On the Capacity of Non-coherent Network Coding

Mahdi Jafari Siavoshani
Soheil Mohajer, Christina Fragouli, Suhas Diggavi

École Polytechnique Fédérale de Lausanne

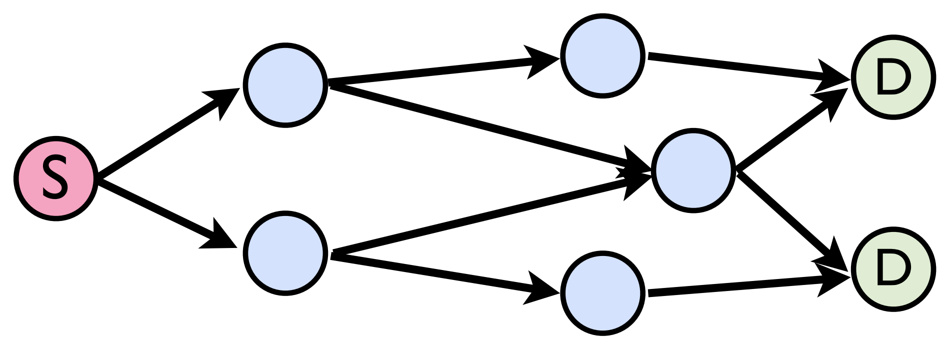
Outline

- Introduction
- Single Source
    - Problem Setup and Model
    - Results
    - Sketch of the Proof
- Multiple Sources
    - Sketch of the Proof
- Compressed Coding Vectors
- Conclusions

Introduction

Randomized Network Coding

- Nodes linearly and uniformly combine the incoming packets.
  - => Sources and destinations are oblivious to the network operation (a non-coherent transmission).



- The standard approach is to append coding vectors to each packet to keep track of the linear operations performed by the network.
  - =>There is a loss of information rate due to coding vector overhead.

Operator Channel - Subspace Coding

Kotter and Kschischang (2008)

- Observation: The linear network coding is vector space preserving.
  - => Information transmission is modeled by the injection of a basis for a vector space Π_S into the network and the collection of a basis for a vector space Π_D by the receiver.
- Network is modeled by the operator channel:

$$\Pi_D = \mathcal{H}_k(\Pi_S) \oplus \Pi_E$$

- KK'08 focused on code construction in $\mathcal{P}\left(\mathbb{F}_q^T\right)$ which is a combinatorial problem.
- They only focused on subspace codes with block length one.

Non-coherent Network Coding

- We may study this problem from information theory point of view by proposing a probabilistic model for the channel.
  - Q 1: What is the maximum achievable rate in such a network with non-coherent assumption when we can use the network many time?
  - Q 2: What is the optimal coding scheme to achieve the capacity?
  - Q 3: How much is the rate loss of using coding vectors compared to the optimal scheme?

Related Work

- R. Koetter and F. Kschischang, "Coding for errors and erasures in random network coding," IT 2008.

- A. Montanari and R. Urbanke, "Coding for network coding", preprint.

- D. Silva, F. R. Kschischang, and R. Koetter, "Communication over finite-field matrix channels," IT 2010.

Problem Setup and Model

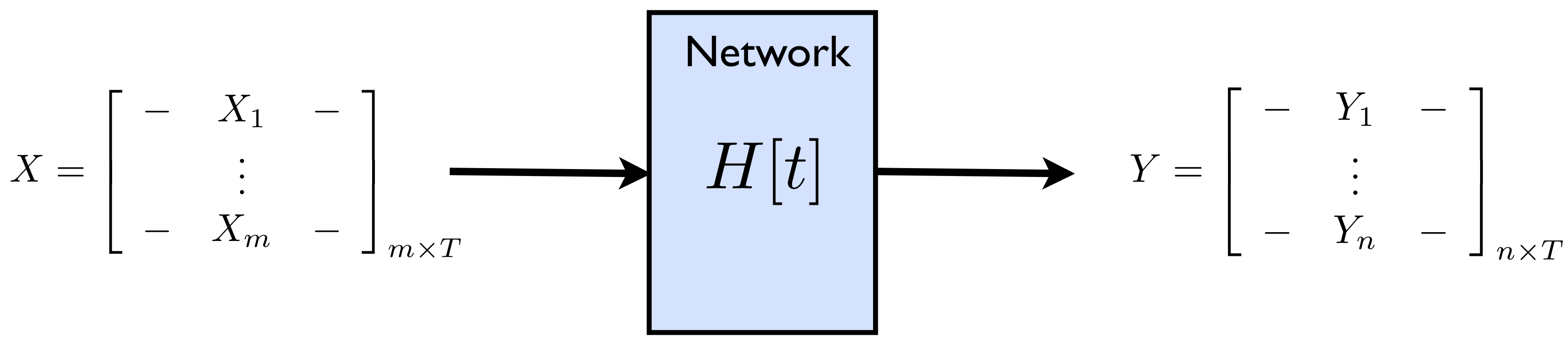
Assumptions

- We assume time is slotted (or we have rounds).
- In each time-slot, the source sends m packets denoted by rows of X, (X is an $m \times T$ matrix over $\mathbb{F}_q$).
- Receiver observes n packets denoted by rows of Y, (an $n \times T$ matrix over $\mathbb{F}_q$).
- Transfer function is unknown to both Tx and Rx, (similar to non-coherent MIMO channel).
- Nodes perform uniform at random randomized network coding over $\mathbb{F}_q$.

$$X = \begin{bmatrix} - & X_1 & - \\ & \vdots & \\ - & X_m & - \end{bmatrix}_{m\times T} \qquad Y = \begin{bmatrix} - & Y_1 & - \\ & \vdots & \\ - & Y_n & - \end{bmatrix}_{n\times T}$$

Channel Model

- The channel model is a block time-varying channel.
- For each time-slot we have: $Y_{n\times T}[t] = H_{n\times m}[t]X_{m\times T}[t]$
- Matrix $H[t]$ is assumed to be uniformly distributed over all possible matrices and independent over different blocks.



