# Cryptographic hash functions

by drd

Project u23, CCC Cologne
19 October 2015

# What is a (cryptographic) hash function?

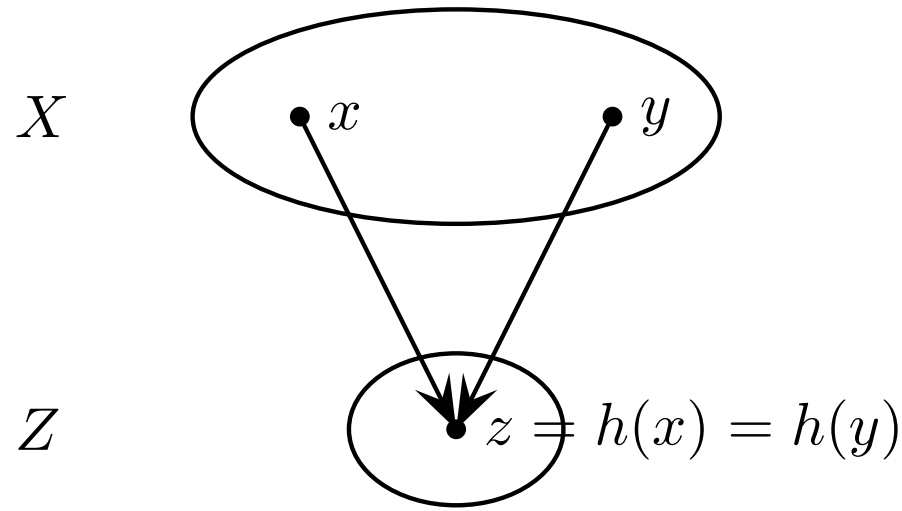
Where are cryptographic hash functions used in practice?

Which properties should cryptographic hash functions have?

Which specific functions are used in practice?

## Definition

Let $h\colon X \longrightarrow Z$ be a mapping between two finite sets $X$ and $Z$. If $x \neq y$ are messages in $X$ with $h(x) = h(y)$, then $x$ and $y$ *collide*, and $(x, y)$ is a *collision*.


$X$
$x$
$y$
$Z$
$z = h(x) = h(y)$

Hash functions are used, among other areas, in connection with signature schemes. This looks for example as follows:

| length | arbitrary | | 256 bits | | 512 bits |
|---|---|---|---|---|---|
| data | message $x$ | $\longrightarrow$ | hash value $h(x)$ | $\longrightarrow$ | signature sig$(h(x))$ |

Collisions are a security problem!

We consider three types of "attackers" on a hash function $h$.

(i) A *collision finder* takes no input and outputs either a collision $(x, y)$, or "failure".

(ii) A *second-preimage finder* takes an input $x$ and outputs either some $y \in X$ that collides with $x$, or "failure".

(iii) An *inverter* takes an input $z \in Z$ and outputs either some $x \in X$ with $h(x) = z$, or "failure".

## Definition

Let $h$ be a family of hash functions. We call $h$

*collision resistant*,
*second-preimage resistant*,
*inversion resistant* (or *one-way*),

if for all probabilistic polynomial-time

collision finders,
second-preimage finders,
inverters,

respectively, for $h$ the success probability is very small.

## Corollary

For a family $h$ of hash functions where the fraction $\#Z/\#X$ is very small, we have

$h$ collision resistant $\Rightarrow$ $h$ second-preimage resistant $\Rightarrow$ $h$ one-way.

How long does it take you to find a collision?

## Birthday paradox

How many randomly chosen people have to be in a room to have a probability of at least 50% that two of them have the same birthday, assuming each birthday occurs with equal probability?

## Birthday paradox

How many randomly chosen people have to be in a room to have a probability of at least 50% that two of them have the same birthday, assuming each birthday occurs with equal probability?

## Surprising answer:

$23$ people are sufficient!

## Theorem:

We consider random choices, with replacement, among $m$ labeled items. The expected number of choices until a collision occurs is $O(\sqrt{m})$.

The time it takes to find a collision (generically) is expectedly the square root of the number of possible hashes!

Exercise: How long does it take you to find a second preimage?

