

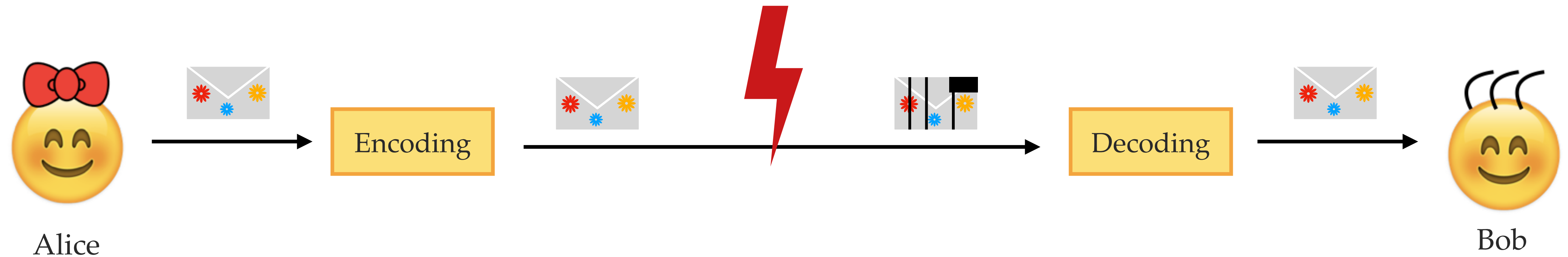
Code-based cryptography I

Monika Trimoska

Selected Areas in Cryptology - Part 1
Spring, 2024

TU/e

Error-correcting codes



- Primary use case: communication over a noisy channel.
- Main idea: introduce some **redundancy** in order to be able to correct the errors.
- Some structured error-correcting codes have efficient decoding algorithms.
- Decoding is, in general, a **hard problem** - so it is hard for *random* codes.

↪ Hard problems (often) find their use in cryptography.

Linear codes

Linear code

An $[n, k]$ **linear code** \mathcal{C} over \mathbb{F}_q is a k -dimensional subspace of \mathbb{F}_q^n .

- The parameter n is called the **length** of the code.
- The parameter k is called the **dimension** of the code.
- The elements in the code are called **codewords**.

Hamming metric

For $\mathbf{x} \in \mathbb{F}_q^n$ the **Hamming weight** of \mathbf{x} is the number of nonzero elements, aka.

$$\text{wt}(\mathbf{x}) = |\{i \in \{1, \dots, n\} \mid x_i \neq 0\}|.$$

Generator matrix

The matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ is called a **generator matrix** of \mathcal{C} , if

$$\mathcal{C} = \{\mathbf{xG} \mid \mathbf{x} \in \mathbb{F}_q^k\}.$$

Linear codes

Linear code

An $[n, k]$ **linear code** \mathcal{C} over \mathbb{F}_q is a k -dimensional subspace of \mathbb{F}_q^n .

Example. $q = 2, n = 5, k = 3$

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Codewords: $\lambda_1(10101) + \lambda_2(11000) + \lambda_3(11110)$

Example. $\mathbf{c}_1 = (111)\mathbf{G} = (10011),$
 $\mathbf{c}_2 = (100)\mathbf{G} = (10101)$

Linear code equivalence

Isometry

An **isometry** (for our purposes) between two codes \mathcal{C} and \mathcal{D} is a **linear map** $\mu : \mathcal{C} \rightarrow \mathcal{D}$ that **preserves the metric**.



In our case: an isometry preserves the **Hamming weight** of codewords.

Which linear transformations preserve the Hamming weight?

→ Multiply a codeword by $\mathbf{A} \in \text{GL}_n$?

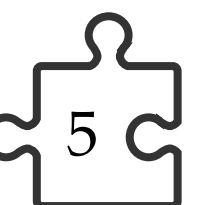
Example. $q = 7, n = 5, k = 3, \mathbf{G} = \begin{pmatrix} 2 & 0 & 4 & 0 & 1 \\ 6 & 1 & 0 & 0 & 0 \\ 3 & 4 & 1 & 3 & 0 \end{pmatrix}$

Let $\mathbf{c} = (100)\mathbf{G} = (20401)$

→ $\mathbf{c}\mathbf{A} = (20401)\mathbf{A} = (12533)$

→ $\text{wt}(\mathbf{c}) \neq \text{wt}(\mathbf{c}\mathbf{A})$

Let $\mathbf{A} = \begin{pmatrix} 0 & 2 & 1 & 5 & 6 \\ 5 & 6 & 4 & 1 & 4 \\ 5 & 0 & 3 & 2 & 1 \\ 6 & 6 & 4 & 6 & 4 \\ 2 & 5 & 6 & 6 & 1 \end{pmatrix}$



Linear code equivalence

Isometry

An **isometry** (for our purposes) between two codes \mathcal{C} and \mathcal{D} is a **linear map** $\mu : \mathcal{C} \rightarrow \mathcal{D}$ that **preserves the metric**.



In our case: an isometry preserves the **Hamming weight** of codewords.

Which linear transformations preserve the Hamming weight?

→ Multiply a codeword by a **permutation** matrix **P** ?

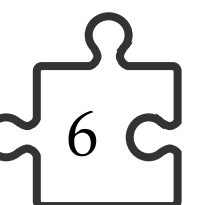
Example. $q = 7, n = 5, k = 3, \mathbf{G} = \begin{pmatrix} 2 & 0 & 4 & 0 & 1 \\ 6 & 1 & 0 & 0 & 0 \\ 3 & 4 & 1 & 3 & 0 \end{pmatrix}$

Let $\mathbf{c} = (100)\mathbf{G} = (20401)$

→ $\mathbf{cP} = (20401)\mathbf{P} = (20014)$

→ $\text{wt}(\mathbf{c}) = \text{wt}(\mathbf{cP})$

Let $\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$



Linear code equivalence

Isometry

An **isometry** (for our purposes) between two codes \mathcal{C} and \mathcal{D} is a **linear map** $\mu : \mathcal{C} \rightarrow \mathcal{D}$ that **preserves the metric**.



In our case: an isometry preserves the **Hamming weight** of codewords.

Which linear transformations preserve the Hamming weight?

→ Multiply a codeword by a **monomial** matrix **Q** ?

Example. $q = 7, n = 5, k = 3, \mathbf{G} = \begin{pmatrix} 2 & 0 & 4 & 0 & 1 \\ 6 & 1 & 0 & 0 & 0 \\ 3 & 4 & 1 & 3 & 0 \end{pmatrix}$

Let $\mathbf{c} = (100)\mathbf{G} = (20401)$

→ $\mathbf{cQ} = (20401)\mathbf{Q} = (20023)$

→ $\text{wt}(\mathbf{c}) = \text{wt}(\mathbf{cQ})$

Let $\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix}$



Linear code equivalence

Isometry

An **isometry** (for our purposes) between two codes \mathcal{C} and \mathcal{D} is a **linear map** $\mu : \mathcal{C} \rightarrow \mathcal{D}$ that **preserves the metric**.

→ In our case: an isometry preserves the **Hamming weight** of codewords.

Which linear transformations preserve the Hamming weight?

→ We can also multiply \mathbf{G} on the **left** by $\mathbf{T} \in \text{GL}_k$.

→ This is just a change of basis (because we defined the code \mathcal{C} as the **row** span of \mathbf{G}).



Linear code equivalence

Isometry

An **isometry** (for our purposes) between two codes \mathcal{C} and \mathcal{D} is a **linear map** $\mu : \mathcal{C} \rightarrow \mathcal{D}$ that **preserves the metric**.



In our case: an isometry preserves the **Hamming weight** of codewords.

Equivalent codes

Two codes \mathcal{C} and \mathcal{D} are **equivalent** if there is an isometry between them.

Linear code equivalence

The Linear Code Equivalence (LCE) problem

Input: Two generator matrices $\mathbf{G}_1, \mathbf{G}_2 \in \mathbb{F}_q^{k \times n}$ for two linear codes \mathcal{C} and \mathcal{D} .

Question: Find - if any - \mathbf{Q} , a monomial matrix, and $\mathbf{T} \in \text{GL}_k(\mathbb{F}_q)$ such that $\mathbf{G}_2 = \mathbf{T}\mathbf{G}_1\mathbf{Q}$.

Matrix (rank-metric) codes

Matrix code

A **matrix code** \mathcal{C} over \mathbb{F}_q is a k -dimensional \mathbb{F}_q -linear subspace of $\mathbb{F}_q^{m \times n}$.

Rank metric

For $\mathbf{C} \in \mathbb{F}_q^{m \times n}$, the **rank weight** of \mathbf{C} is given by the rank of \mathbf{C} , aka.

$$\text{wt}(\mathbf{C}) = \text{rk}(\mathbf{C}).$$

Basis of a matrix code

The basis of a matrix code \mathcal{C} is given by the k -tuple $(\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(k)})$.

