Broadcasting with Side-Information: Applications to Caching and Feedback Communication

Michèle Wigger
Telecom ParisTech
michele.wigger@telecom-paristech.fr

INSA Lyon, 18 September 2015

joint work with T. Laich, T. Oechtering, R. Timo, and Y. Wu

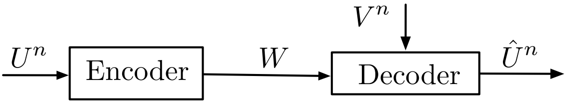
Outline

1. Broadcasting with side-information: source coding and channel coding
2. Caching
3. Broadcast channel with feedback

1. Broadcasting with side-information:

a) source coding

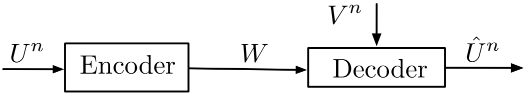
Lossless Source Coding with SI: Single Decoder

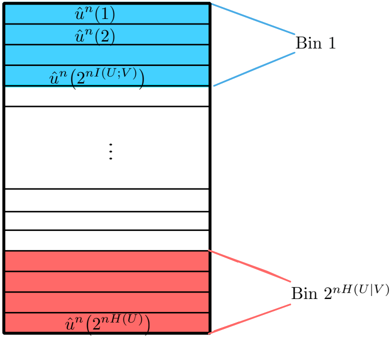


- $\{(U_i, V_i)\}$ IID $\sim P_{UV}$
- Message $W \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$

Slepian-Wolf '73: Lossless reconstruction possible iff $R^\star = H(U|V)$

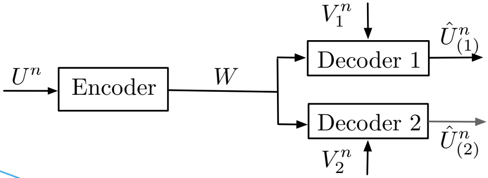
Slepian-Wolf Coding

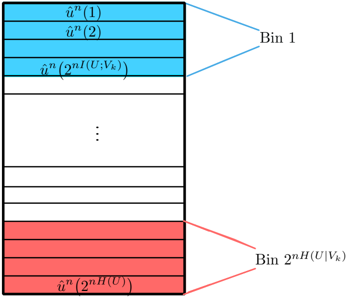




- W =bin-index
- Works for all bin sizes $< 2^{nI(U;V)}$

Sgarro'77: Slepian-Wolf Coding for Many Receivers

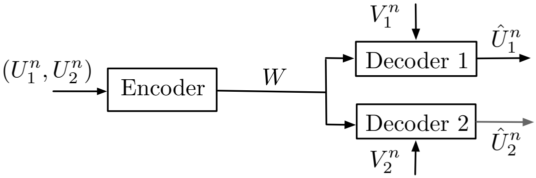




- W =bin-index
- Works for bin sizes $< \min_{k=1,2} 2^{nI(U;V_k)}$

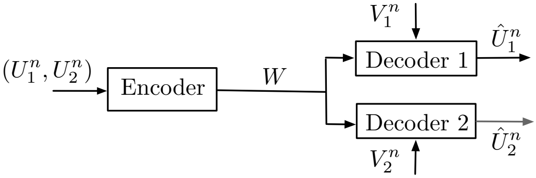
$$R^\star = \max_{k\in\{1,2\}} H(U|V_k)$$

Lossless Heegard-Berger/Kaspi Problem with Two Sources



- Need lossless reconstructions of U_1^n and U_2^n

Coding Scheme for Heegard-Berger/Kaspi Problem



- Lossily describe common part A^n to both decoders
- U_1^n to Decoder 1
- U_2^n to Decoder 2

$$R^\star \leq \min_A \Big\{ \max \{ I(A; U_1, U_2 | V_1), I(A; U_1, U_2 | V_2) \} + H(U_1 | A, V_1) + H(U_2 | A, V_2) \Big\}$$

Physically Degraded Side-Information

$$(U_1^n, U_2^n) \multimap V_1^n \multimap V_2^n$$

- Achievability tight with $A = U_2$ (Kaspi'94, Heegard/Berger'85):

$$R^\star = H(U_2|V_2) + H(U_1|V_1, U_2)$$

Conditionally Less-Noisy SI [Timo, Oechtering, W'14]

Definition

V_1 is conditionally less noisy than V_2 given U_2, $(V_1 \succeq V_2|U_2)$, if

$$I(B; V_1|U_2) \geq I(B; V_2|U_2)$$

for all $B - \circ - (U_1, U_2) - \circ - (V_1, V_2)$.

