Caches as a new PHY Resource in 6G?

(WFF 2023)

Petros ELIA

(Professor - Communication Systems Department)
EURECOM
Sophia Antipolis, France

European Research Council

EURECOM Sophia Antipolis

This talk is about delivering VoD content….

- Video on-demand (VoD): $\geq$ 70% of traffic

Communicating VoD is very expensive

- For wireless network providers
  - Costly bandwidth and infrastructure
- Content providers (Netflix, Amazon, etc.) pay large fees per content

Pros for Caches in PHY

- Most VoD content is cacheable
- Networks are already full of caches (your cell phone)
- Caches are currently used the wrong way (push closer to users)

Consider Caches as a New Resource for PHY

- Use caches to alter <u>structure</u> (not volume) of the PHY problem

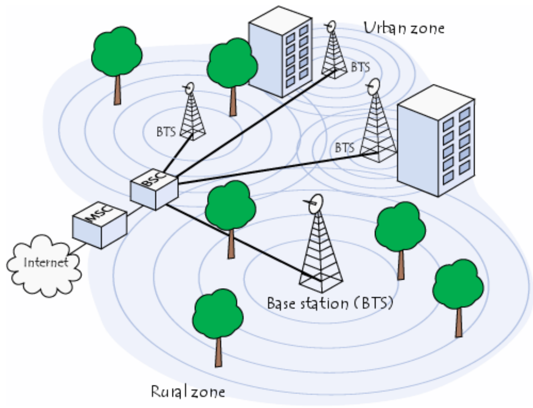
Possible new 6G PHY Resource

Historically a new resource (e.g. MIMO) brings:

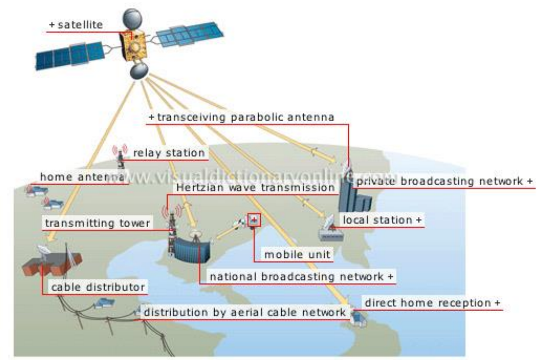
- Big algorithmic gains
- Interesting algorithmic challenges
- New resource must work well with existing solutions

New resource applies in several downlink scenarios

Wired or wireless communications

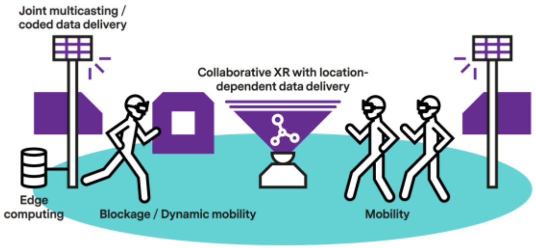


Satellite communications



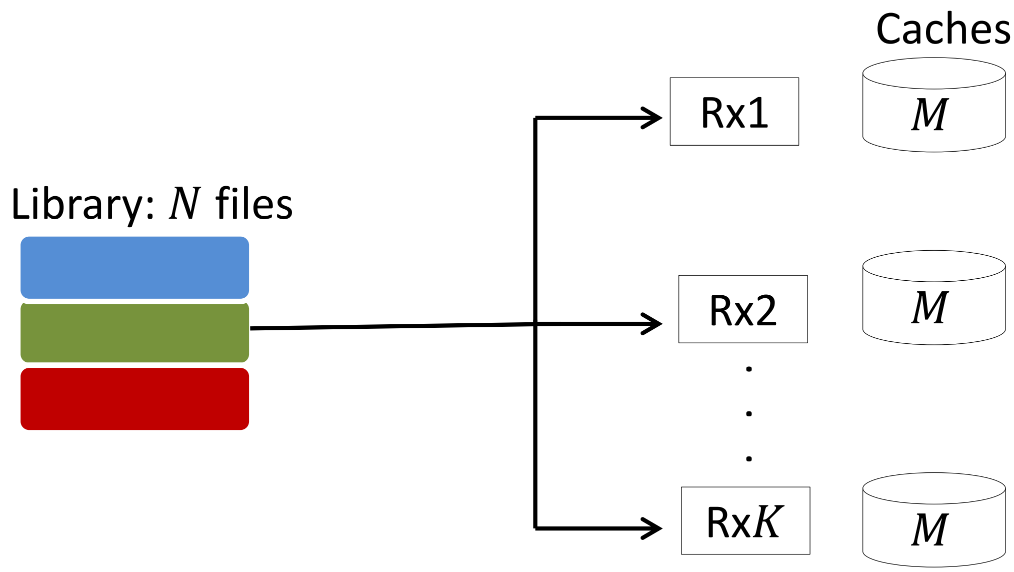
In-flight communications

Virtual Reality



Cloud computing

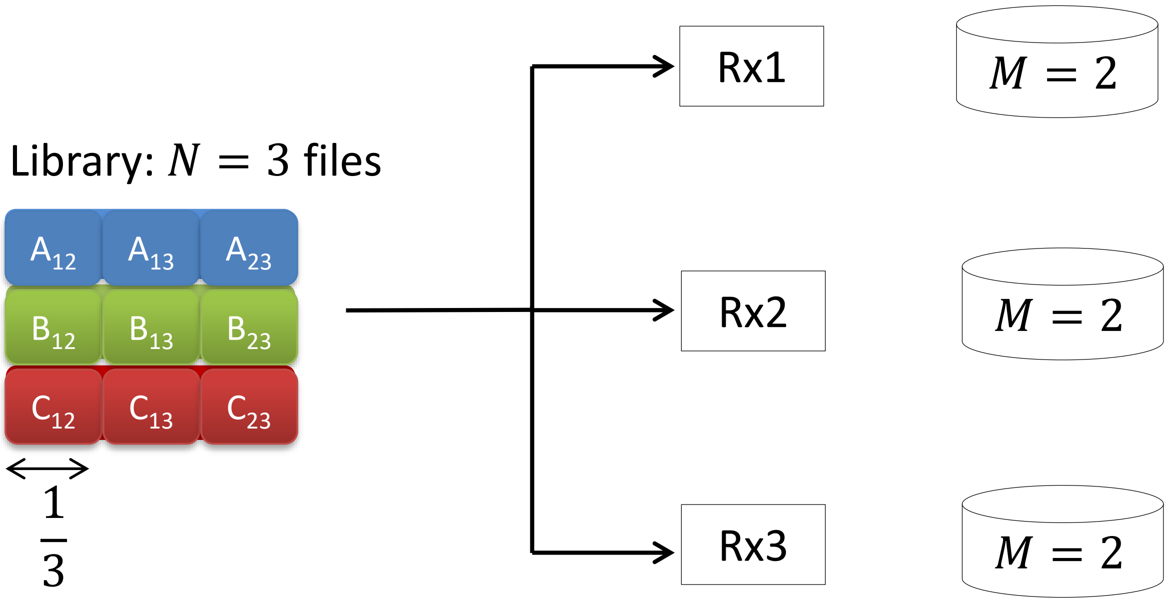
Let’s slow down: Original idea of Coded Caching



