Distributed Decision in Sensor Networks

José M. F. Moura

Department of Electrical and Computer Engineering

Carnegie Mellon University

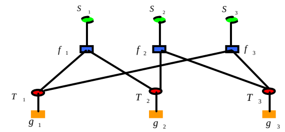
http://www.ece.cmu.edu/~moura

Work with: Saeed Aldosari, Elijah Liu

IBM Watson Research Center
Hawthorne NY
March 23/ 2004

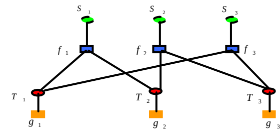
DARPA DSO ACMP Integrated and Sensing Program ARO grant
NSF Integrated Sensing Computation and Networked Systems for Decision and Action

Electrical & Computer ENGINEERING



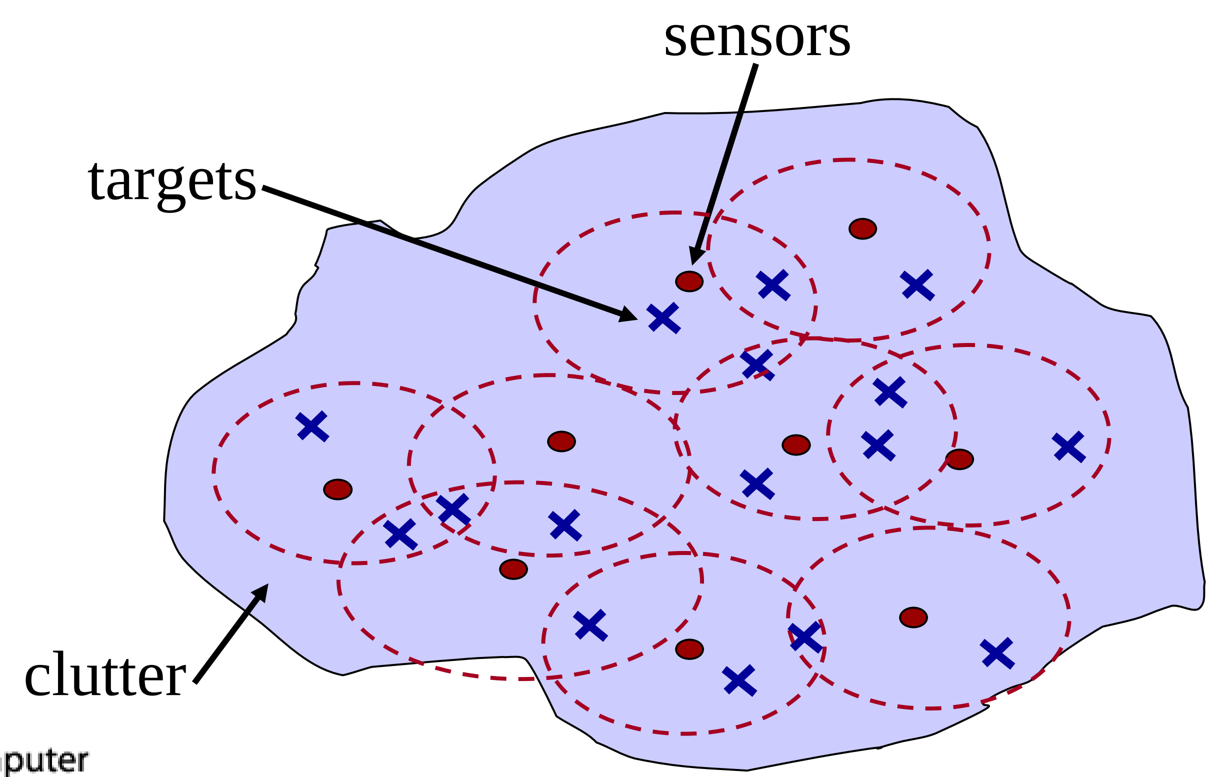
Sensor Networks

- **What about sensor networks:**
  - **Ad-hoc wireless networks? ⇐ constraints**
- **Design issues:**
  - **Inference: decentralized detection**
    - **Structure of detector**
    - **Performance of detector**
    - **Tradeoffs**
  - **Combine data from many sensors:**
    - **Fast fusion algorithms**

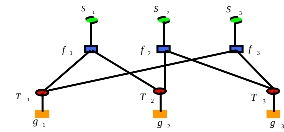


Sensor Networks

- Integrated technology: inexpensive sensors, deployable, multiple modalities (EM, acoustic, IR, magnetic, ...)
- Survey large areas: environmental, security, surveillance, …
- Many sensors/ many targets: global through local

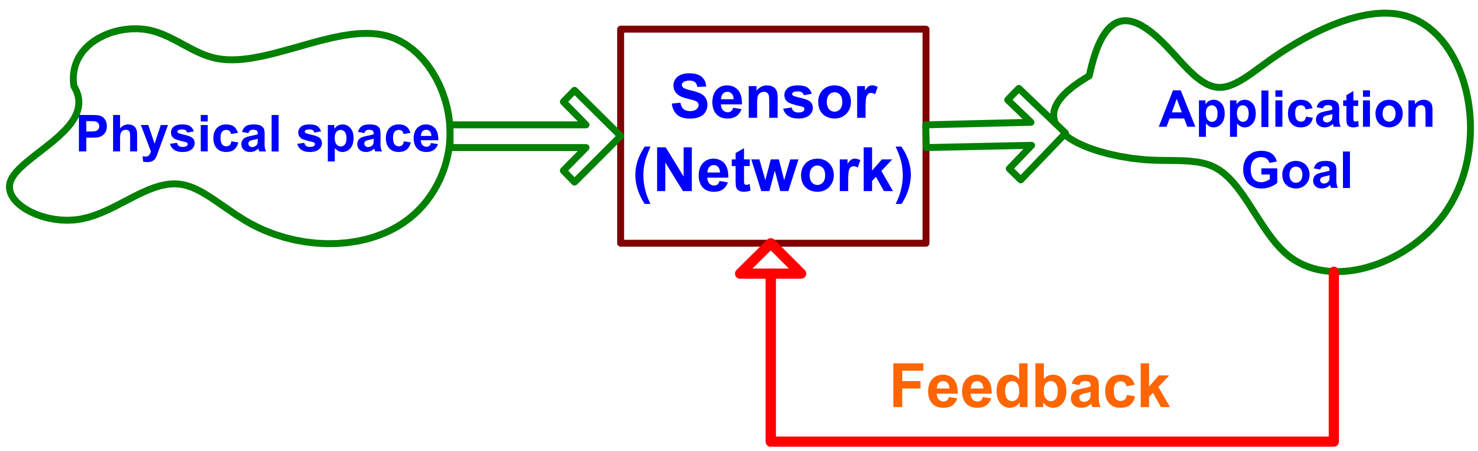


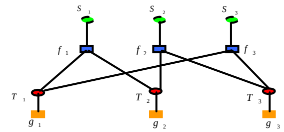
Electrical & Computer ENGINEERING



Sensor Networks

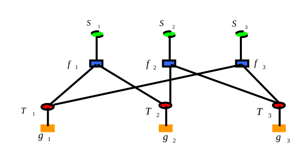
- Distributed, Heterogeneous, Autonomous
- Resource starved: rate constraint
- Network as sensor: design with tradeoffs



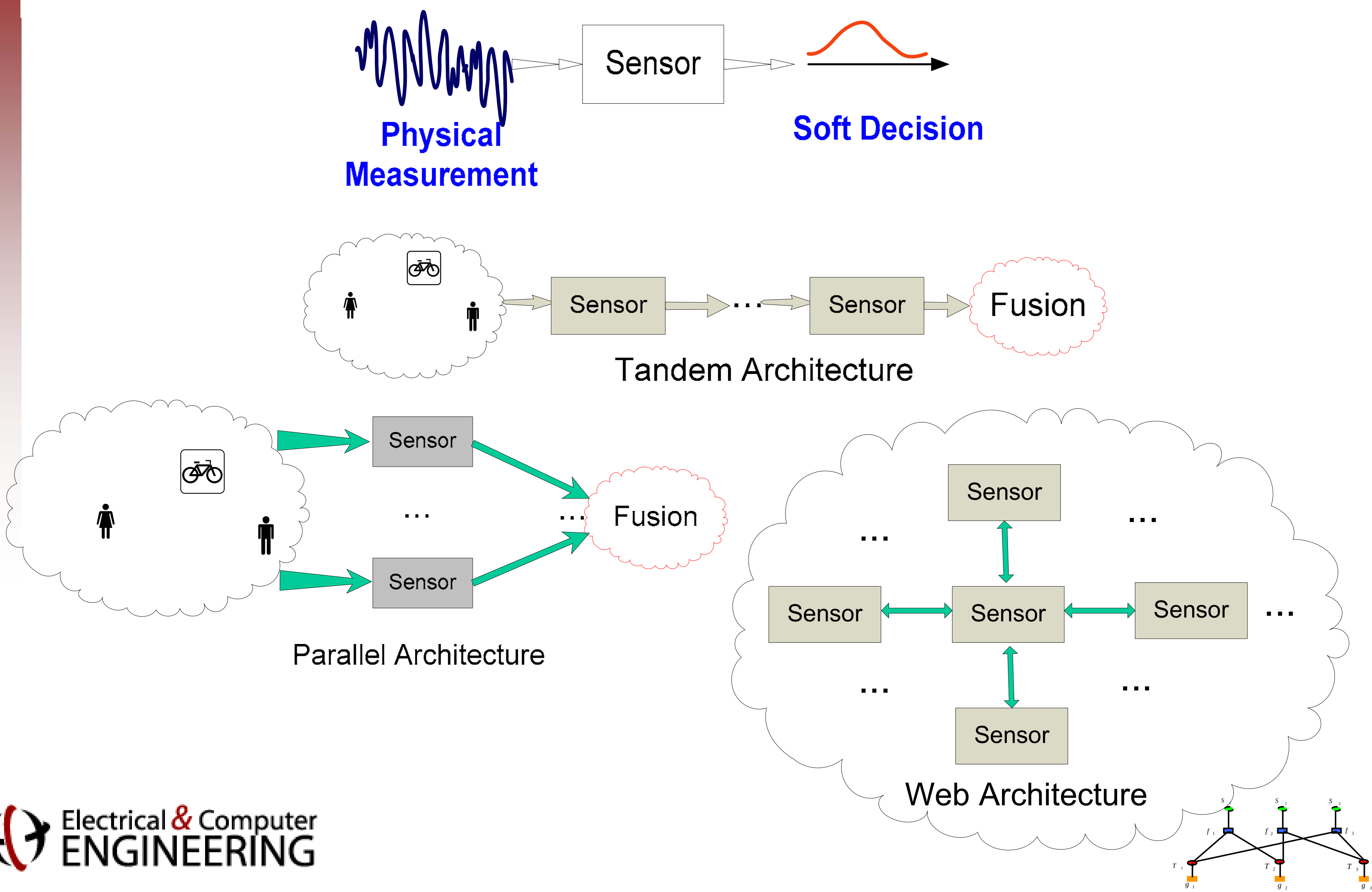


Sensor Networks

- **Main issues:**
  - **Detection: optimal algorithms and performance**
  - **Fusion: fast inference algorithms**
- **Constraints:**
  - **Common access channel rate constraint**
- **Tradeoffs:**
  - **Number of sensors**
  - **Number of bits per sensor**
  - **SNR**