Theorem

If $(V_1 \succeq V_2|U_2)$ *and* $H(U_2|V_1) \leq H(U_2|V_2)$*, then*

$$R^\star = H(U_2|V_2) + H(U_1|V_2, V_1)$$

Converse based on Entropy-Characterization Problem

- Converse for physically degraded SI does not apply/cannot be extended
- Converse for conditionally less-noisy SI relies on:

Lemma (Entropy-Characterization Lemma)

Assume

$$(R^n, {S_1}^n, S_2^n, T^n, L^n) \text{ IID } \sim (R, S_1, S_2, T, L)$$

and

$$J -\!\!\circ\!\!- (R^n, L^n) -\!\!\circ\!\!- (S_1^n, S_2^n, T^n).$$

There exists a W *with cardinality constraint* $|\mathcal{W}| \leq |\mathcal{R}||\mathcal{L}|$ *such that*

$$I(J; S_2^n|L^n) - I(J; S_1^n|L^n) = n\big(I(W; S_2|L) - I(W; S_1|L)\big)$$

and $W -\!\!\circ\!\!- (R, L) -\!\!\circ\!\!- (S_1, S_2, T)$.

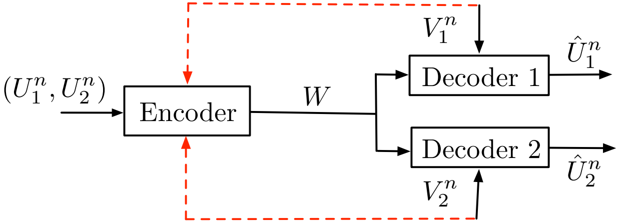
- Proof of lemma by Kramer's telescoping identity or Csiszar's sum-identity

Extensions

We can extend our result on minimum description length to:

- $K \geq 2$ decoders
- Partially lossy case $\rightarrow$ one decoder needs only a lossy reconstruction of its source
- Successive Refinement $\rightarrow$ one decoder obtains an additional private message

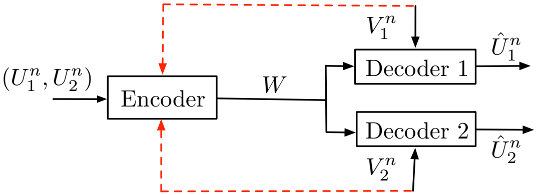
Side-Information also Known at Encoder



- Minimum description rate $R^\star_{\text{cogn}}$

- Encoder SI helps in lossy case [Perron/Diggavi/Telatar'09];
  Not in lossless case when $U_1 = U_2$ [Sgarro'79]

Coding Scheme with Encoder Side-Information

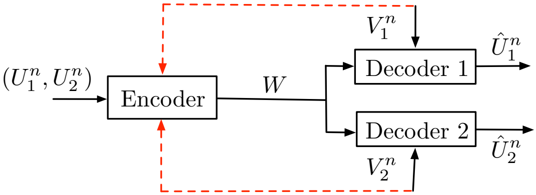


- Describe common part A^n, $g(V_2^n)$ to both decoders
- U_1^n to Decoder 1, U_2^n to Decoder 2

$$R^{\star}_{\text{cogn}} \leq \min_{A} \Big\{ \max\left\{ I(A, g(V_2); U_1, U_2 | V_1), I(A, g(V_2); U_1, U_2 | V_2) \right\} + H(U_1 | A, V_1, g(V_2)) + H(U_2 | A, g(V_2), V_2) \Big\}$$

- Free resources available that we can fill with encoder SI!

Coding Scheme with Encoder Side-Information



- Describe common part A^n, $g(V_2^n)$ to both decoders
- U_1^n to Decoder 1, U_2^n to Decoder 2

$$R^\star_{\text{cogn}} \leq \min_A \Big\{ \max\left\{ I(A, g(V_2); U_1, U_2 | V_1), I(A; U_1, U_2 | V_2) \right\} \\ + H(U_1 | A, V_1, g(V_2)) + H(U_2 | A, V_2) \Big\}$$

- Free resources available that we can fill with encoder SI!

Encoder-SI is Helpful!

- $(V_1 \succeq V_2|U_2)$ but not physically degraded
- $H(U_2|V_2) \geq H(U_2, V_2|V_1) \geq H(U_2|V_1)$
- Without Encoder-SI:

$$R^\star = H(U_2|V_2) + H(U_1|U_2, V_1)$$

- With Encoder-SI, choosing $g(V_2) = V_2$:

$$R^\star_{\text{cogn}} = H(U_2|V_2) + H(U_1|U_2, V_1, V_2) < R^\star$$

Encoder-SI is Helpful!

- $(V_1 \succeq V_2 | U_2)$ but not physically degraded
- $H(U_2|V_2) \geq H(U_2, V_2|V_1) \geq H(U_2|V_1)$
- Without Encoder-SI:

$$R^\star = H(U_2|V_2) + H(U_1|U_2, V_1)$$

- With Encoder-SI, choosing $g(V_2) = V_2$:

