# Security of the STARK-friendly hash functions

Anne Canteaut, Tim Beyne, Itai Dinur, Maria Eichlseder, Gregor Leander, Gaetan Leurent, Maria Naya Plasencia, Léo Perrin, Yu Sasaki, Yosuke Todo, et al.

# Security of the STARK-friendly hash functions

Tim Beyne, Anne Canteaut, Itai Dinur, Maria Eichlseder, Gregor Leander,
Gaëtan Leurent, Léo Perrin, María Naya Plasencia, Yu Sasaki,
Yosuke Todo, Friedrich Wiemer

Dagstuhl seminar - January 20, 2020

# Motivation

ZK-STARK protocol is expected to be deployed on top of the Ethereum blockchain within the next year

$\longrightarrow$ its security and performance highly depend on the underlying hash function.

**Performance.** SFH are specified as sequences of low-degree polynomials or low-degree rational maps over a finite field.

**Security.**

- algebraic attacks based on Gröbner basis [Albrecht et al. 19]...
- all other cryptanalytic techniques.

# MPC-friendly, Snark-friendly and Stark-friendly primitives

**Objectives:**

- minimize the number of multiplications in large fields
- minimize the size of the polynomial relations representing the execution trace over a finite field.

**Examples:**

- Cradic [Knudsen Nyberg 92], Misty [Matsui 97]
- MiMC [Albrecht et al. 16]

# SFH contenders

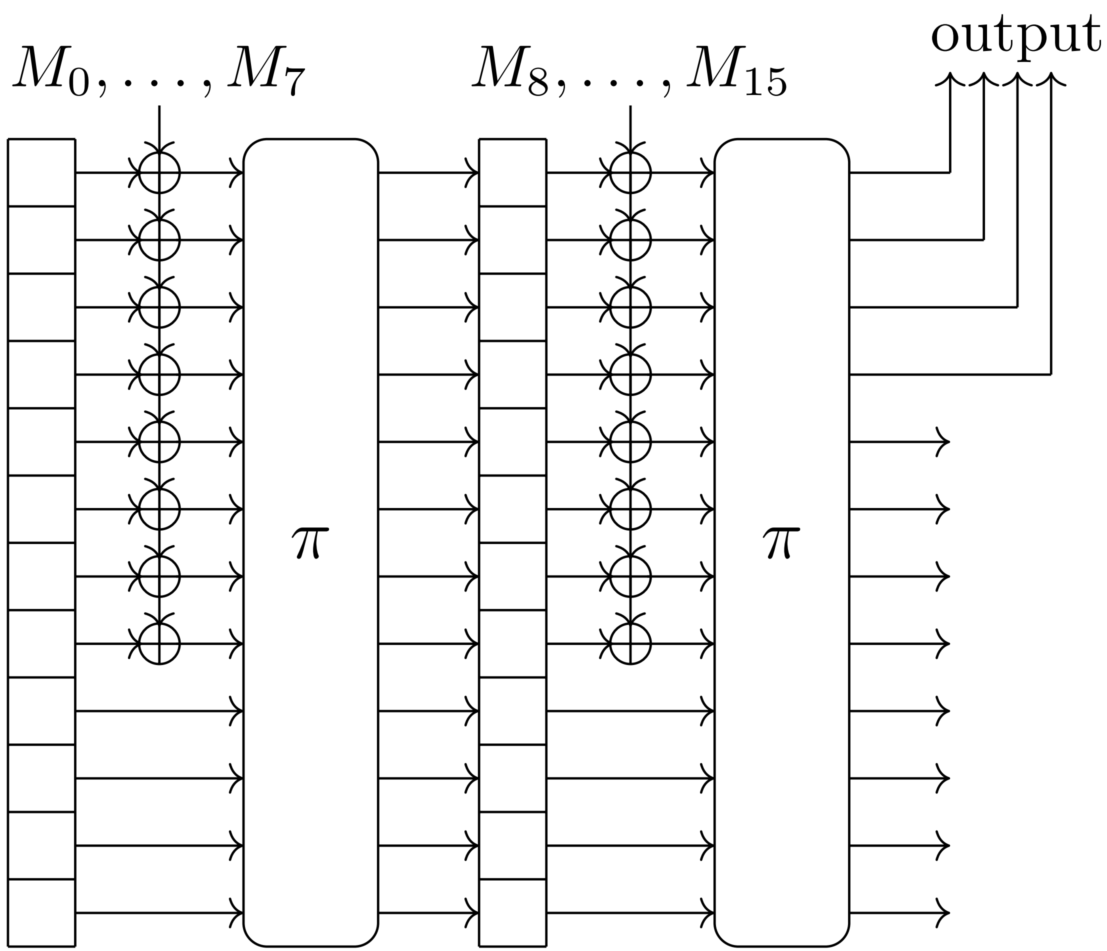
StarkWare challenges `https://starkware.co/hash-challenge/`

Three families of sponges with different permutations

- SPN with large blocks: Vision ($\mathbb{F}_{2^n}$) and Rescue ($\mathbb{F}_p$) [Aly et al. 19]
- HadesMiMC permutation: Starkad ($\mathbb{F}_{2^n}$) and Poseidon ($\mathbb{F}_p$) [Grassi et al. 19]
- GMiMC i.e. $\text{GMiMC}_{\text{erf}}$ over $\mathbb{F}_p$ [Albrecht et al. 19]

# Sponge construction

All candidates follow the same sponge construction with blocksize $t$ and capacity $c$.



