MASSIVE MIMO AND BEYOND

Thomas L. Marzetta
Bell Labs, Alcatel-Lucent

Munich Workshop on Massive MIMO

7 October, 2015

You can always lay down more optical fiber; you can never lay down more spectrum!

Bell Labs

SPECTRUM BELOW 5 GHZ: THE MOST VALUABLE RESOURCE IN THE WORLD!

- FCC AWS-3 spectrum auction, January 2015
  - 65 MHz: 1695-1710 MHz, 1755-1780 MHz, 2155-2180 MHz
  - $41.3 billion

Bell Labs

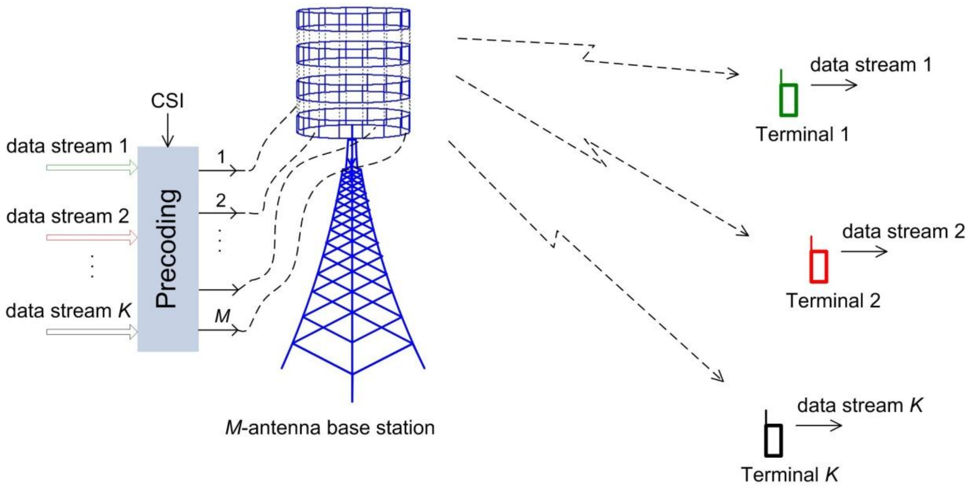
SPECTRUM BELOW 5 GHZ: THE MOST VALUABLE RESOURCE IN THE WORLD!

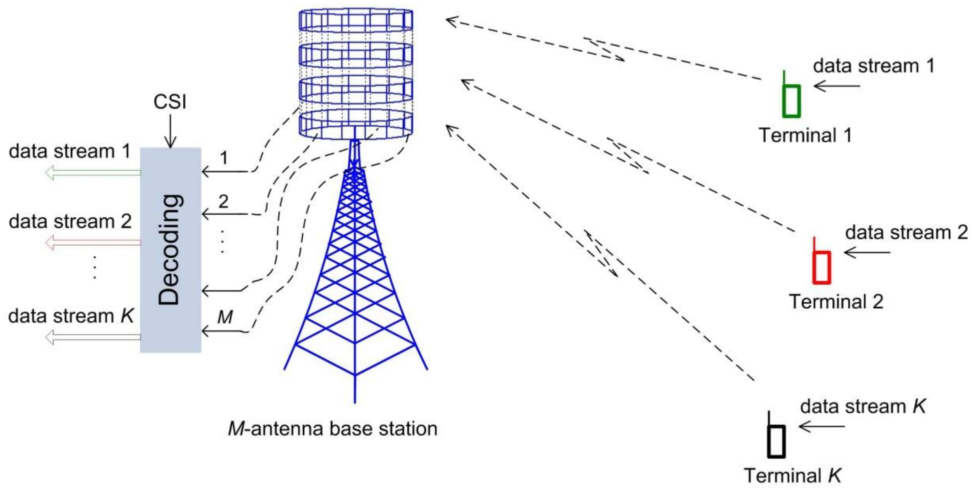
- FCC AWS-3 spectrum auction, January 2015
  - 65 MHz: 1695-1710 MHz, 1755-1780 MHz, 2155-2180 MHz
  - $41.3 billion
- *FCC Commissioner Jessica Rosenworcel, October 2, 2014:*
- “What if we issued a challenge in Washington? … Imagine that we decided to reward the first person who finds a way to make spectrum use below 5 GHz 50 or 100 times more efficient over the next decade. The reward could be something simple—say 10 megahertz of spectrum suitable for mobile broadband.”

Only advancements in the physical layer can meet this challenge

Bell Labs

SPATIAL MULTIPLEXING PUSHED TO AN EXTREME





Massive MIMO serves all users over the *same* time/frequency resources

WHAT IS MASSIVE MIMO?

- Essentials
  - many physically small, low power antennas
  - aggressive spatial multiplexing
  - utilize *measured* channels
- Benefits
  - scalability
  - spectral efficiency
  - simplicity
  - great service to *all* users
  - energy efficiency

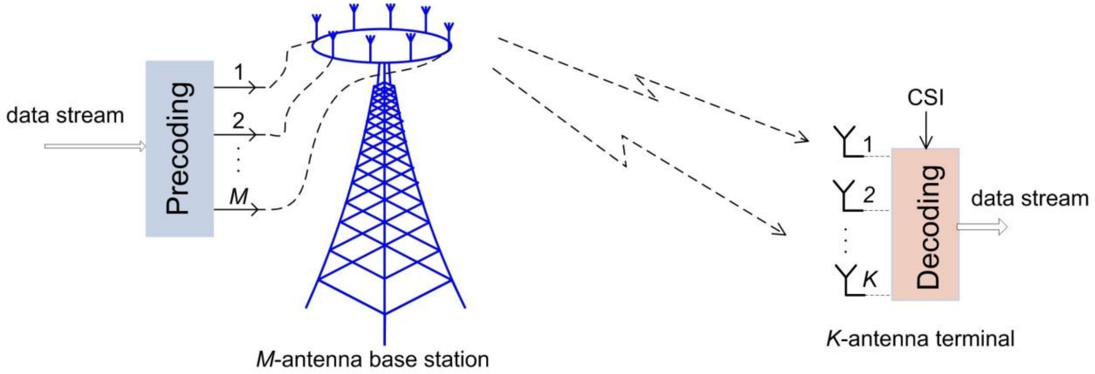
Massive MIMO is a game-changer

OUTLINE

- Information theoretic evolution of MIMO
- Science of Massive MIMO
- Case study
  - Optimum pilot re-use
  - Maximum-ratio vs. zero-forcing

POINT-TO-POINT MIMO

ROY & OTTERSTEN (1991); PAULRAJ & KAILATH (1993); FOSCHINI (1995); RALEIGH & CIOFFI (1998); TELATAR (1999)