$$R^\star_{\text{cogn}} = H(U_2|V_2) + H(U_1|U_2, V_1, V_2) < R^\star$$

→ Sending V_2^n purely beneficial: helps Dec 2 without additional constraint

Ex. of Sources and SI where Encoder-SI strictly helps

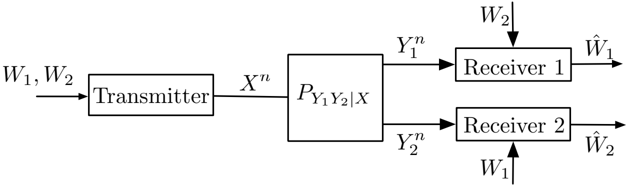
- Source $\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \text{DSBS}(p)$
- $E \sim \mathbb{B}(1/3)$ independent of source
- SI $Y_1 = \begin{cases} U_1, U_2 & \text{if } E = 1 \\ ? & \text{if } E = 0 \end{cases}$ and $V_2 = \begin{cases} ? & \text{if } E = 1 \\ U_1, U_2 & \text{if } E = 0 \end{cases}$

$$R^\star = 2/3 + 1/3 H_b(p) > R^\star_{\text{cogn}} = 2/3$$

1. Broadcasting with side-information: b) channel coding

Complementary Delivery over Discrete Memoryless BC

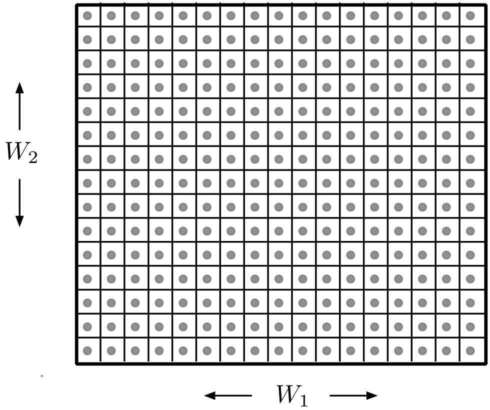
- Receiver k desires $W_k \in \{1, \ldots, 2^{nR_k}\}$
- Capacity region: Pairs (R_1, R_2) s.t. $p(\text{error})$ arbitrarily small



- Capacity: $R_1 \leq I(X;Y_1)$ and $R_2 \leq I(X;Y_2)$ for some P_X
- Given P_X: point-to-point performance to both receivers

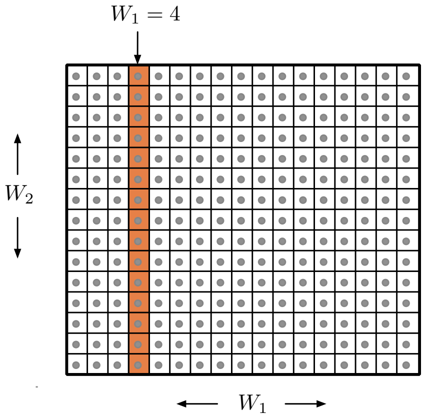
Complementary Delivery: A Capacity-Achieving Scheme

- Product codebook $\{x^n(w_1, w_2)\}$



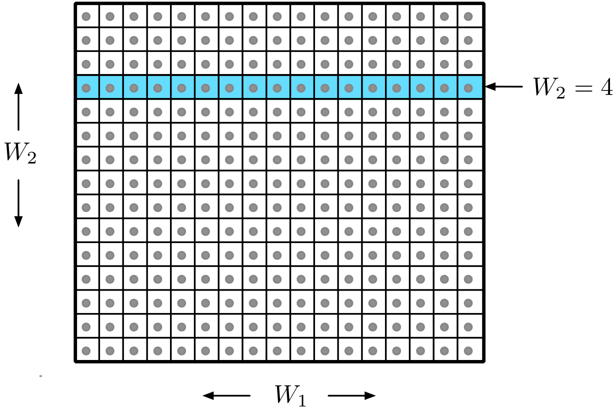
Complementary Delivery: A Capacity-Achieving Scheme

- Product codebook $\{x^n(w_1, w_2)\}$

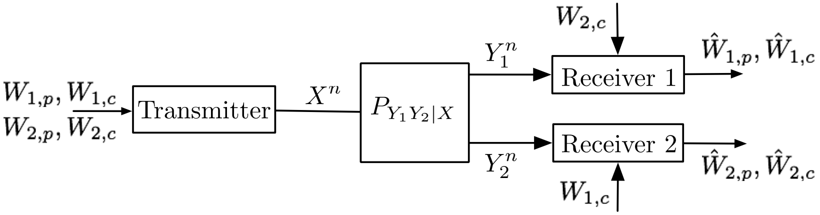


Complementary Delivery: A Capacity-Achieving Scheme

- Product codebook $\{x^n(w_1, w_2)\}$



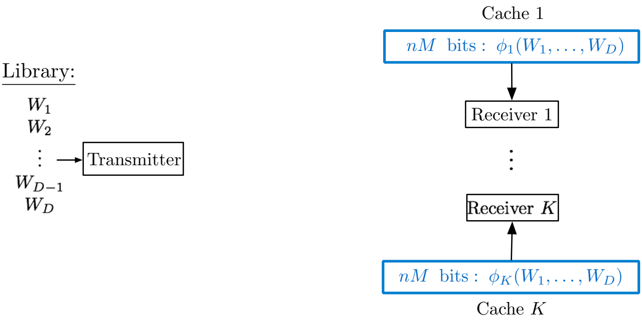
BC with Partial Message Side-Information