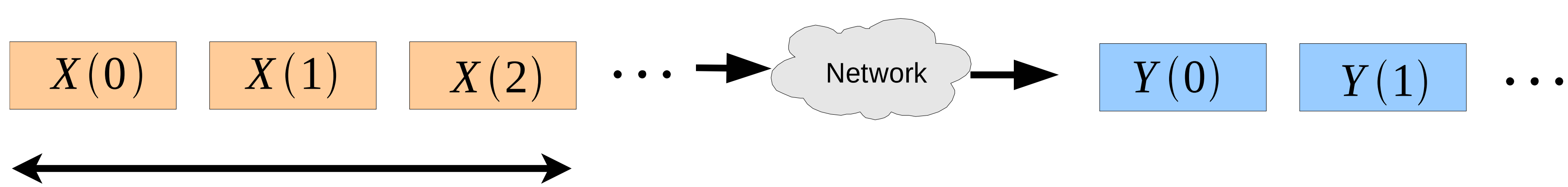
- The packet length T can be interpreted as the coherence time of the channel, during which the transfer matrix remains constant.

Notion of Capacity

- Considering a coding scheme over multiple blocks, the problem becomes an information theoretical problem with channel capacity:

$$C = \max_{P_X} I(X;Y)$$

$$X \in \mathbb{F}_q^{m \times T}, \quad Y \in \mathbb{F}_q^{n \times T}$$



A codeword is a sequence of matrices

Results

Coding over Subspaces is Optimal!

- For the channel transition probability we can show:

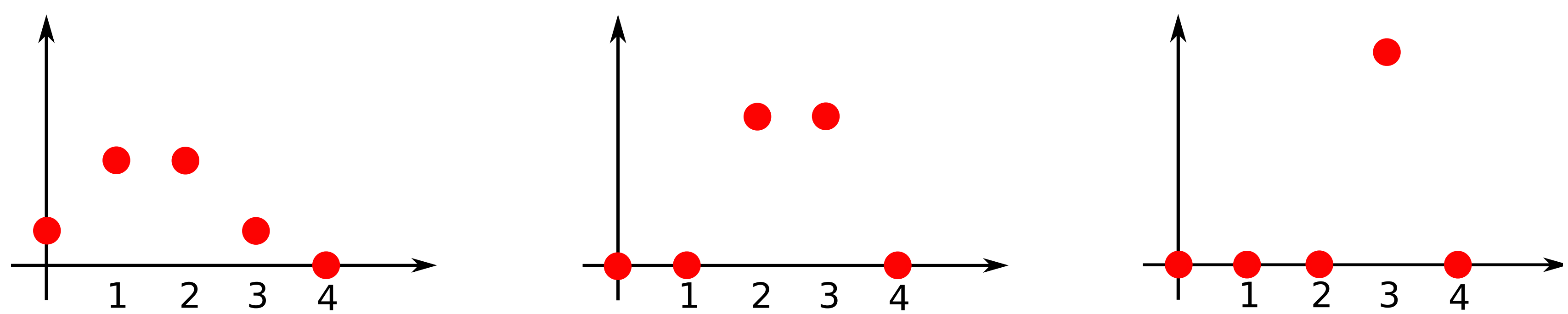
$$\mathbb{P}[Y = y | X = x] = \begin{cases} q^{-n \dim(\langle x \rangle)} & \langle y \rangle \sqsubseteq \langle x \rangle \\ 0 & \text{otherwise} \end{cases}$$

- Conclusions:
  - Coding over subspaces is optimal.
  - Because of the symmetry, the optimal input distribution is uniform over all subspaces having the same dimension.

- Question: What is the optimal input distribution over subspaces with different dimensions?

Illustration of Main Result

- The channel is: $Y_{n\times T} = H_{n\times m} X_{m\times T}$
- There are different regimes, based on relative values of m, n, and T.
- Example: Active subspace dimensions for m = 4, n = 3:



$T \leq n$ $\quad$ $n < T < n + \min[m, n]$ $\quad$ $n + \min[m, n] \leq T$

Main Result

- Theorem:
  - There exists finite q_0 such that for $q > q_0$ the optimal input distribution is non-zero only for the matrices whose rank belongs to the active set:

$$\mathcal{A} = \left\{\min[(T-n)^+, m, n, T], \ldots, \min[m, n, T]\right\}$$

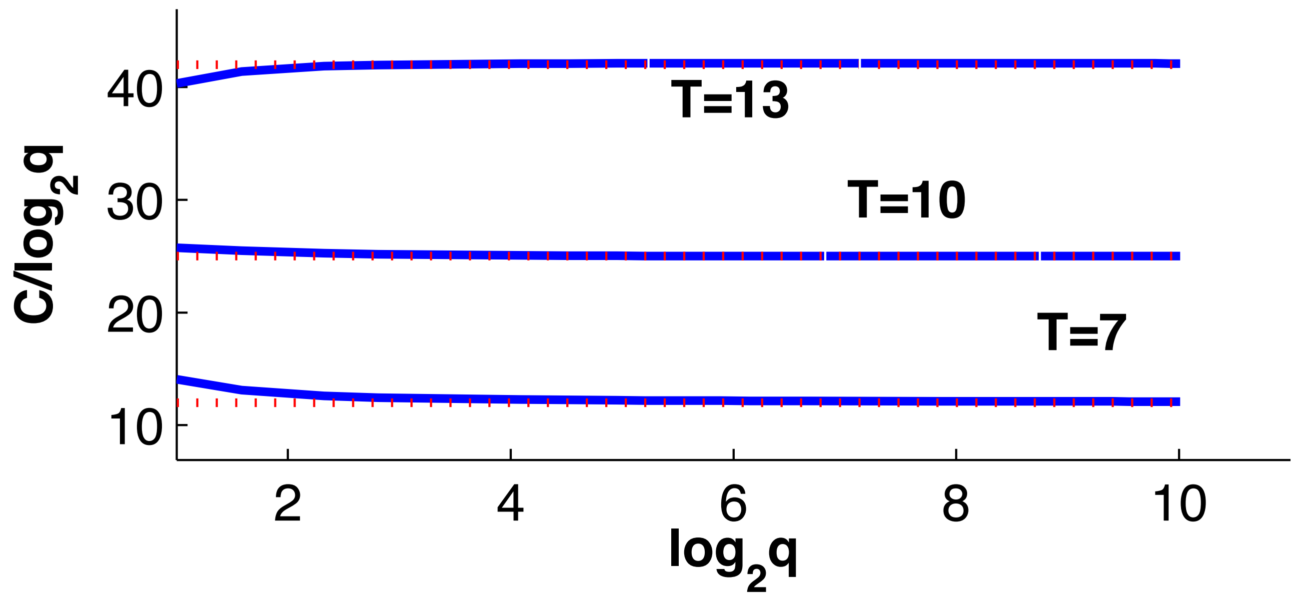
  - The total probability allocated to transmitting matrices of rank i equals:

$$\alpha_i^* \triangleq \mathbb{P}[\text{rank}(X) = i] = 2^{-C} q^{i(T-i)}[1 + o(1)], \quad \forall i \in \mathcal{A}$$

