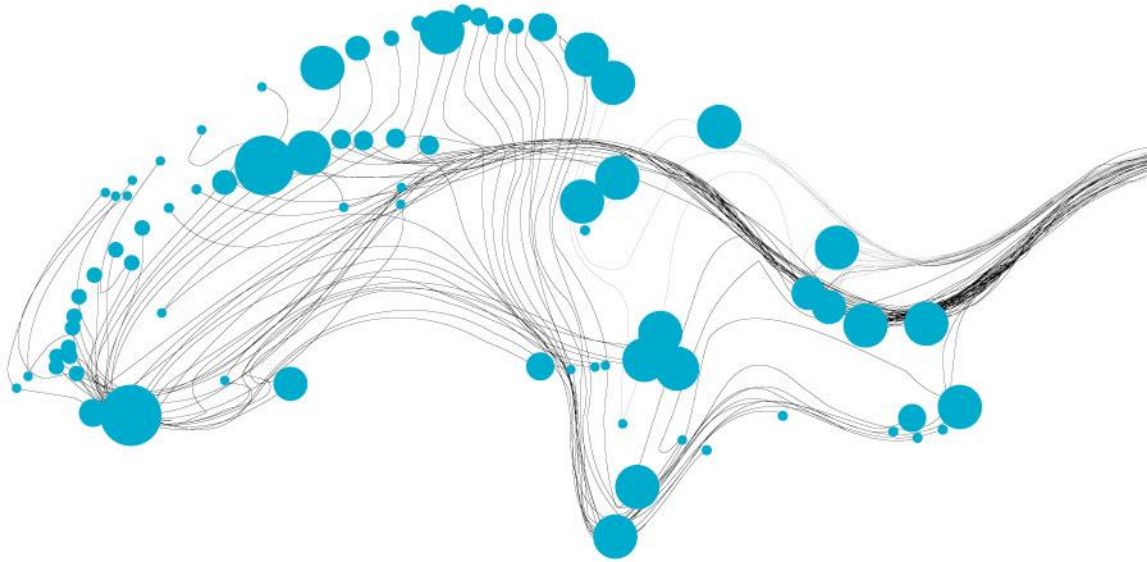
EVOLVING HIGHLY NONLINEAR IDEMPOTENT BOOLEAN FUNCTIONS

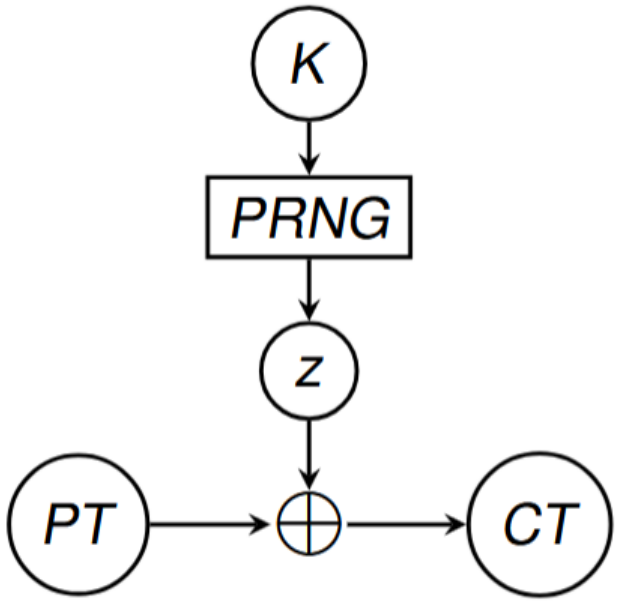
CLAUDE CARLET, MARKO ĐURASEVIĆ, DOMAGOJ JAKOBOVIC, LUCA MARIOT, STJEPAN PICEK



GECCO 2026

SAN JOSE, COSTA RICA, 16 JULY 2026

PRIMITIVES IN SYMMETRIC CRYPTOGRAPHY



(a) Stream cipher

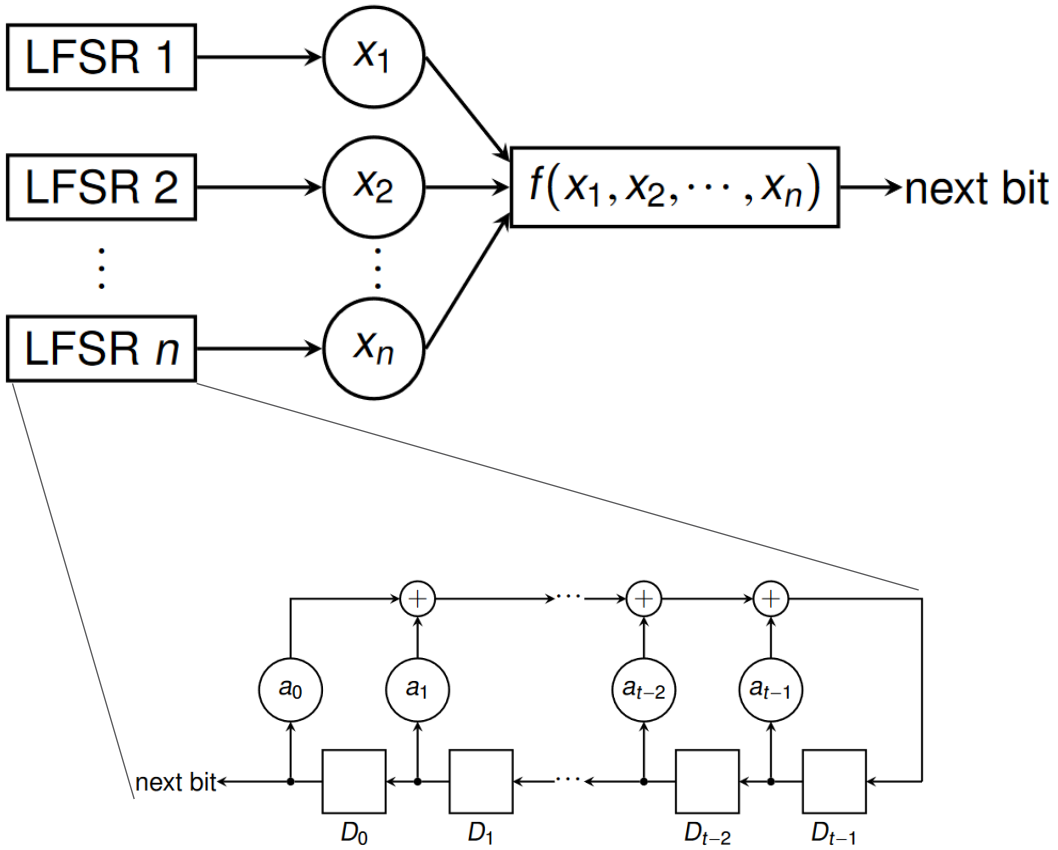
- Symmetric ciphers require several low-level **primitives**:
  - Pseudorandom number generators (PRNG)
  - Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$
  - S-boxes $F : \{0,1\}^n \rightarrow \{0,1\}^n$
- Classic methods to build them: **algebraic constructions** [2]
- Alternative: **(meta)heuristics** (GA, GP, …) [3, 8, 9]

BOOLEAN FUNCTIONS FOR STREAM CIPHERS

- **Idea:** combine the outputs of n Linear Feedback Shift Registers (LFSR) [2]
- LFSRs outputs *combined* using a Boolean function:

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

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



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

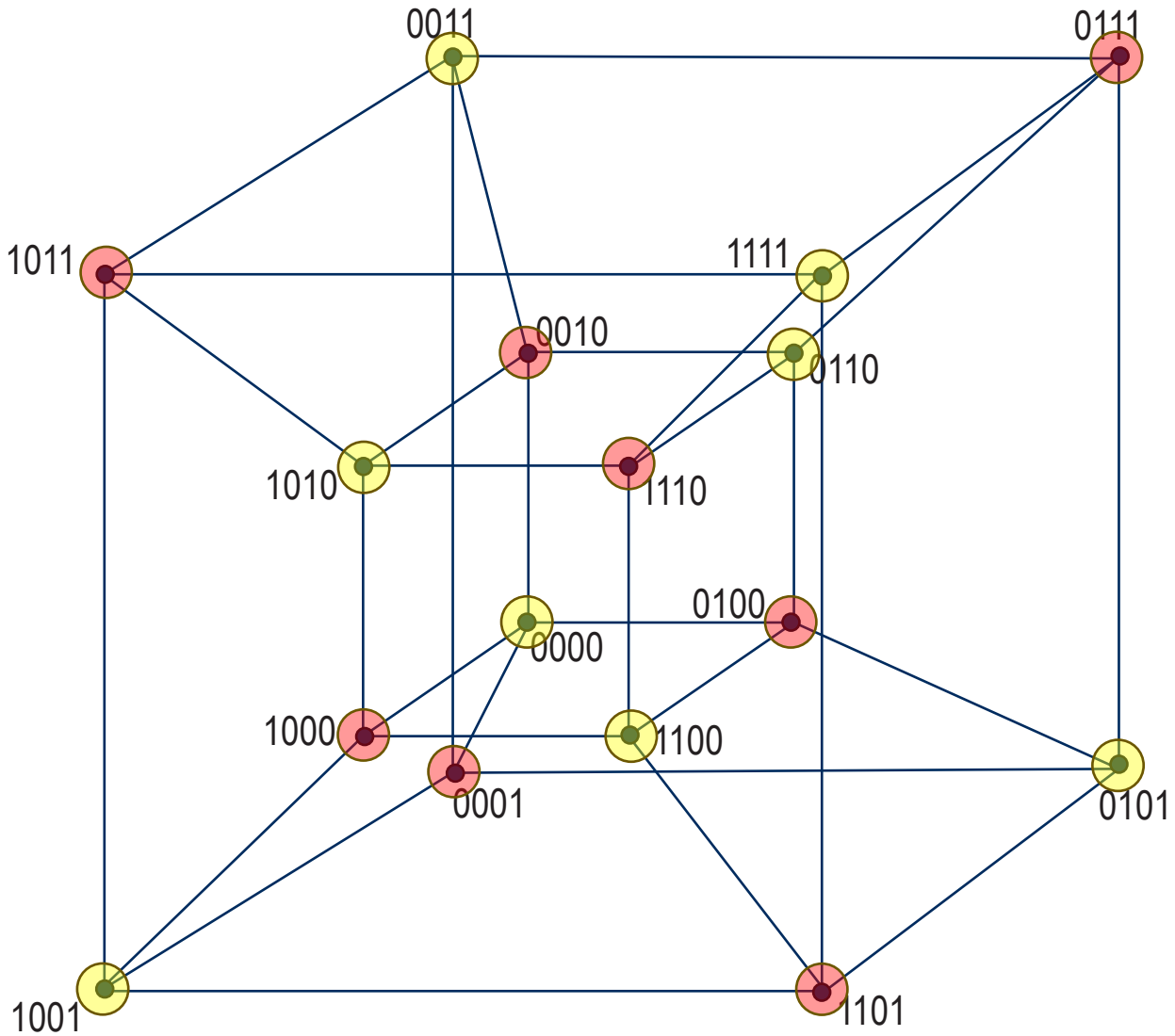
- The function must satisfy some properties to resist specific attacks, e.g.:
  - High **Nonlinearity**: high Hamming distance from **linear** functions:

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

- **Walsh Transform:** Measures correlation with linear functions [2]:

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

NONLINEARITY – GEOMETRIC INTUITION

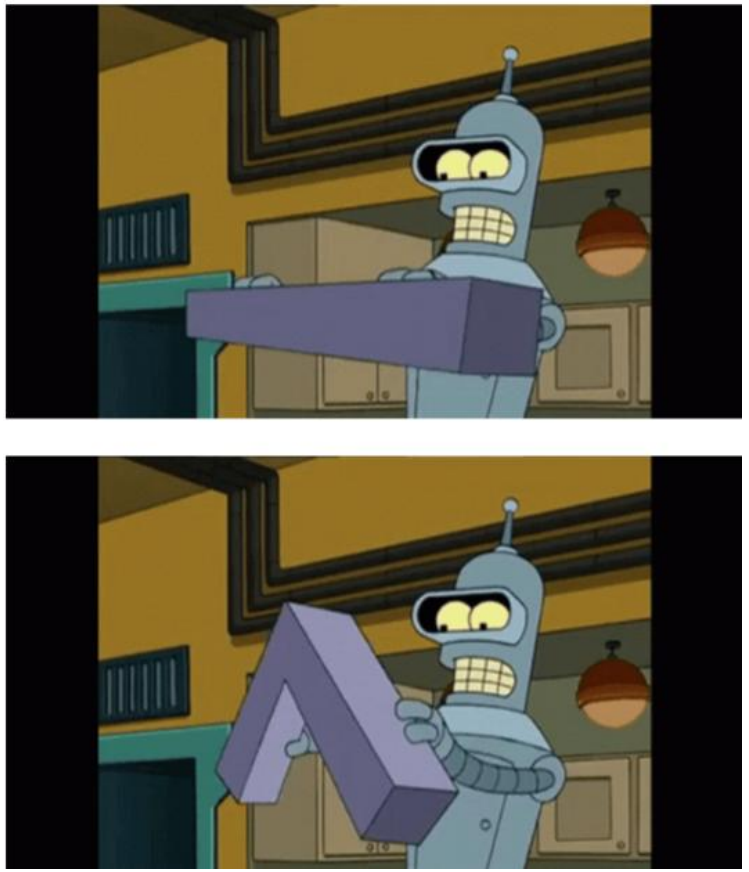


- **Example**: $n = 2$ variables (truth tables of $2^2 = 4$ bits, 16 functions in total)
- Linear functions and complements:

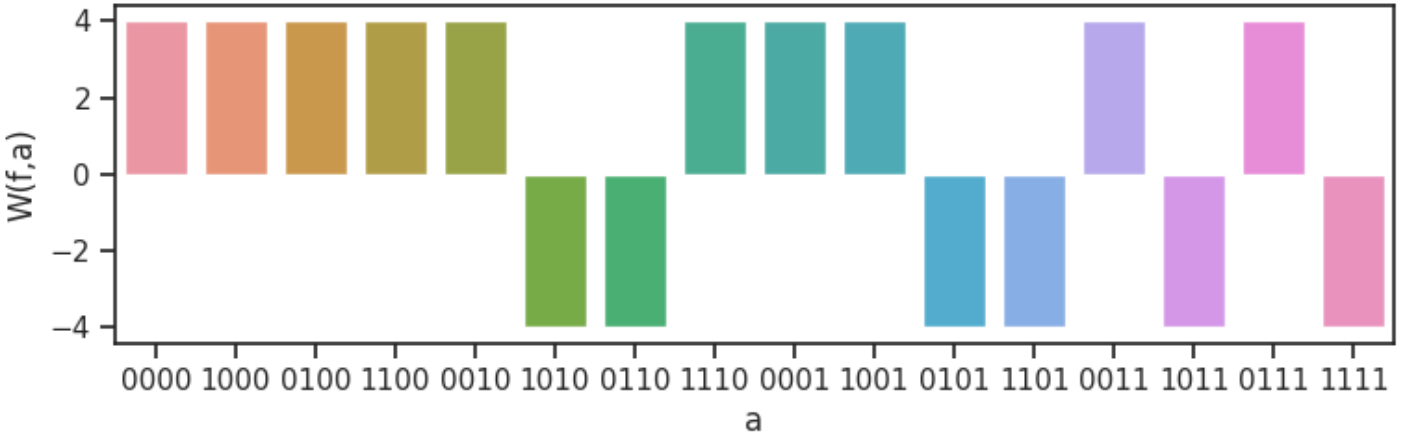
$a_1x_1 \oplus a_2x_2$	TT	TT $\oplus 1$
0	0000	1111
x_1	0011	1100
x_2	0101	1010
$x_1 \oplus x_2$	0110	1001

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

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, 7]



Example with $n = 4$ variables

POLYNOMIAL BASIS

- Represents space of n-bit strings as a finite field $\mathbb{F}_{2^n}$, reducing all elements modulo an irreducible polynomial [2]

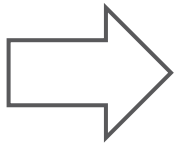
Example with n=3, irreducible $x^3 + x + 1$

Exponent k	α^k	Binary $(x_2x_1x_0)$	Integer
$-\infty$	0	000	0
0	1	001	1
1	α	010	2
2	α^2	100	4
3	$\alpha + 1$	011	3
4	$\alpha^2 + \alpha$	110	6
5	$\alpha^2 + \alpha + 1$	111	7
6	$\alpha^2 + 1$	101	5

IDEMPOTENT FUNCTIONS

- $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$ is **idempotent** if it is invariant under the **Frobenius map** $x \mapsto x^2$
- That is, $f(x) = f(x^2) \qquad \forall x \in \mathbb{F}_{2^n}$
- Essentially: f gets the same output bit for each **orbit** $\mathcal{O}(x) = \{x, x^2, x^{2^2}, \dots\}$

Orbit	Elements	Size
$\mathcal{O}(0)$	$\{0\}$	1
$\mathcal{O}(1)$	$\{1\}$	1
$\mathcal{O}(2)$	$\{2, 4, 6\}$	3
$\mathcal{O}(3)$	$\{3, 5, 7\}$	3



x	0	1	2	3	4	5	6	7
$f(x)$	0	1	1	0	1	0	1	0

Example with n=3, irreducible $x^3 + x + 1$

OPTIMIZATION PROBLEM

- Given $n \in \mathbb{N}$, how do we fill the table so that f is idempotent and highly nonlinear?

