

# STARK-friendly crypto primitives wish-list

---

Eli Ben-Sasson

 April 2023



# Wishlist for crypto primitives

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

## Wish-list for Primitive P:

1. Minimize  $t \times w$ ,  $|F|$ ,  $d$  (also  $s$ ,  $e$ )
2. Minimize CPU time of computing P
3. Make it field agnostic, and safe!

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

## Wish-list for Primitive P:

1. Minimize  $t \times w$ ,  $|F|$ ,  $d$  (also  $s$ ,  $e$ )
2. Minimize CPU time of computing P
3. Make it field agnostic, and safe!

## ethSTARK (Rescue):

1.  $t \times w = 120$ ,  $\log |F| \sim 62$ ,  $d=3$
2. CPU time large, esp in large  $F$  (cube root!)
3. Relatively safe, AES-like (say the experts)

## Poseidon in Starknet/Ex:

1.  $t \times w = 226$ ,  $\log |F| \sim 252$ ,  $d=2$
2. CPU time better than Rescue (100 microsec)
3. Relatively safe (though less than Rescue)

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

## Keccak:

1.  $t \times w = 70,000-90,000$ ,  $\log |F| \sim 252$ ,  $d=2$
2. CPU time amazing (1 microsecond)
3. Very safe

## Wish-list for Primitive P:

1. Minimize  $t \times w$ ,  $|F|$ ,  $d$  (also  $s$ ,  $e$ )
2. Minimize CPU time of computing P
3. Make it field agnostic, and safe!

## ethSTARK (Rescue):

1.  $t \times w = 120$ ,  $\log |F| \sim 62$ ,  $d=3$
2. CPU time large, esp in large  $F$  (cube root!)
3. Relatively safe, AES-like (say the experts)

## Poseidon in Starknet/Ex:

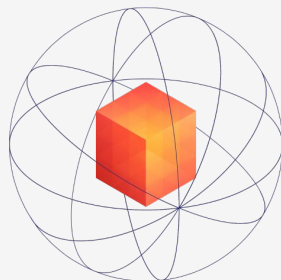
1.  $t \times w = 226$ ,  $\log |F| \sim 252$ ,  $d=2$
2. CPU time better than Rescue (100 microsec)
3. Relatively safe (though less than Rescue)



# Overview

- STARKs - Integrity Through Math
- Arithmetization
- Wish list for crypto primitives

# Integrity Through Math



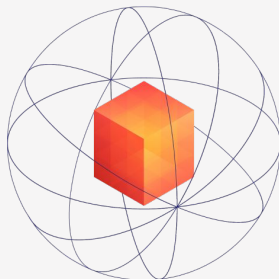
**STARK**

Integrity\* via Math

*\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]*



# Integrity Through Math



## STARK

Integrity\* via Math



### Privacy (Zero Knowledge, ZK)

Prover's private inputs are shielded



### Scalability

Exponentially small verifier running time\*  
Nearly linear prover running time\*



### Universality

Applicability to general computation



### Transparency

No toxic waste (i.e. no trusted setup)

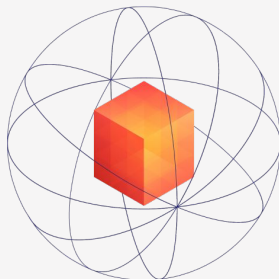


### Lean & Battle-Hardened Cryptography

e.g. post-quantum secure

\*With respect to size of computation

# Integrity Through Math



## STARK

### Integrity\* via Math

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]

1992

*Ákos Babai*<sup>1</sup>  
Univ. of Chicago<sup>6</sup> and  
Univ., Budapest

*Lance Fortnow*<sup>2</sup>  
Dept. Comp. Sci.  
Univ. of Chicago<sup>6</sup>

*Leonid A. Levin*<sup>3</sup>  
Dept. Comp. Sci.  
Boston University<sup>4</sup>

*Mario Szegedy*<sup>5</sup>  
Dept. Comp. Sci.  
Univ. of Chicago

**Abstract.** Motivated by Manuel Blum's concept of *instance checking*, we consider new, very fast and generic mechanisms of checking computations. Our results exploit recent advances in interactive proof protocols [LFKN92], [Sha92], and especially the  $MIP = NEXP$  protocol from [BFL91].

We show that every nondeterministic computational task  $S(x, y)$ , defined as a polynomial time relation between the *instance*  $x$ , representing the input and output combined, and the *witness*  $y$  can be modified to a task  $S'$  such that: (i) the same instances remain accepted; (ii) each instance/witness pair becomes checkable in *polylogarithmic* Monte Carlo time; and (iii) a witness satisfying  $S'$  can be computed in polynomial time from a witness satisfying  $S$ .

Here the instance and the description of  $S$  have to be provided in error-correcting code (since the checker will not notice slight changes). A modification of the  $MIP$  proof was required to achieve polynomial time in (iii); the earlier technique yields  $N^{O(\log \log N)}$  time only.

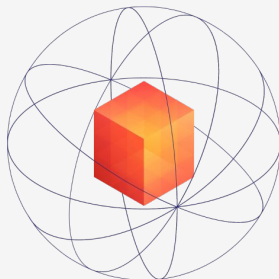
This result becomes significant if software and hardware *reliability* are regarded as a considerable cost factor. The polylogarithmic checker is the only part of the system that needs to be trusted; it can be *hard wired*. (We use just *one Checker* for all problems!) The checker is tiny and so presumably can be optimized and checked off-line at a modest cost.

In this setup, a single reliable PC can monitor the operation of a herd of supercomputers working with possibly extremely powerful but unreliable software and untested hardware.

In another interpretation, we show that in polynomial time every formal mathematical proof can be transformed into a *transparent proof*, i.e. a proof verifiable in polylogarithmic Monte Carlo time, assuming the "theorem-candidate" is given in error-correcting code. In fact, for any  $\varepsilon > 0$ , we can transform any proof  $P$  in time  $\|P\|^{1+\varepsilon}$  into a transparent proof, verifiable in Monte Carlo time  $(\log \|P\|)^{O(1/\varepsilon)}$ .

As a by-product, we obtain a binary error correcting code with very efficient error-correction. The code transforms messages of length  $N$  into codewords of length  $\leq N^{1+\varepsilon}$ ; and for strings within 10% of a valid codeword, it allows to recover any bit of the unique codeword within that distance in polylogarithmic  $((\log N)^{O(1/\varepsilon)})$  time.

# Integrity Through Math



## STARK

### Integrity\* via Math

“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]

1992

*Ákos Babai*<sup>1</sup>  
Univ. of Chicago<sup>6</sup> and  
Univ., Budapest

*Lance Fortnow*<sup>2</sup>  
Dept. Comp. Sci.  
Univ. of Chicago<sup>6</sup>

*Leonid A. Levin*<sup>3</sup>  
Dept. Comp. Sci.  
Boston University<sup>4</sup>

*Mario Szegedy*<sup>5</sup>  
Dept. Comp. Sci.  
Univ. of Chicago

**Abstract.** Motivated by Manuel Blum's concept of *instance checking*, we consider new, very fast and generic mechanisms of checking computations. Our results exploit recent advances in interactive proof protocols [LFKN92], [Sha92], and especially the  $MIP = NEXP$  protocol from [BFL91].