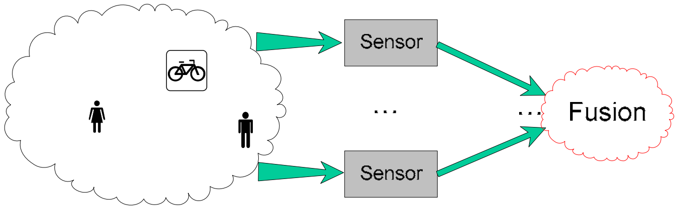
Sensor Network: Architectures



Electrical & Computer ENGINEERING

Sensor Networks: Optimal Detection

- **Decentralized detection:**
  - **Performance: min P_e under rate constraint R**
  - **Decentralized detection:**
    - **Tradeoffs: $R = N \times b$: N, b, SNR**
    - **Optimal detector:**
      - **Global fusion rule**
      - **Local thresholds**
    - **Probability of error**



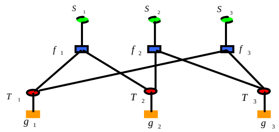
Parallel Architecture

- **Optimal detector: Hard combinatorial problem**

Asymptotic analysis: how good are the results for finite N; how large does N need to be (SNR dependence)

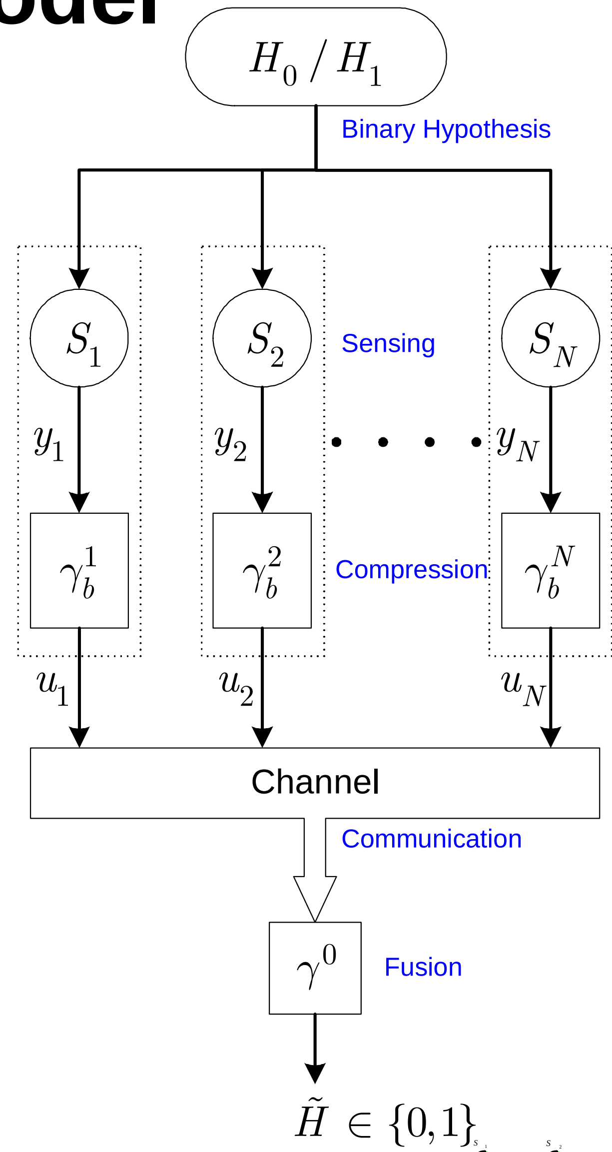
Numerical optimization: fusion rule intuitively pleasing generalization of majority rule for binary quantization

Electrical & Computer ENGINEERING

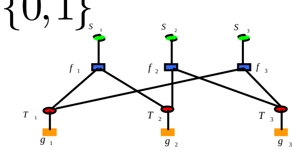


Decentralized Detection: Model

- Source H :
  - Binary Hypothesis H_0 vs. H_1
  - Prior probabilities: π_0 , π_1
- Observations $y_0, y_1, \ldots, y_N$:
  - Conditionally independent given H
  - Identically distributed: $f_i(y) = f(y \mid H_i)$
  - Monotone likelihood ratio $f_1(y) / f_0(y)$
- Compression:
  - b bits per measurement
  - Observation space: $y_n \in \mathbb{R}$
  - Classification space: $u_n \in \mathsf{U} = \{0, 1, \ldots, L-1\}$
  - $\gamma_b^n : \mathbb{R} \to \mathsf{U}$, $|\mathsf{U}| = L = 2^b$
- Communication: Error-free, Bandwidth=R bits/sec
- Fusion: $\gamma^0 : \mathsf{U}^N \to \{0, 1\}$



Standard model: Sandell & Tenney 81, Tsitsiklis 85

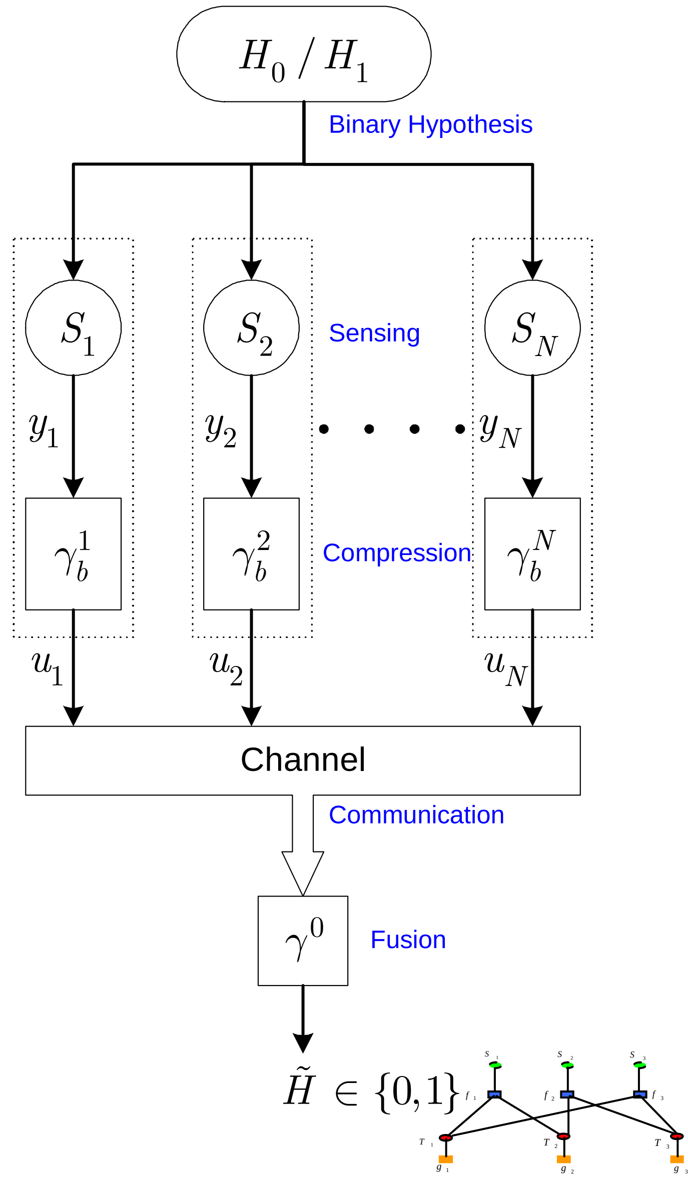


Optimized Decentralized Detection

- Local classifiers: $\gamma_b^1, \gamma_b^2, \ldots, \gamma_b^N$
- Conditional independ. + Monotone likelihood ratio
  - Gauss case, in general: N [L(L-1)/2] thresholds
  - Experiments for $b=2$: optimizing L(L-1)/2 thresholds/local sensor always converges to simpler one with (L-1)/sensor $\Rightarrow$ N(L-1) thresholds
  - $\Rightarrow$ Optimal classifiers scalar quantizers:

$$u = \begin{cases} 0 & -\infty < y < \lambda_1 \\ 1 & \lambda_1 < y < \lambda_2 \\ \vdots & \vdots \\ L-1 & \lambda_{L-1} < y < \infty \end{cases}$$

  - Each classifier characterized by threshold vector: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{L-1})$
    N local classifiers with possibly different $\boldsymbol{\lambda}$
  - Optimal thresholds $\boldsymbol{\lambda}$? Optimal fusion γ^0 ?