- Brilliant invention
- But not scalable

  - unfavorable propagation
  - time required for training grows with system size
  - disappointing multiplexing gains at cell edges
  - 8x4 link, -3.0 dB SNR:

# base station antennas	1	2	4	8
bits/second/Hz	1.51	1.83	2.06	2.19

It's critically important to give uniformly good service throughout the cell

MULTI-USER MIMO

CAIRE & SHAMAI (2003); VISWANATH & TSE (2003); VISHWANATH, JINDAL, & GOLDSMITH (2003)

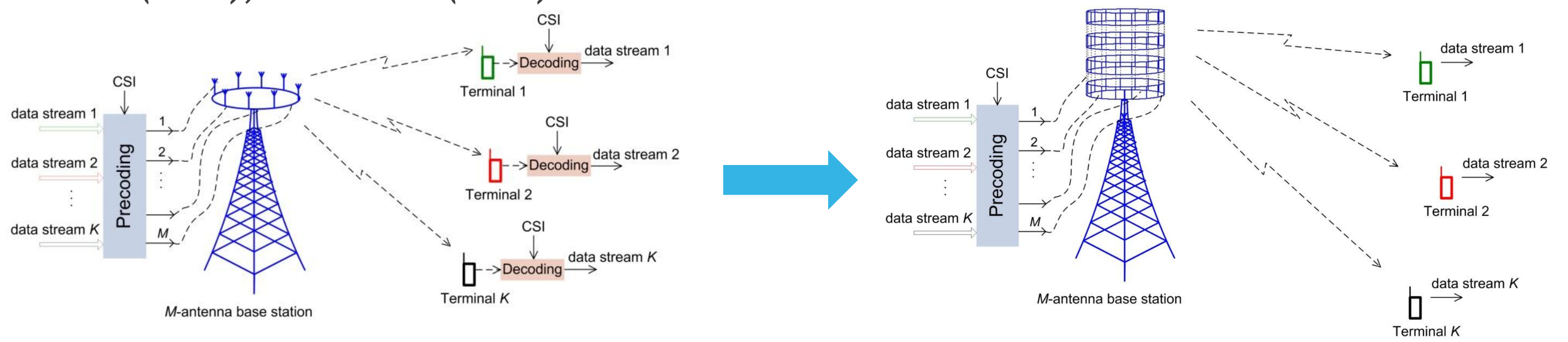
- Splitting the multi-antenna user into autonomous single-antenna users doesn't decrease the sum-throughput!
- Only single-antenna terminals required
- Propagation is almost always favorable
- But not scalable in its original form
  - dirty-paper coding/decoding needed
  - both ends of link have to know channel

A triumph of Shannon information theory, but not really practical as is

Bell Labs

MASSIVE MIMO

MARZETTA (2006); MARZETTA (2010)



- Add many more base station antennas
- Ignore the dictates of Shannon theory
  - channel state information (CSI) only available to the base station
  - use linear pre-coding/de-coding instead of dirty-paper
  - users don't do any signal processing

The large number of antennas paradoxically makes the problem simpler

Bell Labs

MASSIVE MIMO: MORE THAN JUST MANY ANTENNAS

- Using *measured* channels: Beamforming gain grows linearly with number of antennas, irrespective of the noisiness of the measurements
- Frequency-independent power control: Based solely on long-scale (slow) fading, is exceedingly effective
- Pilot contamination: Ultimate limitation in non-cooperative multi-cell systems

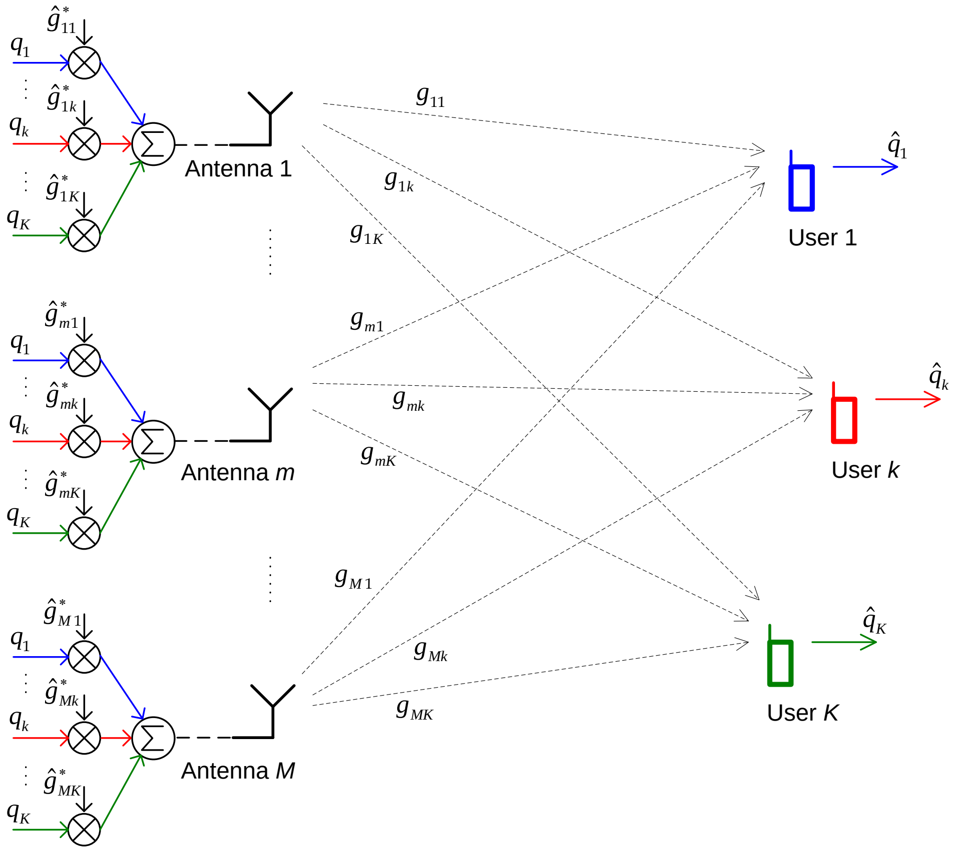
No new mathematics, but a new philosophy!

