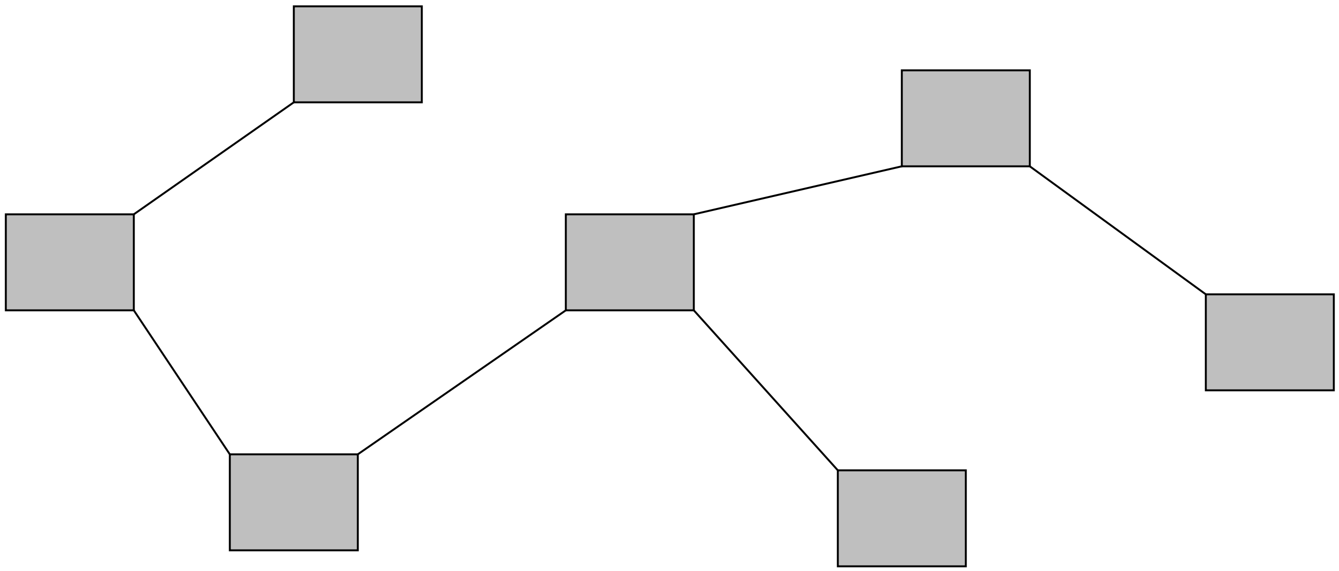
Distributed Functional Compression through Graph Coloring

Vishal Doshi, Devavrat Shah, Muriel Medard

Massachusetts Institute of Technology

Distributed Functional Compression



- Nodes (sensors, cameras) in a network have pieces of information
  - Different nodes have different use of the information in network
  - i.e. they wish to compute different functions of the information
- Question: what is the minimal transmission rates between nodes
  - So as to allow different nodes to compute their functions
  - And, nodes have to do this in a distributed manner

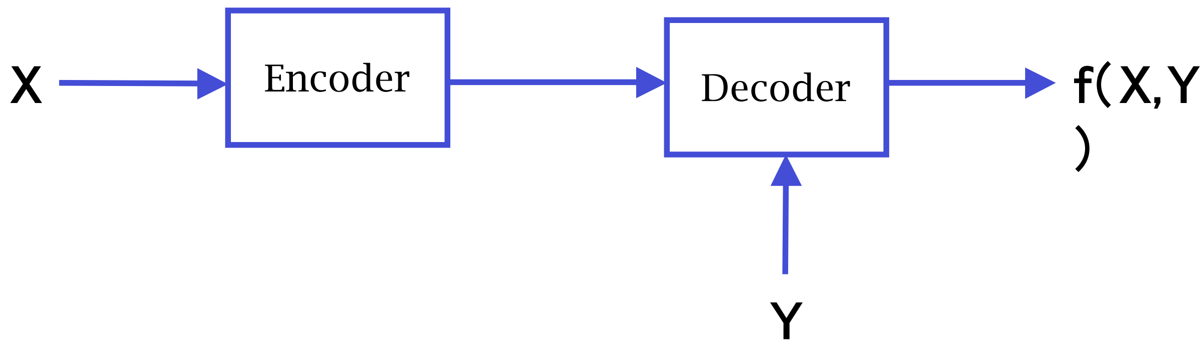
Applications

- Distributed function compression
  - Marriage of information theory, networks and computation
  - Grand challenge for modern information theory

- Algorithmic applications
  - Surveillance cameras gather a lot information
  - But only *small* amount of information is useful
  - Distributed compression can lead to an efficient network

- Policy guidelines
  - Privacy preservation of sensitive information, e.g. Census DB
  - Information theoretic results provide bounds on “information transmission” so as to preserve privacy of people
  - Or, Privacy preservation of people while using cameras for surveillance in public places

First Step : Functional Side Information



- At what rate must **X** be encoded such that the computation of f(**X**,**Y**) is possible at the receiver?
  - Here, f can be any deterministic function
- More importantly,
  - Can we come up with “implementable” coding schemes to achieve (close to) the optimal rate?