- Degraded message sets solved [Kramer & Shamai 2007]
- Semideterministic BC solved [Bracher& Wigger2015]

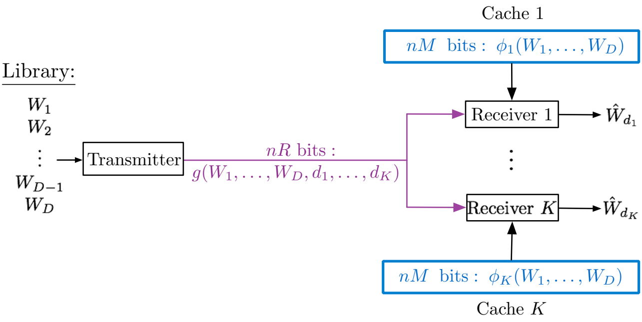
2. Caching

Caching—The Source Coding Version [Maddah-Ali&Niesen 2013]



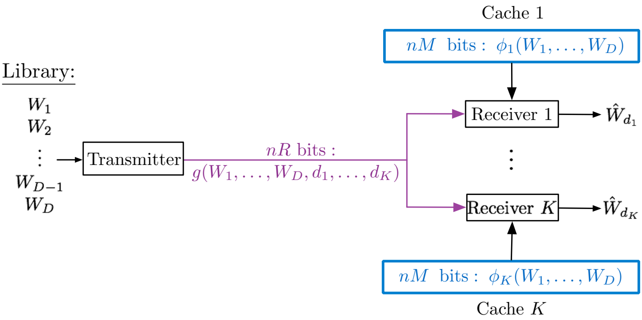
- W_d has $n\rho$ i.i.d. bits
- Placement phase: Tx fills caches without knowing actual demands $d_1, \ldots, d_K$
- Delivery phase: demands known everywhere

Caching—The Source Coding Version [Maddah-Ali&Niesen 2013]



- W_d has $n\rho$ i.i.d. bits
- Placement phase: Tx fills caches without knowing actual demands $d_1, \ldots, d_K$
- Delivery phase: demands known everywhere

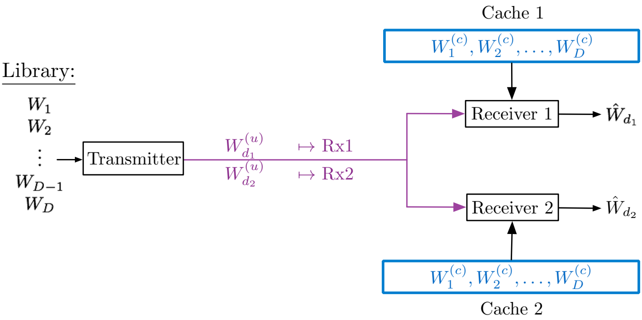
Caching—The Source Coding Version [Maddah-Ali&Niesen 2013]



Rate-Memory Tradeoff

For which description rates and memory M is reconstruction possible?

Uncoded Caching for $K = 2$

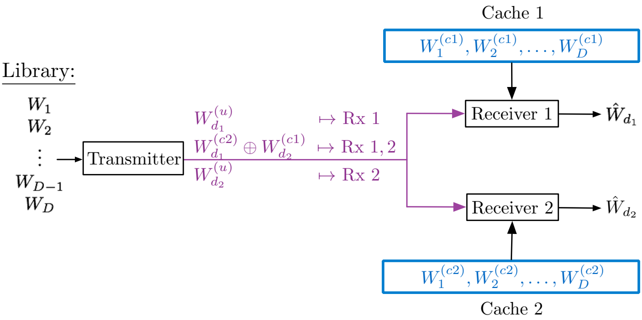


- Split $W_d = (W_d^{(c)}, W_d^{(u)})$ of rates $\left(\frac{\mathsf{M}}{D}, \rho - \frac{\mathsf{M}}{D}\right)$

Rates-Memory Trade-Off (for $D \gg K$)

Reconstruction is possible, if $R = 2\left(\rho - \frac{\mathsf{M}}{D}\right)$

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

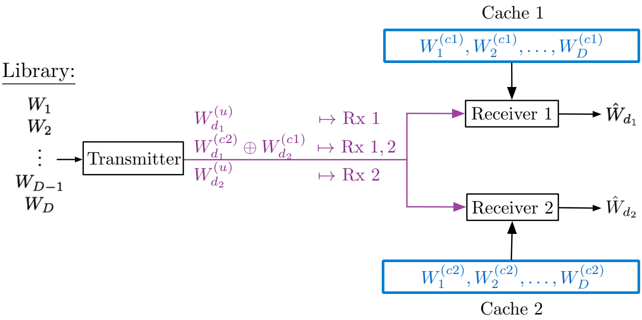


- Split $W_d = (W_d^{(c1)}, W_d^{(c2)}, W_d^{(u)})$ of rates $\left(\frac{\mathsf{M}}{D}, \frac{\mathsf{M}}{D}, \rho - 2\frac{\mathsf{M}}{D}\right)$

