Signal Processing for Large Scale Sensor Networks

Cross Layer Design for Application Specific Networks

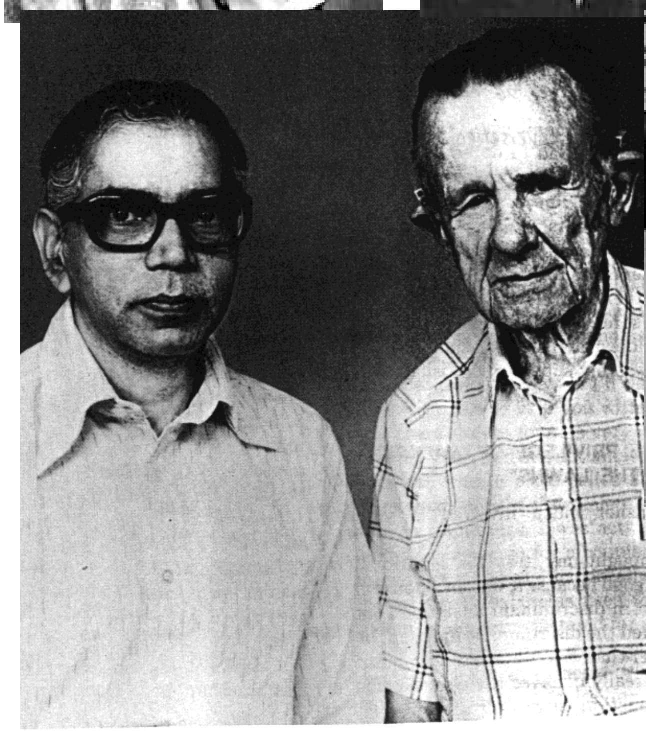
Lang Tong

School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853

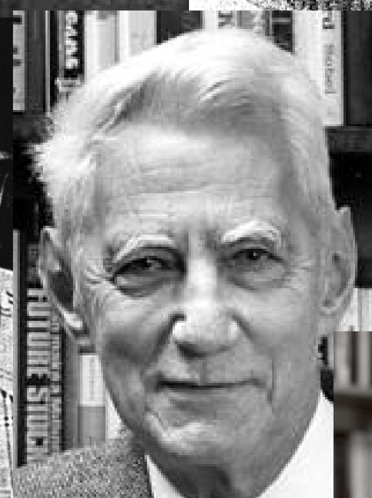
Joint work with G. Mergen, Y. Sung, Anthony Ephremides, and H. Vincent Poor
Supported in part by the NSF, ONR, ARL-CTA on Communication and Networks

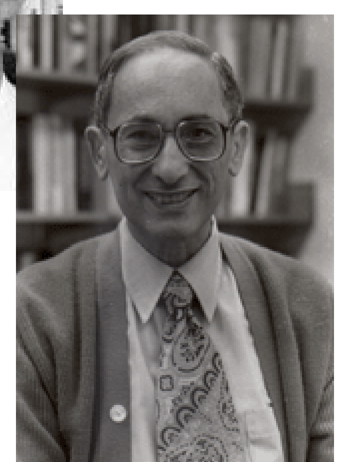
The Google Index

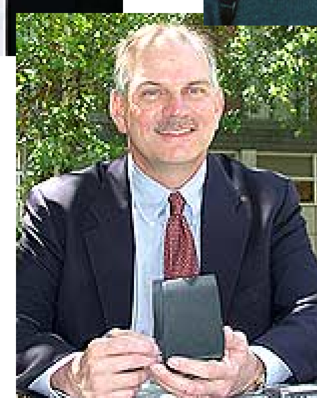
• Sensor Network	4.36M
• Cellular Network	14.5M
• Ad hoc Network	2.9M
• Neural Network	3.82M
• MIMO	6.73M
• Array Processing	6.41M
• Beamforming	0.12M
• Multiuser Detection	0.08M
• Blind Equalization	0.09M



C. R. Rao and Harald Cramér, 1978







Outline

Theme of the Talk

- Dijkstra's Layered Architecture and Cross Layer Design.
- TBMA: Data Centric Medium Access for Parameter Estimation.
- Data Centric Routing for Signal Detection.

Signal Processing Ideas

- Coherent Combining
- Diversity
- Innovation

Dijkastra's Layered Architecture

Operator

User Program

I/O Management

I/O Devices
Memory

Access Schedule

Hardware

Application

Transport

Network

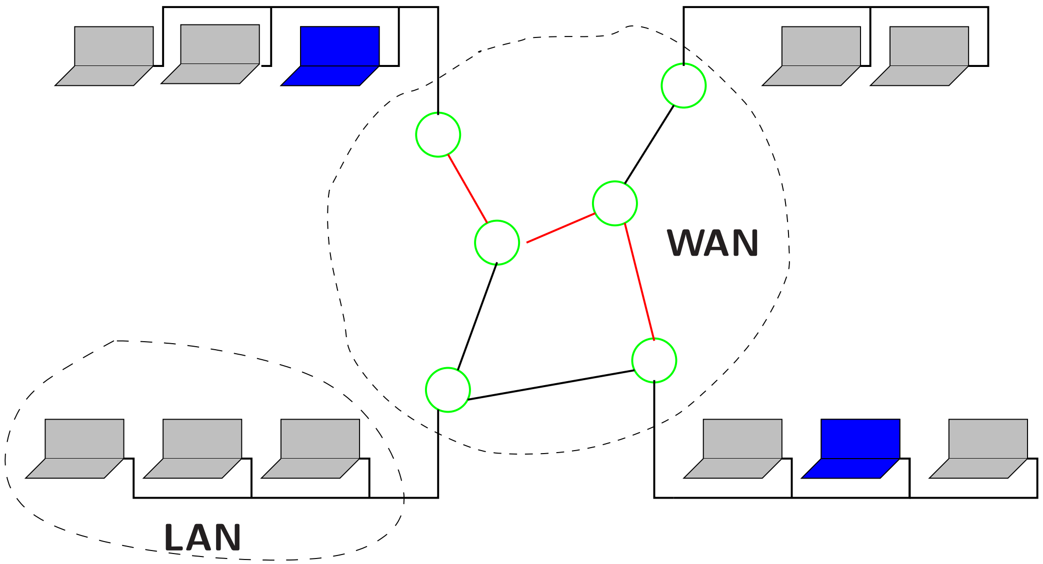
Datalink

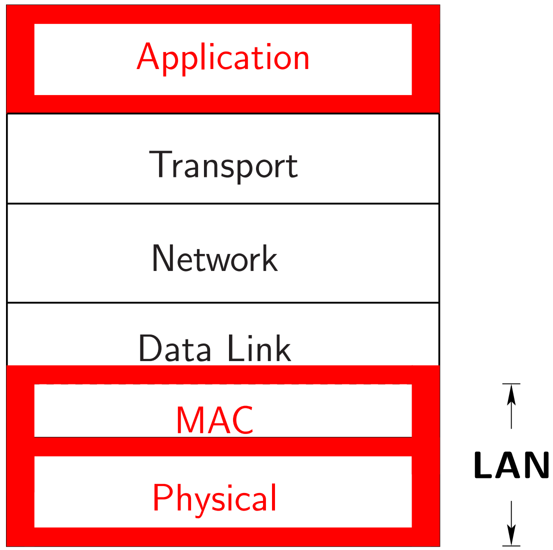
MAC

PHY

EWD 196: The structure of 'THE'-multiprogramming system, *Comm. ACM 11*, 1968, 5:341-346

Why Layered Architecture is Crucial





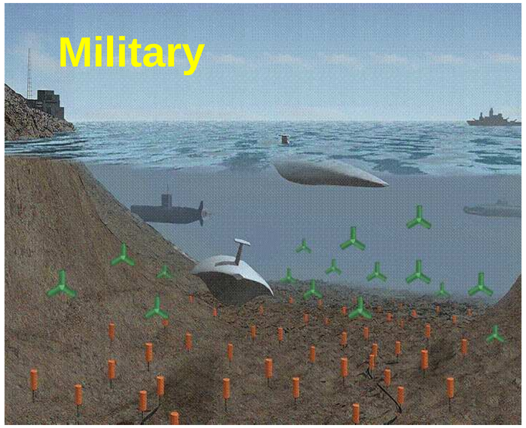
Objectives

- Millions of users
- Thousands of applications
- A growing variety of devices.

The need of layered approaches:

- Partition network functions into layers
- Design each layer separately

Sensor Networks are Application Specific



- Network with purposes.
- "Unconventional" design metrics
- Harsh design constraints

What is the appropriate layering architecture for sensor network?

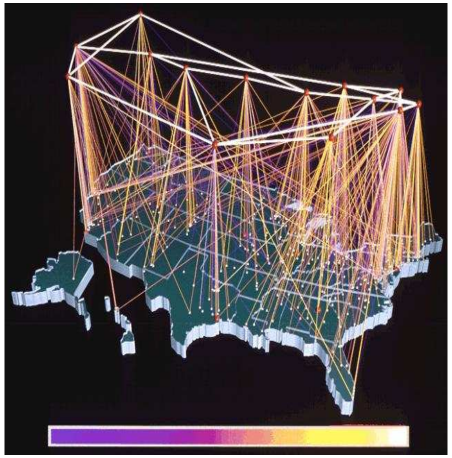
Dijkstra on Layers with Yardsticks

"I can only view a well-structured system as a hierarchy of layers and in the design process, the interface between these layers has to be designed and decided upon each time. I am not so much bothered by designer's willingness and ability to propose such interfaces, I am seriously bothered by the lack of commonly accepted yardsticks along which to compare the evaluate such proposals."

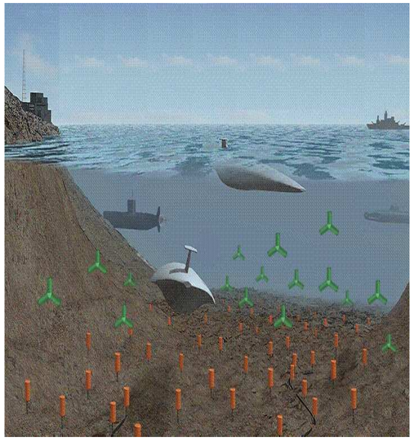
E. W. Dijkstra, "Hierachical Ordering of Sequential Processes," EWD 310.

What should be the yardstick?

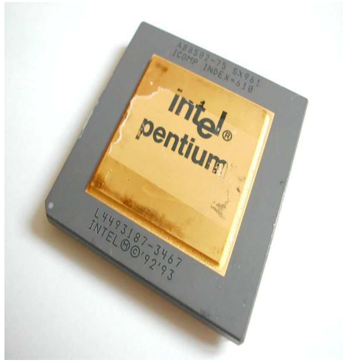
Is There A Parallel in Computing?



