USC Viterbi
School of Engineering

Optimization and Learning for the Next Generation IoT + CPS

Marcos Müller Vasconcelos

mvasconc@usc.edu

mullervasconcelos.github.io

Dept. of Electrical Engineering
University of Southern California

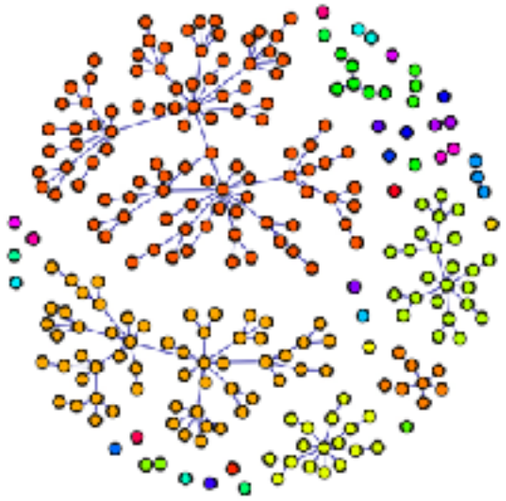
joint work with

Ubli Mitra

Commonwealth Cyber Initiative, Arlington - VA

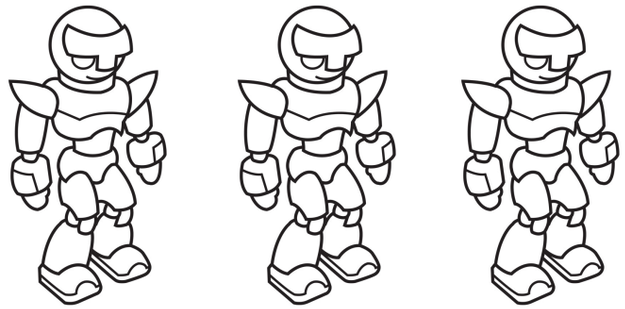
July 16th, 2020

problems of interest

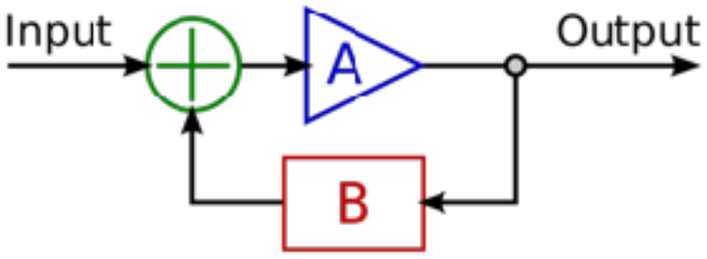


Large scale

Communication



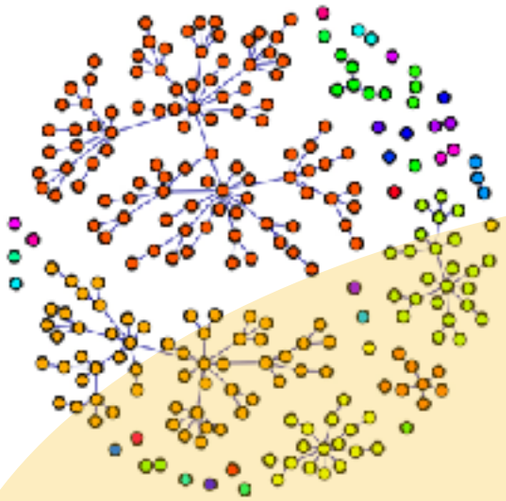
Team



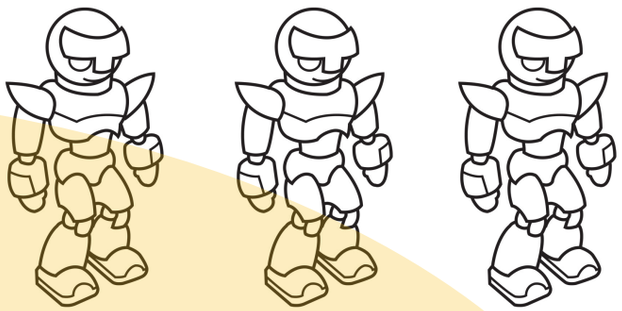
Control

Sensing

our goal

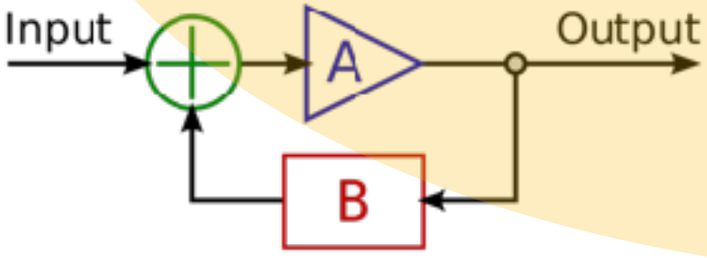


Large scale



Team

Communication

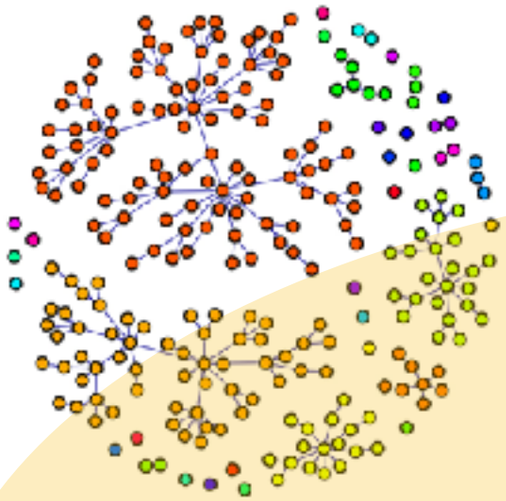


Control

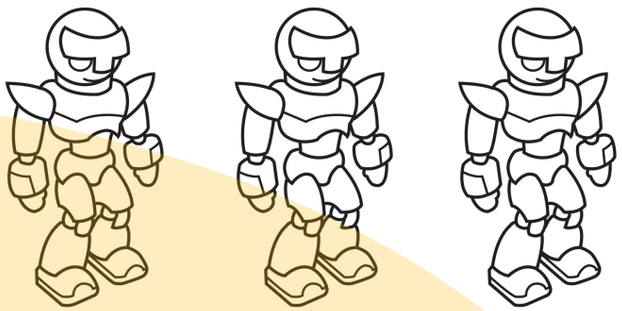
joint optimization

Sensing

our goal

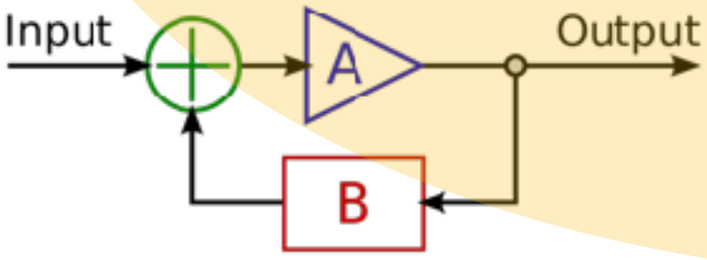


Large scale



Team

Communication



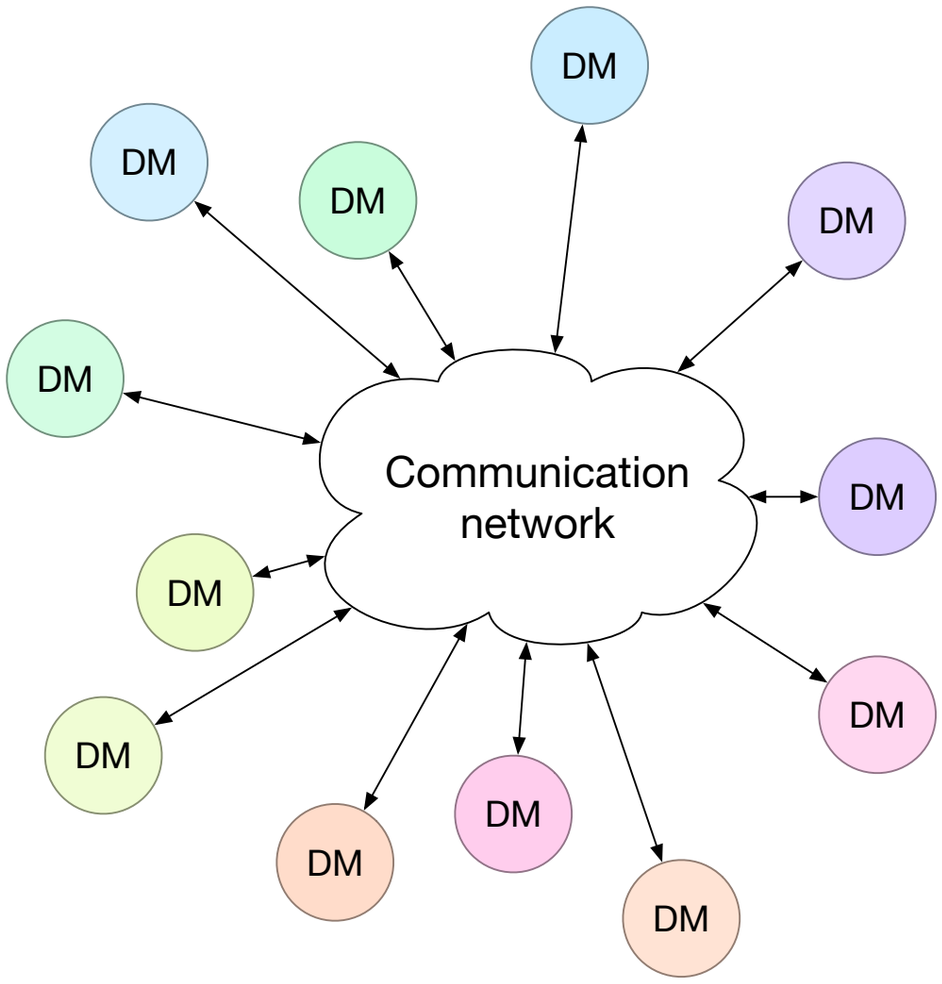
Control

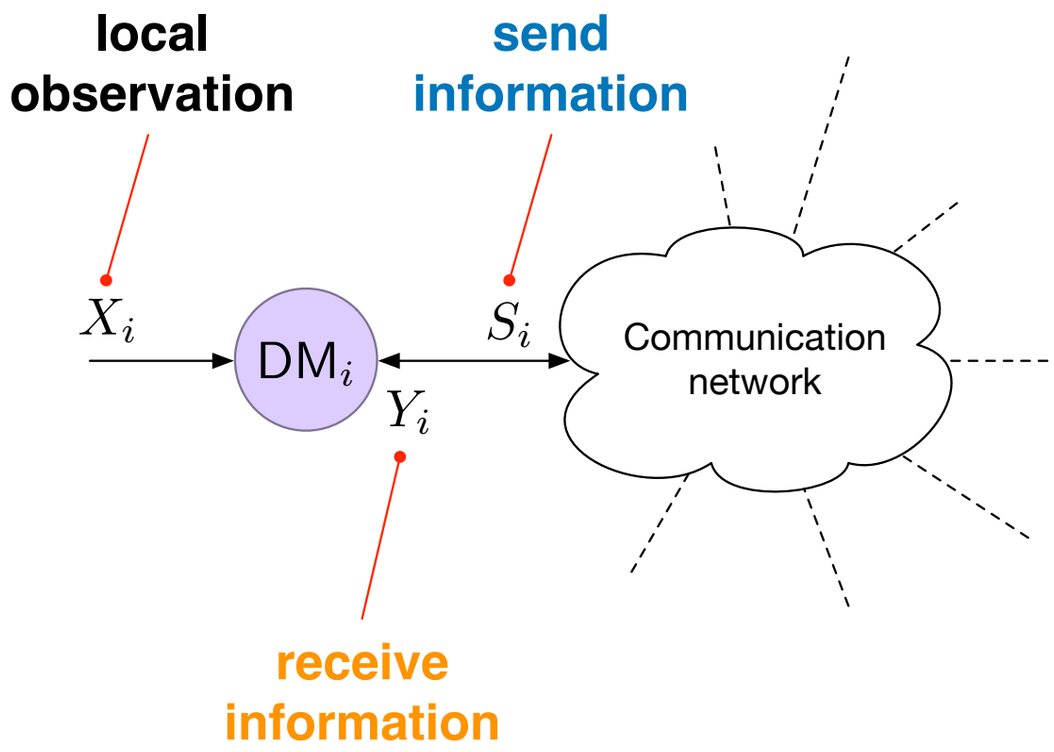
simple

implementation

Sensing

networked decision systems



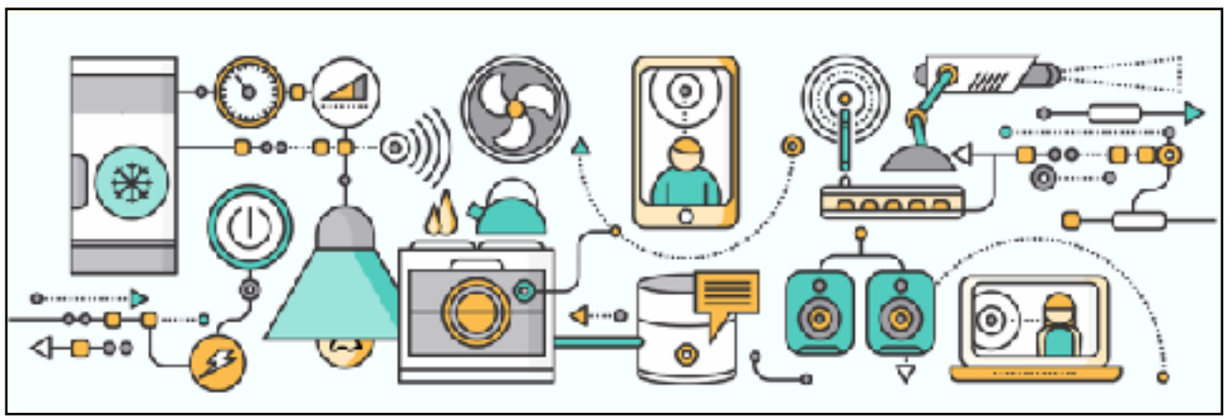


applications

1. **Wireless sensor networks**
2. **Cloud computing**
3. **Cyber-physical systems**
4. ~~**Bacterial colonies**~~