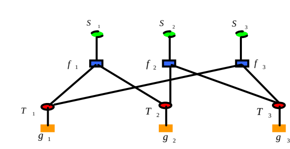
Unquantized vs Quantized Local Decision

- Assuming everything else is fixed, $b \uparrow \Rightarrow P_e(b) \downarrow$ and $P_e(b) \geq P_e(\infty)$
- Problem: $b = ?$ s.t. $P_e(b) \approx P_e(\infty)$
- Difficulty: $P_e(b)$ is hard to evaluate (dependent on fusion rule γ^0)
- Asymptotic analysis: Assume $N \rightarrow \infty$
  - $P_e(b, N) \rightarrow 0$ exponentially fast as $N \rightarrow \infty$
  - The rate of decay is given by:

$$C_b = -\lim_{N \to \infty} \frac{1}{N} \log P_e(b, N)$$

- Assuming everything else is fixed: $b \uparrow \Rightarrow C_b \uparrow$ and $C_b \leq C_\infty$
- Problem: $\boxed{b = ? \text{ s.t. } C_b \approx C_\infty}$

Design optimal detector and study *P_e (b)* as function of *b*



Optimization

- Chernoff Information:

$$C_\infty = -\min_{0 \le s \le 1} \log \int_{-\infty}^{\infty} [f_0(y)]^s [f_1(y)]^{1-s} dy$$

$$C_b(\boldsymbol{\lambda}, s) = -\log \sum_{u=0}^{L-1} [\Pr(u \mid H_0)]^s [\Pr(u \mid H_1)]^{1-s}$$

- Objective function:

$$V_b(\boldsymbol{\lambda}, s) = \log[1 - C_b(\boldsymbol{\lambda}, s) / C_\infty]$$

Optimization: Assume identical local detectors (negligible loss when $N \to \infty$

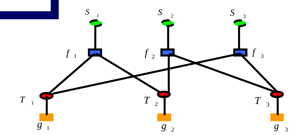
$$\min_{\boldsymbol{\lambda}, s} V_b(\boldsymbol{\lambda}, s) \text{ subject to: } \lambda_1 < \lambda_2 < \cdots < \lambda_{L-1} \text{ and } 0 \le s \le 1$$

- Alternatively: let $\delta = (\delta_1, \delta_2, \ldots, \delta_{L-2})$, where $\delta_1 = \lambda_1$, and $\delta_k = \lambda_k - \lambda_{k-1}, k = 2, 3, \ldots L-2$

$$\min_{\boldsymbol{\delta}, s} V_b(\boldsymbol{\lambda}, s) \text{ subject to: } \delta_k > 0, k = 2, 3, \ldots, L-2 \text{ and } 0 \le s \le 1$$

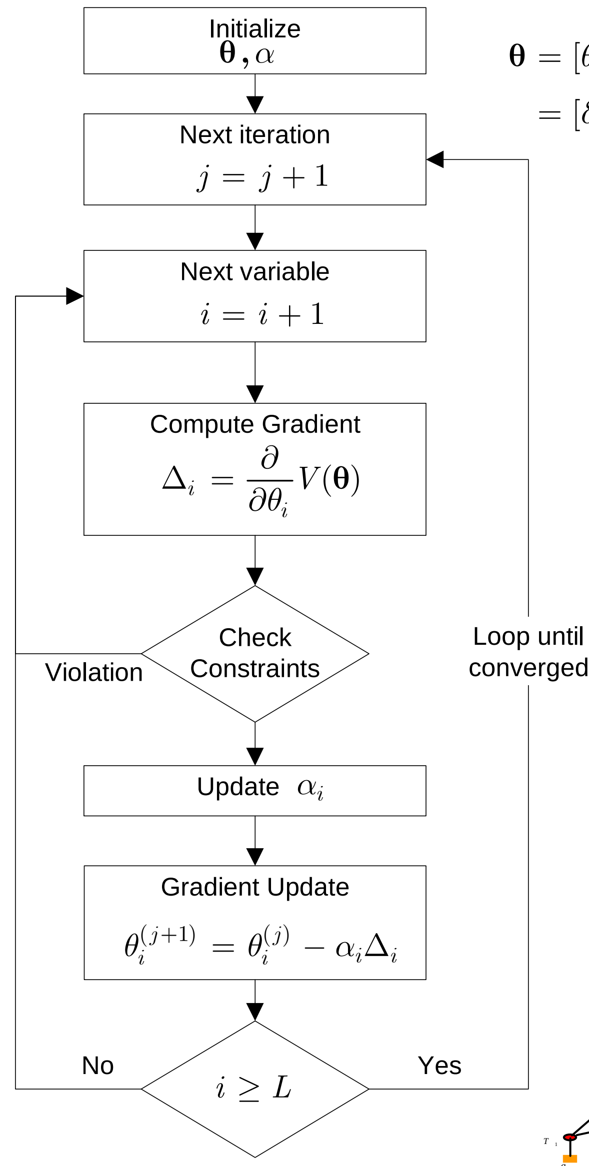
•Asymptotic analysis abstracts out fusion rule

Electrical & Computer ENGINEERING

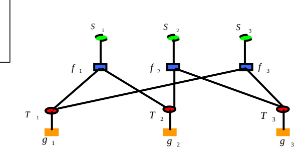


Local Thresholds Algorithm

- Optimization :
  L-Dimensional & nonlinear
  $L = 2^b$
- Algorithm:
  Gradient-descent
- Initialization:
  α_i : convergence vs. speed
- Stopping:
  $\Delta_i < \varepsilon, \;\; i = 1, 2, \ldots, L$



$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_L]$$
$$= [\delta_1, \delta_2, \ldots, \delta_{L-1}, s]$$



Electrical & Computer ENGINEERING

Results : Unquantized vs Quantized

- Observation model:

  under $H_i: \; y = m_i + n$

  $E[n] = 0, \; \mathrm{Var}[n] = \sigma^2$

  m_i are constants representing the signal mean

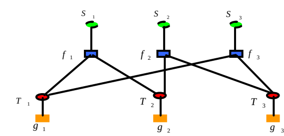
- Case Studies:

  - Noise distributions:Gauss, Laplace, Logistic
  - No. of bits: 1 to 8 bits/sample
  - SNR: 0 to 20dB

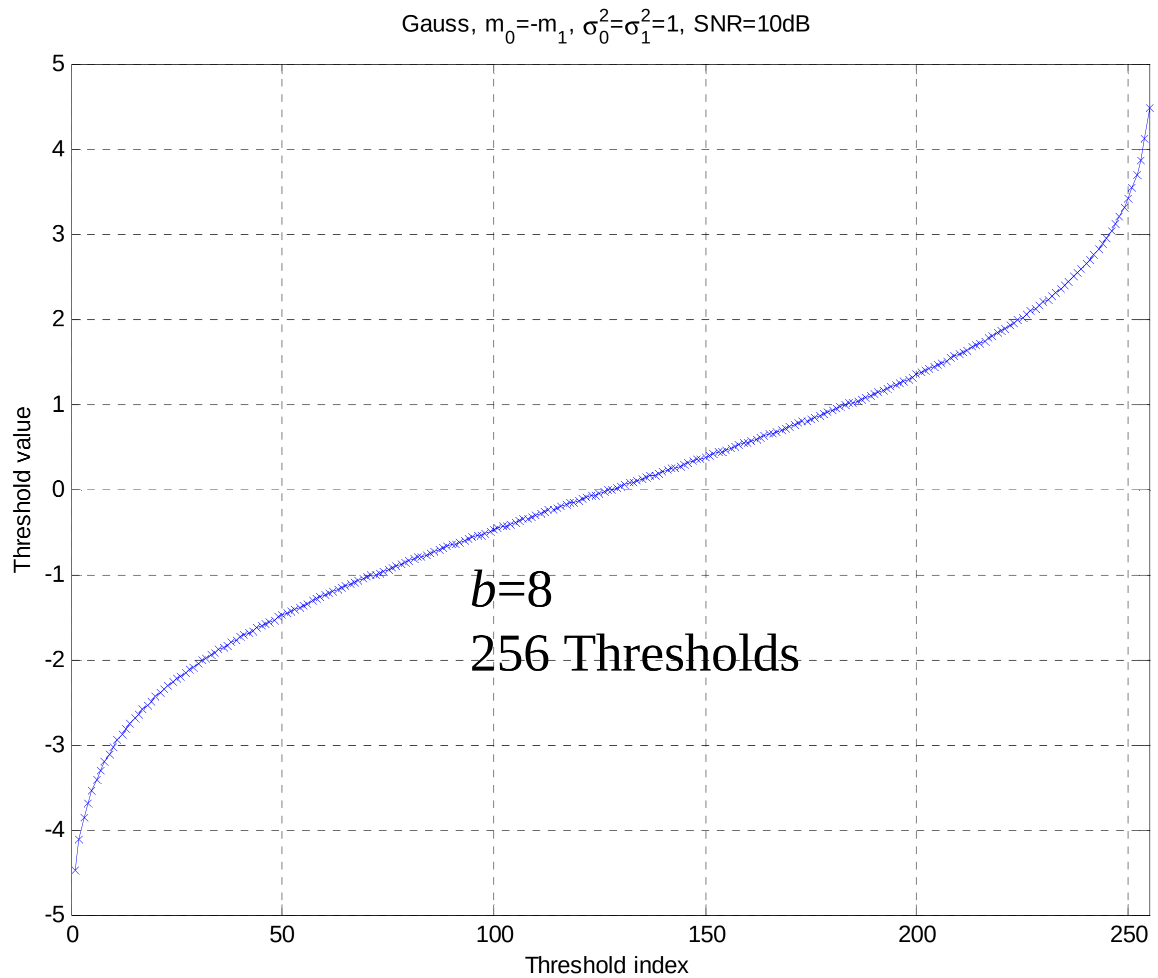
$$f_{\text{logistic}}(y) = \frac{e^{-(y-m)/\rho}}{\rho\left[1 + e^{-(y-m)/\rho}\right]^2}, \;\; \rho = \frac{\sqrt{3}}{\pi}\sigma$$

$$f_{\text{Laplace}}(y) = \frac{1}{2\vartheta} e^{-|y-m|/\vartheta}, \;\; \vartheta = \frac{1}{\sqrt{2}}\sigma$$

