Sparse and Low-Rank Optimization for Dense Wireless Networks

Part I: Models

Jun Zhang

HKUST

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Yuanming Shi

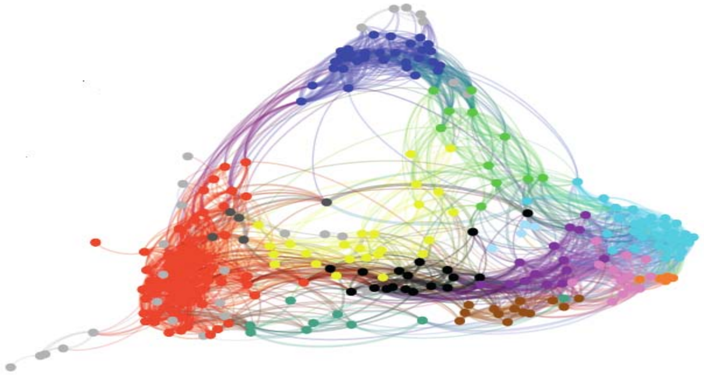
ShanghaiTech University

上海科技大学 ShanghaiTech University

Outline of Part I

- **Motivations**
- **Two Vignettes**
  - Structured Sparse Models
    - Group Sparse Beamforming for Network Power Minimization
    - Sparsity Control for Massive Device Connectivity
  - Generalized Low-Rank Models
    - Low-Rank Matrix Completion for Topological Interference Management
    - Extensions
- **Future Directions**

Motivations: Dense Wireless Networks



Challenge: Ultra mobile broadband

- **Era of mobile data deluge**

18 x

Over past 5 years

429 M

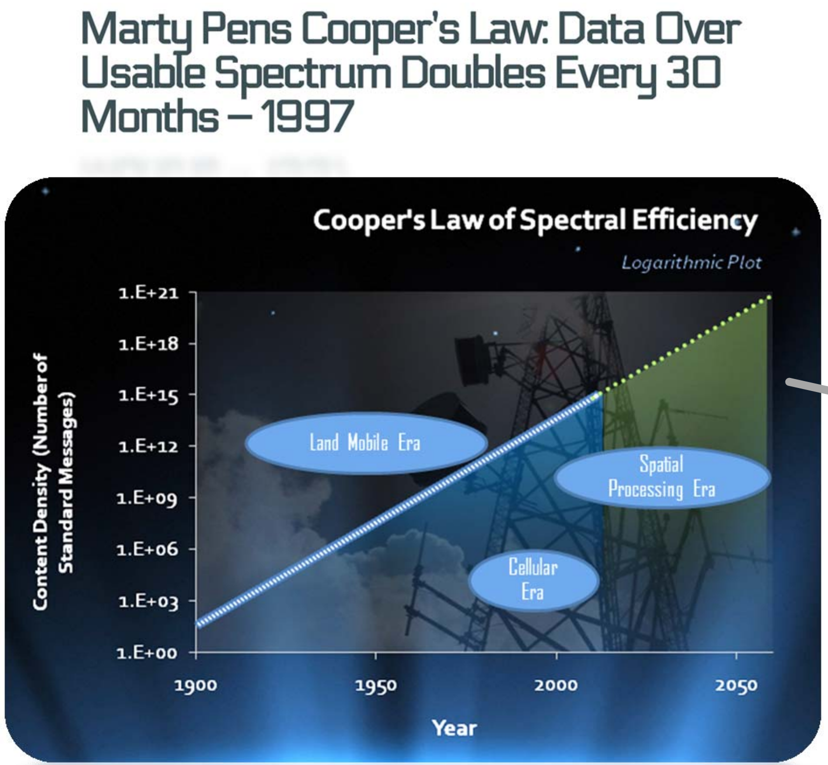
Mobile devices added in 2016

60 %

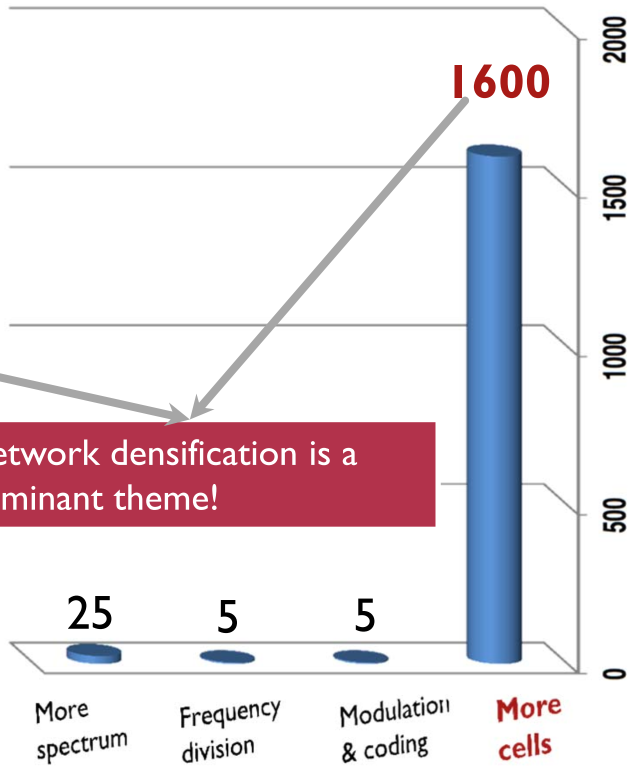
in 2016

Source: Cisco VNI Mobile, 2017

Cooper's Law



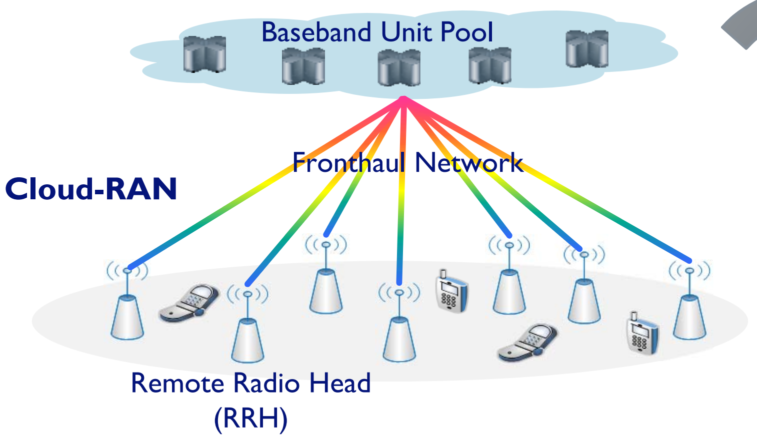
Factor of Capacity Increase since 1950

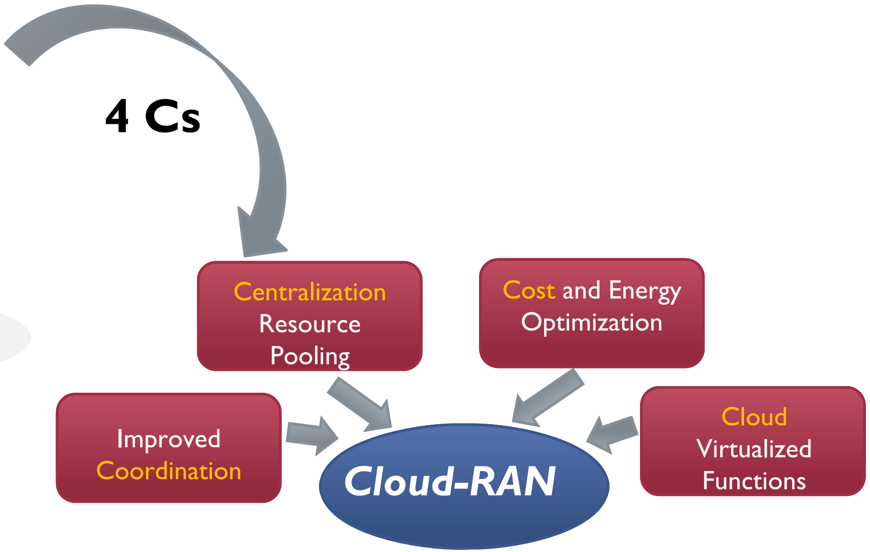


Network densification is a dominant theme!