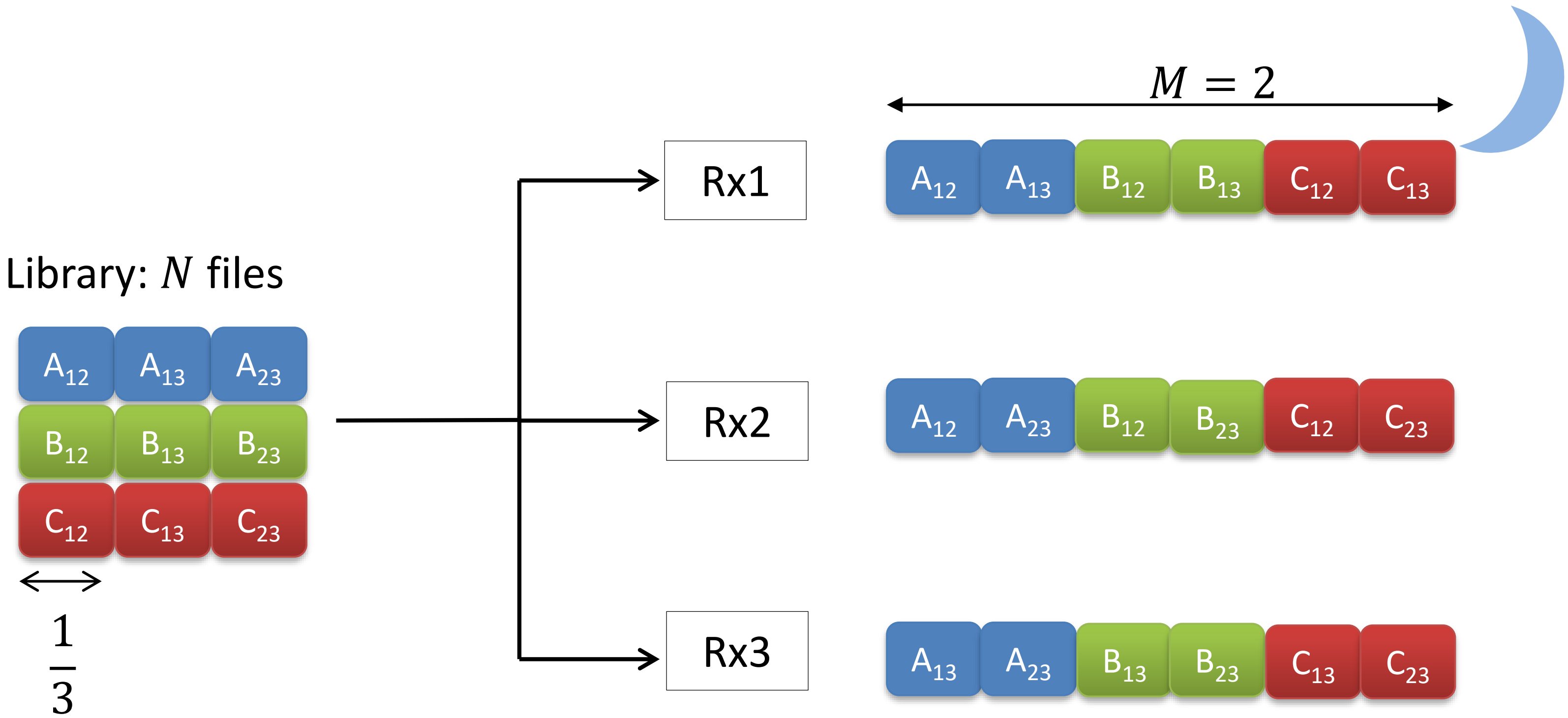
Key breakthrough: USE CACHES TO CANCEL INTERFERENCE

Result: Maddah-Ali, Niesen (Bell Labs - 2013)

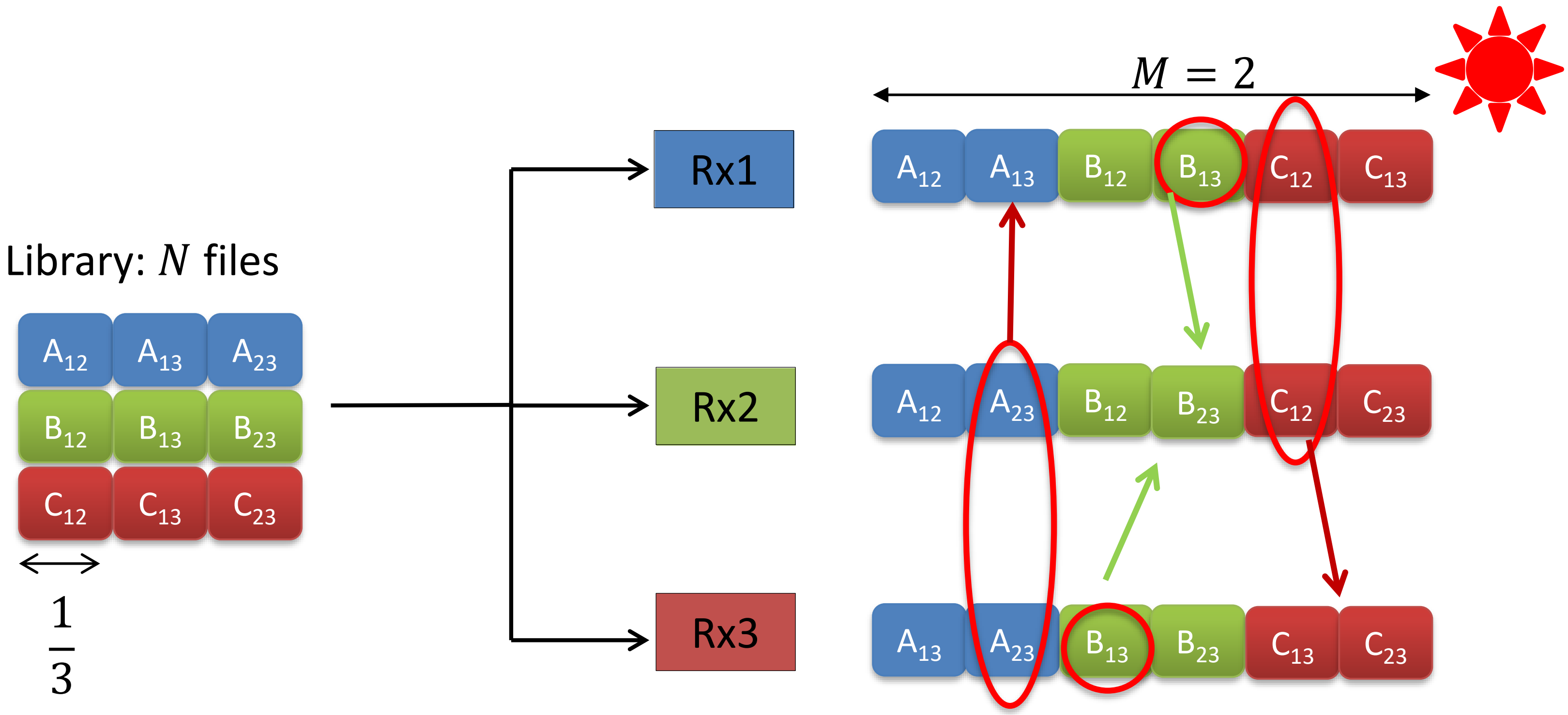
Example: $N = K = 3, M = 2 \quad (\gamma = \frac{2}{3})$



Example: $N = K = 3, M = 2 \quad (\gamma = \frac{M}{N} = \frac{2}{3})$



Example: $N = K = 3, M = 2 \quad (\gamma = \frac{M}{N} = \frac{2}{3})$

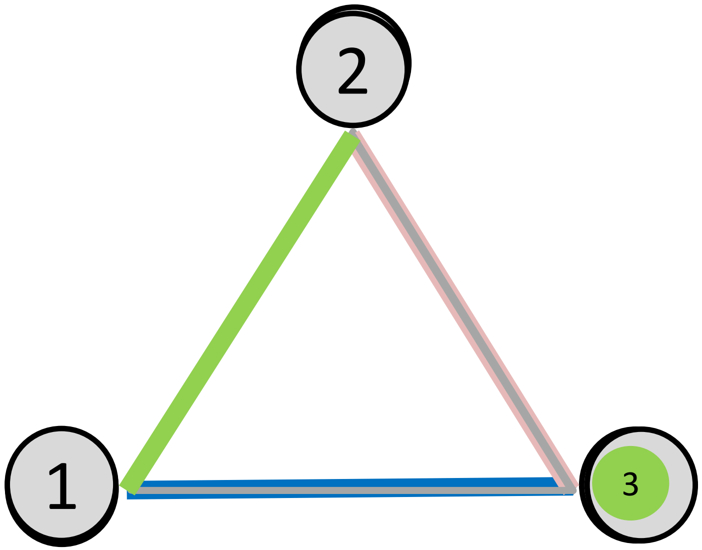


- Transmit : $A_{23} \oplus B_{13} \oplus C_{12}$ (a common message for all)