General Purpose Network



Sensor Network

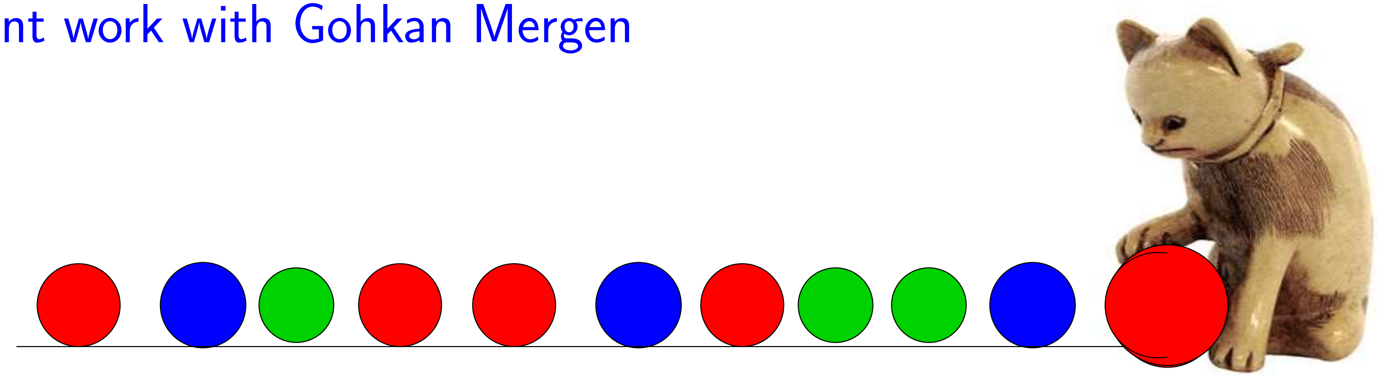


General Purpose Processor

DSP or ASIC

TBMA: A Data Centric Medium Access

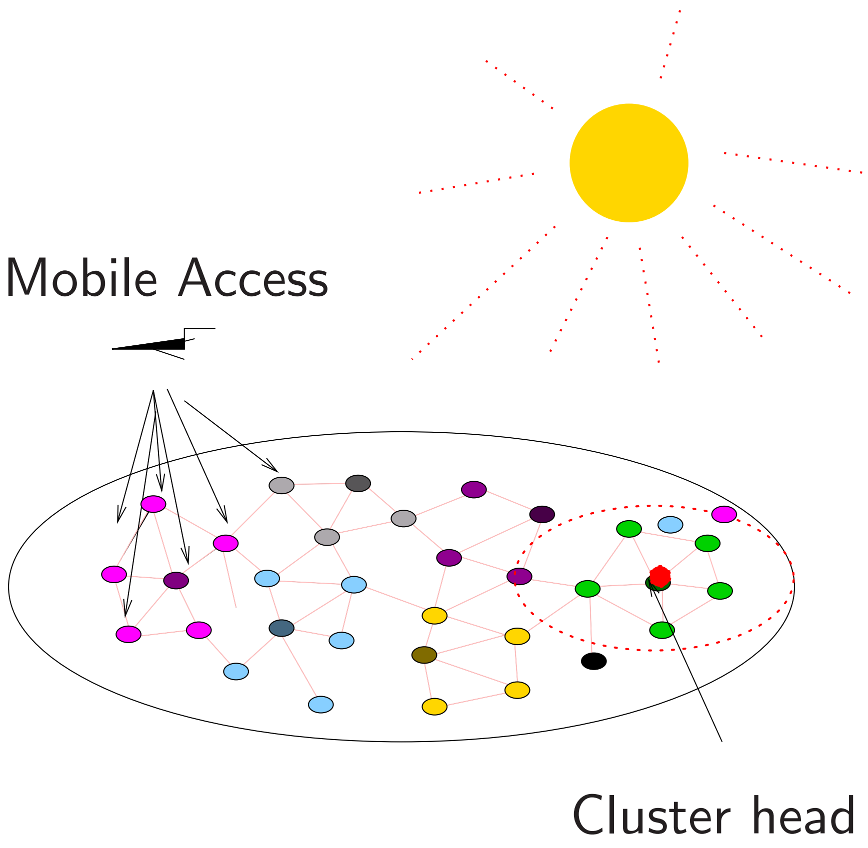
Joint work with Gohkan Mergen

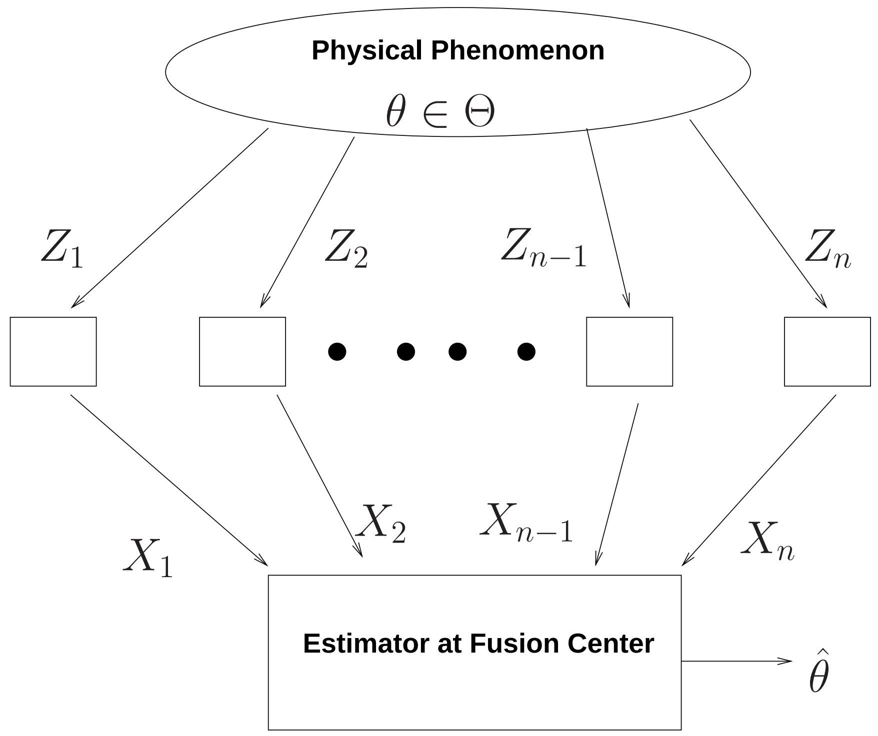


References

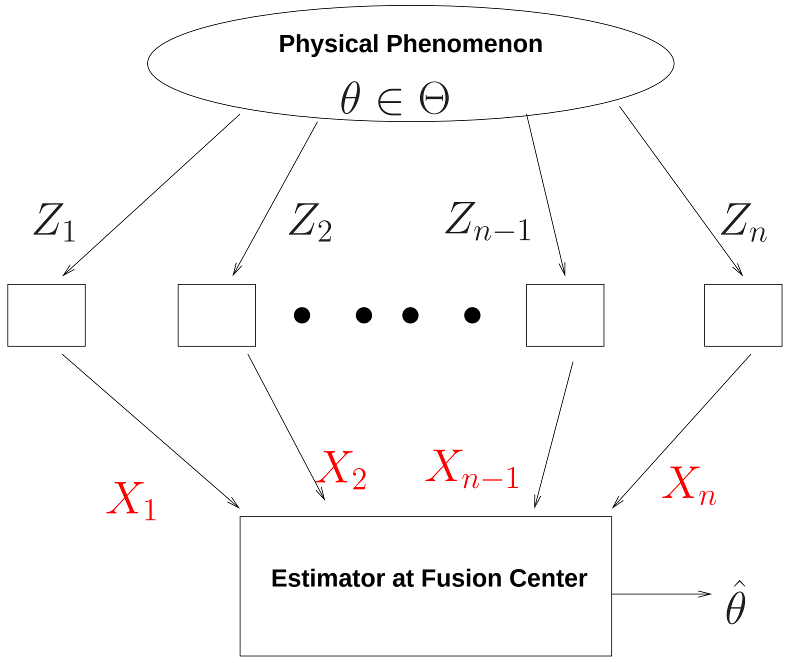
[1] G. Mergen and L. Tong, "Type-based Estimation over Multiaccess Channels," to appear in *IEEE Transactions on Signal Processing*, 2005. See also Allerton 2004.

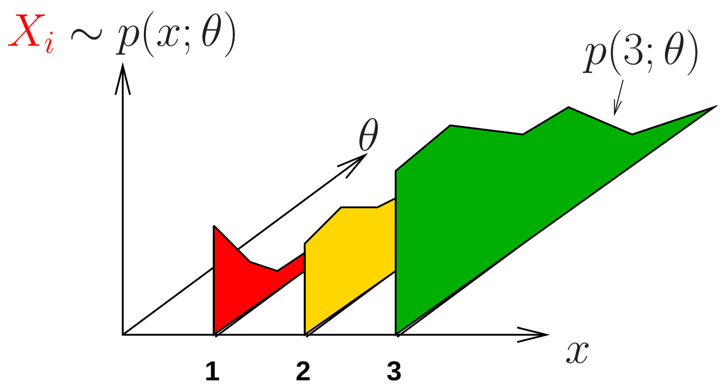
Distributed Parameter Estimation



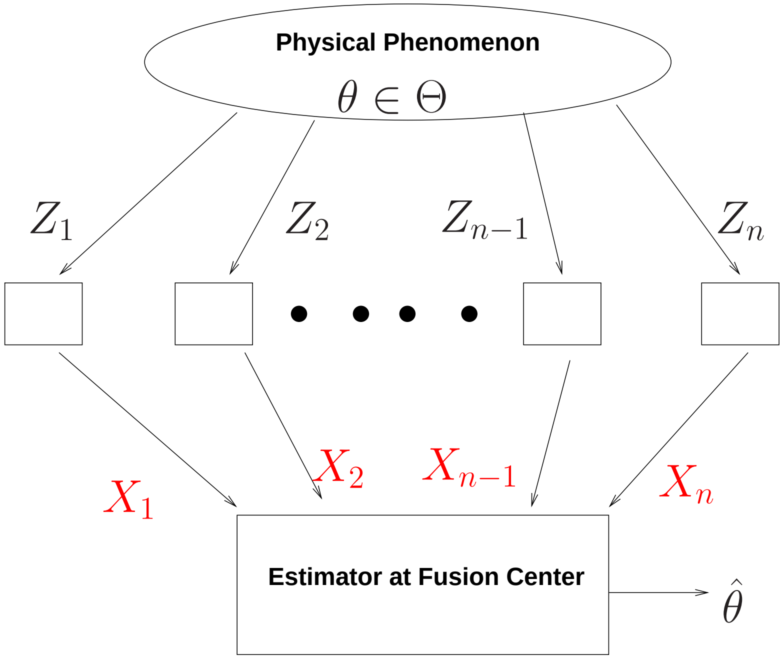


Distributed Parameter Estimation

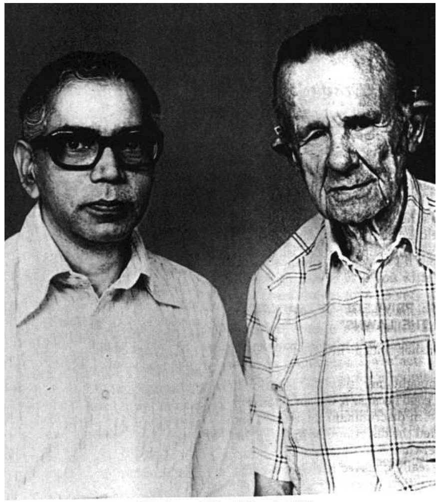




Distributed Parameter Estimation



$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial^2\theta}\ln p(x;\theta)\right)$$



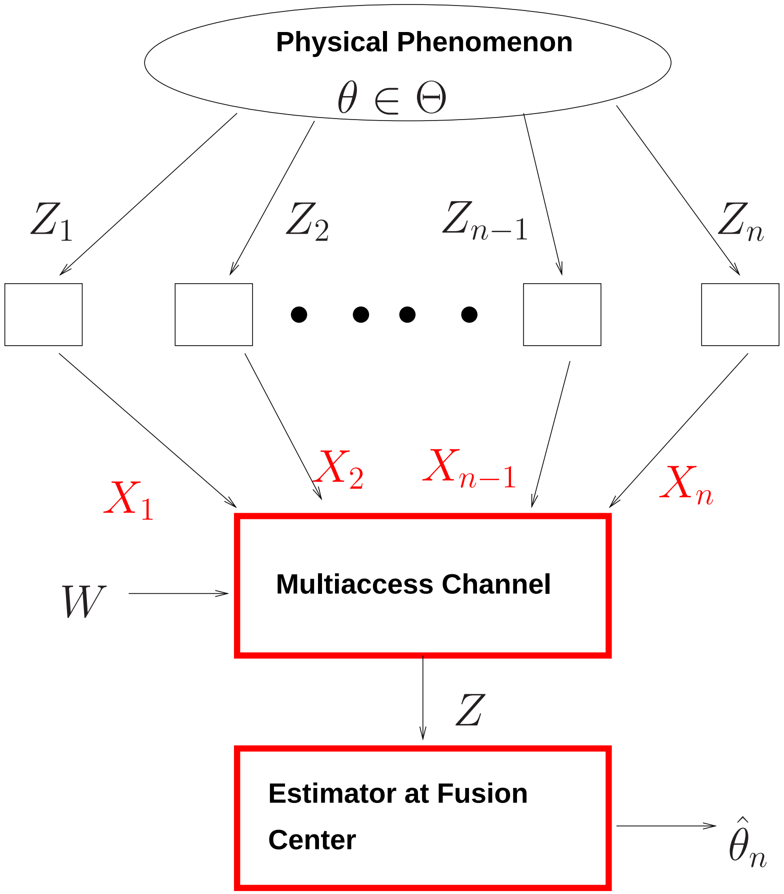
C. R. Rao and Harald Cramér, 1978

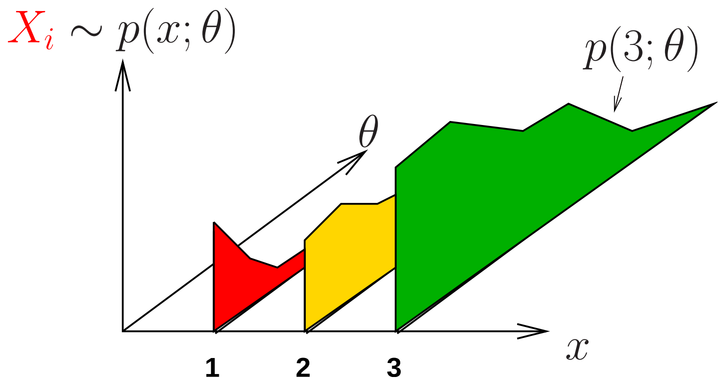
If we have perfect access to $\{X_i\}$,

$$\mathbb{E}\{(\hat{\theta}_n - \theta)^2\} \geq \frac{1}{nI(\theta)},$$

where the Fisher information $I(\theta)$ measures the **efficiency**.

What If No Perfect Access to Sensors?





$$I_Z(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial^2\theta}\ln p(z;\theta)\right) < I_X(\theta)$$

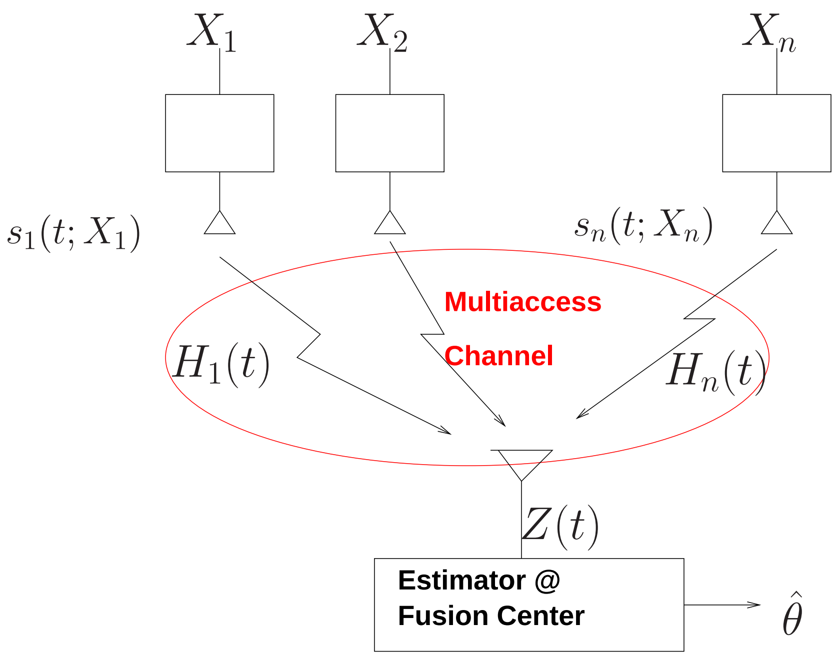
Estimating θ from Z,

$$\mathbb{E}((\hat{\theta}_n - \theta)^2) \geq \frac{1}{nI_Z(\theta)} \geq \frac{1}{nI_X(\theta)}$$

We expect MAC and noise increase MSE.

