SLOW RESOURCE ALLOCATION FOR HETEROGENEOUS NETWORKS

Michael Honig
Department of EECS
Northwestern University

Joint work with Binnan Zhuang, Dongning Guo

Mobile Data Forecast

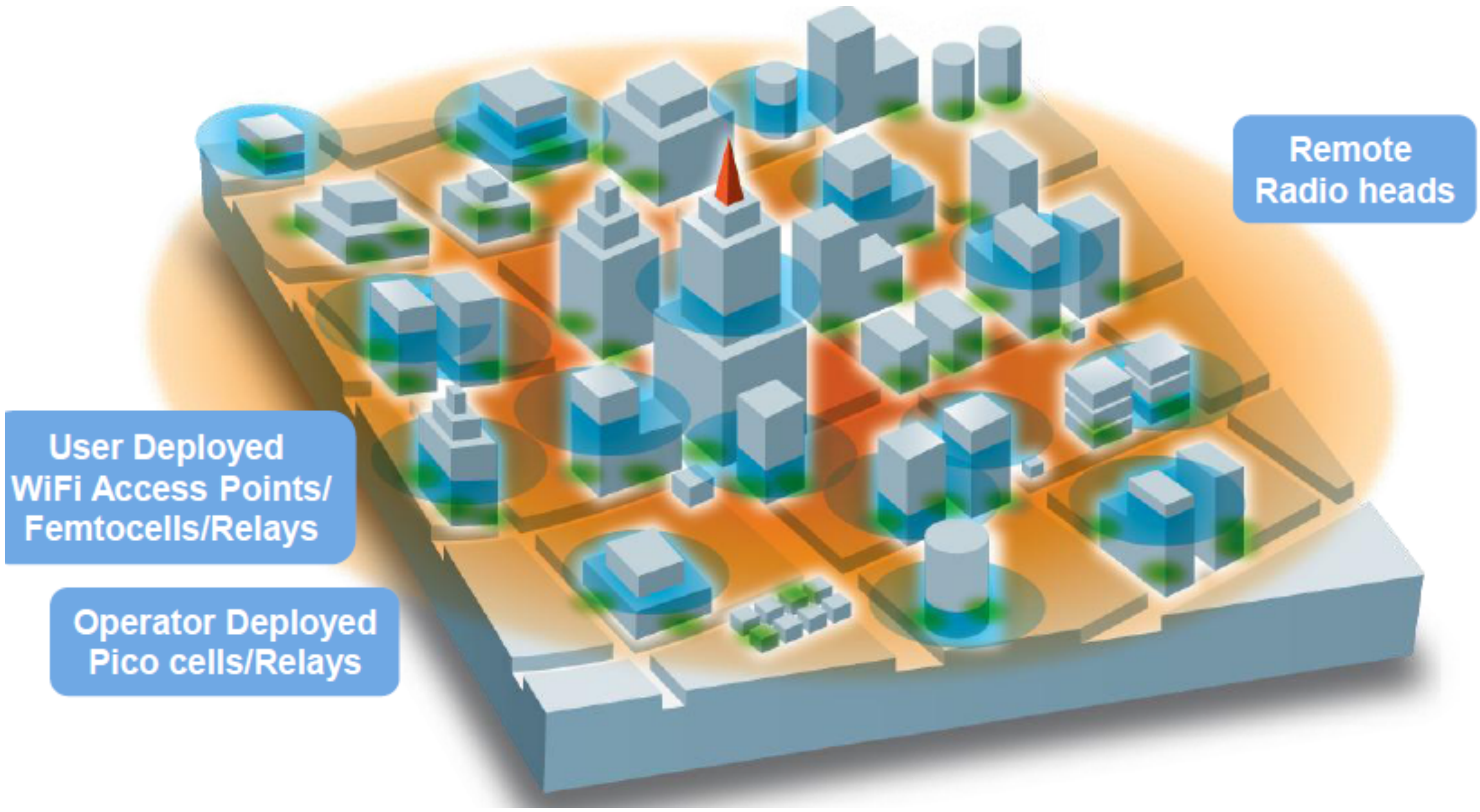
Exa: 10^{18}

Source: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper

Evolution to 5G

- New spectrum (mm-wave, unlicensed)
- Physical layer advances
  - massive MIMO, network coding, cooperation
- Smaller cells

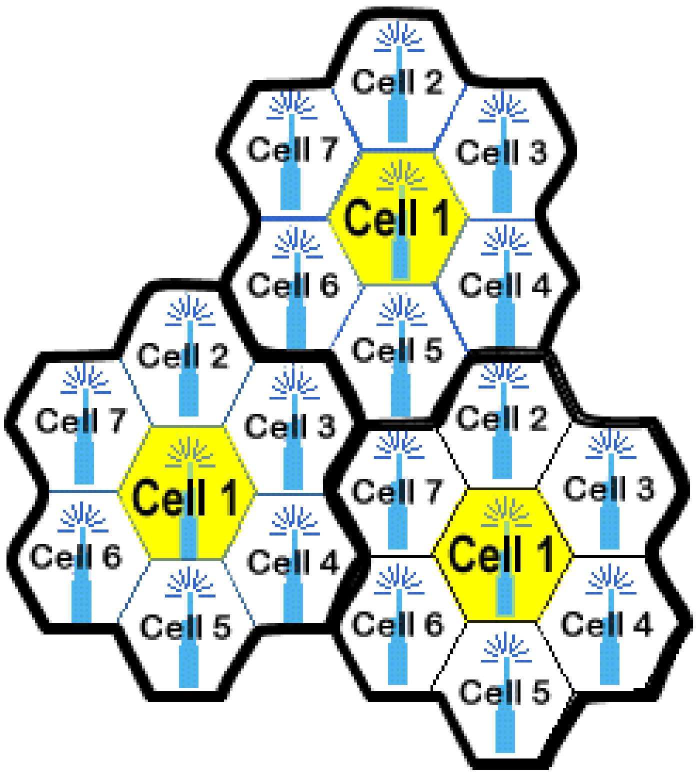
Heterogeneous Network

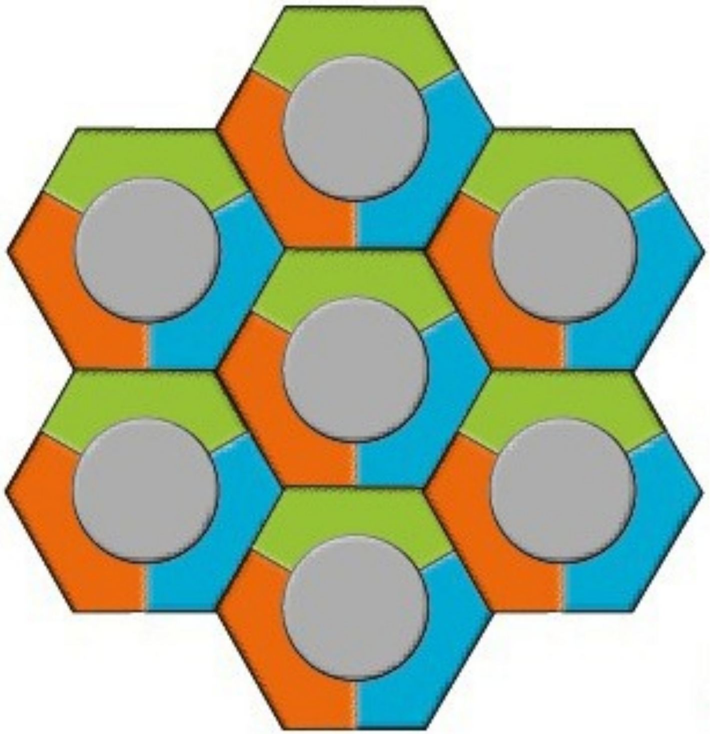


Remote Radio Heads

How to do interference management?

Offline Frequency Planning (1G-4G)



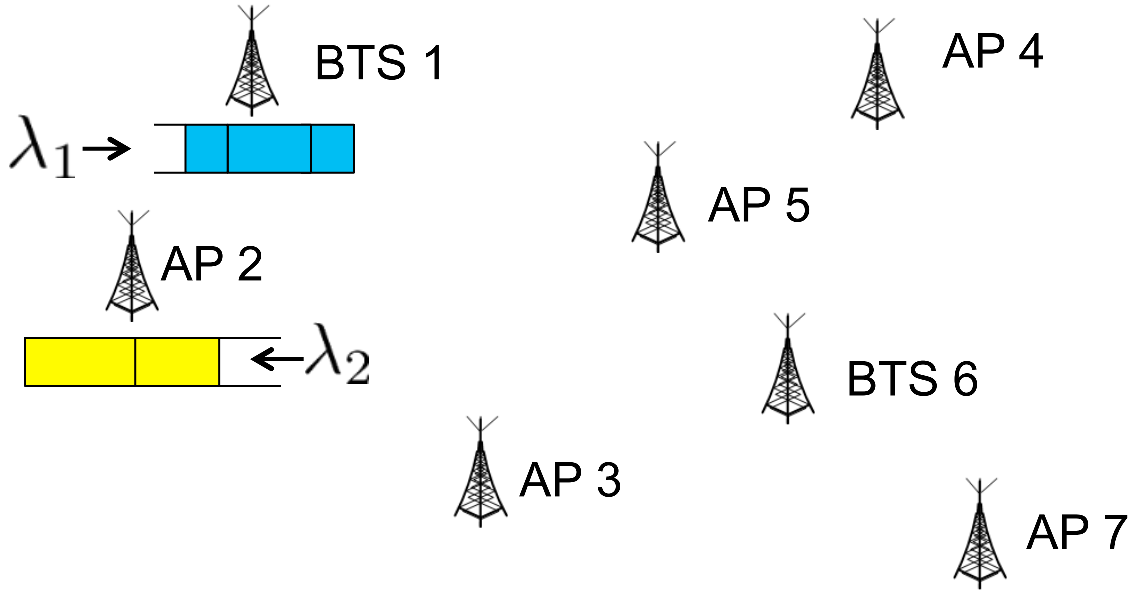


Slow Resource Allocation

- Over many packets (seconds)
  - Average channel gains, offered traffic
- Combined with fast scheduling (milliseconds)
- Traffic varies over space, stationary in time
- Centralized approach

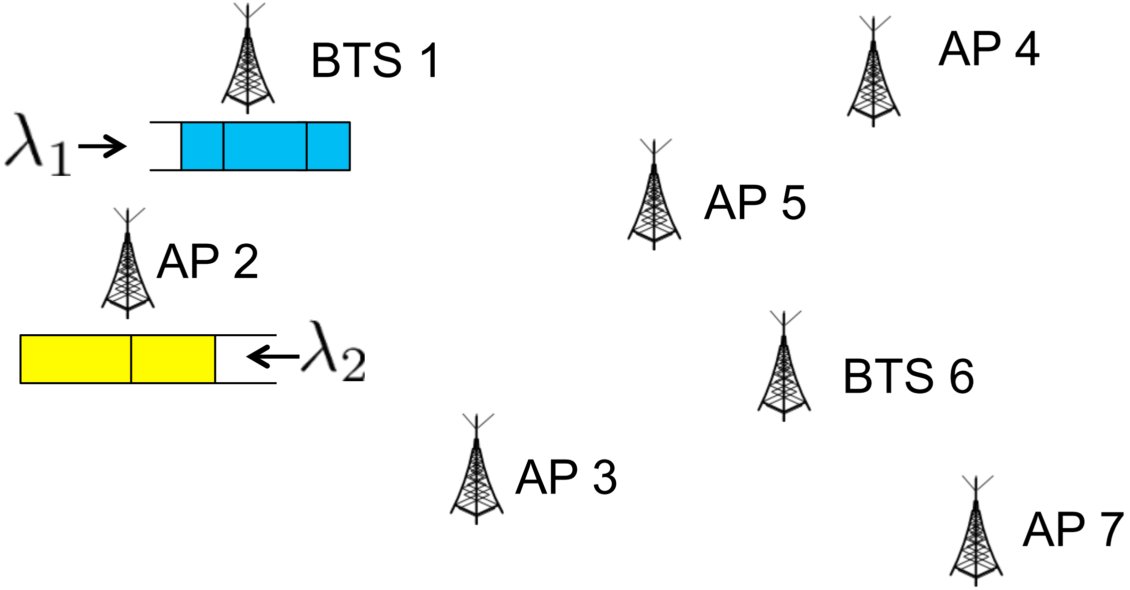
Contribution: general optimization framework

Downlink HetNet Model



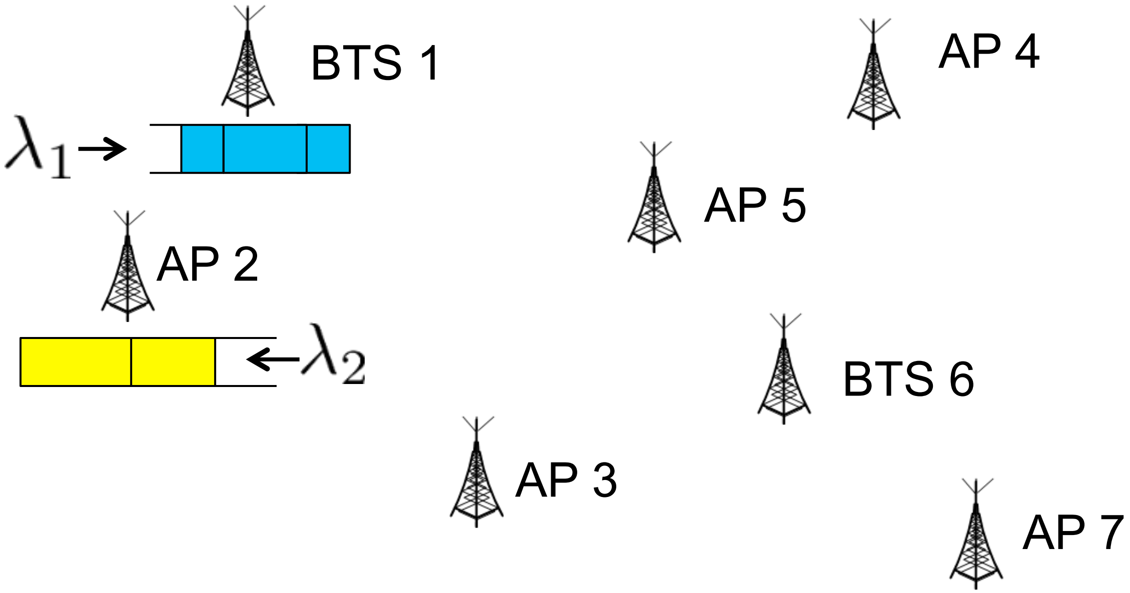
- Cells overlap, traffic varies.
- How to allocate spectrum across cells?

