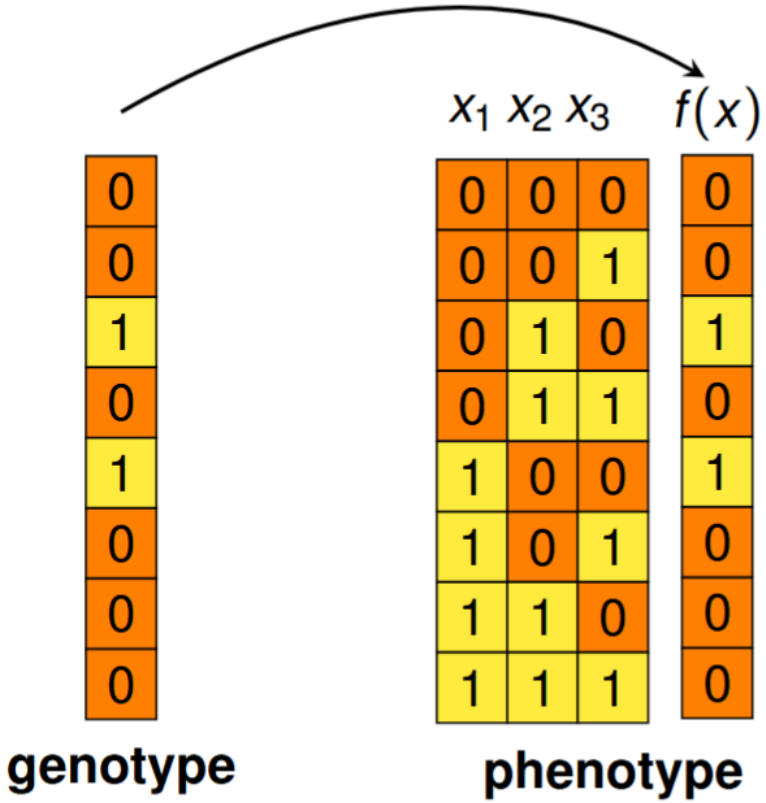
- The truth table has size 2^n so there are 2^{2^n} combinations [2]

n	3	4	5	6	7	8
2^{2^n}	256	65536	$4.3 \cdot 10^9$	$1.8 \cdot 10^{19}$	$3.4 \cdot 10^{38}$	$1.2 \cdot 10^{77}$

