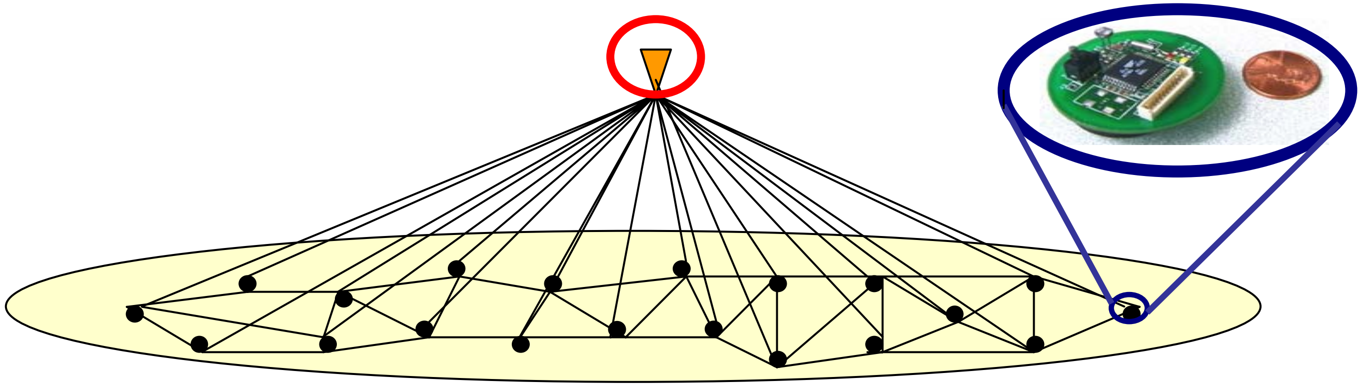
Network Information Flow with Correlated Data

João Barros

University of Porto

Joint work with Sergio D. Servetto, Cornell University

Motivation: Sensor Networks

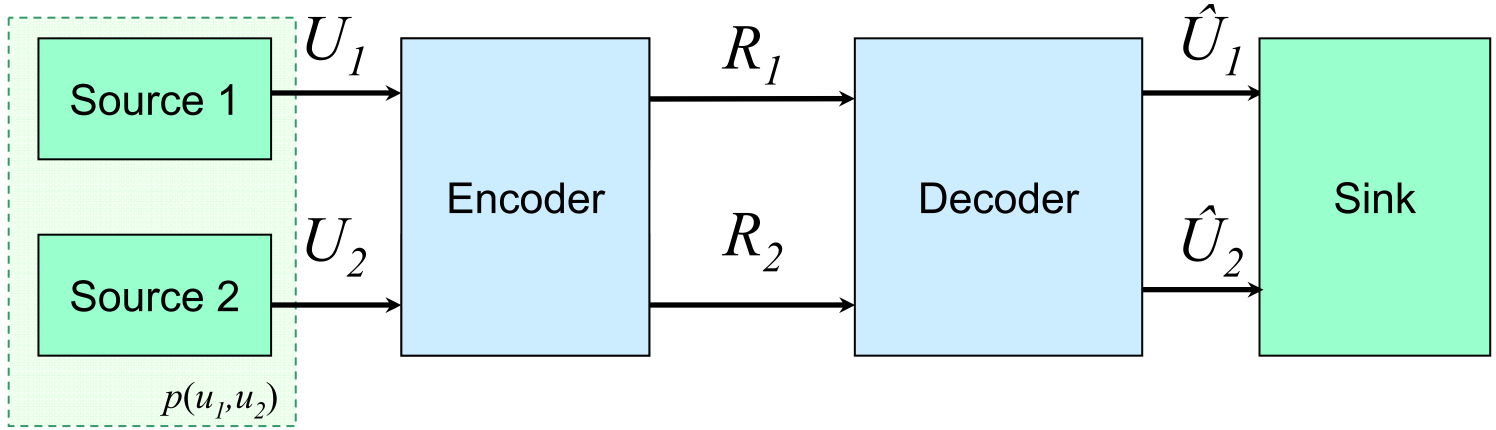


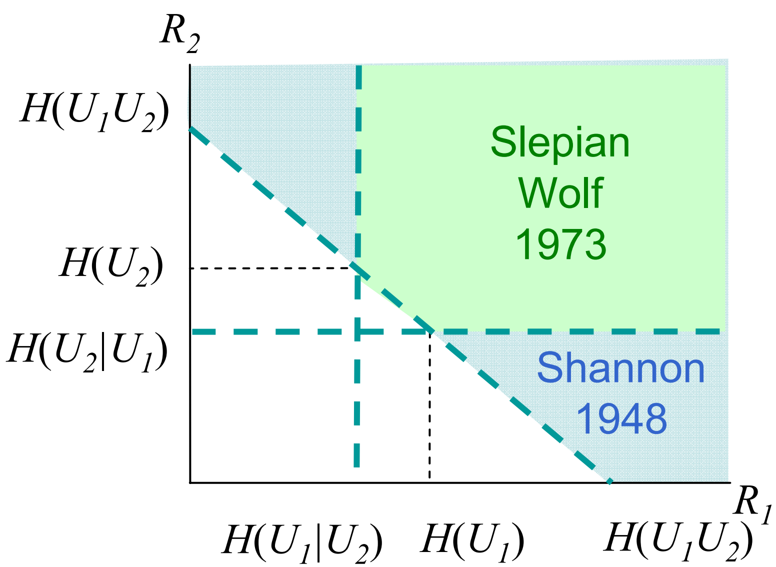
- Data Collection
- Self-organization (adhoc network)
- Data Transmission
- Data Fusion and Post-Processing
- Decision

Research Challenges

- Theoretical limits of data communication in large-scale sensor networks.
- Efficient source and channel codes, modulators, detectors, estimation and data fusion algorithms.
- Networking aspects (topology, routing, flow control, communication protocols)
- Hardware (integrated circuits, low power, wireless radio/optical)
- Economics (applications, products, costs, markets).

