

Collisions for simplified variants of SHA-256

Krystian Matusiewicz and Josef Pieprzyk

`kmatus@ics.mq.edu.au`, `josef@ics.mq.edu.au`

Centre For Advanced Computing, Algorithms and Cryptography,
Department of Computing,
Macquarie University

- Motivation: How secure is SHA-256?
- Description of SHA-256
- Collisions for a linear variant
- Collisions for a linear variant with Boolean functions
- About S-Boxes
- Conclusions and open problems

Motivation: The family tree of MD functions

1990

MD4

Motivation: The family tree of MD functions

1990

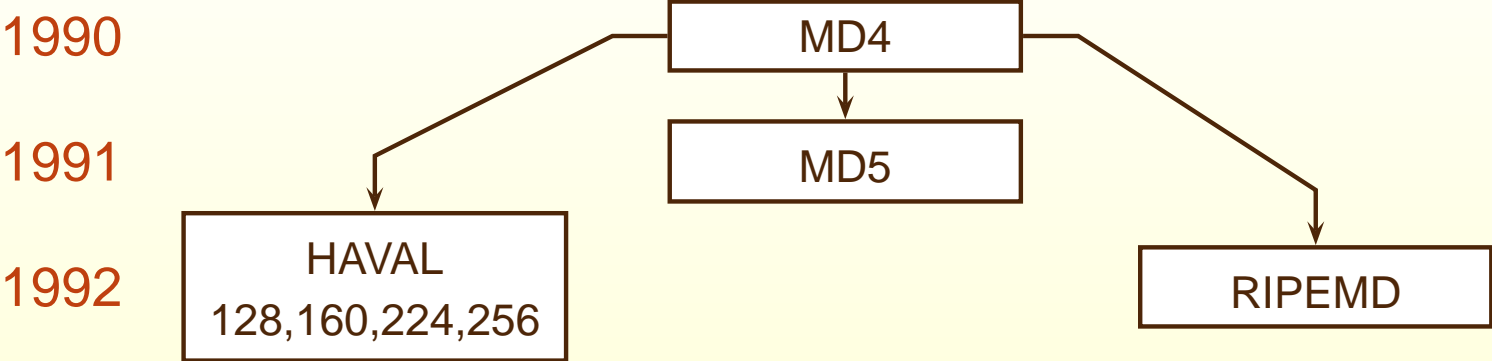
MD4



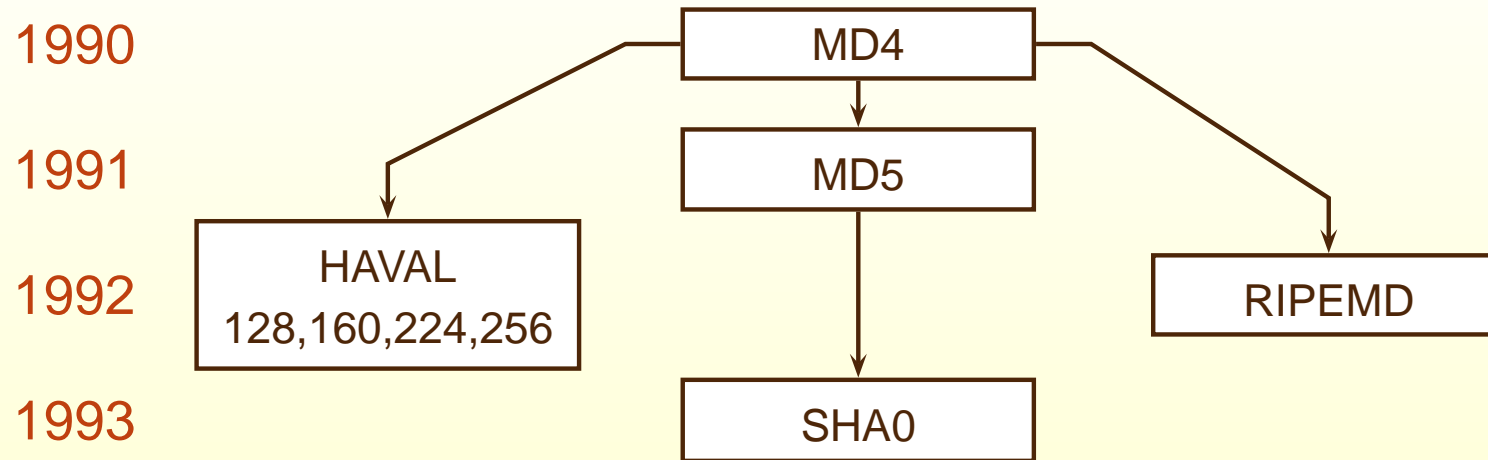
1991

MD5

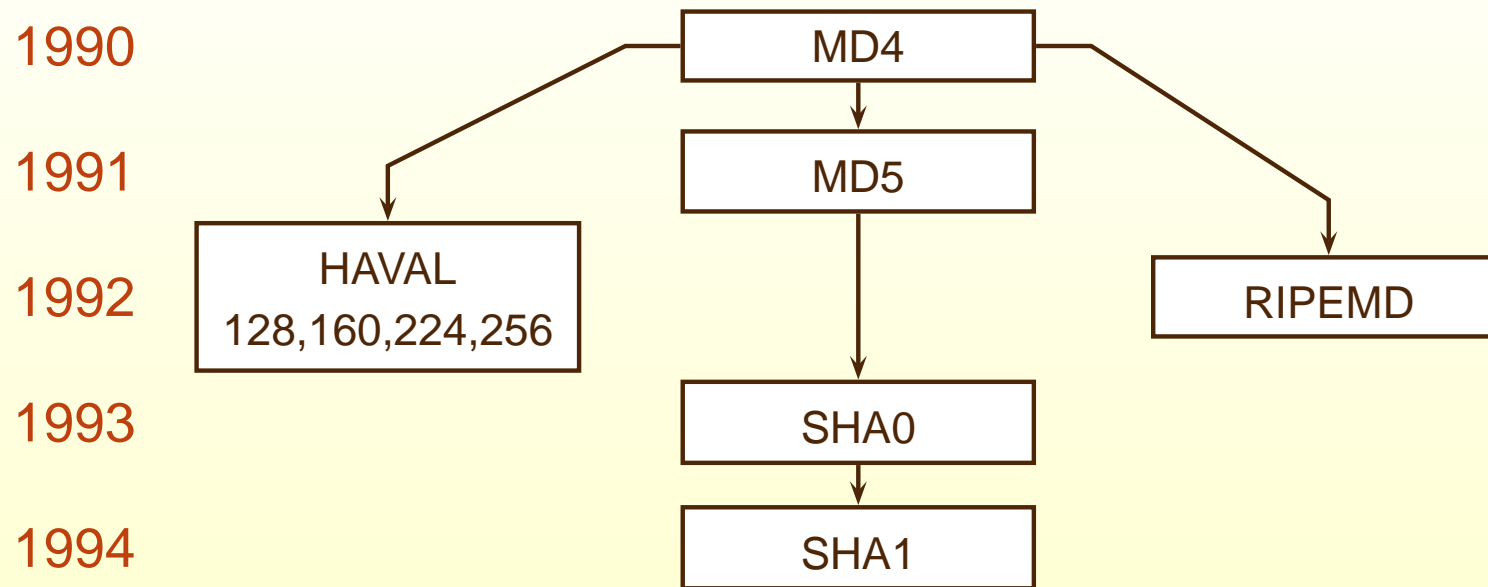
Motivation: The family tree of MD functions



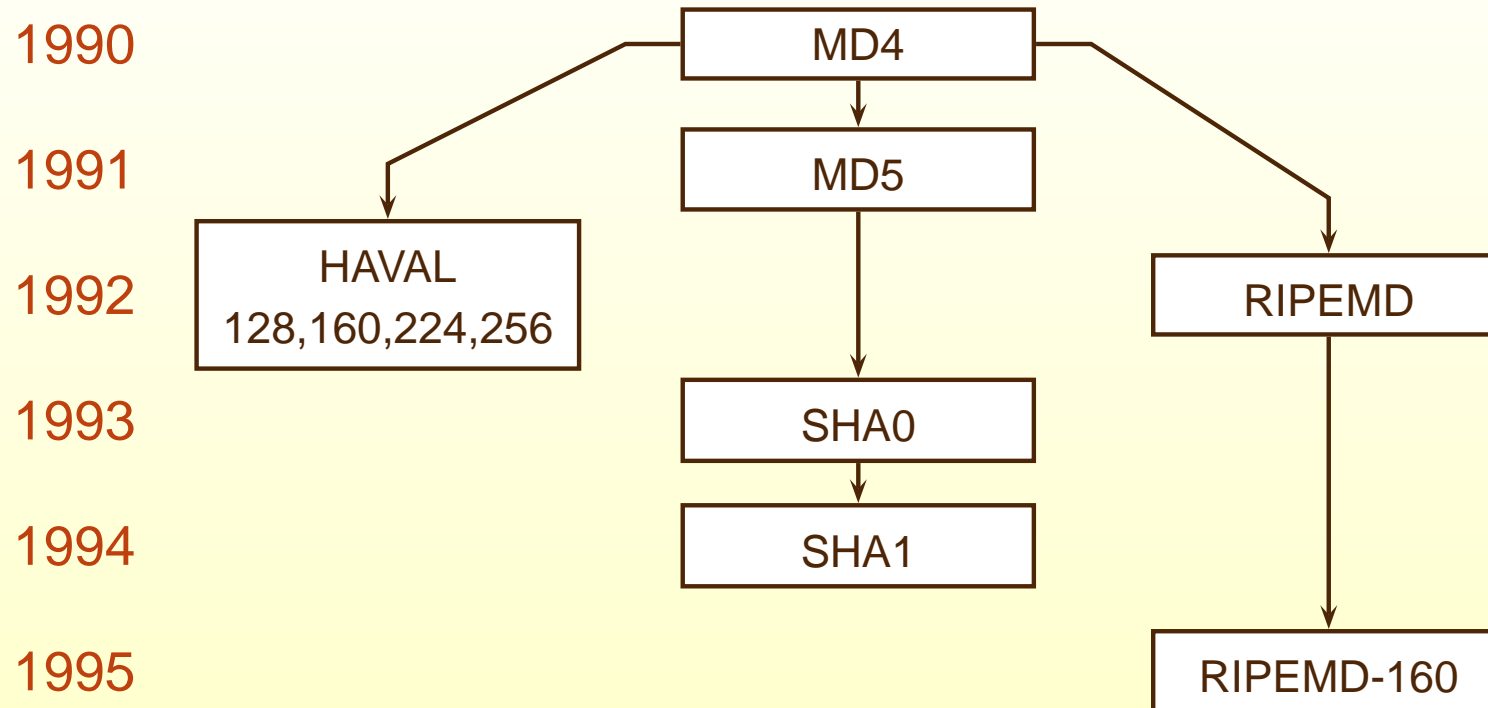
Motivation: The family tree of MD functions