Main Result

- Theorem:
  - The capacity is given by: $C = i^*(T - i^*) \log_2 q + o(1)$
  - where $i^* = \min\left[m, n, \lfloor T/2 \rfloor\right]$

- Numerical calculations show fast convergence of capacity to above result even for small q, (example: $m = 11,\ n = 7$):



Subspace Coding vs. Coding Vectors

- Information rate loss from using coding vectors when $m = n$:

	$T \leq 2m$	$T > 2m$
$C - R_{\text{cv}}$	$o(1)$	$o(1) = (i^* - 1)(T - i^*)\frac{\log_2 q}{q} + O(q^{-1})$

- So in terms of transmission rate, "coding vector" scheme performs well enough if q is not small.
- KK'08 also made a similar observation by proposing an algebraic code construction for fixed dimensional subspace code. However, KK'08 only consider the subspace codes of block length one.

Sketch of the Proof

Proof Sketch

- The matrix channel $\mathrm{ch_m}$ with capacity $C_\mathrm{m} \triangleq C$ is given by:

$$P_{Y|X}(y|x) = \begin{cases} q^{-n\dim(\langle x\rangle)} & \langle y\rangle \sqsubseteq \langle x\rangle \\ 0 & \text{otherwise} \end{cases}$$

- The subspace channel $\mathrm{ch_s}$ with capacity C_s is defined as:

$$P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x) \triangleq \begin{cases} \psi(T,n,\pi_y)q^{-n\dim(\pi_x)} & \pi_y \sqsubseteq \pi_x \\ 0 & \text{otherwise} \end{cases}$$

- Lemma: The channels $\mathrm{ch_m}$ and $\mathrm{ch_s}$ are equivalent in terms of evaluating the mutual information between the input and output. As a result, $C_\mathrm{m} = C_\mathrm{s}$.

Proof Sketch

- Lemma: The input distribution that maximizes for $I(\Pi_X;\Pi_Y)$ is the one which is uniform over all subspaces having the same dimension. So

$$\mathbb{P}[\langle X \rangle = \pi_x] = \mathbb{P}[\Pi_X = \pi_x] = \alpha_r \times \begin{bmatrix} T \\ r \end{bmatrix}_q^{-1}$$

where $r = \dim(\pi_x)$ and $\alpha_r = \mathbb{P}[\dim(\Pi_X) = r]$

- Now, we have to maximize the mutual information $I(\Pi_X;\Pi_Y)$ over different choices of $\alpha_i, \quad i = 0, \ldots, \min(m, T)$.

Proof Sketch

- $I(\Pi_X;\Pi_Y)$ is a concave function of α_i, so we can apply Kuhn-Tucker theorem.
- The optimal values α_i^* should satisfy:

$$\begin{cases} \left.\frac{\partial I(\Pi_X;\Pi_Y)}{\partial \alpha_k}\right|_{\alpha_i^*} = \lambda \quad \forall k \ : \ \alpha_k^* > 0 \\ \left.\frac{\partial I(\Pi_X;\Pi_Y)}{\partial \alpha_k}\right|_{\alpha_i^*} \leq \lambda \quad \forall k \ : \ \alpha_k^* = 0 \end{cases}$$

for $\lambda = C_\mathrm{s} - \log_2 e$ where $\sum_{i=0}^{\min(m,T)} \alpha_i^* = 1.$

- After some manipulations and approximations we can write the Kuhn-Tucker conditions as a linear system:

$$\boldsymbol{A\alpha}^* \succeq 2^{-C_\mathrm{s}+o(1)}\boldsymbol{b}$$

Proof Sketch

- First case: $\delta \triangleq \min(m, T) \leq n$

$$\mathbf{A} = \begin{bmatrix} 1 & q^{-n} & \cdots & q^{-(\delta-1)n} & q^{-\delta n} \\ 0 & q^{-(n-1)} & \cdots & q^{-(\delta-1)(n-1)} & q^{-\delta(n-1)} \\ 0 & 0 & \cdots & q^{-(\delta-1)(n-2)} & q^{-\delta(n-2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q^{-(\delta-1)(n-\delta+1)} & q^{-\delta(n-\delta+1)} \\ 0 & 0 & \cdots & 0 & q^{-\delta(n-\delta)} \end{bmatrix}$$

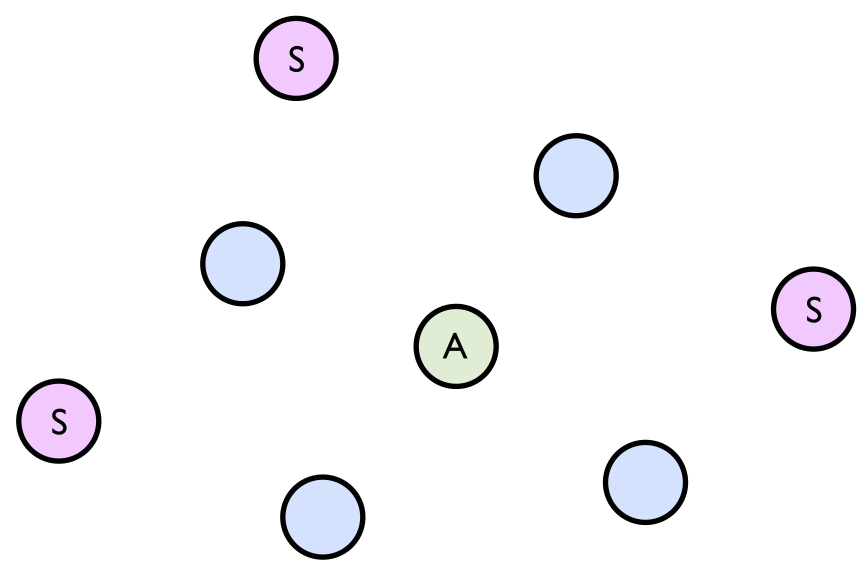
$$\mathbf{b} = \begin{bmatrix} 1 & q^{(T-n)} & \cdots & q^{\delta(T-n)} \end{bmatrix}^{\mathrm{T}}$$

$$\alpha_i^* = \begin{cases} q^{i(T-i)} 2^{-C_{\mathrm{s}}+o(1)} & : \quad \kappa \leq i \leq \delta \\ 0 & : \quad 0 \leq i < \kappa \end{cases}$$

