Argos: Practical Base Stations for Large-scale Beamforming

Clayton W. Shepard

RICE®

Collaborators

Hang Yu

Narendra Anand

Erran Li

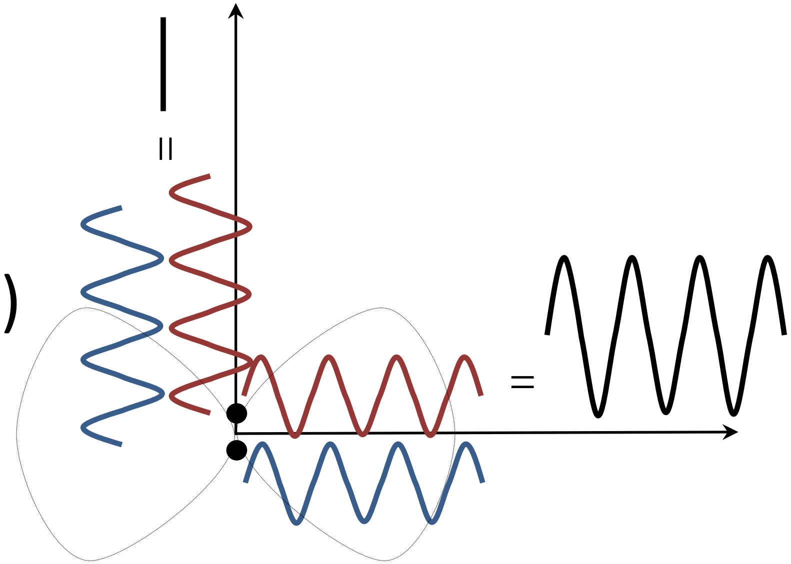
Thomas Marzetta

Richard Yang

Lin Zhong

Background

- Beamforming
  - Power Gain
  - Adjust phase (“beamweights”)
  - Leverages Interference



- Open-loop
  - Pre-compute weights to specify direction
- Closed-loop (adaptive)
  - Use channel state information (CSI) to target receivers

Background

- Single-user beamforming (SUBF)

$$W_{SUBF} = c \cdot H^{*}$$

- Multi-user beamforming (MUBF)

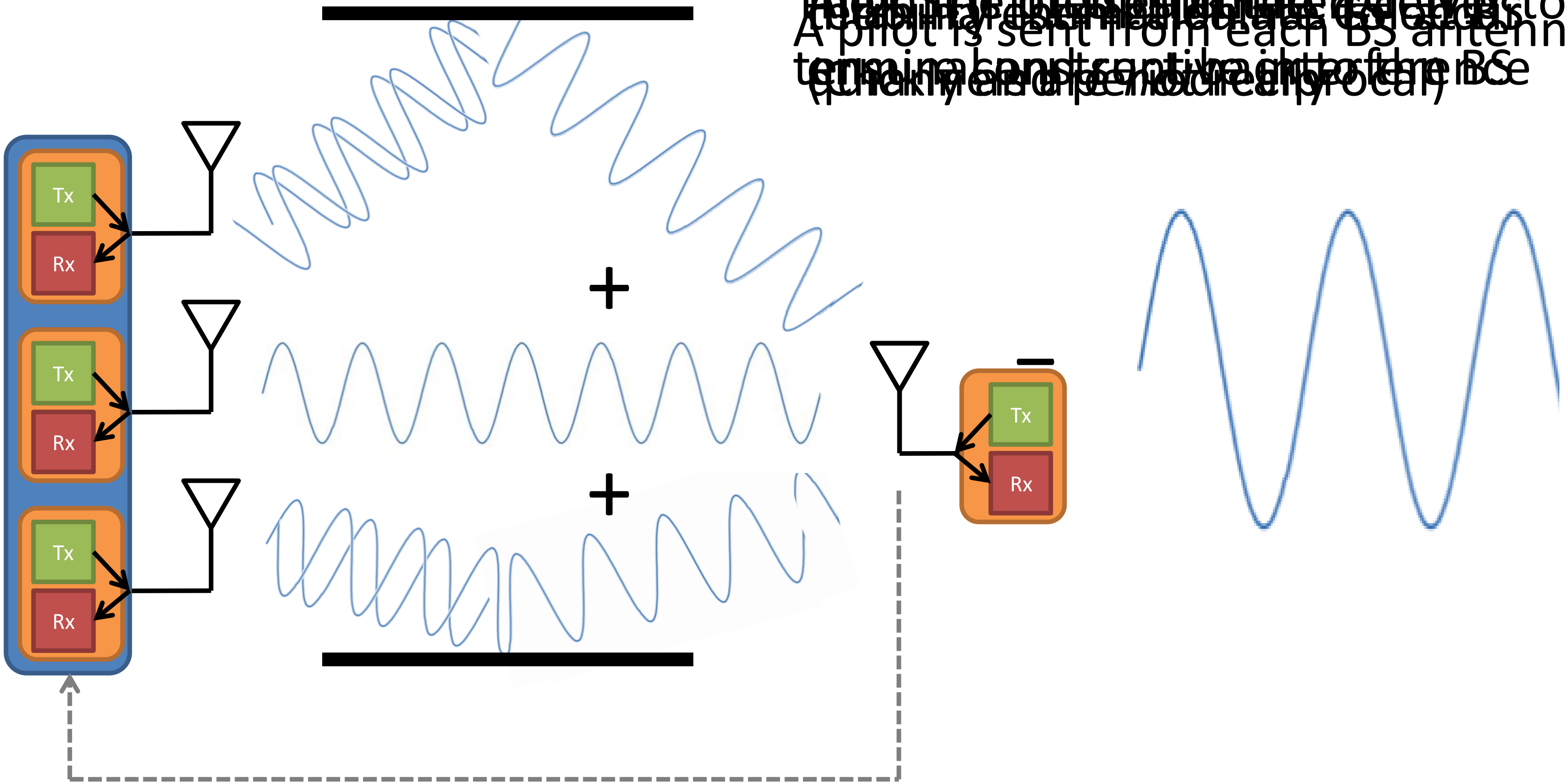
$$W_{MUBF} = c \cdot H^{*}(H^{T}H^{*})^{-1}$$

Background: Channel Estimation

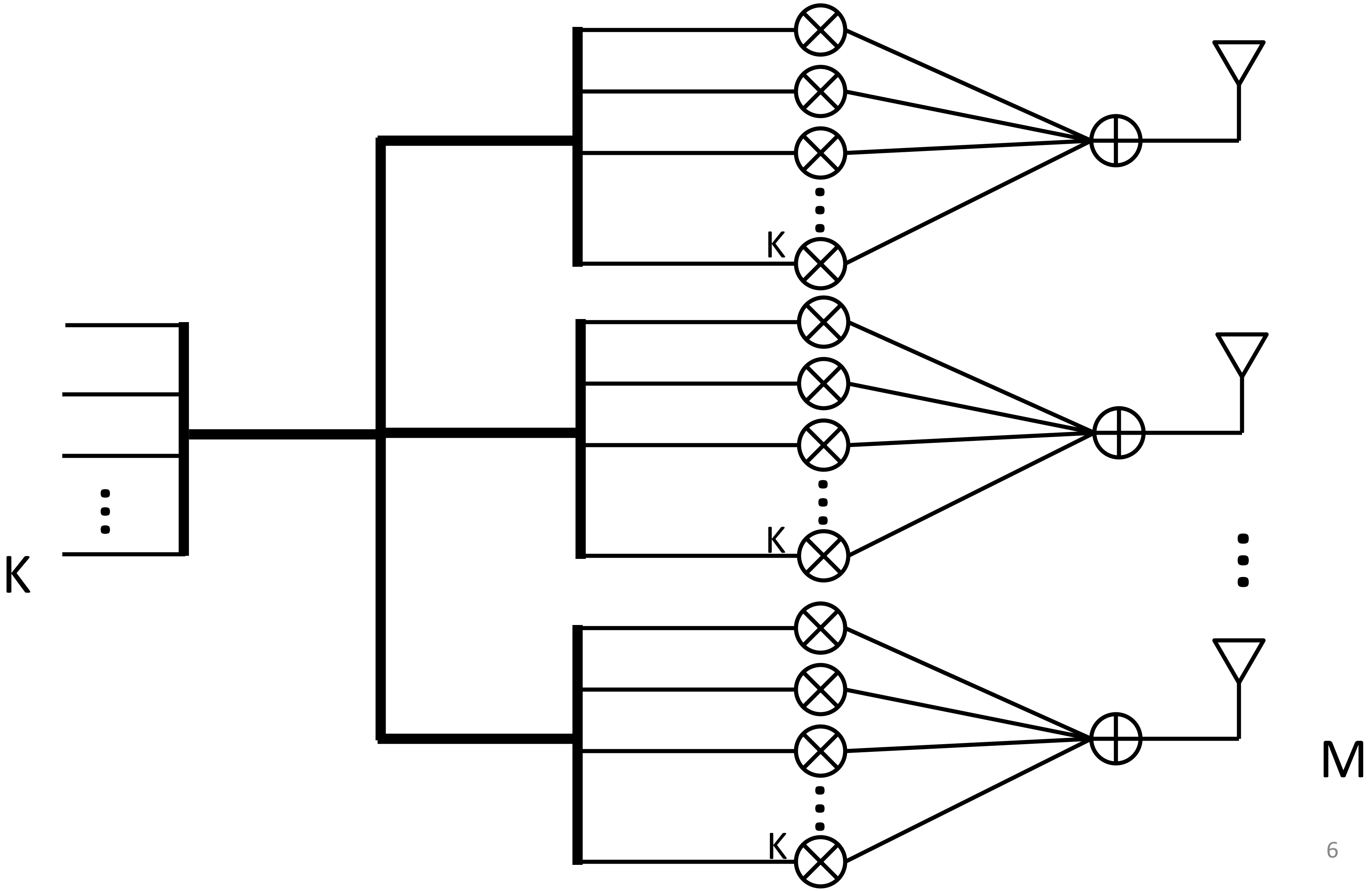
[illegible]

