IH2010

Tardos's Fingerprinting Code over AWGN Channel

Minoru Kuribayashi

Graduate School of Engineering
Kobe University, Japan

E-mail: kminoru@kobe-u.ac.jp

Contents

- Background
- Attack Model
- Hard & Soft Decision Method
- Estimation of Channel
- Proposed Tracing Algorithm
- Experimental Results
- Concluding Remarks

Background

A server distributes personal copies of a content to N users.

c colluders mix their copies to forge a pirated copy.

The objective of a fingerprinting code

Identification of the colluders from the pirated copy.

c−secure code:

If the number of colluders is equal or less than c,

at least one of them can be identified.

Boneh Shaw (1995) concatenation code

Tardos (2003) Probabilistic fingerprinting code

Marking Assumption

Codeword of j-th user $\boldsymbol{x_j} = \{x_{j,1}, x_{j,2}, \ldots, x_{j,L}\}$

$x_{j,i} \in \{0, 1\}$ L: code length

When 4 users are colluded, they can produce a pirated copy under the following condition.

$$\boldsymbol{x_1} = \{0\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\}$$
$$\boldsymbol{x_2} = \{0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\}$$
$$\boldsymbol{x_3} = \{1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\}$$
$$\boldsymbol{x_4} = \{0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 1\}$$

$$\Downarrow$$

$$\boldsymbol{y} = \{y_1, \quad y_2, \ldots, \quad y_L\}$$

- If i-th elements are equal,

  undetectable position

- otherwise

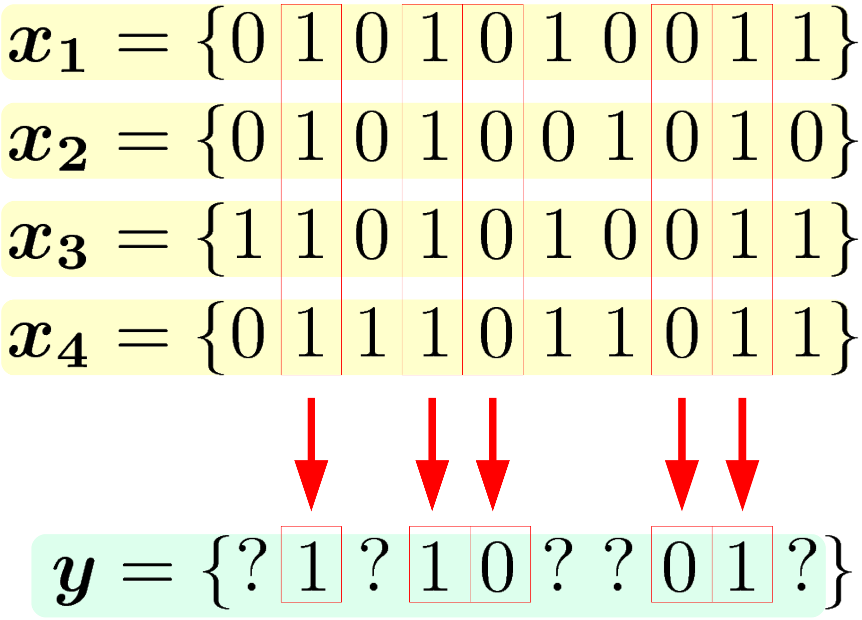
  detectable position

Marking Assumption

Codeword of j-th user $\boldsymbol{x_j} = \{x_{j,1}, x_{j,2}, \ldots, x_{j,L}\}$

$x_{j,i} \in \{0, 1\}$ L: code length

When 4 users are colluded, they can produce a pirated copy under the following condition.



$$\boldsymbol{x_1} = \{0\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\}$$

$$\boldsymbol{x_2} = \{0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\}$$

$$\boldsymbol{x_3} = \{1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\}$$

$$\boldsymbol{x_4} = \{0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 1\}$$

$$\boldsymbol{y} = \{?\ \boxed{1}\ ?\ \boxed{1}\ \boxed{0}\ ?\ ?\ \boxed{0}\ \boxed{1}\ ?\}$$

- If i-th elements are equal,

  undetectable position

- otherwise

  detectable position

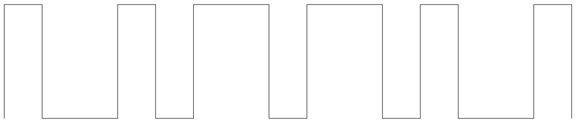
Real Envirnment

Each bit of the codeword is embedded into digital content assisted by a watermarking method.

Even if the watermarking method is robust,
the extracted signal must be distorted by a noise.

Assumption additive white Gaussian noise (AWGN)

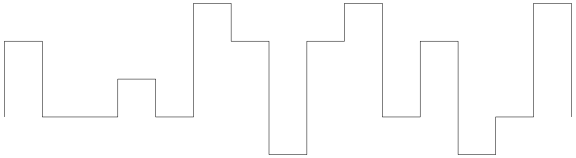
(marking assumption)



y

Optimal Detector

(relaxed assumption)



$y' = y + e$

???

How to design a good detector ?

Tardos Code

Codeword of j-th user $\boldsymbol{x_j} = \{x_{j,1}, x_{j,2}, \ldots, x_{j,L}\}$

$x_{j,i} \in \{0,1\}$ L: code length

$x_{j,i}, (1 \leq j \leq L)$ are independent.

$\Pr(x_{j,i} = 1) = p_i \quad \forall j \in [n]$

i.i.d. random variables

$p_i \in [0,1]$

$p_i \sim f(p)$

Tracing Algorithm $\boldsymbol{y} = \{y_1, y_2, \ldots, y_L\}$ a pirated codeword

If the score S_j exceeds a threshold Z, j-th user is judged guilty.

$$S_j = \sum_{i=1}^{L} y_i U_{j,i}$$

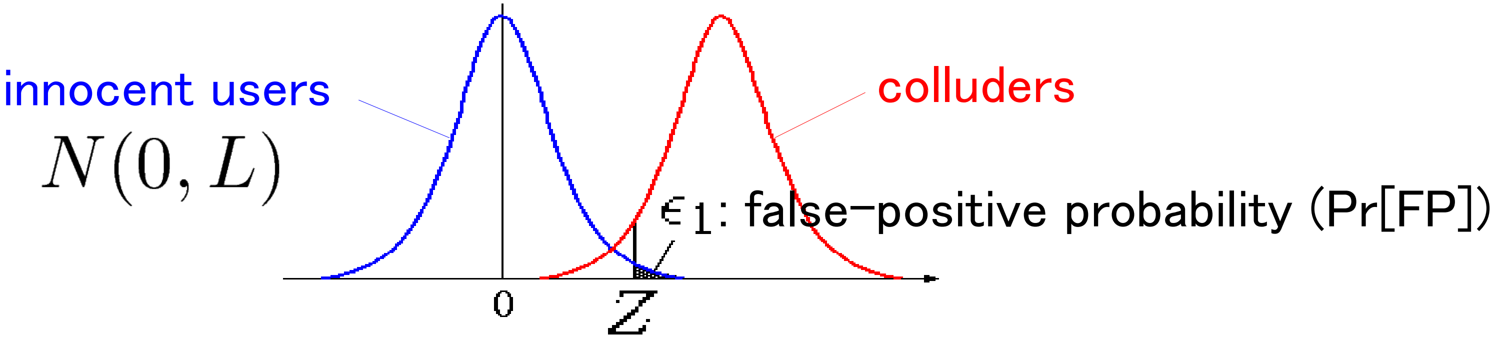
where $U_{j,i} = \begin{cases} \sqrt{\frac{1-p_i}{p_i}} & \text{if } x_{j,i} = 1 \\ -\sqrt{\frac{p_i}{1-p_i}} & \text{if } x_{j,i} = 0 \end{cases}$

