Real-time communication: structure of optimal coding schemes

Aditya Mahajan
McGill University

Acknowledgements: Demo Teneketzis and Ashutosh Nayyar

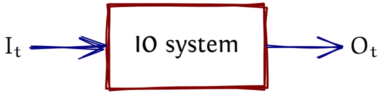
Interactive Information Theory Workshop,
Banff, Canada (January 18, 2012)

Real-time communication: Basic setup

- A stochastic source $\{S_t, t = 1, 2, \dots\}$.
- **Sequential** encoder and **sequential** decoder.
- Different channel models
  - Noiseless channel
  - Noiseless feedback
  - No feedback
  - Noisy feedback
- Finite-delay decoding $\rho_t(S_{t-d}, \hat{S}_t)$
- Fixed rate $X_t \in \mathcal{X}$

  or variable-rate $X_t \in \mathcal{X}_t$ (with a power/quantization cost $c_t(X_t)$).

A sequential strategy



- **Full memory**: $O_t = f_t(I_{1:t}, O_{1:t-1})$.
- **Fixed (not necessarily finite) memory**:

$$O_t = f_t(I_t, M_{t-1}), \quad \text{and} \quad M_t = g_t(I_t, M_{t-1}).$$

- **Sliding window memory**: $O_t = f_t(I_{t-k:t})$.

Solution concept

- Structure of optimal coding schemes

  $O_t = f_t(I_{1:t}, O_{1:t-1})$ vs $O_t = f_t(I_t, \pi_t)$ where $\pi_t = \pi_t(I_{1:t-1}, O_{1:t-1})$.

- Dynamic programming decomposition

  - non-classical information structure
  - Some recent results: Mahajan, 2008, Nayyar, 2010.
  - **Main insight**: Dynamic programming is possible only if structural results exist.

Brief Literature Overview

- Known source statistics (Stochastic control approach)

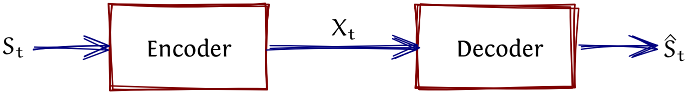
  Ericson (1979), Witsenhausen (1979), Gaarder Slepian (1982), Walrand Varaiya (1983), Borkar Mitter Tatikonda (2001), Teneketzis (2006), Mahajan Teneketzis (2008, 2009), Kaspi Merhav (2010), Nayyar Teneketzis (2011), Asnani Weissman (2011), Yüksel (2011).

- Unknown source statistics (individual sequence approach)

  Linder Lugousi (2001), Weissman Merhav (2002), Gÿorgy, Linder, Lugosi (2004), Matloub Weissman (2006)

- Many related setups . . .
  - **Causal coding**: Neuhoff Gilbert (1982), Linder Zamir (2006)
  - **Sequential coding**: Vishwanathan Berger (2009), Ma Ishwar (2011).
  - **DPCM coding**: Farvardin Modestino (1985), Chang Gibson (1991), Ishwar Ramachandaran (2004), Saxena Rose (2009)

Real-time source coding



- **Source**: First-order Markov process $\{S_t,\ t = 1, 2..., \}$.
- **Full memory encoder**: $X_t = f_t(S_{1:t}, X_{1:t-1})$
- **General decoder**: $\hat{S}_t = g_t(X_t, M_{t-1})$, $\qquad M_t = h_t(X_t, M_{t-1})$, $\qquad M_t \in \mathcal{M}_t$
- **quantization cost**: $c_t(X_t)$ and **distortion cost**: $\rho_t(S_t, \hat{S}_t)$.
- **Objective**: Choose $\mathbf{f} := (f_1, ..., f_T)$ and $\mathbf{g} := (g_1, ..., g_T)$ to minimize

