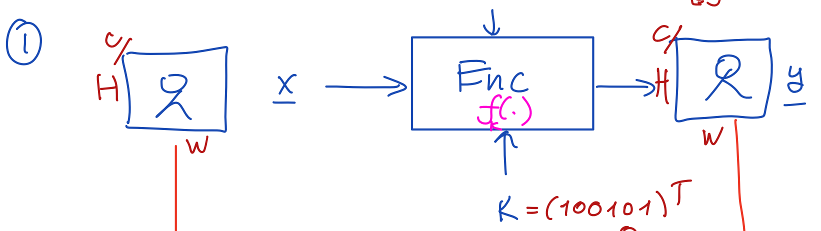
Fundamentals of DWM and DH

06.04.23

Types of modulation

Robust DWM
Steganography
Tamper detection

① $m = (0101\ldots)^T$

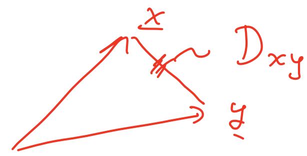


$K = (100101)^T$

$D_{xy} = \|x - y\|_2^2 \le D_{emb}$

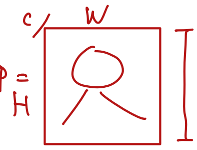
$= (x-y)^T(x-y)$

min



∘ Note: l_2-norm $\neq$ ▷ HVS;

φ-mask



φ = 1 → wm visible; 0 → wm is invisible

$\|x-y\|_\varphi^2 = (x-y)^T \varphi^{-1} (x-y) = \sum_{i=1}^{H}\sum_{j=1}^{W} \varphi(i,j)\cdot(x(i,j)-y(i,j))^2$

$\boxed{y = f_K(x, m)}$ — modulation or WM embedding function.

② Types of $f_K(\cdot)$

— (1) Direct domain $y = f_K(x, m)$

— (2) Transform domain: T, T^{-1}, s.t.: $T T^{-1} = I$

→ (a) $\tilde{x} = T(x) \overset{linear}{=} Tx$ (direct)

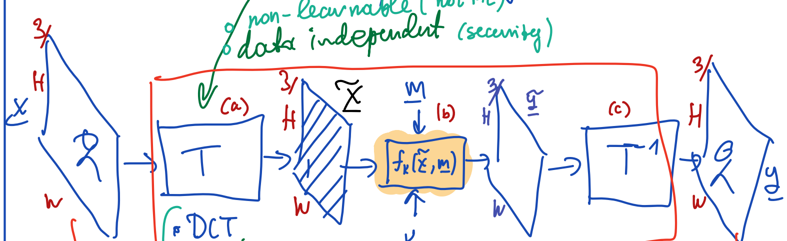
→ (b) $\tilde{y} = f_K(\tilde{x}, m)$ (embedding)

→ (c) $y = T^{-1}(\tilde{y}) = T^{-1}\tilde{y}$ (inverse)

Hand-crafted transform (defined by humans)

∘ non-learnable (not ML)

∘ data independent (security)



public: ∘ DCT, ∘ DFT, ∘ DWT

key-based: ∘ RP (key) → BB...

min D_{xy}

T: $f_\phi(\cdot)$

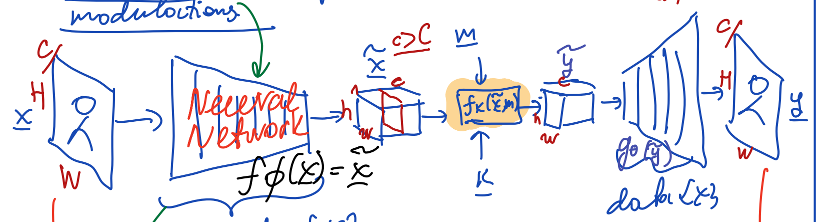
T^{-1}: $g_\theta(\cdot)$

▷ fixed transform

• data agnostic

→ ⫫ {x} indep.

Machine-learning based transform domain modulations

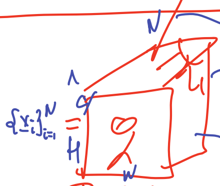


$f_\phi(x) = \tilde{x}$

data {x}

data {x}

min D_{xy}



$f_\phi(x)$

$g_\theta(\tilde{y})$

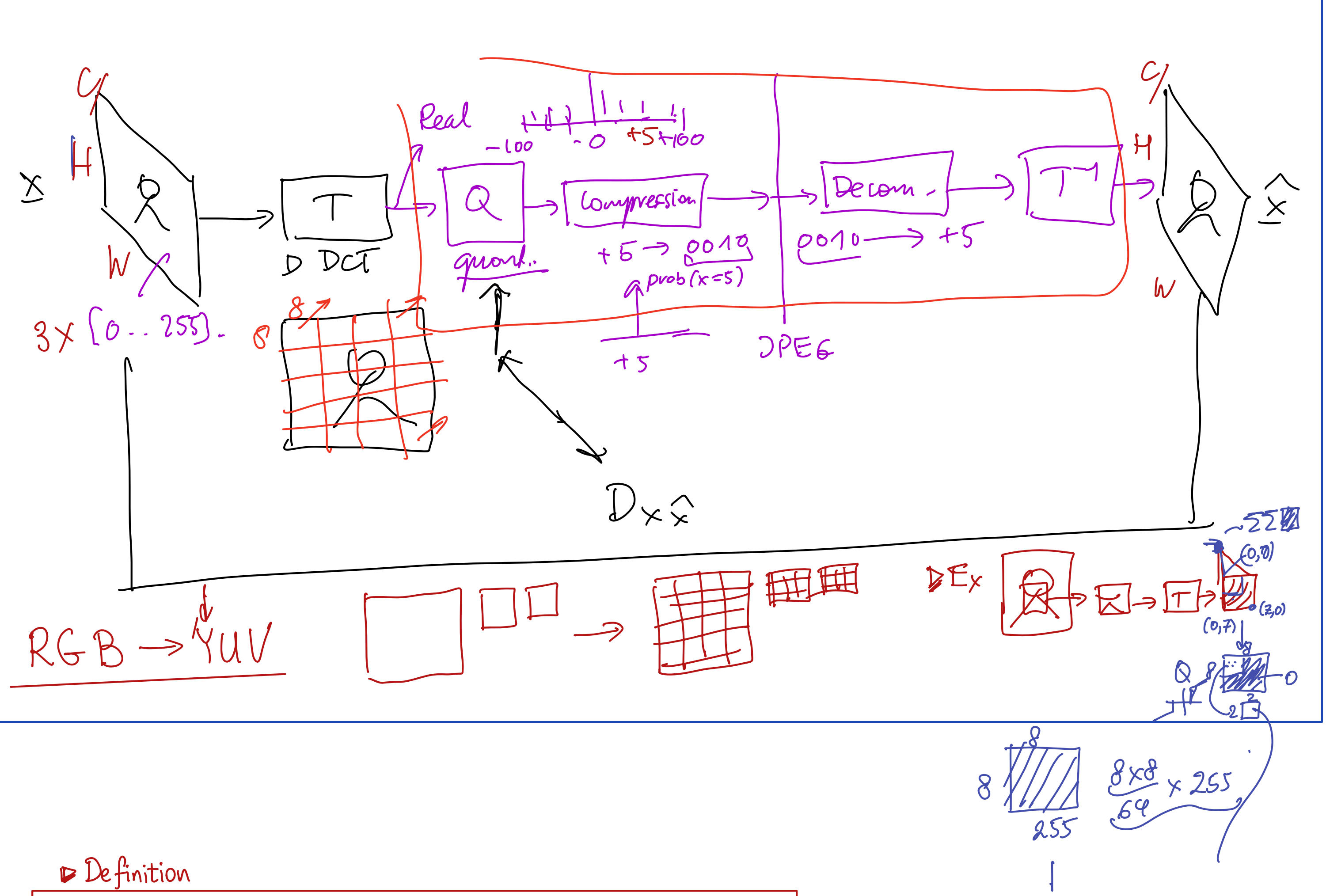
Training data

→ domain specific: ∘ medical ∘ astro ∘ biometrics ⋮ → ML(T) ↑

→ generic $T_{HC} \approx T_{ML}$

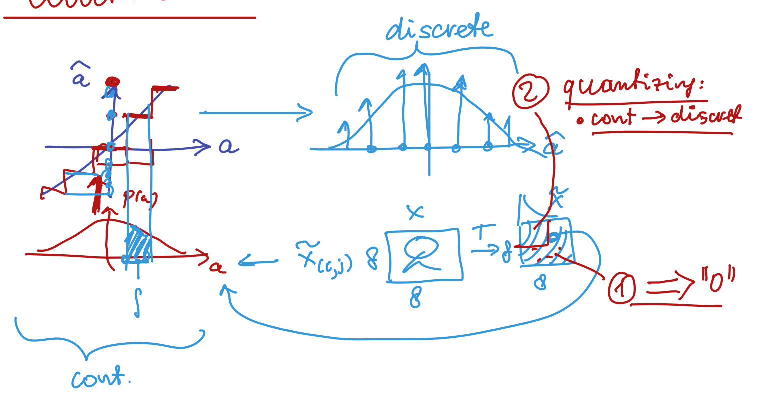
▷ This transform is learned in a function of data {x}.

Link to lossy (JPEG) image compression



Definition

Quantization



$$f_K(\tilde{x}, m) = Q_K(\tilde{x}, m)$$ - quantization based embedding.

• Dither quantizer

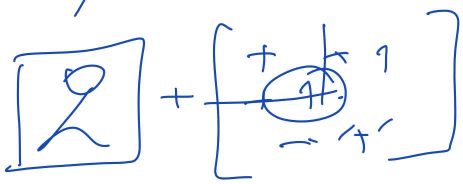
$Q(\tilde{x} + d)$

⧺ ← (K, m)

Recall:

$\boxed{\underline{y} = \underline{x} + \underline{W}}$ $\quad \underline{m}, \underline{K}$ in the direct /image/ domain.

▷ ~~direct domain~~ $(T = I)$
▷ addive modulation
▷ quantization modulation



$\underline{m} = [\,0\ 1\ \ 0\ 1\ 1\ \ ]$
$\underline{K} \to (x_1, y_1)$ $\quad 0 \to -1$
$1 \to +1$

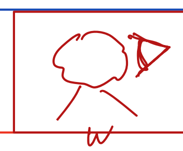
Table: general classification of DH modulation techniques.

$\boxed{\underline{y} = f_{\underline{K}}(\underline{x}, \underline{m})}$

- Direct domain $(I) = T$
  - Additive modulation: $\underline{y} = \underline{x} + \underline{w}$, $\underline{w} \leftarrow \underline{m}, \underline{K}$
  - Quantization modulation
- Transform domain (T): $\tilde{\underline{x}} = T(\underline{x}) \leftarrow \underline{x}$, $\; f_K(\tilde{\underline{x}}, m) \to T^{-1}(.) \to \underline{y}$
  - HC $(T \perp\!\!\!\perp X)$
    - Additive modulation
    - Quantization modulation
  - ML $(T \not\!\perp\!\!\!\perp X)$
    - Additive modulation
    - Quantization modulation

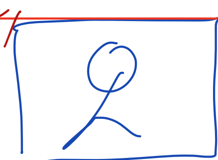
1) Additive modulation.

1.1. Direct domain:

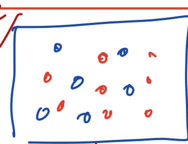


Φ: H

$$\underline{y} = f_{\underline{K}}(\underline{x}, \underline{m}) = \underline{x} + \gamma \underset{\text{masking}}{\varphi} \odot \underline{w} \qquad \underline{w}(\underline{m}, \underline{K})$$



X H $\quad C_1$ $\quad$ W



H $\quad C_1$ $\quad$ W $\quad \underline{w}(\underline{m}, \underline{K})$

$\bullet$ — $+1$
$\circ$ — -1

- $y = x + w$ $\;\leftarrow = x + w$
- $y = T^{-1}\big(T(x + w)\big)$
  $= T^{-1}\big(T(x) + T(w)\big) = \underset{I}{T^{-1}T}(x) + T^{-1}T(w)$

$$\underline{y} = T^{-1}\big[\underbrace{T(\underline{x})}_{\tilde{x}}\big] = \underline{x} \; + T^{-1}\big(\gamma \tilde{\varphi} \odot \tilde{\underline{w}}\big)\big]$$

$T^{-1}[\tilde{x}$

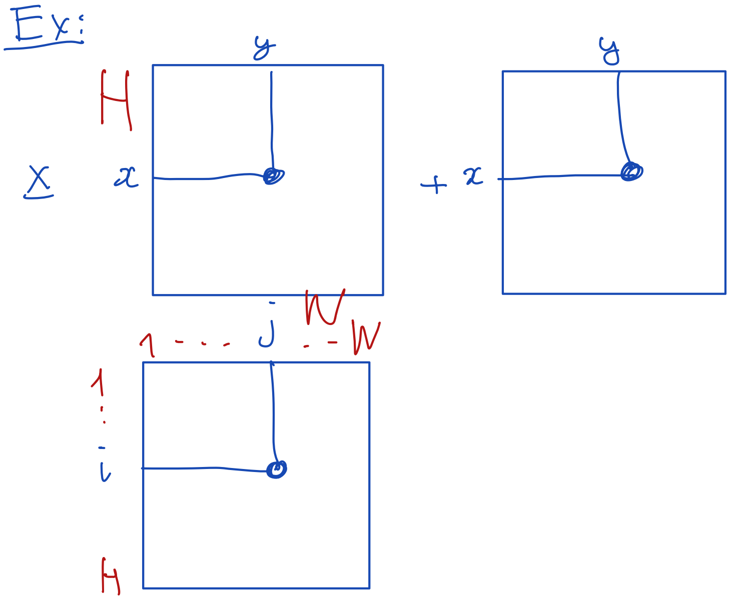
1.2. Transform domain:

$$\underline{y} = f_K(\underline{x}, \underline{m}) = T^{-1}\big(\tilde{T}(\underline{x}) + \tilde{\gamma}\tilde{\varphi} \odot \tilde{\underline{w}}\big) \qquad \underline{w}(\underline{m}, \underline{K})$$

HC $\quad$ ML

- $T(aX + by) = aT(x) + bT(y)$

▷ Ex:



$\underline{y} = \underline{x} + \underline{w}$

$\Updownarrow$

$y[i,j] = x[i,j] + w[i,j]$

$\in [0 \dots 255]$ $\quad \{\pm 1\}$

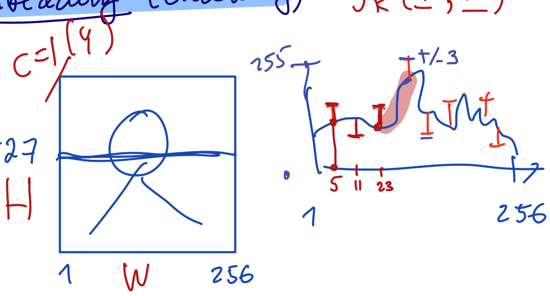
$\times \gamma \cdot \varphi[i,j]$ $\quad \in [0 \dots 1)$

$\gamma > 1$

$\pm 3 \quad \pm 7$

① Embedding (encoding) $f_K(\underline{x}, \underline{m})$

$\varphi <$ JND (Meta) / NVF (Unife)



- position $\equiv \underline{K} \Rightarrow (5, 11, 23 \dots)$
- sign $\equiv \underline{m}$.