Correlation Score

- Symmetric version of the score

$$S_j = \sum_{i=1}^{L} y_i U_{j,i} \longrightarrow S_j = \sum_{i=1}^{L} (2y_i - 1) U_{j,i}$$

- Gaussian approximation of the score's distribution based on CLT



- Design of the threshold Z for a given ϵ_1

$$Z = \sqrt{2L} \cdot \mathrm{erfc}^{-1}\left(\frac{2\epsilon_1}{N}\right)$$

N : number of users

Correlation Score

- Symmetric version of the score

$$S_j = \sum_{i=1}^{L} y_i U_{j,i} \longrightarrow \boxed{S_j = \sum_{i=1}^{L} (2y_i - 1) U_{j,i}}$$

- Gaussian

CLT

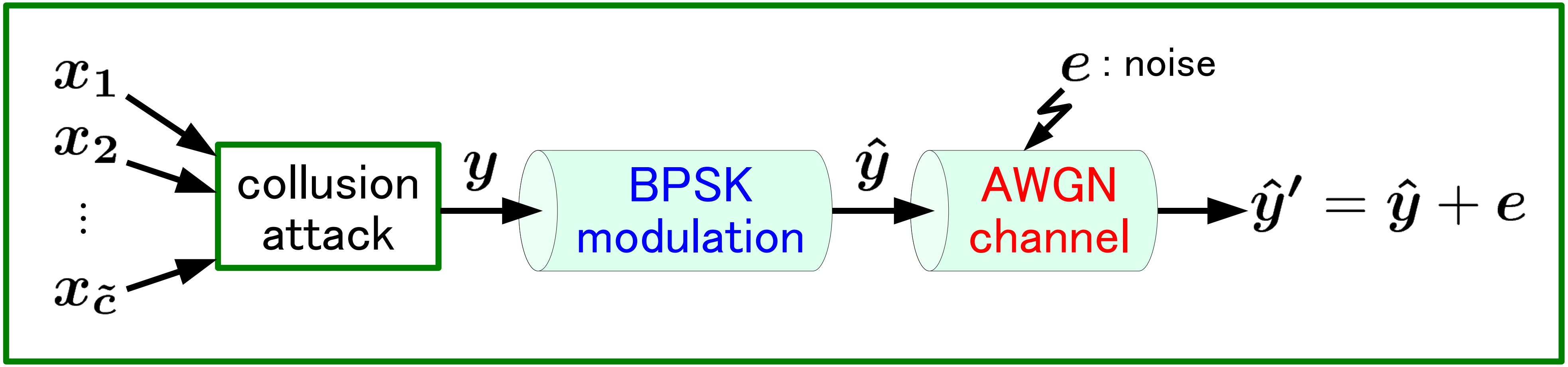
This means **BPSK** (binary phase shift keying) modulation.

$$\underset{\{0,1\}}{y_i} \longrightarrow \underset{\{-1,1\}}{\hat{y}_i = 2y_i - 1}$$

$$\boxed{S_j = \sum_{i=1}^{L} \hat{y}_i U_{j,i}}$$

- De

Attack Model



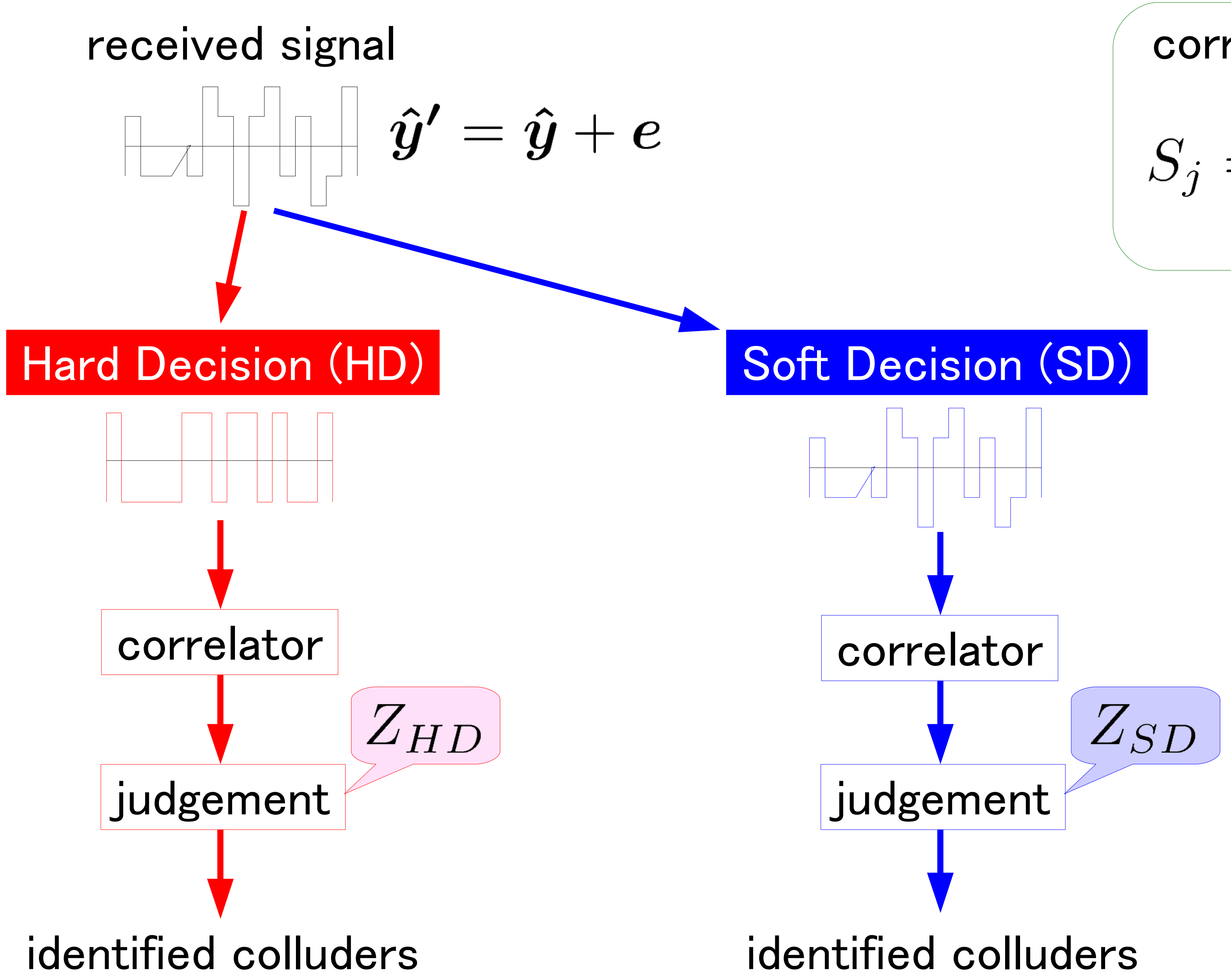
- The number $\tilde{c}$ of colluders is not always equal and less than c.
- The length L of codeword is fixed.