Bell Labs

DOWNLINK DATA TRANSMISSION: CONJUGATE BEAMFORMING

ANTENNAS TRANSMIT THE WEIGHTED MESSAGE-BEARING SYMBOLS TO ARRIVE IN-PHASE AT THE INTENDED USER & OUT-OF-PHASE ELSEWHERE

- Information-bearing symbols combined with *measured* channel characteristics to create transmitted signals
- Decentralized array architecture



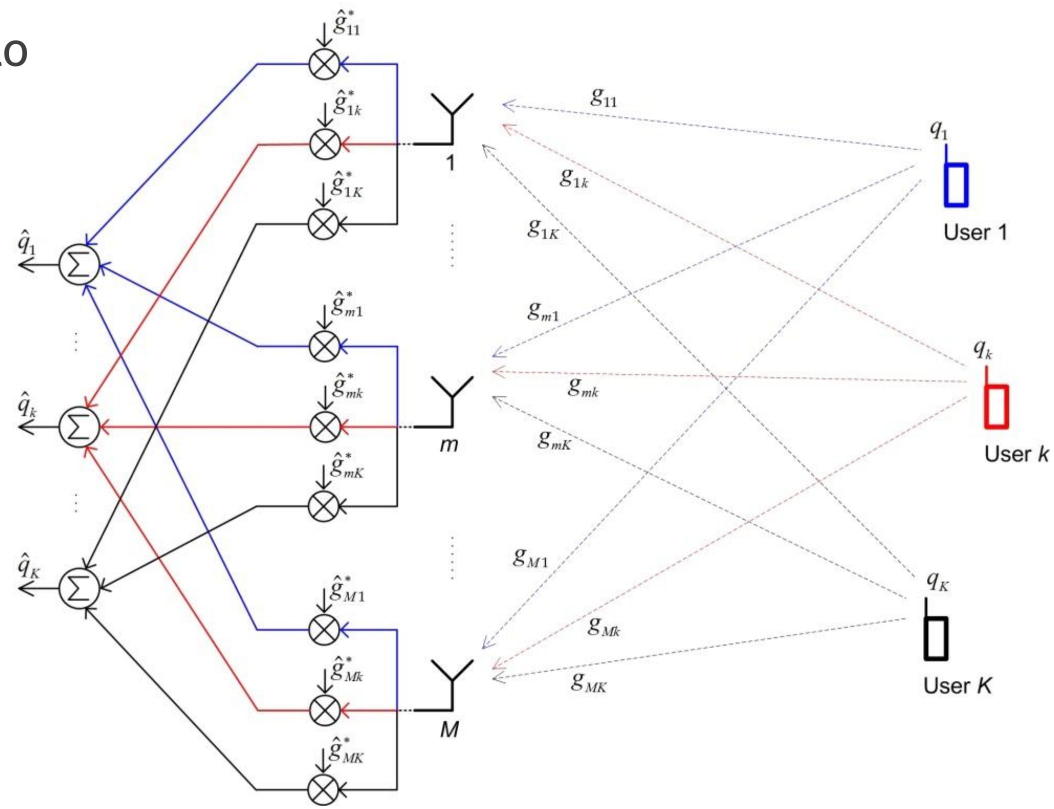
The simplest possible pre-coding, but often very effective

Bell Labs

UPLINK DATA TRANSMISSION: MATCHED FILTERING

BASE STATION WEIGHTS AND ADDS RECEIVED SIGNALS FOR CONSTRUCTIVE REINFORCEMENT OF THE TRANSMISSION FROM EACH USER

- Properties and advantages similar to conjugate beamforming

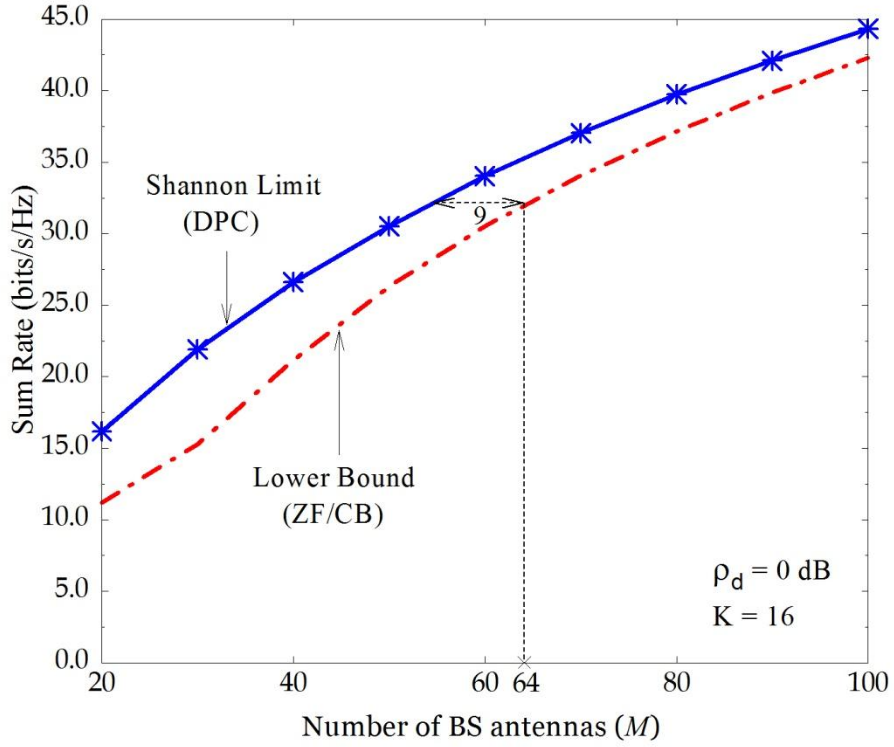


For high SINRs, zero-forcing may outperform conjugate beamforming/matched-filtering