Past Work

- The minimal rate was characterized by Orlitsky-Roche (2001):

$$H_G(X|Y) = \min_{\substack{W-X-Y \\ X \in W \in \Gamma(G)}} I(W;X|Y)$$

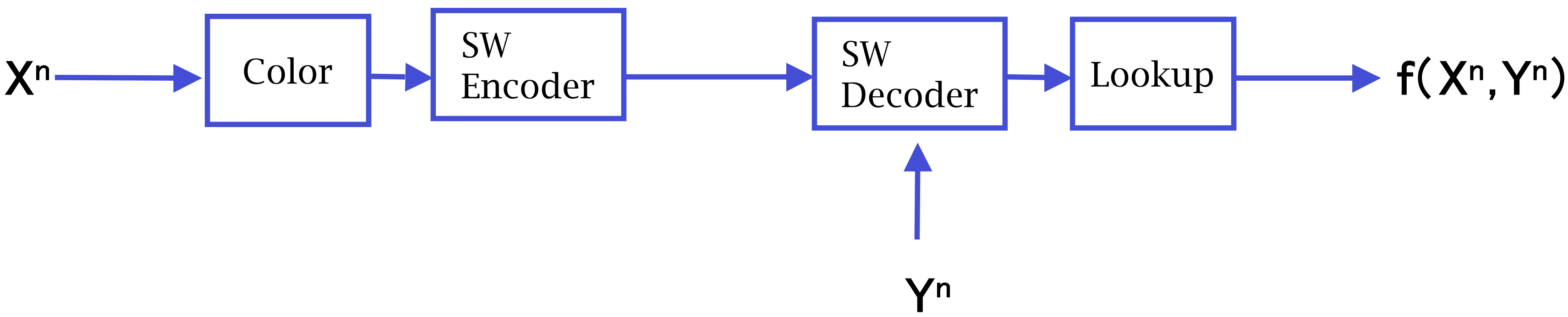
  where G is a graph defined by the function f(X,Y) and the distribution p(X,Y) called the characteristic graph.

- $H_G(X|Y) < H(X|Y)$ and the disparity is determined by the complexity of the function
  - Thus, functional compression definitely provides gains

- However, the above scheme is somewhat complicated
  - We provide coding scheme using notion of graph coloring
  - It can achieve the optimal rate
  - Leads to design of good simple-to-implement heuristics

Our Coding Scheme

- The following scheme achieves the optimal rate:

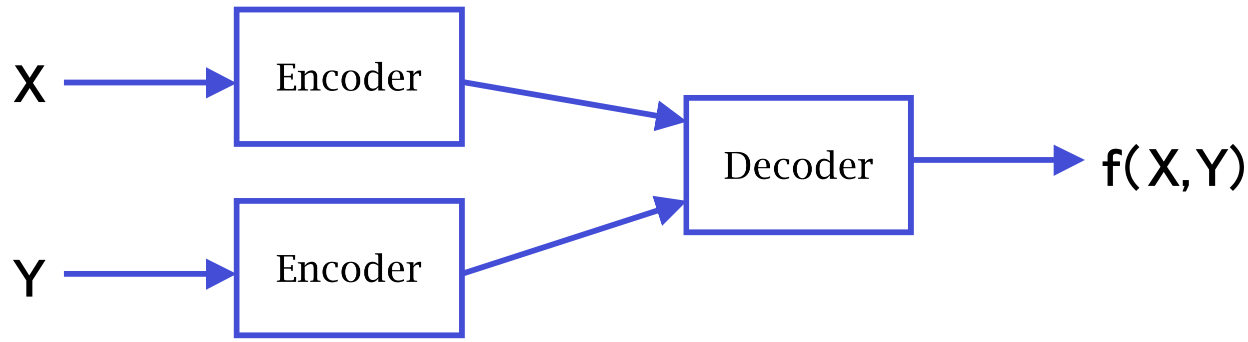


1. Construct $\mathbf{G^n}$ and color it (gains even without min entropy colorings)
2. Color each source symbol
3. Do Slepian-Wolf style coding to recover $\mathbf{c(X^n)}$ and $\mathbf{Y^n}$ at the receiver
4. With high probability, $\mathbf{c(X^n)}$ and $\mathbf{Y^n}$ uniquely determine $f(X^n,Y^n)$