Encoding Correlated Sources



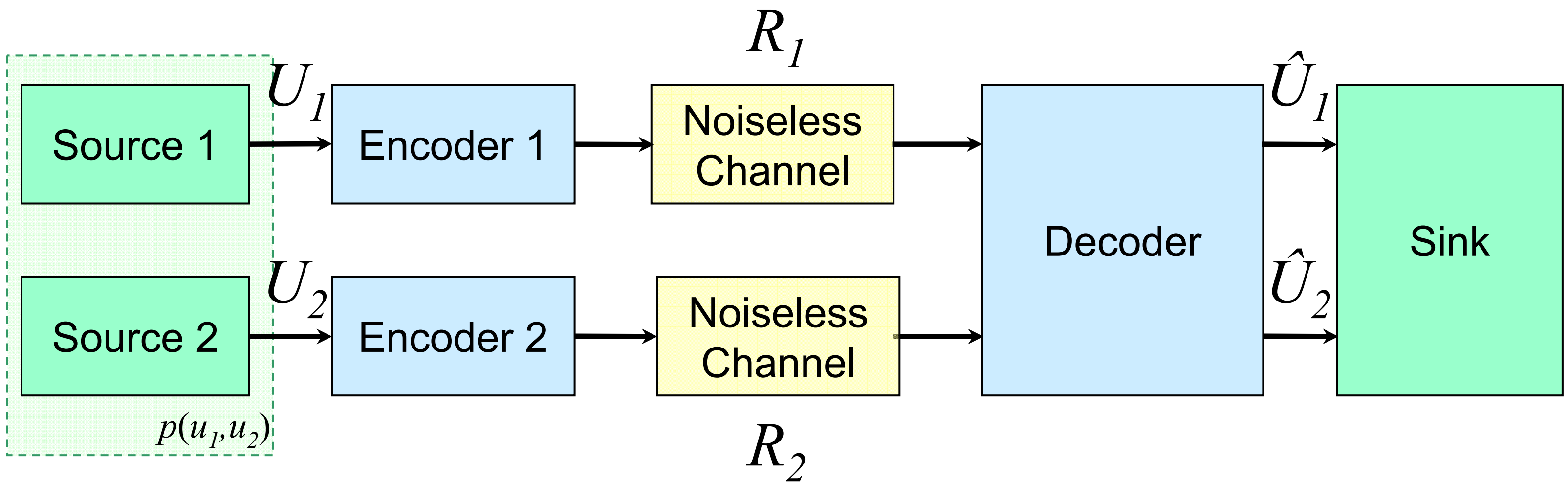


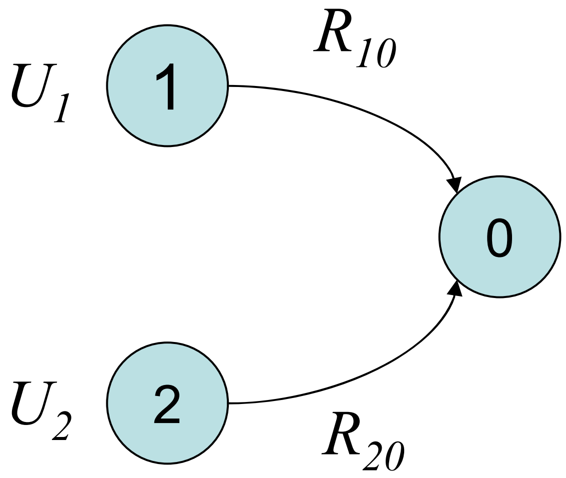
$$R_1 > H(U_1|U_2)$$

$$R_2 > H(U_2|U_1)$$

$$R_1+R_2 > H(U_1U_2)$$

A network flow interpretation...





A three-node graph with nodes 1 and 2 as the sources and node 0 as the sink.

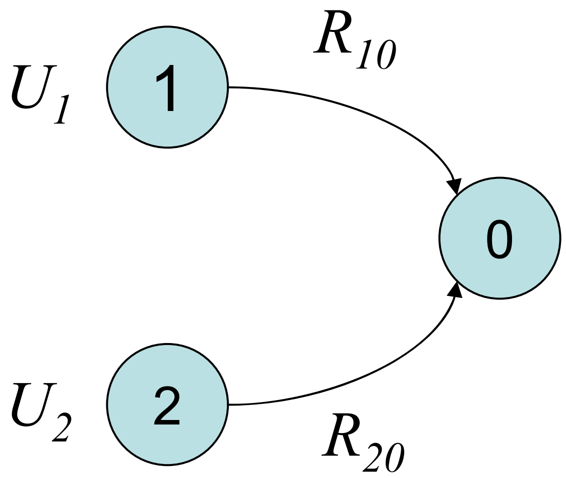
Network flow

Let *(G,V)* be a directed graph in which each edge E_{ij} has capacity *c(i,j)* ≥0. Let *s* be the source and *t* be the destination.