A pilot is sent from each BS antenna

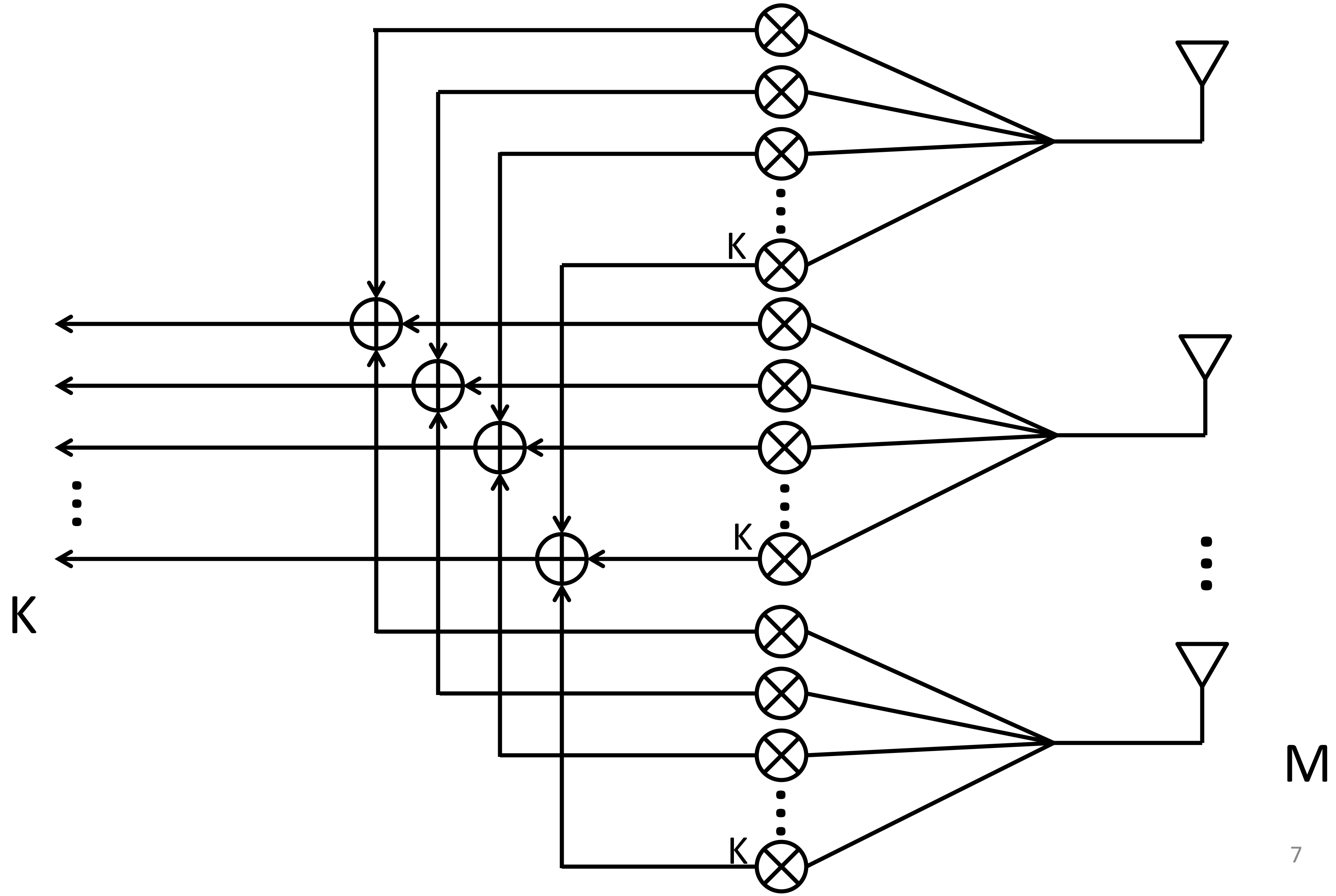
[illegible]



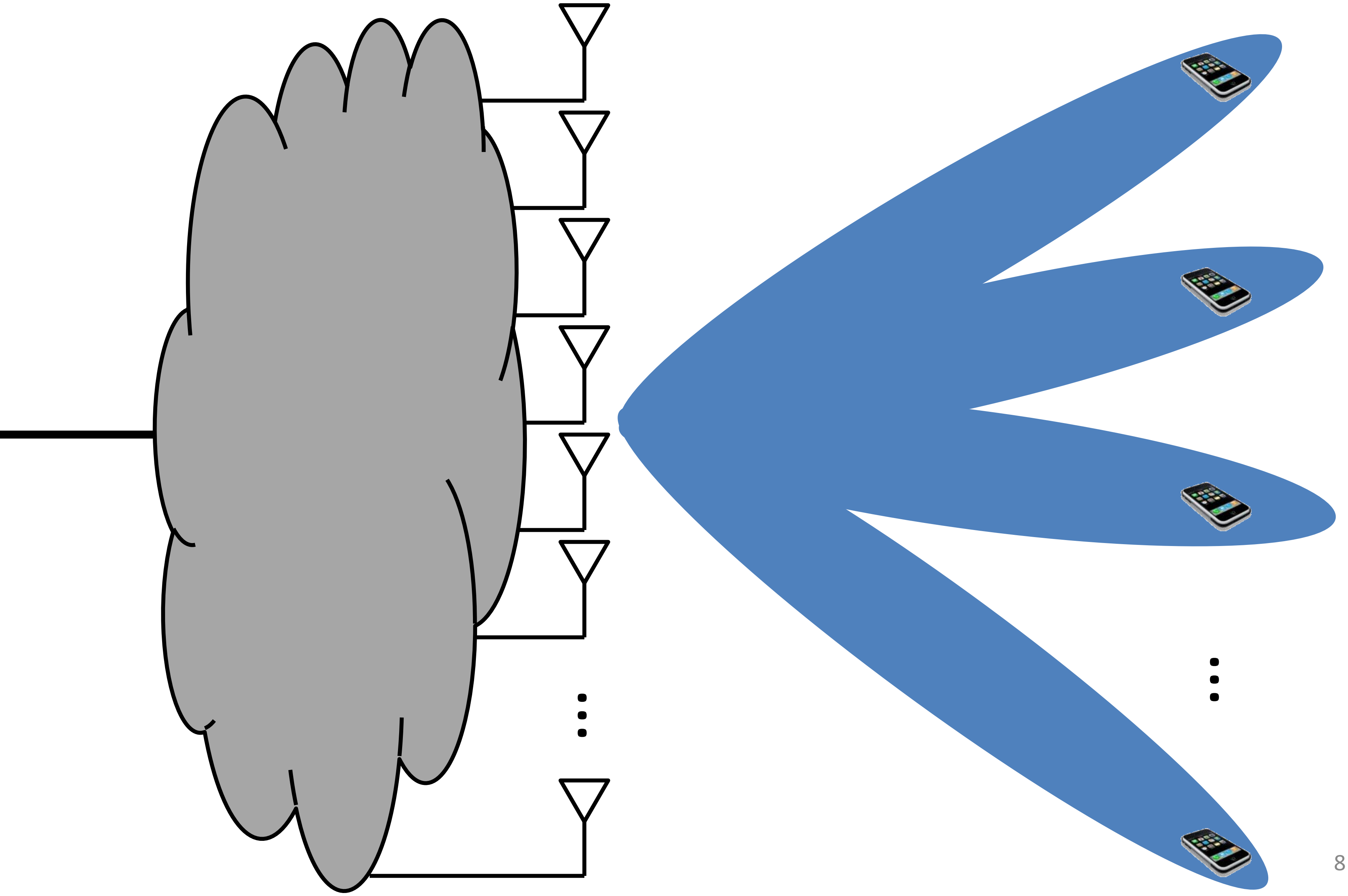
MUBF linear pre-coding: downlink



MUBF linear pre-coding: uplink



Our vision



Prior Work

- Large-scale beamforming theory
  - T.L. Marzetta. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. IEEE Transactions on Wireless Communications, Nov. 2010.
  - Fredrik Rusek and Daniel Persson and Buon Kiong Lau and Erik G. Larsson and Thomas L. Marzetta and Ove Edfors and Fredrik Tufvesson Scaling up MIMO: Opportunities and Challenges with Very Large Arrays. arXiv, Jan. 2012.
- Real-world beamforming
  - E. Aryafar, N. Anand, T. Salonidis, and E. Knightly. Design and Experimental Evaluation of Multi-userBeamforming in Wireless LANs. In Proceedings of MobiCom, 2010
- Reciprocal calibration
  - F. Kaltenberger, H. Jiang, M. Guillaud, R. Knopp. Relative channel reciprocity calibration in MIMO/TDD systems. Future Network and Mobile Summit, June 2010.

First large-scale beamforming base station

Overview of contributions

- Scalable architecture
- Internal reciprocity calibration
- Novel fully distributed beamforming method

Can beamforming scale with the number of base station antennas?