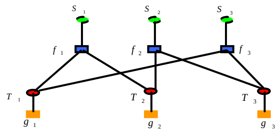
Electrical & Computer ENGINEERING



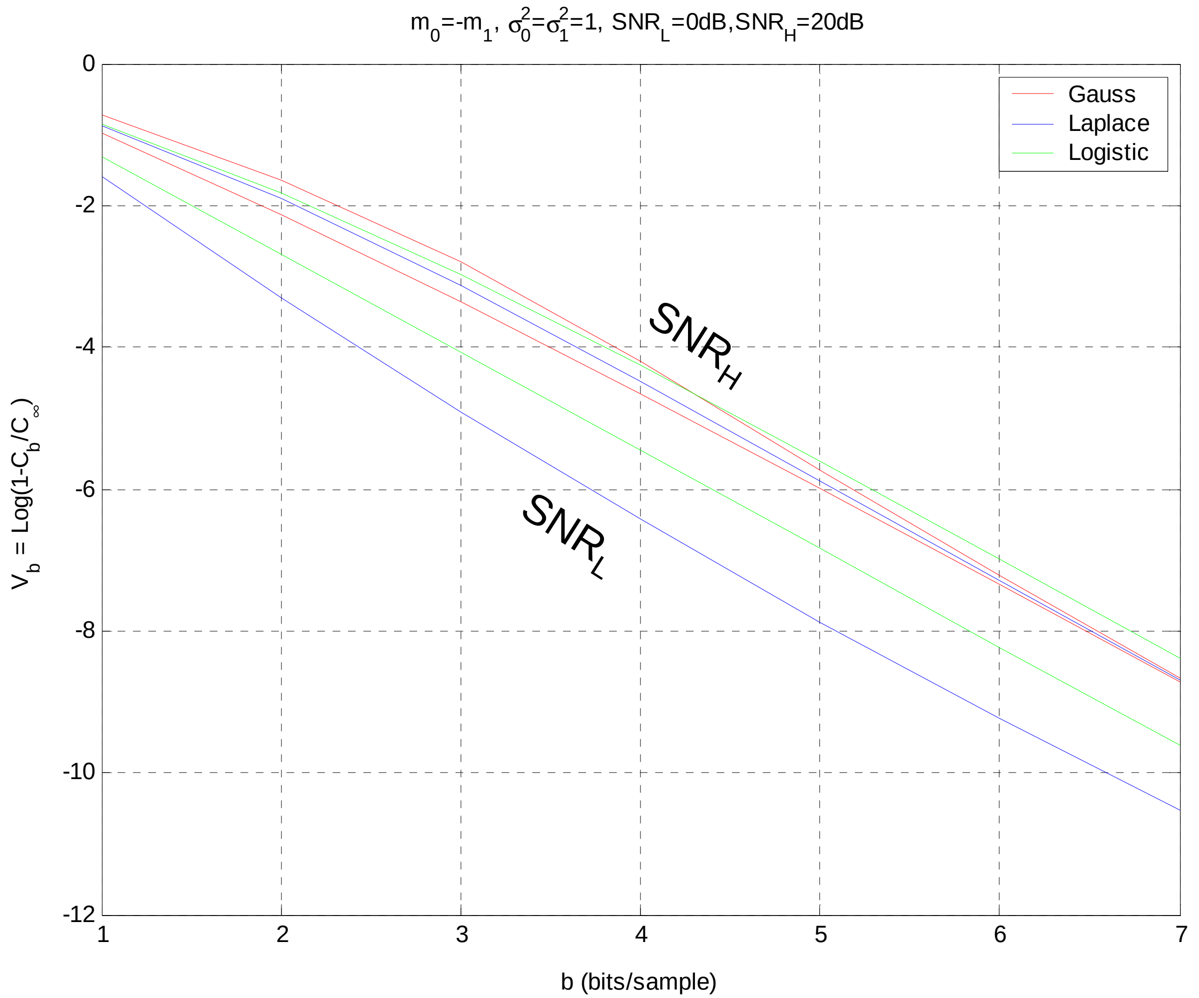
Results : Unquantized vs Quantized

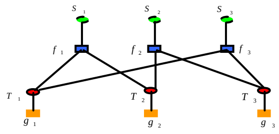


Electrical & Computer ENGINEERING



Results : Unquantized vs Quantized





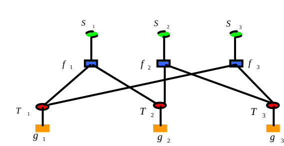
Conclusions: Unquantized vs Quantized

- $C_b \to C_\infty$ exponentially fast as $b \uparrow$
  $\Rightarrow$ Little gain if we go to higher number of bits b.

- Threshold distribution:
  is such that threshold points are concentrated around the boundary between $f_0(y)$ and $f_1(y)$ (area where it is most hard to discriminate between H_0 and H_1).

- In all cases studied, s converges to 0.5

- For high b, the ratio C_b / C_∞ is less sensitive to SNR when the noise Gaussian.



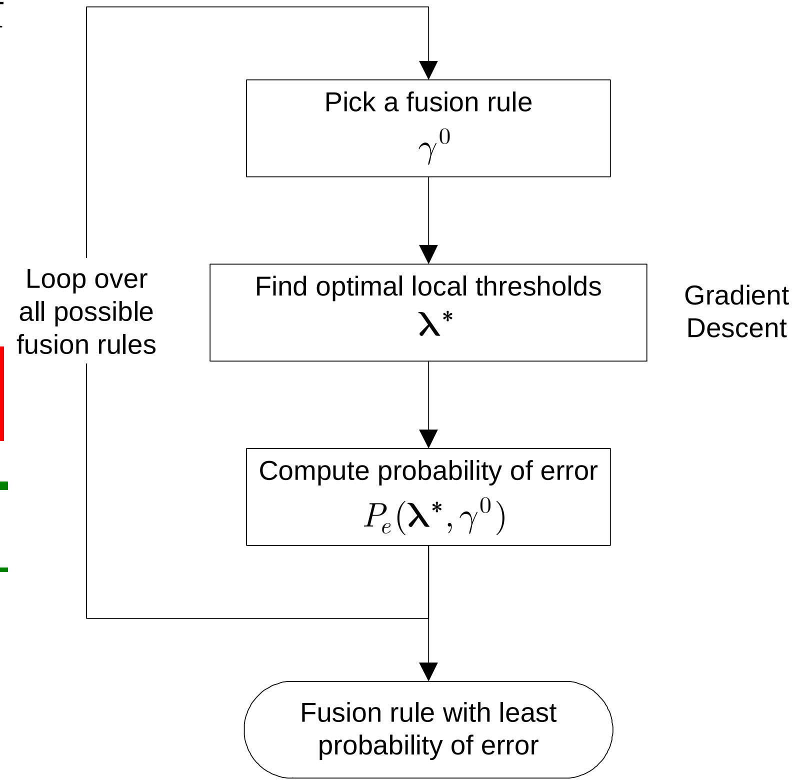
Fusion Rule

- Asymptotic studies abstracts the role of the fusion rule ⇒ consider finite number of sensors N
- In general, for finite N, optimal local classifiers $\gamma_b^1, \gamma_b^1, \ldots \gamma_b^N$ might not be identical ⇒ $(L-1)N$ thresholds
- No. of possible fusion rules: $2^{L^N} = 2^{2^{Nb}}$

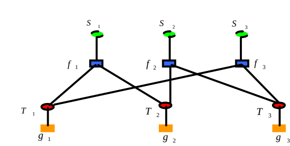
u_1	u_2	$\cdots$	u_N	$\tilde{H}$
0	0	$\cdots$	0	h_1
0	0	$\cdots$	1	h_2
$\vdots$	$\vdots$	$\ddots$	$\vdots$	
$L-1$	$L-1$	$\cdots$	$L-1$	h_{L^N}