Extension for Multiple Sources

Motivation

- Consider sensor network applications where multiple nodes want to report their data to one or multiple access points.



$$Y[t] = \sum_{i=1}^{s} H_i[t]X_i[t] = \left[\begin{array}{c|c|c} H_1[t] & \cdots & H_s[t] \end{array}\right] \begin{bmatrix} X_1[t] \\ \vdots \\ X_s[t] \end{bmatrix}$$

$$= H_{\mathrm{MAC}}[t]X_{\mathrm{MAC}}[t]$$

$$X_i \in \mathbb{F}_q^{m_i \times T}, \quad H_i \in \mathbb{F}_q^{n \times m_i}, \quad Y \in \mathbb{F}_q^{n \times T}$$

- We only consider the two sources problem. However, the same technique can be extended to more than two sources.
- We only characterize the asymptotic behavior of the rate region when q is large and $T \geq 2(m_1 + m_2)$
- The channel transition probability is given by:

$$P_{Y|X_1X_2}(y|x_1, x_2) = \begin{cases} q^{-n\dim(\langle x_1\rangle + \langle x_2\rangle)} & \langle y\rangle \sqsubseteq \langle x_1\rangle + \langle x_2\rangle \\ 0 & \text{otherwise} \end{cases}$$

- Again coding over subspaces is an optimal scheme.

Main Result

- Theorem:

For $T \geq 2(m_1 + m_2)$, the asymptotic (in the field size q) rate region of the MAC $\mathrm{ch}_{\mathrm{m-MAC}}$ is given by:

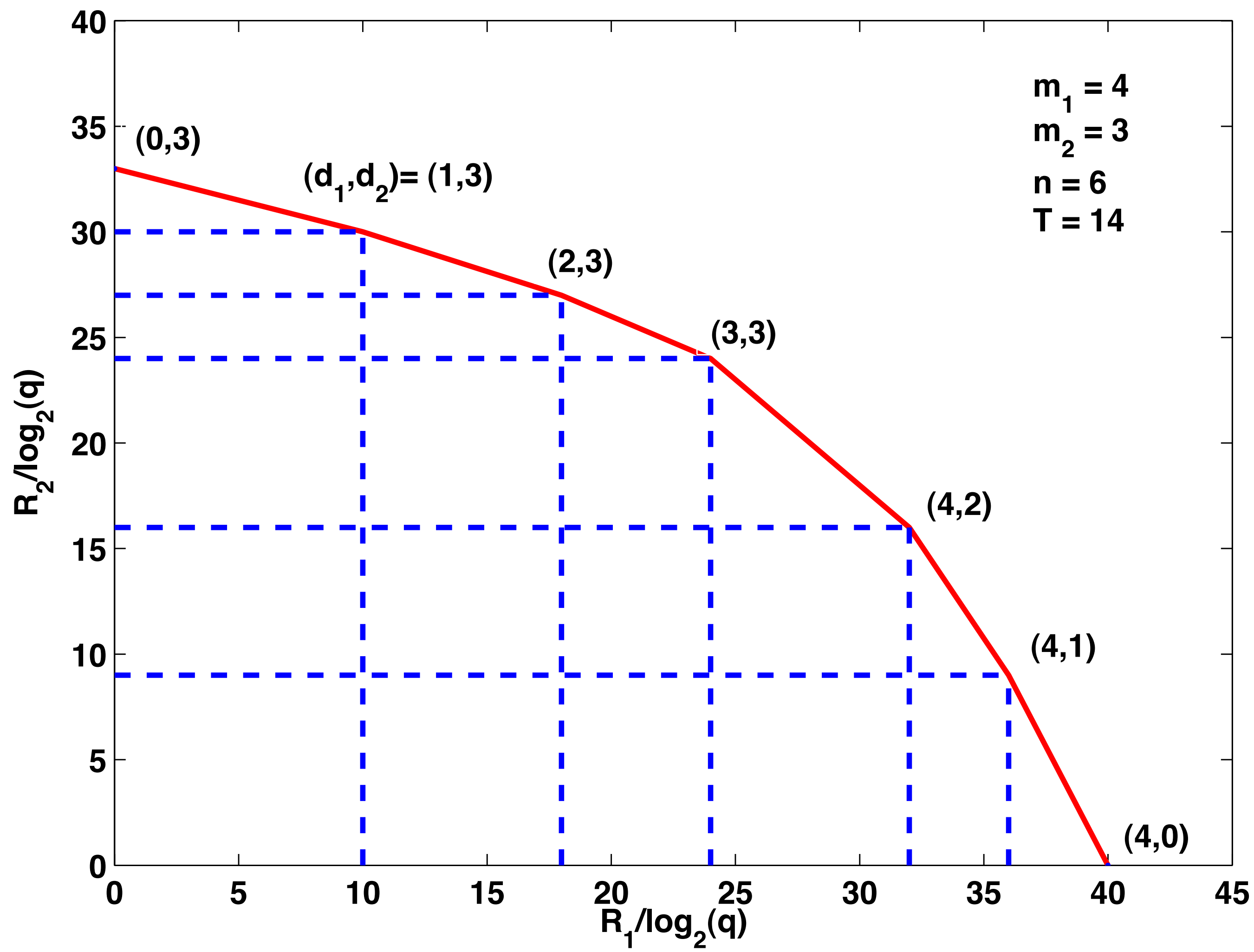
$$\mathcal{R}^* \triangleq \text{convex hull} \bigcup_{(d_1,d_2)\in\mathcal{D}^*} \mathcal{R}(d_1, d_2)$$

$$\mathcal{R}(d_1, d_2) \triangleq \{(R_1, R_2) : \; R_i \leq d_i(T - d_1 - d_2) \log_2 q, \; i = 1, 2\}$$

$$\mathcal{D}^* \triangleq \{(d_1, d_2) : \; 0 \leq d_i \leq \min[n, m_i], \; i = 1, 2, \\ 0 \leq d_1 + d_2 \leq \min[n, m_1 + m_2]\}$$

Illustration of the Result

- Example:



$$\mathcal{D}^* = \{(0,3),\ (1,3),\ (2,3),\ (3,3),\ (4,2),\ (4,1),\ (4,0)\}$$

- $(4,3) \notin \mathcal{D}^*$ because of the cooperative upper bound.

