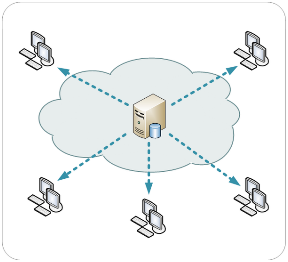
An Information-Theoretic View of Cache-Aided Cellular Networks

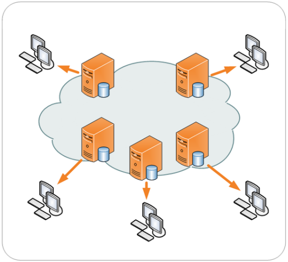
Michèle Wigger

Joint work with Shirin Saeedi Bidokthi, Shlomo Shamai (Shitz), and Roy Timo.

IWCIT, Tehran, Iran, 3 May 2016

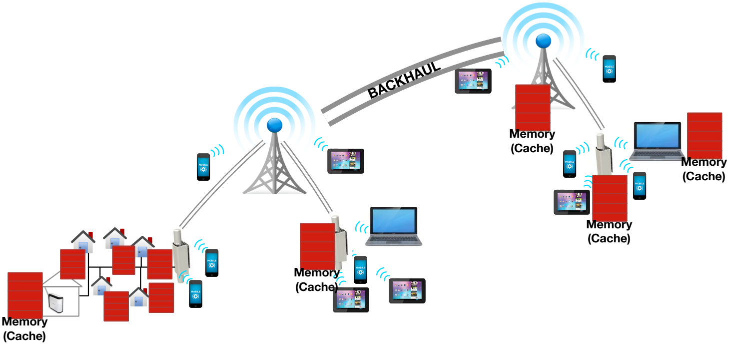
Content Delivery Networks





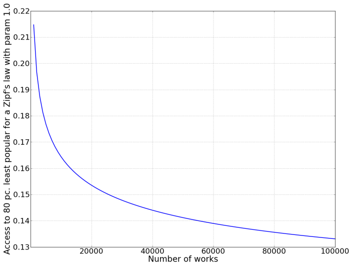
- Store contents in caches before file demands even known
- Reduce network load and latency during high-congestion periods
- Idea useful if certain files very popular and known in advance

Distributed Caches: Promising Solution for Cellular Networks



- Can cache at main BSs, picoBSs, femtoBSs, or directly at end users

File Popularities

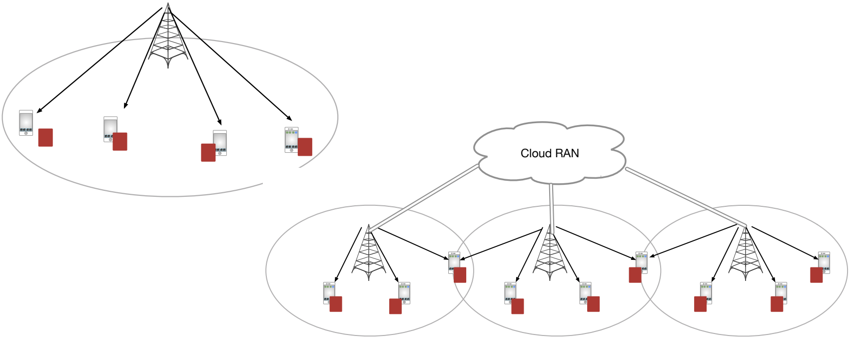


- Static file popularity follows a Zipf distribution $P(x) = Cx^{-\alpha}$

- Evolution of file popularities (youtube videos) can also be predicted

Use pro-active caching to improve cellular systems!

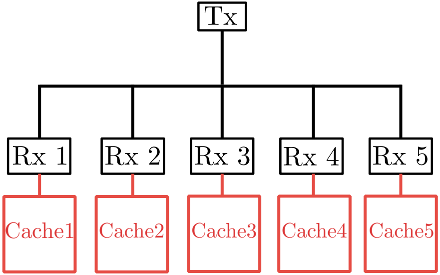
Cellular Scenarios



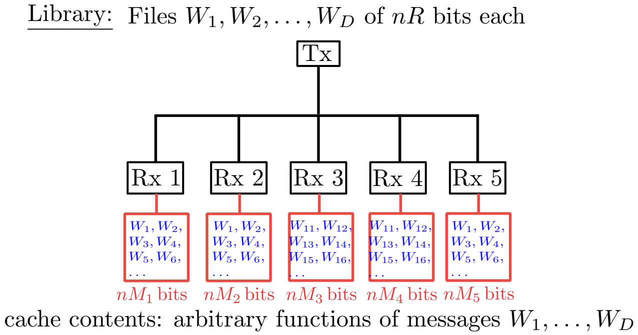
- All files equally popular → interested in worst-case performance
- Centralized protocol on how to fill caches
- Caches filled during nights when demands not yet known

Maddah-Ali & Niesen Source Coding Setup

Library: Files $W_1, W_2, \ldots, W_D$ of nR bits each (no popularities)