Estimation over Multiaccess Channels

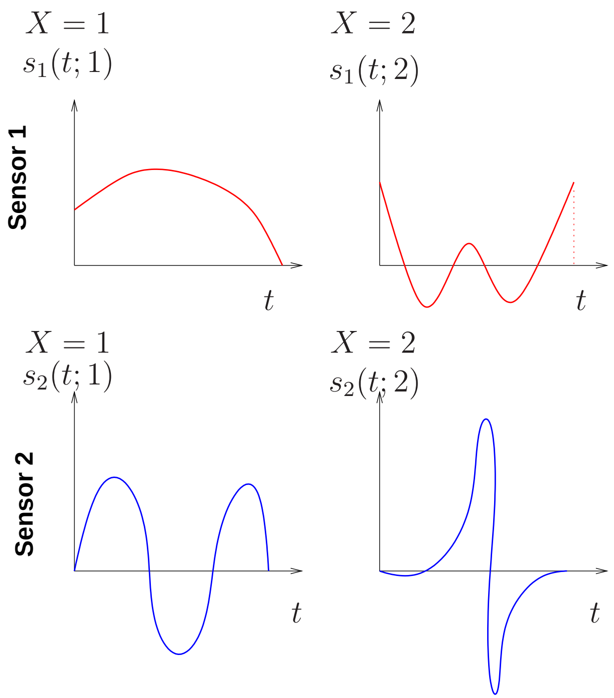
The Multiaccess Channel



$$Z(t) = \sum_i H_i(t) * s_i(t; X_i) + N(t)$$

Sensor Signaling Design

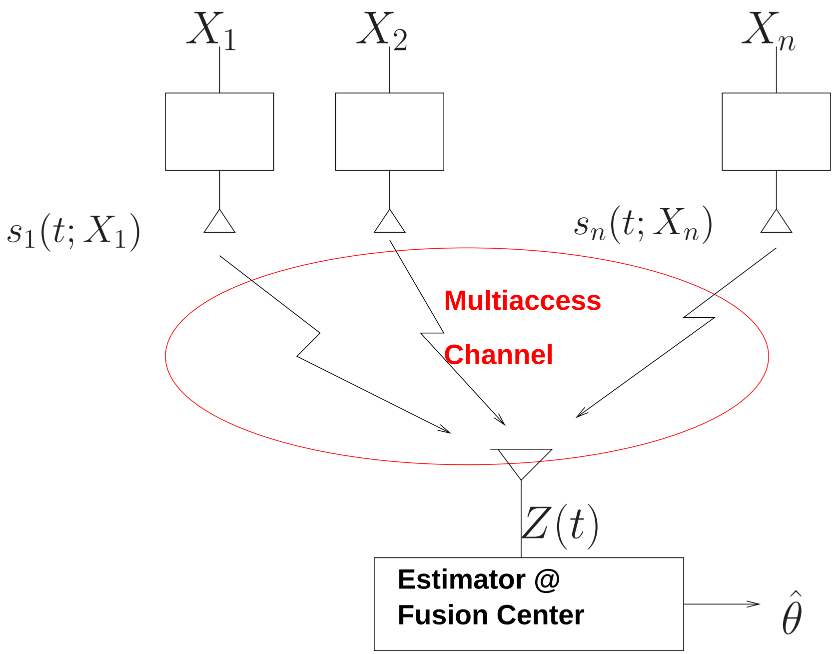
Encode $X_i = x$ to waveform $s_i(t;x)$ subject energy constraint.



Estimation over multiaccess channel is a joint design of signaling, multiaccess, and signal processing.

Estimation over MAC: The Layered Approach

The Multiaccess Channel



$$Z(t) = \sum_i s_i(t; X_i) + N(t)$$

A Layered Approach

- Encode each X_i into bits.
- Efficient modulation + Error control
- Medium Access Control (MAC)
  TDMA, FDMA, CDMA.
- Demodulation and Estimation

$$Z(t) \rightarrow \{\hat{X}_i\} \rightarrow \hat{\theta}$$

Advantages:

- Modular, well understood, simple.

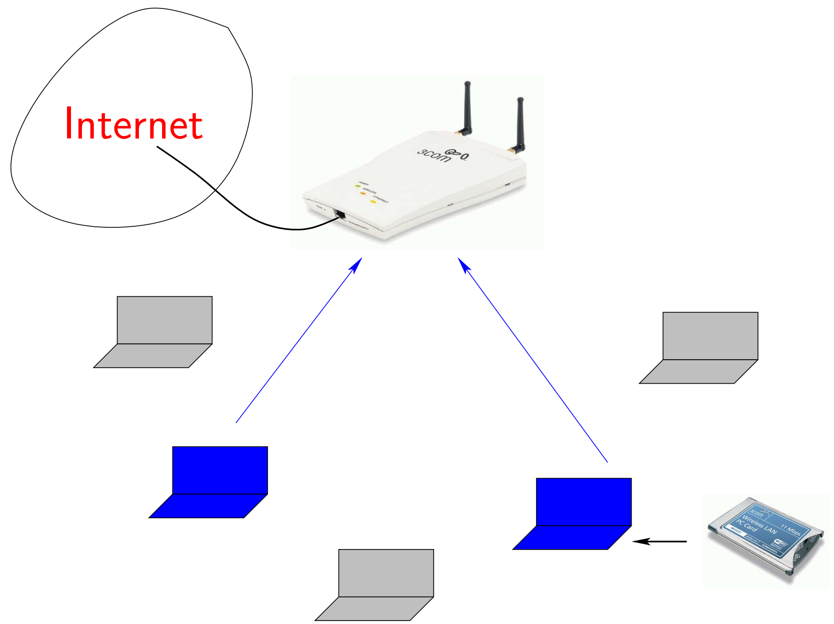
Caveat:

- Scalability: BW $\propto$ # users.
- More importantly, ...

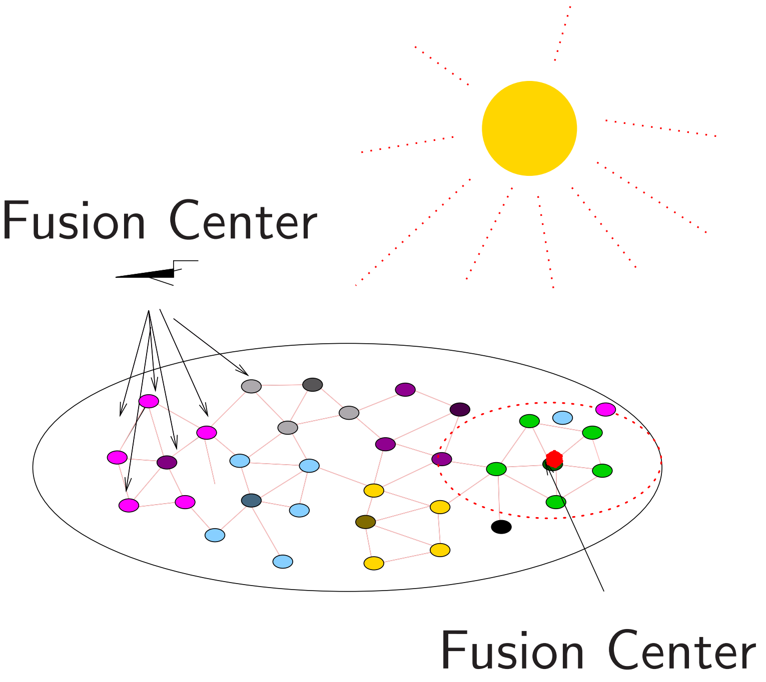
The layered approach ignores data dependencies.

From User Centric to Data Centric

User centric MAC



Data Centric MAC



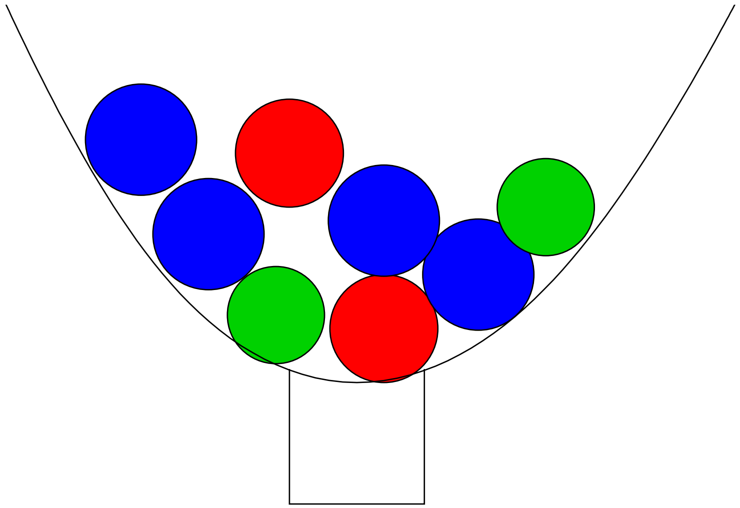
Design paradigm

- ☐ Allocate resources to users: time, frequency, etc....
- ☐ Maximize the rate region.

Design paradigm

- ☐ Allocate resources to data: time, frequency, etc....
- ☐ Optimize inference performance.

Urns and Balls

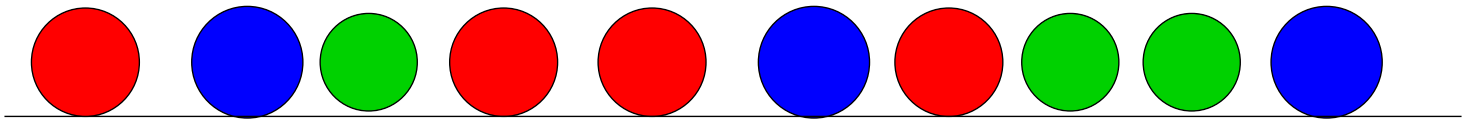


$\Pr(\text{Red}) = \theta_1$

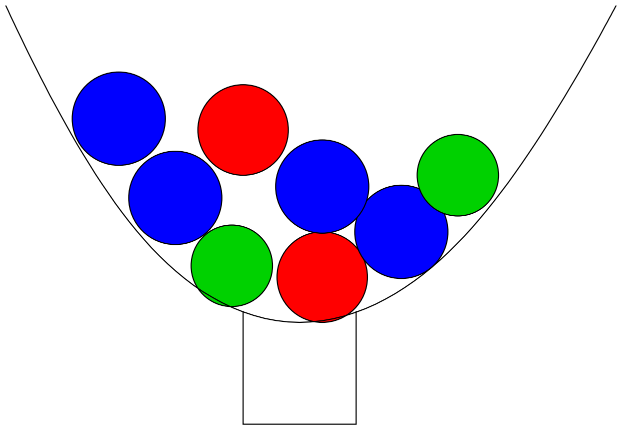
$\Pr(\text{Blue}) = \theta_2$

$\Pr(\text{Green}) = 1 - \theta_1 - \theta_2$

Realization $\mathbf{x}$

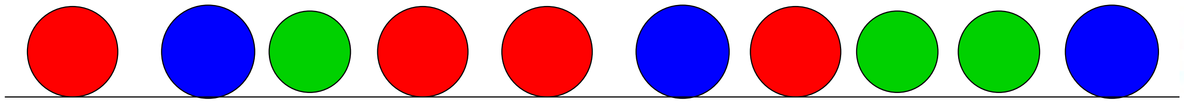


Empirical Measure and Type



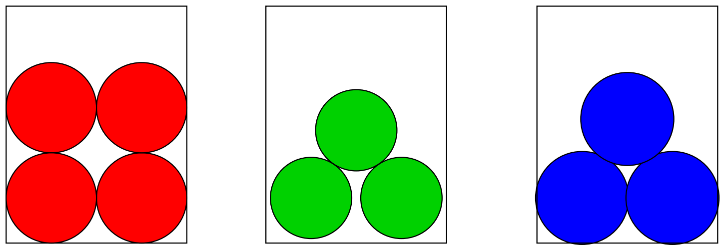
$\Pr(\text{Red}) = \theta_1$
$\Pr(\text{Blue}) = \theta_2$
$\Pr(\text{Green}) = 1 - \theta_1 - \theta_2$