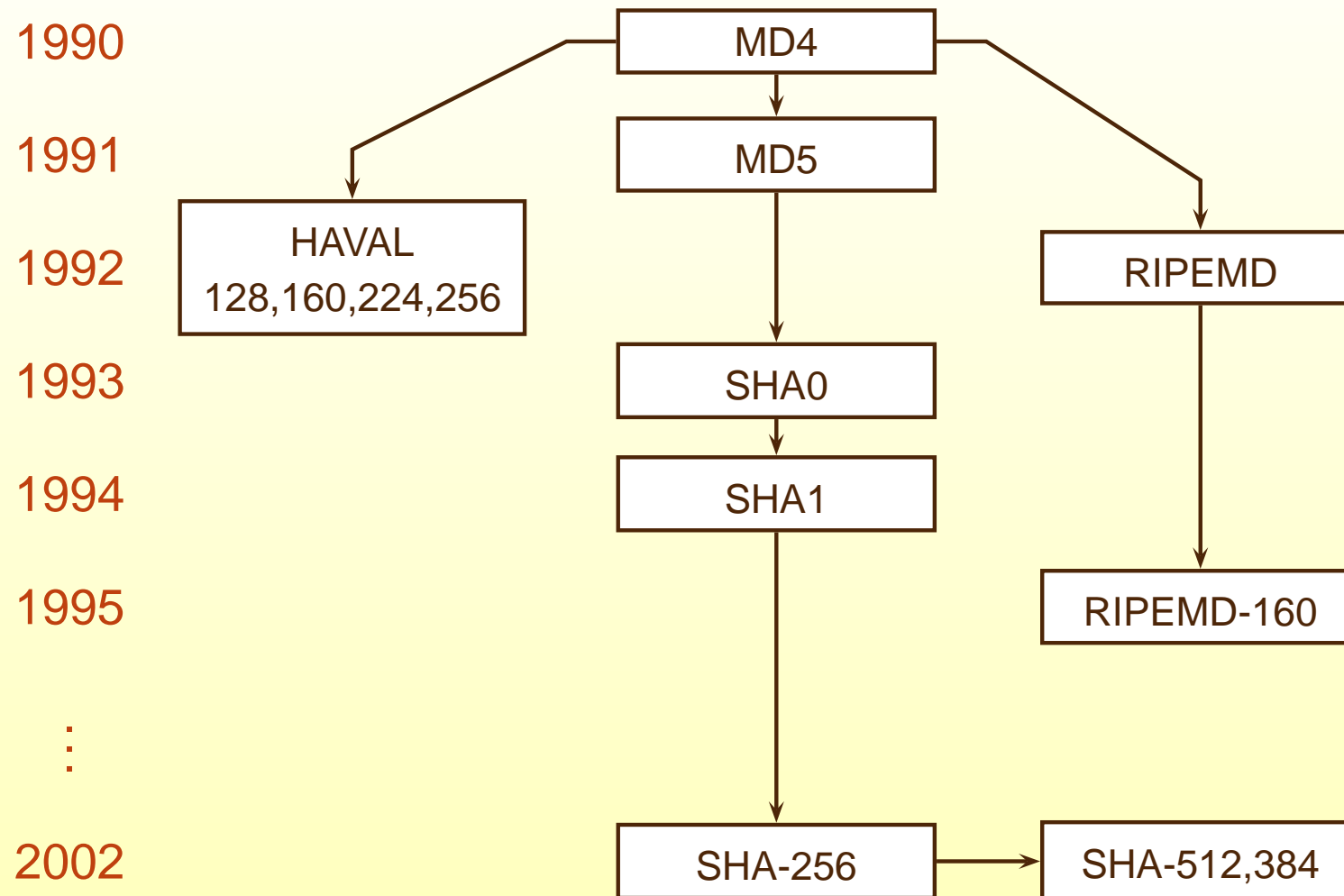
Motivation: The family tree of MD functions



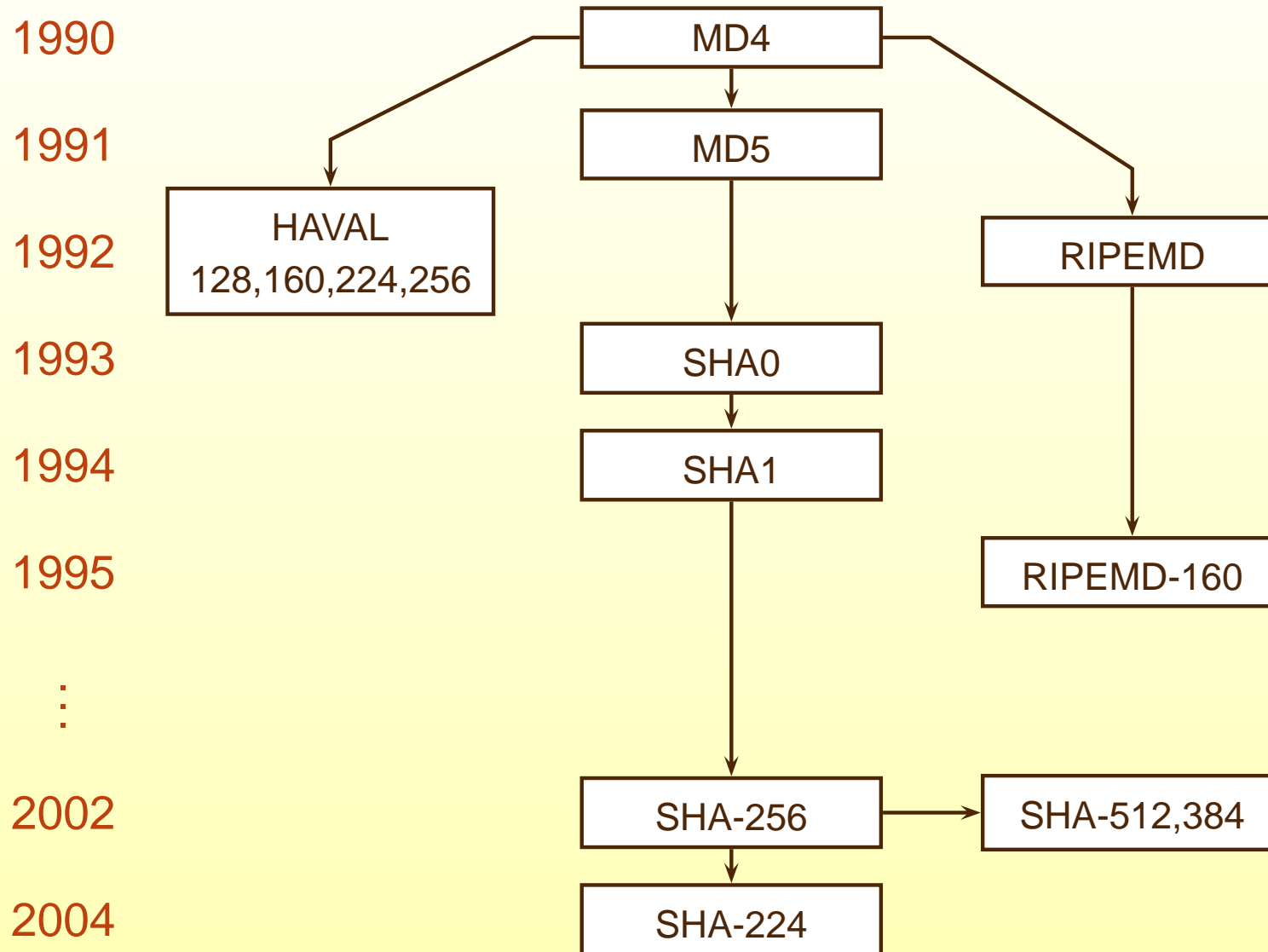
Motivation: The family tree of MD functions