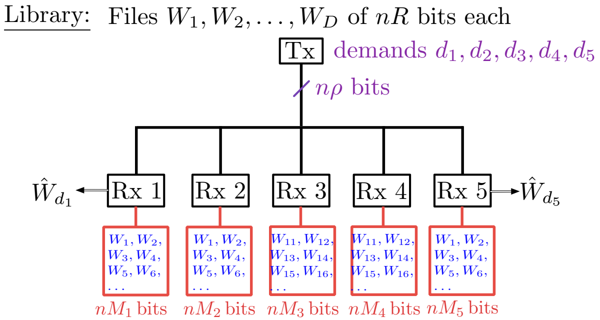
Maddah-Ali & Niesen Source Coding Setup



Communication in two phases:

- Caching phase: Tx fills caches without knowing demands $d_1, \ldots, d_5$

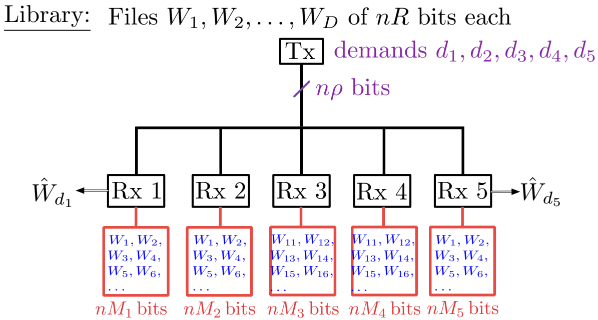
Maddah-Ali & Niesen Source Coding Setup



Communication in two phases:

- Caching phase: Tx fills caches without knowing demands $d_1, \ldots, d_5$
- Delivery phase: Tx describes $W_{d_1}, \ldots, W_{d_5}$ to Rxs $1, \ldots, 5$, respectively, through $n\rho$ common bits

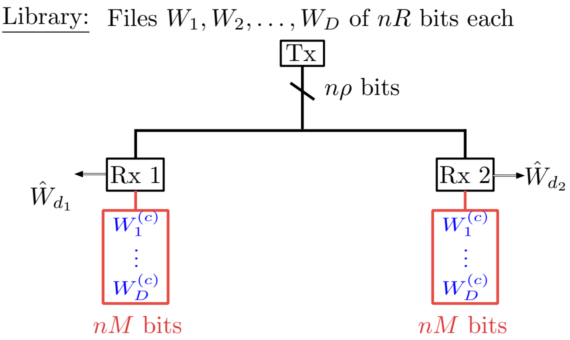
Maddah-Ali & Niesen Source Coding Setup



Rates-Memories Tradeoff

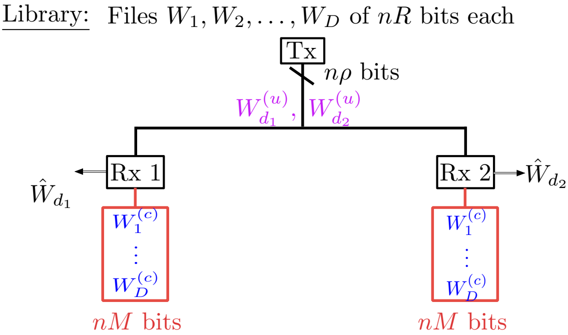
For which $(\rho, R, M_1, \ldots, M_K)$ is error-free data transmission possible?

Naive Uncoded Caching for $K = 2$ Receivers



- Split $W_d = (W_d^{(c)}, W_d^{(u)})$ of rates $\frac{M}{D}$ and $R - \frac{M}{D}$

Naive Uncoded Caching for $K = 2$ Receivers

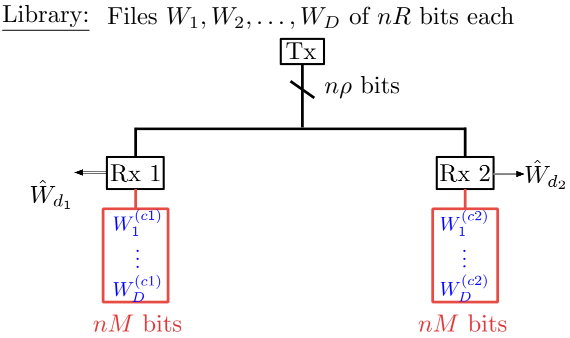


- Split $W_d = (W_d^{(c)}, W_d^{(u)})$ of rates $\frac{M}{D}$ and $R - \frac{M}{D}$

Rates-Memory Trade-Off

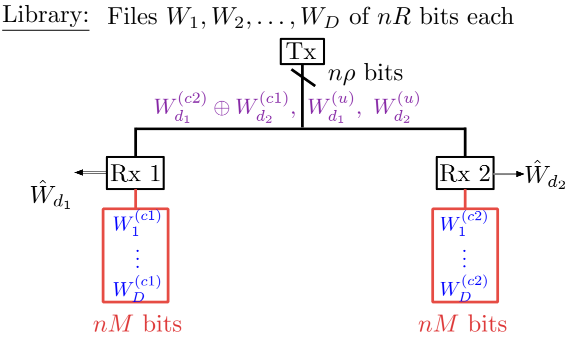
Reconstruction is possible, if $R \leq \frac{1}{2}\rho + \frac{M}{D}$