Solution: cloud radio access networks

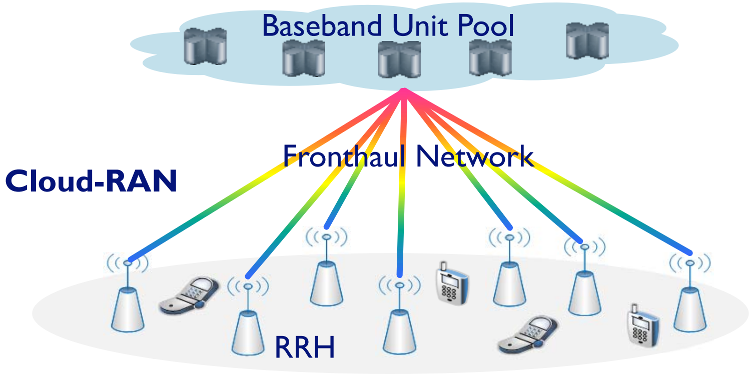
- **Dense Cloud-RAN:** a cost-effective way for wireless network densification and cooperation





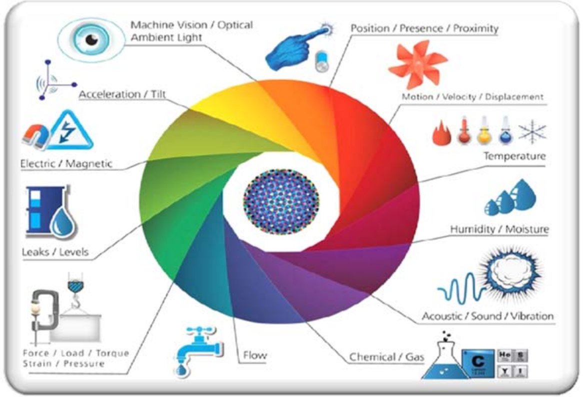
Dense Cloud-RAN

- Higher network capacity
  - Denser deployment
- Scalable connectivity
  - Flexible resource management
- Higher energy efficiency
  - Low-power RRHs, flexible energy management
- Higher cost efficiency
  - Low-cost RRHs, efficient resource utilization



Intelligent things for smart city

- A smart city highly depends on intelligent technology: connected sensors, intelligent devices and IoT networks become wholly integrated



Challenge: Intelligent IoT

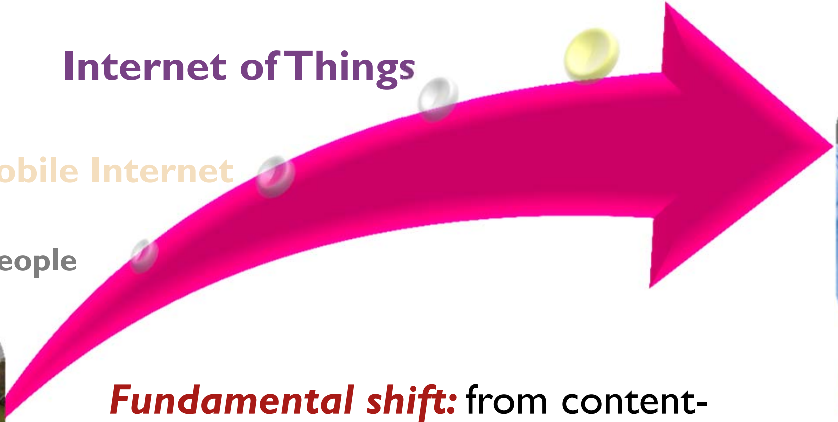
(internet of skills)

Tactile Internet

Internet of Things

Mobile Internet

People to People



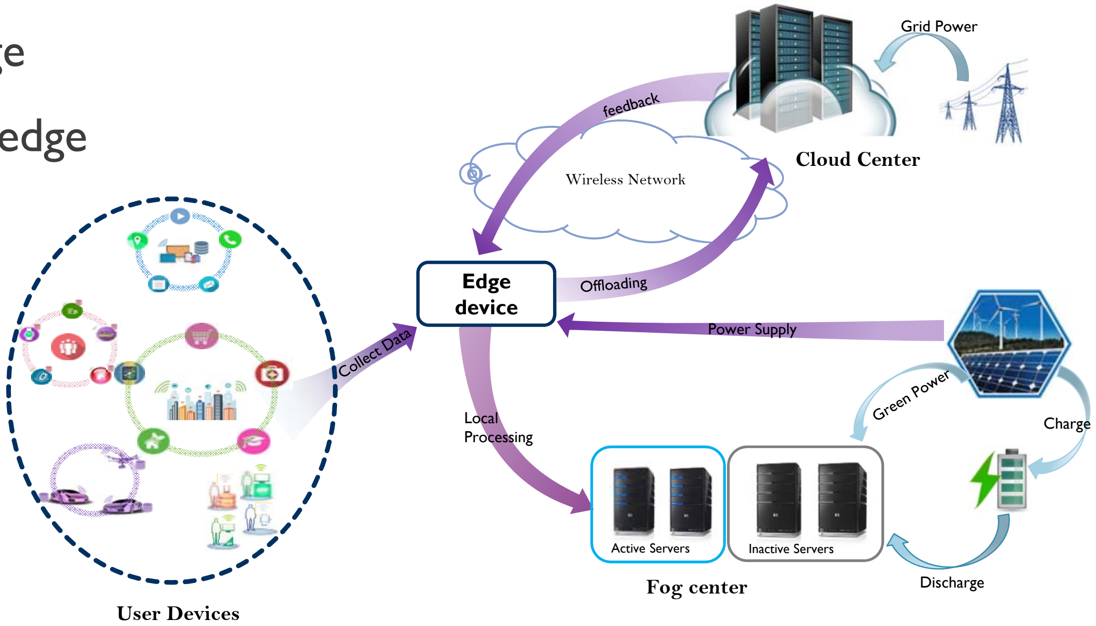
Fundamental shift: from content-delivery to skillset-delivery networks

- **Low Latency**
- **High Computation Intensity**
- **Massive Connectivity**
- …

Solution: fog radio access networks

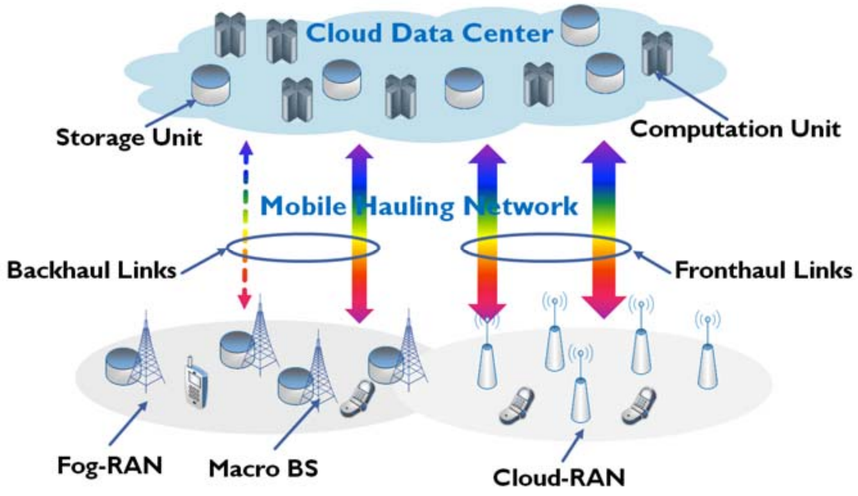
- **Dense Fog-RANs:** push computation and storage resources to network edge – Overcome the **long distance** problem
  - Caching at the edge
  - Computing at the edge

share computation, storage, communication resources across the whole network



A new paradigm for wireless networking

- **Goal:** support ultra-low latency, reliable, Gbps communications, massive device connectivity, massive data analytics, edge-AI…



Difficulties

- **Networking issues:**
  - Huge network power consumption
  - Massive device connectivity
  - Severe network interference

Source: Alcatel-Lucent, 2013

- **Computing issues:**
  - Complicated (non-convex) problem structures
  - Limited computational resources

Sparse and low-rank optimization

- **Successful Stories**
  - **Compressed sensing/matrix completion:** Collect random measurements; reconstruct via optimization
  - **Statistical machine learning:** Random data models; fit model via optimization

- **Advantages**
  - **Modeling flexibility:** Low-dimensional structures in high-dimensional data
  - **Fundamental bounds:** Computational and statistical tradeoffs

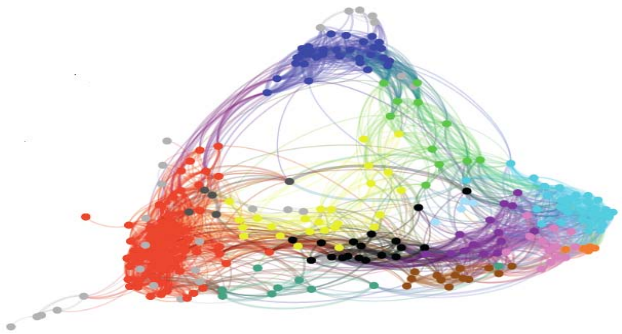
Sparse and low-rank optimization

- **Emerging examples in wireless**
  - **Structured sparse models:** Group sparse beamforming, user admission control, massive device connectivity…
  - **Generalized low-rank models:** Topological interference management, mobile edge caching, wireless distributed computing, index coding…
- **Motivations**
  - **Modeling flexibility:** Structured models in dense and complex networks
  - **Computational scalability:** Convex optimization, manifold optimization…
  - **Theoretical guarantees:** Convex geometry, differential geometry…

Vignette A: Structured Sparse Models

Case I: Group Sparse Beamforming for Network Power Minimization

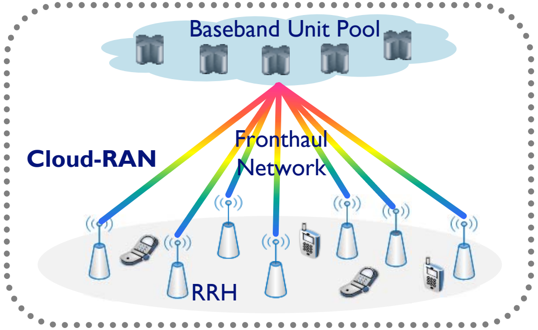
Case II: Sparsity Control for Massive Device Connectivity



