University of Minnesota
Driven to Discover℠

Cartography for Cognitive Networks

Georgios B. Giannakis

Acknowledgments: Prof. G. Mateos, MURI Grant No. AFOSR FA9550-10-1-0567
NSF Grants ECCS-1202135, AST-1247885, ECCS-1002180

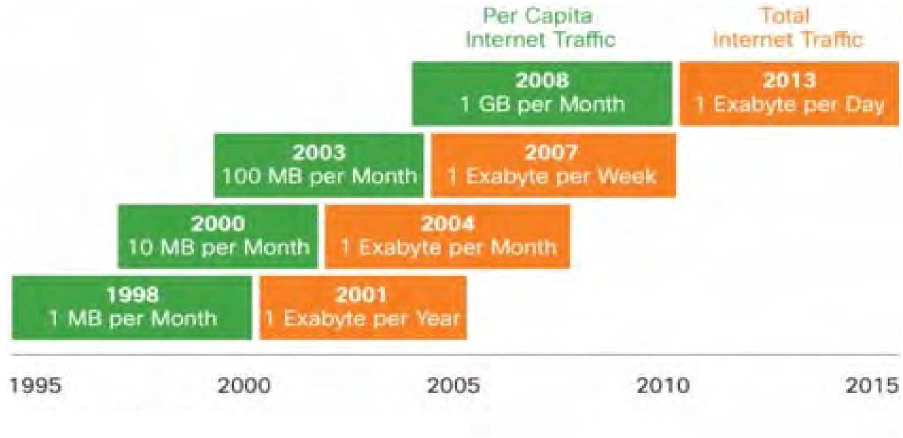
A Coruña, Spain
June 22, 2014

SAM 2014

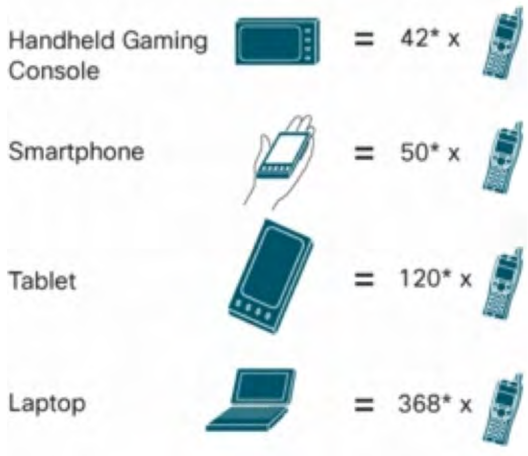
Network traffic growth

Projected IP traffic in Exabytes/month

IP traffic is growing explosively



"Smart" devices multiply traffic



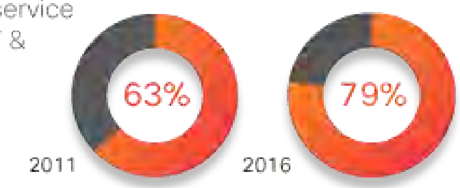
- Communication networks today
  - Large-scale interconnection of "smart" devices
  - Commercial, consumer-oriented, heterogeneous

Source: CISCO Visual Networking Index Global Mobile Data Traffic Forecast Update, 2012-2017

Service diversification

Residential services

By 2016, Online Music will have 1.7 billion users, ranking as the highest penetrated service among all residential service categories (TV & Internet).



Digital TV will be the fastest growing service across all residential service categories (TV & Internet) from 2011 to 2016. ▲ 13.8% CAGR

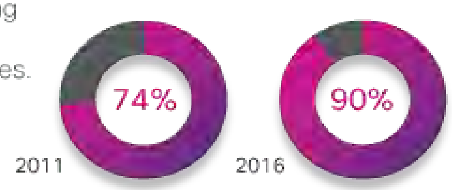
Voice over IP (VoIP) will be the fastest growing residential Internet service from 2011 to 2016.
▲10.6% CAGR

By 2016, Video on Demand (VoD) will be the highest penetrated residential TV service with 258 million subscribers.
14% of digital TV households.

The average household will have 2.5 fixed devices/connections by 2016.

Mobile services

By 2016, Mobile SMS will have 4.1 billion users, ranking as the highest penetrated service among all consumer mobile services.



Mobile Video will be the fastest growing consumer mobile service from 2011 to 2016.
▲ 42.9% CAGR

Mobile Banking & Commerce will be the second fastest growing consumer mobile service from 2011 to 2016.
▲ 42.7% CAGR

By 2016, Mobile Social Networking will be the second highest penetrated consumer mobile service with 2.4 billion users.
53% of consumer mobile users.

By 2016, mobile consumers will have an average of 1.6 devices per person.

Source: CISCO Visual Networking Index Global Mobile Data Traffic Forecast Update, 2012-2017

Dynamic network cartography

- Accurate network diagnosis and statistical analysis tools
  - Seamless end-user experience in dynamic environments
  - Secure and stable network operation
- **Network cartography:** succinct depiction of the *network state*
  - Tool for statistical modeling, monitoring and management
  - Offers situational awareness of the network landscape

Network state

Interference
Path delays
Link utilization
Anomalous flows
Coverage
Traffic volume
Hierarchy/Reputation
Topology
Vulnerability

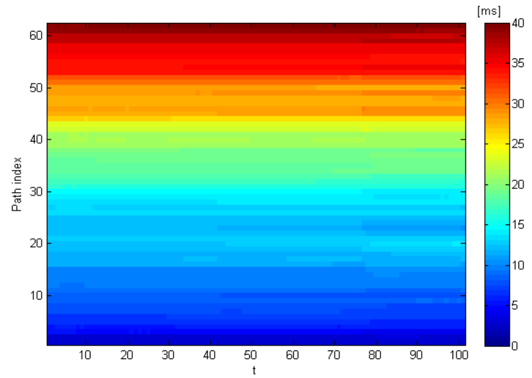
Impact

Information dissemination
Routing
QoS
Congestion control
Network health monitoring
Risk analysis
Security assurance
Spectrum allocation

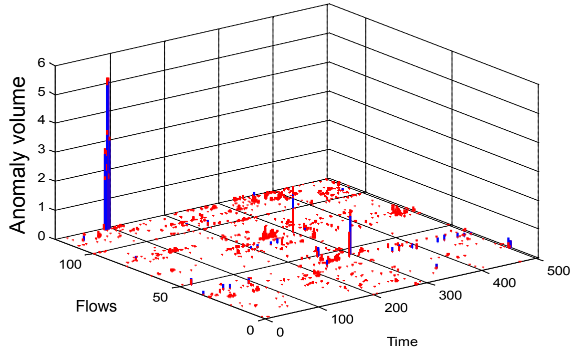
G. Mateos, K. Rajawat, and G. B. Giannakis ”Dynamic network cartography,” *IEEE Signal Processing Magazine*, May 2013.

Tutorial outlook

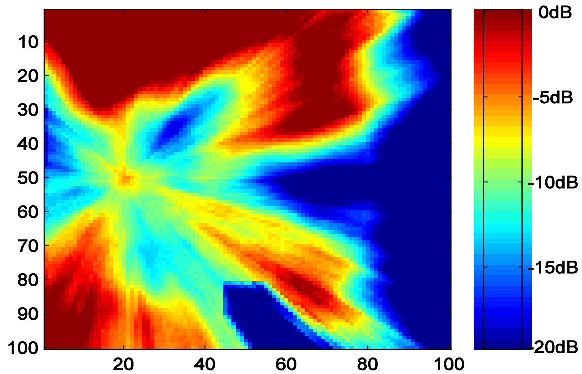
- Dynamic network delay and traffic cartography
  - Map network state via limited measurements
  - Monitor network health



- Dynamic anomalography for IP networks
  - Reveal where and when traffic anomalies occur
  - Leverage sparse anomalies and low-rank traffic

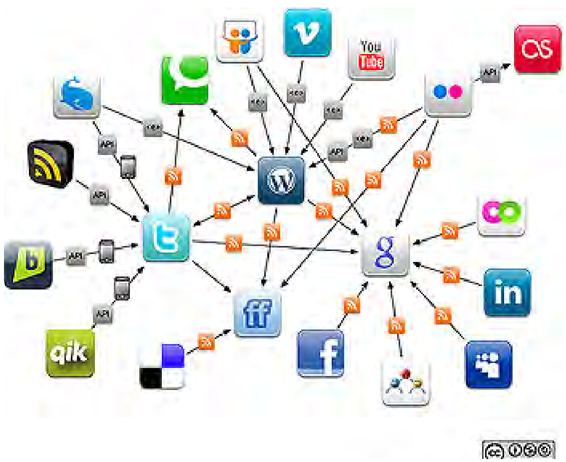


- RF cartography for cognition at the PHY
  - Map ambient RF power in space-time-frequency
  - Identify "crowded" regions to be avoided

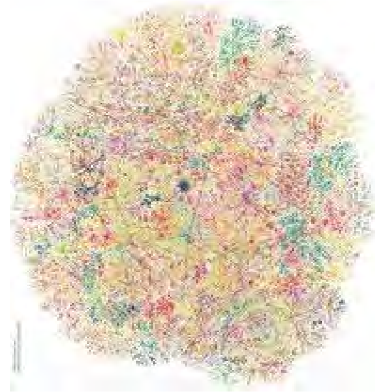


General context: NetSci analytics

Online social media



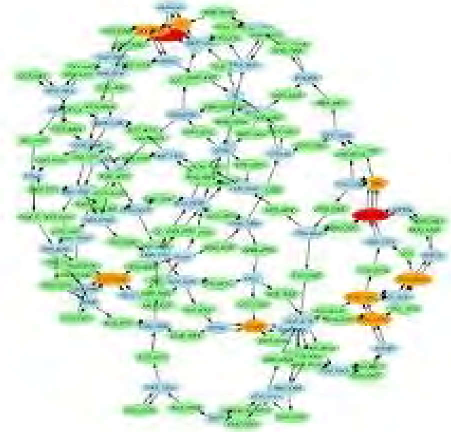
Internet



Clean energy and grid analytics

Robot and sensor networks

Biological networks



Square kilometer array telescope

- **General tools:** process, analyze, and learn from large pools of network data

Roadmap

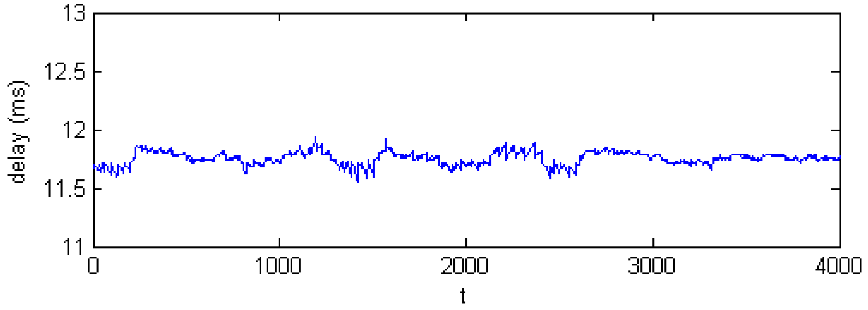
- Dynamic network delay cartography
  - Kriged Kalman filter predictor
  - Optimal network sampling
  - Empirical validation: Internet2 and NZ-AMP data
- Unveiling network anomalies via sparsity and low rank
- Network-wide link count prediction
- RF cartography for cognition at the PHY
- Conclusions and future research directions

Why monitor delays?

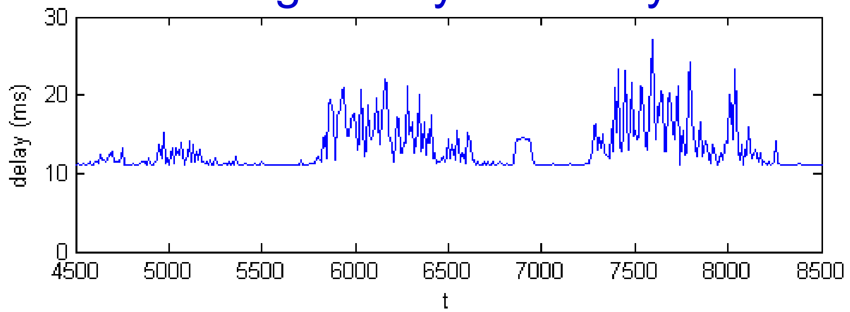
- Motivating reasons
  - Assess network health
  - Fault diagnosis
  - Network planning

- Application domains
  - Old 8-second rule for WWW
  - Content delivery networks
  - Peer-to-peer networks
  - Multiuser games
  - Dynamic server selection

Low delay variability



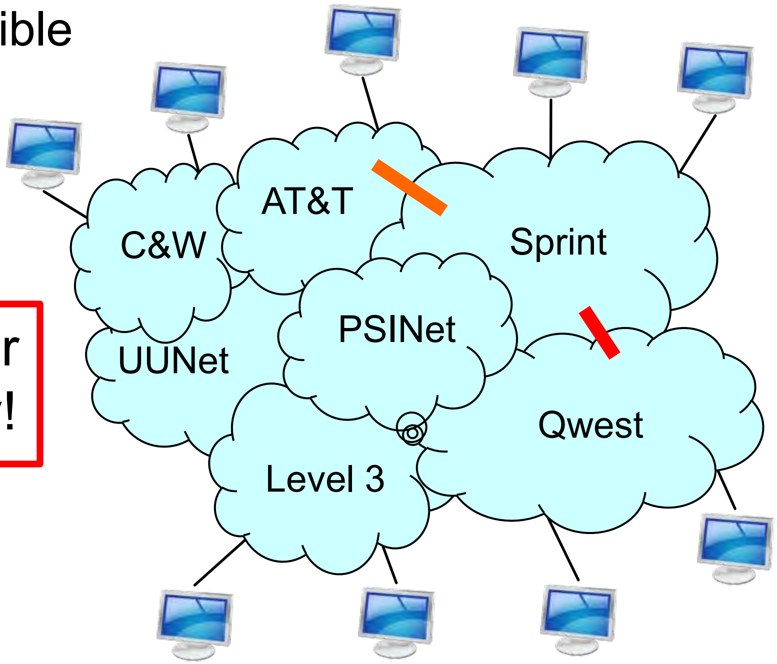
High delay variability



Research issues and goal

- Few tools are widely supported, e.g., `traceroute, ping`
- Additional tools from CAIDA[1]
  - Require software installation at intermediate routers
  - Useless if intermediate routers not accessible

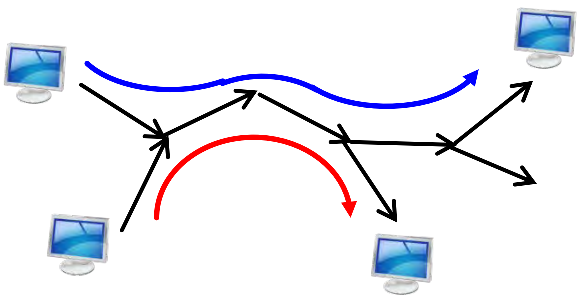
Desiderata: infer delays from a limited number of end-to-end measurements only!



[1]Cooperative Association for Internet Data Analysis. [Online]. www.caida.org

Problem statement

- Consider a network graph with links, nodes, and paths
- Challenges
  - Overhead: # paths ($=: P$) ~ $O(\#\text{ nodes}^2)$
  - Heavily congested routers may drop packets



- **Q:** Can fewer measurements suffice?
  - Most paths tend to share a lot of links [Chua'06]
- Inference task
  - Measure $\mathbf{y}_s$ on subset $\mathcal{S} \subset \mathcal{P}$
  - Predict $\mathbf{y}_{\bar{s}}$ on remaining paths $\bar{\mathcal{S}} := \mathcal{P} \setminus \mathcal{S}$

Network Kriging prediction

- Given $\mathbf{V}_{ss} := \text{cov}(\mathbf{y}_s)$, $\mathbf{V}_{\bar{s}s} := \text{cov}(\mathbf{y}_{\bar{s}}, \mathbf{y}_s)$, universal Kriging:

$$\hat{\mathbf{y}}_{\bar{s}} = \mathbf{V}_{\bar{s}s}\mathbf{V}_{ss}^{-1}\mathbf{y}_s$$

- To obtain $\mathbf{V}_{ss}$, $\mathbf{V}_{\bar{s}s}$ adopt a linear model for path delays

$$\mathbf{y} = \mathbf{G}\mathbf{x} \qquad [\mathbf{G}]_{pl} = \begin{cases} 1 & \text{link } l \in \text{path } p \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_{\bar{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{S} \\ \bar{\mathbf{S}} \end{bmatrix} \mathbf{G}\mathbf{x} \qquad \text{cov}(\mathbf{y}) = \begin{bmatrix} \mathbf{V}_{ss} & \mathbf{V}_{s\bar{s}} \\ \mathbf{V}_{\bar{s}s} & \mathbf{V}_{\bar{s}\bar{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{S} \\ \bar{\mathbf{S}} \end{bmatrix} \mathbf{G}\mathbf{\Sigma}\mathbf{G}^T \begin{bmatrix} \mathbf{S}^T & \bar{\mathbf{S}}^T \end{bmatrix}$$

- Sampling matrix ***S*** known (selected via heuristic algorithms)

D. B. Chua, E. D. Kolaczyk, and M. Crovella, "Network kriging," *IEEE J. Sel. Areas Communications,* vol. 24, no. 12, pp. 2263–2272, Dec. 2006.

Spatio-temporal prediction

- Wavelet-based approach [Coates'07]
  - Diffusion wavelet matrix constructed using network topology
  - Can capture temporal correlations, but for τ time slots
  - High complexity ($O(\tau^3 P^3)$) $\Rightarrow$ cannot have $\tau > 10$

- **Q:** Should the same set of paths be measured per time slot?
  - Load balancing? Measurement on random paths?

- Prior art does not jointly offer
  - Spatio-temporal inference with online path selection, at low complexity

M. Coates, Y. Pointurier, and M. Rabbat, "Compressed network monitoring for IP and all-optical networks," in Proc. *ACM Internet Measurement Conf.*, San Diego, CA, Oct. 2007.

Simple delay model

- Delay measured on path $p \in \mathcal{P}$

$$y_p(t) = \chi_p(t) + \nu_p(t) + \epsilon_p(t)$$

Component due to traffic queuing: random-walk with noise cov. $\mathbf{C}_{\boldsymbol{\eta}}$

$$\boldsymbol{\chi}(t) = \boldsymbol{\chi}(t-1) + \boldsymbol{\eta}(t)$$

Component due to processing, transmission, propagation: Traffic independent, temporally white, w/ cov. $\mathbf{C}_{\boldsymbol{\nu}} = \alpha \mathbf{G}\mathbf{G}^T$

Measurement noise i.i.d. over paths and time with known variance

$$\mathbb{E}[\epsilon_p(t)\epsilon_p^T(t)] = \sigma^2$$

K. Rajawat, E. Dall'Anese, and G. B. Giannakis, "Dynamic network delay cartography," *IEEE Transactions on Information Theory,* 2013.

Kriged Kalman Filter: Formulation

- Path measured on subset $\mathcal{S} \in \mathcal{P}$

$$\mathbf{y}_s(t) = \mathbf{S}(t)\boldsymbol{\chi}(t) + \boldsymbol{\nu}_s(t) + \boldsymbol{\epsilon}_s(t)$$

$$\boldsymbol{\nu}_s(t) := \mathbf{S}(t)\boldsymbol{\nu}(t)$$

- KKF:

$$\begin{aligned} \boldsymbol{\chi}(t) &= \boldsymbol{\chi}(t-1) + \boldsymbol{\eta}(t) \\ \mathbf{y}_s(t) &= \mathbf{S}(t)\boldsymbol{\chi}(t) + \boldsymbol{\nu}_s(t) + \boldsymbol{\epsilon}_s(t) \end{aligned}$$

Goal: Given history $\mathcal{H}(t) := \{\mathbf{y}_s(\tau)\}_{\tau=1}^{t}$ find $\hat{\mathbf{y}}_{\bar{s}}(t)$

KKF updates

- State and covariance recursions

$$\hat{\boldsymbol{\chi}}(t) := \mathbb{E}^*[\boldsymbol{\chi}(t)|\mathcal{H}(t)]$$
$$= \hat{\boldsymbol{\chi}}(t-1) + \mathbf{K}(t)(\mathbf{y}_s(t) - \mathbf{S}(t)\hat{\boldsymbol{\chi}}(t-1))$$

$$\mathbf{M}(t) := \mathbb{E}[(\hat{\boldsymbol{\chi}}(t) - \boldsymbol{\chi}(t))(\hat{\boldsymbol{\chi}}(t) - \boldsymbol{\chi}(t))^T]$$
$$= (\mathbf{I} - \mathbf{K}(t)\mathbf{S}(t))(\mathbf{M}(t-1) + \mathbf{C}_{\boldsymbol{\eta}})$$

- KKF gain

$$\mathbf{K}(t) := (\mathbf{M}(t-1) + \mathbf{C}_{\boldsymbol{\eta}})\mathbf{S}^T(t)\left[\mathbf{S}(t)(\mathbf{M}(t-1) + \mathbf{C}_{\boldsymbol{\eta}} + \mathbf{C}_{\boldsymbol{\nu}})\mathbf{S}^T(t) + \sigma^2\mathbf{I}\right]^{-1}$$

- Kriging predictor

$$\hat{\mathbf{y}}_{\bar{s}}(t) = \bar{\mathbf{S}}(t)\hat{\boldsymbol{\chi}}(t) + \bar{\mathbf{S}}(t)\mathbf{C}_{\boldsymbol{\nu}}\mathbf{S}^T(t)\left[\mathbf{S}(t)\mathbf{C}_{\boldsymbol{\nu}}\mathbf{S}^T(t) + \sigma^2\mathbf{I}\right]^{-1}(\mathbf{y}_s(t) - \mathbf{S}(t)\hat{\boldsymbol{\chi}}(t))$$

Which paths to measure?

- KKF can model and track network-wide delays
  - Practical sampling of paths? Optimal measurements? Criterion?

- Error covariance matrix

$$\mathbf{M}^{\mathbf{y}}_{\bar{s}}(t) := \mathbb{E}\left\{(\mathbf{y}_{\bar{s}}(t) - \hat{\mathbf{y}}_{\bar{s}}(t))(\mathbf{y}_{\bar{s}}(t) - \hat{\mathbf{y}}_{\bar{s}}(t))^T\right\}$$
$$= \sigma^2\mathbf{I}_{\bar{S}} + \sigma^2\bar{\mathbf{S}}(t)\left[\mathbf{\Phi}^{-1} + \mathbf{S}^T(t)\mathbf{S}(t)\right]^{-1}\bar{\mathbf{S}}^T(t)$$

$$\mathbf{\Phi} = (\mathbf{M}(t-1) + \mathbf{C}_{\boldsymbol{\eta}} + \mathbf{C}_{\boldsymbol{\nu}})/\sigma^2$$

- **Online experimental design:** minimize $\log\det(\mathbf{M}^{\mathbf{y}}_{\bar{s}}(t)) =: -f_t(\mathcal{S})$

$$\mathcal{S}^*(t) := \arg\max_{\mathcal{S}\subset\mathcal{P}} f_t(\mathcal{S}) \quad \text{subject to } |\mathcal{S}| = S$$

  - Log-det: D-optimal design (entropy of a Gaussian r. v.)

Greedy algorithm

Algorithm

Repeat *S* times

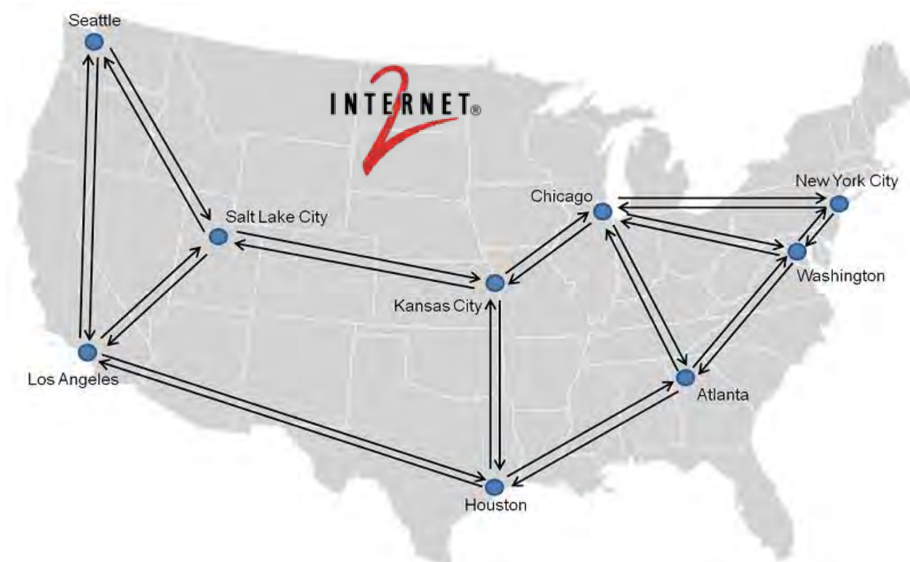
$$\mathcal{S} \leftarrow \mathcal{S} \cup \arg\max_{p \notin \mathcal{S}} \delta_{\mathcal{S}}(p)$$

- Submodular + monotonic $\Rightarrow$ greedy solution $\left(1 - \frac{1}{e}\right)$ optimal [Nemhauser'78]
- Increments can be evaluated efficiently: $O(PS^3)$ with $P \gg S$
  - Operational complexity can be reduced further [Krause'11]
- Can be modified to handle cases when
  - Each node measures delay on all paths – which *S* nodes to choose?
  - All nodes measure delay on only one path – which path to choose?