Coded caching for $K = 2$ Receivers [Maddah-Ali&Niesen 2013]



- Split $W_d = \left(W_d^{(c1)}, W_d^{(c2)}, W_d^{(u)}\right)$ of rates $\frac{M}{D}$, $\frac{M}{D}$, and $R - 2\frac{M}{D}$

Coded caching for $K = 2$ Receivers [Maddah-Ali&Niesen 2013]

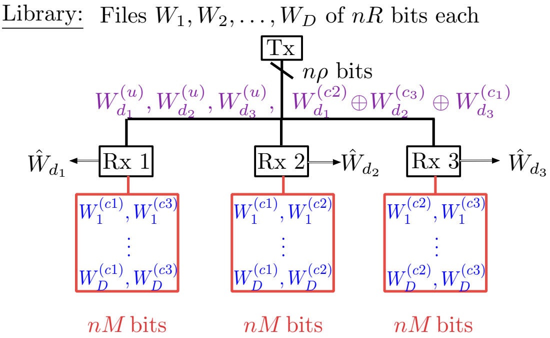


- Split $W_d = \left(W_d^{(c1)}, W_d^{(c2)}, W_d^{(u)}\right)$ of rates $\frac{M}{D}$, $\frac{M}{D}$, and $R - 2\frac{M}{D}$

Rates-Memory Trade-Off

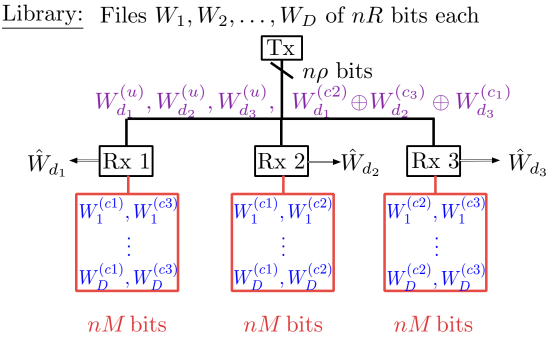
Reconstruction possible, if $R \leq \frac{1}{2}\rho + \frac{M}{D} + \frac{M}{2D}$

Coded caching for $K = 3$ Receivers [Maddah-Ali&Niesen 2013]



- Split $W_d = (W_d^{(c1)}, W_d^{(c2)}, W_d^{(c3)}, W_d^{(u)})$ of rates $\frac{M}{2D}$, $\frac{M}{2D}$, $\frac{M}{2D}$, $R - \frac{3M}{2D}$
- Save two parts at each receiver

Coded caching for $K = 3$ Receivers [Maddah-Ali&Niesen 2013]

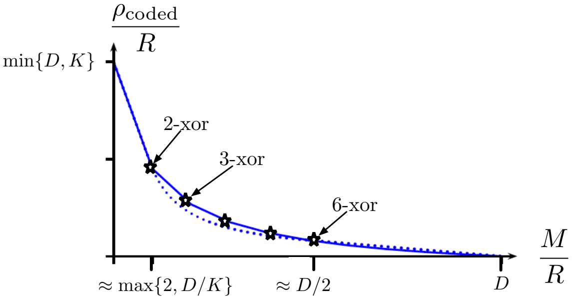


- Split $W_d = (W_d^{(c1)}, W_d^{(c2)}, W_d^{(c3)}, W_d^{(u)})$ of rates $\frac{M}{2D}, \; \frac{M}{2D}, \; \frac{M}{2D}, \; R - \frac{3M}{2D}$
- Save two parts at each receiver

Rates-Memory Trade-Off

Reconstruction possible, if $R \leq \frac{1}{3}\rho + \frac{M}{D} + \frac{M}{3D}$

Local and Global Caching Gains $K \geq 2$ [Maddah-Ali&Niesen 2013]



Coded caching achieves

Reconstruction possible, if $\rho_{\text{coded}} \geq K(R - \frac{M}{D}) \cdot \min\left\{\frac{1}{1+KM/R/D}, \frac{D}{K}\right\}$

$$1 \leq \frac{\rho^{\star}(R, M)}{\rho_{\text{coded}}(R, M)} \leq 12, \qquad \forall K, \rho, D, M.$$

Improvements

- Lower Bound with Gap 4.7

  [C.-Y. Wang, S.-H. Lim, and M. Gastpar, "A New Converse Bound for Coded Caching", Arxiv, Jan. 2016]

- Upper Bound with Coded Caching Information

  [M. Mohammadi. Amiri, D. Gündüz "Fundamental Limits of Caching: Improved Delivery Rate-Cache Capacity Trade-off", Arxiv, Apr. 2016]

Extensions

- Decentralized caching

  [M. A. Maddah-Ali, U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff"]

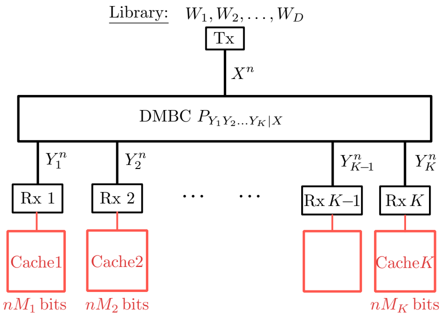
- Nonuniform or random demands

  [U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands"]
  [Ji, Tulino, Llorca, and Caire, "Order-optimal rate of caching and coded multicasting with random demands"]