# Parameters of the sponge

| security level | $\log_2 q$ | $c$ | $t$ | |
|---|---|---|---|---|
| 128 bits | 64 | 4 | 12 | variant 128-d |
| | 128 | 2 | 4 | variant 128-a |
| | 128 | 2 | 12 | variant 128-c |
| | 256 | 1 | 3 | variant 128-b |
| | 256 | 1 | 11 | variant 128-e |
| 256 bits | 128 | 4 | 8 | variant 256-a |
| | 128 | 4 | 14 | variant 256-b |

# Performance for $128$-bit security

**Best candidate:**

Variant 128-d:

$$t = 12 \text{ and } c = 4 \text{ over } \mathbb{F}_q$$

$$q = \begin{cases} 2^{63} \\ 2^{61} + 20 \times 2^{32} + 1 \end{cases}$$

## Compared performance for these parameters

prime fields are more STARK-friendly than binary fields

### Prime field:

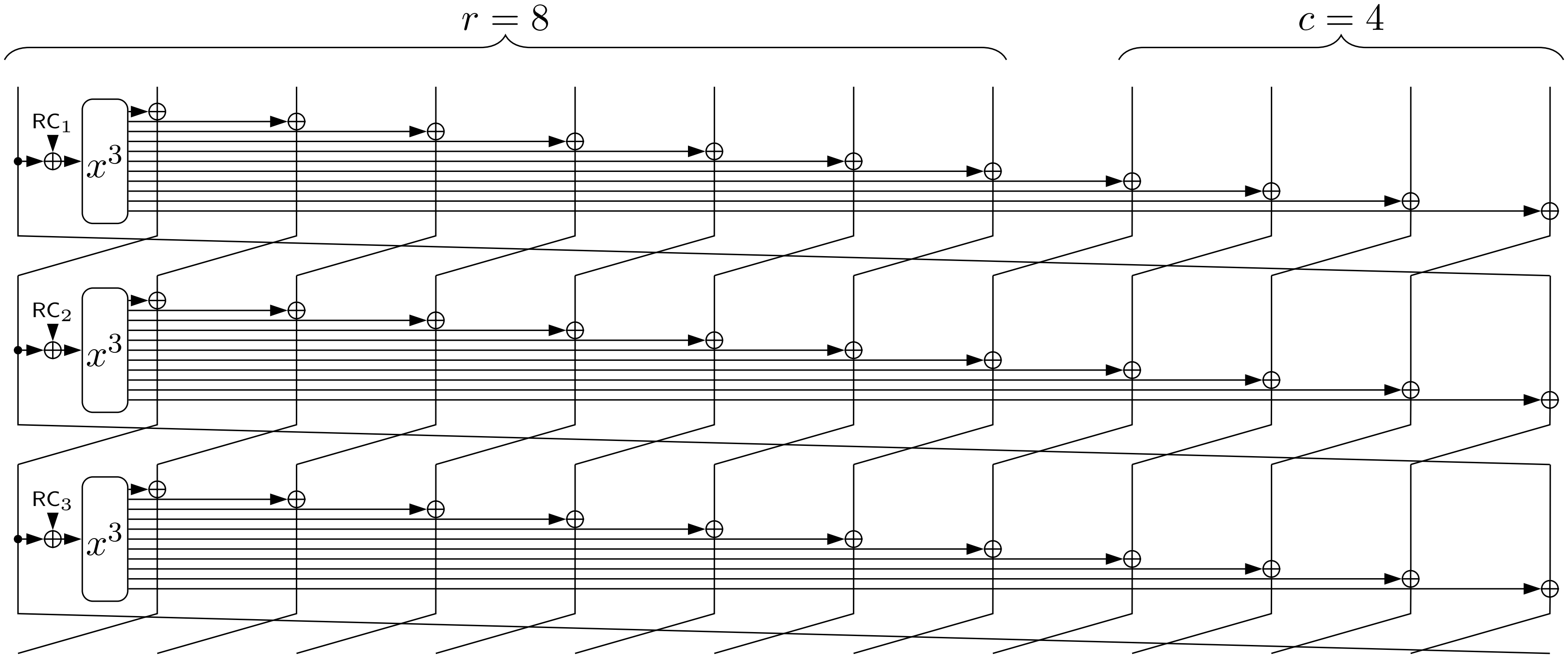
1. GMiMC
2. Rescue
3. Poseidon

### Binary field:

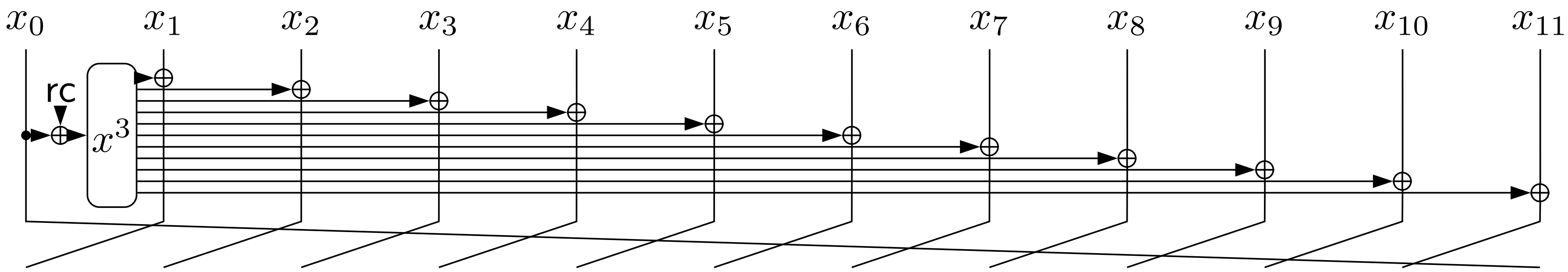
1. Vision
2. Starkad

# GMiMC

# GMiMC with $101$ rounds



## A differential distinguisher




### Original analysis:

best attack with a characteristic over $(t+1)$ rounds with probability $\left(2q^{-1}\right)^2$.