We show that every nondeterministic computational task  $S(x, y)$ , defined as a polynomial time relation between the instance  $x$ , representing the input and output combined, and the witness  $y$  can be modified to a task  $S'$  such that: (i) the same instances remain accepted; (ii) each instance/witness pair becomes checkable in *polylogarithmic* Monte Carlo time; and (iii) a witness satisfying  $S'$  can be computed in polynomial time from a witness satisfying  $S$ .

Here the instance and the description of  $S$  have to be provided in error-correcting code (since the checker will not notice slight changes). A modification of the  $MIP$  proof was required to achieve polynomial time in (iii); the earlier technique yields  $N^{O(\log \log N)}$  time only.

This result becomes significant if software and hardware *reliability* are regarded as a considerable cost factor. The polylogarithmic checker is the only part of the system that needs to be trusted; it can be *hard wired*. (We use just *one Checker* for all problems!) The checker is tiny and so presumably can be optimized and checked off-line at a modest cost.

In this setup, a single reliable PC can monitor the operation of a herd of supercomputers working with possibly extremely powerful but unreliable software and untested hardware.

In another interpretation, we show that in polynomial time every formal mathematical proof can be transformed into a *transparent proof*, i.e. a proof verifiable in polylogarithmic Monte Carlo time, assuming the “theorem-candidate” is given in error-correcting code. In fact, for any  $\varepsilon > 0$ , we can transform any proof  $P$  in time  $\|P\|^{1+\varepsilon}$  into a transparent proof, verifiable in Monte Carlo time  $(\log \|P\|)^{O(1/\varepsilon)}$ .

As a by-product, we obtain a binary error correcting code with very efficient error-correction. The code transforms messages of length  $N$  into codewords of length  $\leq N^{1+\varepsilon}$ ; and for strings within 10% of a valid codeword, it allows to recover any bit of the unique codeword within that distance in polylogarithmic  $((\log N)^{O(1/\varepsilon)})$  time.

# Integrity Through Math

**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program **P**, reached state hash  $y$

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

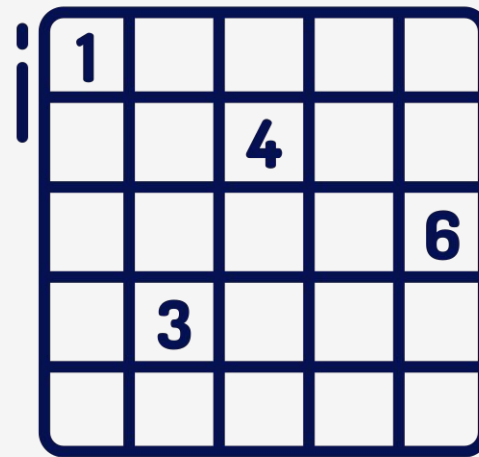
# Integrity Through Math

**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and **#tx (=1,000,000)**



**PCP**



# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

*Claim: Starting @ state hash  $x$ , after 1,000,000 txs processed by program  $P$ , reached state hash  $y$*

1	1	3	2	6
4	1	4	7	3
8	5	8	3	6
6	3	5	9	7
5	2	8	7	6

PCP

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

Prover submits solution

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

*Claim: Starting @ state hash  $x$ , after 1,000,000 txs processed by program  $P$ , reached state hash  $y$*

1	1	3	2	6
4	1	4	7	3
8	5	8	3	6
6	3	5	9	7
5	2	8	7	6

PCP

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

**Prover** submits solution

**Verifier** samples and checks a single constraint

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

## Magic (aka Math)

- Sampling constraints takes exponentially small time!
- Good proofs satisfy ALL constraints!
- A “proof” of a false claim satisfies < 1% of constraints!

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]

1	1	3	2	6
4	1	4	7	3
8	5	8	3	6
6	3	5	9	7
5	2	8	7	6

PCP

**Prover** submits solution

**Verifier** samples and checks a single constraint

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

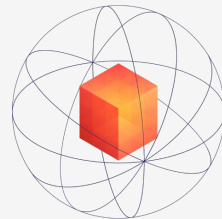
## Magic (aka Math)

- Sampling constraints takes exponentially small time!
- Good proofs satisfy ALL constraints!
- A “proof” of a false claim satisfies < 1% of constraints!

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

1	1	3	2	6
4	1	4	7	3
8	5	8	3	6
6	3	5	9	7
5	2	8	7	6

# STARK



**Prover** submits solution

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

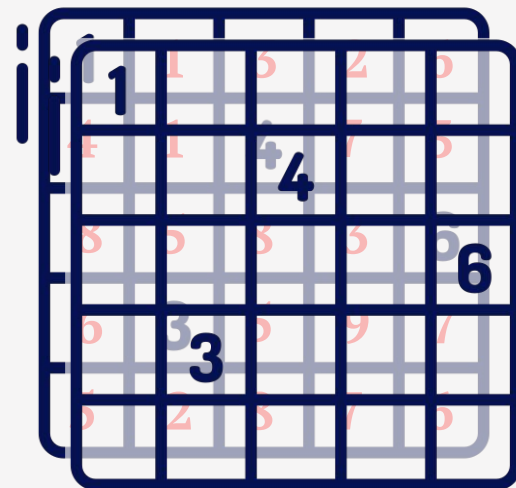
**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

## Magic (aka Math)

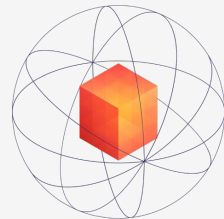
- Sampling constraints takes exponentially small time!
- Good proofs satisfy ALL constraints!
- A “proof” of a false claim satisfies  $< 1\%$  of constraints!

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]



# STARK



**Prover** submits solution

**Verifier** posts another (random) sudoku puzzle

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

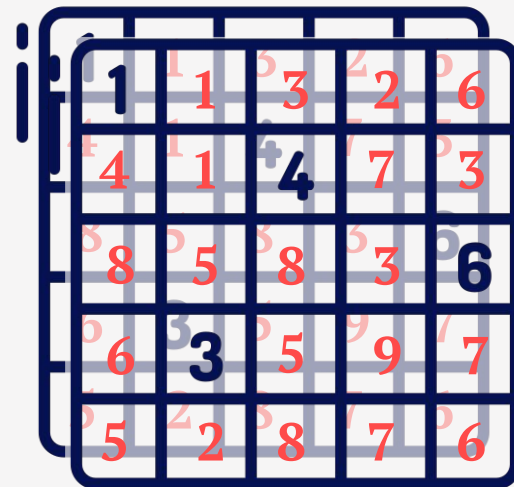
**Claim:** Starting @ state hash  $x$ , after 1,000,000 txs processed by program  $P$ , reached state hash  $y$

## Magic (aka Math)

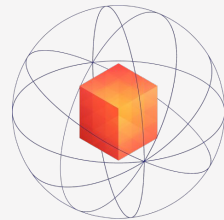
- Sampling constraints takes exponentially small time!
- Good proofs satisfy ALL constraints!
- A “proof” of a false claim satisfies  $< 1\%$  of constraints!