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

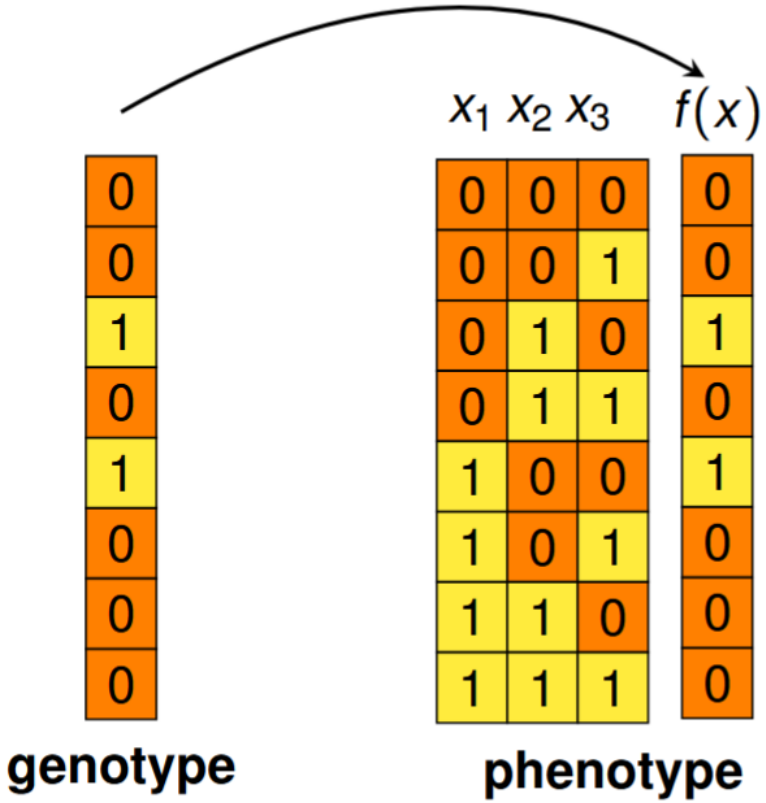
$$f^* = \text{argmax}_{f:\{0,1\}^n \to \{0,1\}}(nl_f), \quad nl_f \leq 2^{n-1} - 2^{\frac{n}{2}-1} \qquad f(x) = f(x^2)$$

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

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

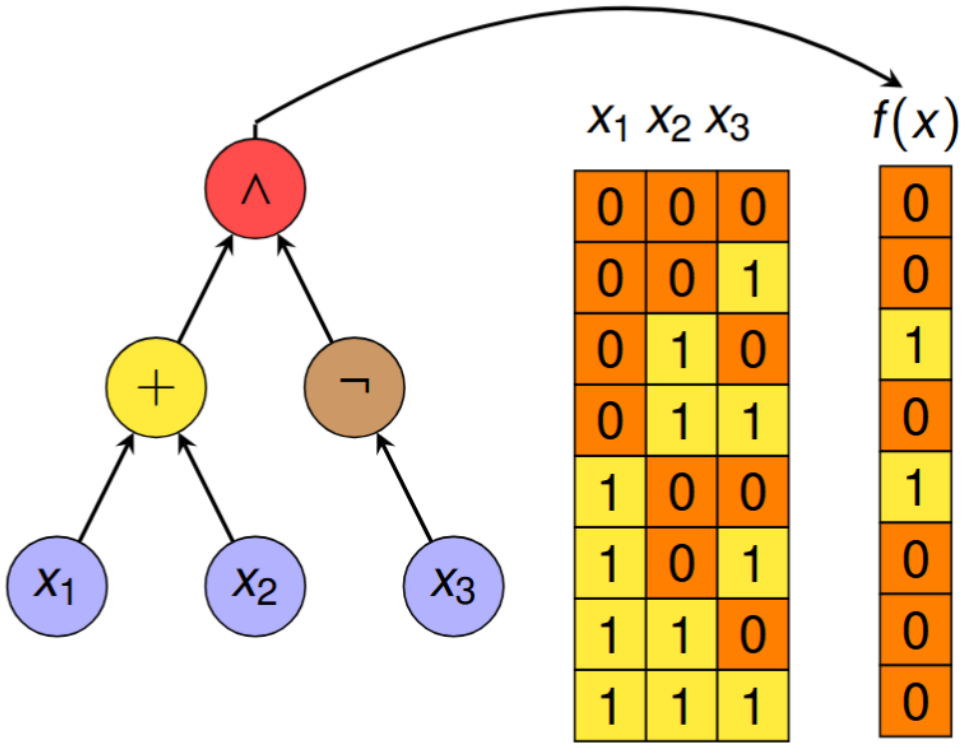
- Exhaustive search unfeasible for $n > 5$, optimum unknown for $n > 7$ odd [2, 3]