$$\boxed{Speedup\ Factor = 3 = K\gamma + 1 \quad users\ at\ a\ time}$$

Coded Caching: Intuition - Clique

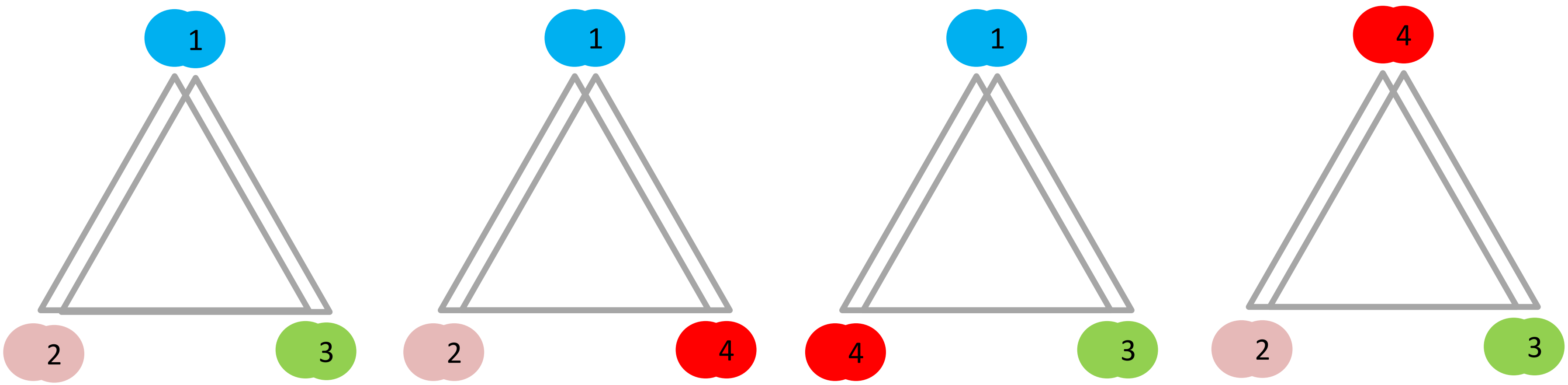
$$K = 3, \gamma = \frac{2}{3}, K\gamma = 2$$