Bell Labs

EXCESS ANTENNAS MAKE MASSIVE MIMO NEARLY OPTIMAL

0.0 dB SNR   16 users

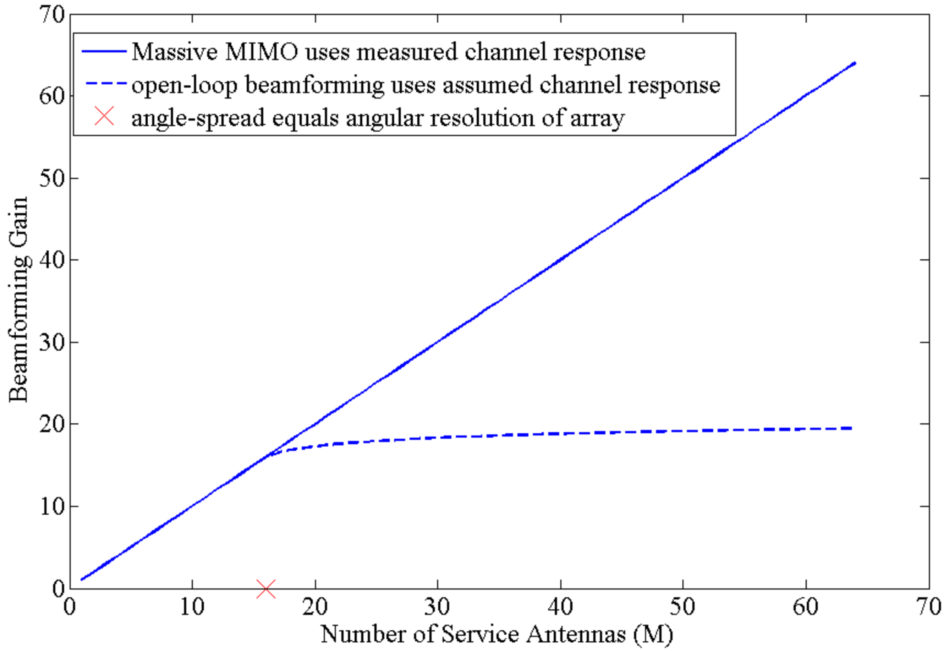


Any sub-optimality can be compensated for with more antennas

Bell Labs

WHY SO IMPORTANT TO DO BEAMFORMING WITH *MEASURED* PROPAGATION?

- *Measured* channels
  - scalable
  - gain grows linearly with number of antennas
    - irrespective of noisiness of CSI
    - no tightening of array tolerance required
- *Assumed* channels
  - not scalable
  - gain eventually grows only logarithmically

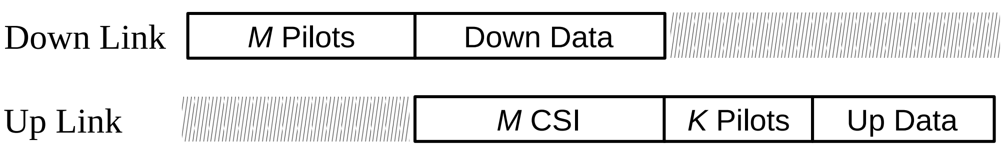


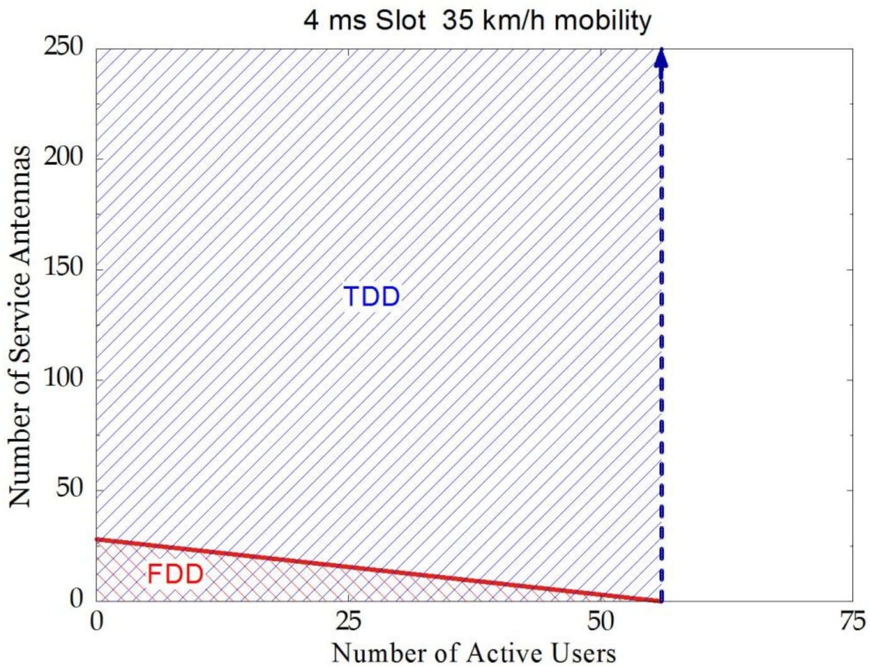
If open-loop beamforming, then not Massive MIMO!

TDD SLOT STRUCTURE ENSURES TIMELY CHANNEL-STATE INFORMATION: M SERVICE-ANTENNAS, K USERS

- TDD slot: training time $\propto K$

- FDD slot: training time $\propto 2M + K$





Mobility limits the number of active users; FDD is a disaster!

PILOT CONTAMINATION

- For mobile users, there is a limited number of orthogonal pilots
- When the same pilot is transmitted by more than one user:
  - base station obtains a *linear combination* of channels
    - extra pilot power doesn't help
  - coherent interference
    - doesn't disappear with more antennas

Pilot contamination has always existed, but was never noticed!

Bell Labs

WHAT DOES SHANNON THEORY HAVE TO SAY?

- Assume block fading and no available CSI
- Downlink: *One for all, and all for one!*
  - Assume successful multiplexing so each user receives his own intended data
  - The strongest user can decode everybody's data stream, so the sum-throughput is no better than the MISO throughput to that user
  - Therefore no multiplexing gains possible
- Uplink: throughput is upper-bounded by Point-to-Point MIMO capacity
  - Assume i.i.d. Rayleigh fading, and no near-far effects

$$\overbrace{\mathbf{Y}}^{M\times T} = \sqrt{\frac{\rho}{M}}\overbrace{\mathbf{G}}^{M\times K}\overbrace{\mathbf{X}}^{K\times T} + \mathbf{W}, \quad K > T$$