applications

Smart grid

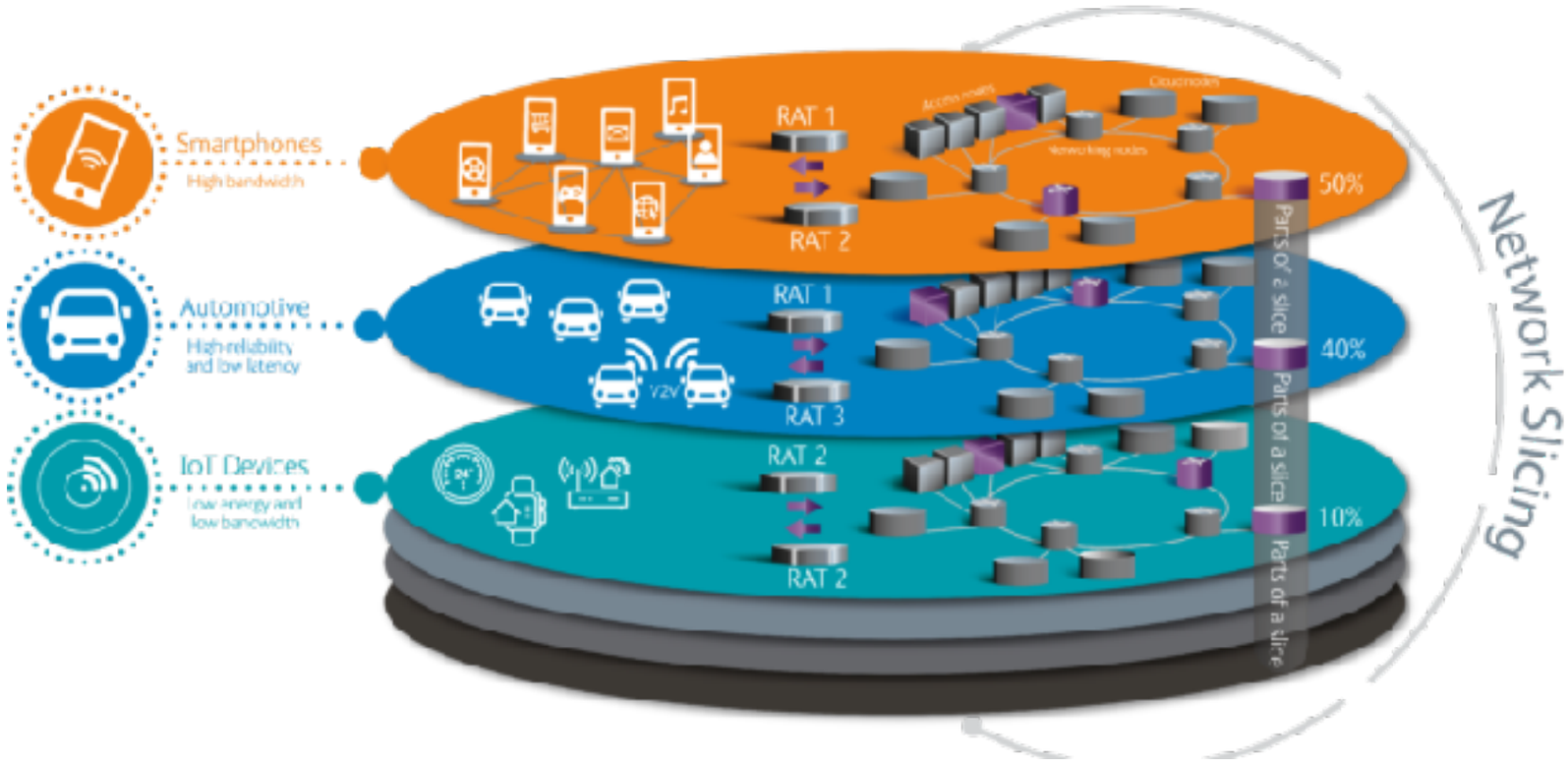


Internet of Things

Vehicle to vehicle networks

Social networks

5G - opportunities and challenges



massive connectivity (billions of devices)

very high data rates

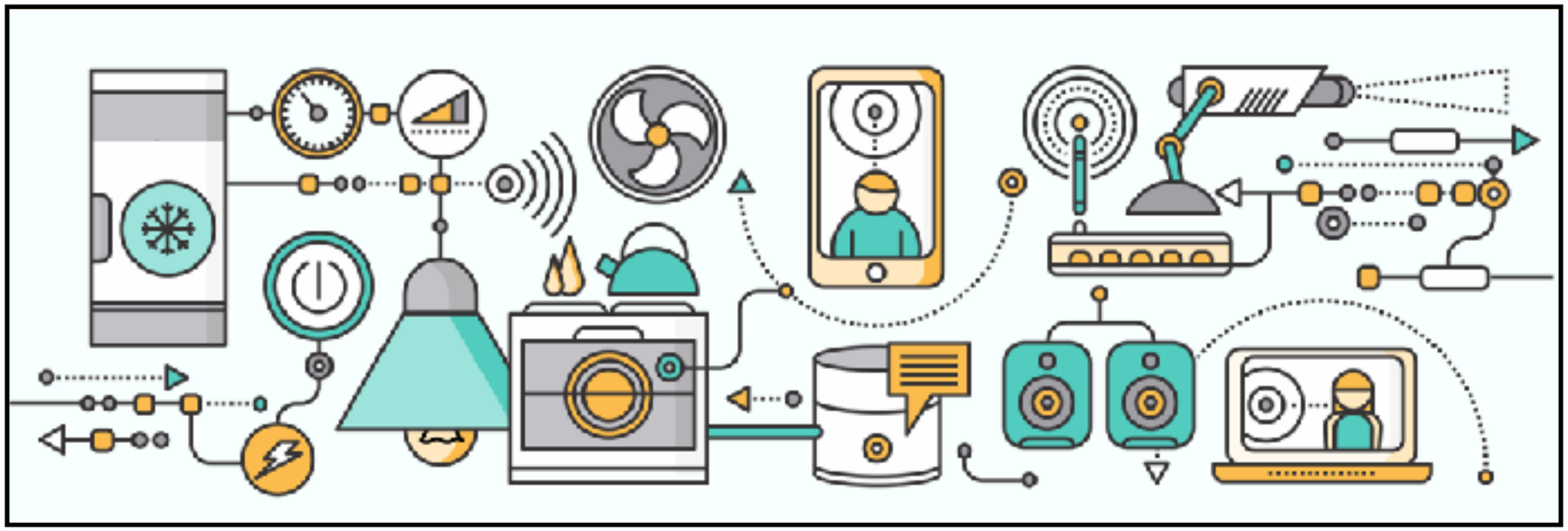
ultra reliable

mmWave implies in low range communication

costly infrastructure

massive number of devices requires resource allocation

internet-of-things



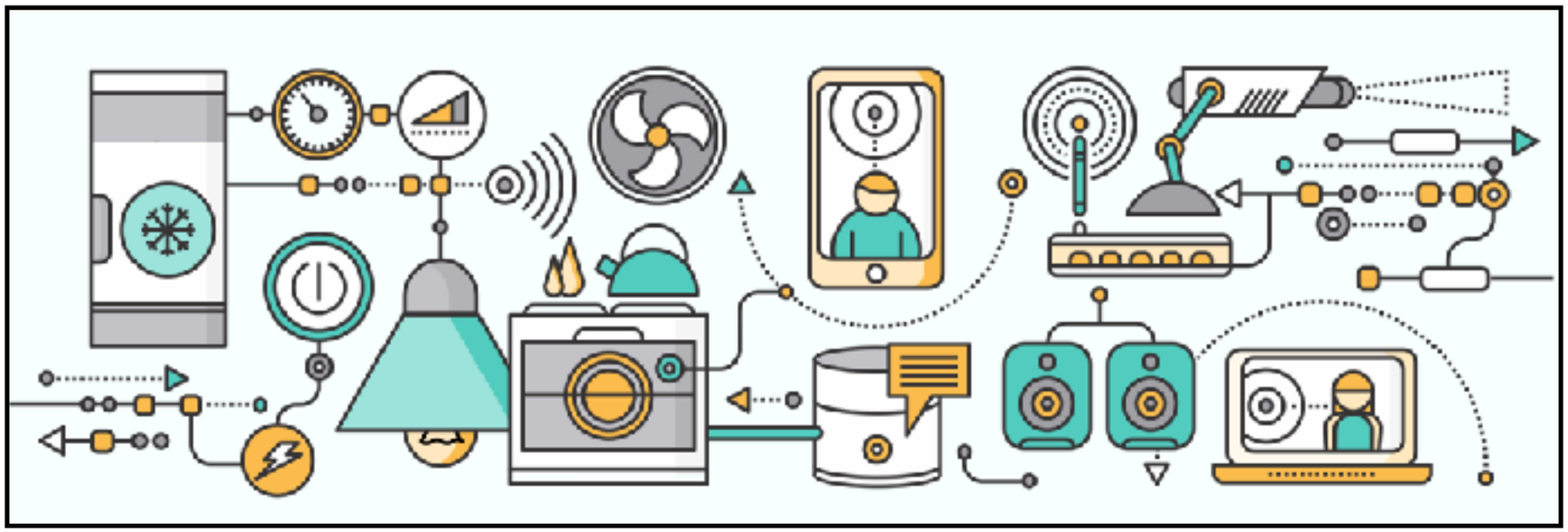
Massive number of devices sharing the network

Real-time wireless networking

Traditional MAC* schemes require feedback or introduce delay

***MAC = Medium Access Control (ALOHA, CSMA, TDMA, FDMA, etc…)**

internet-of-things

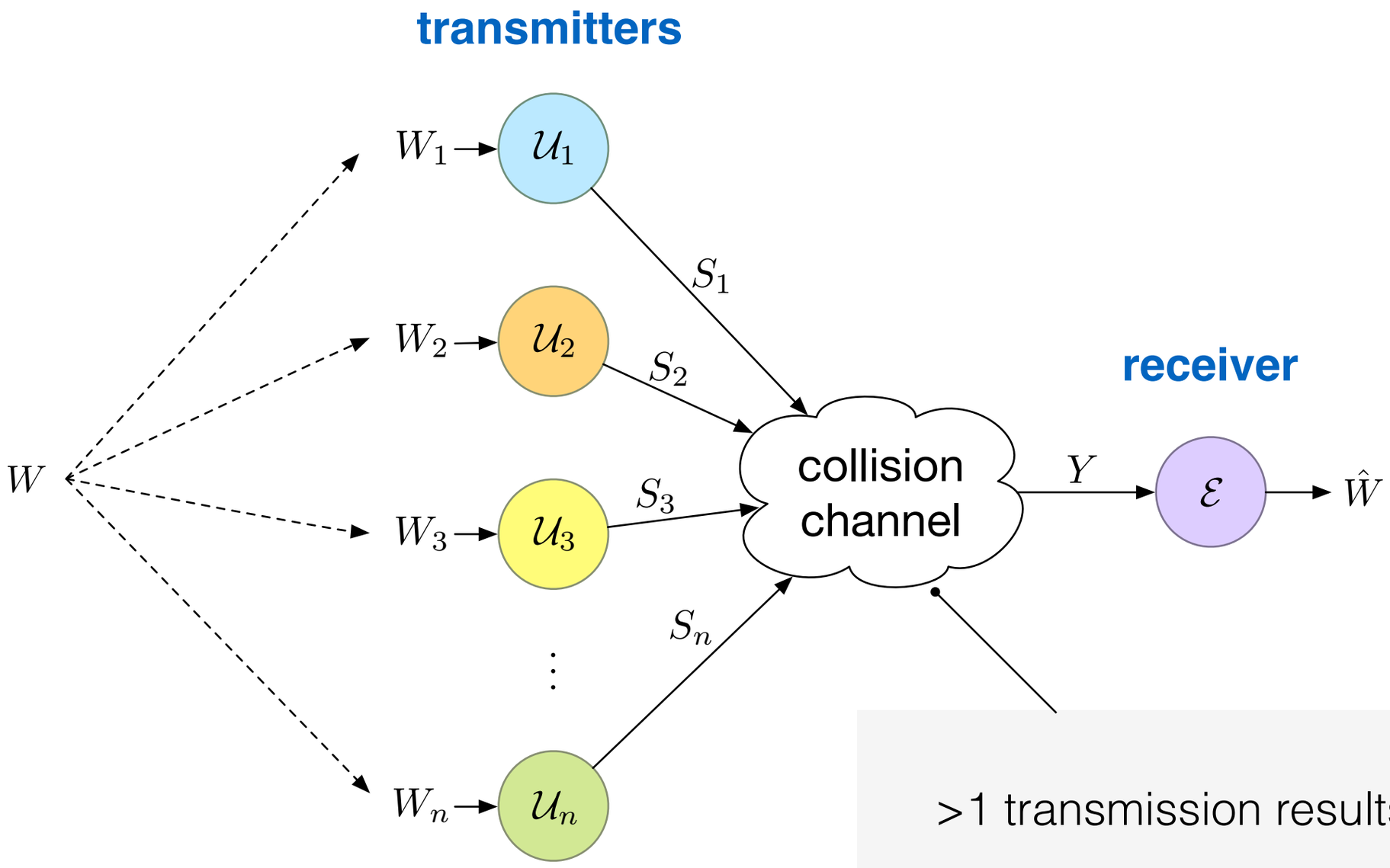


Massive number of devices sharing the network

Real-time wireless networking

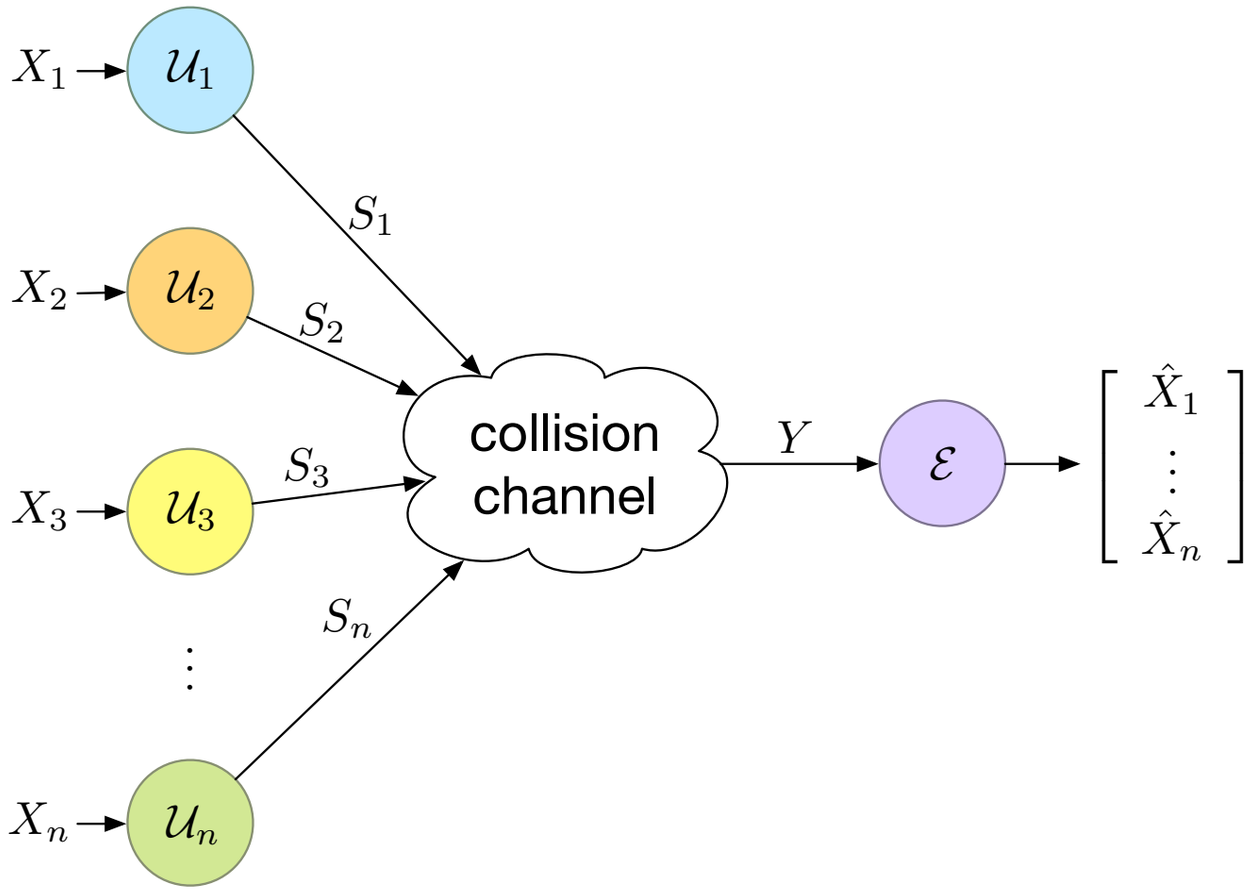
How do we rethink MAC schemes for 5G?

rethinking random access



What transmission policies will achieve optimal performance?