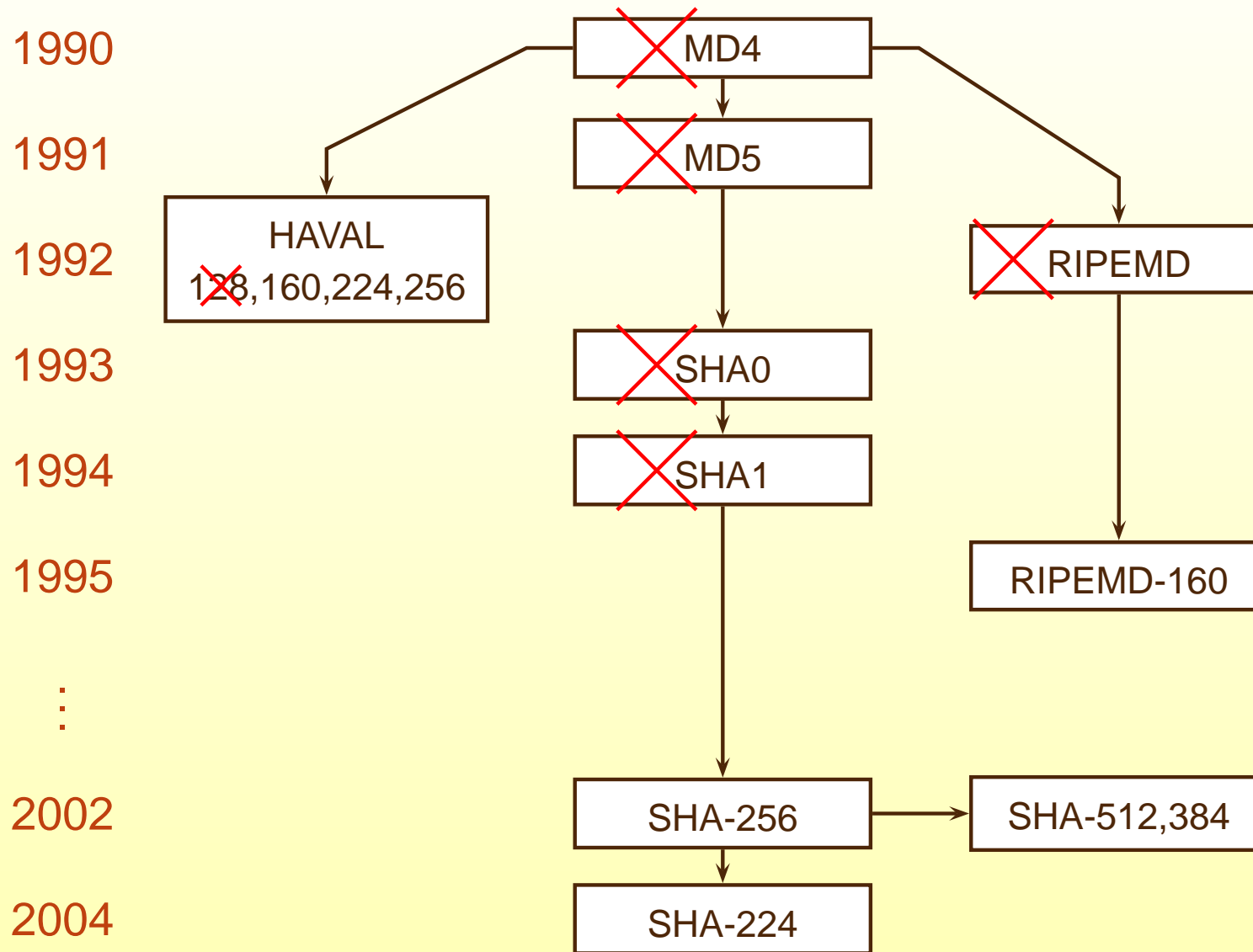
Motivation: The family tree of MD functions



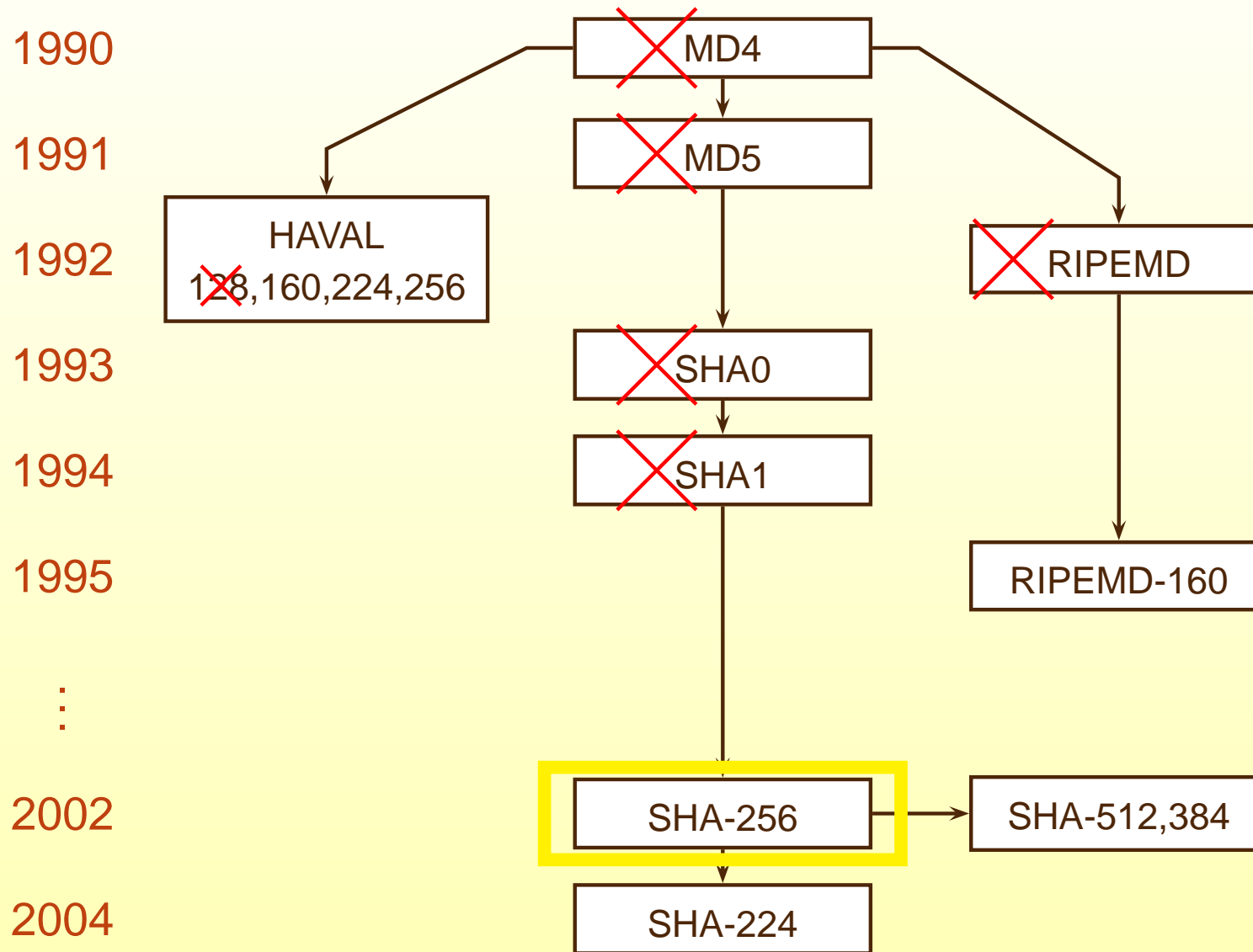
Motivation: The family tree of MD functions



Motivation: The family tree of MD functions



Motivation: The family tree of MD functions



Motivation: Security of SHA-256

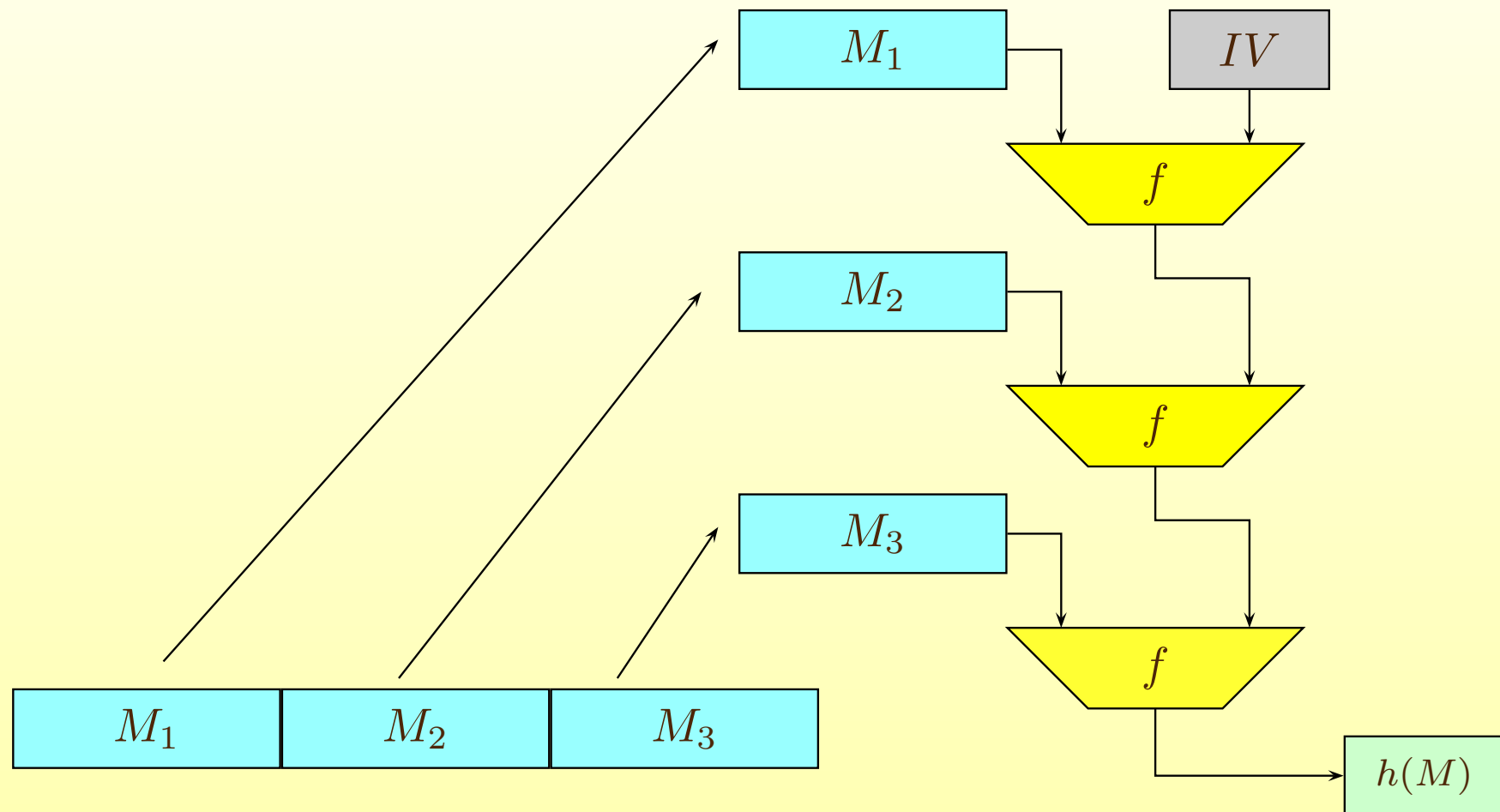
- What is the role of the components of SHA-256?
- How do they contribute to the security of the function?