$$\sum_{t=1}^{T} \left[\rho_t(S_t, \hat{S}_t) + c_t(X_t)\right]$$

Simple generalizations of the model

Higher-order Markov source

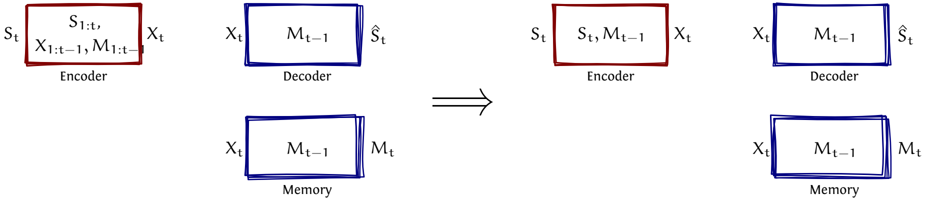
If $\{S_t, t = 1, 2, \ldots\}$ is k-th order Markov, the results hold for $\tilde{S}_t = S_{t-k+1:t}$ which is first-order Markov.

Fixed-finite delay

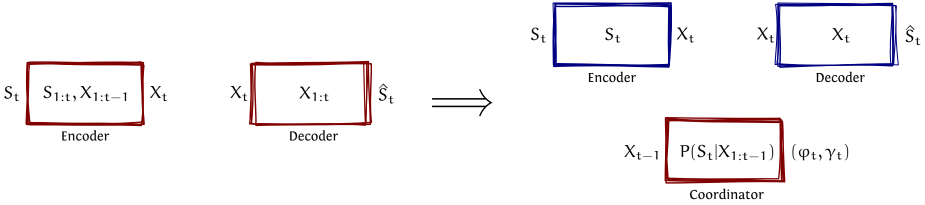
To reconstruct after delay d, i.e., $\rho_t(S_{t-d}, \hat{S}_t)$ (also called finite **look-ahead**), the results hold for $\tilde{S}_t = S_{t-d:t}$ and an appropriately defined distortion $\tilde{\rho}_t(\tilde{S}_t, \hat{S}_t) = \rho_t(S_{t-d}, \hat{S}_t)$.

Types of structural result

MDP-type structural result



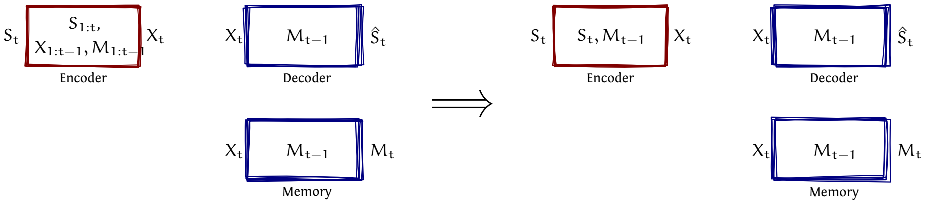
POMDP-type structural result



MDP-type of structural result

Without loss of optimality, we may restrict attention to encoders of the form

$$X_t = f_t(S_t, M_{t-1})$$

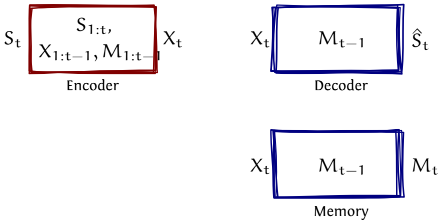


MDP-type of structural result (cont.)

Originally proved in Witsenhausen (1979) (and Kapsi Merhav (2010) for quantization cost). Simpler proof based on Teneketzis (2006).

- Note that $\{M_t, t = 1, 2, \dots\}$ is a filtration of $\{X_t, t = 1, 2, \dots\}$.

- **Arbitrarily** fix decoder g and memory update h. Optimal design of the **best-response** encoder is a **centralized** stochastic control problem.