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]



# STARK



**Prover** submits solution

**Verifier** posts another (random) sudoku puzzle

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx$  (=1,000,000)

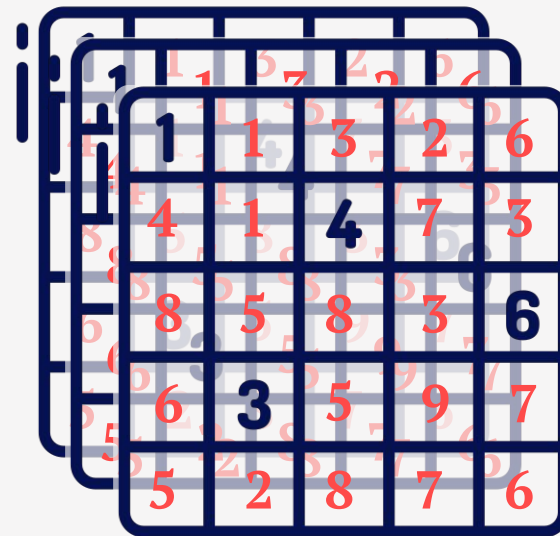
**Claim:** Starting @ state hash  $x$ , after 1,000,000 txs processed by program  $P$ , reached state hash  $y$

## Magic (aka Math)

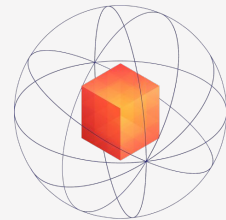
- Sampling constraints takes exponentially small time!
- Good proofs satisfy ALL constraints!
- A “proof” of a false claim satisfies < 1% of constraints!

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]



# STARK



**Prover** submits solution

**Verifier** posts another (random) sudoku puzzle

# Integrity Through Math

A sudoku-like set of constraints is implied by the statement proved, by  $x$ ,  $y$ ,  $P$ , and  $\#tx (=1,000,000)$

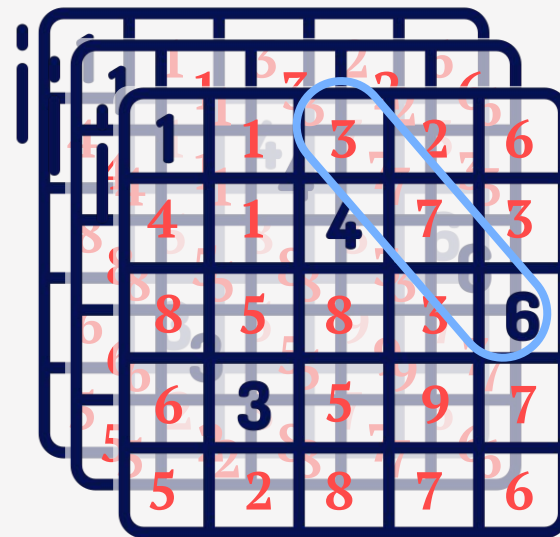
**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

## Magic (aka Math)

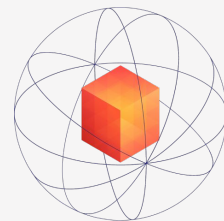
- Sampling constraints takes exponentially small time!
- Good proofs satisfy ALL constraints!
- A “proof” of a false claim satisfies  $< 1\%$  of constraints!

*“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”*

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]



# STARK



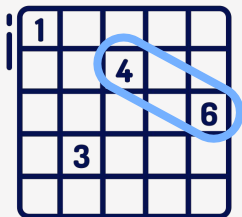
**Prover** submits solution

**Verifier** posts another (random) sudoku puzzle

**Verifier** samples and checks a single constraint



# Integrity Through Math



PCP

Integrity\* via Math  
(impractical)

“...a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”

\* Integrity means doing the right thing, even when no one is watching [C.S. Lewis]

1992

Ászló Babai<sup>1</sup>  
Univ. of Chicago<sup>6</sup> and  
Univ., Budapest

Lance Fortnow<sup>2</sup>  
Dept. Comp. Sci.  
Univ. of Chicago<sup>6</sup>

Leonid A. Levin<sup>3</sup>  
Dept. Comp. Sci.  
Boston University<sup>4</sup>

Mario Szegedy<sup>5</sup>  
Dept. Comp. Sci.  
Univ. of Chicago

## PCP

**Abstract.** Motivated by Manuel Blum's concept of *instance checking*, we consider new, very fast and generic mechanisms of checking computations. Our results exploit recent advances in interactive proof protocols [LFKN92], [Sha92], and especially the  $MIP = NEXP$  protocol from [BFL91].

We show that every nondeterministic computational task  $S(x, y)$ , defined as a polynomial time relation between the *instance*  $x$ , representing the input and output combined, and the *witness*  $y$  can be modified to a task  $S'$  such that: (i) the same instances remain accepted; (ii) each instance/witness pair becomes checkable in *polylogarithmic* Monte Carlo time; and (iii) a witness satisfying  $S'$  can be computed in polynomial time from a witness satisfying  $S$ .

Here the instance and the description of  $S$  have to be provided in error-correcting code (since the checker will not notice slight changes). A modification of the  $MIP$  proof was required to achieve polynomial time in (iii); the earlier technique yields  $N^{O(\log \log N)}$  time only.

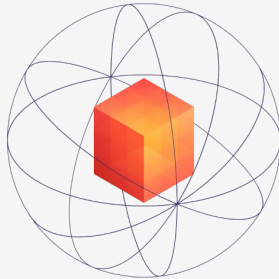
This result becomes significant if software and hardware *reliability* are regarded as a considerable cost factor. The polylogarithmic checker is the only part of the system that needs to be trusted; it can be *hard wired*. (We use just *one Checker* for all problems!) The checker is tiny and so presumably can be optimized and checked off-line at a modest cost.

In this setup, a single reliable PC can monitor the operation of a herd of supercomputers working with possibly extremely powerful but unreliable software and untested hardware.

In another interpretation, we show that in polynomial time every formal mathematical proof can be transformed into a *transparent proof*, i.e. a proof verifiable in polylogarithmic Monte Carlo time, assuming the “theorem-candidate” is given in error-correcting code. In fact, for any  $\varepsilon > 0$ , we can transform any proof  $P$  in time  $\|P\|^{1+\varepsilon}$  into a transparent proof, verifiable in Monte Carlo time  $(\log \|P\|)^{O(1/\varepsilon)}$ .

As a by-product, we obtain a binary error correcting code with very efficient error-correction. The code transforms messages of length  $N$  into codewords of length  $\leq N^{1+\varepsilon}$ ; and for strings within 10% of a valid codeword, it allows to recover any bit of the unique codeword within that distance in polylogarithmic  $((\log N)^{O(1/\varepsilon)})$  time.