Assumptions



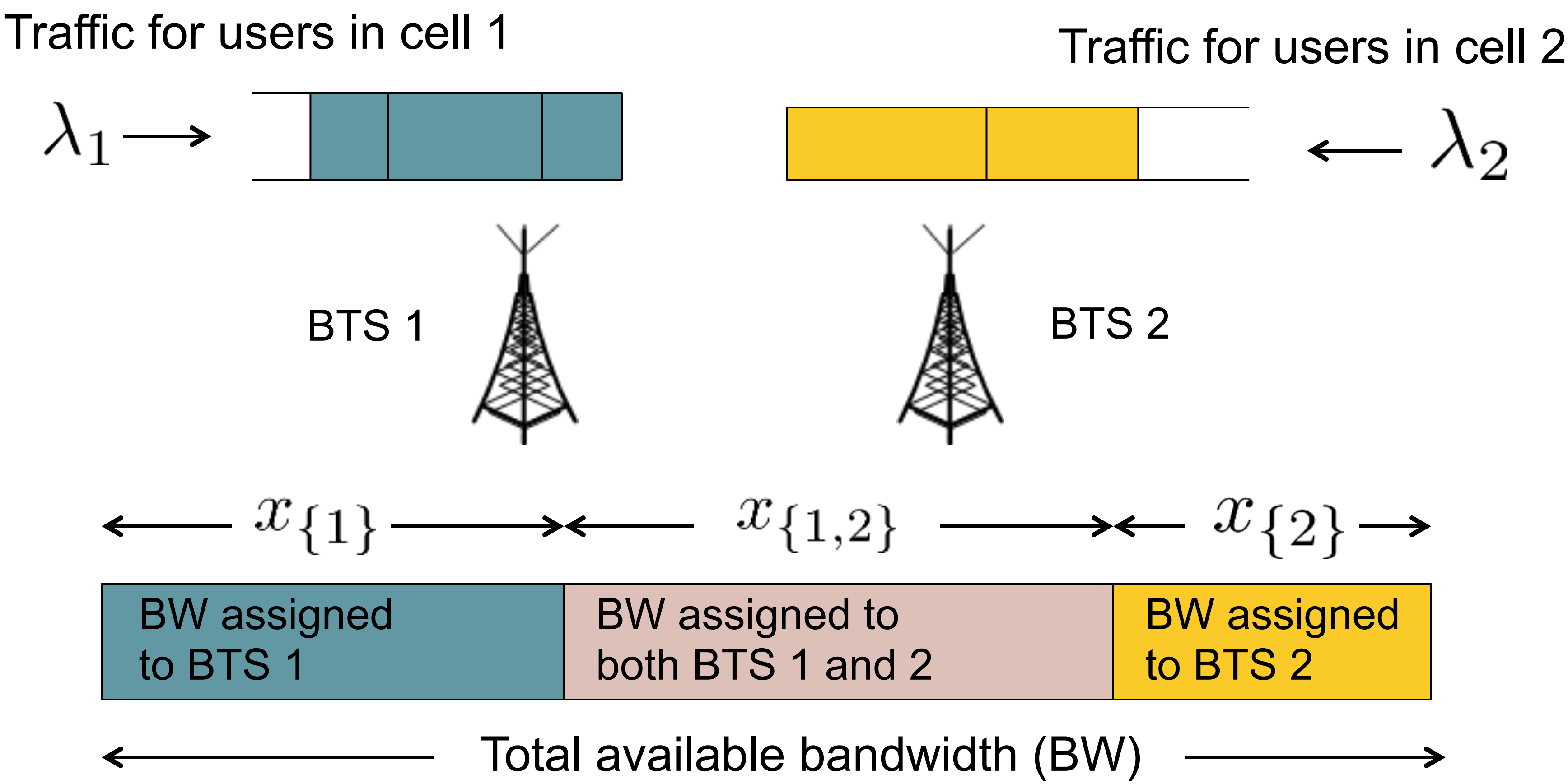
- Resources within each cell are allocated via fast scheduling.
- Resources across cells are allocated over a slower time-scale.
- Centralized controller knows average traffic, average channels.

Traffic-Driven Resource Allocation

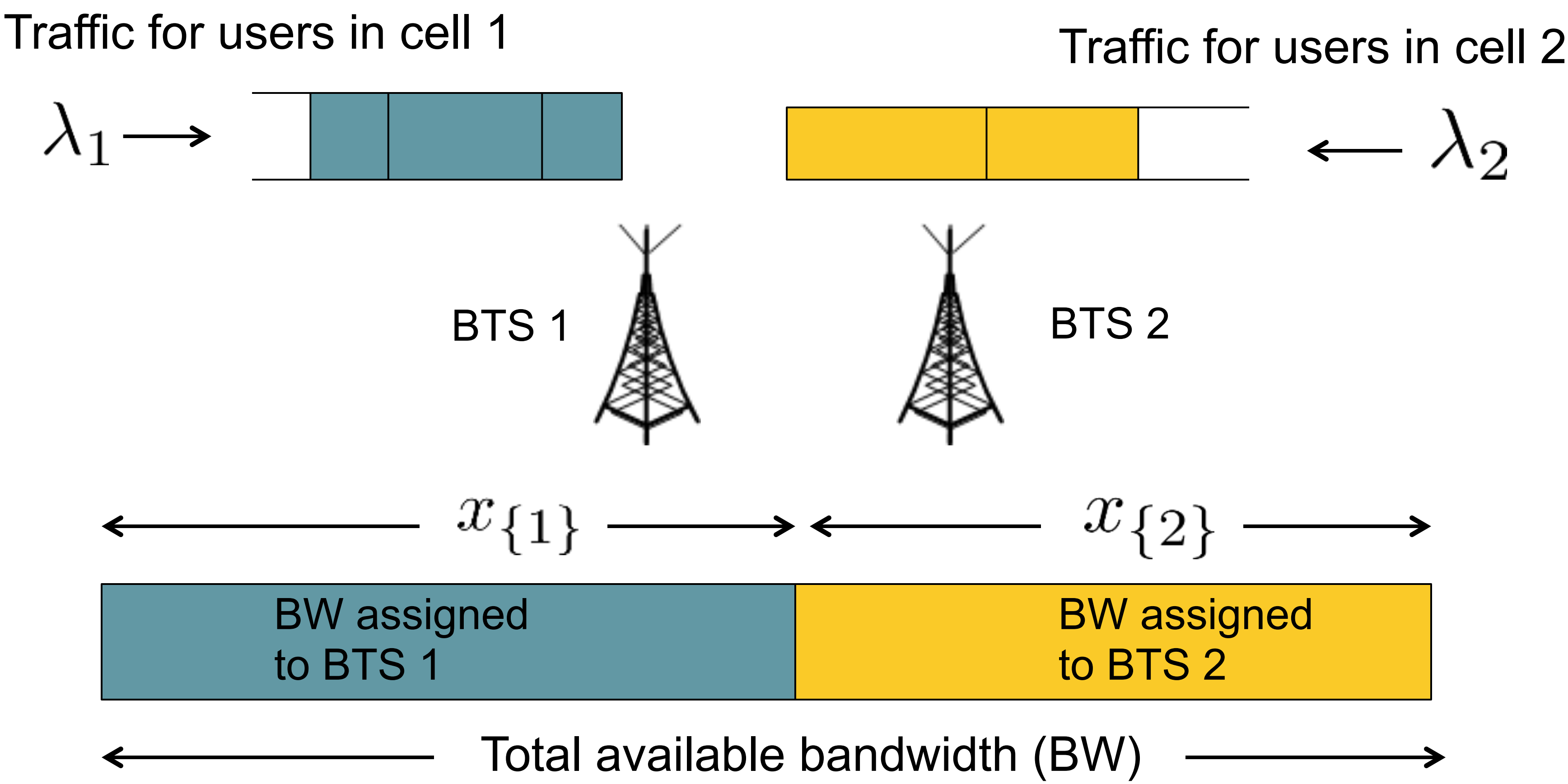


- Consider all possible ways the spectrum can be partitioned among BTS's.
- Optimize over this partition.

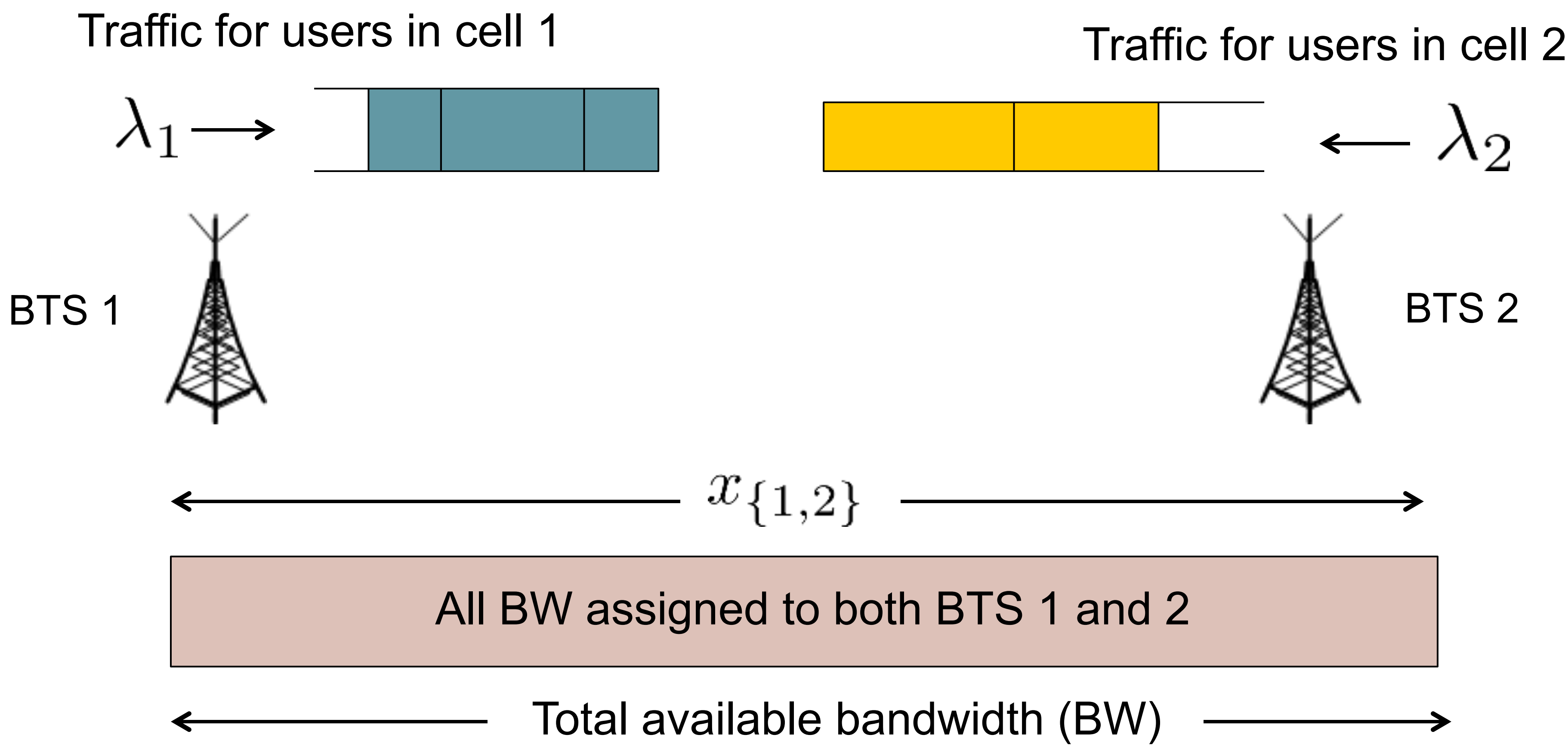
Two Base Stations



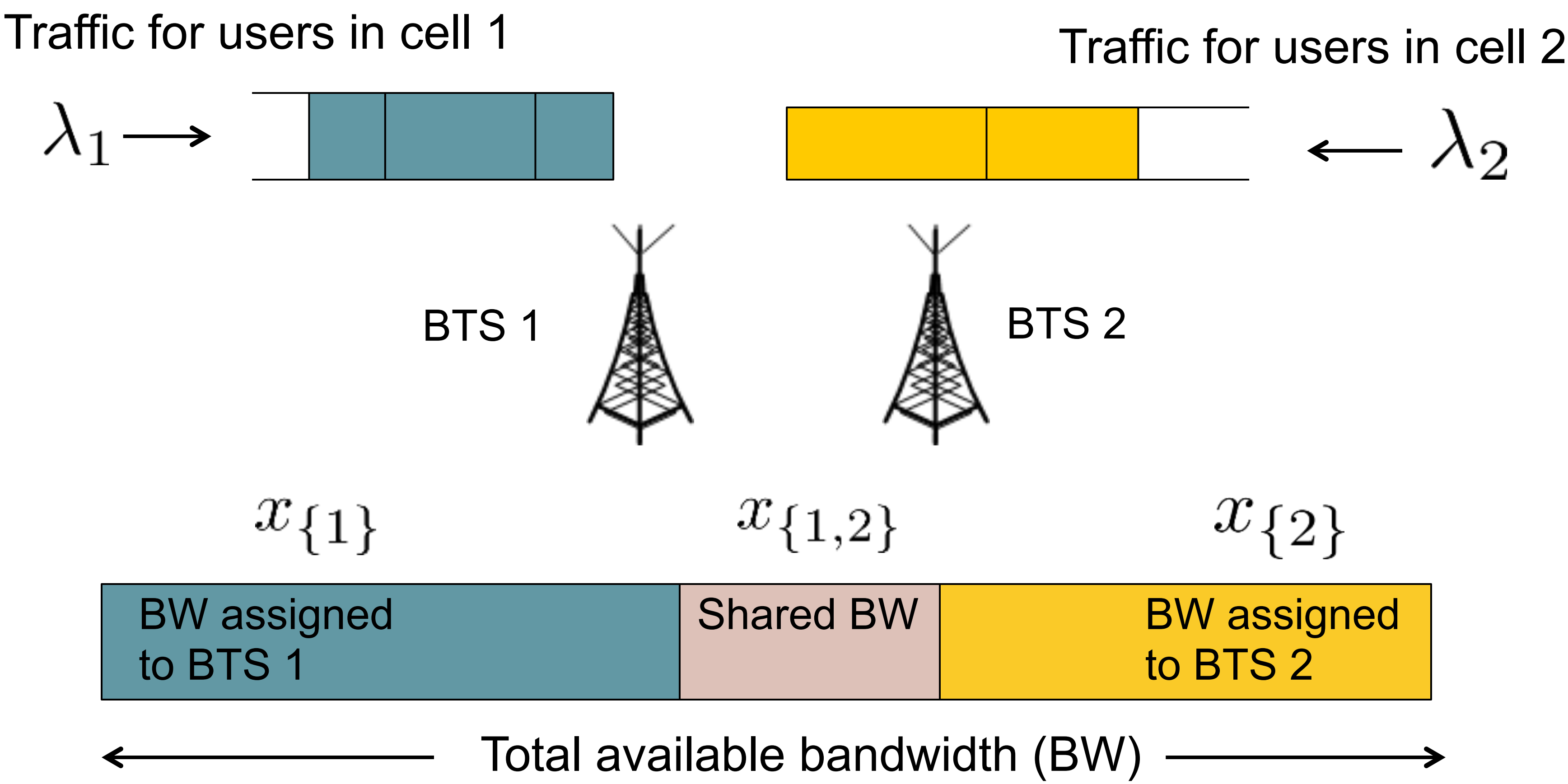
Orthogonal Allocation



Full Frequency Reuse



Partial Sharing



Orthogonal Allocation

Traffic for users in cell 1

Traffic for users in cell 2

λ_1 → $\leftarrow \lambda_2$

“I would build a GREAT wall!”

$\longleftarrow x_{\{1\}} \longrightarrow$ $\longleftarrow x_{\{2\}} \longrightarrow$

BW assigned to BTS 1	BW assigned to BTS 2

$\longleftarrow$ Total available bandwidth (BW) $\longrightarrow$

Full Frequency Reuse

Traffic for users in cell 1

Traffic for users in cell 2

$\lambda_1 \rightarrow$ $\leftarrow \lambda_2$

“Tear down this wall!”