A. Krause, C. Guestrin. "Submodularity and its Applications in Optimized Information Gathering: An Introduction", *ACM Transactions on Intelligent Systems and Technology*, vol. 2 (4), July 2011

Empirical validation: Internet2

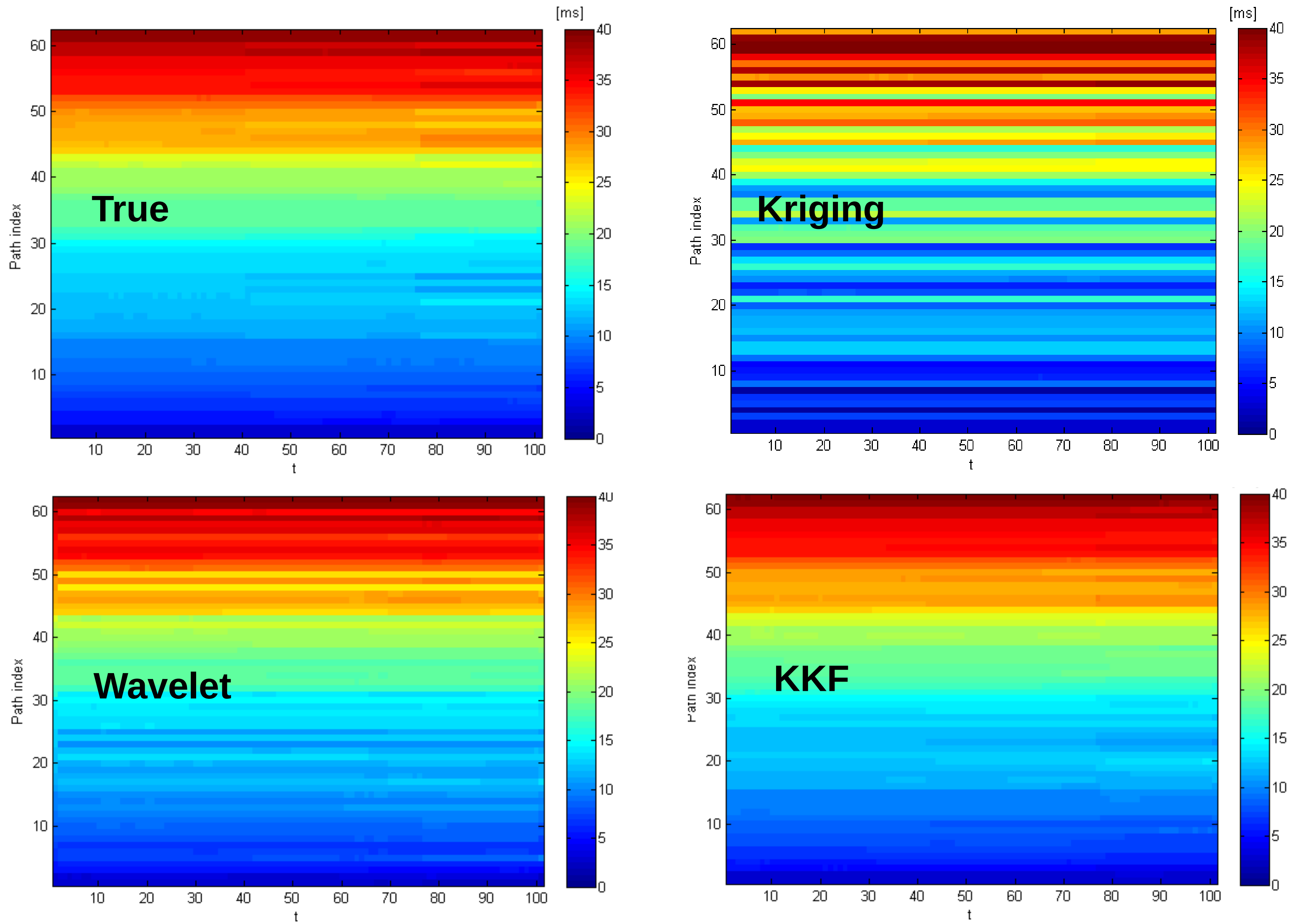
- Internet2 backbone
  - 72 paths
  - Lightly loaded



- One-way delay measurements using `OWAMP`
- Measurements every minute for 3 days in July 2011 ~ 4500 samples
- Training phase employed to estimate C_η, α
  - Empirical estimates; see e.g., [Myers'76]
  - Techniques modified to handle measurements on subset of paths
  - First 1000 samples used for training; 50 random paths used for training

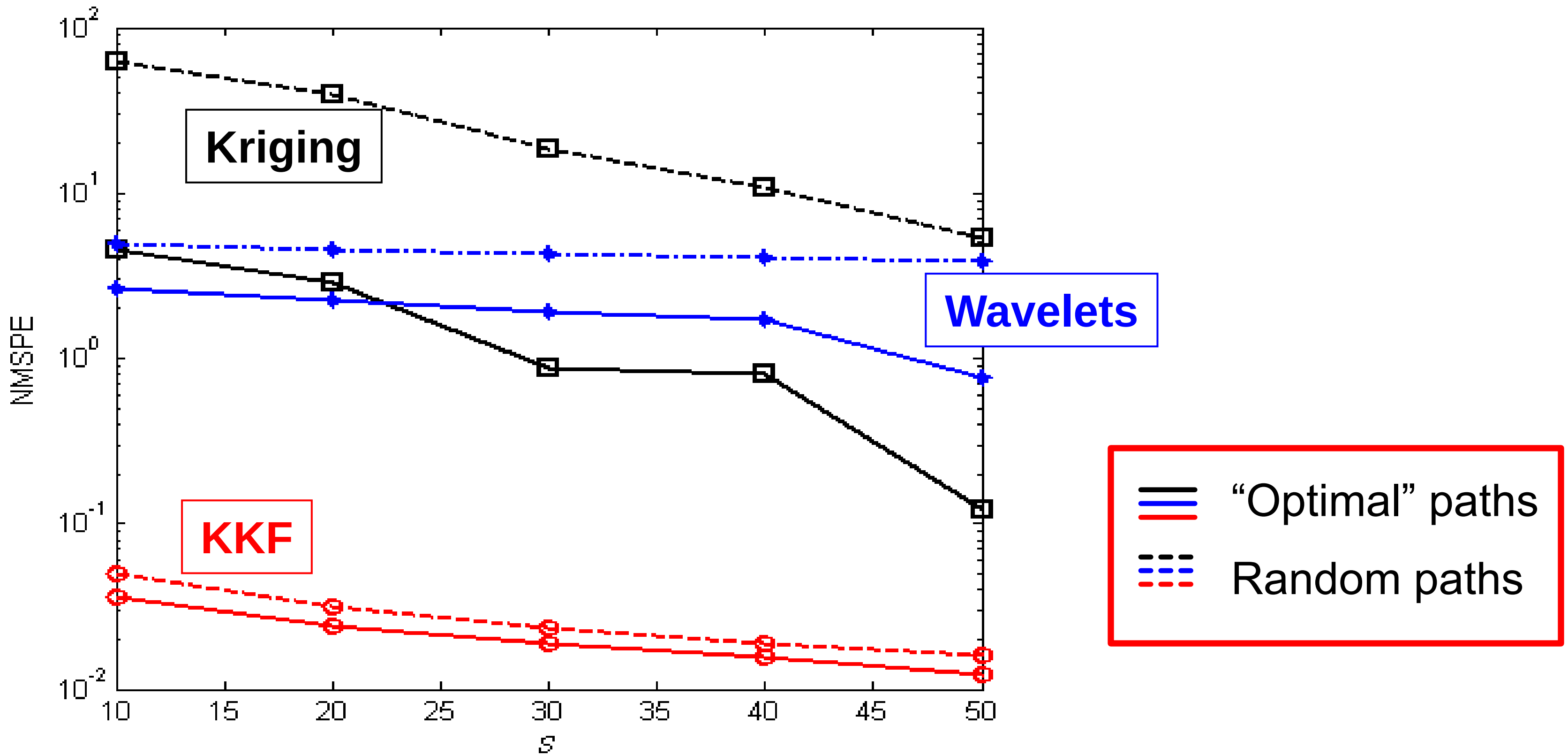
Data: http://internet2.edu/observatory/archive/data-collections.html

Network delay cartography (Internet2)



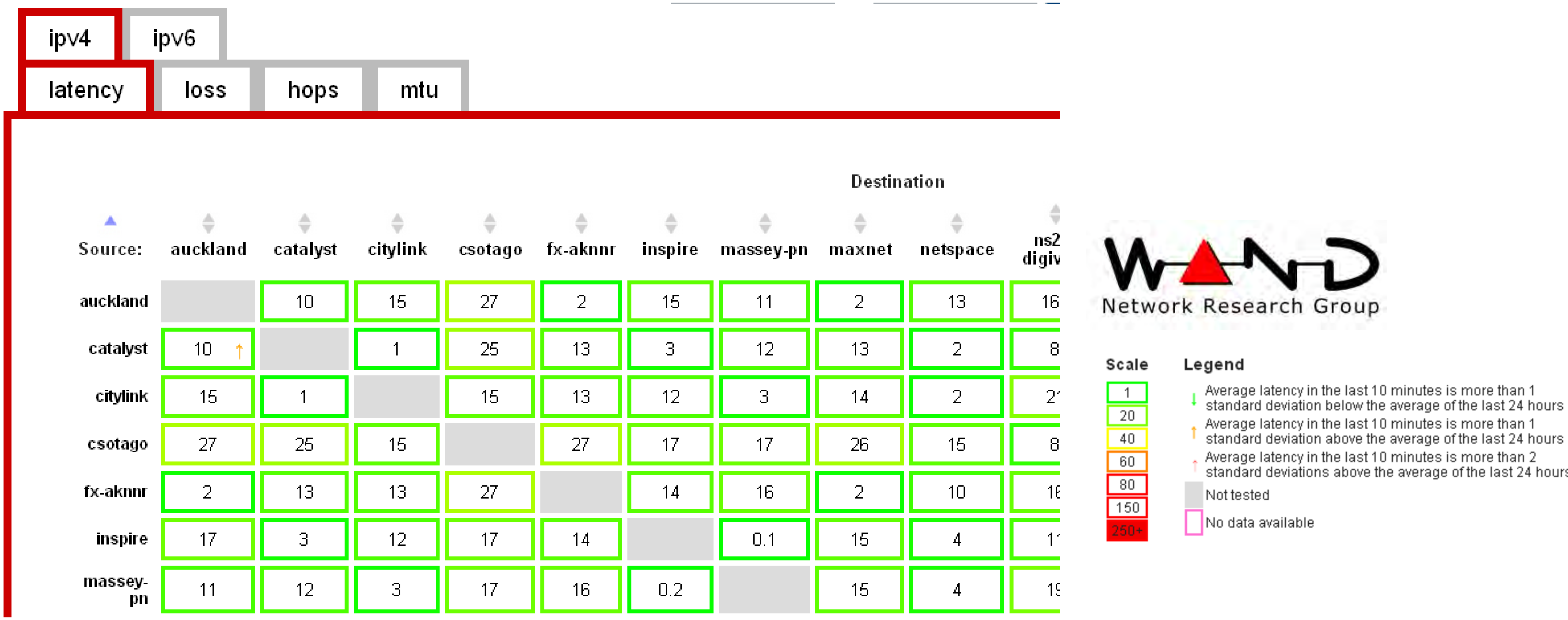
Normalized MSPE (Internet2)

- Normalized MSPE $:= \frac{1}{T(P-S)} \sum_{t=1}^{T} \|\hat{\mathbf{y}}_{\bar{s}}(t) - \mathbf{y}_{\bar{s}}(t)\|^2$



Empirical validation: NZ-AMP

- Delays measured on NZ-AMP, part of NLANR project
  - 186 paths, heavily loaded network

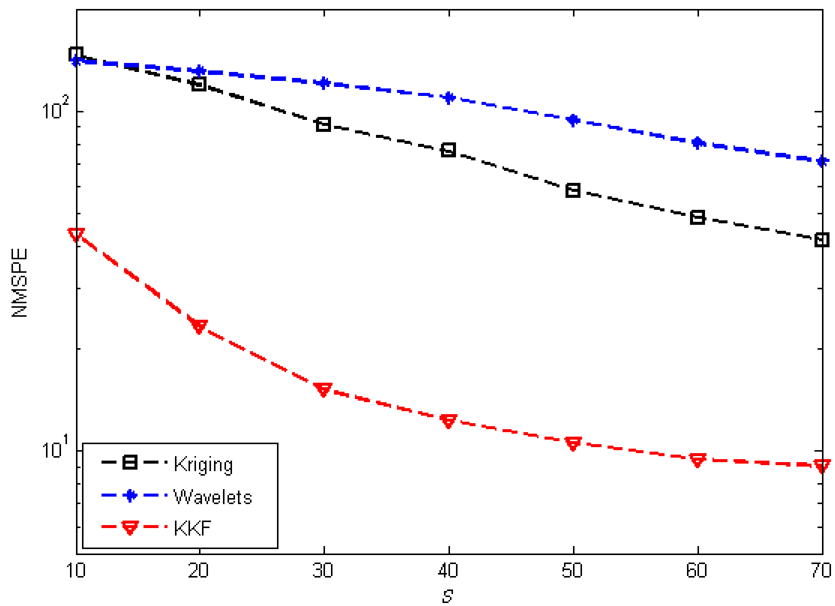


Source:	auckland	catalyst	citylink	csotago	fx-aknnr	inspire	massey-pn	maxnet	netspace	ns2 digiv
auckland		10	15	27	2	15	11	2	13	16
catalyst	10 ↑		1	25	13	3	12	13	2	8
citylink	15	1		15	13	12	3	14	2	2
csotago	27	25	15		27	17	17	26	15	8
fx-aknnr	2	13	13	27		14	16	2	10	1
inspire	17	3	12	17	14		0.1	15	4	1
massey-pn	11	12	3	17	16	0.2		15	4	1

- Measurements every 10 minutes during August 2011 ~ 4500 samples
- Round-trip times measured using `ICMP`, paths via `scamper`

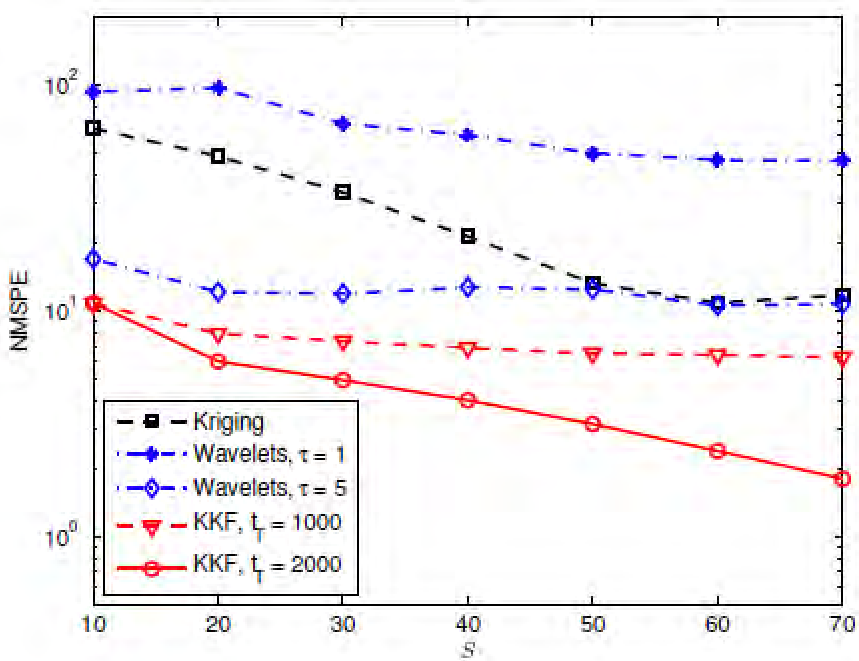
Data: http://erg.wand.net.nz

Normalized MSPE (NZ-AMP)

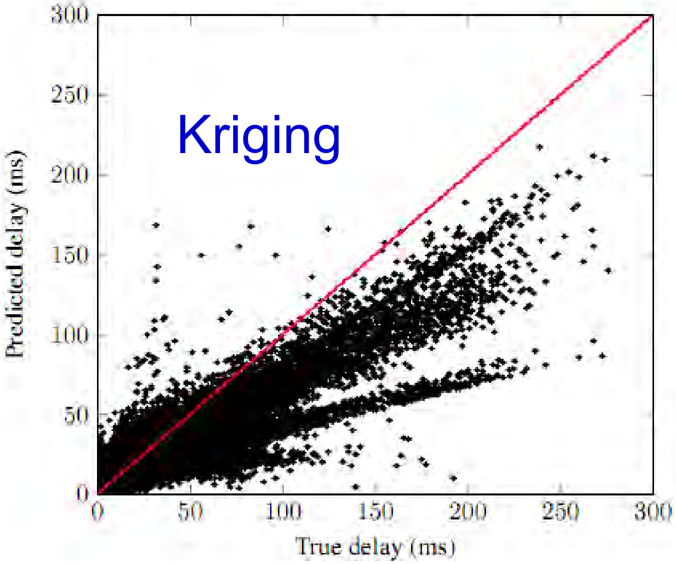


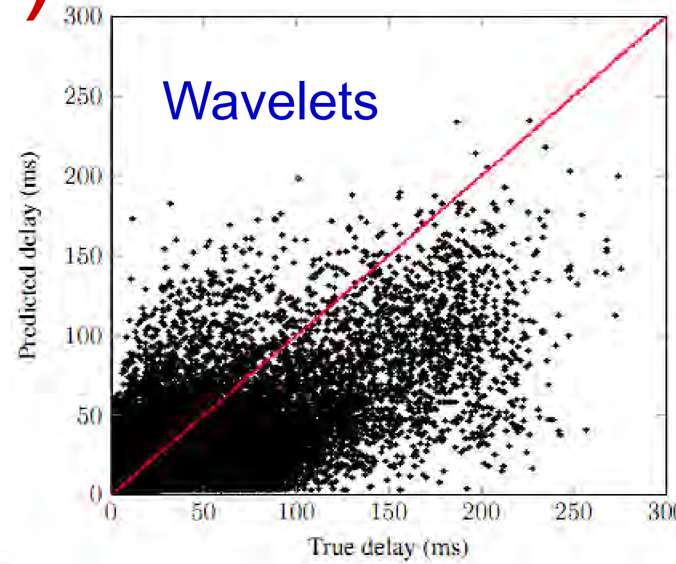
Random path selection

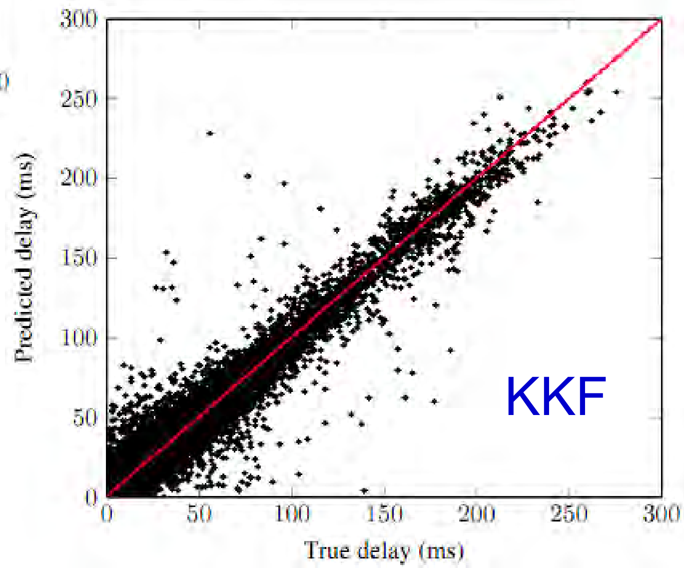
"Optimal" path selection



Scatter plots (NZ-AMP)







$\hat{\mathbf{y}}_{\bar{S}}(t)$ vs. $\mathbf{y}_{\bar{S}}(t)\ \forall\, t$

$S = 30$

Takeaways

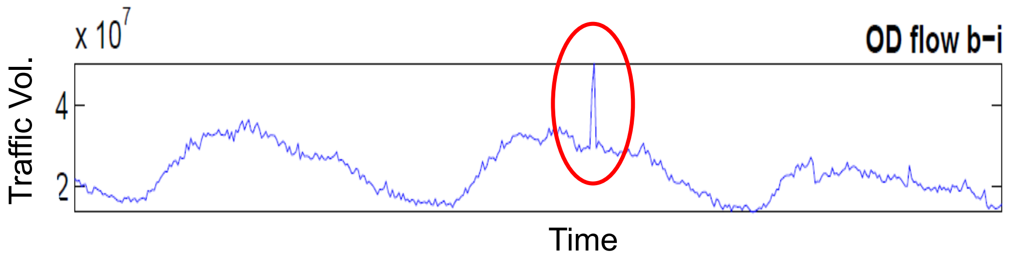
- Spatio-temporal inference useful for network health monitoring
- Dynamic network delay cartography via Kriged Kalman filtering
- Near-optimal path selection by utilizing submodularity
- Empirical validation on Internet2 and NZ-AMP datasets

Roadmap

- Dynamic network delay cartography
- Unveiling network anomalies via sparsity and low rank
  - Traffic modeling and identifiability
  - (De-) centralized and online algorithms
  - Numerical tests
- Network-wide link count prediction
- RF cartography for cognition at the PHY
- Conclusions and future research directions

Traffic anomalies

- Backbone of IP networks
- Traffic anomalies: changes in origin-destination (OD) flows
  - Failures, transient congestions, DoS attacks, intrusions, flooding



- Motivation: Anomalies $\Rightarrow$ congestion $\Rightarrow$ limits end-user QoS provisioning

Objective: Measuring superimposed OD flows per link, identify anomalies by leveraging sparsity of anomalies and low-rank of traffic.