$$A_{23} \oplus B_{13} \oplus C_{12}$$

Coded Caching: Intuition - Cliques

$$K = 4, \qquad \gamma = \frac{2}{4}, \qquad K\gamma = 2$$



$$A_{23} \oplus B_{13} \oplus C_{12}$$

$$A_{24} \oplus B_{14} \oplus D_{12}$$

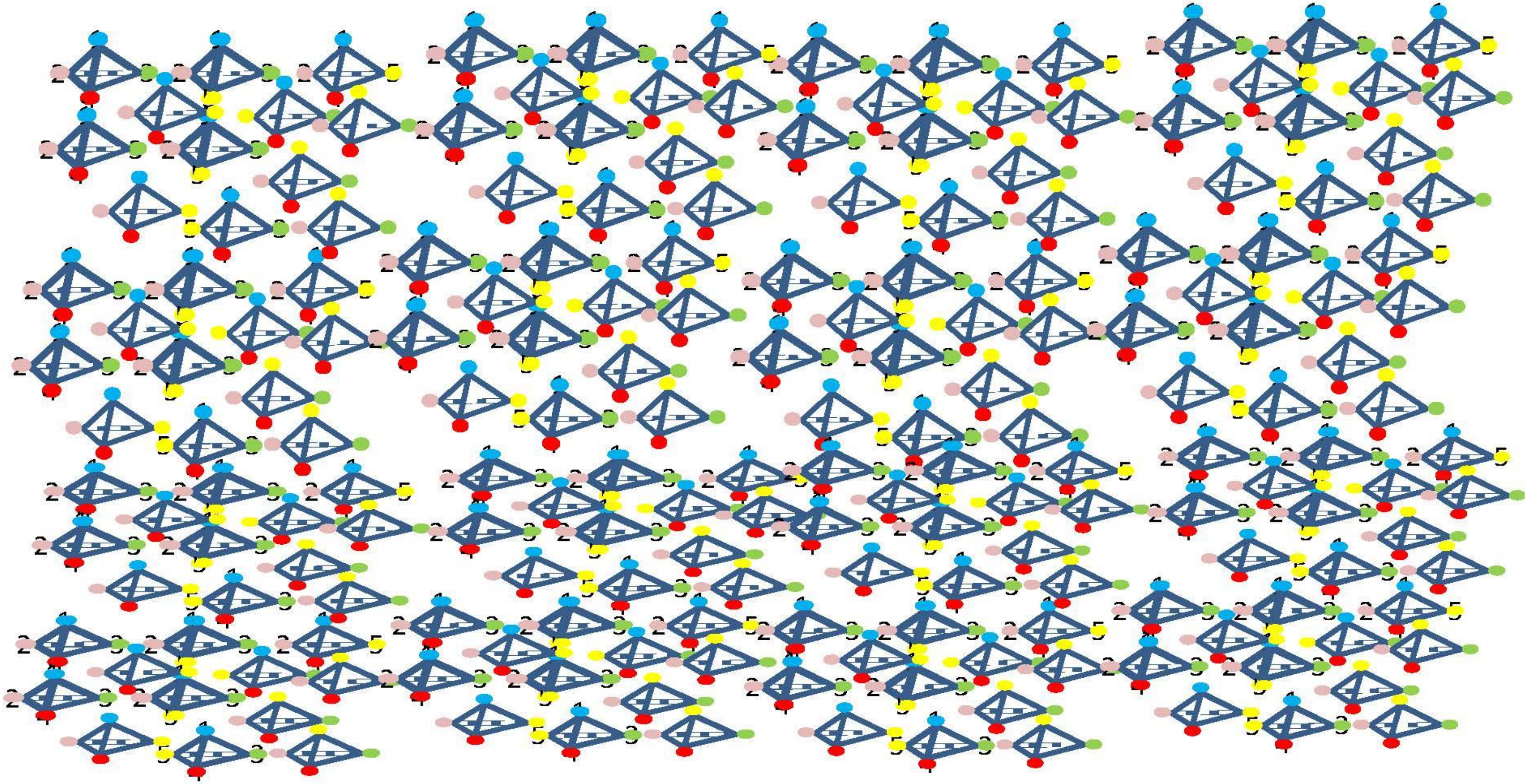
$$A_{34} \oplus D_{13} \oplus C_{14}$$

$$D_{23} \oplus B_{34} \oplus C_{24}$$

Coded Caching: Intuition – Problematically Many Cliques

$$K = 100, \qquad \gamma = \frac{9}{100} \qquad K\gamma = 9$$

$$10 \; users \; at \; a \; time \,, \qquad 2 \cdot 10^{13} \;\; cliques$$

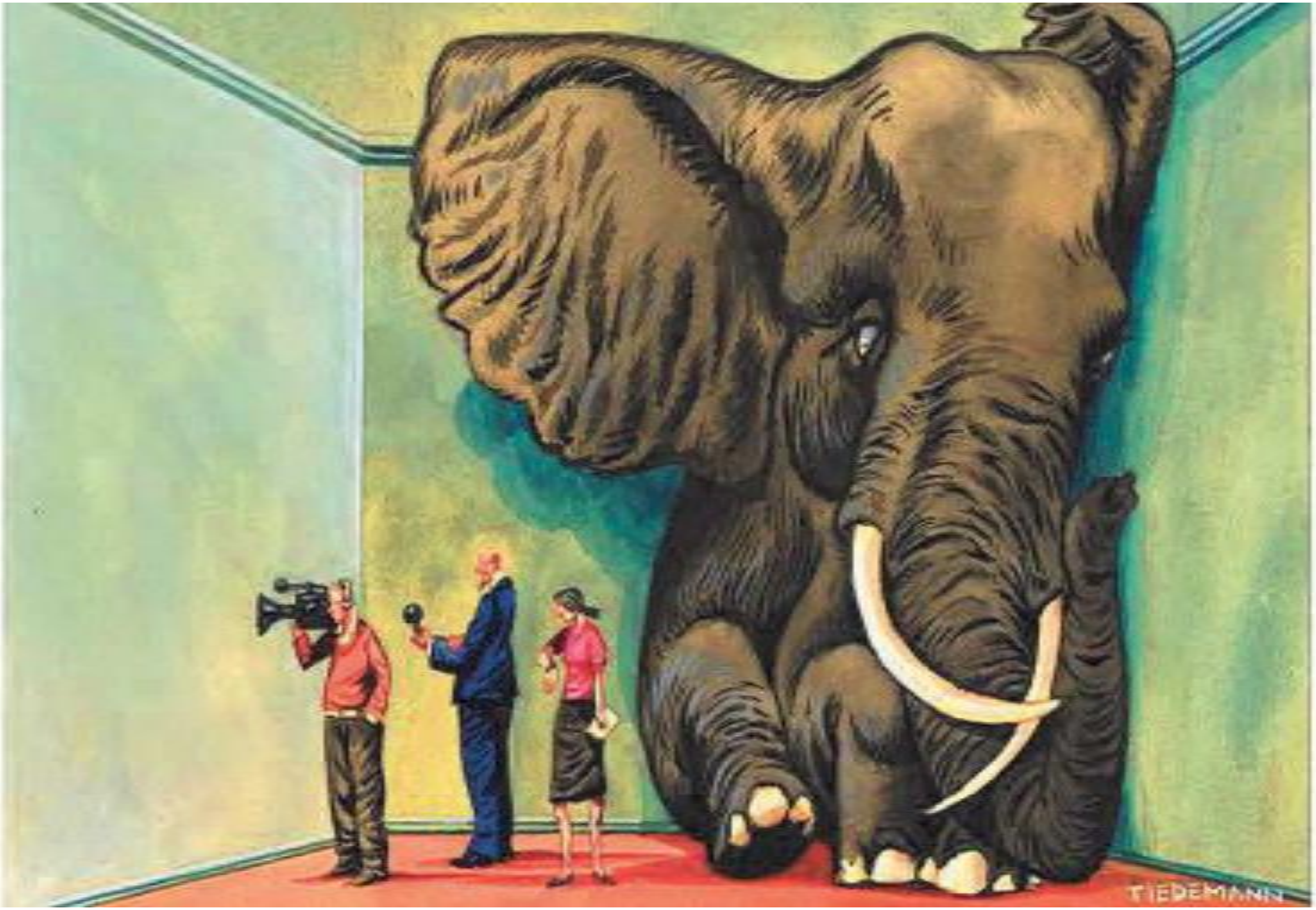


Bad news: There exist `fundamental’ bottlenecks

- In Theory: Speedup = $K\gamma + 1$

- *In theory, theory and practice are the same. In practice they are not.”* Albert Einstein

Bad news: There exist `fundamental' bottlenecks



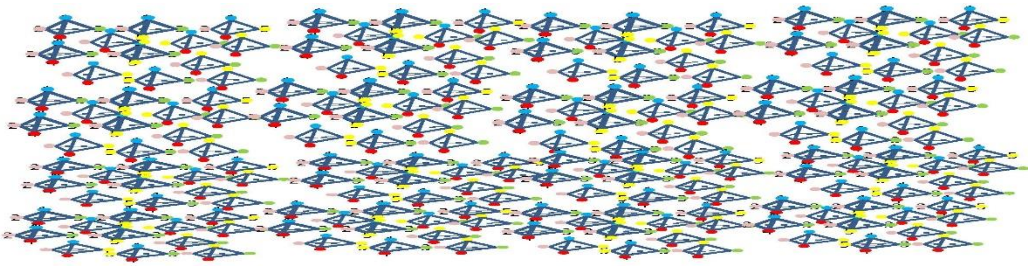
Resolve fundamental limitations.

- Coded Caching requires infinite SNR
  - Gain goes to **zero** for smaller SNR
  - Reason: Worst-user multicasting
- Requires ASTRONOMICAL file sizes
  - Gain very small for reasonable file sizes
- Works for one transmit antenna
  - Discard multi-antenna systems? Absurd.

Elevating to multi-antennas: Resolving file-size problem

- Gains vanish if files are not astronomically large!

Because we need astronomically many cliques



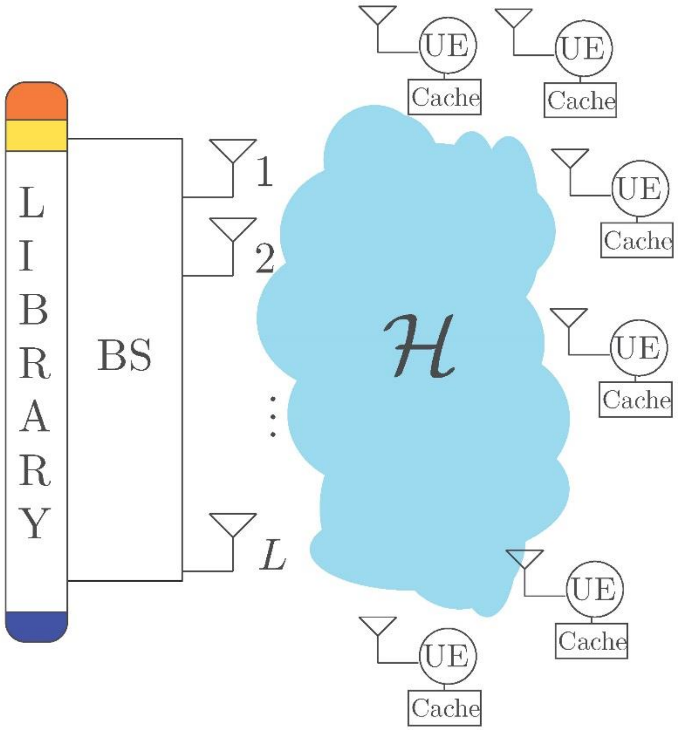
If you don’t have astronomical file sizes, the gains almost VANISH

Effective (MAX) $\boldsymbol{K\gamma + 1 \rightarrow\approx 5 - 7}$

Simple Solution: two birds with one stone

Massive CACHING gains

Resolving file-size problem (Lampiris-Elia JSAC)