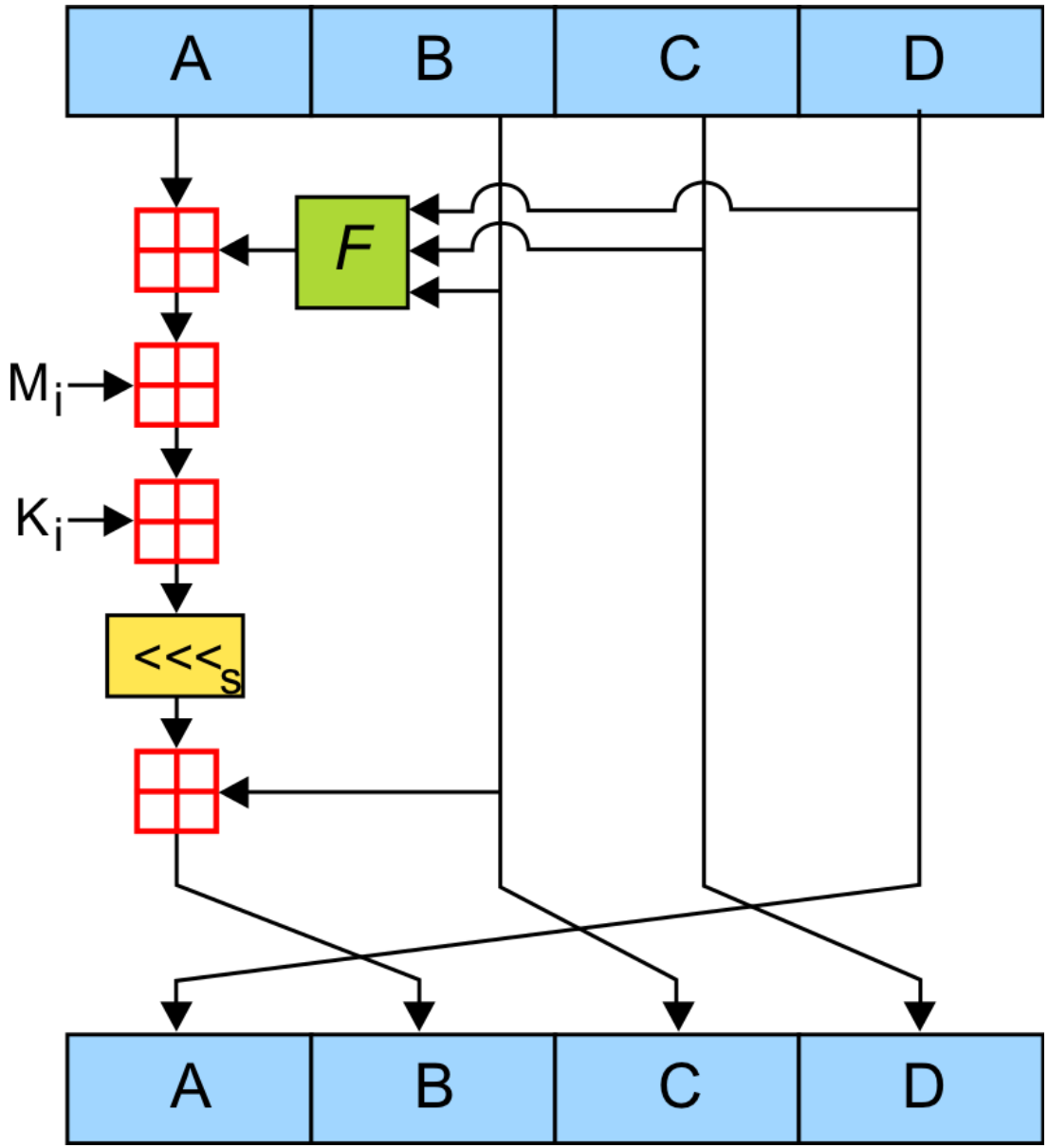



Figure: One round of MD5.*

*https://de.wikipedia.org/wiki/Message-Digest_Algorithm_5

## Theoretical result

"How to break MD5 and other hash functions"
by Wang et al. (2004)

## Theoretical result

"How to break MD5 and other hash functions"
by Wang et al. (2004)

## Practical break

- Prediction of the US presidential election in 2008.
- In 2007, the MD5 hash

  `0x3D515DEAD7AA16560ABA3E9DF05CBC80`,

  of a document with a prediction was published.
- After the election, a pdf with this MD5 sum was published.
- It contained the correct outcome.

TU/e technische universiteit eindhoven

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Prediction of the next President of the United States

Marc Stevens[1], Arjen Lenstra[2], and Benne de Weger[3]

[1] CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
[2] EPFL IC LACAL, Station 14, and Bell Laboratories
CH-1015 Lausanne, Switzerland
[3] TU Eindhoven, Faculty of Mathematics and Computer Science
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

We predict that the winner of the 2008 election for
President of the United States
will be:

**Barack Obama**

TU/e technische universiteit eindhoven

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Prediction of the next President of the United States

Marc Stevens[1], Arjen Lenstra[2], and Benne de Weger[3]

[1] CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
[2] EPFL IC LACAL, Station 14, and Bell Laboratories
CH-1015 Lausanne, Switzerland
[3] TU Eindhoven, Faculty of Mathematics and Computer Science
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

We predict that the winner of the 2008 election for
President of the United States
will be:

**Paris Hilton**

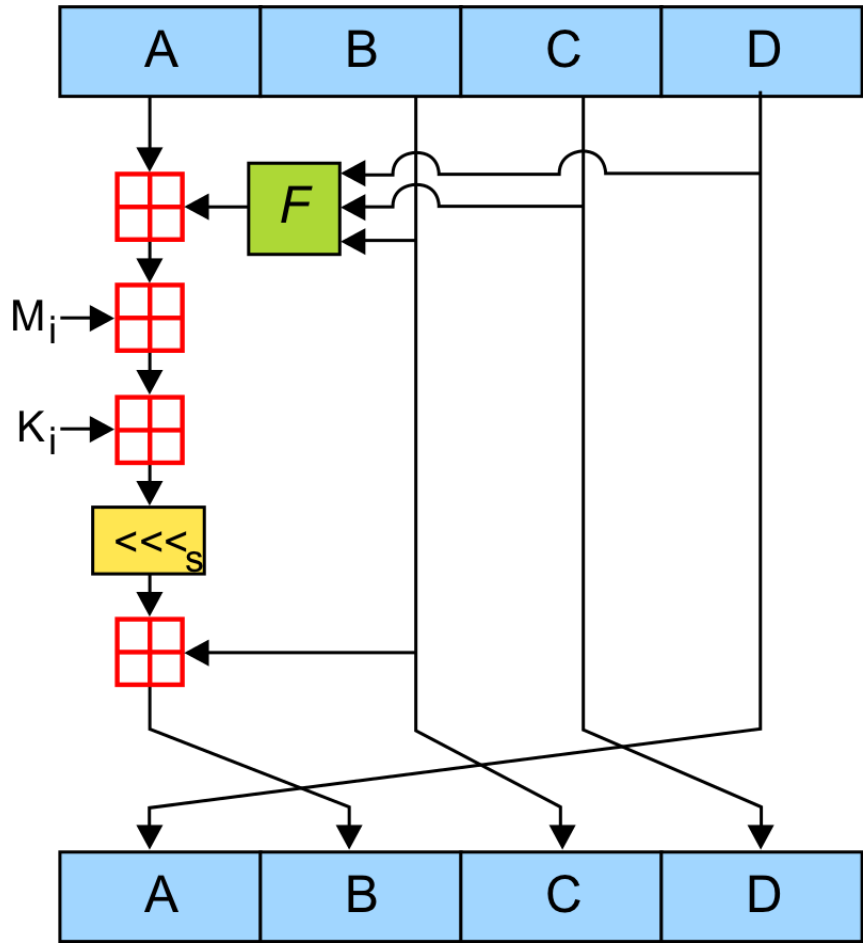



Figure: One round of MD5.*

*https://de.wikipedia.org/wiki/Message-Digest_Algorithm_5

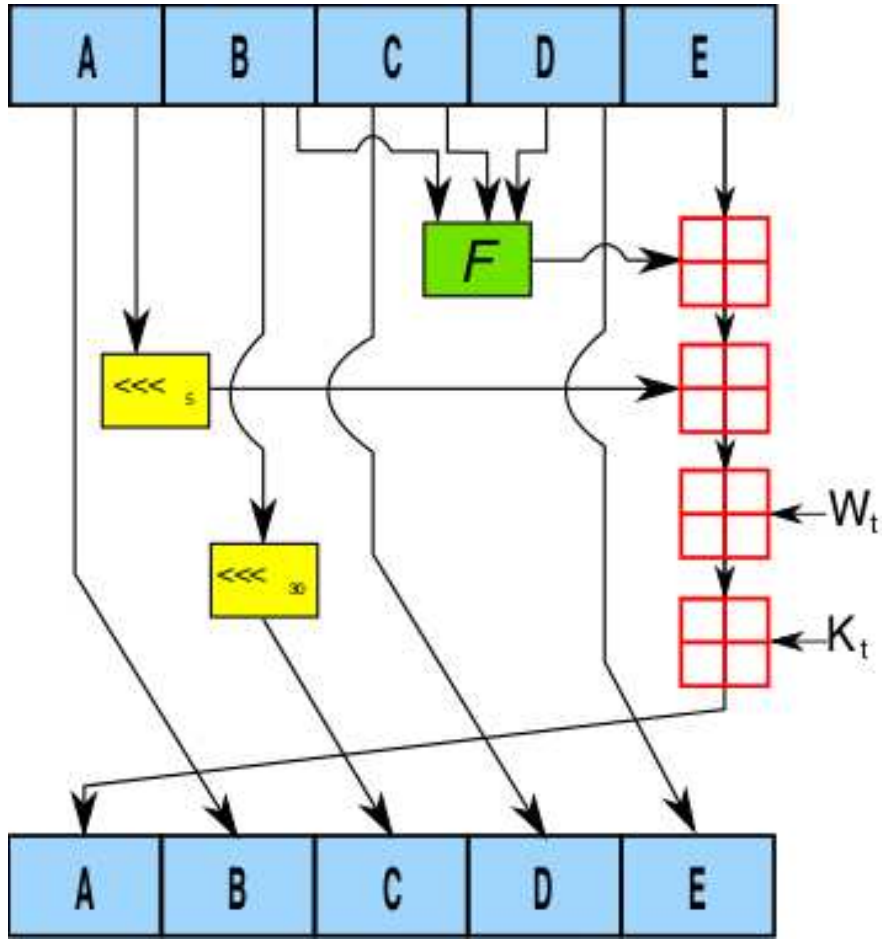



Figure: One round of SHA-1.*

*https://de.wikipedia.org/wiki/Secure_Hash_Algorithm

## Theoretical result

- "Efficient Collision Search Attacks on SHA-0" by Wang et al. (2005)
- "Freestart collision on full SHA-1" by Stevens et al. (2015)

## Theoretical result

- "Efficient Collision Search Attacks on SHA-0" by Wang et al. (2005)
- "Freestart collision on full SHA-1" by Stevens et al. (2015)

## Practical break

None *yet*. (!)