Matrix (rank-metric) codes

Example. $q = 13$, $m = 4$, $n = 6$, $k = 5$

$$\mathbf{C} = \lambda_1 \cdot \begin{pmatrix} 2 & 8 & 10 & 4 & 5 & 7 \\ 1 & 11 & 7 & 9 & 6 & 12 \\ 3 & 0 & 13 & 5 & 4 & 8 \\ 9 & 6 & 3 & 2 & 10 & 11 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 12 & 0 & 4 & 11 & 9 & 3 \\ 5 & 6 & 8 & 13 & 2 & 1 \\ 10 & 7 & 3 & 9 & 4 & 6 \\ 2 & 5 & 11 & 8 & 1 & 10 \end{pmatrix} + \lambda_3 \cdot \begin{pmatrix} 5 & 2 & 9 & 11 & 4 & 8 \\ 3 & 7 & 1 & 10 & 12 & 0 \\ 6 & 9 & 2 & 13 & 11 & 8 \\ 1 & 5 & 6 & 3 & 10 & 7 \end{pmatrix} + \lambda_4 \cdot \begin{pmatrix} 9 & 4 & 6 & 1 & 13 & 2 \\ 8 & 0 & 5 & 12 & 6 & 11 \\ 3 & 7 & 10 & 9 & 4 & 5 \\ 2 & 8 & 11 & 3 & 7 & 1 \end{pmatrix} + \lambda_5 \cdot \begin{pmatrix} 7 & 10 & 4 & 6 & 8 & 3 \\ 1 & 5 & 2 & 11 & 9 & 0 \\ 13 & 7 & 6 & 4 & 12 & 2 \\ 8 & 3 & 1 & 9 & 5 & 10 \end{pmatrix} \quad \lambda_i \in \mathbb{F}_q$$

Matrix code equivalence

Isometry

An **isometry** (for our purposes) between two codes \mathcal{C} and \mathcal{D} is a **linear map** $\mu : \mathcal{C} \rightarrow \mathcal{D}$ that **preserves the metric**.



In this case: an isometry preserves the **rank weight** of codewords.

Which linear transformations preserve the rank?

- Multiply a codeword on the right by any $\mathbf{M} \in \mathbb{F}_q^{n \times r}$ ✗
- Multiply a codeword on the right by $\mathbf{B} \in GL_n$ ✓
- Multiply a codeword on the left by $\mathbf{A} \in GL_m$ ✓
- Take the transposition of a codeword (only when $m = n$, does not make the equivalence problem harder) ✓

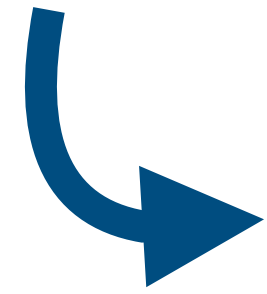
Matrix code equivalence

The Matrix Code Equivalence (MCE) problem

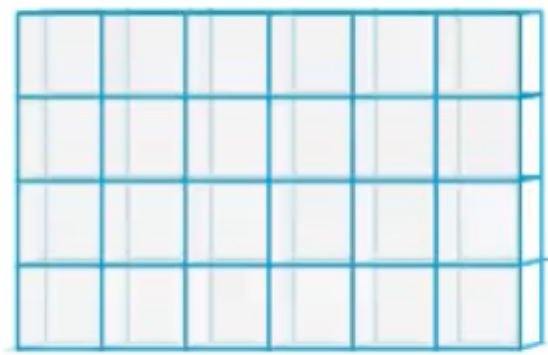
Input: Two k -dimensional matrix codes $\mathcal{C}, \mathcal{D} \subset \mathbb{F}_q^{m \times n}$ for two matrix codes \mathcal{C} and \mathcal{D} .

Question: Find - if any - a map (\mathbf{A}, \mathbf{B}) , where $\mathbf{A} \in \text{GL}_m(\mathbb{F}_q)$ and $\mathbf{B} \in \text{GL}_n(\mathbb{F}_q)$ such that for all $\mathbf{C} \in \mathcal{C}$, it holds that $\mathbf{ACB} \in \mathcal{D}$.

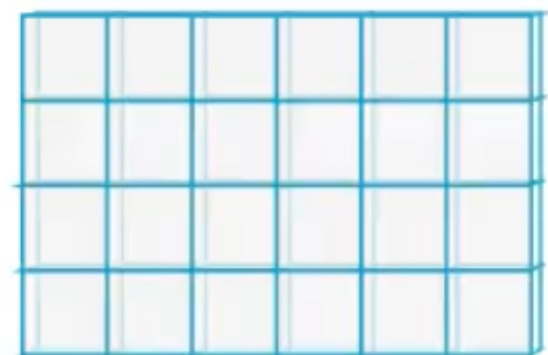
From matrix codes to 3-tensors



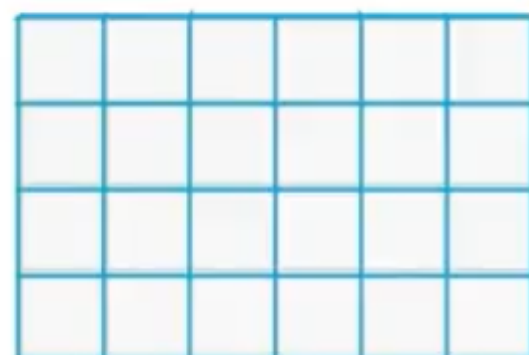
We can think of a matrix code as a 3-tensor over \mathbb{F}_q .



C_1



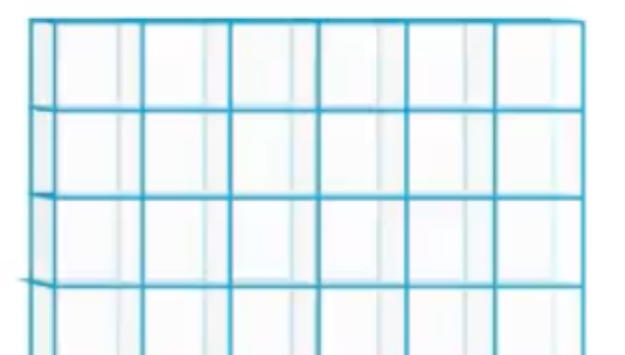
C_2



C_3

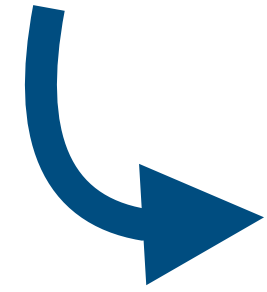


C_4



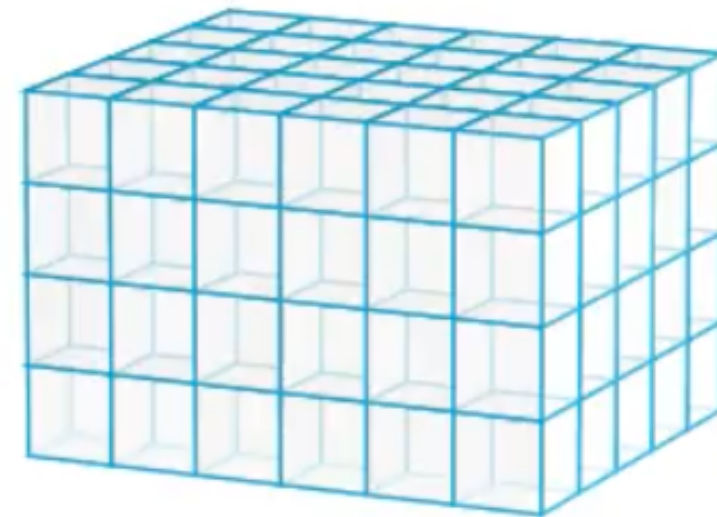
C_5

From matrix codes to 3-tensors

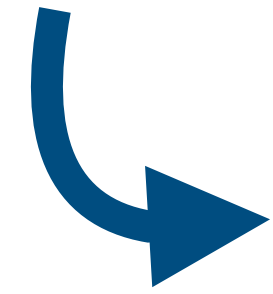


We can think of a matrix code as a 3-tensor over \mathbb{F}_q .