A flow in *G* is a real-valued function satisfying the following constraints: $f : V \times V \longrightarrow \Re$

- **Capacity constraint:** For all *i* and *j* in *V*, $f(i,j) \leq c_{ij}$
- **Skew symmetry:** For all *i* and *j* in *V*, $f(i,j) = -f(j,i)$
- **Flow conservation:** For all *i* in *V* \ {s,t} $\sum_{i \in V} f(i,j) = 0$

Back to the Slepian-Wolf Problem



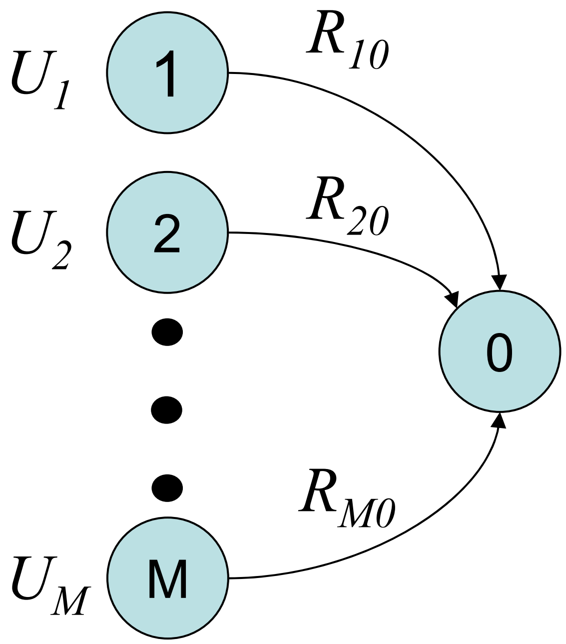
$$R_{10} > H(U_1|U_2)$$

$$R_{20} > H(U_2|U_1)$$

$$R_{10}+R_{20} > H(U_1U_2)$$

The Slepian-Wolf Theorem gives necessary and sufficient conditions for feasible flows that guarantee perfect reconstruction at node 0.

Many correlated sources



Perfect reconstruction is possible if and only if

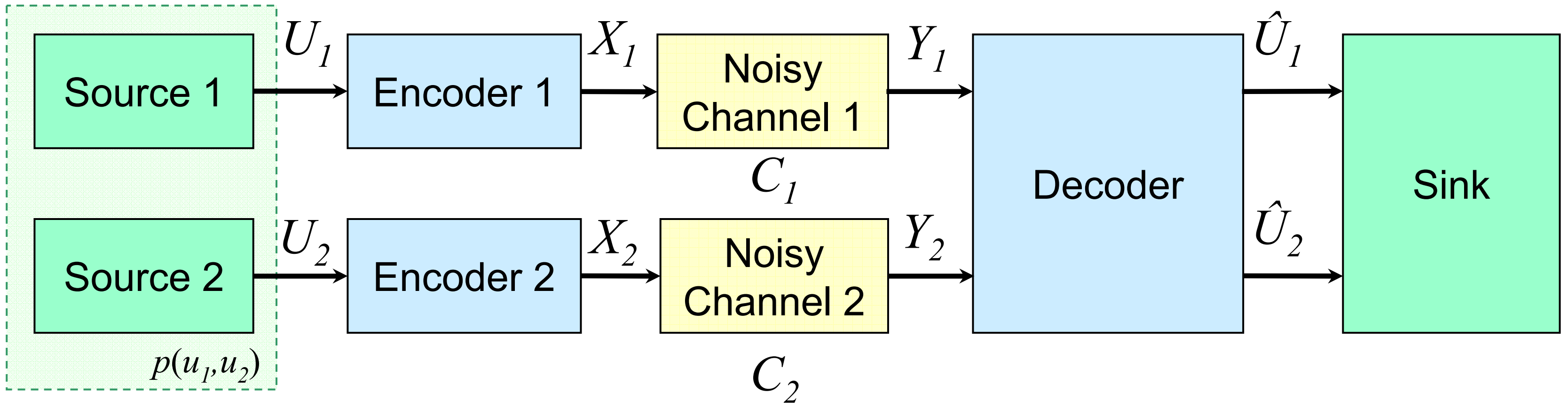
$$\sum_{i \in S} R_{i0} > H(U(S) \mid U(S^c))$$