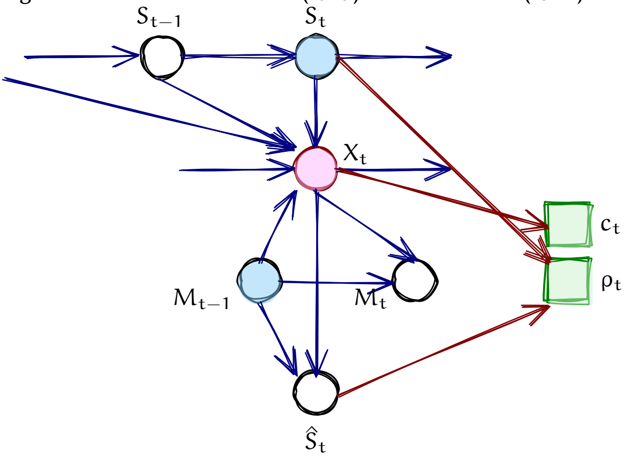


- The process $\{(S_t, M_{t-1}), t = 1, \dots\}$ is a controlled Markov chain controlled by X_t.

- The conditional expected cost $E[c_t(X_t) + \rho_t(S_t, \hat{S}_t) \mid \text{data at controller}]$ depends only on (S_t, M_{t-1}, X_t).

MDP-type of structural result (cont.)

Alternate visual proof based on graphical modeling approach of Mahajan Tatikonda (2010) which generalizes Witsenhausen (1979) and Blackwell (1964)



Conditioned on (S_t, M_{t-1}) and X_t, “past” is independent of “future”.

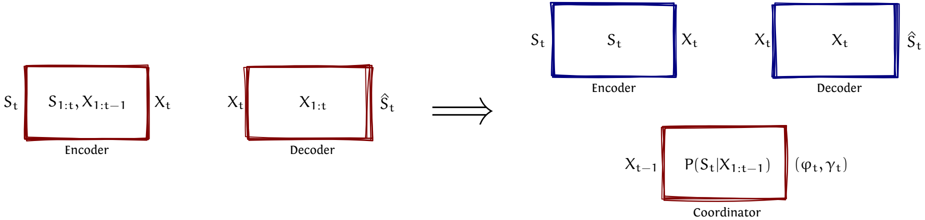
Allows for algorithmic verification of MDP type structural results.

POMDP-type of structural result

Consider the case of full memory at decoder, i.e., $\mathcal{M}_t = \prod_{\tau=1}^{t} \mathcal{X}_\tau$.

Define $\Pi_t = P(S_t \mid X_{1:t-1})$. Then, **without loss of optimality**, we may restrict attention to encoders and decoders of the form

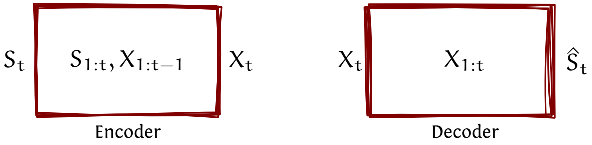
$$X_t = f_t(S_t, \Pi_t) \quad \text{and} \quad \hat{S}_t = g_t(X_t, \Pi_t).$$



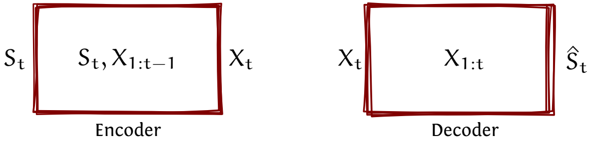
POMDP-type of structural result (cont.)

Originally proved in Walrand Varaiya (1983). Direct proof based on the general approach of Nayyar Mahajan Teneketzis (2011). (Lipster Shariyayev (1977) showed a similar result for transmitting linear Markov processes over AWGN channels with noiseless feedback.)

- Initial structure of encoders and decoders

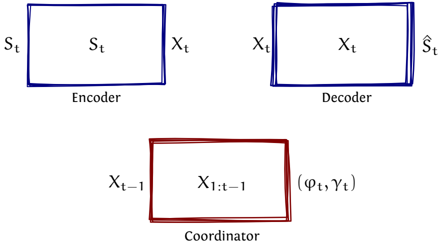


- Use the MDP-type structure result



POMDP-type of structural result (cont.)