# Integrity Through Math



## STARK

Integrity\* via Math  
(~~i~~mpractical)

“..a single reliable PC can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”

\* *Integrity means doing the right thing, even when no one is watching* [C.S. Lewis]

1992

SIAM J. COMPUT.  
Vol. 38, No. 2, pp. 551-607

SHORT PCPS WITH POLYLOG QUERY COMPLEXITY

ELI BEN-SASSON<sup>†</sup> AND MADHU SUDAN<sup>‡</sup>

2015

Interactive Oracle Proofs\*

Eli Ben-Sasson<sup>1</sup>, Alessandro Chiesa<sup>2</sup> and Nicholas Spooner<sup>3</sup>

2004

© 2008 Society for Industrial and Applied Mathematics

Fast Reed-Solomon Interactive Oracle Proofs of Proximity

Eli Ben-Sasson\* Iddo Bentov<sup>†</sup> Yinon Horesh\* Michael Riabzev\*

January 12, 2018

2018

Scalable, transparent, and post-quantum secure computation with integrity

Eli Ben-Sasson\* Iddo Bentov<sup>†</sup> Yinon Horesh\* Michael Riabzev\*

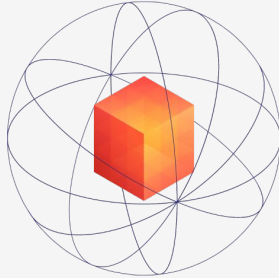
March 6, 2018

Abstract

Human dignity demands that personal information, like medical and forensic data, be hidden from the public. But veils of secrecy designed to preserve privacy may also be abused to cover up lies and deceit by institutions entrusted with Data, unjustly harming citizens and eroding trust in central institutions. Zero knowledge (ZK) proof systems are an ingenious cryptographic solution to this tension between the ideals of personal *privacy* and institutional *integrity*, enforcing the latter in a way that does not compromise the former. Public trust demands *transparency* from ZK systems, meaning they be set up with no reliance on any trusted party, and have no trapdoors that could be exploited by powerful parties to bear false witness. For ZK systems to be used with Big Data, it is imperative that the public verification process scale sublinearly in data size. Transparent ZK proofs that can be verified *exponentially* faster than data size were first described in the 1990s but early constructions were impractical, and no ZK system realized thus far in code (including that used by crypto-currencies like Zcash<sup>TM</sup>) has achieved *both* transparency and exponential verification speedup, simultaneously, for general computations.

Here we report the first realization of a transparent ZK system (ZK-STARK) in which verification scales exponentially faster than database size, and moreover, this exponential speedup in verification is observed concretely for meaningful and sequential computations, described next. Our system uses several recent advances on interactive oracle proofs (IOP), such as a “fast” (linear time) IOP system for error correcting codes.

# Integrity Through Math



## STARK

Integrity\* via Math  
(~~i~~mpractical)

“...a **single reliable PC** can monitor the operation of a herd of supercomputers with powerful but unreliable software and untested hardware ...”

\* *Integrity means doing the right thing, even when no one is watching* [C.S. Lewis]

1992

Checking Computations in Polylogarithmic Time  
Aszlo Babai<sup>1</sup> Lance Fortnow<sup>2</sup> Leonid A. Levin<sup>3</sup>  
Univ. of Chicago<sup>6</sup> and Dept. Comp. Sci. Dept. Comp. Sci.  
Univ., Budapest Univ. of Chicago<sup>6</sup> Boston University<sup>4</sup>  
Mario Szegedy<sup>5</sup>  
Dept. Comp. Sci.  
Univ. of Chicago

2004

SIAM J. COMPUT.  
Vol. 38, No. 2, pp. 551–607

© 2008 Society for Industrial and Applied Mathematics

SHORT PCPS WITH POLYLOG QUERY COMPLEXITY

ELI BEN-SASSON<sup>†</sup> AND MADHU SUDAN<sup>‡</sup>

2015

Interactive Oracle Proofs\*

Eli Ben-Sasson<sup>1</sup>, Alessandro Chiesa<sup>2</sup> and Nicholas Spooner<sup>3</sup>

Fast Reed-Solomon Interactive Oracle Proofs of Proximity

Eli Ben-Sasson\* Iddo Bentov<sup>†</sup> Yinon Horesh\* Michael Riabzev\*

January 12, 2018

2018

Scalable, transparent, and post-quantum secure computation with integrity

Eli Ben-Sasson\* Iddo Bentov<sup>†</sup> Yinon Horesh\* Michael Riabzev\*

March 6, 2018

Abstract

Human dignity demands that personal information, like medical and forensic data, be hidden from the public. But veils of secrecy designed to preserve privacy may also be abused to cover up lies and deceit by institutions entrusted with Data, unjustly harming citizens and eroding trust in central institutions. Zero knowledge (ZK) proof systems are an ingenious cryptographic solution to this tension between the ideals of personal *privacy* and institutional *integrity*, enforcing the latter in a way that does not compromise the former. Public trust demands *transparency* from ZK systems, meaning they be set up with no reliance on any trusted party, and have no trapdoors that could be exploited by powerful parties to bear false witness. For ZK systems to be used with Big Data, it is imperative that the public verification process scale sublinearly in data size. Transparent ZK proofs that can be verified *exponentially* faster than data size were first described in the 1990s but early constructions were impractical, and no ZK system realized thus far in code (including that used by crypto-currencies like Zcash<sup>TM</sup>) has achieved *both* transparency and exponential verification speedup, simultaneously, for general computations.

Here we report the first realization of a transparent ZK system (ZK-STARK) in which verification scales exponentially faster than database size, and moreover, this exponential speedup in verification is observed concretely for meaningful and sequential computations, described next. Our system uses several recent advances on interactive oracle proofs (IOP), such as a “fast” (linear time) IOP system for error correcting codes.

# Arithmetization

# Arithmetization

Arithmetization Converts (“reduces”) Computational Integrity problems to problems about local relations between a bunch of polynomials

**Claim:** Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

# Arithmetization

Arithmetization Converts (“reduces”) Computational Integrity problems to problems about local relations between a bunch of polynomials

## Pre-arithmetization claim

*Starting @ state hash  $x$ ,  
after **1,000,000** txs  
processed by program  
 $P$ , reached state hash  $y$*

# Arithmetization

Arithmetization Converts (“reduces”) Computational Integrity problems to problems about local relations between a bunch of polynomials

## Pre-arithmetization claim

*Starting @ state hash  $x$ ,  
after **1,000,000** txs  
processed by program  
**P**, reached state hash  $y$*

## Reduction

*produces 2  
polynomials:  
 **$Q(X,Y,T,W)$ ,  $R(X)$**  and  
degree bound  **$d$***

# Arithmetization

Arithmetization Converts (“reduces”) Computational Integrity problems to problems about local relations between a bunch of polynomials

## Pre-arithmetization claim

*Starting @ state hash  $x$ , after **1,000,000** txs processed by program **P**, reached state hash  $y$*