0 → −1
1 → +1

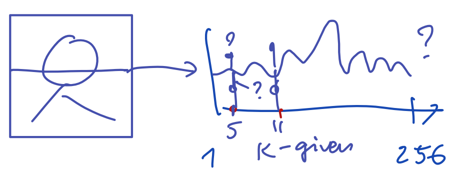
1 0 1 1 0 1 1 0

$\gamma = 3$ || (+1) (−1) +1 +1

$y = x + w \odot \varphi$

② Extraction (decoding)

⚠ The extraction of WM is NOT a trivial problem.

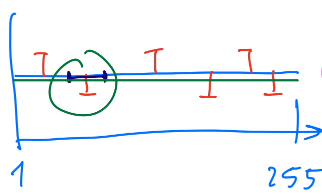


DHider (defender) Attacker

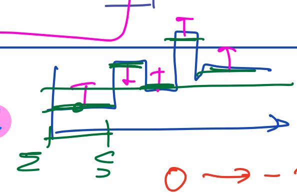
Why?

▷ Intuition.

> ▷ Suppose!



Const signal



0 → −1
1 → +1

$y = x + \gamma \varphi w$

scalar

$[0, .5]$

↓

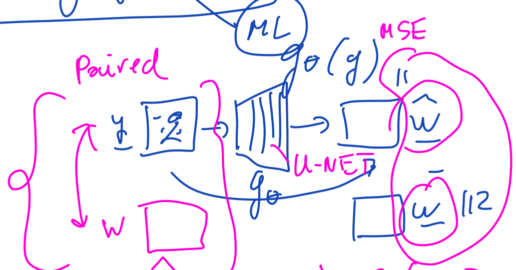
$\hat{w} = g(y)$

MC

ML

$g_\theta(y)$ MSE

Paired



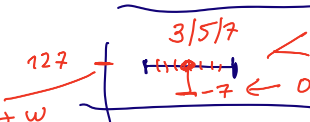
$\hat{\theta} = \arg\min_{\theta} \mathcal{L}(y, w) = \sum_{m=1}^{M} \| w_m - \hat{w}_m \|_2^2$

$\llcorner g_\theta(y_m)$

- extraction on the window → "local mean".

Local mean

(a)



mean (a)

median.

$y = x + \underset{(-7)}{w}$

3: $\frac{127 + (127-7) + 127}{3}$

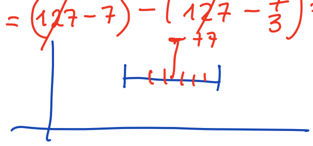
↓

- $\hat{x} = \frac{1}{3} \sum_{i=1}^{3} x_i$, $\quad \frac{\cancel{3} \times 127}{\cancel{3}} - \frac{7}{3} = \hat{x}$

↓

$\hat{w} = y - \hat{x} = (\cancel{127} - 7) - (\cancel{127} - \frac{7}{3}) = -7 + \frac{7}{3} = \int$

(−1)



Local media. (5)

Watermark w faces a strong interference from the side of host image x.

1) $y = x + w$

2) $\hat{x} = pred(y)$

↳ denoiser

HC { ↳ wiener2 $= \bar{y} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_w^2}(y - \bar{y})$

↳ Σ - mean value: $\bar{y}$, σ_y^2

↳ median

↳ ML (U-NET)

127

$y_w = [(127, (127-7), 127)]$

↓ med. $(127-7, 127, 127)$

"local mean" → local media

$\hat{x}_{med} = 127.$

$\hat{w} = y - \hat{x} = (127-7) - 127 = -7.$

127 +7

$y_w = (127, (127+7), 127)$

↓ $\hat{x}_{med}$ 127 127 127+7

127.

$\hat{w} = y - \hat{x}_{med} = (127+7) - 127 = +7.$

1) x + w = y $\{\pm 1\}$ γ 3/5/7

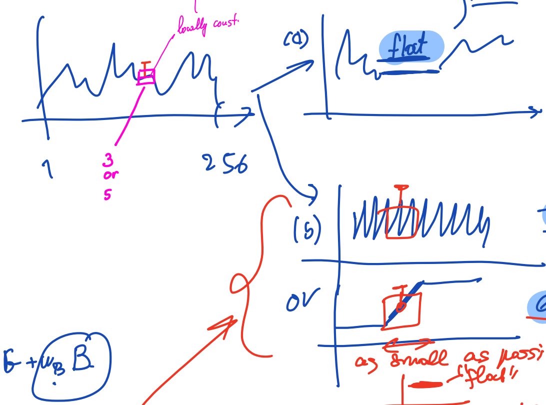
$y = \hat{x} + w$

2) wiener2(y) [3 3] $= \hat{x}$

3) $\hat{w} = (y - \hat{x})$ sign) vs

$W \in \{\pm 1\}$

▷ Real signals



R G B

+/-28 γB

±2/±3 γG γR

YUV

$w_R R + w_G G + w_B B$

no accurate prediction.

±3 → ±7 (±10)

▷ accurate extraction.

still invisible.

attacker cannot do the accurate prediction

security

* ▷ In case of image:

$y = x + \underset{[m,k]}{w} = x + \underset{noise}{\epsilon}$

o[IP]: denoiser: $w \equiv \epsilon$ as noise.

(a) $\boxed{\hat{x} = denoiser(y)}$

Matlab: wiener2(y, [3 3], σ_ϵ^2) 5 5 / 3x3 auto.

(b) $\boxed{\hat{w} = y - \hat{x}} = [x + w] - \hat{x}$

b.1. if $\hat{x} = x$: $\hat{w} = w$.

b.2: if $\hat{x} = x + z_{est}$:

$\hat{w} = y - \hat{x} = \cancel{x} + w - \cancel{x} - z_{est} = w - z_{est}$.

→ repetitive embedding.

one bit +1

multiple times

[+3.5 +5.7 -1.3 +2.5] = Σ/4

window size for the prediction.

flat. vs edges & texture.

Recap of DWM

$m = \boxed{0\,1\,0 \cdots 0\,1\,0\,1}$ (positions $1 \ldots L$)

$K = \boxed{1\,0\,1\,0\,1\,1 \cdots 0}$ (positions $1 \ldots M$)

(iid) $K_{8,m}$

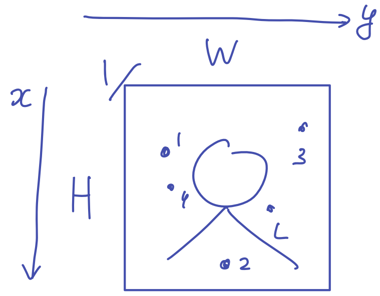
$m \oplus K$ (xor) $\rightarrow \boxed{0\,1\,0\,1\,0\,0\,1\,1} \rightarrow \boxed{ECC} \rightarrow$

- (security)
- diversity.

$(K_{8,m}, K_{8,p}) \rightarrow K_{8,P}$

$\downarrow$

L bits



Allocation of bits

$m = \boxed{0\,1\,0 \cdots 0\,1\,0\,1}$ (positions $1, 2, \ldots, L$)

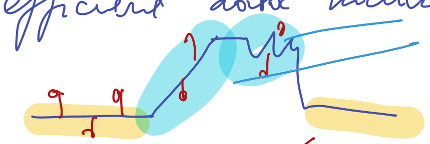
(x_1, y_1) (x_2, y_2) (x_L, y_L)

$\boxed{\text{" 1 bit = 1 position "}}$

Note: $H \times W$ vs L

$\underbrace{3k \times 4k}_{12 \text{ Mio}}$ 64 or 128 bits.

- ⊖ $H \times W$ "redundancy" is not used for efficient data hiding.
- ⊖



the acc. of prediction is low.

ok → predict "bits"!

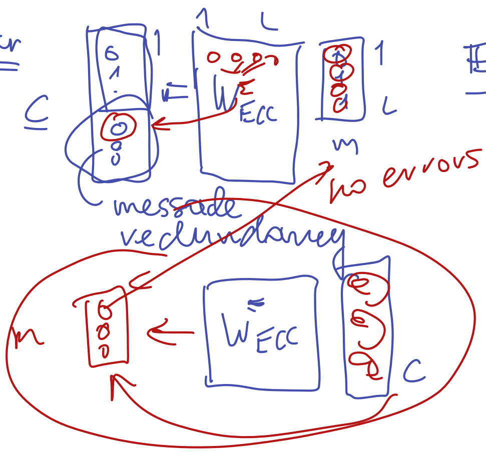
▷ Solution.

① ECC

② – w ECC ($R = 1/\bigcirc$) → complexity
– w/o ECC → repetition.

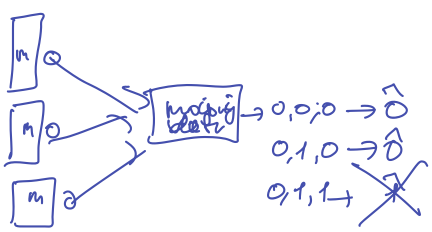
$m = \boxed{0\,1\,0\,1} \rightarrow \boxed{ECC} \rightarrow$

- Turbo → Polar
- RS
- BCH



Rate 1/3:

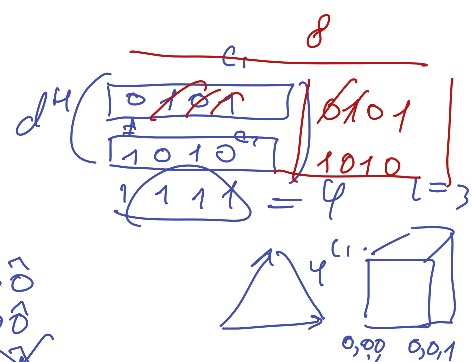
Ex:



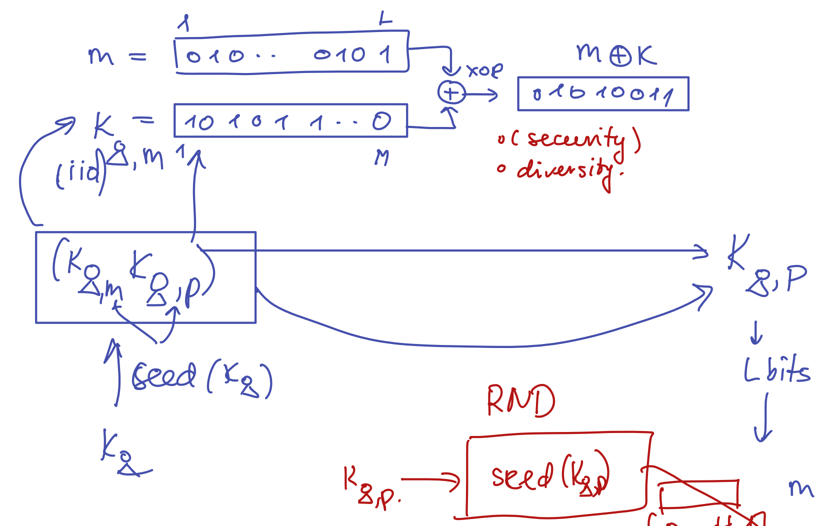
Repetitive codes → not efficient

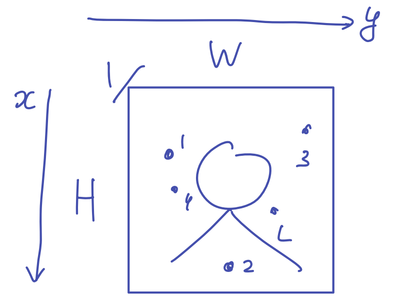
$\boxed{M} \longrightarrow \boxed{}$

$K \longrightarrow \boxed{}$

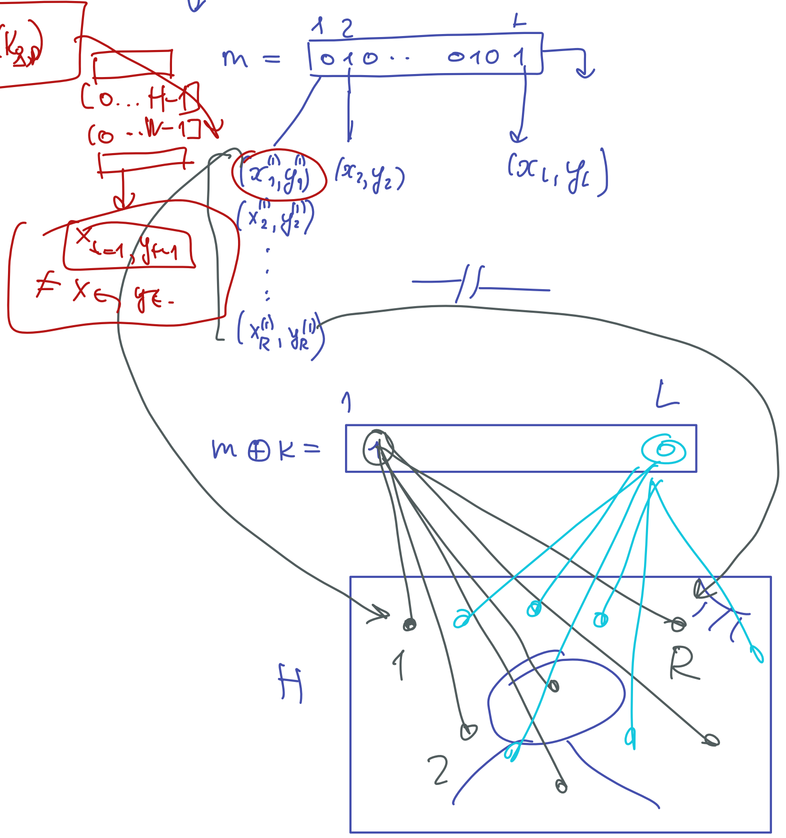


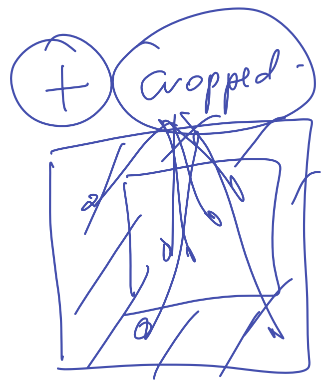
w/o ECC : "1 bit ⇒ multiple positions!"





Allocation of bits



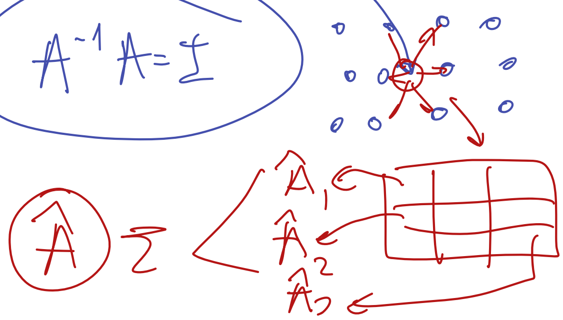


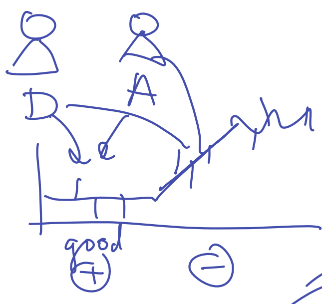
(+) Multiple → special way.