- Motivation: How secure is SHA-256?
- **Description of SHA-256**
- Collisions for a linear variant
- Collisions for a linear variant with Boolean functions
- About S-Boxes
- Conclusions and open problems

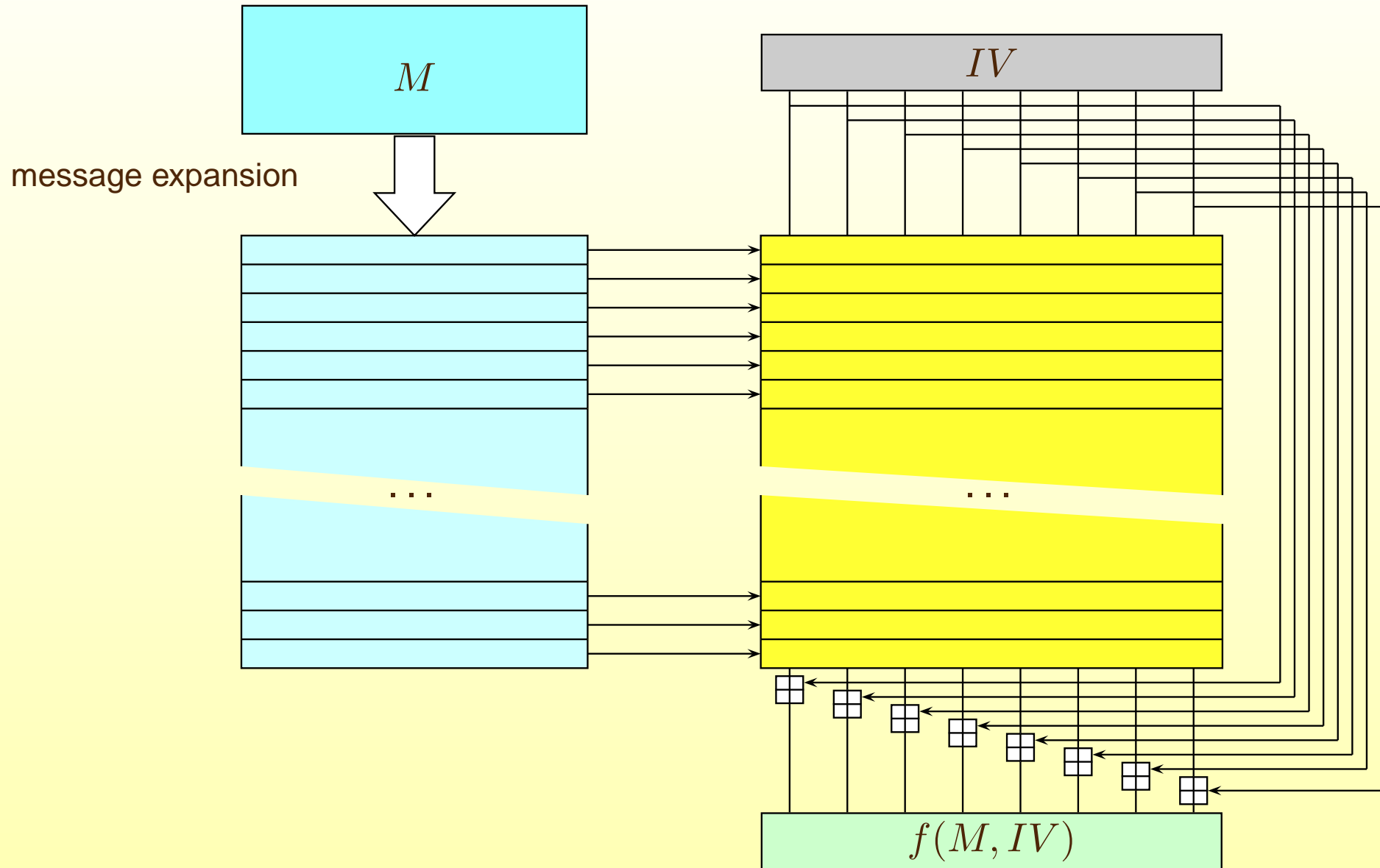
Description of SHA-256

Iterated hash function using a compression function

$$f : \{0, 1\}^{512} \times \{0, 1\}^{256} \rightarrow \{0, 1\}^{256}$$



SHA-256 compression function



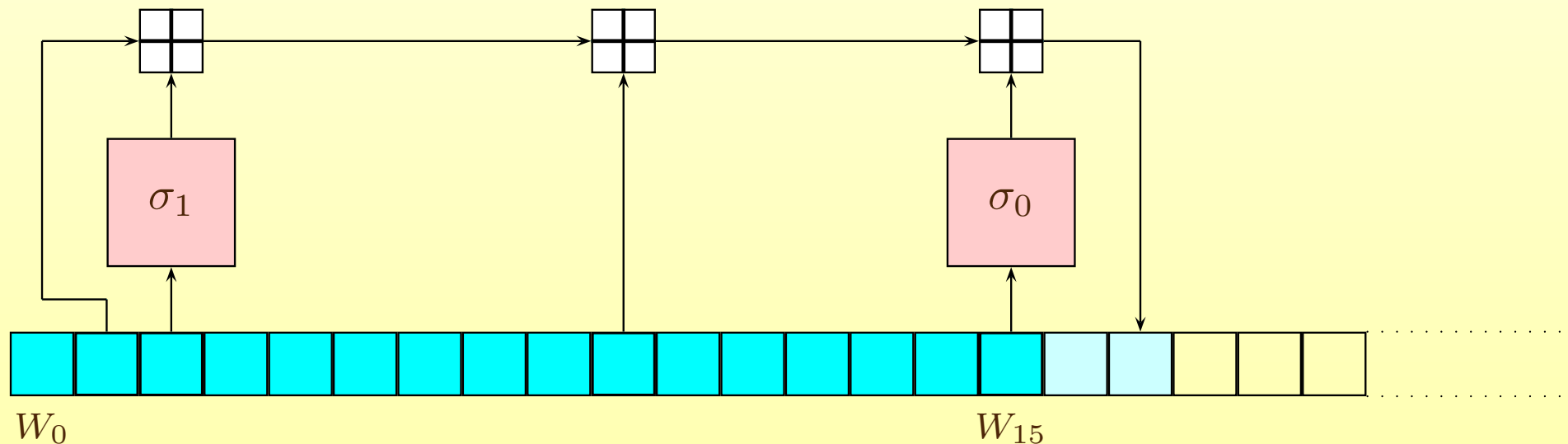
Message expansion of SHA-256

$$W_i = \begin{cases} M_i & \text{for } 0 \leq i < 16, \\ \sigma_1(W_{i-2}) + W_{i-7} + \sigma_0(W_{i-15}) + W_{i-16} & \text{for } 16 \leq i < 64. \end{cases}$$