MMSE estimation over the collision channel



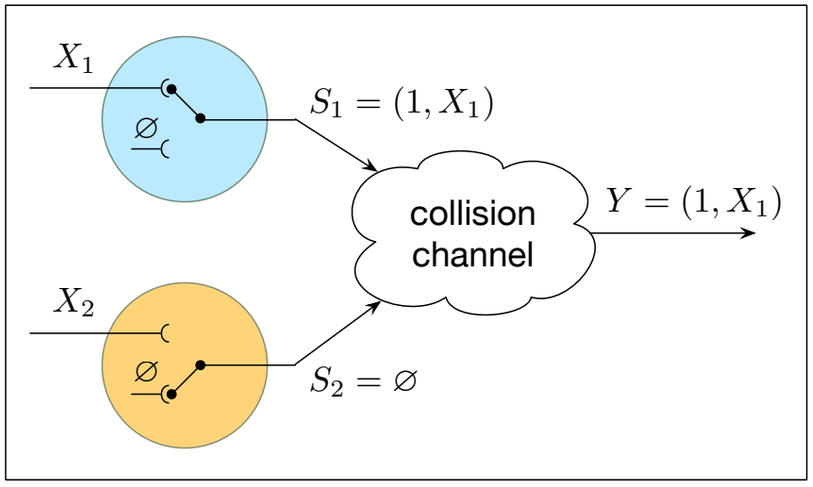
Find a strategy $(\mathcal{U}_1^\star, \cdots, \mathcal{U}_n^\star)$ that jointly minimizes the following cost

$$\mathcal{J}(\mathcal{U}_1, \ldots, \mathcal{U}_n) = \mathbf{E}\left[\sum_{i=1}^{n}(X_i - \hat{X}_i)^2\right]$$

rethinking the collision channel

single transmission

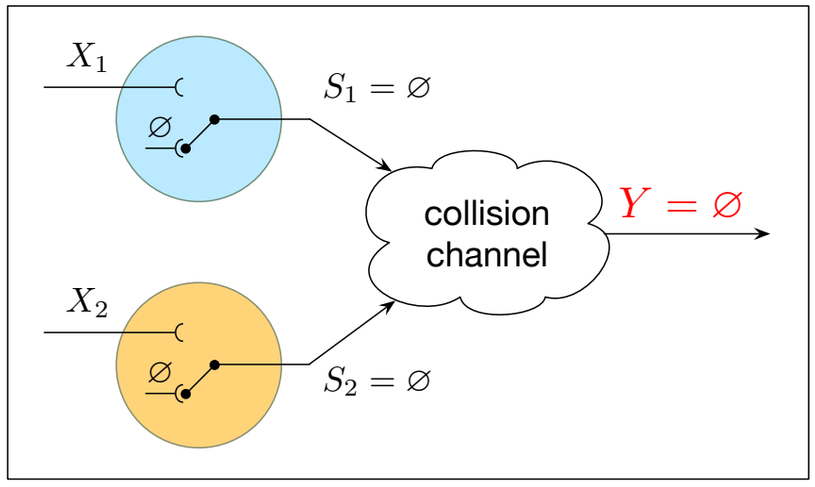
$U_1 = 1, U_2 = 0$



success!

no transmissions

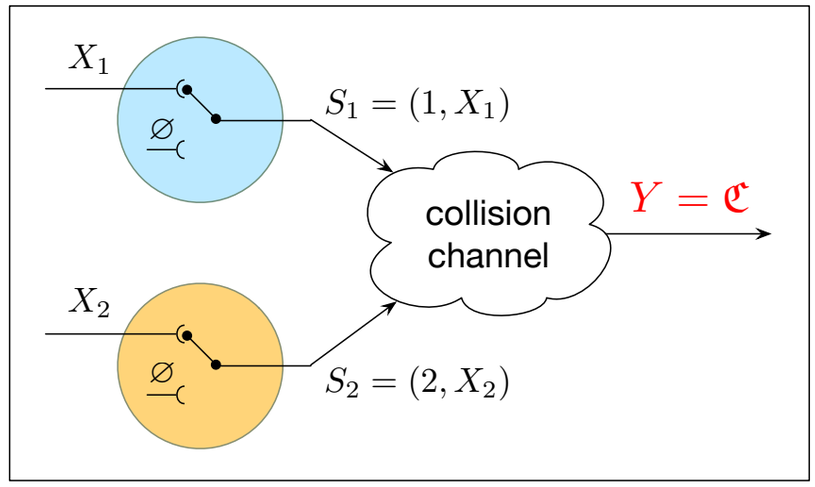
$U_1 = 0, U_2 = 0$



silence $\varnothing$

>1 transmissions

$U_1 = 1, U_2 = 1$

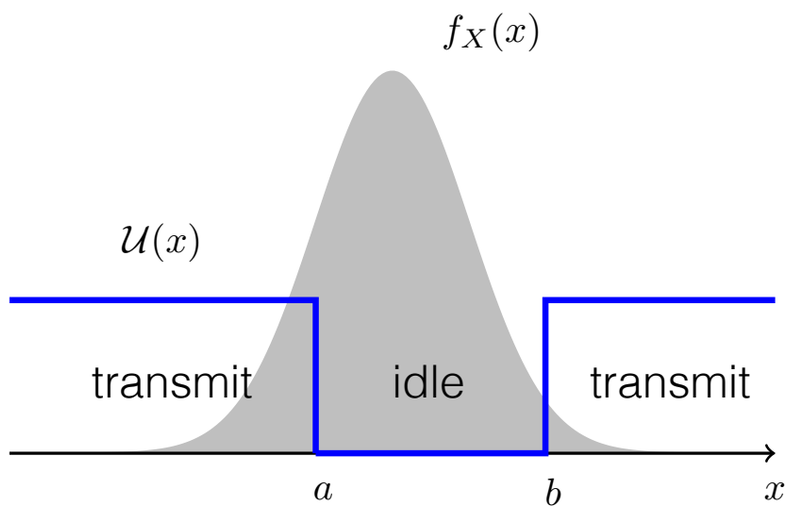


collision $\mathfrak{C}$

From the channel output we can always recover U_1 and U_2

characterization of optimal policies

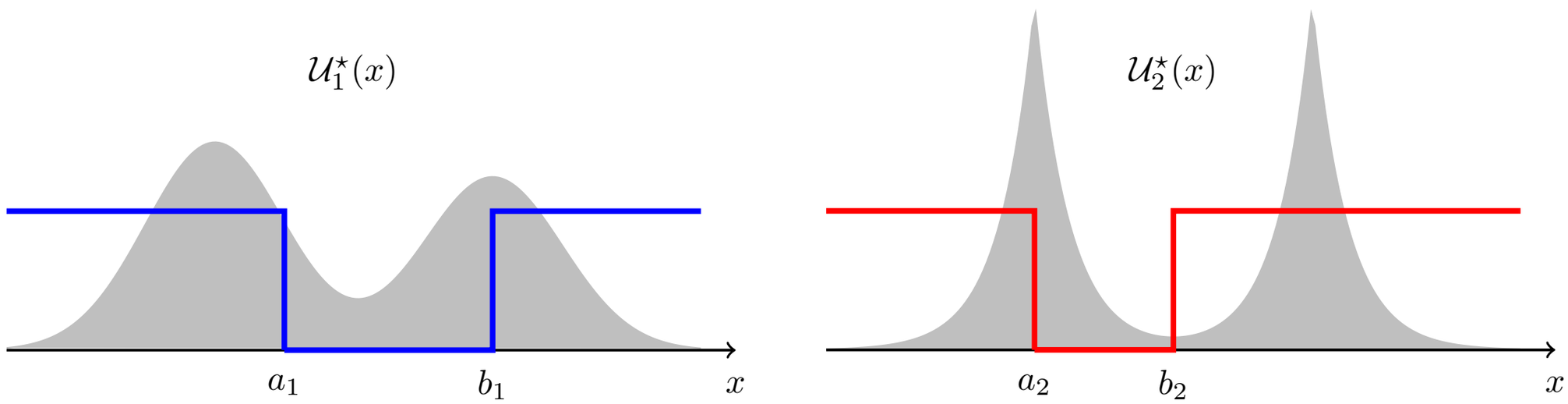
There exists an **optimal strategy** where each sensor uses a **threshold policy**



The optimal strategy implicitly uses silence and collisions to convey information

remarks

1. Valid for **any continuous probability distributions**

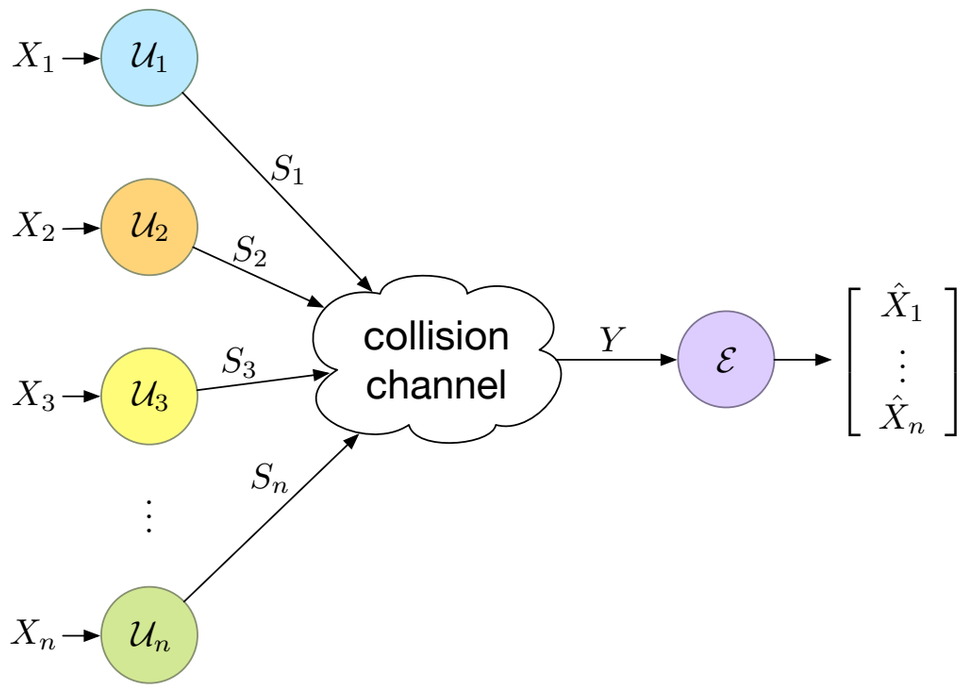


2. **Vector observations**
3. **Any number of sensors**

fusion center decodes ids of colliding packets

open question

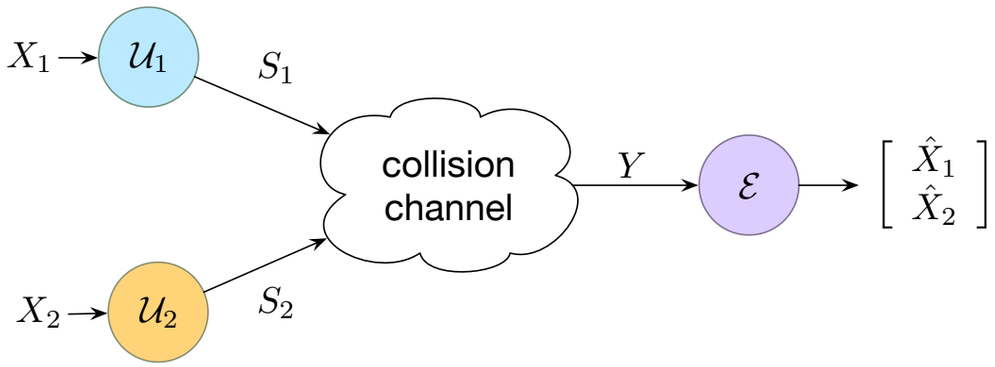
can we implement this idea?