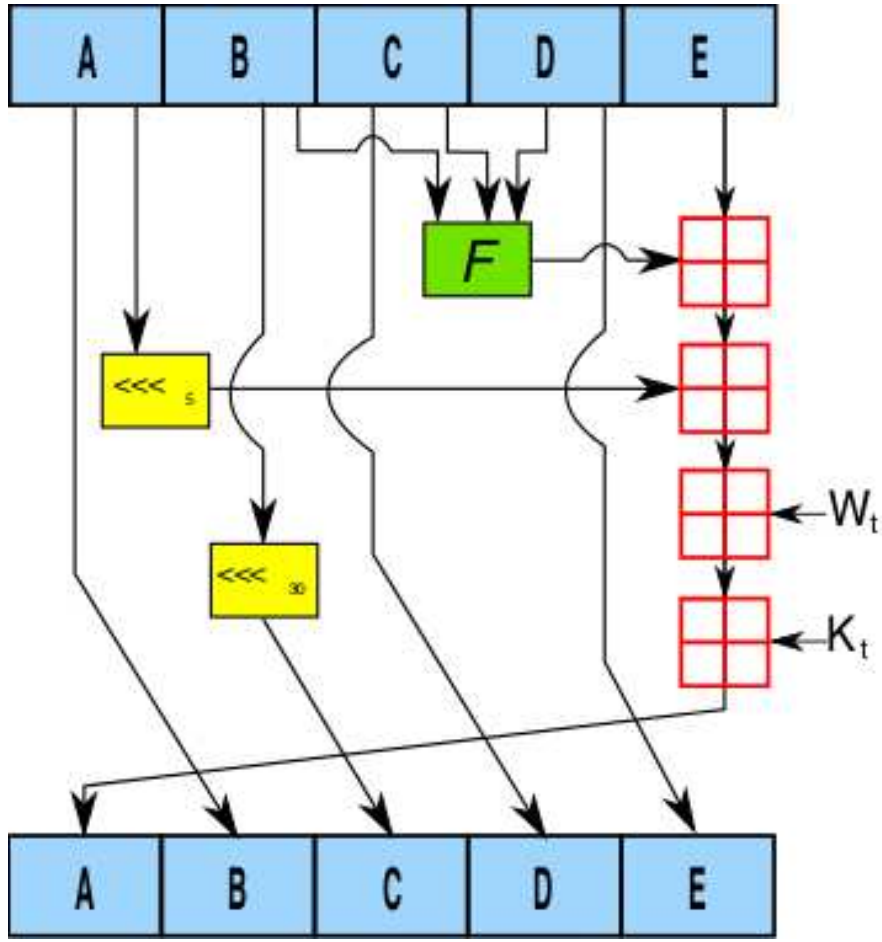



Figure: One round of SHA-1.*

*https://de.wikipedia.org/wiki/Secure_Hash_Algorithm

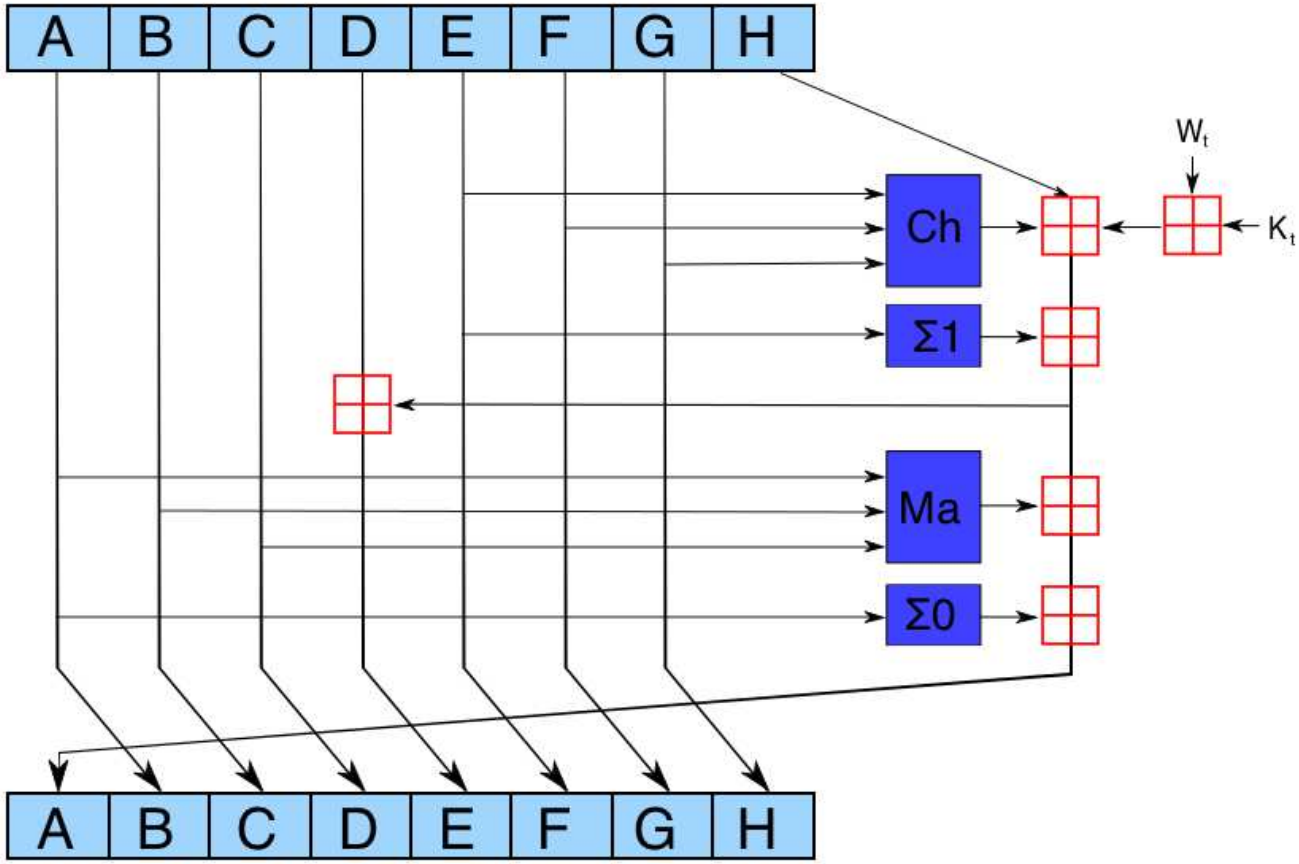



Figure: One round of SHA-256.*

*https://de.wikipedia.org/wiki/SHA-2

Convinced?

To overcome this hash crisis, a competition was started in 2007. In 2012, the winner was announced: Keccak, now known as SHA-3.