- Online caching phase

  [R. Pedarsani, M. A. Maddah-Ali and U. Niesen, "Online coded caching"]

Delivery over Noisy Broadcast Channel (BC) [Saeedi, Timo, Wigger 2015-2016]

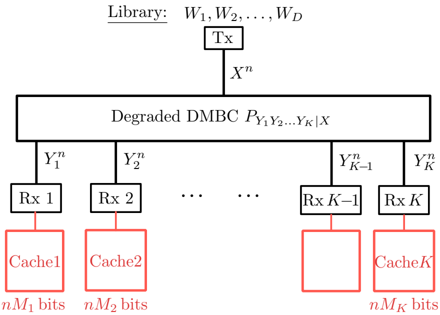


Capacity-Memory Tradeoff

$C(M_1, \ldots, M_K)$: supremum of achievable rate of messages R

- New achievability: joint cache-channel scheme based on piggyback coding
- New converse for degraded BCs

Converse for Degraded BCs with Arbitrary Caches



Theorem (Saeedi,Timo,Wigger16)

$$C(M_1, \ldots, M_K) \leq \min_{\mathcal{S} \subseteq \{1,\ldots,K\}} \left(R_{\text{sym},\mathcal{S}}(M_1, \ldots, M_K) + \frac{M_{\mathcal{S}}}{D} \right),$$

- *$R_{\text{sym},\mathcal{S}}$ and $M_{\mathcal{S}}$: symmetric capacity and total cache at receivers in $\mathcal{S}$*

Proof Outline

- Step 1:

$$C(M_1,\dots,M_K) \leq I\big(U_1;Y_1\big)+\alpha_1,$$
$$C(M_1,\dots,M_K) \leq I\big(U_k;Y_k\big|U_1,\dots,U_{k-1}\big)+\alpha_k, \qquad \forall k\in\{2,\dots,K\}.$$

for $(U_1,U_2,\dots,U_K)$ s.t.

$$U_1 \to U_2 \to \cdots \to U_K \to X \to Y_K \to Y_{K-1} \to \cdots \to Y_1$$

and real numbers $\alpha_1,\dots,\alpha_K \geq 0$ s.t.

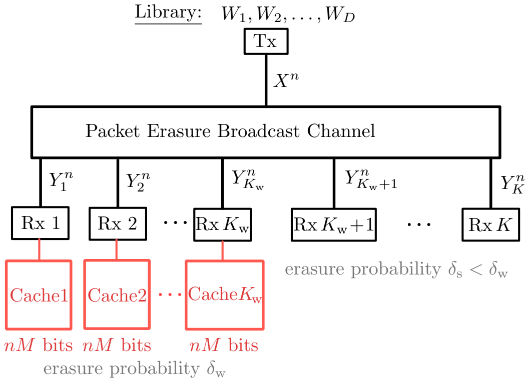
$$\alpha_{k'} \leq \alpha_k, \qquad k,k'\in\{1,\dots,K\},\ k'\leq k,$$
$$\sum_{k=1}^{K}\alpha_k \leq \frac{K}{D}\sum_{k\in\{1,\dots,K\}} M_k,$$

- Step 2: Show equal αs are optimal.

Our New Joint Cache-Channel Scheme for Packet-Erasure BCs

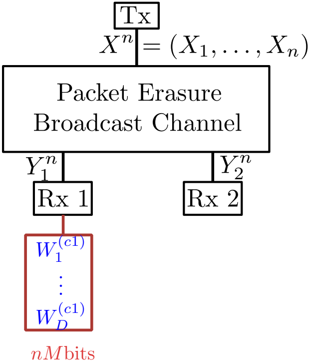
- Receiver k gets erasure with probability δ_k where $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_K$

$$Y_k^n = (X_1, X_2, \Delta, X_4, \Delta, \ldots, X_{n-1}, \Delta)$$



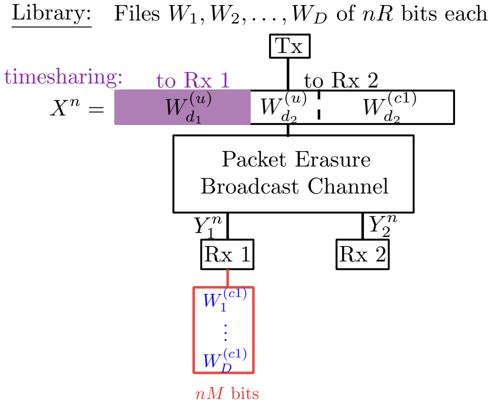
Example: Asymmetric Caches and Separate Channel Coding

Library: Files $W_1, W_2, \ldots, W_D$ of nR bits each



- $W_d = (W_d^{(c1)},\ W_d^{(u)})$ of rates $(\frac{\mathsf{M}}{D},\ R - \frac{\mathsf{M}}{D})$