## Reduction

*produces 2 polynomials:  **$Q(X,Y,T,W)$**  and degree bound  **$d$***

## Post-arithmetization claim

*I know 4 polynomials of degree  **$d$**  -  $A(x)$ ,  $B(x)$ ,  $C(x)$ ,  $D(X)$  - such that:*

$$Q(X, A(X), B(X+1), C(2*X)) = D(X) * R(X)$$



# Arithmetization

Arithmetization Converts (“reduces”) Computational Integrity problems to problems about local relations between a bunch of polynomials

## Pre-arithmetization claim

Starting @ state hash  $x$ , after **1,000,000** txs processed by program  $P$ , reached state hash  $y$

## Reduction

produces 2 polynomials:  
 $Q(X,Y,T,W)$ ,  $R(X)$  and degree bound  $d$

## Post-arithmetization claim

I know 4 polynomials of degree  $d$  -  $A(x)$ ,  $B(x)$ ,  $C(x)$ ,  $D(x)$  - such that:

$$Q(X, A(X), B(X+1), C(2*X)) = D(X) * R(X)$$

## Theorem

If  $A$ ,  $B$ ,  $C$ ,  $D$  do not satisfy **THIS**,

then nearly all  $x$  expose Bob's lie

# Arithmetization

Assuming Theorem, we get a scalable proof system for Bob's original claim:

1. Apply reduction, ask Bob to **provide access to  $A, B, C, D$  of degree- $d$**
2. Sample random  $x$  and accept Bob's claim iff equality holds for this  $x$

**New problem:** Force Bob to

- (1) commit to degree  $d$  polynomials, then
- (2) answer queries to the precommitted polys

## Post-arithmetization claim

*I know 4 polynomials of degree  $d$  -  $A(x), B(x), C(x), D(x)$  - such that:*

$$Q(X, A(X), B(X+1), C(2*X)) = D(X) * R(X)$$

## Theorem

*If  $A, B, C, D$  do not satisfy THIS,*

*then nearly all  $x$  expose Bob's lie*

# Arithmetization

Assuming Theorem, we get a scalable proof system for Bob's original claim:

1. Apply reduction, ask Bob to provide access to  $A, B, C, D$  of degree- $d$
2. Sample random  $x$  and accept Bob's claim iff equality holds for this  $x$

**New problem:** Force Bob to

- (1) commit to degree  $d$  polynomials, then
- (2) answer queries to the precommitted polys

## Post-arithmetization claim

*I know 4 polynomials of degree  $d$  -  $A(x), B(x), C(x), D(x)$  - such that:*

$$Q(X, A(X), B(X+1), C(2*X)) = D(X) * R(X)$$

## Theorem

*If  $A, B, C, D$  do not satisfy THIS,*

*then nearly all  $x$  expose Bob's lie*

# AIR (Algebraic Intermediate Representation)

## Computation:

A set of rules for how to sequentially evolve a state, starting with an input and ending up with an output.

# AIR (Algebraic Intermediate Representation)

python:

```
a = 1
b = 0
for i in range(n):
    a, b = a + b, a
return a
```

# AIR (Algebraic Intermediate Representation)

python:

```
a = 1
b = 0
for i in range(n):
    a, b = a + b, a
return a
```

trace:

$a_0=1$	$b_0=0$
$a_1=1$	$b_1=1$
$a_2=2$	$b_2=1$
$\vdots$	$\vdots$
$a_n$	$b_n$

# AIR (Algebraic Intermediate Representation)

python:

```
a = 1
b = 0
for i in range(n):
    a, b = a + b, a
return a
```

trace:

$a_0=1$	$b_0=0$
$a_1=1$	$b_1=1$
$a_2=2$	$b_2=1$
$\vdots$	$\vdots$
$a_n$	$b_n$

constraints:

$$a_0 = 1$$

$$b_0 = 0$$

$$a_{i+1} = a_i + b_i$$

$$b_{i+1} = a_i$$

$$a_n = fib(n)$$

# AIR

python:

```
a = 1
b = 0
for i in range(n):
    a, b = a + b, a
return a
```

trace:

$a_0=1$	$b_0=0$
$a_1=1$	$b_1=1$
$a_2=2$	$b_2=1$
$\vdots$	$\vdots$
$a_n$	$b_n$

constraints:

$$a_0 = 1$$

$$b_0 = 0$$

$$a_{i+1} = a_i + b_i$$

$$b_{i+1} = a_i$$

$$a_n = fib(n)$$

domain:

First row

First row

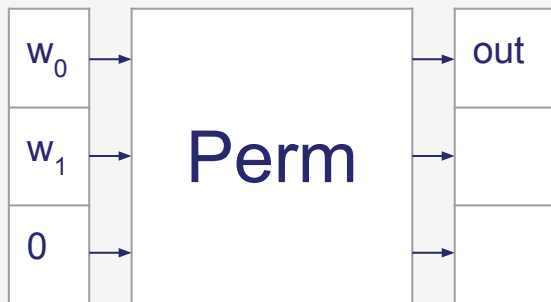
$0 \leq i < n$

$0 \leq i < n$

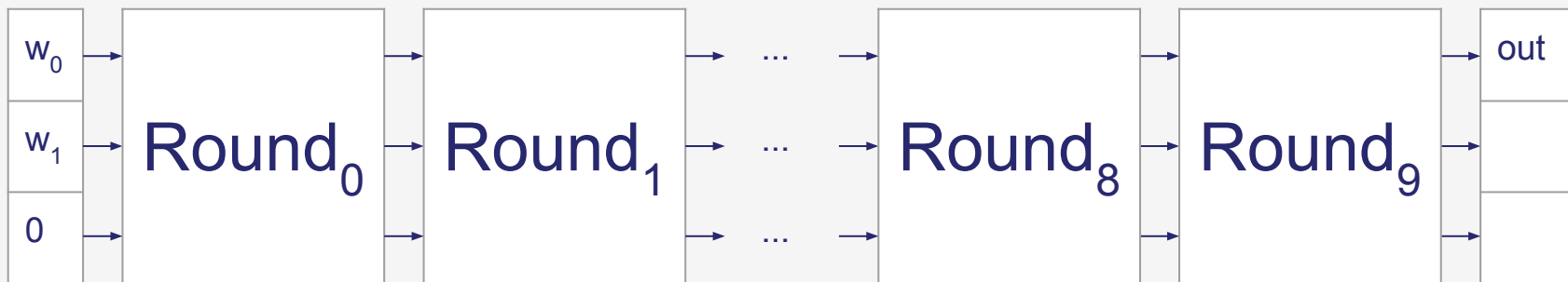
Row n



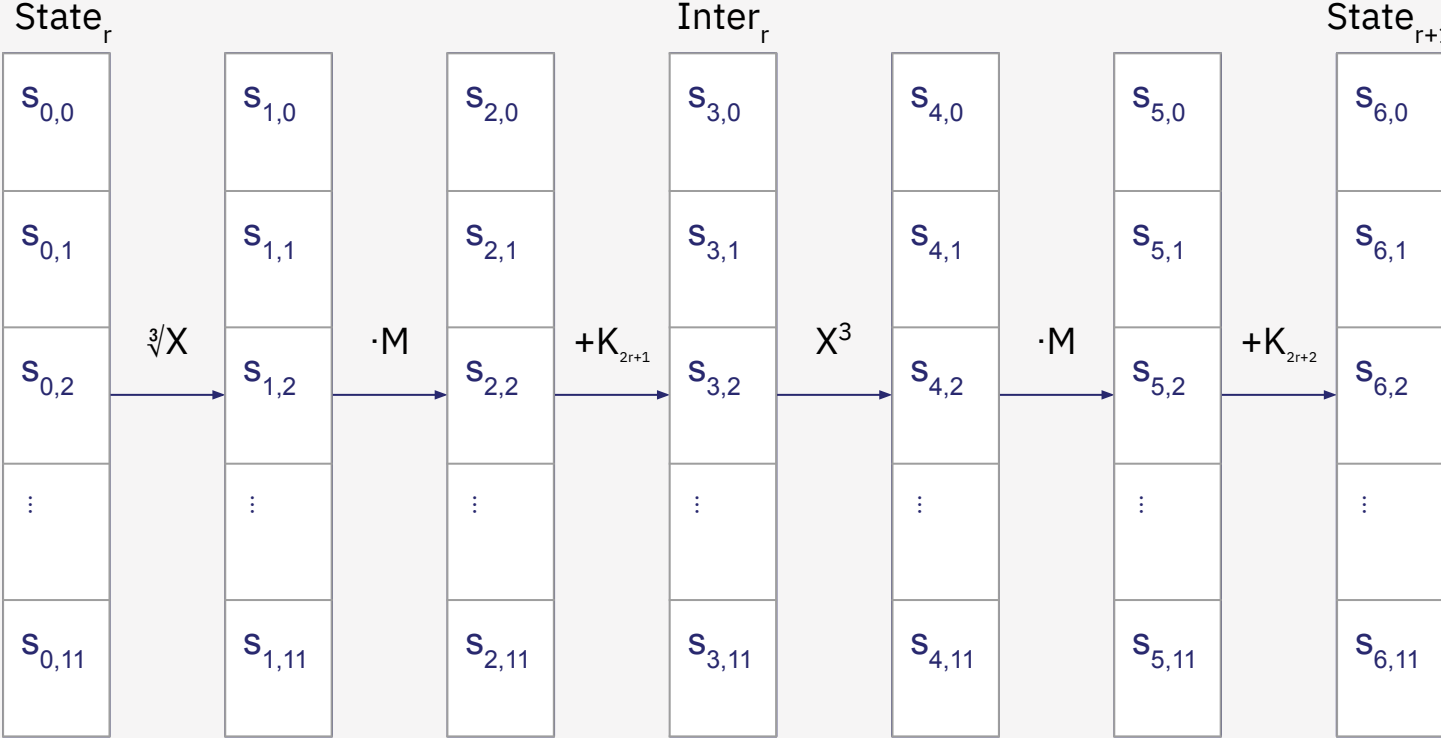
# Rescue hash function



# Rescue hash function



# Round r

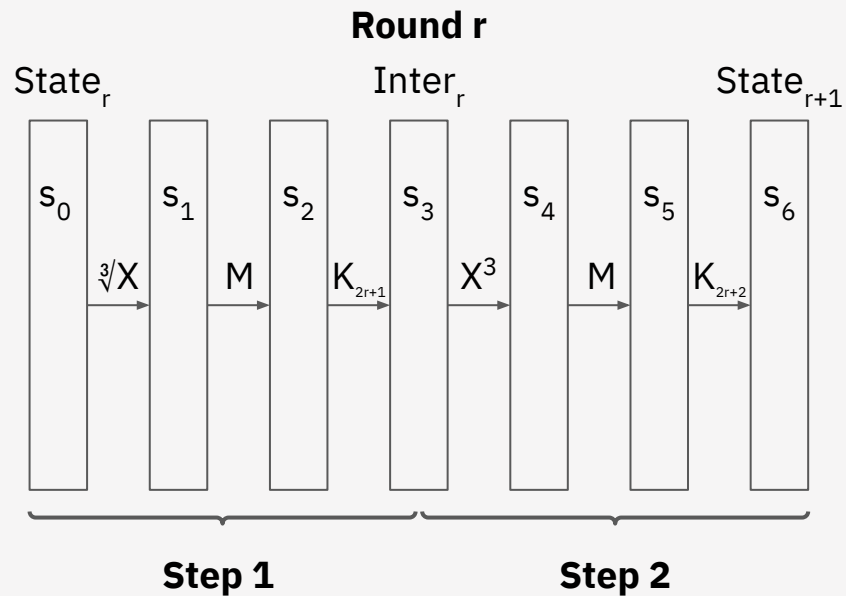


Step 1

Step 2

# Rescue AIR

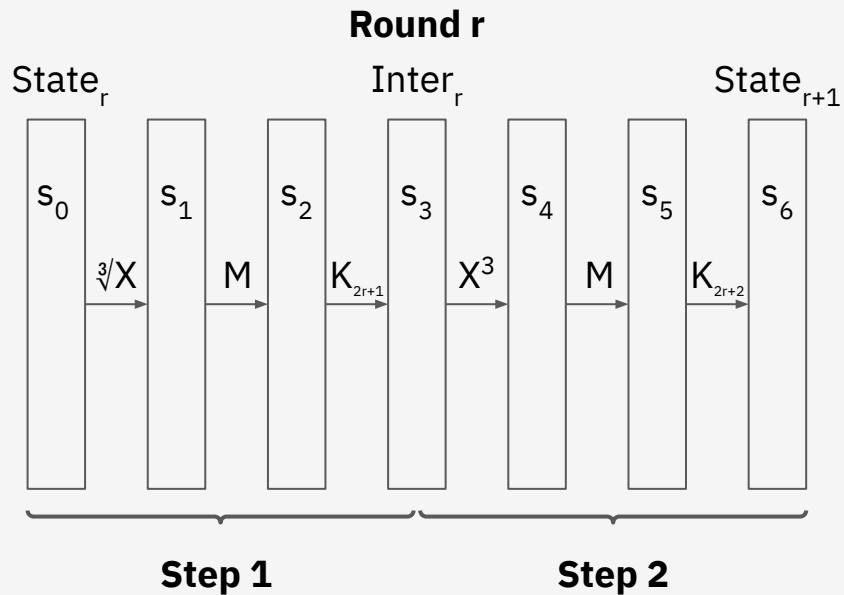
$$s_{1,i}^3 = s_{0,i}$$



# Rescue AIR

$$s_{1,i}^3 = s_{0,i}$$

$$s_{2,i} = \sum_j M_{ij} s_{1,j} \quad (s_2 = M \cdot s_1)$$

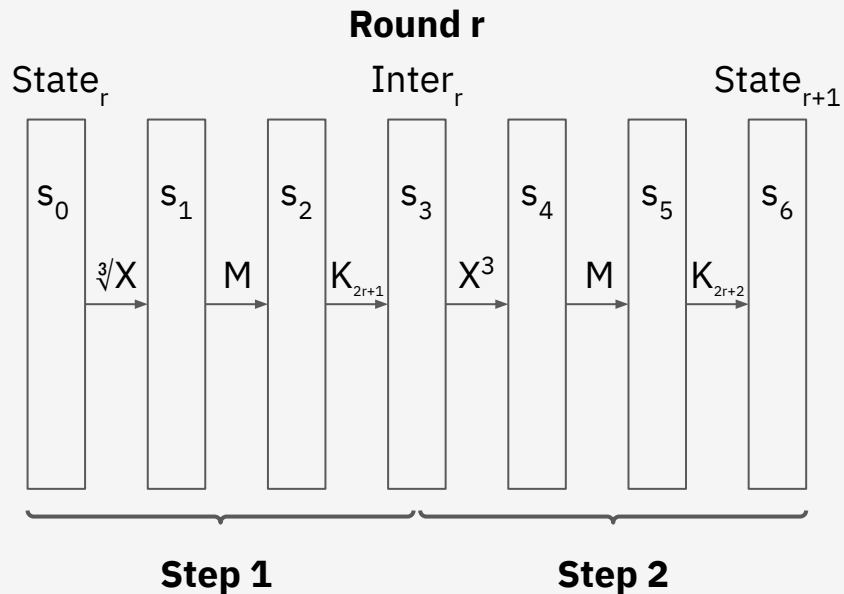


# Rescue AIR

$$s_{1,i}^3 = s_{0,i}$$

$$s_{2,i} = \sum_j M_{ij} s_{1,j} \quad (s_2 = M \cdot s_1)$$