### A better differential:

$$(0\ldots 0, \alpha, \alpha') \overset{\mathcal{R}^{t-2}}{\rightarrow} (\alpha, \alpha', 0\ldots 0) \overset{\mathcal{R}}{\rightarrow} (\alpha'+\beta, \beta \ldots \beta, \alpha) \overset{\mathcal{R}}{\rightarrow} (\beta+\beta' \ldots \beta+\beta', \alpha+\beta', \alpha'+\beta)$$

For $\beta' = -\beta$ , we get an iterative differential

$$(0\ldots 0, \alpha, \alpha') \xrightarrow{\mathcal{R}^t} (0, \ldots, 0, \alpha+\beta, \alpha'+\beta) \text{ with probability } 2q^{-1}$$

# A differential distinguisher

With this $t$-round differential with proba $2q^{-1}$

- A differential characteristic over the $101$ rounds with probability $2^{-480}$ for a $732$-bit blocksize.
- With structures, we get valid pairs with complexity $2^{359}$ (full permutation) and valid pairs with complexity less than $2^{128}$ for 58 rounds.
- With a rebound-like technique, we expect to get valid pairs conforming with the differential over $58$ rounds with complexity close to $2^{64}$ (on-going work).

# Impossible differentials

**Original analysis:** best impossible differential over $(2t-2)$ rounds

**A better impossible differential over $(3t-4)$:**

$$(0,\ldots,0,\alpha) \overset{\mathcal{R}^{3t-4}}{\nrightarrow} (\beta,0,\ldots,0)$$

for any nonzero $\alpha \neq \beta$.

**When $q = 2^n$.**

For any (affine) subspace $V \subset \mathbb{F}_2^n$ with $\dim V > \deg F$,

$$\sum_{x \in V} F(x) = 0.$$

# Integral attacks over $\mathbb{F}_q$

**When $q = 2^n$.**

For any (affine) subspace $V \subset \mathbb{F}_2^n$ with $\dim V > \deg F$,

$$\sum_{x \in V} F(x) = 0.$$

Because, for $V = b + \langle a_1, \ldots, a_v \rangle$,

$$D_{a_1} D_{a_2} \ldots D_{a_v} F(b) = \sum_{x \in V} F(x)$$

Not valid in odd characteric.

## But for any $q$

For any exponent $k$ with $0 \leq k \leq q-2$,

$$\sum_{x \in \mathbb{F}_q} x^k = 0$$

**General result.**

For any $F : \mathbb{F}_q \to \mathbb{F}_q$ with $\deg(F) \leq q-2$,

$$\sum_{x \in \mathbb{F}_q} F(x) = 0 \; .$$

## But for any $q$

For any exponent $k$ with $0 \le k \le q-2$,

$$\sum_{x \in \mathbb{F}_q} x^k = 0$$

**General result.**

For any $F : \mathbb{F}_q \to \mathbb{F}_q$ with $\deg(F) \le q-2$,

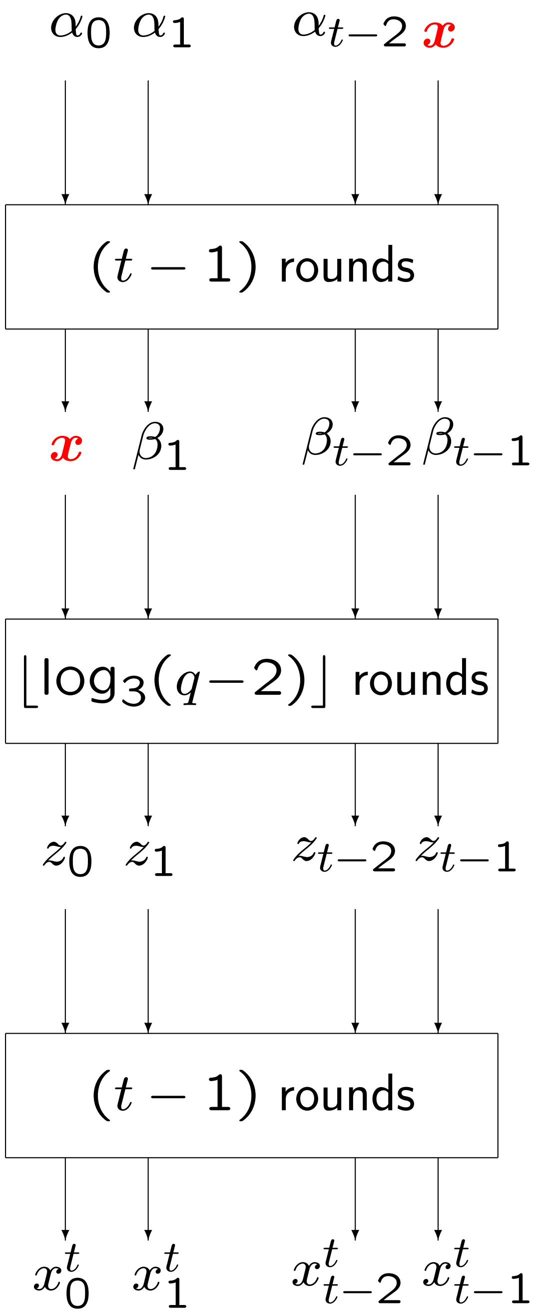
$$\sum_{x \in \mathbb{F}_q} F(x) = 0 \; .$$