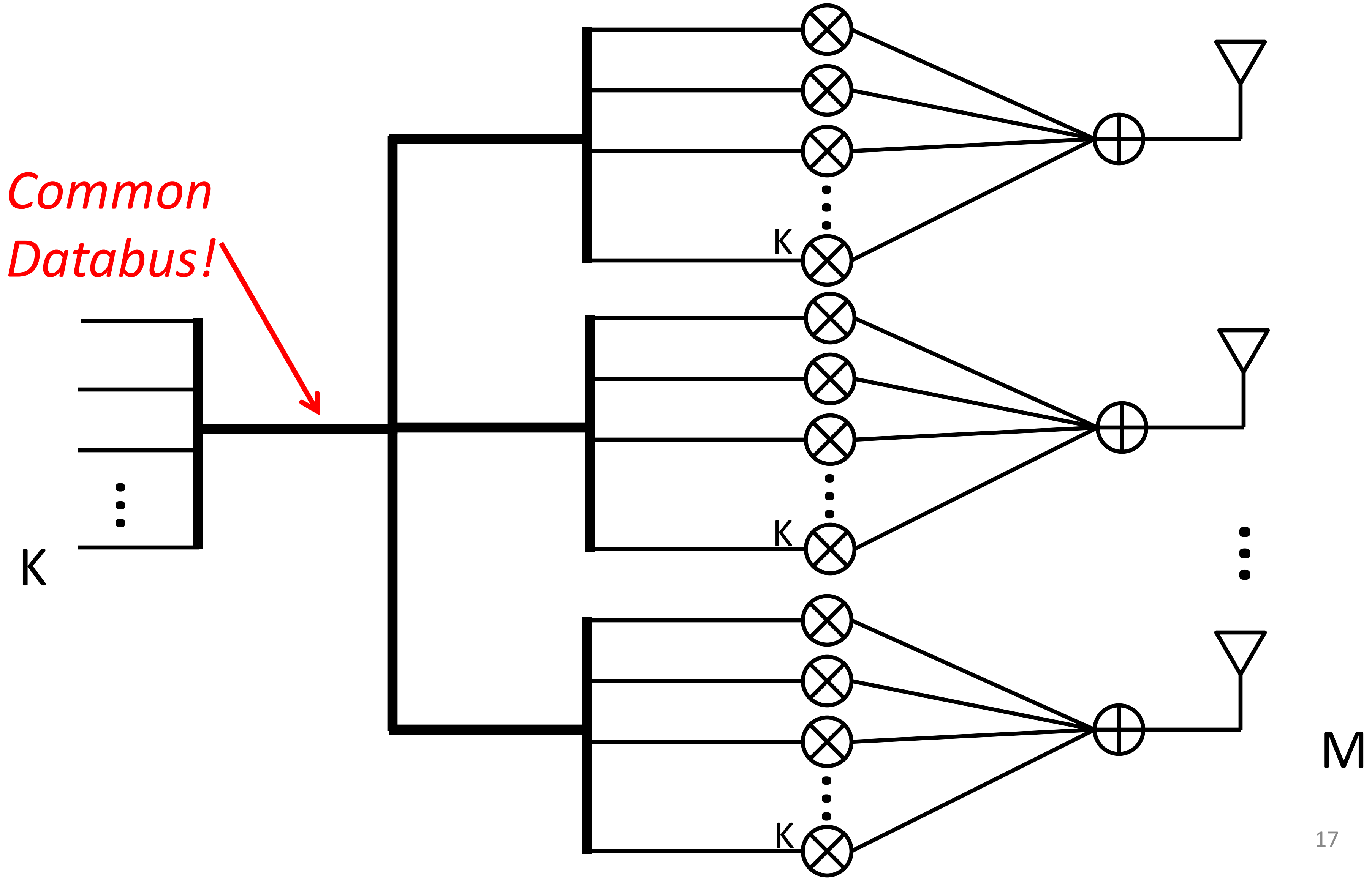
Not with current techniques!

- CSI acquisition
  - Typically requires # of base station (BS) antennas (M) + # of terminals (K) pilots
- Weight calculation
  - All existing methods have centralized data dependency
  - Requires M*K channel estimates and produces M*K weight values
- Linear pre-coding
  - Produces M data streams

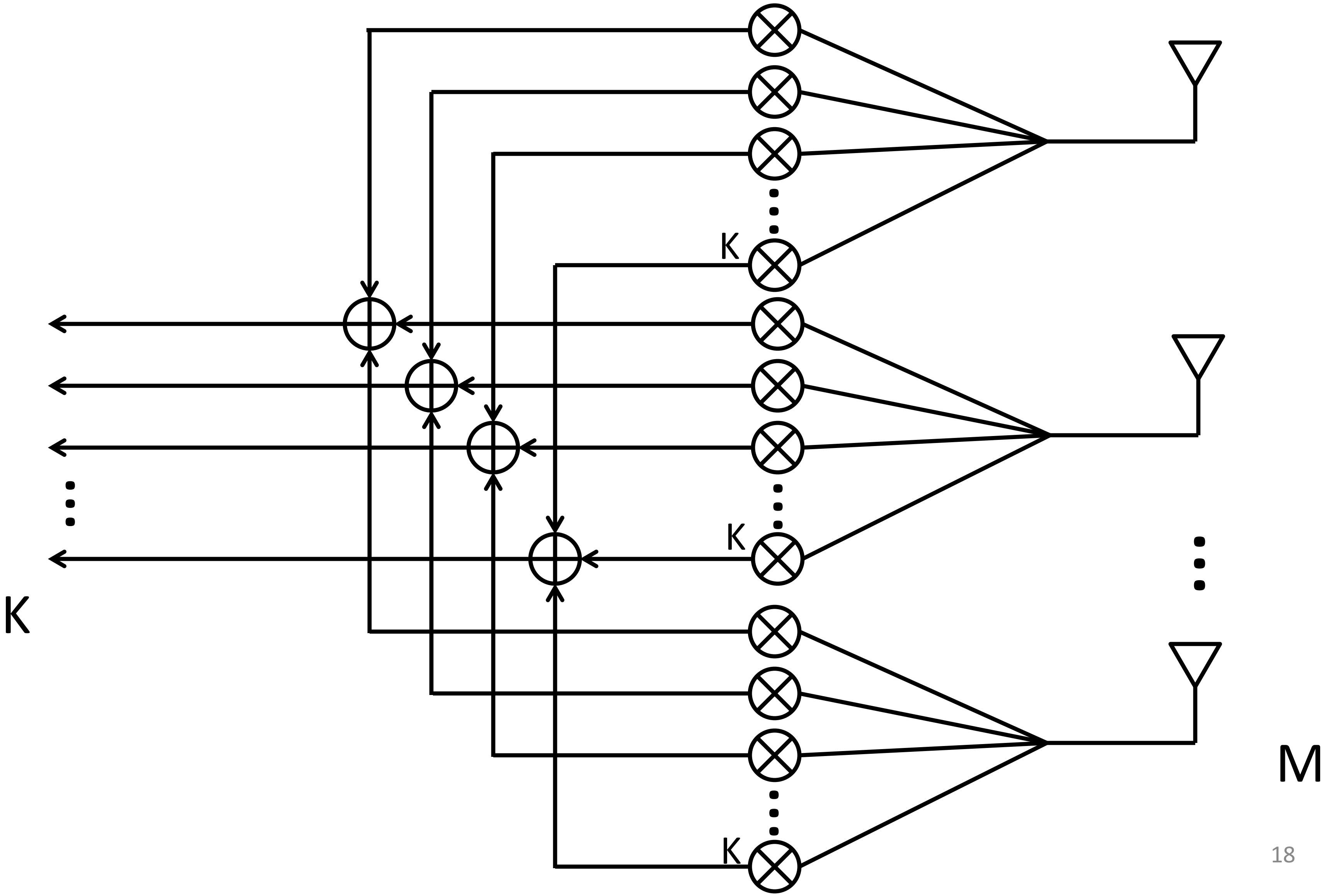
With careful design and new techniques it can!

- CSI Acquisition
  - Leverage TDD reciprocity to limit pilots to K
  - Requires calibration
- Weight Calculation
  - Novel decentralized weight calculation
- Linear Pre-coding
  - Apply weights at radio
  - For uplink combine streams any time they meet