Model

- Graph $G(N, L)$ with N nodes, L links, and F flows ($F >> L$)

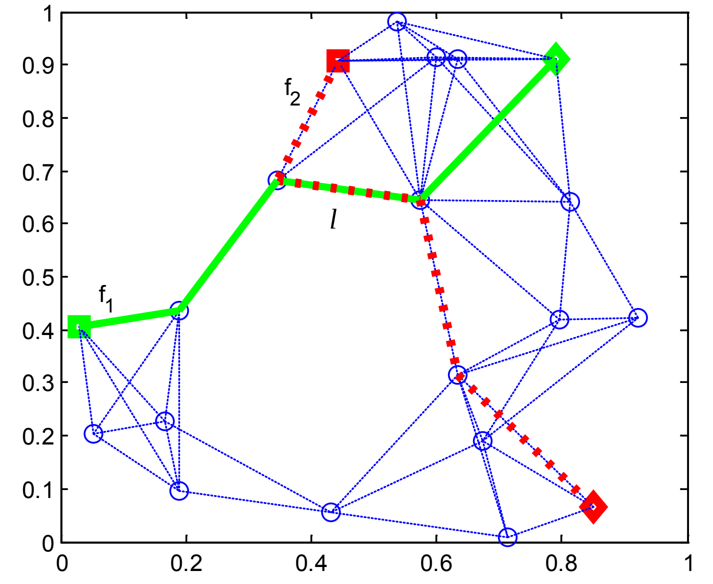
  (as) Single-path per OD flow $z_{f,t}$

- Packet counts per link l and time slot t

Anomaly

$$y_{l,t} = \sum_{f=1}^{F} r_{l,f}(z_{f,t} + a_{f,t}) + v_{l,t}$$

$r_{l,f} \in \{0,1\}$



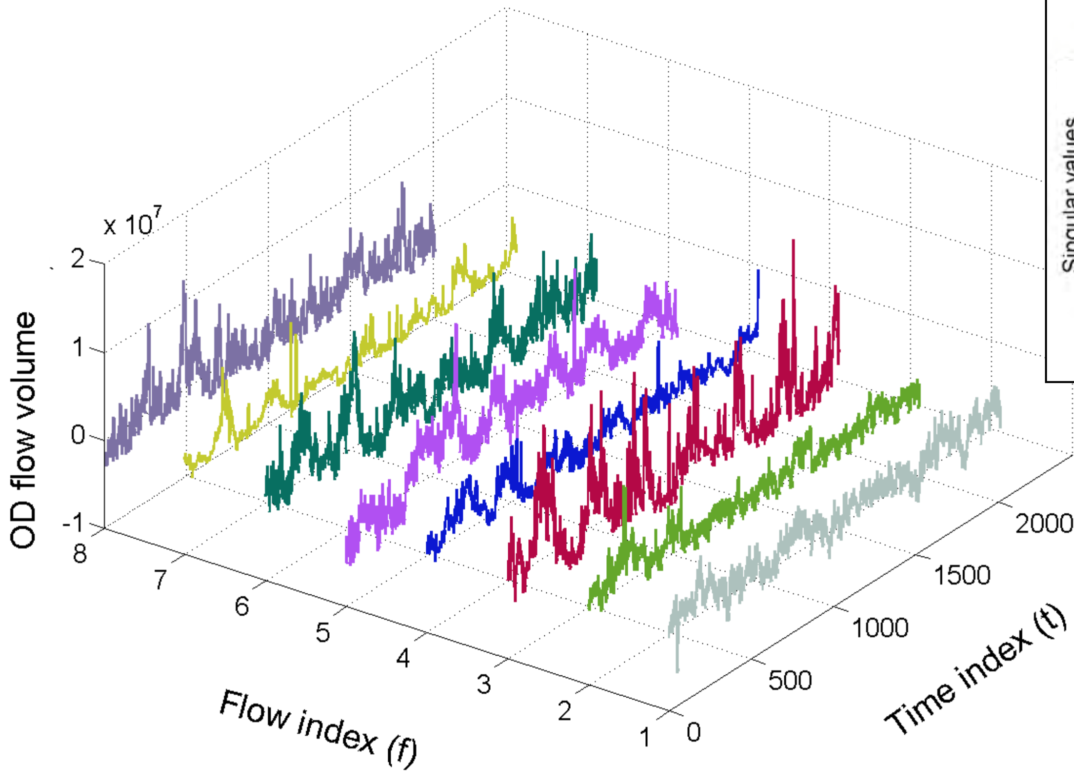
- Matrix model across T time slots: $\mathbf{Y} = \mathbf{R}(\mathbf{Z} + \mathbf{A}) + \mathbf{V}$

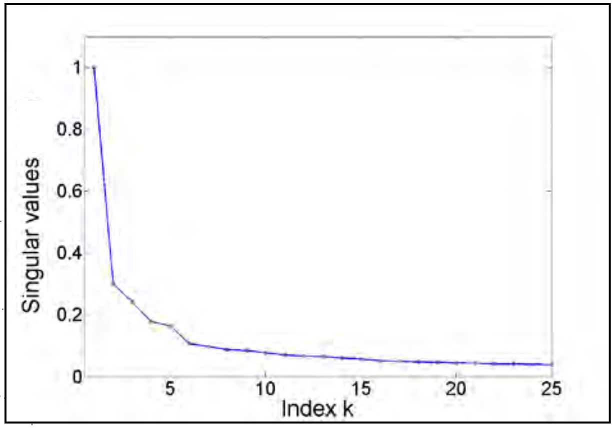
  $\mathbf{Y}$: LxT, $\mathbf{R}$: LxF

Low rank of traffic matrix

$$\mathbf{Y} = \mathbf{R}(\mathbf{Z} + \mathbf{A}) + \mathbf{V}$$

- **Z**: traffic matrix has low rank, e.g., [Lakhina et al'04]





Data: http://math.bu.edu/people/kolaczyk/datasets.html

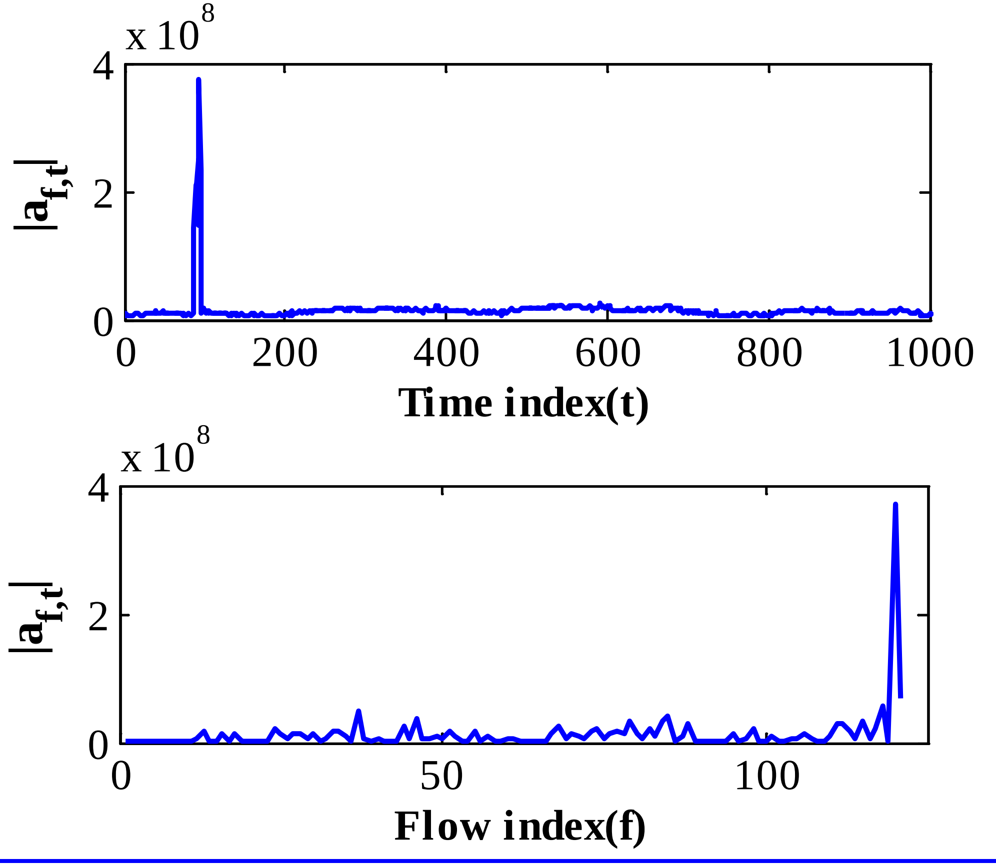
Sparsity of anomaly matrix

$$\mathbf{Y} = \mathbf{R}(\mathbf{Z} + \mathbf{A}) + \mathbf{V}$$

- **A**: anomaly matrix is sparse across both time and flows

Time

Flows



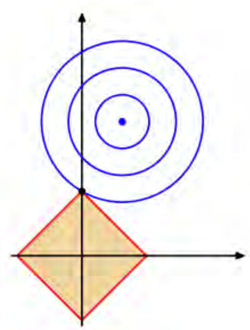
Problem statement

$$\mathbf{Y} = \underbrace{\mathbf{RZ}}_{:=\mathbf{X}} + \mathbf{RA} + \mathbf{V}$$

- Given $\mathbf{Y}$ and routing matrix $\mathbf{R}$, identify sparse $\mathbf{A}$ when $\mathbf{Z}$ is low rank
  - $\mathbf{R}$ fat but $\mathbf{X}$ still low rank

$$\{\hat{\mathbf{X}}, \hat{\mathbf{A}}\} = \arg\min_{\{\mathbf{X},\mathbf{A}\}} \frac{1}{2}\|\mathbf{Y} - \mathbf{X} - \mathbf{RA}\|_F^2 + \lambda_1\|\mathbf{A}\|_1 + \lambda_*\|\mathbf{X}\|_* \quad \text{(P1)}$$

$\sum_{i,j} |a_{i,j}|$



$\sum_i \sigma_i(\mathbf{X})$

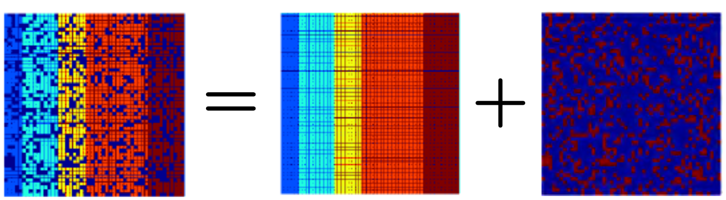
- Low-rank $\Rightarrow$ sparse vector of SVs $\Rightarrow$ nuclear norm $\|\cdot\|_*$ and ℓ_1 norm

Prior art

- Anomaly identification
  - Change detection on per-link time series [Brutlag'00], [Casas et al'10]
  - Spatial PCA [Lakhina et al'04]
  - Network anomography [Zhang et al'05]

- Rank minimization with the nuclear norm, e.g., [Recht-Fazel-Parrilo'10]
  - Matrix decomposition [Candes et al'10], [Chandrasekaran et al'11]

$$\mathbf{M} = \mathbf{L}_0 + \mathbf{S}_0$$



Observed Low rank Sparse

Principal Component Pursuit

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{s. to} \quad \mathbf{M} = \mathbf{L} + \mathbf{S} \qquad \text{(PCP)}$$

Challenges and importance

$$\mathbf{Y} = \mathbf{X} + \mathbf{RA} + \mathbf{V}$$

- $\mathbf{RA}$ not necessarily sparse and $\mathbf{R}$ fat $\Rightarrow$ PCP not applicable
- $\underbrace{LT}_{X} + \underbrace{FT}_{A} \gg \underbrace{LT}_{Y}$

STRUCTURE

- Important special cases
  - ***R*** = ***I*** : matrix decomposition with PCP [Candes et al'10]
  - ***X*** = ***0*** : compressive sampling with basis pursuit [Chen et al'01]
  - $\boldsymbol{X} = \boldsymbol{C}_{L x \rho} \boldsymbol{W}'_{\rho x T}$ and ***A*** = ***0*** : PCA [Pearson 1901]
  - ***X*** = ***0, R*** = ***D*** unknown: dictionary learning [Olshausen'97]

Exact recovery

- Noise-free case

$$\mathbf{Y} = \mathbf{X}_0 + \mathbf{R}\mathbf{A}_0 = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}' + \mathbf{R}\mathbf{A}_0$$

$$r = \text{rank}[\mathbf{X}_0], \quad s = \|\mathbf{A}_0\|_0$$

$$\min_{\{\mathbf{X},\mathbf{A}\}} \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.to } \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A} \qquad \text{(P0)}$$

Q: Can one recover sparse $\mathbf{A}_0$ and low-rank $\mathbf{X}_0$ exactly?

A: Yes! Under certain conditions on $\{\mathbf{X}_0, \mathbf{A}_0, \mathbf{R}\}$

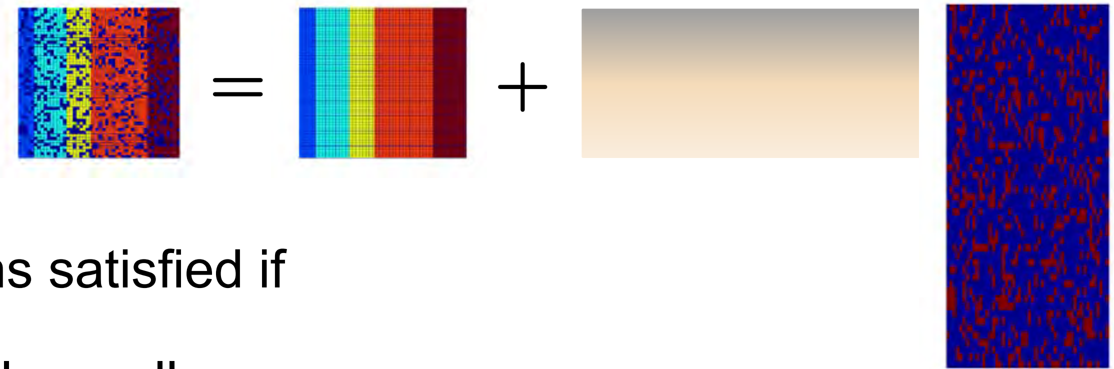
Theorem: Given $\mathbf{Y}$ and $\mathbf{R}$, assume every row and column of $\mathbf{A}_0$ has at most *k*<*s* non-zero entries, and $\mathbf{R}$ has full row rank. If C1)-C2) hold, then with $\lambda \in (\lambda_{\min}, \lambda_{\max})$ (P0) exactly recovers $\{\mathbf{X}_0, \mathbf{A}_0\}$

C1) $(1 - \mu(\Phi, \Omega_R))^2(1 - \delta_k(\mathbf{R})) > \alpha$

C2) $\lambda_{\min} := \beta\|\mathbf{R}'\mathbf{U}\mathbf{V}'\|_\infty < \lambda_{\max} := \sqrt{s^{-1}}[\gamma^{-1} - \mu(\Phi, \Omega_R)\sqrt{r(1 + \delta_k(\mathbf{R}))}]$

M. Mardani, G. Mateos, and G. B. Giannakis,``Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Trans. Information Theory*, 2013.

Intuition



- Exact recovery conditions satisfied if
  - r and s are sufficiently small
  - Nonzero entries of $\boldsymbol{A}_0$ are “sufficiently spread out”
  - Incoherent rank and sparsity- preserving subspaces
  - $\boldsymbol{R}$ satisfies a restricted isometry property
- Remarks
  - Amplitude of non-zero entries of $\boldsymbol{A}_0$ irrelevant
  - Conditions satisfied for certain random ensembles w.h.p.

Numerical validation

- Setup

  L=105, F=210, T = 420

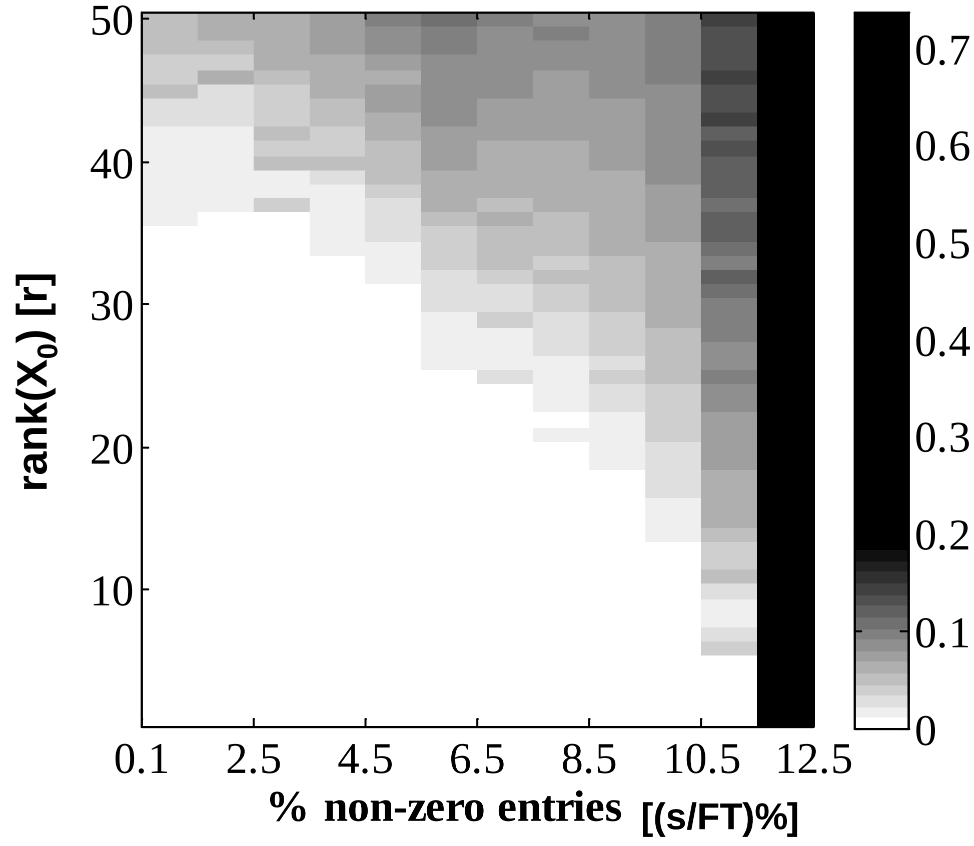
  $\mathbf{R}$ ~ Bernoulli(1/2)

  $\mathbf{X}_o$ = $\mathbf{RPQ}$', $\mathbf{P}$, $\mathbf{Q}$ ~ N(0, 1/FT)

  a_{ij} ϵ {-1,0,1} w.p. {π/2, 1-π, π/2}

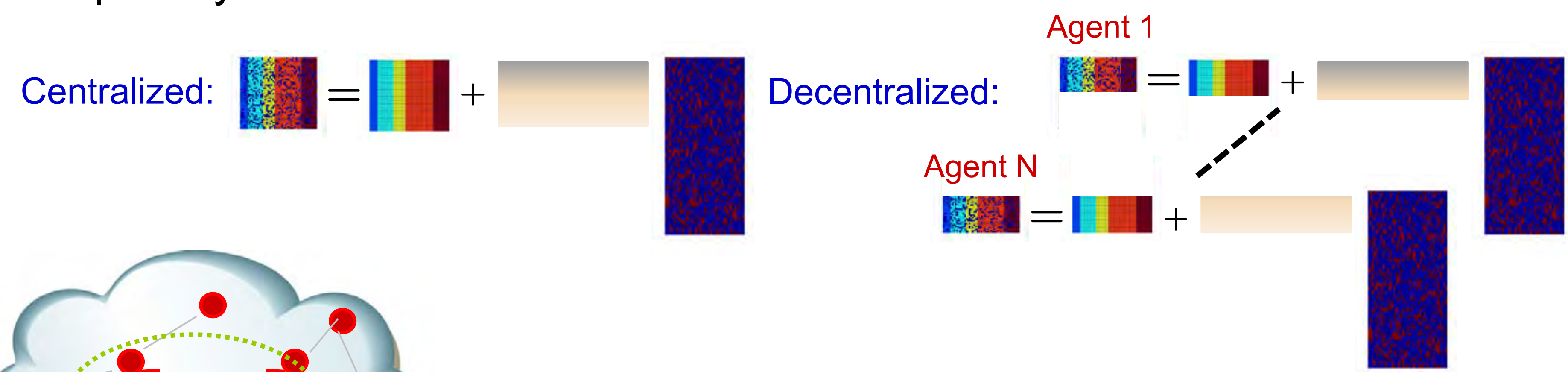
- Relative recovery error

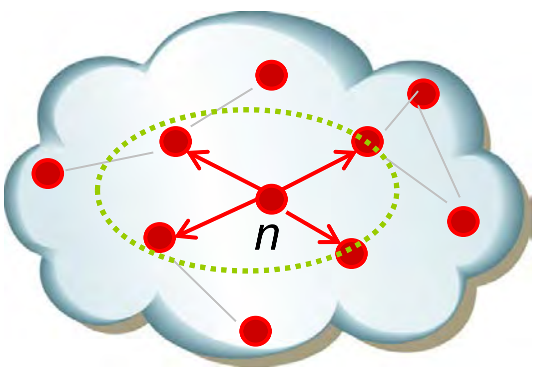
$$e = \frac{\|\hat{\mathbf{A}} - \mathbf{A}_0\|_F}{\|\mathbf{A}_0\|_F}$$



In-network processing

- Spatially-distributed link count data





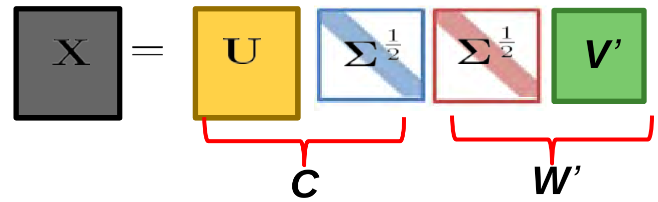
- Local processing and single-hop communications

Goal: Given local link counts per agent, unveil anomalies in a distributed fashion by leveraging low-rank of the nominal data matrix and sparsity of the outliers.

- Challenge: $\|\cdot\|_*$ not separable across rows (links/agents)

Separable regularization

- Key property



$$\|\mathbf{X}\|_* := \min_{\{\mathbf{C},\mathbf{W}\}} \frac{1}{2}\left\{\|\mathbf{C}\|_F^2 + \|\mathbf{W}\|_F^2\right\}, \text{ s.to } \mathbf{X} = \mathbf{C}\mathbf{W}'$$

Lxρ

$\geq$rank[$\boldsymbol{X}$]

- Separable formulation equivalent to (P1)

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{A}\}} \frac{1}{2}\|\mathbf{Y} - \mathbf{C}\mathbf{W}' - \mathbf{R}\mathbf{A}\|_F^2 + \lambda_1\|\mathbf{A}\|_1 + \frac{\lambda_*}{2}\{\|\mathbf{C}\|_F^2 + \|\mathbf{W}\|_F^2\} \quad \text{(P2)}$$

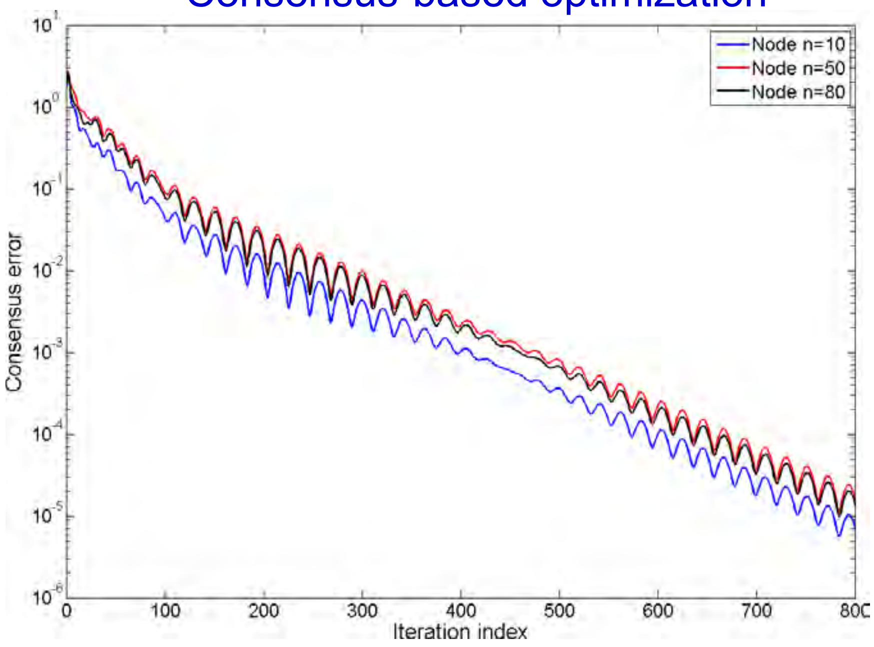
  - Nonconvex; less variables: $LT \Rightarrow \rho(L+T)$

Proposition 3: If $\{\bar{\mathbf{C}}, \bar{\mathbf{W}}, \bar{\mathbf{A}}\}$stat. pt. of (P2) and $\|\mathbf{Y} - \bar{\mathbf{C}}\bar{\mathbf{W}}' - \mathbf{R}\bar{\mathbf{A}}\| \leq \lambda_*$, then $\{\hat{\mathbf{X}} := \bar{\mathbf{C}}\bar{\mathbf{W}}', \hat{\mathbf{A}} := \bar{\mathbf{A}}\}$ is a *global optimum* of (P1).

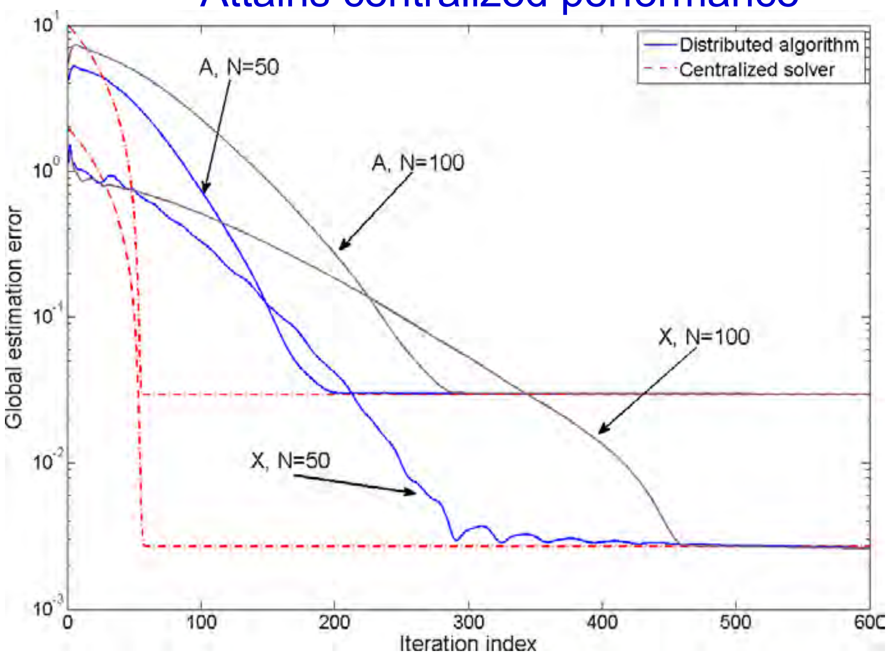
Distributed algorithm

- Alternating-direction method of multipliers (ADMM) solver for (P2)
  - Method [Glowinski-Marrocco'75], [Gabay-Mercier'76]
  - Learning over networks [Schizas-Ribeiro-Giannakis'07]

Consensus-based optimization



Attains centralized performance



M. Mardani, G. Mateos, and G. B. Giannakis, "In-network sparsity regularized rank minimization: Algorithms and applications," *IEEE Transactions on Signal Processing*, 2013.