- Identify a coordinator for the system



- The coordinator observes the **common information** $X_{1:t-1}$.
- ...and chooses prescriptions to the encoder and decoder:

$$(\varphi_t, \gamma_t) = h_t(X_{1:t-1})$$

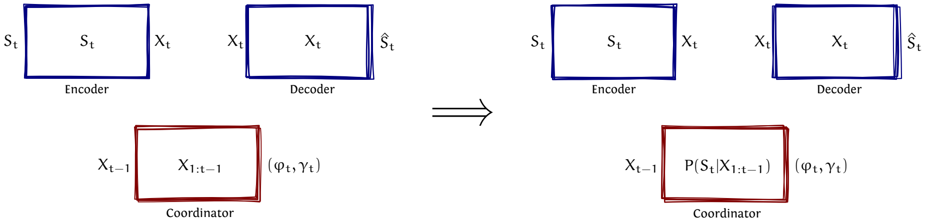
- Then encoder and decoder passively use the prescription:

$$X_t = \varphi_t(S_t) \quad \text{and} \quad \hat{S}_t = \gamma_t(X_t)$$

POMDP-type of structural result (cont.)

The coordinated system

- Is a centralized, partially-observed system.
- Use POMDP results to find structure of optimal strategies.
- Define **belief state** $\Pi_t = P(\text{state} \mid \text{obs}) = P(S_t \mid X_{1:t-1})$.
  Then Π_t is a **sufficient statistic** or **info-state** for $X_{1:t-1}$.



POMDP-type of structural result (cont.)

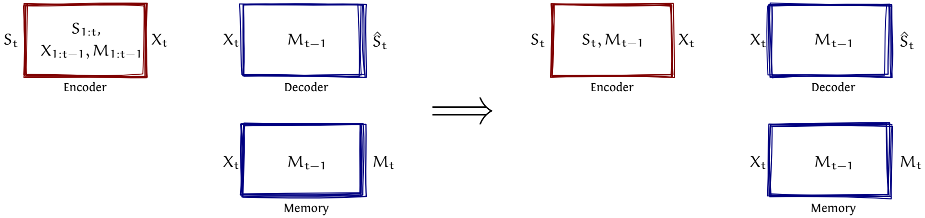
Define $\Pi_t = P(S_t \mid X_{1:t-1})$. Then, **without loss of optimality**, we may restrict attention to encoders and decoders of the form

$$X_t = f_t(S_t, \Pi_t) \quad \text{and} \quad \hat{S}_t = g_t(X_t, \Pi_t).$$

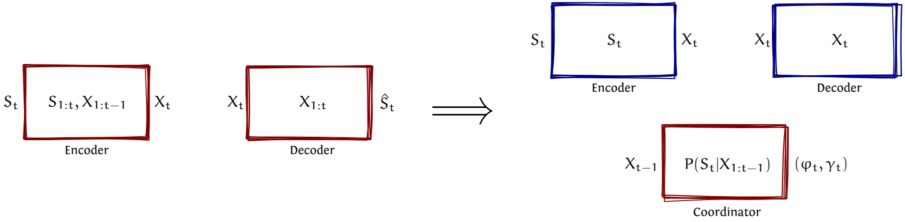
- The structure of the decoder can be slightly simplified.
- The coordinator approach also gives a dynamic programming decomposition.
- Even for the finite memory-setup, the coordinator approach can be used to get a dynamic programming decomposition. In this setup, the common information is empty; hence the coordinator is equivalent to a system designer choosing the design before the system starts operating.

Types of structural result

MDP-type structural result

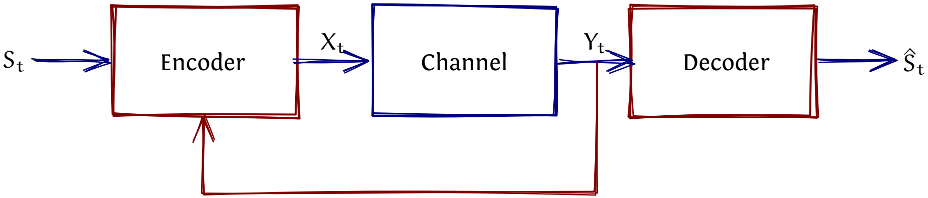


POMDP-type structural result



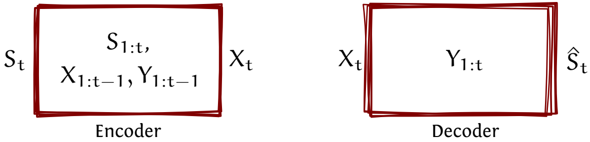
Noisy channel with noiseless feedback

Originally considered in Walrand Varaiya (1983)