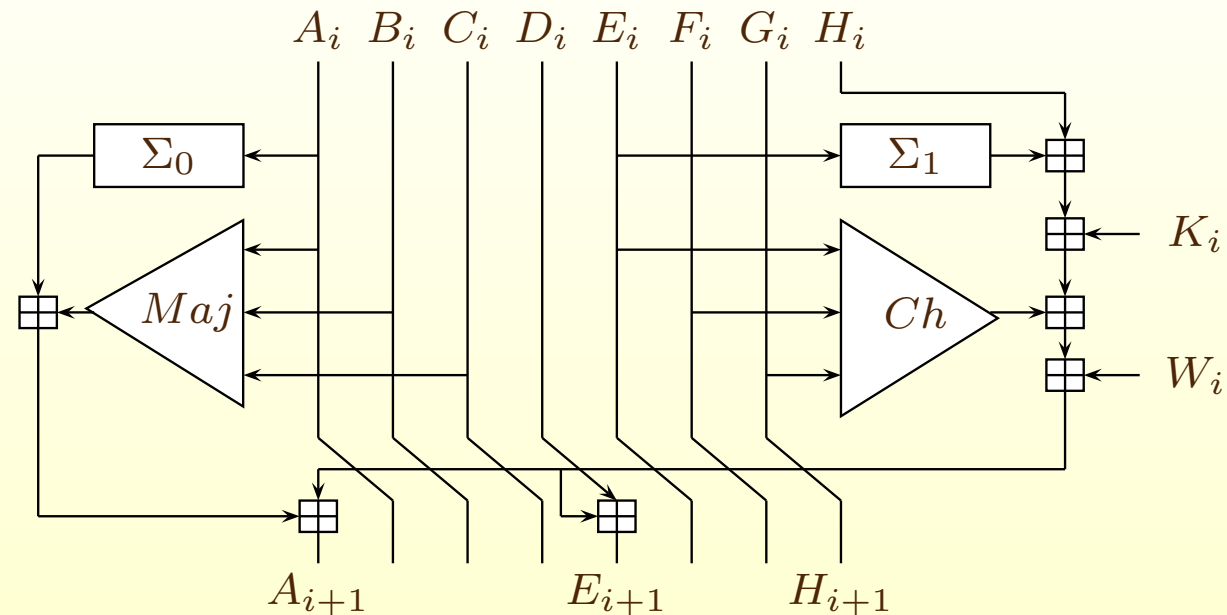
where

$$\sigma_0(x) = ROTR^2(x) \oplus ROTR^{18}(x) \oplus SHR^3(x)$$

$$\sigma_1(x) = ROTR^{17}(x) \oplus ROTR^{19}(x) \oplus SHR^{10}(x)$$



Step transformation of SHA-256



$$\Sigma_0(x) = ROTR^2(x) \oplus ROTR^{13}(x) \oplus ROTR^{22}(x)$$

$$\Sigma_1(x) = ROTR^6(x) \oplus ROTR^{11}(x) \oplus ROTR^{25}(x)$$

$$Maj(A, B, C) = (A \wedge B) \vee (A \wedge C) \vee (B \wedge C)$$

$$Ch(E, F, G) = (E \wedge F) \vee (\neg E \wedge G)$$

- Motivation: How secure is SHA-256?
- Description of SHA-256
- **Collisions for a linear variant**
- Collisions for a linear variant with Boolean functions
- About S-Boxes
- Conclusions and open problems

SHA-256 contains three types of functions:

- \mathbb{F}_2 – linear: $\sigma_0, \sigma_1, \Sigma_0, \Sigma_1$
- $\mathbb{Z}_{2^{32}}$ – linear: addition modulo 2^{32} : $+$
- nonlinear in respect of both structures: bitwise Boolean functions

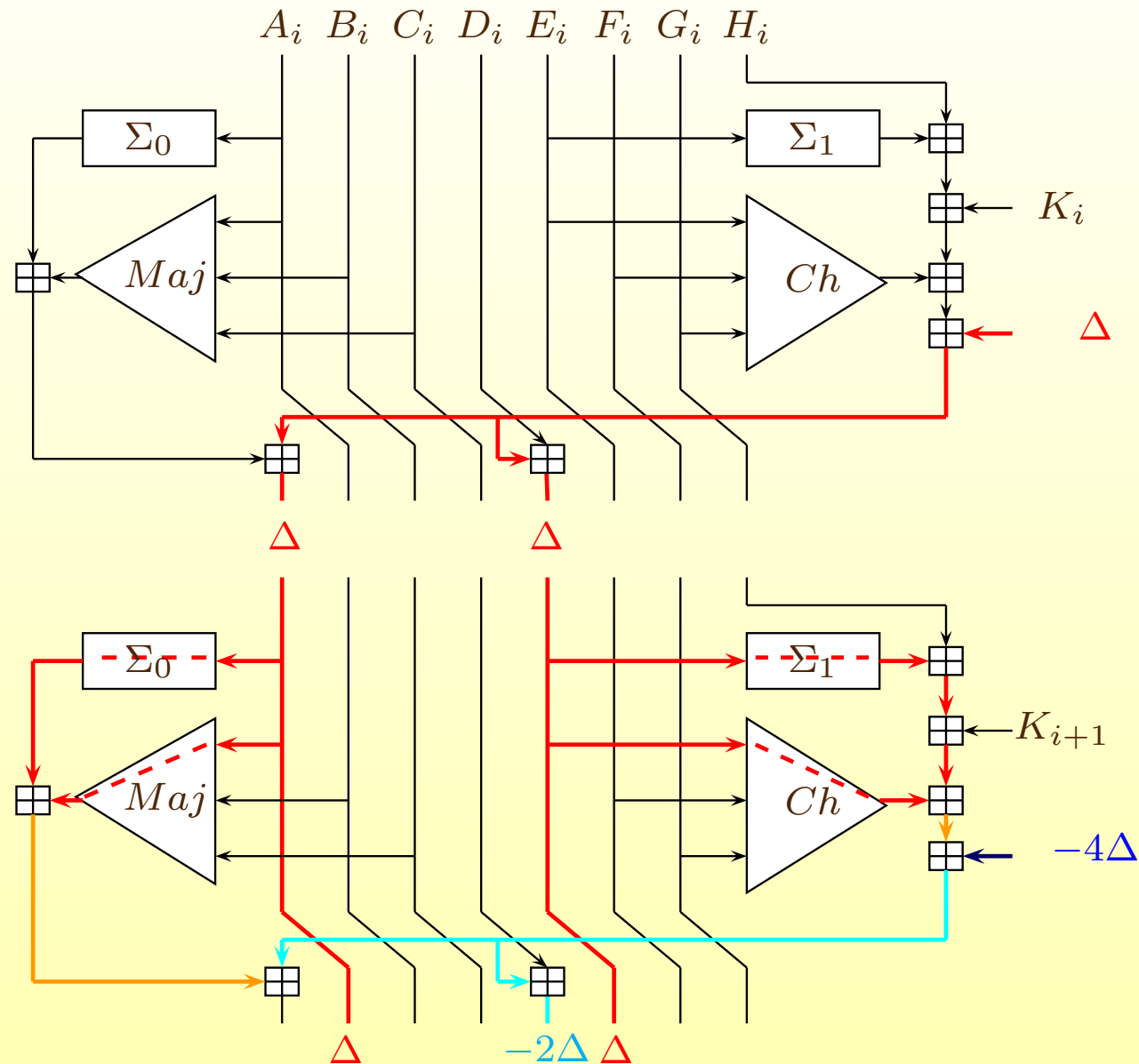
Simplified variant 1:

- replace $\sigma_0, \sigma_1, \Sigma_0, \Sigma_1$ with *id*,
 $\sigma_0(x) = \sigma_1(x) = \Sigma_0(x) = \Sigma_1(x) = x$,
- replace Boolean functions with addition:
 $Maj(x, y, z) = Ch(x, y, z) = x + y + z$

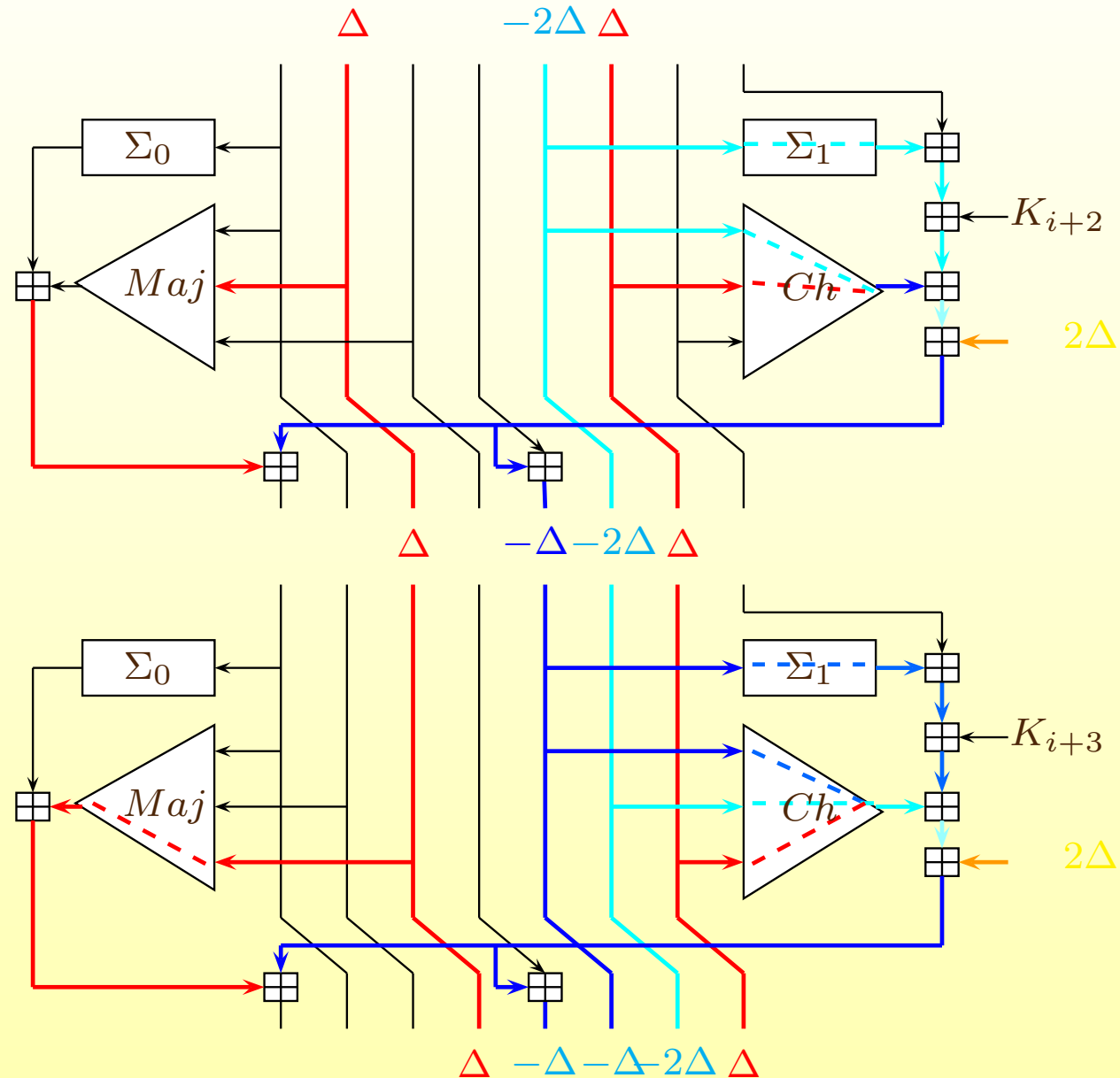
We get fully $\mathbb{Z}_{2^{32}}$ –linear function.