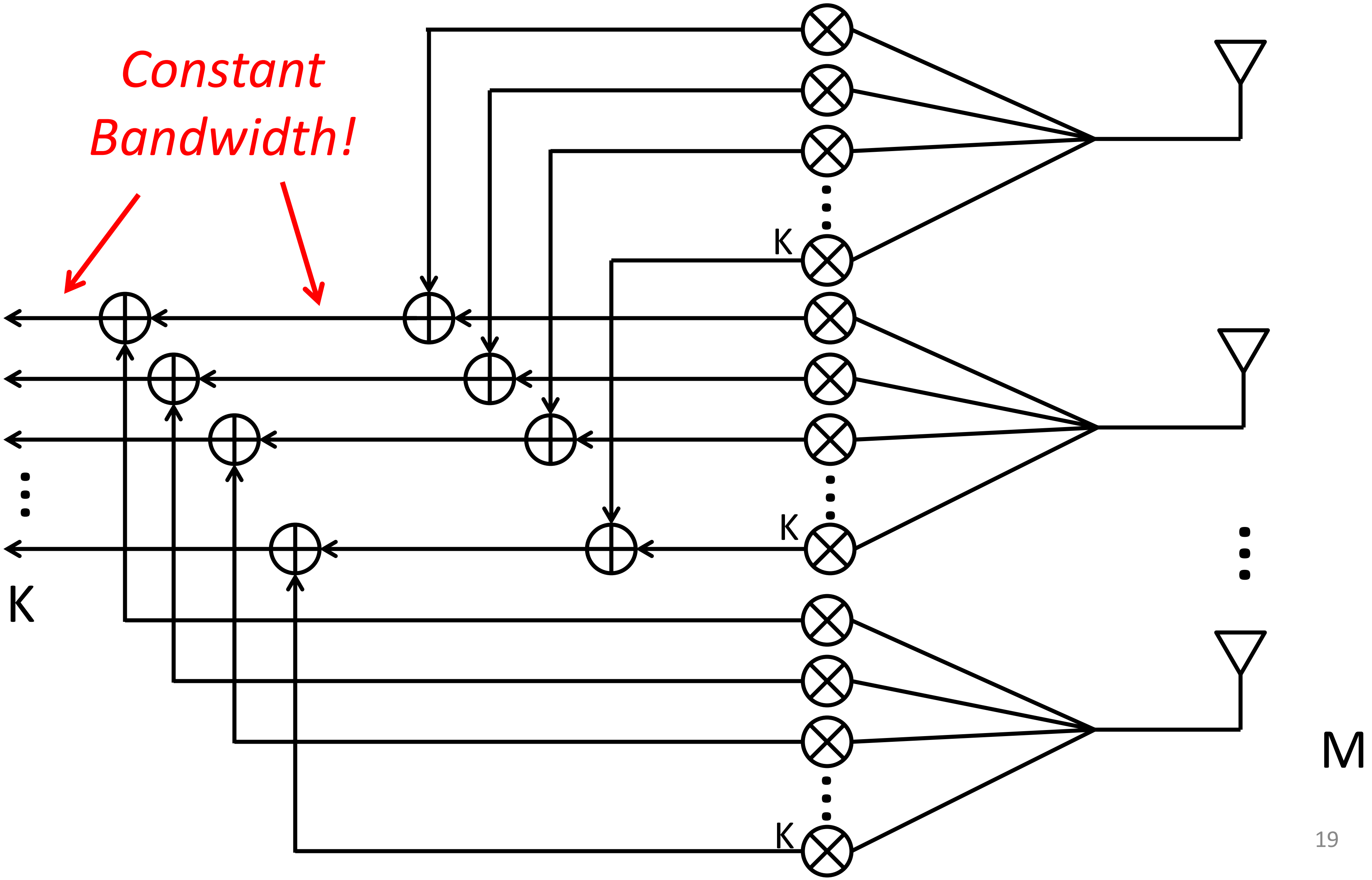
Scalable linear pre-coding



MUBF linear pre-coding: uplink



Scalable linear pre-coding



Constant Bandwidth!

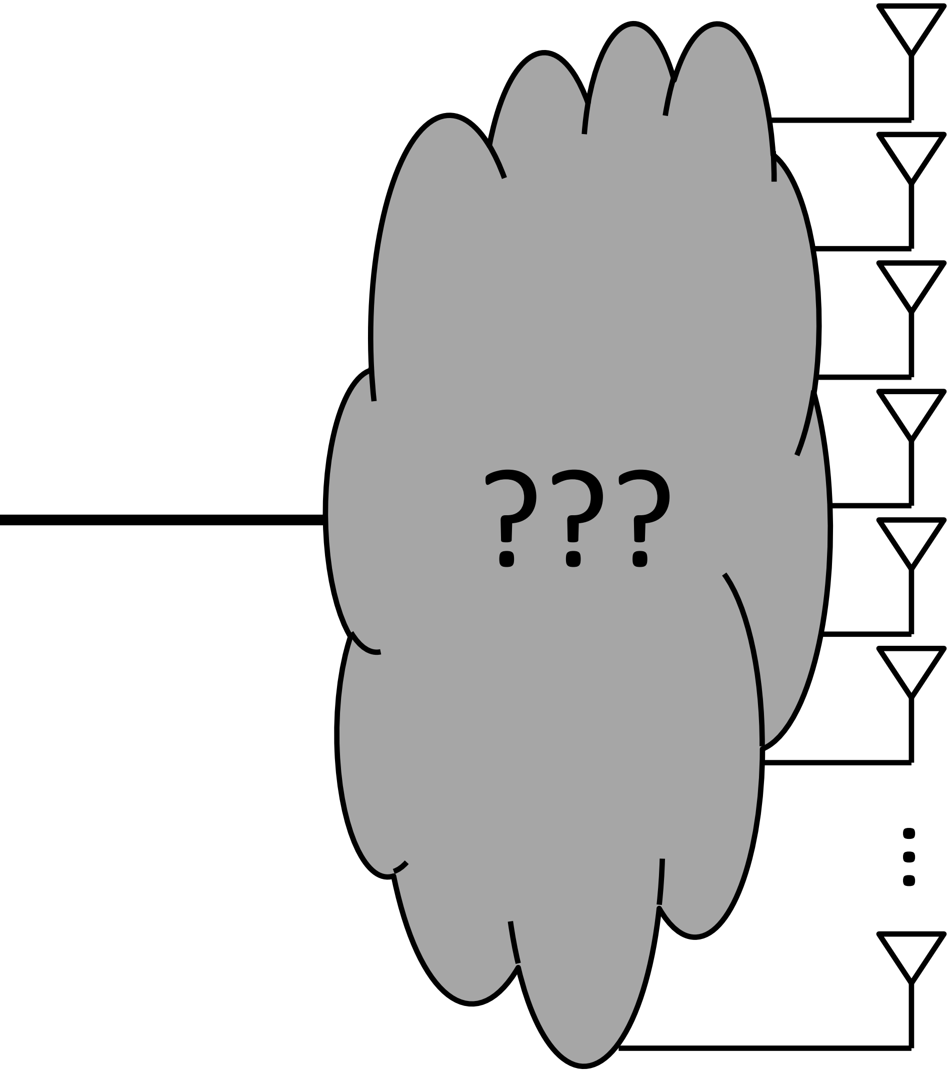
Ramifications

- CSI and weights are computed and applied (linear pre-coding) locally at each BS radio
  - No overhead for additional BS radios

- No central data dependency
  - No latency from data transport
  - No stringent latency requirements
  - Constant data rate common bus (no switching!)

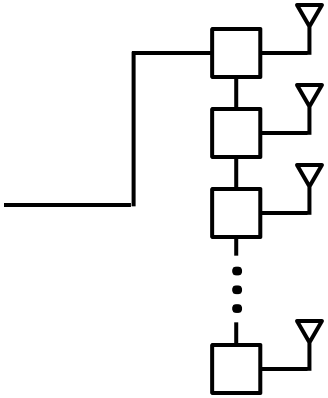
- Unlimited scalability!

Design goals

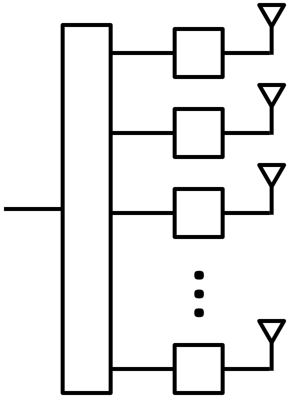


- Scalable
  - Support thousands of BS antennas
- Cost-effective
  - Cost scales linearly with # of antennas
- Reliable

How do we design it?

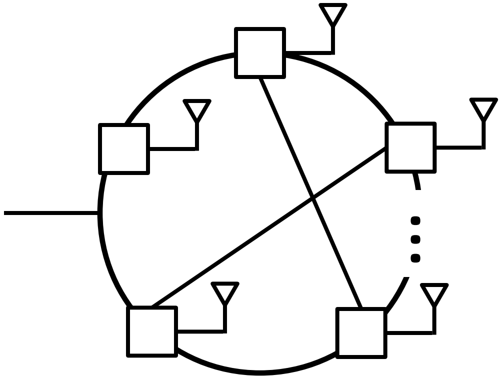


- Daisy-chain (series)
  - Unreliable
  - Large end to end latency



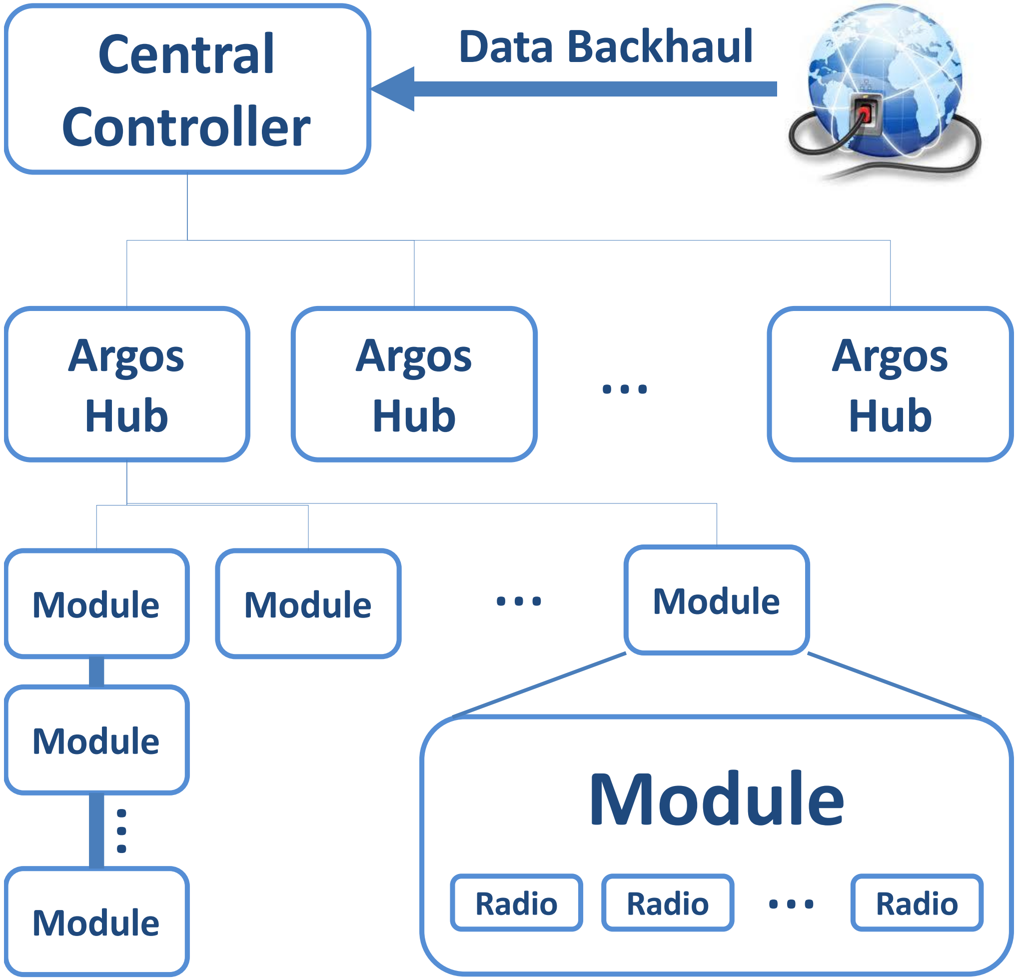
- Flat structure
  - Un-scalable
  - Expensive, with large fixed cost