$\longleftarrow \quad x_{\{1,2\}} \quad \longrightarrow$

All BW assigned to both BTS 1 and 2

$\longleftarrow$ Total available bandwidth (BW) $\longrightarrow$

Partial Sharing

Traffic for use

$\lambda_1 \rightarrow$

Traffic for users in cell 2

$\leftarrow \lambda_2$

"one country, two systems"

$x_{\{1\}}$ $x_{\{1,2\}}$ $x_{\{2\}}$

BW assigned to BTS 1	Shared BW	BW assigned to BTS 2

← Total available bandwidth (BW) →

Partial Sharing

Traffic for use

Traffic for users in cell 2

$\lambda_1 \longrightarrow$

$\longleftarrow \lambda_2$

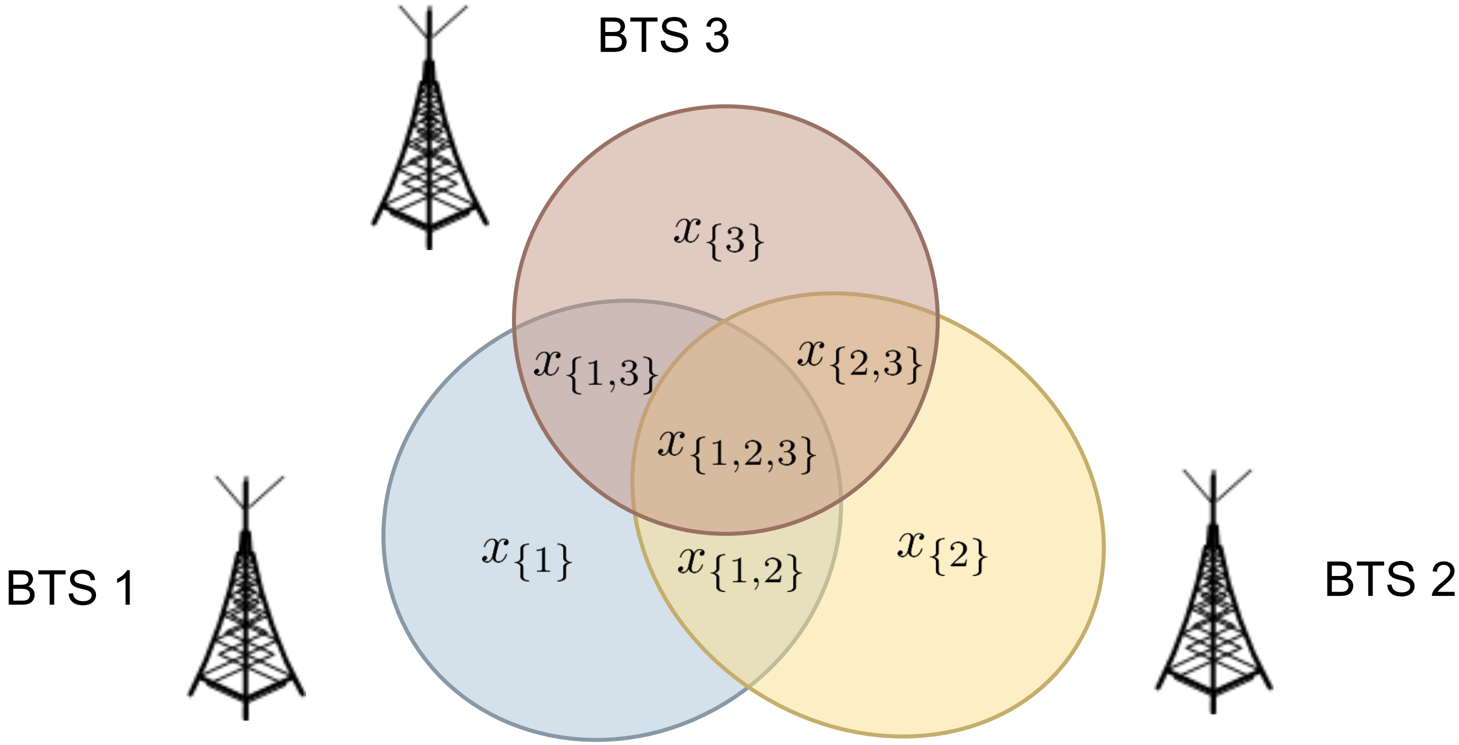
"one country, two systems"

$x_{\{1\}}$ $x_{\{1,2\}}$ $x_{\{2\}}$

BW assigned to BTS 1	Shared BW	BW assigned to BTS 2

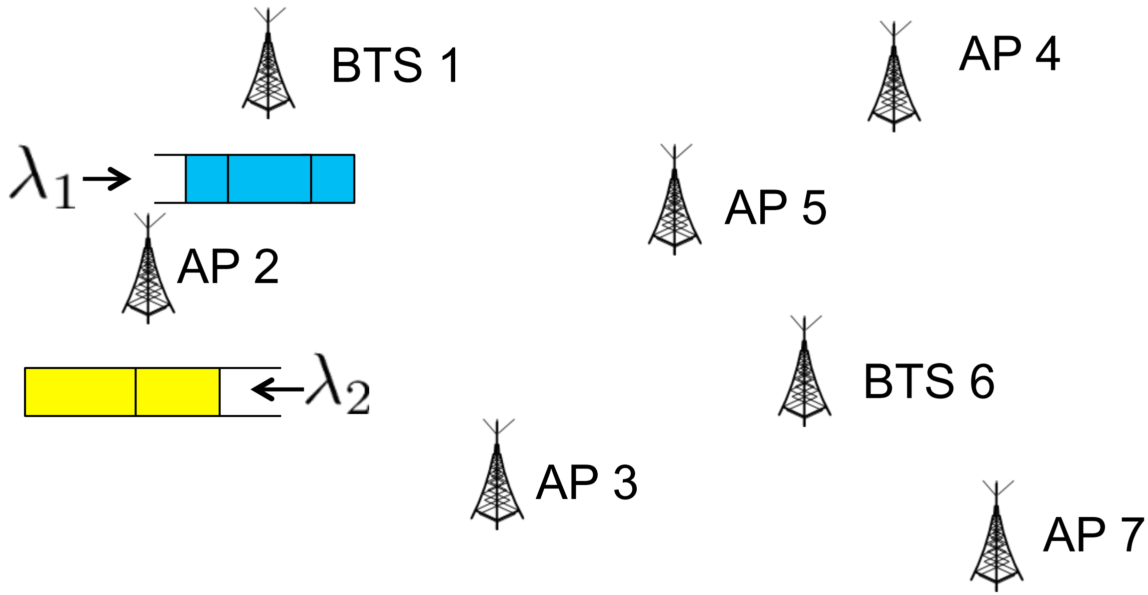
Partition should depend on traffic!

3-BTS Example

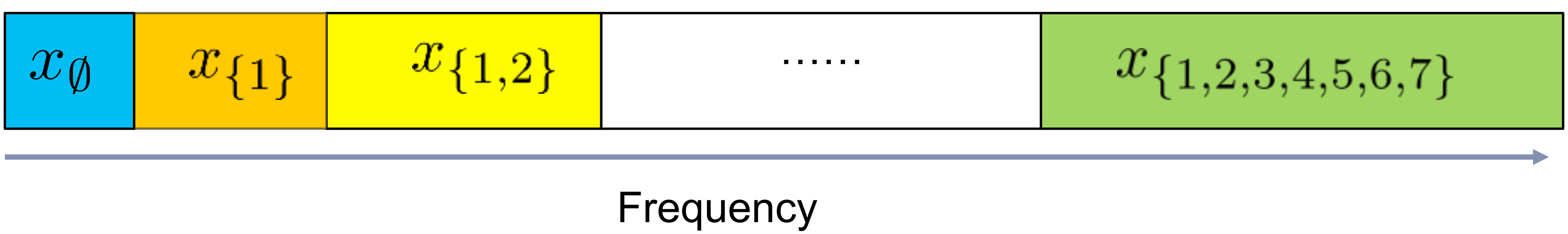


$x_{\{1\}}$	$x_{\{2\}}$	$x_{\{3\}}$	$x_{\{1,2\}}$	$x_{\{1,3\}}$	$x_{\{2,3\}}$	$x_{\{1,2,3\}}$

K Base Stations

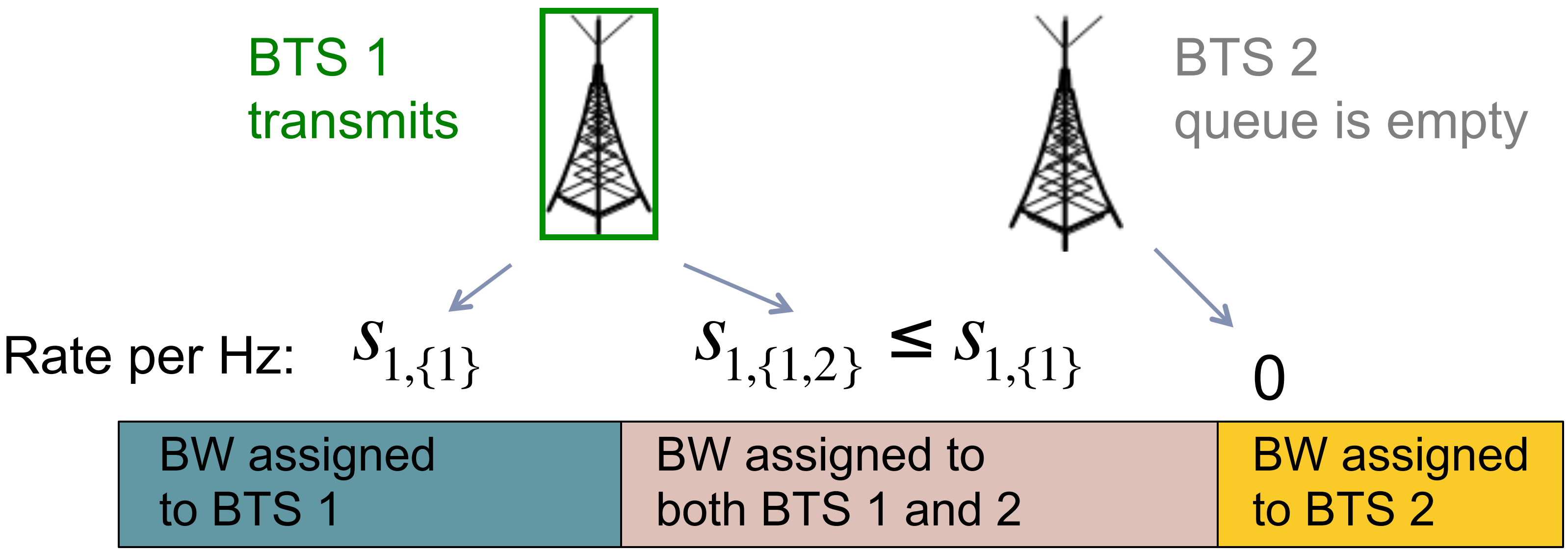


spectrum allocation: 2^k reuse patterns (variables)



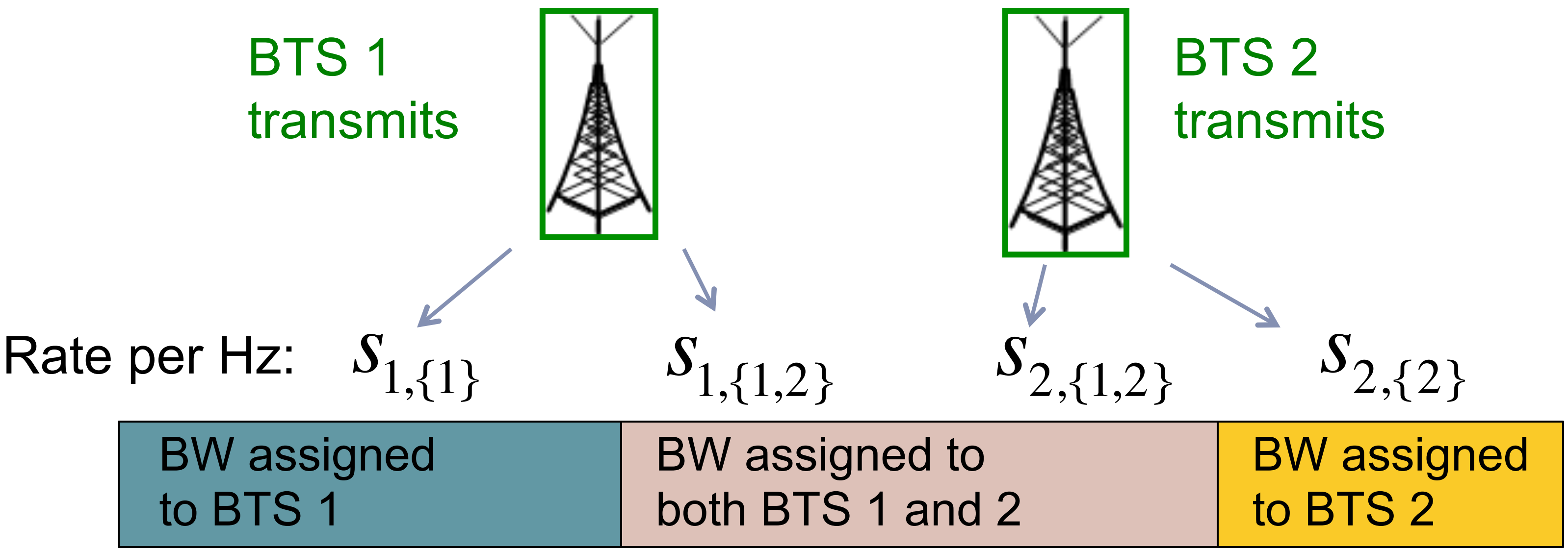
Bandwidth Optimization Problem

- Adjust partition to minimize average latency
- Take into account queuing delays and interference
- Interference affects achievable rates



Bandwidth Optimization Problem

- Adjust partition to minimize average latency
- Take into account queuing delays and interference
- Interference affects achievable rates

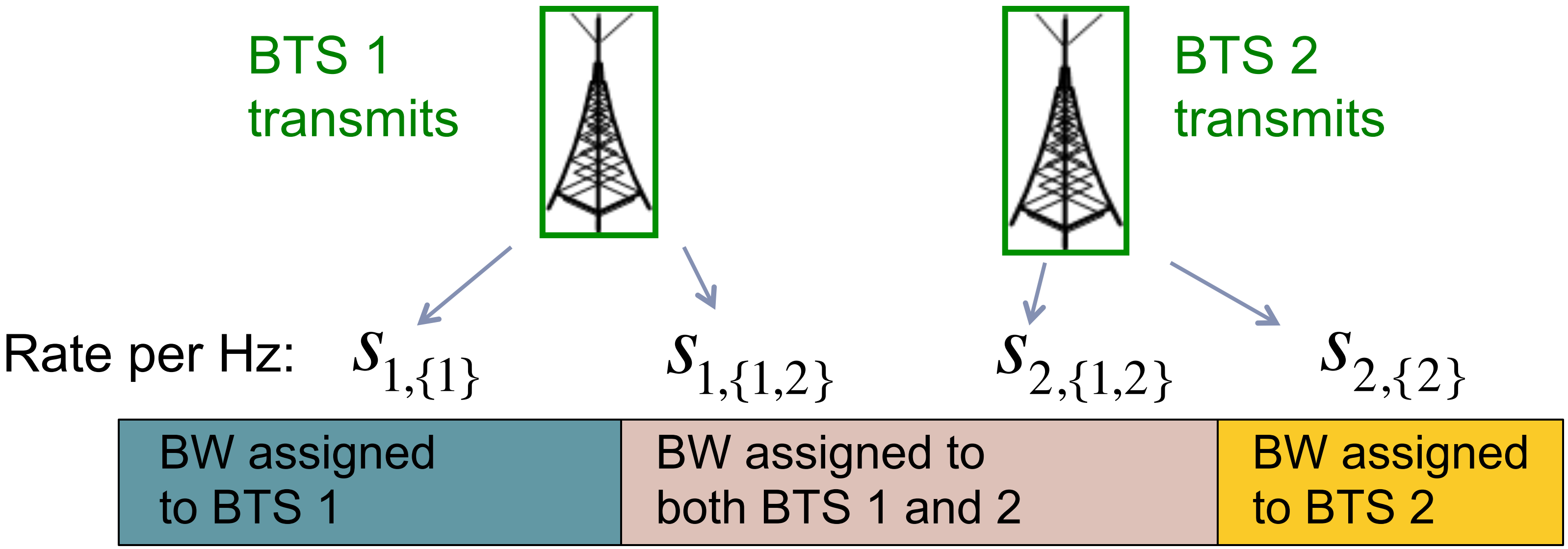


Spectral Efficiency

$$s_A^{i\to j} = \mathbf{1}_{i\in A}\frac{W}{L}\log\left(1+\frac{p^{i\to j}}{I^{A\to j}+\sigma^2}\right)$$