GENOTYPE ENCODINGS & FITNESS FUNCTIONS

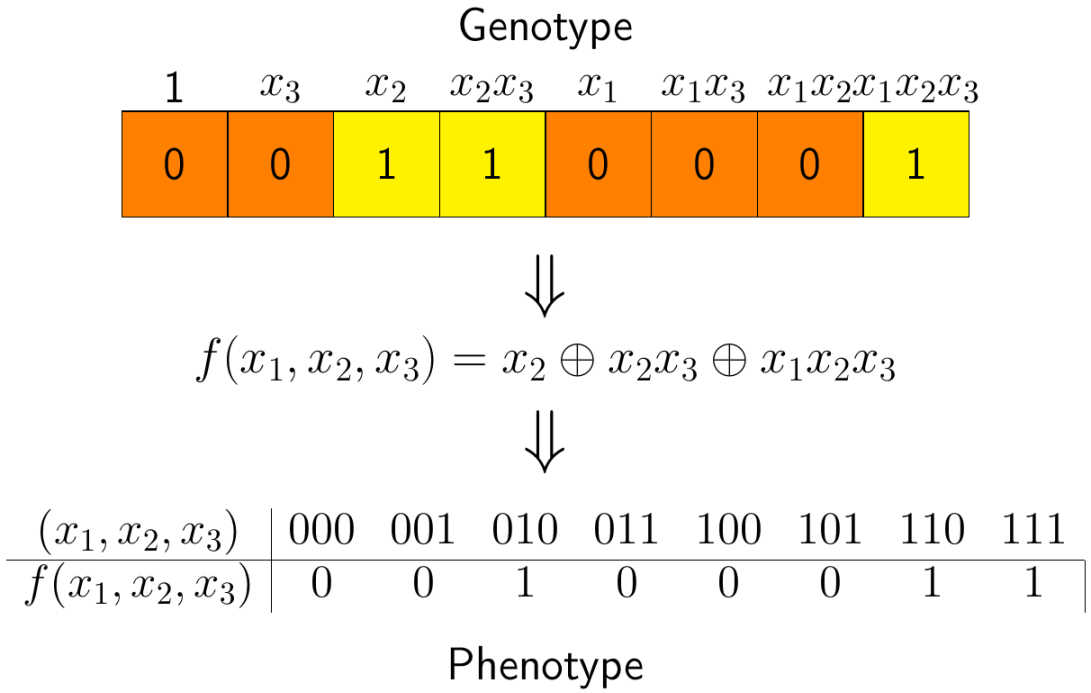


BITSTRING ENCODING FOR GENETIC ALGORITHMS



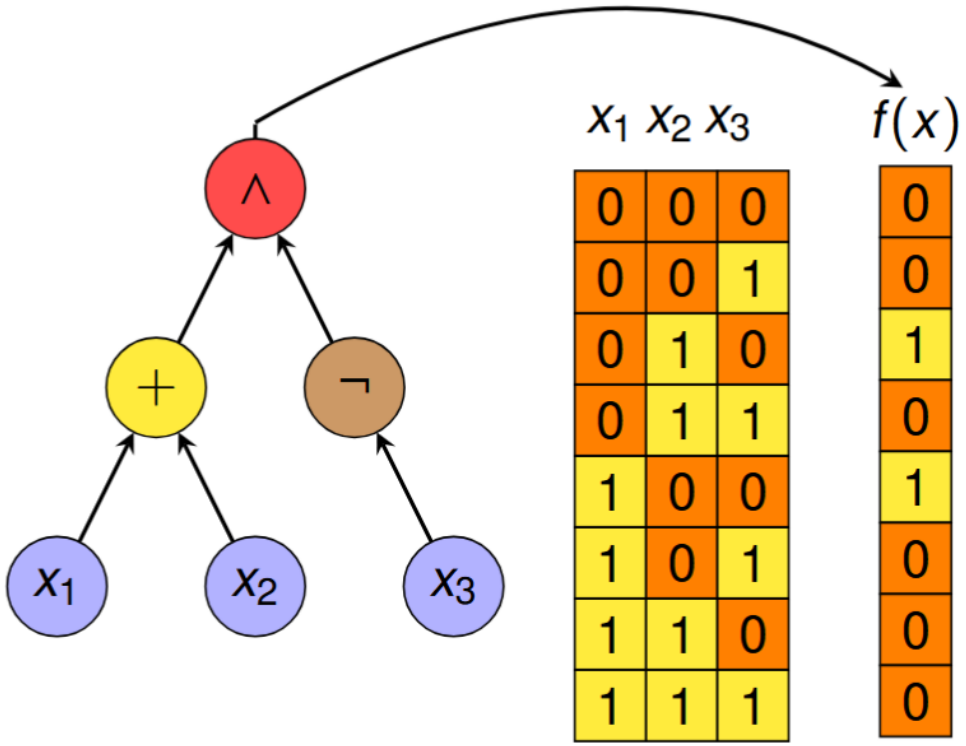
- **GA genotype**: fixed-length bitstrings [9, 13]
- **phenotype**: truth table of f
- **Assumption:** the input vectors of $\{0,1\}^n$ are sorted lexicographically
- Usual variation operators of GA (one-point crossover, bit-flip mutation) [12, 14]

ALGEBRAIC NORMAL FORM ENCODING (ANF)



- Represents a function f as a sum (XOR) of products (AND) [2]
- **Genotype:** ANF can also be encoded by a bitstring of length 2^n
- **Phenotype:** mapping to truth table done with the (Fast) *Moebius Transform*
- Same set of GA operators used for the bitstring encoding [9, 13]

SYMBOLIC ENCODING WITH GP



- **GP Genotype**: a syntactic tree
  - Leaf nodes: input variables
  - Internal nodes: operators (e.g. AND, OR, NOT, XOR, …)
- **Phenotype:** evaluate the tree for all possible assignments of the leaf nodes
- Classic GP operators (subtree crossover and mutation, …) [7, 14]

FIRST FITNESS FUNCTION

- **Three-stage** fitness:
  - Step 1: optimize “5-valuedness”: $fit_1 = \frac{1}{1+|\#values-5|}$
  - Step 2: optimize balancedness: $fit_2 = |2^{n-1} - \#1_f|$
  - Step 3: optimize nonlinearity: $fitness = nl_f + \frac{2^n - \#max\_values}{2^n}$.
- **Step 3 Idea:** do not simply maximize nonlinearity
- **Rationale:** optimize a “smoother” fitness