- Token-ring / Interconnected
  - Not amenable to linear pre-coding
  - Variable Latency
  - Routing overhead



Solution: Argos

- Modular
  - Daisy-chainable
  - 1 or more radios
- Hierarchal
  - Increases Reliability
  - Constrains Latency
  - Cost-effective



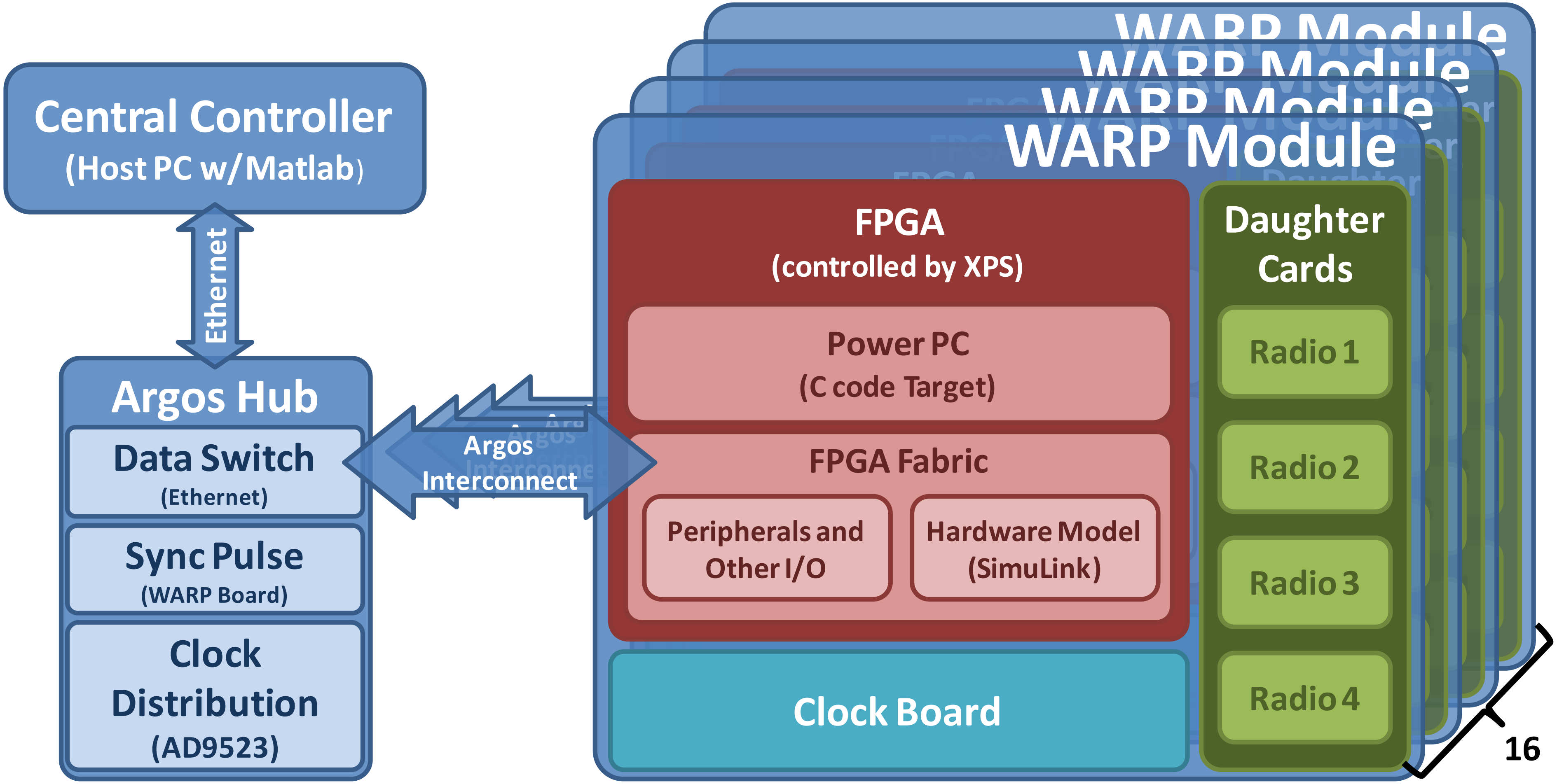
Scalability of Argos

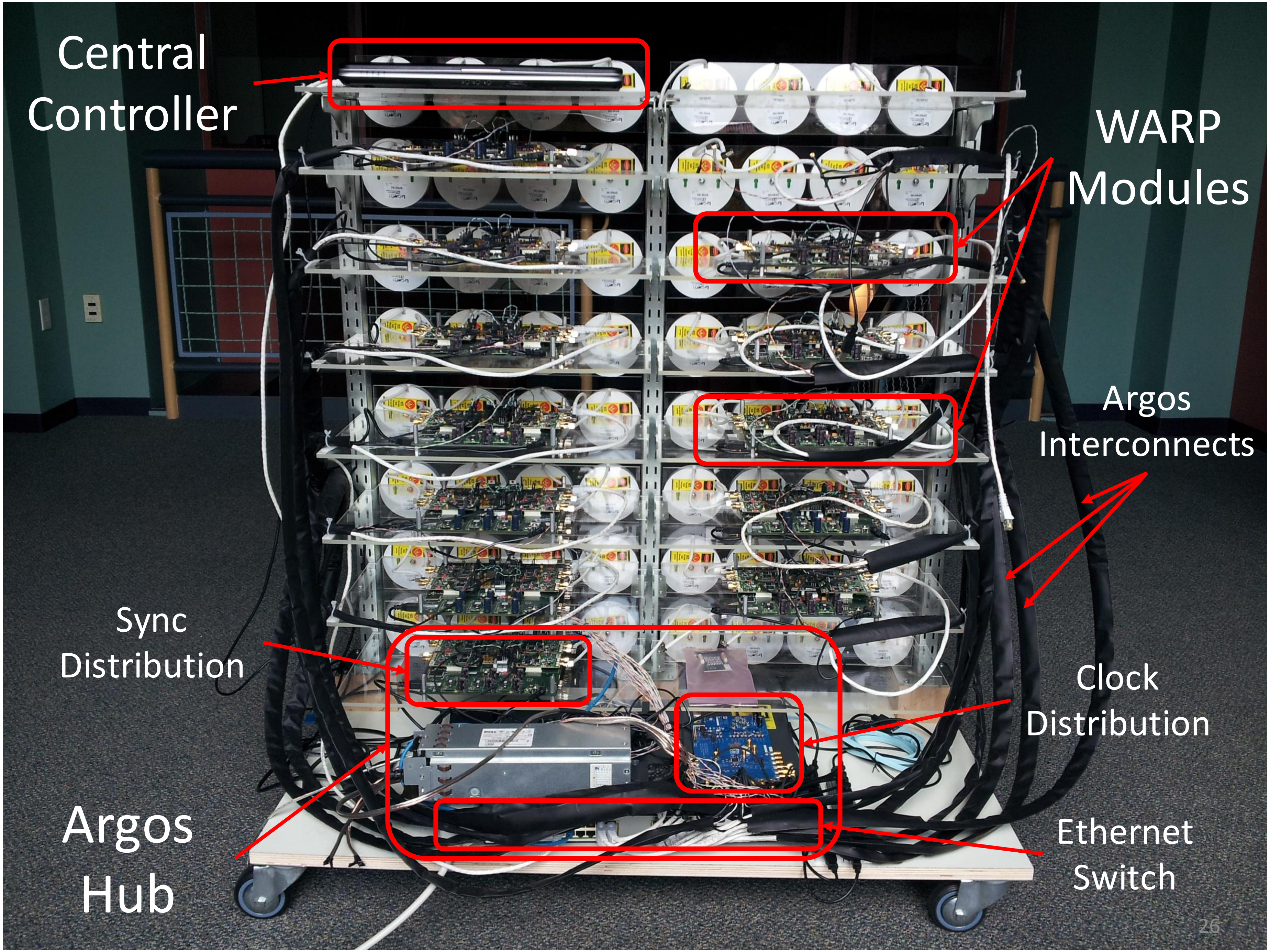
- Scalable in 4 directions:
  - # of Radios per Module
  - # of Modules per Chain
  - # of ports per Hub
  - # of Hubs (and levels)

- Reliable
  - Branches can fail without affecting other branches
  - Central hubs can be easily made redundant

- Accommodates linear pre-coding
  - Add samples together at every junction

Implementation




Central Controller
WARP Modules
Argos Interconnects
Sync Distribution
Clock Distribution
Argos Hub
Ethernet Switch