Case I: Group Sparse Beamforming for Network Power Minimization

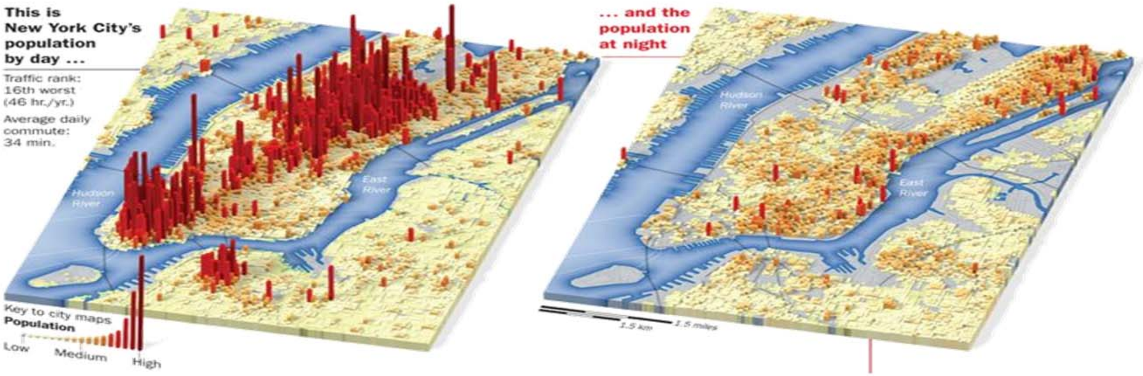
Network power consumption

- **Goal:** Design green dense Cloud-RANs
- **Prior works:** Physical-layer *transmit power consumption*
  - Wireless power control: [Chiang, *et al.*, FT 08], [Qian, *et al.*, TWC 09], [Sorooshyari, *et al.*, TON 12], …
  - Transmit beamforming: [Sidiropoulos and Luo, TSP 2006], [Yu and Lan, TSP 07], [Gershman, et al., SPMag 10],…
- **Challenge:**
  - Network power consumption:
    - Radio access units, fronthaul links, etc.



Network adaptation

- **Goal:** Provide a holistic approach to minimize network power consumption (including RRHs, fronthaul links, etc.)
- **Key observation:** Spatial and temporal mobile data traffic variation



Network adaptation: adaptively switch off network entities to save power

System model

- The received signal at the k-th MU is given by

$$y_k = \sum_{l=1}^{L} \mathbf{h}_{kl}^{\mathsf{H}} \mathbf{v}_{lk} s_k + \sum_{i \neq k} \sum_{l=1}^{L} \mathbf{h}_{kl}^{\mathsf{H}} \mathbf{v}_{li} s_i + n_k, \forall k$$

  - $\mathbf{h}_{kl} \in \mathbb{C}^{N_l}$: channel propagation between MU k and RRH l
  - $\mathbf{v}_{lk} \in \mathbb{C}^{N_l}$: transmit beamforming vector from the l-th RRH to k-th MU
  - Per-RRH transmit power constraint: $\mathcal{C} = \left\{ \mathbf{v}_{lk} : \sum_{k=1}^{K} \|\mathbf{v}_{lk}\|_2^2 \leq P_l, \forall l \right\}$
- The signal-to-interference-plus-noise ratio (SINR) for MU

$$\mathsf{SINR}_k = \frac{|\mathbf{h}_k^{\mathsf{H}} \mathbf{v}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^{\mathsf{H}} \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k, \forall k.$$

$$\mathbf{h}_k \triangleq [\mathbf{h}_{k1}^T, \dots, \mathbf{h}_{kL}^T]^T \in \mathbb{C}^N$$

$$\mathbf{v}_k \triangleq [\mathbf{v}_{1k}^T, \mathbf{v}_{2k}^T, \dots, \mathbf{v}_{Lk}^T]^T \in \mathbb{C}^N$$

$$N = \sum_{l=1}^{L} N_l$$

Network power consumption

- **Continuous function:** Transmit power consumption

$$T(\mathbf{v}) = \sum_{l=1}^{L} \sum_{k=1}^{K} \frac{1}{\eta_l} \|\mathbf{v}_{lk}\|_2^2$$

  - $\eta_l > 0$: Drain inefficiency coefficient of the radio frequency power amplifier

- **Combinatorial function:** Relative fronthaul link power consumption

$$F(\mathcal{T}(\mathbf{v})) = \sum_{l=1}^{L} P_l^c I(\mathrm{Supp}(\mathbf{v}) \cap \mathcal{V}_\ell \neq \emptyset)$$

  - $\mathcal{V}_l = \{K\sum_{i=1}^{l-1} N_i + 1, \ldots, K\sum_{i=1}^{l} N_i\}, \forall l$: a partition of $\mathcal{V} = \{1, \ldots, KN\}$
  - $P_l^c \geq 0$: relative fronthaul link power consumption, i.e., the static power saving when both the fronthaul link and the corresponding RRH are switched off
  - Aggregative beamformer: $\mathbf{v} = [\underbrace{\mathbf{v}_{11}^T, \ldots, \mathbf{v}_{1K}^T}_{\tilde{\mathbf{v}}_1^T}, \ldots, \underbrace{\mathbf{v}_{L1}^T, \ldots, \mathbf{v}_{LK}^T}_{\tilde{\mathbf{v}}_L^T}]^T$

Problem formulation

- **Goal:** Minimize network power consumption in Cloud-RAN

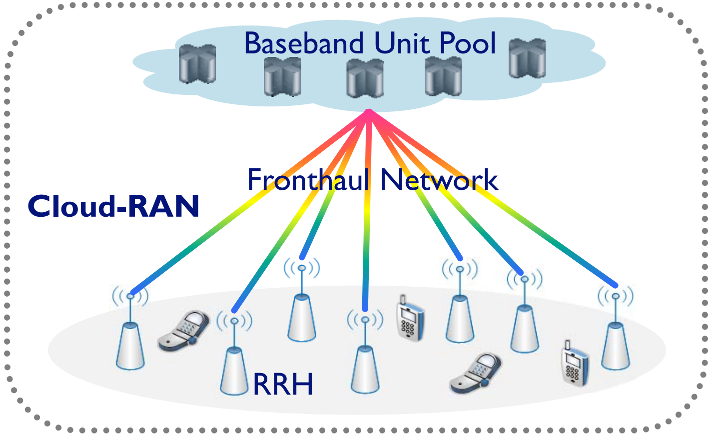
$$\underset{\mathbf{v}\in\mathcal{C}}{\text{minimize}} \quad F(\mathrm{Supp}(\mathbf{v})) + T(\mathbf{v}) \qquad \text{combinatorial composite function}$$

$$\text{subject to} \quad \frac{|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k|^2}{\sum_{i\neq k}|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_i|^2+\sigma_k^2} \geq \gamma_k, \forall k.$$

  - Simultaneously control both the combinatorial function F and the continuous function T
  - **Challenges:** Non-convex, high-dimensional
- **Prior algorithms:** heuristic or computationally expensive [Philipp, et. al, TSP 13], [Luo, et. al, JSAC 13], [Quek, et. al, TWC 13],…

Finding structured solutions

- **Proposal:** group sparse beamforming



$$\mathbf{v} = [\underbrace{\mathbf{v}_{11}^T,\ldots,\mathbf{v}_{1K}^T}_{\tilde{\mathbf{v}}_1^T},\ldots,\underbrace{\mathbf{v}_{L1}^T,\ldots,\mathbf{v}_{LK}^T}_{\tilde{\mathbf{v}}_L^T}]^T$$

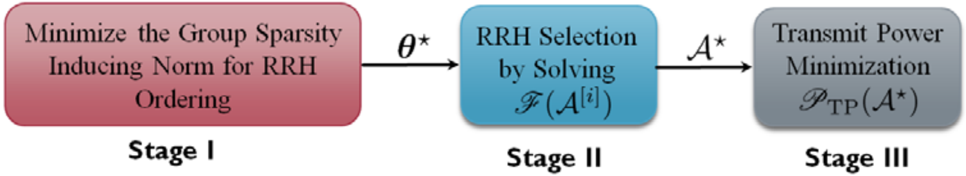
Beamforming coefficients of the first RRH, forming a **group**

- Switch off the l-th RRH $\longrightarrow \tilde{\mathbf{v}}_l = \mathbf{0}$, i.e., group sparsity structure in $\mathbf{v}$

[Ref] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809-2823, May 2014. 2014. **(The 2016 Marconi Prize Paper Award)**

Group sparse beamforming algorithm

- Adaptive RRH selection: Switch off the RRHs with small coefficients in the aggregative beamformers