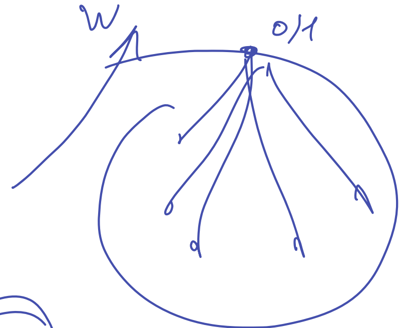
* Note:

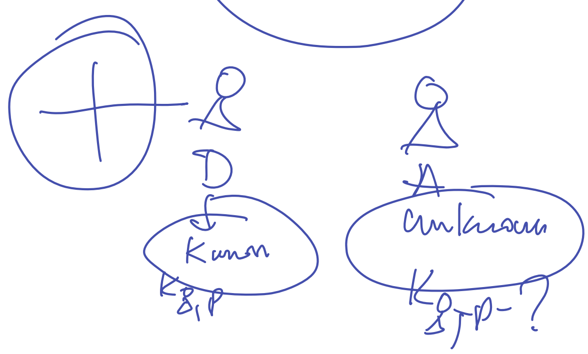
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

$$A^{-1}A = I$$

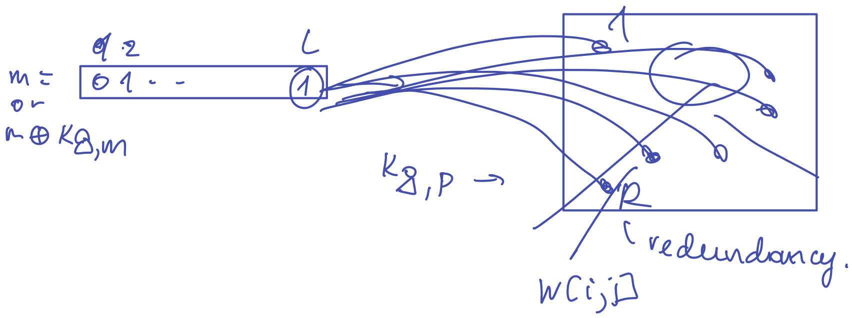








Message decoding via aggregation.



▷ "Place" a bit?

1) Modulation

$\{m_i\}_{i=1}^{L}$ $0 \longrightarrow -1$

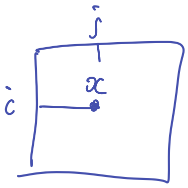
$1 \longrightarrow +1$

$c_i = \boxed{2 \cdot m_i - 1}$

$m_i = 1: \; 2 \cdot 1 - 1 = 1$

$m_i = 0: \; 2 \cdot 0 - 1 = -1$

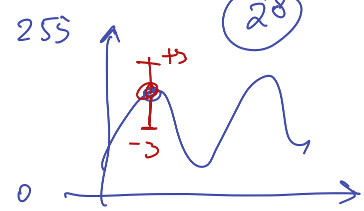
2) Embedding



$x[i,j] + w[i,j]$

$(\in \{0, \dots, 255\}$ 2^8 $\in -1 \uparrow \; +1 \uparrow$

-3 $+3$

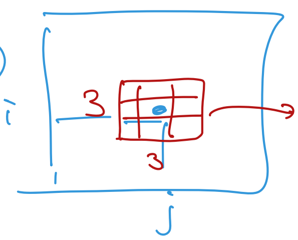


▷ Decoding.

① Bit extraction

$K_{\delta,p} \to (i,j)$

$y =$



$\to$ denoising (wiener2(y))

$x(i,j) + w(i,j)$

0 — 255 $+/- \, 3$

② . Denoising:

$\hat{x}[i,j] = \text{wiener2}[y(i,j)]$

$\hookrightarrow x[i,j] + w[i,j]$

③ Prediction:

$\hat{w}[i,j] = y[i,j] - \hat{x}[i,j]$

$= \cancel{x[i,j]} + w[i,j] - \cancel{\hat{x}[i,j]}$

(a) if $\hat{x}[i,j] = x[i,j]$:

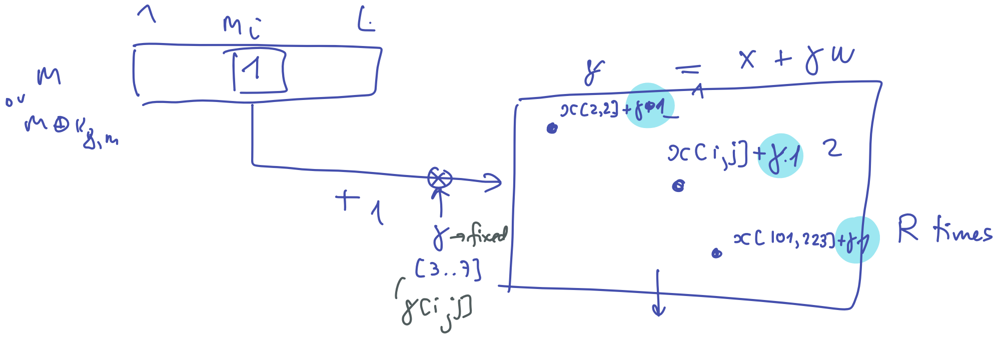
$\hat{w}[i,j] = w[i,j]$

(b) $\to$ real: $\hat{x}[i,j] = x[i,j] + e[i,j]$

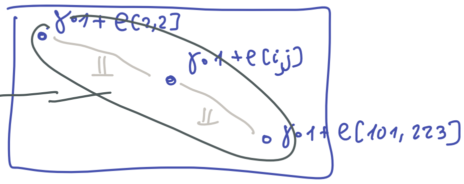
flat: $e[i,j] \longrightarrow 0$

edge: $e[i,j] \longrightarrow \uparrow$

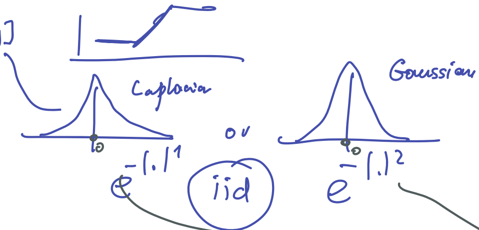
15 · 21 $> |w|$



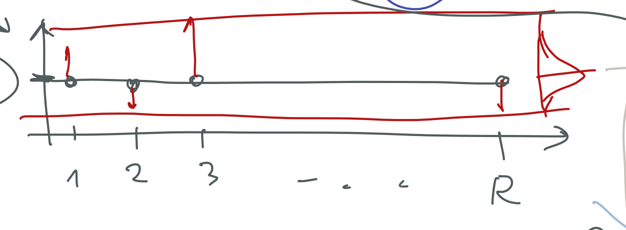
$\hat{w} = y - \text{denoiser}(y)$



Note: $e[i,j]$ ~



Imagine: $1\cdot\delta$



$$\frac{1}{R}\sum_{k=1}^{R}(1\cdot\delta + e[k]) = \frac{1}{R}\sum_{k=1}^{R}\boxed{1}\delta + \frac{1}{R}\sum_{k=1}^{R}e[k]$$

$+1$ const; $\frac{1}{R}R\cdot\boxed{1}\delta \to \delta$

$R\to\infty$: $\to 0$; $\frac{\sigma_e^2}{R}$, $R\to\infty$, $\to 0$

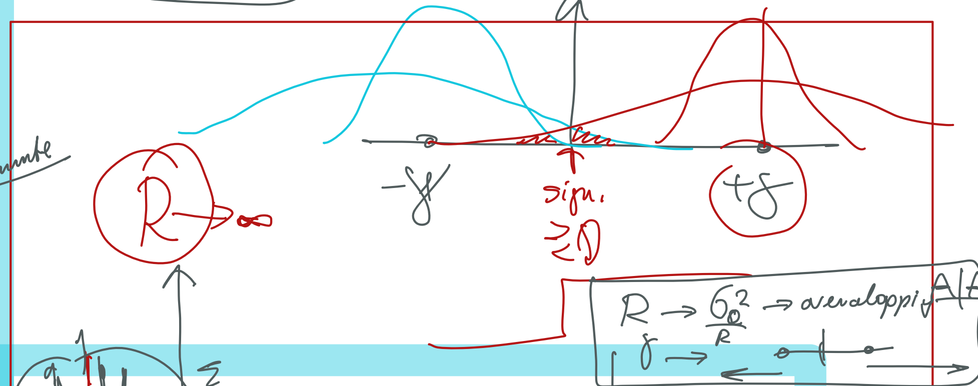
2'000'000 / 100 bits; 120k; 100 ÷ 200; 2..100k

$$\text{sign}\left[\frac{1}{R}\sum_{k=1}^{R}(\boxed{1}\delta + e[k])\right] \to +1,\ -1$$

+ / −

$\binom{12M_{10}}{100}$ $\binom{12M_{10}}{20}$

δ – fixed.



Conclusions: impact of redund. R

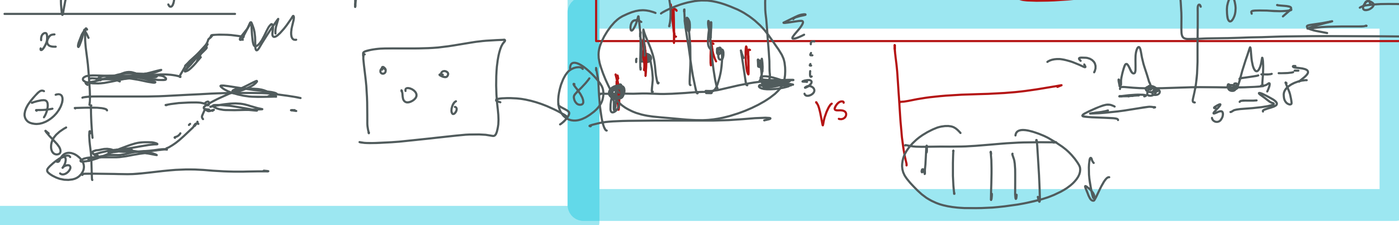
① 1) $H\times W(x) \to R$

2) $R \to \frac{\sigma_e^2}{R} \to$ $\frac{\sigma_e^2}{R}$ $\frac{\sigma_e^2}{R}$ accumulate

3) A vs D in adv. (position)

can't accumulate.

② impact of δ / str. of DWM).



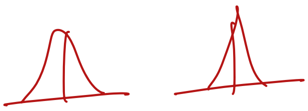
Sample mean

- Consider a sample mean of N independent and identically distributed random variables $X_1, X_2, \cdots, X_N$ with $E[X_i] = \mu_X$ and $Var[X_i] = \sigma_X^2$
- The **sample mean** is

$$M_N(X) = \frac{S_N(X)}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

- The moments of **sample mean**
  - **Expected value:** $E[M_N(X)] = E[X] = \mu_X$
  - **Variance:** $Var[M_N(X)] = \frac{Var[X]}{N} = \frac{\sigma_X^2}{N}$

σ_e^2



$N = R$

$$\frac{1}{R}\sum_{k=1}^{R} e[k]$$

Sample mean: moments

- **Proof**
- **Expected value**: sampling mean is an unbiased estimator

$$E[M_N(X)] = \frac{1}{N}(E[X_1] + \cdots + E[X_N]) = \frac{1}{N}(E[X] + \cdots + E[X]) = E[X] = \mu_X$$

- **Variance**: suppose that $Var[X_i] = \sigma_X^2$

$Var[e_1] \cdots + Var[e_R]$

$$\boxed{Var[M_N(X)]} = \frac{1}{N^2}(Var[X_1] + \cdots + Var[X_N])$$

$$= \frac{1}{N^2}(Var[X] + \cdots + Var[X]) = \frac{1}{N}Var[X] = \frac{\sigma_X^2}{N}$$

R R

$Var[a \cdot X] = a^2 Var[X]$

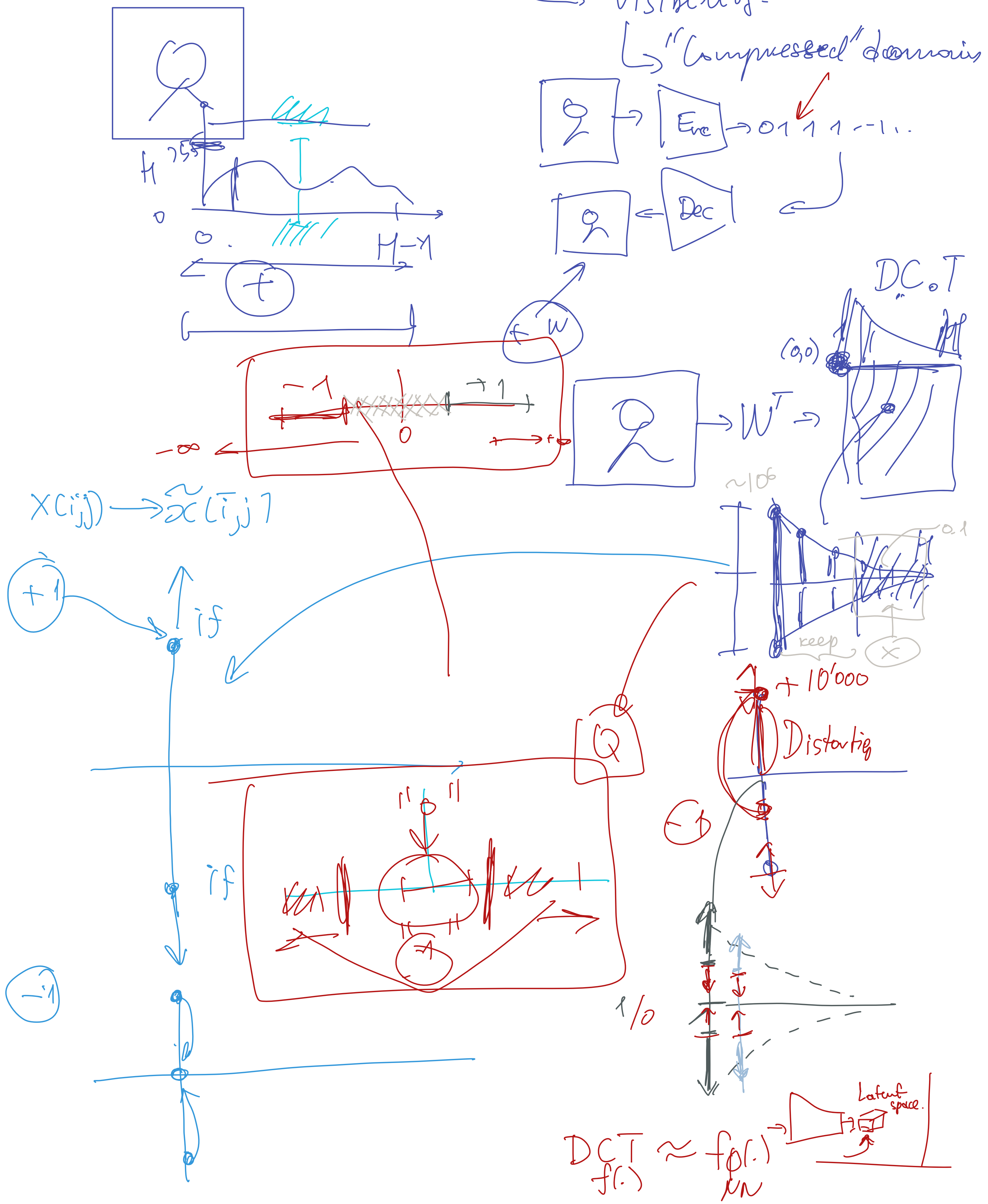
$\frac{1}{R}(e[1] + e[2] + e[R])$

$$\frac{\sigma_e^2}{R}$$

The variance of sample mean approaches zero as the number of samples increases (a.k.a. efficient estimator)

1) "Pixel" domain

2) → Complexity
→ Compliancy with compression.
→ Visibility.
↳ "Compressed" domain

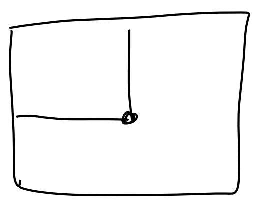


"Distributed" watermark embedding into multiple pixels: the redundancy again

$\underline{m} = [0\ 1\ 0\ \dots\ 1]$ bits.