SECOND FITNESS FUNCTION

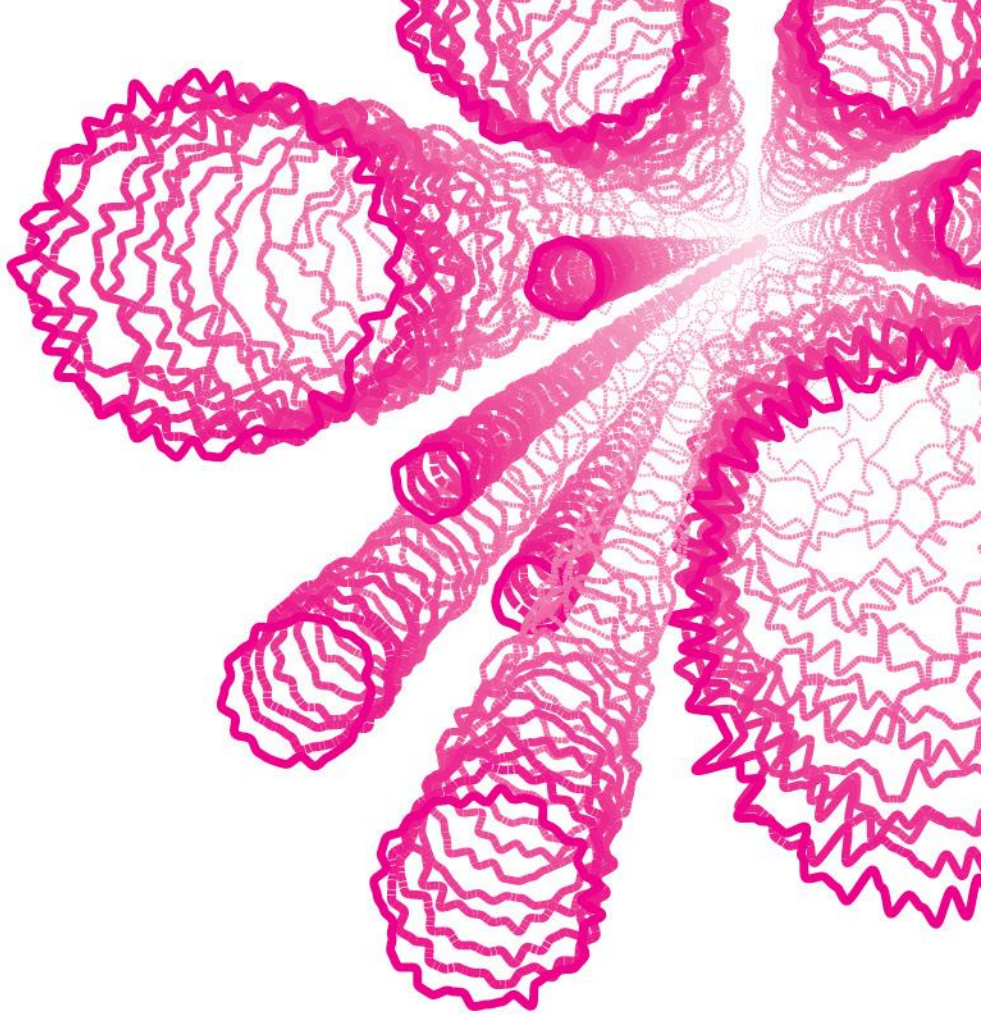
- **Penalty factor for 5-valuedness:**

$$pen(f) = \begin{cases} |\{a \in \{0,1\}^n : W_f(a) \notin \{0, \pm 2^{\frac{n-1}{2}}, \pm 2^{\frac{n+1}{2}}\}\}| & , \text{ if n is odd,} \\ |\{a \in \{0,1\}^n : W_f(a) \notin \{0, \pm 2^{\frac{n}{2}}, \pm 2^{\frac{n+2}{2}}\}\}| & , \text{ if n is even.} \end{cases}$$

- **Fitness function:** $fitness_2(f) = \frac{nl_f}{1+pen(f)}$

- **Rationale:** might allow the algorithm to reach 5-valuedness by traversing a wider portion of the search space

