# On optimal block Markov coding schemes for multiple-access channel with feedback

Aditya Mahajan

McGill University

Conference on Information Science and Systems, 2012

# Motivation

The capacity of multi-terminal communication systems is often characterized by multi-letter directed information expressions which are difficult to compute.

# Motivation

The capacity of multi-terminal communication systems is often characterized by multi-letter directed information expressions which are difficult to compute.

## Capacity of MAC with feedback

$$C = \bigcup_{\text{code trees}} \left\{ (R_1, R_2) : \begin{aligned} R_1 &\le I_\infty(X^1 \to Y \mid X^2) \\ R_2 &\le I_\infty(X^2 \to Y \mid X^1) \\ R_1 + R_2 &\le I_\infty(X^1 X^2 \to Y) \end{aligned} \right\}$$

[Kramer-03]

# Motivation

The capacity of multi-terminal communication systems is often characterized by multi-letter directed information expressions which are difficult to compute.

## Capacity of MAC with feedback

$$C = \bigcup_{\text{code trees}} \left\{ (R_1, R_2) : \begin{array}{r} R_1 \le I_\infty(X^1 \to Y \mid X^2) \\ R_2 \le I_\infty(X^2 \to Y \mid X^1) \\ R_1 + R_2 \le I_\infty(X^1 X^2 \to Y) \end{array} \right\}$$

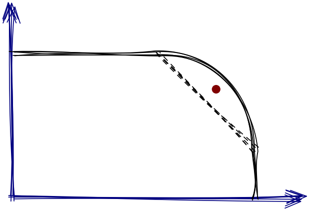
[Kramer-03]

Block Markov coding schemes are a class of achievable schemes that give inner bounds on capacity

# Idea for acheivability: cooperative comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in two phases

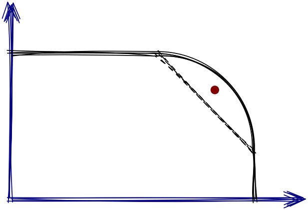
[Gaarder-Wolf-75]

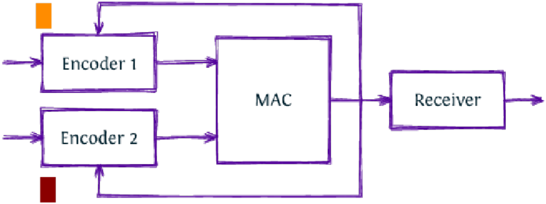


# Idea for acheivability: cooperative comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in two phases

[Gaarder-Wolf-75]

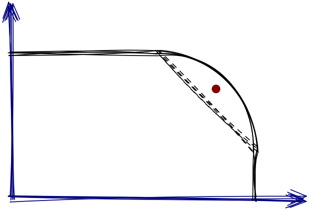


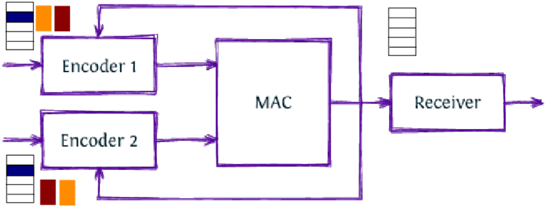


# Idea for acheivability: cooperative comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in two phases

[Gaarder-Wolf-75]

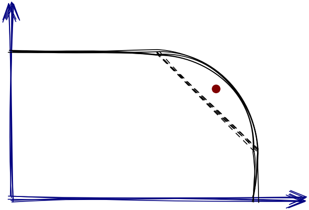


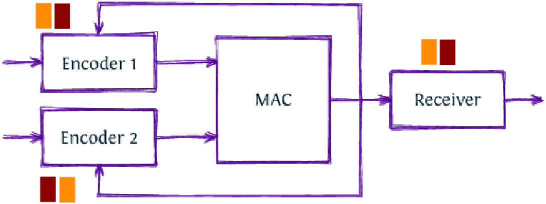


# Idea for acheivability: cooperative comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in two phases

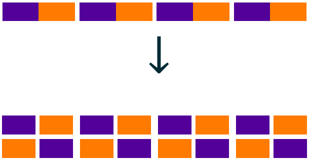
[Gaarder-Wolf-75]





# Block Markov coding scheme

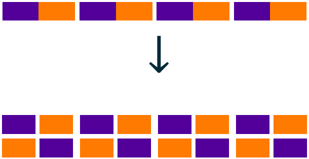
## Cover-Leung Scheme: decode after one block



[Cover-Leung-81]