**Less general than the property over $\mathbb{F}_{2^n}$:**

For any (affine) subspace $V \subset \mathbb{F}_2^n$ with $\dim V > \deg F$,

$$\sum_{x \in V} F(x) = 0$$

# Integral distinguisher on GMiMC




$x \in \mathbb{F}_q$

polynomial in $x$
of degree $\leq q-2$

$$Q(x) = \sum_{i=1}^{t-1} x_i^t - (t-2)x_0^t$$

## Until the degree does not exceed $(q-2)$

**Input set.**

$$\mathcal{X} = \{(\alpha_0, \ldots, \alpha_{t-2}, x), x \in \mathbb{F}_q\}$$

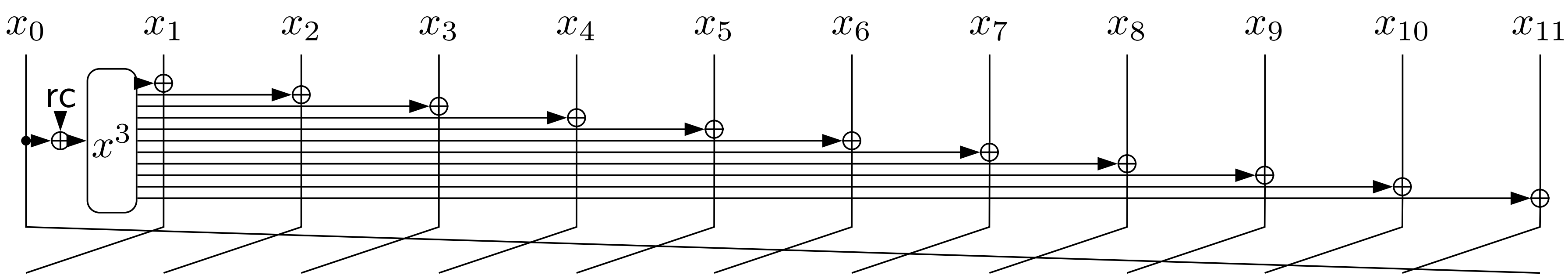
**After $(t-1)$ rounds.**

$$\mathcal{X}' = \{(x, \beta_1, \ldots, \beta_{t-1}), x \in \mathbb{F}_q\}$$

After $r$ rounds, the degree in $x$ of each branch is at most $3^r$.

$\Rightarrow$ all branches are balanced if $3^r \leq q-2$.

## Adding $(t-2)$ rounds




The inputs and outputs of Round $\ell$ satisfy

$$x_i^{\ell} - x_{i+1}^{\ell} = x_{i+1}^{\ell-1} - x_{i+2}^{\ell-1}, \quad \forall i \leq t-3$$

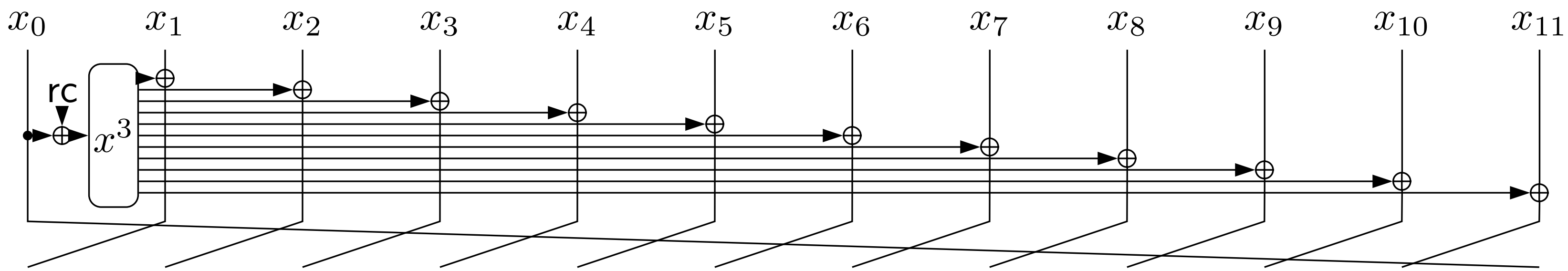
Over $(t-2)$ rounds,

$$x_0^{t-1} - x_1^{t-1} = x_{t-2}^{1} - x_{t-1}^{1}$$

is a polynomial in $x$ of degree $\leq (q-2)$.

$\Rightarrow$ Distinguisher with complexity $q$ on $(2t-3+\lfloor \log_3(q-2) \rfloor)$ rounds (59 rounds)

# Adding one more round




The inputs and outputs of Round $\ell$ satisfy

$$x_i^\ell = x_{i+1}^{\ell-1} + (x_j^\ell - x_{j+1}^{\ell-1}) \text{ and } x_{t-1}^\ell = x_0^{\ell-1}$$

$$\Rightarrow \sum_{i-0}^{t-1} x_i^{\ell-1} - (t-1)x_j^t = \sum_{i=0}^{t-1} x_i^\ell - (t-1)x_{j-1}^\ell$$