① $0 \to -1$, $1 \to +1$ — $[-1\ +1\ -1\ \dots\ +1]$ bits.

② key $\underline{k} \to$



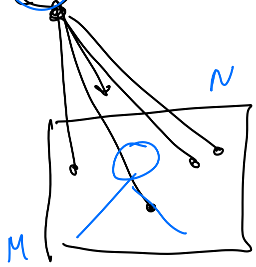
(a) $(x_1, y_1)\ (x_2, y_2)\ \dots\ (x_L, y_L)$.

(no redundancy).

1 bit $\equiv$ 1 location. L

$[-1\ +1\ +1\ \dots\ +1]$ bits

(b).



$(x_1^{(1)}, y_1^{(1)})\ (x_2^{(1)}, y_1^{(1)})\ \dots$

R = # of repetitions.

$\boxed{R \times L}$

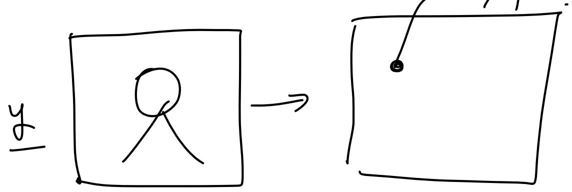
(Label A)

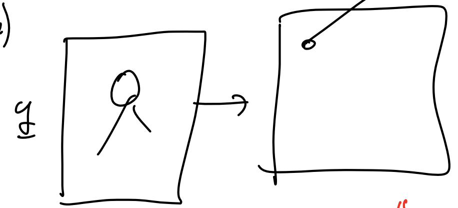
$\boxed{RL}$ positions

Attacker — **Defender**

$+/-1$ – ? or $\emptyset$ (not modified)

$+/-1$ – ? or $\emptyset$





- has no info where the same bit was placed.

$[+1\ \dots\ ]$

R

?

Σ

$+/- \Rightarrow$ CLT $\Rightarrow$ -1, $+1$

$P_r^0 = 0.5$ — $+1$, -1

▷ has an info about "R times" locations.

$[+1\ \dots\ ]$

$+/-1$

locations.

① $\emptyset \to$ is not active.

$(+/-1)$

② R — Σ — $\geq$ — $+1$ / -1

$+0.2$, $+1.3$, $+0.75$, $+3.5$, -0.1, … R

aggregat. $\times$ w_1, w_2, …, w_R

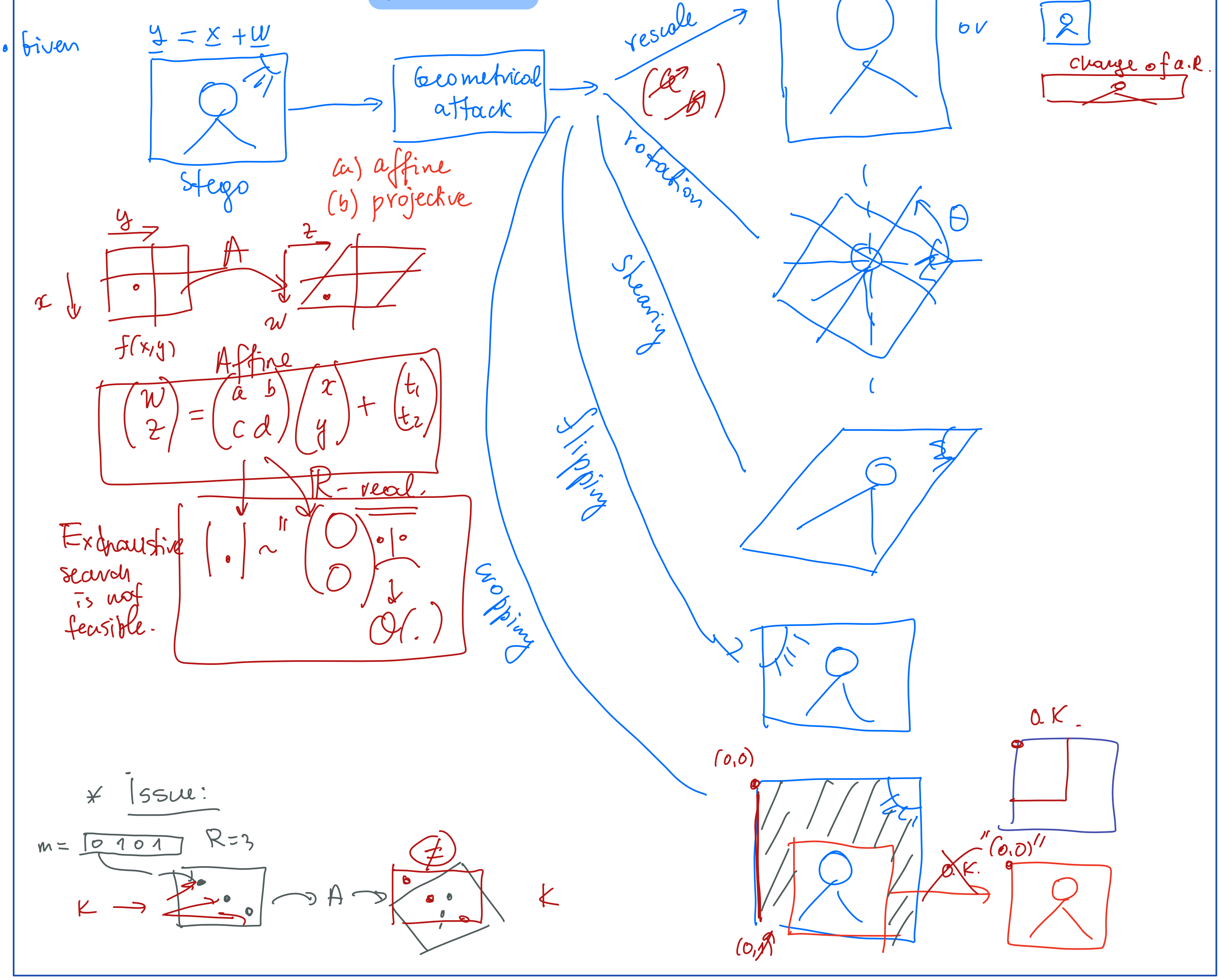
$\boxed{\text{aggregat.}}$

IID

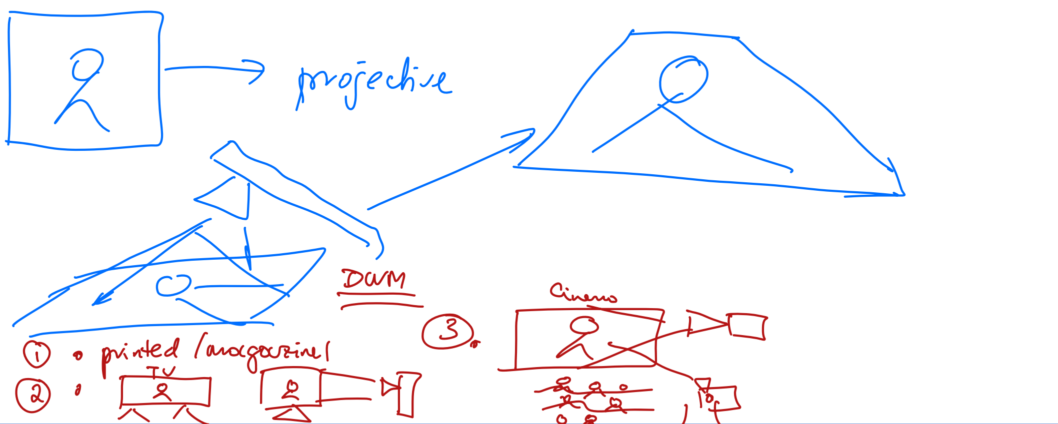
• IJ smd.

σ_x^2

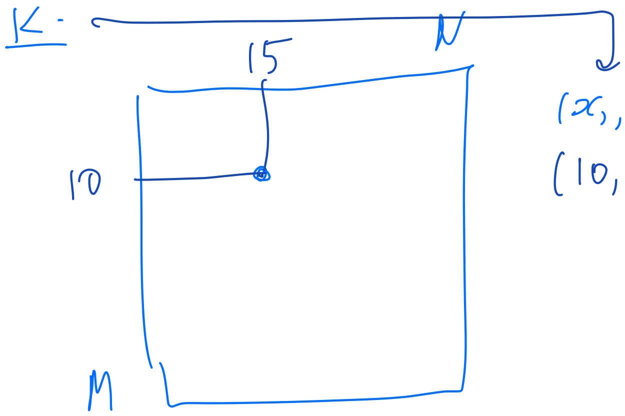
Robustness to geometrical attacks.



Practical use case:



① $\underline{K}$:

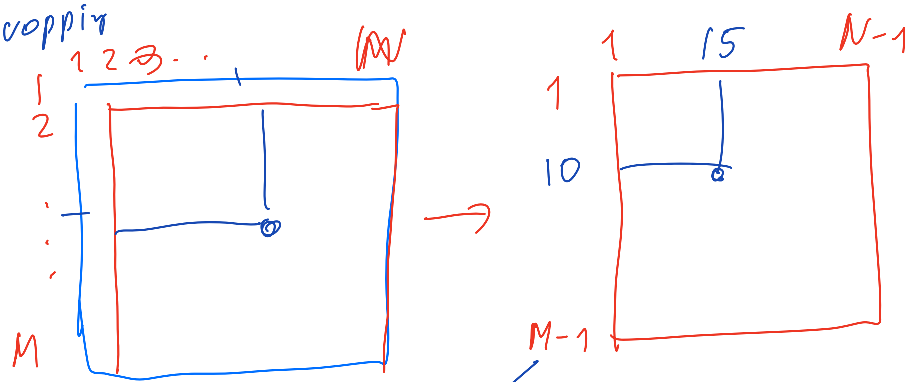


(x_1, y_1) (x_2, y_2)

$(10, 15)$

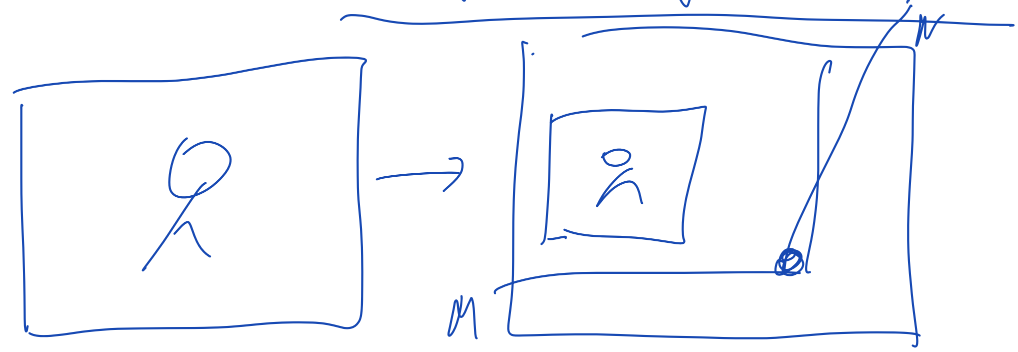
② Extraction

(a) Cropping

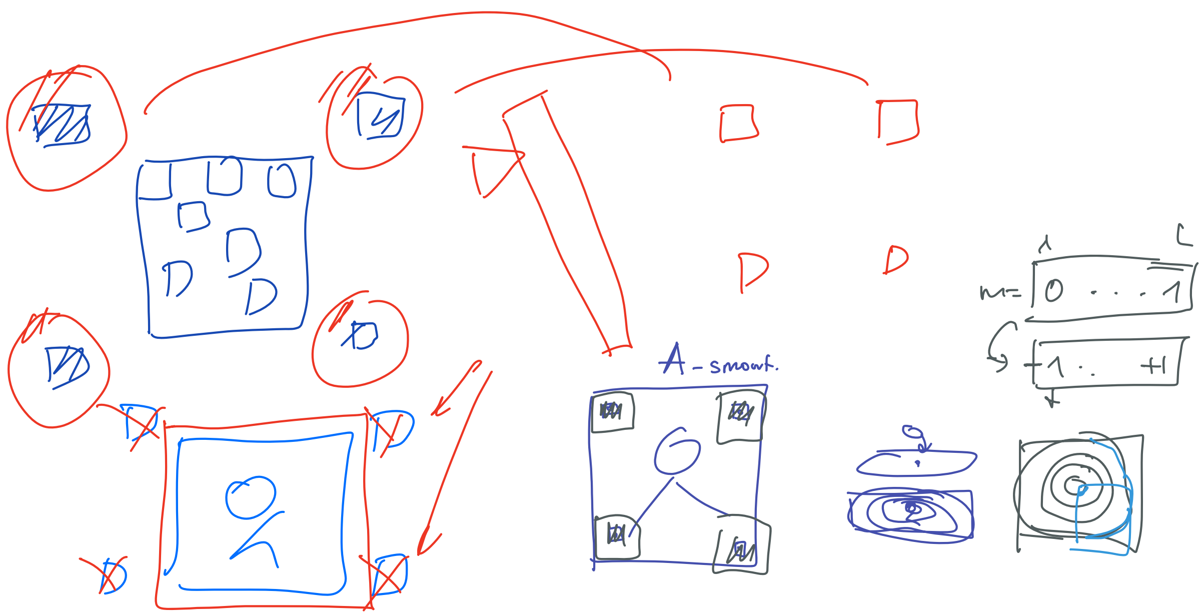


$\underline{K} \longrightarrow (x_1, y_1)$ (x_2, y_2) - - - -

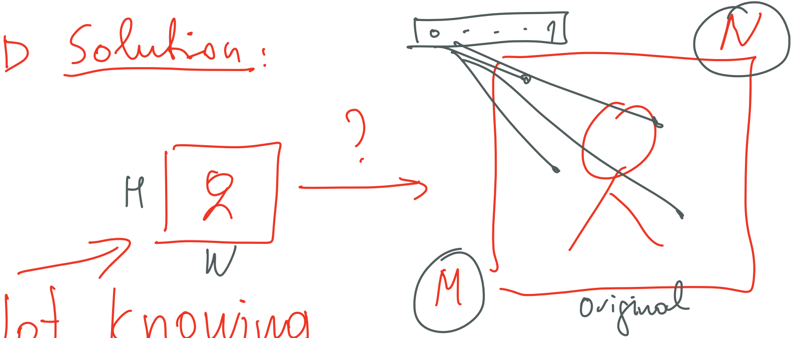
(b)



QR codes



D Solution:

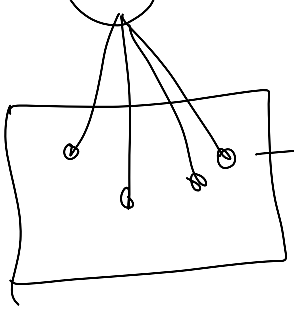


(M,N) - ?

Not knowing what was done with this image?)

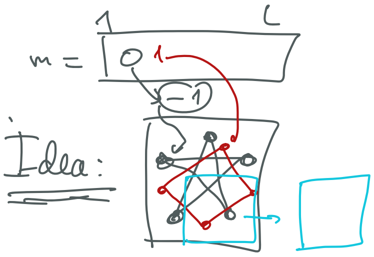
1) label A

$\underline{m} = [+1 +1 \ldots]$ L bits.



R times ≡ For redundancy. (errors of prediction in view of interference with $\underline{x}$).