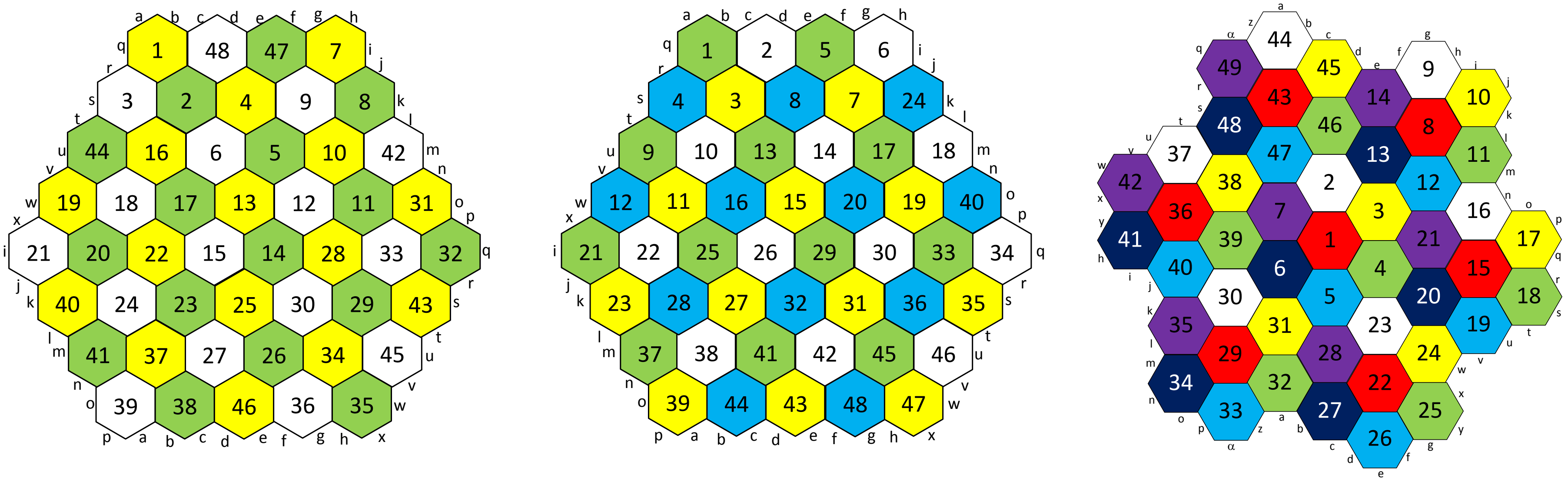
$$= \sqrt{\frac{\rho}{M}}\overbrace{\mathbf{G}}^{M\times K}\overbrace{\mathbf{\Phi}}^{K\times K}\overbrace{\mathbf{D}_{\mathbf{v}}}^{K\times T}\overbrace{\mathbf{\Psi}^{\mathrm{H}}}^{T\times T} + \mathbf{W} = \sqrt{\frac{\rho}{M}}\overbrace{\mathbf{G}'}^{M\times K}\overbrace{\mathbf{D}_{\mathbf{v}}}^{K\times T}\overbrace{\mathbf{\Psi}^{\mathrm{H}}}^{T\times T} + \mathbf{W}$$

  - No point in powering more than K=T users

We need a real breakthrough here!

Bell Labs

PILOT RE-USE FACTOR 3, 4, 7: PUSH CONTAMINATING CELLS FARTHER AWAY FROM HOME CELL

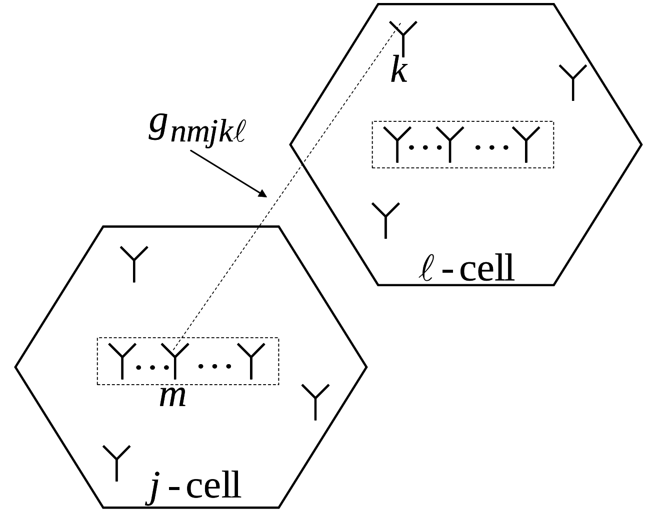


The cost: extra overhead

Bell Labs

•Propagation Model

$$g_{mk} = \underbrace{\beta_k^{1/2}}_{\text{slow}} \cdot \underbrace{h_{mk}}_{\text{fast}}, \quad m: \text{ antenna} \quad k: \text{ terminal}$$



- Slow fading and fast fading
- Slow fading comprises geometric attenuation (Hata Model) combined with log-normal shadow fading
  - Constant with respect to frequency and service antenna
  - Easy to estimate
  - Assumed known a-priori
- Fast fading
  - Rayleigh CN(0,1), iid with respect to antenna, terminal
  - Piecewise constant and iid from one frequency smoothness interval to another
  - Unknown a-priori
  - Estimated from up link pilots & TDD reciprocity

Bell Labs

DOWN LINK DATA: LINEAR PRE-CODING

$$\underbrace{\mathbf{s}}_{M\times 1} = \underbrace{\mathbf{A}}_{M\times K} \underbrace{\mathbf{D}_{\boldsymbol{\eta}}^{1/2}}_{K\times 1} \underbrace{\mathbf{q}}_{K\times 1}$$

$\{q_k\}$: iid, CN(0,1) $\qquad E\{\mathbf{s}^{\mathrm{H}}\mathbf{s}\}=1 \qquad$ power-control : $\boldsymbol{\eta} \geq \mathbf{0},\ \mathbf{1}^{\mathrm{T}}\boldsymbol{\eta}=1$

- Conjugate beam-forming

$$\mathbf{A} = \frac{1}{\sqrt{M}}\hat{\mathbf{G}}^{*}\mathbf{D}_{\boldsymbol{\gamma}}^{-1/2}$$

  - permits de-centralized architecture and processing

- Zero-forcing

$$\mathbf{A} = \sqrt{M-K}\,\hat{\mathbf{G}}^{*}\left(\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}}^{*}\right)^{-1}\mathbf{D}_{\boldsymbol{\gamma}}^{1/2}$$

  - Implementing linear pre-coding takes more computations than QR factorization!

DOWN LINK: CONJUGATE BEAM-FORMING

$$\mathbf{x} = \sqrt{\frac{\rho_{\mathrm{f}}}{M}}\mathbf{G}^{\mathrm{T}}\hat{\mathbf{G}}^{*}\mathbf{D}_{\boldsymbol{\gamma}}^{-1/2}\mathbf{D}_{\boldsymbol{\eta}}^{1/2}\mathbf{q}+\mathbf{w} = \sqrt{\frac{\rho_{\mathrm{f}}}{M}}\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}}^{*}\mathbf{D}_{\boldsymbol{\gamma}}^{-1/2}\mathbf{D}_{\boldsymbol{\eta}}^{1/2}\mathbf{q}+\mathbf{w} - \sqrt{\frac{\rho_{\mathrm{f}}}{M}}\tilde{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{G}}^{*}\mathbf{D}_{\boldsymbol{\gamma}}^{-1/2}\mathbf{D}_{\boldsymbol{\eta}}^{1/2}\mathbf{q}$$

$$x_k = \sqrt{\frac{\rho_{\mathrm{f}}\eta_k}{M\gamma_k}}\hat{\mathbf{g}}_k^{\mathrm{T}}\hat{\mathbf{g}}_k^{*}q_k + w_k - \sqrt{\frac{\rho_{\mathrm{f}}}{M}}\tilde{\mathbf{g}}_k^{\mathrm{T}}\hat{\mathbf{G}}^{*}\mathbf{D}_{\boldsymbol{\gamma}}^{-1/2}\mathbf{D}_{\boldsymbol{\eta}}^{1/2}\mathbf{q} + \hat{\mathbf{g}}_k^{\mathrm{T}}\cdot\sum_{n\neq k}\sqrt{\frac{\rho_{\mathrm{f}}\eta_n}{M\gamma_n}}\hat{\mathbf{g}}_n^{*}q_n$$