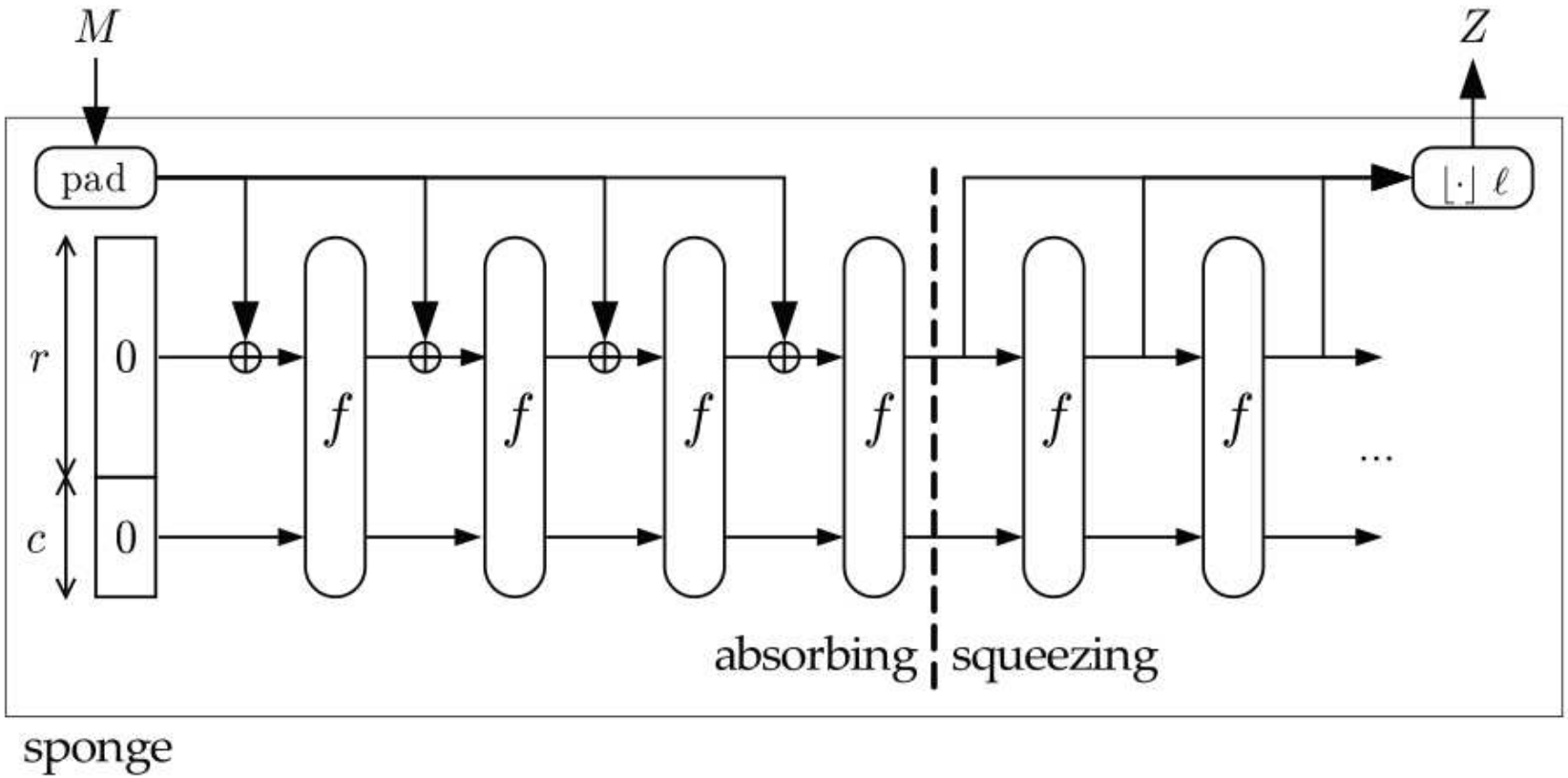



Figure: The SHA-3 sponge construction.*

*http://sponge.noekeon.org/CSF-0.1.pdf

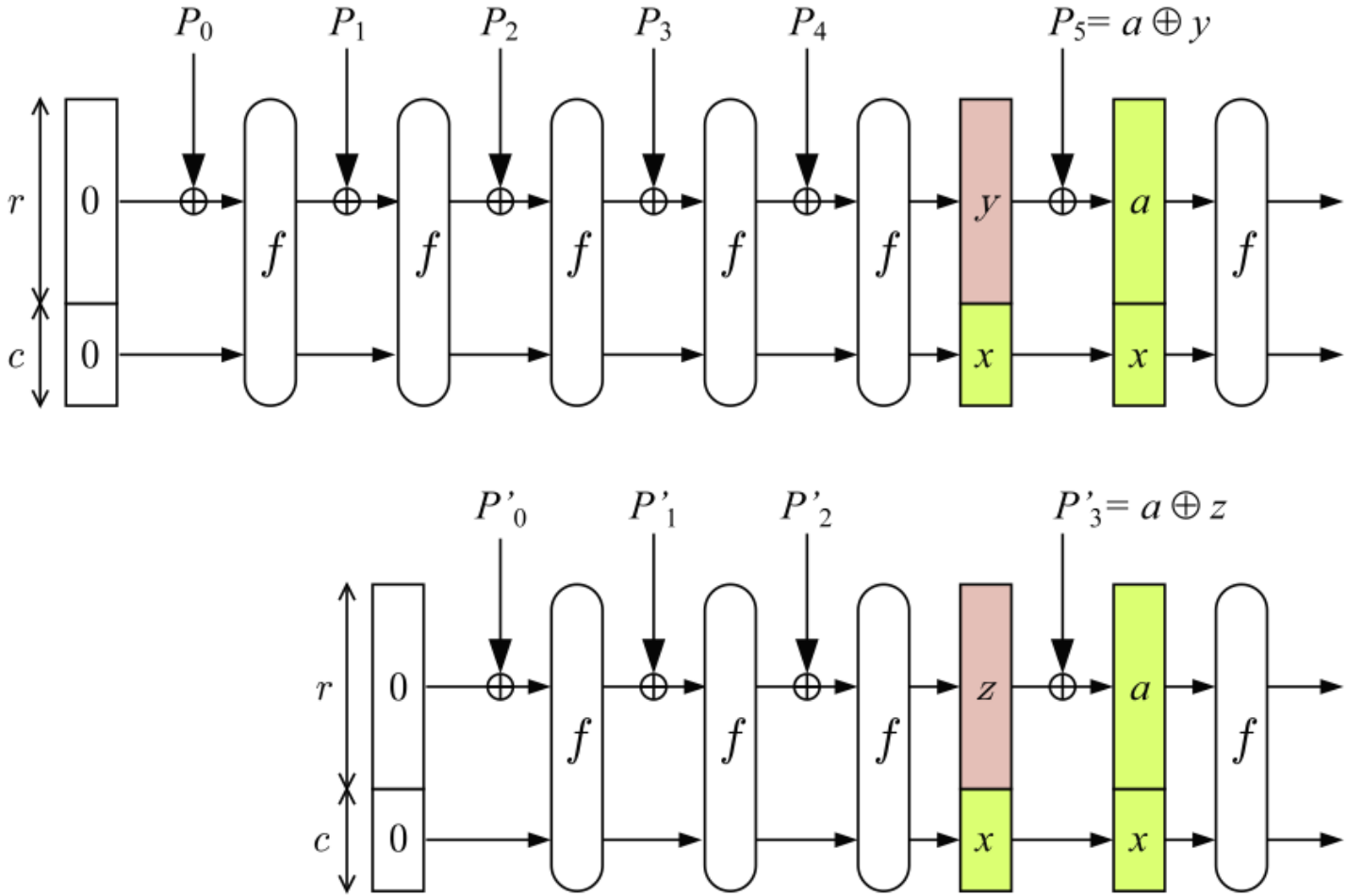



Figure: Collisions in the sponge construction.*

*http://sponge.noekeon.org/CSF-0.1.pdf