# Block Markov coding scheme

## Cover-Leung Scheme: decode after one block



[Cover-Leung-81]

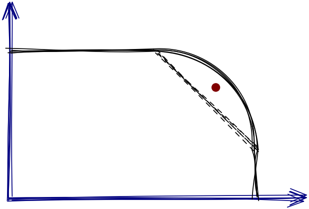
$$R_{CL} = \bigcup_{P_U P_{X^1|U} P_{X^2|U}} \left\{ (R_1, R_2) : \begin{array}{r} R_1 \le I(X^1 \wedge Y \mid U X^2) \\ R_2 \le I(X^2 \wedge Y \mid U X^1) \\ R_1 + R_2 \le I(X^1 X^2 \wedge Y) \end{array} \right\}$$

Not tight in general

# Idea for achievability: two way comm

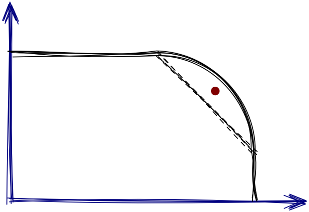
Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in three phases

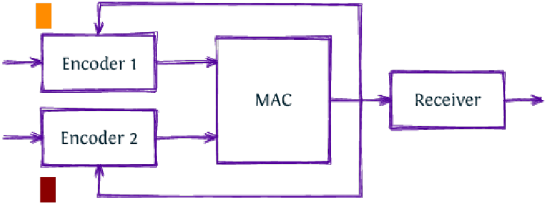
[Bross-Lapidoth-05]



# Idea for achievability: two way comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in three phases [Bross-Lapidoth-05]

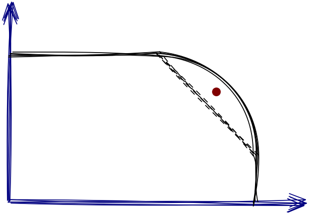


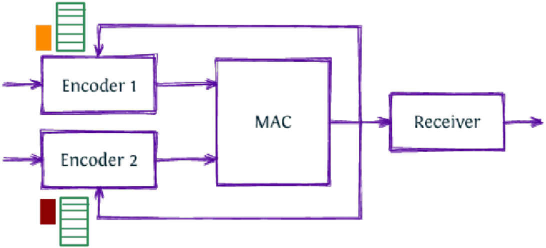


# Idea for achievability: two way comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in three phases

[Bross-Lapidoth-05]

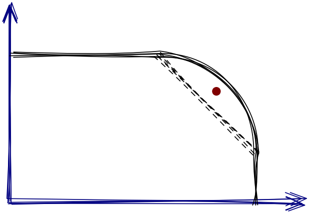


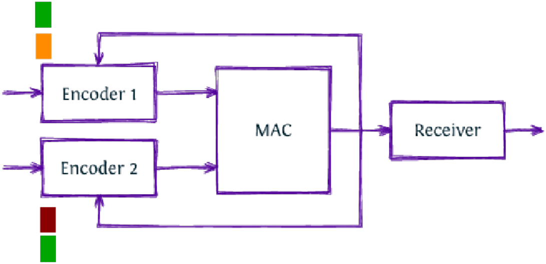


# Idea for achievability: two way comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in three phases

[Bross-Lapidoth-05]

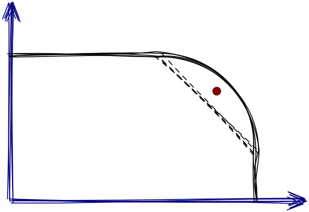


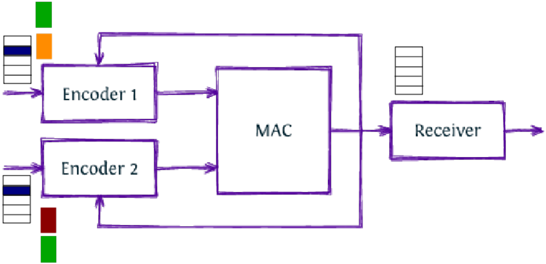


# Idea for achievability: two way comm

Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in three phases

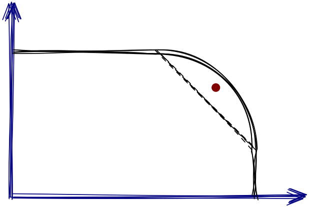
[Bross-Lapidoth-05]