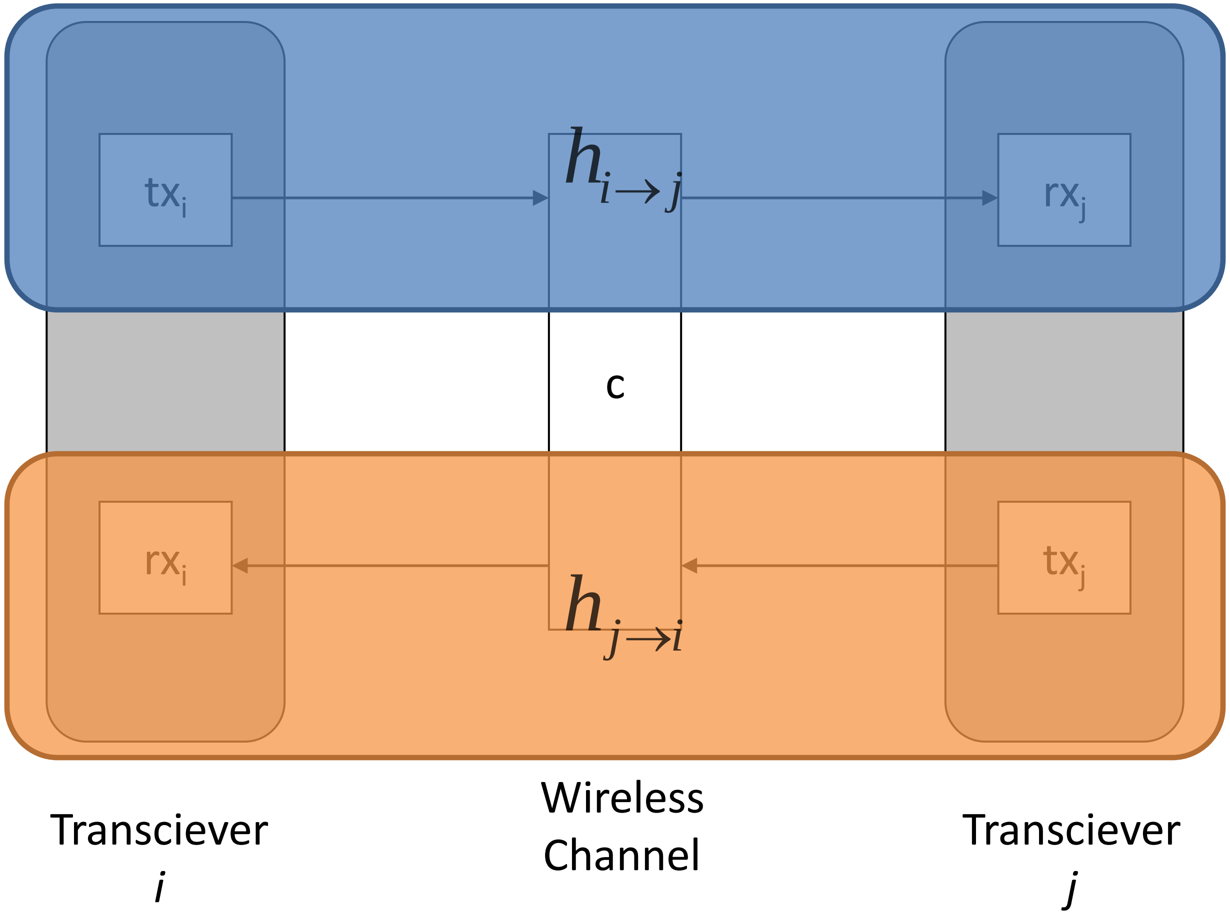
Overview of contributions

- Scalable architecture
- **Internal reciprocity calibration**
- Novel fully distributed beamforming method

Channel reciprocity

$$h_{i\rightarrow j} = tx_i \cdot c \cdot rx_j$$

$$h_{j\rightarrow i} = tx_j \cdot c \cdot rx_i$$



Calibration coefficients

- Given the complete channel: $h_{i\to j} = tx_i \cdot c \cdot rx_j$

- We define a calibration coefficient as:

$$A_{i\to j} = \frac{h_{i\to j}}{h_{j\to i}} = \frac{tx_i \cdot \cancel{c} \cdot rx_j}{tx_j \cdot \cancel{c} \cdot rx_i} = \frac{tx_i \cdot rx_j}{tx_j \cdot rx_i} = \frac{1}{A_{j\to i}}$$

- Thus:

$h_{i\to j} = A_{i\to j} h_{j\to i}$ and $A_{i\to j} = \frac{A_{1\to j}}{A_{1\to i}}$

Applying to large-scale BS

- Find A between each BS antenna and a reference antenna (1) $A_{1\to m}$
- Every BS radio listens to terminal pilot $h_{t\to m}$
- Find A between reference and terminal $A_{1\to t}$
- We can derive $A_{m\to t} = \dfrac{A_{1\to t}}{A_{1\to m}}$
- Now every *h* can be found via $h_{m\to t} = A_{m\to t} h_{t\to m}$

Key observation

- But this requires K+1 pilots…
  - Even worse, it requires feedback

- A constant phase shift across the entire array does not alter the beampattern!

$$h_{m\to t} = A_{m\to t} h_{t\to m} = \frac{A_{1\to t}}{A_{1\to m}} h_{t\to m} \Rightarrow \frac{1}{A_{1\to m}} h_{t\to m}$$

- Assuming $A_{1\to t} = 1$ results in a constant phase offset, and thus does not affect radiation pattern

Internal calibration

- We find all $A_{1\to m}$ offline
  - They are static, and can be found quickly

- Send K orthogonal pilots to find all $h_{t_k\to m}$
  - Used for uplink beamforming directly

- Use $h_{m\to t} = \dfrac{h_{t\to m}}{A_{1\to m}}$ for downlink beamforming

Overview of contributions

- Scalable architecture
- Internal reciprocity calibration
- **Novel fully distributed beamforming method**

Problem with existing methods

- Central data dependency
- Transport latency causes capacity loss
- Can not scale
  - Becomes exorbitantly expensive then infeasible

Conjugate beamforming

- Requires global power scaling by constant:

$$\mathbf{W}_{conj} = c \cdot \mathbf{H}^*$$

- Where, e.g.:

$$c = \left(\sum_{k=1}^{K} \sum_{m=1}^{M} \|\mathbf{h}_{m,k}^2\| \right)^{-1}$$

- This creates a central data dependency

Local conjugate beamforming

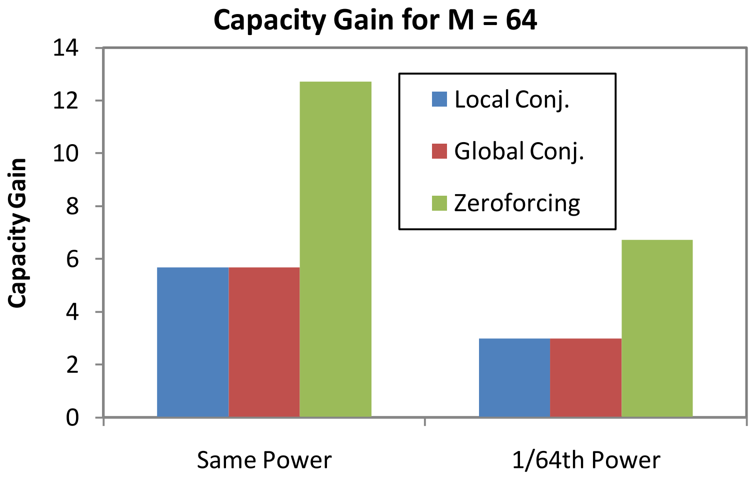
- Scale power locally:

$$c_m = \left(\sum_{k=1}^{K} \|\mathbf{h}_{m,k}^2\| \right)^{-1} \quad (m = 1, 2, \ldots M)$$

- Maximizes utilization of every radio
  - More appropriate for real-world deployments
- Quickly approaches optimal as K increases
  - Channels are independent and uncorrelated