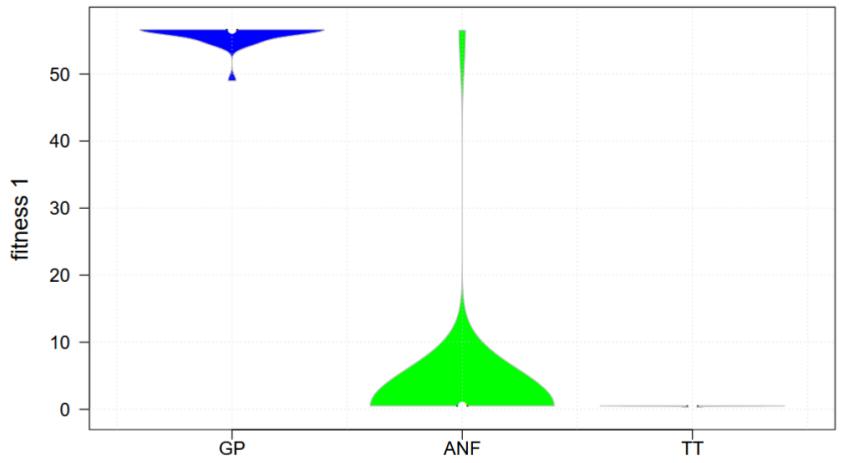
EXPERIMENTAL EVALUATION



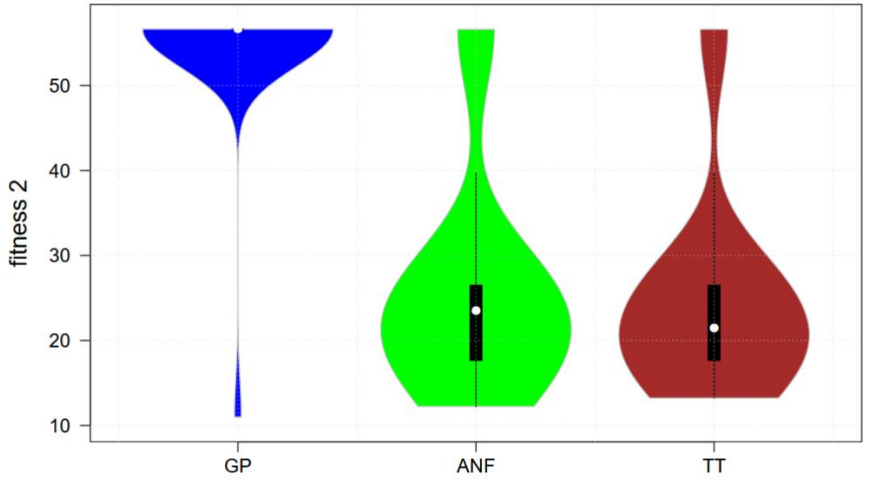
EXPERIMENTAL SETTING

- **EA used**: **GA** for bitstring/ANF encoding, **GP** for symbolic encoding
- **Breeding strategy**: 3-tournament Steady-State selection (SST)
- **Mutation probability:** 0.5
- **Problem sizes:** from $n = 5$ to $n = 16$
- **Fitness budget:** 10^6 fitness evaluations
- **Repetitions:** 30 independent runs

RESULTS – DISTRIBUTIONS FOR SIZE 7



(a) Results for $fitness_1$.



(b) Results for $fitness_2$.

- **Main finding**: GA fails to find 5-valued spectra functions already for $n = 7$
- GP, instead, finds 5-valued spectra functions for all considered problem sizes

RESULTS – GP STATISTICS

Size	$fitness_1$			$fitness_2$		
	avg	stdev	max	avg	stdev	max
5	12.63	0.00	12.63	12.62	0.02	12.63
6	24.94	0.01	24.94	24.94	0.00	24.94
7	56.00	1.52	56.63	55.10	8.33	56.63
8	112.96	0.01	112.97	112.96	0.01	112.97
9	235.68	6.21	240.63	67.04	79.01	240.63
10	480.98	0.01	480.98	148.30	169.41	480.98
11	985.13	11.97	992.63	110.17	3.20	115.40
12	1984.97	0.02	1984.99	219.51	8.05	230.60
13	3996.07	50.47	4032.63	419.61	5.51	422.60
14	8064.99	0.02	8065.00	838.17	13.33	870.60
15	16183.77	56.77	16256.60	1645.43	11.51	1664.20
16	31958.31	2990.04	32513.00	3292.20	23.32	3328.20