$$\mathcal{C} \subseteq \mathbb{F}_q^{m \times n \times k}$$

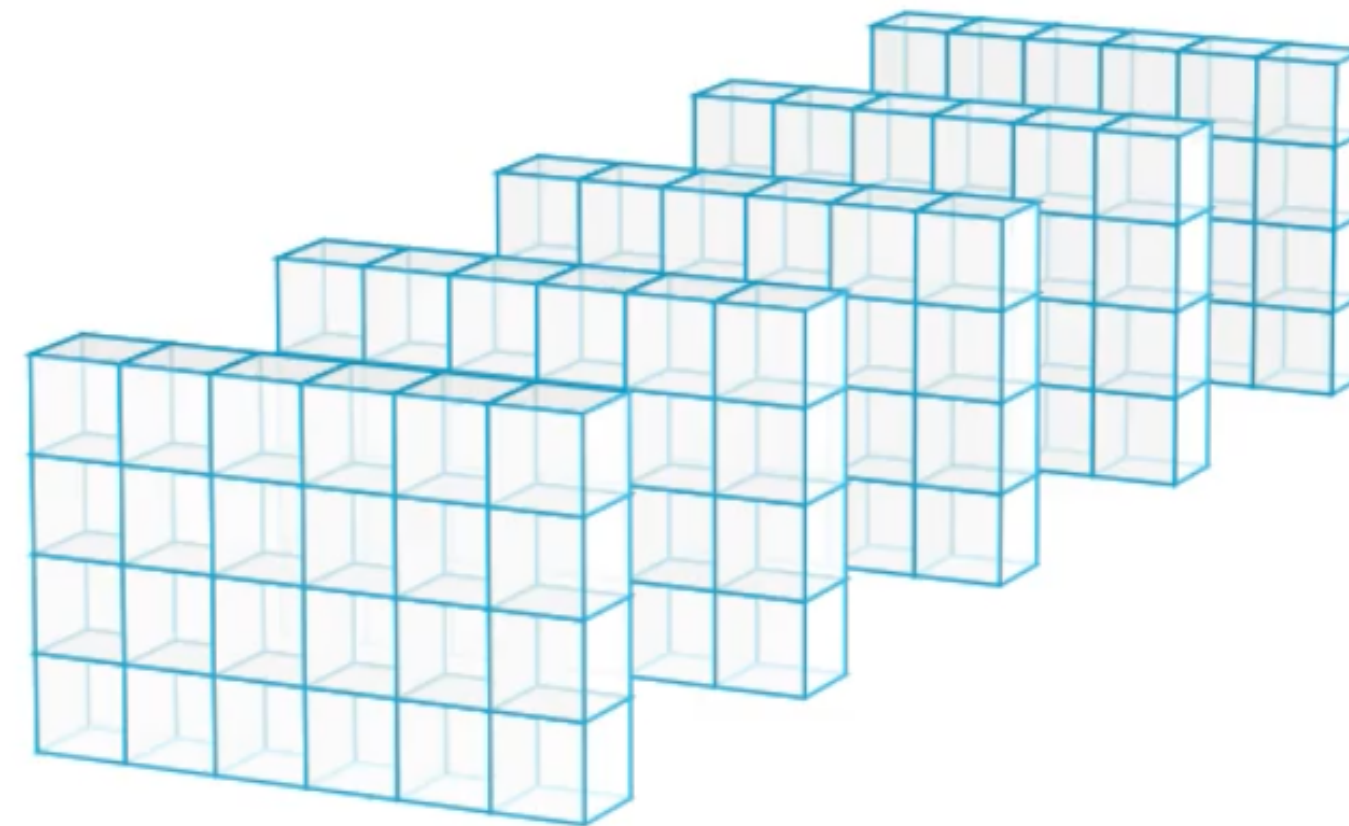


From matrix codes to 3-tensors

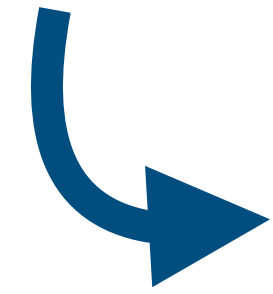


Viewed as a 3-tensor, we can see \mathcal{C} from three directions

- a k -dimensional code in $\mathbb{F}_q^{m \times n}$
- an m -dimensional code in $\mathbb{F}_q^{n \times k}$
- an n -dimensional code in $\mathbb{F}_q^{m \times k}$

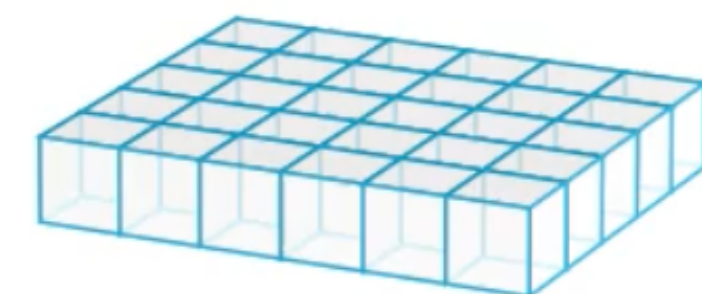
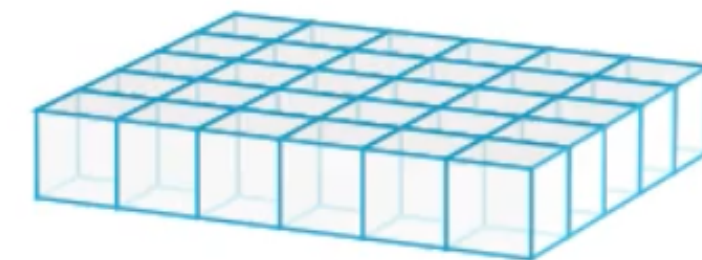
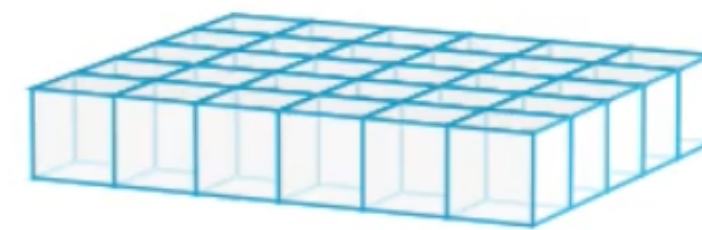
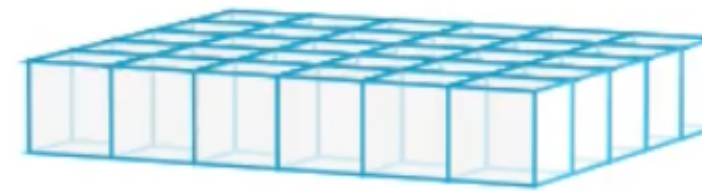


From matrix codes to 3-tensors

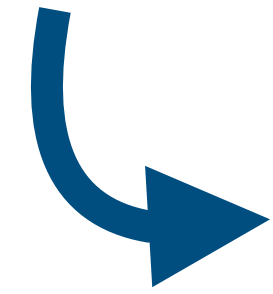


Viewed as a 3-tensor, we can see \mathcal{C} from three directions

- a k -dimensional code in $\mathbb{F}_q^{m \times n}$
- an m -dimensional code in $\mathbb{F}_q^{n \times k}$
- an n -dimensional code in $\mathbb{F}_q^{m \times k}$

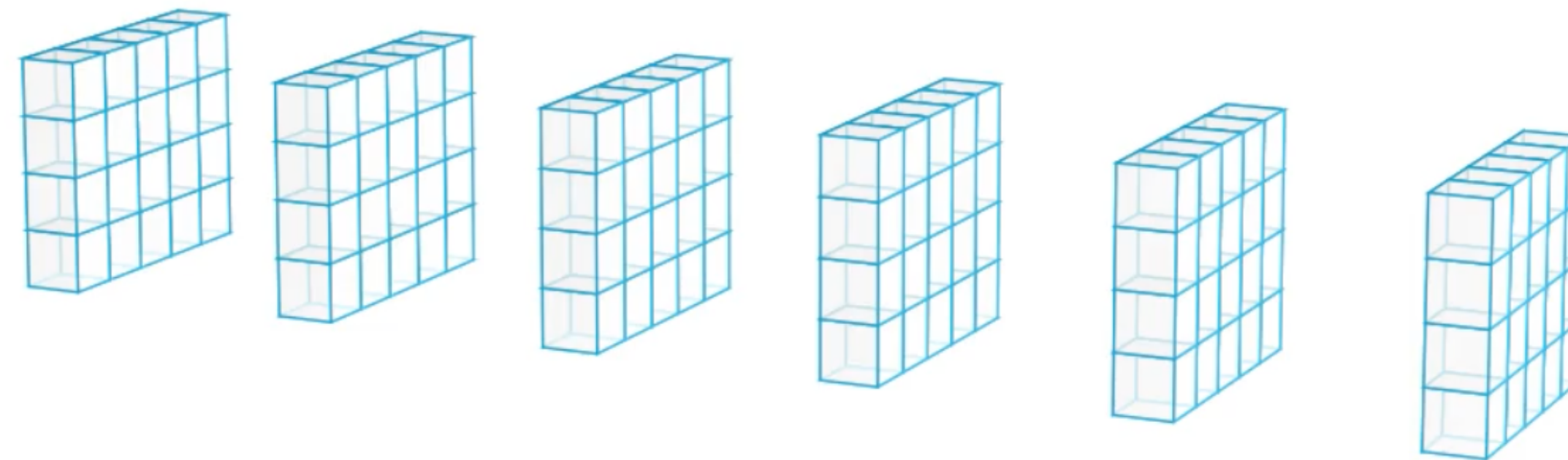


From matrix codes to 3-tensors



Viewed as a 3-tensor, we can see \mathcal{C} from three directions

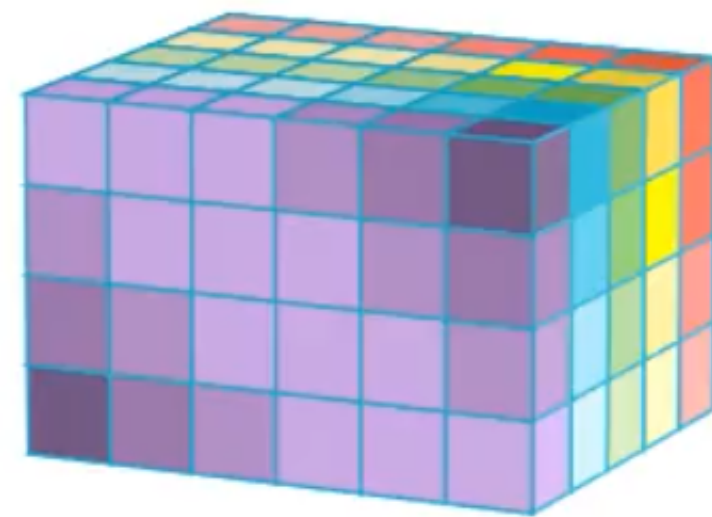
- a k -dimensional code in $\mathbb{F}_q^{m \times n}$
- an m -dimensional code in $\mathbb{F}_q^{n \times k}$
- an n -dimensional code in $\mathbb{F}_q^{m \times k}$



Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

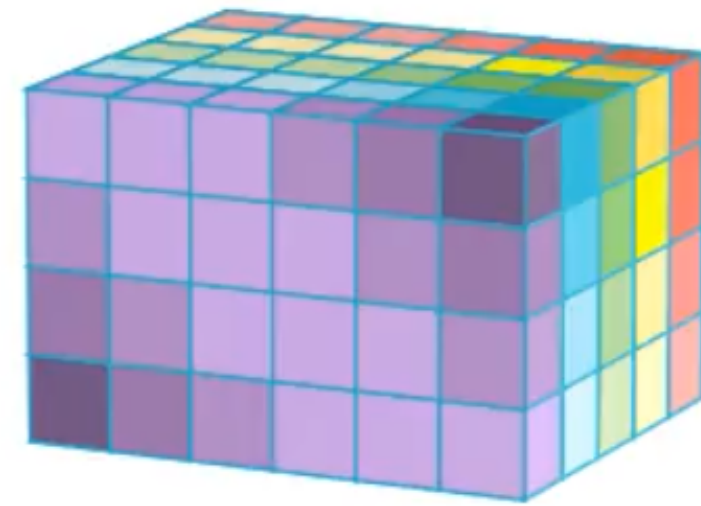
$$\mathcal{C} \subseteq \mathbb{F}_q^{m \times n \times k}$$



Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathbf{T} \in \text{GL}_k(q)$$



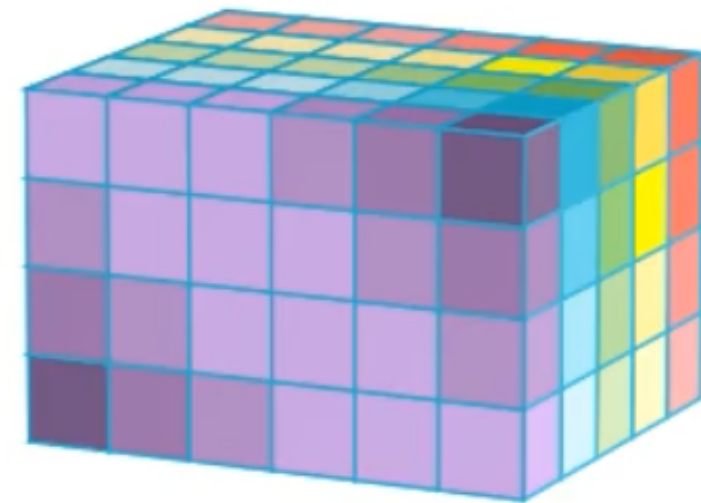
$$\mathbf{A} \in \text{GL}_m(q)$$

$$\mathbf{B} \in \text{GL}_n(q)$$

Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathbf{T} \in \text{GL}_k(q)$$



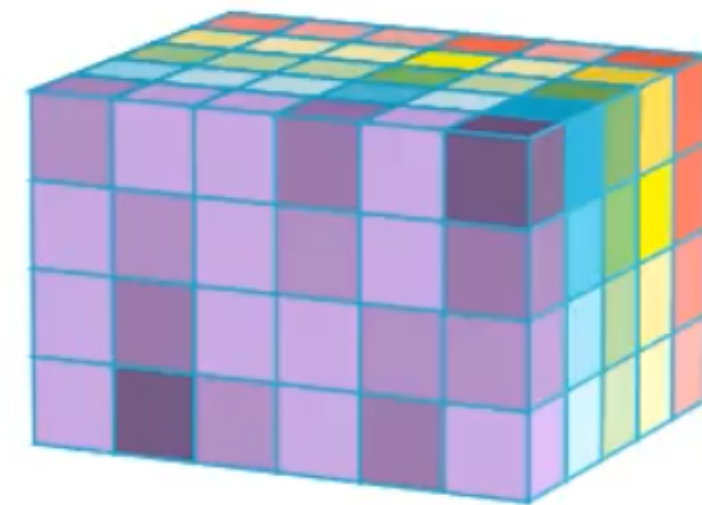
$$\mathbf{A} \in \text{GL}_m(q)$$

$$\mathbf{B} \in \text{GL}_n(q)$$

Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathbf{T} \in \text{GL}_k(q)$$



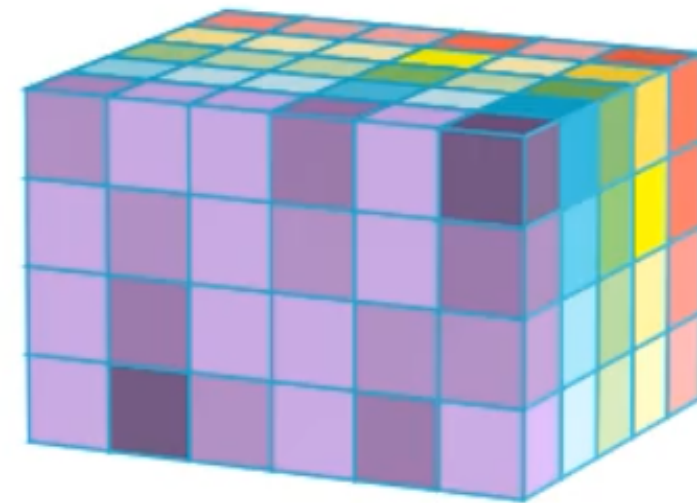
$$\mathbf{A} \in \text{GL}_m(q)$$

$$\mathbf{B} \in \text{GL}_n(q)$$

Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathbf{T} \in \text{GL}_k(q)$$



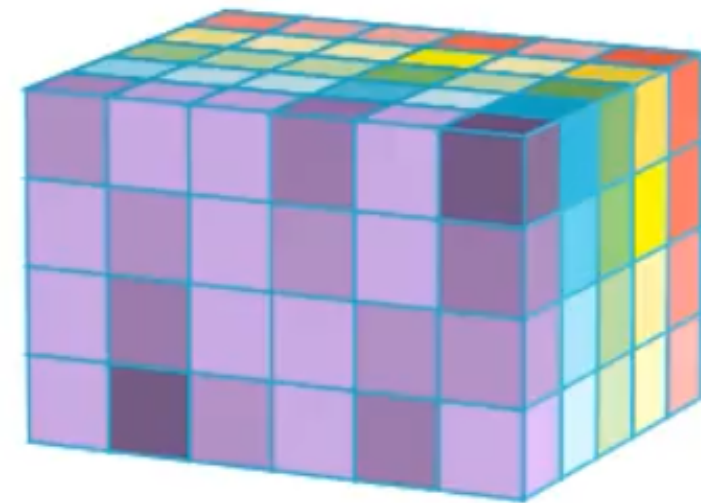
$$\mathbf{A} \in \text{GL}_m(q)$$

$$\mathbf{B} \in \text{GL}_n(q)$$

Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathbf{T} \in \text{GL}_k(q)$$



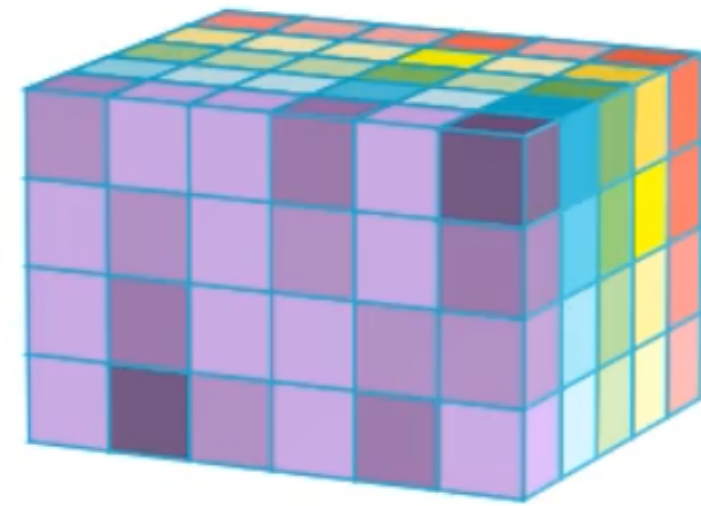
$$\mathbf{A} \in \text{GL}_m(q)$$

$$\mathbf{B} \in \text{GL}_n(q)$$

Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathbf{T} \in \text{GL}_k(q)$$



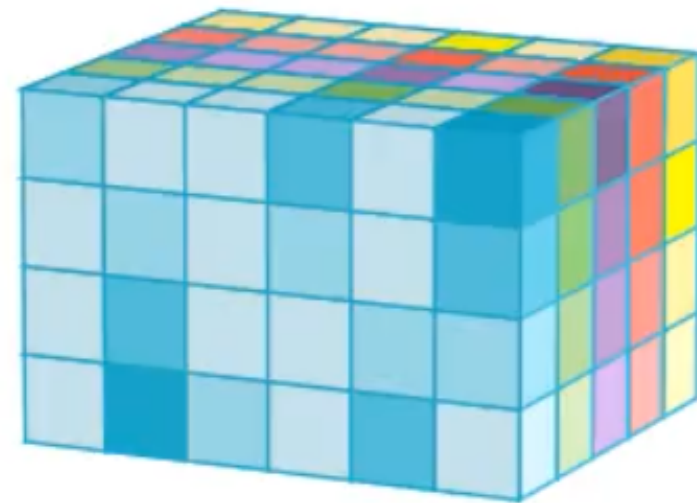
$$\mathbf{A} \in \text{GL}_m(q)$$

$$\mathbf{B} \in \text{GL}_n(q)$$

Tensor isomorphism

↪ The equivalence then becomes **tensor isomorphism**.

$$\mathcal{D} \subseteq \mathbb{F}_q^{m \times n \times k}$$



The slide features a solid red background. In the corners, there are decorative wireframe cubes. The top-left and top-right corners contain large, light-red wireframe cubes. The bottom-left and bottom-right corners contain smaller, light-red wireframe cubes. In the center of the slide, the title "Cryptanalysis" is written in a large, white, serif font. Below it, the subtitle "(The MCE case)" is written in a smaller, white, serif font.