L transmit antennas

Able to multiplicatively boost multiplexing gains

Subpacketization

$$S_{new} < \sqrt[L]{S_{old}}$$

$$\text{e.g. } L = 5$$

$$10^8 GBytes \rightarrow 120\ Bytes$$

Massive boost of speedup factor

Harsh Reality

Limited File Sizes

Adding Antennas

$$1 + 7 \to 8$$

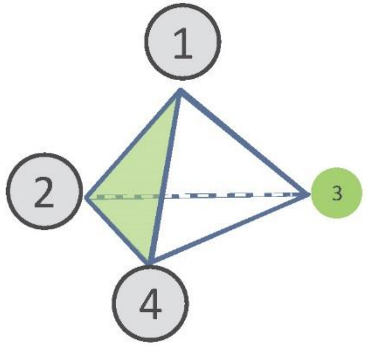
$$2 + 7 \to 16$$

$$3 + 7 \to 24$$

$$4 + 7 \to 32$$

$$\vdots$$

New algorithm: Pyramid of vectors

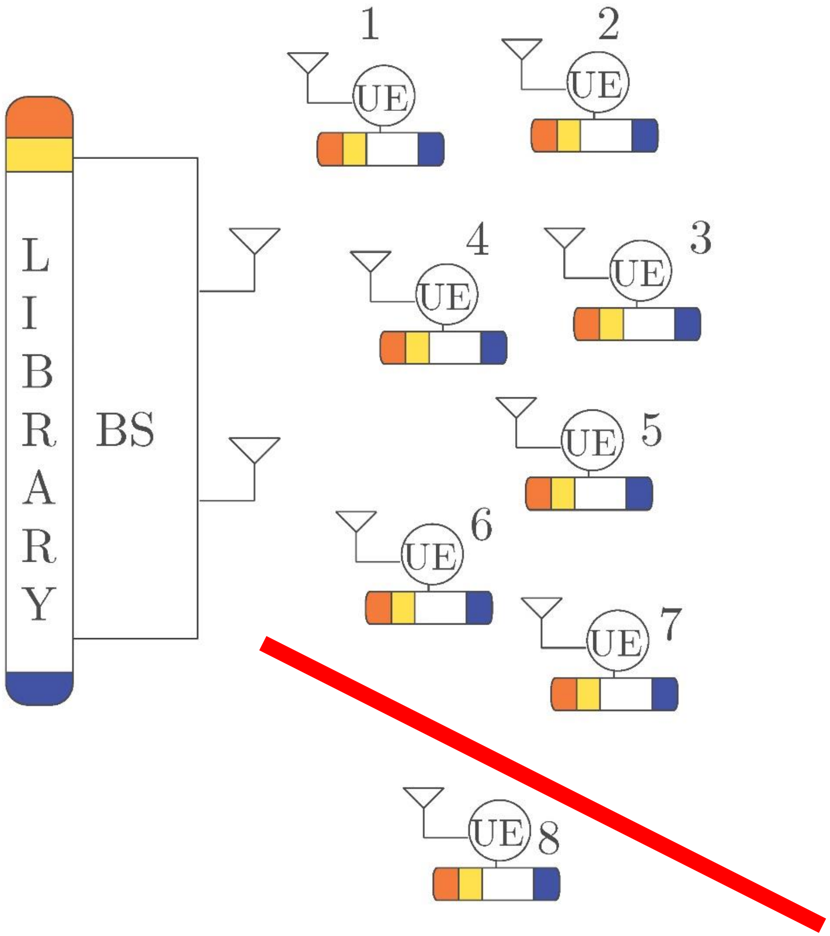


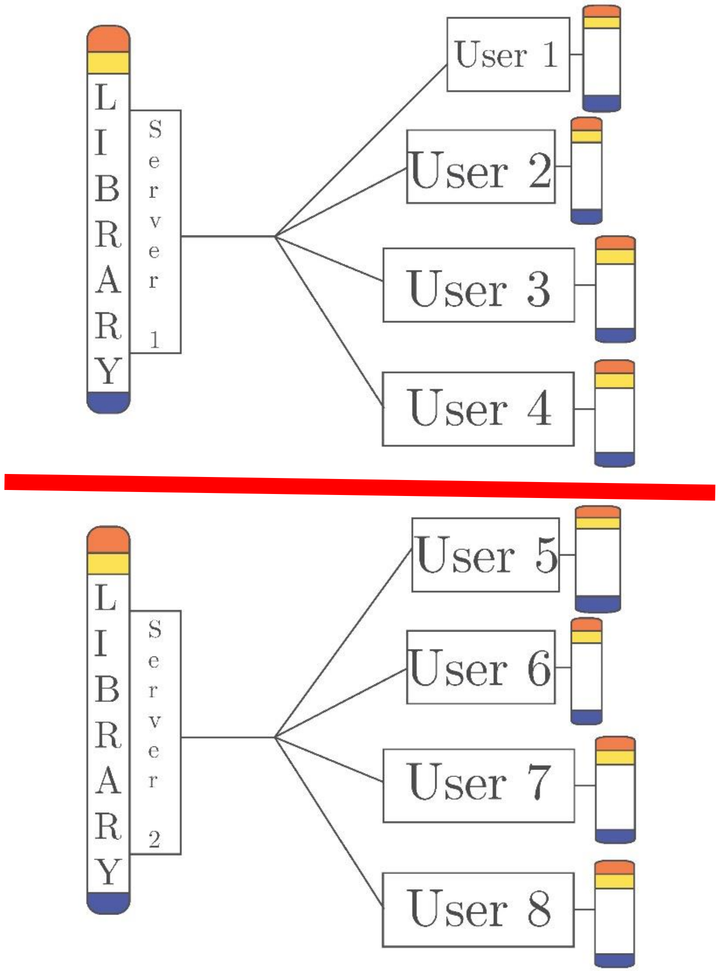
$D_{123} \oplus A_{234} \oplus B_{134} \oplus C_{124}$

$$\mathbf{x}_{1234} = \mathcal{H}_{\mathcal{G}_1}^{-1} \begin{bmatrix} W_1^{234} \\ W_{11}^{234} \\ W_{21}^{234} \\ W_{31}^{234} \\ W_{41}^{234} \end{bmatrix} + \mathcal{H}_{\mathcal{G}_2}^{-1} \begin{bmatrix} W_2^{134} \\ W_{12}^{134} \\ W_{22}^{134} \\ W_{32}^{134} \\ W_{42}^{134} \end{bmatrix} + \mathcal{H}_{\mathcal{G}_3}^{-1} \begin{bmatrix} W_3^{124} \\ W_{13}^{124} \\ W_{23}^{124} \\ W_{33}^{124} \\ W_{43}^{124} \end{bmatrix} + \mathcal{H}_{\mathcal{G}_4}^{-1} \begin{bmatrix} W_4^{123} \\ W_{14}^{123} \\ W_{24}^{123} \\ W_{34}^{123} \\ W_{44}^{123} \end{bmatrix}$$

$$\mathbf{y}_{1234}^{(1)} = \underbrace{\mathcal{H}_{\mathcal{G}_1} \mathcal{H}_{\mathcal{G}_1}^{-1}}_{I_{L\times L}} \begin{bmatrix} W_1^{234} \\ W_{11}^{234} \\ W_{21}^{234} \\ W_{31}^{234} \\ W_{41}^{234} \end{bmatrix} + \cancel{\mathcal{H}_{\mathcal{G}_1} \mathcal{H}_{\mathcal{G}_2}^{-1} \begin{bmatrix} W_2^{134} \\ W_{12}^{134} \\ W_{22}^{134} \\ W_{32}^{134} \\ W_{42}^{134} \end{bmatrix}} + \cancel{\mathcal{H}_{\mathcal{G}_1} \mathcal{H}_{\mathcal{G}_3}^{-1} \begin{bmatrix} W_3^{124} \\ W_{13}^{124} \\ W_{23}^{124} \\ W_{33}^{124} \\ W_{43}^{124} \end{bmatrix}} + \cancel{\mathcal{H}_{\mathcal{G}_1} \mathcal{H}_{\mathcal{G}_4}^{-1} \begin{bmatrix} W_4^{123} \\ W_{14}^{123} \\ W_{24}^{123} \\ W_{34}^{123} \\ W_{44}^{123} \end{bmatrix}}$$

Intuition: multiple decompositions with few antennas

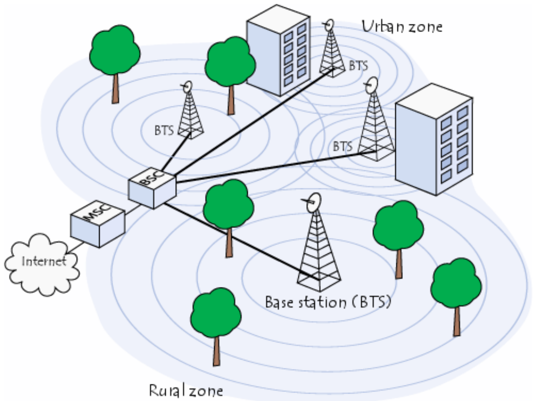




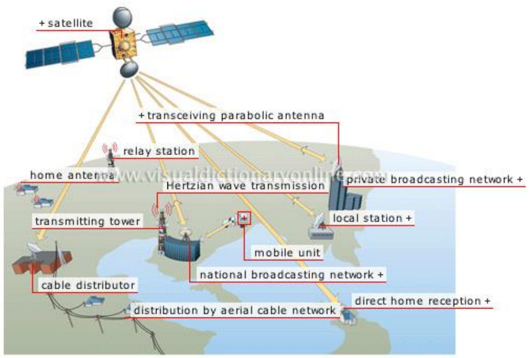
This has been Theory

- Theoretical gains are large.
- What happens when we move closer to practice
- ERC Proof of Concept – LIGHT

Wired or wireless (Cellular-Wifi)