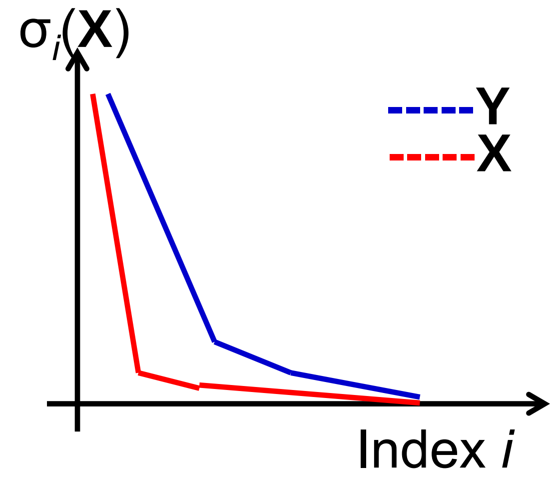
Benchmark: PCA-based methods

- Idea: anomalies increase considerably rank(**Y**)

Algorithm

i) Form subspace $\mathcal{N}$ via *r*-dominant left singular vectors of ***Y*** (resp. $\mathcal{N}^c$)

ii) Infer anomalies from $\mathcal{P}_{\mathcal{N}^c}(\mathbf{Y})$



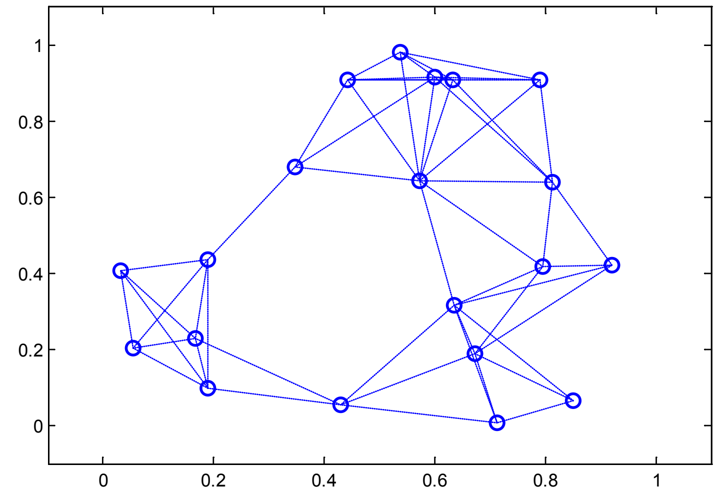
- Assumes knowledge of r:=rank(***X***)

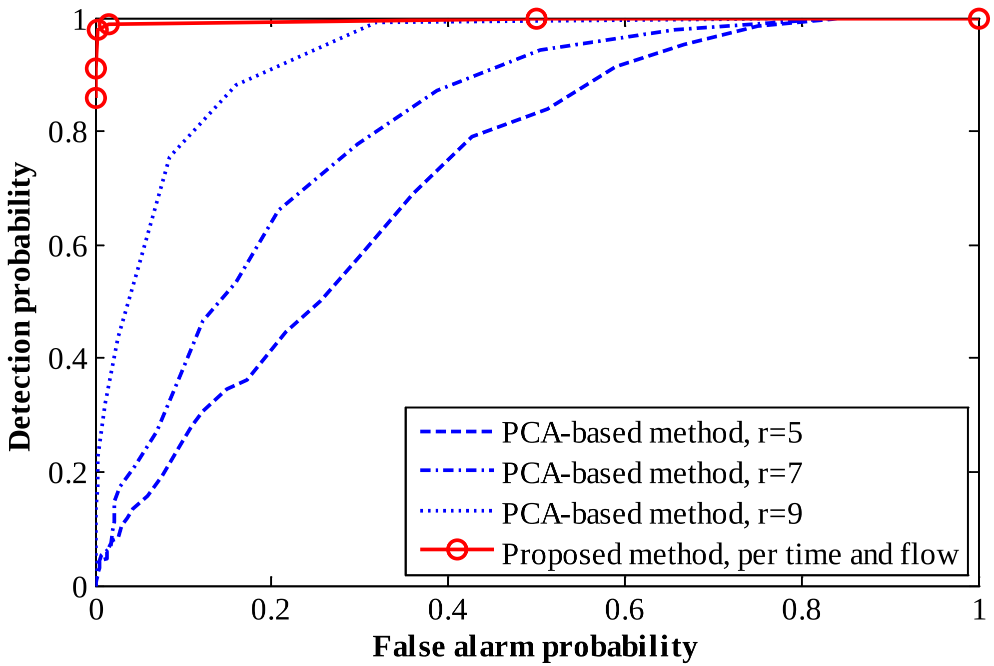
- [Lakhina et al'04] For *t* = 1,…,*T* $\|\mathcal{P}_{\mathcal{N}^c}(\mathbf{y}_t)\|_2 \gtrless_{H_0}^{H_1} \tau$

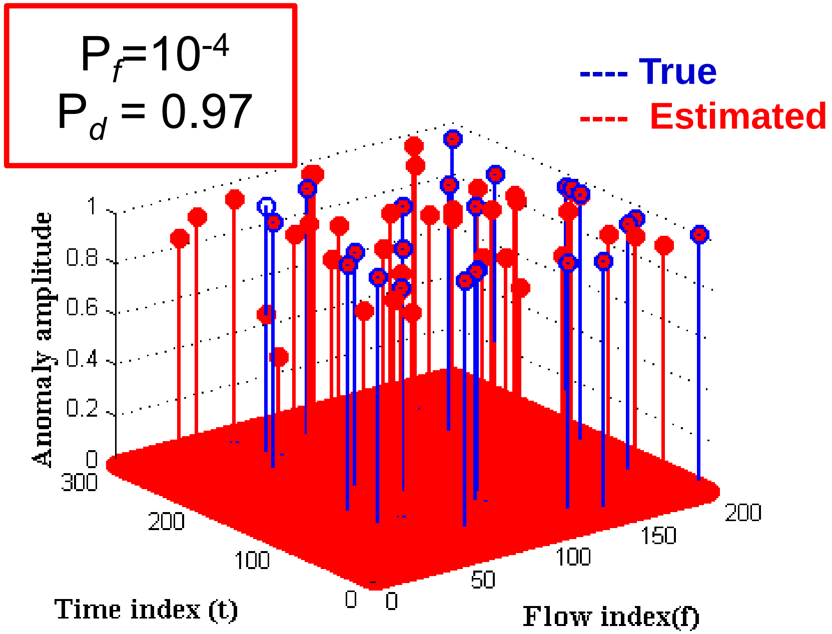
- [Zhang et al'05] Sparse anomalies $\hat{\mathbf{A}} = \arg \min_{\mathcal{P}_{\mathcal{N}^c}(\mathbf{Y})=\mathbf{RA}} \|\mathbf{A}\|_1$

Synthetic data

- Random network topology
  - N=20, L=108, F=360, T=760
  - Minimum hop-count routing

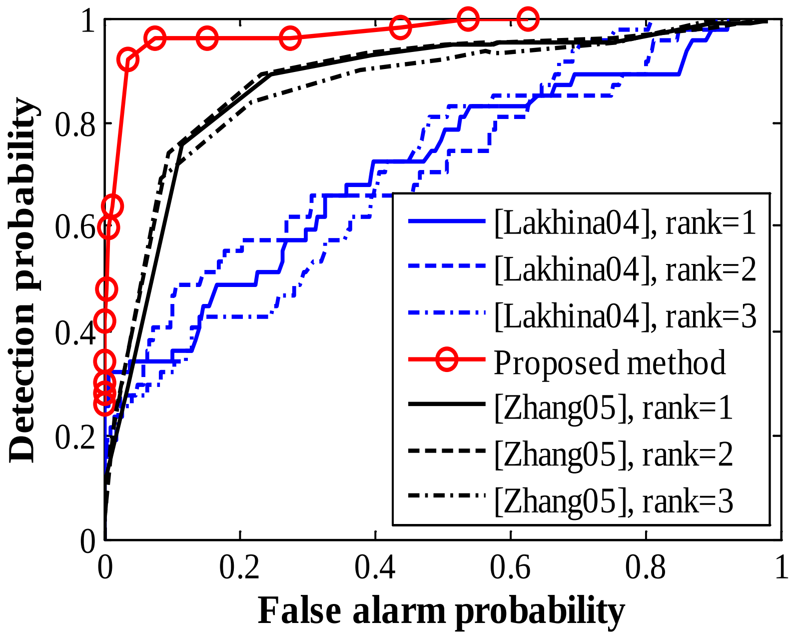


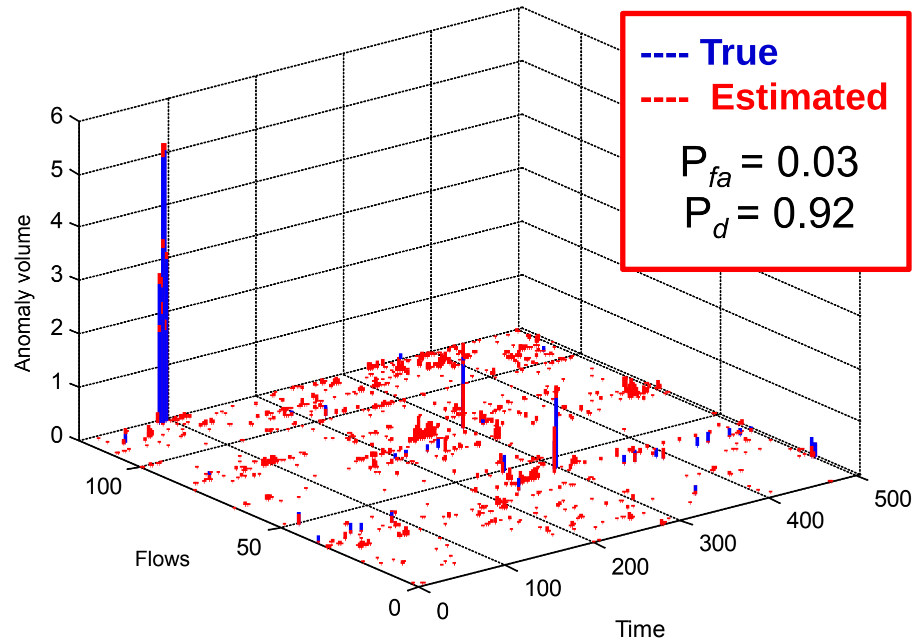




Internet2 data

- Real network data
  - Dec. 8-28, 2008
  - N=11, L=41, F=121, T=504





Data: http://www.cs.bu.edu/~crovella/links.html

Dynamic anomalography

- Construct an estimated map of anomalies in real time
- Streaming data model:

$$\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t) = \mathcal{P}_{\mathcal{S}_t}(\mathbf{x}_t + \mathbf{R}_t\mathbf{a}_t + \mathbf{v}_t), \;\; t = 1, 2, \ldots \qquad \mathbf{x}_t := \mathbf{R}_t\mathbf{z}_t$$

Goal: Given $\{\mathcal{P}_{\mathcal{S}_i}(\mathbf{y}_i), \mathbf{R}_i\}_{i=1}^{t}$ estimate $(\mathbf{x}_t, \mathbf{a}_t)$ online when $\{\mathbf{x}_t\}$ is in a low-dimensional space and $\{\mathbf{a}_t\}$ is sparse

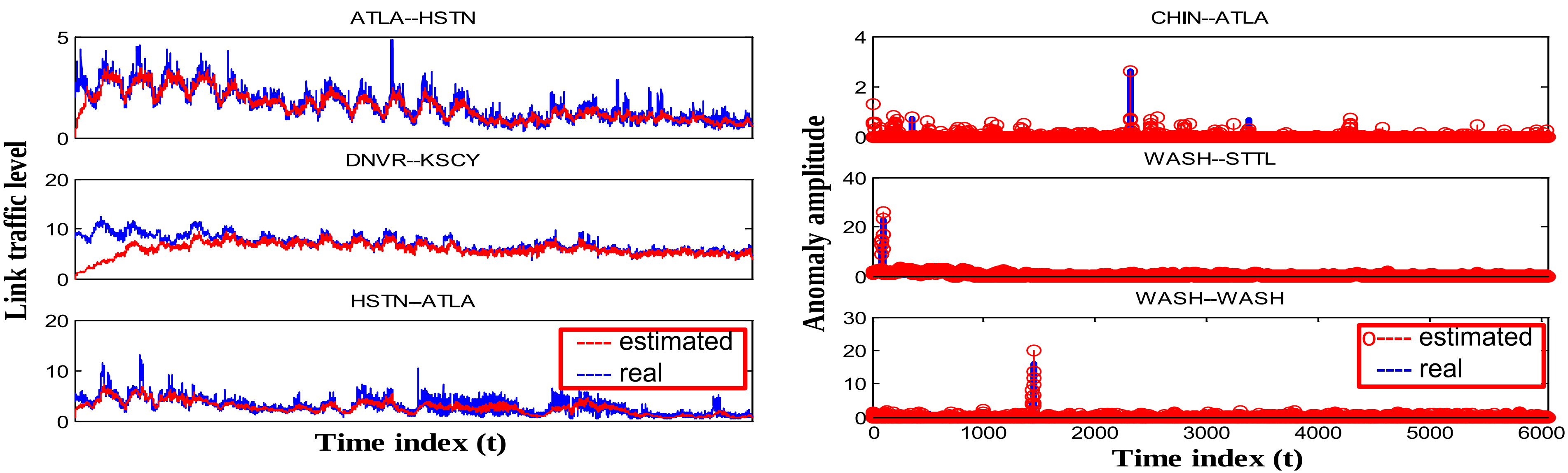
- (Robust) subspace tracking
  - Projection approximation (PAST) [Yang'95]
  - Missing data: GROUSE [Balzano et al'10], PETRELS [Chi et al'12]
  - Outliers: [Mateos-Giannakis'10], GRASTA [He et al'11]
- Compressed "outliers" challenge identifiability

M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE Journal of Selected Topics in Signal Processing*, pp. 50-66, Feb. 2013.

Online estimator

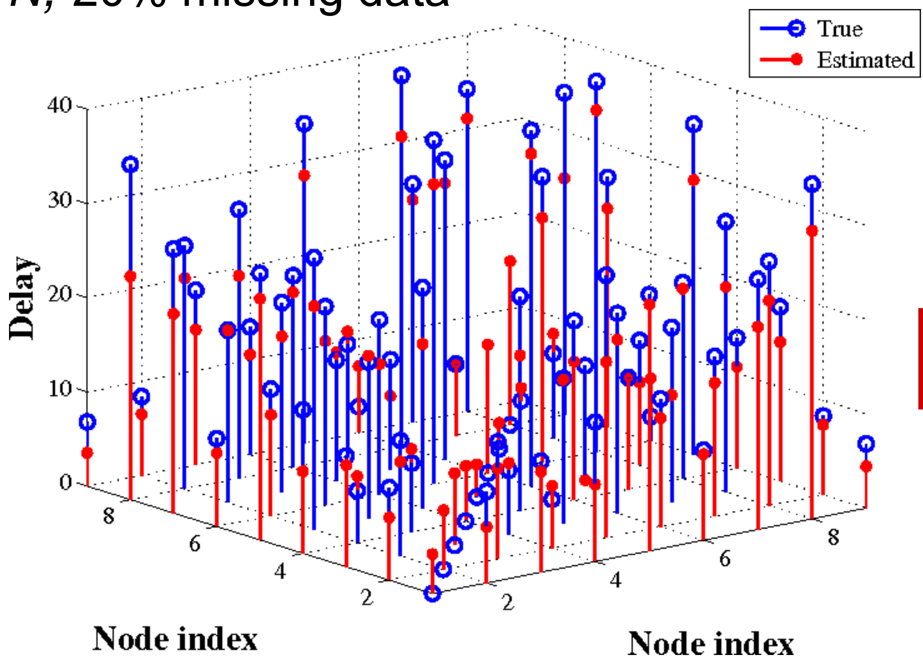
- Challenge: $\|\cdot\|_*$ not separable across columns (time) $\Rightarrow \mathbf{x}_t = \mathbf{C}\mathbf{w}_t$
- Approach: regularized exponentially-weighted LS formulation

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{A}\}} \sum_{\tau=1}^{t} \beta^{t-\tau}\left[\frac{1}{2}\|\mathcal{P}_{\mathcal{S}_\tau}(\mathbf{y}_\tau - \mathbf{C}\mathbf{w}_\tau - \mathbf{R}_\tau\mathbf{a}_\tau)\|_2^2 + \frac{\lambda_*}{2\sum_{u=1}^{t}\beta^{t-u}}\|\mathbf{C}\|_F^2 + \frac{\lambda_*}{2}\|\mathbf{w}_\tau\|_2^2 + \lambda_1\|\mathbf{a}_\tau\|_1\right]$$



Delay cartography

- Network distance prediction [Liau et al'12]
- Approach: distributed low-rank matrix completion
- Internet2 data (Aug 18-22,2011)
  - End-to-end latency matrix
  - *N*=9, *L*=*T*=*N;* 20% missing data



Relative error: 10%

Data: http://internet2.edu/observatory/archive/data-collections.html

Takeaways

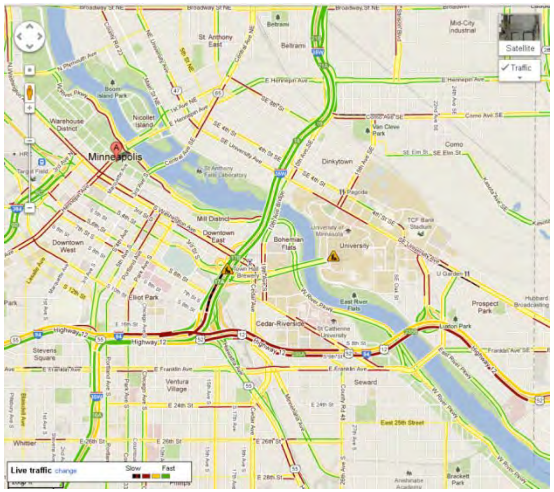
- Unveiling network traffic anomalies via convex optimization
  - Leveraging sparsity and low rank
- Reveal when and where anomalies occur
- Exact recovery of low-rank plus compressed sparse matrices
- Distributed/online algorithms with guaranteed performance

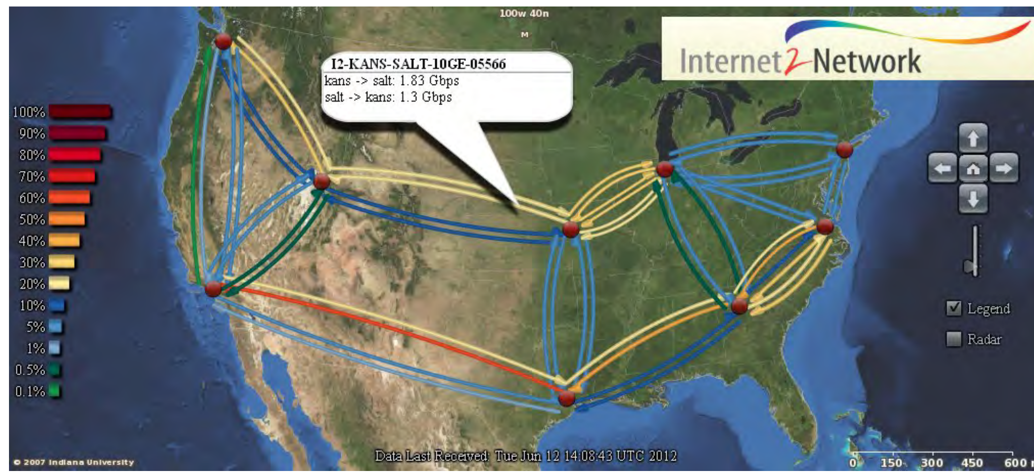
Roadmap

- Dynamic network delay cartography
- Unveiling network anomalies via sparsity and low rank
- Network-wide link count prediction
  - Semi-supervised learning for traffic maps
  - Batch and online processing
  - Empirical validation: Internet2 data
- RF cartography for cognition at the PHY
- Conclusions and future research directions

A commuting conundrum

- Objective: map a "good" route for packet delivery
  - Measure traffic at few roads/links only



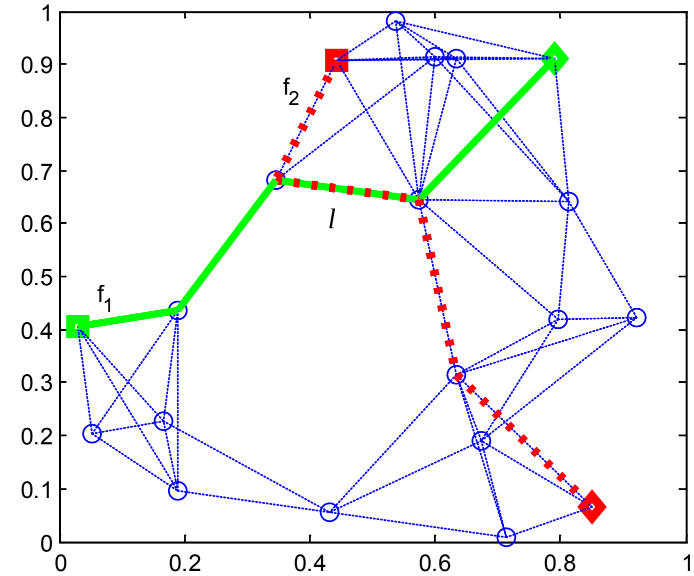


- Application domains
  - Transportation networks [Gastner-Newman'04]
  - Communication networks [Soule et al'05]
  - Sensor networks [Abrams et al'04]

Model

- Graph $G(N, L)$ with N nodes, L links, and F flows ($F >> L$)

(as) Single-path per OD flow $z_{f,t}$



- Packet counts per link l and time slot t

$$y_{l,t} = \sum_{f=1}^{F} r_{l,f} x_{f,t} + v_{l,t}$$

$r_{l,f} \in (0,1)$

- Incomplete, noisy measurements on a subset of links $l \in \mathcal{S}_t$

$$\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t) = \mathcal{P}_{\mathcal{S}_t}(\mathbf{R}\mathbf{x}_t + \mathbf{v}_t)$$

$$[\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t)]_l = \begin{cases} y_{l,t}, & l \in \mathcal{S}_t \\ 0, & l \notin \mathcal{S}_t \end{cases}$$

Problem statement

Goal: Given $\mathcal{P}_{\mathcal{S}_{t'}}(\mathbf{y}_{t'})$, $\mathbf{R}$ and historical data $\{\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t)\}_{t=1}^{T}$, find $\hat{\mathbf{y}}_t, t' > T$

- **Prior art**
  - Traffic estimation $\hat{\mathbf{y}}_{t'} = \mathbf{R}\hat{\mathbf{x}}_{t'}$ [Zhang et al'05]
  - Kriging [Chua et al'06], plus traffic modeling [Vaughn et al'10]
  - Topology-driven basis expansion [Crovella-Kolaczyk'03], [Coates et al'07]

- **Impact**
  - Ability to handle missing data
  - Online prediction capturing spatio-temporal correlations
  - Computationally-efficient link traffic prediction

P. A. Forero, K. Rajawat, and G. B. Giannakis, "Prediction of partially observed dynamical processes over networks via dictionary learning," *J. Machine Learning Research*, 2013.

Data-driven model of link counts

- Sparse representation of link counts $\mathbf{y}_t = \mathbf{D}\mathbf{s}_t$

  $L \times Q,\ (L \leq Q)$

- Notation: $\mathcal{D} := \{\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_Q] \in \mathbb{R}^{L \times Q} : \|\mathbf{d}_q\| \leq 1,\ \forall q\}$

  $\mathbf{S} := [\mathbf{s}_1, \ldots, \mathbf{s}_T]$

Dictionary Learning (DL) [Olshausen-Field'97]

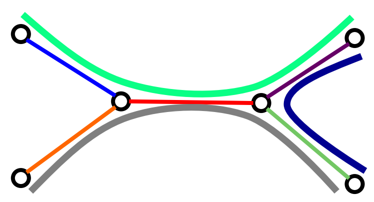
Given $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]$, find dictionary (basis) $\mathbf{D}$ and sparse $\mathbf{S}$

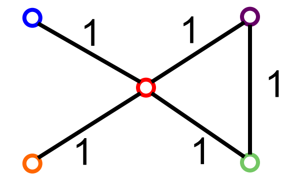
$$(\hat{\mathbf{S}}, \hat{\mathbf{D}}) = \arg \min_{\mathbf{S}, \mathbf{D} \in \mathcal{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_1$$

- **Q:** How about DL from incomplete data $\{\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t)\}_{t=1}^{T}$?