Example: Asymmetric Caches and Separate Channel Coding



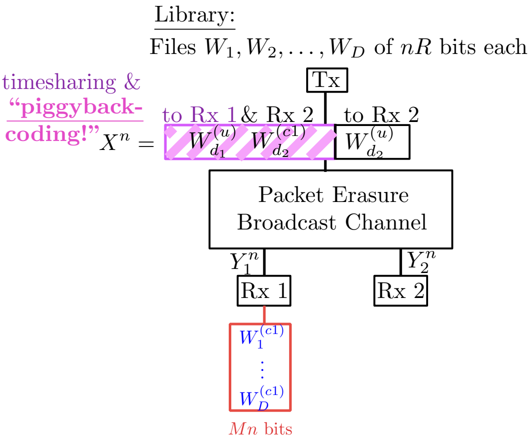
- $W_d = (W_d^{(c1)},\ W_d^{(u)})$ of rates $\left(\frac{\mathsf{M}}{D},\ R - \frac{\mathsf{M}}{D}\right)$

Separate Cache-Channel Coding → No Global Caching Gain

$p(\text{error}) \to 0$ if: $$\frac{R - \frac{\mathsf{M}}{D}}{F(1-\delta_1)} + \frac{R}{F(1-\delta_2)} \leq 1$$

Standard Erasure BC: $p(error)$ if : $$\frac{R_1}{F(1-\delta_1)} + \frac{R_2}{F(1-\delta_2)} \leq 1$$

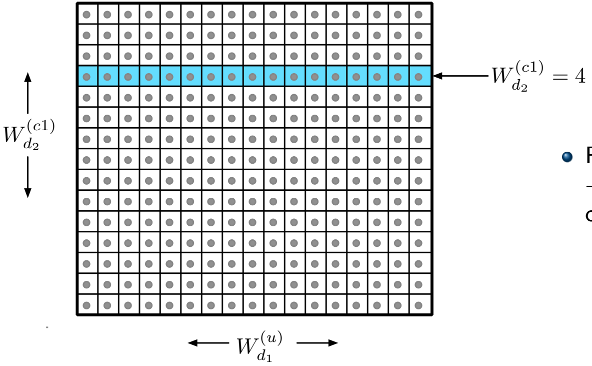
Our Joint Cache-Channel Scheme for this Example



- $W_d = (W_d^{(c1)},\ W_d^{(u)})$ of sub-rates $(\frac{\mathsf{M}}{D},\ \rho - \frac{\mathsf{M}}{D})$

Piggyback Coding to Send $(W_{d_1}^{(u)}, W_{d_2}^{(c1)})$ to Both Rxs

codebook of codewords $X^{n'}(W_{d_1}^{(u)}, W_{d_2}^{(c1)})$

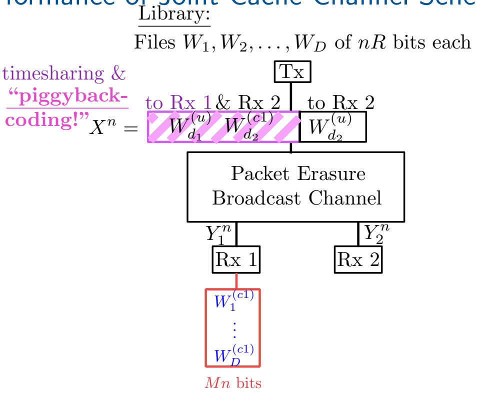


- Rx 1 knows $W_{d_2}^{(c1)}$ $\rightarrow$ restrict decoding to corresponding row

Transmission of $W_{d_2}^{(c1)}$ not affecting Rx 1 at all!

$$p(error) \rightarrow 0 \text{ as } n \rightarrow \infty: \qquad \max\left\{ \frac{R - \frac{M}{D}}{F(1-\delta_1)}, \frac{R}{F(1-\delta_2)} \right\} \leq \frac{n'}{n}$$

Performance of Joint Cache-Channel Scheme for Example

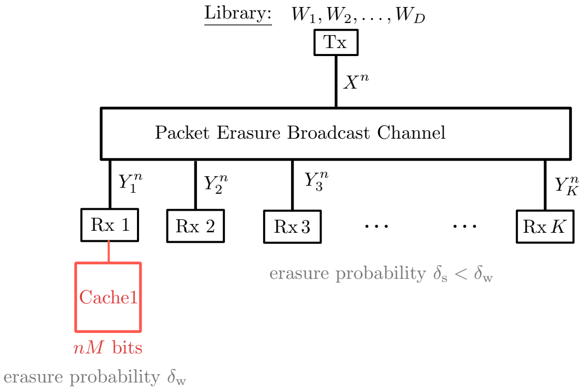


- $W_d = (W_d^{(c1)}, \; W_d^{(u)})$ of sub-rates $(\frac{M}{D}, \; R - \frac{M}{D})$

Joint Cache-Channel Coding → Global Caching Gain!