Rates-Memory Trade-Off (for $D \gg K$)

Reconstruction possible, if $R = 2\left(\rho - \frac{\mathsf{M}}{D}\right) - \frac{\mathsf{M}}{D}$

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



- Split $W_d = (W_d^{(c1)}, W_d^{(c2)}, W_d^{(u)})$ of rates $(\mathsf{M}/D, \mathsf{M}/D, \rho - 2\mathsf{M}/D)$

Rate-Memory Trade-Off

$$R^\star \leq 2(\rho - \mathsf{M}/D) - \mathsf{M}/D$$

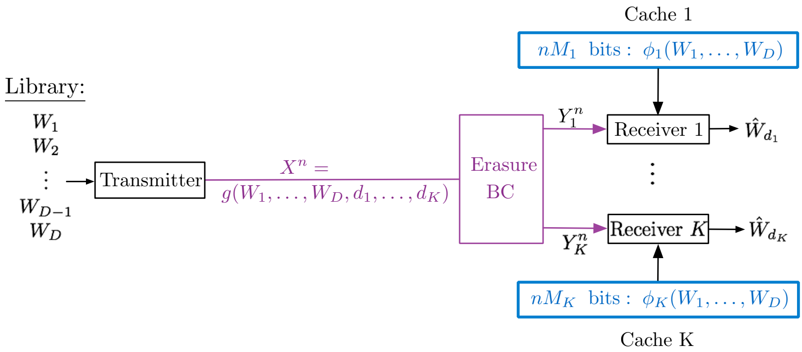
Local and Global Caching Gains $K \geq 2$ [Maddah-Ali&Niesen 2013]

- Choose $t \triangleq \frac{MK}{D\rho}$ (assumed integer)
- Split every W_d into $\binom{K}{t}$ submessages and cache each of them at t receivers
- Send x-or of $t+1$ submessages to the group of $t+1$ receivers that know all except for one submessage

Coded caching achieves (for $D \gg K$)

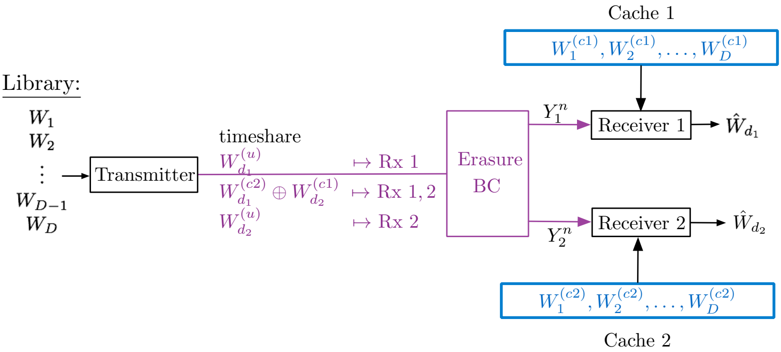
Reconstruction possible, if $R = K(\rho - \frac{\mathsf{M}}{D}) \cdot \frac{1}{1+K\mathsf{M}/\rho/D}$

Caching over Packet Erasure BCs [Timo&W'2015]



- Input X: F bits
- Receiver k gets erasure with probability δ_k where $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_K$

Scheme 1: Coded Caching and Separate Channel Coding

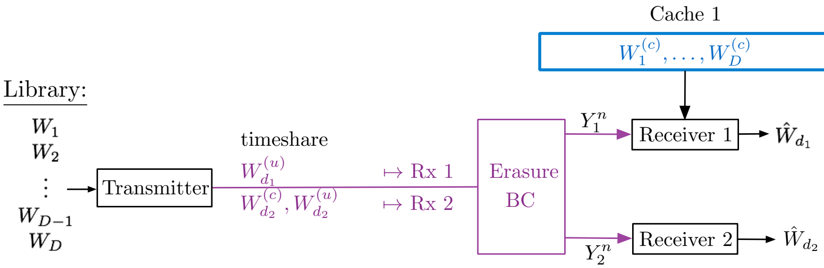


- Split $W_d = (W_d^{(c1)}, W_d^{(c_2)}, W_d^{(u)})$ of rates $(\frac{\mathsf{M}}{D}, \frac{\mathsf{M}}{D}, \rho - 2\frac{\mathsf{M}}{D})$

Achievable rates

p(error) $\to 0$ possible if: $$\frac{\rho - \frac{\mathsf{M}}{D}}{F(1-\delta_1)} + \frac{\rho - 2\frac{\mathsf{M}}{D}}{F(1-\delta_2)} \leq 1$$