Capturing spatial link dependence

- Auxiliary graph $\mathcal{G}$ with vertices = links in *G*
  - Edge weights $w_{l,l'}$ = number of OD flows common to links l, l'
  - Adjacency matrix: $\mathbf{W} = \mathbf{R}\mathbf{R}'$, graph Laplacian $\mathbf{L} = \mathrm{diag}(\mathbf{W}\mathbf{1}_L) - \mathbf{W}$





- Cost function to learn *D*

$$C_t(\mathbf{D}, \mathbf{s}_t) := \|\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t - \mathbf{D}\mathbf{s}_t)\|_2^2 + \lambda_1\|\mathbf{s}_t\|_1 + \lambda_2 \mathbf{s}_t'\mathbf{D}'\mathbf{L}\mathbf{D}\mathbf{s}_t$$

- Regularizers effect sparsity and *smoothness* over $\mathcal{G}$

$$\mathbf{s}_t'\mathbf{D}'\mathbf{L}\mathbf{D}\mathbf{s}_t = \frac{1}{2}\sum_{l=1}^{L}\sum_{l'=1}^{L} w_{l,l'}(x_{l,t} - x_{l',t})^2$$

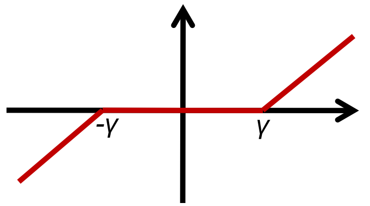
Semi-supervised DL

Semi-supervised Dictionary Learning (SSDL)

Given $\{\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t)\}_{t=1}^{T}$, $\mathbf{R}$, find dictionary (basis) $\mathbf{D}$ and sparse $\mathbf{S}$

$$(\hat{\mathbf{S}}, \hat{\mathbf{D}}) = \arg \min_{\mathbf{S},\mathbf{D}\in\mathcal{D}} \sum_{t=1}^{T} C_t(\mathbf{D}, \mathbf{s}_t)$$

- SSDL biconvex, block-coordinate descent (BCD) solver
  - Update $\{\mathbf{s}_t\}_{t=1}^{T}$ via parallel entry-wise soft-thresholding
  - Update each $\mathbf{D}$ via QP + projection onto the Euclidean ball



Proposition: BCD's iterates converge to a stationary point of SSDL

Link load prediction

- Given $\mathcal{P}_{\mathcal{S}_{t'}}(\mathbf{y}_{t'})$ and learnt dictionary $\hat{\mathbf{D}}$, solve

$$\hat{\mathbf{s}}_{t'} := \arg\min_{\mathbf{s}} \|\mathcal{P}_{\mathcal{S}_{t'}}(\mathbf{y}_{t'} - \hat{\mathbf{D}}\mathbf{s})\|_2^2 + \lambda_1\|\mathbf{s}\|_1 + \lambda_2\mathbf{s}'\hat{\mathbf{D}}'\mathbf{L}\hat{\mathbf{D}}\mathbf{s}$$

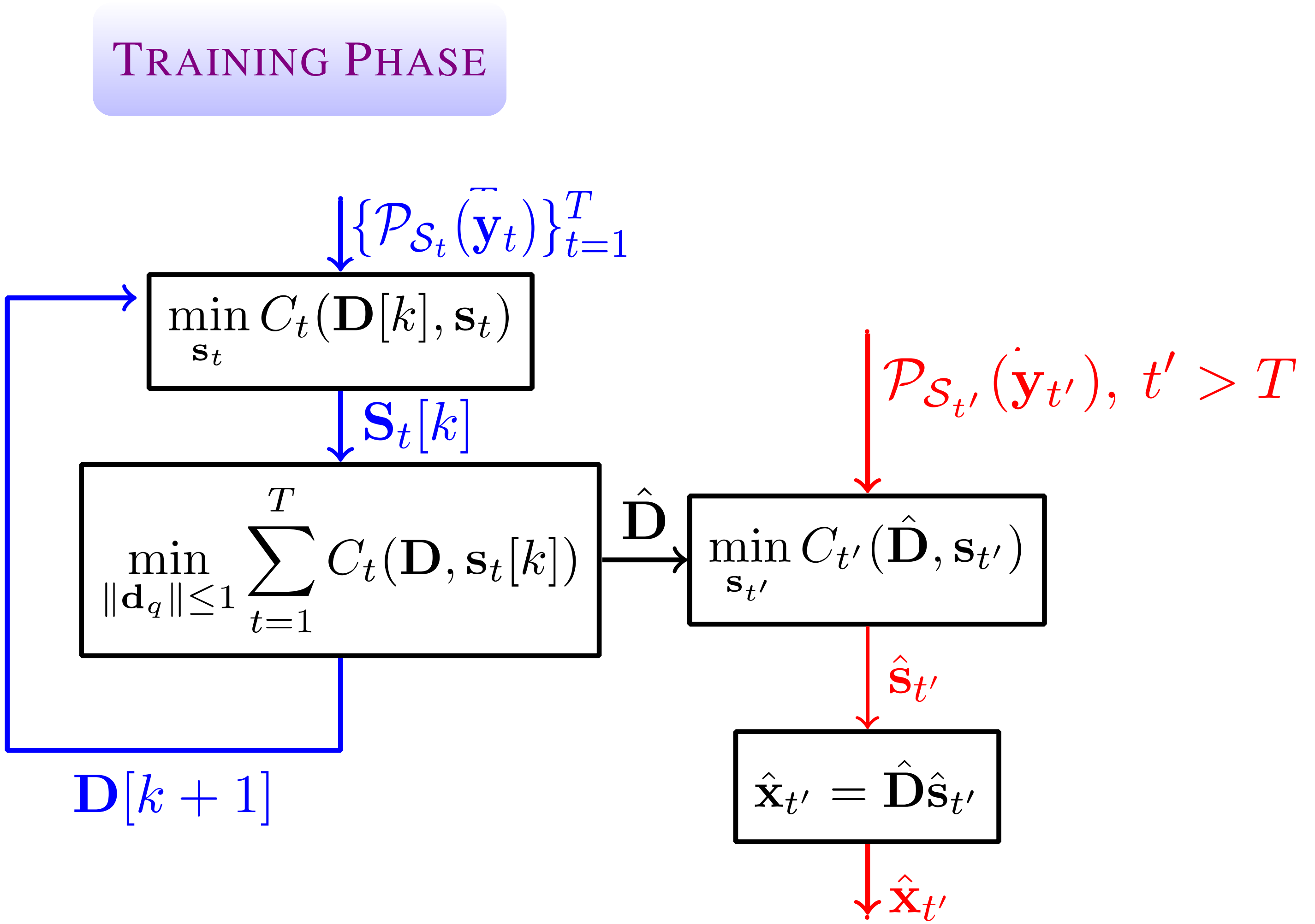
  - Captures sparsity of $\mathbf{s}_{t'}$ and smoothness of link loads over $\mathcal{G}$

- Predict $\mathbf{y}_{t'}$ based on $\hat{\mathbf{s}}_{t'}$

$$\boxed{\hat{\mathbf{y}}_{t'} = \hat{\mathbf{D}}\hat{\mathbf{s}}_{t'}}$$

  - Scaling factor $(1+\lambda_2)$ reduces bias in $\hat{\mathbf{y}}_{t'}$ [Zou-Hastie'05]

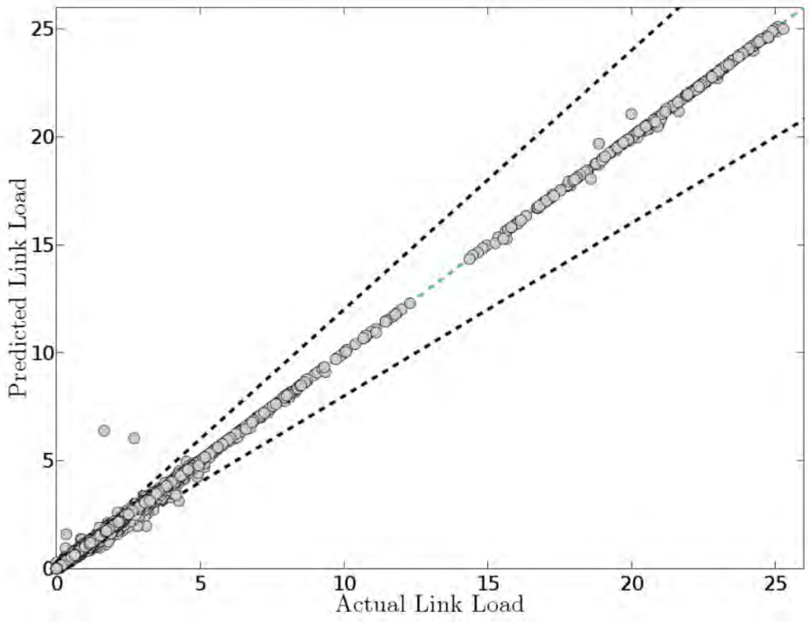
Batch processing summary



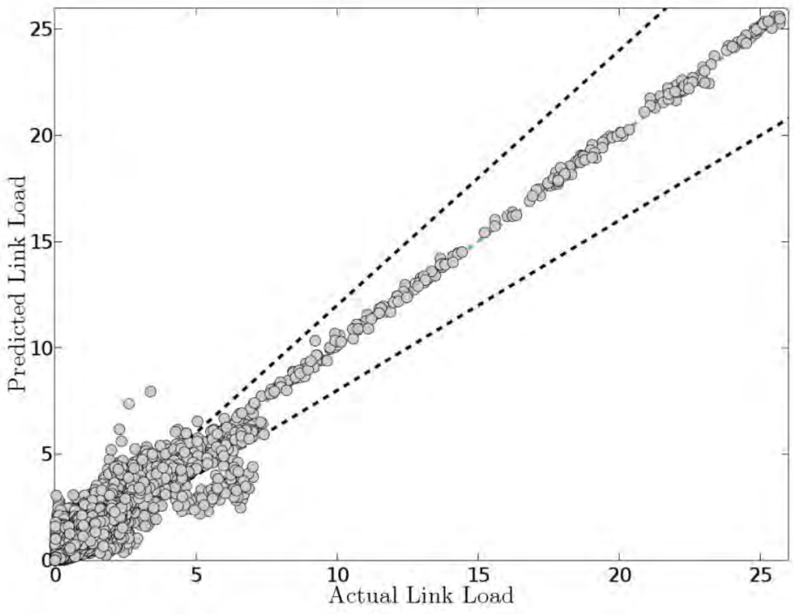
Test case: Internet2

- Internet2 measurement archive
  - L=54, *T=2000*

Training phase – 30 links measured



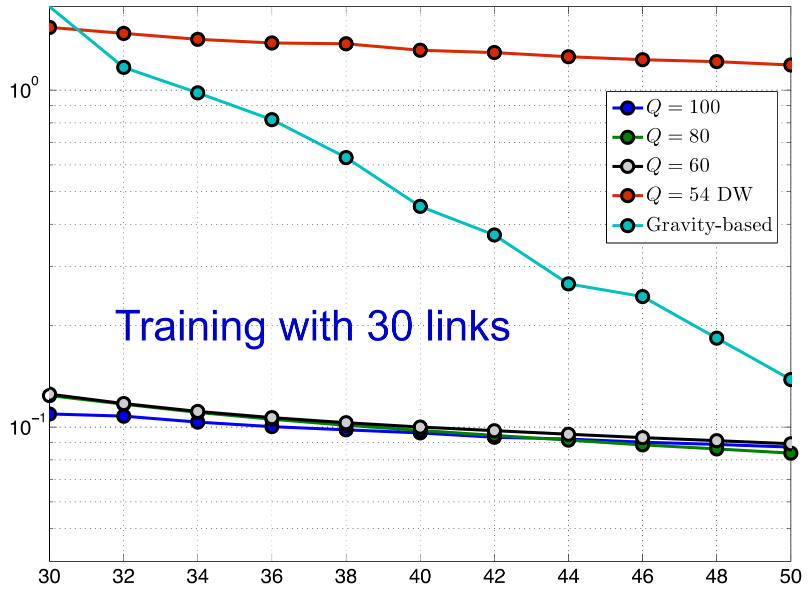
Operational phase – 30 links measured

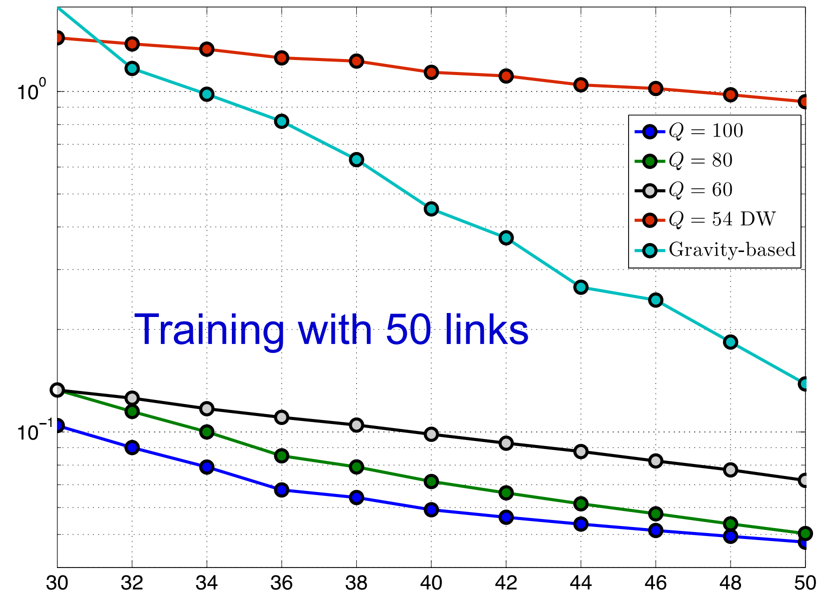


- Prediction improves as link load increases

Prediction error (Internet2)

- Normalized prediction error: $\text{NPE} := \frac{1}{Lt_0}\sum_{\tau=1}^{t_0}\|\mathbf{y}_\tau - \hat{\mathbf{y}}_\tau\|_2^2$
  - *Q* = number of columns of *D*; t_0=*2000*
- Gravity-based [Zhang et al'05]; Diffusion wavelets [Coifman-Maggioni'07]





- SSDL outperforms competing alternatives

Online processing

- Capture temporal correlations on $\{\mathbf{s}_\tau\}$

$$C_\beta^t(\mathbf{D}_t, \mathbf{s}) := \sum_{\tau=1}^{t} \beta^{t-\tau} \|\mathcal{P}_{\mathcal{S}_\tau}(\mathbf{y}_\tau - \mathbf{D}_t\mathbf{s})\|_2^2 + \lambda_1\|\mathbf{s}\|_1 + \lambda_2\mathbf{s}'\mathbf{D}_t'\mathbf{L}\mathbf{D}_t\mathbf{s}$$

- Given $\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t)$ and dictionary $\mathbf{D}_t$, solve $\mathbf{s}_t := \arg\min_{\mathbf{s}} C_\beta^t(\mathbf{D}_t, \mathbf{s})$

- Predict $\mathbf{y}_t$ based on $\mathbf{s}_t$

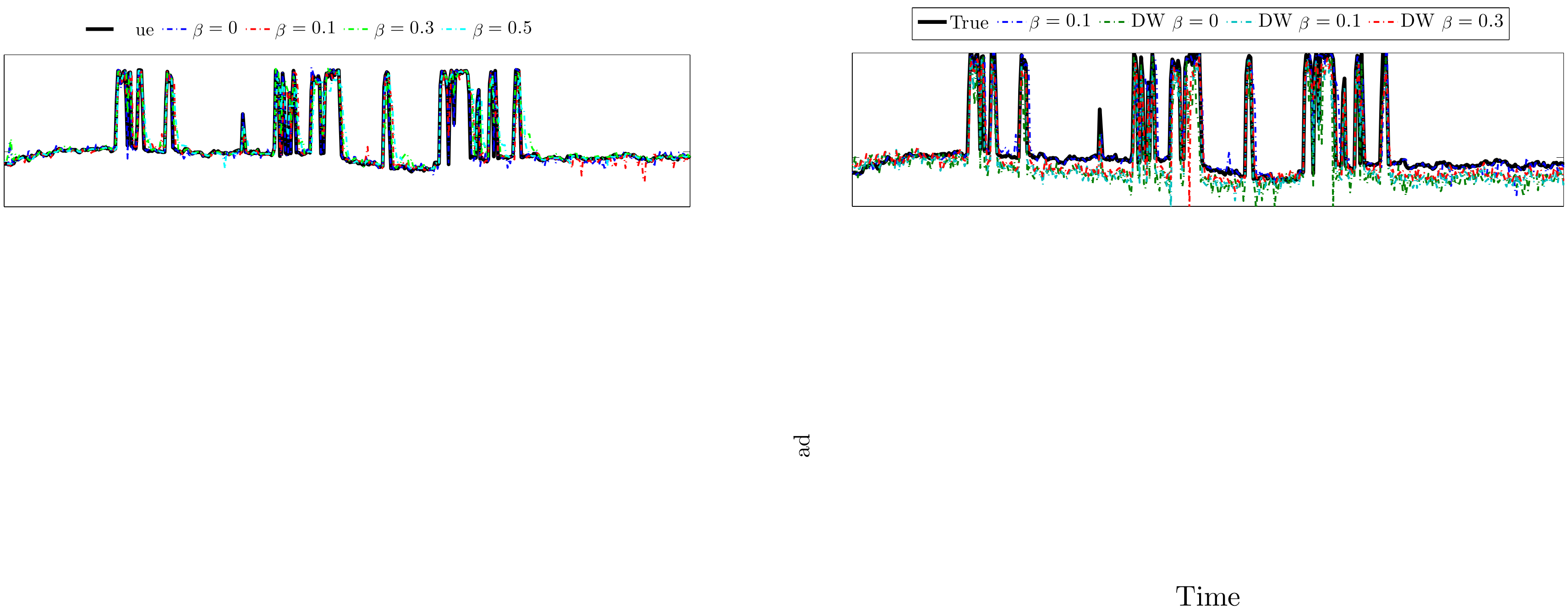
$$\boxed{\hat{\mathbf{y}}_t = (1 + \lambda_2)\mathbf{D}_t\mathbf{s}_t}$$

- Dictionary update

$$\mathbf{D}_{t+1} = \arg\min_{\mathbf{D} \in \mathcal{D}} \frac{1}{t} \sum_{\tau=1}^{t} C_\beta^\tau(\mathbf{D}, \mathbf{s}_\tau)$$

Real-time prediction (Internet2)

- *Q=60*, different values of the forgetting factor β
  - Measure traffic at *30* links only



- SSDL-based tracker outperforms diffusion wavelets

Takeaways

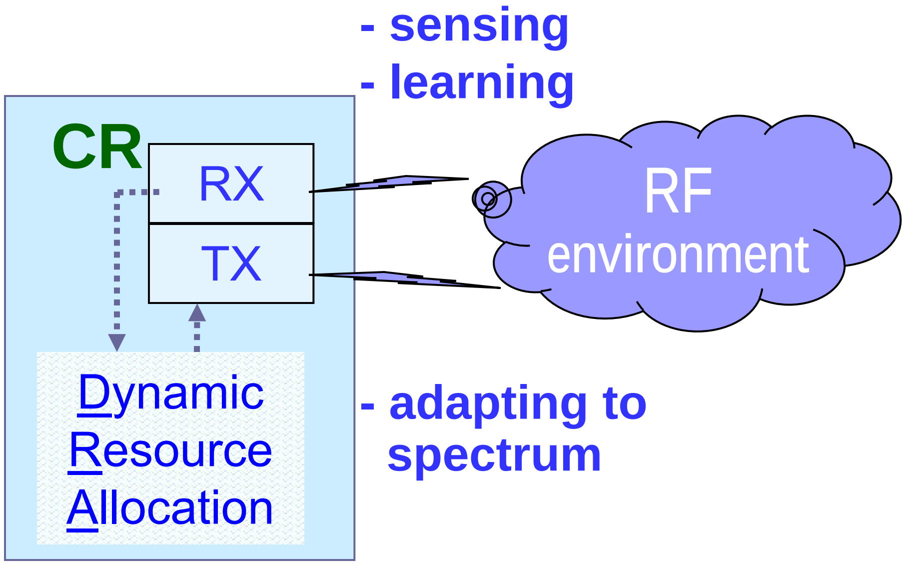
- Prediction of network processes from incomplete observations
  - Link count prediction based on dictionary learning
- Spatial correlation of link counts via Laplacian regularization
  - Semi-supervised learning
- Online algorithms capturing temporal correlations

Roadmap

- Dynamic network delay cartography
- Unveiling network anomalies via sparsity and low rank
- Network-wide link count prediction
- RF cartography for cognition at the PHY
  - Interference spectrum cartography
  - Channel gain cartography
- Conclusions and future research directions

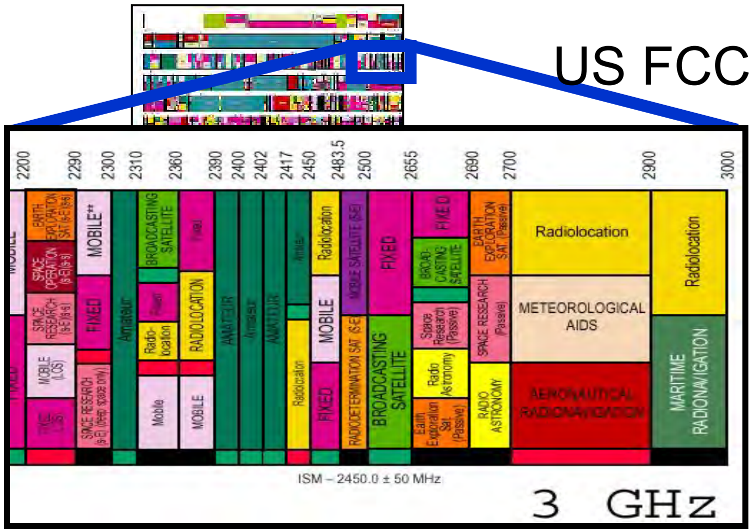
What is a cognitive radio?