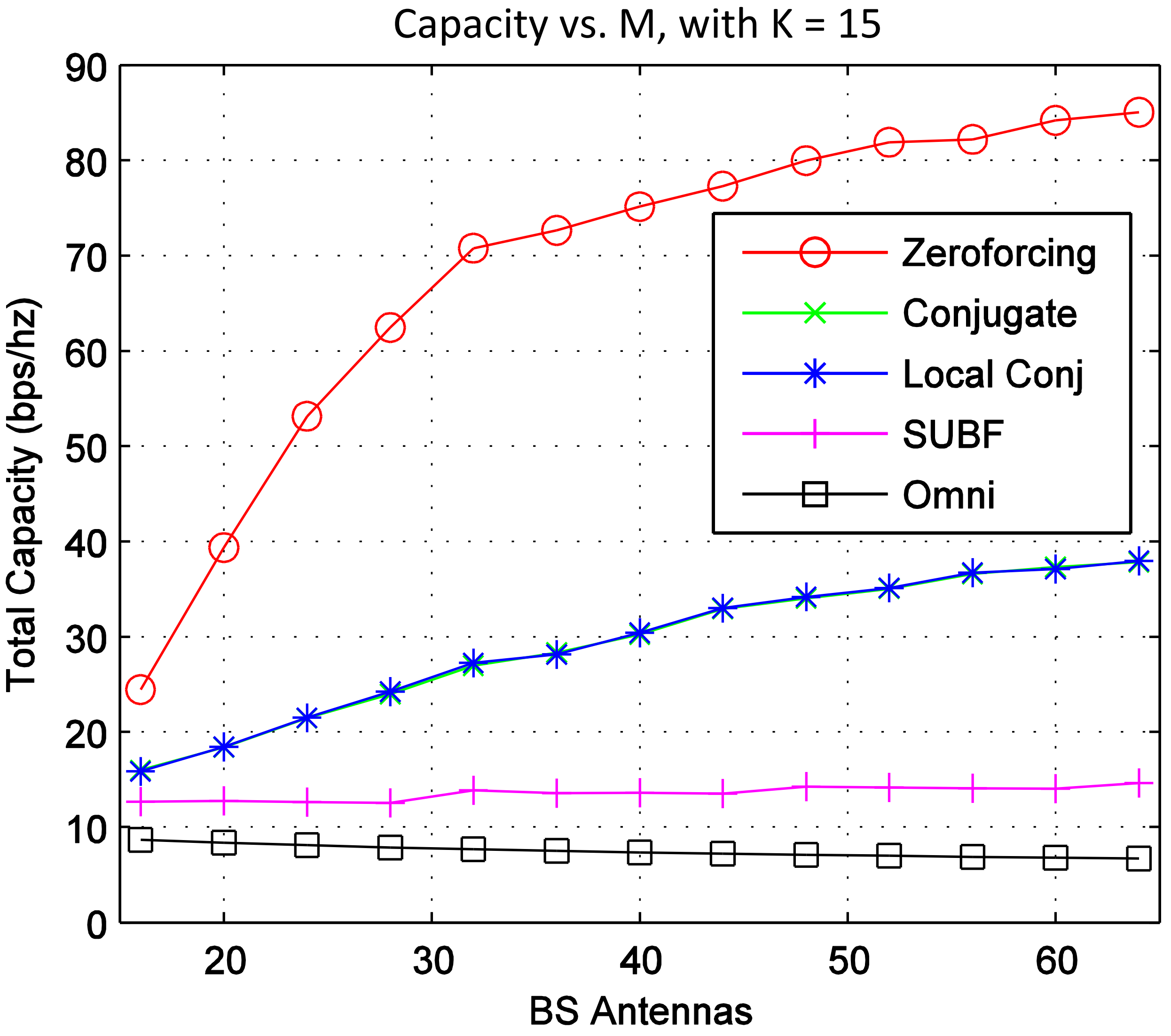
Results

- Huge Capacity Gains

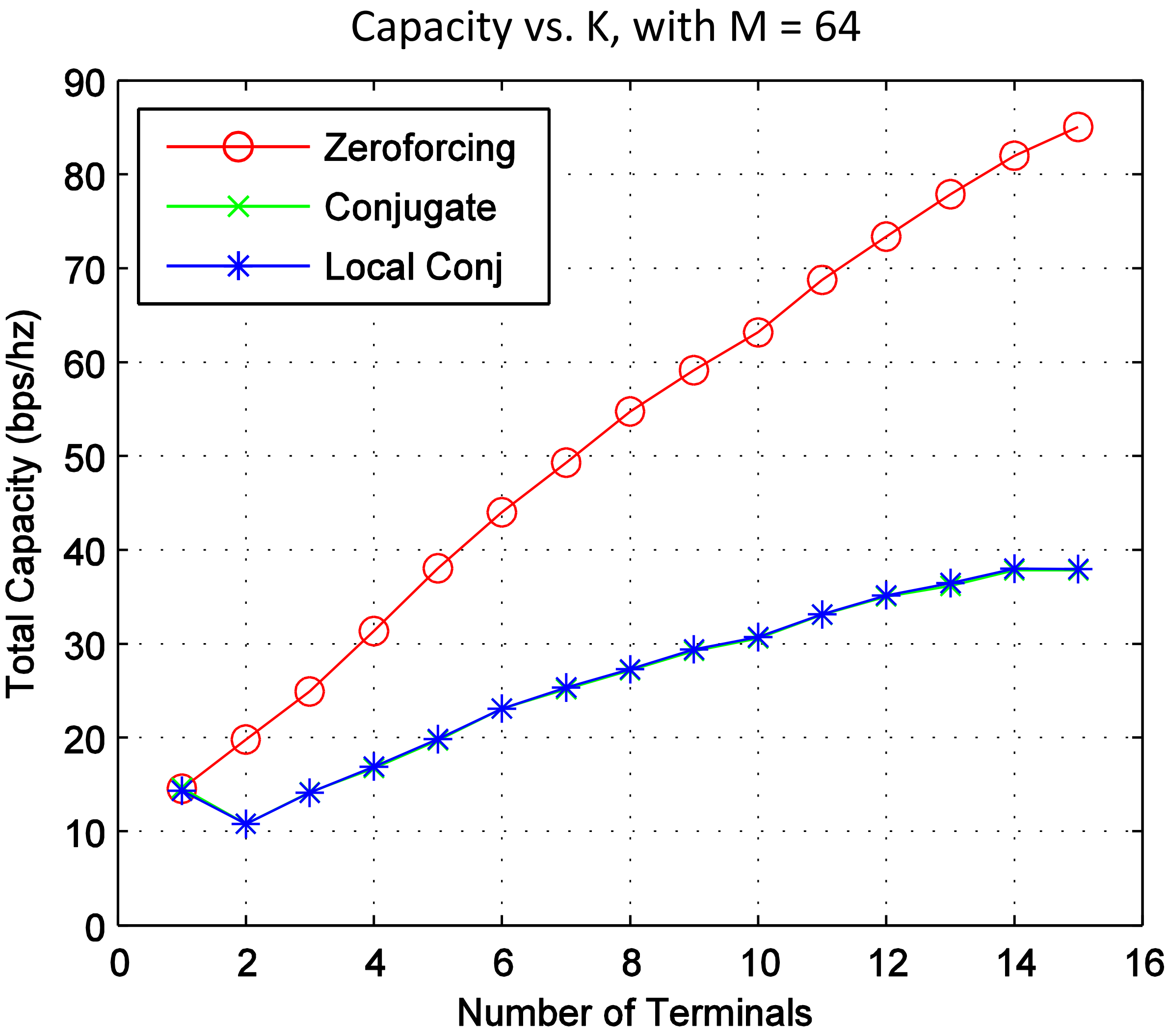


- Performance linear with M and K
- Channel Calibration Stable
- Local conjugate indistinguishable from global
  - Approaches optimality quickly with K