BITSTRING ENCODING FOR GENETIC ALGORITHMS



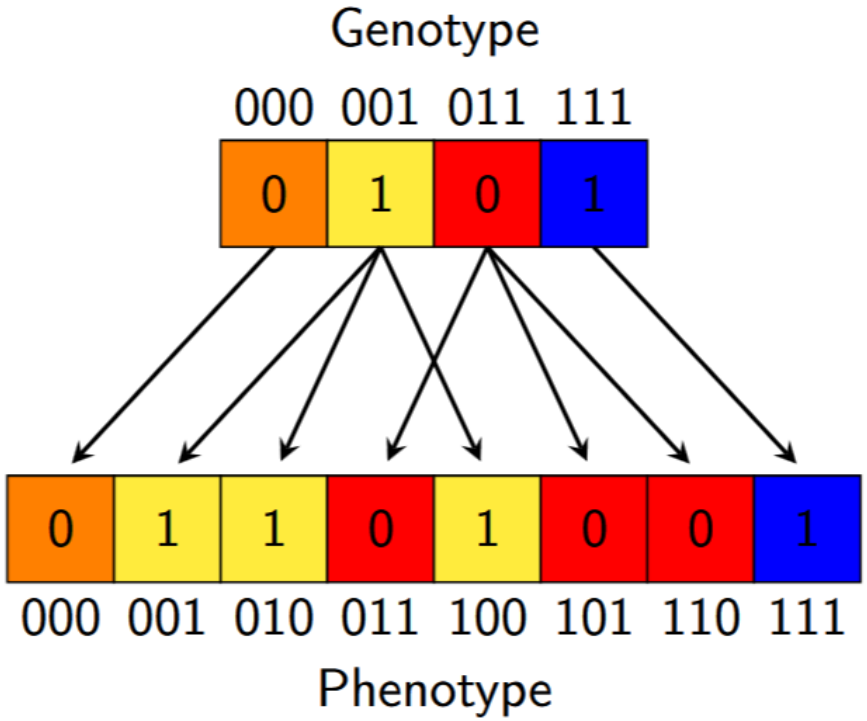
- **GA genotype**: fixed-length bitstrings [10]
- **phenotype**: truth table of f
- **Assumption:** the input vectors of $\{0,1\}^n$ are sorted lexicographically
- Usual variation operators of GA (one-point crossover, bit-flip mutation) [6, 11]

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, …) [5, 11]

RESTRICTED ENCODING



- Forces f to be idempotent
- Greatly reduce the search space: 2^{g_n}, where

$$g_n = \frac{1}{n} \sum_{t|n} \phi(t) 2^{\frac{n}{t}}$$

variables	4	5	6	7	8	9	10	11	12
orbits	6	8	14	20	36	60	108	188	352

- **GA genotype**: bitstring of length g_n
- **GP genotype**: evaluate the tree only for a single representative of each orbit [1]

FITNESS FUNCTION

- Nonlinearity computation: through the **Walsh-Hadamard Transform** [2]:

$$W_f(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus a \cdot x}, \qquad nl_f = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |W_f(a)|$$

- **Penalty term** for idempotency
- **Fitness functions:** optimize directly idempotency penalty and nonlinearity
  - $fitness_1 : -PEN + \delta_{PEN,0} \cdot (nl_f)$
  - $fitness_2 : -PEN + \delta_{PEN,0} \cdot \left(nl_f + \frac{2^n - \#max\_values}{2^n} \right)$

EXPERIMENTAL SETTING

Parameter	Value
Selection	Steady-state, 3-tournament elimination
Population size	500
Stopping criterion	10^6 evaluations
Mutation probability	$p_{mut} = 0.5$
Independent runs	30
Bitstring crossover	One-point, uniform (randomly selected)
Bitstring mutation	Simple bit mutation, shuffle mutation
GP function set	OR, XOR, AND, IF, NOT
GP crossover	Simple tree, uniform, size fair, one-point, context preserving
GP mutation	Subtree mutation
Local search: solutions treated	1% of population size
Local search: trials per solution	25 (random mutations)
Framework	ECF (C++)