Cryptanalysis

(The MCE case)

The slide features a solid red background. In the four corners, there are decorative wireframe cubes. The top-left and top-right cubes are large and partially cut off by the edges. The bottom-left and bottom-right cubes are also large and partially cut off. In the center of the slide, there are three smaller wireframe cubes arranged in a horizontal line, with the middle one being the smallest and the two on either side being slightly larger.

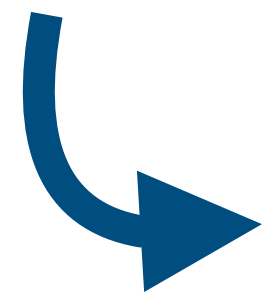
Algebraic attack

Algebraic attack

The MCE problem in matrix form

Let $(\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(k)})$ be a basis of code \mathcal{C} and let $(\mathbf{D}^{(1)}, \dots, \mathbf{D}^{(k)})$ be a basis of code \mathcal{D} . Find $\mathbf{A} \in \text{GL}_m(\mathbb{F}_q)$, $\mathbf{B} \in \text{GL}_n(\mathbb{F}_q)$ and $\mathbf{T} \in \text{GL}_k(\mathbb{F}_q)$ such that

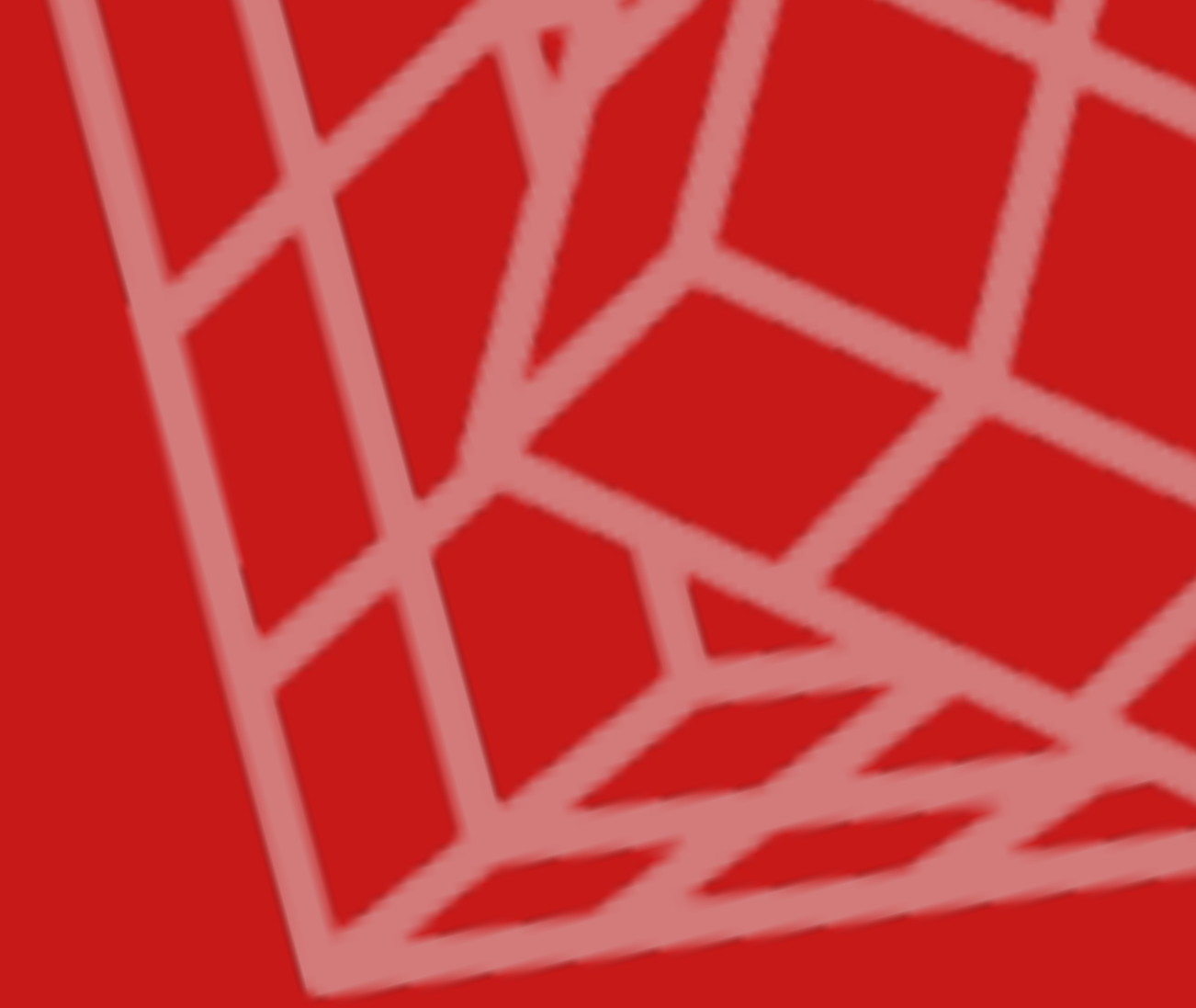
$$\mathbf{D}^{(i)} = \sum_{1 \leq j \leq k} t_{j,i} \mathbf{A} \mathbf{C}^{(j)} \mathbf{B}, \quad \forall 1 \leq i \leq k$$



Alternatively, this gives a better modelisation:

$$\sum_{1 \leq j \leq k} t_{j,i} \mathbf{D}^{(j)} = \mathbf{A} \mathbf{C}^{(i)} \mathbf{B}, \quad \forall 1 \leq i \leq k$$

Combinatorial attack



Collision


 We have a collision when we know a codeword \mathbf{C} in \mathcal{C} that maps to a codeword \mathbf{D} in \mathcal{D} .

 We can then infer linear constraints from

$$\mathbf{A}^{-1}\mathbf{D} = \mathbf{CB}$$

If we add these linear constraints to the system obtained from the algebraic attack, we can efficiently solve the system of equations and recover the isometry (the resolution being efficient - close to polynomial - is an empirical result, not yet proven).

Collision

 With two collisions, we get the following system

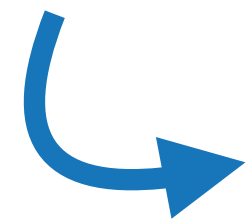
$$\begin{aligned}\mathbf{A}^{-1}\mathbf{D}_1 &= \mathbf{C}_1\mathbf{B} \\ \mathbf{A}^{-1}\mathbf{D}_2 &= \mathbf{C}_2\mathbf{B}\end{aligned}$$

- Results in a **linear** system with the same number of variables and equations.
- If $\mathbf{C}_1, \mathbf{C}_2, \mathbf{D}_1, \mathbf{D}_2$ are all full rank, we should have a unique solution.
- We can easily recover \mathbf{A} from \mathbf{A}^{-1} .

The birthday paradox

What is the probability that, in a set of N randomly chosen people, at least two will share a birthday?

How big should N be to get a probability of 50% ?

 $N = 23$

- The **birthday problem** in collision search algorithms - - - - -
We draw, randomly, elements from a set of size N .
How many times do we **expect** to draw an element before we get the same element twice?

The birthday paradox

- The **birthday problem** in collision search algorithms

We draw, randomly, elements from a set of size N .
How many times do we **expect** to draw an element before we get the same element twice?

- Probability that there is no collision when the first element is drawn: 1
- Probability that there is no collision when the second element is drawn: $1 - \frac{1}{N}$
- Probability that there is no collision when the third element is drawn: $1 - \frac{2}{N}$
- ...

The birthday paradox

- The **birthday problem** in collision search algorithms

We draw, randomly, elements from a set of size N .

How many times do we **expect** to draw an element before we get the same element twice?

$$P(X > T) = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{T-1}{N}\right)$$

Using a first-order Taylor approximation $e^x \approx 1 + x$, this simplifies to

$$\begin{aligned} P(X > T) &\approx e^{-\frac{1}{N}} \cdot e^{-\frac{2}{N}} \cdot \dots \cdot e^{-\frac{T-1}{N}} \approx \\ &\approx e^{-(1+2+\dots+(T-1))/N} \approx \\ &\approx e^{-\frac{T(T-1)}{2N}} \end{aligned}$$



For $P(X > T) \approx 63\%$, we get $T \approx 1.41\sqrt{N}$.

General collision attack

Algorithm 1 General Birthday-based Equivalence Finder

```
1: function SAMPLESET( $S, \mathbb{P}, \ell$ )
2:    $L \leftarrow \emptyset$ 
3:   repeat
4:      $a \xleftarrow{\$} S$ 
5:     if  $\mathbb{P}(a)$  then  $L \leftarrow L \cup \{a\}$ 
6:     end if
7:   until  $|L| = \ell$ 
8:   return  $L$ 
9: end function

10: function COLLISIONFIND( $S_1, S_2$ )
11:    $L_1 \leftarrow \text{SAMPLESET}(S_1, \mathbb{P}, \ell)$ 
12:    $L_2 \leftarrow \text{SAMPLESET}(S_2, \mathbb{P}, \ell)$ 
13:   for all  $(a, b) \in L_1 \times L_2$  do
14:      $\phi \leftarrow \text{FINDFUNCTION}(a, b)$ 
15:     if  $\phi \neq \perp$  then
16:       return solution  $\phi$ 
17:     end if
18:   end for
19:   return  $\perp$ 
20: end function
```

Collision attack : complexity

↳ Depends on the **choice of the predicate \mathbb{P}** . The choice is made such that we obtain the optimal **balance** between the two parts of the algorithm, aka. they take approximately the same time (whenever possible).