Note: $h_k \in \{0,1\}$

Exhaustive Search

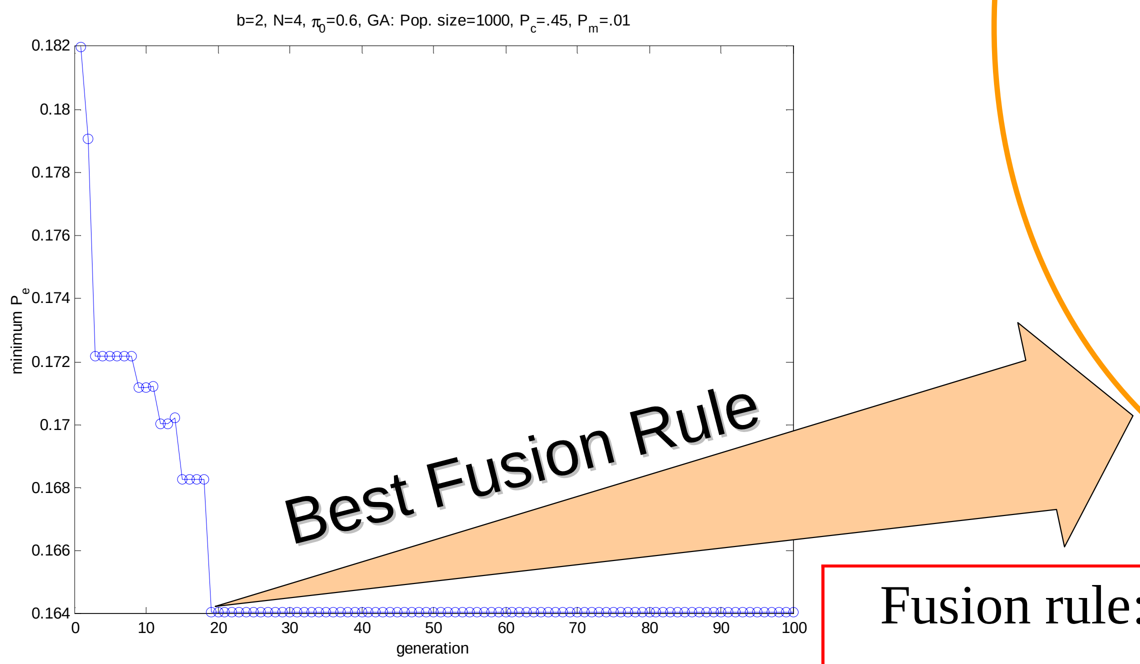


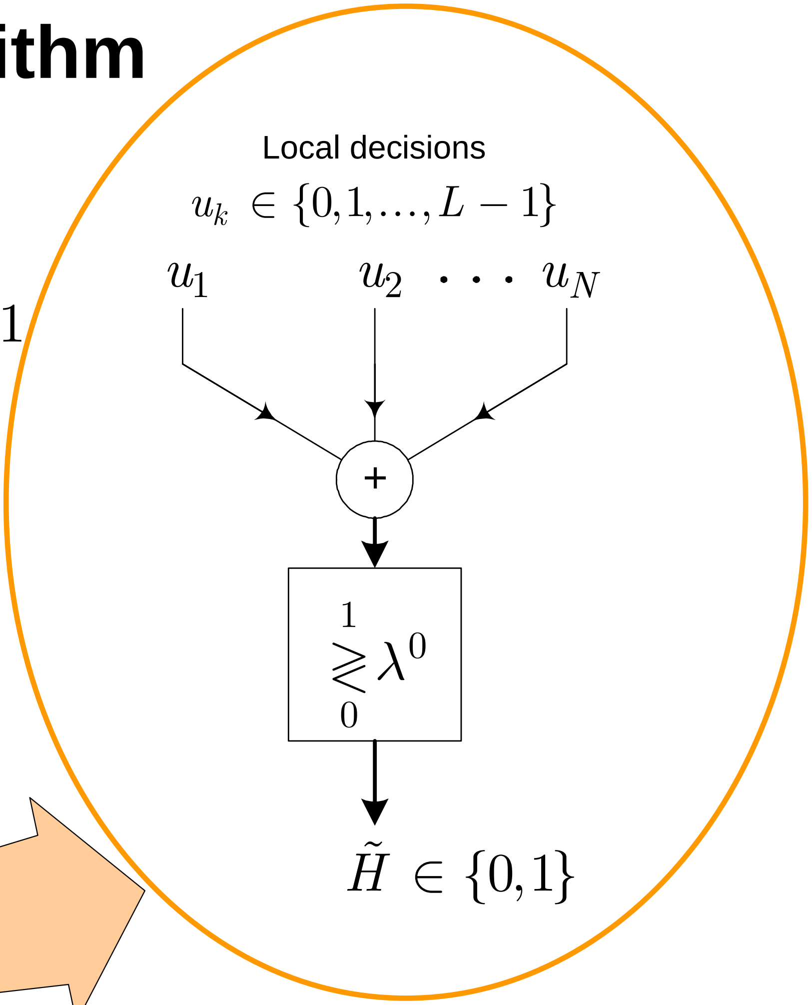
Electrical & Computer ENGINEERING



Fusion Rule: Genetic Algorithm

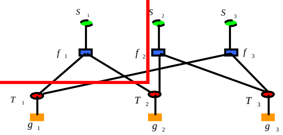
- Setup:
  - Priors: $\pi_0, \pi_1 = 1 - \pi_0$
  - Noise: Gauss: $m_0 = 0, m_1 = 1, \sigma^2 = 1$
  - (b, N): (1,N), (2,2), (2,3), (2,4), (3,2)





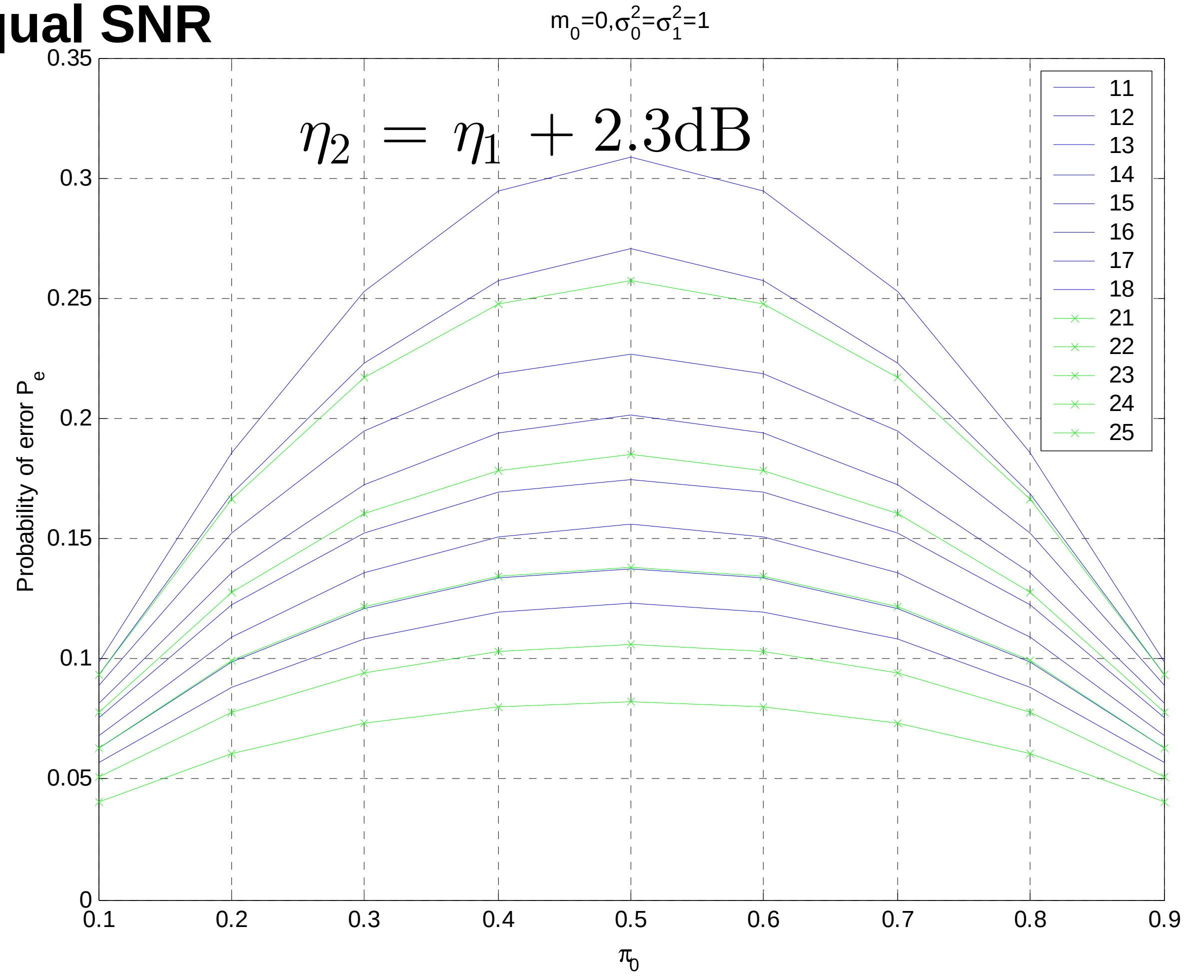
Fusion rule: regardless of priors

$$\lambda^0 \simeq \frac{1}{2} N(L-1)$$



2N Binary vs *N* Quaternary Sensors

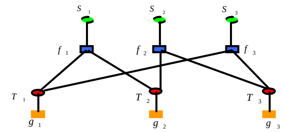
- **Rate Constraint: *R= N x b***
- **Unequal SNR**



P_e Ranking

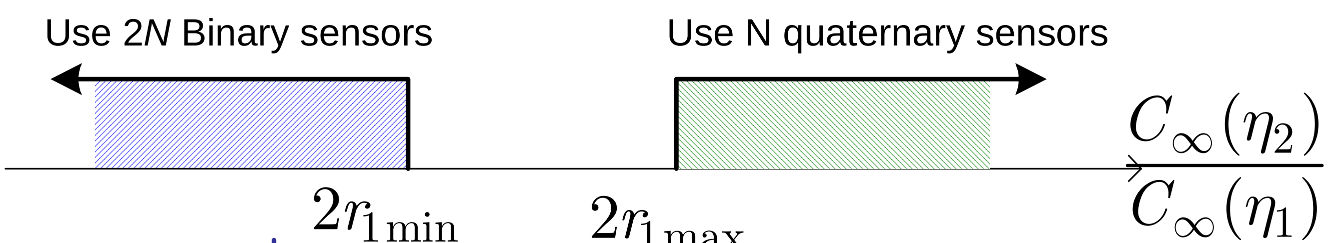
b	*N*	Total Bits
1	1	1
1	2	2
2	1	2
1	3	3
1	4	4
2	2	4
1	5	5
1	6	6
2	3	6
1	7	7
1	8	8
2	4	8
2	5	10

Electrical & Computer ENGINEERING



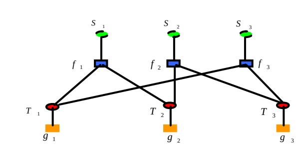
2N Binary vs *N* Quaternary Sensors

$$r_b = \frac{C_b(\eta_b)}{C_\infty(\eta_b)}, \quad r_{b\,\min} < r_b < r_{b\,\max},$$