Satellite communications

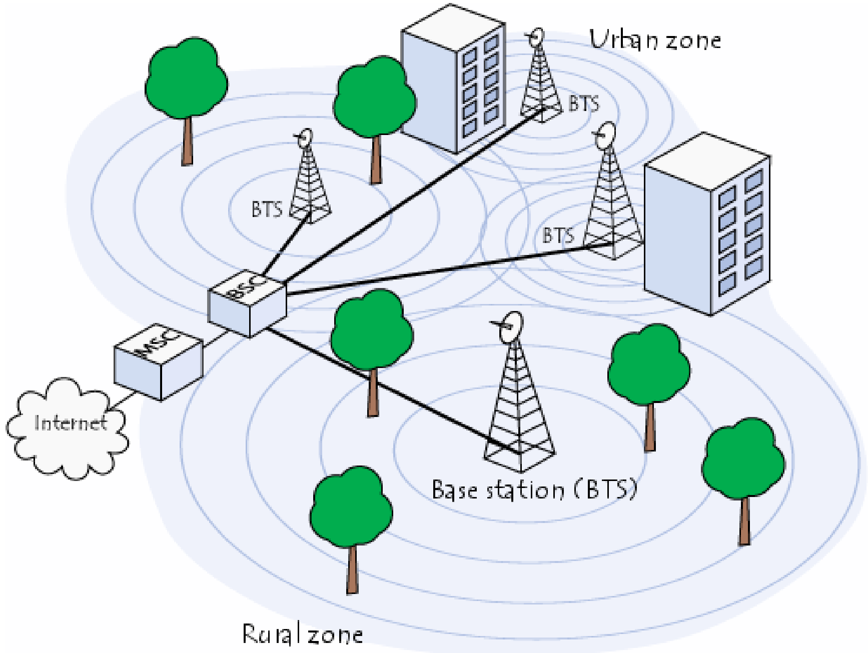


Cloud computing

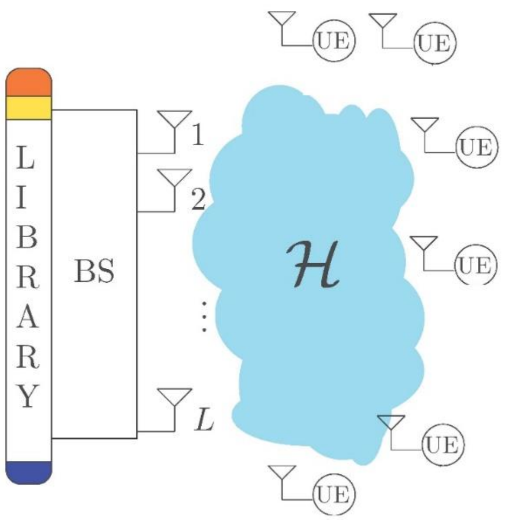
This has been Theory

- Focus on Multi-User MISO – Cellular, WiFi, etc.

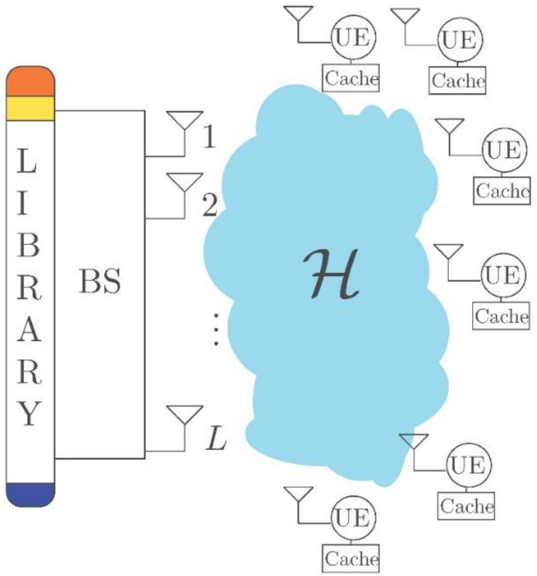
Wired or wireless (Cellular-Wifi) - DOWNLINK



Gains over OPTIMIZED MISO BC – Analysis and Sims



Add Caches



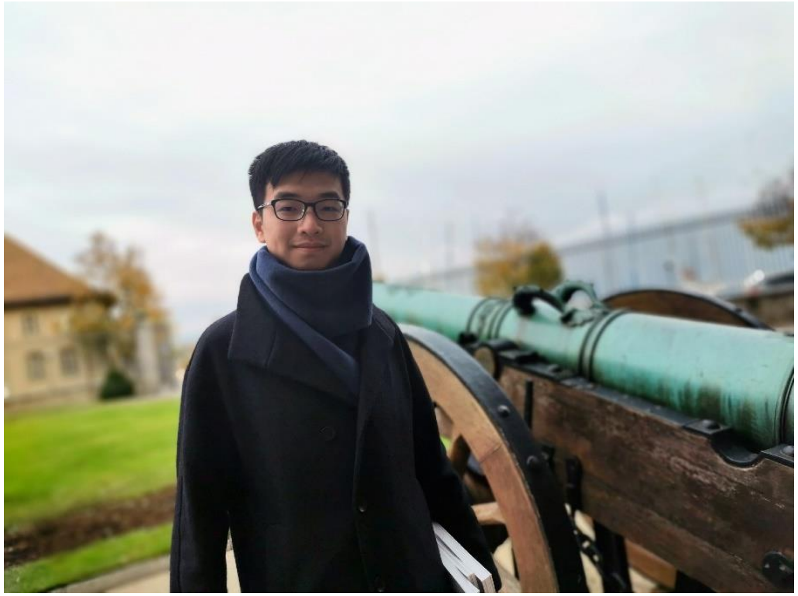
- Optimize as you wish
- Account for CSI costs

Optimize
Account for CSI Cost

$$Gain = \frac{R_{\Sigma}^{*}(\text{cache} - \text{aided})}{R_{\Sigma}^{*}(\text{no} - \text{cache})}$$

"Vector Coded Caching Multiplicatively Boosts the Throughput of Realistic Downlink Systems," Zhao-Bazco-Elia

Large Matrix Analysis and Sims of Cache-Aided MIMO-BC



Hui Zhao

Antonio Bazco-Nogueras

$$Gain = \frac{R_\Sigma^*(\mathrm{cache-aided})}{R_\Sigma^*(\mathrm{no-cache})}$$

"Vector Coded Caching Multiplicatively Boosts the Throughput of Realistic Downlink Systems," Zhao-Bazco-Elia

Justifying Theoretical Gains $G: 5 \rightarrow 7$

Netflix movies - Zipf $\approx$ 1.4
90% of traffic speeds-up by <u>theoretical</u> factor of G

$$\boldsymbol{G = 7}$$

- Cache: 25GB
- Movies: HD, 1.3GB, 90 minutes
- Latency: 2 min. (small buffer OK)
- Comm Packet: 50 bytes
- Subpacketization: 600K

$$\boldsymbol{G = 6}$$

- Cache: 25GB
- Movies: Full-HD, 2.47GB, 90 minutes

$$\boldsymbol{G = 5}$$

- Cache: 5GB
- Movies: SD-480p
- Comm Packet: 200 bytes
- Subpacketization: 600K

Gains Across Various Precoders (MISO)