$$p(error) \to 0 \quad \text{if :} \qquad \underbrace{\max\left\{ \frac{R - \frac{M}{D}}{F(1-\delta_1)}, \; \frac{R}{F(1-\delta_2)} \right\}}_{\text{piggyback coding}} + \frac{R - \frac{M}{D}}{F(1-\delta_2)} \leq 1$$

Extension to Single Cache and K_s strong receivers



- Piggyback cached-parts for *all* strong receivers on weak receiver's uncached part

Performance for Single Cache and K_s strong receivers

- Joint cache-channel coding

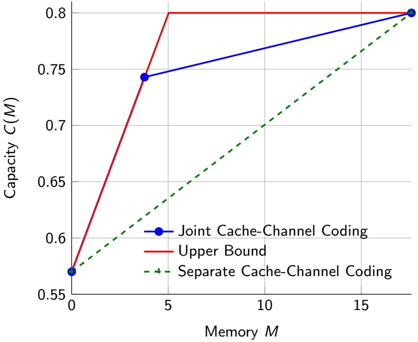
$$R(M) = \begin{cases} F\left(\frac{1}{1-\delta_\mathrm{w}} + \frac{K_\mathrm{s}}{1-\delta_\mathrm{s}}\right)^{-1} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_1] \qquad \textit{tight}! \\ F\frac{(1-\delta_\mathrm{s})}{1+K_\mathrm{s}} + \frac{M}{(1+K_\mathrm{s})D}, & \text{if } \frac{M}{D} \in (\Gamma_1, \Gamma_2], \\ F(1-\delta_\mathrm{s}), & \text{if } \frac{M}{D} \geq \Gamma_2, \end{cases}$$

with $\Gamma_1 := F\frac{(1-\delta_\mathrm{s})}{K_\mathrm{s}}\frac{(\delta_w-\delta_\mathrm{s})}{(K_\mathrm{s}(1-\delta_w)+(1-\delta_\mathrm{s}))}$ and $\Gamma_2 := \frac{(1-\delta_\mathrm{s})}{K_\mathrm{s}}F$.

- Separate cache-channel coding:

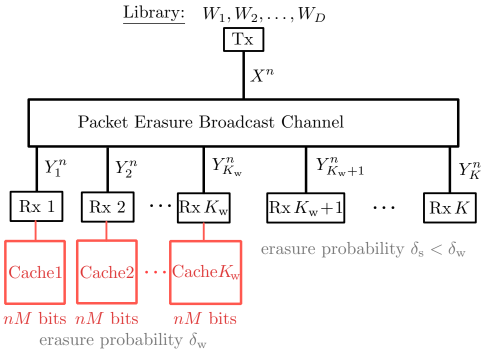
$$R(M) = \begin{cases} F\left(\frac{1}{1-\delta_\mathrm{w}} + \frac{K_\mathrm{s}}{1-\delta_\mathrm{s}}\right)^{-1} + \frac{1-\delta_\mathrm{s}}{1-\delta_\mathrm{s}+K_\mathrm{s}(1-\delta_\mathrm{w})}\frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_2], \\ F(1-\delta_\mathrm{s}), & \text{if } \frac{M}{D} \geq \Gamma_2, \end{cases}$$

Numerical Comparison for Single Cache and K_s strong receivers



$$\delta_w = 0.8, \quad \delta_s = 0.2, \quad K_s = 10, \quad D = 22, \quad F = 10$$

Joint Cache-Channel Scheme with Many Caches

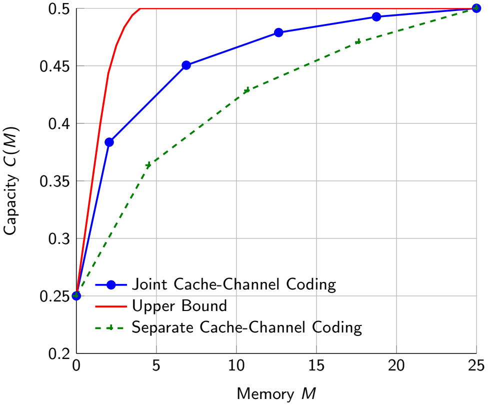


- Split $W_d = (W_d^{(t)}, W_d^{(t-1)})$

- Weak receivers: Maddah-Ali&Niesen for $W_d^{(t)}$ with $(t+1)$-XORs and $W_d^{(t-1)}$ with t-XORs

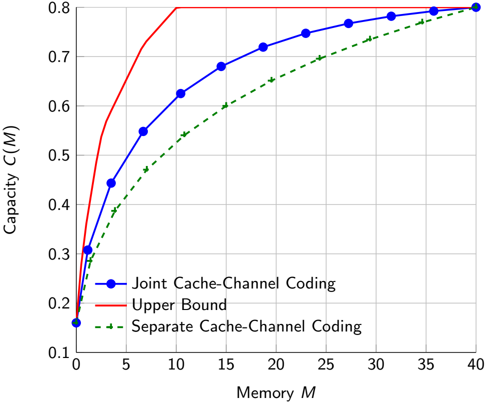
- Piggyback $W_d^{(t)}$ for strong receivers on t-XORs to weak receivers