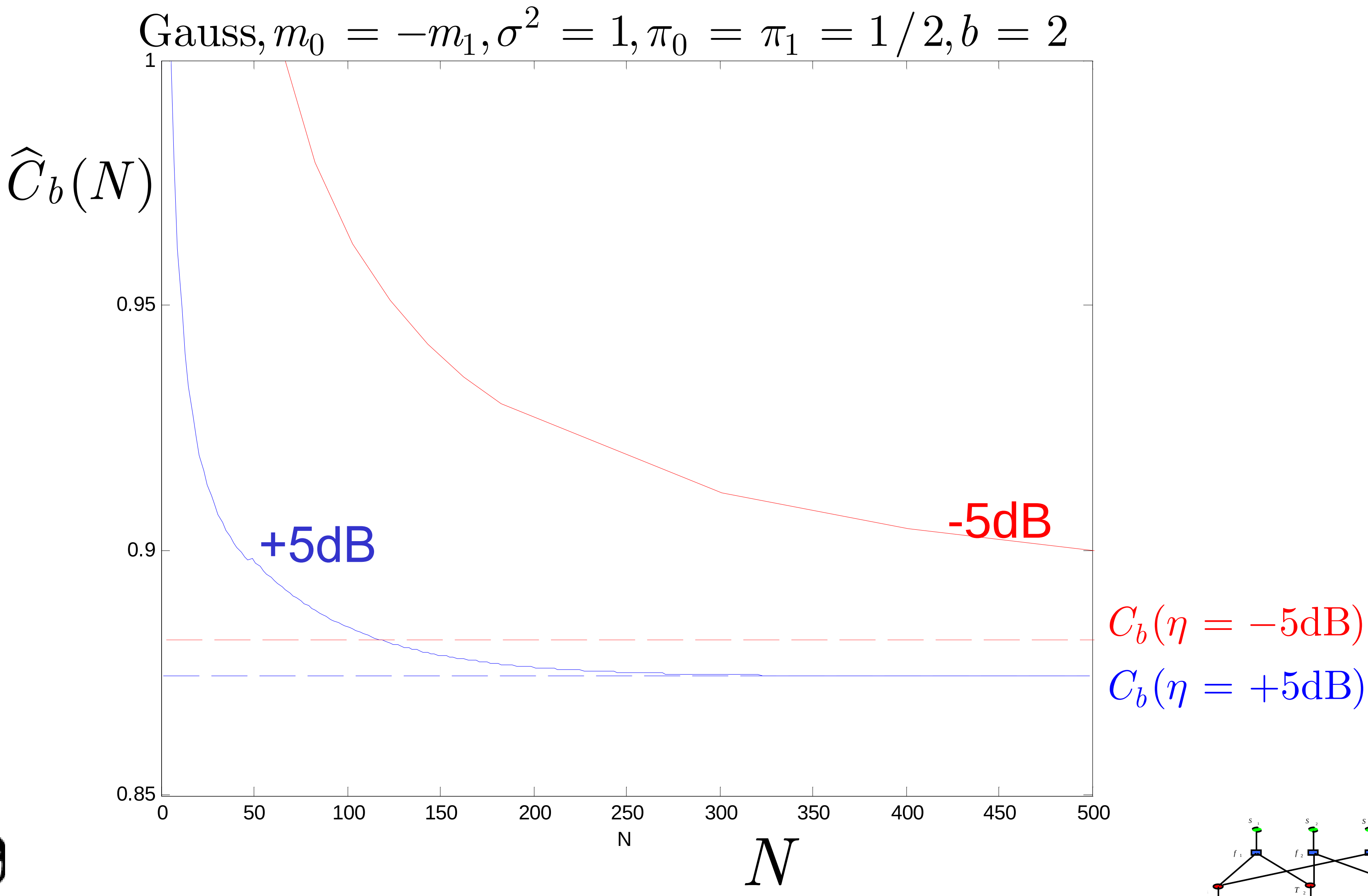


	$\frac{2r_{1\,\min}}{r_{2\,\max}}$	$\frac{2r_{1\,\max}}{r_{2\,\min}}$
Gauss	1.14	1.62
Logistic	1.06	2.0
Laplace	1	1.5

$$m_0 = -m_1, \sigma^2 = 1$$

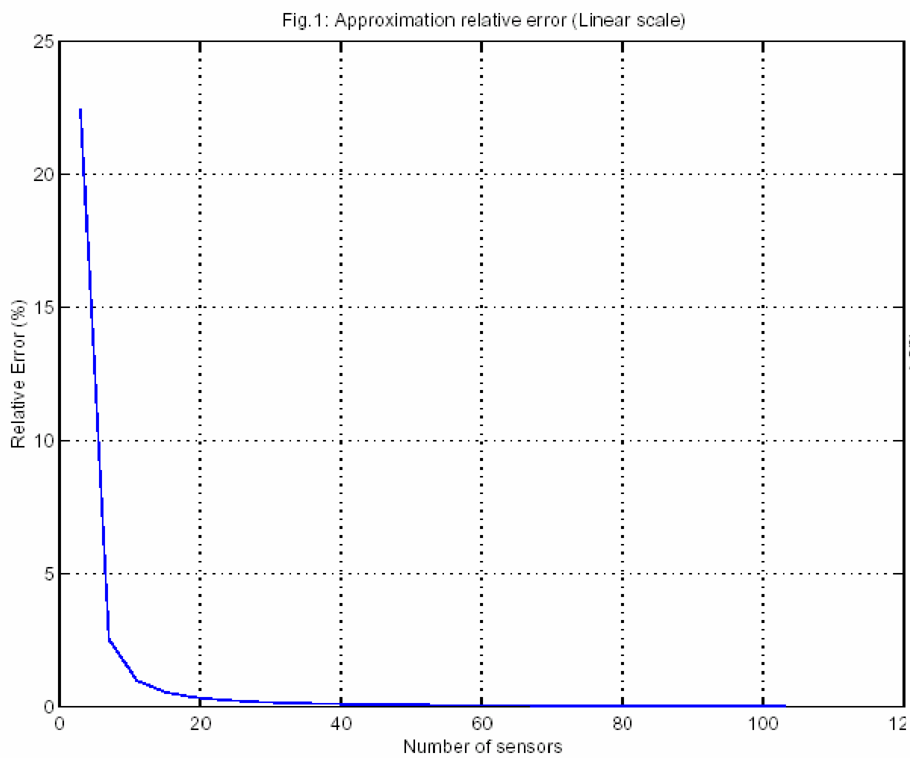


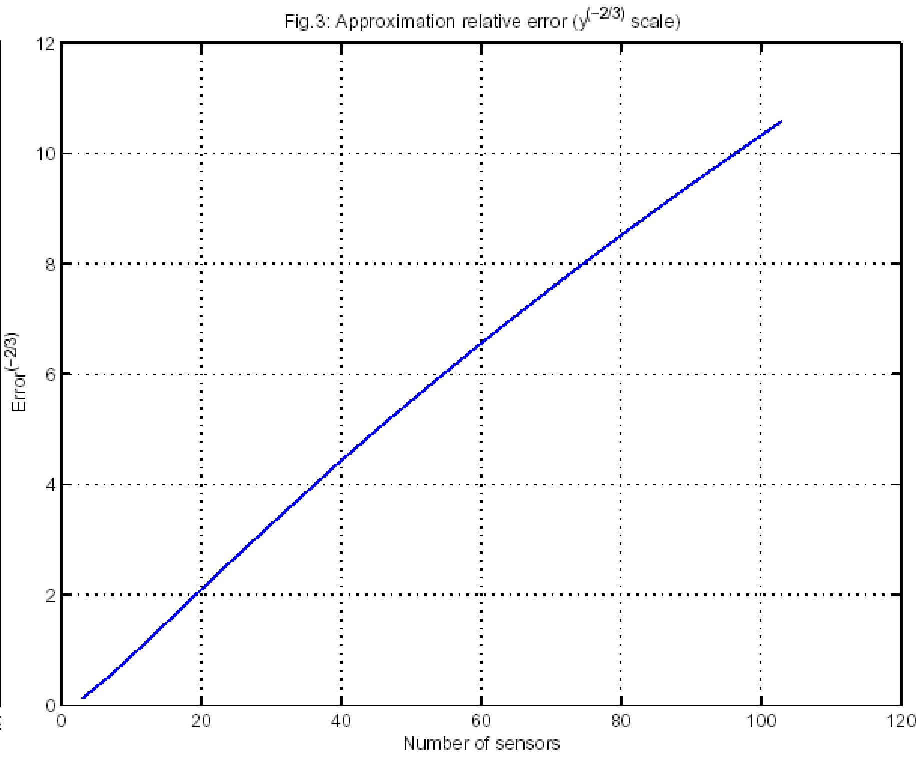
Finite *N* Sensors vs Asymptotic Analysis



Performance Analysis: Finite large *N*

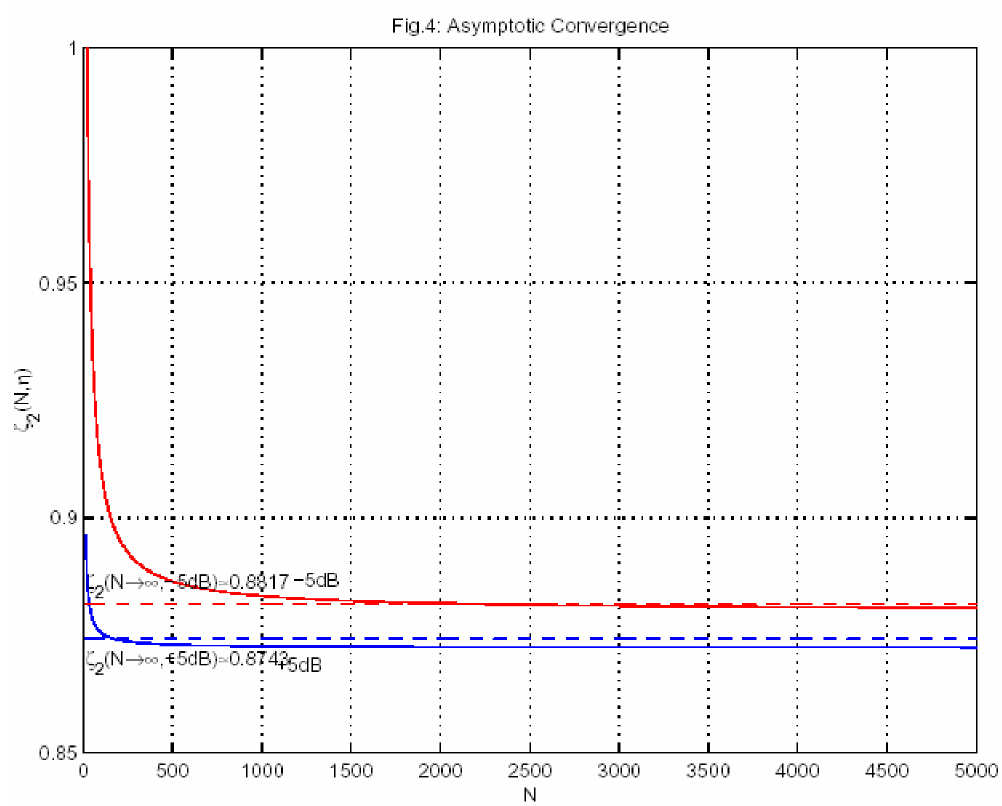
Lugannani-Rice approximation: $(P_e^{LR} - P_e^{exact})/P_e^{exact} \times 100$