- **Stage I:** The *tightest* convex positively homogeneous lower bound of the combinatorial composite objective function *(network power)*

$$\Omega(\mathbf{v}) = 2\sum_{l=1}^{L}\sqrt{\frac{P_l^c}{\eta_l}}\|\tilde{\mathbf{v}}_l\|_2 \quad \Rightarrow \quad \underset{\mathbf{v}\in\mathcal{C}}{\text{minimize}}\ \Omega(\mathbf{v}) \quad \Rightarrow \quad \boldsymbol{\theta}^\star = [\|\tilde{\mathbf{v}}_l^\star\|_2]_{l=1}^{L}$$

mixed ℓ_1/ℓ_2-norm — induce group sparsity — RRH ordering

Group sparse beamforming algorithm

- **Stage II:** Find the optimal active RRHs via solving a sequence of following feasibility detection problems (e.g., bi-section search)

$$\mathscr{F}(\mathcal{A}^{[i]}): \text{find} \quad \mathbf{v}_1, \ldots, \mathbf{v}_K$$

$$\text{subject to} \quad \frac{|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k|^2}{\sum_{i\neq k}|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_i|^2+\sigma_k^2} \geq \gamma_k, \forall k$$

$$\mathbf{v}_k = [\mathbf{v}_{lk}] \in \mathbb{C}^{|\mathcal{A}|N}$$

$$\mathbf{h}_k = [\mathbf{h}_{lk}] \in \mathbb{C}^{|\mathcal{A}|N}$$

- **Stage III:** Transmit power minimization via coordinated beamforming

$$\underset{\mathbf{v}\in\mathcal{C}}{\text{minimize}} \quad T(\mathbf{v};\mathcal{A}^{\star})$$

Active RRH set

$$\text{subject to} \quad \frac{|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k|^2}{\sum_{i\neq k}|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_i|^2+\sigma_k^2} \geq \gamma_k, \forall k.$$

$$\mathbf{v}_k = [\mathbf{v}_{lk}] \in \mathbb{C}^{|\mathcal{A}^{\star}|N}$$

$$\mathbf{h}_k = [\mathbf{h}_{lk}] \in \mathbb{C}^{|\mathcal{A}^{\star}|N}$$

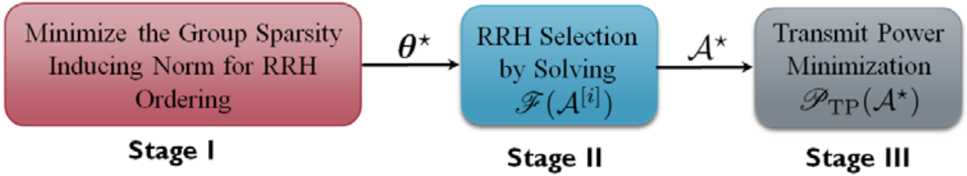
Summary of group sparse beamforming

- SINR constraints can be reformulated as second-order cone constraints

$$f_k(\mathbf{v}) = \sqrt{\sum_{i\neq k} |\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_i|^2 + \sigma_k^2} - \frac{1}{\sqrt{\gamma_k}}\Re(\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k) \leq 0, \forall k.$$ convex

  - Key observation: phases of $\mathbf{v}_k$'s do not change objective and constraints

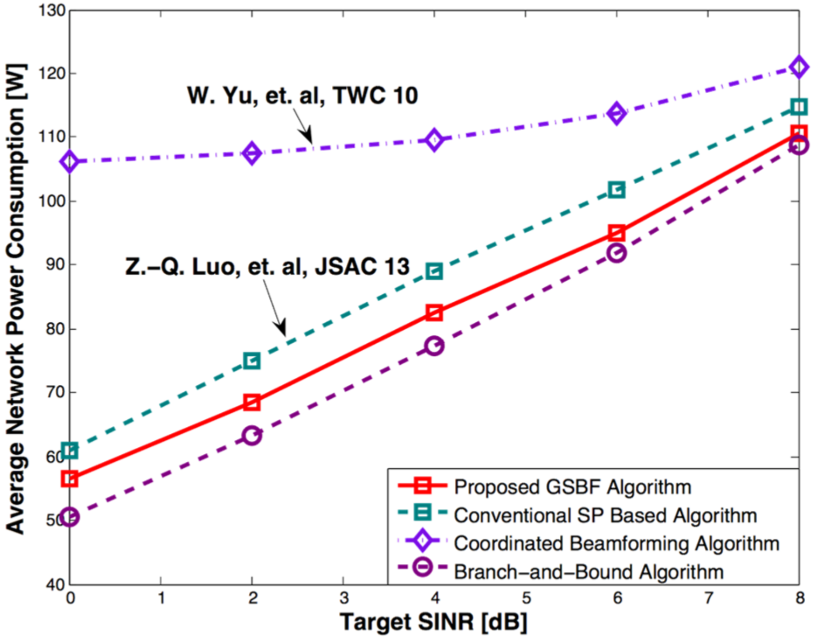
- Group sparse beamforming via convex programming



  - Stage I: Group sparsity inducing via solving one convex program
  - Stage II: **A sequence of convex feasibility problems** need to be solved
  - Stage III: Coordinated beamforming via solving one convex program

The power of group sparse beamforming

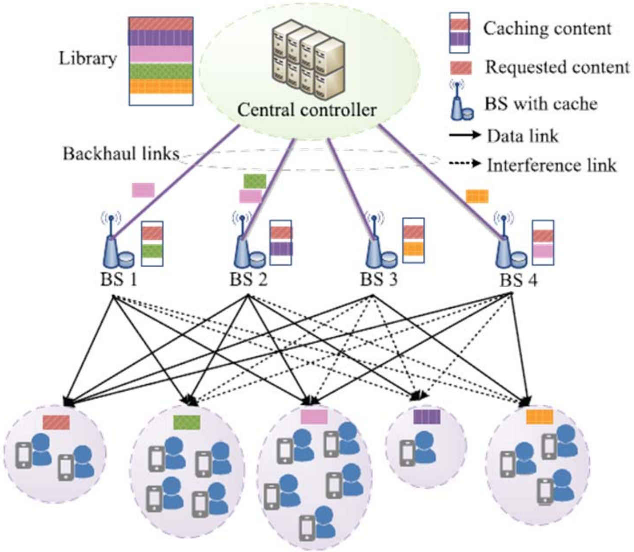
- Group spare beamforming for green Cloud-RAN (10 RRHs, 15 MUs)



Conclusions:
1) Enabling flexible network adaptation;
2) Offering efficient algorithm design via convex programming
3) Empowering wide applications

Extension: Wireless cooperative networks

- A comprehensive consideration: 1) Active BS selection; 2) Transmit beamforming; 3) Backhaul data assignment

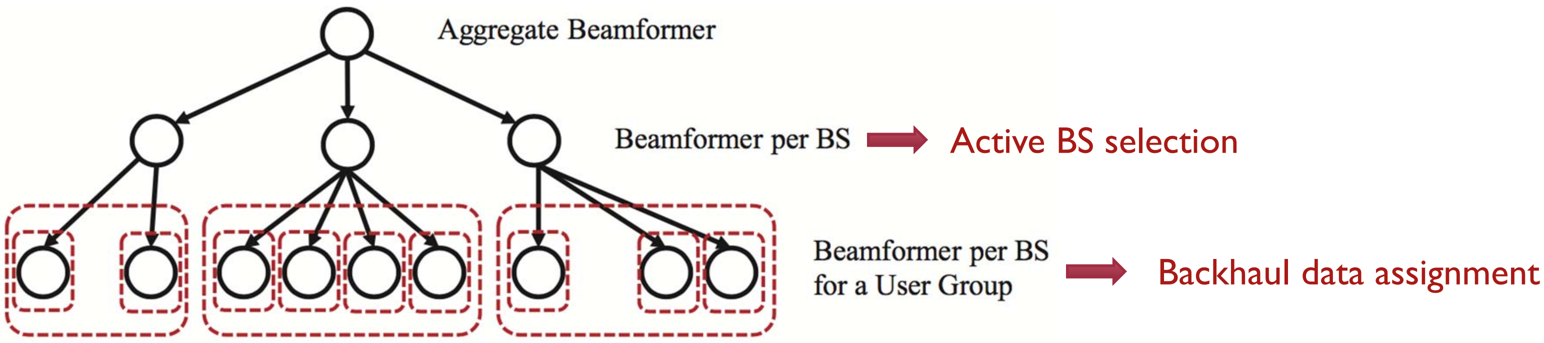


Network power consumption:

1) Static power consumption at BSs
2) Transmit power consumption from BSs
3) Traffic-dependent backhaul power consumption

Layered group sparse beamforming

- **Proposal:** A generalized layered group sparse beamforming (LGSBF) modeling framework
  - To induce the layered sparsity structure in the beamformers



[Ref] X. Peng, Y. Shi, J. Zhang, and K. B. Letaief, "Layered group sparse beamforming for cache-enabled wireless networks," *IEEE Trans. Commun.*, to appear.