↳ We will get an intuition for the complexity with an exercise in the assignment.

The slide features a solid red background. In the corners, there are decorative wireframe cubes. The top-left and top-right corners contain large, light-red wireframe cubes. The bottom-left and bottom-right corners contain smaller, light-red wireframe cubes. In the center of the slide, the text "Digital signatures from equivalence problems" is written in a white, serif font, arranged in three lines.

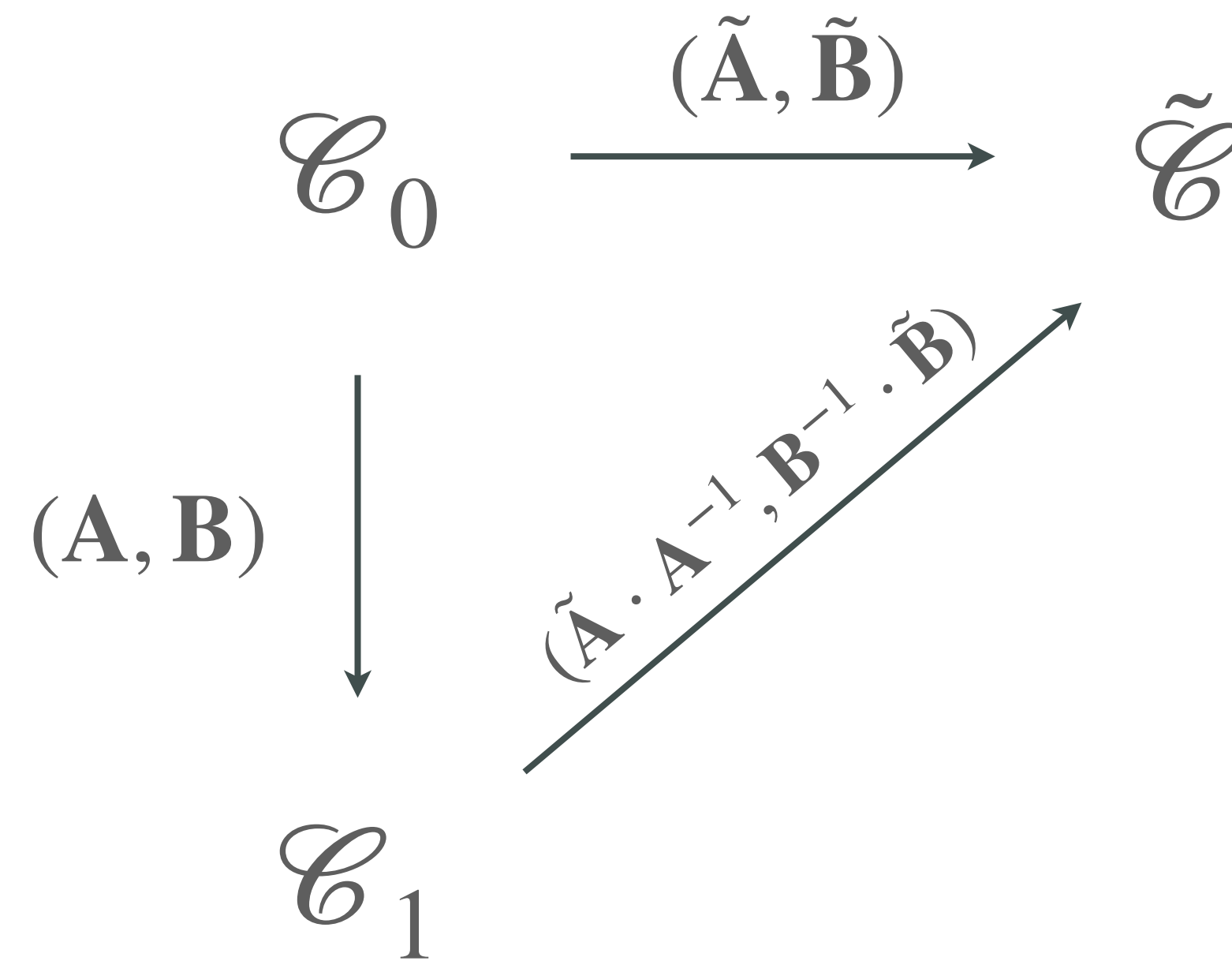
Digital signatures
from equivalence
problems

Zero-knowledge proof of knowledge

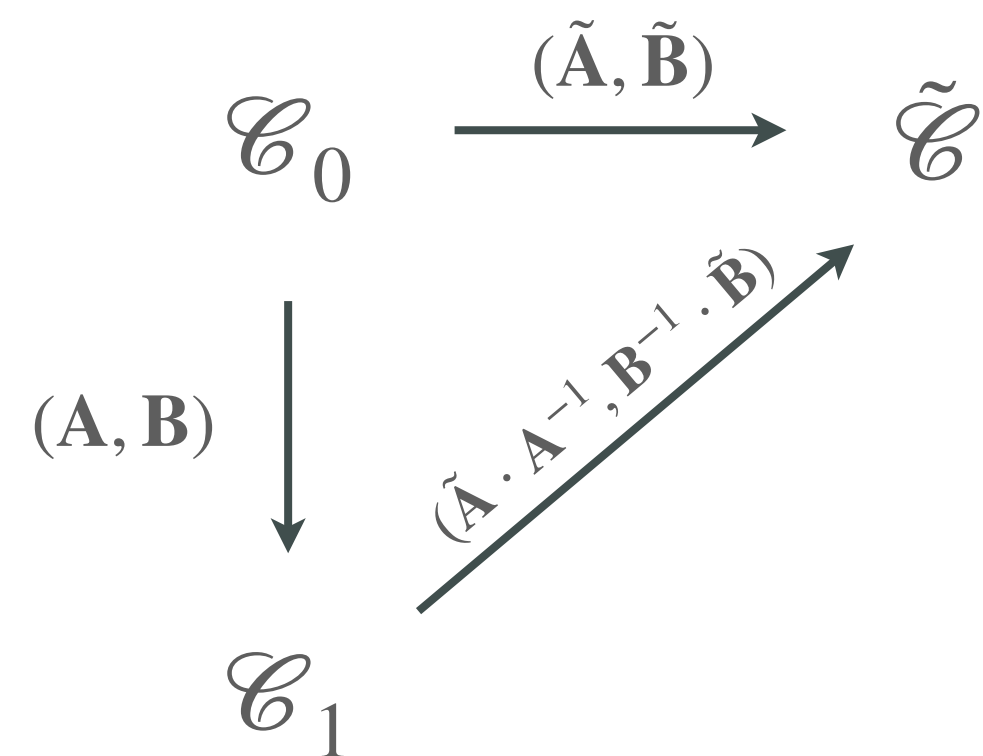


The **prover** needs to prove to the **verifier** that they know a secret, without revealing the secret or anything about the secret.

ZK identification scheme



ZK identification scheme



Prover



A

(A, B)

Commit to ephemeral code $\tilde{\mathcal{C}}$



Verifier

Pick a challenge $b \in \{0,1\}$

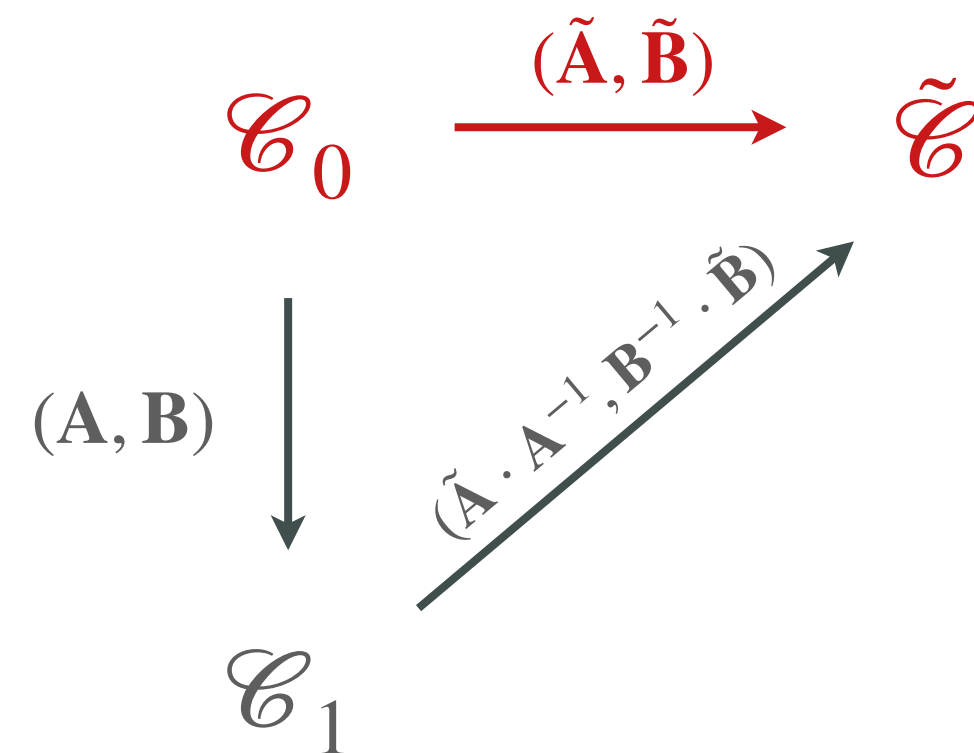


A

\mathcal{C}_0 \mathcal{C}_1

Response

ZK identification scheme



Prover



A

(A, B)