Idea:

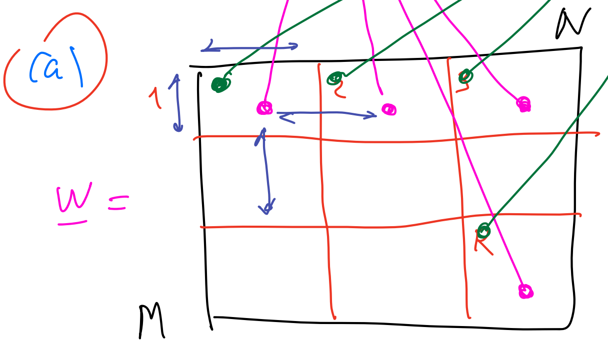


2). Embedding

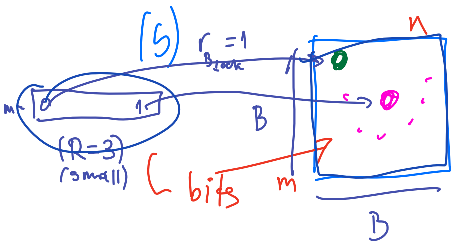
• $\underline{m} = [+1 +1 -1 \ldots -1]$

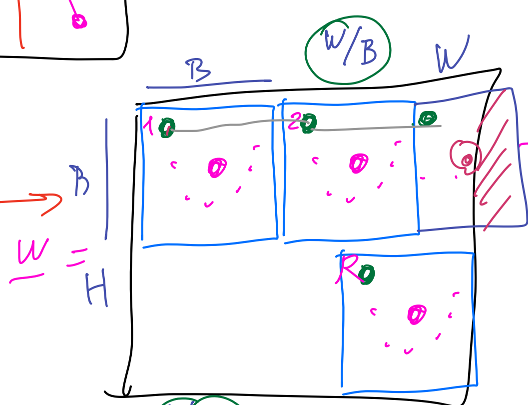
• R times.

(ordered placement).



R-blocks.

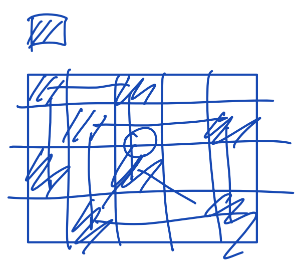




repeated R times

$H/B \times W/B = R$

$$\boxed{\underline{y} = \underline{x} + \underline{w}}$$



Extraction

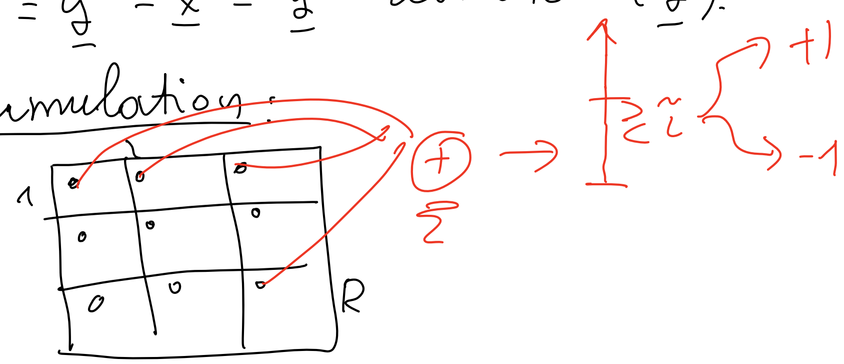
No geom. attacks

1) ▷ $\underline{y} = \underline{x} + \underline{w}$

2) Prediction of $\hat{\underline{w}}$:

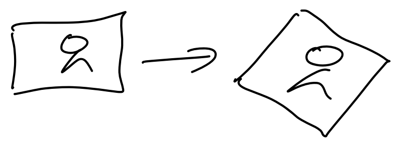
$$\hat{\underline{w}} = \underline{y} - \hat{\underline{x}} = \underline{y} - \text{denoiser}(\underline{y}) \quad | \ \text{wiener2}(.)$$

3) Accumulation:



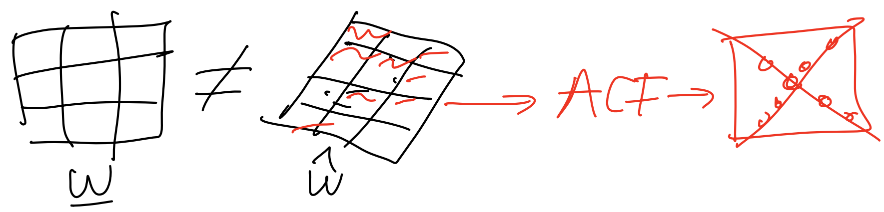
With geom. attacks

1) $y = G(\underline{x} + \underline{w})$

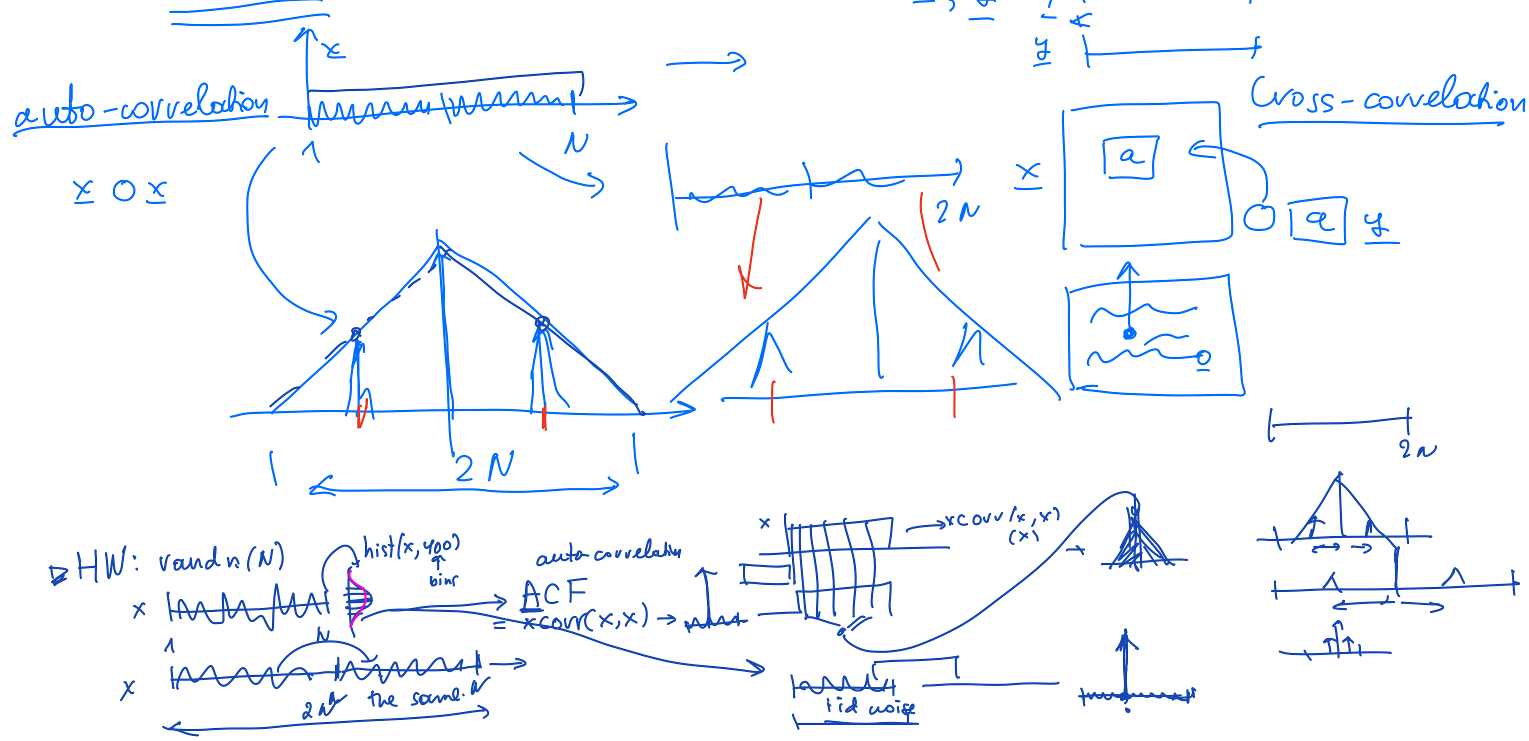


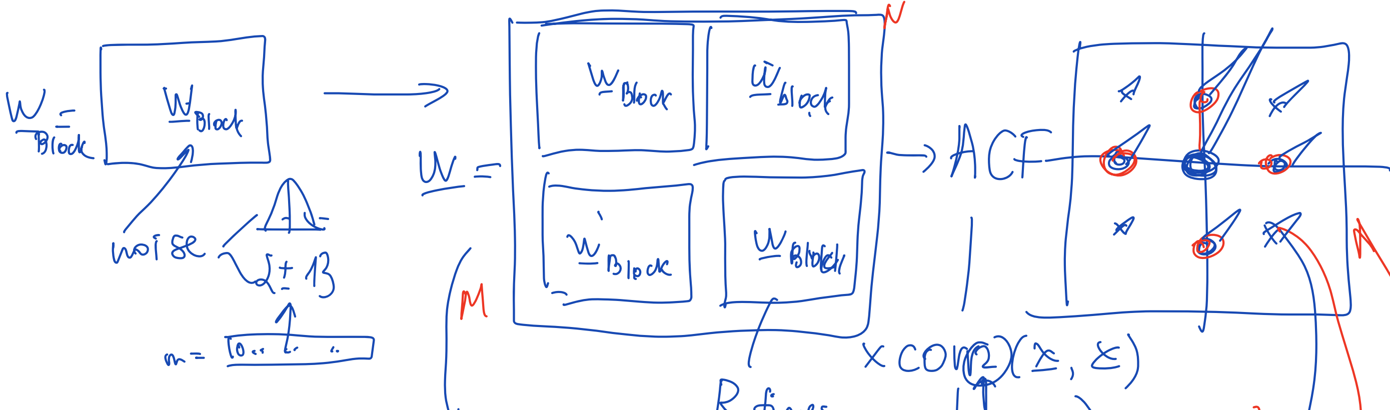
2) Prediction

$$\hat{\underline{w}} = \underline{y} - \hat{\underline{x}} = \underline{y} - \text{denoiser}(\underline{y}) \quad | \ \text{wiener2}(.)$$

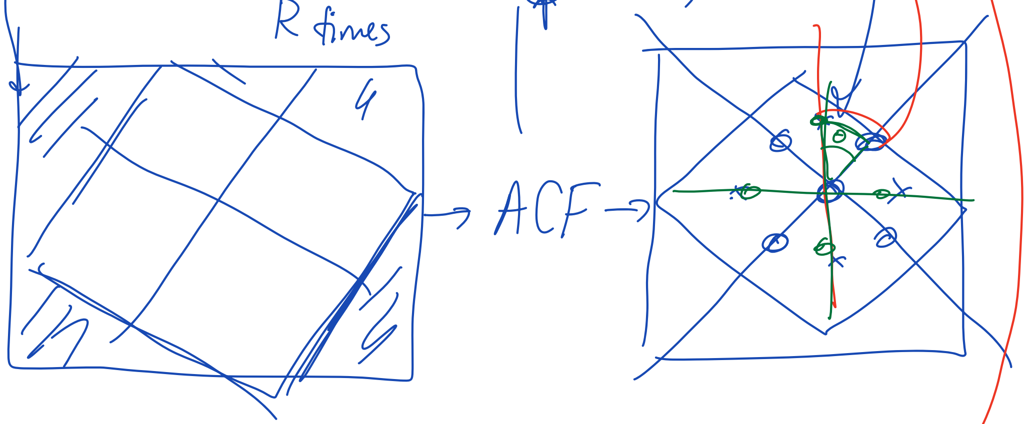


▷ Recall:



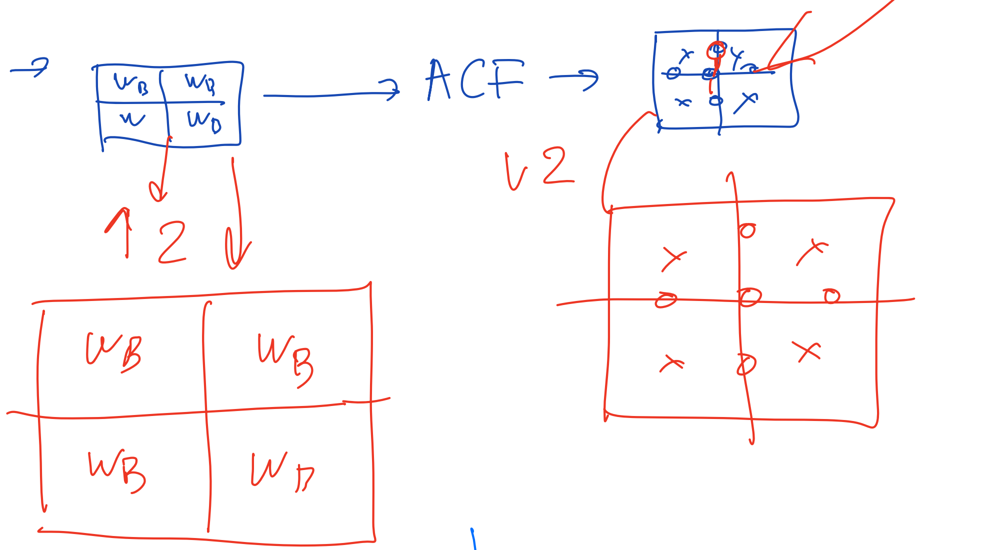


▷ rotate(W, 45)



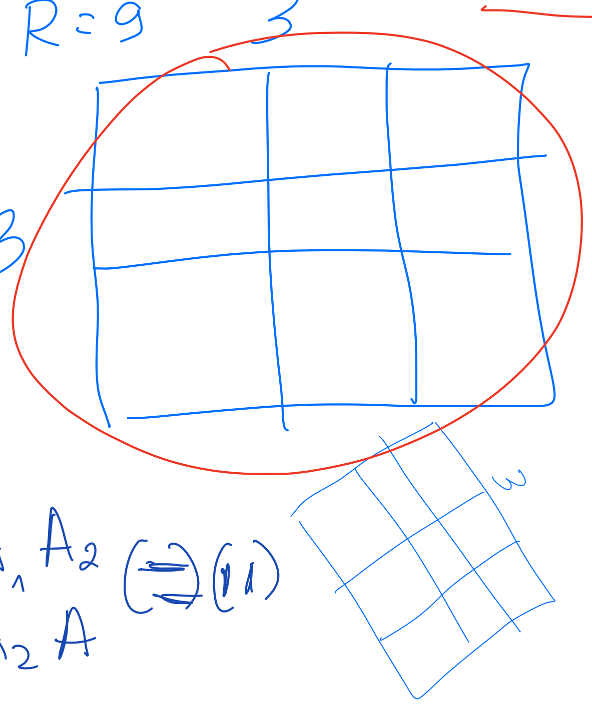
▷ <u>scale</u>

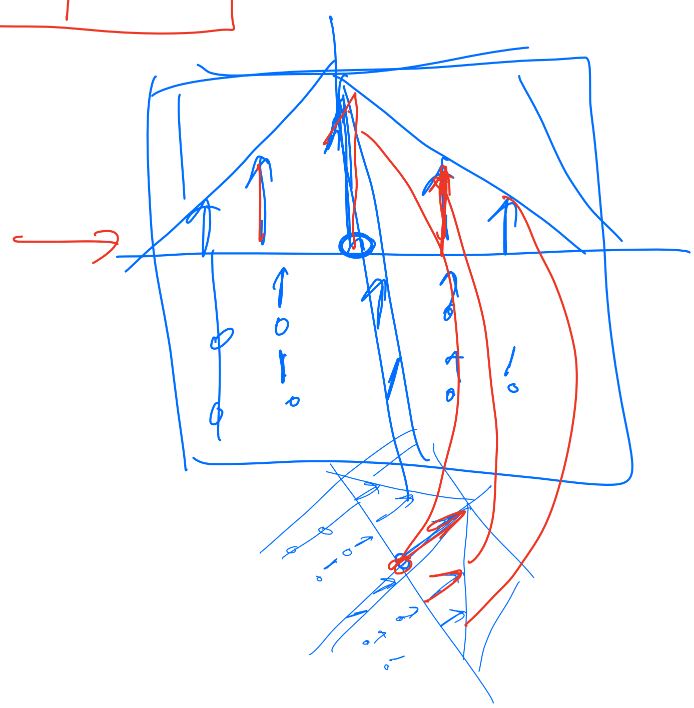
imresize(W, 0.5^2) →



Ex: R = 9

W 3





$A = A_1 A_2$ (≠) (?) $A_2 A$

DWM

$y = f(x, m, k)$

Additive DWM

$y = x + w_{(m,k)}$

ACF

geometric synchronization

Quantization based DWM (MIT)