Case II: Sparsity Control for Massive Device Connectivity

Motivation

- Downlink transmission with massive devices: **user admission control**
- Uplink machine-type communication (e.g., IoT devices) with sporadic traffic: **massive device connectivity**

Sporadic traffic: only a small fraction of potentially large number of devices are active

Downlink user admission control

- Coordinated beamforming for transmission power minimization

$$\begin{aligned} \underset{\mathbf{v}\in\mathcal{C}}{\text{minimize}} \quad & \|\mathbf{v}\|_2^2 \\ \text{subject to} \quad & \frac{|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k|^2}{\sum_{i\neq k}|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_i|^2+\sigma_k^2} \geq \gamma_k, \forall k. \end{aligned}$$

- SINR constraints can be reformulated as second-order cone constraints

$$f_k(\mathbf{v}) = \sqrt{\sum_{i\neq k}|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_i|^2+\sigma_k^2} - \frac{1}{\sqrt{\gamma_k}}\Re(\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k) \leq 0, \forall k.$$

  - Key observation: phases of $\mathbf{v}_k$'s do not change objective and constraints

Infeasibility

- Set of convex inequalities:

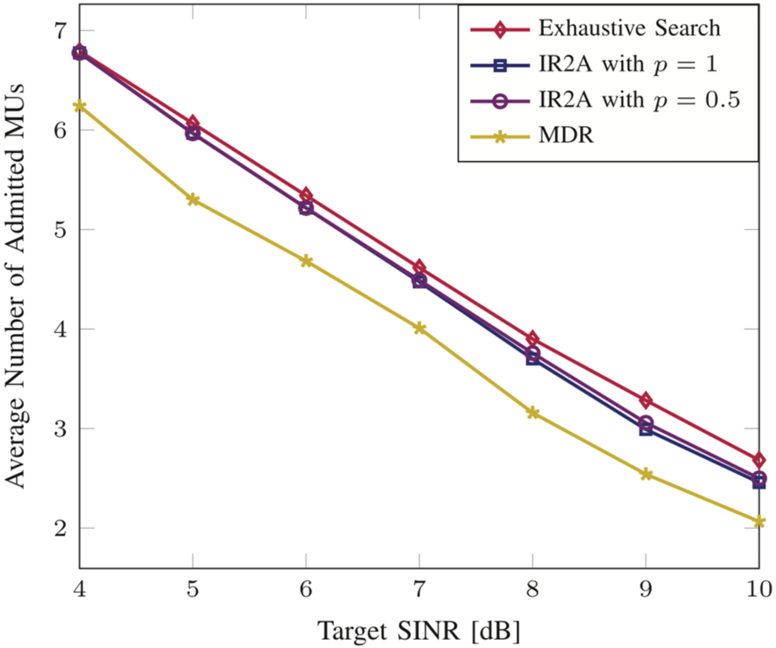
$$f_1(\mathbf{v}) \le 0, \ldots, f_m(\mathbf{v}) \le 0, \qquad \mathbf{v} \in \mathcal{C}$$

- Power minimization problem is generally **infeasible**: large number of users, unfavorable channel conditions, high data rate requirements,…
- **Goal:** Maximize the user capacity, i.e., the number of admitted users
- **Solution:** Choose $\mathbf{v}$ to minimize the number of violated inequalities

$$\begin{aligned} \underset{\mathbf{v}\in\mathcal{C},\boldsymbol{s}}{\text{minimize}} \quad & \|\boldsymbol{s}\|_0 \\ \text{subject to} \quad & f_i(\mathbf{v}) \le s_i, i = 1, \ldots, m \\ & s_i \ge 0, i = 1, \ldots, m \end{aligned}$$

Sparse optimization for user admission control

- Average number of admitted mobile users versus target SINR



IR2A: iterative reweighted ℓ_2-algorithm

$$\|\boldsymbol{s}\|_0 \rightarrow \|\boldsymbol{s}\|_p (0 < p < 1)$$

MDR: membership deflation by convex relaxation ($\|\boldsymbol{s}\|_0 \rightarrow \|\boldsymbol{s}\|_1$)

Massive device connectivity in uplink

- Cellular network with a massive number of devices
  - Single-cell uplink with a BS with M antennas; Block-fading channel with coherence time T ; Total N single-antenna devices, $\mathcal{S} \subset \{1, 2, \ldots, N\}$ devices are active (sporadic traffic)

$$\boldsymbol{y}(\ell) = \sum_{i \in \mathcal{S}} \boldsymbol{h}_i q_i(\ell) + \boldsymbol{n}(\ell), \ell = 1, \ldots, L$$

- Define diagonal activity matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ with $|\mathcal{S}|$ non-zero diagonals

$$\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{H} + \boldsymbol{N}$$

  - $\boldsymbol{Y} = [\boldsymbol{y}(1), \ldots, \boldsymbol{y}(L)]^T \in \mathbb{C}^{L \times M}$ denotes the received signal across M antennas
  - $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N]^T \in \mathbb{C}^{N \times M}$: channel matrix from all devices to the BS
  - $\boldsymbol{Q} = [\boldsymbol{q}(1), \ldots, \boldsymbol{q}(L)]^T \in \mathbb{C}^{L \times N}$: known transmit pilot matrix from devices

Challenges of massive connectivity

- **Sporadic traffic**
  - User activity detection is a key requirement
  - Massive number of devices mean pilot sequences cannot be orthogonal
  - Device identification is a sparse optimization problem
- Prior works on compressed sensing for massive connectivity: 1) Without channel estimation [Zhang-Luo-Guo'13]; 2) Joint user activity detection and channel estimation [Xu-Rao-Lau'15]; 3) Approximate Message Passing (AMP) [Wei'16]
- **Proposal:** User activity detection and channel estimation based on the compressed sensing techniques (without channel distribution prior)

Group sparsity estimation

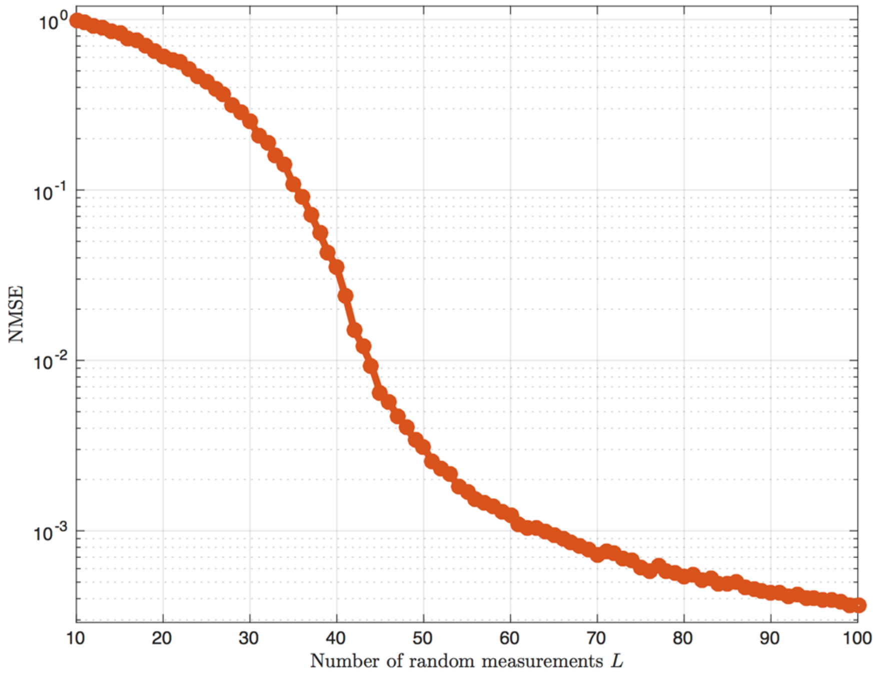
- Let $\boldsymbol{\Theta}^{\natural} = \boldsymbol{A}\boldsymbol{H} \in \mathbb{C}^{N \times M}$ (unknown): group sparsity in rows $\boldsymbol{\theta}^{[i]}$ of matrix $\boldsymbol{\Theta}^{\natural}$
  - Simultaneous user activity detection and channel estimation
- Let $\boldsymbol{Q} \in \mathbb{C}^{L \times N}$ be a known measurement operator (pilot matrix)
- **Observe** $\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{\Theta}^{\natural} + \boldsymbol{N}$
- Find estimate $\hat{\boldsymbol{\Theta}}$ by solving a **convex program**

$$\underset{\boldsymbol{\Theta} \in \mathbb{C}^{N \times M}}{\text{minimize}} \quad f(\boldsymbol{\Theta}) \quad \text{subject to} \quad \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{\Theta}\|_F \leq \epsilon$$

  - $f(\boldsymbol{\Theta}) = \sum_{i=1}^{N} \|\boldsymbol{\theta}^{[i]}\|_2$ is mixed ℓ_1/ℓ_2-norm to reflect group sparsity structure
- **Hope:** $\hat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}^{\natural}$

Sparse estimation for massive connectivity

- Normalized MSE versus pilot matrix length L



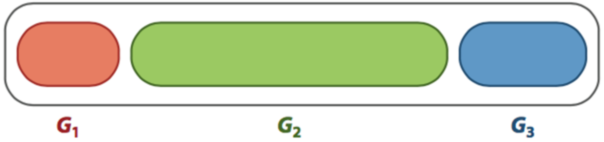
$$N = 100, M = 2, |\mathcal{S}| = 20$$

Summary: structured sparse models

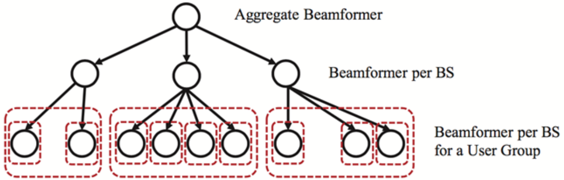
- Generalized structured sparse optimization for dense networks

$$\underset{\boldsymbol{z}\in\mathbb{C}^n}{\text{minimize}} \quad f(\boldsymbol{z}) := \alpha f_1(\text{Supp}(\boldsymbol{z})) + \beta f_2(\boldsymbol{z}) \quad \text{subject to} \quad \boldsymbol{z} \in \mathcal{C}$$