Sketch of the Proof

Achievability Scheme

- For given $(d_1, d_2) \in \mathcal{D}^*$, define the following subspace codebooks:

$$\widetilde{\mathcal{C}}_1 \triangleq \left\{ \langle X_1 \rangle : X_1 = \left[\begin{array}{c|c|c} \mathbf{I}_{d_1 \times d_1} & \mathbf{0}_{d_1 \times d_2} & \mathbf{U}_1 \\ \hline \mathbf{0}_{(m_1 - d_1) \times d_1} & \mathbf{0}_{(m_1 - d_1) \times d_2} & \mathbf{0}_{(m_1 - d_1) \times (T - d_1 - d_2)} \end{array} \right], \mathbf{U}_1 \in \mathbb{F}_q^{d_1 \times (T - d_1 - d_2)} \right\}$$

$$\widetilde{\mathcal{C}}_2 \triangleq \left\{ \langle X_2 \rangle : X_2 = \left[\begin{array}{c|c|c} \mathbf{0}_{d_2 \times d_1} & \mathbf{I}_{d_2 \times d_2} & \mathbf{U}_2 \\ \hline \mathbf{0}_{(m_2 - d_2) \times d_1} & \mathbf{0}_{(m_2 - d_2) \times d_2} & \mathbf{0}_{(m_2 - d_2) \times (T - d_1 - d_2)} \end{array} \right], \mathbf{U}_2 \in \mathbb{F}_q^{d_2 \times (T - d_1 - d_2)} \right\}$$

- The receiver receives:

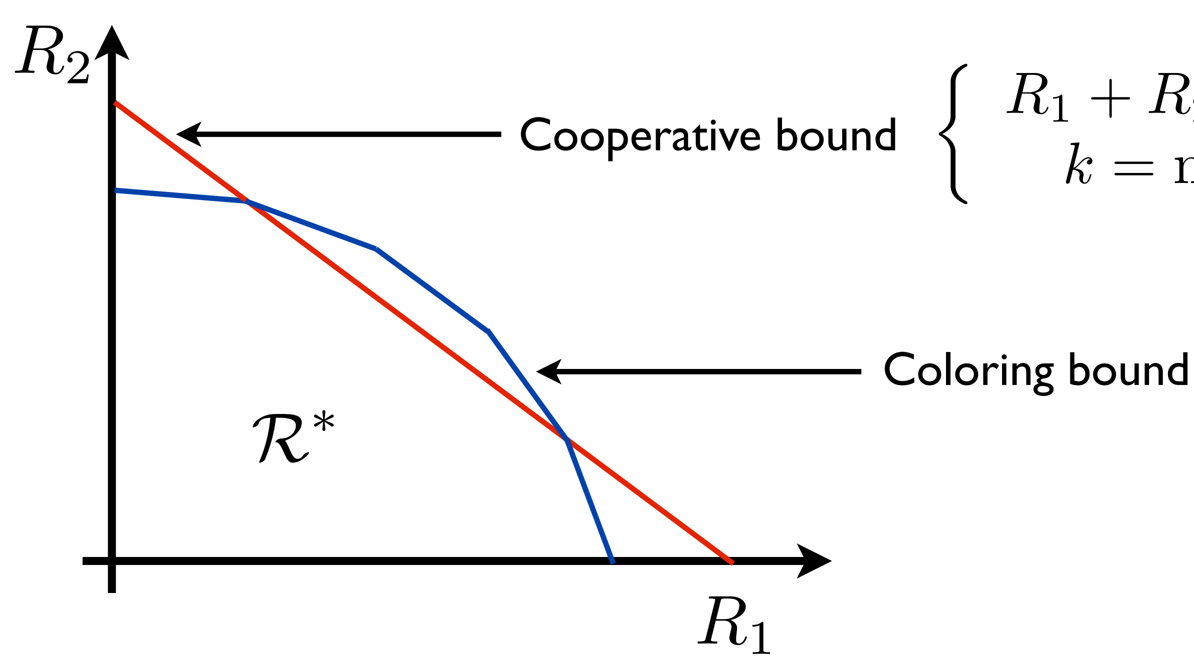
$$Y = H_1 X_1 + H_2 X_2 = \left[\begin{array}{c|c|c} \hat{H}_1 & \hat{H}_2 & \hat{H}_1 \mathbf{U}_1 + \hat{H}_2 \mathbf{U}_2 \end{array} \right]$$

- Since $d_1 + d_2 \leq n$, the matrix $[\hat{H}_1 \;\; \hat{H}_2]$ is full-rank with high probability, and therefore the decoder is able to decode $\mathbf{U}_1$ and $\mathbf{U}_2$.

- The remaining non-integer points in the rate region can be achieved using time-sharing.