$$s_3 = s_2 + K_{2r+1}$$



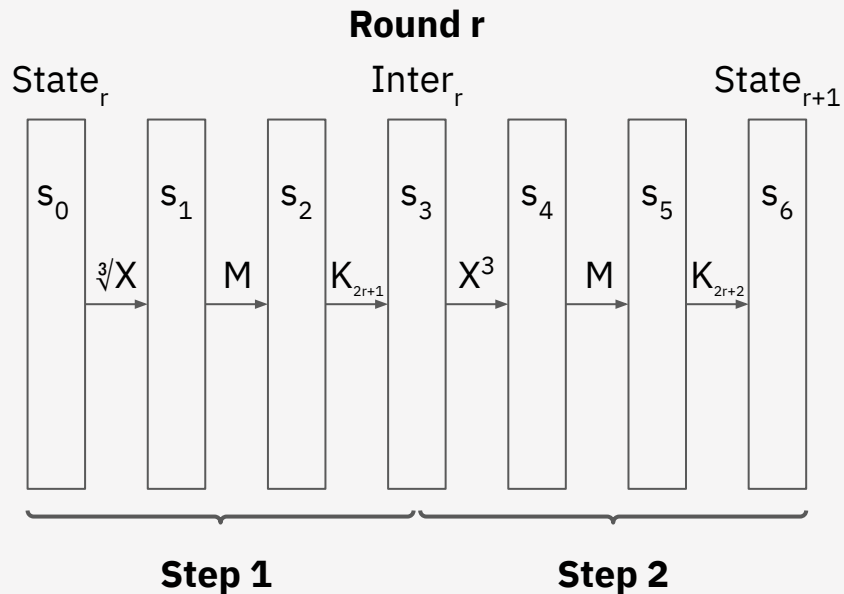
# Rescue AIR

$$s_{1,i}^3 = s_{0,i}$$

$$s_{2,i} = \sum_j M_{ij} s_{1,j} \quad (s_2 = M \cdot s_1)$$

$$s_3 = s_2 + K_{2r+1}$$

$$s_{4,i} = s_{3,i}^3$$



# Rescue AIR

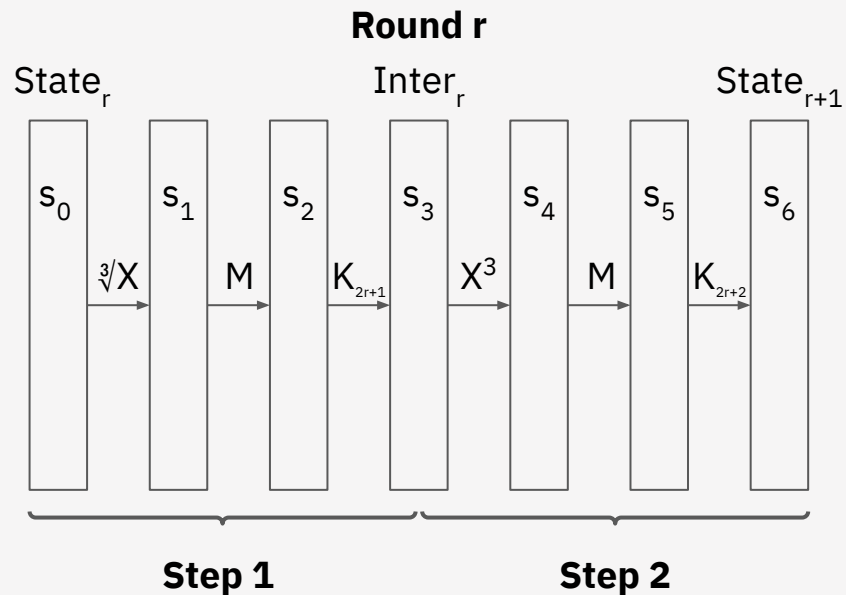
$$s_{1,i}^3 = s_{0,i}$$

$$s_{2,i} = \sum_j M_{ij} s_{1,j} \quad (s_2 = M \cdot s_1)$$

$$s_3 = s_2 + K_{2r+1}$$

$$s_{4,i} = s_{3,i}^3$$

$$s_5 = M \cdot s_4$$





# Rescue AIR

$$s_{1,i}^3 = s_{0,i}$$

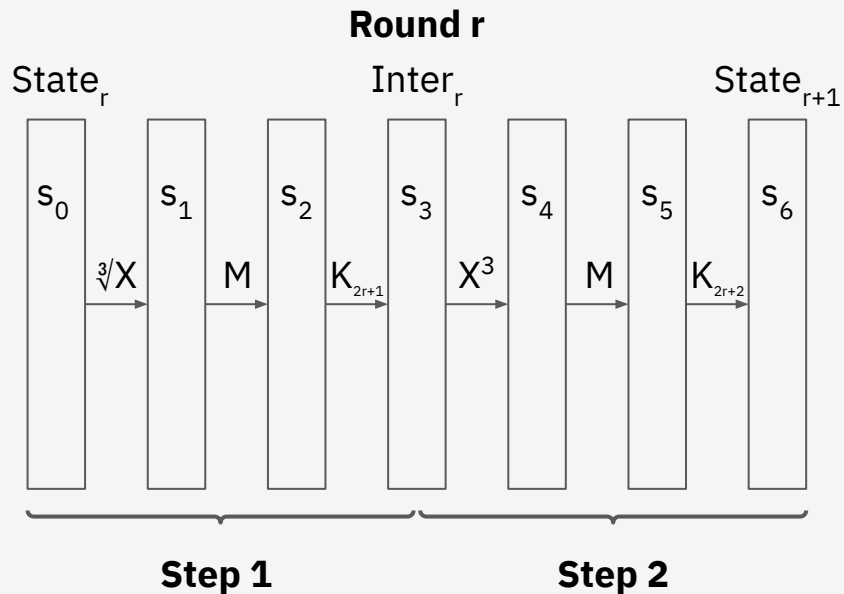
$$s_{2,i} = \sum_j M_{ij} s_{1,j} \quad (s_2 = M \cdot s_1)$$