Is it possible to use disturbance-corrections strategy to find collisions for this model?

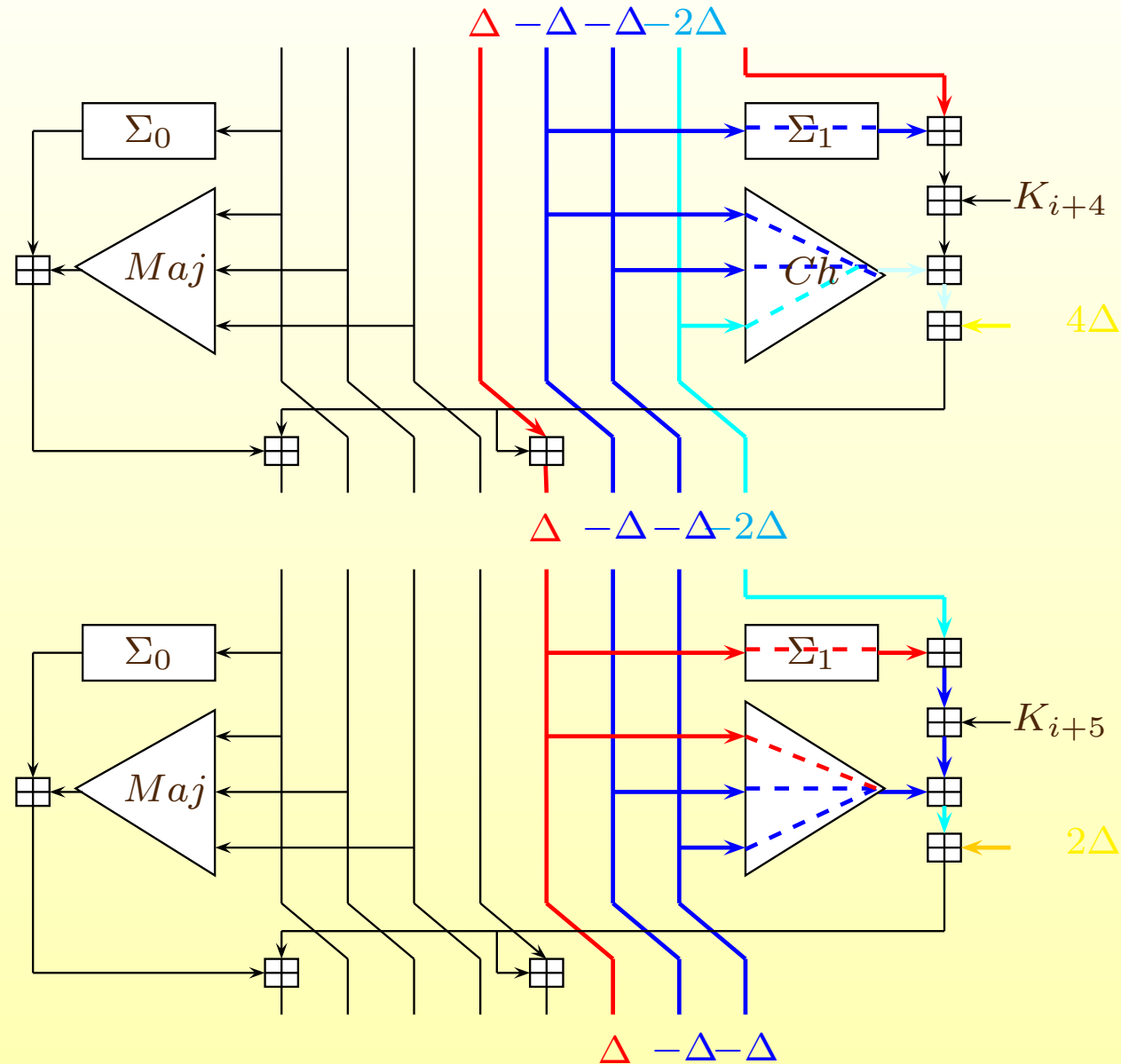
Correcting single disturbance: steps 1 – 2



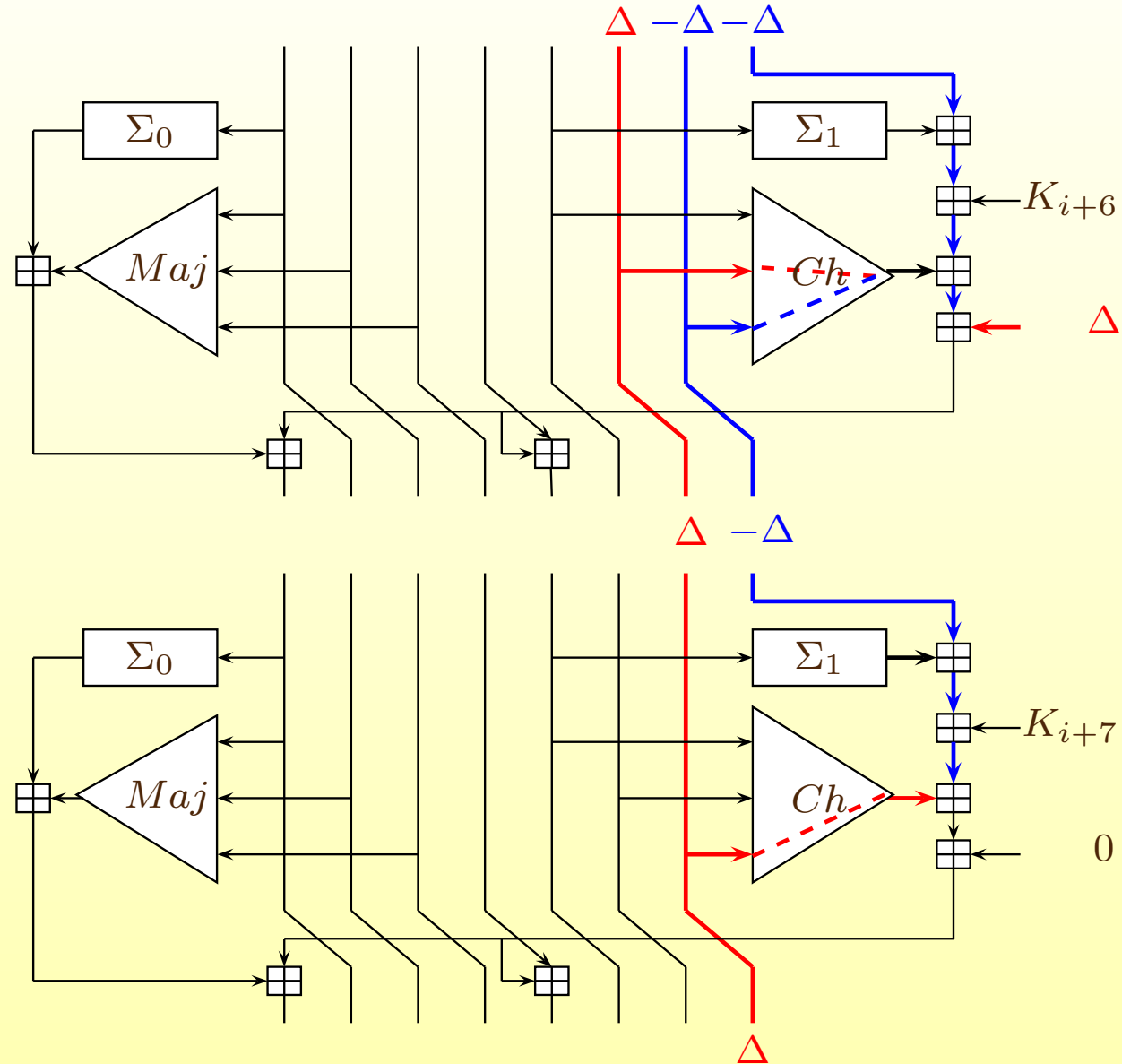
Correcting single disturbance: steps 3 – 4