Realization $\mathbf{x}$



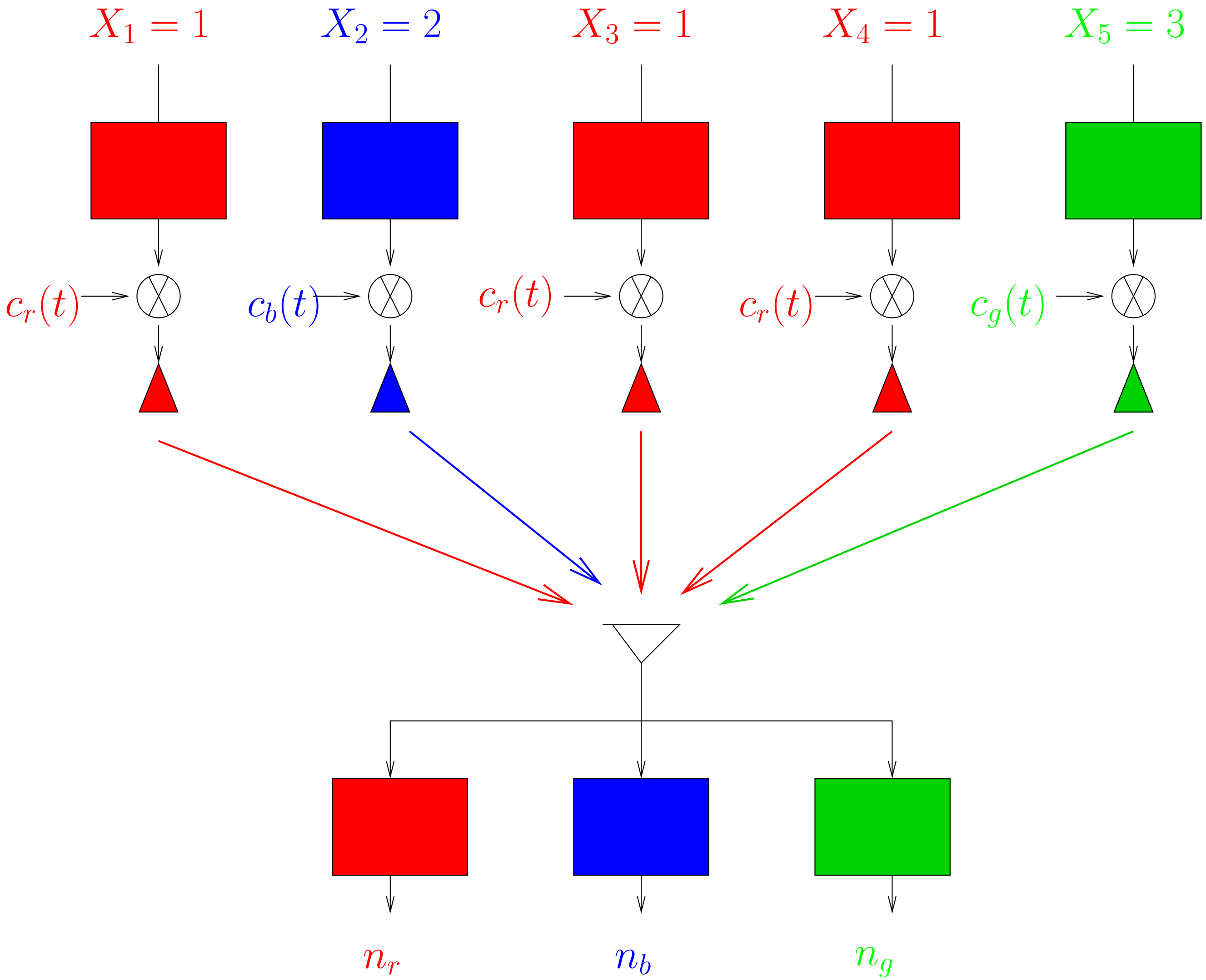
Type (Empirical Measure)

$\mathbf{P}_{\mathbf{x}} = (\frac{4}{10}, \frac{3}{10}, \frac{3}{10})$



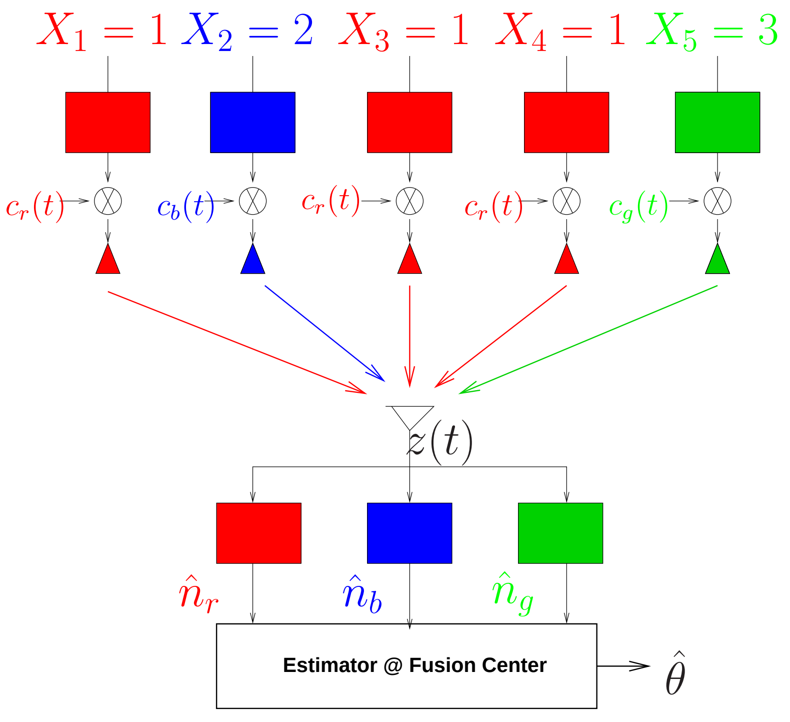
Type P_{x} gives sufficient statistics. Thus it suffices to transmit P_{x}.

TBMA: Type Based Multiple Access



TBMA delivers the (noisy) empirical distribution

Advantages and Caveats of TBMA



References

Liu-Sayeed: Allerton'04

Mergen-Tong: Allerton'04, TSP'04
ICASSP'05

TDMA vs. TBMA

□ Scalability:

$$\text{BW} \propto n \quad \text{vs.} \quad \text{BW} \propto K$$

□ Asymptotic Optimality:

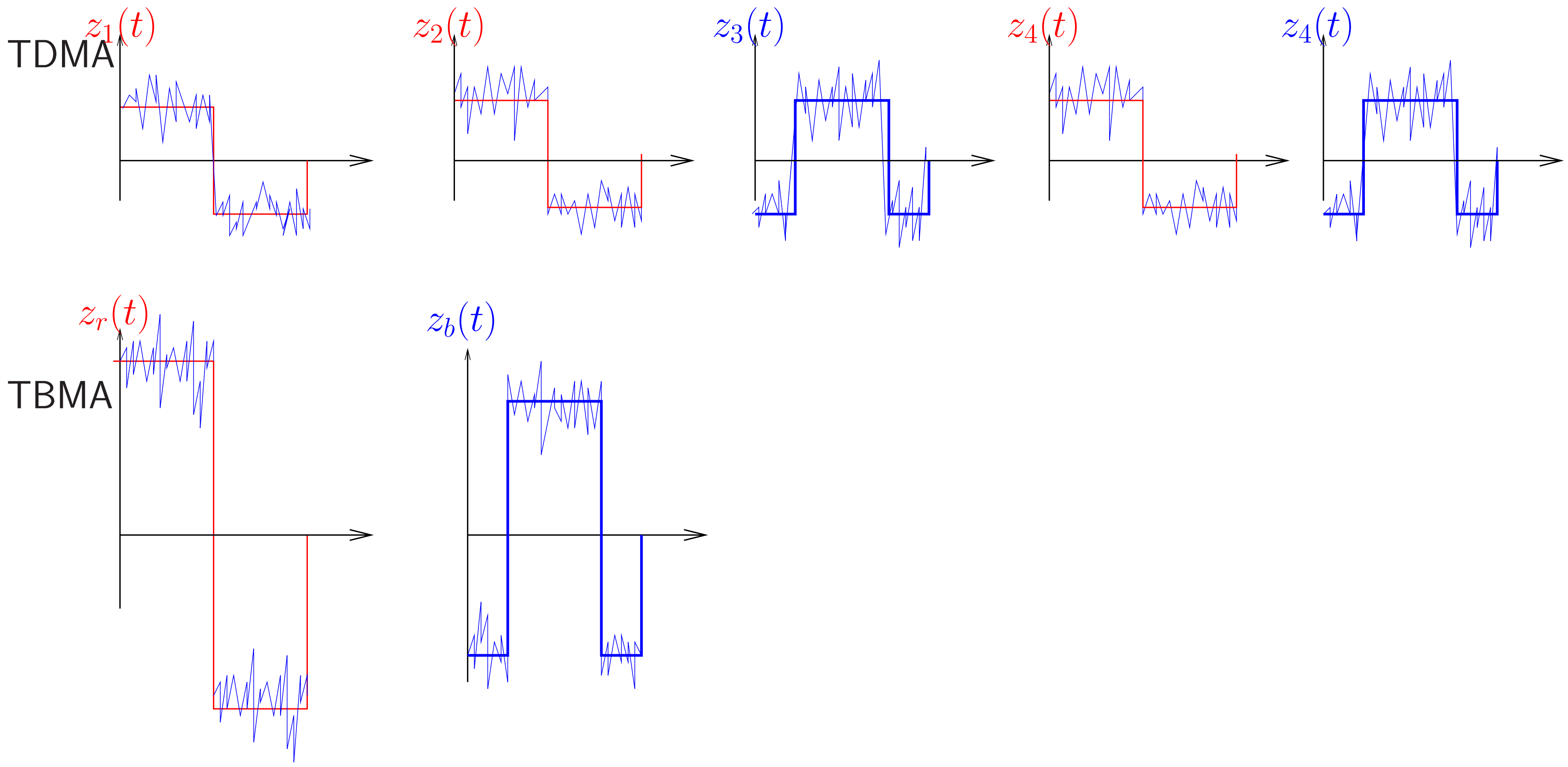
$$\hat{\theta}_n \sim \mathcal{N}(0, \frac{1}{nI_Z(\theta)}) \quad \text{vs.} \quad \hat{\theta}_n \sim \mathcal{N}(0, \frac{1}{nI_X(\theta)})$$

Performs as if $\{X_i\}$ are accessible directly.

Caveats

□ Data types must add coherently.
□ Must gain synchronization.
□ Need to deal with fading.

The Power of Coherent Combining



Steps to Optimality

- TBMA delivers noisy types:

$$\mathbf{Z} = \mathbf{P}_x + \mathbf{W} \sim \mathcal{N}(\mathbf{p}(\cdot;\theta), \frac{1}{n}\mathbf{\Sigma}(\theta))$$

- The Likelihood Function

$$f(\mathbf{z}|\ \theta) = \exp\left(-\frac{n}{2}\sum_{i=1}^{K}\frac{(p(i;\theta)-z_i)^2}{p(i;\theta)} + \log\sqrt{\prod_{i=1}^{K}p(i;\theta)}\right) g(\mathbf{z})$$

- An Asymptotic ML Estimator

$$\hat{\theta}_n = \arg\min_{\theta}\sum_{i=1}^{k}\frac{(p(i;\theta)-z_i)^2}{p(i;\theta)}$$

- Convergence

$$\hat{\theta}_n \xrightarrow{p} \theta, \qquad \sqrt{n}(\hat{\theta}_n - \theta) \rightarrow \mathcal{N}(0, \frac{1}{I_X(\theta)})$$

Observations

- Less likely samples weight more!

$$\hat{\theta}_n = \arg\min_{\theta} \sum_{i=1}^{k} \frac{(p(i;\theta) - z_i)^2}{p(i;\theta)}$$

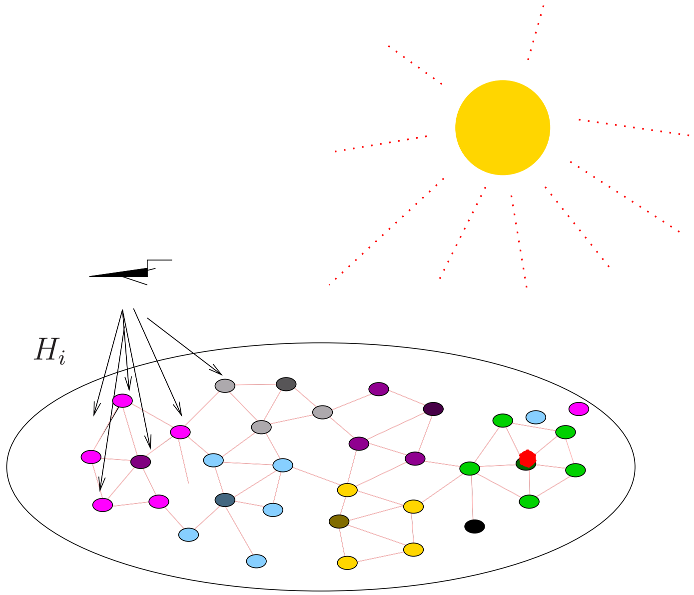
- Works as if having direct access to X_i's

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow \mathcal{N}(0, \frac{1}{I(\theta)})$$

- The theorem holds for any noise power.
  The σ^2 determines the speed of convergence to the asymptotic MSE.

TBMA over Fading Channels

Multiaccess Fading Channel

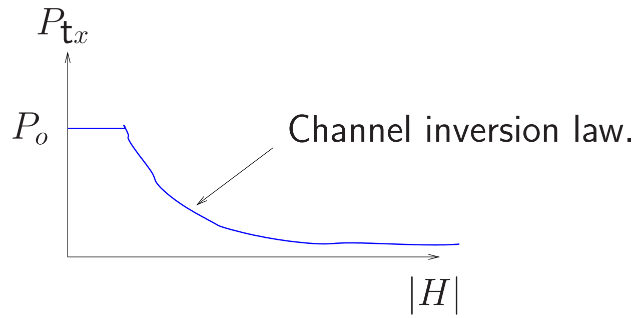


$$Z(t) = \sum_i H_i s(t; X_i) + N(t)$$

Fading Characteristics

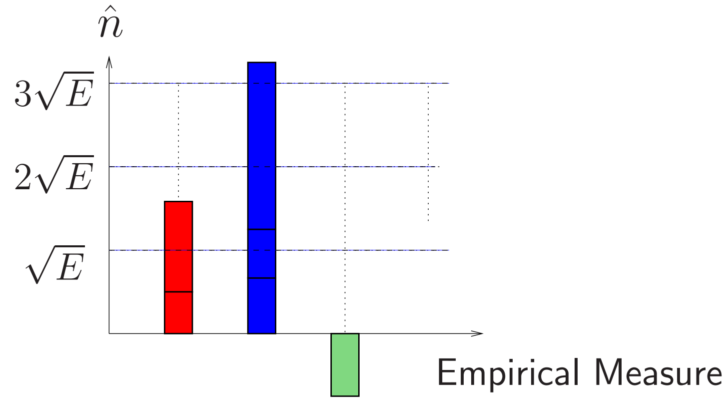
- Random vs. Deterministic
- Ergodic vs. Nonergodic
- Knowledge of channel state.