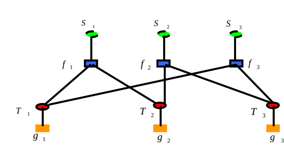
Performance Analysis

Lugannani-Rice approximation:



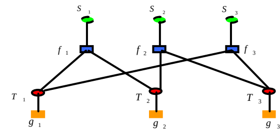
$$\xi_2\left(N,\eta\right)=\frac{\log\dfrac{P_e\left(N\right)}{P_e\left(N+2\right)}}{2C_\infty}=\log\frac{C_b\left(N,\eta\right)}{C_\infty}$$

Electrical & Computer ENGINEERING



Network as Sensor: Detection

- Fusion rule: **majority**-rule
- Under MAC constraint: ***R*= *N x b***
  - Prefer quaternary over binary sensors if SNR 1.5 to 2 dB larger for quaternary sensors
- $P_e \to 0$ exponentially fast as $N \to \infty$
- The rate of decay $\widehat{C}_b(N)$ approaches that predicted by asymptotic studies C_b as $N \to \infty$
- Convergence rate of $\widehat{C}_b(N) \to C_b$ depends on system parameters, especially SNR: slower at lower SNR
- Slow convergence $\Rightarrow$ very large number of sensors N is required to approach asymptotes (over thousands)
- Local detectors tradeoffs: How many levels of quantization of local decisions
  - Few bits/ decision at high SNR
  - More bits/ decision at low SNR – usually 1 to 2 bits more than at low SNR
  - Asymptotic analysis may lead to wrong decisions at low SNR

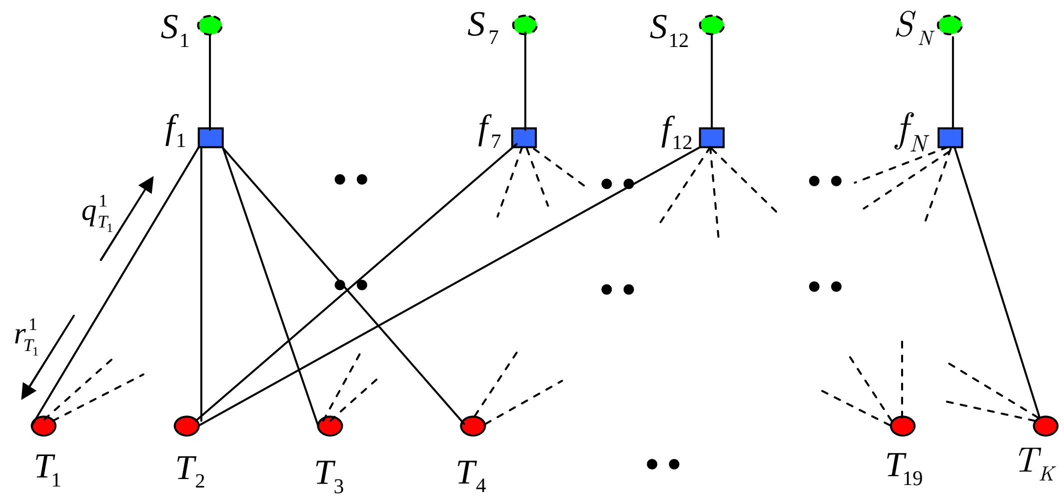


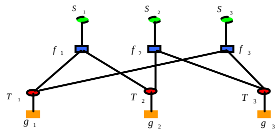
Fast Fusion: Fast Sum-Product

$$p(T_1,T_2,T_3|S_1,S_2,S_3) = \underbrace{p(T_1,T_2|S_1)}_{f_1}\underbrace{p(T_2,T_3|S_2)}_{f_2}\underbrace{p(T_1,T_3|S_3)}_{f_3}\underbrace{\frac{1}{p(T_1)}}_{g_1}\underbrace{\frac{1}{p(T_2)}}_{g_2}\underbrace{\frac{1}{p(T_3)}}_{g_3}$$

$$p(T_1 \mid S_1,S_2,S_3) = \sum_{T_2,T_3} p(T_1,T_2,T_3 \mid S_1,S_2,S_3)$$

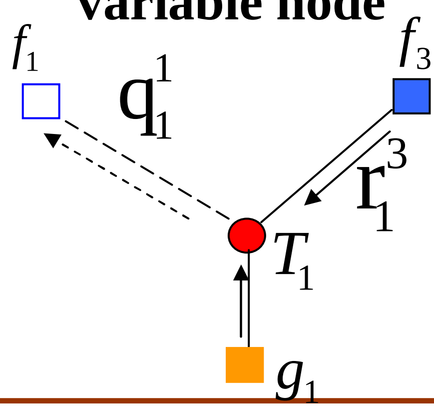
- **Map fusion of N sensors detecting K targets on bipartite graph**
- **Fuse sensor soft info. with sum-product algorithm**
- **Message flow alg.**
- **Compute marginals from 1D marginals**





Sum-product algorithm (Compute function marginals)

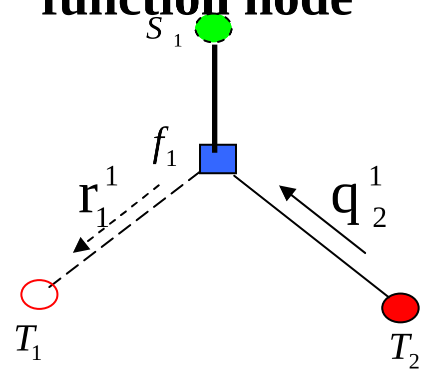
Message updating around variable node



$$q_1^1 = r_1^3 \cdot g_1(T_1)$$

Send out only extrinsic info.

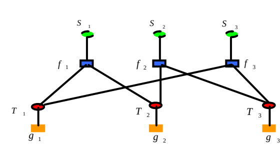
Message updating around function node



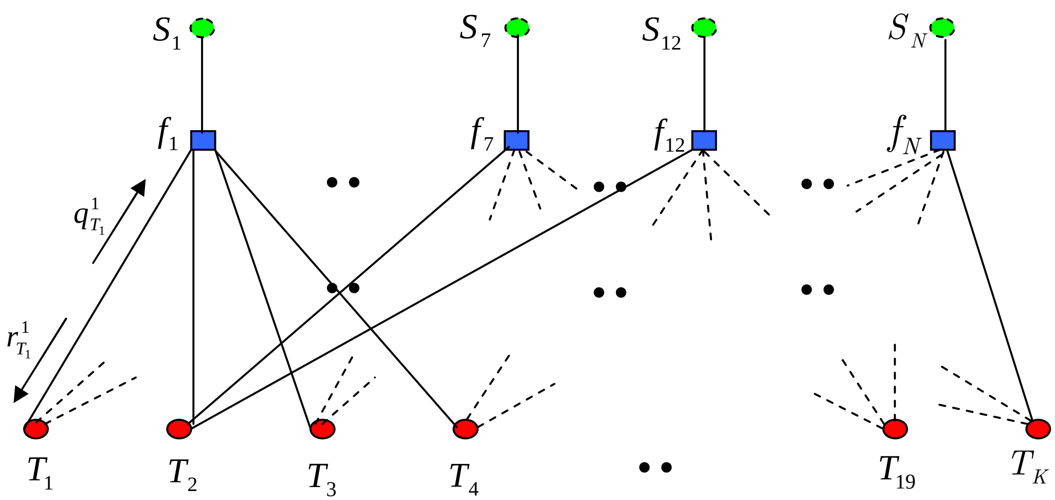
$$r_1^1 = \sum_{T_2} \left\{ q_2^1 \cdot f_1(T_1, T_2, S_1) \right\}$$

Message updating rules of sum-product algorithm

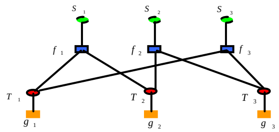
Electrical & Computer ENGINEERING



Fast Fusion: Fast Sum-Product



- Fast implementation of sum-product algorithm:
- Divide and conquer, divide and multiply (reduce complexity by K/2)
- Approximate sensors soft info. by sums of Gauss
- Propagate means and covariances
- Convergence: if covariances converge, means converge to their correct values
- Fast algorithm: scenarios with hundreds of sensors and hundreds of targets



Fast Fusion: Convergence of Variances

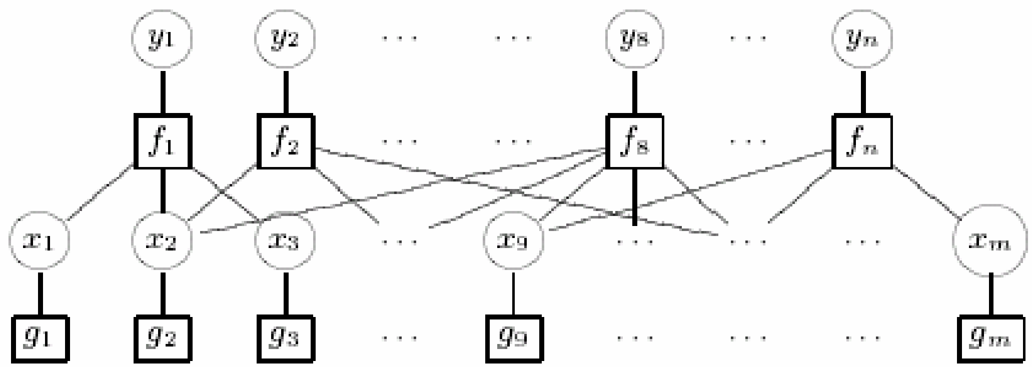


Fig. 3. Sensor Network of General Topology

$$\left(C_1^{(k+1)},\ldots,C_n^{(k+1)}\right)=\left(\mathcal{F}_1\left(\left\{C_j^{(k)}\right\}_{j\neq 1}\right),\ldots,\mathcal{F}_n\left(\left\{C_j^{(k)}\right\}_{j\neq n}\right)\right)$$
$$=\mathcal{F}\left(C_1^{(k)},\ldots,C_n^{(k)}\right)\ . \qquad (14)$$

Theorem 1. *The operator $\mathcal{F}$ possesses a fixed point in $\mathcal{D}^n$. Furthermore, denoting this fixed point by $(C_1^*,\ldots,C_n^*)$*

$$\lim_{k\to\infty}\mathcal{F}^k\left(C_1^{(0)},\ldots,C_n^{(0)}\right)=(C_1^*,\ldots,C_n^*) \qquad (15)$$

for all positive-definite diagonal matrices $C_1^{(0)},\ldots,C_n^{(0)}$.