RESULTS – BEST FITNESS VALUES

n	6	7	8	9	10	11	12
GP, $fit1$	28	16	64	2	4	2	1024
GP, $fit2$	24	16	64	2	4	2	4
GP_R, $fit1$	28	56	120	240	488	984	1983
GP_R, $fit2$	28	56	120	240	488	986	1988
GP_R, LS, $fit1$	28	56	120	240	487	984	1984
GP_R, LS, $fit2$	28	56	120	240	486	986	1988
TT, $fit1$	28	56	112	-14	-64	-184	-474
TT, $fit2$	28	56	114	-37	-105	-271	-595
TT_R, $fit1$	28	56	120	240	488	986	1992
TT_R, $fit2$	28	56	120	240	488	988	1992
TT_R, LS, $fit1$	28	56	120	240	488	986	1991
TT_R, LS, $fit2$	28	56	120	240	488	987	1992
Best-known	28	56	120	242	496	996	2016

- **Main finding**: restricted encoding has the best performance overall
- GA with the TT/restricted representation fares better than GP
- Optimum reached only up to 8 variables

RESULTS – FITNESS DISTRIBUTIONS

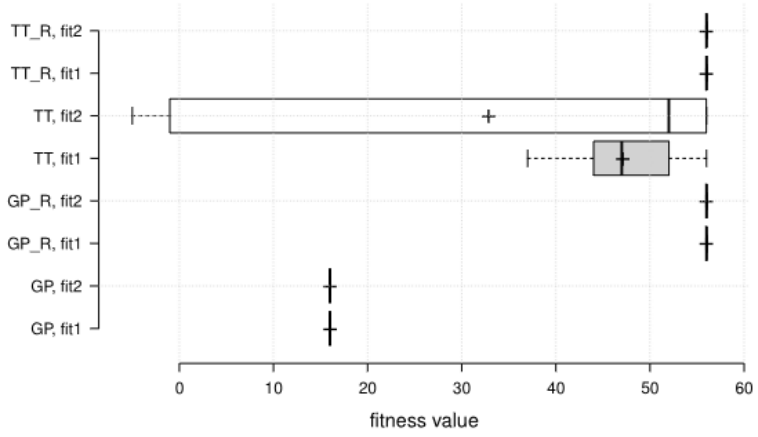


Figure 1: Comparison for $n = 7$ variables

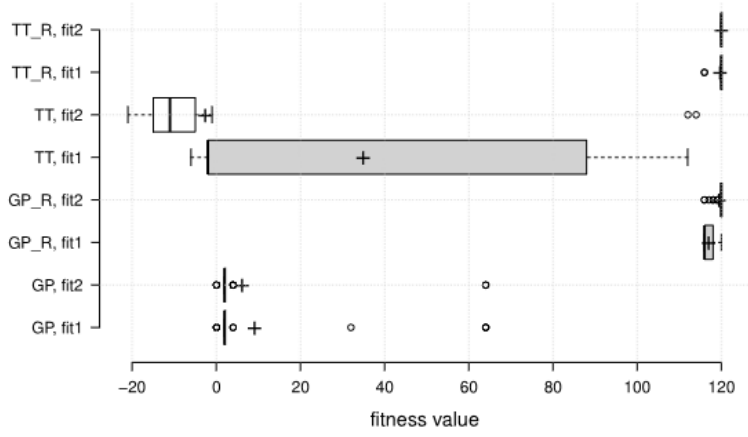


Figure 2: Comparison for $n = 8$ variables

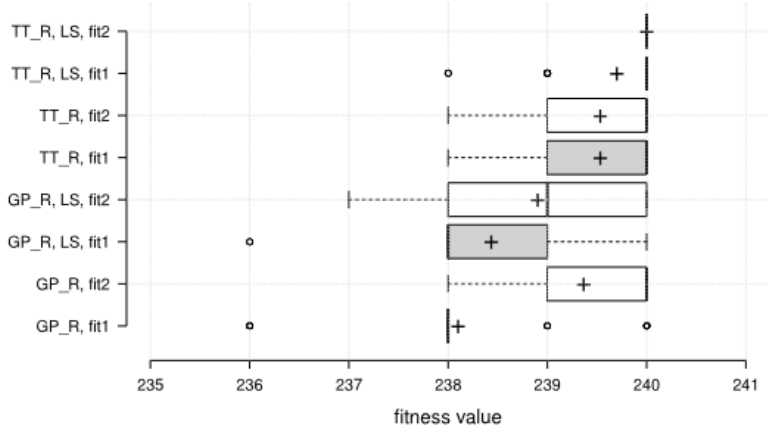
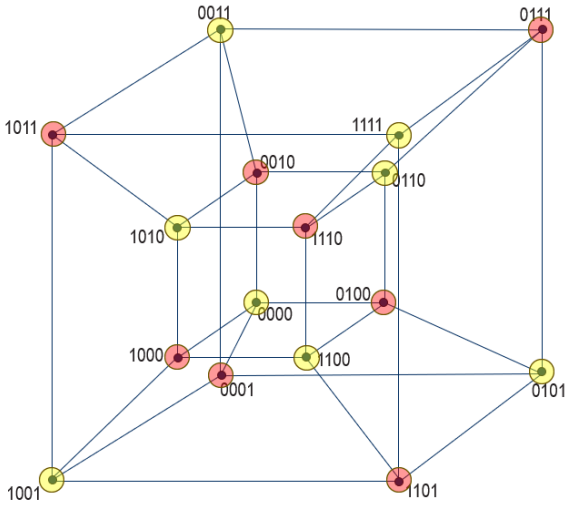