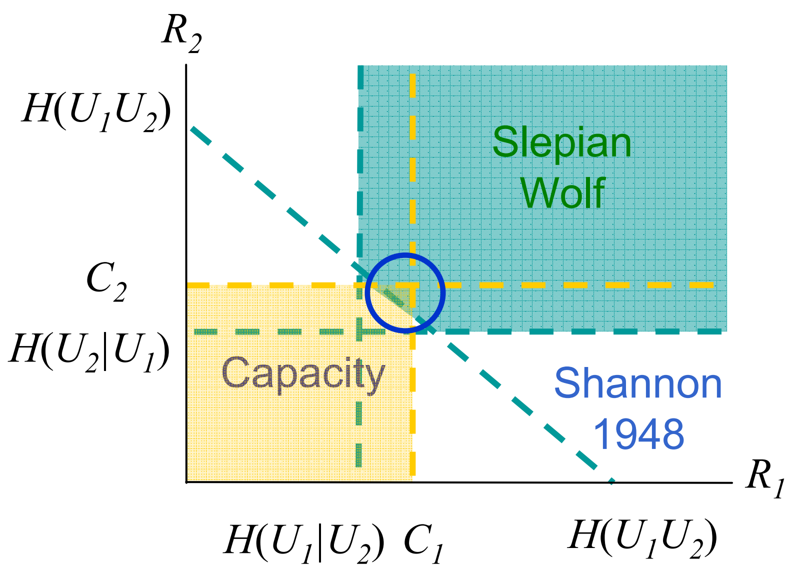
for all sets $S \subset \{1,2,\ldots.,M\}$,

$S \cap S^c = 0$,

$S \neq 0$

Theorem 1





Barros, Servetto 2002:

- Perfect reconstruction is possible if and only if

$$H(U_1|U_2) < C_1$$

$$H(U_2|U_1) < C_2$$

$$H(U_1U_2) < C_1+C_2$$

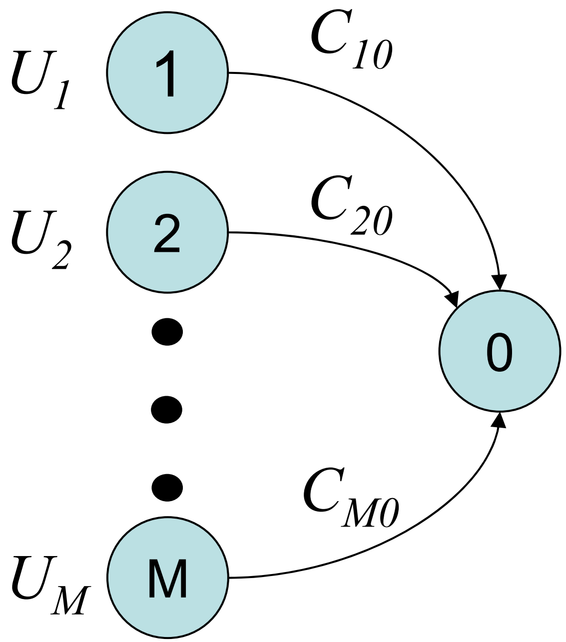
Sketch of Proof

- Begin with the converse and show that there do not exist codes outside the region of intersection.
- Exploit the properties of the following Markov chain:

$$Y_1^N - X_1^N - U_1^N - U_2^N - X_2^N - Y_2^N$$

- Use Fano's inequality to prove that the conditions of the theorem are necessary for arbitrarily small probability of error.
- Show that all points inside the intersection can be achieved using a cascade of Slepian Wolf codes and optimal point-to-point channel codes.

Multiple Sources and Channels



Perfect reconstruction is possible if and only if

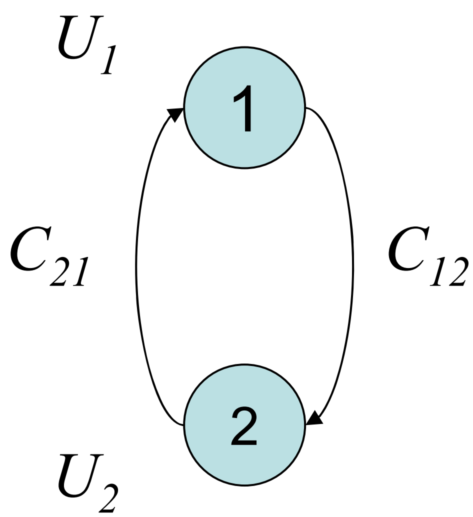
$$H(U(S)|U(S^c)) < \sum_{i \in S} C_{i0}$$

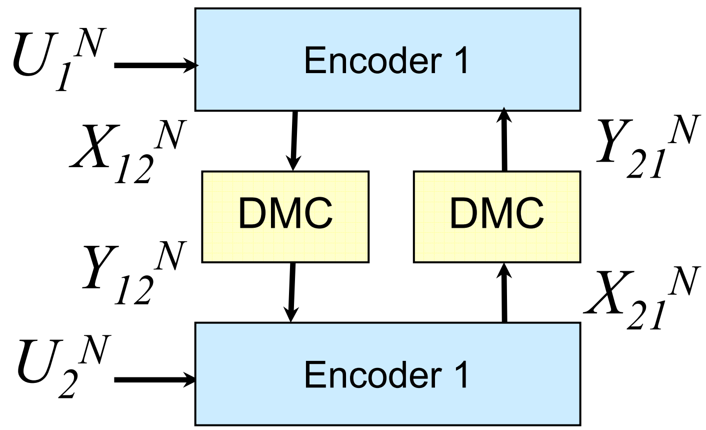
for all sets $S \subset \{1,2,\ldots.,M\}$,