Upper Bound

- Finding the upper bound goes along the following steps:
  - We use two different upper bounds:
    - A cooperative upper bound $\mathcal{R}_{\text{coop}}$
    - A combinatorial coloring upper bound $\mathcal{R}_{\text{col}}$
  - Find $\mathcal{R}_{\text{col}} \cap \mathcal{R}_{\text{coop}}$ and show that $\mathcal{R}_{\text{col}} \cap \mathcal{R}_{\text{coop}} \subseteq \mathcal{R}^*$

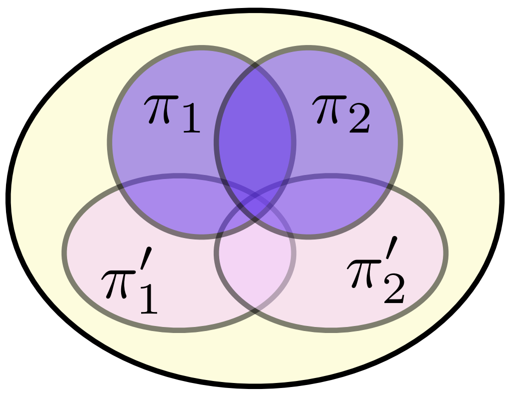


$$\begin{cases} R_1 + R_2 \leq k(T-k)\log_2 q \\ k = \min[m_1 + m_2, n] \end{cases}$$

Coloring Bound

- For channel transition probability we have:

$$P_{\Pi_Y|\Pi_{X_1}\Pi_{X_2}} = P_{\Pi_Y|\Pi_{X_1}+\Pi_{X_2}}$$



$$\pi_1 + \pi_2 = \pi'_1 + \pi'_2$$

- So, the receiver cannot distinguish between:

$$\pi_1 + \pi_2 \text{ and } \pi'_1 + \pi'_2$$

- What is the maximum number of distinguishable subspace sequences which can be conveyed through the channel?

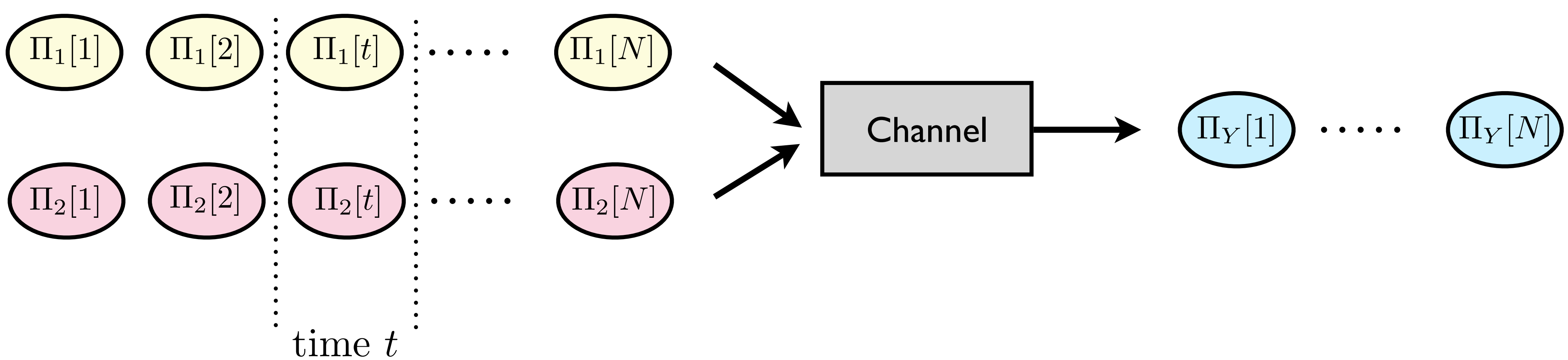
Coloring Bound

- From the proof of the outer bound for MAC we have:

$$R_1 \leq \frac{1}{N} I(\Pi_{X_1}^N ; \Pi_Y^N | \Pi_{X_2}^N) \leq \frac{1}{N} \sum_{t=1}^{N} I(\Pi_{X_1 t} ; \Pi_{Y t} | \Pi_{X_2 t})$$

$$R_2 \leq \frac{1}{N} I(\Pi_{X_2}^N ; \Pi_Y^N | \Pi_{X_1}^N) \leq \frac{1}{N} \sum_{t=1}^{N} I(\Pi_{X_2 t} ; \Pi_{Y t} | \Pi_{X_1 t})$$

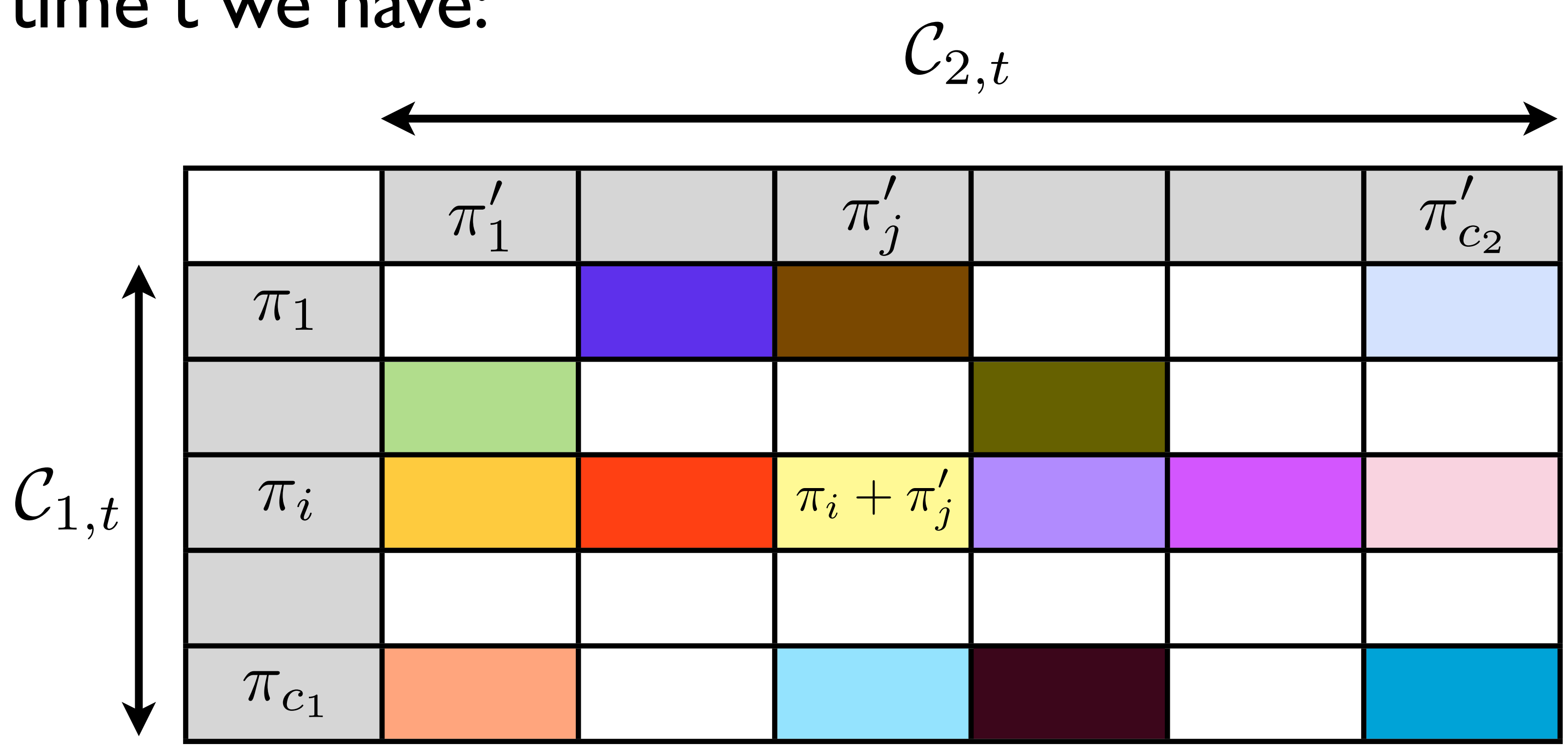
$$R_1 + R_2 \leq \frac{1}{N} I(\Pi_{X_1}^N , \Pi_{X_2}^N ; \Pi_Y^N) \leq \frac{1}{N} \sum_{t=1}^{N} I(\Pi_{X_1 t} , \Pi_{X_2 t} ; \Pi_{Y t})$$