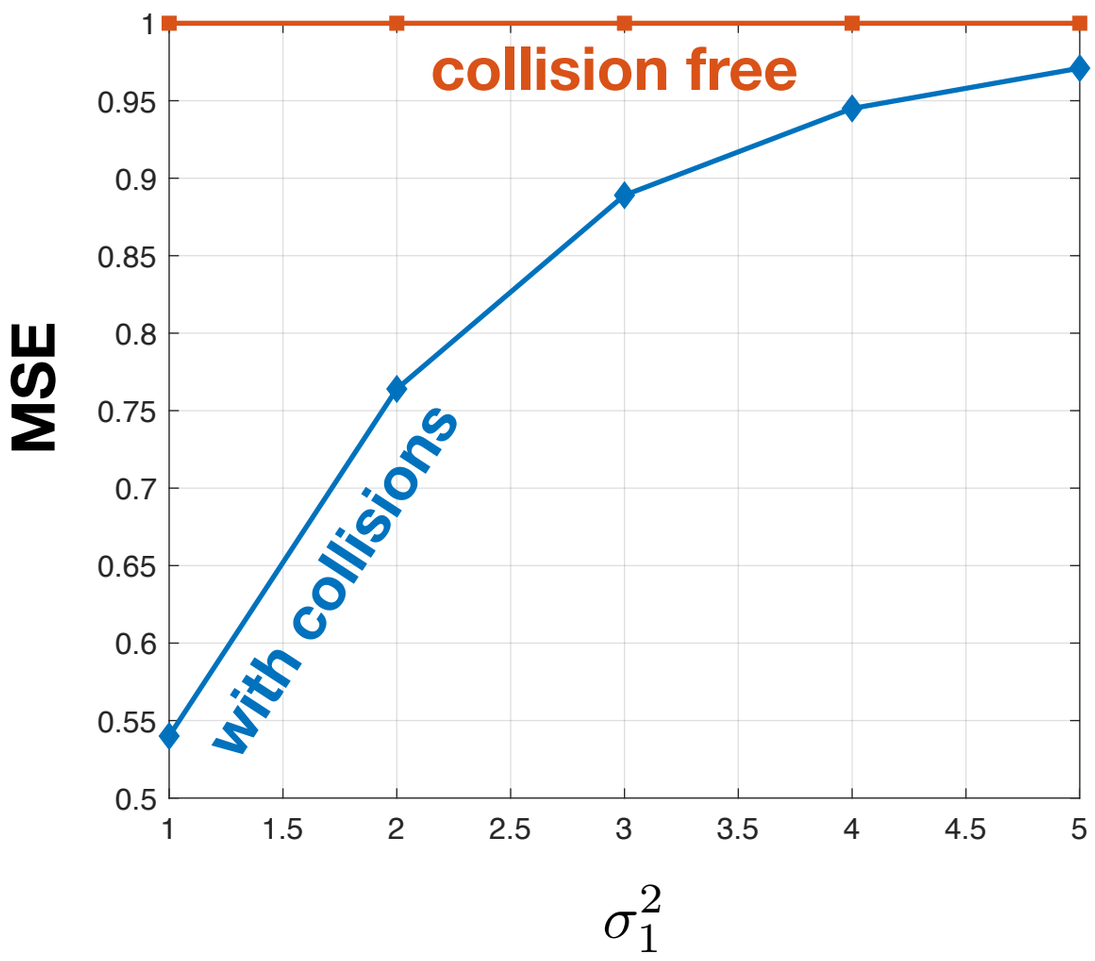


performance

$$X_1 \sim \mathcal{N}(0, \sigma_1^2)$$

$$X_2 \sim \mathcal{N}(0, 1)$$

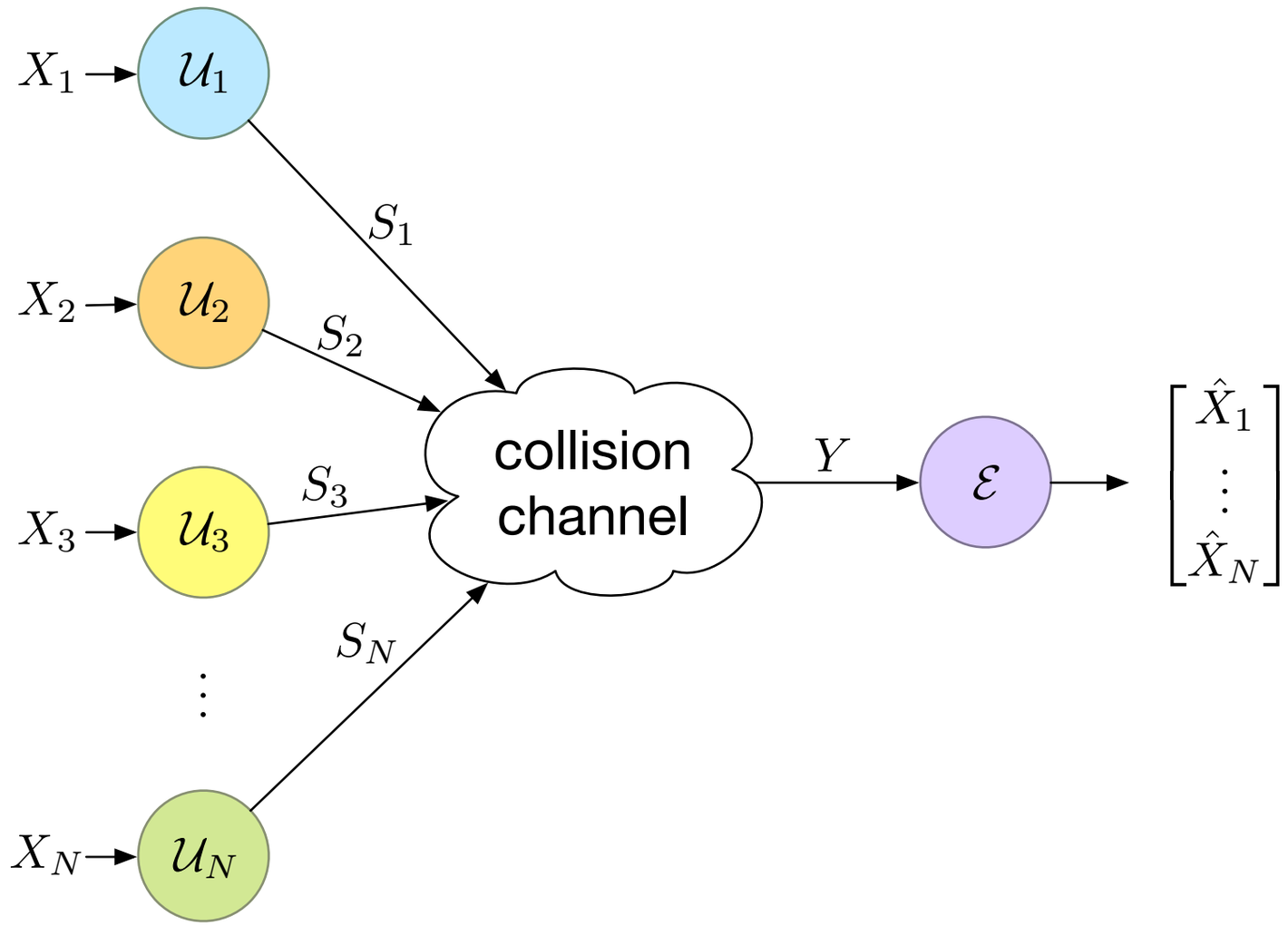




Collisions can be used to improve estimation performance!

Vasconcelos & Martins (*IEEE TAC* 2017)

discrete random variables



Find a strategy $(\mathcal{U}_1^\star, \cdots, \mathcal{U}_N^\star)$ that minimizes the following cost:

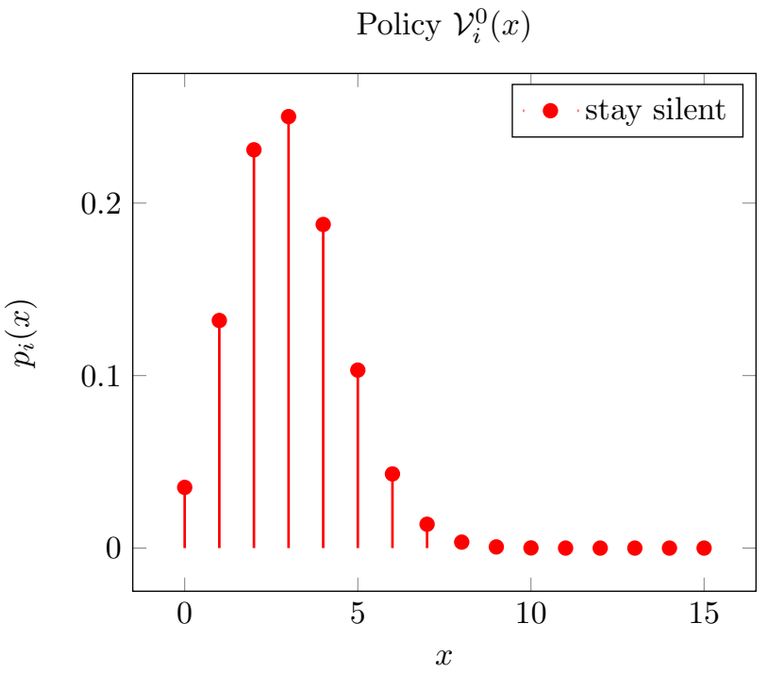
$$\mathcal{J}(\mathcal{U}_1, \cdots, \mathcal{U}_N) = \sum_{k=1}^{N} \eta_k \mathbf{P}(X_k \neq \hat{X}_k)$$

results

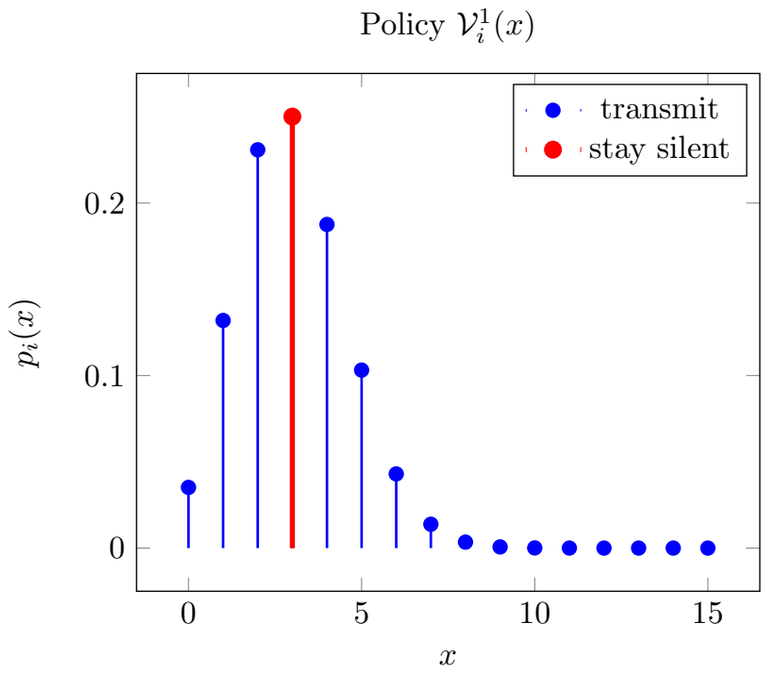
There exists a globally optimal solution $(\mathcal{U}_1^\star, \cdots, \mathcal{U}_N^\star)$ where:

$$\mathcal{U}_i^\star \in \left\{\mathcal{V}_i^0, \mathcal{V}_i^1, \mathcal{V}_i^2\right\}$$

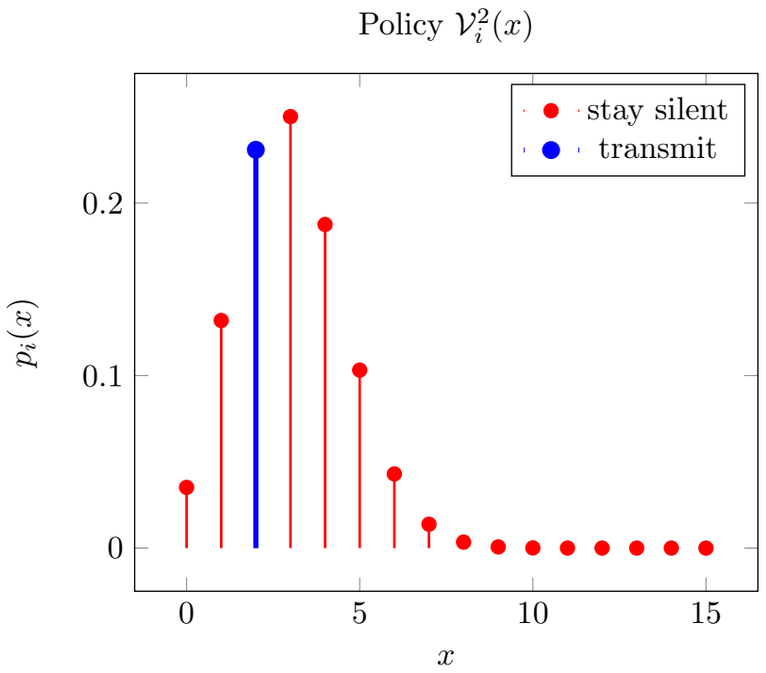
Never transmit



Transmit all, except the most likely symbol

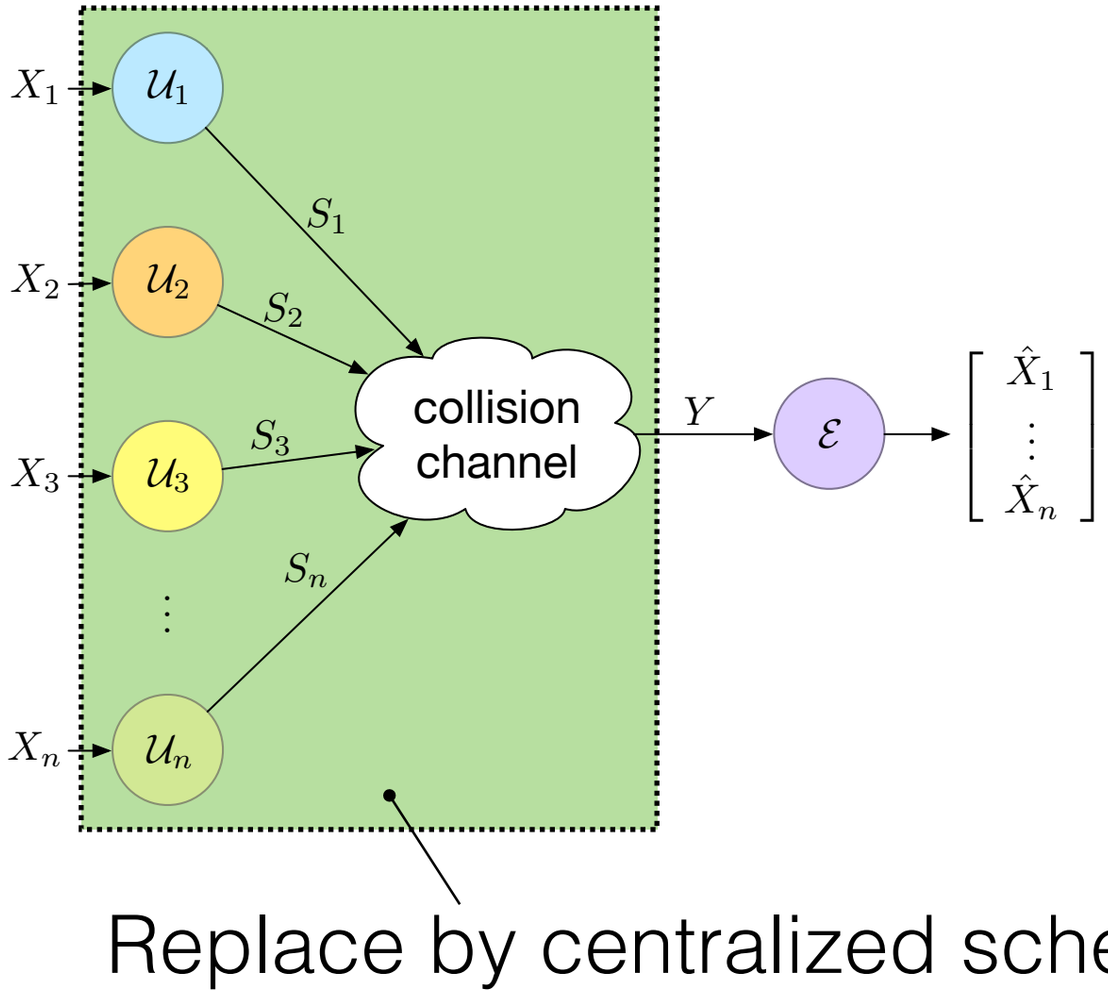


Transmit only the second most likely symbol



Vasconcelos & Martins (*IEEE TAC* 2019)

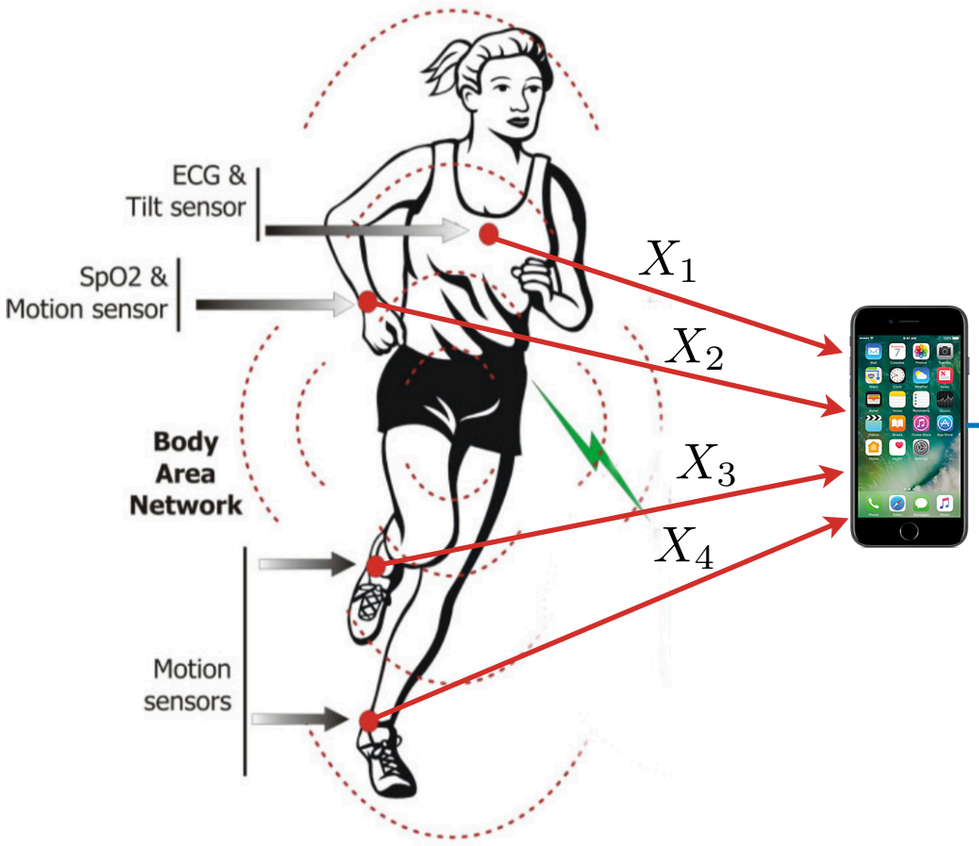
rethinking scheduling



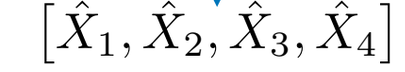
The optimal performance of this system
is a lower bound to the decentralized problem

application: remote health monitoring

Wireless body area networks



KNOW-ME
Mitra et al. (2012)