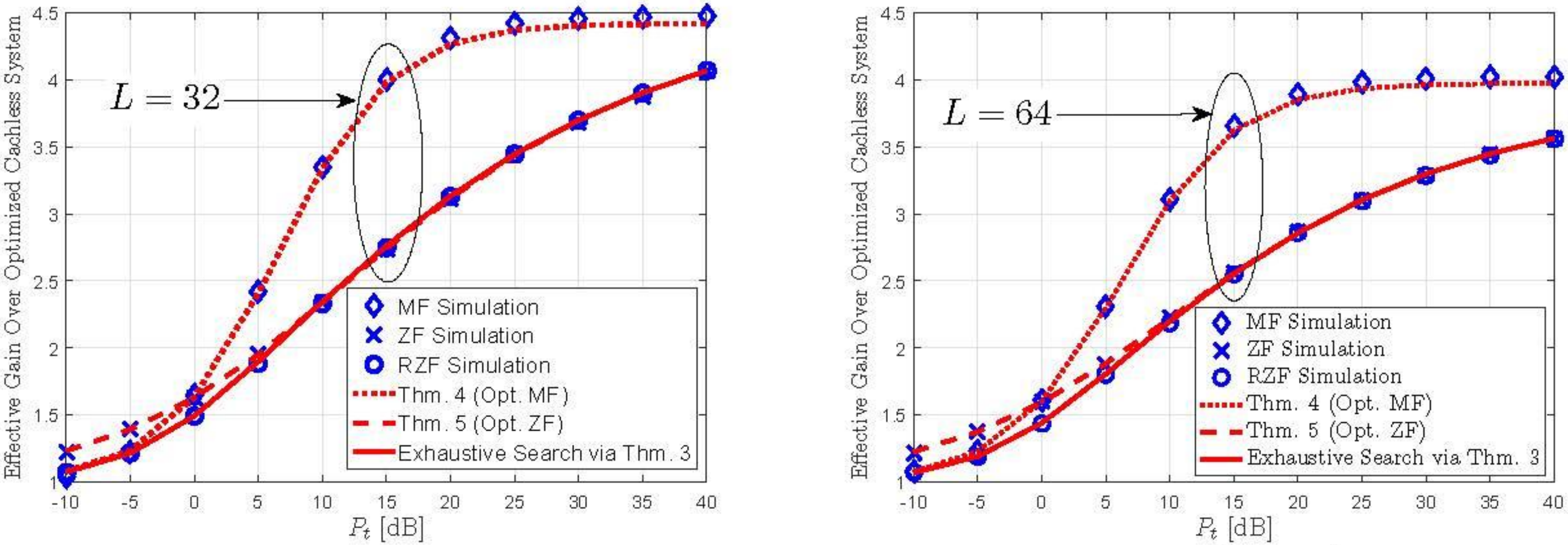
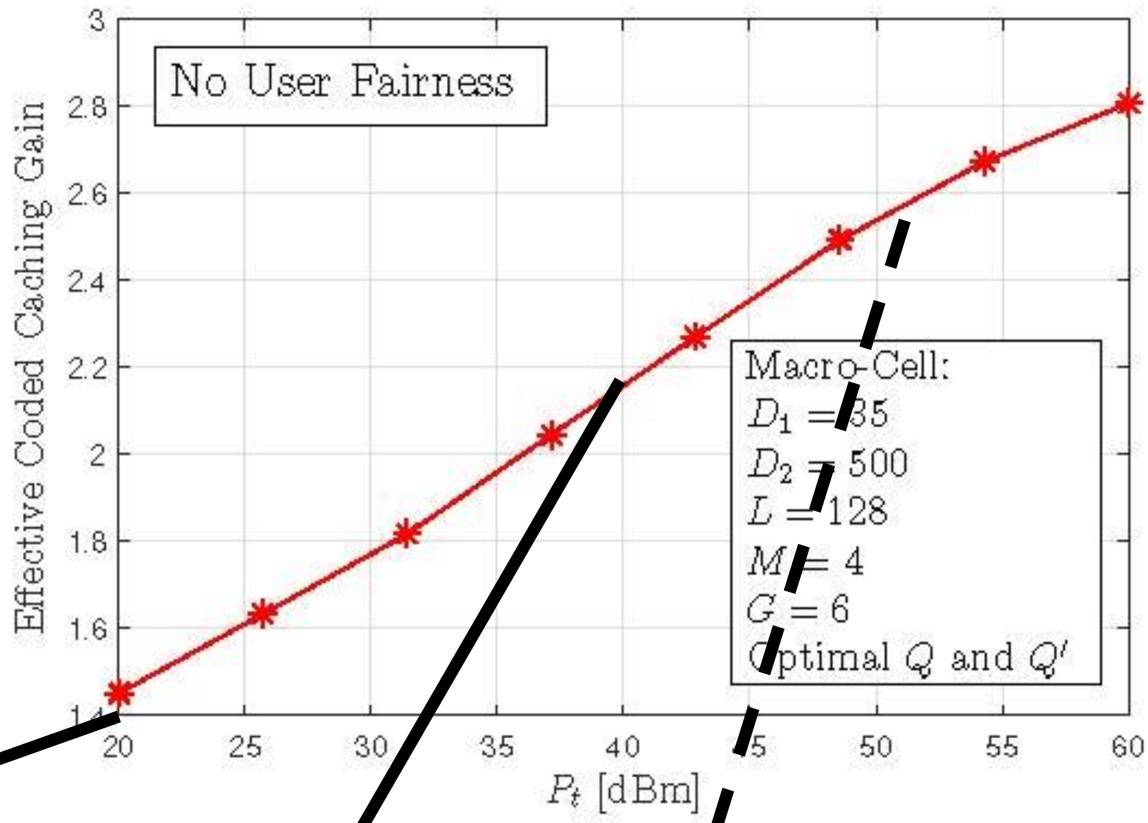


Fig. 2: Effective gain $\mathcal{G}^\star$ over optimized cacheless system for $L \in \{32, 64\}$ and $G = 6$.