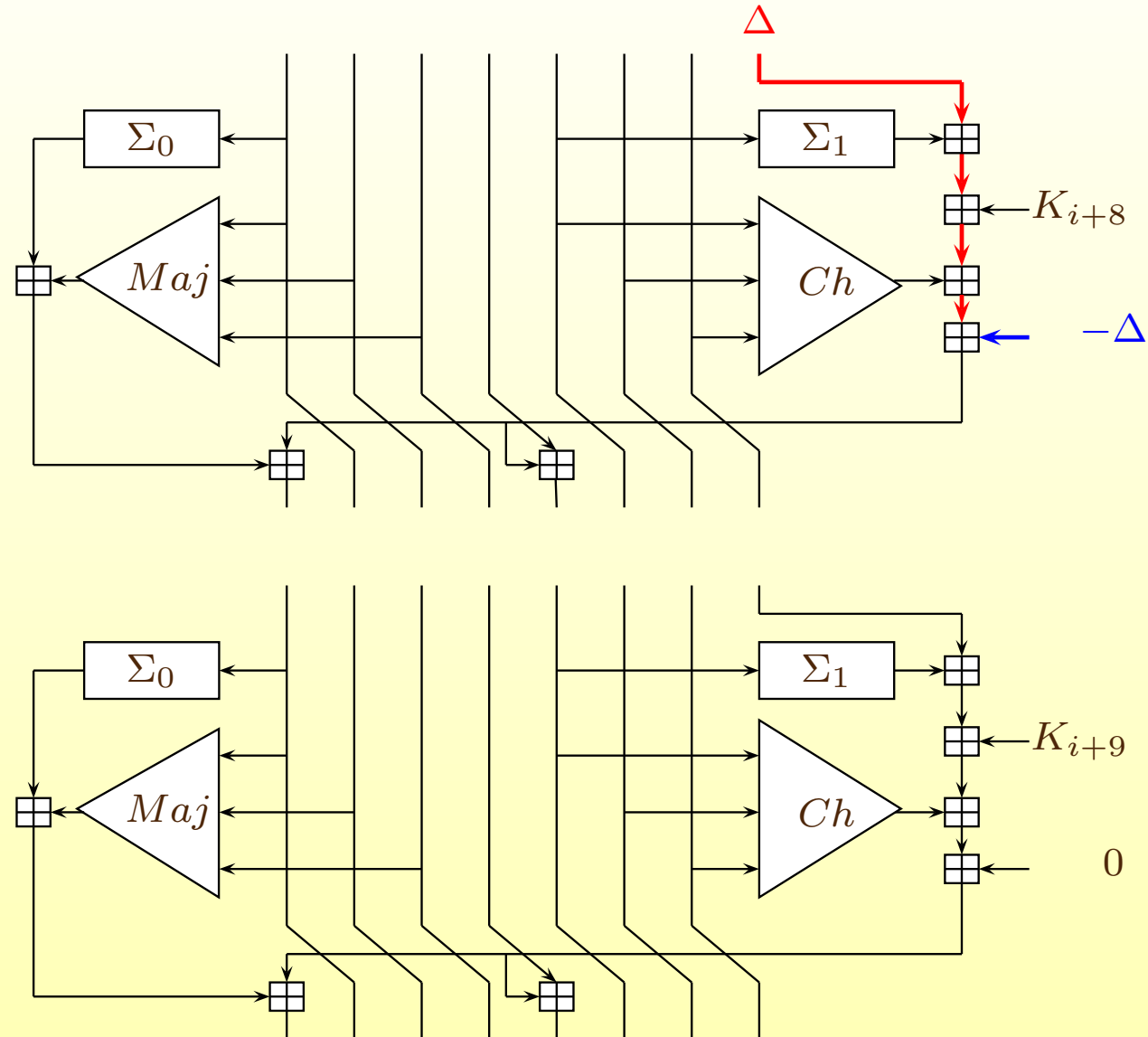
Correcting single disturbance: steps 5 – 6



Correcting single disturbance: steps 7 – 8



Correcting single disturbance: step 9



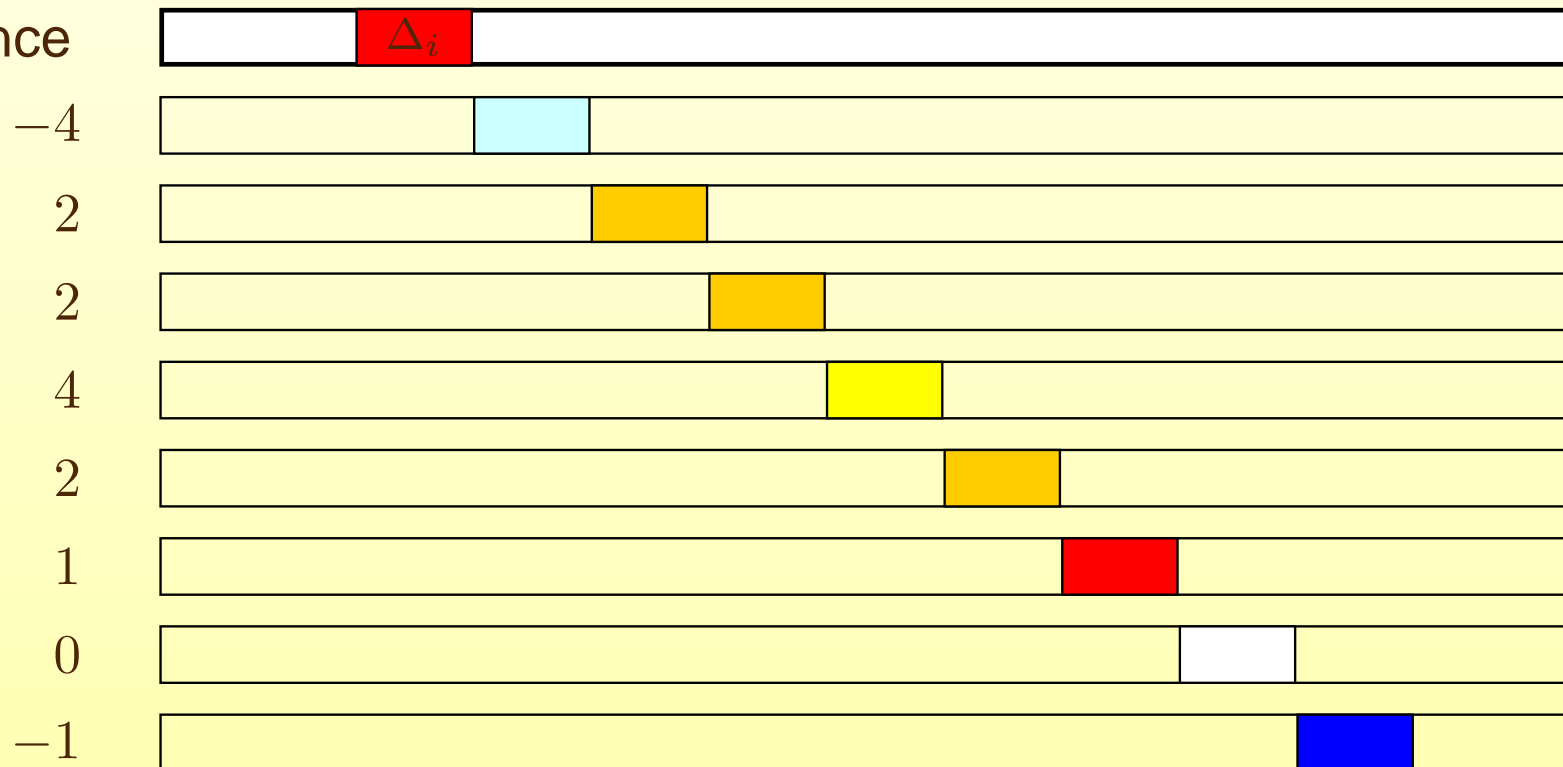
Single corrective pattern

Disturbance in i -th word Δ_i is corrected by the following sequence

$$\Delta_i, -4\Delta_i, 2\Delta_i, 2\Delta_i, 4\Delta_i, 2\Delta_i, \Delta_i, 0, -\Delta_i .$$



disturbance



Conditions for a disturbance vector

We treat expanded messages as vectors $W \in \mathbb{Z}_2^{64}$

A difference $\Delta = W' - W$ is a valid disturbance pattern if two conditions are satisfied:

- C1. the last 8 words of Δ are zero,
- C2. Δ with prepended 8 zero block must also be the result of the expansion process.

C1 is necessary to allow enough time to correct the last difference as 8 steps are needed to correct each disturbance.

C2 is necessary for constructing a corrective pattern as a linear combination of Δ and “delayed” disturbance vectors.

For disturbance pattern

$$\Delta = [\Delta_0, \dots, \Delta_{63}]^T$$

the full corrective pattern is computed as

$$\begin{aligned} C &= \Delta - 4 \cdot [0, \Delta_0, \dots, \Delta_{62}]^T \\ &\quad + 2 \cdot [0, 0, \Delta_0, \dots, \Delta_{61}]^T \\ &\quad + 2 \cdot [0, 0, 0, \Delta_0, \dots, \Delta_{60}]^T \\ &\quad + \dots \\ &\quad - 1 \cdot [0, 0, 0, 0, 0, 0, 0, 0, \Delta_0, \dots, \Delta_{55}]^T. \end{aligned}$$

„Delayed” pattern $[0, 0, 0, 0, 0, 0, 0, 0, \Delta_0, \dots, \Delta_{55}]^T$ has to be the result of the expansion.

Message expansion as a linear transform

Message expansion with $\sigma_0 = \sigma_1 = id$ is $\mathbb{Z}_{2^{32}}$ -linear, so it can be represented as 64×16 matrix

$$E = \begin{bmatrix} I_{16} \\ A \\ A^2 \\ A^3 \end{bmatrix},$$

where A is a linear transform producing 16 new words out of 16 old ones according to the recurrence relation.

Then we have

$$W = E \cdot M$$

where $M \in \mathbb{Z}_{2^{32}}^{16}$ is the initial message and $W \in \mathbb{Z}_{2^{32}}^{64}$ is the expanded message.

Finding disturbance patterns

We are looking for such message differences $\Delta_M = M' - M$ that expanded differences $\Delta = E(\Delta_M)$ satisfy conditions C1 and C2.

This can be written as

$$0 = A^3[8 :: 16] \cdot \Delta_M \quad \text{the last 8 elements of } \Delta \text{ are zero}$$

$$0 = A^{-1}[8 :: 16] \cdot \Delta_M \quad \text{8 prepended elements of } \Delta \text{ would be zero}$$

where $M[a :: b]$ means a matrix consisting of rows of matrix M from a -th row to b -th row, inclusive.