Scheme 2: Asymmetric Caching and Separate Channel Coding

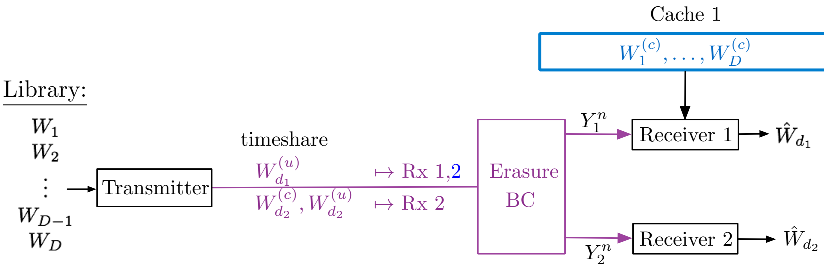


- Split $W_d = (W_d^{(c)}, W_d^{(u)})$ of rates $(2\frac{\mathsf{M}}{D}, \rho - 2\frac{\mathsf{M}}{D})$

Asymmetric Cache Assignment Can Help

$p(\text{error}) \to 0$ if: $$\frac{\rho - 2\frac{\mathsf{M}}{D}}{F(1-\delta_1)} + \frac{\rho}{F(1-\delta_2)} \leq 1$$

Scheme 2: Asymmetric Caching and Separate Channel Coding

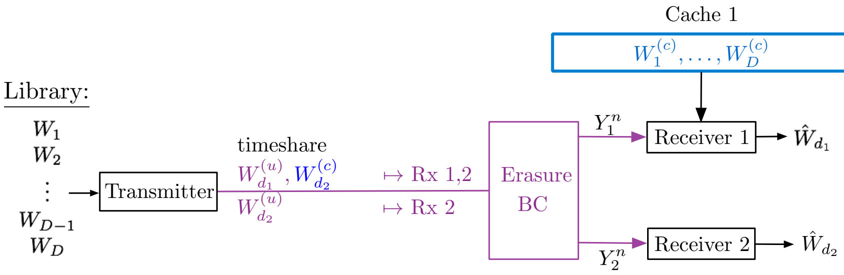


- Split $W_d = (W_d^{(c)}, W_d^{(u)})$ of rates $(2\frac{\mathsf{M}}{D}, \rho - 2\frac{\mathsf{M}}{D})$

Asymmetric Cache Assignment Can Help

$p(\text{error}) \to 0$ if: $\quad \dfrac{\rho - 2\frac{\mathsf{M}}{D}}{F(1-\delta_1)} + \dfrac{\rho}{F(1-\delta_2)} \leq 1$ No Global Caching Gain

Our Joint-Cache Channel Scheme for $K = 2$



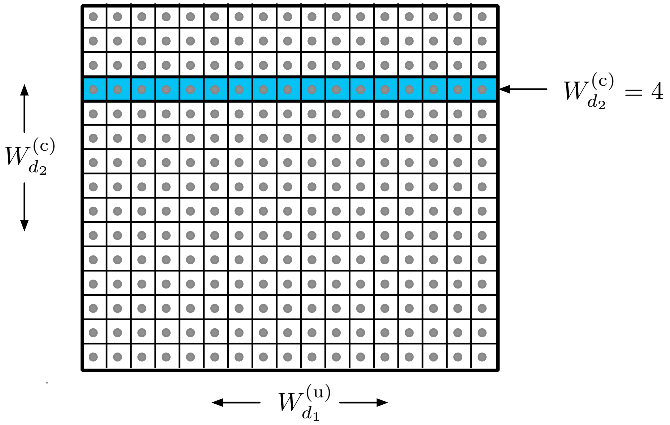
- Split $W_d = (W_d^{(c)}, W_d^{(u)})$ of rates $(2\frac{\mathsf{M}}{D}, \rho - 2\frac{\mathsf{M}}{D})$

Joint Cache-Channel Coding with Piggyback coding $\rightarrow$ Global Caching Gain

$p(\text{error}) \rightarrow 0$ if: $\quad \dfrac{\rho - 2\frac{\mathsf{M}}{D}}{F(1-\delta_1)} + \dfrac{\rho - 2\frac{\mathsf{M}}{D}}{F(1-\delta_2)} \leq 1 \quad$ and $\quad \dfrac{2\rho - 2\frac{\mathsf{M}}{D}}{F(1-\delta_2)} \leq 1$

Piggyback Coding to send $(W_{d_1}^{(u)}, W_{d_2}^{(c)})$

- Rx 1 already knows $W_{d_2}^{(c)}$ $\rightarrow$ Transmission of $W_{d_2}^{(c)}$ does not bother Rx 1