Our Goal Under a fixed length L, we want to catch as many colluders as possible without increasing the false-positive probability ϵ_1 no matter how many colluders are involved in.

Hard & Soft Decision Method

correlation score

$$S_j = \sum_{i=1}^{L} \hat{y}_i U_{j,i}$$



received signal

$$\hat{\boldsymbol{y}}' = \hat{\boldsymbol{y}} + \boldsymbol{e}$$

Hard Decision (HD)

correlator

judgement

Z_{HD}

identified colluders

Soft Decision (SD)

correlator

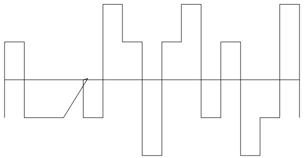
judgement

Z_{SD}

identified colluders

Hard & Soft Decision Method

received signal



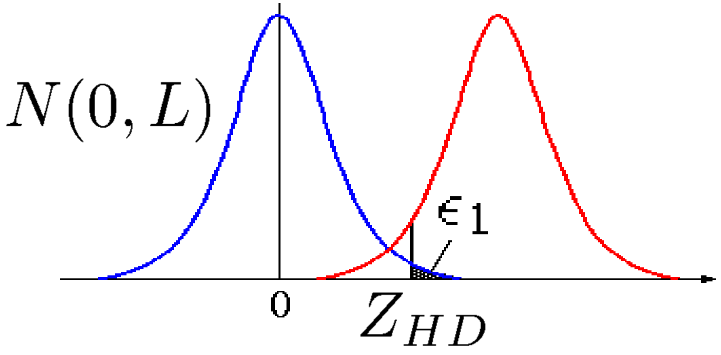
$$\hat{\boldsymbol{y}}' = \hat{\boldsymbol{y}} + \boldsymbol{e}$$

correlation score

$$S_j = \sum_{i=1}^{L} \hat{y}_i U_{j,i}$$

Hard Decision (HD)

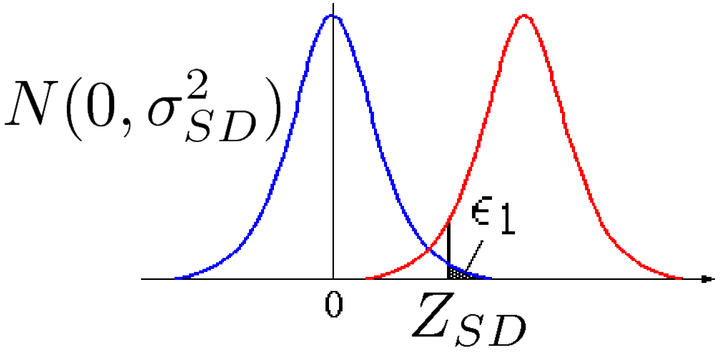
Based on the CLT, Z_{HD} is calculated by a given ϵ_1.



$$Z_{HD} = \sqrt{2L} \cdot \mathrm{erfc}^{-1}\left(\frac{2\epsilon_1}{N}\right)$$

Soft Decision (SD)

Based on the CLT, Z_{SD} is calculated by a given ϵ_1.



$$Z_{SD} = \sqrt{2\sigma_{SD}^2} \cdot \mathrm{erfc}^{-1}\left(\frac{2\epsilon_1}{N}\right)$$

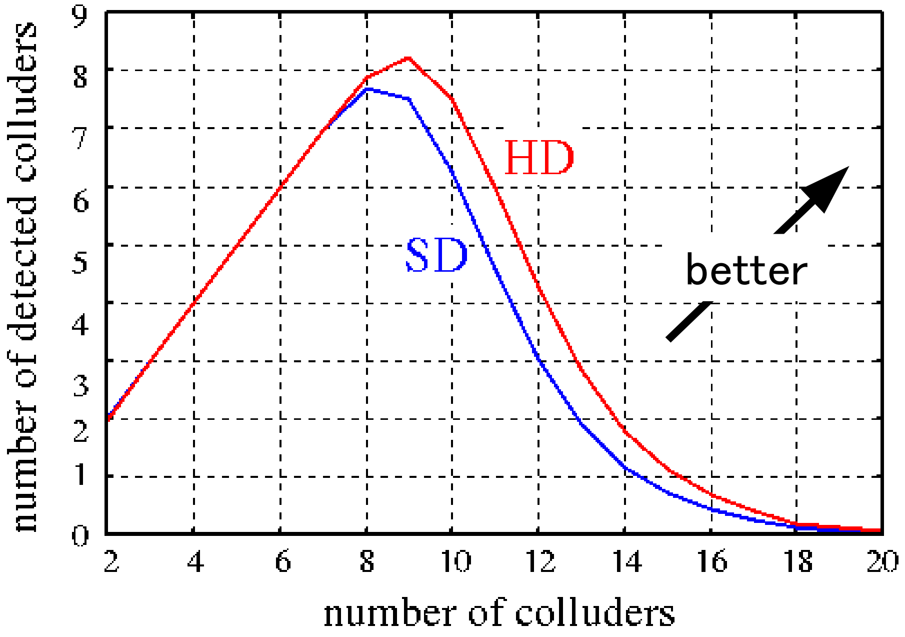
Comparison (SNR: 8dB)