- Fixed radio
  - *Policy-based*: parameters set by operators
- Software-defined radio (SDR)
  - *Programmable*: can adjust parameters to intended link
- Cognitive radio (CR)
  - *Intelligent: sense* the environment & *learn* to *adapt* [Mitola'00]



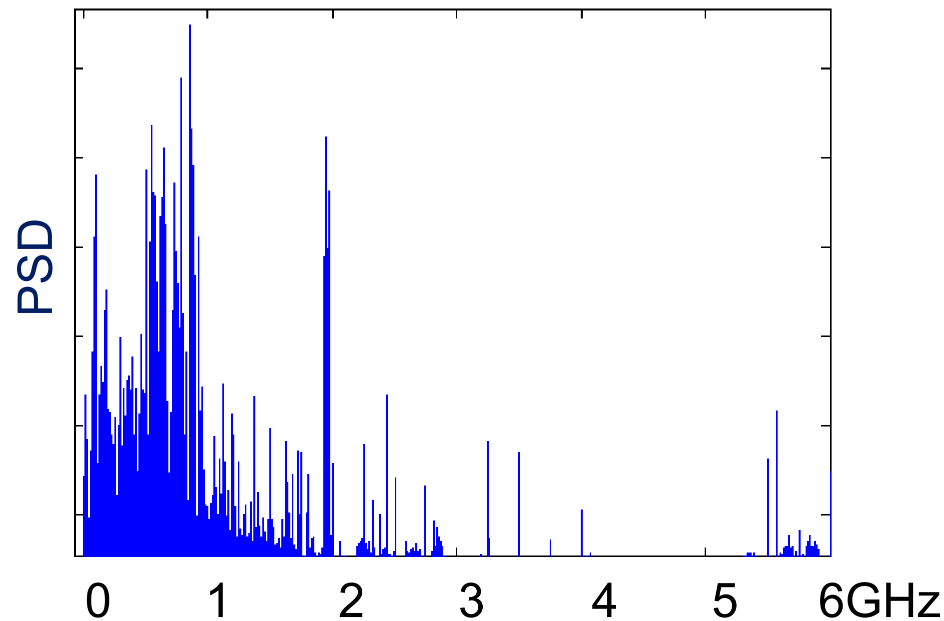
- *Cognizant transceiver:* sensing
- *Agile transmitter:* adaptation
- *Intelligent DRA:* decision making
  - Radio reconfiguration decisions
  - Spectrum access decisions

Spectrum scarcity problem



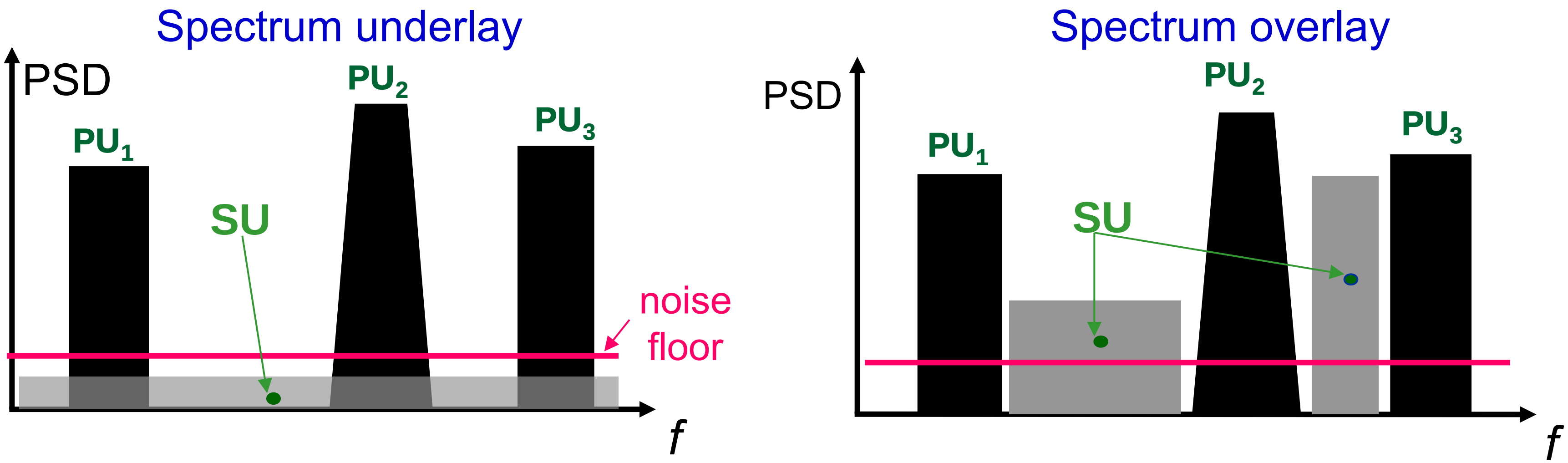
- Fixed spectrum access policies
  - Useful radio spectrum **pre-assigned**

Inefficient occupancy



Dynamical access under user hierarchy

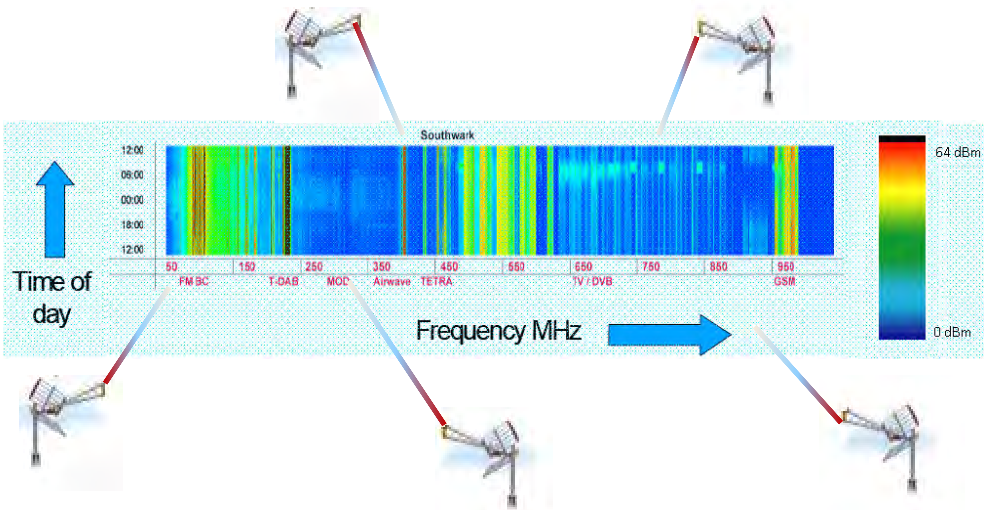
- Primary users (PUs) versus secondary users (SUs/CRs)



- Spectrum underlay
  - Restriction on transmit power levels
  - Operation over ultra wide bandwidths
- Spectrum overlay
  - Constraints on when and where to transmit
  - Avoid interference to Pus via sensing and adaptive allocation

Cooperative sensing for efficient sharing

- Multiple CRs jointly detect the spectrum [Ganesan-Li'06][Ghasemi-Sousa'07]



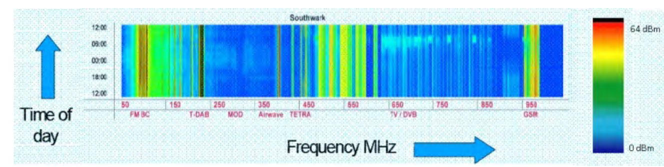
Source: Office of Communications (UK)

- Benefits of cooperation
  - Spatial diversity gain mitigates multipath fading/shadowing
  - Reduced sensing time and local processing
  - Ability to cope with hidden terminal problem
- Limitation: existing approaches do not exploit space-time dimensions

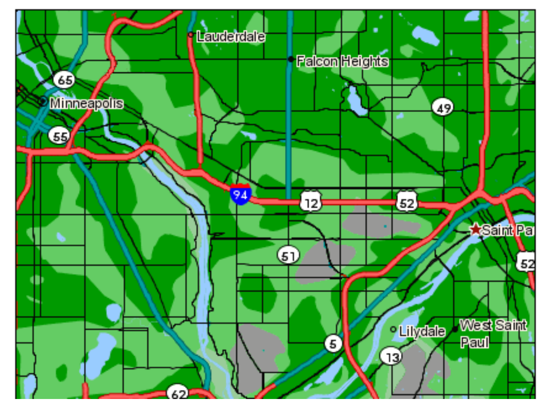
Cooperative PSD cartography

- **Idea:** CRs collaborate to form a spatial map of the RF spectrum

Goal: Find PSD map $\Phi(\boldsymbol{x}, f)$ across space $\boldsymbol{x} \in \mathbb{R}^2$ and frequency $f \in \mathbb{R}$



- **Specifications:** coarse approx. suffices
- **Approach:** basis expansion of $\Phi(\boldsymbol{x}, f)$



J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Transactions on Signal Processing,* pp. 1847-1862, March 2010.

Modeling

- Transitters

  $\text{Tx}_s,\ \ s = 1, \ldots, N_s$

- Sensing CRs

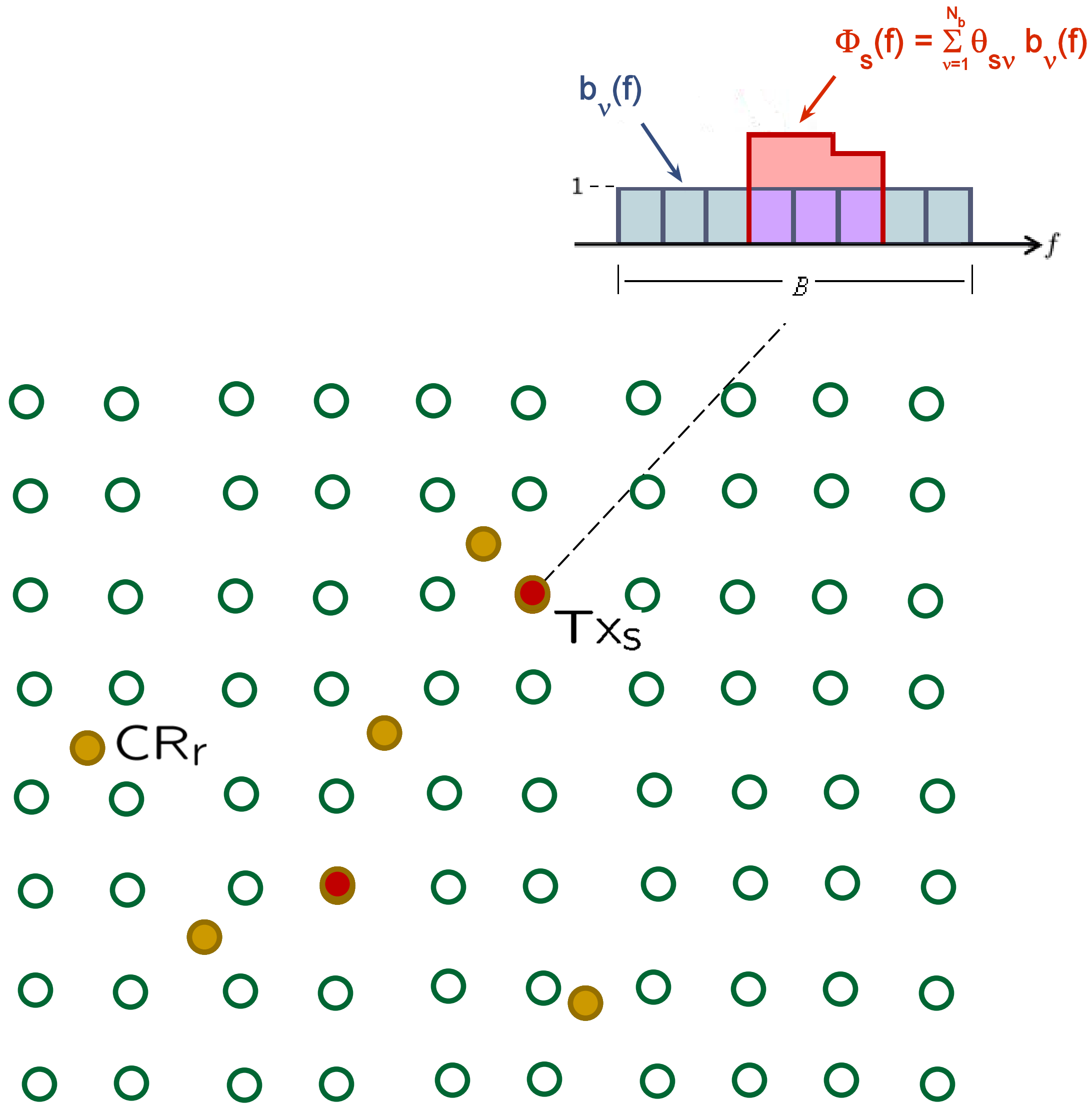
  $\text{CR}_r,\ \ r = 1 : N_r$

- Frequency bases

  $b_\nu(f),\ \ \nu = 1 : N_b$

- Sensed frequencies

  $f_k,\ \ k = 1 : K$

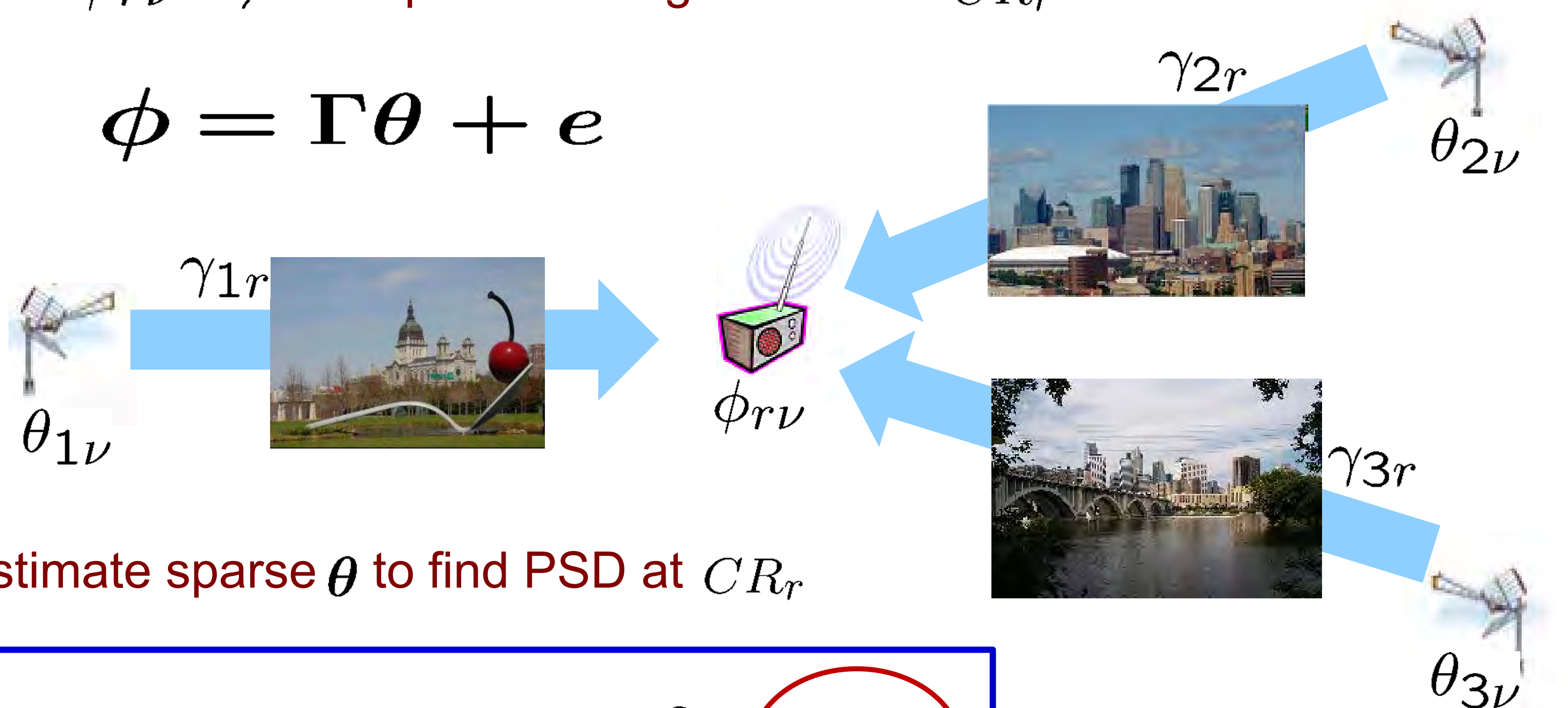


➤ Sparsity present in space and frequency

Space-frequency basis expansion

- Find $\theta_{s\nu} \Rightarrow$ Tx-power of source *s* over frequency band ν
- Data $\phi_{r\nu} \Rightarrow$ Rx-power at cognitive radio CR_r

$$\boldsymbol{\phi} = \boldsymbol{\Gamma}\boldsymbol{\theta} + \boldsymbol{e}$$

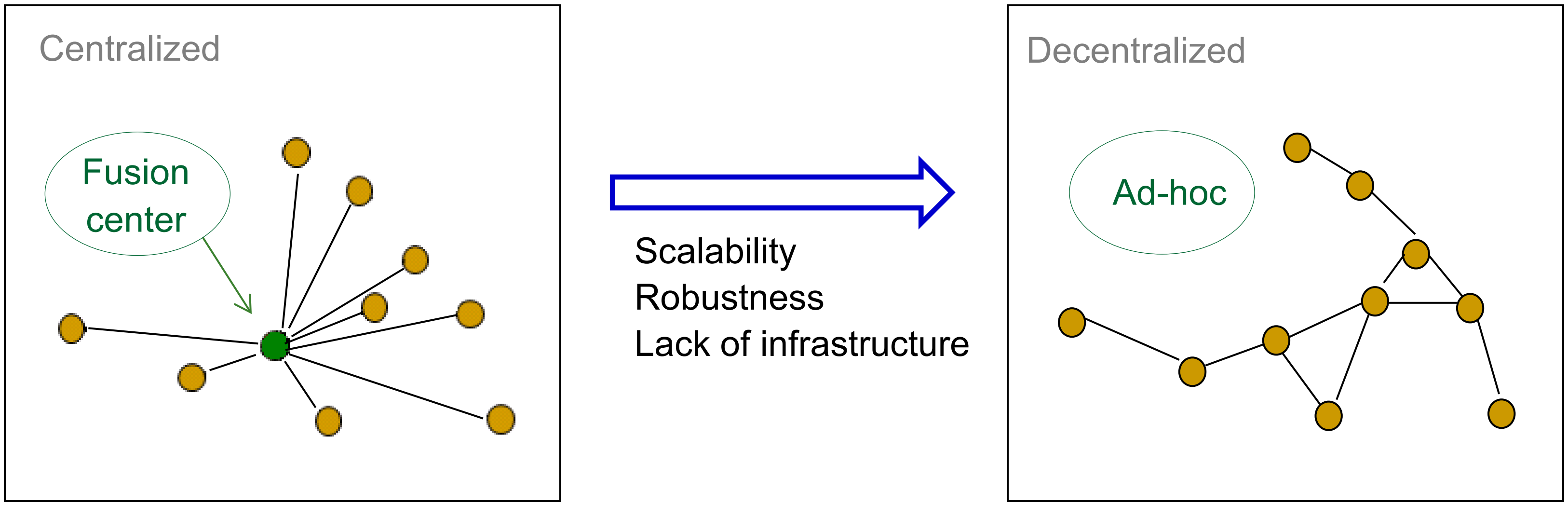


- Estimate sparse $\boldsymbol{\theta}$ to find PSD at CR_r

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; ||\boldsymbol{\phi} - \boldsymbol{\Gamma}\boldsymbol{\theta}||_2^2 + \lambda||\boldsymbol{\theta}||_1$$

Sparsity-promoting regularization

Distributed recursive implementation



- Consensus-based approach
  - Solve locally

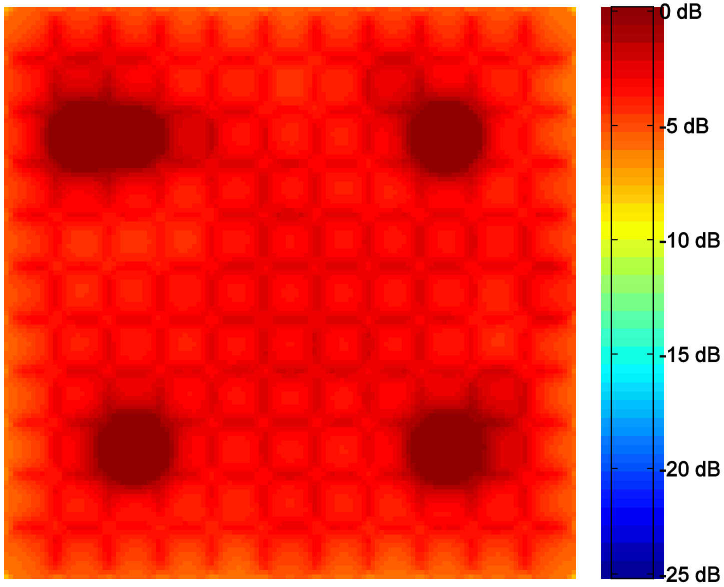
$$\hat{\theta} = \arg\min_{\boldsymbol{\theta}_r} \; ||\phi_r - \boldsymbol{\Gamma}_r \boldsymbol{\theta}_r||_2^2 + \frac{\lambda}{M} ||\boldsymbol{\theta}_r||_1$$
$$\text{s.to} \quad \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r'}, \quad \forall r' \in \mathcal{N}_r$$

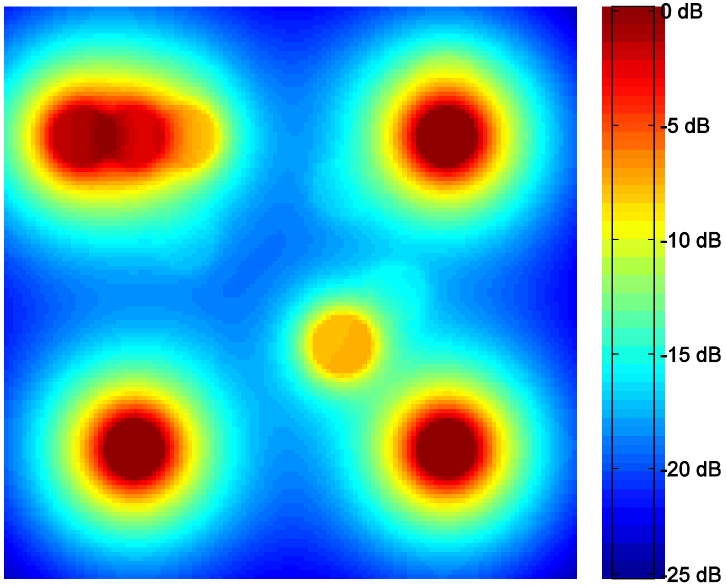
- Constrained optimization using ADMM

Exchange of local $\boldsymbol{\theta}_r$ estimates

RF spectrum cartography

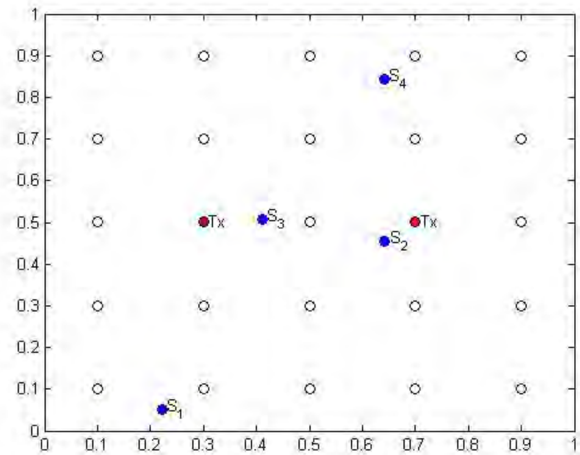
- 5 sources
- $N_s = 121$ candidate locations, $N_r = 50$ CRs



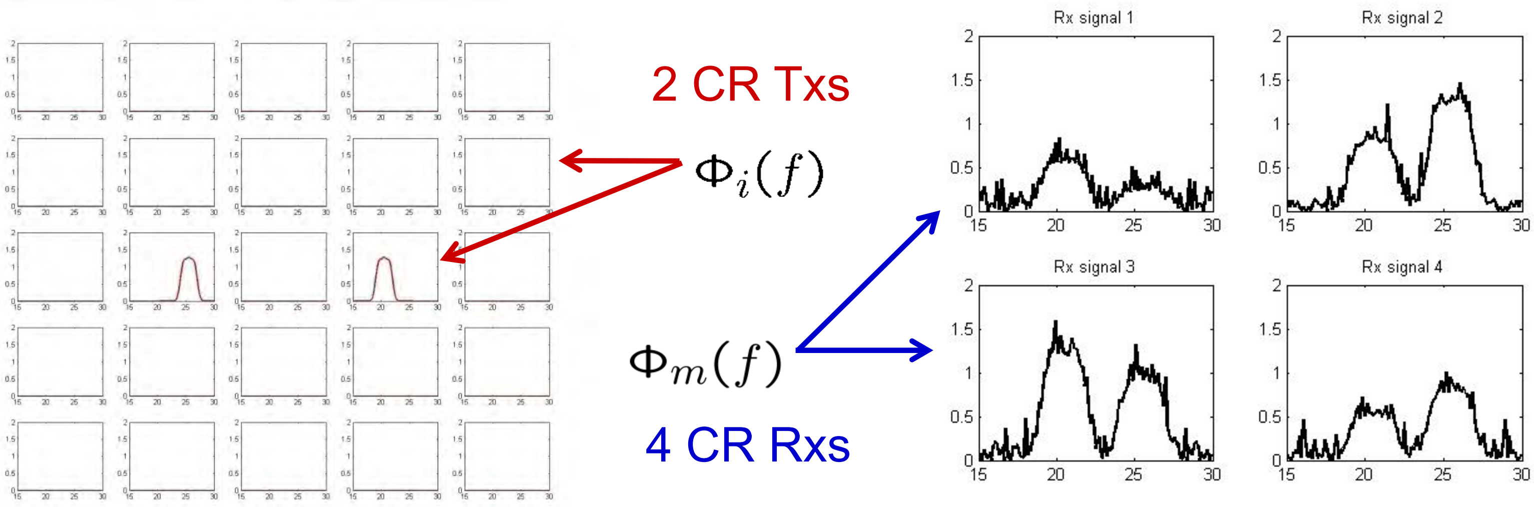


- As a byproduct, Lasso localizes all sources via variable selection

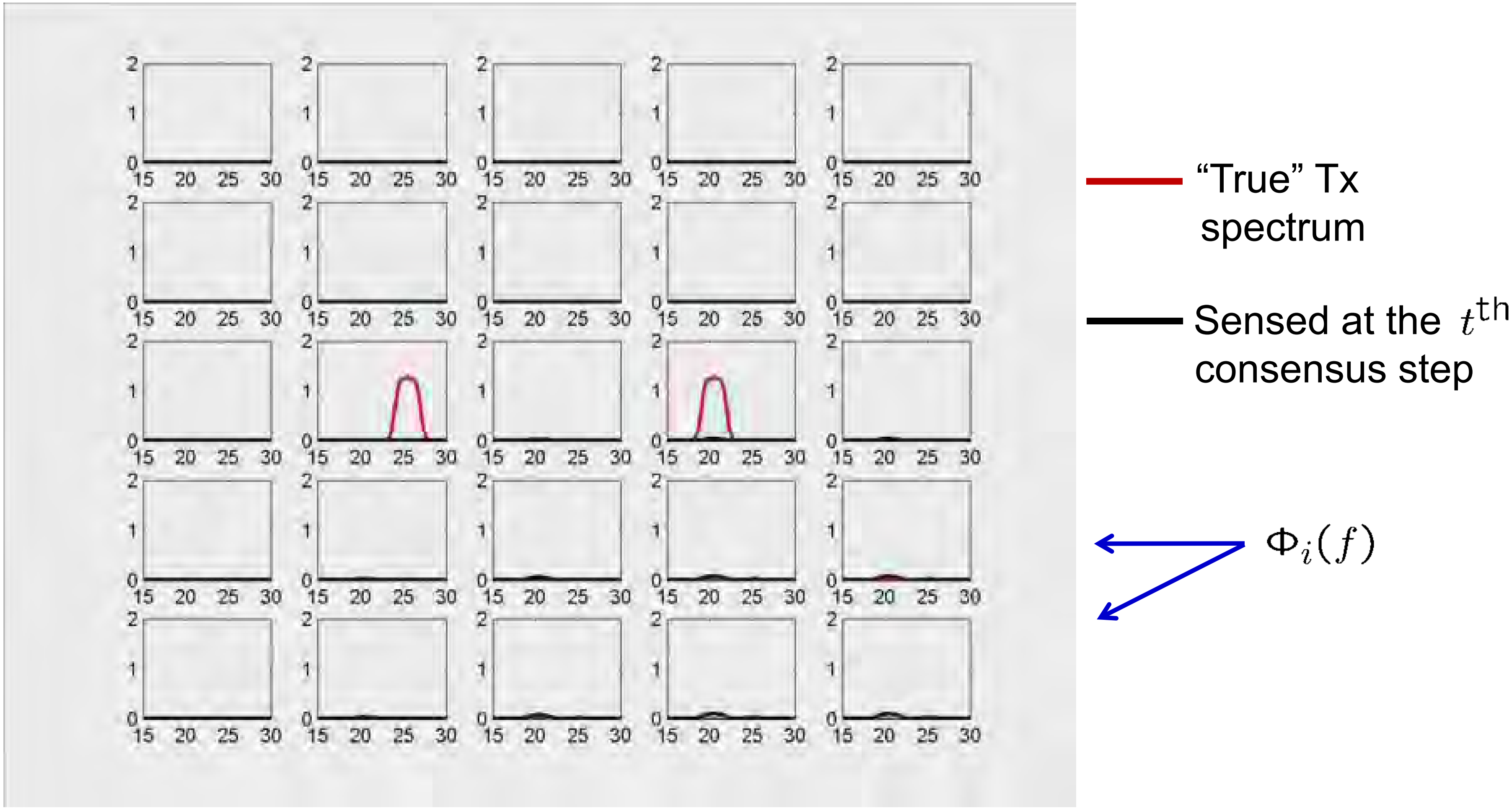
Simulated test: PSD map estimation



- Centralized sensing
- No fading
- N_s=25



Distributed consensus with fading

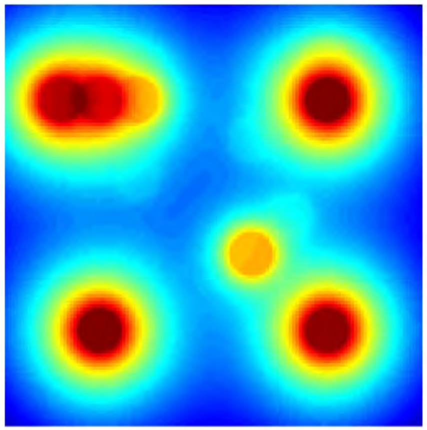


- Starting from a local estimate, sensors reach consensus

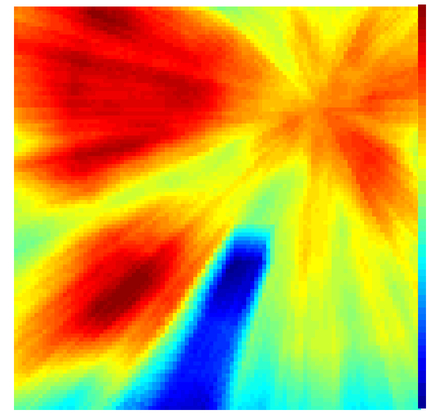
Spline-based PSD cartography

- **Q:** How about shadowing?