$S \cap S^c = 0$,

$S \neq 0$

Cooperating nodes



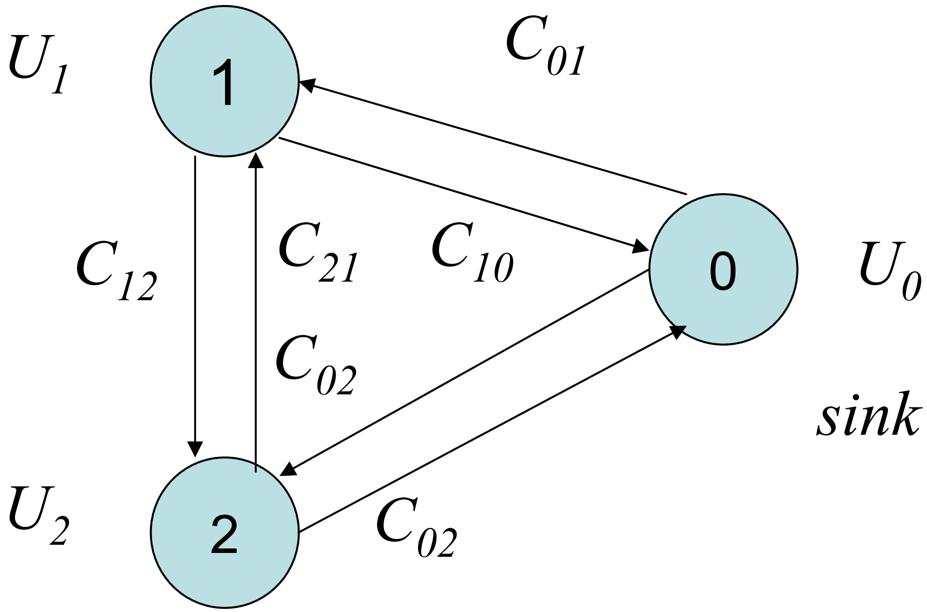


- Cooperation between encoders occurs in K rounds and is specified by 2K functions:

$$h_{1k} : \mathrm{U}_1^{\mathrm{N}} \times \mathrm{Y}_{21}^{\mathrm{N}}(1) \times \ldots \times \mathrm{Y}_{21}^{\mathrm{N}}(k-1) \longrightarrow \mathrm{X}_{12}^{\mathrm{N}}(k)$$

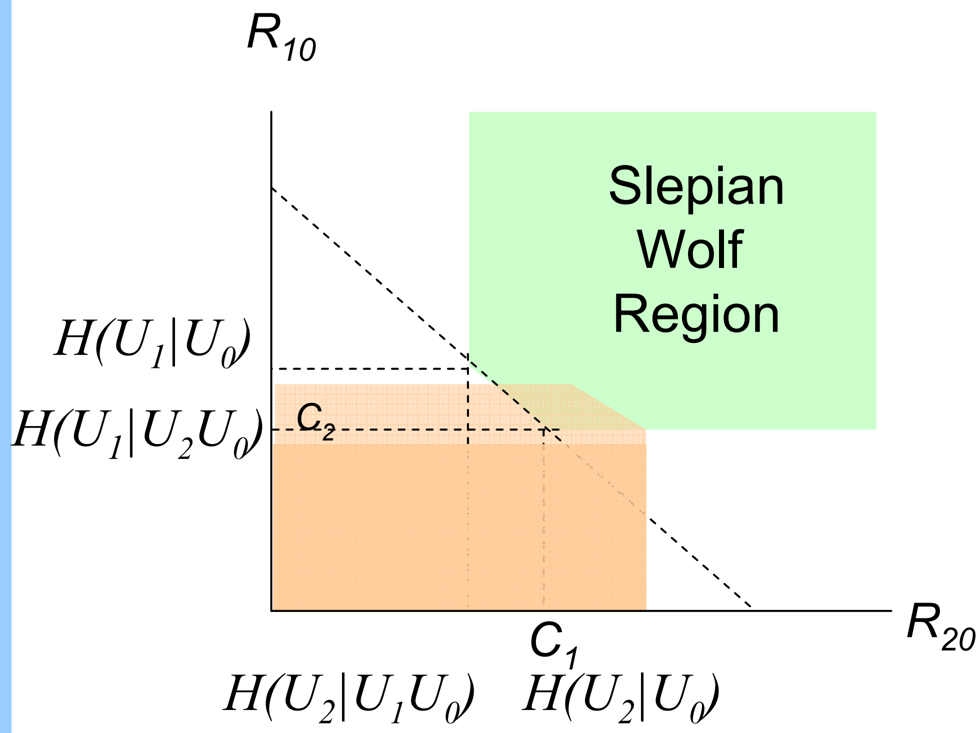
$$h_{2k} : \mathrm{U}_2^{\mathrm{N}} \times \mathrm{Y}_{12}^{\mathrm{N}}(1) \times \ldots \times \mathrm{Y}_{12}^{\mathrm{N}}(k-1) \longrightarrow \mathrm{X}_{2\mathrm{k}}^{\mathrm{N}}(k)$$

Slepian-Wolf Networks



- Each node observes one of the correlated sources.
- After K rounds of communication node 0 must produce a perfect reconstruction of all sources.