$y = Q_m(x, k)$

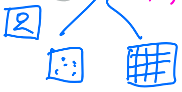
FP

Digital   Physical

Privacy Protection

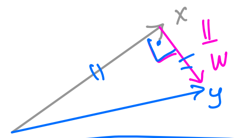
↓

Biometrics

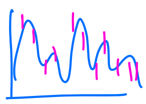


$X \to [x_1, x_2, \dots, x_N] \to x$

$p(x, w) = p(x) \cdot p(w)$



$\text{s.t. } \|x\|_2^2 \gg \|w\|_2^2$



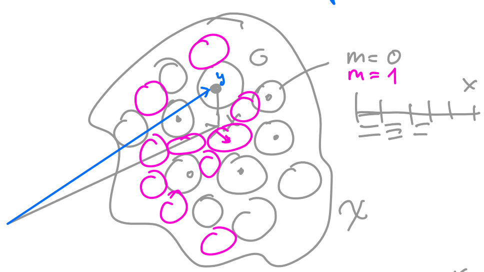
(a) denoising: $\sigma_x^2 = 0 \to$; $\sigma_x^2 \to \infty \to$

$$\boxed{E[\|w - \hat{w}\|_2^2] = N \cdot \frac{\sigma_x^2 \sigma_w^2}{\sigma_x^2 + \sigma_w^2}}$$

(b) R-repetitions → ↑ robustness

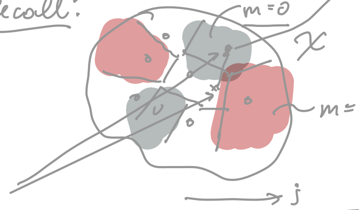
↑ geomer. synch.

1) Embedding



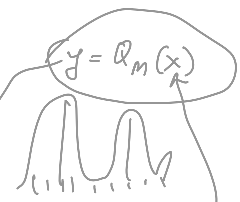
Recall:

• K-means



$y = Q_m(x)$

$y = f(x, m, k)$



$K \to K \sim (i, j)$

pixel → $x(i, j)$ [0 … 255]

• Distortion-compensated Quantization Index Modulation (DC-QIM) [Joakim Eggers] PhD. Bernd Girod.

↳ Teddy Furon ⊕ Hervé Jegou.

• Product Quantization
↳ fast search.
FAISS.

Ex: 0, m=1, 156, 144, 255, m=0

Addive $y = Q(x)$ QIM

$y = Q_m(x, k)$

$\propto Q_m(x)$

$= x + \alpha(Q_m(x) - x)$

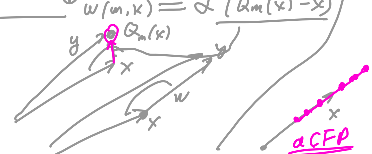
$\alpha = 1$: $y = Q_m(x)$ (QIM)

$0 < \alpha \le 1$: DC-QIM.

$y = x + w(m, x)$

$\|w\|_2^2 \le T$

$w(m, x) \triangleq \alpha(Q_m(x) - x)$

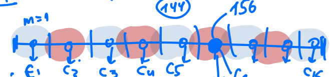


2) Extraction

$y = $ … $y(i_1, j_1)$

$K \to (i_1, j_1)$

• Secret (K) (wo noise)



$\Rightarrow m=0$

$\Rightarrow m=1$

$\mathcal{C}_{Q_0} = \{c_1, c_3, c_5, c_7 \dots\}$

$\mathcal{C}_{Q_1} = \{c_2, c_4, c_6 \dots\}$

Secret.

• Decoding:

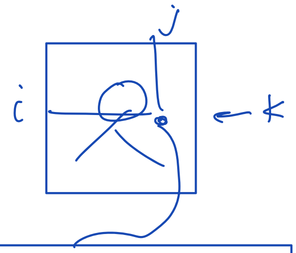
$\hat{Q}_0 \leftarrow \min_{c \in \mathcal{C}_{Q_0},\ c \in \mathcal{C}_{Q_1}} [|y - c|]$ 156

m=1

Multi-bit DWM

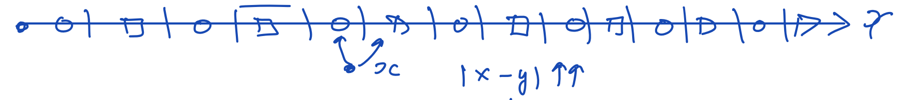
$N = 1$

○ — $m = 0$
□ — $m = 1$



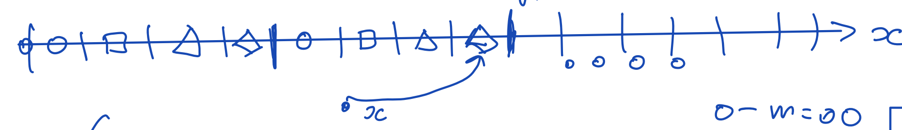
1-bit (a)

Dist. $|x - y|$



$|x - y| \uparrow\uparrow$

2-bit (b)



○ — $m = 00$
□ — $m = 01$
△ — $m = 10$
◇ — $m = 11$

$m = 11$

△ → $D_{nis} = |x - y|$

$Q_m(\cdot) \leftarrow x$
→ 0/1 (a)
→ 00, 01, 10, 11 (b)

D dist = as before → robustness ↓

▷ To increase robustness for the same Ddist,

m = ○

aggregate {majority voting} → {0 0 0 0 1 0 0 1} → 0
1 2 3 4 5 6 7 8

R = 8

hist[x, 256]

K ↑ → ~~large enc.~~

Ddist ↓ Security

robustness

redundancy

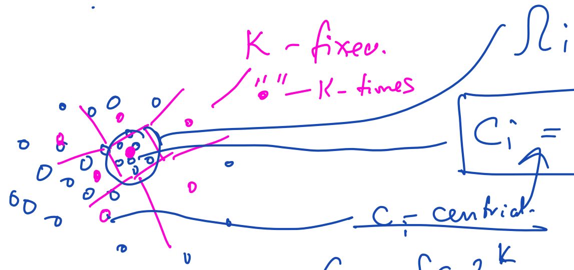
"Formal intro"

Product quantization (PQ).

proposed by H. Jegou (2010).

- QIM (DC-QIM).
- K-means
  - Data Science
  - Compression.
  - fast indexing.

1) $D = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix}$ (rows $1 \dots N$), $x_i \in \mathbb{R}^{d_x}$ (dim → d_x)



$\boxed{C_i = \frac{1}{K_i} \sum_{j \in \Omega_i} x_j}$ = "mean" of region.

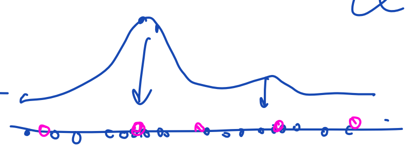
c_i - centroid.

$\mathcal{C} = \{c_i\}_{i=1}^{K}$

in DWM.

- K-means
- based on key.
  - K-means
  - based on key. (K-NN)

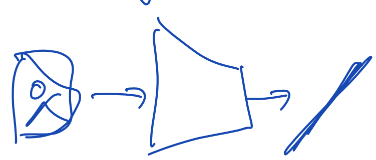
$d_x = 1$



2) 

If N is large $N > 8'000$ (dead)

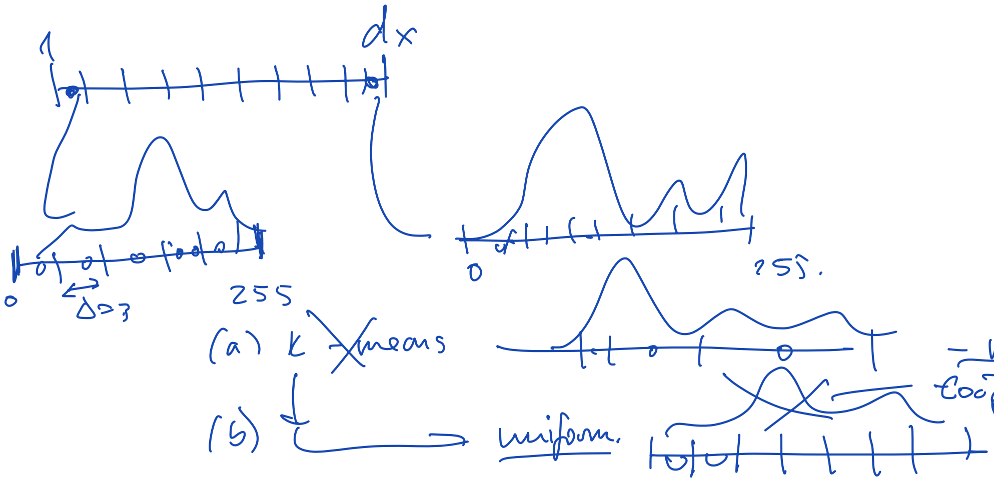
& d_x is large



"800 → 2400!"

3) → Jegou → PQ.

1D.



(a) K ~~means~~ — non uniform.

— Coupled

(b) → uniform. — uniform

o Problem of ~~d~~ 1D — d_x

not optimal for min D dist.

1D ① x 1D

1D $d_x = 1$ 0 255

Ⅱ

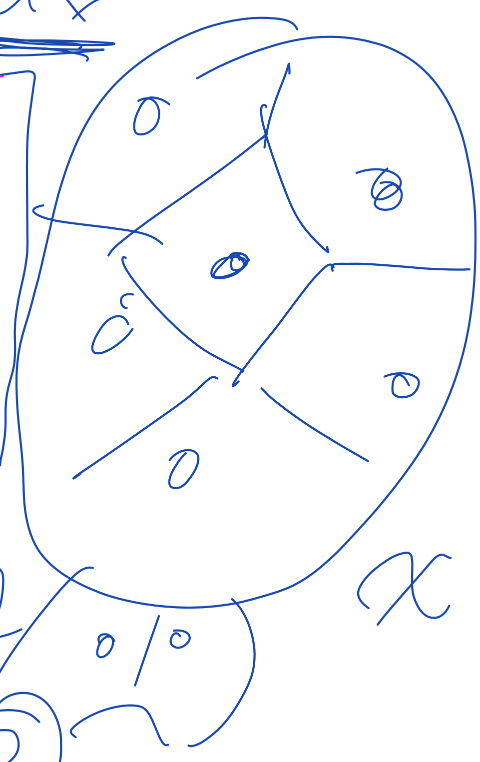
o Alter: ② d_x 1 d_x K-mea-

$x = M \cdot N$

$d_x = M \times N$

1 2 3

K



$K < 8'000$

③ PQ 1 2 ... B

L $1 \le d_x \le N$

$d_x = 8'000$

$\frac{d_x}{L} = B$

$L \le 8$ 8

K_c

FAISS

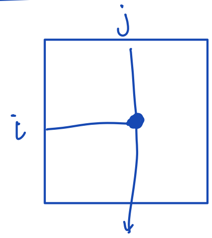
"Local" K means.

$K_c \cdot B$ — $\frac{d_x}{L} = \frac{8'000}{8} = 1000$

256

256...

2) Quantization modulation.



$x[i,j]$

$[0 \dots 255]$

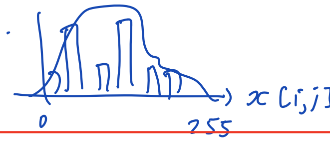
▷ additive modulation

• Direct domain:

$$\boxed{y[i,j] = x[i,j] + w[i,j] \hookrightarrow (\underline{m}, \underline{K}).}$$

▷ quantization modulation

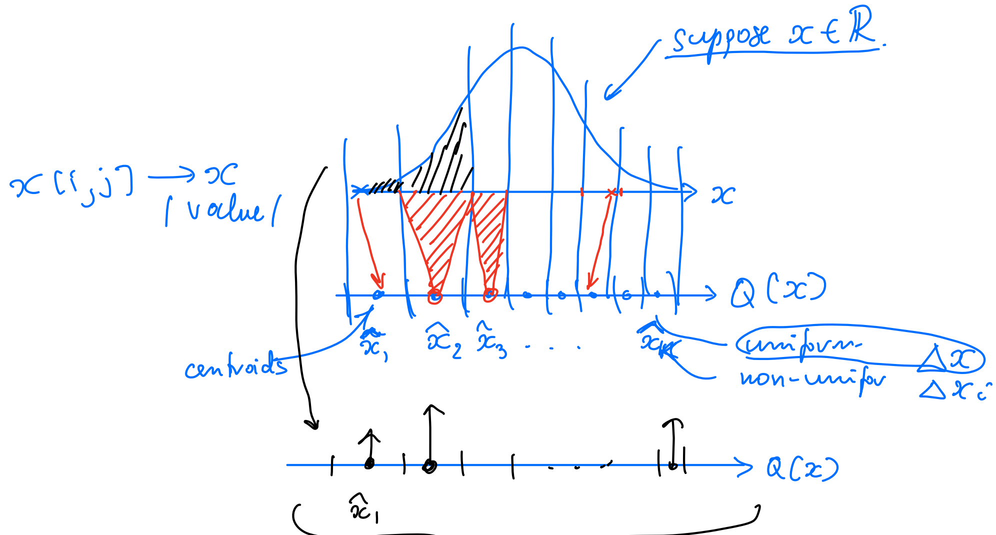
• histogram.



$$\boxed{y[i,j] = \begin{cases} Q_0[x[i,j]], & \text{for } m = 0, \\ Q_1[x[i,j]], & \text{for } m = 1 \end{cases}}$$

▷ where $Q_0[\cdot]$, $Q_1[\cdot]$ for the quantizer for bit 0 or bit 1, respectively.

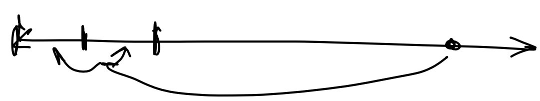
▷ Recall: quantization



quantized:

$$\boxed{Q[\cdot] : \underset{x}{\mathbb{R}} \longrightarrow \underset{\hat{x}}{\mathcal{O}}}$$

$\left[Q : \left\lfloor \frac{x}{\text{quant table}} \right\rceil \right.$.