code length : $L = 10^4$

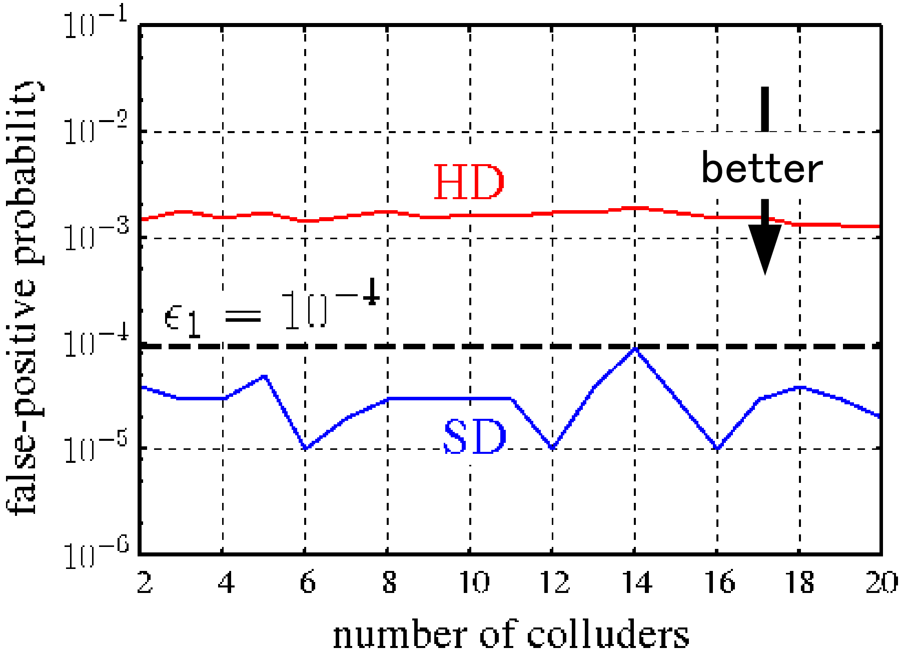
Pr[FP]: $\epsilon_1 = 10^{-4}$

# of users : $N = 10^4$

# of detected colluders



False positive probability



Although the HD method catches more colluders,
the false-positive probability is increased.

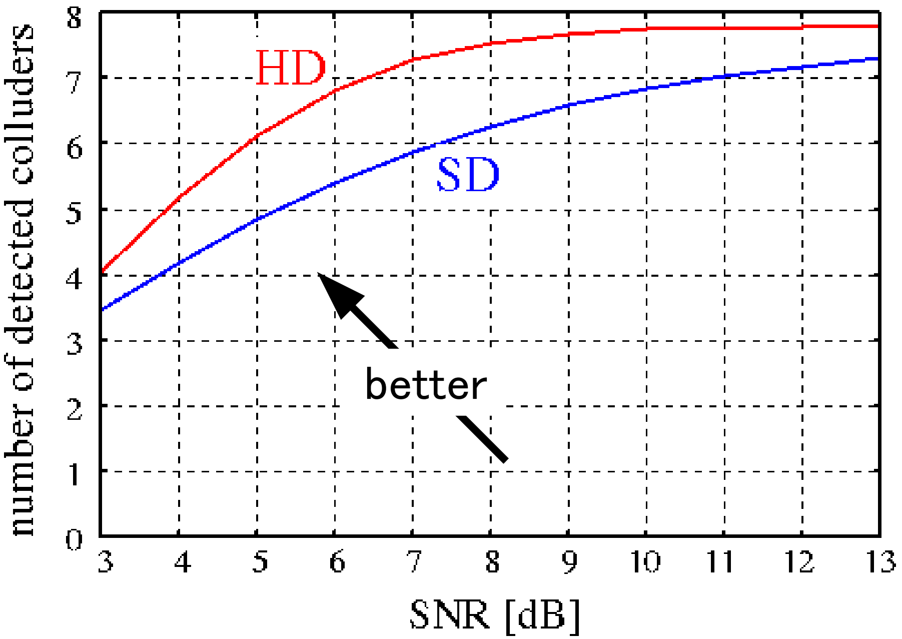
Comparison ($\tilde{c}$ = 10)

code length : $L = 10^4$

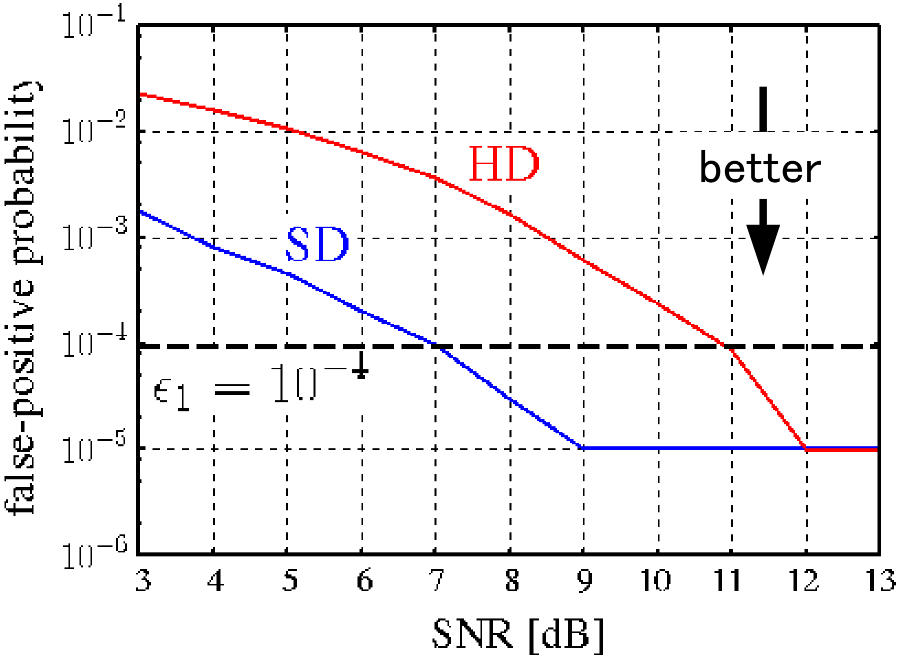
Pr[FP]: $\epsilon_1 = 10^{-4}$

# of users : $N = 10^4$

# of detected colluders



False positive probability

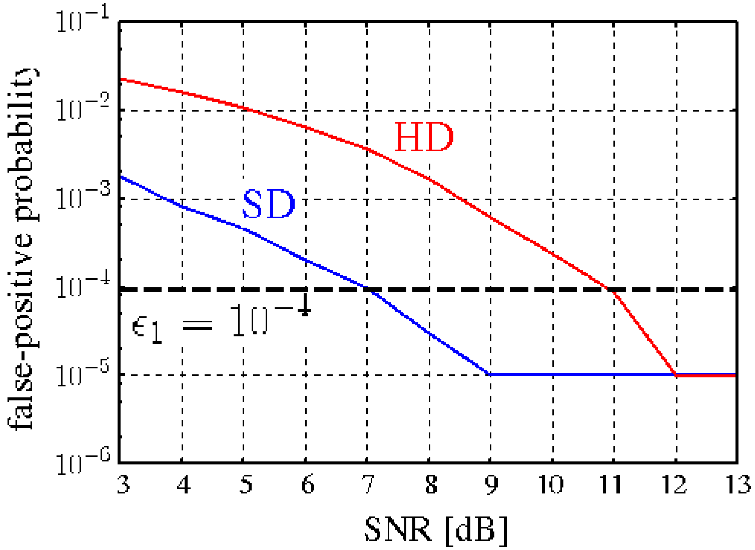


When the amount of noise is very small,
the use of CLT seems to be valid for both methods.

Consideration

With the increase of noise, the false positive probability is increased.