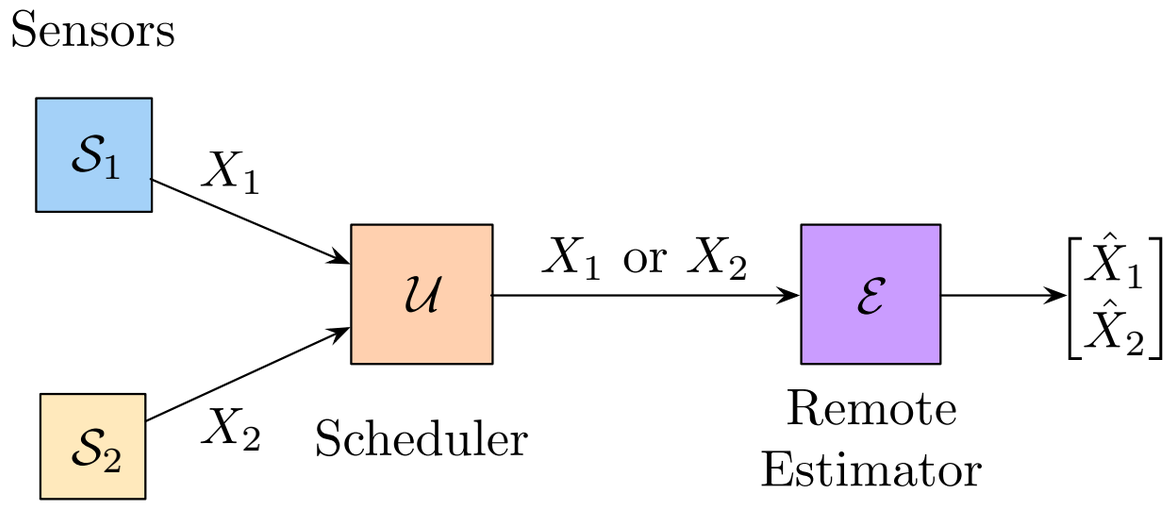
Diagnosis
or
Feedback

Design challenges

1. **Data heterogeneity**
2. **Communication constraints**
3. **Energy constraints***

***of the mobile phone, not the sensors!**

simplest case: two sensors



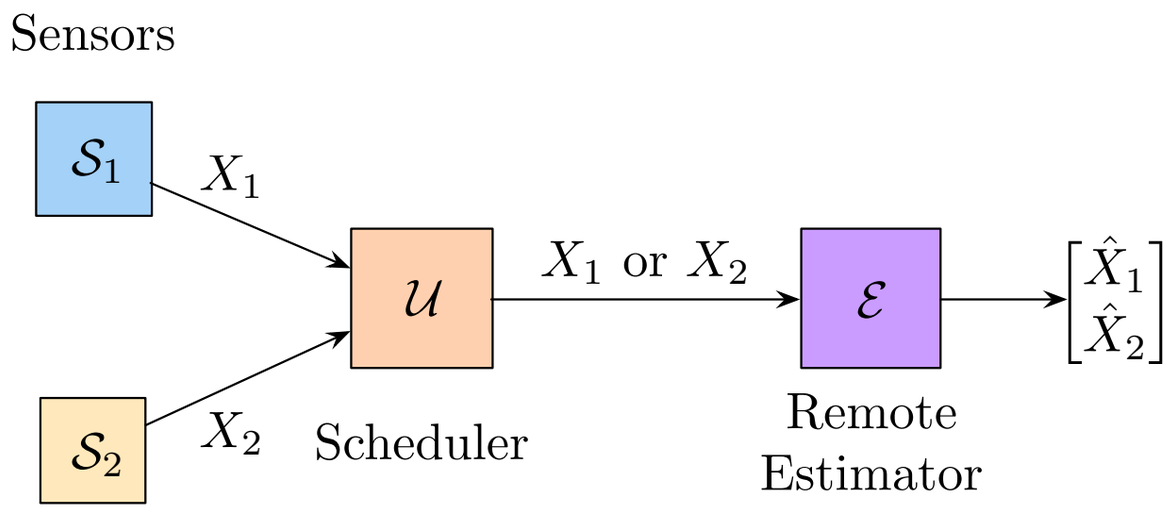
$$\min_{(\mathcal{U},\mathcal{E})\in\mathbb{U}\times\mathbb{E}} \quad \mathcal{J}(\mathcal{U},\mathcal{E}) = \mathbf{E}\left[(X_1-\hat{X}_1)^2+(X_2-\hat{X}_2)^2\right]$$

“Dimensionality reduction”
or
“Subset selection”

1. independent Gaussian observations

Observations

$$X_i \sim \mathcal{N}(0, \sigma_i^2)$$



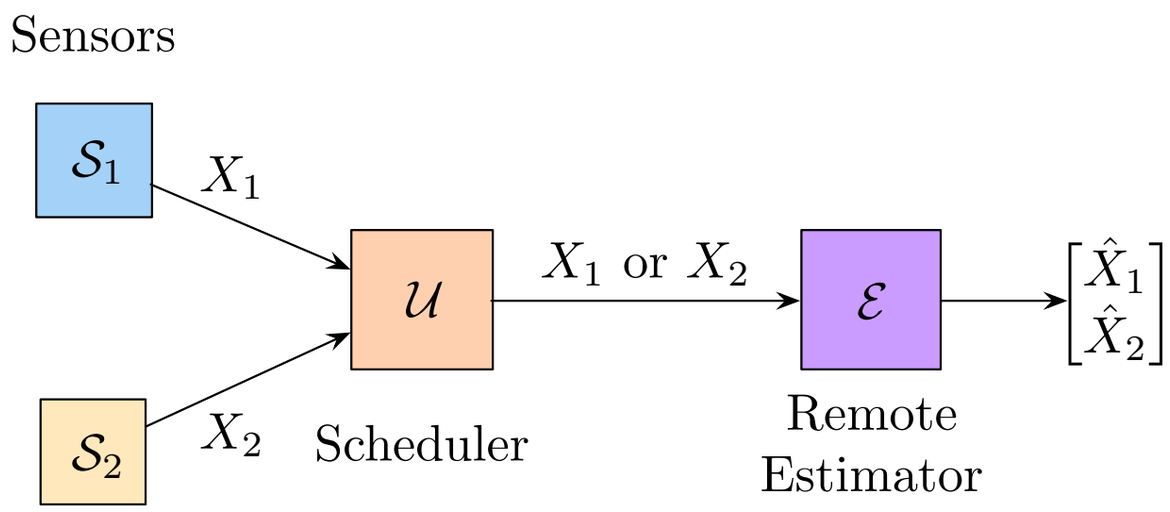
$$X_1 \perp\!\!\!\perp X_2 \quad \Longrightarrow \quad (\mathcal{U}^{\max}, \mathcal{E}^{\text{mean}}) \text{ is person-by-person optimal}$$

2. symmetrically correlated observations

Observations

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$



$$\sigma_1^2 = \sigma_2^2 \quad \Longrightarrow \quad (\mathcal{U}^{\max}, \mathcal{E}^{\text{soft}}) \text{ is person-by-person optimal}$$

Soft-threshold estimation policy

$$\mathcal{E}^{\text{soft}}(1, x_1) = \begin{bmatrix} x_1 \\ \eta(x_1) \end{bmatrix}$$

$$\mathcal{E}^{\text{soft}}(2, x_2) = \begin{bmatrix} \eta(x_2) \\ x_2 \end{bmatrix}$$

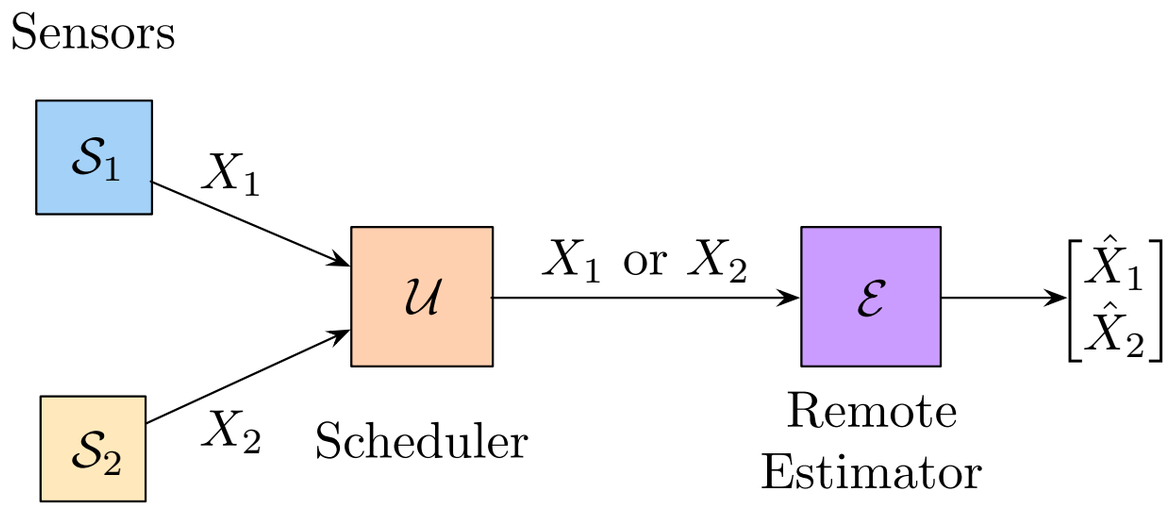
$$\eta(\xi) = \frac{\int_{-|\xi|}^{|\xi|} \tau \exp\left(-\frac{(\tau-\rho\xi)^2}{2\sigma^2(1-\rho^2)}\right) d\tau}{\int_{-|\xi|}^{|\xi|} \exp\left(-\frac{(\tau-\rho\xi)^2}{2\sigma^2(1-\rho^2)}\right) d\tau}$$

2. symmetrically correlated observations

Observations

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

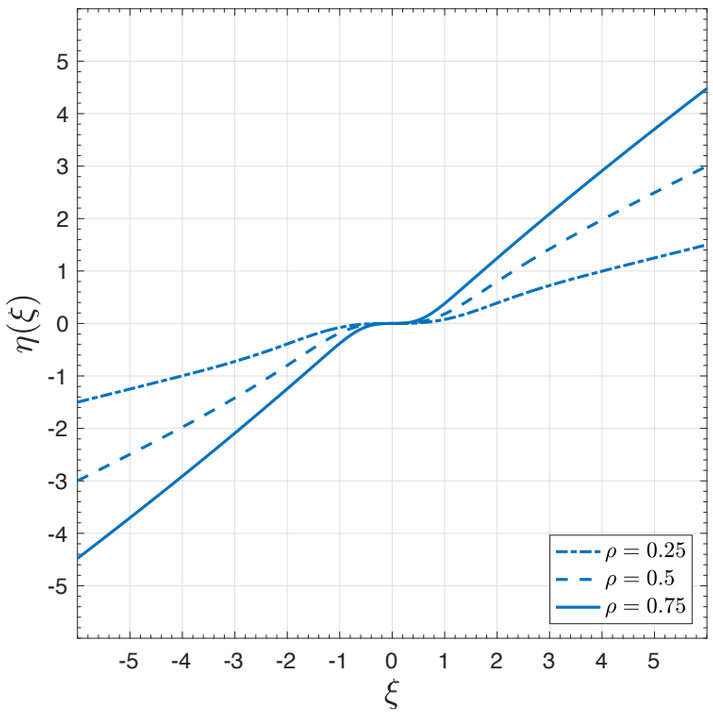


$$\sigma_1^2 = \sigma_2^2 \quad \Longrightarrow \quad (\mathcal{U}^{\max}, \mathcal{E}^{\text{soft}}) \text{ is person-by-person optimal}$$

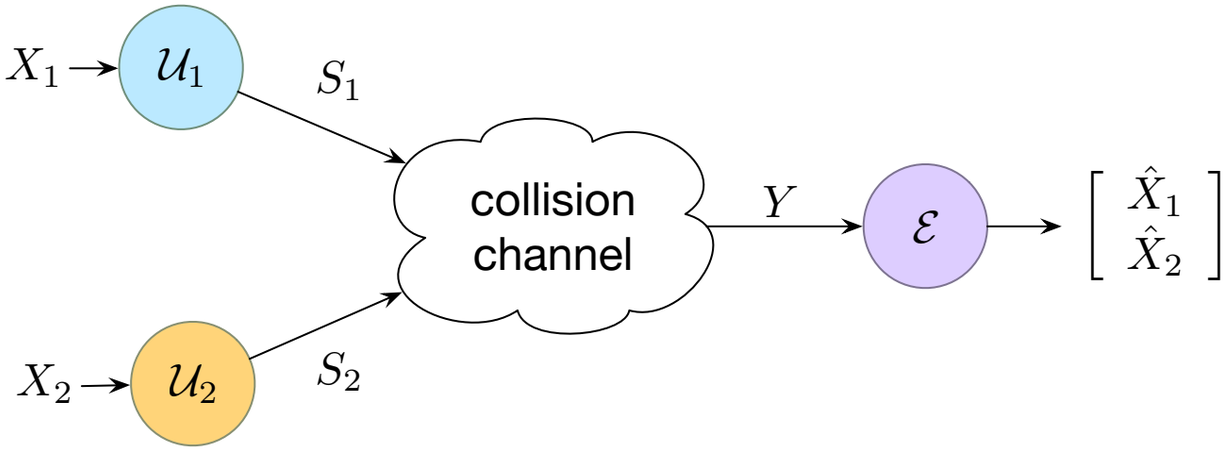
Soft-threshold estimation policy

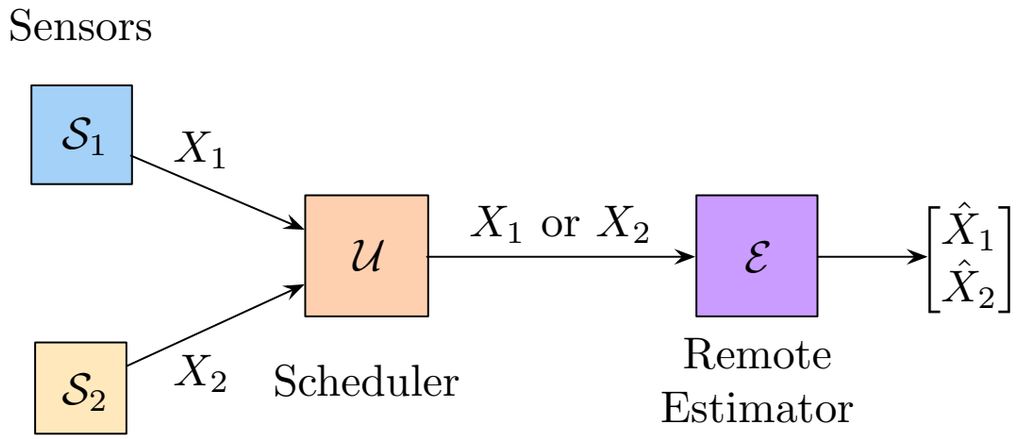
$$\mathcal{E}^{\text{soft}}(1, x_1) = \begin{bmatrix} x_1 \\ \eta(x_1) \end{bmatrix}$$

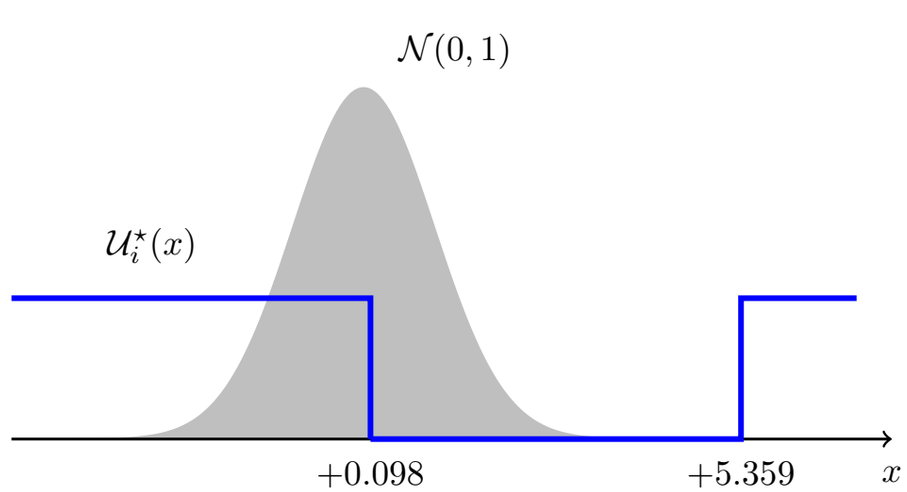
$$\mathcal{E}^{\text{soft}}(2, x_2) = \begin{bmatrix} \eta(x_2) \\ x_2 \end{bmatrix}$$