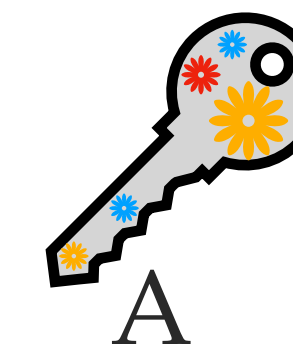
Commit to ephemeral code $\tilde{\mathcal{C}}$



Verifier

Pick a challenge $b \in \{0,1\}$

$b = 0$



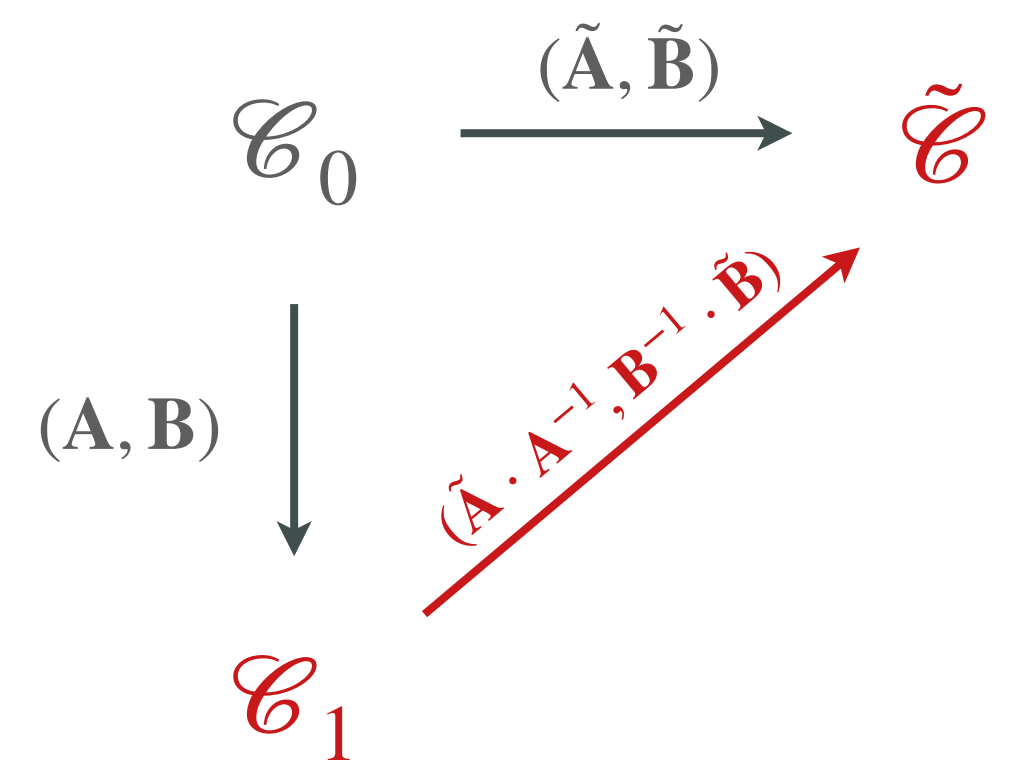
A

$\mathcal{C}_0 \quad \mathcal{C}_1$

Response

(\tilde{A}, \tilde{B})

ZK identification scheme



Prover



A

(A, B)



Verifier



A

\mathcal{C}_0 \mathcal{C}_1

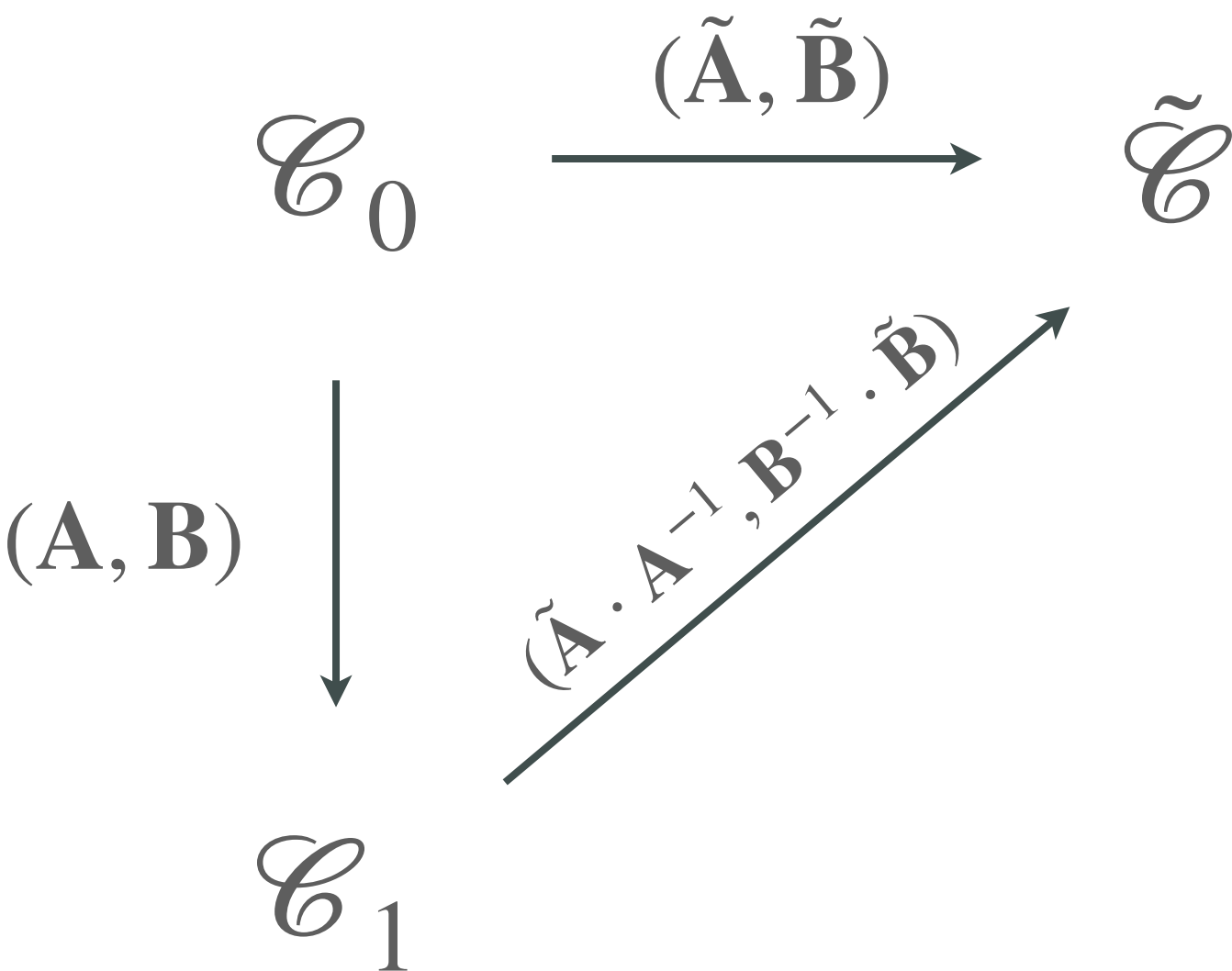
Commit to ephemeral code $\tilde{\mathcal{C}}$