False positive probability



Remark

In this experiment, the threshold Z_{HD} is calculated by a given ϵ_1 based on the CLT.

$$Z_{HD} = \sqrt{2L} \cdot \mathrm{erfc}^{-1}\left(\frac{2\epsilon_1}{N}\right)$$

$$= 3.97\sqrt{2L}$$ (constant)

$$\epsilon_1 = 10^{-4}, \ N = 10^4$$

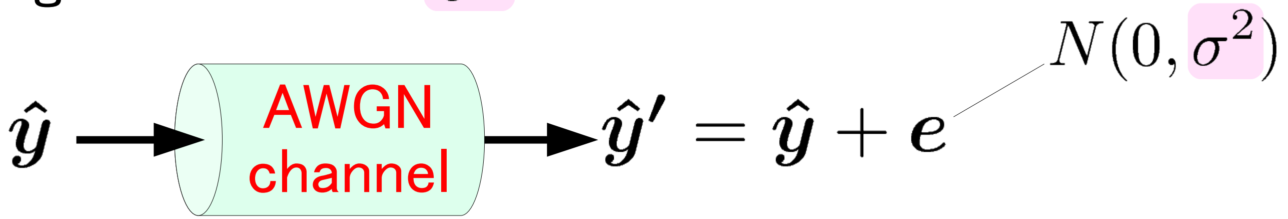
In the related works,

$$Z \sim Z_{HD}$$ (constant)

The false positive probability must be increased in the same fashion.

Proposed Method

- Estimating the variance σ^2 of noise



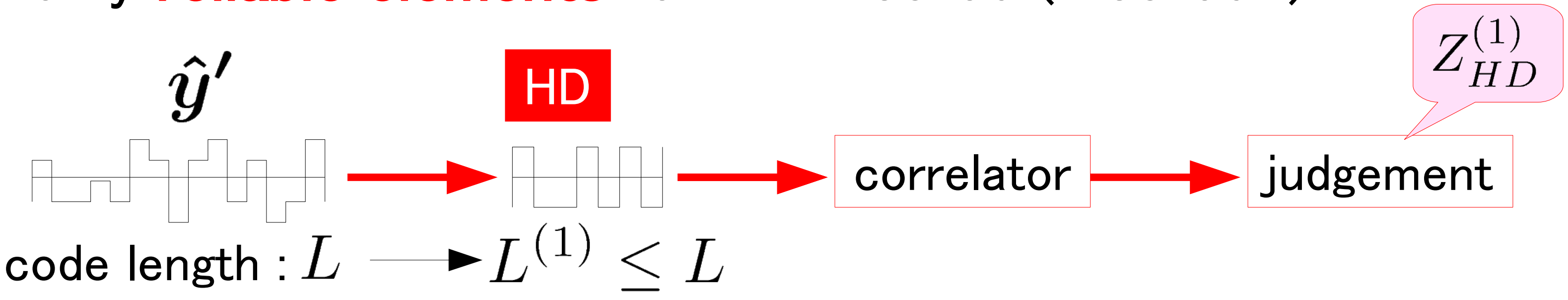
$\hat{y}$

AWGN channel

$\hat{y}' = \hat{y} + e$

$N(0, \sigma^2)$

- Use only **reliable elements** for HD method (method I)



$\hat{y}'$

HD

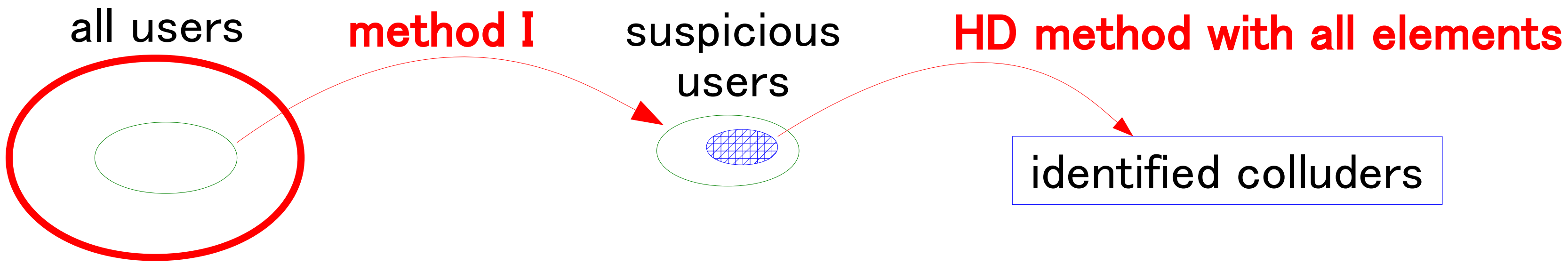
correlator

judgement

$Z_{HD}^{(1)}$

code length : $L \longrightarrow L^{(1)} \leq L$

- method I + unreliable elements (method II)



all users

method I

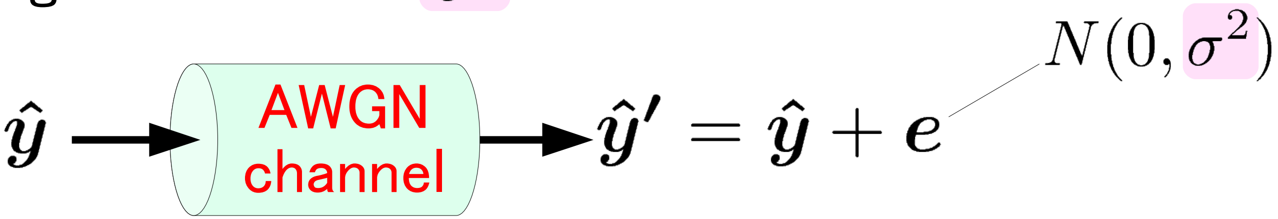
suspicious users

HD method with all elements

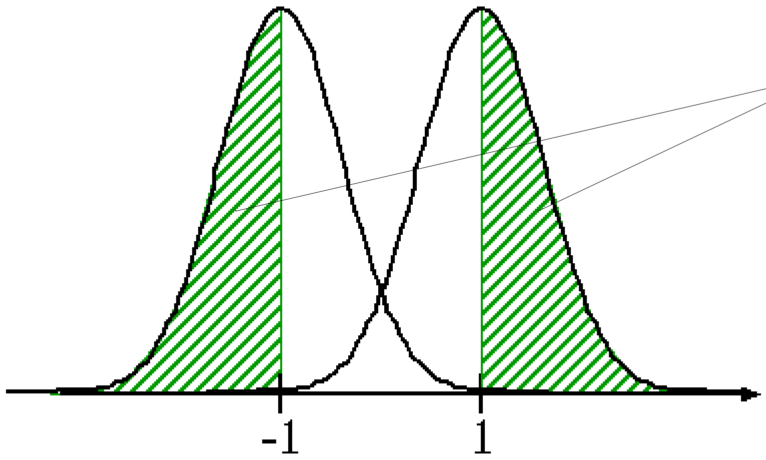
identified colluders

Estimation of Channel

- Estimating the variance σ^2 of noise


$\hat{\boldsymbol{y}}$
AWGN channel
$\hat{\boldsymbol{y}}' = \hat{\boldsymbol{y}} + \boldsymbol{e}$
$N(0, \sigma^2)$