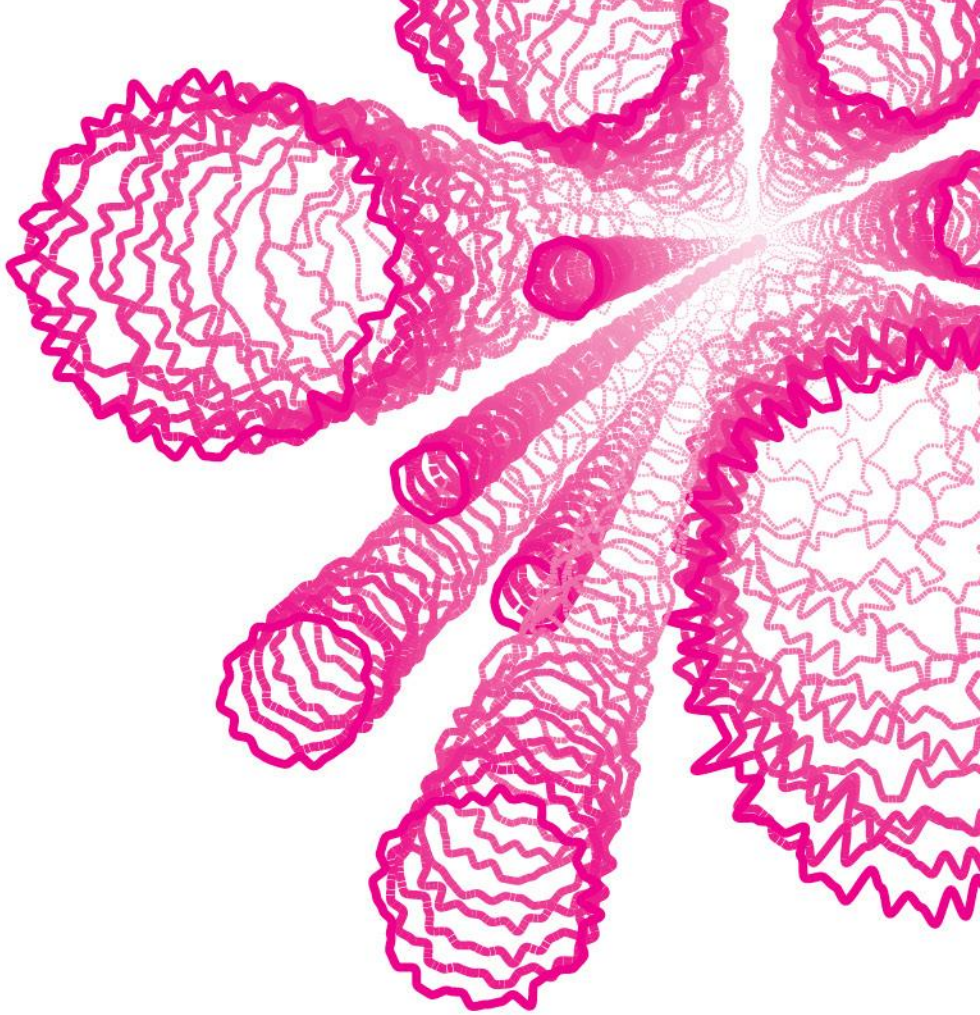
- **Main finding**: difference between odd and even sizes
- **Odd size**: reaches (with difficulty) best nl values
- **Even size**: gets stuck in local optima

RESULTS – BEST NL VALUES

Size	$fitness_1$	$fitness_2$	best-known nonlinearity
	max NL	max NL	
5	**12**	**12**	12
6	24	24	26
7	**56**	**56**	56
8	112	112	116
9	**240**	240	240
10	480	480	492
11	**992**	114	992
12	1984	230	2010
13	4032	422	4036
14	8064	870	8120
15	16256	1664	16272
16	32512	3328	32638

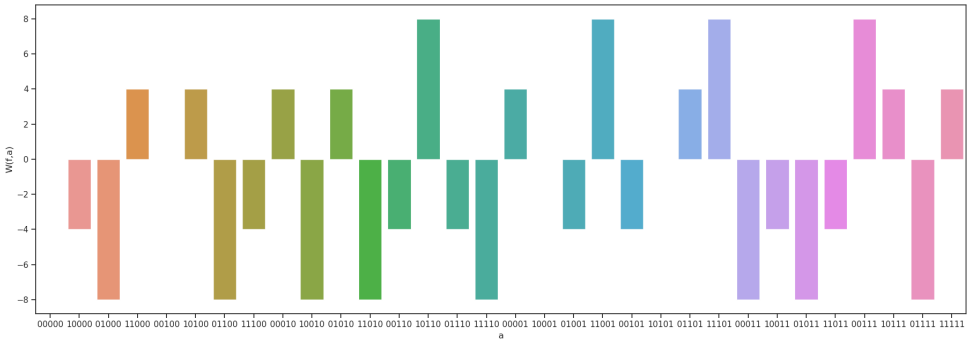
- Fitness 1 performs much better than fitness 2 for larger sizes
- Best-known values of NL reached for odd sizes up to n=11