Most Challenging Scenario – very large L, Large Cells, Lower power

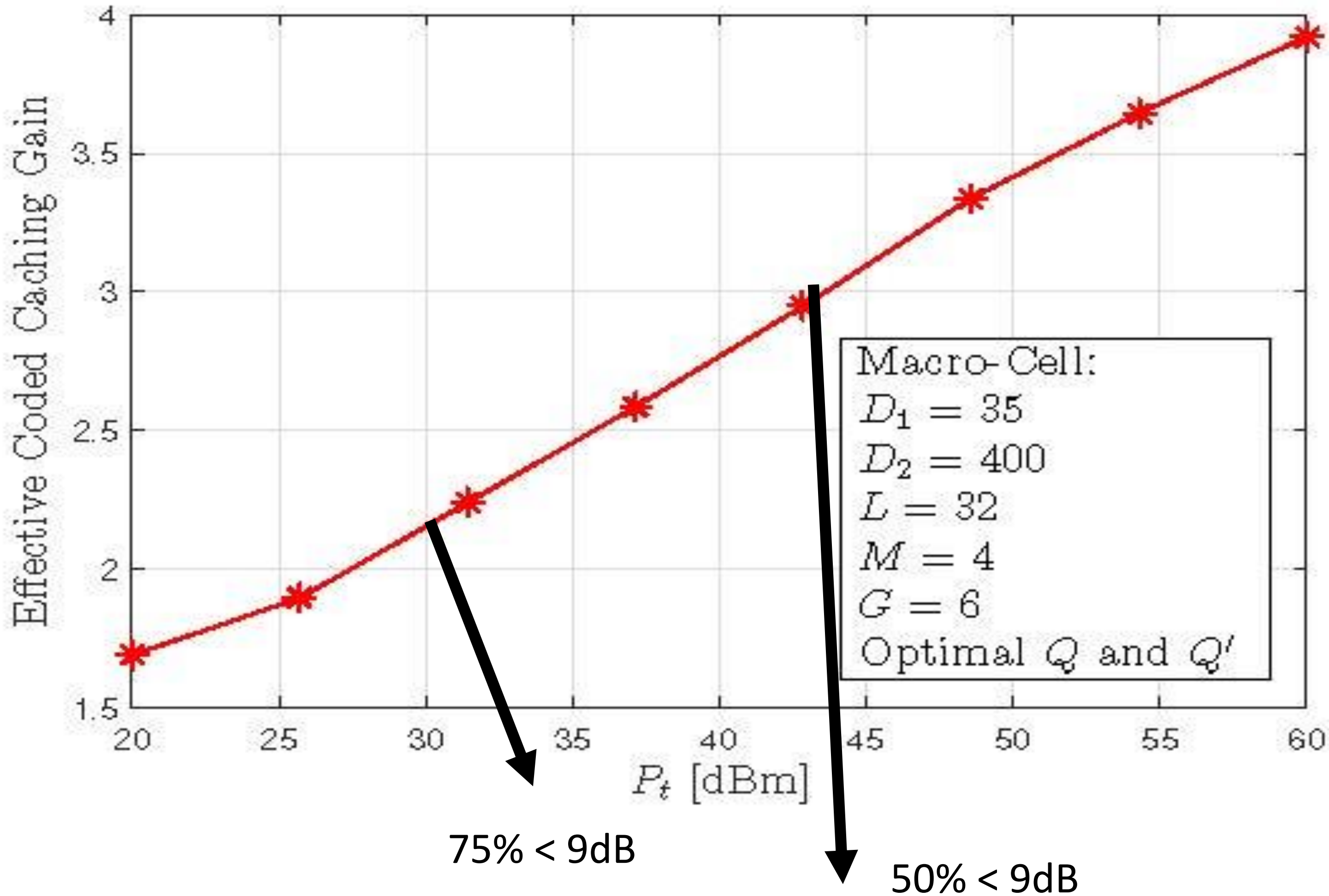


75% < -3dB
40% < -10dB

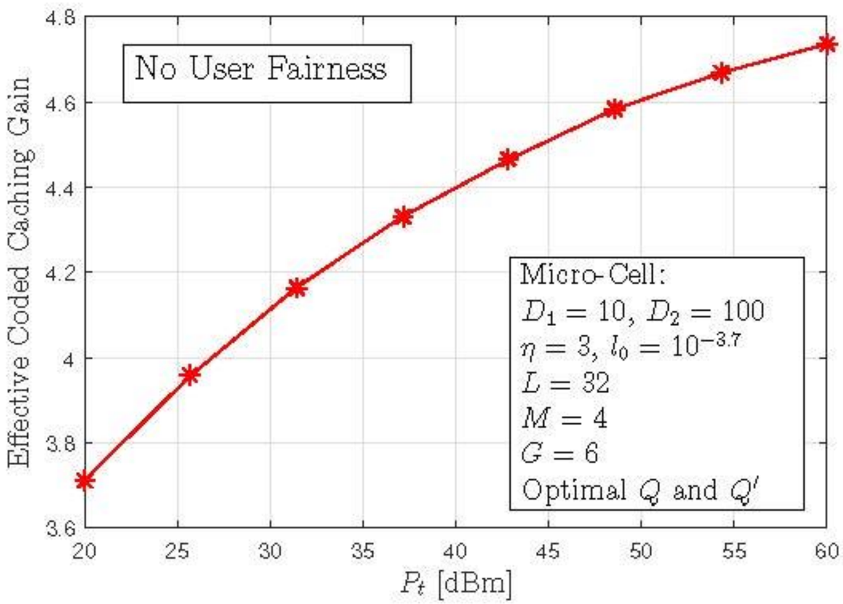
60% < 9dB

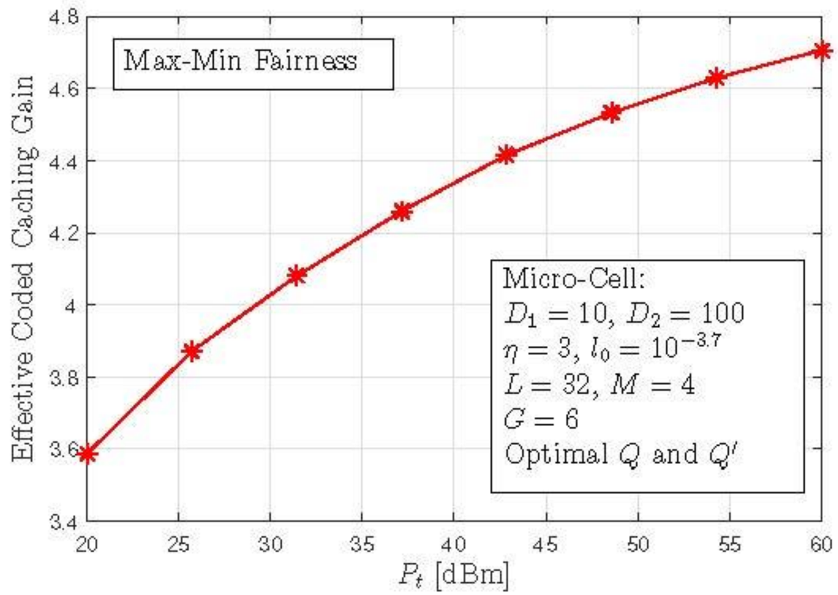
75% < 27dB
44% < 21dB

Large cell – 32 Tx-Antennas



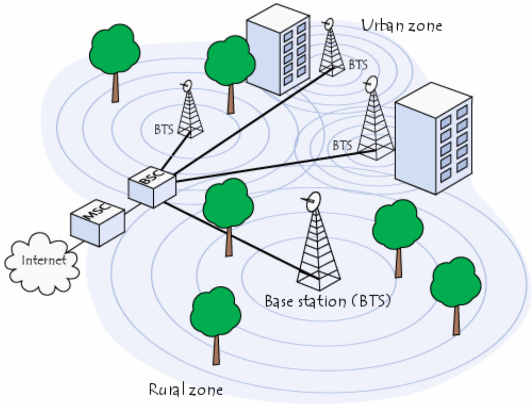
Smaller cells – w. & w/o Max-Min Fairness MU-MIMO



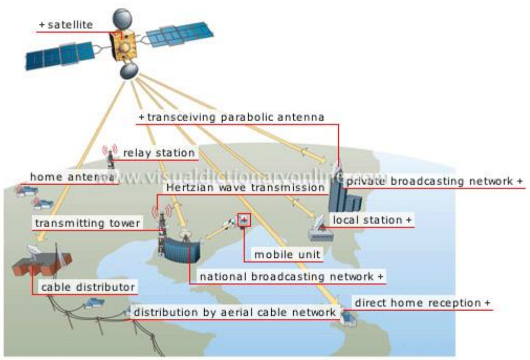


Many Applications to be Explored...

Wired or wireless communications

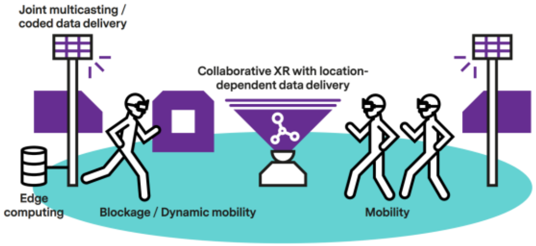


Satellite communications



In-flight communications

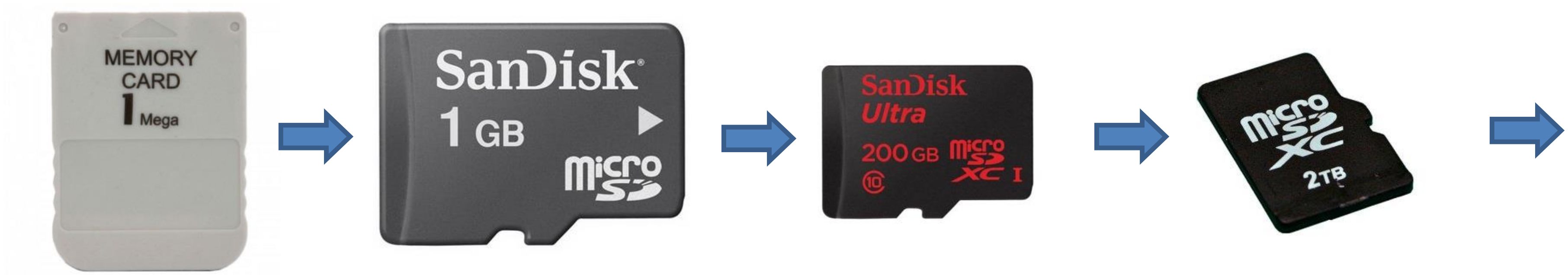
Virtual Reality



Cloud computing

Need and opportunity

- ✓ Abundant new resource
- ✓ Works well with other resources (very simple precoding)
- ✓ Very substantial gains



- ✓ Some VERY Simple algorithms
  - ✓ E.g. Superimpose ZF transmissions and read from memory
- ✓ Some very playful challenges (Satellite, Large Cells, Many users)

Thank You!