  - $\text{Supp}(\boldsymbol{z})$ is the index set of non-zero coefficients of a vector $\boldsymbol{z}$
  - f_1: combinatorial positive-valued set-function to control sparsity in $\boldsymbol{z}$
  - f_2: continuous convex function to represent the system performance
  - $\mathcal{C}$: to model system constraints, e.g., QoS constraints



group-structured sparsity

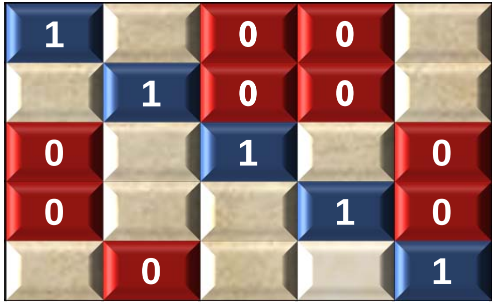


layered group sparsity

Vignette B: Generalized Low-Rank Models

Case I: Low-Rank Matrix Completion for Topological Interference Management

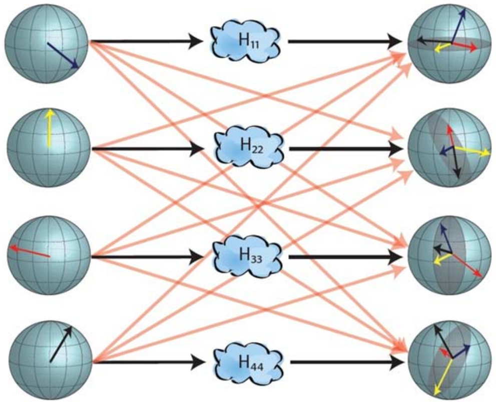
Case II: Extensions to Mobile Edge Caching, Distributed Computing and Index Coding



Case I: Topological Interference Management

Interference channel

- **Channel model:** $y_i = h_{ii}x_i + \sum_{j \neq i} h_{ij}x_j + z_j, i = 1, \ldots, n$



capacity is unknown

- **Degrees-of-freedom:** simplify the analysis; lead to physical insights

$$\mathsf{DoF} = \lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log(\mathsf{SNR})}$$

Interference alignment

- Assume the channel coefficients change over time:

$$y_i(t) = h_{ii}(t)x_i(t) + \sum\nolimits_{j\neq i} h_{ij}(t)x_j(t) + z_j(t)$$

- Consider T channel uses: $X_i = [x_i(1), \ldots, x_i(T)]^\dagger, Z_i = [z_i(1), \ldots, z_i(T)]^\dagger$

$$Y_i = H_{ii}X_i + \sum_{i\neq j} H_{ij}X_j + Z_i \qquad H_{ij} = \begin{bmatrix} h_{ij}(1) & & \\ & \ddots & \\ & & h_{ij}(T) \end{bmatrix}$$

- Transmitter j sends m information symbols S_j across T channel uses

$$X_j = V_j S_j \qquad V_j \in \mathbb{C}^{T\times m}$$ represents the precoding matrix

- The i-th interference term $\sum_{j\neq i} H_{ij}V_jS_j$ lives in the range space of matrix

$$[H_{i1}V_1 \ \ldots \ H_{i,i-1}V_{i-1} \ H_{i,i+1}V_{i+1} \ \ldots \ H_{in}V_n]_{T\times(n-1)m}$$

Interference alignment

- **Interference alignment condition:** find precoding matrices and decoding matrices $V_j \in \mathbb{C}^{T\times m}, U_i \in \mathbb{C}^{m\times T}$ such that

$$\mathrm{rank}(U_i H_{ii} V_i) = m, \forall i = 1, \dots, n$$
$$U_i[H_{i1}V_1 \ \dots \ H_{i,i-1}V_{i-1} \ H_{i,i+1}V_{i+1} \ \dots \ H_{in}V_n] = 0, \forall i = 1, \dots, n$$

- Each user can send m symbols: interference free across T channel uses

$$\text{DoF} = m$$

- **Intuition:** The interference has aligned onto a $T - m$ dimensional subspace at each receiver.

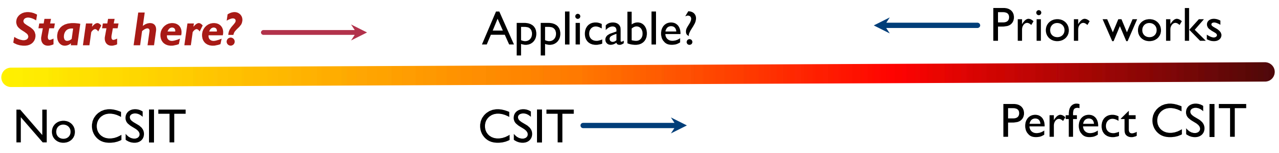
Interference alignment

- **Everyone gets half the cake** **[Cadambe-Jafar'08]:** DoF $= n/2$
  - **Diagonal** H_{ij} are time-varying and generic, $T \to \infty$, $m = T/2$ is almost surely asymptotically achievable
- **Remarks:**
  - Require very long block lengths
  - Require the channels to vary generically over time
  - Require full knowledge of the channel coefficients of *every link* in the network, at *each transmitter* and for *all times*!

Can we exploit the interference alignment principle in practical systems?

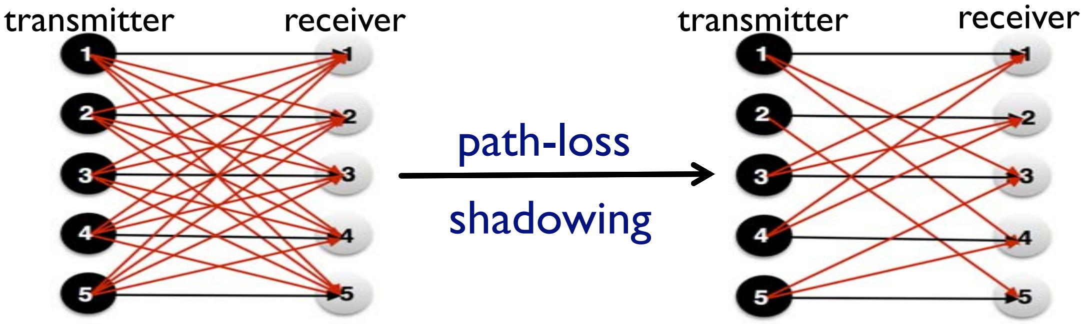
Practical interference management

- **Goal:** Exploit the IA principle under realistic assumptions on CSIT
- **Prior works:** Abundant CSIT ➡ relaxed CSIT
  - Perfect CSIT [Cadambe and Jafar, TIT 08]
  - Delayed CSIT [Maddah-Ali and Tse, TIT 12]
  - Alternating CSIT [Tandon, et al., TIT 13], partial and imperfect CSIT [Shi, et al., TSP 14],…
- **Curses:** CSIT is rarely abundant (due to training & feedback overhead)



Topological interference alignment

- **Blessings:** partial connectivity in dense wireless networks



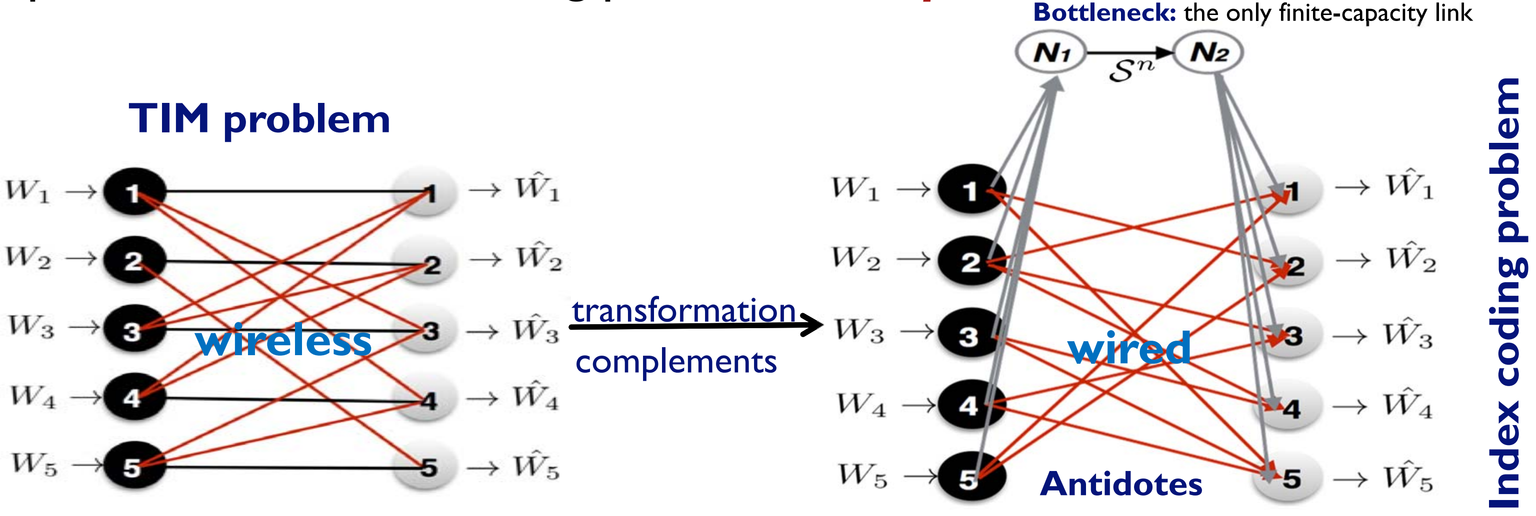
Degrees of Freedom?

$$\mathrm{DoF} = \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log(\mathrm{SNR})}$$

- **Approach:** topological interference management (TIM) [Jafar, TIT 14]
  - **Maximize the achievable DoF:** only based on the network topology information **(no CSIT)**

Index coding approach

- **Theorem [Jafar, TIT 14]:** under linear (vector space) solutions, TIM problem and index coding problem are ***equivalent***



Only a few index coding problems have been solved!