The distribution of received signal $\hat{y}'_i = \hat{y}_i + e_i$



$\hat{y}_i \in \{-1, 1\}$

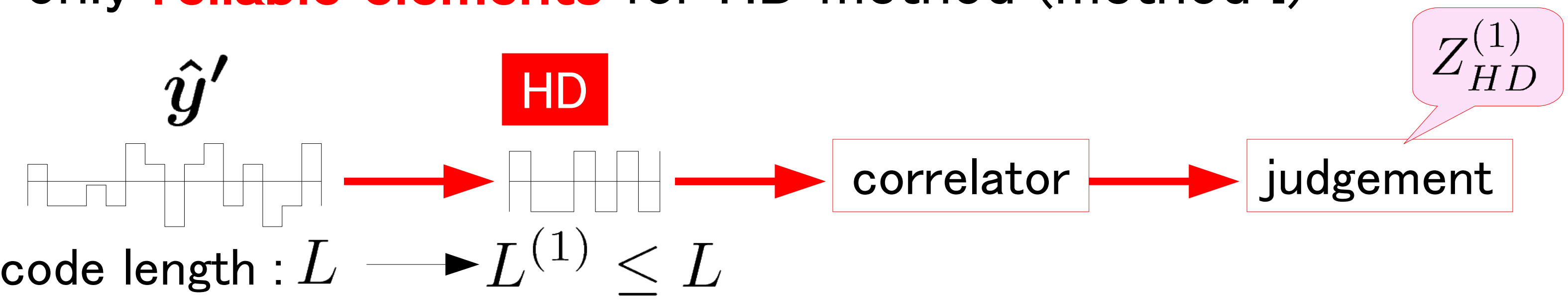
We estimate the variance σ^2 only from these elements;

$$|\hat{y}'_i| > 1$$

The number of such elements is $L/2$ in average.

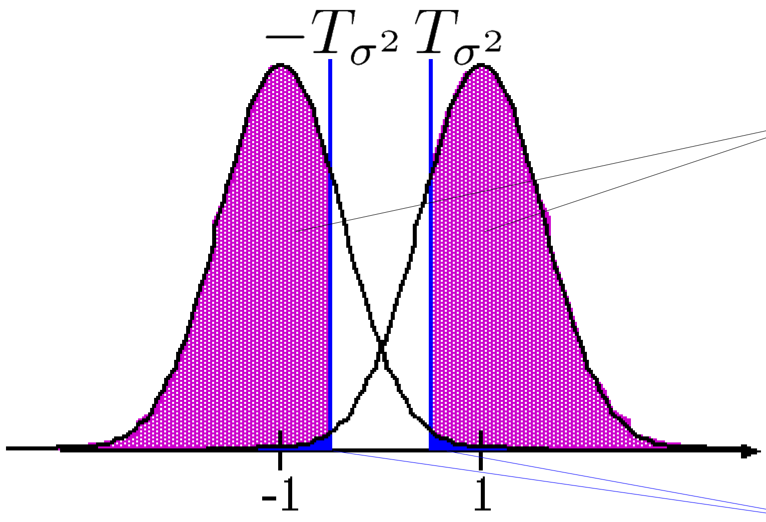
Proposed Method I

- Use only **reliable elements** for HD method (method I)