- Average powers, channels
- Known to the optimizer

Bandwidth Optimization Problem

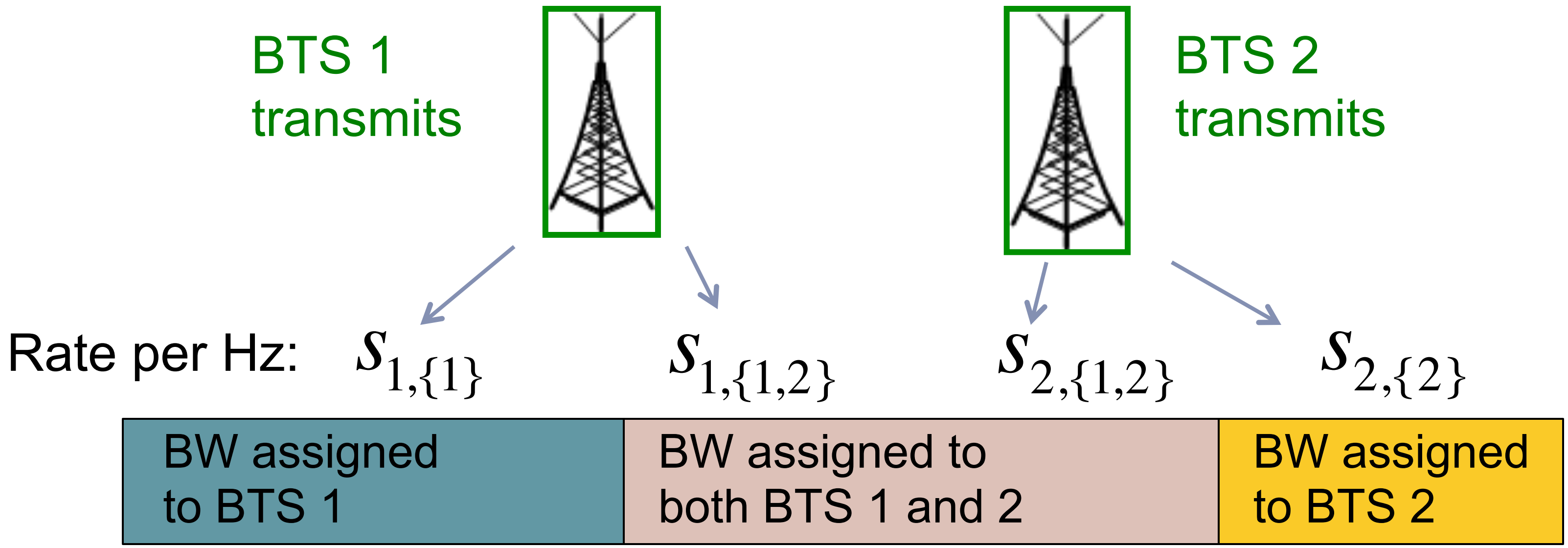


Total rates:

$$r_1 = s_{1,\{1\}} x_{\{1\}} + s_{1,\{1,2\}} x_{\{1,2\}}$$

$$r_2 = s_{2,\{2\}} x_{\{2\}} + s_{2,\{1,2\}} x_{\{1,2\}}$$

Bandwidth Optimization Problem



Total rate from BTS *i*:

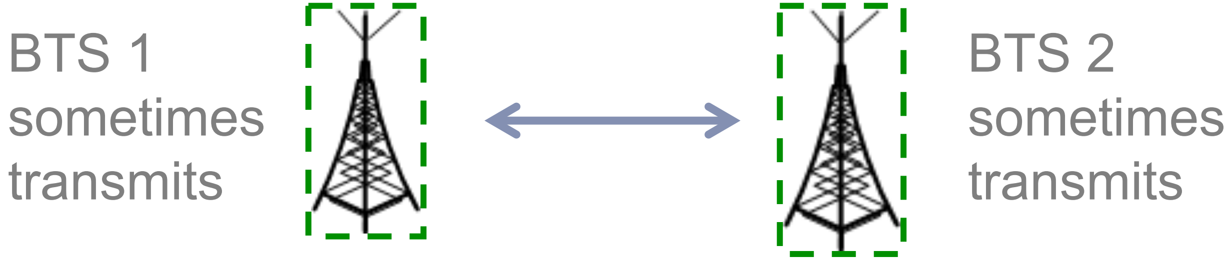
$$r_i = \sum_{B \subset \mathcal{N}} s_{i,B} x_B$$

← sum over all reuse patterns

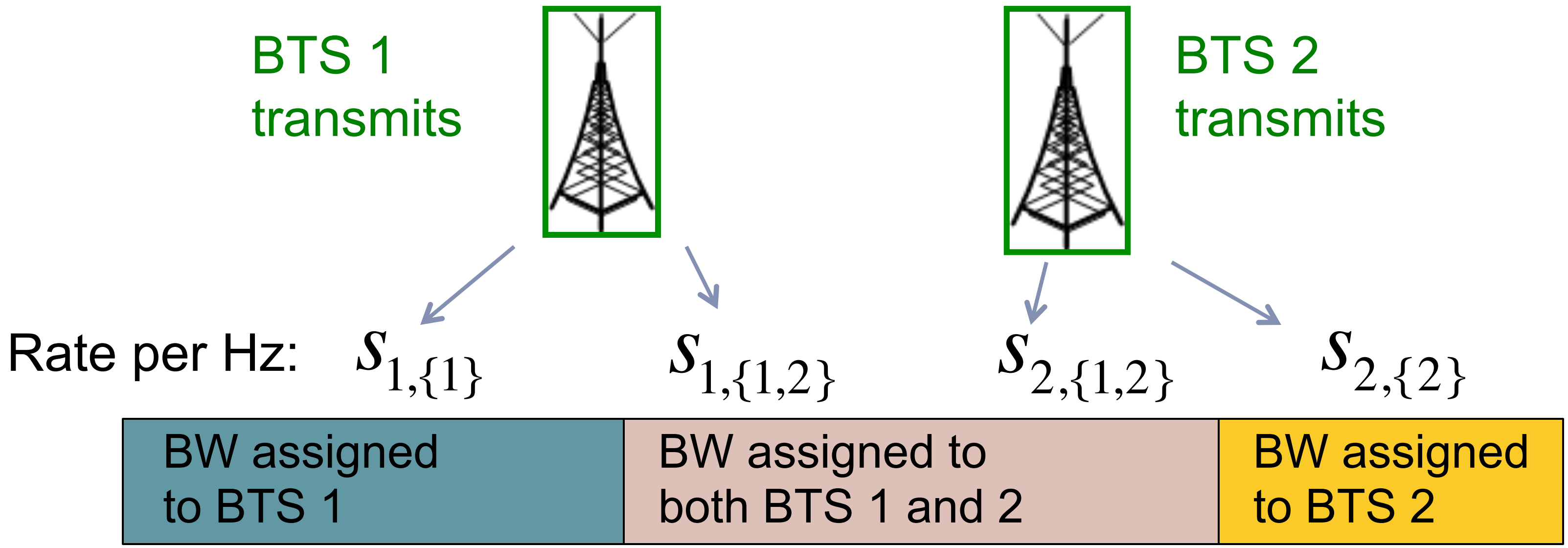
$\mathcal{N} = \{1, 2, \cdots, N\}$ set of BTSs

Bandwidth Optimization Problem

- Adjust partition to minimize average latency
- Take into account queuing delays and interference
- Interference affects achievable rates
- Queues at different BTS's are dependent – complicates optimization!



Backlogged Traffic: Delay



Average packet sojourn time (M/M/1): $t_i = \dfrac{1}{r_i - \lambda_i}$

Conservative Optimization

$$\min_{\{\mathbf{x},\mathbf{r}\}} \sum_{i=1}^{N} \left(\frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i} \right) \frac{1}{r_i - \lambda_i}$$

Subject to:

$$r_i > \lambda_i$$

$$r_i = \sum_{B \subset \mathcal{N}} s_{i,B} x_B \quad \forall i \in \mathcal{N}$$

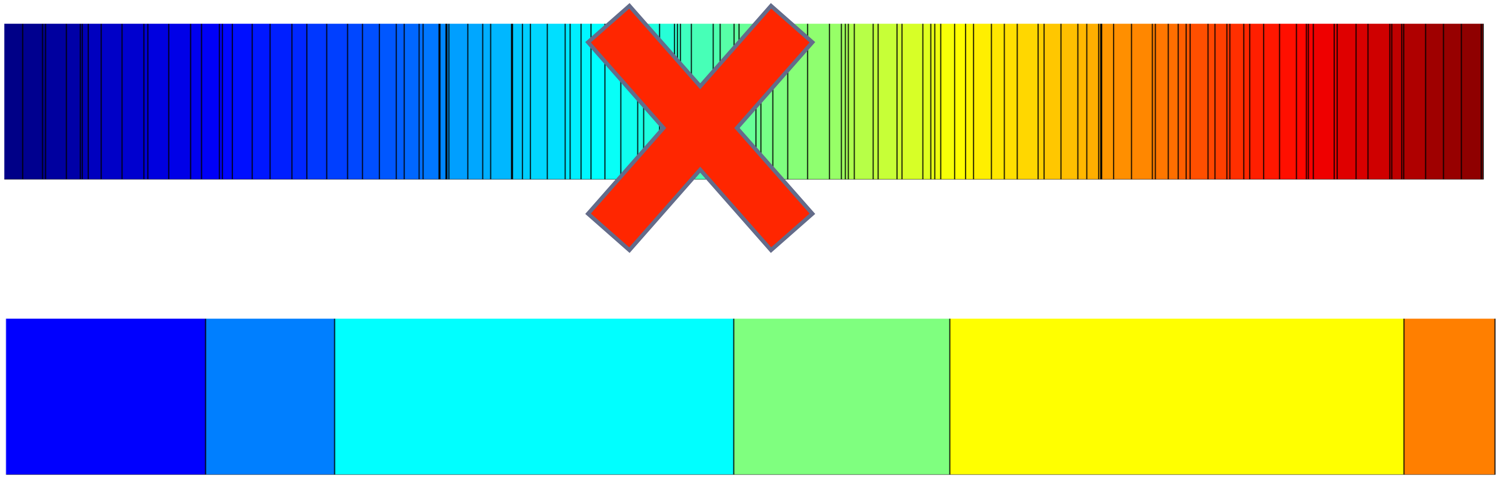
$$x_B \geq 0 \quad \forall B \subset \mathcal{N}$$

$$\sum_{B \subset \mathcal{N}} x_B = 1$$

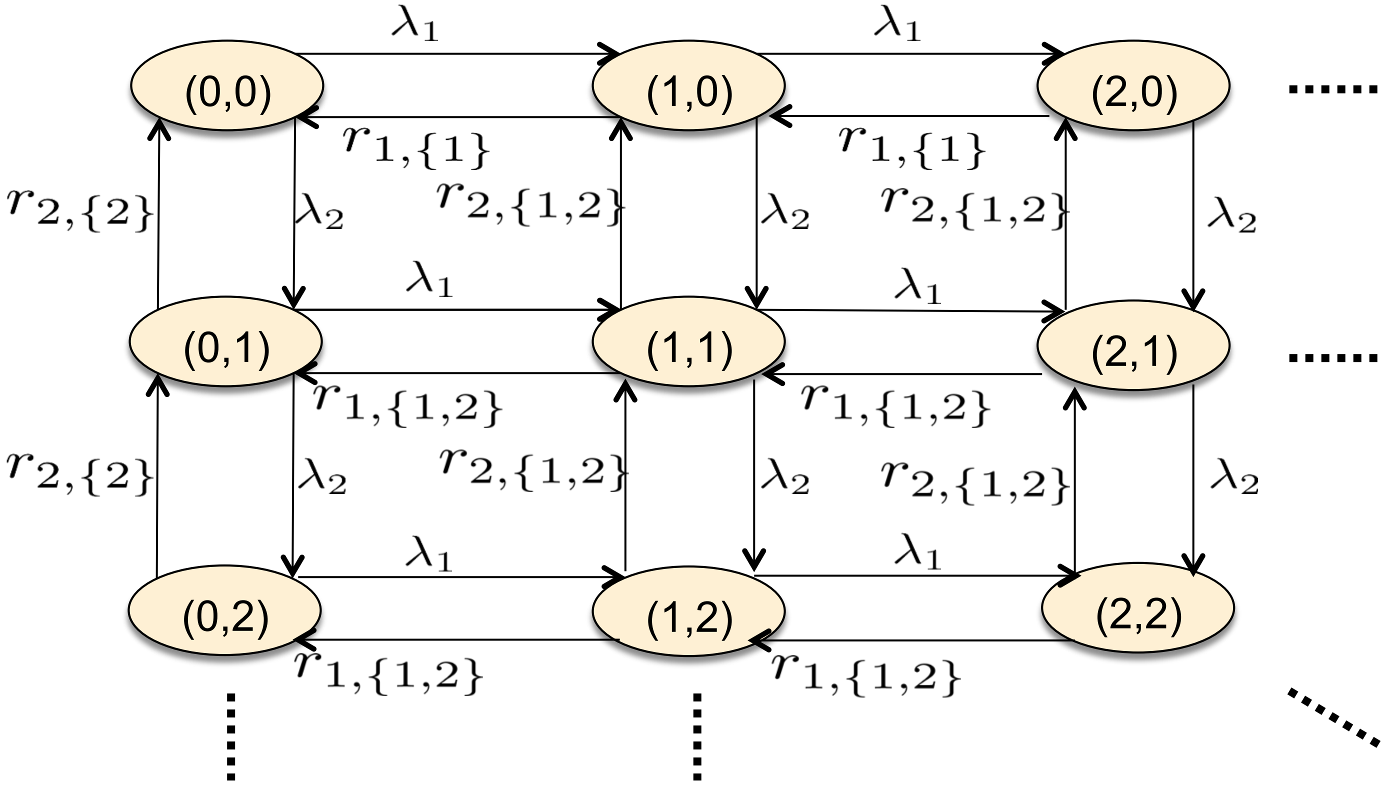
- Convex, 2^N-1 variables
- The solution achieves the maximum throughput region.