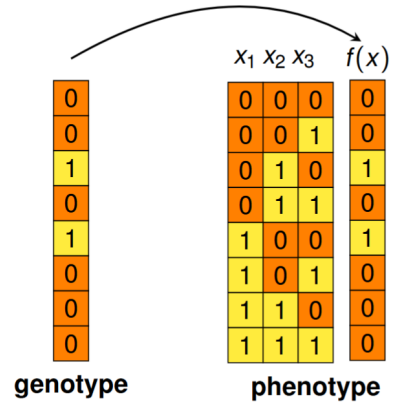
Figure 3: Comparison for $n = 9$ variables

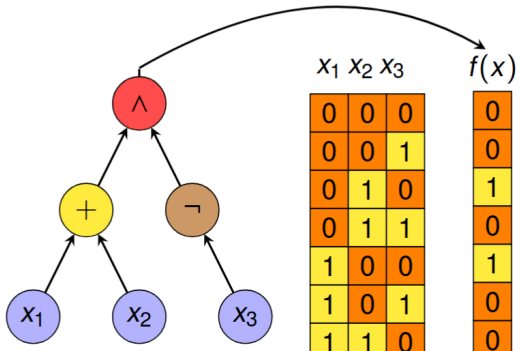
- Unrestricted encoding show huge variance: crossover and mutation are disruptive
- TT_R outperforms GP_R statistically
- Local search does not bring any benefit

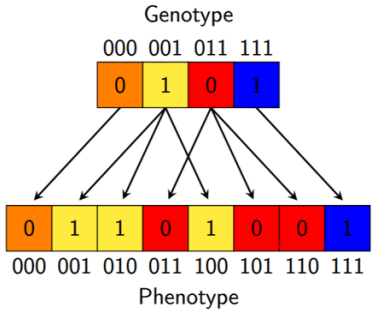
CONCLUSIONS

- **Summary:**
  - Evolving idempotent functions is generally difficult for EA
  - Restricted encoding prevents disruptive nature of crossover and mutation
- **Future directions:**
  - Study the LONs of idempotent functions [4]
  - Consider additional properties (e.g. balancedness) [6]









UNIVERSITY OF TWENTE.

This presentation is based upon work from COST Action Randomised Optimisation Algorithms Research Network (ROAR-NET), CA22137, supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.cost.eu

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

ROAR—NET

REFERENCES

1. C. Carlet, M. Durasevic, B. Gasperov, D. Jakobovic, L. Mariot, S. Picek: A new angle: On evolving rotation symmetric Boolean functions. In: Proc. of EvoApps 2024, pp. 287–302 (2024)
2. C. Carlet: Boolean functions for cryptography and coding theory. Cambridge University Press (2021)
3. 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)
4. D. Jakobovic, S. Picek, M.S.R. Martins, M. Wagner: Toward more efficient heuristic construction of Boolean functions. Appl. Soft Comput. 107:107327 (2021)
5. 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)
6. L. Manzoni, L. Mariot, E. Tuba: Balanced crossover operators in Genetic Algorithms. Swarm Evol. Comput. 54: 100646 (2020)
7. L. Mariot, M. Saletta, A. Leporati, L. Manzoni: Heuristic search of (semi-)bent functions based on cellular automata. Nat. Comput. 21(3): 377-391 (2022)
8. 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)
9. L. Mariot, S. Picek, A. Leporati, D. Jakobovic: Cellular automata based S-boxes. Cryptogr. Commun. 11(1): 41-62 (2019)
10. W. Millan, J. Clark, E. Dawson: Heuristic Design of Cryptographically Strong Balanced Boolean Functions. Proc. of EUROCRYPT 1998, pp. 489–499 (1998)
11. 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)