code length : $L \longrightarrow L^{(1)} \leq L$

The distribution of received signal $\hat{y}'_i = \hat{y}_i + e_i$



variance σ^2

reliable elements: $|\hat{y}'_i| \geq T_{\sigma^2}$

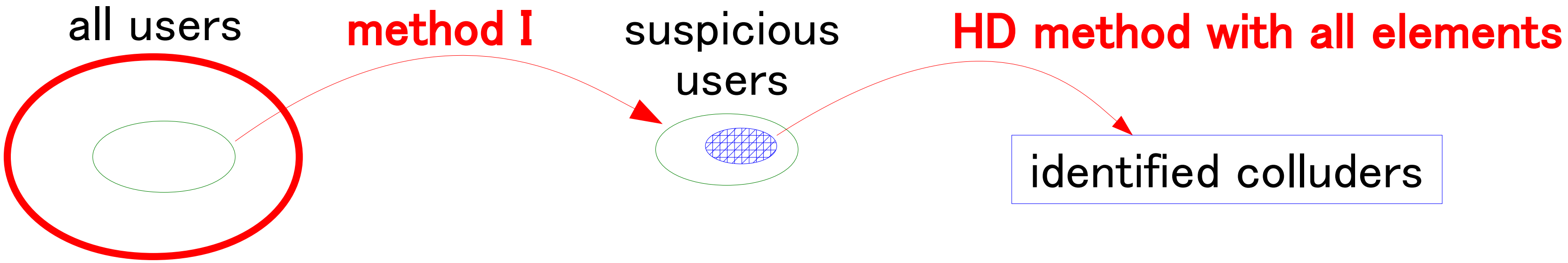
# of bit flips: $P_{flip}L$

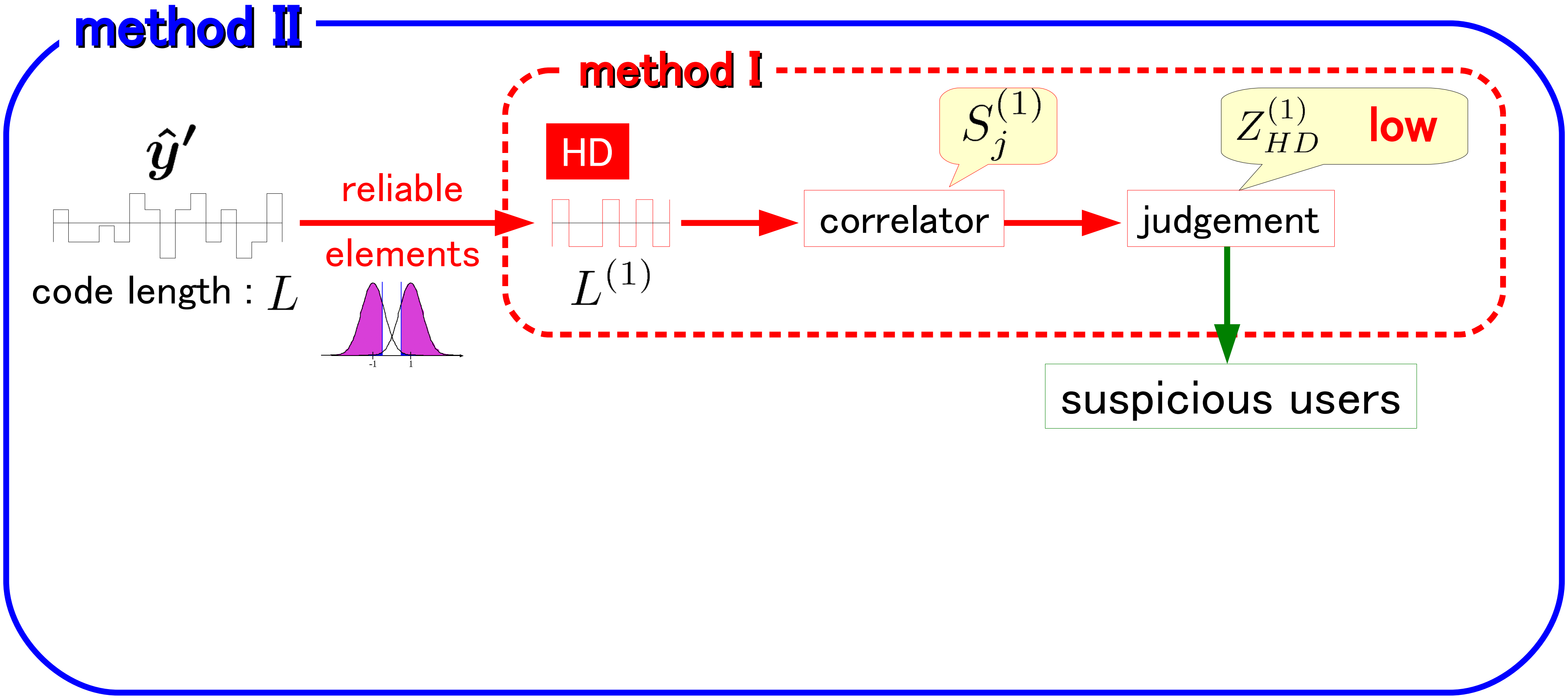
where T_{σ^2} is a threshold and

$$P_{flip} = \frac{1}{2}\mathrm{erfc}\left(\frac{T_{\sigma^2}}{2\sigma^2}\right)$$

Proposed Method II

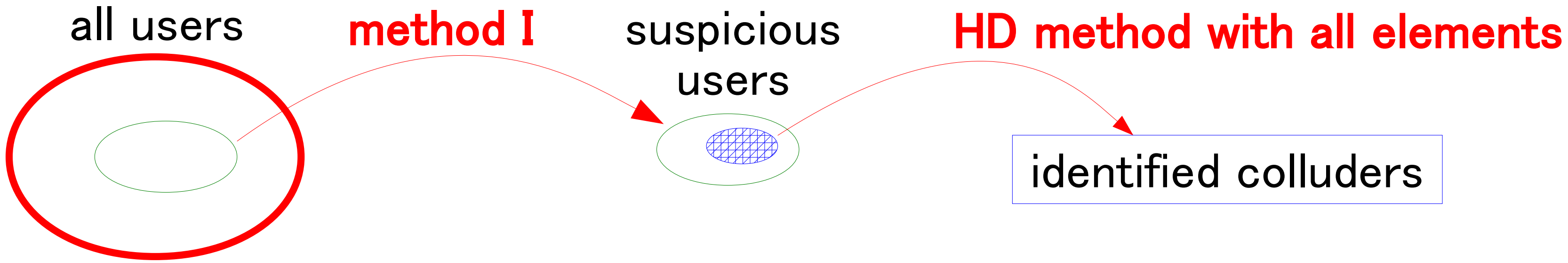
- method I + unreliable elements (method II)

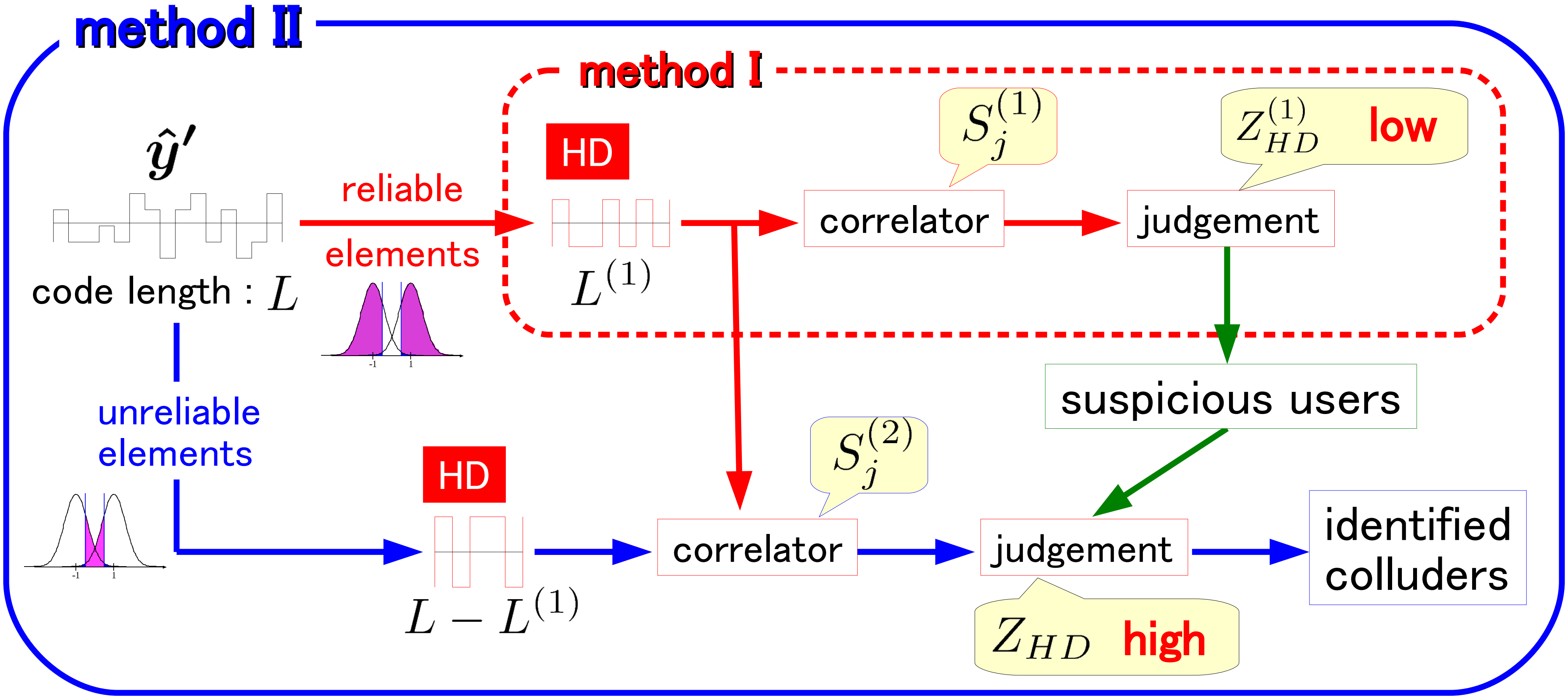




Proposed Method II

- method I + unreliable elements (method II)





Comparison ($\tilde{c} = 10$)