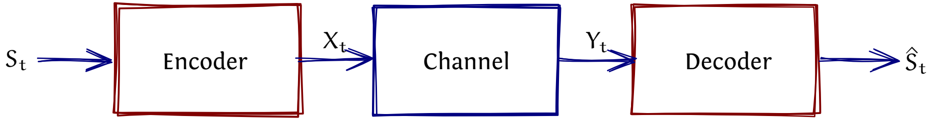
Effectively the same information-structure as in case of source coding.

Same results!



Noisy channel with no feedback

Originally considered in Teneketzis (2006). Dynamic programming decomposition presented in Mahajan Teneketzis (2009).



Let $\Xi_t = P(M_{t-1} \mid X_{1:t})$. Then, **without loss of optimality**, we can restrict attention to encoders of the form

$$X_t = f_t(S_t, \Xi_t)$$

Some generalizations

- **Side information at decoder**

  Considered in Teneketzis (2006) (for noisy channels) and in Kapsi Merhav (2010).

- **Noisy observations of the source**

  Considered in Borkar Mitter Tatikonda (2001) and Yüksel (2011)

- **Channels with memory**

  Considered in Mahajan Teneketzis (2009)

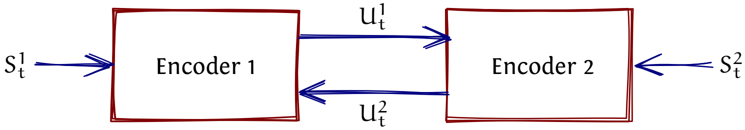
- **Fixed finite delay decoding of i.i.d. sequences**

  Considered in Asnani Weissman (2011) (for noisy channels with noiseless feedback).

> The **structural results** for these generalization can be worked out using the above described ideas.

An exercise: multi-round communication

Originally considered last night after a few drinks!



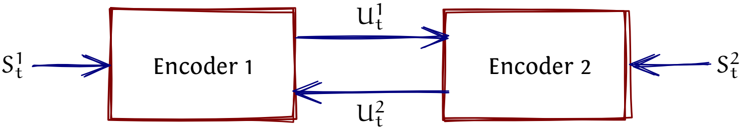
$$U_t^1 = f_t^1(S_{1:t}^1, U_{1:t-1}^1, U_{1:t-1}^2) \quad \hat{S}_t^2 = g_t^1(S_{1:t}^1, U_{1:t}^1, U_{1:t}^2)$$

$$U_t^2 = f_t^2(S_{1:t}^2, U_{1:t-1}^1, U_{1:t-1}^2) \quad \hat{S}_t^1 = g_t^2(S_{1:t}^2, U_{1:t}^1, U_{1:t}^2)$$

- **Distortion**: $\rho_t(S_t^1, S_t^2, \hat{S}_t^1, \hat{S}_t^2)$
- Independent Markov sources.

An exercise: multi-round communication

Originally considered last night after a few drinks!



$$U_t^1 = f_t^1(S_{1:t}^1, U_{1:t-1}^1, U_{1:t-1}^2) \quad \hat{S}_t^1 = g_t^1(S_{1:t}^1, U_{1:t}^1, U_{1:t}^2)$$

$$U_t^2 = f_t^2(S_{1:t}^2, U_{1:t-1}^1, U_{1:t-1}^2) \quad \hat{S}_t^2 = g_t^2(S_{1:t}^2, U_{1:t}^1, U_{1:t}^2)$$

- **Distortion**: $\rho_t(S_t^1, S_t^2, \hat{S}_t^1, \hat{S}_t^2)$
- Independent Markov sources.
- **Control sharing** info structure.