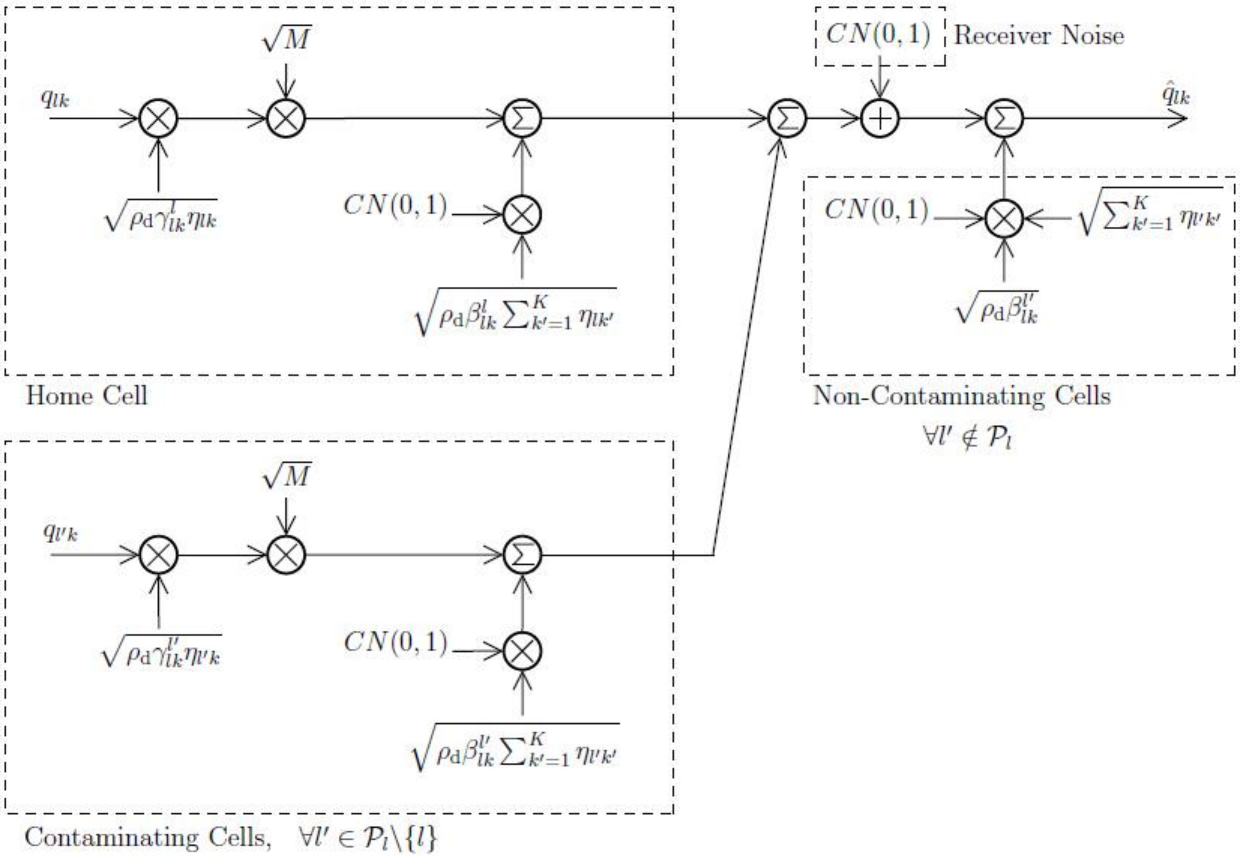
$$= \underbrace{\sqrt{\frac{\rho_{\mathrm{f}}\eta_k}{M\gamma_k}}\mathrm{E}\left\{\hat{\mathbf{g}}_k^{\mathrm{T}}\hat{\mathbf{g}}_k^{*}\right\}q_k}_{(0)} + \underbrace{w_k}_{(1)} - \underbrace{\sqrt{\frac{\rho_{\mathrm{f}}}{M}}\tilde{\mathbf{g}}_k^{\mathrm{T}}\hat{\mathbf{G}}^{*}\mathbf{D}_{\boldsymbol{\gamma}}^{-1/2}\mathbf{D}_{\boldsymbol{\eta}}^{1/2}\mathbf{q}}_{(2)}$$

$$+ \underbrace{\hat{\mathbf{g}}_k^{\mathrm{T}}\cdot\sum_{n\neq k}\sqrt{\frac{\rho_{\mathrm{f}}\eta_n}{M\gamma_n}}\hat{\mathbf{g}}_n^{*}q_n}_{(3)} + \underbrace{\sqrt{\frac{\rho_{\mathrm{f}}\eta_k}{M\gamma_k}}\left(\hat{\mathbf{g}}_k^{\mathrm{T}}\hat{\mathbf{g}}_k^{*} - \mathrm{E}\left\{\hat{\mathbf{g}}_k^{\mathrm{T}}\hat{\mathbf{g}}_k^{*}\right\}\right)q_k}_{(4)}$$

(0) desired signal	(1) receiver noise	(2) channel estimation error	(3) channel non-orthogonality	(4) beam-forming gain uncertainty
$M\rho_{\mathrm{f}}\eta_k\gamma_k$	1	$\rho_{\mathrm{f}}(\beta_k-\gamma_k)$	$\rho_{\mathrm{f}}\gamma_k\sum_{n\neq k}\eta_n$	$\rho_{\mathrm{f}}\gamma_k\eta_k$

$$\mathrm{SINR}_k = \frac{\eta_k M\rho_{\mathrm{f}}\gamma_k}{1+\rho_{\mathrm{f}}(\beta_k-\gamma_k)+\rho_{\mathrm{f}}\gamma_k\sum_{n\neq k}\eta_n+\rho_{\mathrm{f}}\gamma_k\eta_k} = \frac{\eta_k M\rho_{\mathrm{f}}\gamma_k}{1+\rho_{\mathrm{f}}\left(\sum_{n=1}^{K}\eta_n\right)\beta_k}$$