Example for $K = 2$, $\delta_1 = 4/5$ and $\delta_2 = 1/5$ and M$\leq \rho 3D/8$

1. Symmetric caches $\mathsf{M}_1 = \mathsf{M}_2 = \mathsf{M}$ & coded caching with separate source-channel coding

$$\rho \leq \frac{4}{5} F(1-\delta_1) + \frac{6}{5}\frac{\mathsf{M}}{D}$$

2. Asymmetric caches $\mathsf{M}_1 = 2\mathsf{M}$ and $\mathsf{M}_2 = 0$ & separate source-channel coding

$$\rho \leq \frac{4}{5} F(1-\delta_1) + \frac{8}{5}\frac{\mathsf{M}}{D}$$

3. Asymmetric caches $\mathsf{M}_1 = 2\mathsf{M}$ and $\mathsf{M}_2 = 0$ & joint cache-channel coding

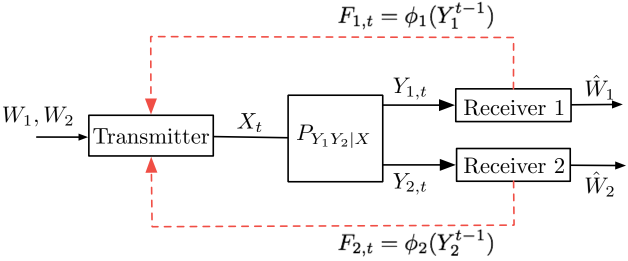
$$\rho \leq \frac{4}{5} F(1-\delta_1) + 2\frac{\mathsf{M}}{D}$$

Our General Joint Cache Channel Scheme

- Assign larger cache sizes to weaker receivers
- Do Maddah-Ali&Niesen coded caching:
  - Cache submessages to groups of t receivers
  - Deliver x-ors of submessages to $t + 1$ receivers
- For the delivery phase piggyback information to stronger receivers on x-or messages to weaker receivers!

3. Broadcast channel with feedback

Noise-Free Rate-Limited Feedback



- Feedback rate constraint: $|\mathcal{F}_{i,1}| \cdots |\mathcal{F}_{i,n}| \leq nR_{\text{fb},i}, \quad i = 1, 2$
- $X_t = f_t(W_1, W_2, F_1^{t-1}, F_2^{t-1})$

Previous Results: Perfect two-sided feedback $R_{\text{fb},1} = R_{\text{fb},2} = \infty$

- El Gamal'78: No feedback-gain for physically degraded BCs

- Dueck'80, Ozarow'85, Kramer'00, Wang'09, Tassiulas&Georgiadis'10, Shayevitz&W'10, Maddah-Ali&Tse'10: Feedback-gain for some BCs

- Kramer'00, Shayevitz-W'13, Venkataramanan-Pradhan'13: achievable region for general memoryless BCs (difficult to evaluate)

For most BCs: don't know whether feedback helps

(Strictly) Less-Noisy DMBCs

Less-Noisy DMBC $Y_2 \succeq Y_1$

For every auxiliary $U \multimap X \multimap (Y_1, Y_2)$:

$$I(U;Y_2) \geq I(U;Y_1)$$

- Binary Symmetric BC
- Binary Erasure BC
- Binary Symmetric/Erasure BC for certain parameters

(Strictly) Less-Noisy DMBCs

Strictly Less-Noisy DMBC $Y_2 \succ Y_1$

For every auxiliary $U -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1, Y_2)$ with $I(U; Y_1) > 0$:

$$I(U; Y_2) > I(U; Y_1)$$

- Asymmetric Binary Symmetric BC
- Asymmetric Binary Erasure BC
- Binary Symmetric/Erasure BC for certain parameters

(Strictly) Less-Noisy DMBCs

Strictly Less-Noisy DMBC $Y_2 \succ Y_1$

For every auxiliary $U \to X \to (Y_1, Y_2)$ with $I(U; Y_1) > 0$:

$$I(U; Y_2) > I(U; Y_1)$$

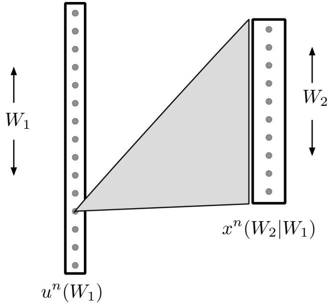
- Asymmetric Binary Symmetric BC
- Asymmetric Binary Erasure BC
- Binary Symmetric/Erasure BC for certain parameters

Nofeedback-Capacity of Less-Noisy DMBCs, $(I(U;Y_2) \geq I(U;Y_1))$

- Capacity: all rate pairs (R_1, R_2) where for some $U \to X \to (Y_1, Y_2)$

$$R_1 \leq I(U;Y_1)$$
$$R_2 \leq I(X;Y_2|U)$$

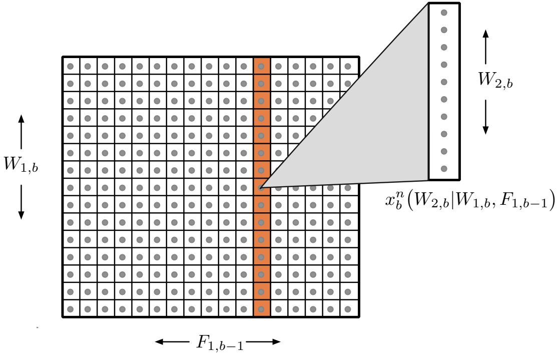
- Achieved by superposition coding



- If $I(U;Y_2) > I(U;Y_1)$, Rx 2 could even decode an extra message in cloud center
- Problem: Rx 1 cannot decode, unless it knows this extra message...