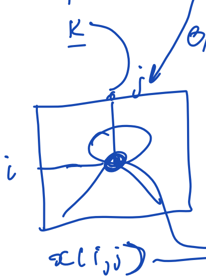
▷ Modulation

$m = [0, 1, 0 \ldots]$

$x \rightarrow \boxed{f_K(x, m)} \rightarrow y$

K

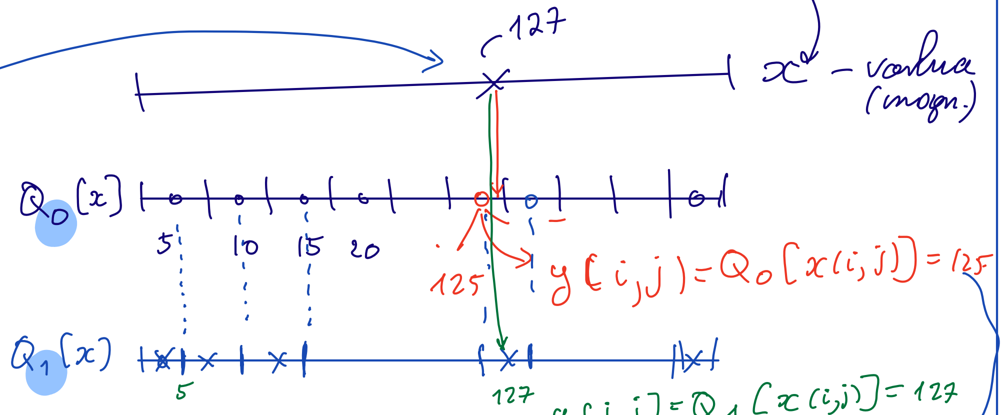
0/1



$x(i,j) = 125$

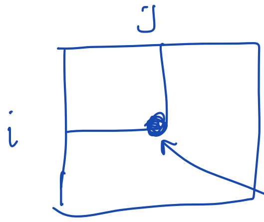
$m = 0$: $Q_0[x(i,j)]$

$m = 1$: $Q_1[x(i,j)]$

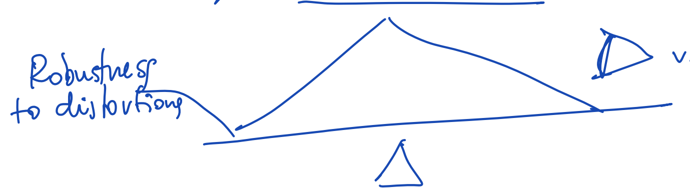


$y(i,j) = Q_0[x(i,j)] = 125$

$y[i,j] = Q_1[x(i,j)] = 127$

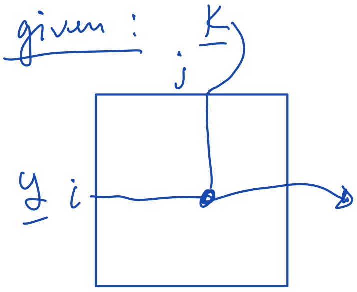


Selection of interval Δx



▷ Extraction.

∘ given : K



value $\equiv y = 125 \rightarrow \hat{m}$ (bit) $\rightarrow$ 0 / 1

(a) (no noise)

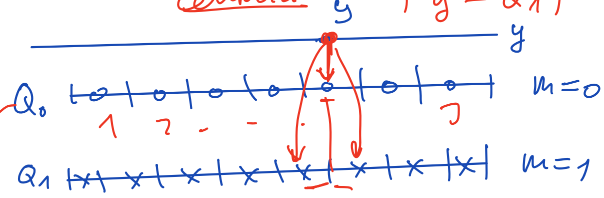
(LSB).

centroid

$j = 1, \ldots, J$

$|y - Q_0^j|$

$|y - Q_1^j|$



$\hat{m} = 0.$

▷ Decision rule :

$$\boxed{|y - Q_0^j| \gtrless |y - Q_1^j|}$$

0 / 1

Additive:

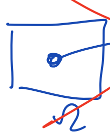
flat: T → w ⇒ $\hat{m}$ → accurate.

e/t T → $\hat{m}$ → not very accurate

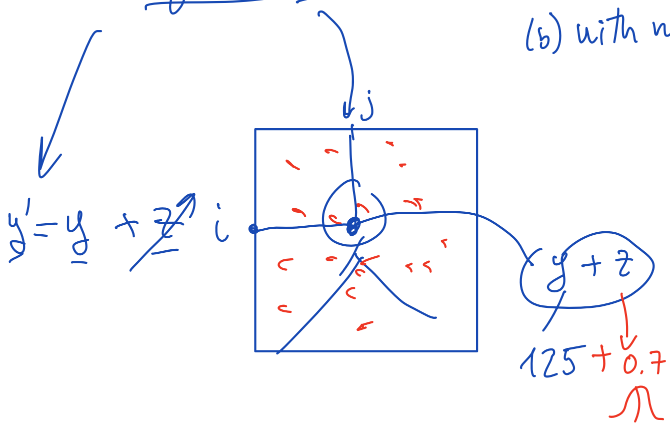
Q: is it important x is flat / e & t. ?

if is

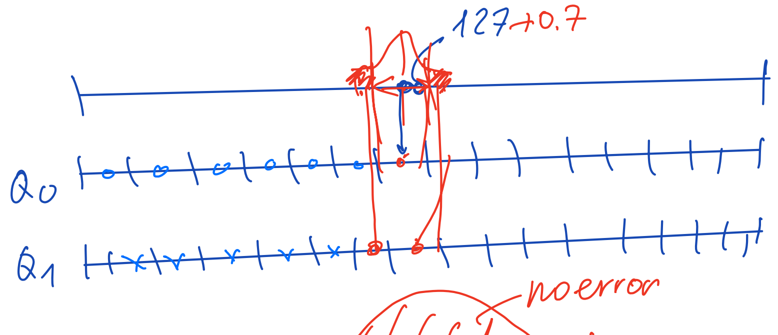
~~$\hat{x} \rightarrow y - \hat{x} = \hat{w}$ ↳ Q → 0 / 1~~



• given K

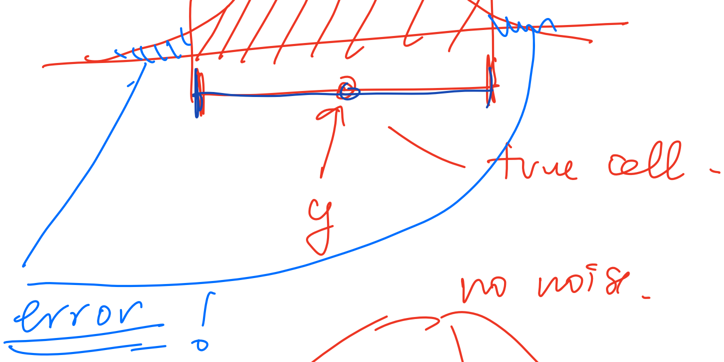


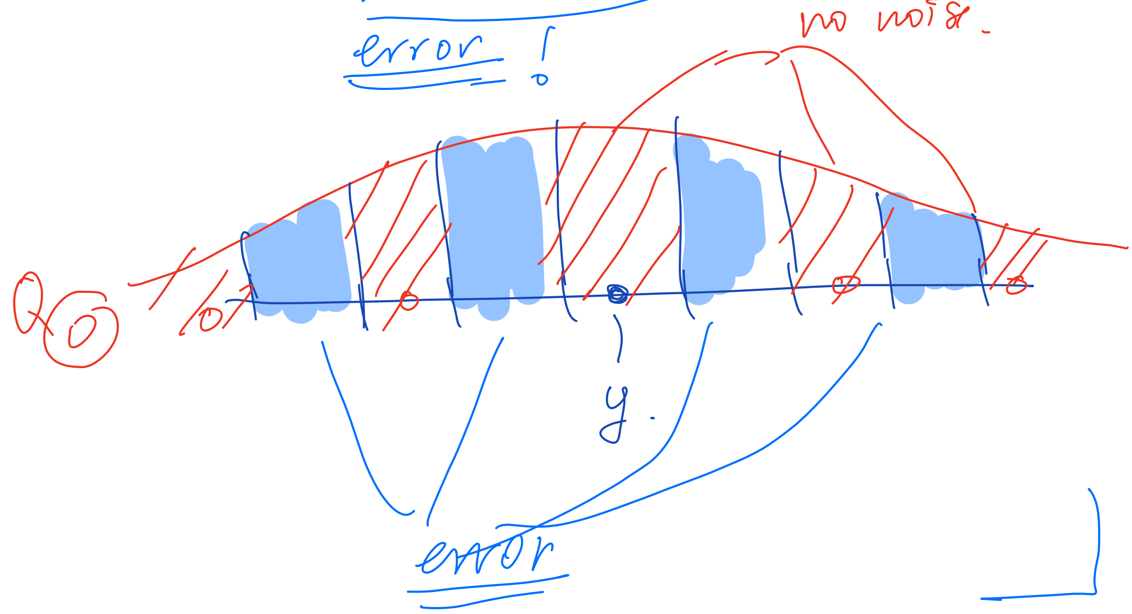
(b) with noise.



▷ Robustness.

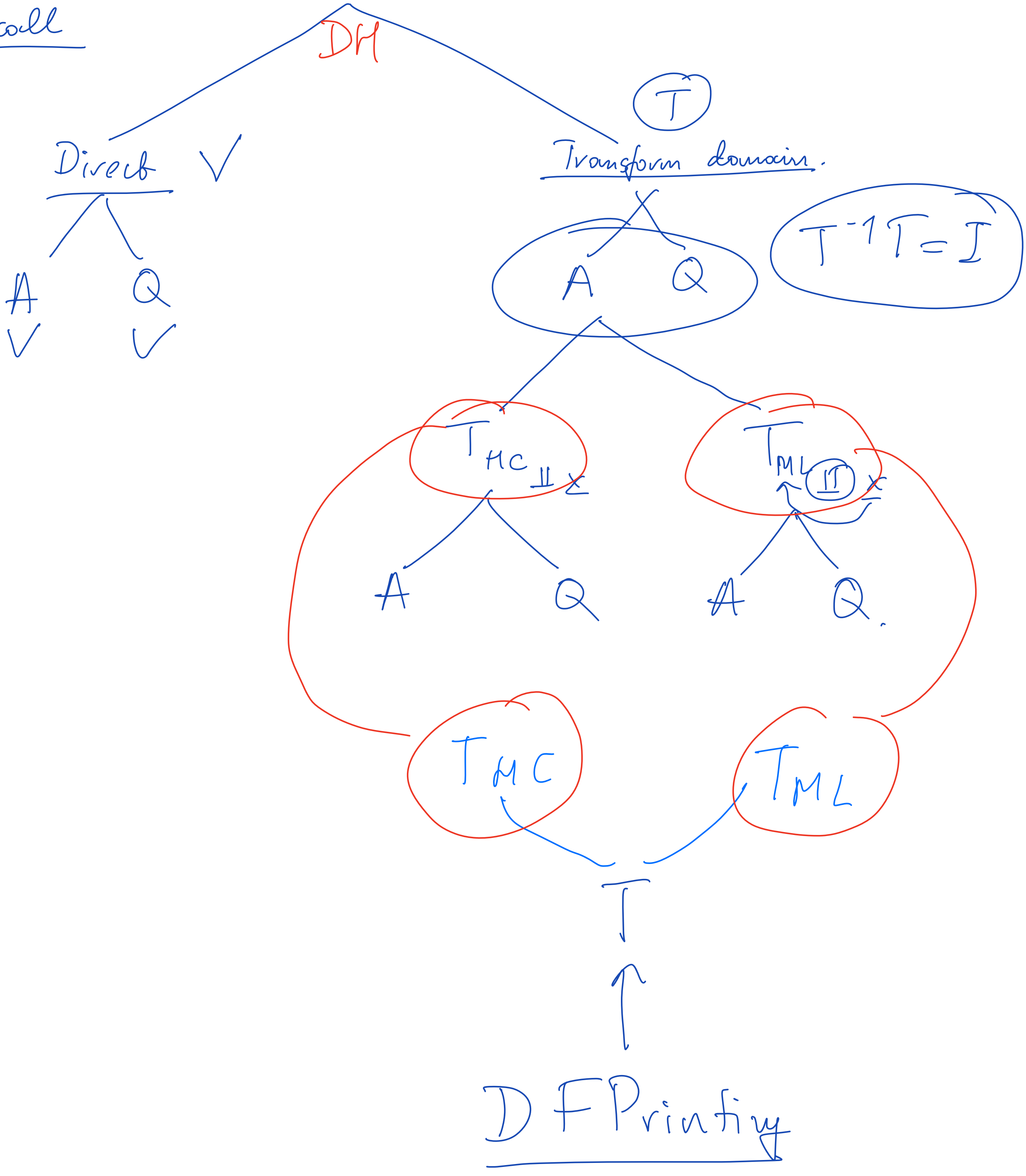
1





▷ LSB: a part. case of quant. modulation.

① Recall


DH
Direct ✓
A ✓
Q ✓
T
Transform domain.
A
Q
$T^{-1}T = I$
T_{MC} II ≤
T_{ML} II ≤
A
Q
A
Q.
T_{MC}
T_{ML}
T
DFPrinting

1) Transforms (Defender)

$$\underline{x} \longrightarrow \boxed{T} \longrightarrow \tilde{\underline{x}} \longrightarrow \boxed{W} \longrightarrow \hat{\underline{x}}$$

K (into T), K (into W)

- direct transform
- encoder

- inverse transform
- decoder

stop

▷ DWM (DDH)

$\underline{x} \xrightarrow{f(\cdot)} \tilde{\underline{x}} \xrightarrow{f(\cdot)} \hat{\underline{x}}$

▷ DFPting

$\underline{x} \xrightarrow{f(\cdot)} \begin{bmatrix} | \\ | \end{bmatrix} \Rightarrow \hat{\underline{x}}$ ✓

$L < MN$

no need in inver.

[T.H. forensic analysis]

$\underline{x} \xrightarrow{f} \tilde{\underline{x}}$ – features → ~~f^{-1}~~

→ classifier

Yes, we need to chose $T \Longrightarrow W = T^{-1}$

"T / W"

HC

ML → Sect. A → see below

i) HC (hand-crafter = engineered) transform

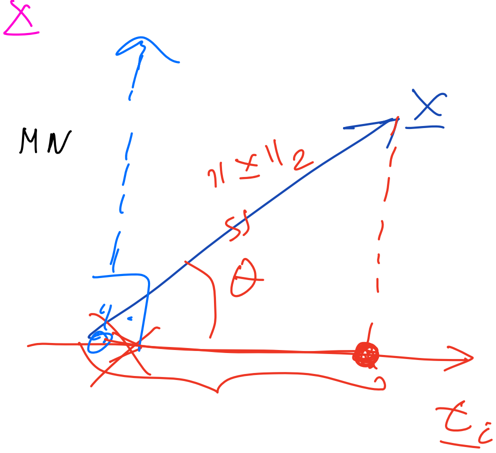
▷ Recall. $\underline{X}_{M \times N} \to \begin{bmatrix} | \\ \underline{x} \\ | \end{bmatrix}_{MN \times 1}$

basis vectors

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_i \\ \vdots \\ \tilde{x}_{MN} \end{bmatrix} = \begin{bmatrix} | \\ \tilde{\underline{x}} \\ | \end{bmatrix}_{MN \times 1} = \underbrace{\begin{bmatrix} — t_1^T — \\ — t_2^T — \\ \vdots \\ — t_{MN}^T — \end{bmatrix}}_{T} \begin{bmatrix} | \\ \underline{x} \\ | \end{bmatrix}_{MN \times 1}$$

$$\boxed{\tilde{\underline{x}} = T\underline{x}}$$

$$\boxed{\tilde{x}_i = t_i^T \underline{x} = \underline{x}^T t_i = \sum_{j=1}^{M \cdot N} x_j \cdot t_{ji}.}$$



▷ HC Transform: a selection of T (resp. W).