¬ Our scheme is built on existing distributed source coding

  ¬ Provides separation between “functional coding” and “data

Step II: Distributed Functional Compression



- Next we consider the distributed source coding set up

- Again, we wish to determine minimal transmit rates for **X** and **Y** such that f(**X**,**Y**) can be recovered at the receiver?
  - The sources cannot communicate with each other

- Can we extend our result for Step I to this more general problem?
  - Answer: Yes, but with some conditions

Distributed Functional Compression

- Condition: the joint distribution of sources is such that
  - $p(\mathbf{x_1},\mathbf{y_1})>0$, $p(\mathbf{x_2},\mathbf{y_2})>0$ implies $p(\mathbf{x_1},\mathbf{y_2})>0$
  - This is a reasonable assumption for a wide array of applications
- Under this condition, our coding scheme is optimal
- The rate region is the closure of the set of all $(\mathbf{R_X},\mathbf{R_Y})$ such that for some $\mathbf{n}$ and colorings $\mathbf{c_X}$ and $\mathbf{c_Y}$ of $\mathbf{G^n}$ and $\mathbf{H^n}$, where $\mathbf{H^n}$ is the characteristic graph of $\mathbf{Y}$ with respect to $\mathbf{X}$, the following holds:

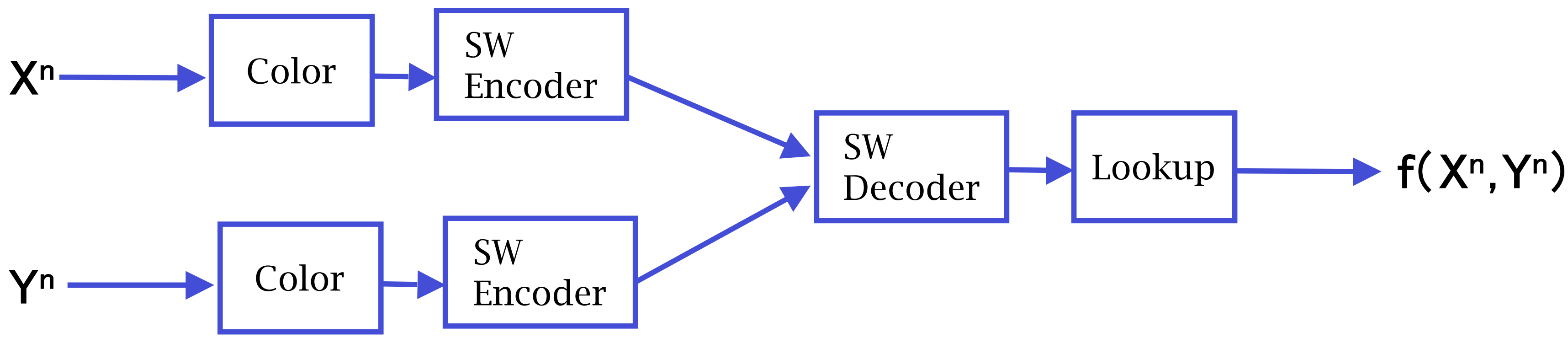
$$\mathbf{nR_X} \geq H(\mathbf{c_X}(\mathbf{X})|\mathbf{c_Y}(\mathbf{Y}))$$

$$\mathbf{nR_Y} \geq H(\mathbf{c_Y}(\mathbf{Y})|\mathbf{c_X}(\mathbf{X}))$$

$$\mathbf{n(R_X+R_Y)} \geq H(\mathbf{c_X}(\mathbf{X}),\mathbf{c_Y}(\mathbf{Y}))$$