Low-rank matrix completion approach

- **Goal:** Deliver one data stream per user over N time slots
  - Transmitter i transmits $\mathbf{v}_i s_i$, receiver i receives

$$\mathbf{y}_i = \mathbf{v}_i h_{ii} s_i + \sum_{(i,j)\in\mathcal{S}, i\neq j} \mathbf{v}_j h_{ij} s_j + \mathbf{n}_i$$

$\mathcal{S}$: network connectivity pattern

  - Receiver decodes symbol s_i by projecting $\mathbf{y}_i$ onto the space $\mathbf{u}_i \in \mathbb{C}^N$

$$\mathbf{u}_i^{\mathsf{H}} \mathbf{y}_i = \mathbf{u}_i^{\mathsf{H}} \mathbf{v}_i h_{ii} s_i + \sum_{(i,j)\in\mathcal{S}, i\neq j} \mathbf{u}_i^{\mathsf{H}} \mathbf{v}_j h_{ij} s_j + \mathbf{u}_i^{\mathsf{H}} \mathbf{n}_i$$

- **Topological interference alignment condition**

$$M_{ij} = \begin{cases} \mathbf{u}_i^{\mathsf{H}} \mathbf{v}_i = 1, & \forall i, \\ \mathbf{u}_i^{\mathsf{H}} \mathbf{v}_j = 0, & \forall i \neq j, (i,j) \in \mathcal{S}, \\ \star, & \text{otherwise.} \end{cases} \quad \Rightarrow \quad \begin{array}{l} \mathbf{u}_i^{\mathsf{H}} \mathbf{y}_i = h_{ii} s_i + \mathbf{u}_i^{\mathsf{H}} \mathbf{n}_i \\ \mathrm{DoF} = \frac{1}{\mathrm{rank}(\boldsymbol{M})} = \frac{1}{N} \end{array}$$

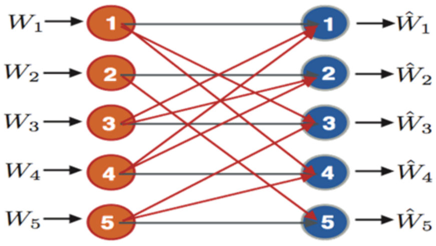
Generalized low-rank model

- Generalized low-rank optimization with network side information

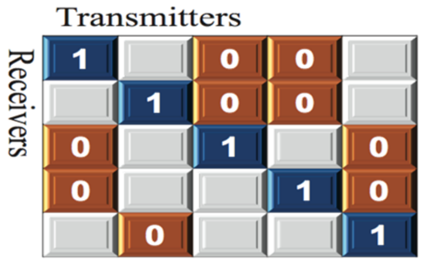
$$\begin{aligned} \underset{\boldsymbol{M}\in\mathbb{C}^{K\times K}}{\text{minimize}} \quad & \text{rank}(\boldsymbol{M}) \\ \text{subject to} \quad & M_{ii}=1, i=1,\ldots,K \\ & M_{ij}=0, \forall(i,j)\in\mathcal{S} \end{aligned}$$

topological interference alignment condition

  - $\boldsymbol{M}=[\boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{v}_j]$: precoding vectors and decoding vectors $\boldsymbol{u}_k,\boldsymbol{v}_k\in\mathbb{C}^N$
  - $\text{rank}(\boldsymbol{M})$ equals the inverse of achievable degrees-of-freedom (DoF) $\frac{1}{N}$



(a) Topological interference alignment



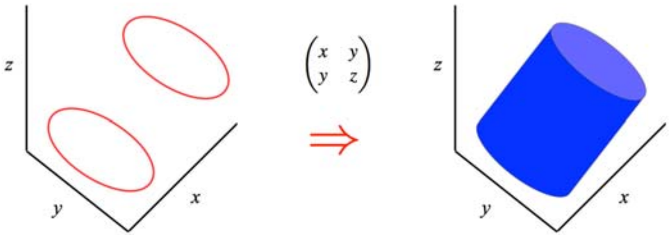
(b) Side information modeling matrix $\boldsymbol{M}$

side information $\mathcal{S}$

Nuclear norm fails

- **Convex relaxation fails:** always returns the identity matrix!

$$\begin{aligned} \underset{M \in \mathbb{C}^{K\times K}}{\text{minimize}} \quad & \|M\|_* \\ \text{subject to} \quad & M_{ii} = 1, i = 1, \ldots, K \\ & M_{ij} = 0, \forall (i,j) \in \mathcal{S} \end{aligned}$$



  - **Fact:** $\mathrm{Trace}(M) \le \|M\|_*$

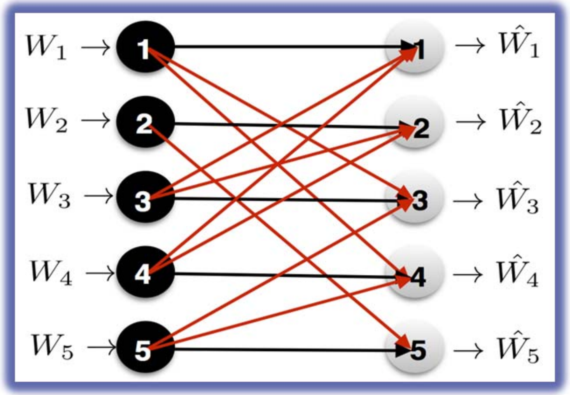
- **Proposal:** Solve the nonconvex problems directly with rank adaptivity

$$\begin{aligned} \underset{M \in \mathbb{C}^{K\times K}}{\text{minimize}} \quad & f(M) := \|\mathcal{A}(M) - z\|_F^2 \\ \text{subject to} \quad & \mathrm{rank}(M) = r \end{aligned}$$

manifold constraint

Riemannian manifold optimization problem

Numerical results



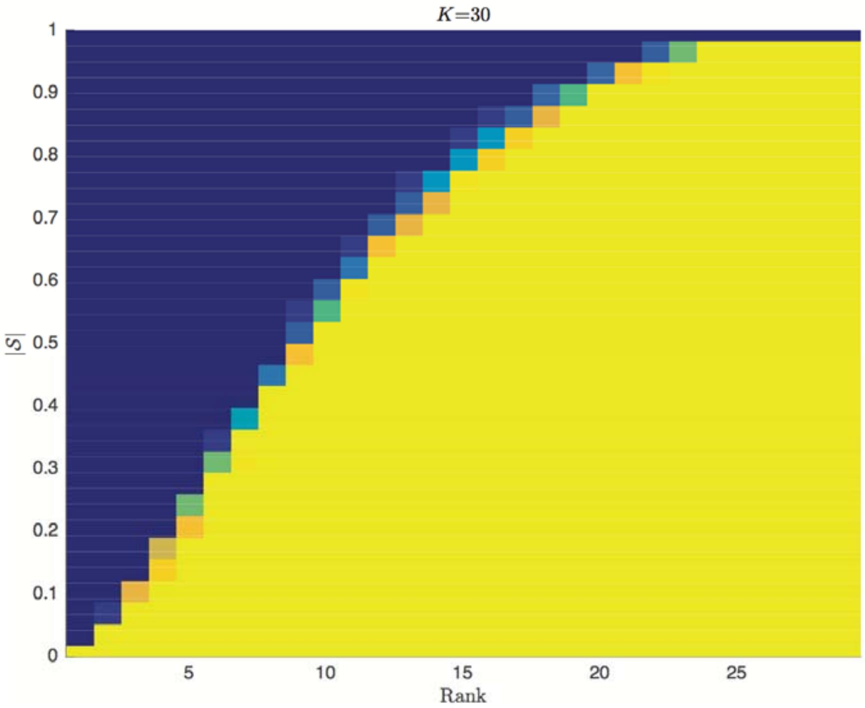
1		0	0	
	1	0	0	
0		1		0
0			1	0
	0			1

Recover all the optimal DoF results for the special TIM problems in [Jafar '14]