TBMA with TX CSI



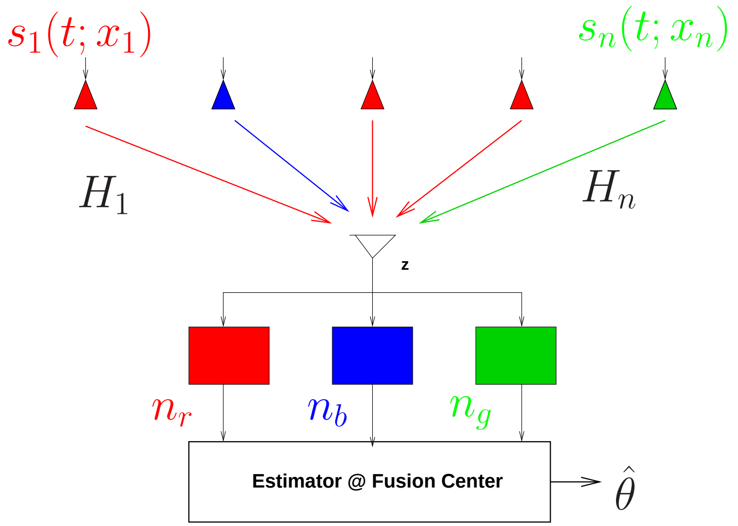
Optimality can be retained.

TBMA without CSI



There is a price to pay.....

TBMA over Fading Channels



We consider flat fading channel

$$Z(t) = \sum_i H_i s(t; X_i) + N(t)$$

where H_i are i.i.d., known neither at the transmitter nor the receiver.

Loss due to Fading

If $\mathbb{E}(H_i) \neq 0$, then

$$\hat{\theta}_n \rightarrow \theta \text{ in p}$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow \mathcal{N}(\theta, (1 + \frac{\text{var}(H_i)}{\mathbb{E}^2(H_i)})I(\theta))$$

The MSE increases by a factor

$$G = 1 + \frac{\text{var}(H_i)}{\mathbb{E}^2(H_i)}$$

The Good and Bad

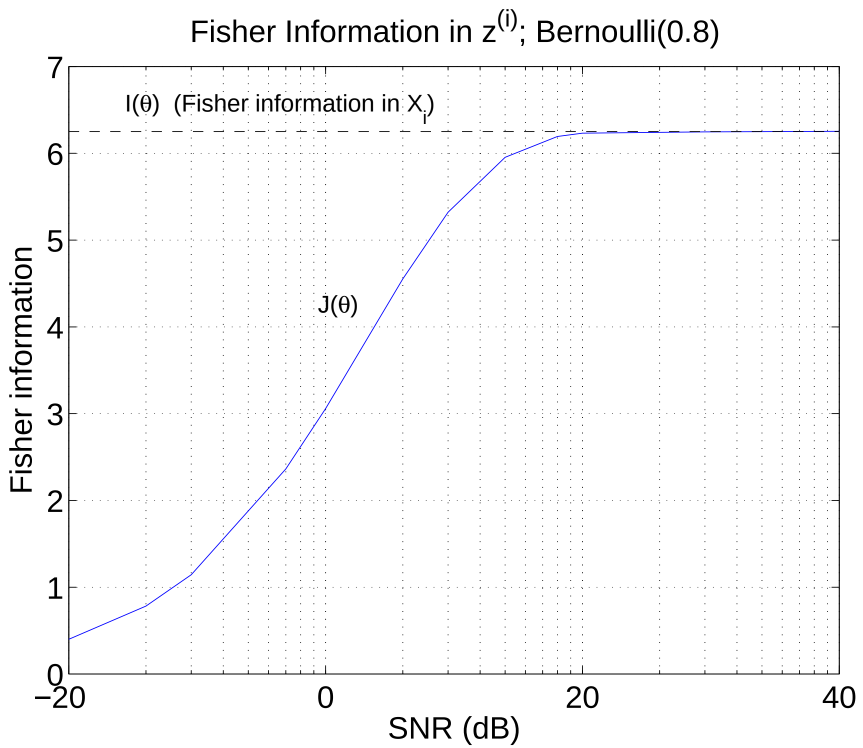
- ☐ Noncoherent estimation provides

$$\text{MSE}(\hat{\theta}) \sim O(\frac{1}{n})$$

- ☐ When $\mathbb{E}(H_i) = 0$, TBMA does not give consistent estimate.

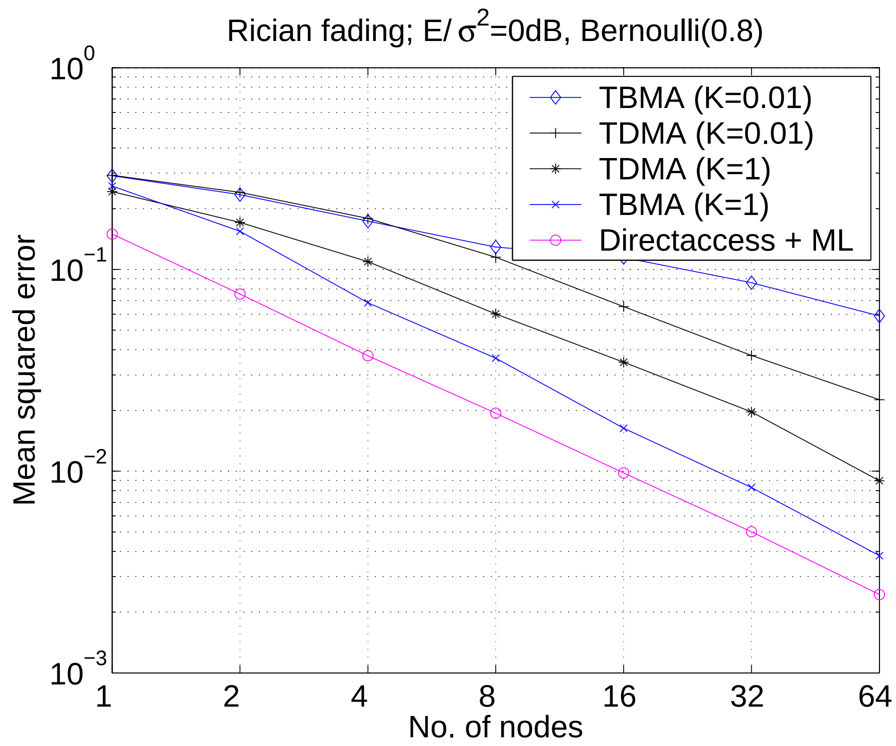
My Curve (TBMA) vs. Your Curve (TDMA)

Mine is Better



The gain is substantial at low SNR

Mine may be worse...



As $\mathbb{E}(H) \to 0$, TBMA deteriorates...

Route Selection for Detection in Sensor Networks

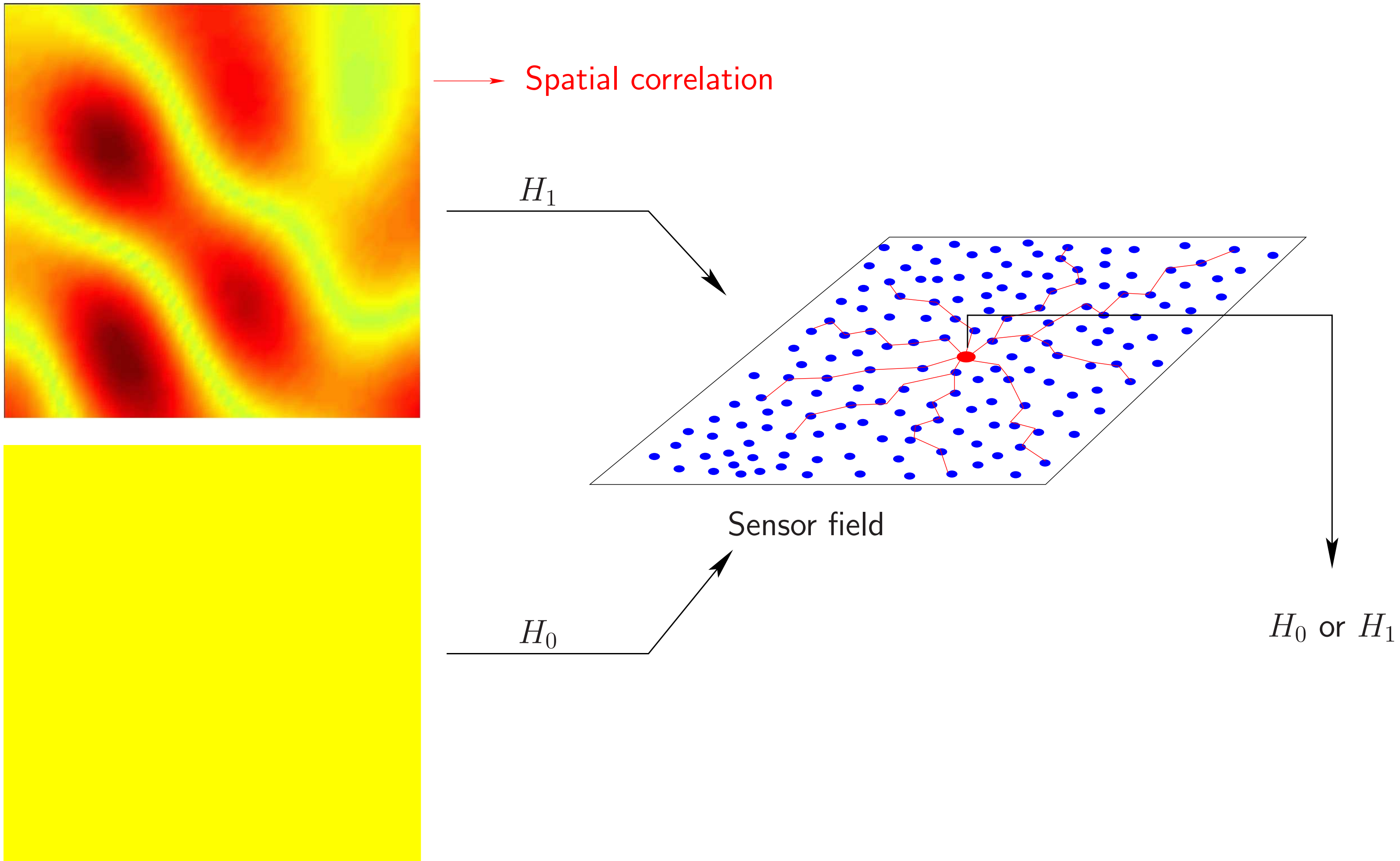
Joint work with
Youngchul Sung, A. Ephremides, and H. Vince Poor

References

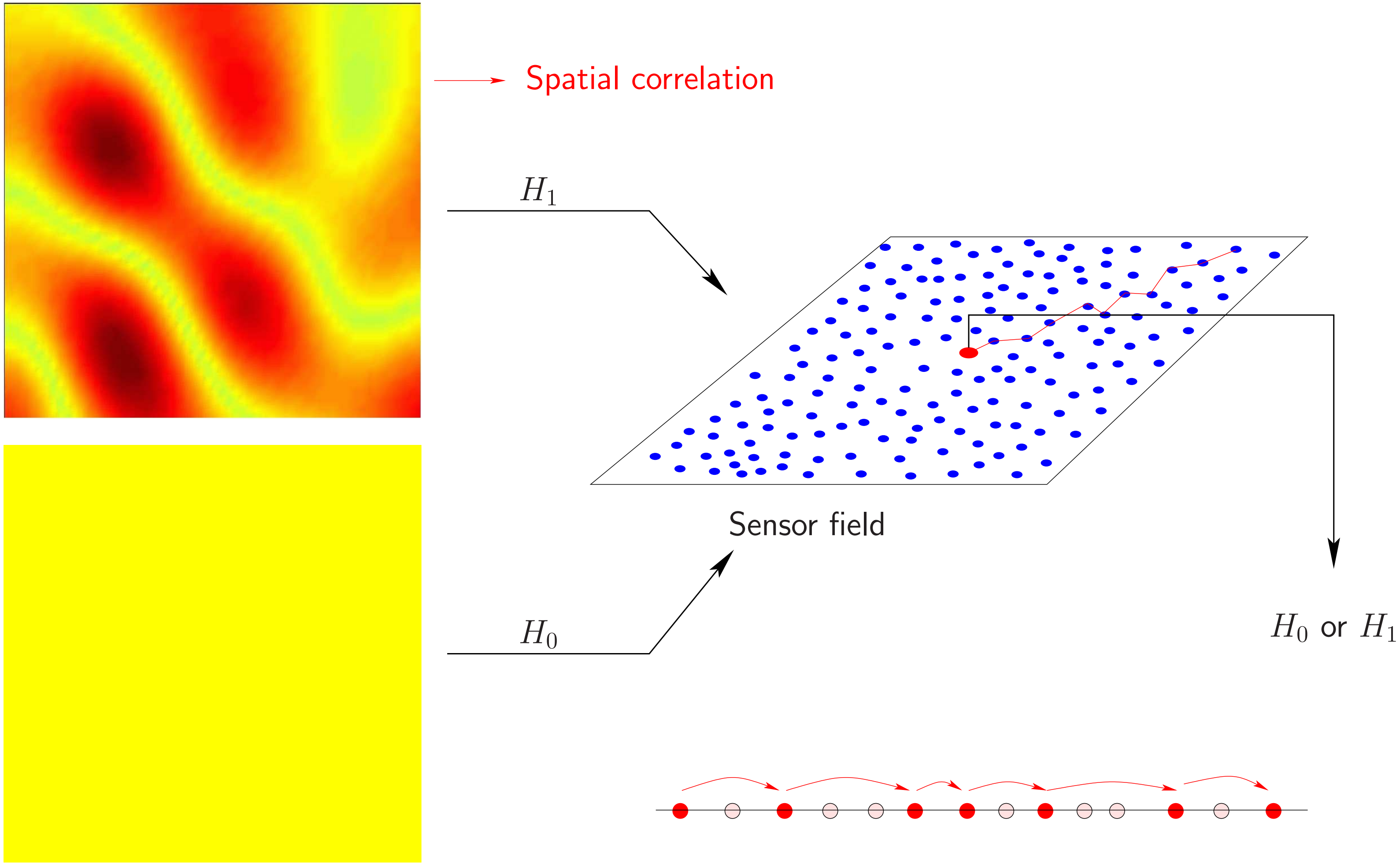
[1] Y. Sung, L. Tong, and A. Ephremides, "Routing for detection of correlated random fields in large sensor networks," CISS'05, to be submitted to *IEEE Transactions on Information Theory*.