Coloring Bound

- $\mathcal{C}_{i,t}$ denotes the projection of the codebook of user i to its t'th element.

- At time t we have:



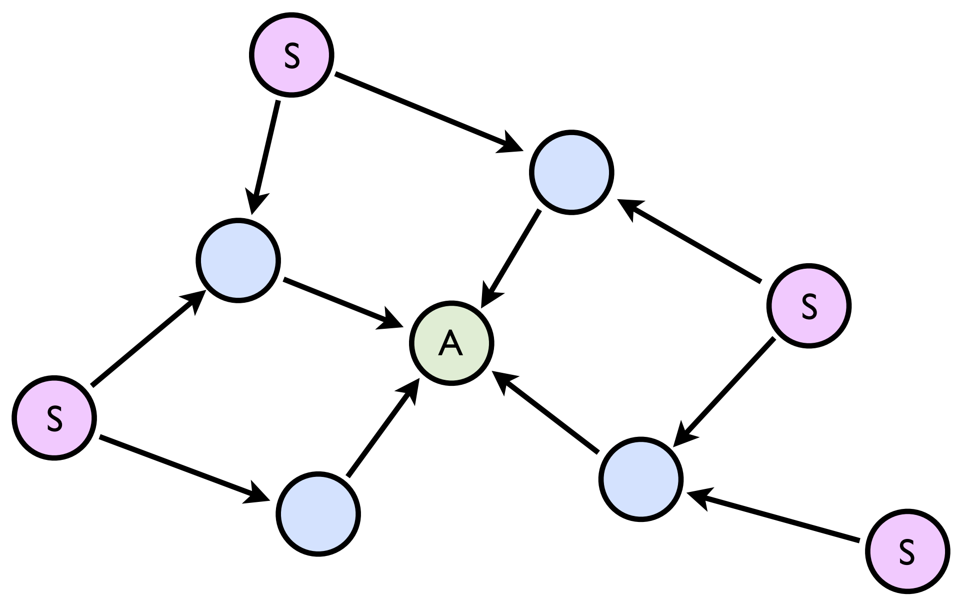
- Theorem: There exists integer numbers $0 \leq \delta_i(t) \leq m_i$ such that

$$c_{i,t} = |C_{i,t}| \dot{\leq} q^{\delta_i(t)[T-\delta_1(t)-\delta_2(t)]}$$

Compressed Network Coding Vectors

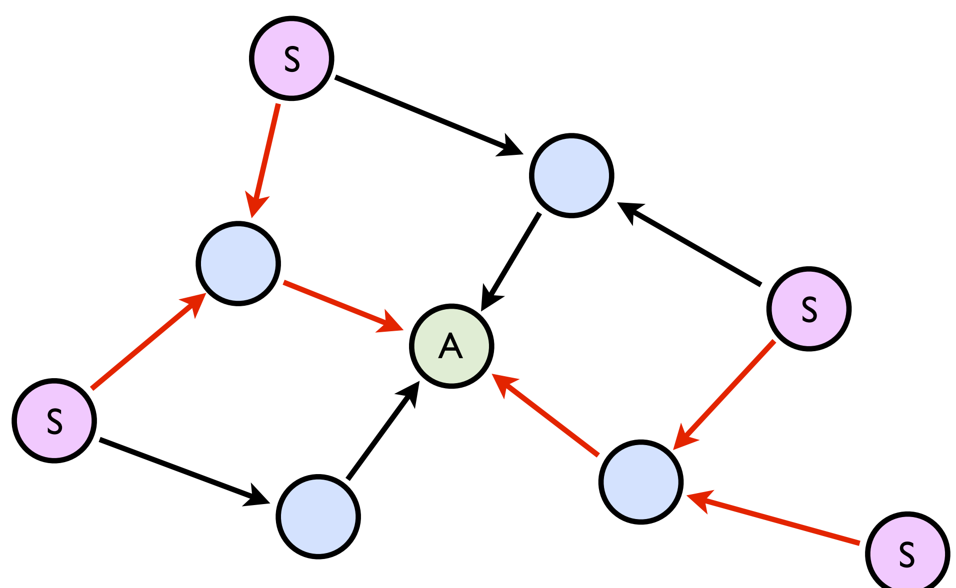
Motivation

- Motivation: Combining network coding with data collecting protocols in sensor networks where N sources send information to an access point.



Motivation

- In the previous approaches: an underlying assumption is that, all sources packets may get combined in the network.
- Compressed coding vectors: assume that each coded packets contains a linear combination of at most M out the N source packets.
  - => This allows us to use coding vectors whose length grows sub-linearly with N.
    - => more efficient network communication.