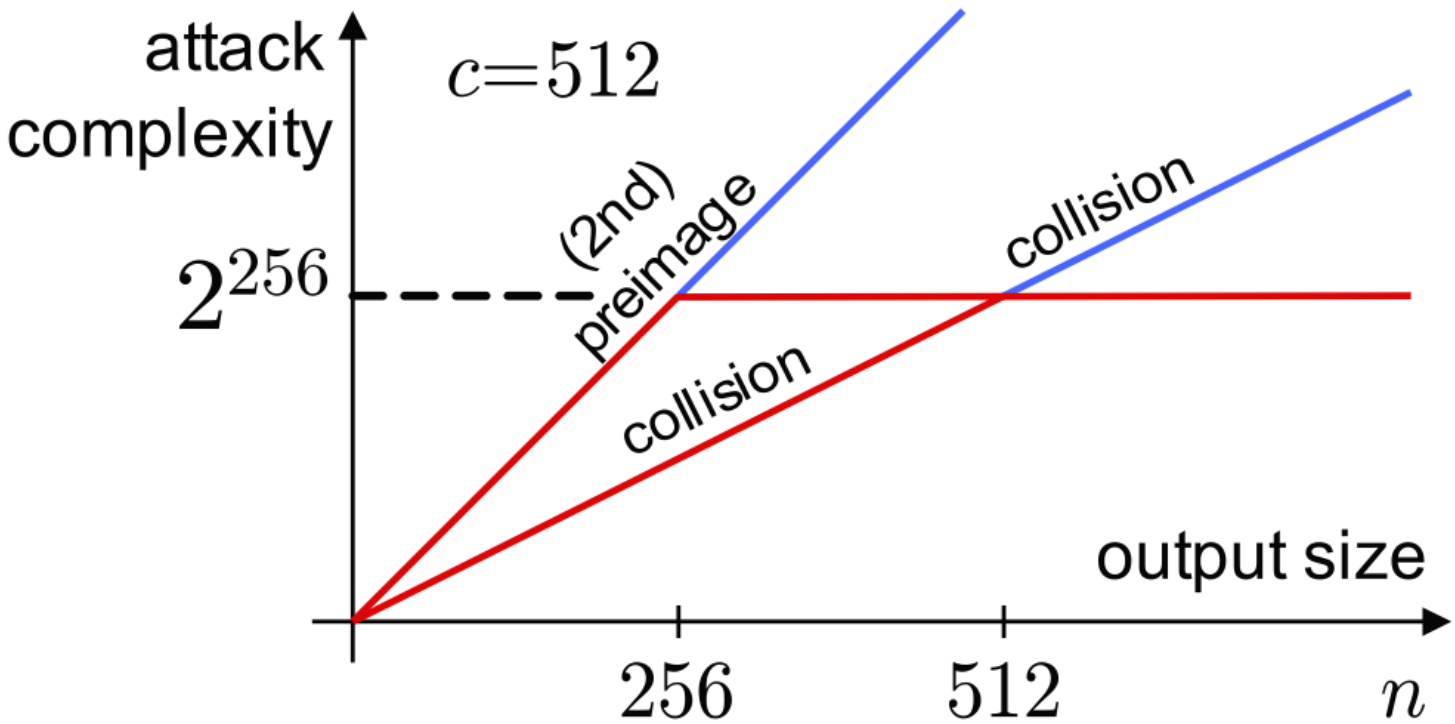



Figure: The sponge claim.*

*http://sponge.noekeon.org/CSF-0.1.pdf

One round of the SHA-3 $f$ function consists of five steps

$$R = \iota \circ \chi \circ \pi \circ \rho \circ \theta,$$

where $\iota$ is addition by some round specific constant.