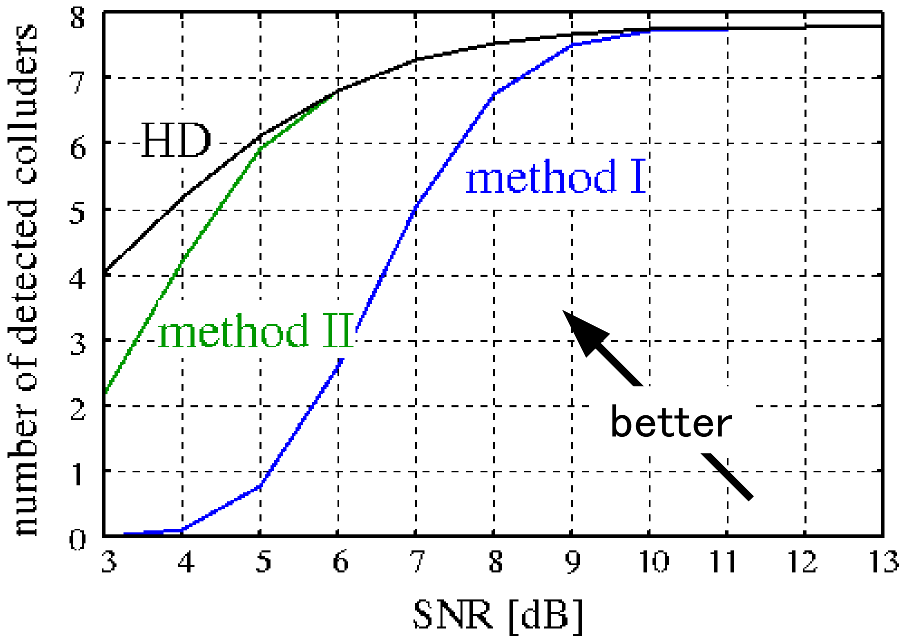
code length : $L = 10^4$

Pr[FP] : $\epsilon_1 = 10^{-4}$

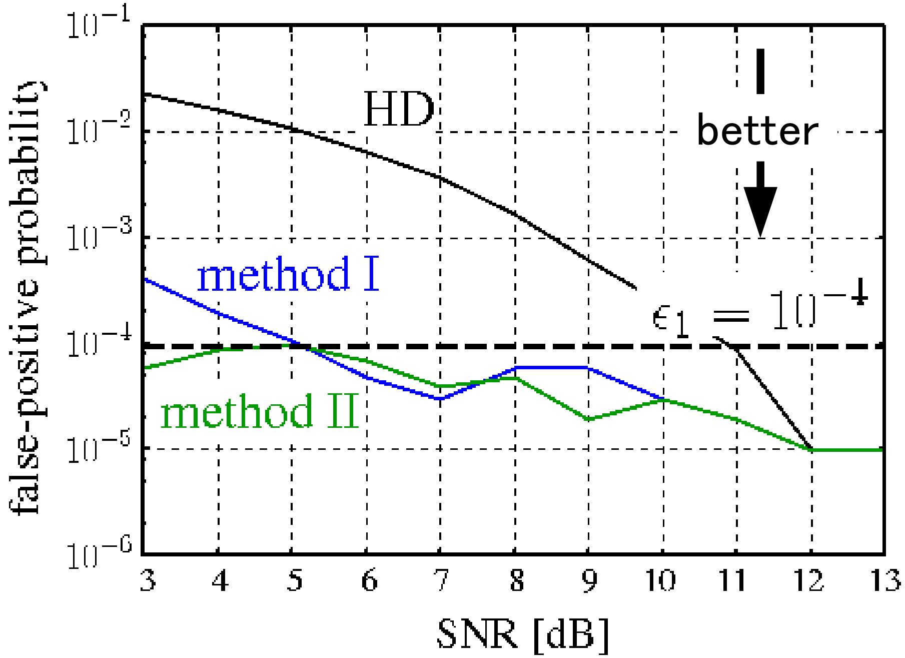
# of users : $N = 10^4$

# of bit flips : $P_{flip}L = 1$

# of detected colluders



False positive probability



The method II can detect many colluders without increasing the false positive probability.

False Positive Probability

# of colluders : $\tilde{c} = 10, 20, 100$

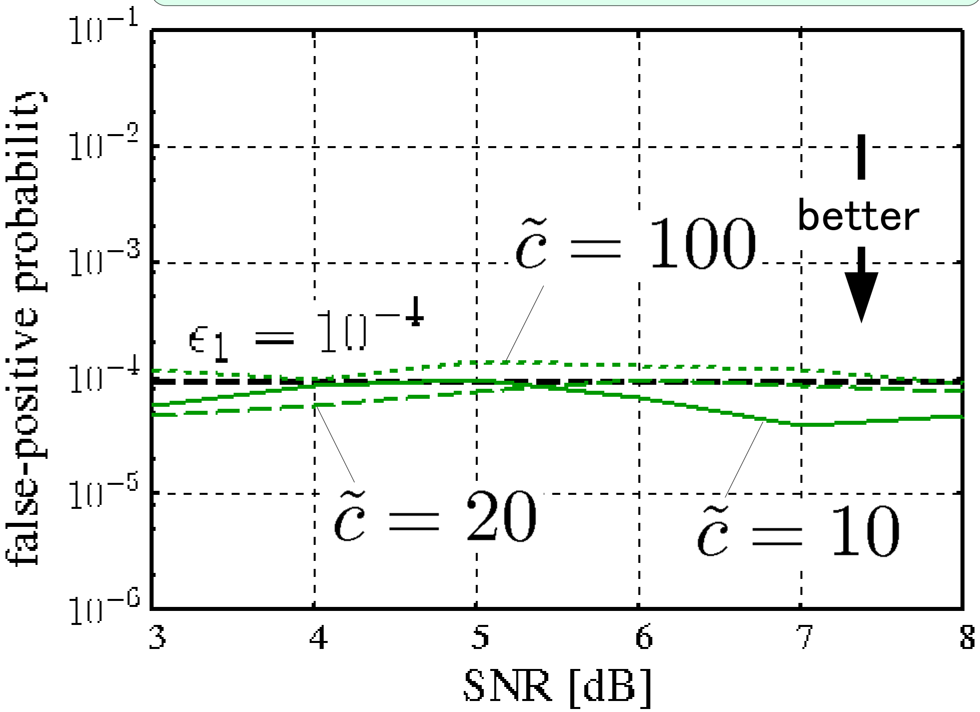
code length : $L = 10^4$

# of users : $N = 10^4$

Pr[FP] : $\epsilon_1 = 10^{-4}$

# of bit flips : $P_{flip} L = 1$

False positive probability



The false positive probability is independent on the number of colluders.

Concluding Remarks

- When the amount of noise is very small, the use of CLT is valid.
- The false positive probability is increased with the amount of noise.

  It is advisable to design a threshold considering the noise.

- We can identify colluders only from reliable elements without the serious increase of false positive probability. (method I)
- Among the suspicous users detected by the method I, the HD method can catch many colluders with less innocents.
- The false positive probability does not increase with the number of colluders for the method II.