Our Coding Scheme

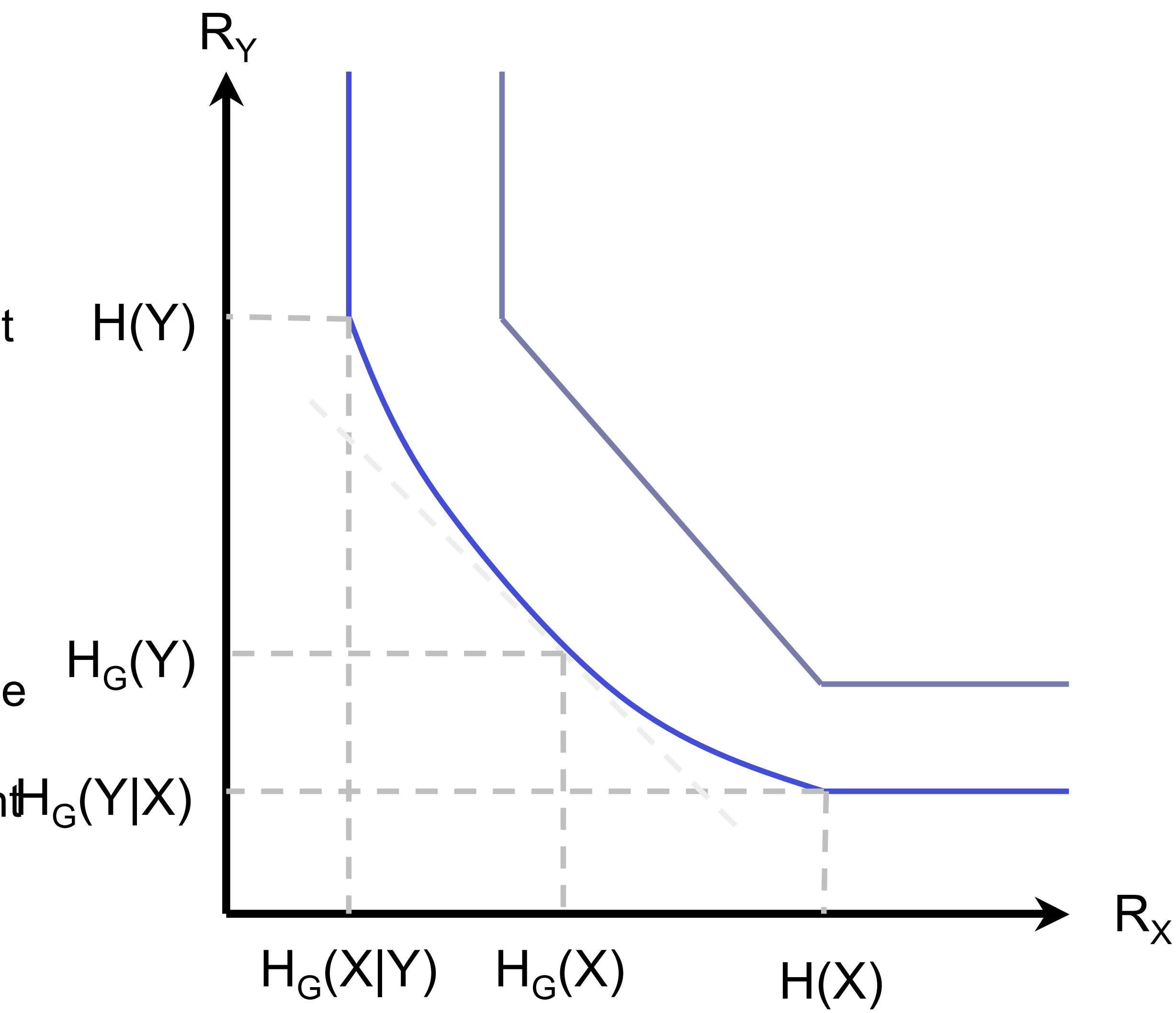
- The following scheme achieves the optimal rate:



1. Construct graphs $\mathbf{G}^n$ and $\mathbf{H}^n$ and color them
2. Color each source symbol
3. Do Slepian-Wolf style coding to recover $\mathbf{c}_X(\mathbf{X}^n)$ and $\mathbf{c}_Y(\mathbf{Y}^n)$ at receiver
4. With high probability, $\mathbf{c}_X(\mathbf{X}^n)$ and $\mathbf{c}_Y(\mathbf{Y}^n)$ uniquely determine $f(X^n,Y^n)$

Rate Region for Distributed Coding

- The inner region is the classical Slepian-Wolf rate region
- We can achieve rates out to the blue curve using the function information.
- We have single letter characterizations for three points on the region including the minimal joint rate point.



Summary

- We are developing novel coding schemes that utilize graph coloring as a means to achieve optimal rates
- Our results extends to multiple sources and one receiver
- Our approach leads to layered architecture
  - Add “functional layer” on existing distributed compression schemes
- There is a lot of literature on Approximation algorithms for graph coloring
  - This allows for heuristics based on these algorithms
- *And, it furthers the current project in the correct direction !*