Compressed Coding Vectors

- The sources packets are of the form: $[\boldsymbol{e}_i \mid \boldsymbol{x}_i]$
- A packet in the network is represented as: $\boldsymbol{p} \triangleq [\boldsymbol{p}^C \mid \boldsymbol{p}^I]$
- Consider a linear code $\mathcal{C} = [N, N-r, d]_q$ with parity check matrix $\mathbf{H}_\mathcal{C}$ where $d = \min(2M+1, N+1)$
- As coding vector, assign to source packet $\boldsymbol{x}_i$ the ith column of the matrix $\mathbf{H}_\mathcal{C}$: $\boldsymbol{h}_i = \boldsymbol{e}_i \cdot \mathbf{H}_\mathcal{C}^T$
- => compressed coding vectors:

$$\hat{\boldsymbol{p}}^C = \boldsymbol{p}^C \cdot \mathbf{H}_\mathcal{C}^T$$

- Because $\mathrm{wt}(\boldsymbol{p}^C) \le M$ so if $\boldsymbol{p}_1^C \neq \boldsymbol{p}_2^C$ then $\hat{\boldsymbol{p}}_1^C \neq \hat{\boldsymbol{p}}_2^C$
- For each packet, recovering $\boldsymbol{p}^C$ from $\hat{\boldsymbol{p}}^C$ reduces to a decoding problem.

Bounds on the Length of CCV

- From the Gilbert-Varshamov bound we have an upper bound for the length of compressed coding vectors:

$$r \leq NH_q\left(\frac{2M}{N}\right)$$

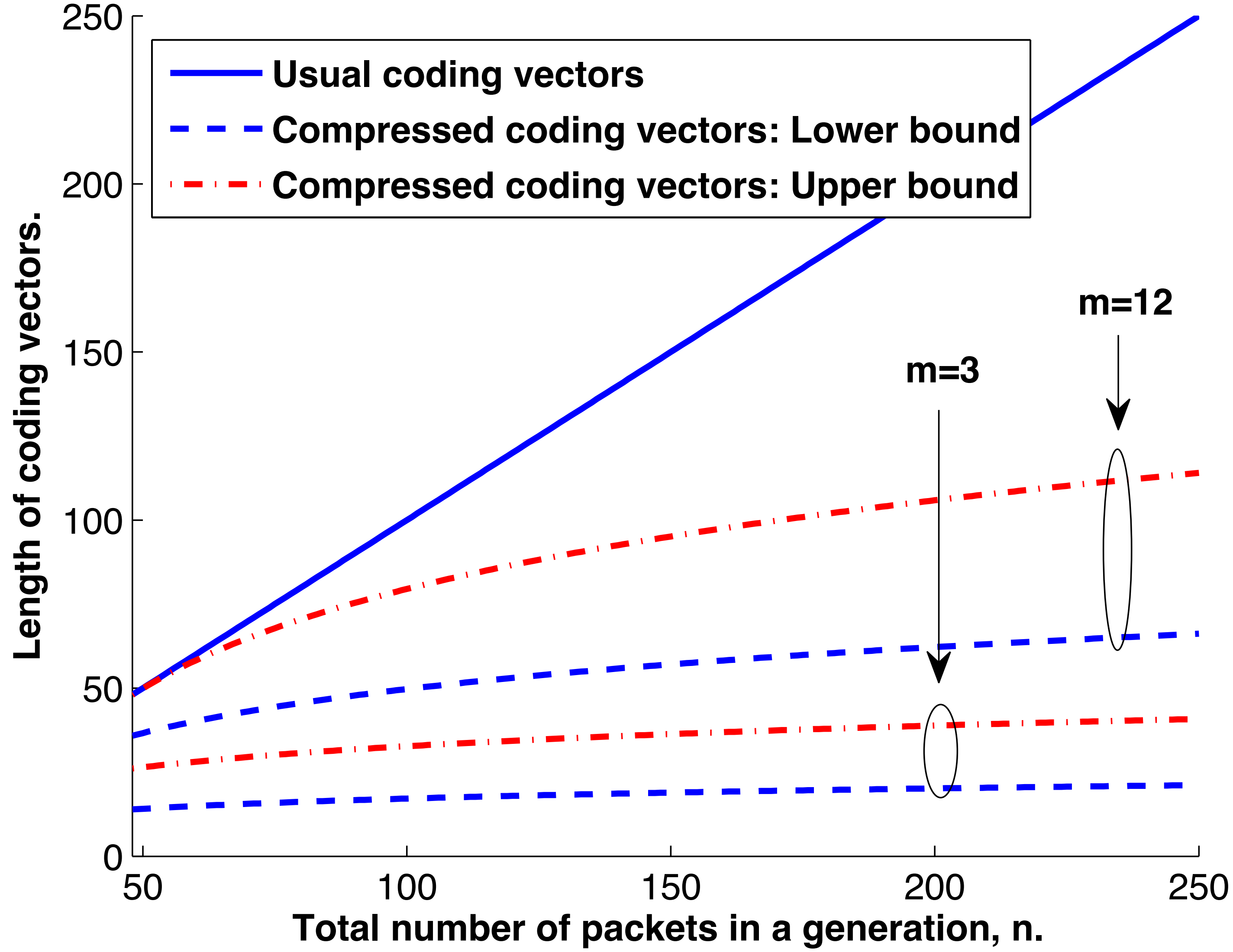
- From the Sphere packing bound we have a lower bound on the length of compressed coding vectors:

$$r \geq NH_q\left(\frac{M}{N}\right) - \frac{1}{2}\log_q\left(8M\left(1 - \frac{M}{N}\right)\right)$$

- For fixed M and growing N we have:

$$M\log_q N + O(1) \leq r \leq 2M\log_q N + O(1)$$

Bounds on the Length of CCV



Conclusions

- We proposed a matrix channel model for non-coherent randomized network coding and characterized its capacity.
- Using coding vectors is not far from optimal scheme if the field size is large.
- Motivated by sensor network application, we also looked at the multi-source non-coherent network coding problem and characterize the asymptotic (in filed size) rate region.
- In terms of rate improvement, subspace coding does not offer a significant difference.

Thank you!