Property of Solution

- Theorem: The optimal allocation divides the spectrum into at most N segments (instead of 2^N).
  - Follows from Carathéodory's theorem.
- 7-BTS example:

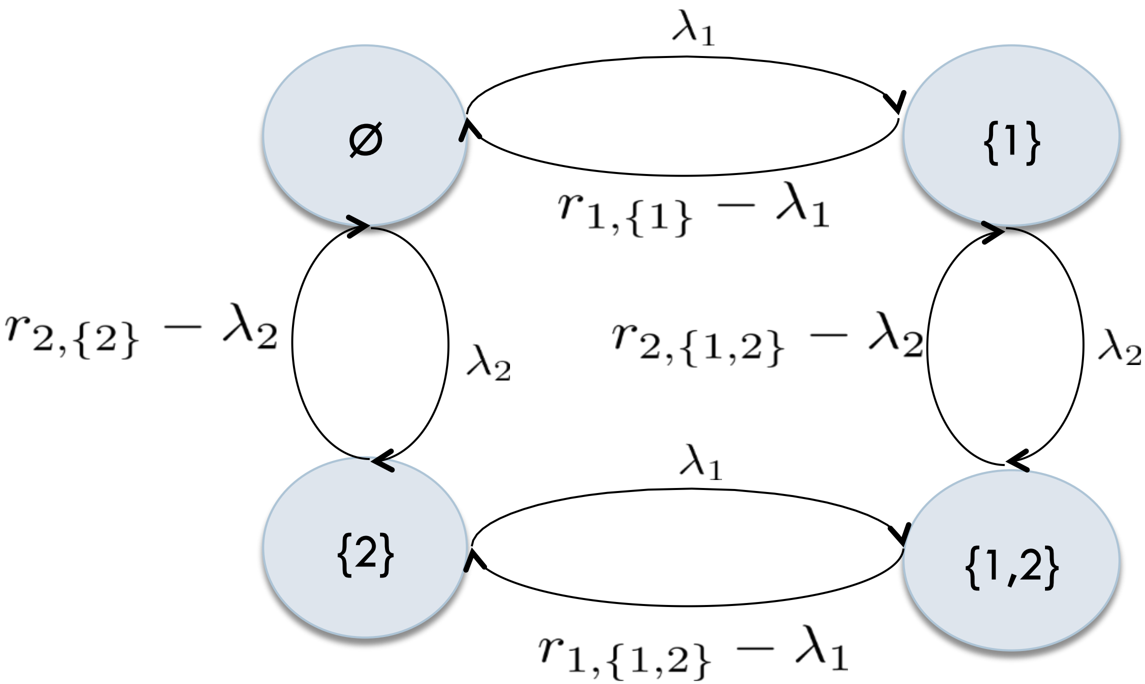


Interactive Queues (Two BTSs)



State Aggregation

Aggregate states with the same set of active BTSs.



Assumptions:

- The N queues are independent conditioned on the pattern.
- For the transition $A \rightarrow A'$, the new state is chosen according to the steady-state distribution.

Refined Optimization

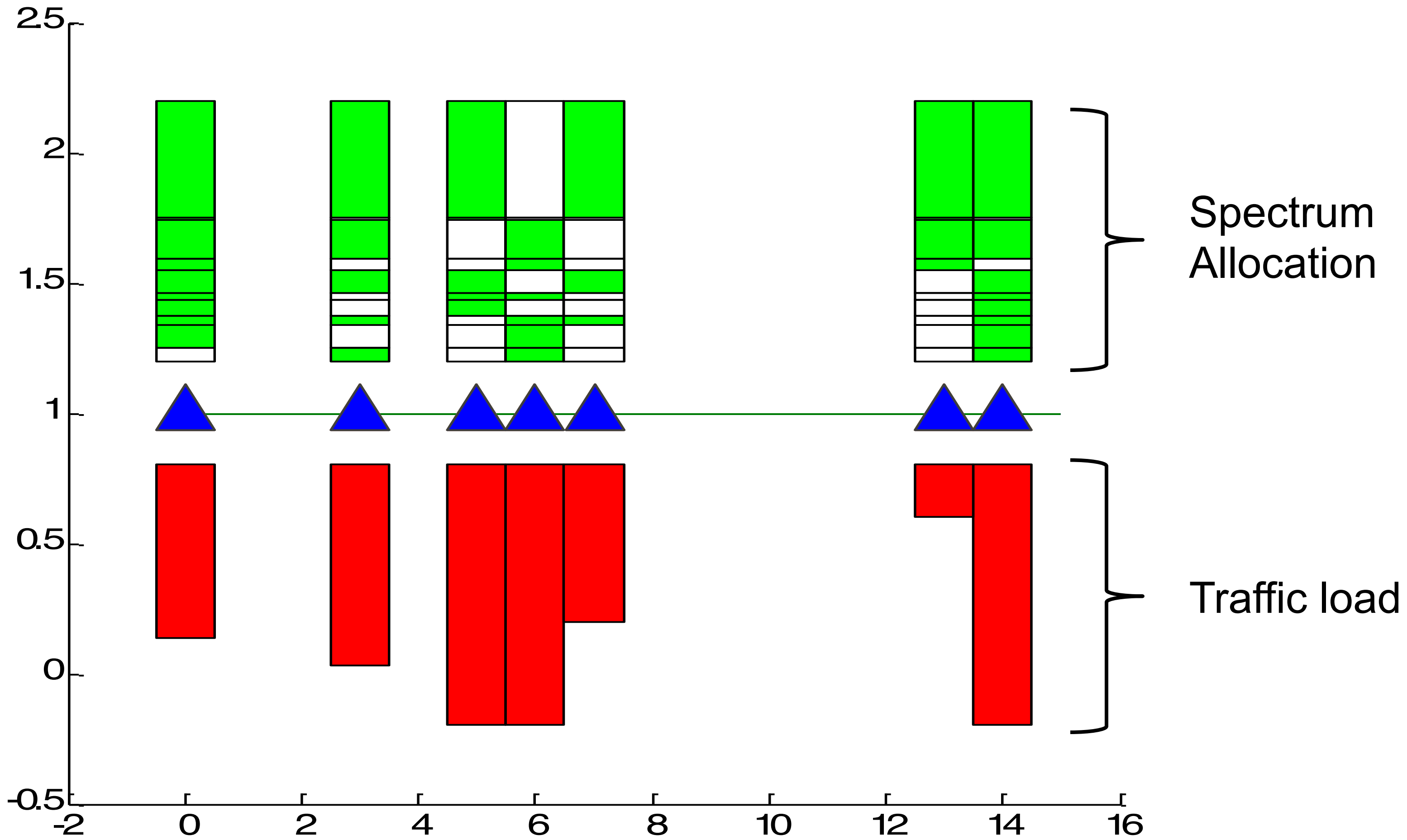
$$\min_{\{\mathbf{x},\mathbf{r},\mathbf{t}\}} \sum_{i=1}^{N} \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} t_i$$

Subject to: $$t_i = \sum_{A \ni i} \frac{p(A) r_{i,A}}{\lambda_i (r_{i,A} - \lambda_i)}$$

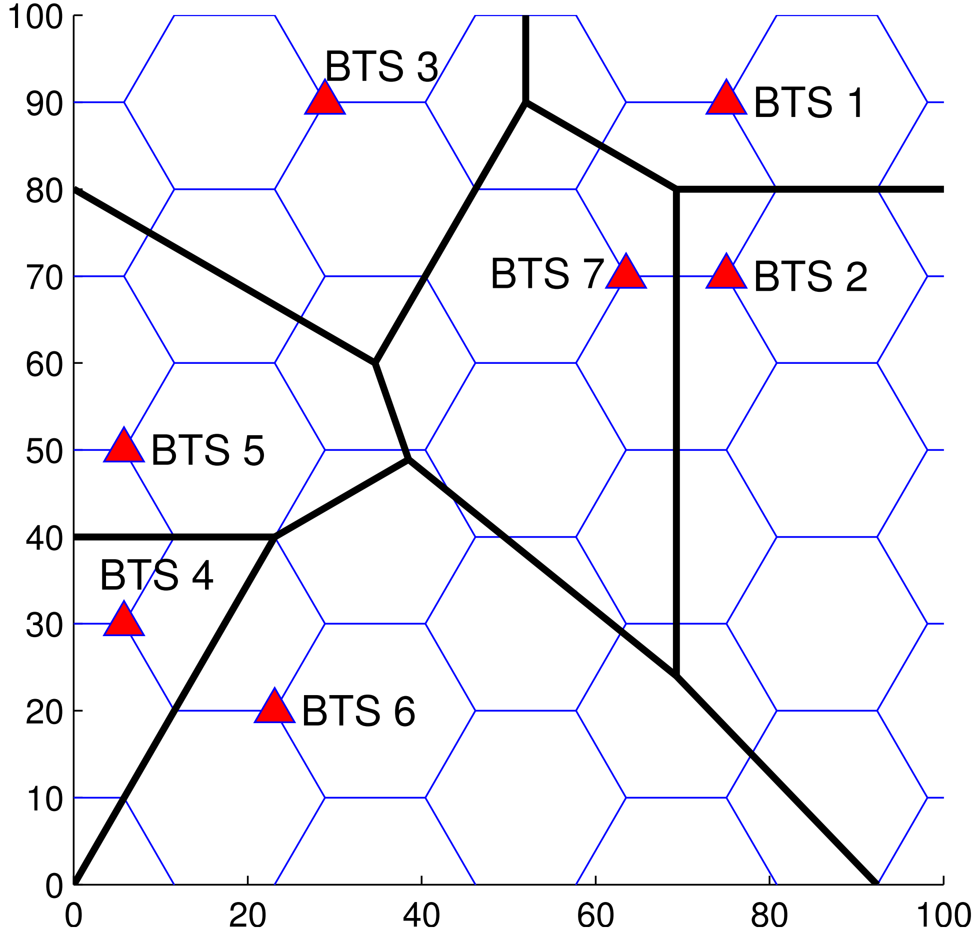
$$r_{i,A} = \sum_{B \subset \mathcal{N}} s_{i,B\cap A} x_B, \quad r_{i,N} > \lambda_i$$

- Not convex
- The solution achieves the maximum throughput region.

One-Dimensional Example

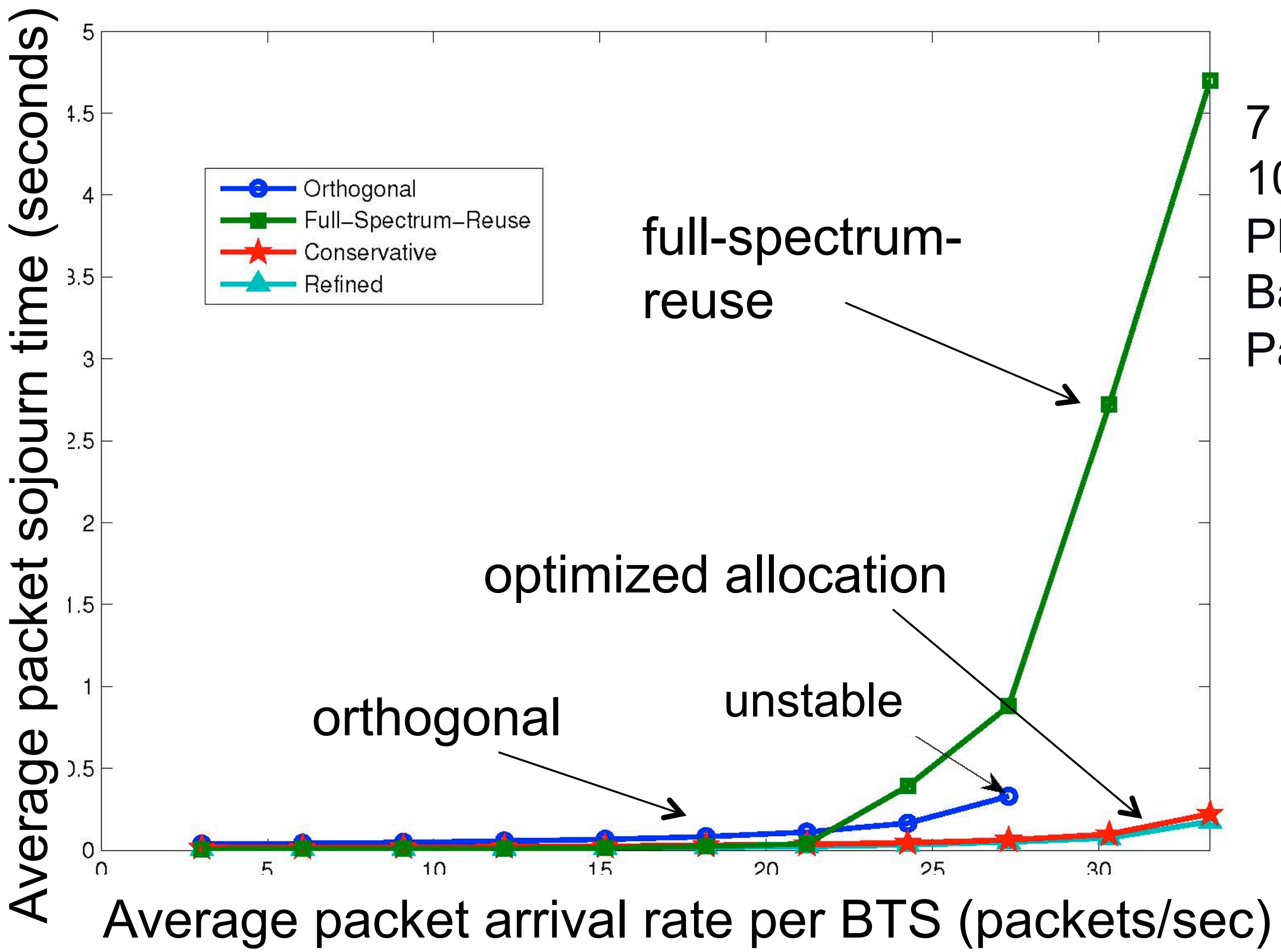


Small-Cell Network



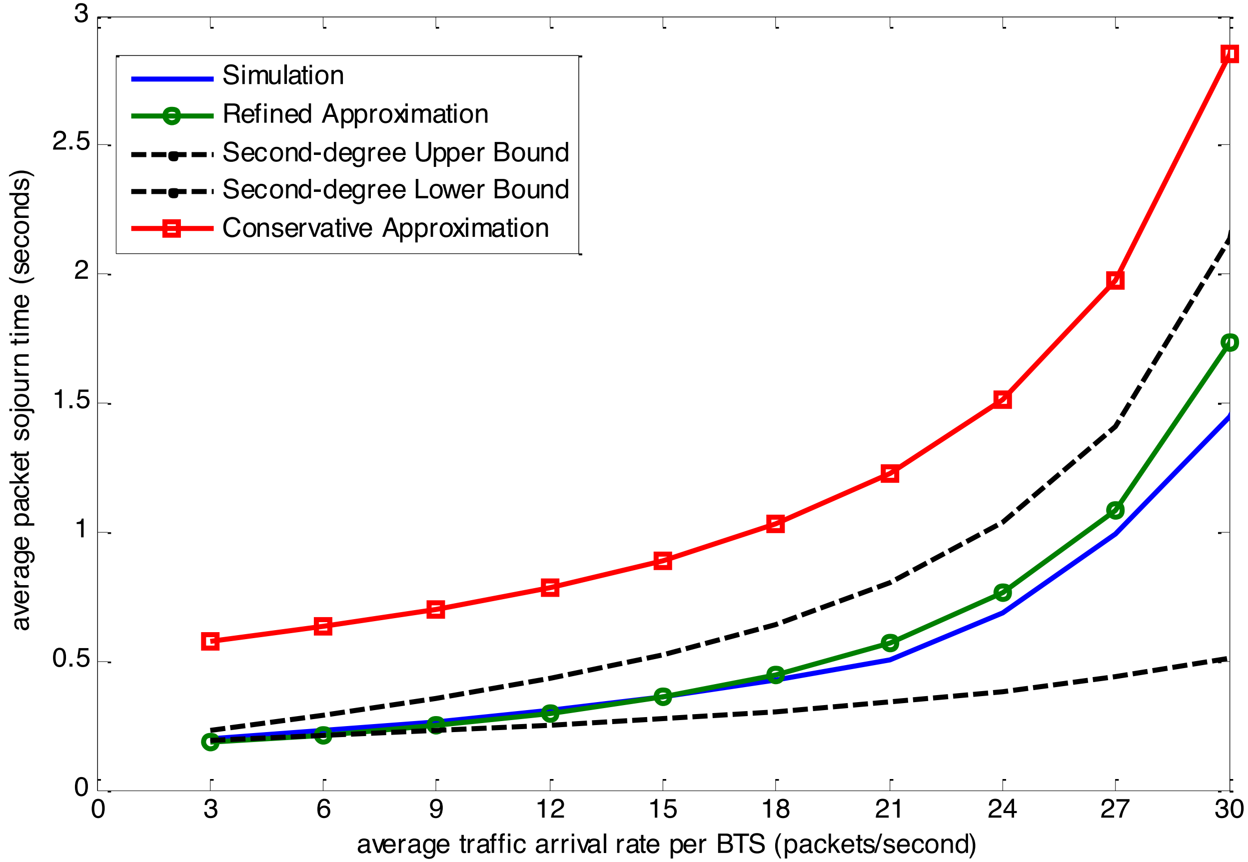
7 BTS's
100x100 m^2
Pkt. length 1MB
Bandwidth 160MHz
Pathloss exponent 3