$$s_3 = s_2 + K_{2r+1}$$

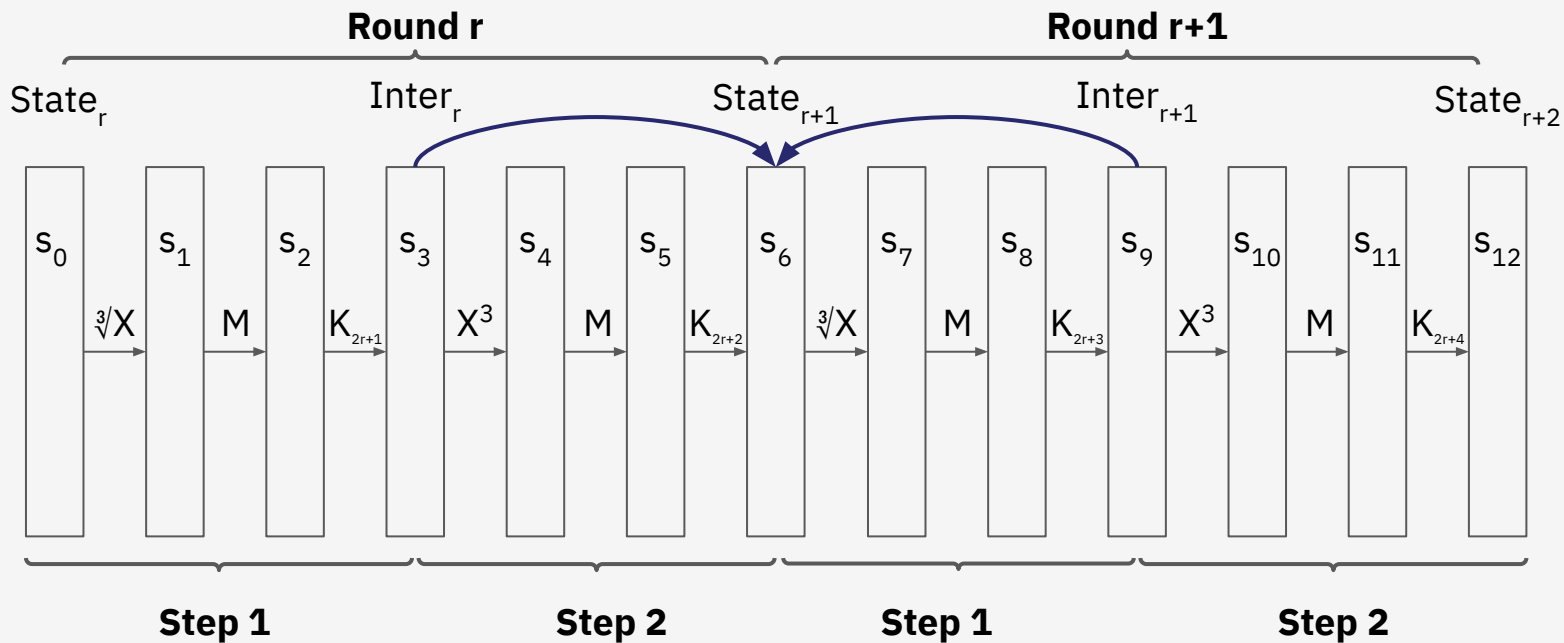
$$s_{4,i} = s_{3,i}^3$$

$$s_5 = M \cdot s_4$$

$$s_6 = s_5 + K_{2r+2}$$



# Rescue AIR



$$\left(\sum_j M_{ij} s_{3,j}^3\right) + K_{2r+2,i} = s_{6,i} = \left(\sum_j (M^{-1})_{ij} (s_{9,j} - K_{2r+3,j})\right)^3$$

# Wishlist for crypto primitives

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

## Wish-list for Primitive P:

1. Minimize  $t \times w$ ,  $|F|$ ,  $d$  (also  $s$ ,  $e$ )
2. Minimize CPU time of computing P
3. Make it field agnostic, and safe!

# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

- Execution trace size:  $t$  rows,  $w$  columns
  - Fib(n):  $t=n$ ,  $w=2$
  - Rescue:  $t= \text{\#rounds}$ ,  $w = |\text{state}|$
- Trace entries in field  $F$ . Which field?
  - **Any** field! [BCKL22] (FFT-friendly better)
- Constraints
  - Maximal degree:  $d$
  - # constraints:  $s$
  - Enforcement domain complexity:  $e$

## Wish-list for Primitive P:

1. Minimize  $t \times w$ ,  $|F|$ ,  $d$  (also  $s$ ,  $e$ )
2. Minimize CPU time of computing P
3. Make it field agnostic, and safe!

## ethSTARK (Rescue):

1.  $t \times w = 120$ ,  $\log |F| \sim 62$ ,  $d=3$
2. CPU time large, esp in large  $F$  (cube root!)
3. Relatively safe, AES-like (say the experts)

## Poseidon in Starknet/Ex:

1.  $t \times w = 226$ ,  $\log |F| \sim 252$ ,  $d=2$
2. CPU time better than Rescue (100 microsec)
3. Relatively safe (though less than Rescue)



# AIR Parameters: $t$ , $w$ , $F$ , $d$ , $s$ , $e$

## Keccak:

1.  $t \times w = 70,000-90,000$ ,  $\log |F| \sim 252$ ,  $d=2$
2. CPU time amazing (1 microsecond)
3. Very safe

## Wish-list for Primitive P:

1. Minimize  $t \times w$ ,  $|F|$ ,  $d$  (also  $s$ ,  $e$ )
2. Minimize CPU time of computing P
3. Make it field agnostic, and safe!

## ethSTARK (Rescue):

1.  $t \times w = 120$ ,  $\log |F| \sim 62$ ,  $d=3$
2. CPU time large, esp in large  $F$  (cube root!)
3. Relatively safe, AES-like (say the experts)

## Poseidon in Starknet/Ex:

1.  $t \times w = 226$ ,  $\log |F| \sim 252$ ,  $d=2$
2. CPU time better than Rescue (100 microsec)
3. Relatively safe (though less than Rescue)

# STARK-friendly crypto primitives wish-list

---

Eli Ben-Sasson

📅 April 2023