Capacity Region



Theorem 3

$$H(U_1|U_2U_0) < C_{10} + C_{12}$$

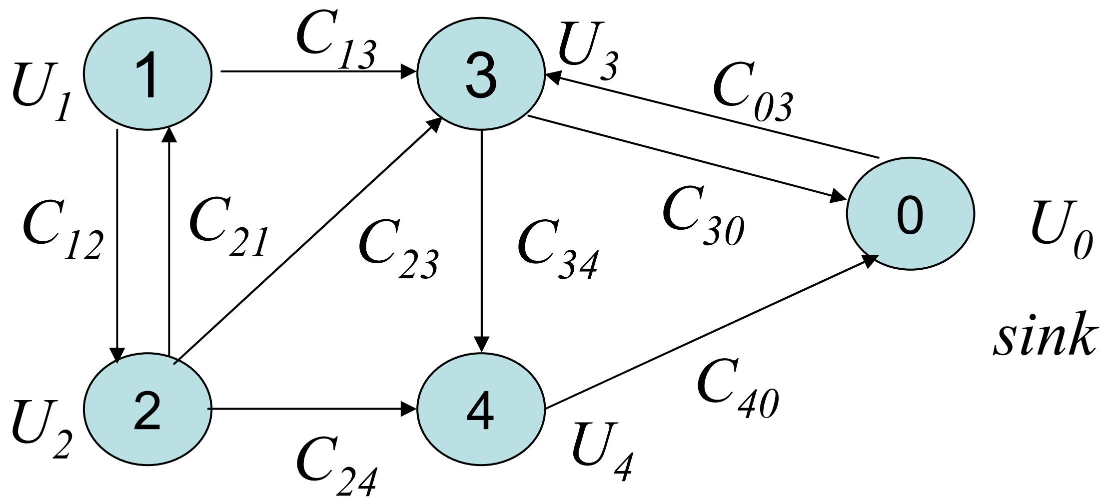
$$H(U_2|U_1U_0) < C_{20} + C_{21}$$

$$H(U_1U_2|U_0) < C_{10} + C_{20}$$

(converse proof !)

- Cooperation increases the capacity region of the channels.

Networks of Independent Channels

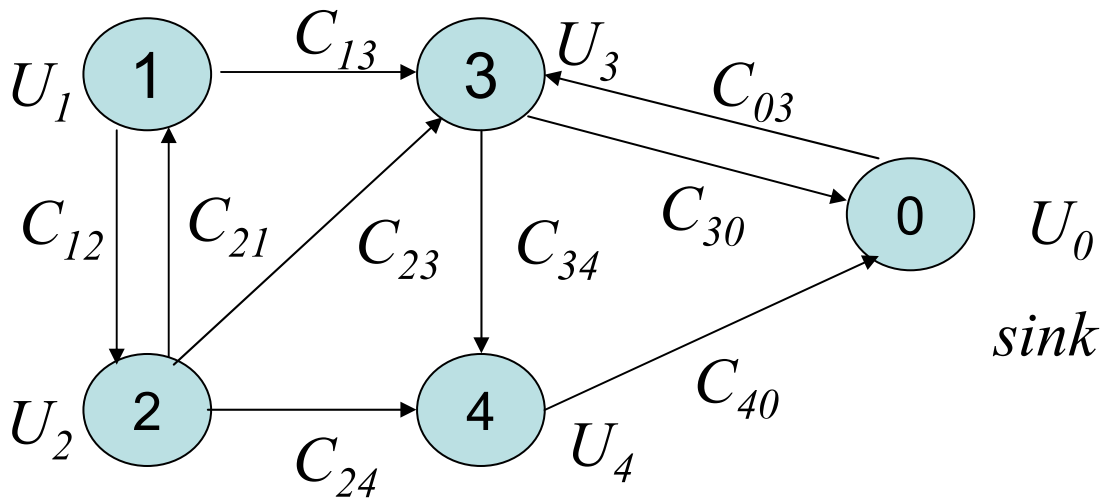


- The network is described by a directed graph.
- Again, after K rounds of communication node 0 must produce a perfect reconstruction of all sources.
- In each round the sent codewords depend on all previously received channel outputs.

Coding Strategy

- Use capacity-achieving channel codes to turn the noisy network into a noiseless network.
- Use network source codes for (1) distributed compression and (2) data delivering to the destination.

Network Source Codes



- Use classical Slepian-Wolf codes at some operating point $(R_1, R_2 \ldots, R_M)$.
- View this as a flow network and consider a flow *f* with *M* sources and demands $(R_1, R_2 \ldots, R_M)$ *at node 0.*
- If *f* exists, then *f* determines the number of bits that each node must send to its neighbours.

Network Source Codes

- Perform random binning to send the required information to each neighbour at the rates specified by *f.*
- The decoder collects all the bin indices at node 0 and applies standard Slepian-Wolf decoding.

Final Result

- Slepian Wolf Theorem: $$\sum_{i \in S} R_i > H(U(S) \mid U(S^c))$$

- Elementary flow concepts: a flow is feasible if $$\sum_{i \in S} R_i < \sum_{\substack{i \in S \\ j \in S^c}} C_{ij}$$

  i.e. the total amount of flow injected on one side of the cut has to be lower than the capacity of the links carrying that quantity of flow to the other side.