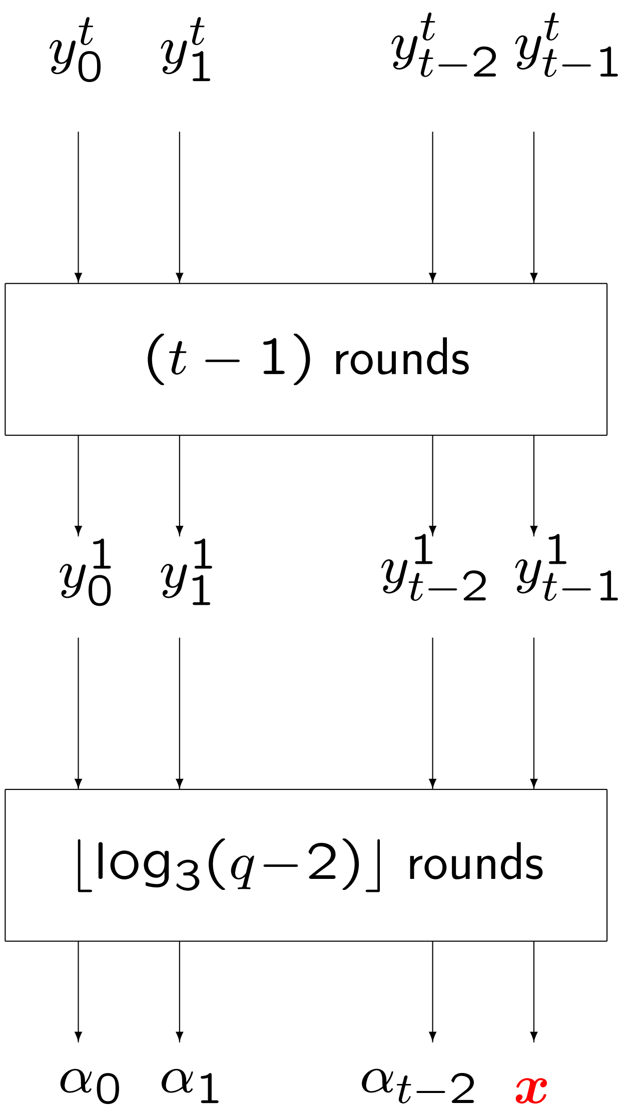
**Over $(t-1)$ rounds,**

$$\sum_{i=0}^{t-1} x_i^1 - (t-1)x_{t-1}^1 = \sum_{i=0}^{t-1} x_i^t - (t-1)x_0^t$$

$\Rightarrow$ Distinguisher with complexity $q$ on $(2t-2+\lfloor\log_3(q-2)\rfloor)$ rounds (60 rounds)

A few more rounds with two active branches (on-going work).

## Computing backwards

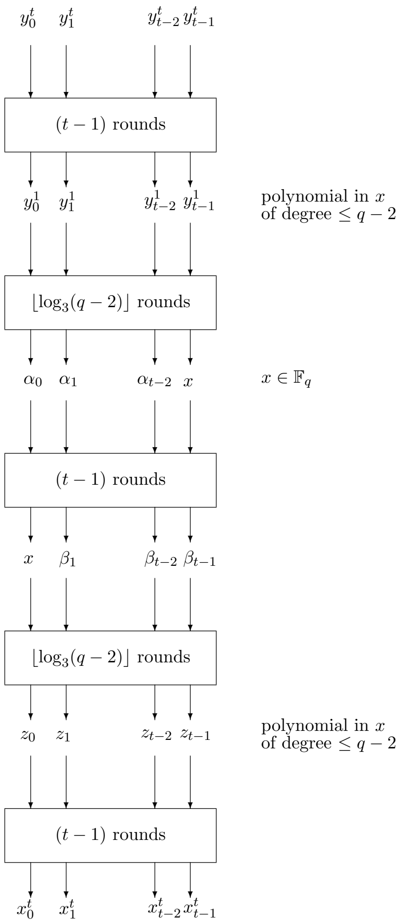



$$Q'(x) = \sum_{i=0}^{t-2} y_i^t - (t-2)y_{t-1}^t$$

$x \in \mathbb{F}_q$

## Zero-sum partition on GMiMC on $(3t-3+2\lfloor\log_3(q-2)\rfloor)$ rounds (109)

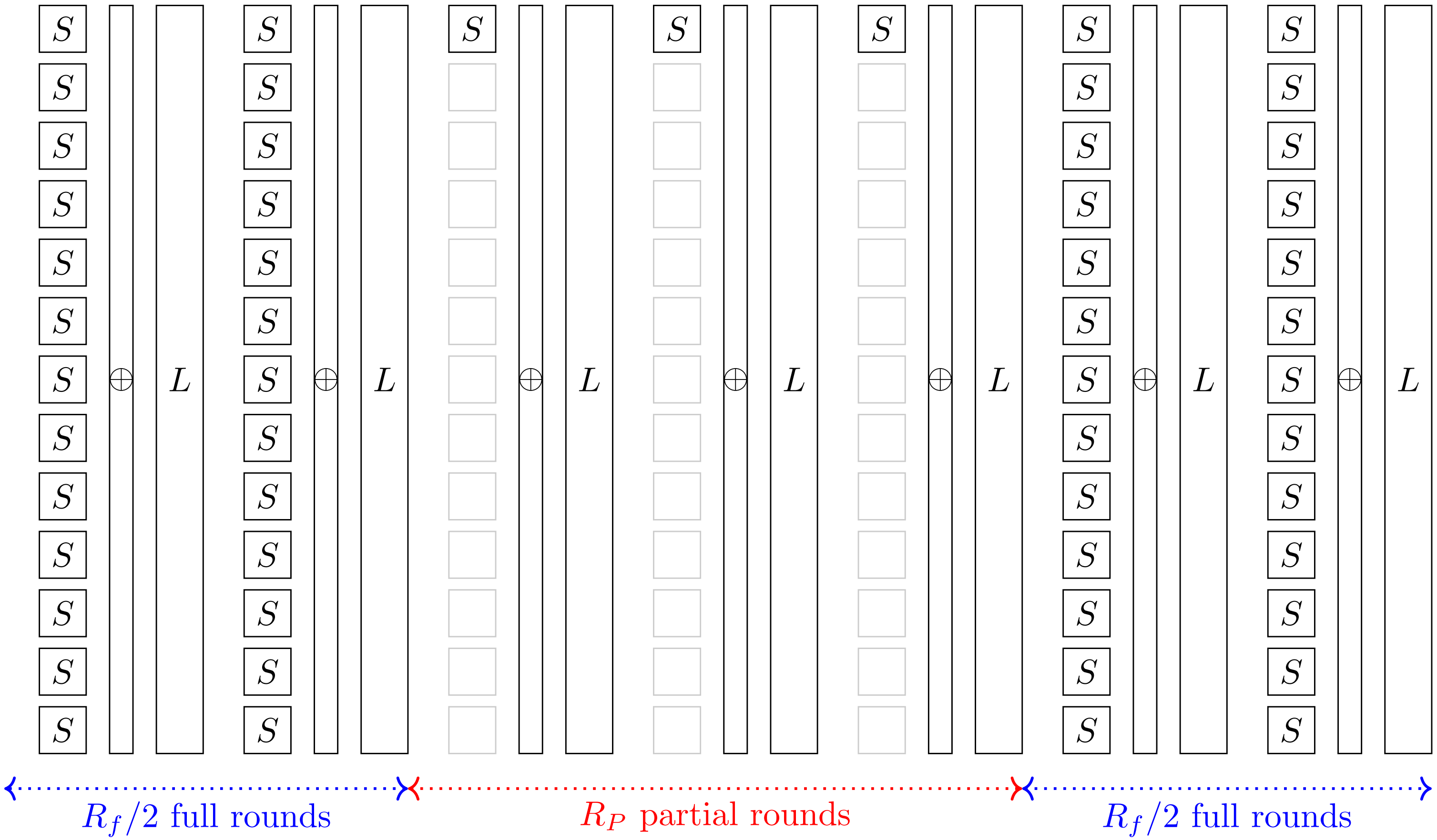
$\ell(x_0,\ldots,x_{t-1}) = \sum_{i=1}^{t-1} x_i - (t-2)x_0$ sum to $0$