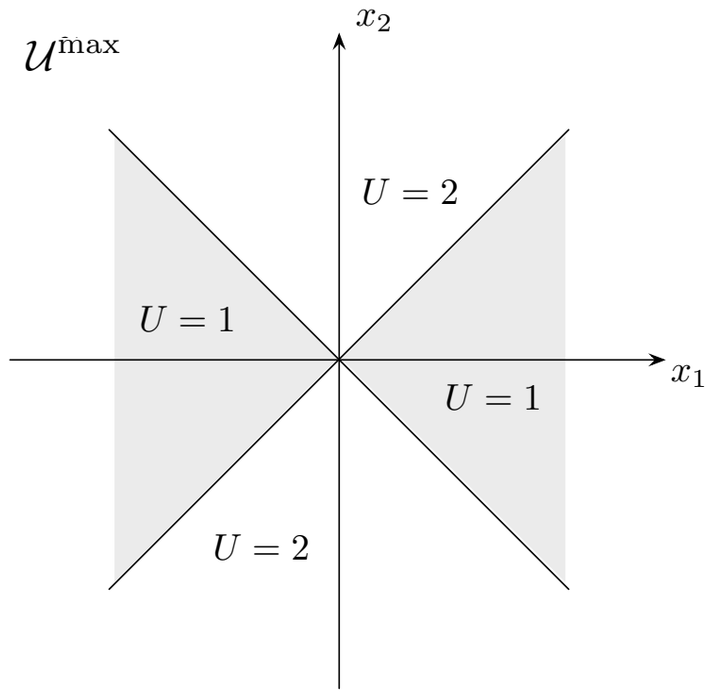


collision vs. scheduling









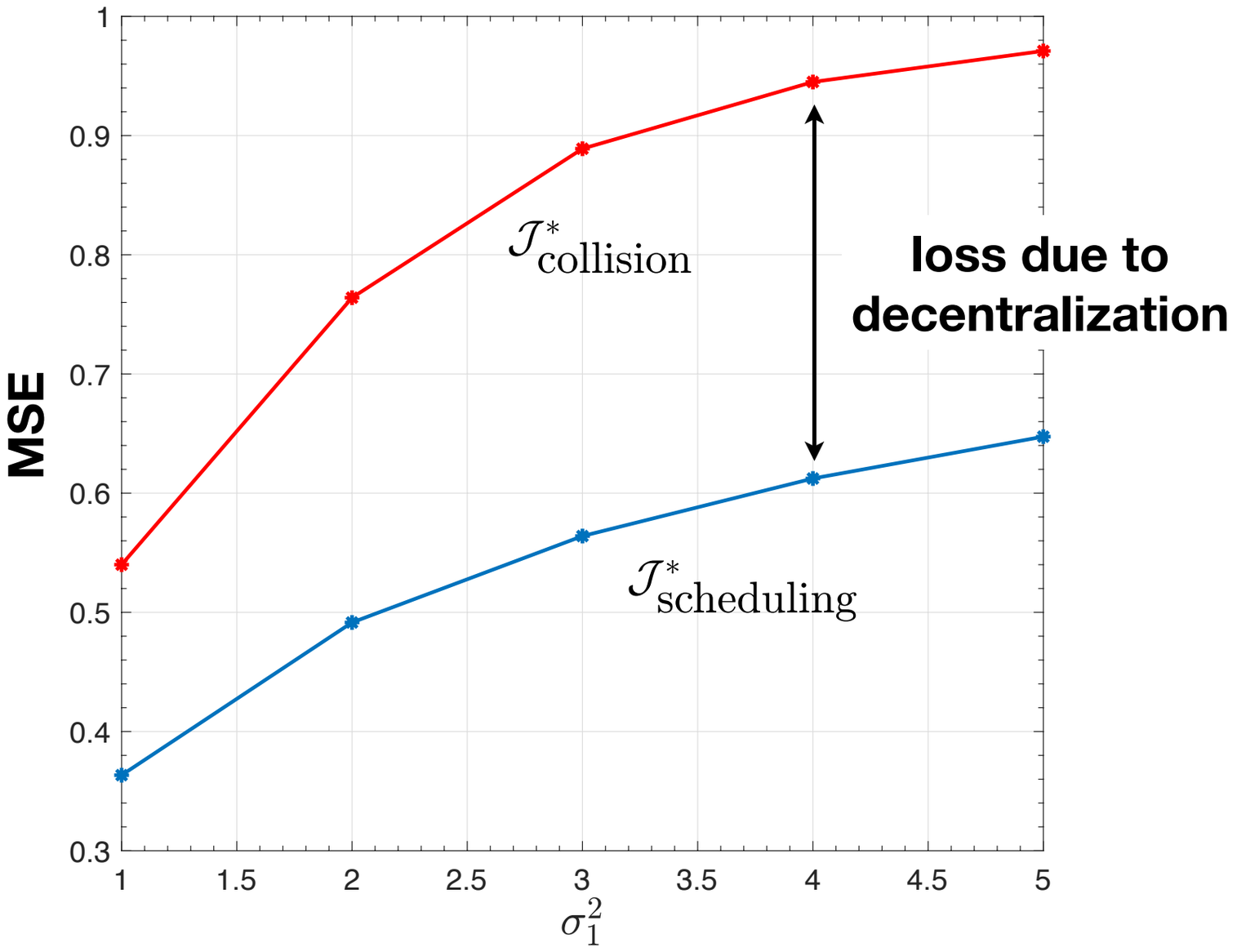
Threshold policies + collision channel ≈ “decentralized max function”

collision vs. scheduling

$X_1 \perp\!\!\!\perp X_2$

$X_1 \sim \mathcal{N}(0, \sigma_1^2)$

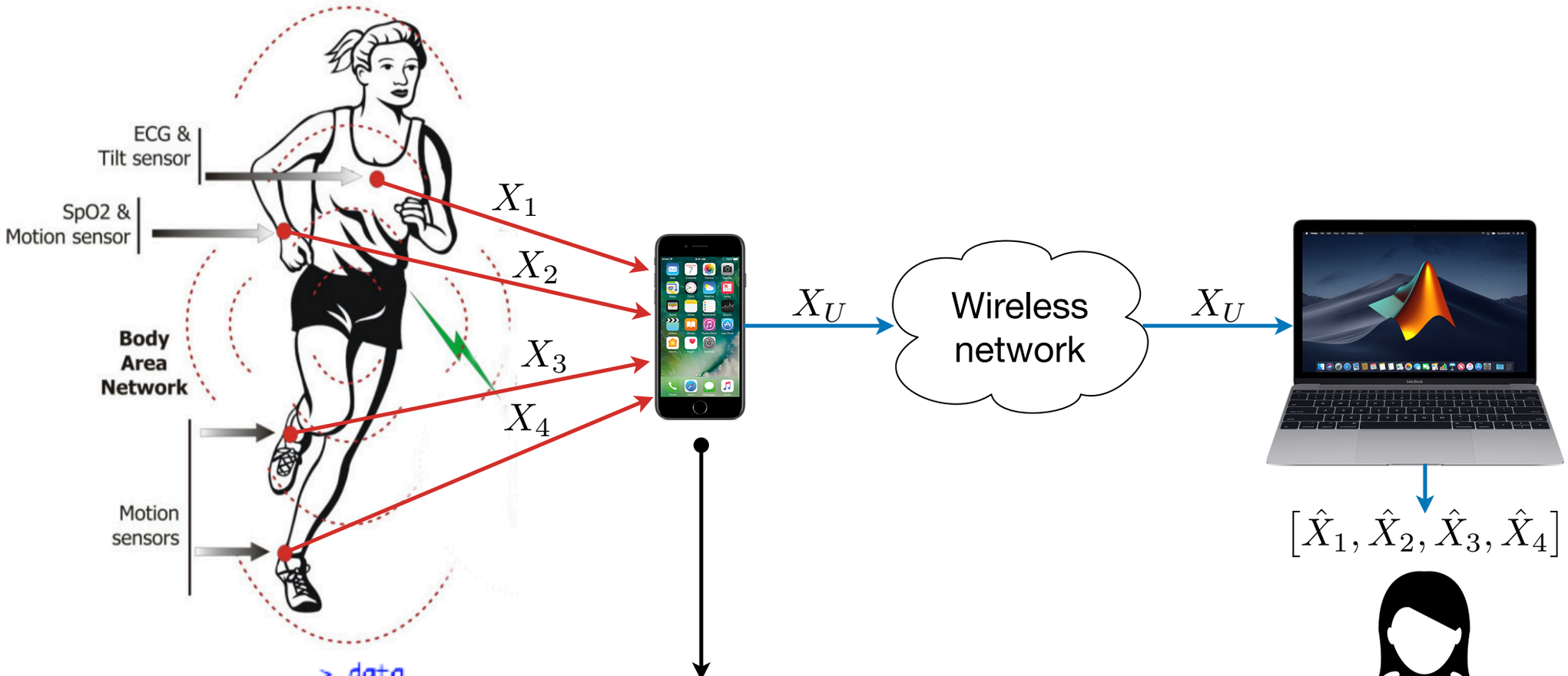
$X_2 \sim \mathcal{N}(0, 1)$



Vasconcelos & Mitra (*IEEE TCNS* 2020)

data-driven scheduling

“In practice we don’t know the joint distribution”




```
> data
          X1           X2          X3          X4
1  0.2215982 -0.040513953 -0.11320355 -0.928056469
2  0.2612376 -0.001308288 -0.10454418 -0.977229008
3  0.2789176 -0.016137590 -0.11060182 -0.995759902
4  0.2773308 -0.017383819 -0.11114810 -0.283740259
5  0.2891883 -0.009918505 -0.10756619  0.030035338
6  0.2554617 -0.023953149 -0.09730200 -0.354708025
7  0.2813734 -0.018158740 -0.10724561 -0.974059465
8  0.2770874 -0.015687994 -0.10921827 -0.986822280
9  0.2779115 -0.018420827 -0.10590854 -0.987271889
10 0.2764266 -0.018594920 -0.10550036 -0.423642838
11 0.2776153 -0.022661416 -0.11681294  0.046366681
```

data-driven scheduling

$$J(\theta) = \mathbf{E}\Big[\min\Big\{\big(X_1 - (w_{12}X_2 + b_{12})\big)^2, \big(X_2 - (w_{21}X_1 + b_{21})\big)^2\Big\}\Big]$$

$$\theta = \begin{bmatrix} w_{21} & b_{21} & w_{12} & b_{12} \end{bmatrix}$$

$$\mathcal{D} = \Big\{\big(x_1(k), x_2(k)\big),\ k = 1, \cdots, N\Big\}$$

non-convex
non-smooth
linear regression

Replace expectations with the **empirical mean**

$$J_{\mathcal{D}}(\theta) = \frac{1}{N} \sum_{(x_1,x_2)\in\mathcal{D}} \min\Big\{\big(x_1 - (w_{12}x_2 + b_{12})\big)^2, \big(x_2 - (w_{21}x_1 + b_{21})\big)^2\Big\}$$

Vasconcelos & Mitra (*IEEE TCNS* under review)

approximate convex-concave procedure

$$J_{\mathcal{D}}(\theta) = F_{\mathcal{D}}(\theta) - G_{\mathcal{D}}(\theta)$$ **empirical risk**

$$F_{\mathcal{D}}(\theta) = \frac{1}{N} \sum_{(x_1,x_2)\in\mathcal{D}} \big(x_1 - (w_{12}x_2 + b_{12})\big)^2 + \big(x_2 - (w_{21}x_1 + b_{21})\big)^2$$

$$G_{\mathcal{D}}(\theta) = \frac{1}{N} \sum_{(x_1,x_2)\in\mathcal{D}} \max\Big\{\big(x_1 - (w_{12}x_2 + b_{12})\big)^2, \big(x_2 - (w_{21}x_1 + b_{21})\big)^2\Big\}$$

$$\theta^{(k+1)} = \mathbf{A}_{\mathcal{D}}^{-1} g_{\mathcal{D}}(\theta^{(k)}) + \mathbf{b}_{\mathcal{D}}$$

$$\theta^{(k+1)} \to \hat{\theta}^{\star}$$ **a local minimum of the empirical risk**