▷ "manually" → expertise of person.

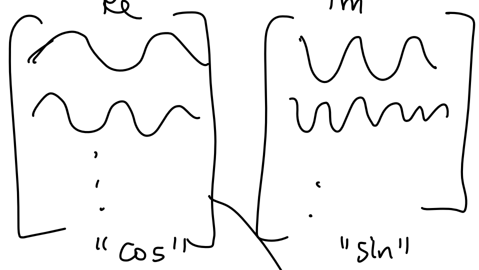
▷ selection is $\perp$ to data $\underline{X}$. (no ML)

◦ In practice:

T to be:

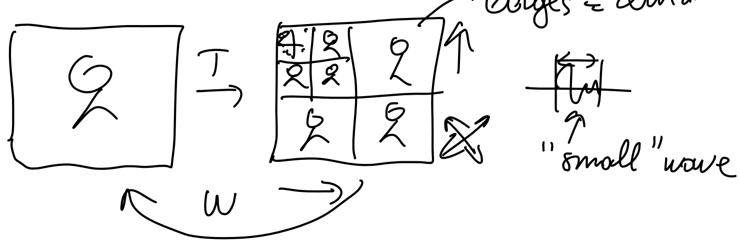
invertable

◦ DFT



◦ DCT

◦ DWT ← wavelet



◦ Hadamard.

Invertability

◦ it depends on the construction of matrix $T_{L\times MN}$

◦ Random projection

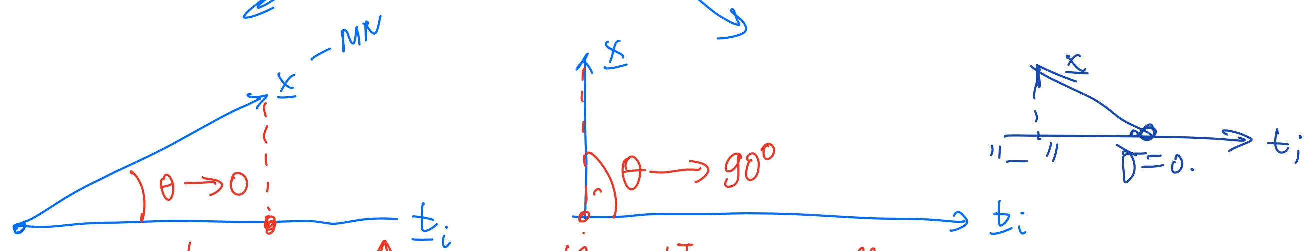
$\mathcal{N}(0, I_{MN})$

$$T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_{MN}^T \end{bmatrix} \begin{bmatrix} | \\ x \\ | \end{bmatrix}$$

① $T \perp X$

② $T \leftarrow$ keep K.
◦ seed (K).
(security).

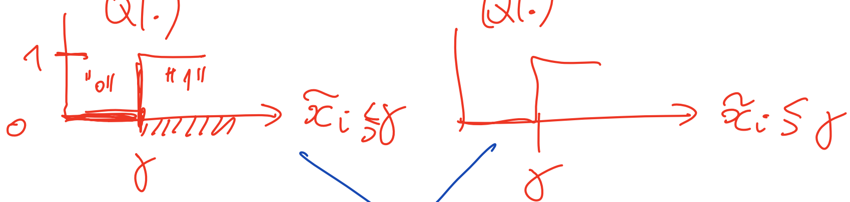
random & $\perp$ x.

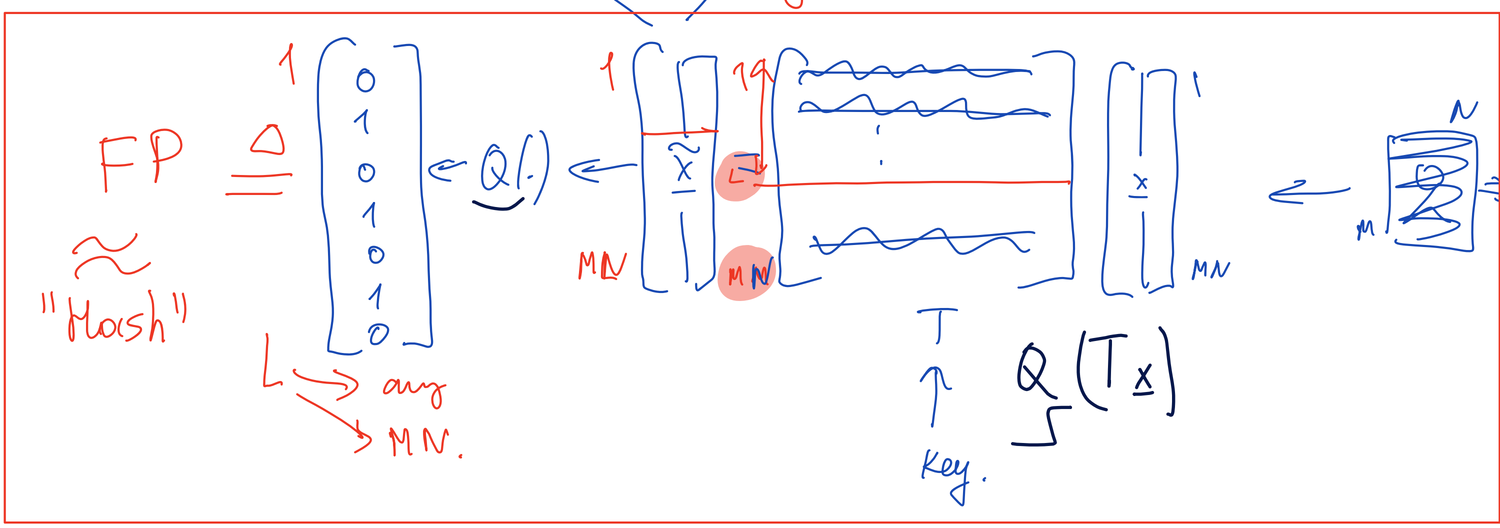


$\tilde{x}_i = t_i^T x$ ↑ large

$\tilde{x}_i = t_i^T x \to$ small.

$Q(.)$ $\quad$ $Q(.)$



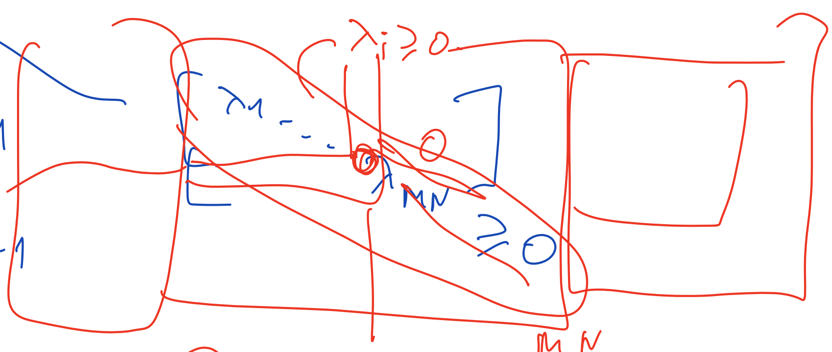


• $L < ML \to T \longrightarrow T^{-1}$ — ?

SVD: $T = U S V^{-1}$

$T^{-1} = (\quad)^{-1}$

$= V S^{-1} U^{-1}$



① $L < MN$ to avoid a perfect invertability

② $\tilde{x} \to Q(\tilde{x}) \to 1$, $\to 0.$

Defender (FP)

▷ Attacker : $Q(\tilde{x}) \longmapsto \underline{\hat{x}} \longleftrightarrow \underline{x}$.

• Invertability

MC (math) ☹

ML

15:13

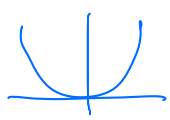
"SVD":

→ approximate solution (exact).

$\boxed{\tilde{x} = T x}$ — linear. equations.

$\varphi(x)$ or $\mathcal{L}(x) = \frac{1}{2}\|\tilde{x} - Tx\|_2^2$

$$\boxed{\begin{aligned}\hat{x} &= \underset{x\in\mathbb{R}^{MN}}{\operatorname{argmin}}\ \mathcal{L}(x)\\ &= \underset{x}{\operatorname{argmin}}\ \underbrace{\frac{1}{2}\|\tilde{x} - Tx\|_2^2}_{}\quad \ell_2 \neq\end{aligned}}$$



$\nabla_x \mathcal{L}(x) = 0.$

W_θ

$$\boxed{\hat{x} = \boxed{(T^T T)^{-1} T}\,\tilde{x}}$$

$T^{-1} \to (T^T T)^{-1}$

• $\{x_i, Q(\tilde{x}_i)\}_{i=1}^N$

y_i; $L = T x_i$

$x_i \to T \to \tilde{x}_i \to Q \to y_i$

1 → $\begin{bmatrix}0\\1\\1\end{bmatrix}_L$ →

2 → ⋮ →

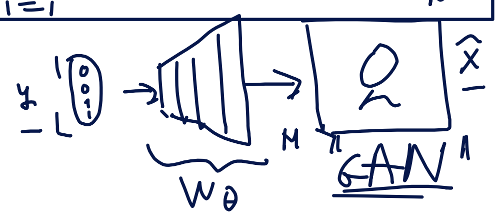
⋮

N → $\begin{bmatrix}1\\ \vdots\\0\end{bmatrix}_L$ →

x y

? Key-? $\to W_\theta \to \hat{x}$

$$\boxed{\mathcal{L}(\theta) = \sum_{i=1}^{N} \|x_i - W_\theta(y_i)\|_2^2}$$



"GAN"

$$\boxed{\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\ \mathcal{L}(\theta)}$$

$L \sim MN$

☹

① T not invertable $L \ll MN$.

② $Q(\tilde{x}) \to$ loss of inform.

→ stoch. (M.A.P.)

$$\mathcal{L}(x) = \frac{1}{2}\|\tilde{x} - Tx\|_2^2 + \lambda\,\underbrace{\Omega(x)}_{\text{prior}}$$

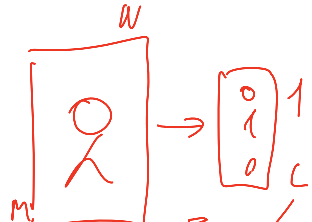
$\Omega(x)$

ℓ_2: $\|x\|_2 \to$ ridge

ℓ_1: $\|x\|_1 \to$ lasso.

$\nabla_x \mathcal{L}(x) = 0$. .

$Q(\tilde{x}) \to$ No



• Not feasible. for the attacker. ☹

▷ Sec A. T/W based on ML (Defender)

DWM /DH/ — DFP

$$\hat{X} = X$$

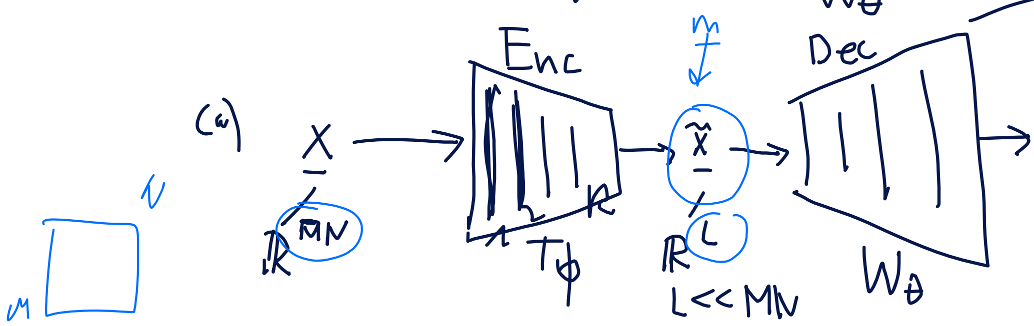
$$\underline{X} \longrightarrow \boxed{T} \longrightarrow \tilde{\underline{X}} \longrightarrow \boxed{W} \longrightarrow \hat{\underline{X}}$$

∘ ML: the transf.: $T \longrightarrow T_\phi$

$W \longrightarrow W_\theta$

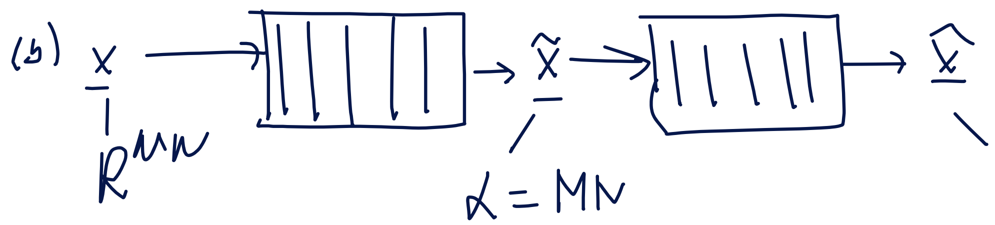
Auto-encoder (AE)

(a)



·MLP

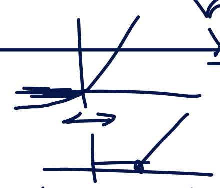
(b)



·FLOWs
·ResNet
·Unet

Ex: Enc:

$$\boxed{\tilde{X} = T_\phi(X) = \sigma_R\Big(T_R \cdots \sigma_2\big(T_2 \cdot \underbrace{\sigma_1(T_1 \underline{X} + \underline{b}_1)}_{\tilde{X}_1} + \underline{b}_2\big) + \cdots \underline{b}_R\Big)}$$



$$\phi = \{T_1, T_2, \ldots, T_{kj}, \underline{b}_1, \ldots \underline{b}_R\}$$

Dec:

$$\boxed{\hat{X} = W_\theta(\tilde{X}) = \sigma_R\big(W_R - \cdots - \sigma_1(W_1 \tilde{X} + \underline{a}_1) - \cdots + \underline{a}_R\big)}$$

▷ <u>Training</u>

$$\boxed{\mathcal{L}(\phi,\theta) = \sum_{i=1}^{N} \| \underline{x}_i - \hat{\underline{x}}_i \|_2 = \sum_{i=1}^{N} \| \underline{x}_i - W_\theta(T_\phi(x_i)) \|_2^2}$$

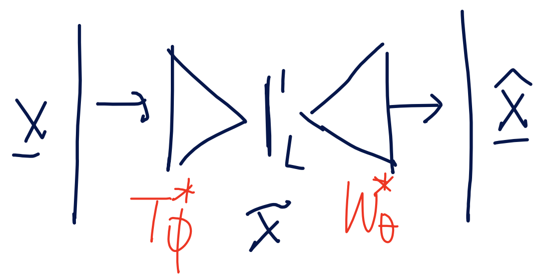
$$\hat{\underline{x}}_i = W_\theta\big(\underbrace{T_\phi(\underline{x})}_{\tilde{x}}\big)$$

1) $\nabla_{\phi,\theta} = 0$

2) (SGD) : $$\boxed{(\hat{\phi},\hat{\theta}) = (\phi,\theta)^K - \beta \nabla_{\phi,\theta}\, \mathcal{L}(\phi,\theta)}$$

<u>Remarks:</u>

1. $\#$ in $\phi,\theta \longrightarrow N \uparrow$

2. 

DWM
(T_ϕ^* & W_θ^*)

<u>DFP</u>