Results: scaling M

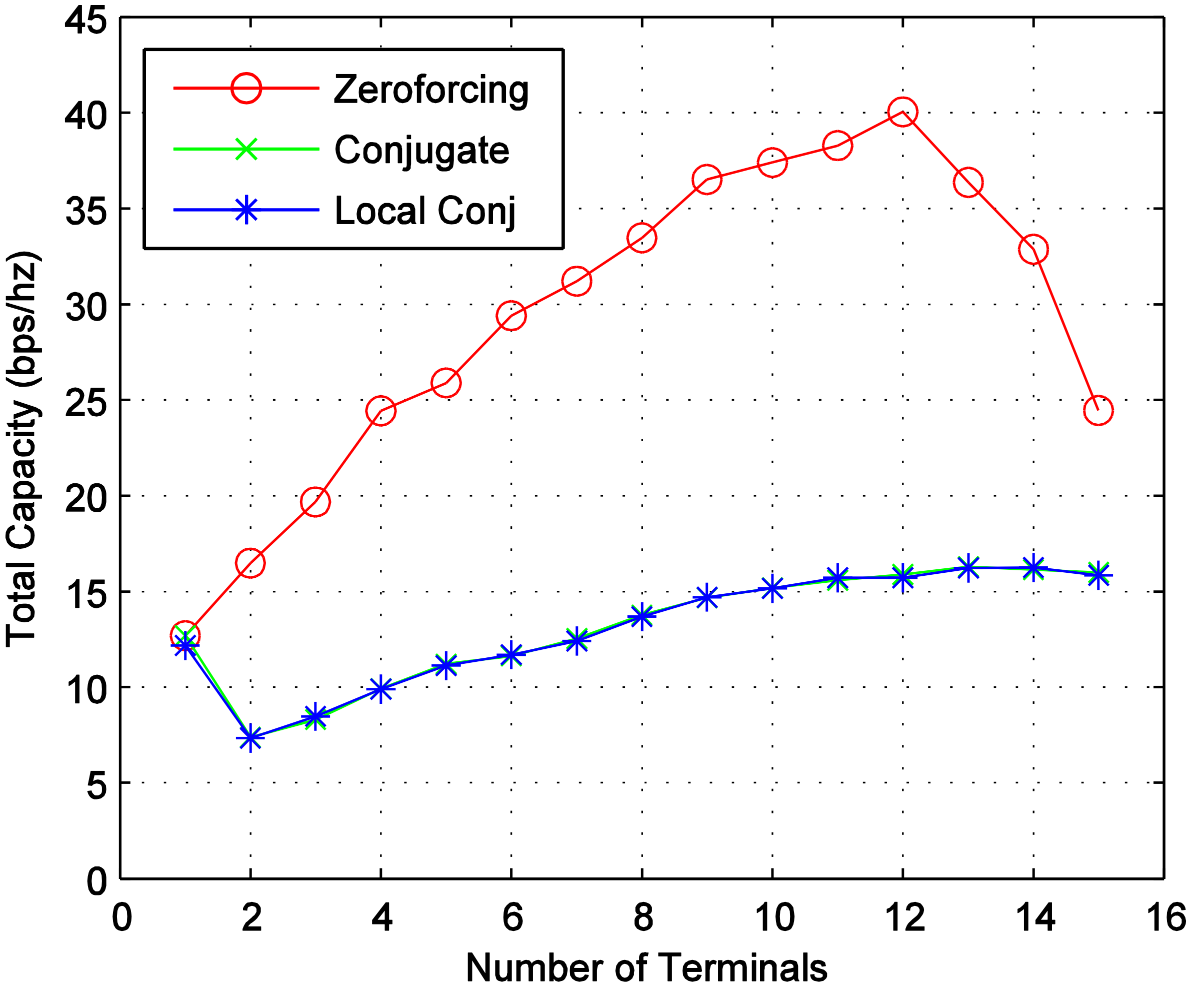


Results: scaling K

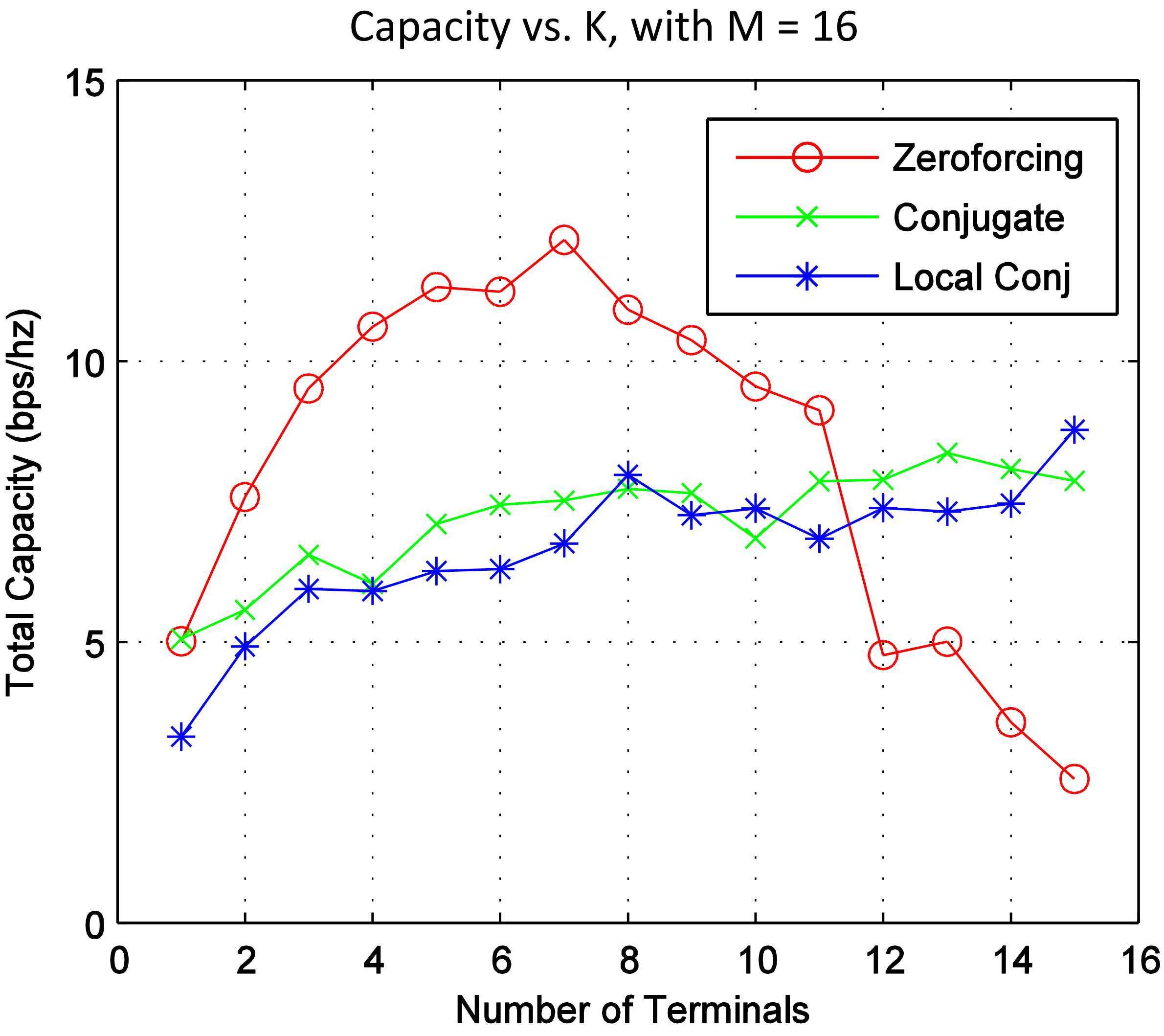


Results: scaling K

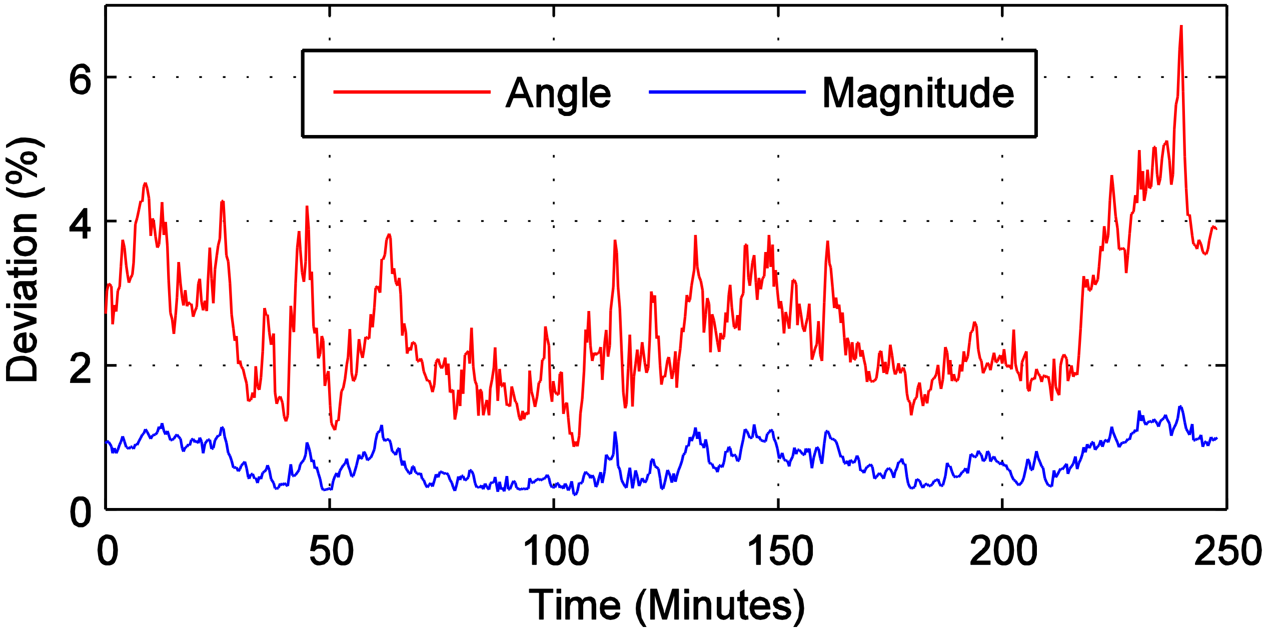
Capacity vs. K, with M = 16



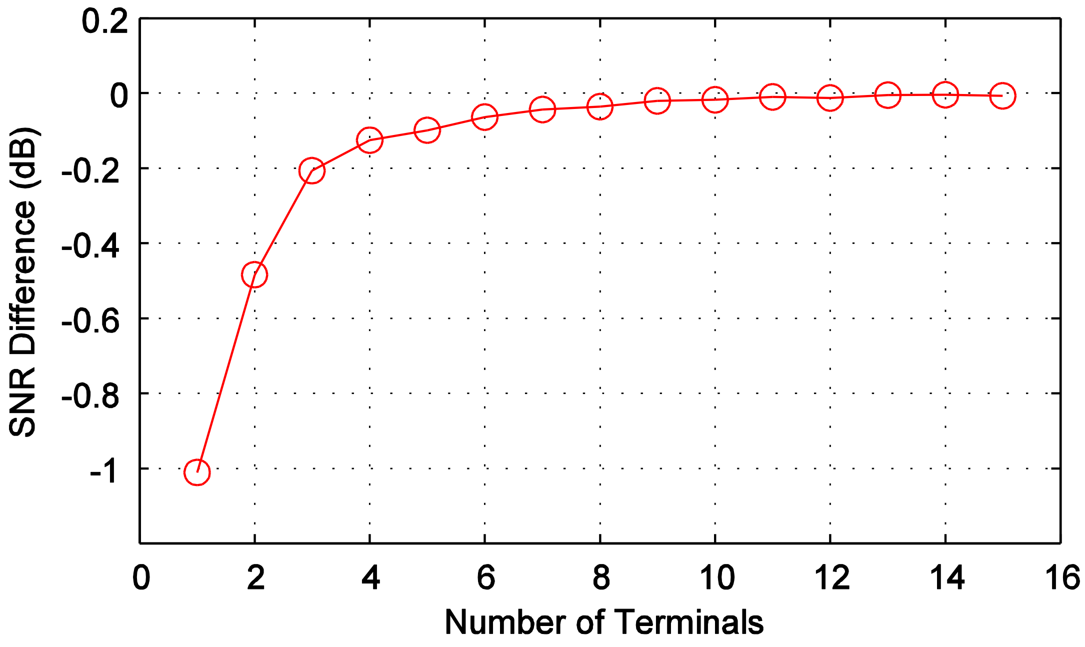
Results: low power



Results: calibration stability



Results: local conjugate



Future directions

- Find optimal tradeoff between zeroforcing and conjugate
- Demonstrate network optimality
  - Lower power reduces other-cell interference
  - Leverage cooperative beamforming
- Investigate promising match with full duplex
  - Leverage huge EIRP gains

Conclusion

- First large-scale beamforming platform
  - Real-world demonstration of manyfold capacity increase
- Devised novel architecture and techniques
  - Unlimited Scalability

Acknowledgements

- Theoretical Discussion and Background
  - Ashutosh Sabharwal
- WARP Support
  - Patrick Murphy, Gaurav Patel, Chris Hunter, Sidharth Gupta
- Platform Construction
  - Nathan Zuege, Chris Harris, Azalia Mirhoseini, Danny Eaton, Paul Williams