An exercise: multi-round communication (cont.)

An exercise: multi-round communication (cont.)

$S^1_{1:t}$, $U^1_{1:t-1}, U^2_{1:t-1}$

Encoder 1

$S^2_{1:t}$, $U^1_{1:t-1}, U^2_{1:t-1}$

Encoder 2

Lemma: The sources are conditionally independent.

$$P(S^1_{1:t}, S^2_{1:t} \mid U^1_{1:t-1}, U^2_{1:t-1}) = P(S^1_{1:t} \mid U^1_{1:t-1}, U^2_{1:t-1})\, P(S^2_{1:t} \mid U^1_{1:t-1}, U^2_{1:t-1})$$

An exercise: multi-round communication (cont.)

S^1_t,
$U^1_{1:t-1}, U^2_{1:t-1}$

Encoder 1

$S^2_{1:t}$,
$U^1_{1:t-1}, U^2_{1:t-1}$

Encoder 2

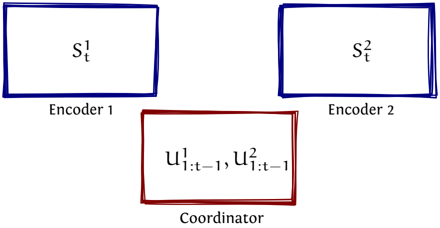
Past $S^1_{1:t-1}$ is redundant at encoder 1.

An exercise: multi-round communication (cont.)

Encoder 1	Encoder 2
S^1_t, $U^1_{1:t-1}, U^2_{1:t-1}$	S^2_t, $U^1_{1:t-1}, U^2_{1:t-1}$

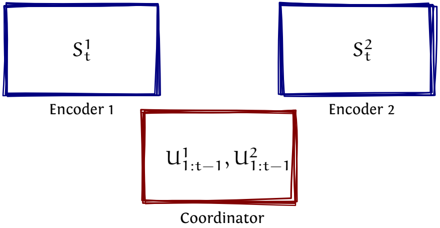
By symmetry $S^2_{1:t-1}$ is redundant at encoder 2.

An exercise: multi-round communication (cont.)

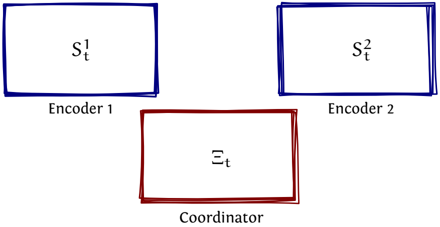


Consider a **coordinator** that observes **common information**

An exercise: multi-round communication (cont.)



Define $\Xi_t = \mathbb{P}(S^1_t, S^2_t \mid U^1_{1:t-1}, U^2_{1:t-1})$. Then,



An exercise: multi-round communication (cont.)

Define $\Pi_t^i = P(S_t^i \mid U_{1:t-1}^1, U_{1:t-1}^2)$. From conditional independence of sources:

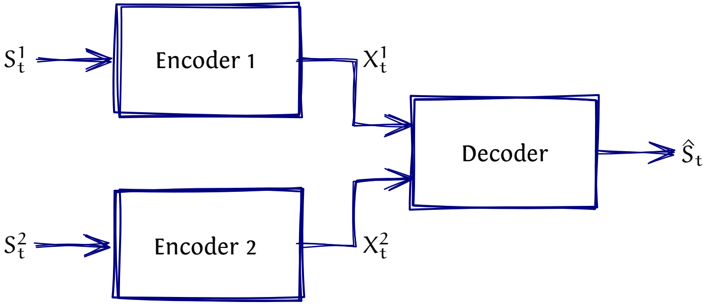
$$\Xi_t \equiv (\Pi_t^1, \Pi_t^2)$$

There is no loss of optimality in restricting:

$$U_t^i = f_t^i(S_t^i, \Pi_t^1, \Pi^2) \quad \text{and} \quad \hat{S}_t^j = g_t^i(U_t^j, \Pi_t^1, \Pi_t^2)$$

Multi-terminal systems

Originally considered in Nayyar Teneketzis (2011)