convergence rates

CCP

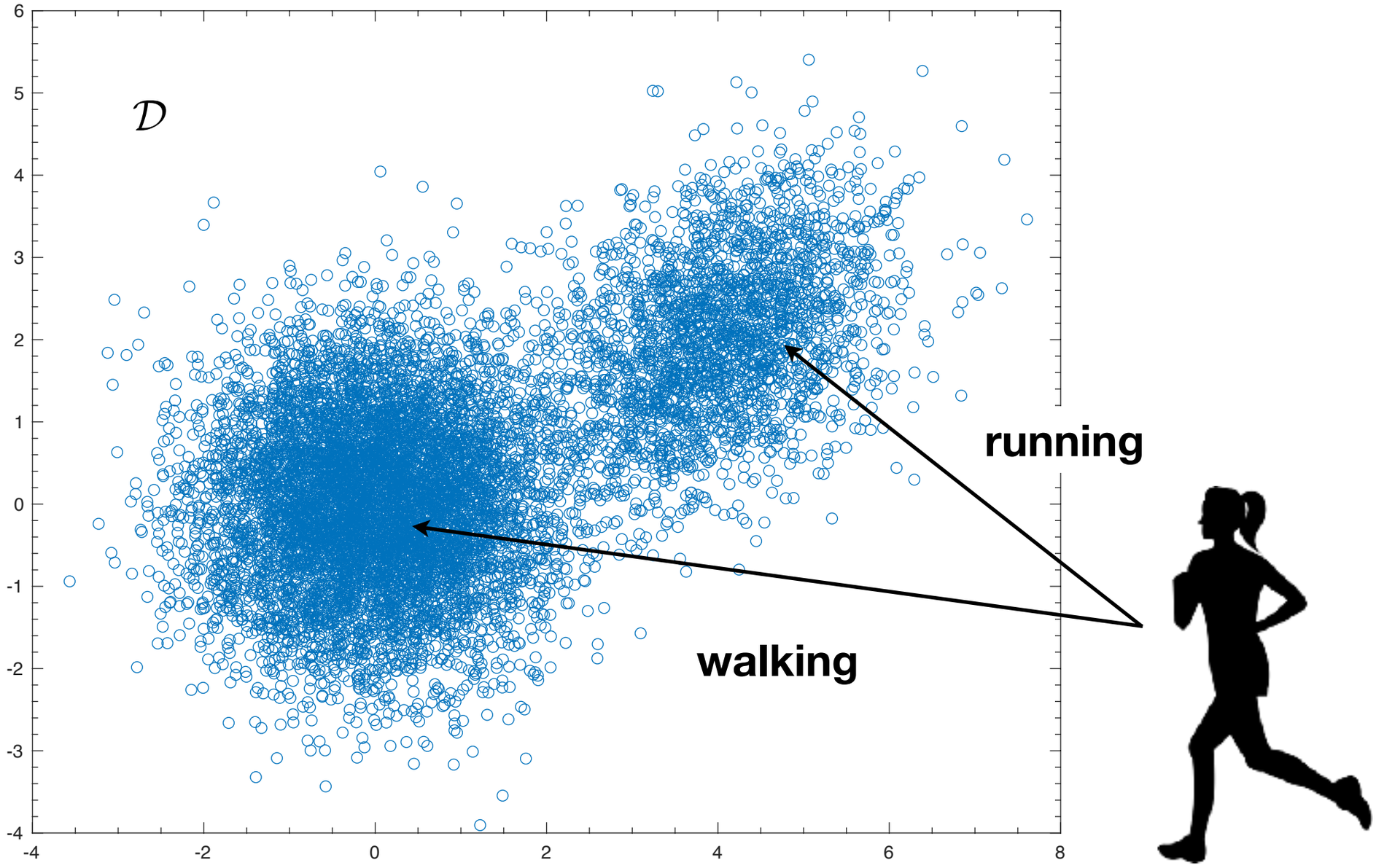
$$\mathcal{O}(1)$$

typical SGD

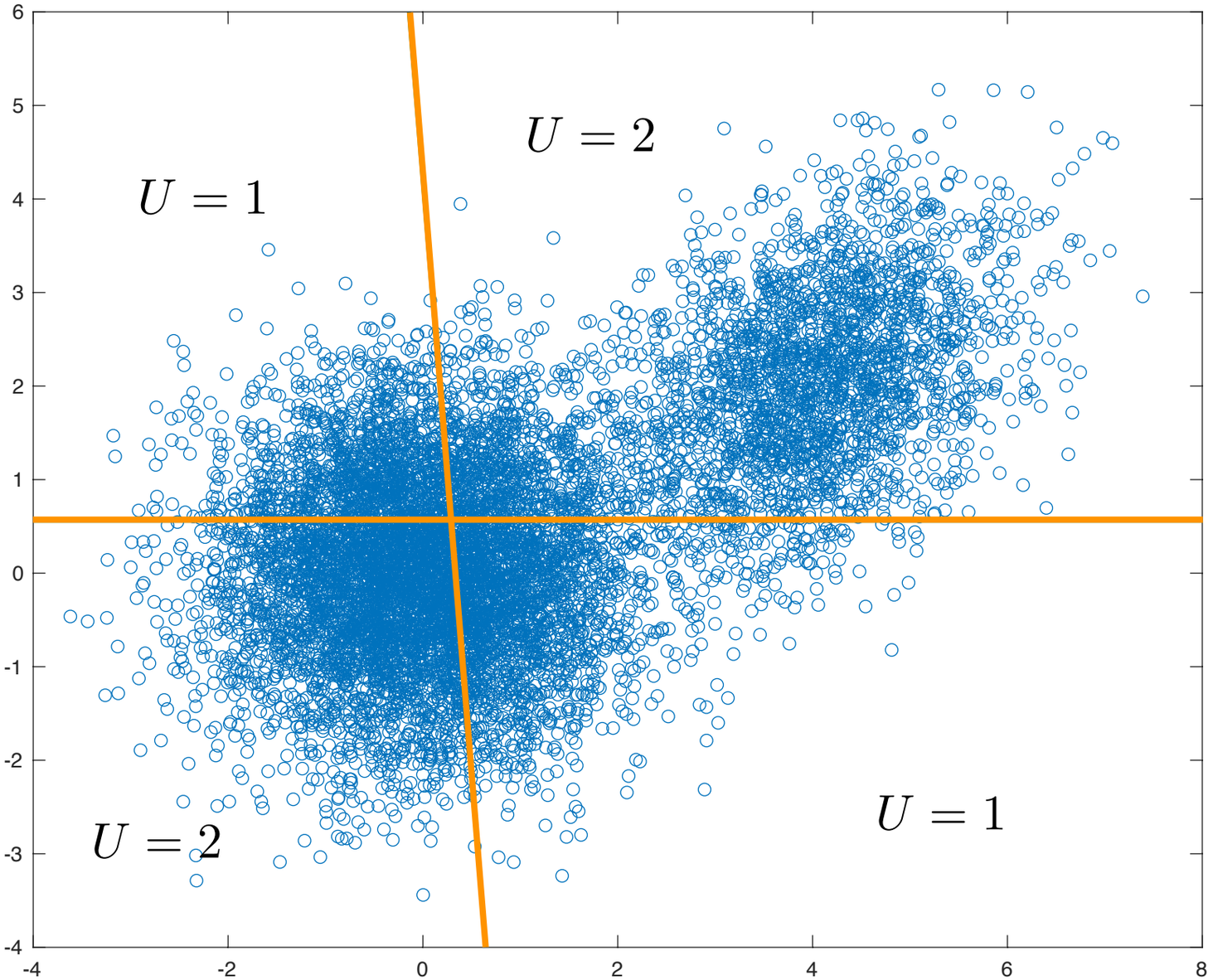
$$\mathcal{O}(1/\sqrt{k})$$

Gaussian mixture data

$$(X_1, X_2) \sim 0.75 \cdot \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + 0.25 \cdot \mathcal{N}\left(\begin{bmatrix}4\\2\end{bmatrix}, \begin{bmatrix}1 & 0.4\\0.4 & 1\end{bmatrix}\right)$$

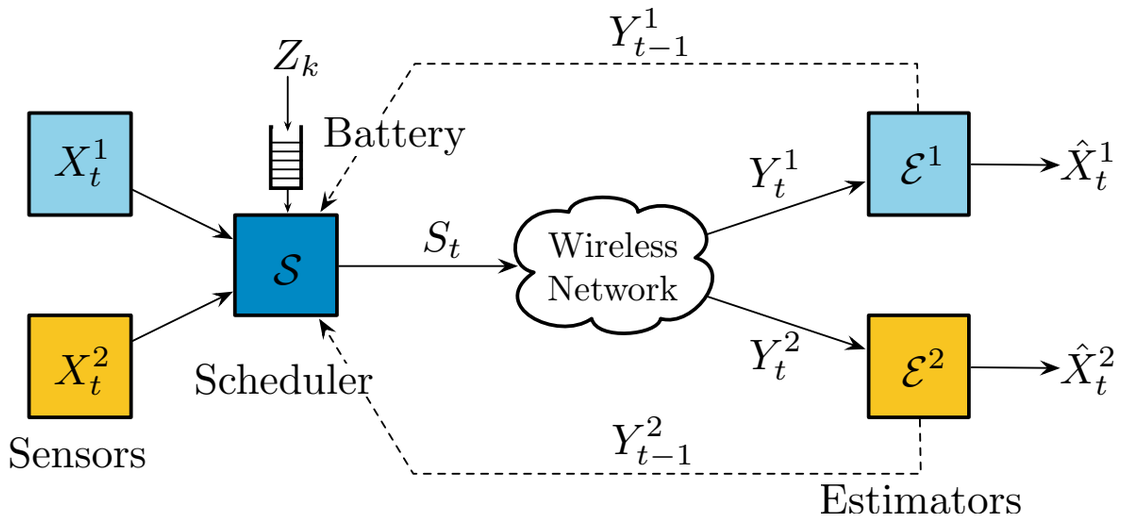


data-driven scheduler



sequential scheduling with energy harvesting

characterization of optimal policies when pdfs are symmetric and unimodal

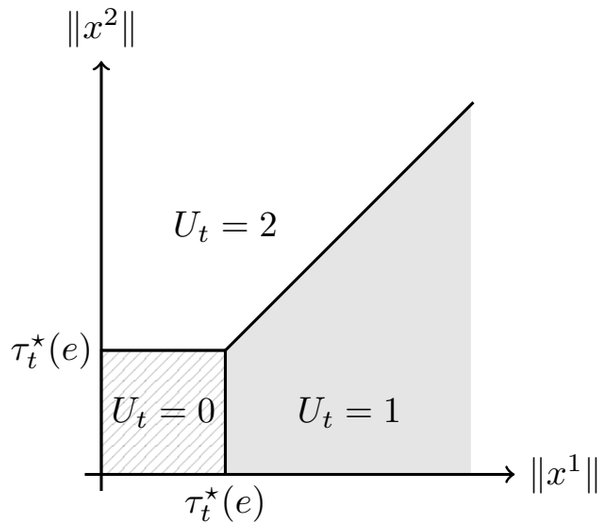


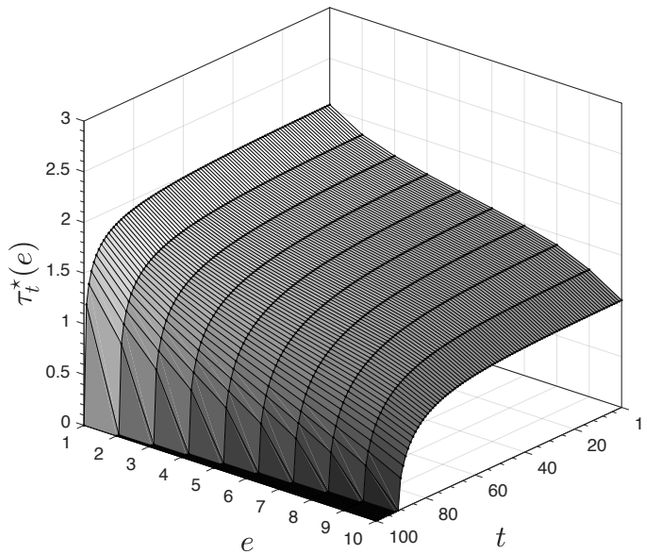
(A) Information expansion/relaxation

(B) Common information approach

(C) POMDP analysis

now at Qualcomm AI





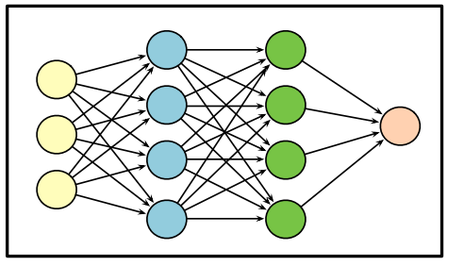
Vasconcelos, Gagrani, Nayyar & Mitra (*IEEE TCNS* 2021)

more rethinking

privacy preserving collaborative machine learning

$$\min_{\theta} L(\theta) = \min_{\theta} \mathbb{E}\big[\ell(\theta, X)\big]$$

θ is a vector of model parameters



deep learning server

stochastic gradient descent

$$\theta^{(k+1)} = \theta^{(k)} - \alpha_k \nabla L(\theta^{(k)})$$

ideally

$f_X(x)$ is known

$\ell(\theta, x)$ is convex in θ for every x

reality

$$\mathcal{D} = \{x_k\}_{k=1}^{m}$$

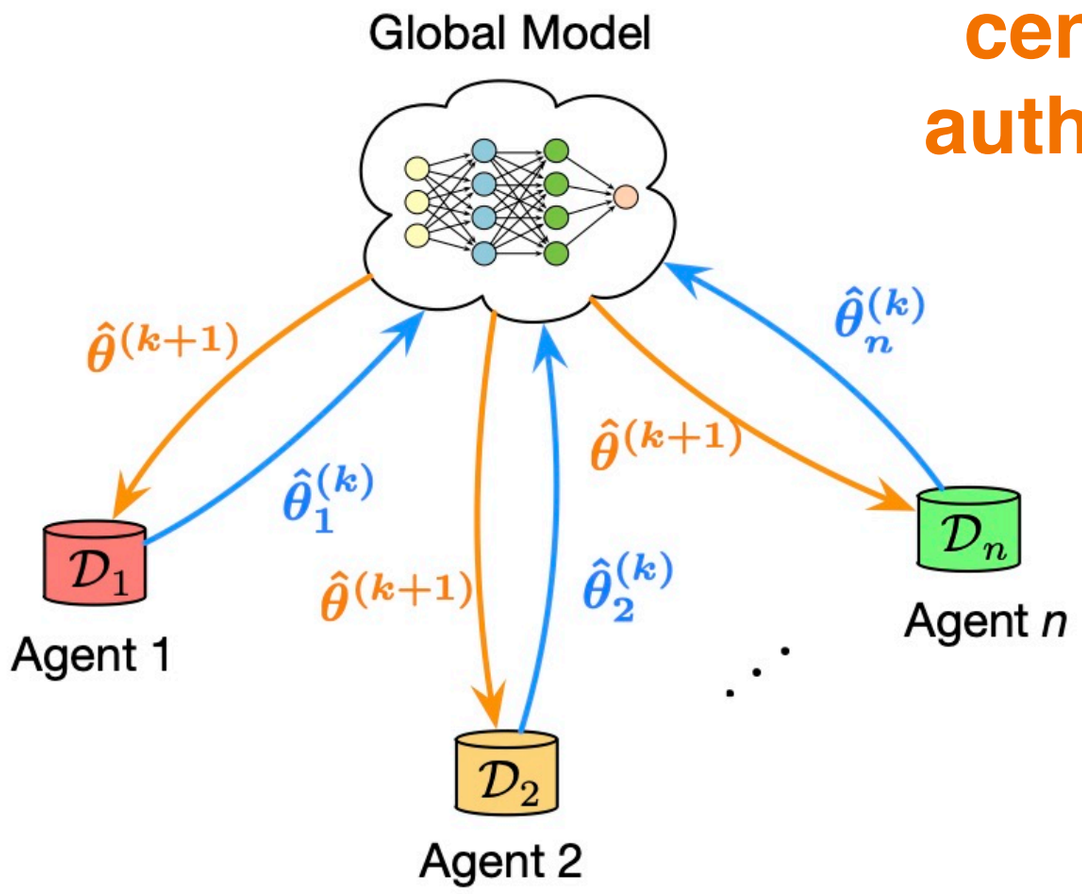
$\ell(\theta, x)$ is not convex in θ

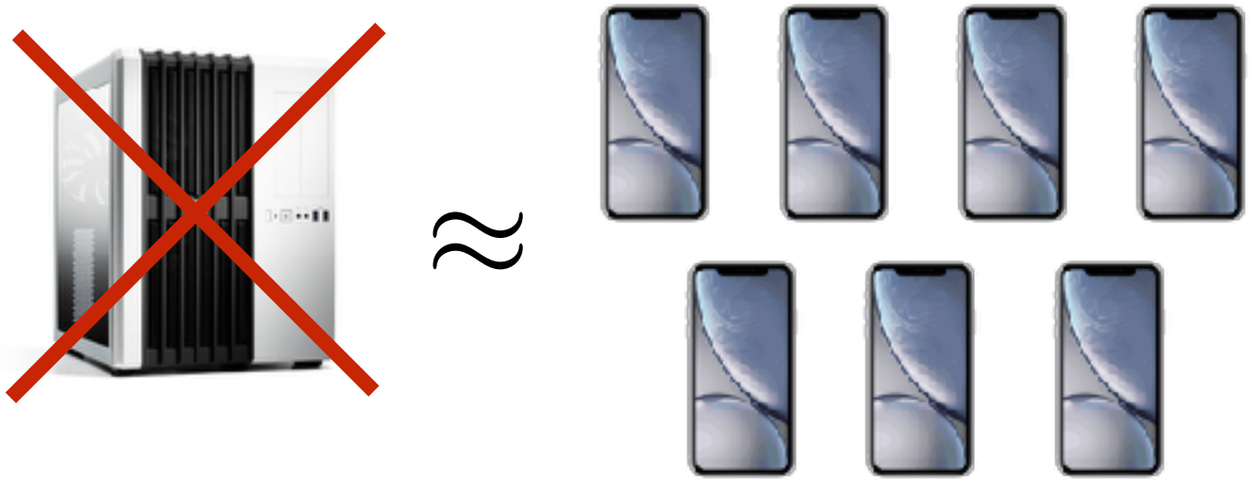
federated learning

requires a central authority

$$\mathcal{D} = \bigcup_{i=1}^{n} \mathcal{D}_i \qquad \mathcal{D}_i = \{x_k^i\}_{k=1}^{m}$$