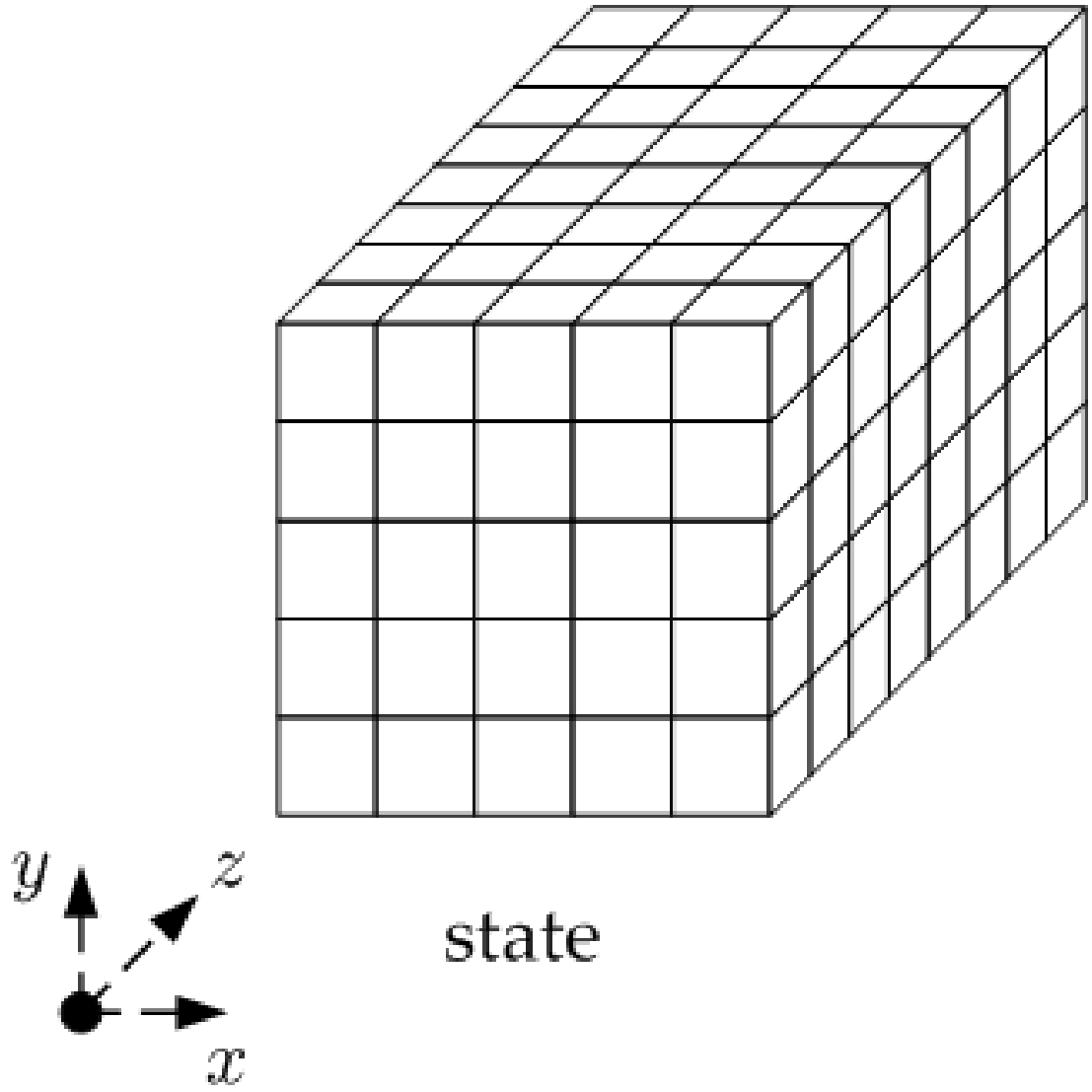



Figure: A state of the SHA-3 $f$ function.*

*http://keccak.noekeon.org/Keccak-reference-3.0.pdf

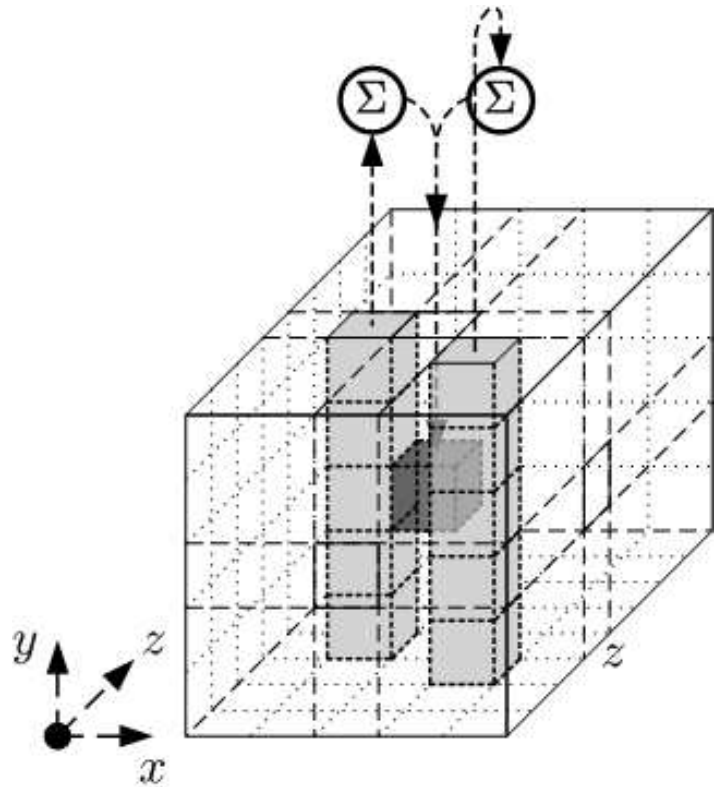



Figure: The step $\theta$ in the SHA-3 $f$ function*

Compute

$$a[x][y][z] \leftarrow a[x][y][z] + \sum_{y'=0}^{4} a[x-1][y'][z] + \sum_{y'=0}^{4} a[x+1][y'][z-1].$$

*http://keccak.noekeon.org/Keccak-reference-3.0.pdf

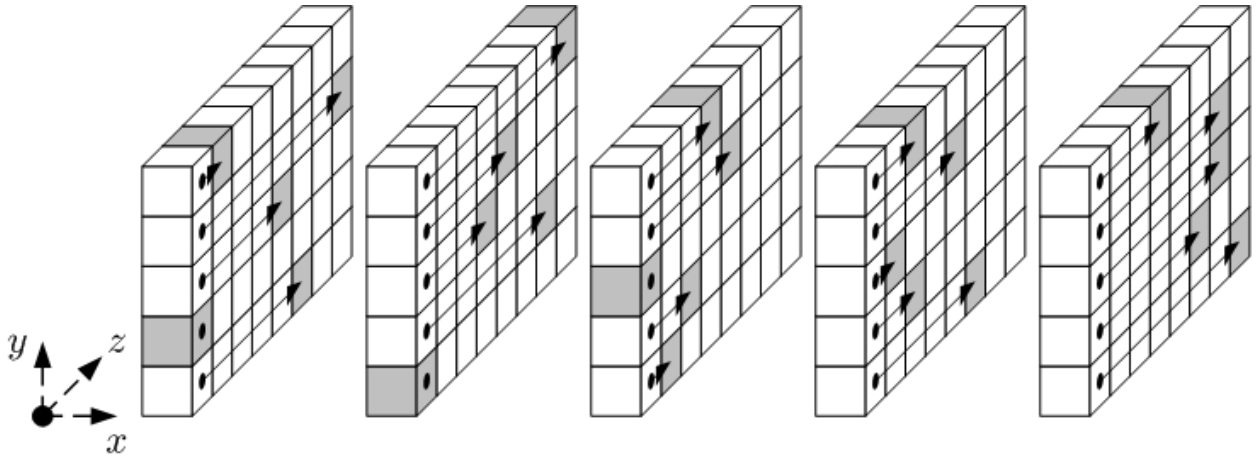



Figure: The step $\rho$ in the SHA-3 $f$ function*

Compute

$$a[x][y][z] \leftarrow a[x][y][z - (t+1)(t+2)/2]$$

for some suitably selected $0 \leq t < 24$.