BC-scheme with feedback from the weaker Receiver 1

- Block-Markov coding with a piggyback superposition code in each block b:



- Choose $F_{1,b-1}$ as a Wyner-Ziv message to compress $Y_{1,b-1}^n \to \tilde{Y}_{1,b-1}^n$

- Rx 2 decodes $W_{2,b}$ based on $(\tilde{Y}_{1,b}^n, Y_{2,b}^n)$

A Simple Achievable Region

Theorem

(R_1, R_2) achievable, if for some $P_U P_{X|U} P_{\hat{Y}_1|UY_1}$:

$$R_1 \leq I(U;Y_1)$$
$$R_2 \leq I(X;\tilde{Y}_1,Y_2|U) = I(X;Y_2|U) + I(X;\tilde{Y}_1|U,Y_2)$$

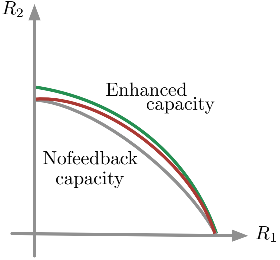
and $$I(\tilde{Y}_1;Y_1|U,Y_2) \leq \min\{R_{\mathsf{FB},1}, I(U;Y_2) - I(U;Y_1)\}.$$

- Sending $\tilde{Y}_1$ is purely beneficial: not bothering Rx 1 and helping Rx 2

"Duality" to Cover-Leung Scheme for MAC with Feedback

	Cover-Leung on MAC	Wu-Wigger on BC
New path	*tx* $\rightarrow$ *rx* $\rightarrow$ *tx*	*rx* $\rightarrow$ *tx* $\rightarrow$ *rx*
Side-Information	*cognitive txs*	*clustered decoding*
Common Info	*shared messages*	*shared outputs*

If $R_{\mathsf{FB},1} > 0$, Entire Capacity Region Increased when $Y_2 \succ Y_1$



Theorem: For Any DMBC $Y_2 \succ Y_1$, when $R_{\mathsf{FB},1} > 0$

Feedback improves all $(R_1 > 0, R_2 > 0)$ of the nofeedback capacity,

unless (R_1, R_2) lies on boundary of capacity of enhanced channel

- Ex.: Asymmetric Binary Symmetric, Binary Erasure, Gaussian BC

Extension to Two-Sided Feedback

- Marton-coding
- Send feedback messages $F_{1,b-1}$ and $F_{2,b-1}$ in cloud center of block b using two-sided piggyback-coding
- Feedback messages compress outputs $Y_{1,b}^n$ or $Y_{2,b}^n$

$F_{1,b-1}$ “transparent” for Receiver 1, $F_{2,b-1}$ for Receiver 2

$\rightarrow$ like “double-booking” resources in cloud-center

Achievable Region with Backward Decoding

Achievable Region Backward Decoding

(R_1, R_2) achievable, if for some
$P_Q P_{U_0 U_1 U_2|Q} P_{\tilde{Y}_1|Y_1 Q} P_{\tilde{Y}_2|Y_2 Q}$ and $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$:

$$\begin{aligned}
R_1 &\leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \\
R_2 &\leq I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) \\
R_1 + R_2 &\leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \\
&\quad + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - I(U_1; U_2|U_0) \\
R_1 + R_2 &\leq I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) \\
&\quad + I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - I(U_1; U_2|U_0) \\
R_1 + R_2 &\leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \\
&\quad + I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) - I(U_1; U_2|U_0)
\end{aligned}$$

and

$$\begin{aligned}
I(\tilde{Y}_1; Y_1|Y_2, Q) &\leq R_{\mathsf{FB},1} \\
I(\tilde{Y}_2; Y_2|Y_1, Q) &\leq R_{\mathsf{FB},2}
\end{aligned}$$

Discussion of our Two-Sided Feedback Schemes

- Can "double-book" resources in cloud center
- Can improve over nofeedback even when $I(U_0; Y_1) = I(U_0; Y_2)$
  (e.g., Blackwell BC with states by Kim/Chia/El Gamal'13)
- Improves over nofeedback for semideterministic BC [Bracher&W'2015]
- Further improvement: Encoder could process feedback info:
  reconstruct outputs $\tilde{Y}_{1,b}, \tilde{Y}_{2,b}$ and compress $(U_{0,b}, U_{2,b}, U_{2,b}, \tilde{Y}_{1,b}, \tilde{Y}_{2,b})$

Noisy Feedback Channel (Instead of a Feedback Pipe)

Our rates remain achievable for noisy feedback channels of capacities $R_{\mathsf{FB},1}$ and $R_{\mathsf{FB},2}$ if receivers can code over feedback links!

Summary

- Lots of open problems in broadcasting with side-information
- Side-information allows to piggyback information for strong receivers
- Joint source-channel coding required
- Caching over erasure broadcast networks: use asymmetric cache sizes and joint cache-channel coding
- BC with feedback:
  - Free resources available in BCs that can be exploited with feedback
    $\rightarrow$ Feedback increases entire capacity for strictly less-noisy BCs
  - Exchange outputs using joint source-channel coding