$\ell'(y_0,\ldots,y_{t-1}) = \sum_{i=0}^{t-2} y_i - (t-2)y_{t-1}$ sum to $0$

# HadesMiMC

# HadesMiMC




$\boldsymbol{R_f} = \mathbf{8}$ full rounds and $\boldsymbol{R_P} = \mathbf{43}$ (binary) and $\boldsymbol{R_P} = \mathbf{40}$ (prime)

# Resistance against statistical attacks

Analysed without the partial rounds.

## Differential cryptanalysis:

$x^3$ has differential uniformity $2$ over $\mathbb{F}_q$.

The best differential characteristic satisfies

$$\mathrm{EDP} \leq \left(\frac{2}{q}\right)^{(t+1)R_f/2}$$

$\rightarrow R_f = 6$ are enough.

# Degree of the permutation over $\mathbb{F}_q$

Each coordinate is seen as a multivariate polynomial over $\mathbb{F}_q$

**After $r$ rounds:**

$$\sum_{u=(u_1,\ldots,u_t)} \lambda_u \left( \prod_{i=1}^{t} x_i^{u_i} \right) \text{ where } u_i \leq 3^r$$

$\Rightarrow$ 39 rounds are enough for Poseidon (40 for Starkad) to reach degree $(q-1)$ in each variable

$\Rightarrow$ $\lceil \log_3(t) \rceil$ more rounds are enough to get total degree $(q-1)t$.

**Remark:** StarkWare challenges with $q \simeq 2^{256}$ and $96$ rounds have degree at most $2^{152}$ in each variable.

# Zero-sum partition over $\mathbb{F}_q$

**State after the last full Sbox layer before the partial rounds.**

$$\mathcal{X} = \{(\alpha_0, \ldots, \alpha_{t-2}, x), x \in \mathbb{F}_q\}$$

**After 38 rounds forwards.**

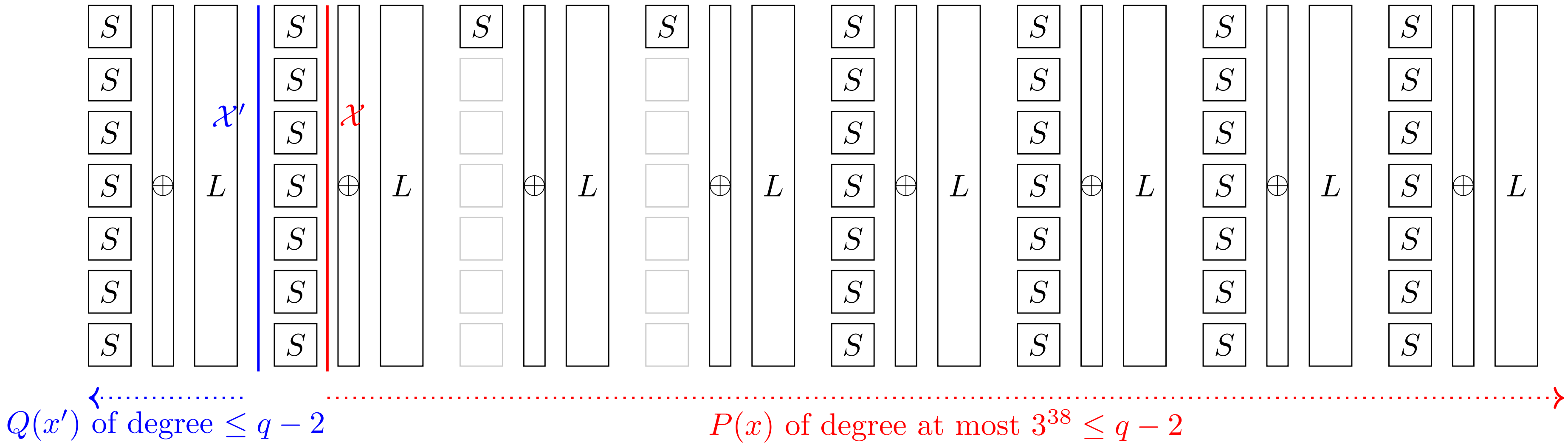
each coordinate has degree at most $(q-2)$.

**Computing backwards.**

$$S^{-1} : x \mapsto x^s \text{ with } s = \frac{2q-1}{3}$$

$\Rightarrow$ Zero-sum for $R_f = 2 + 4$ and $R_P = 34$ ($35$ for Starkad).