DOWNLINK MAXIMUM-RATIO: EFFECTIVE CHANNEL



$$\mathrm{SINR}_{\ell k} = \frac{M\rho_\mathrm{f}\eta_{k\ell}\gamma_{\ell k\ell}}{1+\sum_{j\in \mathrm{all}} \rho_\mathrm{f}\left(\sum_{n=1}^{K}\eta_{nj}\right)\beta_{jk\ell} + \sum_{j\in \mathrm{PC}_\ell} M\rho_\mathrm{f}\eta_{kj}\gamma_{jk\ell}}, \quad \beta : \text{m.s. channel}, \gamma : \text{m.s. estimate}$$

Massive MIMO creates a flat channel to each terminal

Bell Labs

SINR FOR K-TH TERMINAL IN ELL-TH CELL

β: m.s. channel γ: m.s. estimate $\beta-\gamma$: m.s. error η: power control

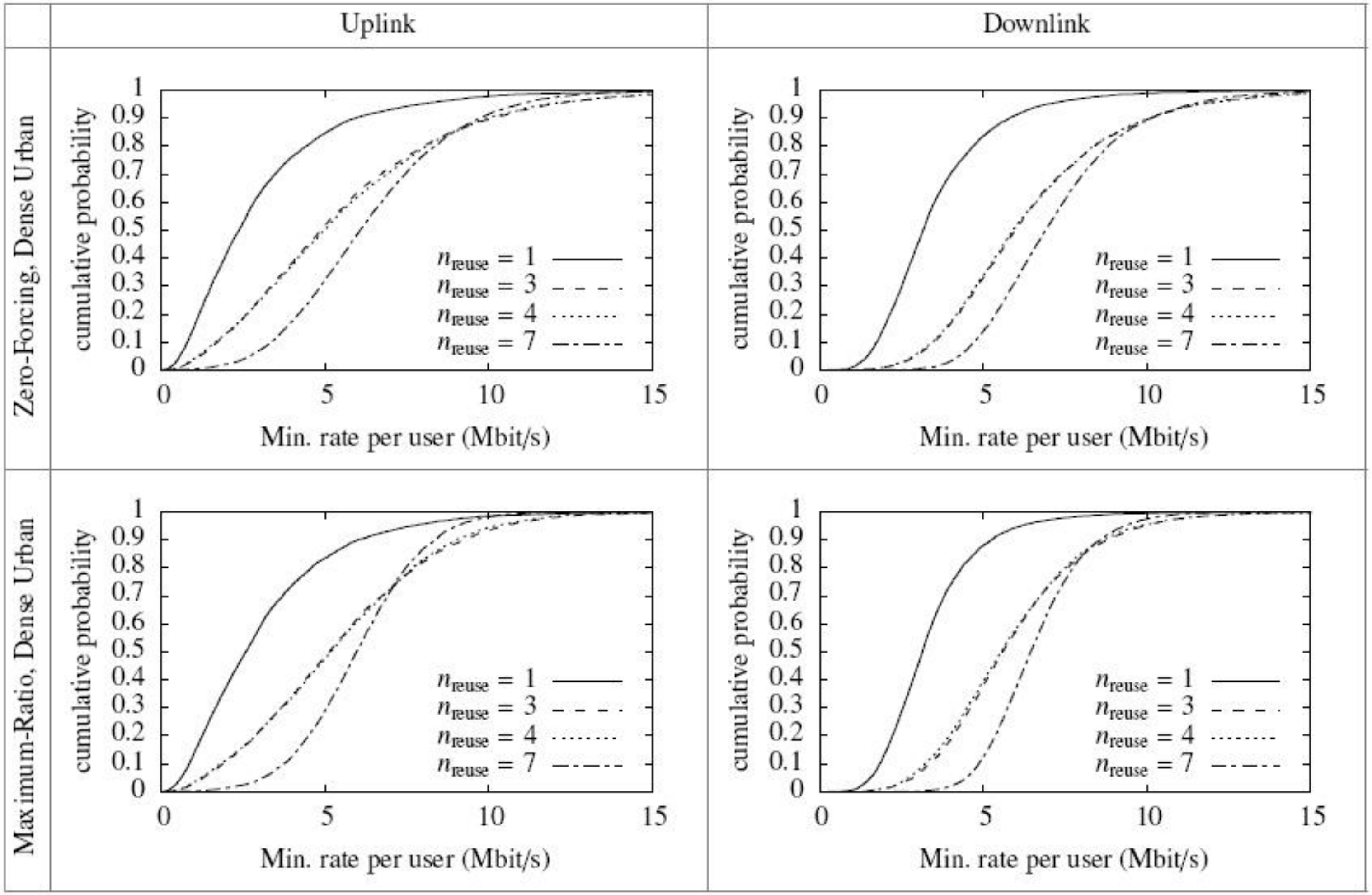
	Conjugate Beam Forming / Matched Filtering	Zero Forcing
Down Link	$\dfrac{M\rho_{\mathrm{f}}\eta_{k\ell}\gamma_{\ell k\ell}}{1+\sum_{j\in \mathrm{all}}\rho_{\mathrm{f}}\left(\sum_{n=1}^{K}\eta_{nj}\right)\beta_{jk\ell}+\sum_{j\in \mathrm{PC}_\ell}M\rho_{\mathrm{f}}\eta_{kj}\gamma_{jk\ell}}$	$\dfrac{(M-K)\rho_{\mathrm{f}}\eta_{k\ell}\gamma_{\ell k\ell}}{1+\sum_{j\notin \ell\cup \mathrm{PC}_\ell}\rho_{\mathrm{f}}\left(\sum_{n=1}^{K}\eta_{nj}\right)\beta_{jk\ell}+\sum_{j\in \ell\cup \mathrm{PC}_\ell}\rho_{\mathrm{f}}\left(\sum_{n=1}^{K}\eta_{nj}\right)(\beta_{jk\ell}-\gamma_{jk\ell})+\sum_{j\in \mathrm{PC}_\ell}(M-K)\rho_{\mathrm{f}}\eta_{kj}\gamma_{jk\ell}}$
Up Link	$\dfrac{M\rho_{\mathrm{r}}\eta_{k\ell}\gamma_{\ell k\ell}}{1+\sum_{j\in \mathrm{all}}\rho_{\mathrm{r}}\sum_{n=1}^{K}\eta_{nj}\beta_{\ell nj}+\sum_{j\in \mathrm{PC}_\ell}M\rho_{\mathrm{r}}\eta_{kj}\gamma_{\ell kj}}$	$\dfrac{(M-K)\rho_{\mathrm{r}}\eta_{k\ell}\gamma_{\ell k\ell}}{1+\sum_{j\notin \ell\cup \mathrm{PC}_\ell}\rho_{\mathrm{r}}\sum_{n=1}^{K}\eta_{nj}\beta_{\ell nj}+\sum_{j\in \ell\cup \mathrm{PC}_\ell}\rho_{\mathrm{r}}\sum_{n=1}^{K}\eta_{nj}(\beta_{\ell nj}-\gamma_{\ell nj})+\sum_{j\in \mathrm{PC}_\ell}(M-K)\rho_{\mathrm{r}}\eta_{kj}\gamma_{\ell kj}}$