$$\min_{\theta} \frac{1}{nm} \sum_{x \in \mathcal{D}} \ell(\theta, x)$$



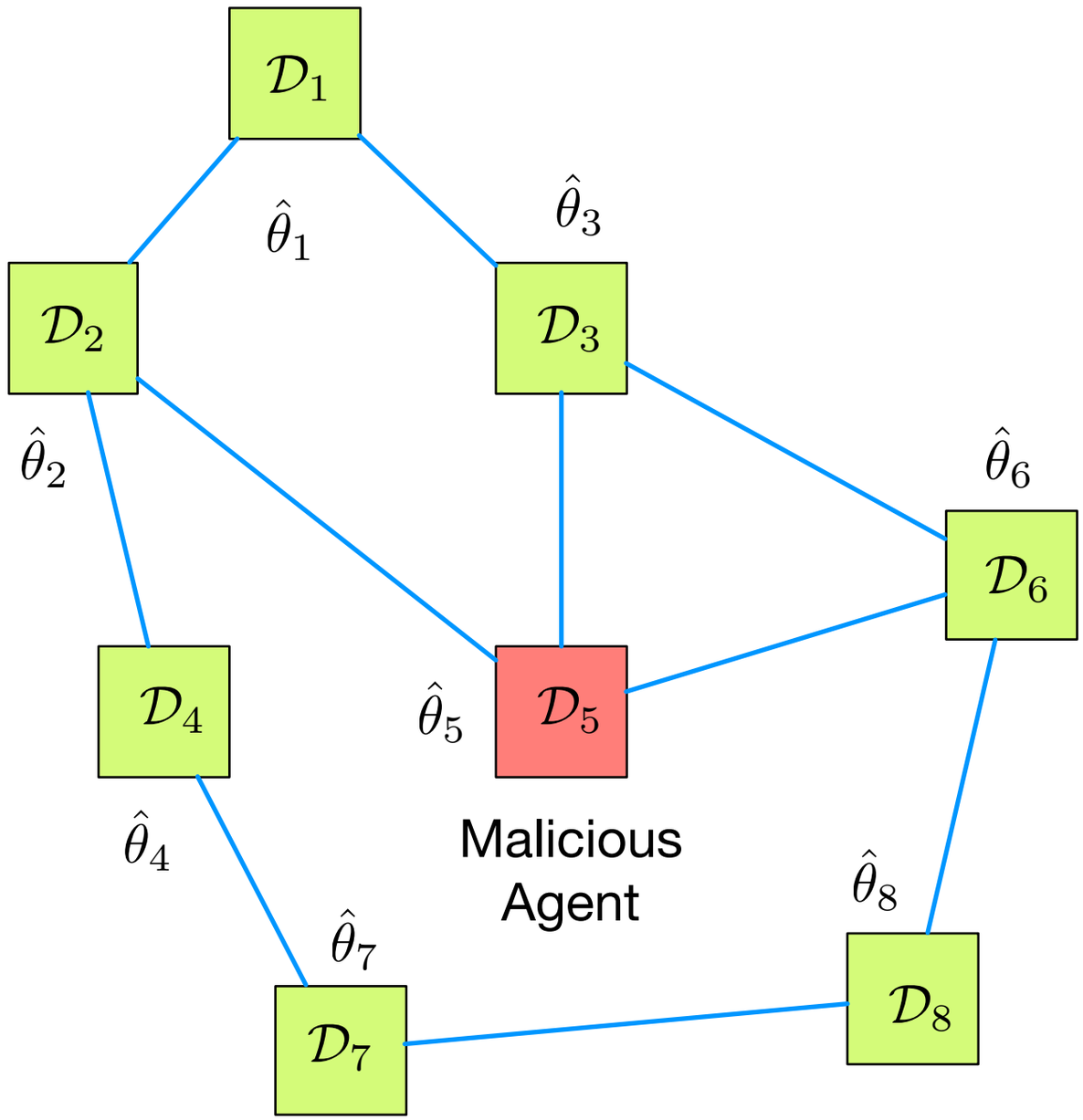


agents do not exchange local data

local processing power is limited

no provable privacy guarantees

distributed edge learning



$$\min_{\theta} \frac{1}{nm} \sum_{x \in \mathcal{D}} \ell(\theta, x)$$

$$\frac{1}{n} \sum_{i=1}^{n} f_i(\theta)$$

$$f_i(\theta) = \frac{1}{m} \sum_{x \in \mathcal{D}_i} \ell(\theta, x)$$

consensus

+

differential privacy

+

non-observability

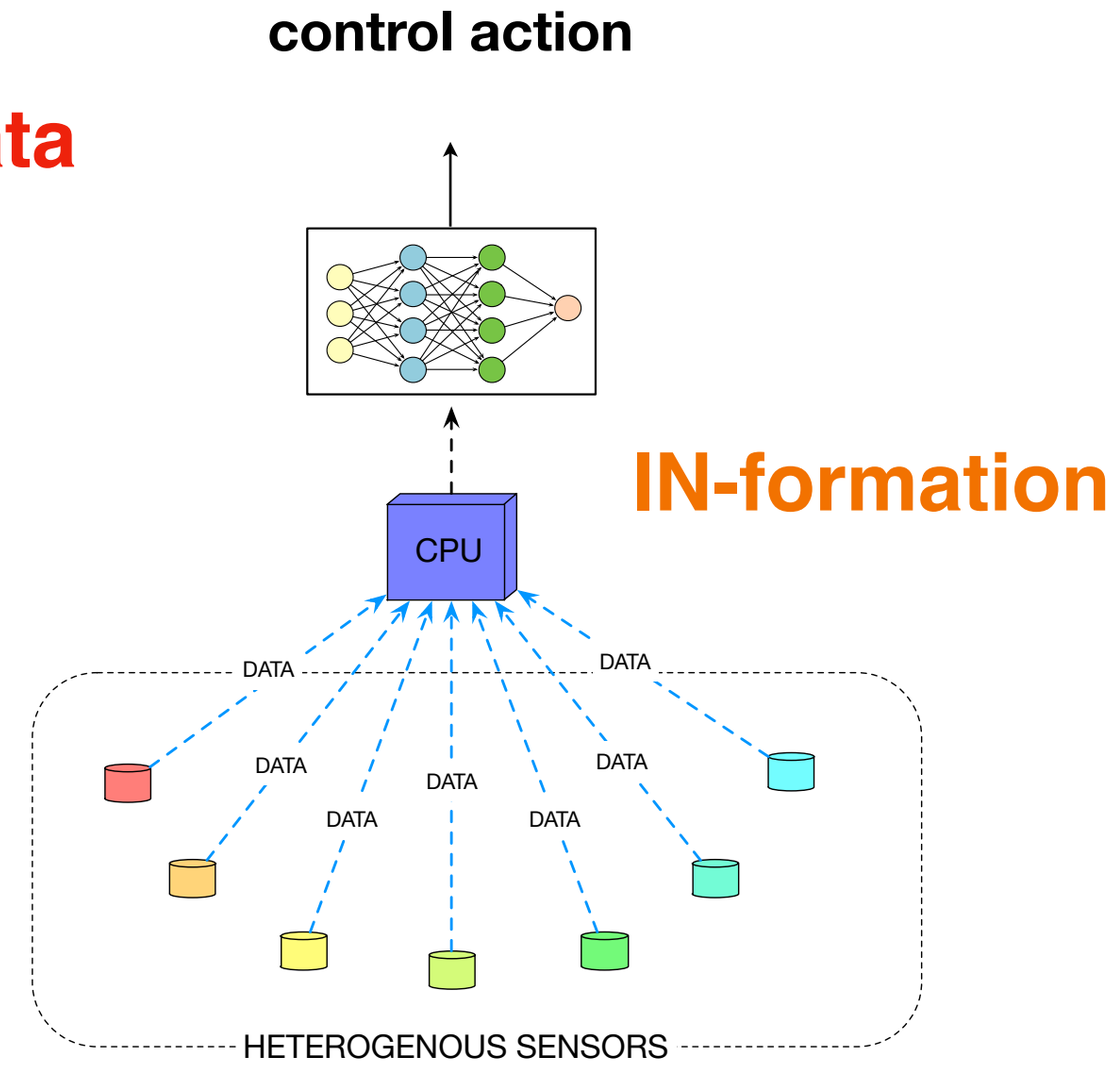
CPS architectures inspired by the human brain

pushing high bandwidth sensor data

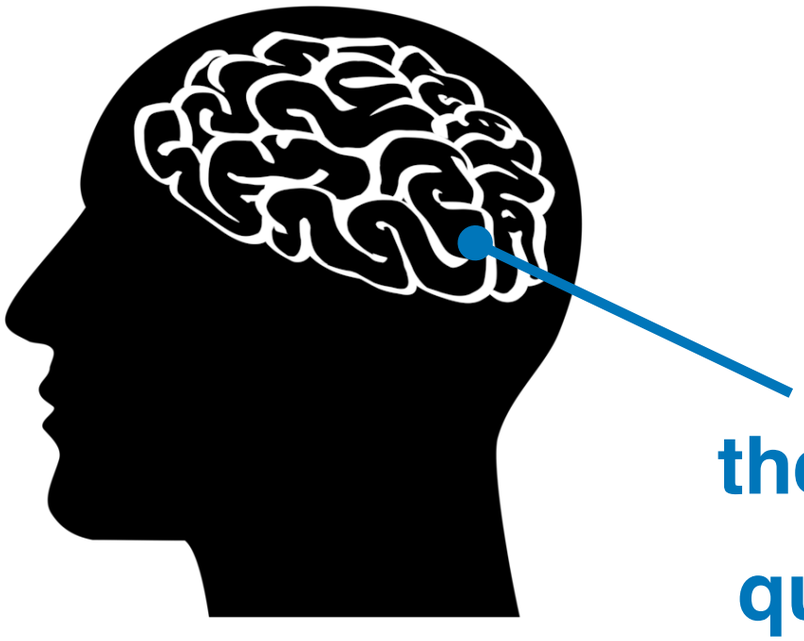
to the CPU is expensive

processing that data and

send it to the devices is slow



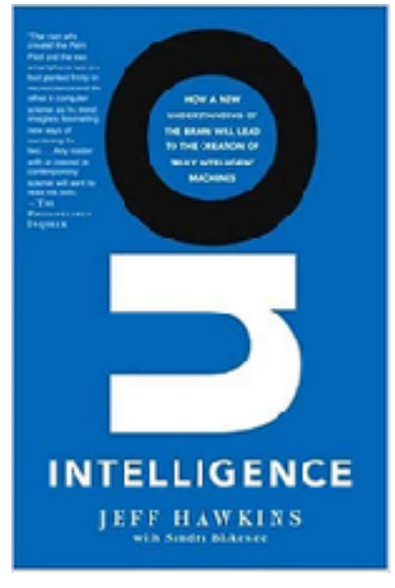
IN-formation



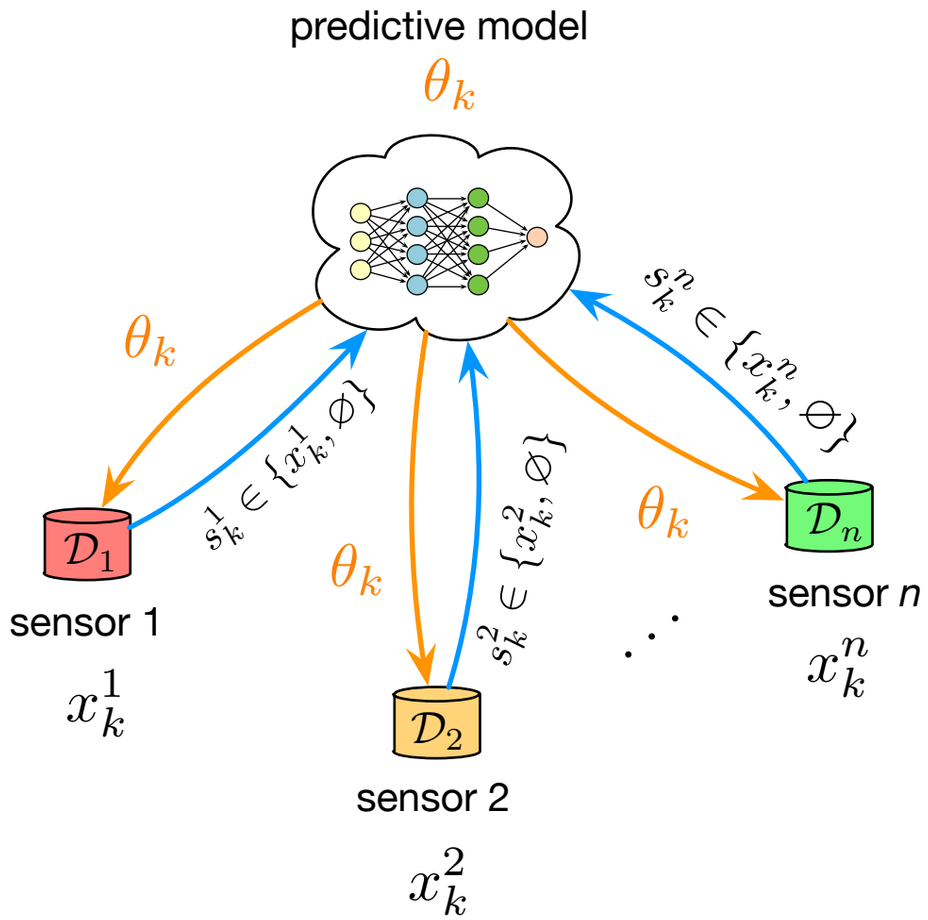
the human brain can react

quickly to sudden events

because it ignores (a lot) of data



CPS architectures inspired by the human brain

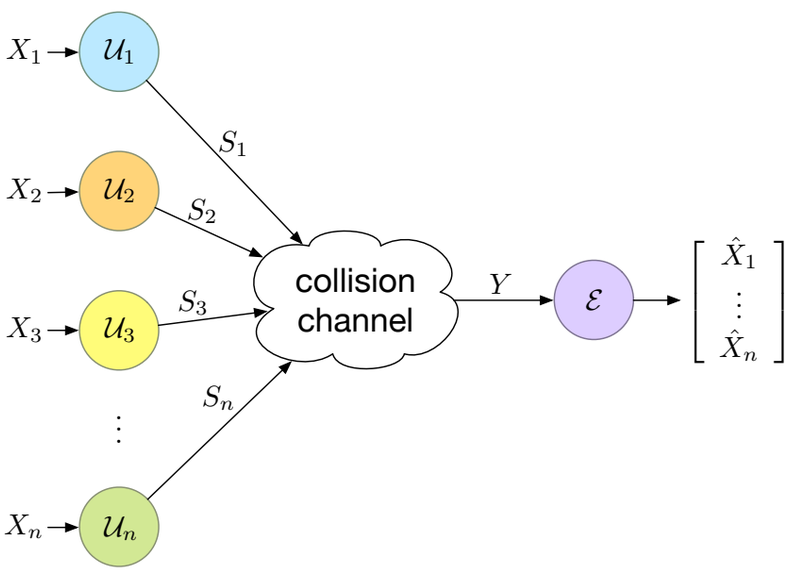


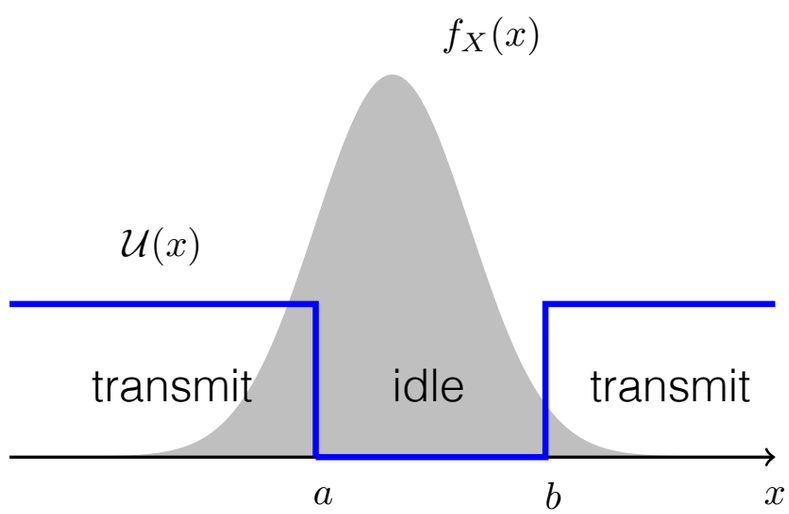
the CPU forms a predictive model of the state of the system

sends model parameters to the sensors

sensors only send information that deviates from what the brain already knows

OUT-formation





less redundant data = reduction in latency!

conclusion + open questions

5G will power the IoT and CPS revolution

5G will require rethinking PHY/MAC layers with the goal is control + decision-making

high data rates and low probability of error is not the end of the story

1. how to train ML models efficiently while guaranteeing privacy?

2. can we leverage the structure of human decision-making to create low-latency (safe) autonomous vehicles?

Thank you!

mullervasconcelos.github.io

mvasconc@usc.edu