Delay vs. Traffic Intensity

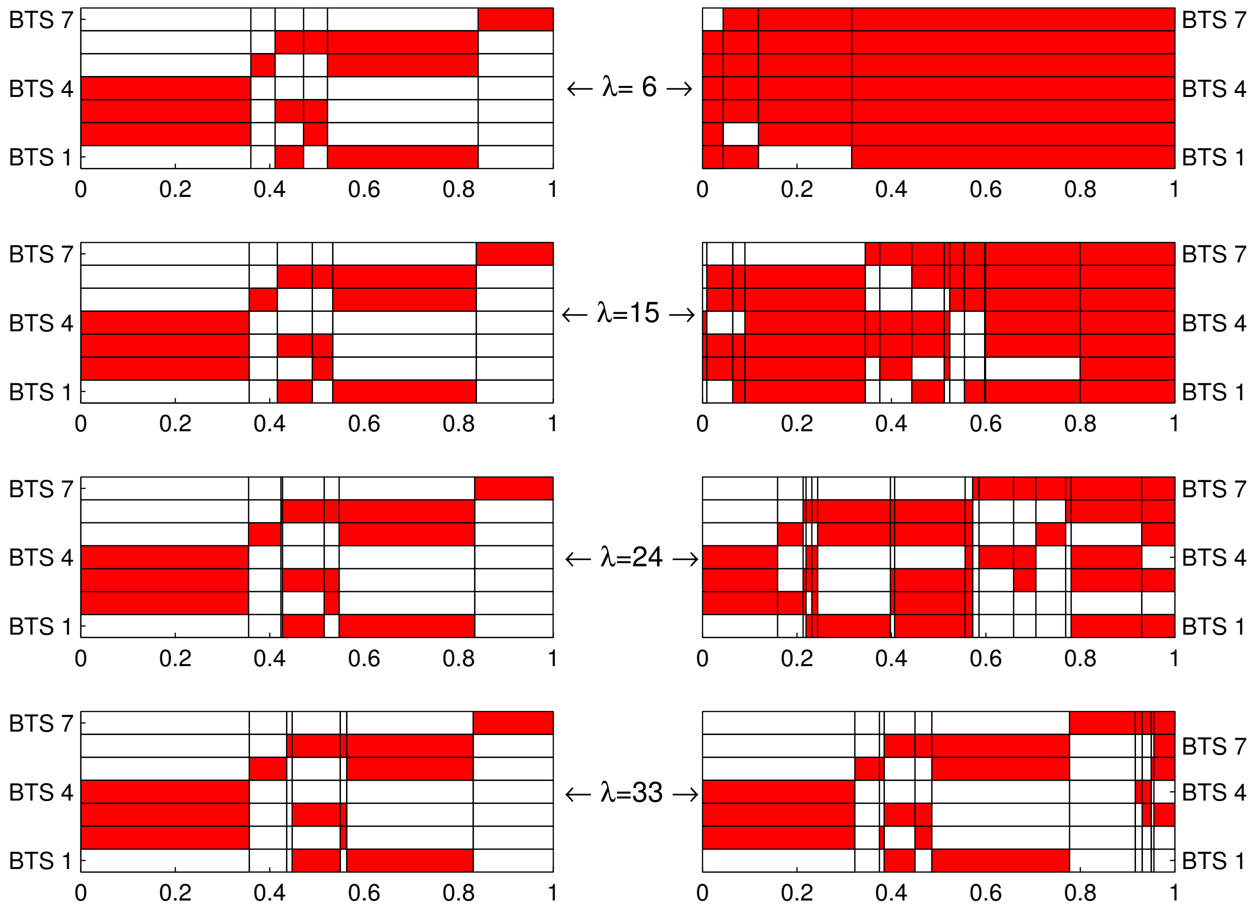


7 BTS’s
100x100 m^2
Pkt. length 1MB
Bandwidth 160MHz
Pathloss exponent 3

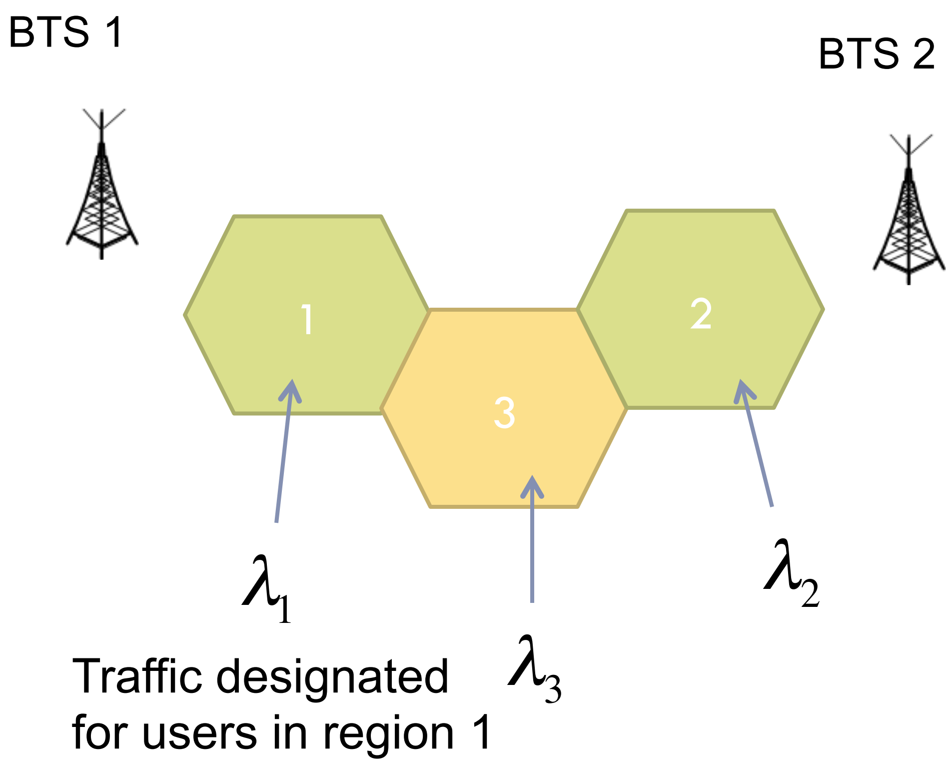
Approximation vs. Bounds



Conservative vs. Refined Allocations



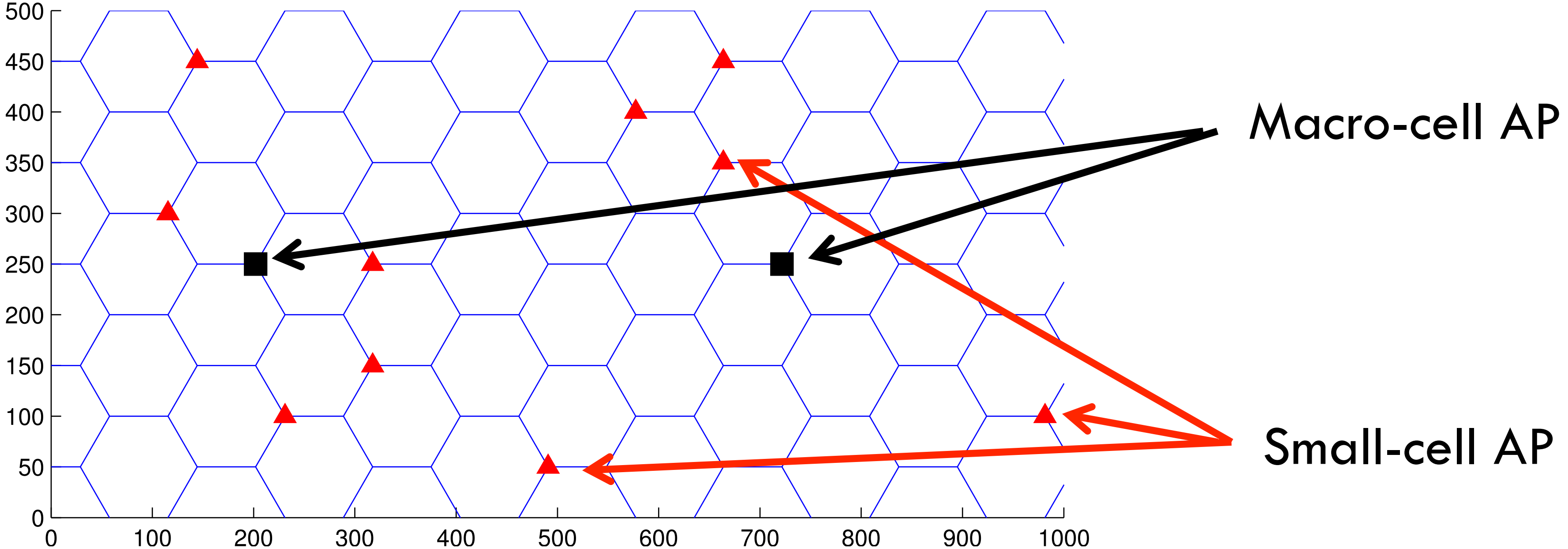
Load to BTS Assignment



Optimization variables: Spectrum used by BTS *n* to serve hexagon *k* under reuse pattern *A*

Problem: Jointly allocate traffic and bandwidth across base stations.

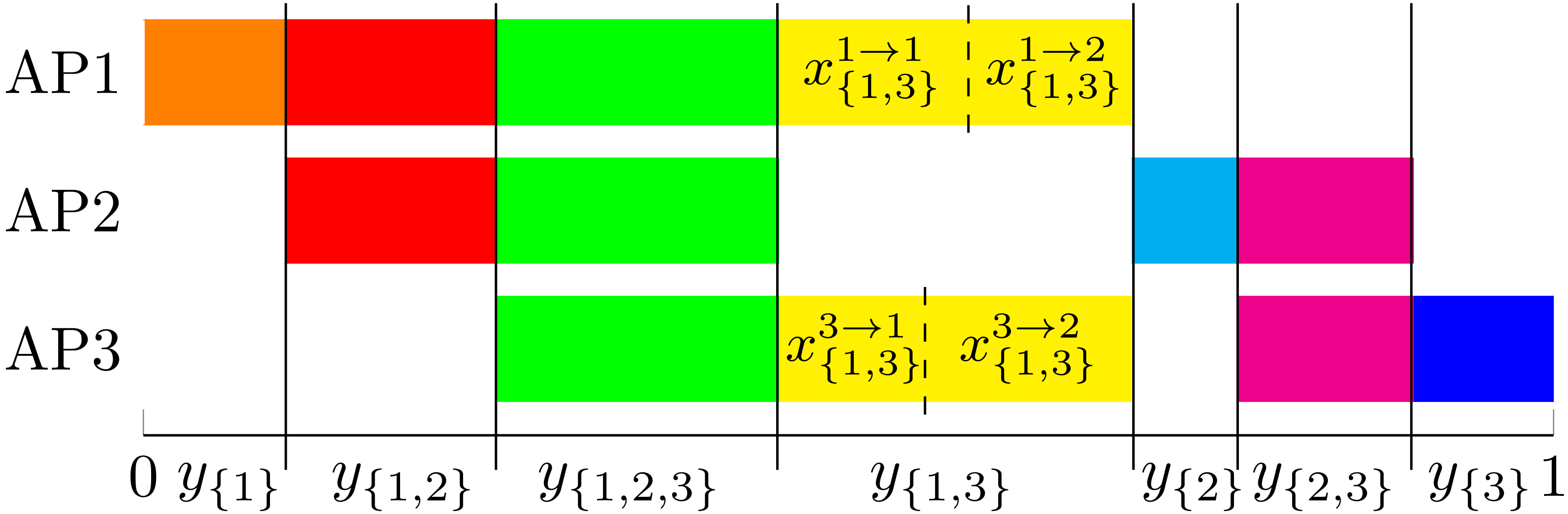
Load to BTS Assignment: Notation



- Set of BTSs: $\mathcal{N} = \{1, 2, \cdots, N\}$
- Set of UE groups: $\mathcal{K} = \{1, 2, \cdots, K\}$
- λ_k : packet arrival rate for group *k*

Load to AP Assignment: Notation

- $s_A^{i \to j}$: spectral efficiency of BTS *i* serving group *j* under reuse pattern *A*.
- $x_A^{i \to j}$: spectrum resource used by BTS *i* to serve group *j* under reuse pattern *A*.
- y_A : fraction of spectrum resources allocated to reuse pattern *A*.



Conservative Optimization (Original)

$$\min_{\{\mathbf{x},\mathbf{r}\}} \sum_{i=1}^{N} \left(\frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i} \right) \frac{1}{r_i - \lambda_i}$$

Subject to:

$$r_i > \lambda_i$$

$$r_i = \sum_{B \subset \mathcal{N}} s_{i,B} x_B \quad \forall i \in \mathcal{N}$$

$$x_B \geq 0 \quad \forall B \subset \mathcal{N}$$

$$\sum_{B \subset \mathcal{N}} x_B = 1$$

Conservative Optimization (Modified)

$$\max_{\mathbf{x},\mathbf{r}} U(\mathbf{x},\mathbf{r})$$

Subject to:

$$r_i > \lambda_i$$

$$r_i = \sum_{B\subset\mathcal{N}} s_{i,B}x_B \quad \forall i \in \mathcal{N}$$

$$x_B \geq 0 \quad \forall B \subset \mathcal{N}$$

$$\sum_{B\subset\mathcal{N}} x_B = 1$$

Conservative Optimization (Modified)

$$\max_{\mathbf{x},\mathbf{r}} U(\mathbf{x}, \mathbf{r})$$

Subject to: $$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

$$x_B \geq 0 \qquad \forall B \subset \mathcal{N}$$

$$\sum_{B \subset \mathcal{N}} x_B = 1$$

Conservative Optimization (Modified)

$$\max_{\mathbf{x},\mathbf{r}} U(\mathbf{x},\mathbf{r})$$

Subject to:

$$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

$$\sum_{j=1}^{K} x_B^{i \to j} = y_B \qquad \forall i \in \mathcal{N}$$

- Convex for concave *U*, O($KN2^N$) variables
- The solution achieves the maximum throughput region.