Inequality constraints on SINR equivalent to linear inequality constraints on power control variables

CASE STUDIES: OPTIMUM PILOT RE-USE FACTOR; MAXIMUM-RATIO VS. ZERO-FORCING

	Dense Urban	Suburban
Carrier frequency(GHz)	1.9	1.9
TDD spectral bandwidth (MHz)	20	20
Slot duration (ms)	2	1
User allowed mobility (km/h)	71	142
Uplink radiated power/user (mW)	200	200
Number of service antennas	64	256
Total downlink radiated power (W)	1	1
Active users/cell	18	18
Cell radius (km)	.50	2.0
Power control	Max/min	Max/min
95% likely throughput/terminal Mb/s zero-forcing	4.1 down, 2.6 up	3.1 down, 1.1 up
95% likely throughput/terminal Mb/s maximum-ratio	4.5 down, 3.1 up	3.2 down, 1.1 up

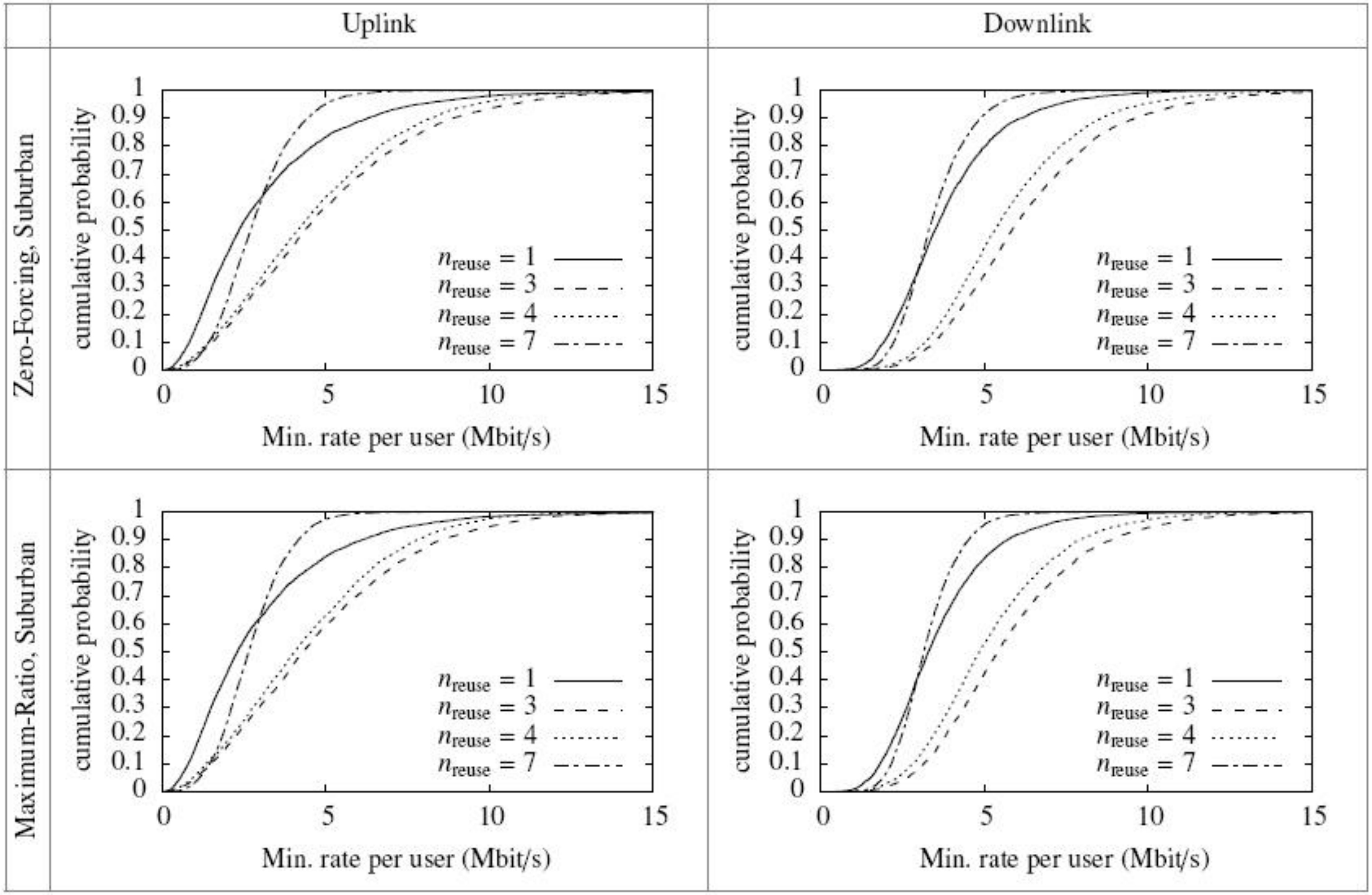
NET THROUGHPUT PER USER: DENSE URBAN



Maximum-ratio + pilot re-use 7 best for Dense Urban

Bell Labs

NET THROUGHPUT PER USER: SUBURBAN



Maximum-ratio + pilot re-use 3 best for Suburban

Bell Labs

NON-CELLULAR MASSIVE MIMO

- Backhaul for small-cells
- Cell-Free Massive MIMO
  - M randomly distributed access points serve K users over an entire city
- Fixed wireless access to homes
- Multicasting
  - Deliberately create and take advantage of pilot contamination!

MASSIVE MIMO IN NON-ELECTROMAGNETIC MEDIA

- Hyperbolic
  - acoustic waves
  - elastic waves
- Parabolic
  - heat diffusion
- Eliptical
  - electric current: *Ground Telegraphy* (Richard Courant, Arnold Sommerfeld, Lee de Forest)

MASSIVE SENSOR TELEMETRY

- Outstanding examples of signal processing
  - 3D reflection seismology
  - Computer tomography
  - Synthetic aperture radar
- Essential to sample data spatially and temporally at Nyquist rate
  - Proper sampling and huge data sets make physics-based signal processing easier!
- Uplink Massive MIMO transports data intact and in real time

Massive MIMO means more than entertainment enablement

IS MASSIVE MIMO THE END OF THE LINE?