1	.1	0	0	9.5
6.8	1	0	0	64
0	.1	1	-1	0
0	-.1	-1	1	0
.1	0	-.1	.1	1

Provide numerical insights (optimal/lower-bound) for the general TIM problems

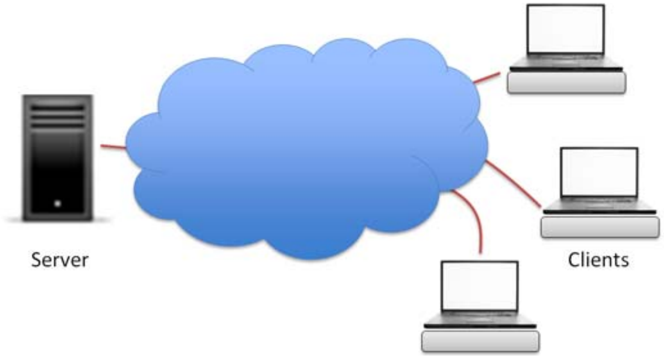
Phase transitions for topological IA



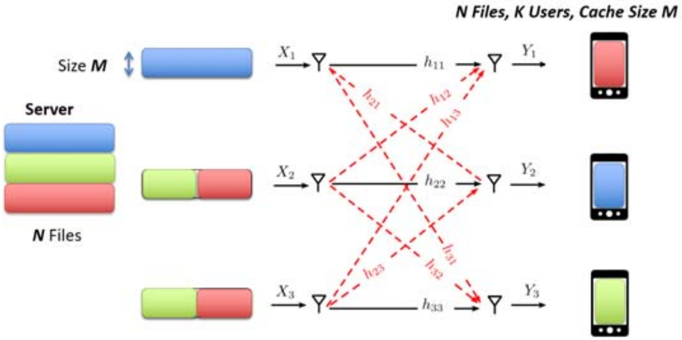
The heat map indicates the empirical probability of success (blue=0%; yellow=100%)

Extension to cache networks

- **Cache gains:** load balancing, interference cancellation/alignment, cooperative transmission, …
- **Placement phase:** populate caches (prefetching)
- **Delivery phase:** reveal request, deliver content



wired cache network



wireless cache network

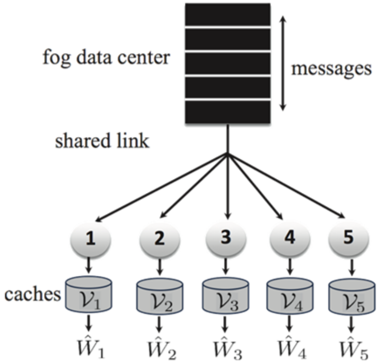
Caching at receivers

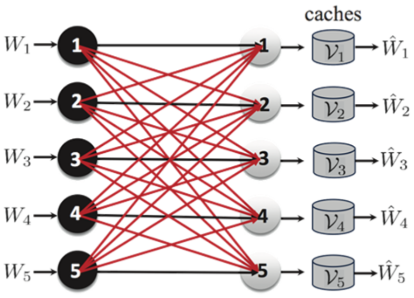
- **Cached receivers:** topological interference alignment

$$\boxed{\begin{aligned} \boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{v}_i &= 1, \quad i = 1, \cdots, K \\ \boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{v}_j &= 0, \quad \forall j \notin \mathcal{V}_i, (i,j) \in \mathcal{S} \end{aligned}}$$

$$\Rightarrow$$

$$\boxed{\begin{aligned} &\text{minimize} && \operatorname{rank}(\boldsymbol{M}) \\ &\text{subject to} && M_{ii} = 1, i = 1, \ldots, K \\ & && M_{ij} = 0, \forall j \in \mathcal{V}_i, (i,j) \in \mathcal{S} \end{aligned}}$$





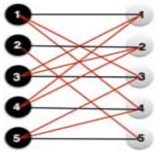
Side information:

1) Cached files $\mathcal{V}_i$
2) Network topology $\mathcal{S}$

When index coding meets low-rank matrices

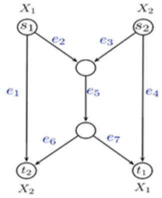
Interference Alignment

[Jafar '14]



Network Coding

[Rouayheb et al. '10, '15]

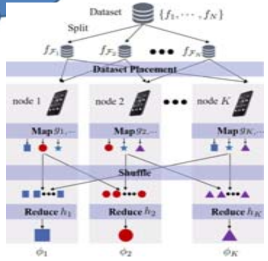


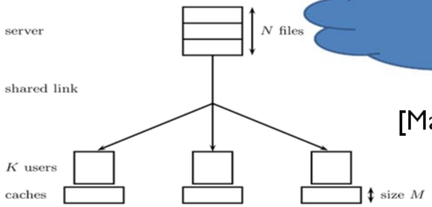
Index Coding

[Birk, Kol, INFOCOM'98]

Distributed Computing

Li-Maddah-Ali-Avestimehr'14





Caching

[Maddah-Ali & Niesen '13]

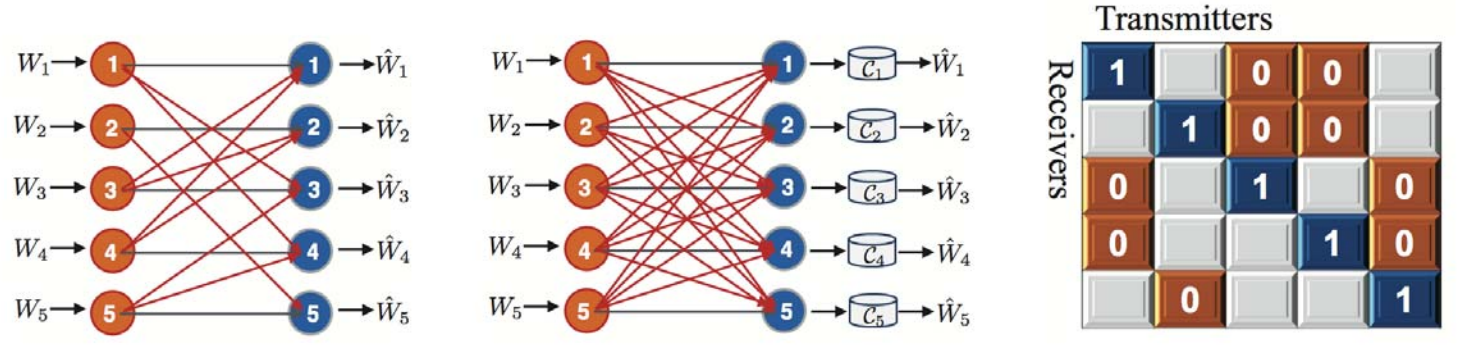
Low-rank model offers a new way to investigate these problems!

Summary: generalized low-rank models

- Generalized low-rank optimization for dense edge networks

$$\underset{\boldsymbol{M}\in\mathbb{C}^{m\times n}}{\text{minimize}} \quad \text{rank}(\boldsymbol{M}) \quad \text{subject to} \quad \boldsymbol{M}\in\mathcal{D}$$

  - $\mathcal{D}$ encodes network side information, e.g., cached files, network topology, computed intermediate values for data shuffling



(a) TIM problem. (b) Cache-aided interference channel. (c) Side information modeling matrix.

Concluding remarks

- **Structured sparse models**
  - Group sparse optimization offers a principled way for network adaptation, e.g., to minimize network power consumption
  - Sparsity control and estimation is powerful to support massive device connectivity

- **Future directions:**
  - More application scenarios: IoTs, V2X …

Concluding remarks

- **Generalized low-rank models**
  - Low-rank matrix completion provides a systematic approach to investigate the topological interference alignment problem
  - Low-rank model is powerful for performance optimization in mobile edge caching and distributed computing systems

- **Future directions:**
  - More applications: blind deconvolution for IoT, big data analytics (e.g., ranking)

To learn more...

- **Web:** http://shiyuanming.github.io/sparserank.html
- **Papers:**
- Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809-2823, May 2014. (The 2016 Marconi Prize Paper Award)
- Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729-4743, Sept. 2015. t. 2015. (The 2016 IEEE Signal Processing Society Young Author Best Paper Award)
- Y. Shi, J. Zhang, K. B. Letaief, B. Bai and W. Chen,"Large-scale convex optimization for ultra-dense Cloud-RAN," *IEEE Wireless Commun. Mag.*, pp. 84-91, Jun. 2015.
- Y. Shi, J. Zhang, W. Chen, and K. B. Letaief, "Generalized sparse and low-rank optimization for ultra-dense networks," *IEEE Commun. Mag.*, to appear.

To learn more...

- Y. Shi, J. Zhang, and K. B. Letaief, "Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty," *IEEE Trans. Signal Process.*, vol. 63,, no. 4, pp. 960-973, Feb. 2015.
- Y. Shi, J. Zhang, and K. B. Letaief, "Robust group sparse beamforming for multicast green Cloud-RAN with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4647-4659, Sept. 2015.
- Y. Shi, J. Cheng, J. Zhang, B. Bai, W. Chen and K. B. Letaief, "Smoothed L_p-minimization for green Cloud-RAN with user admission control," *IEEE J. Select. Areas Commun.*, vol. 34, no. 4, Apr. 2016.
- Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion for topological interference management by Riemannian pursuit," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, Jul. 2016.
- Y. Shi, B. Mishra, and W. Chen, "Topological interference management with user admission control via Riemannian optimization," *IEEE Trans. Wireless Commun.*, to appear.
- X. Peng, Y. Shi, J. Zhang, and K. B. Letaief, "Layered group sparse beamforming for cache-enabled wireless networks," *IEEE Trans. Commun.*, to appear.