# Zero-sum partition over $\mathbb{F}_q$



## Improvement when $q = 2^n$

Each Boolean coordinate is seen as a multivariate polynomial in $nt$ variables over $\mathbb{F}_2$

**Degree over $\mathbb{F}_{2^n}$ vs. binary degree.**

$$P(x) = \sum_{u \leq 2^n - 1} \lambda_u x^u$$

has binary degree

$$\max\{wt(u) : 0 \leq u < 2^n \text{ and } \lambda_u \neq 0\}$$

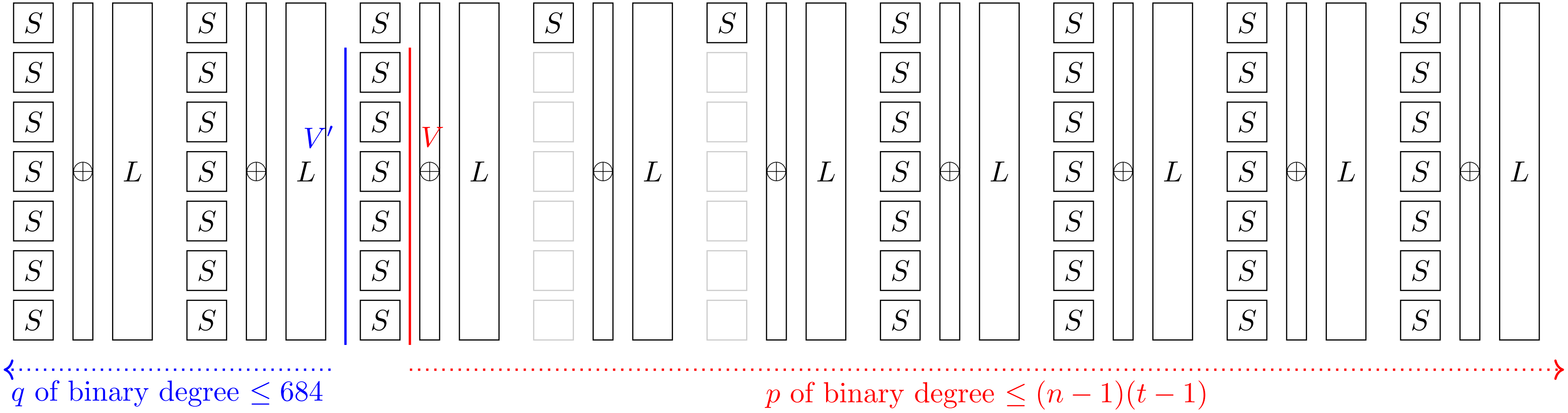
$\Rightarrow$ The inverse Sbox has binary degree $\frac{n+1}{2}$.

**Several rounds backwards [Boura, C. 13].**

- Two rounds backwards have binary degree $\leq 684$
- Three rounds backwards have binary degree $\leq 748$

## Zero-sum partition over $\mathbb{F}_2$ with $R_f = 3 + 4$ and $R_P = 35$

$$V = \{(0, x_1, \ldots, x_{t-1}), x_i \in \mathbb{F}_{2^n}\}\ .$$



## When the MDS matrix has a small order

### How to propagate a subspace through all partial rounds?

Choose $V$ such that all elements in each coset of $L(V)$ have the same value on the first coordinate.

$$L(V) \subset H_0 = \{(0, x_1, \ldots, x_{t-1}), x_i \in \mathbb{F}_q\}$$

or equivalently

$$V \subset \langle M_0 \rangle^{\perp} .$$

We can iterate this $R_P$ times if

$$\mathcal{V} = H_0 \cap \bigcap_{r=0}^{R_P-1} L^r \left( \langle M_0 \rangle^{\perp} \right) \neq \{0\}$$

This holds if $L^r = \mathbf{Id}$ for some $r \leq t - 2$.

## When the MDS matrix is an involution

The internal states after each partial layer form a coset of $V$ or of $W = L(V)$.

### Special choice for $V$.

$$V = \{(xv_0, \ldots, xv_{t-1}), x \in \mathbb{F}_q\}$$

with $v \in \mathcal{V}$.

$\Rightarrow$ The outputs of the partial rounds vary in a coset of

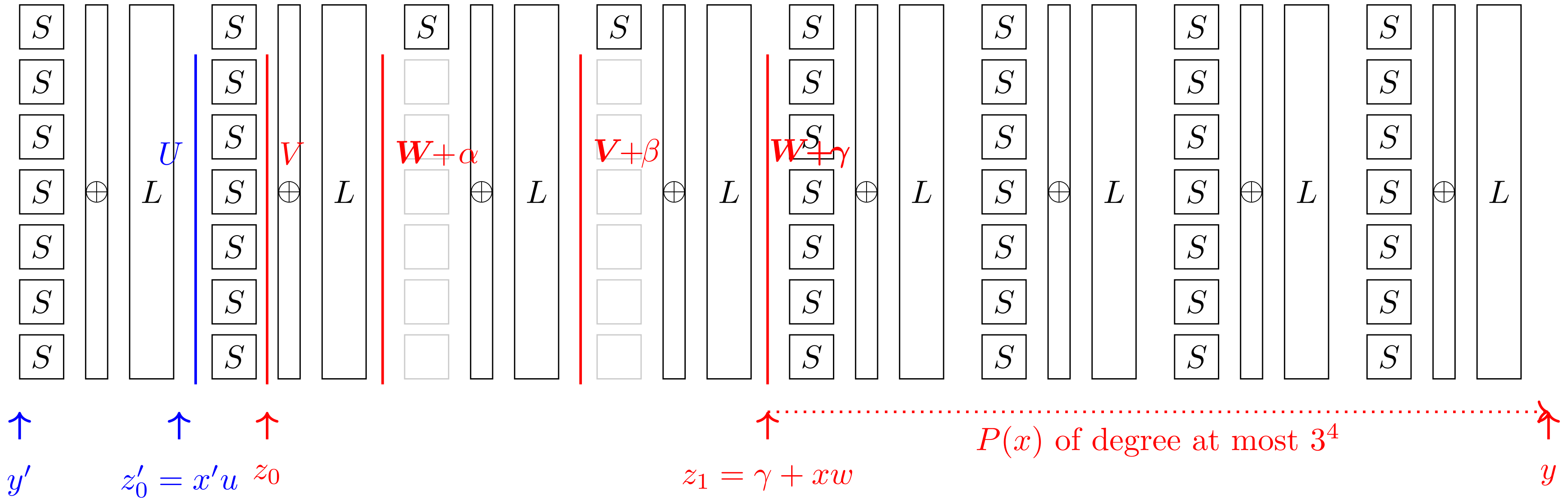
$$\{(xw_0, \ldots, xw_{t-1}), x \in \mathbb{F}_q\}$$