Numerical Comparison with Multiple Caches, Example I



$$\delta_{\mathrm{w}} = 0.8, \;\; \delta_{\mathrm{s}} = 0.2, \;\; K_{\mathrm{w}} = 4, \; K_{\mathrm{s}} = 16, \;\; D = 50, \;\; F = 10$$

Numerical Comparison with Multiple Caches, Example II



$$\delta_{\mathrm{w}} = 0.8, \quad \delta_{\mathrm{s}} = 0.2, \quad K_{\mathrm{w}} = 10, \; K_{\mathrm{s}} = 10, \quad D = 50, \quad F = 10$$

Most Significant Gains for Low Cache Sizes $M/D \leq \Gamma_1$

$$C(M) \geq R_0 + \frac{M}{D} \cdot \gamma_{\text{local}} \cdot \gamma_{\text{global,sep}} \cdot \gamma_{\text{global,joint}}, \qquad M/D \leq \Gamma_1,$$

where R_0 is the symmetric capacity without caches and

$$\gamma_{\text{local}} := \frac{K_{\text{w}}(1-\delta_{\text{s}})}{K_{\text{w}}(1-\delta_{\text{s}}) + K_{\text{s}}(1-\delta_{\text{w}})},$$

$$\gamma_{\text{global,sep}} := \frac{1+K_{\text{w}}}{2},$$

$$\gamma_{\text{global,joint}} := 1 + \frac{2K_{\text{w}}}{1+K_{\text{w}}} \cdot \frac{K_{\text{s}}(1-\delta_{\text{w}})}{K_{\text{w}}(1-\delta_{\text{s}})}.$$

Insights and Intuition

- Important to consider noisy communication channel:
  - Joint cache-channel coding (piggyback coding)
  - Caching gains combine with feedback gains

    [A. Ghorbel, M. Kobayashi, S. Yang, "Cache-enabled broadcast packet erasure channels with state feedback"]
  - Interplay between caching gains and CSI gains

    [J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: interplay of coded-caching and CSIT feedback"]
- Larger caches for weak receivers $\rightarrow$ even more important with joint cache-channel coding
- Piggyback coding useful whenever info for strong Rx in cache of weak Rx!

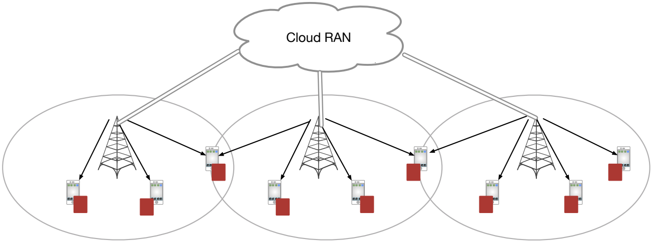
An Information-Theoretic Result on Interference Channels

- Interference networks with tx-caches

  [Maddah-Ali and Niesen, "Cache-aided interference channels," in Proc. of *ISIT15*]

  - Allows interference cancellation
  - Allows loadbalancing to avoid bottlenecks
  - Allows improved interference alignment (X-channel)

Multi-Cell Model with 3 Communication Phases

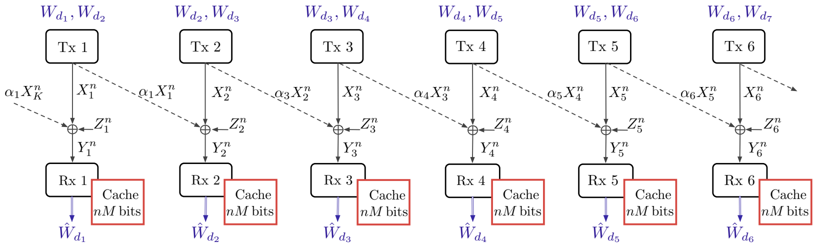


1. Caching Phase: Server fills caches without knowing demands $d_1, \ldots, d_K$

2. Download from Server: Txs download messages of connected receivers from server

3. Delivery to Users: Txs communicate their messages to the receivers

Wyner's Cellular Networks [Shamai, Timo, Wigger 2016]

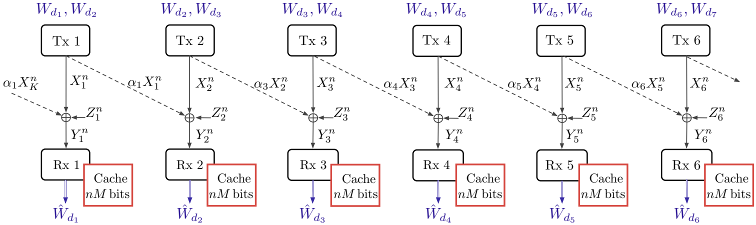


Per-User DoF Rate-Memory Tradeoff

$S(M)$: largest S so that rate $R = S \cdot \frac{1}{2}\log(1+P)$ achievable for large $P \gg 1$ and $M = \mu \cdot \frac{1}{2}\log(1+P)$

- Upper and lower bounds for symmetric and asymmetric Wyner network
- Complete interference mitigation possible