[2] Y. Sung, L. Tong, and H. V. Poor, "Neyman-Pearson detection of Gauss-Markov signals in noise: Closed-form error exponent and properties," submitted to *IEEE Transactions on Information Theory*, Nov., 2004

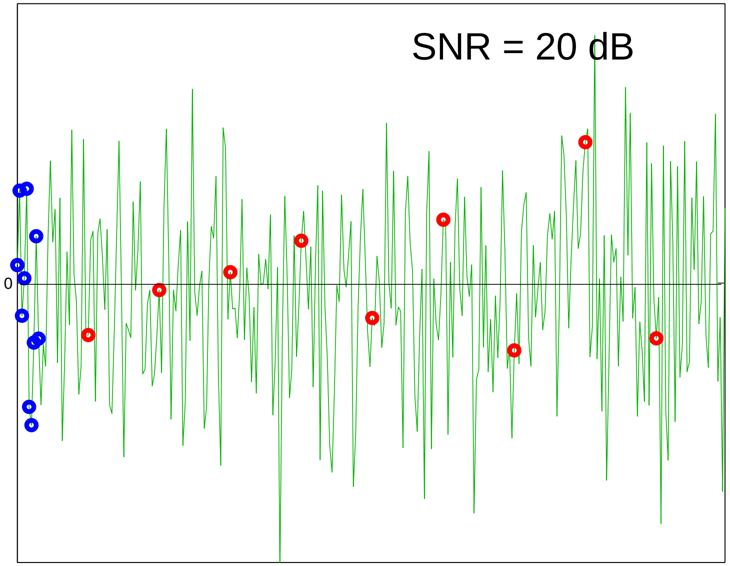
Sensor Route Selection



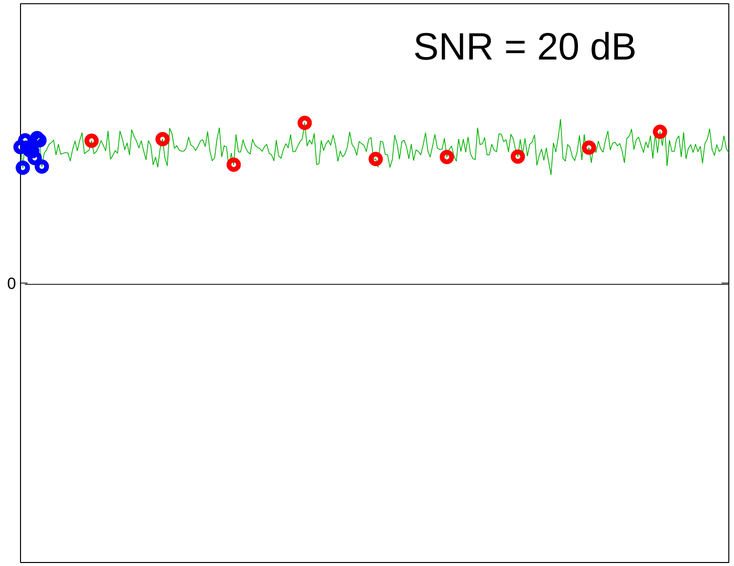
Sensor Relay and Data Aggregation



Where Should We Collect Data?

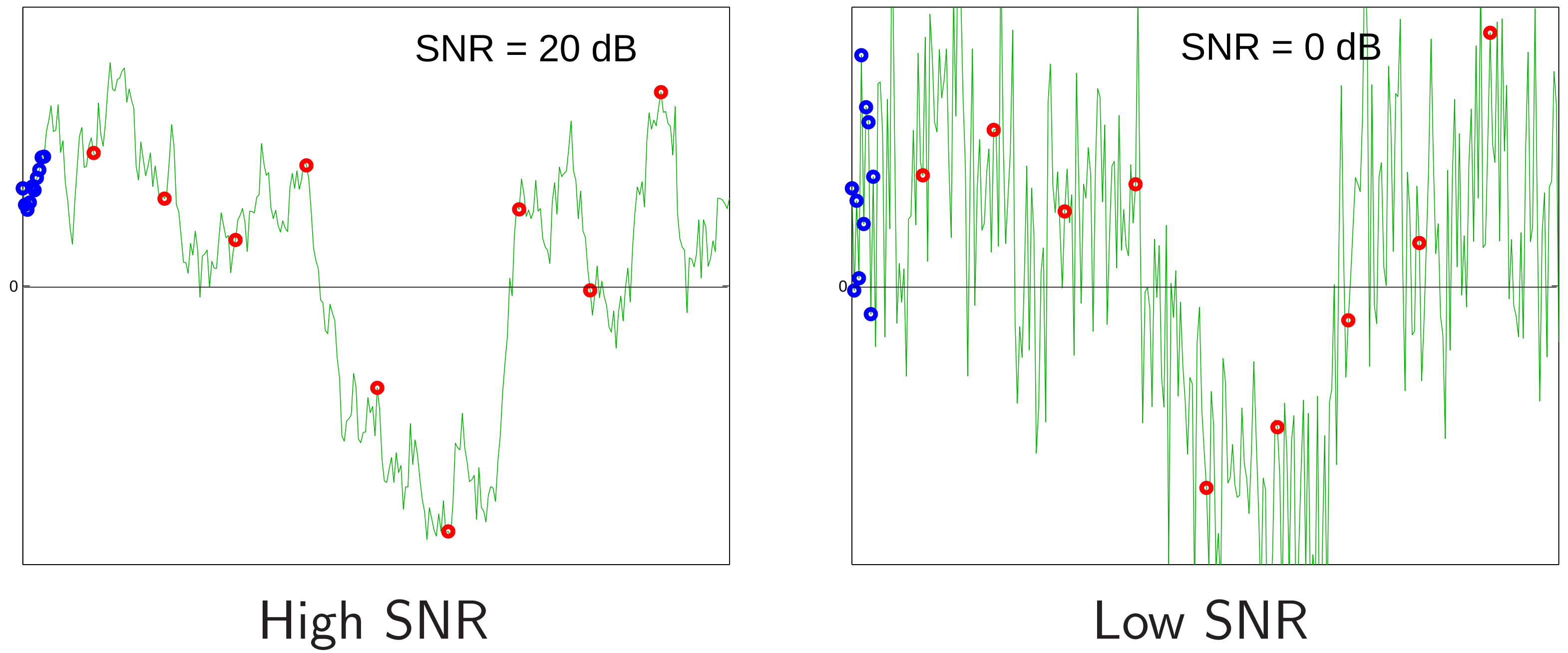


I.i.d. signal



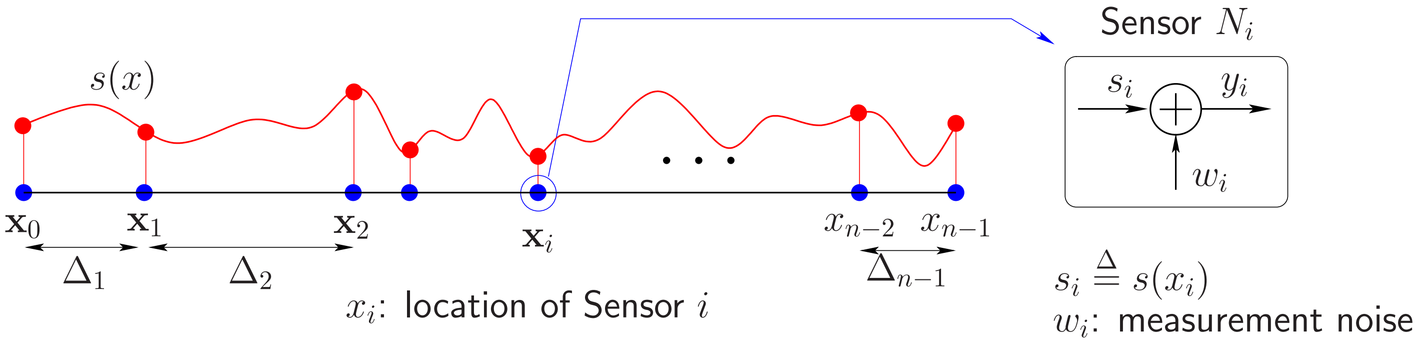
Perfectly correlated signal

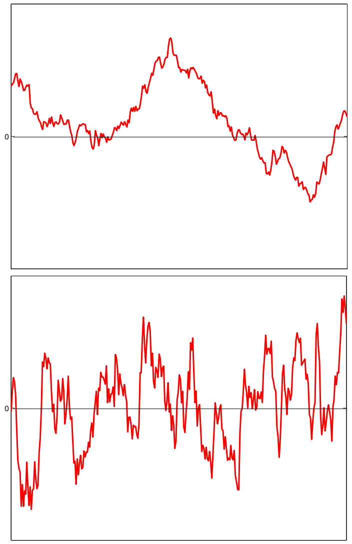
Energy-Performance Trade-off



It is a tradeoff between **diversity** and **coherency**

Hypotheses on a Given Route





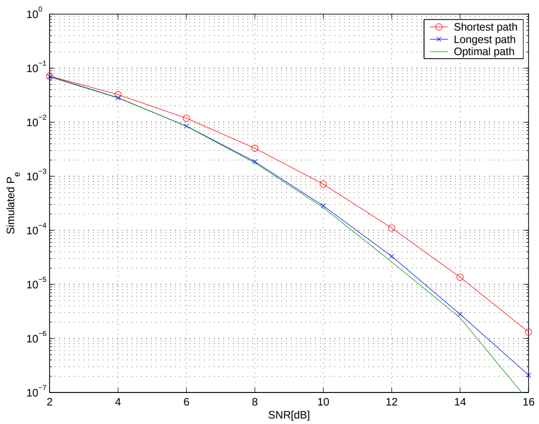
$$\mathcal{H}_0 : \quad y_i = w_i \quad \text{vs.} \quad \mathcal{H}_1 : \begin{cases} s_{i+1} = a_i s_i + u_i, \\ y_i = s_i + w_i, \end{cases}$$

where $a_i = e^{-A\Delta_i}$ characterizes the signal correlation under $\mathcal{H}_1$.

Performance and Route Selection

The Bayesian Detector

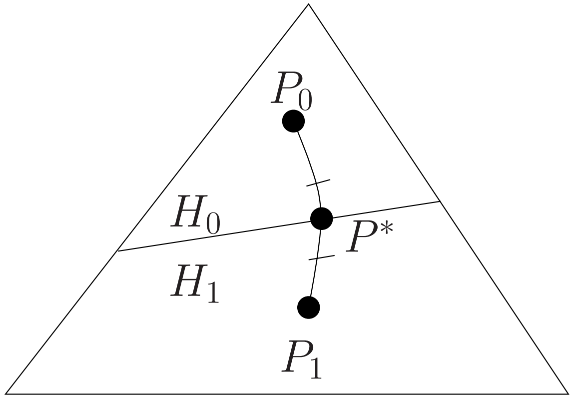
$$\log \frac{p_{1,n}}{p_{0,n}}(y_0, \cdots, y_{n-1}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau_n,$$
$$P_e = \pi_0 P_F + \pi_1 P_M \quad (1)$$



- Number of sensors along the route
- Geometry of the route
- Field correlation
- Signal-to-noise ratio (SNR)

How do we incorporate detection performance into route metric?

Chernoff Bound and Information



The Chernoff bound is given by

$$\begin{aligned} P_e &= \pi_0 P(\mathcal{E}|H_0) + \pi_1 P(\mathcal{E}|H_1) \\ &\leq \mathbb{E}(-C(P_0, P_1)) \end{aligned}$$

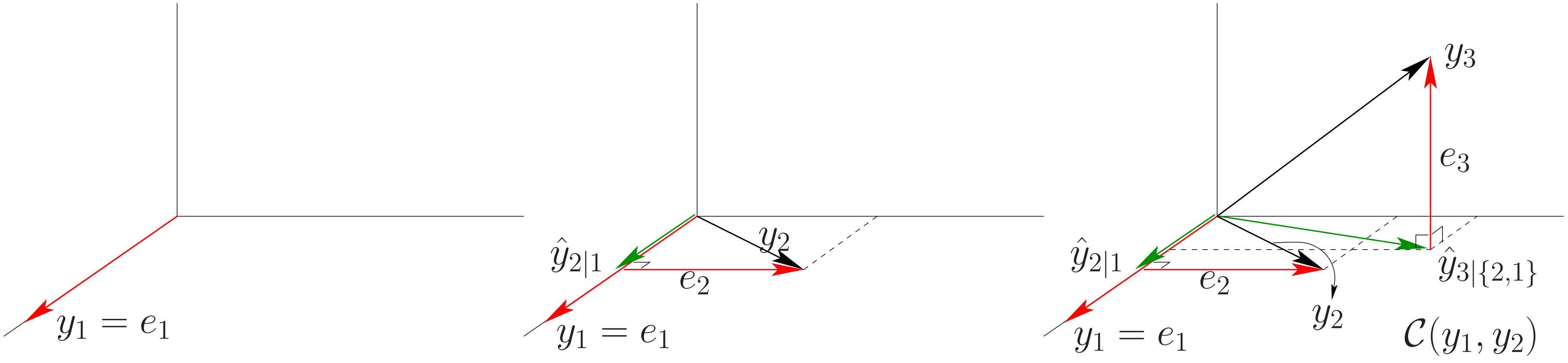
where $C(P_0, P_1)$ is the Chernoff information

$$\begin{aligned} C(P_0, P_1) &\triangleq \sup_{0\leq s\leq 1} -\log \mathbb{E}_0 \left\{ \mathbb{E}\{s \log \frac{p_1(y_0, \cdots, y_{n-1})}{p_0(y_0, \cdots, y_{n-1})}\} \right\} \\ &= D(P^*||P_0) \end{aligned}$$

Remark:

The direct calculation of $C(P_0, P_1)$ is based on the eigenvalues of covariance matrix $\mathbf{R} = \mathbb{E}\{\mathbf{y}_n \mathbf{y}_n^T\}$, which does not give an additive link metric.