Pick a challenge $b \in \{0,1\}$
 $b = 1$

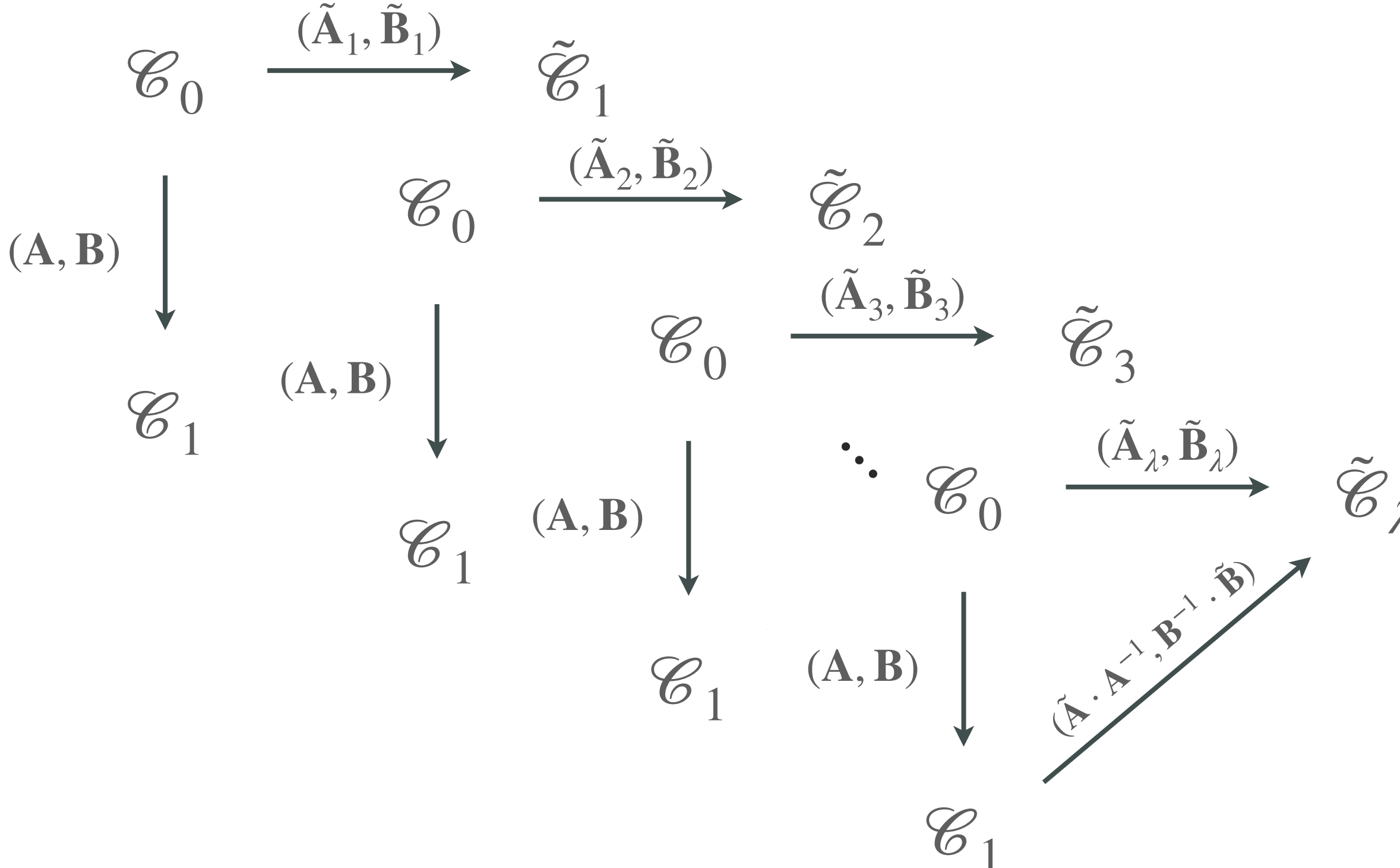
Response
 $(\tilde{\mathbf{A}} \cdot \mathbf{A}^{-1}, \tilde{\mathbf{B}} \cdot \mathbf{B}^{-1})$

ZK identification scheme

→ To get a security level of 2^λ



→ Repeat λ times



Properties : completeness



If the statement is true, an **honest prover** is always able to convince an **honest verifier**.

Properties : soundness



A **dishonest prover** cannot convince an honest verifier other than with a **small probability**.

2-Special soundness

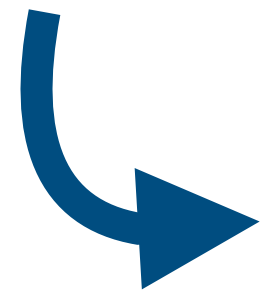
Having obtained two valid transcripts with the **same commitment** and a **different challenge**, we can extract a solution for the underlying problem.

Properties : zero-knowledge

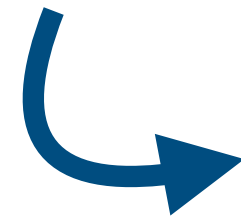


Anyone observing the transcript (including the verifier) **learns nothing** other than the fact that the statement is true.

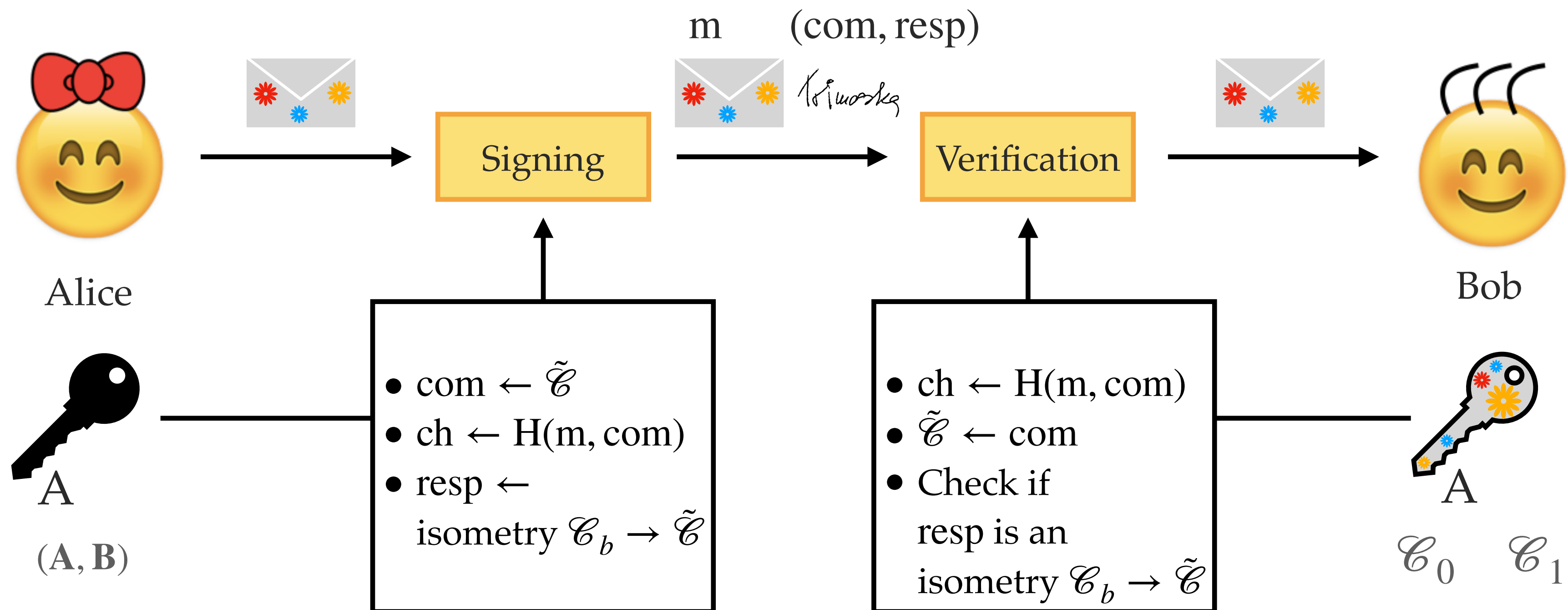
The Fiat-Shamir transform



The goal is to transform an **interactive** identification scheme into a digital signature scheme.

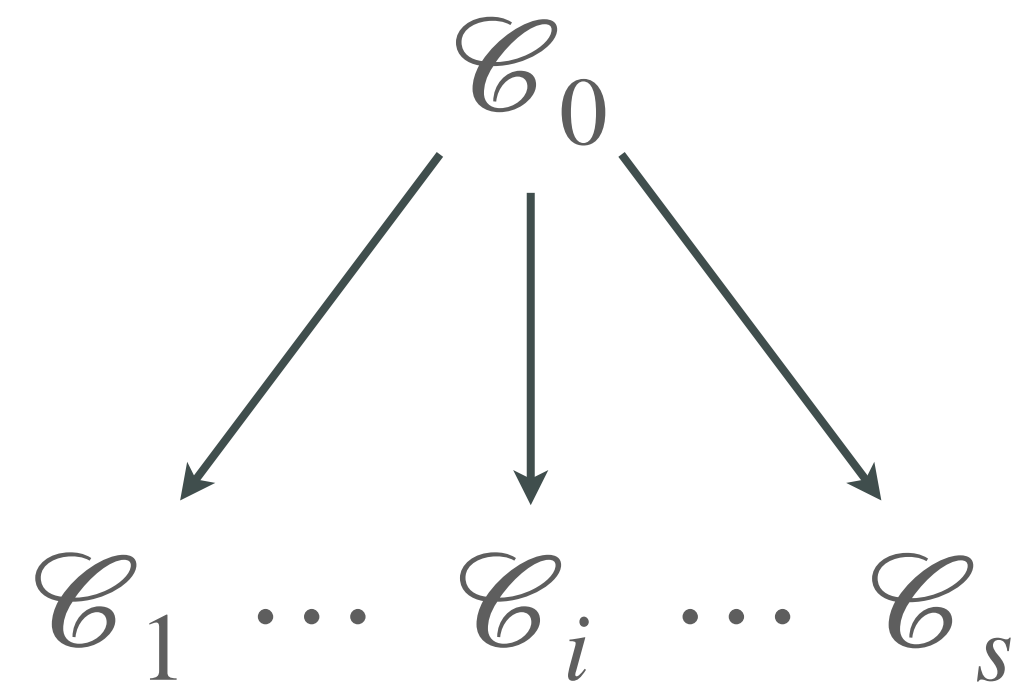


Instead of the prover choosing a challenge, the challenge is determined by the hash of the message and commitments.



Soundness amplification

Multiple public keys



- Provide s public keys
- Challenge is $b \in \{0, \dots, s\}$
- Response is an isometry $\mathcal{C}_b \rightarrow \tilde{\mathcal{C}}$

Signatures from equivalence problems

Equivalence-based digital signature schemes in the NIST competition (and elsewhere):

LESS

Linear code equivalence

MEDS

Matrix code equivalence

ALTEQ

Alternating trilinear form equivalence

Patarin's signature scheme: Isomorphism of polynomials (seen in previous lecture)

SeaSign, SQISign: Isogeny between elliptic curves

...