Bounds on Per-User DoF Rate-Memory Tradeoff for Asymmetric Model

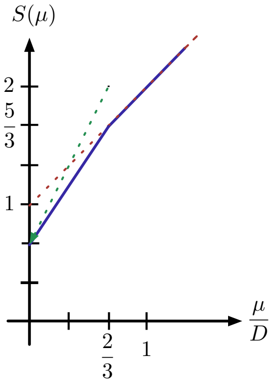


Theorem (Shamai, Timo, Wigger'16)

$$\min\left\{\frac{2}{3}+\frac{3\mu}{D}\ ,\ 1+\frac{\mu}{D}\right\} \geq S \geq \begin{cases} \frac{2}{3}+\frac{3}{2}\frac{\mu}{D}, & \text{if } 0 \leq \frac{\mu}{D} \leq \frac{2}{3} \\ 1+\frac{\mu}{D}, & \text{if } \frac{\mu}{D} \geq \frac{2}{3}. \end{cases}$$

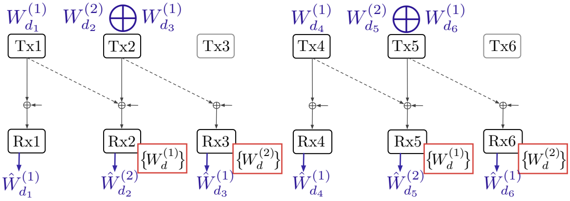
- $S = 1 + \frac{\mu}{D}$ performance of interference-free P2P links with caches
- Factor $\frac{3}{2}$ in lower bound: need two cache memories to free up one DoF

Interference can Completely be Canceled when $\mu \geq 2/3D$

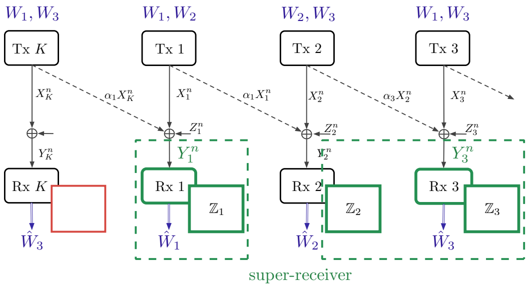


Scheme achieving $\mu = 2/3$ and $S = 5/3$ for Asymmetric Network

- Split $W_d = (W_d^{(1)}, W_{d_2}^{(2)})$ at DoF $(1, 1)$
- Round-rob scheme among all users
- Need $\mu/D = 2/3$ and achieve $S = 5/3$

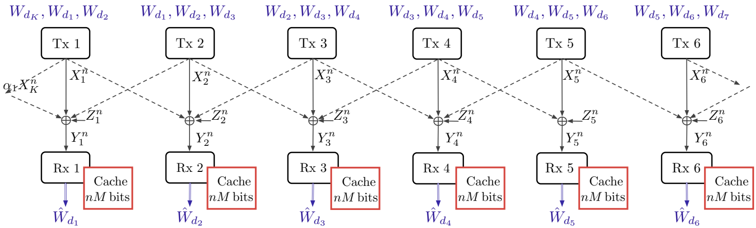


Converse for Asymmetric Network



- Choose an arbitrary demand vector (worst-case error)
- Consider triples of receivers $\rightarrow$ global caching gain limited by 3
- MAC-type bound with caching

Bounds on Per-User DoF Rate-Memory Tradeoff for Symmetric Model

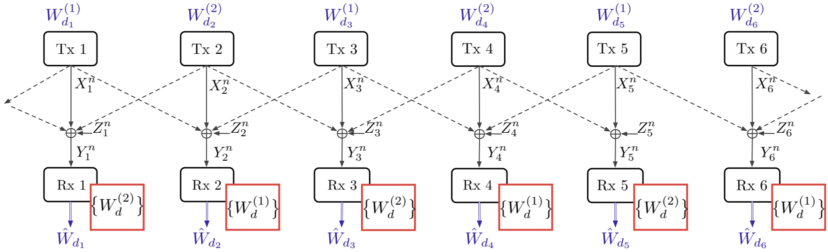


Theorem (Shamai,Timo,Wigger'16)

$$\min\left\{\frac{2}{3}+\frac{6\mu}{D}\ ,\ 1+\frac{\mu}{D}\right\} \geq S \geq \begin{cases} \frac{2}{3}+\frac{4}{3}\frac{\mu}{D}, & \text{if } 0 \leq \frac{\mu}{D} \leq 1 \\ 1+\frac{\mu}{D}, & \text{if } \frac{\mu}{D} \geq 1. \end{cases}$$

- For $\frac{\mu}{D}$: performance of interference-free P2P links with caches
- Factor $\frac{\mu}{D} \geq \frac{4}{3}$ in lower bound: need 3 cache memories to free up 1 DoF

Coding Scheme for Symmetric Network



- Cache contents allow receivers to first cancel interference
- Needs $\mu = 1$ for $S = 2$

Summary

- Important to consider noisy channels $\rightarrow$ joint cache-channel coding
- New piggyback coding idea
- Receiver-caching allows to completely cancel interference in cellular networks
  $\rightarrow$ smart code design significantly reduces required cache memories
- Outer bounds for degraded BCs and sparse interference networks with receiver caching