Average Delay Minimization

$$\min \sum_{j=1}^{K} \lambda^j \frac{1}{(r^j - \lambda^j)^+}$$

Subject to: $$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

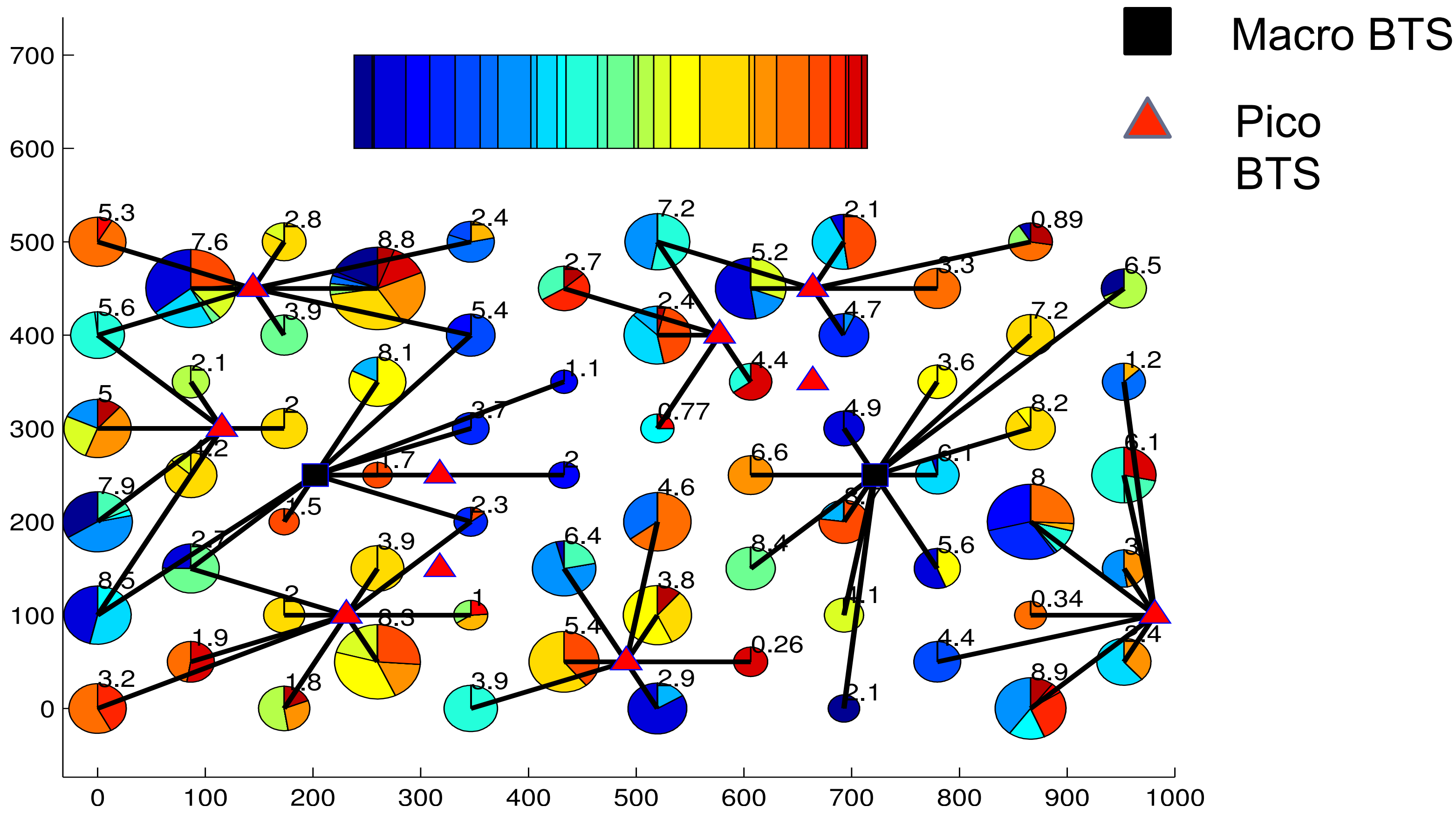
$$\sum_{j=1}^{K} x_B^{i \to j} = y_B \qquad \forall i \in \mathcal{N}$$

- Convex for concave *U*, O(*KN*2^N) variables
- The solution achieves the maximum throughput region.

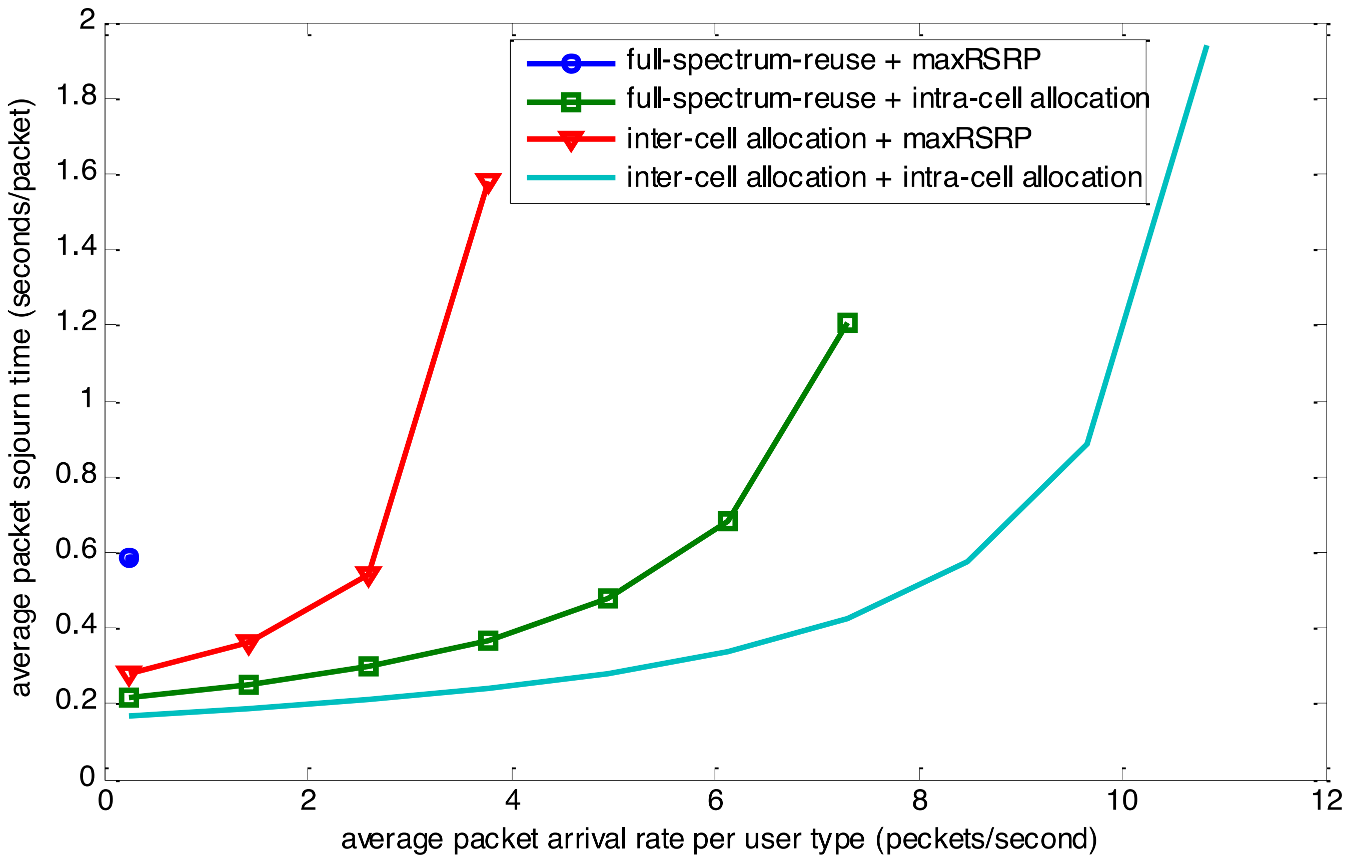
Properties of the Solution

- Uses at most K of the 2^N reuse patterns
- At most N-1 groups are jointly served by ≥ 1 AP.
- Throughput optimal

Spectrum Allocation



Delay (2 macros, 8 small cells)



Energy Conservation

- Tradeoff:
  - Turning off an AP saves energy.
  - Turning off an AP increases the load for neighbors.
- Problem: serve the offered traffic with the minimal number of active APs.
- Related work:
  [Pollakis, Cavalcante and Stanczak, '12]
  (no spectrum optimization)

Average Delay Minimization

$$\min \sum_{j=1}^{K} \lambda^j \frac{1}{(r^j - \lambda^j)^+}$$

Subject to: $$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

$$\sum_{j=1}^{K} x_B^{i \to j} = y_B$$

$$\sum_{B \subset \mathcal{N}} y_B = 1, \qquad \mathbf{x} \geq 0$$

Weighted Energy Minimization

$$\min_{\mathbf{z}} \sum_{i \in \mathcal{N}} c^i |z^i|_0$$

Subject to:

total bandwidth assigned to AP *i*

$$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

$$\sum_{j=1}^{K} x_B^{i \to j} = y_B, \qquad \sum_{B \subset \mathcal{N}} y_B = 1$$

$$\mathbf{x} \geq 0$$

Weighted Energy Minimization

$$\min_{\mathbf{z}} \sum_{i \in \mathcal{N}} c^i |z^i|_0$$

Subject to:

$$\sum_{B \subset \mathcal{N}} \sum_{j \in \mathcal{K}} x_B^{i \to j} \leq z^i \qquad \forall i \in \mathcal{N}$$

$$r^j - \lambda^j \geq 1/\tau^j$$

$$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

$$\sum_{j=1}^{K} x_B^{i \to j} = y_B, \qquad \sum_{B \subset \mathcal{N}} y_B = 1$$

$$\mathbf{x} \geq 0$$

Weighted Energy/Utility Minimization

$$\min_{\mathbf{z},\mathbf{r}} \sum_{i \in \mathcal{N}} c^i |z^i|_0 + U(\mathbf{r})$$

Subject to:

$$\sum_{B \subset \mathcal{N}} \sum_{j \in \mathcal{K}} x_B^{i \to j} \leq z^i \qquad \forall i \in \mathcal{N}$$

$$r^j - \lambda^j \geq 1/\tau^j$$

$$r^j = \sum_{i=1}^{n} \sum_{B \subset \mathcal{N}} s_B^{i \to j} x_B^{i \to j} \qquad \forall j \in \mathcal{K}$$

$$\sum_{j=1}^{K} x_B^{i \to j} = y_B, \qquad \sum_{B \subset \mathcal{N}} y_B = 1$$

$$\mathbf{x} \geq 0$$

Reweighted ℓ_1 Minimization

$$\min_{\mathbf{z},\mathbf{r}} \sum_{i \in \mathcal{N}} c^i |z^i|_0 \quad \Rightarrow \quad \min_{\mathbf{z},\mathbf{r}} \sum_{i \in \mathcal{N}} w^i c^i z^i$$

1. $w^i \leftarrow 1$
2. Iterate:
   1. Solve the linear program;
   2. Update the weights $w^i \leftarrow (z^i + \varepsilon)^{-1}$
3. Terminate after convergence or a maximum number of iterations.

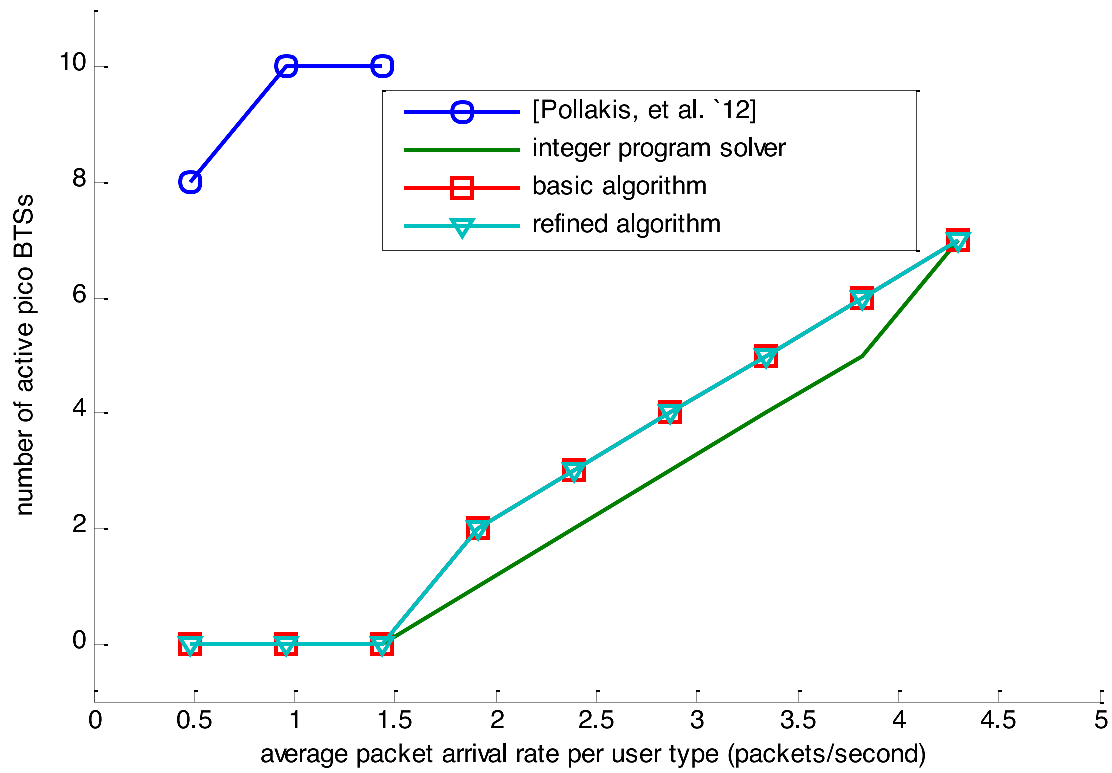
Reweighted ℓ_1 Minimization

$$\min_{\mathbf{z},\mathbf{r}} \sum_{i \in \mathcal{N}} c^i |z^i|_0 \quad \Rightarrow \quad \min_{\mathbf{z},\mathbf{r}} \sum_{i \in \mathcal{N}} w^i c^i z^i$$

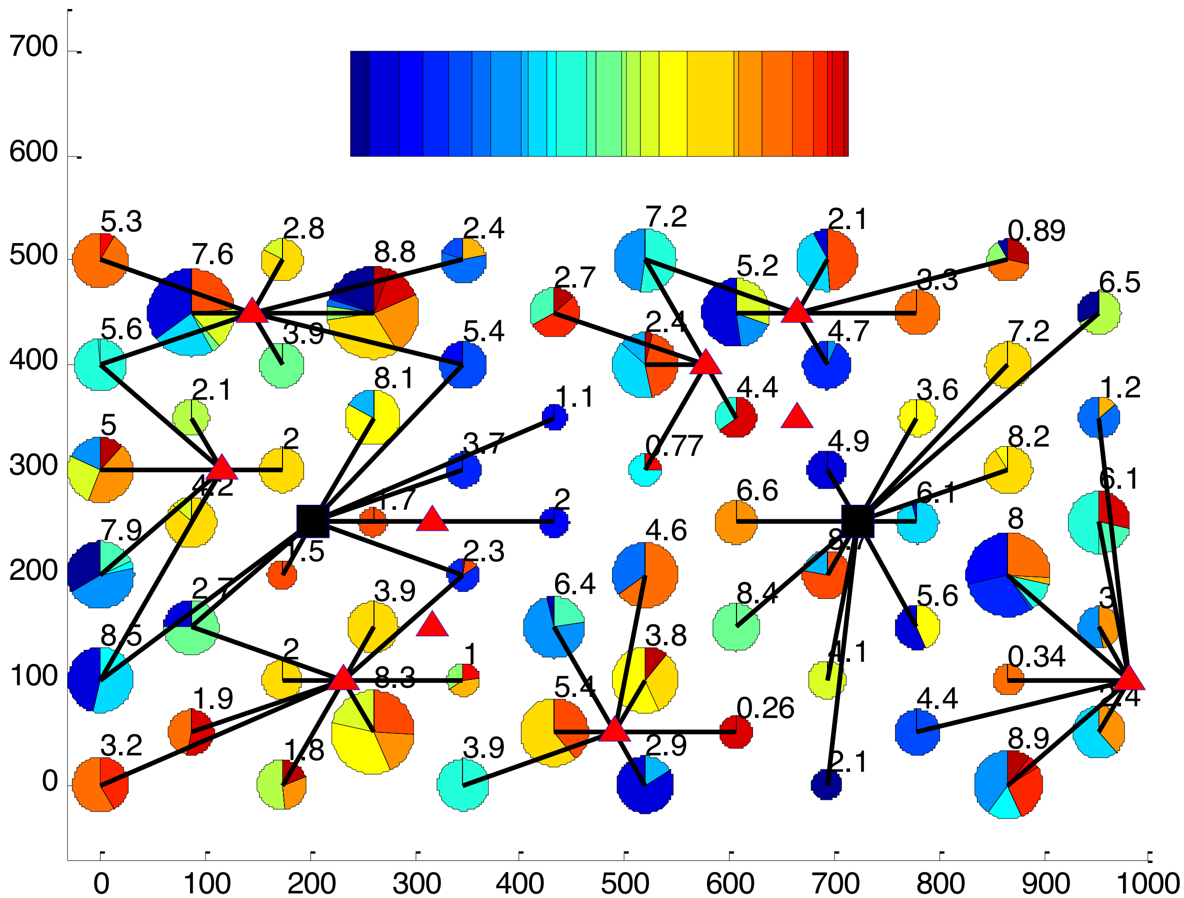
1. $w^i \leftarrow 1$
2. It
   1.
   2.
3. Terminate after convergence or a maximum number of iterations.