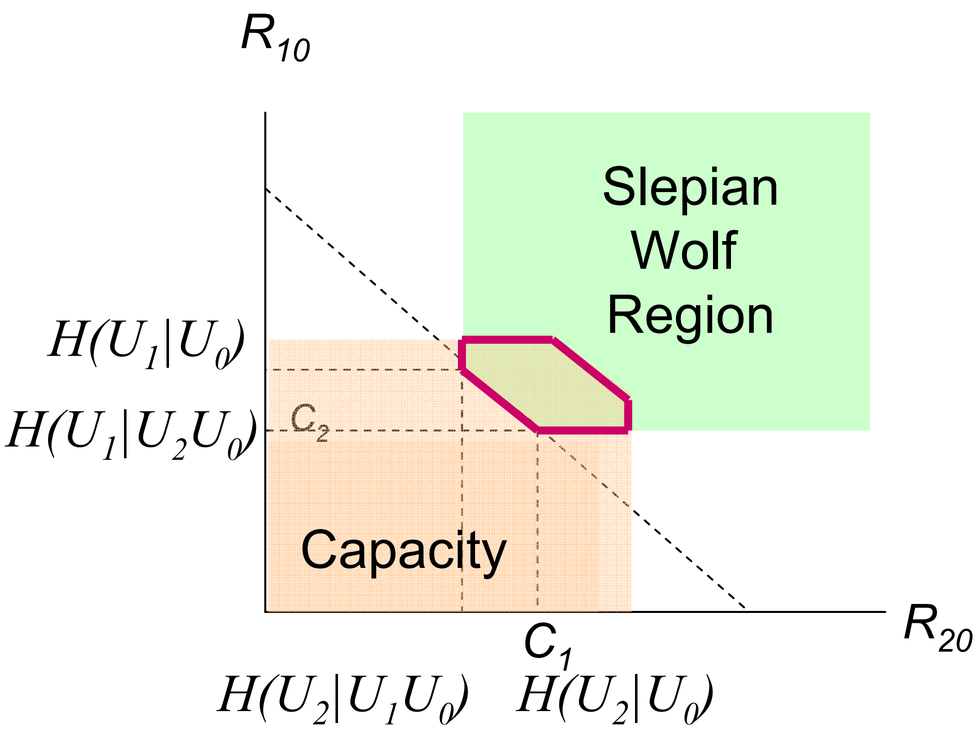
- Thus, $$H(U(S) \mid U(S^c)) < \sum_{\substack{i \in S \\ j \in S^c}} C_{ij}$$

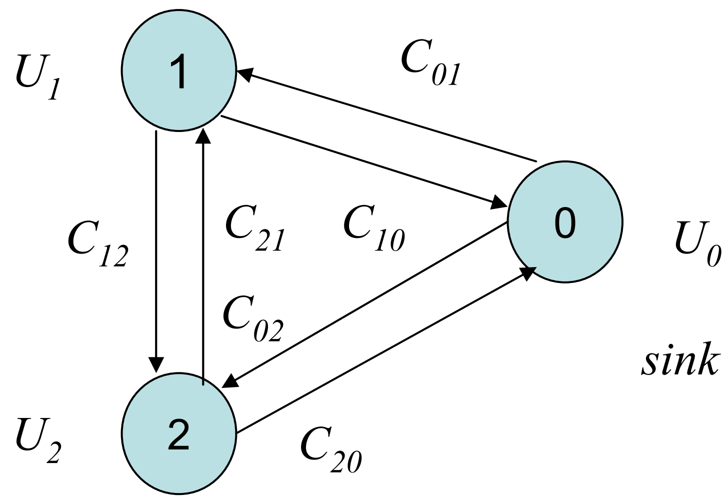
- The converse is too ugly for this talk.

An Optimal Protocol Stack

Interface across layers	**Layers**	**Protocols**
Rate control	Transport Layer	Data compression Rate control
Packets Local topology	Network Layer	Routing Fault tolerance Rate allocation
Bits	MAC Layer	Neighborhood discovery, power allocation **independent channels**
Waveforms	Physical Layer	Exchange of waveforms with neighbors

Network Optimization





- Is the rate polytope non-empty?
- If yes, what is an optimal flow?

Linear Programming

- Linear cost model (e.g. energy per bit)

$$\min \kappa(f) = \sum_{(v_i,v_j)\in E} w(v_i,v_j)\cdot f(v_i,v_j)$$

subject to:

$$f(i,j) \le c_{ij} \qquad 0 \le i,j \le M$$

flow constraints

$$f(i,j) = -f(j,i) \qquad 0 \le i,j \le M$$

$$\sum_{i\in V} f(i,j) = 0 \qquad 1 \le i \le M$$

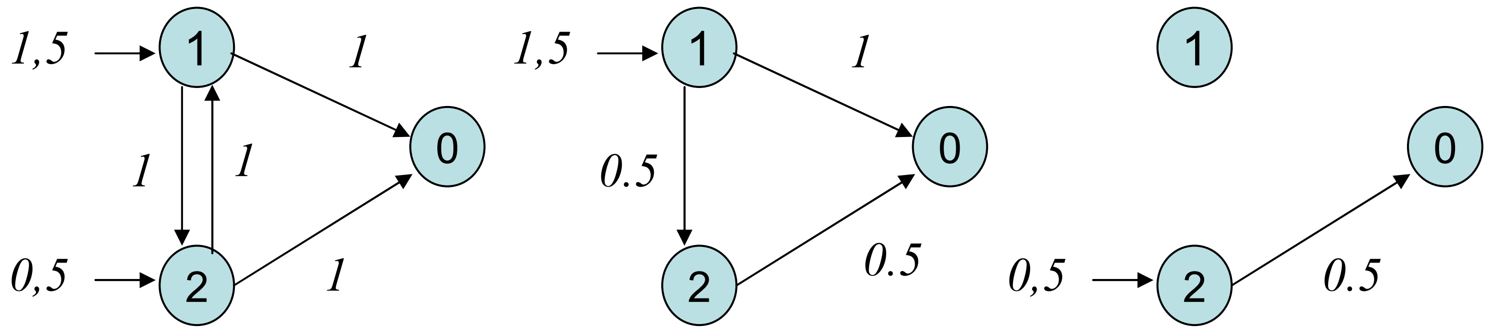
coding constraints

$$H(U_S \mid U_{S^c}) \le \sum_{i\in S} f(s,i) \le \sum_{i\in S,\, j\in S^c} c_{ij}$$

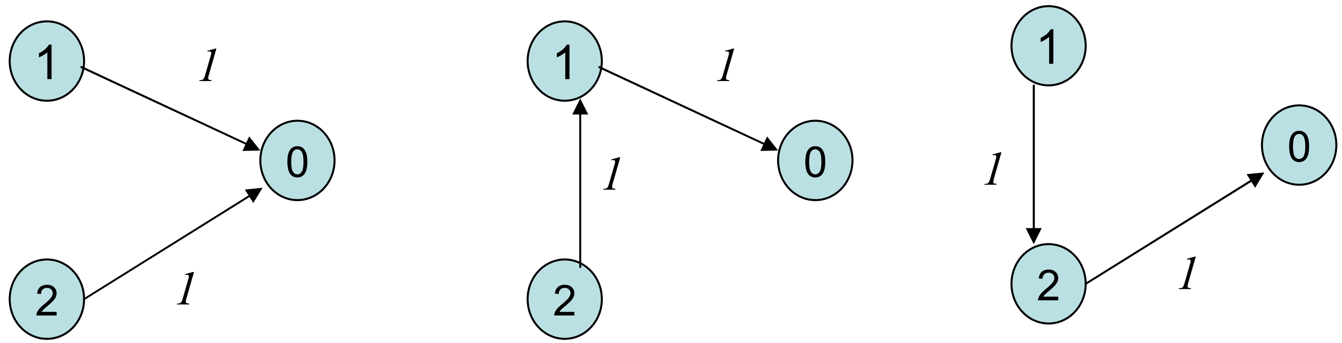
$$f(s,i) = R_i \qquad 1 \le i \le M$$

Data Gathering Trees

- This example is solvable…

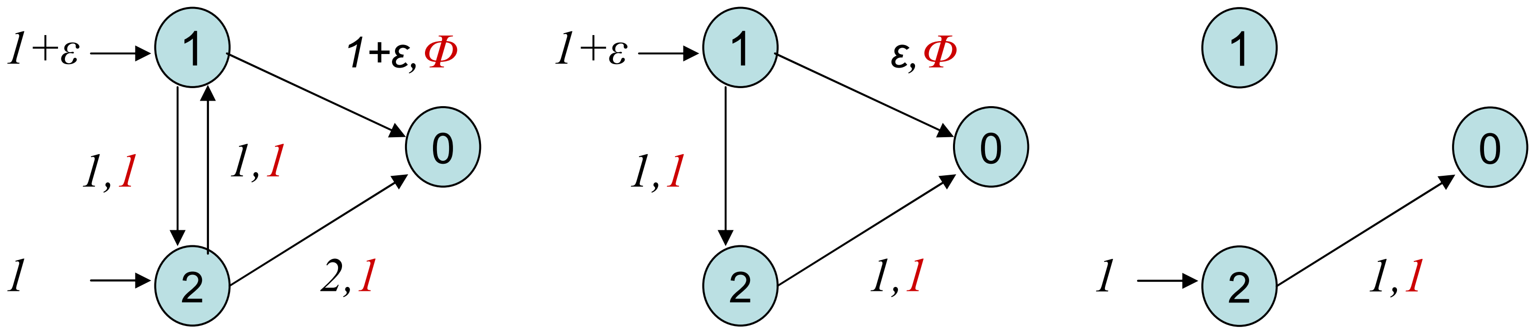


- …but not with trees!

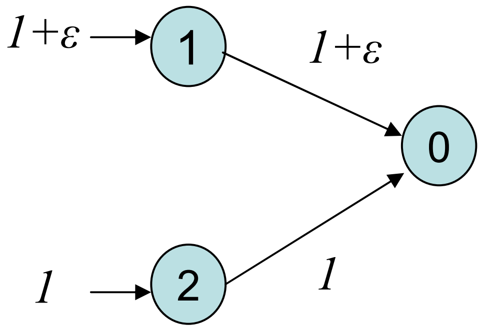


Data Gathering Trees (2)

- This example is also solvable…



- …but the cost of using a tree is huge!



Ratio Φ(1+ε)+1) / (ε Φ+3)

for large Φ, we get about 1+1/ε, unbounded for small ε!

Conclusions

- Network information flow can become a very powerful and useful concept.
- No doubt, the validity and usefulness of the separation principle in networks is a question of *when* and **not** *if* it holds.
- When does information transported through a network behave like water in pipes?
- How do we do *information-theoretically optimal networking*?

Merci!