The Innovation Process



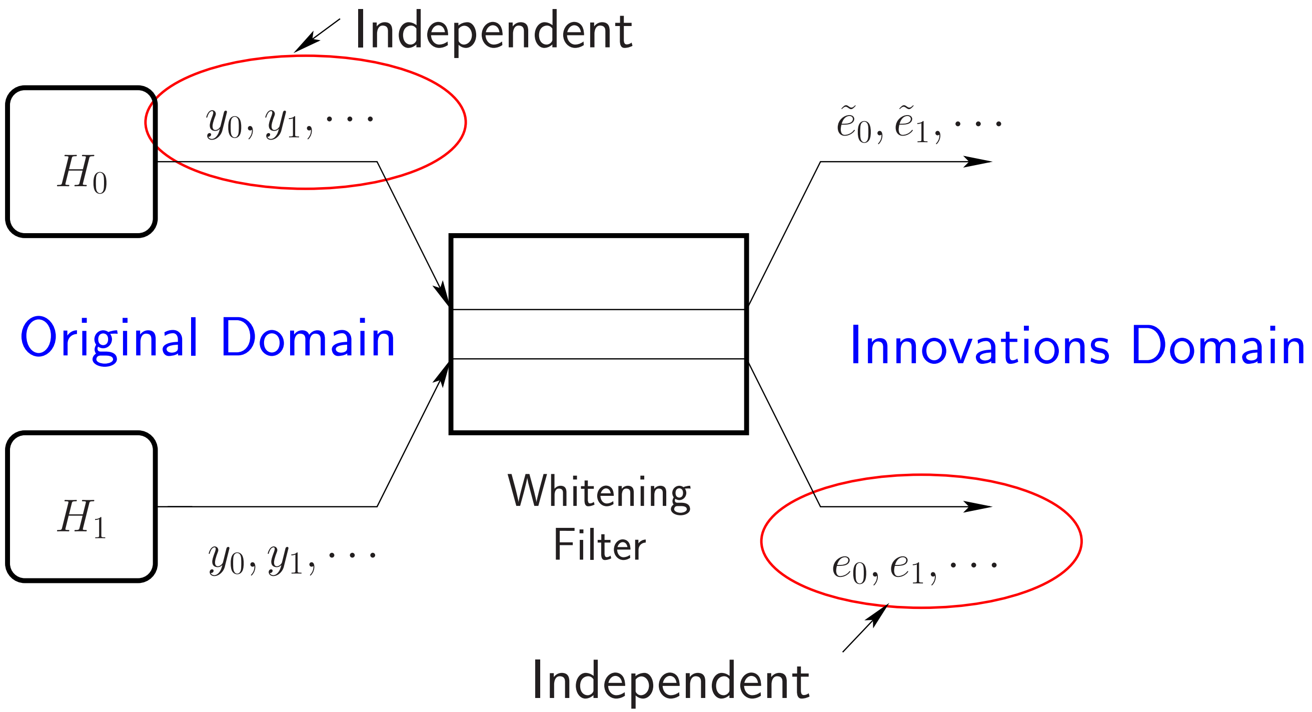
The prediction error

$$e_i \stackrel{\Delta}{=} y_i - \hat{y}_{i|\{1,\cdots,i-1\}}, \qquad e_i \sim \mathcal{N}(0, R_{e,i})$$

forms the innovation sequence, and $R_{e,i}$ the error variance.

$$\{y_1, y_2, \cdots, y_n\} \;\Leftrightarrow\; \{e_1, e_2, \cdots, e_n\}, \quad e_i \perp e_j, \quad \text{for all } i \neq j$$

Chernoff Information via Innovations Approach



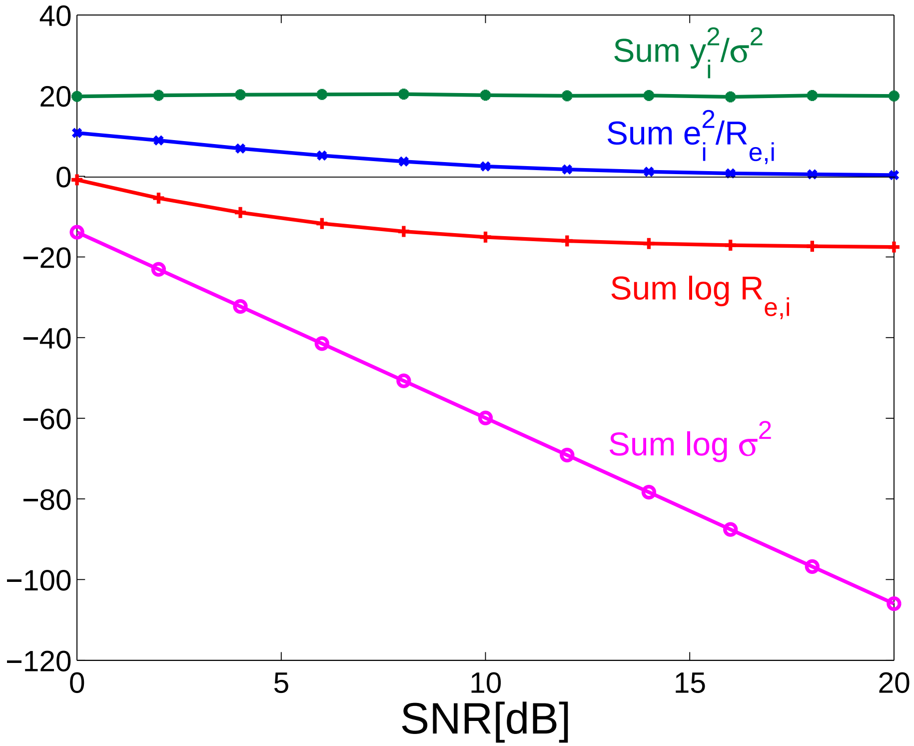
$$\log p_1(y_0, \cdots, y_{n-1}) = -\frac{1}{2}\sum_{i=0}^{n-1}\left(\log(2\pi R_{e,i}) - \frac{e_i^2}{R_{e,i}}\right),$$

$$\log p_0(y_0, \cdots, y_{n-1}) = -\frac{1}{2}\sum_{i=0}^{n-1}\left(\log(2\pi \sigma^2) - \frac{y_i^2}{\sigma^2}\right)$$

(Schweppe, 1965)

Chernoff Information via Innovations Approach

$$C(P_0,P_1) = \sup_{0\le s\le 1} -\log \mathbb{E}_0\left\{\mathbb{E}\left[s\left(-\frac{1}{2}\sum_{i=0}^{n-1}\log(2\pi R_{e,i}) - \frac{1}{2}\sum_{i=0}^{n-1}\frac{e_i^2}{R_{e,i}} + \frac{1}{2}n\log(2\pi\sigma^2) + \underbrace{\frac{1}{2}\sum_{i=0}^{n-1}\frac{y_i^2}{\sigma^2}}_{\to n/2}\right)\right]\right\}$$



Chernoff Information via Innovations Approach

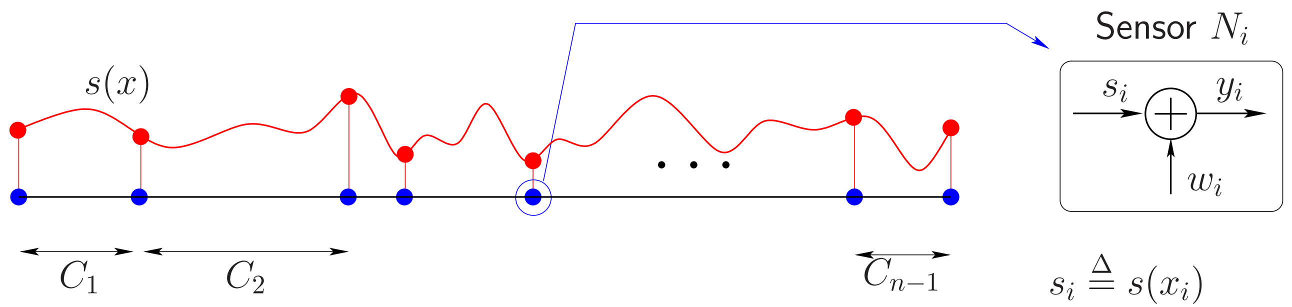
Chernoff bound at high SNR

$$P_e \leq B_c \approx \mathbb{E}\left\{-\sum_{i=0}^{n-1}\left[\frac{1}{2}\log\left(1+\frac{P_{e,i}}{\sigma^2}\right)-\frac{1}{2}\right]\right\}$$

$$P_{e,i} = \text{Variance of signal innovation } (s_i - \hat{s}_{i|i-1})$$

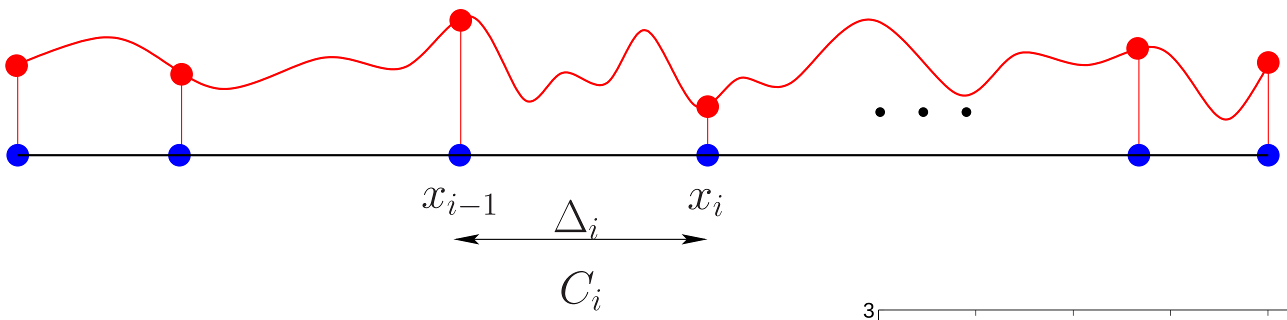
The Link Metric and Optimal Routing

$$C_i \triangleq \frac{1}{2}\log\left(1+\frac{P_{e,i}}{\sigma^2}\right)$$



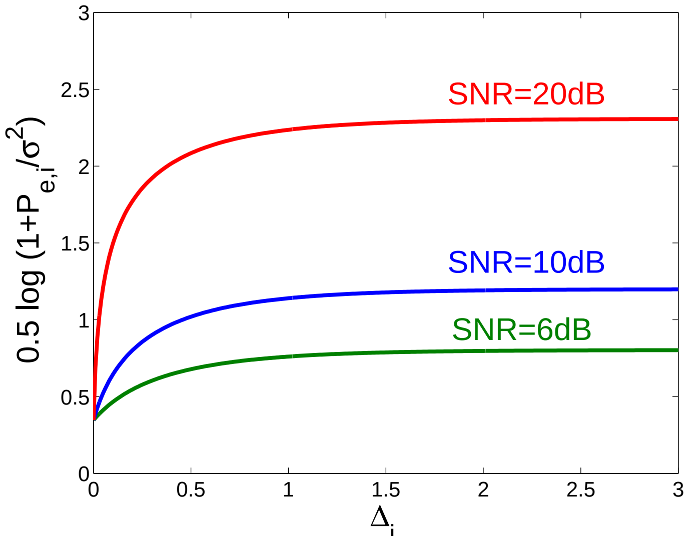
$s_i \triangleq s(x_i)$
w_i: measurement noise