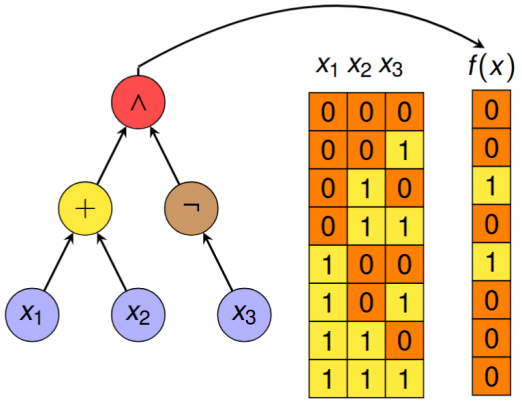
CONCLUSIONS AND FUTURE DIRECTIONS

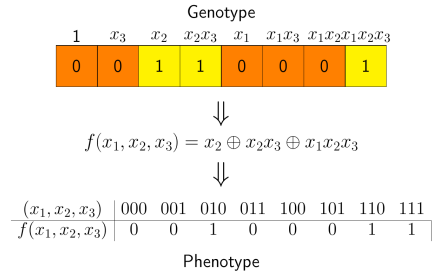


SUMMARY

- **Conclusions:**
  - GP is the best heuristic to evolve 5-valued spectra functions
  - Still, the optimization problem is challenging
- **Future directions:**
  - Experiment with GP variants (e.g. CGP) [5]
  - Adapt *spectral inversion* for plateaued functions to the 5-valued case [3, 11, 15]







UNIVERSITY OF TWENTE.

REFERENCES

1. C. Carlet, M. Djurasevic, D. Jakobovic, L. Mariot, S. Picek: Evolving constructions for balanced, highly nonlinear Boolean functions. Proc. of GECCO 2022, pp. 1147–1155 (2022)
2. C. Carlet: Boolean functions for cryptography and coding theory. Cambridge University Press (2021)
3. J. A. Clark, J. L. Jacob, S. Maitra, P. Stanica: Almost Boolean functions: The design of Boolean functions by spectral inversion. Comput. Intell. 20(3): 450–462 (2004)
4. M. Djurasevic, D. Jakobovic, L. Mariot, S. Picek: A survey of metaheuristic algorithms for the design of cryptographic Boolean functions. Cryptogr. Commun. 15(6): 1171-1197 (2023)
5. R. Hrbacek, V. Dvorak: Bent function synthesis by means of Cartesian Genetic Programming. In: Proc. of PPSN XIII, pp. 414–423 (2014)
6. J. Katz, Y. Lindell: Introduction to Modern Cryptography. Third Edition, CRC Press (2021)
7. J.R. Koza: Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems (Vol. 34). Stanford, CA: Stanford University, Department of Computer Science (1990)
8. S. Maitra, P. Sarkar: Cryptographically significant Boolean functions with five valued Walsh spectra. Theor. Comp. Sci. 276(1–2):133–146 (2002)
9. L. Manzoni, L. Mariot, E. Tuba: Balanced crossover operators in Genetic Algorithms. Swarm Evol. Comput. 54: 100646 (2020)
10. L. Mariot, D. Jakobovic, T. Bäck, J.C. Hernandez-Castro: Artificial Intelligence for the Design of Symmetric Cryptographic Primitives. Security and Artificial Intelligence, pp. 3-24 (2022)
11. L. Mariot, A. Leporati: A genetic algorithm for evolving plateaued cryptographic Boolean functions. In: Proc. of TPNC 2015, pp. 33–45 (2015)
12. L. Mariot, D. Jakobovic, A. Leporati, S. Picek: Hyper-bent Boolean Functions and Evolutionary Algorithms. Proc. of EuroGP 2019, pp. 262-277 (2019)
13. W. Millan, J. Clark, E. Dawson: Heuristic Design of Cryptographically Strong Balanced Boolean Functions. Proc. of EUROCRYPT 1998, pp. 489–499 (1998)
14. S. Picek, D. Jakobovic, J. F. Miller, L. Batina, M. Cupic: Cryptographic Boolean functions: One output, many design criteria. Appl. Soft Comput. 40: 635-653 (2016)
15. L. Rovito, A. D. Lorenzo, L. Manzoni: Discovering non-linear Boolean functions by evolving Walsh transforms with genetic programming. Algorithms 16(11): 499 (2023)