Path-loss



Shadowing



- **A:** Basis expansion with coefficient functions

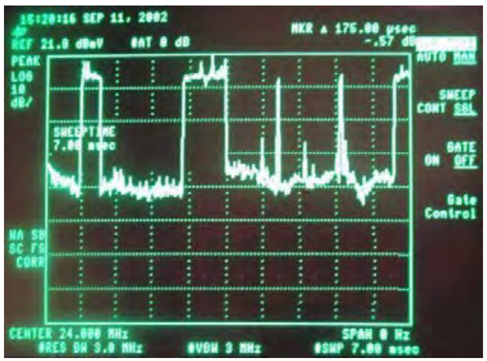
$$\Phi(\mathbf{x}, f) = \sum_{\nu=1}^{N_b} g_\nu(\mathbf{x}) b_\nu(f)$$

  - $g_\nu(\mathbf{x})$: unknown dependence on spatial variable ***x***

J. A. Bazerque, G. Mateos, and G. B. Giannakis, ``Group-Lasso on Splines for Spectrum Cartography,'' *IEEE Transactions on Signal Processing,"* pp. 4648-4663, October 2011.

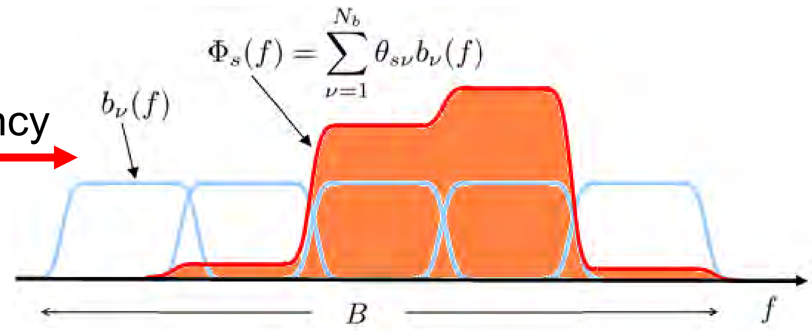
Frequency basis expansion

- PSD of Tx source $s \in \{1, \ldots, N_s\}$ is $\Phi_s(f)$



Basis expansion in frequency

$$\Phi_s(f) = \sum_{\nu=1}^{N_b} \theta_{s\nu} b_\nu(f)$$



- **Basis functions**
  - Accommodate prior knowledge $\Longrightarrow$ raised-cosine
  - Sharp transitions (regulatory masks) $\Longrightarrow$ rectangular, non-overlapping
  - Overcomplete basis set (large N_b) $\Longrightarrow$ robustness

Spatial PSD model

- Spatial loss function $l_s(\boldsymbol{x}) : \mathbb{R}^2 \rightarrow \mathbb{R} \Rightarrow$ Unknown

- BEM: $$\Phi(\boldsymbol{x}, f) = \sum_{\nu=1}^{N_b} g_\nu(\boldsymbol{x}) b_\nu(f)$$ $\left(g_\nu(\boldsymbol{x}) := \sum_{s=1}^{N_s} \theta_{s\nu} l_s(\boldsymbol{x})\right)$

  - Per sub-band factorization in space and frequency (indep. of N_s)

- Goal: estimate PSD atlas as $\hat{\Phi}(\boldsymbol{x}, f) = \sum_{\nu=1}^{N_b} \hat{g}_\nu(\boldsymbol{x}) b_\nu(f)$

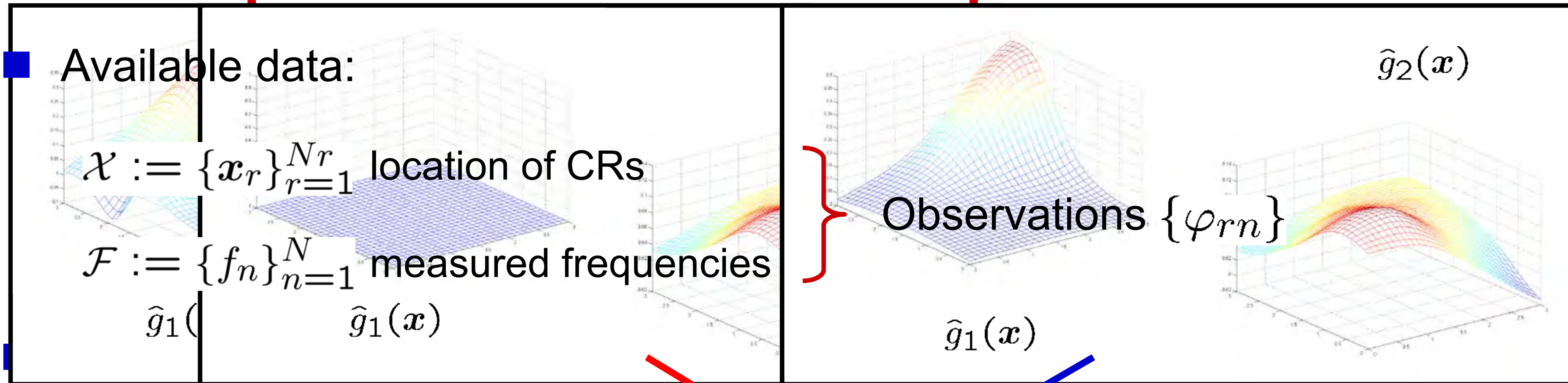
Nonparametric basis pursuit

- Available data:

$\mathcal{X} := \{\boldsymbol{x}_r\}_{r=1}^{N_r}$ location of CRs

$\mathcal{F} := \{f_n\}_{n=1}^{N}$ measured frequencies

} Observations $\{\varphi_{rn}\}$



$$\{\hat{g}_\nu\}_{\nu=1}^{N_b} = \arg \min_{\{g_\nu \in \mathcal{S}\}} \sum_{r=1}^{N_r} \sum_{n=1}^{N} \left(\varphi_{rn} - \sum_{\nu=1}^{N_b} g_\nu(\boldsymbol{x}_r) b_\nu(f_n) \right)^2 + \lambda \sum_{\nu=1}^{N_b} \int_{\mathbb{R}^2} ||\nabla^2 g_\nu(\boldsymbol{x})||_F^2 d\boldsymbol{x} + \mu \sum_{\nu=1}^{N_b} \|[g_\nu(\boldsymbol{x}_1), \ldots, g_\nu(\boldsymbol{x}_{N_r})]'\|_2 \quad \text{(P1)}$$

- Avoid overfitting by promoting smoothness
- Nonparametric basis selection ($\hat{g}_\nu \not\equiv 0 \Rightarrow b_\nu(f)$ selected)

Thin-plate splines solution

Proposition 1: Estimates $\{\hat{g}_\nu\}_{\nu=1}^{N_b}$ (P1) are thin-plate splines [Duchon' 77]

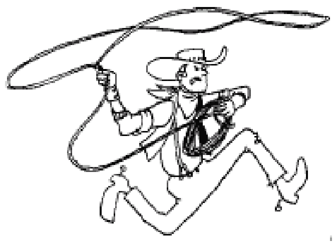
$$\hat{g}_\nu(\boldsymbol{x}) = \sum_{r=1}^{N_r} \hat{\beta}_{\nu r} K(||\boldsymbol{x} - \boldsymbol{x}_r||) + \hat{\boldsymbol{\alpha}}'_{\nu 1}\boldsymbol{x} + \hat{\alpha}_{\nu 0}$$

where $K(\rho)$ is the radial basis function $K(\rho) = \rho^2 \log(\rho)$, and

$$\hat{\boldsymbol{\beta}}_\nu := [\hat{\beta}_{\nu 1}, \ldots, \hat{\beta}_{\nu N_r}] \in \mathcal{B} := \left\{ \boldsymbol{\beta} : \sum_r \beta_r = 0, \sum_r \beta_r \boldsymbol{x}_r = \boldsymbol{0},\ \boldsymbol{x}_r \in \mathcal{X} \right\}.$$

- Unique, closed-form, finitely-parameterized minimizers!
- Q1: How to estimate $\{\boldsymbol{\alpha}_\nu, \boldsymbol{\beta}_\nu\}_{\nu=1}^{N_b}$ based on φ?
- Q2: How does (P1) perform basis selection?

Lassoing bases



- (P1) equivalent to group Lasso estimator [Yuan-Lin' 06]

  - Matrices ($\mathcal{X}$ and $\mathcal{F}$ dependent)

i) $\mathbf{T} := \begin{bmatrix} 1 & \boldsymbol{x}'_1 \\ \vdots & \vdots \\ 1 & \boldsymbol{x}'_{N_r} \end{bmatrix} = [\mathbf{Q}_1 \ \mathbf{Q}_2][\mathbf{R}' \ \mathbf{0}]'$, ii) $[\mathbf{K}]_{rl} := K(\|\boldsymbol{x}_r - \boldsymbol{x}_l\|)$, iii) $[\mathbf{B}]_{n\nu} := b_\nu(f_n)$

Proposition 2: Minimizers $\{\hat{\alpha}_\nu, \hat{\beta}_\nu\}_{\nu=1}^{N_b}$ (P1) are fully determined by

$$\hat{\zeta} := \arg\min_{\zeta} \|\mathbf{y} - \mathbf{X}\zeta\|_2^2 + \mu \sum_{\nu=1}^{N_b} \|\zeta_\nu\|_2 \quad \text{w/} \ \mathbf{y} := \begin{bmatrix} \boldsymbol{\varphi} \\ \mathbf{0} \end{bmatrix}, \ \mathbf{X} := \begin{bmatrix} \mathbf{B} \otimes \mathbf{I} \\ \mathbf{I} \otimes \mathbf{F}(\lambda, \mathbf{T}, \mathbf{K}, \mathbf{B}) \end{bmatrix}$$

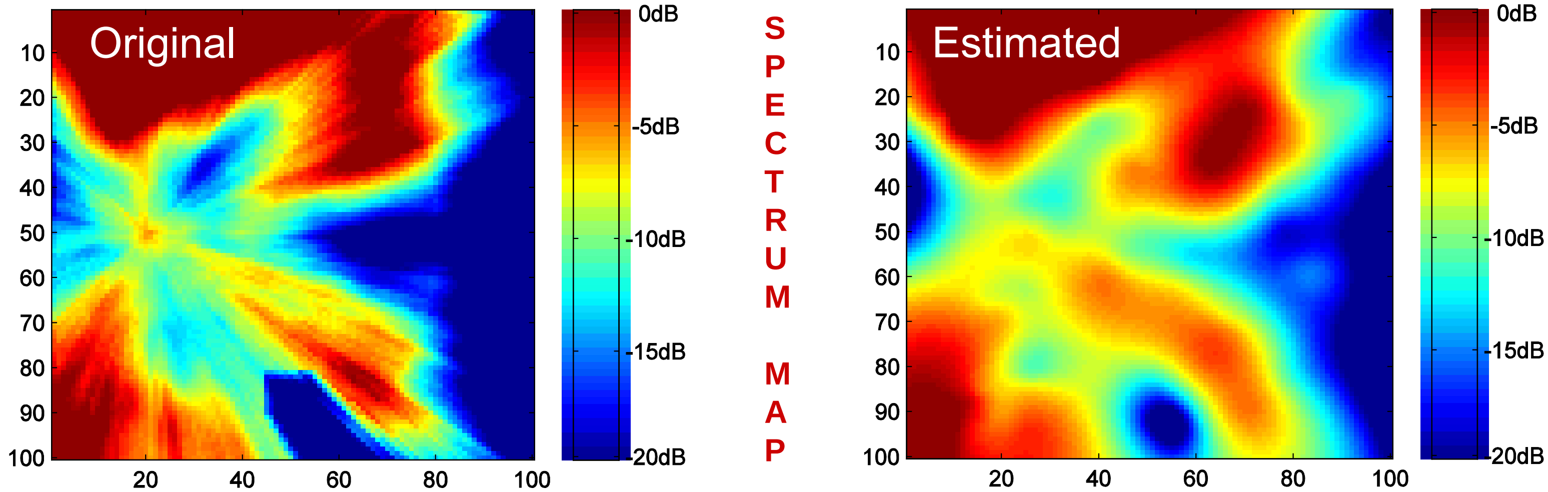
as $[\hat{\beta}'_\nu, \hat{\alpha}'_\nu]' = \mathrm{bdiag}(\mathbf{Q}_2, \mathbf{I})[\mathbf{K}\mathbf{Q}_2 \ \mathbf{T}]^{-1}\hat{\zeta}_\nu$.

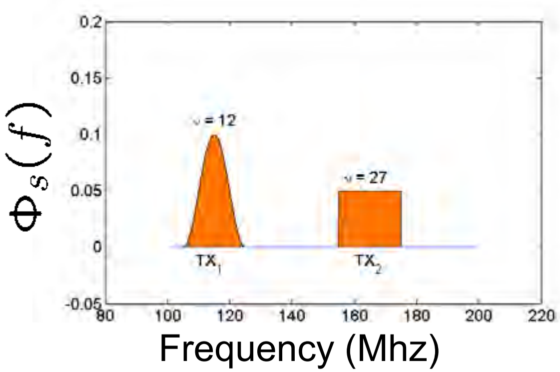
- Group Lasso encourages sparse factors $\hat{\zeta}_\nu$

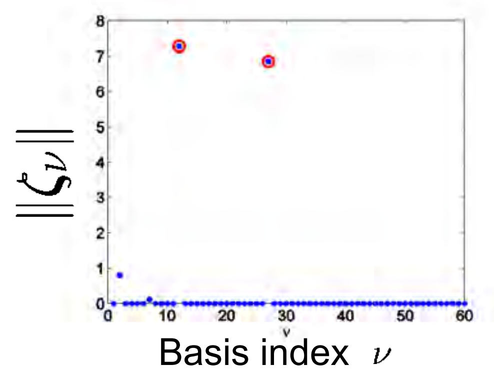
  - Full-rank mapping: $\hat{\zeta}_\nu = \mathbf{0} \Rightarrow \hat{g}_\nu(\boldsymbol{x}) \equiv 0$

Simulated test

- $N_s = 2$ sources; raised cosine pulses
- $N_r = 50$ sensing CRs, $N = 64$ sampling frequencies
- $N_b = (2 \times 15 \times 2) = 60$; (roll off x center frequency x bandwidth)

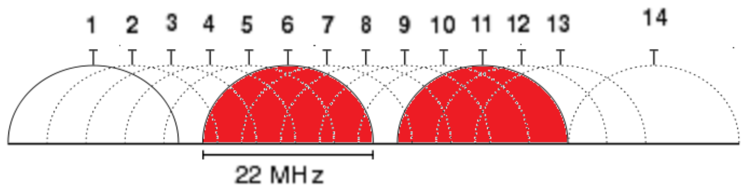






Numerical test IEEE 802.11

- $N_b = 14$

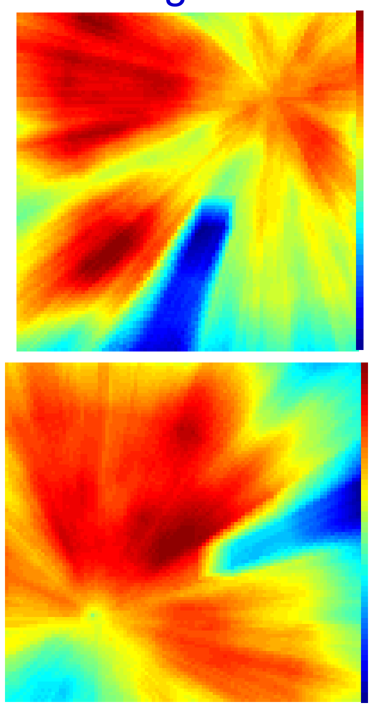


PUs $N_s = 2$

Original

Channel 6

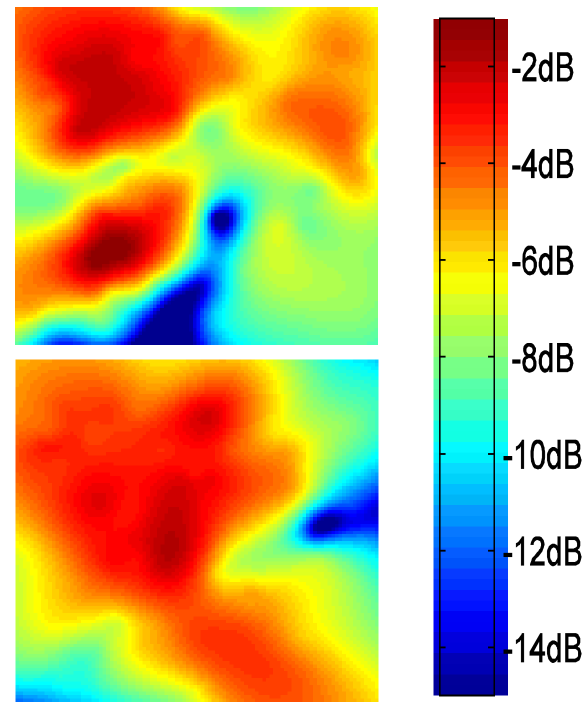
Channel 11



PSD cartography

$N_r = 100$ CRs

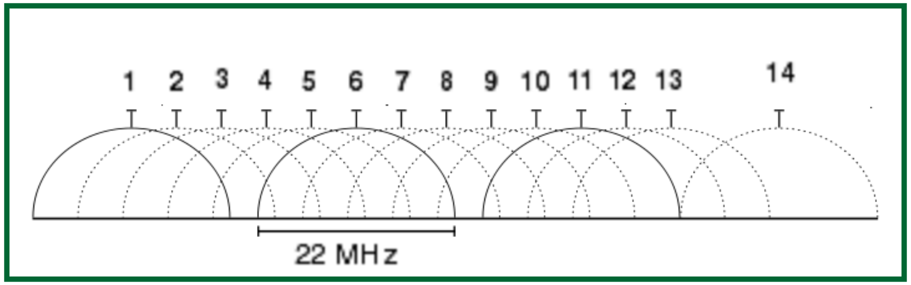
Estimated



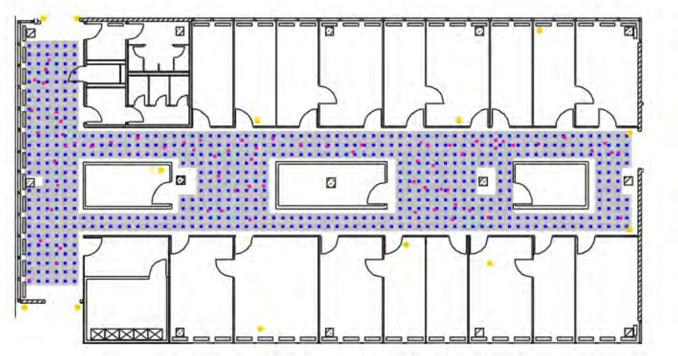
- Maps estimated under fading + shadowing + overlapping bases