Link Metric as a Function of Δ_i



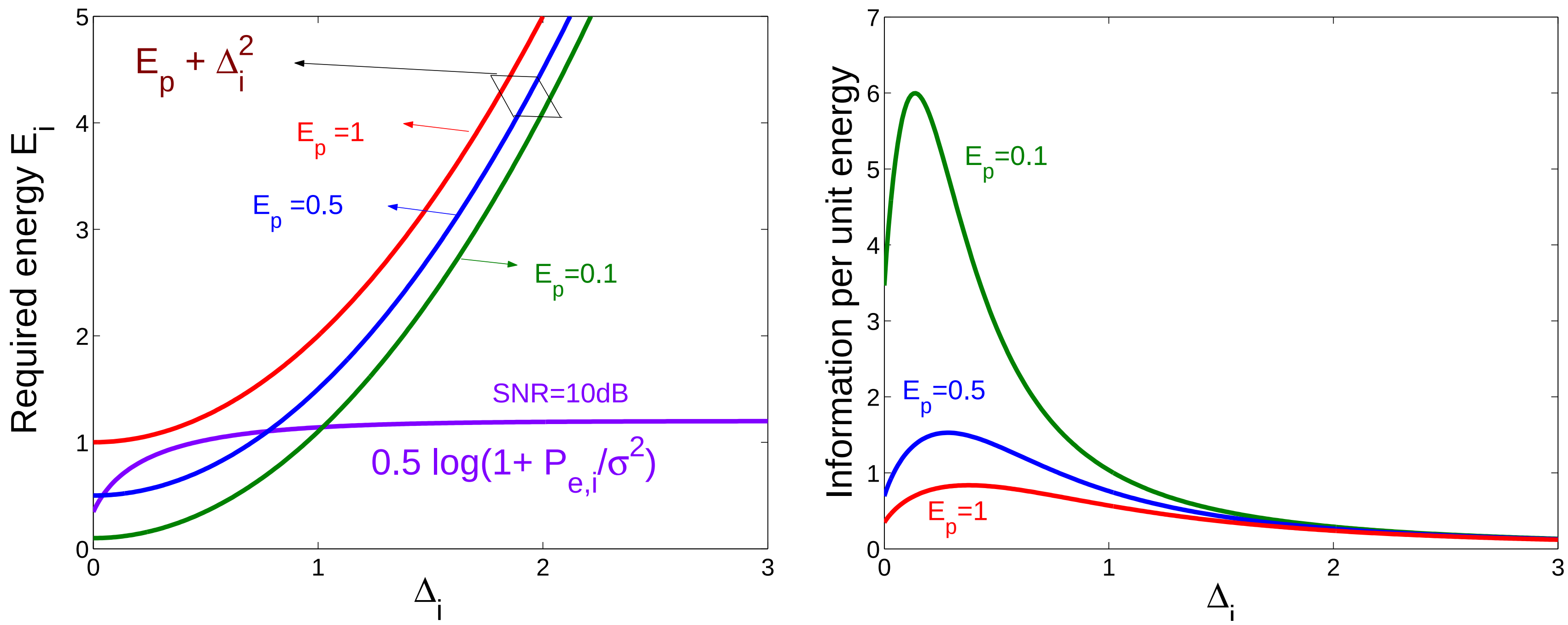
$$C_i(\Delta_i) = \frac{1}{2}\log\left\{\text{SNR} + 1 - (\text{SNR} - K_{i-1})e^{-2A\Delta_i}\right\}$$
$$\approx \frac{1}{2}\log\left\{\text{SNR} + 1 - (\text{SNR} - 1)e^{-2A\Delta_i}\right\}$$

$C_i(\Delta_i)$ is strictly increasing and concave.



Without energy constraint, maximize hop size.

Information Efficiency Per Link

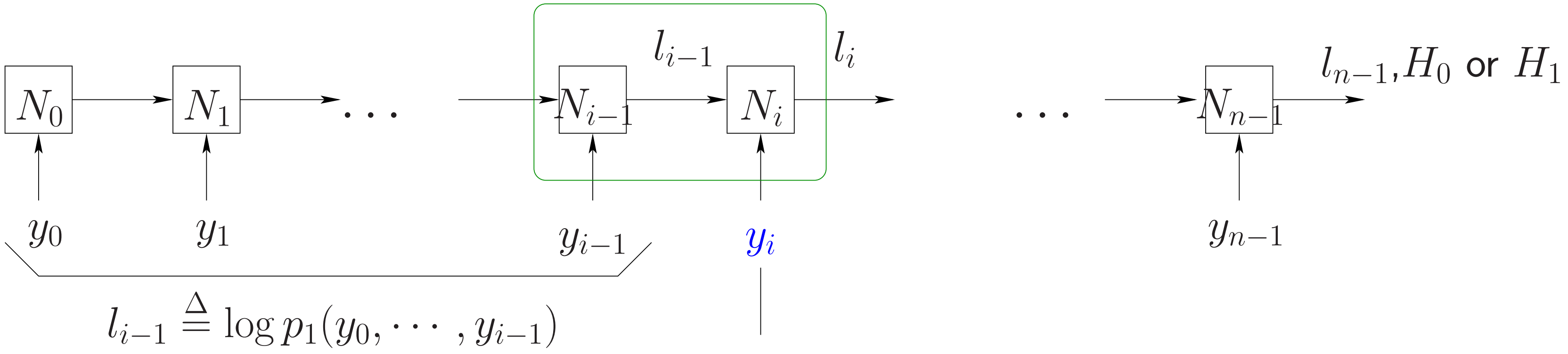


$$\eta = \frac{C}{E_p + \Delta^{\nu}}$$

E_p = Processing energy

ν = Propagation decay coefficient

Schweppe's LR Recursion



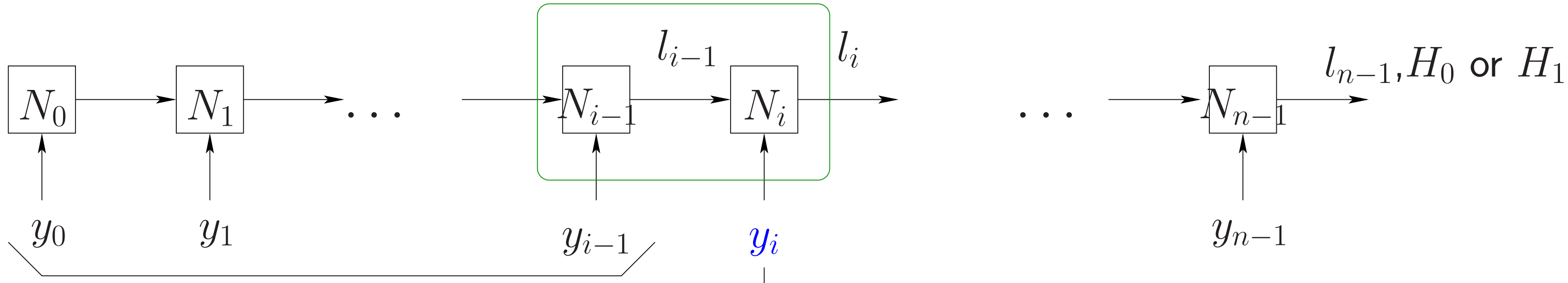
$$l_i = \log p_1(y_0, \cdots, y_i)$$

$$l_i = l_{i-1} + \log p_1(y_i|y_0, \cdots, y_{i-1})$$

$$p_1(y_i|y_1, \cdots, y_{i-1}) = \frac{1}{\sqrt{2\pi R_{e,i}}}\mathbb{E}(-\frac{1}{2}\frac{(y_i - \hat{y}_{i|i-1})^2}{R_{e,i}})$$

$$\begin{aligned} \hat{y}_{i|i-1} &\triangleq \mathbb{E}_1(y_i|y_0^{i-1}), && \text{linear MMSE estimate of } y_i \\ e_i &= y_i - \hat{y}_{i|i-1}, && \text{innovation of } y_i \\ R_{e,i} &= \mathbb{E}e_i^2, && \text{innovation variance.} \end{aligned}$$

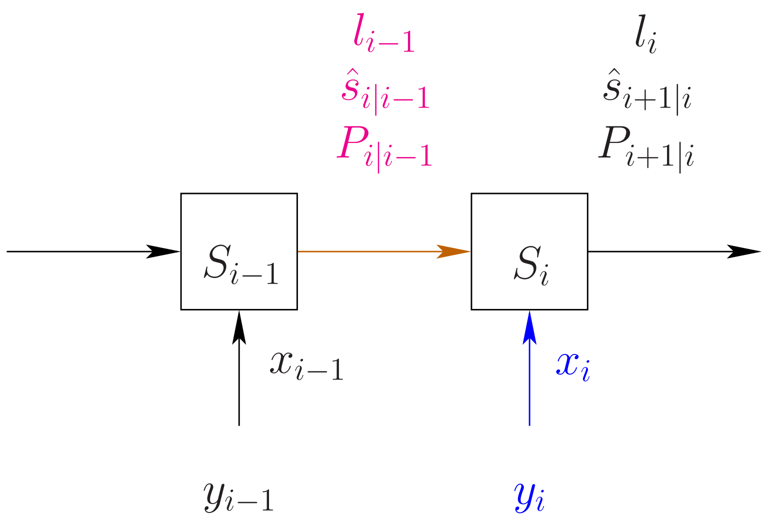
Kalman Recursion in Sensor Networks



$$l_{i-1} \triangleq \log p_1(y_0, \cdots, y_{i-1})$$

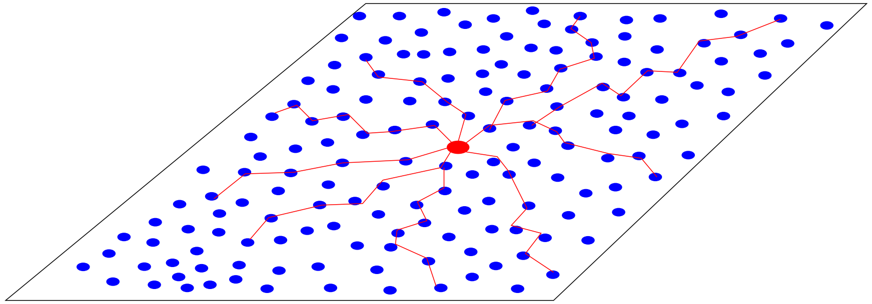
$$l_i = \log p_1(y_0, \cdots, y_i)$$

$$l_i = l_{i-1} + \log p_1(y_i | y_0, \cdots, y_{i-1})$$



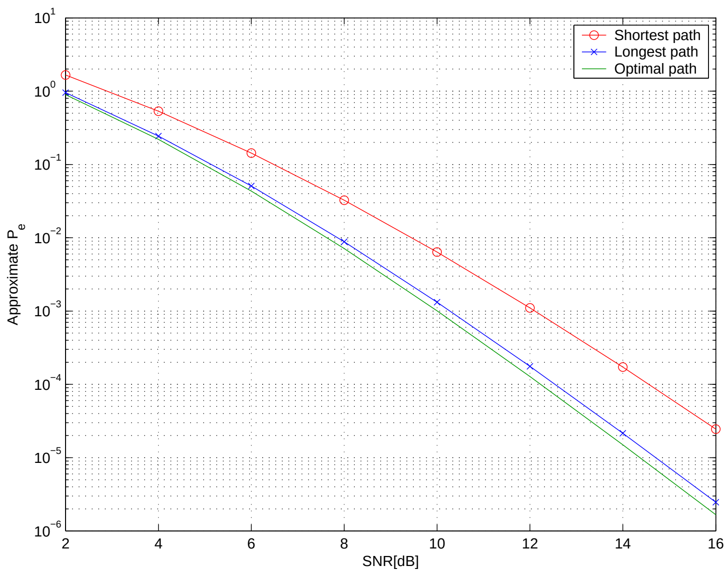
$$\begin{aligned}
e_i &= y_i - \hat{s}_{i|i-1}, \\
R_{e,i} &= P_{i|i-1} + \sigma^2, \\
l_i &= l_{i-1} - \frac{1}{2}\left(\log(2\pi R_{e,i}) + \frac{e_i^2}{R_{e,i}}\right), \\
K_{p,i} &= (a_i P_{i|i-1})/R_{e,i}, \\
\hat{s}_{i+1|i} &= a_i \hat{s}_{i|i-1} + K_{p,i} e_i, \\
P_{i+1|i} &= a_i^2 P_{i|i-1} + Q_i - K_{p,i}^2 R_{e,i}.
\end{aligned}$$

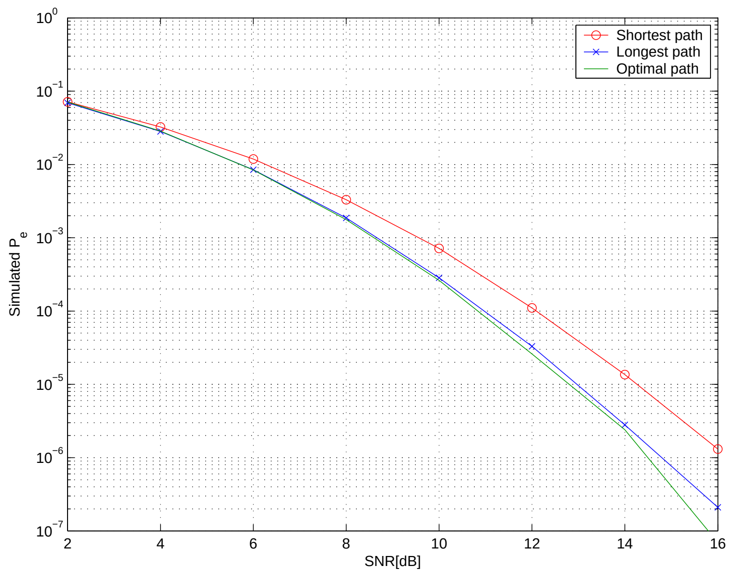
A Numerical Simulation



Sensor Field

- Random generation of 30 independent paths with 20 nodes and $\Delta_i \overset{i.i.d.}{\sim} \mathbb{E}(1)$
- Selection of shortest path, longest path, and optimal path in the proposed metric





Conclusion: Remarks

On Cross Layer Design

- Sensor networks are often application specific.
- The design application specific networks calls for cross layer strategies.
- The crucial step is to derive application specific metric.

Cautionary Remarks

- The analytical results depend on the specific signal model.
- Results should be viewed as insights.

References

[1] G. Mergen and L. Tong, "Type-based estimation over multiaccess channels," to appear in IEEE Trans. on Signal Processing, 2005.

[2] Y. Sung, L. Tong, and A. Ephremides, "Routing for detection of correlated random fields in large sensor networks," CISS'05, to be submitted to *IEEE Transactions on Information Theory*.

[3] Y. Sung, L. Tong, and H. V. Poor, "Neyman-Pearson detection of Gauss-Markov signals in noise: Closed-form error exponent and properties," submitted to *IEEE Transactions on Information Theory*, Nov., 2004