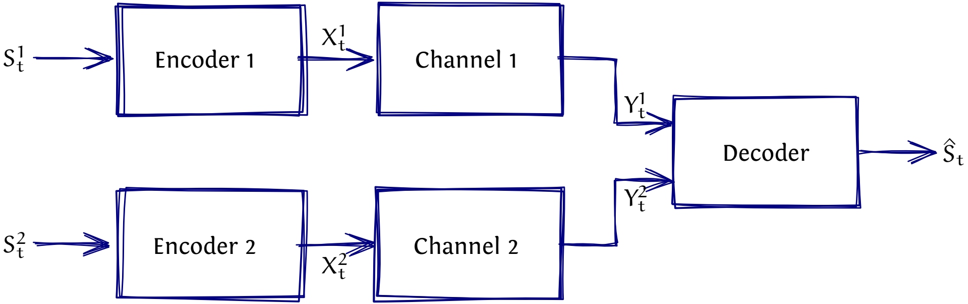


Assumption: $\exists A$ such that $P(S^1_{1:t}, S^2_{1:t} \mid A) = P(S^1_{1:t} \mid A)\, P(S^2_{1:t} \mid A)$

> Define $B^i_t = P(A \mid S^1_{1:t})$ and $\Pi_t = P(S^1_t, S^2_t \mid X^1_{1:t-1}, X^2_{1:t-1})$. Then, **wloo**:
>
> $$X^i_t = f_t(S^i_t, B^i_t, \Pi_t) \quad \text{and} \quad \hat{S}_t = g_t(X^1_t, X^2_t, \Pi_t)$$

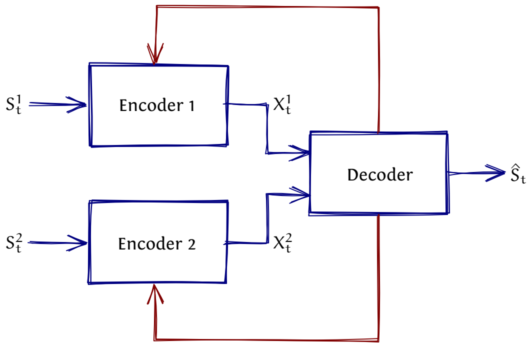
Multi-terminal systems

Originally considered in Nayyar Teneketzis (2011)



Multi-terminal systems

Originally considered in Yüksel (2011)



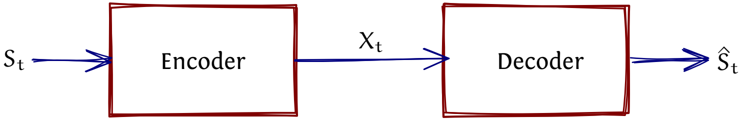
An encoder of the form

$$X_t^i = P(S_t^i, P(S_t^i \mid X^1_{1:t-1}, X^2_{1:t-1}))$$

is **not optimal**

Another flavor of results
Finite memory vs sliding window memory

Originally showed in Kaspi Merhav (2010) using ideas from Merhav Ziv (2006).



Variable code setup

- Quantization cost: $c(\cdot) = -\lambda \log P(\cdot)$
- System 1: **Finite memory decoder** with memory $\mathcal{M}$
- System 2: **Sliding window decoder** with window size k

$$\mathcal{J}(\text{Finite memory decoder}) \geqslant \mathcal{J}(\text{Sliding window decoder}) - \lambda \log |\mathcal{M}|/k$$

Implications for interactive communication

- Identify "easy" and "hard" problems based on info struct
  - This classification is not always consistent with that of information theory (no-feedback vs noiseless feedback; MAC with feedback vs BC with feedback, etc.).
- Optimal structure of block Markov coding scheme
  - For example, for MAC with feedback, do different achievable schemes (Cover-Leung, Bross-Lapidoth, Venkataramanan-Pradhan) have optimal structure at block level.
  - Might be useful for relay networks as well.
- Relation between auxiliary random variables in info theory
  - I think that there is a relation . . . but cannot explain it formally.