### Forward direction.

Each output coordinate is a polynomial in $x$ of degree at most $3^{R_f/2} \leq q - 2$.

$\Rightarrow$ The output coordinates sum to zero.

# Zero-sum partition with $R_f = 2 + 4$ and any $R_P$ with complexity $q$



# Open question on the complexity of algebraic attacks

**Input:** $(a_1, \ldots, a_{t-k}) \in \mathbb{F}_q^{t-k}$ and $(b_1, \ldots, b_k) \in \mathbb{F}_q^k$
**Find** $x_1, \ldots, x_k \in \mathbb{F}_q^k$ such that
$\pi(a_1, \ldots, a_{t-k}, x_1, \ldots, x_k) = (b_1, \ldots, b_k, y_1, \ldots y_{t-k})$ for some $y_1, \ldots y_{t-k}$

**Degree of the univariate polynomial of the lexicographical Gröbner basis [Faugère-Perret].**

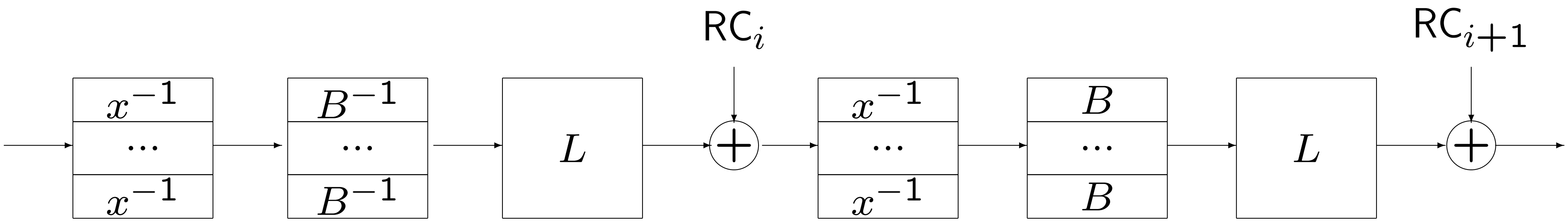
$$D = 3^{kR_f + R_P - 2k + 1}$$

Complexity for solving the system $= D^2$.

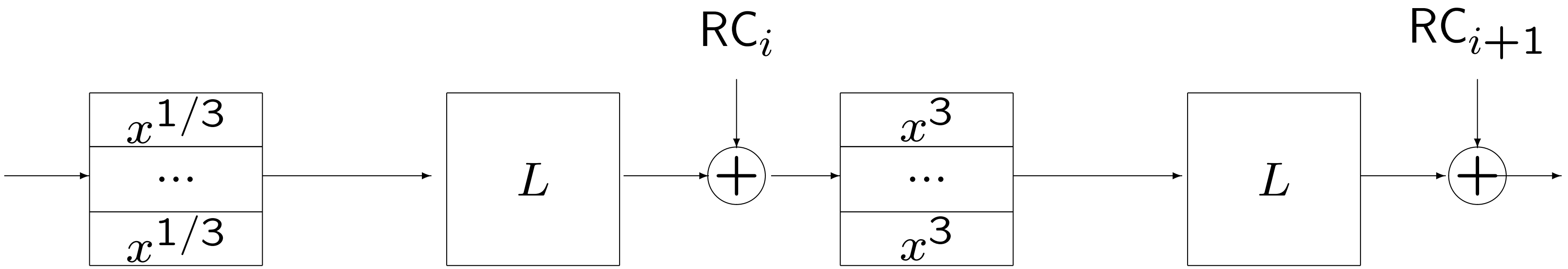
Variants aiming at $256$-bit security have $D \simeq 2^{170}$.

# Vision and Rescue

## Vision (20 rounds)



# Rescue (20 rounds)



# Degree of Rescue

Activate one input coordinate $x \in \mathbb{F}_p$

**After one round.**

$$\lambda x^{1/3} + \mu$$

$\Rightarrow$ degree $\frac{2p-1}{3}$

**After the second Sbox layer.**

$$(\lambda x^{1/3} + \mu)^3$$

which contains only monomials $x^{1/3}$, $x^{2/3}$, $x$ and a constant term.

$x^{2/3}$ has degree $\frac{p+1}{3}$.

$\Rightarrow$ The degree does not increase between the first and second round.

**But even by activiting more inputs,** we cannot find an integral attack on more than 4 rounds.

# Conclusions

We need to find the right tools for analyzing symmetric primitives over non-binary fields:

- linear attacks and their variants?
- more general integral attacks?

**Open question:**

Does the form of $q$ affect the security?
For instance, if $p = 2^{2^n} + 1$?