These two matrix equations form a linear system over the ring $\mathbb{Z}_{2^{32}}$.

Finding disturbance patterns: solving the system

The system

$$0 = A^3[8 :: 16] \cdot \Delta_M$$

$$0 = A^{-1}[8 :: 16] \cdot \Delta_M$$

has one-dimensional solution space given by

$$\Delta_M = [0x10000000, 0xA0000000, 0xC0000000, 0xA0000000, \\ 0xE0000000, 0x20000000, 0x40000000, 0x40000000, \\ 0x80000000, 0xD0000000, 0x10000000, 0x60000000, \\ 0x50000000, 0x40000000, 0x70000000, 0x30000000]^T.$$

Any nonzero multiple of this vector constitutes a valid disturbance pattern for linearized version of SHA-256 – we can use it to find collisions.

- Motivation: How secure is SHA-256?
- Description of SHA-256
- Collisions for a linear variant
- **Collisions for a linear variant with Boolean functions**
- About S-Boxes
- Conclusions and open problems

The next step: Incorporating Boolean functions

Let us consider a variant still without S-boxes $\sigma_0, \sigma_1, \Sigma_0, \Sigma_1$ but with Boolean functions *Maj* and *Ch*.

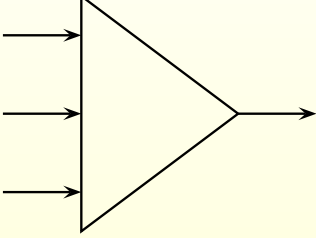
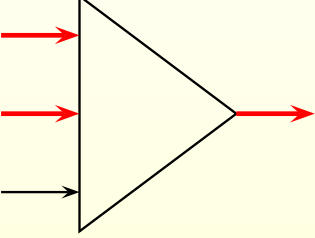
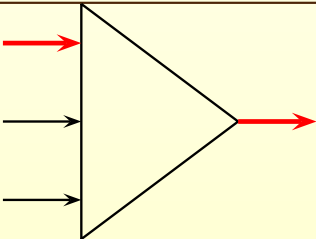
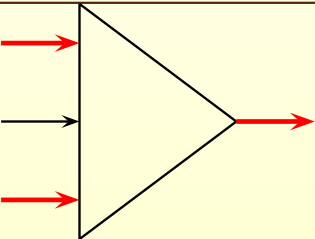
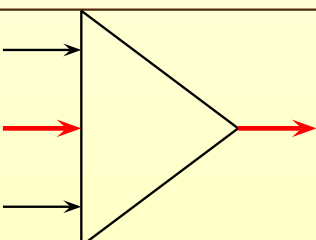
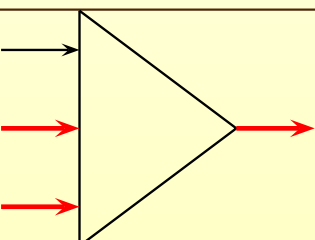
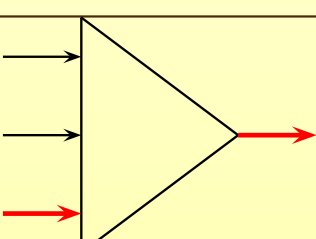
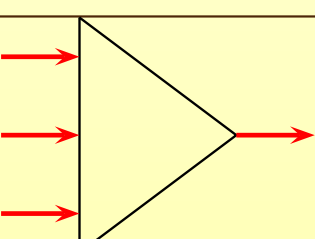
If we multiply the pattern by 8, we get a valid disturbance pattern with 1's in the most significant bits only.

```
1000000001101011 1011100110100110
0000011100101111 1011100000000000
```

There are only 27 nonzero bits in this pattern.

We can approximate Boolean functions with certain probabilities.

Approximation of Boolean functions

$(\delta_x, \delta_y, \delta_z)$	illustration	$(\delta_x, \delta_y, \delta_z)$	illustration
$(0,0,0)$		$(1,1,0)$	
$(1,0,0)$		$(1,0,1)$	
$(0,1,0)$		$(0,1,1)$	
$(0,0,1)$		$(1,1,1)$	

Approximation of Boolean functions

We can approximate both Boolean functions with probability at least $1/2$ by a “function” that produces output difference each time input difference is nonzero.

input difference $(\delta_x, \delta_y, \delta_z)$	<i>Ch</i> function		<i>Maj</i> function	
	conditions	Prob	conditions	Prob
(1,0,0)	$y + z = 1$	1/2	$y + z = 1$	1/2
(0,1,0)	$x = 1$	1/2	$x + z = 1$	1/2
(0,0,1)	$x = 0$	1/2	$x + y = 1$	1/2
(1,1,0)	$x + y + z = 0$	1/2	$x + y = 0$	1/2
(1,0,1)	$x + y = 0$	1/2	$x + z = 0$	1/2
(0,1,1)	—	1	$y + z = 0$	1/2
(1,1,1)	$y + z = 0$	1/2	—	1

Corrective pattern for this variant

For the variant with Boolean functions approximated by „always output difference” and disturbance pattern with nonzero bits in only the most significant position, single corrective sequence has the following form

$$\Delta_i, 0, 0, \Delta_i, \Delta_i, 0, 0, 0, \Delta_i$$

After obtaining the full corrective pattern we can estimate the probability of a successful correction.

Probabilities of successful corrections in each step

s	Maj	Ch	e	s	Maj	Ch	e	s	Maj	Ch	e	s	Maj	Ch	e
0	000	000	0	16	110	010	2	32	011	100	2	48	111	110	1
1	100	100	2	17	111	101	1	33	001	010	2	49	111	011	0
2	010	010	2	18	011	010	2	34	000	001	1	50	011	101	2
3	001	101	2	19	101	001	2	35	000	100	1	51	101	010	2
4	000	110	1	20	110	100	2	36	000	010	1	52	110	101	2
5	000	111	1	21	111	110	1	37	000	001	1	53	111	110	1
6	000	011	0	22	011	011	1	38	100	100	2	54	011	011	1
7	000	001	1	23	001	101	2	39	110	110	2	55	001	101	2
8	000	000	0	24	100	110	2	40	111	011	0	56	000	010	1
9	000	000	0	25	110	011	1	41	011	001	2	57	000	101	1
10	100	100	2	26	011	101	2	42	001	100	2	58	000	010	1
11	110	110	2	27	101	110	2	43	100	110	2	59	000	001	1
12	011	111	2	28	010	011	1	44	010	111	2	60	000	000	0
13	101	111	2	29	001	001	2	45	101	011	1	61	000	000	0
14	010	011	1	30	100	000	1	46	110	001	2	62	000	000	0
15	101	101	2	31	110	000	1	47	111	100	1	63	000	000	0

Straightforward result

$$e = \prod_{i=0}^{63} e_i = 84$$

$$\text{Prob}[\text{collision}] = 2^{-e} = \mathbf{2^{-84}}$$

We can do better. By appropriate selection of message words in 16 first steps we can eliminate probabilistic behaviour in these steps and get better probability

$$e' = \prod_{i=16}^{63} e_i = 64$$

$$\text{Prob}[\text{collision}] = 2^{-e'} = \mathbf{2^{-64}}$$

- Motivation: How secure is SHA-256?
- Description of SHA-256
- Collisions for a linear variant
- Collisions for a linear variant with Boolean functions
- **About S-Boxes**
- Conclusions and open problems

The role of S-Boxes : the full SHA-256 structure

- S-Boxes provide diffusion of differences
- one bit input difference gives 2–3 bit output difference
- still possible to use modular differentials

$$Prob[\Sigma(x + \delta) - \Sigma(x) = \Sigma(\delta)] = 2^{-3}$$

for one-bit input differences δ ,

- we also need another difference

$$Prob[\Sigma(x + \gamma) - \Sigma(x) = \Sigma(\gamma)] \approx 2^{-9}$$

for input difference $\gamma = \Sigma(\delta)$,

- using these differentials, single correction sequence for full round structure has a probability of 2^{-42}
- Hawkes, Paddon, Rose using some additional optimizations achieved 2^{-39}

- Motivation: How secure is SHA-256?
- Description of SHA-256
- Collisions for a linear variant
- Collisions for a linear variant with Boolean functions
- About S-Boxes
- **Conclusions and open problems**

Conclusions and open problems

- it is possible to use disturbance-corrections strategy for SHA-256-like architecture

Conclusions and open problems

- it is possible to use disturbance-corrections strategy for SHA-256-like architecture
- mixing provided by modular additions and Boolean functions alone is not sufficient for building a secure hash function

Conclusions and open problems

- it is possible to use disturbance-corrections strategy for SHA-256-like architecture
- mixing provided by modular additions and Boolean functions alone is not sufficient for building a secure hash function
- S-Boxes are vital for the security of SHA-256

Conclusions and open problems

- it is possible to use disturbance-corrections strategy for SHA-256-like architecture
- mixing provided by modular additions and Boolean functions alone is not sufficient for building a secure hash function
- S-Boxes are vital for the security of SHA-256
- can we force the full message expansion process to produce differences that follow disturbance-corrections patterns?

Conclusions and open problems

- it is possible to use disturbance-corrections strategy for SHA-256-like architecture
- mixing provided by modular additions and Boolean functions alone is not sufficient for building a secure hash function
- S-Boxes are vital for the security of SHA-256
- can we force the full message expansion process to produce differences that follow disturbance-corrections patterns?
- are there any other high probability differentials for SHA-256?

The End

Thank you!

Lemma [Hawkes, Paddon, Rose]

Let $\lambda = \Delta X = X' \oplus X$ and $\delta X = X' - X$. Having ΔX we can determine δX if we know all $X[i]$ for all $i < 31$ such that $\lambda[i] = 1$