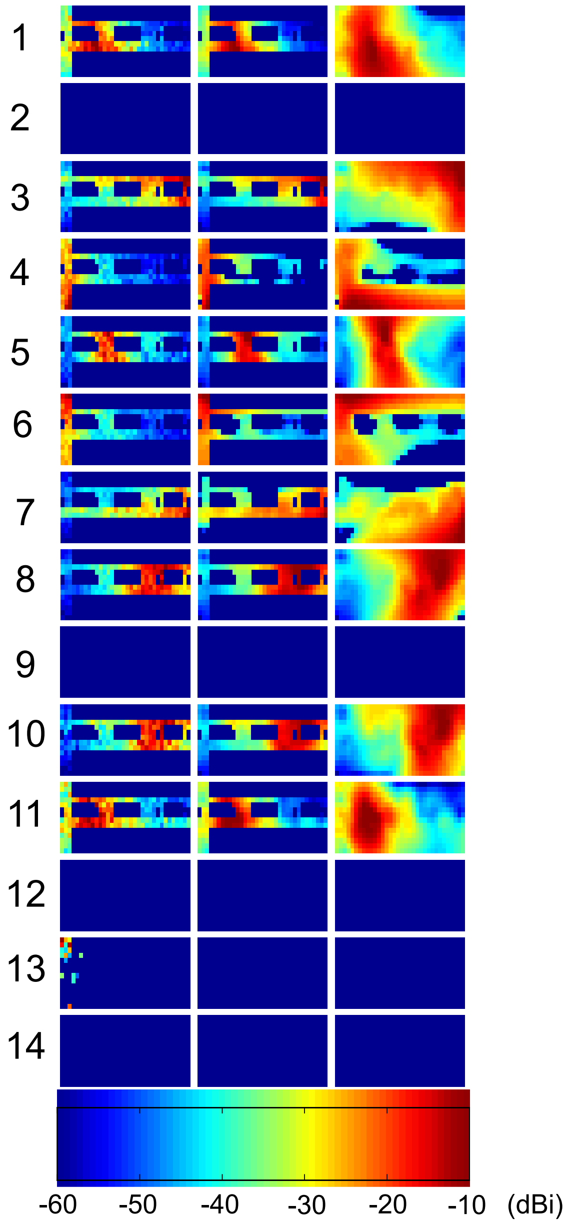
Real RF data

- IEEE 802.11 WLAN activity sensed



$N_r = 166$ CRs

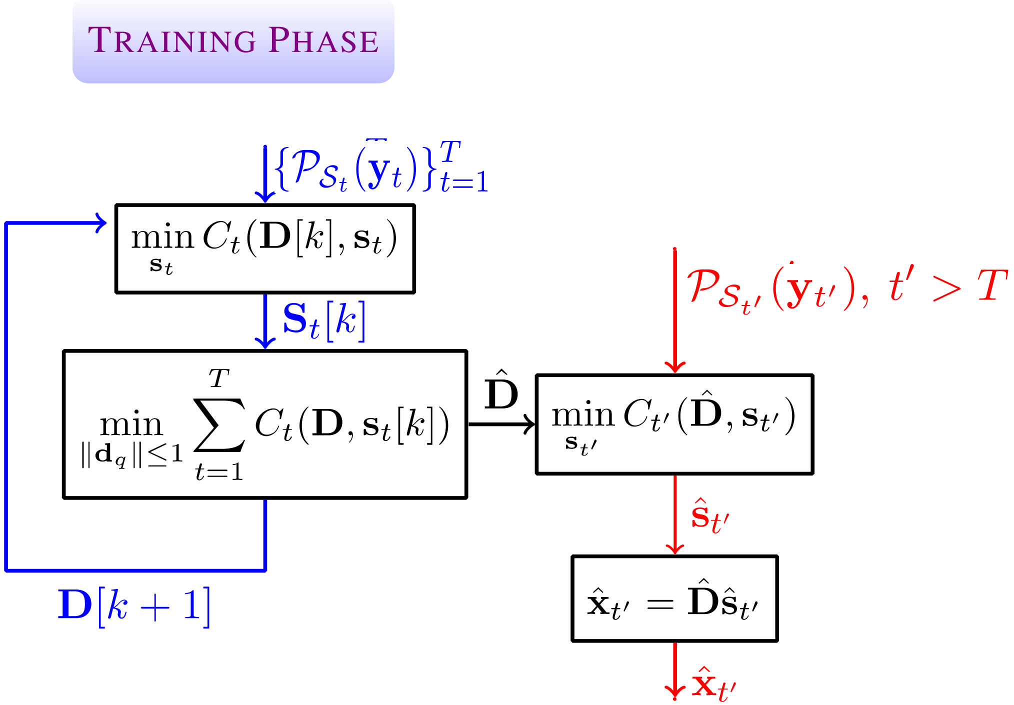




- Frequency bases identified
- Maps recovered and extrapolated

Semi-supervised DL for PSD maps

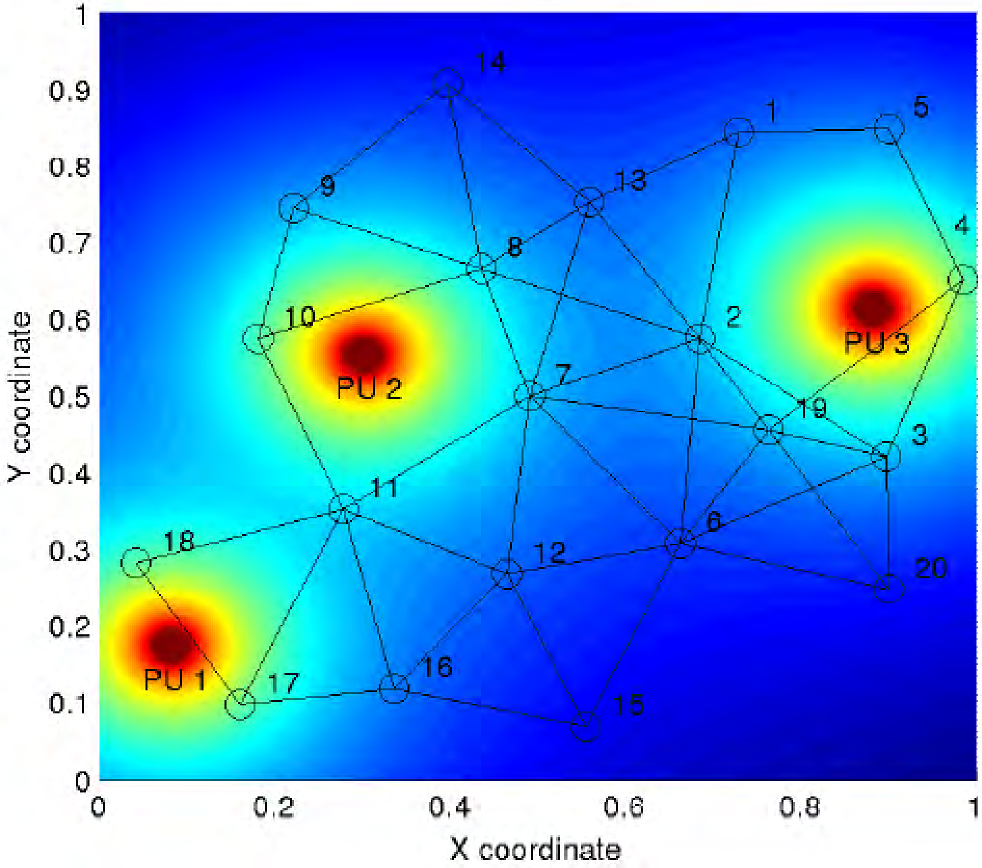
- Signal model $\boldsymbol{\pi}_t = \mathbf{G}_t \mathbf{p}_t \quad \Rightarrow \quad \boldsymbol{\pi} = \mathbf{D}\mathbf{s}$
  - Rx-power measured by a few CRs $\mathcal{P}_{\mathcal{S}_t}(\mathbf{y}_t) = \mathcal{P}_{\mathcal{S}_t}(\pi_t + \mathbf{v}_t)$
- Batch formulation

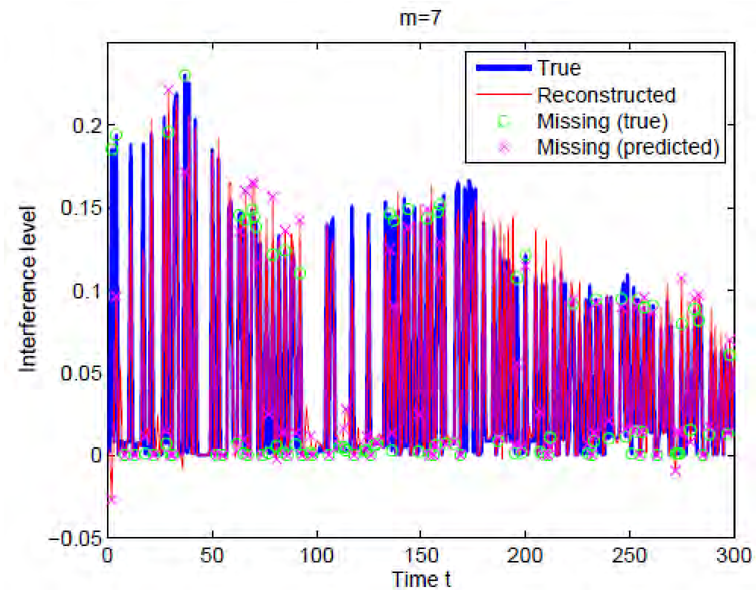


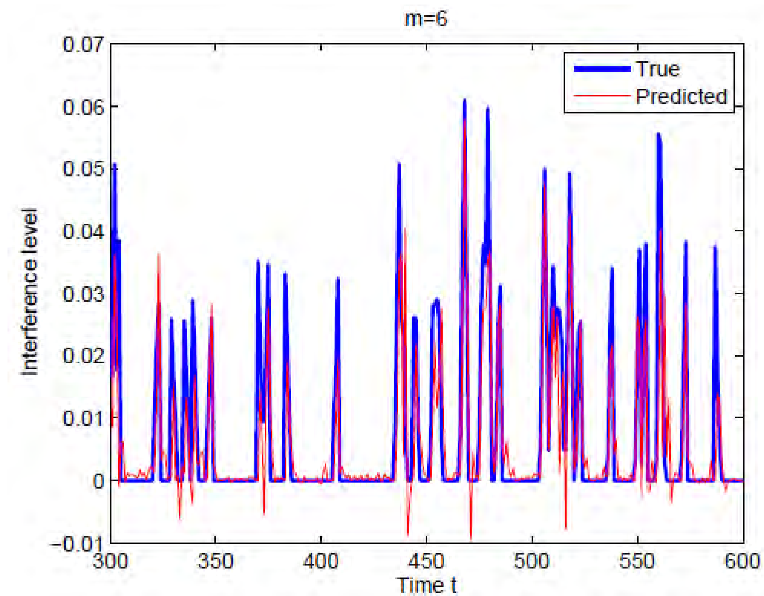
- Online algorithm via exponentially weighted criterion

S.-J. Kim and G. B. Giannakis, "Cognitive Radio Spectrum Prediction using Dictionary Learning," *Proc. of Globecom Conf.*, Atlanta, GA, 2013.

Numerical tests

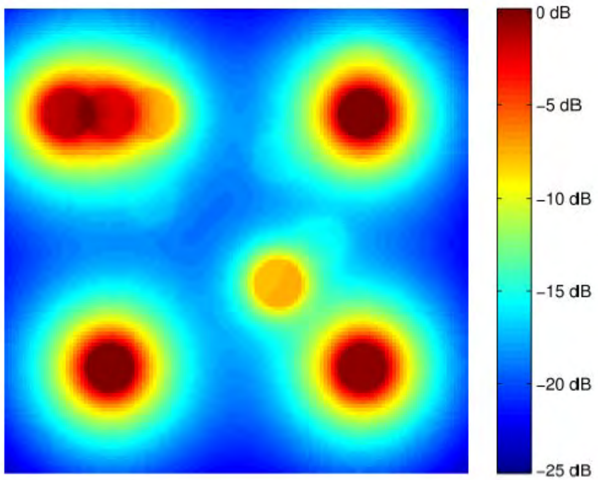




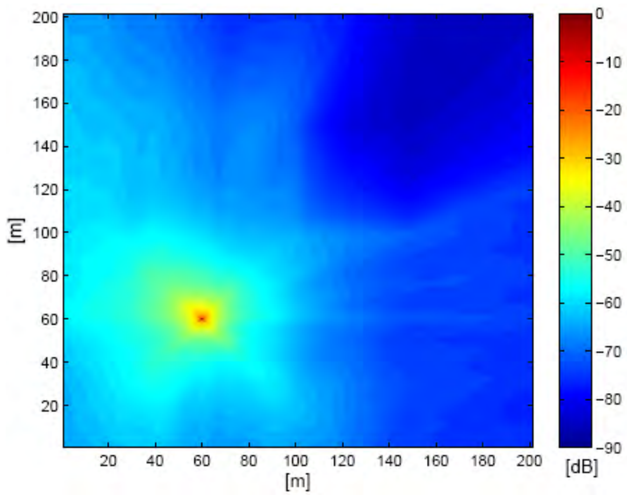


Recap: PHY sensing via RF cartography

- Power spectral density (PSD) maps
  - Capture ambient power in space-time-frequency
  - Can identify “crowded” regions to be avoided



- Channel gain (CG) maps
  - Time-frequency channel from any-to-any point
  - CRs adjust Tx power to min. PU disruption



S.-J. Kim, E. Dall’Anese, and G. B. Giannakis,“Cooperative Spectrum Sensing for Cognitive Radios using Kriged Kalman Filtering,” *IEEE J. Selected Topics in Signal Processing*, pp. 24-36, Feb. 2011.

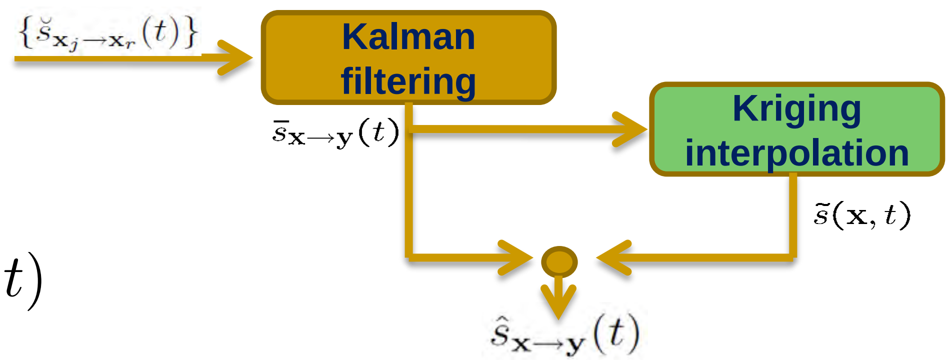
Channel gain cartography

- CG after averaging small-scale fading (dB)

$$G_{\mathbf{x}\to\mathbf{y}}(t) := \Gamma_{\mathbf{x}\to\mathbf{y}}(t) + s_{\mathbf{x}\to\mathbf{y}}(t)$$

- State-space model for shadowing

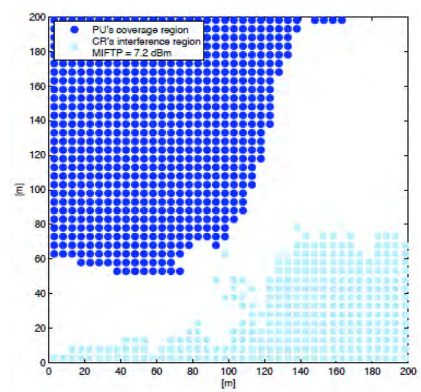
$$s_{\mathbf{x}\to\mathbf{y}}(t) := \bar{s}_{\mathbf{x}\to\mathbf{y}}(t) + \tilde{s}_{\mathbf{x}\to\mathbf{y}}(t)$$

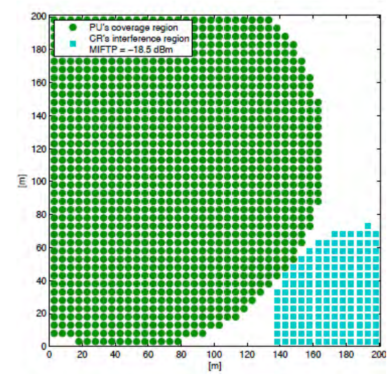


Approach: spatial LMMSE interpolation (Kriging) + KF for tracking channel dynamics

Payoffs: tracking PU activities;
accurate interference models;
efficient resource allocation

Outlook: jointly optimal PHY
CR sensing and access



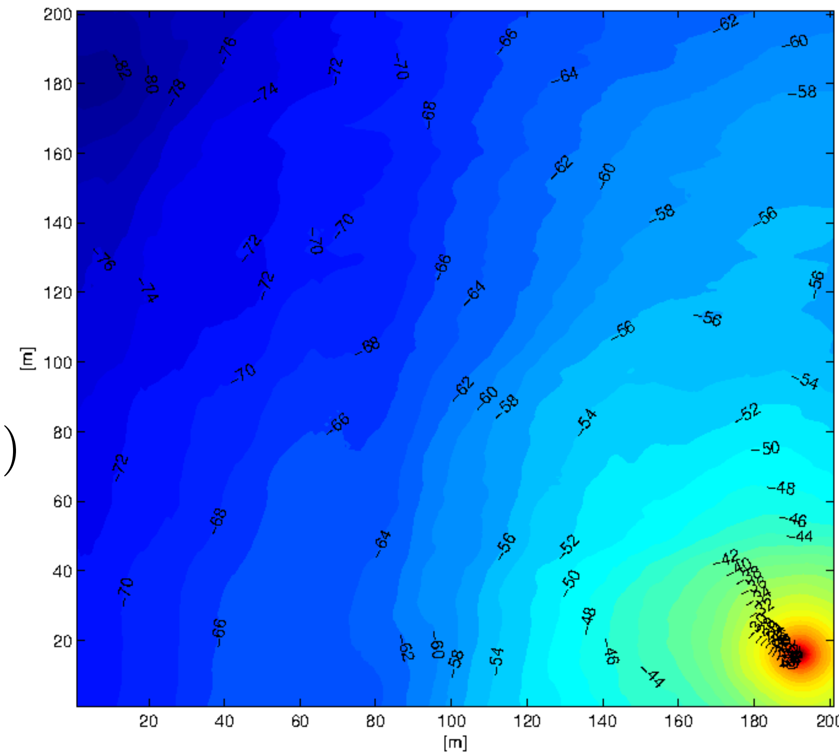


E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel Gain Map Tracking via Distributed Kriging," *IEEE Trans. on Vehicular Technology*, pp. 1205-1211, March 2011.

Any-to-any CG estimation

- Shadowing model-free approach
  - Slow variations in shadow fading
  - Low-rank any-to-any CG matrix $\hat{\mathbf{G}}$

Estimated CG map



Approach: low-rank matrix completion

$$\min_{\mathbf{C},\mathbf{W}} \|\mathcal{P}_{\mathcal{S}}(\mathbf{G} - \mathbf{C}\mathbf{W}')\|_F^2 + \lambda(\|\mathbf{C}\|_F^2 + \|\mathbf{W}\|_F^2)$$

Payoffs: global view of any-to-any CGs;
real-time propagation metrics;
efficient resource allocation

Outlook: kernel-based extrapolator for missing CR-to-PU measurements,
or future time intervals

S.-J. Kim and G. B. Giannakis, "Dynamic Network Learning for Cognitive Radio Spectrum Sensing," *Proc. of Intl. Workshop on Comp. Advances in Multi-Sensor Adaptive Process.*, Saint Martin, 2013.

PU power and CR-PU link learning

- Reduce overhead in any-to-any CG mapping
  - Learn CGs only between CRs and PUs
  - Online detection of active PU transmitters

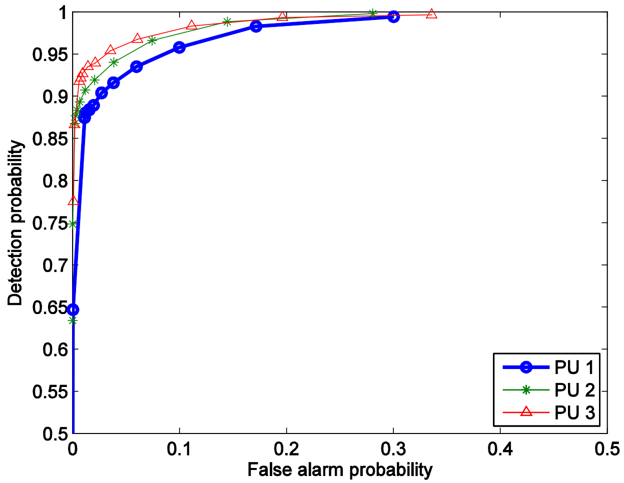
Approach: blind dictionary learning

$$\min_{\mathbf{G},\mathbf{P}} \|\mathbf{\Pi} - \mathbf{GP}\|_F^2 + \lambda_1 \|\mathbf{P}\|_1$$

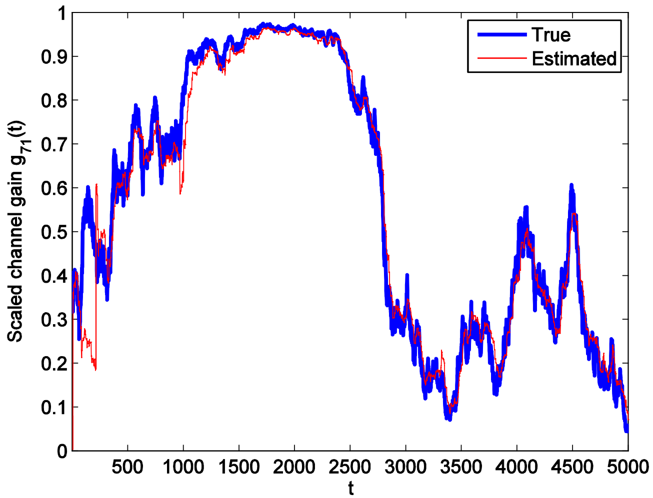
Payoffs: tracking PU activities;
efficient resource allocation

Outlook: missing data due to limited sensing;
distributed and robust algorithms

Detection of PU activity



Estimated CG



S.-J. Kim, N. Jain, and G. B. Giannakis, "Joint Link Learning and Cognitive Radio Sensing," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2011.

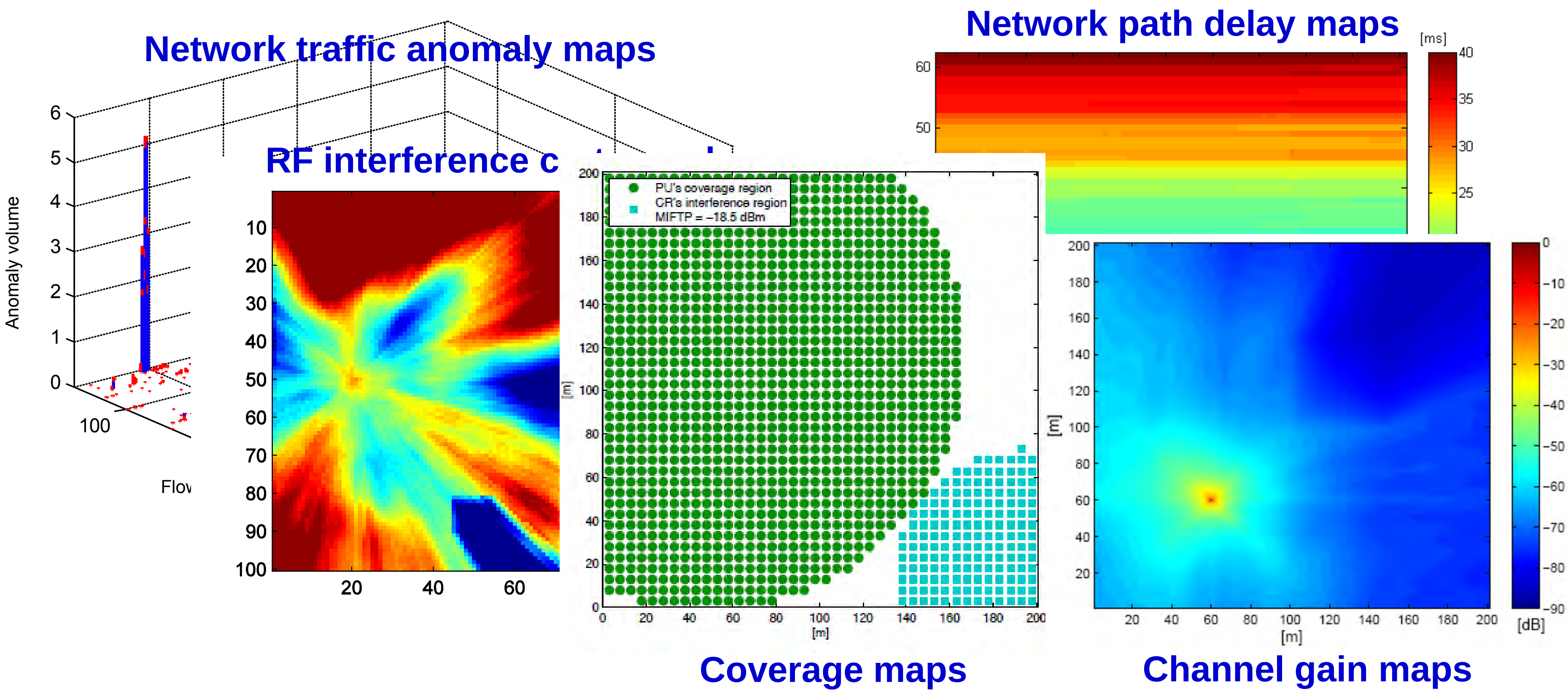
Takeaways

- PHY layer spatiotemporal sensing via RF cartography
  - Space-time-frequency view of interference and channel gains
- Identify idle bands across space
  - Aid dynamic spectrum access policies
- PU/source localization and tracking
- Parsimony via sparsity and distribution via consensus
  - Lasso, group Lasso on splines, and method of multipliers

Roadmap

- Dynamic network delay cartography
- Unveiling network anomalies via sparsity and low rank
- Network-wide link count prediction
- RF cartography for cognition at the PHY
- Conclusions and future research directions

The big picture ahead…



- Network cartography: succinct depiction of the *network state*
- **Vision:** use *atlas* to enable spatial re-use, hand-off, localization, Tx-power tracking, resource allocation, health monitoring, and routing

Concluding summary

- Dynamic network cartography
  - Framework to construct maps of the dynamic network state
  - Real-time, distributed scalable algorithms for large-scale networks
- Global state mapping from incomplete and corrupted data
  - Path delay and link traffic maps
  - Prompt and accurate identification of traffic anomalies
  - PHY layer sensing in wireless CR networks via RF cartography
- Statistical SP toolbox
  - Sparsity-cognizant learning, low-rank modeling
  - Kriged Kalman filtering of dynamical processes over networks
  - Semi-supervised dictionary learning
  - Distributed optimization via the ADMM

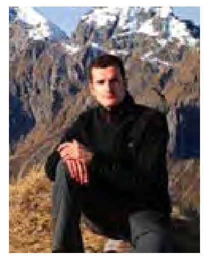
Thank you!

Questions?

University of Minnesota

http://spincom.umn.edu

Dr. J. A. Bazerque
UofM



Dr. E. Dall'Anese
UofM

Dr. P. A. Forero
SPAWAR

Dr. S. J. Kim
UofM

M. Mardani
UofM

Prof. K. Rajawat
IIT Kanpur