Fast Fusion: Convergence of Means

$$M_i^{(k+1)} = \mathcal{H}_i\left(\left\{M_j^{(k)}, C_j^{(k)}\right\}_{j \neq i}\right)$$

$$A_{\Sigma_i,\{D_j\}_{j\neq i}} = \left(\Sigma_i^{-1} + \sum_{j\neq i} \xi_i\left(\lambda_{ij}\left(D_j^{-1}\right)\right) - \sum_{j\neq i} \xi_i\left(\lambda_{ij}\left(I_0\right)\right)\right)^{-1} . \qquad (11)$$

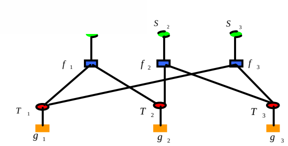
$$T_{\Sigma_1,\ldots,\Sigma_n} = \begin{pmatrix} 0 & \Theta_{1*} & \cdots & \cdots & \Theta_{1*} \\ \Theta_{2*} & 0 & \Theta_{2*} & \cdots & \Theta_{2*} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Theta_{(n-1)*} & \cdots & \Theta_{(n-1)*} & 0 & \Theta_{(n-1)*} \\ \Theta_{n*} & \cdots & \Theta_{n*} & \Theta_{n*} & 0 \end{pmatrix} . \qquad (18)$$

Each block matrix Θ_{i*} in $T_{\Sigma_1,\ldots,\Sigma_n}$ takes the form of

$$\Theta_{i*} = \Omega\left(A^{-1}_{C_i^*,\{C_j^*)\}_{j\neq i}} \tau_i(A_{\Sigma_i,\{C_j^*\}_{j\neq i}}) - I_i\right) , \qquad (19)$$

where the matrix $A_{\Sigma_i,\{C_j^*\}_{j\neq i}}$ is defined as before in (11) and the matrix $A^{-1}_{C_i^*,\{C_j^*)\}_{j\neq i}}$ is defined as

$$\sum_{j=1}^{n} \tau_i\left((C_j^*)^{-1}\right) - \sum_{j=1, j\neq i}^{n} \tau_i\left(\lambda_{ij}(I_0)\right) .$$



Fast Fusion: Convergence of Means

Theorem 2. *If* $\rho(T_{\Sigma_1,\ldots,\Sigma_n}) < 1$, *then* $\exists$ *vectors* $M_1^*,\ldots,M_n^*$ *such that, for any* $M_1^{(0)},\ldots,M_n^{(0)}$ *and any* $C_1^{(0)},\ldots,C_n^{(0)} \in \mathcal{D}$,

i. the sequence $(M_1^{(k)},\ldots,M_n^{(k)})$ *converges to* $(M_1^*,\ldots,M_n^*)$.

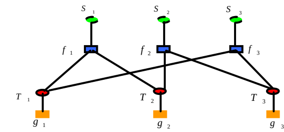
ii. the estimated means for the marginal densities obtained from the message statistics are the true marginal means.

Lemma 9 *If* $\max_{i\in\{1,\ldots,n\}} \rho(\Theta_{i*}) < \frac{1}{n-1}$, *then* $\rho(T_{\Sigma_1,\ldots,\Sigma_n}) < 1$.

$$\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}}_{n} - \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}}_{n}.$$

$$\begin{aligned} \rho(T_{\Sigma_1,\ldots,\Sigma_n}) &= \rho\left((\Theta_{1*}\oplus\ldots\oplus\Theta_{n*})\cdot\left(\widetilde{I}\otimes I_\Omega\right)\right) \\ &\leq \rho(\Theta_{1*}\oplus\ldots\oplus\Theta_{n*})\cdot\rho\left(\widetilde{I}\otimes I_\Omega\right) \\ &= \max_{i\in\{1,\ldots,n\}} \rho(\Theta_{i*})\cdot\rho\left(\widetilde{I}\right)\rho(I_\Omega) \\ &= \max_{i\in\{1,\ldots,n\}} \rho(\Theta_{i*})\cdot(n-1). \end{aligned}$$

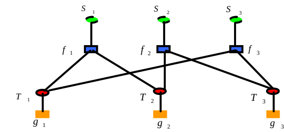
Electrical & Computer ENGINEERING

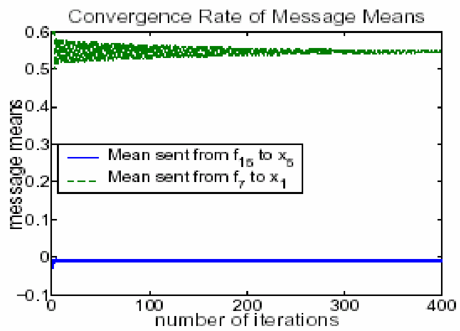


Fast Fusion: Convergence of Means

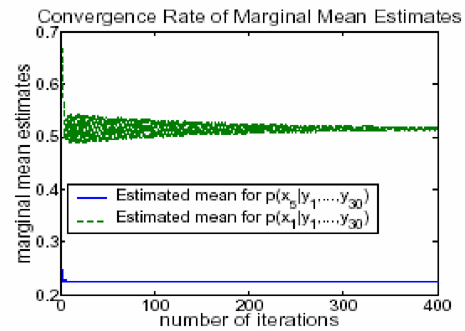
Lemma 11 *If* $\Sigma_1 = \ldots = \Sigma_n = \Sigma$, *then the sequence of covariance matrics converge to a unique fixed point* $(C, \ldots, C)$ *in* $\mathcal{D}^n$ *and* $C \leq \delta(\Sigma)$. *Moreover,* $\rho(T_{\Sigma_1, \ldots, \Sigma_n}) < 1$.

Proposition 2 *For any symmetric matrix* $\Sigma > 0$ *such that* $\Sigma^{-1} - (n-1)I_0 > 0$, *if* $\Sigma_i = \left(\Sigma^{-1} + \gamma_i\, I_0\right)^{-1}$ *for* $i = 1, \ldots, n$ *then* $(\Sigma_1, \ldots, \Sigma_n) \in \mathcal{C}$ *for all* $\sum_{i=1}^{n} \gamma_i = 0$ *and* $-\frac{1-(n-1)\lambda_{max}(\Sigma)}{\lambda_{max}(\Sigma)} < \gamma_1, \ldots, \gamma_n < \frac{1-(n-1)\lambda_{max}(\Sigma)}{\lambda_{max}(\Sigma)}$.



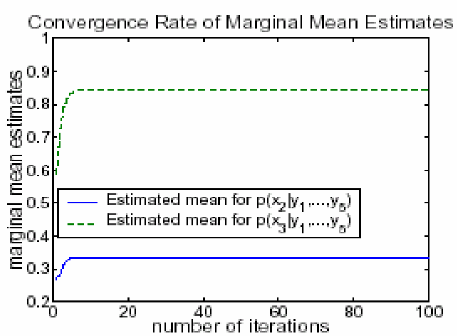


(a) Message Means Converge

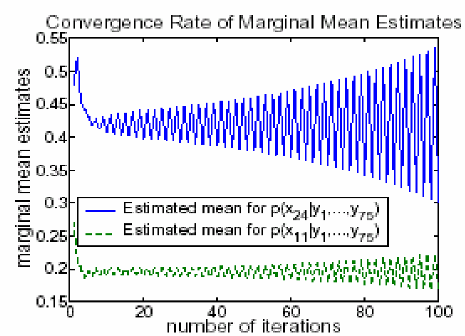


(b) Marginal Mean Estimates Converge

Figure 6: Regular Sensor Network of 30 Sensors and 30 Targets

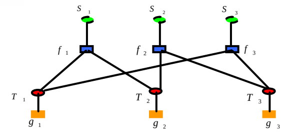


(a) $\rho(T_{\Sigma_1,\ldots,\Sigma_5}) = .7133$



(b) $\rho(T_{\Sigma_1,\ldots,\Sigma_{75}}) = 1.0214$

Figure 7: Effect of the Proximity of the Spectral Radius to One



Conclusions

- **Sensor networks:**
  - **Optimal design: tough combinatorial problem**
  - **Decentralized detection (parallel network):**
    - ➢ **Surprising fusion rule: majority rule**
    - ➢ **Tradeoffs under rate constraint:**
      - ✓ **N versus b as a function of SNR: may prefer N/2 w/ b=2, rather than N and b=1 if more reliable sensors (1.5 dB)**
- **Fusion:**
  - **Sum product algorithm**
  - **Convergence: under appropriate initial conditions on covariances, means converge to correct means**
    - ❑ This generalizes to arbitrary sensor network configurations result of Rusmevichientong and Van Roy (Feb 01) for a fully connected graph and 2 factor nodes