Convergence proved in [Pollakis et al ‘12] via majorization-minimization.

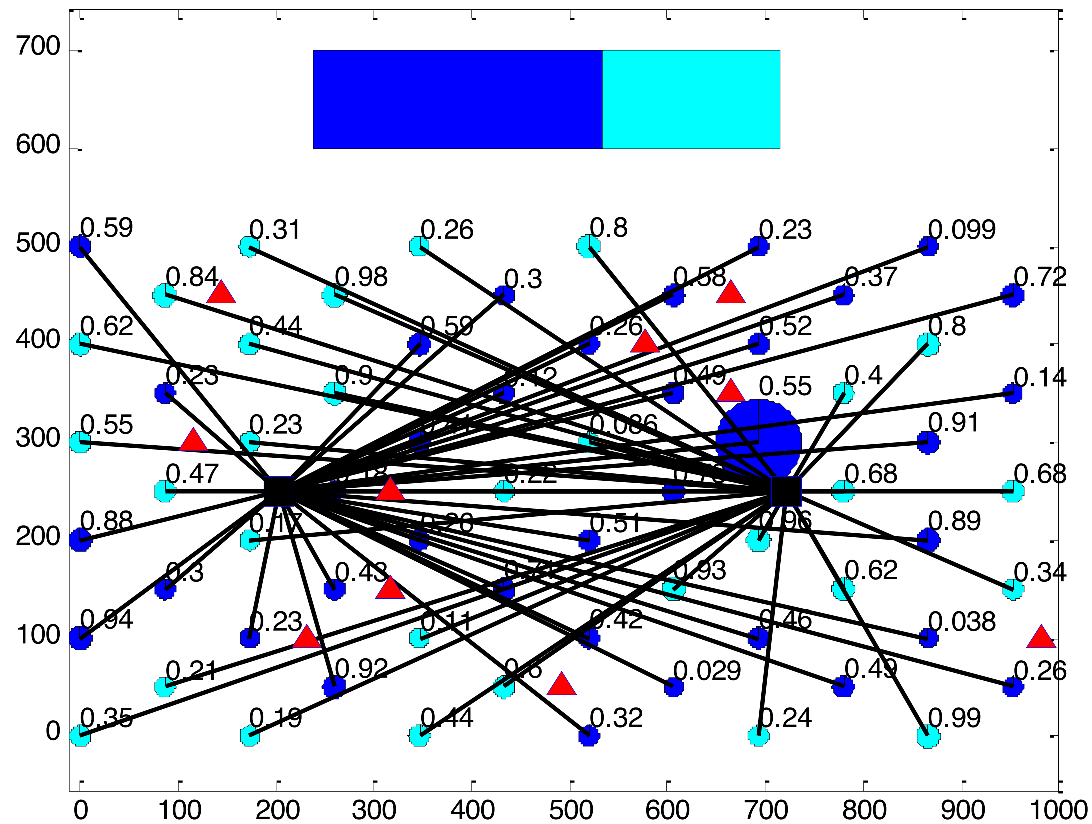
Energy vs. Traffic



Spectrum Allocation (Heavy Traffic)



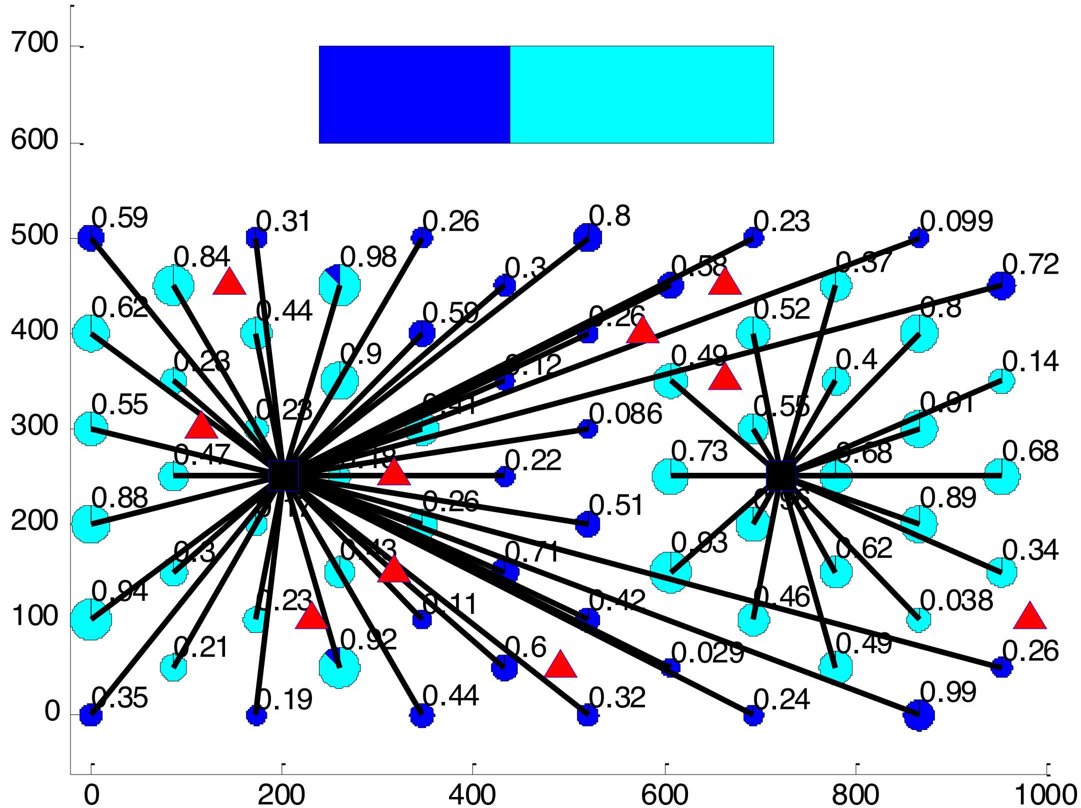
Spectrum Allocation (Light Traffic)



Post Processing

- Minimizing energy only finds a feasible solution.
- Once the set of active APs is determined, can further minimize average delay as before.

Post Processing (Light Traffic)



Scalability

- Number of variables increases as $O(KN2^N)$
- Infeasible to find optimal allocation for $N >> 20$.
- To scale to large networks can exploit
  - Path loss: radio signals cause negligible interference over large enough distances;
  - Small node degrees: typically bounded by a constant

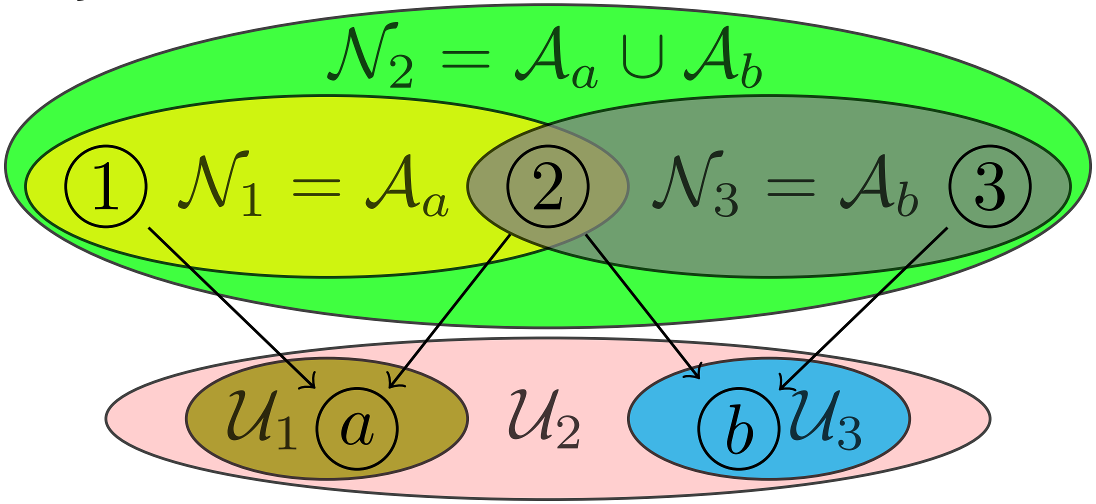
Node Neighborhood

- E is the set of network links with non-negligible gain

$$\mathcal{A}_j = \{i | (i \to j) \in E\}$$

$$\mathcal{U}_i = \{j | (i \to j) \in E\}$$

$$\mathcal{N}_i = \{\cup_{j \in \mathcal{U}_i} \mathcal{A}_j\}$$



Local Patterns and Variables

$$r^j = \sum_{A \subset \mathcal{N}} \sum_{i \in A} s_A^{i \to j} x_A^{i \to j} = \sum_{i \in \mathcal{A}_j} \sum_{B \subset \mathcal{N}_i} s_B^{i \to j} z_B^{i \to j}$$

- Local variables $z_B^{i \to i}$ are only defined for links in E and B in $\mathcal{N}_i$.
- Introduce local variables y_B^i defined for B in $\mathcal{N}_i$.
- Number of local variables is O(N).
- Consistency constraint in overlapping neighborhoods:

$$\sum_{B \subset \mathcal{N}_i : B \cap \mathcal{N}_m = C} y_B^i = \sum_{B \subset \mathcal{N}_m : B \cap \mathcal{N}_i = C} y_B^m, \quad \forall i, m \in \mathcal{N}, \forall C \neq \emptyset$$

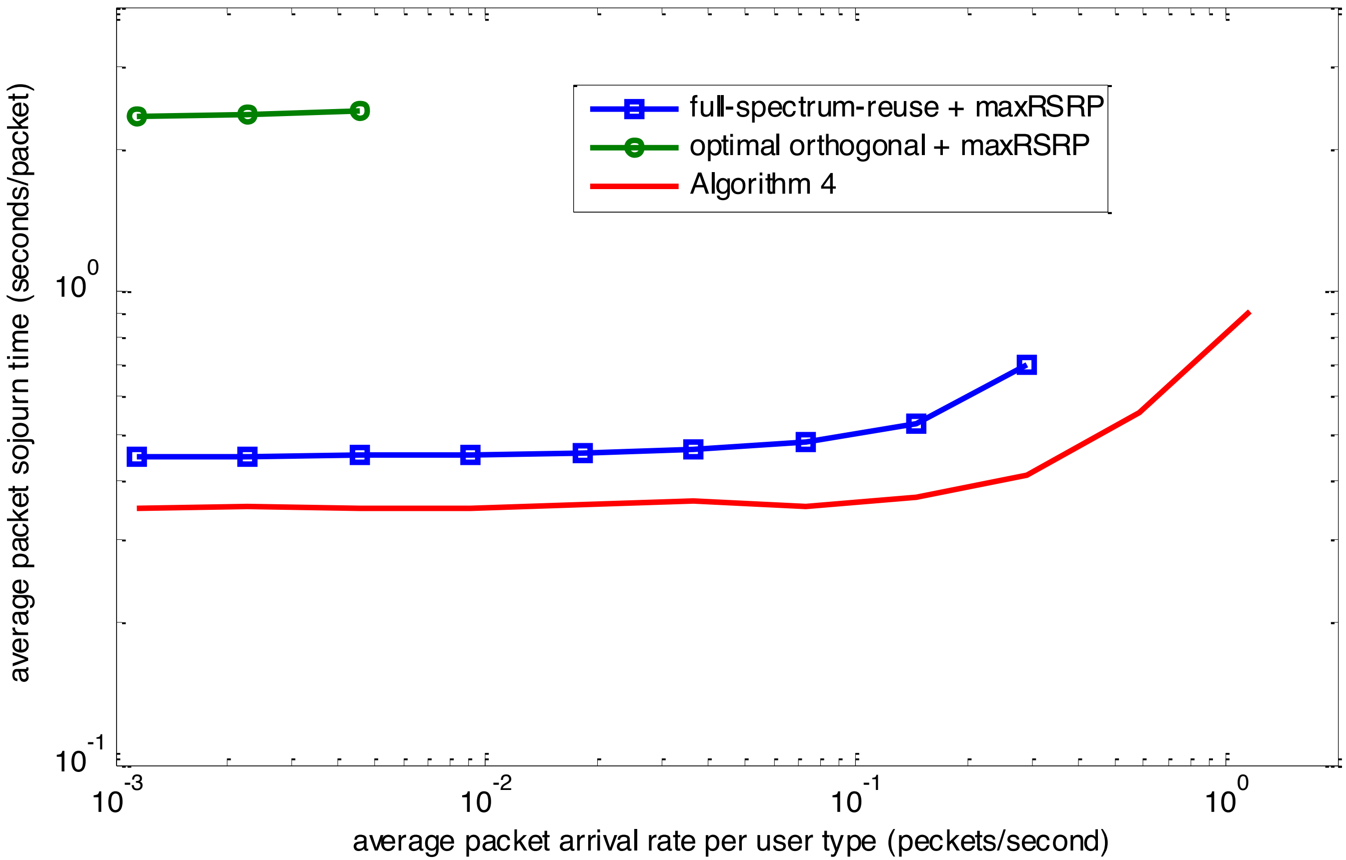
Relaxed Optimization

- Add previous constraint, optimize over **z, y**
- Relax total bandwidth constraint:

$$\sum_{B \subset \mathcal{N}} y_B \leq 1$$

- Scale back bandwidth assignments to meet constraint
- Need to satisfy additional alignment constraints
  → strong vertex coloring problem on hypergraph

Delay Example



N=25
K=126

Concluding Remarks

- Slow resource allocation can exploit spatial traffic variations.
- Centralized optimization
  - Requires gathering traffic statistics across cells
  - Re-optimize periodically
- Network size limited by computational complexity
  - Number of variables increases exponentially
  - Scalability facilitated by optimizing over local neighborhoods