*`http://keccak.noekeon.org/Keccak-reference-3.0.pdf`

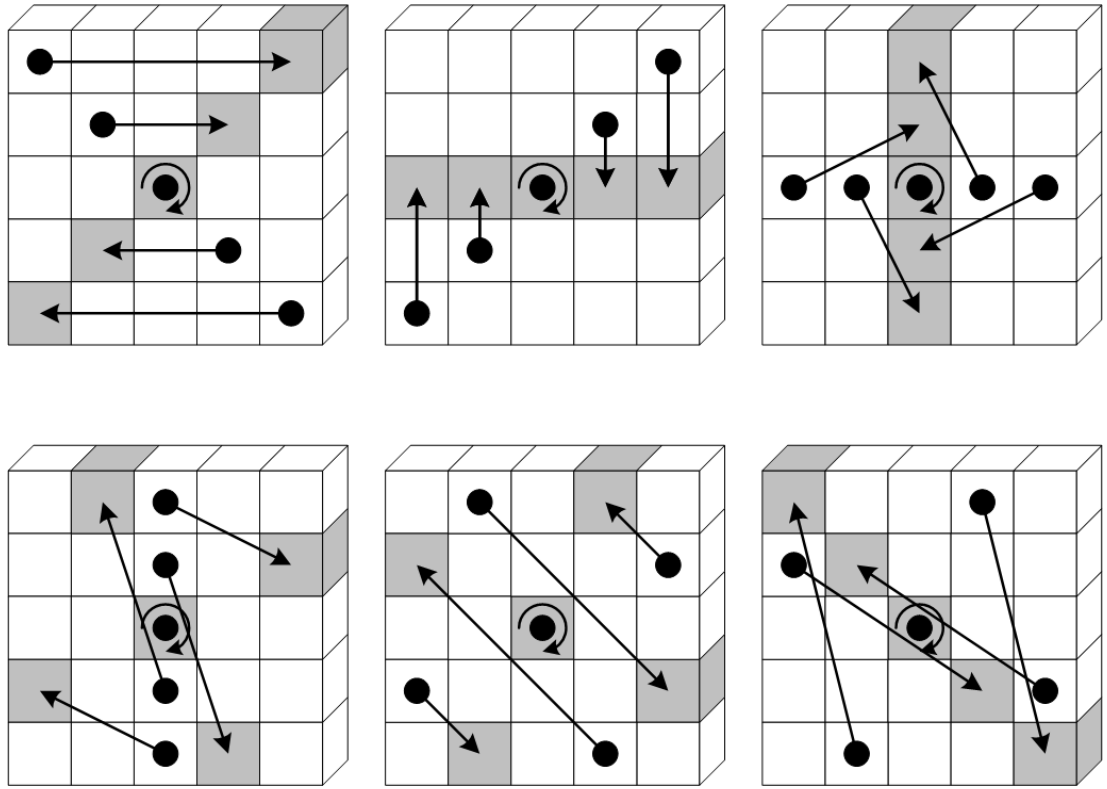


Figure: The step $\pi$ in the SHA-3 $f$ function*

Compute

$$a[x][y] \leftarrow a[x'][y']$$

for some suitably selected $x', y'$.

*`http://keccak.noekeon.org/Keccak-reference-3.0.pdf`

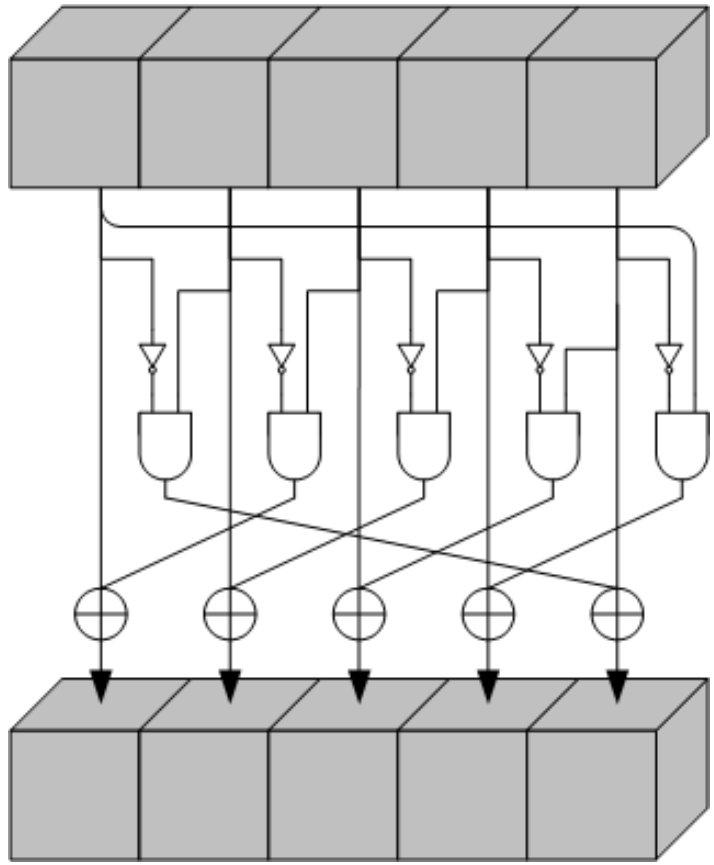


Figure: The step $\chi$ in the SHA-3 $f$ function*

Compute

$$a[x] \leftarrow a[x] + (a[x+1] + 1)a[x+2]$$

*`http://keccak.noekeon.org/Keccak-reference-3.0.pdf`

## Hash functions are essential for security:

- Central property: Collision resistance.
- Several examples known, many broken.
- This lead to a severe crisis in the crypto community.

## Hash crisis resolved for the moment:

- NIST hashing competition 2007 – 2012.
- SHA-3 comes with a thorough security analysis.
- It seems to be the future.

Thanks =)