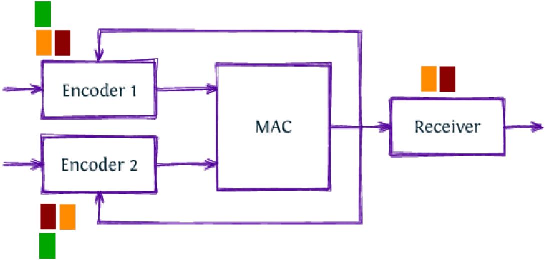




# Idea for achievability: two way comm

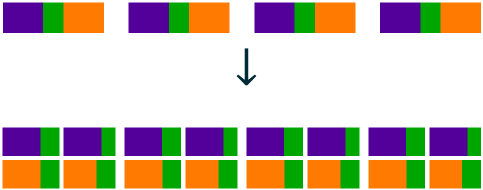
Suppose we want to transmit at a rate point outside the no-feedback capacity region. Communicate in three phases [Bross-Lapidoth-05]





# Block Markov coding scheme

## Bross-Lapidoth Scheme: decode after two blocks



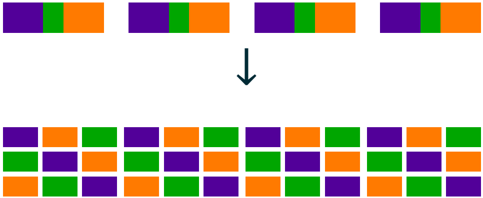
[Bross-Lapidoth-05]

# Block Markov coding scheme

$$R_{BL} = \bigcup_{\substack{P_U P_{X^1|U} P_{X^2|U} \\ g_1 : \mathcal{X}^1 \times \mathcal{Y} \to \mathcal{V}^1, g_2 : \mathcal{X}^2 \times \mathcal{Y} \to \mathcal{V}^2 \\ \text{stuff}}} \left\{ (R_1, R_2) : \begin{aligned} R_1 &\leq (1+\eta)^{-1} I(X^1 \wedge Y \mid U X^2 V^1) \\ R_2 &\leq (1+\eta)^{-1} I(X^2 \wedge Y \mid U X^1 V^2) \\ R_1 + R_2 &\leq (1+\eta)^{-1} I(X^1 X^2 \wedge Y \mid V^1 V^2) - R_L \\ R_L &\leq \text{complicated expression} \\ \eta &\geq \text{complicated expression} \end{aligned} \right\}$$

# Another block Markov scheme

## Venkataramanan Pradhan Scheme



[Venkataramanan-Pradhan-11]

# Another block Markov scheme

$$R_{VP} = \bigcup_{\substack{\tilde{S}=(\tilde{U},\tilde{V}^1,\tilde{V}^2,\tilde{Y}),S=(U,V^1,V^2,Y) \\ P_{\tilde{S}\tilde{X}^1,\tilde{X}^2} \\ P_{V^1|\tilde{S},\tilde{X}^1}P_{V^2|\tilde{S},\tilde{X}^2} \\ P_{UX^1X^2Y|V^1V^2} \\ \text{stuff}}} \left\{ (R_1, R_2) : \begin{aligned} R_1 &\le I(X^1 \wedge Y \mid X^2V^2U\tilde{S}\tilde{X}^2) - [I(V^1 \wedge X^2 \mid YV^2U\tilde{S}\tilde{X}^2) - I(U \wedge Y \mid \tilde{U}\tilde{Y})]^+ \\ R_2 &\le I(X^2 \wedge Y \mid X^1V^1U\tilde{S}\tilde{X}^1) - [I(V^2 \wedge X^1 \mid YV^1U\tilde{S}\tilde{X}^1) - I(U \wedge Y \mid \tilde{U}\tilde{Y})]^+ \\ R_1 + R_2 &\le I(X^1X^2 \wedge Y \mid U\tilde{S}) + I(U \wedge Y \mid \tilde{U}\tilde{Y}) \end{aligned} \right\}$$

Not tight in general

# Comparison of the different schemes

- BL and VP schemes (decode after two blocks) better than CL (decode after one block)

- Relation between BL and VP schemes is not clear.

  The achievable rate region of both schemes is given by the solutions of complicated non-linear optimization problems. As such, it is hard to evaluate the entire region.

# Comparison of the different schemes

- BL and VP schemes (decode after two blocks) better than CL (decode after one block)
- Relation between BL and VP schemes is not clear.

  The achievable rate region of both schemes is given by the solutions of complicated non-linear optimization problems. As such, it is hard to evaluate the entire region.

> Is there another block Markov coding scheme that **decodes after two blocks** and beats both BL and VP?

What is the structure of optimal block Markov coding schemes that decode after a fixed delay?

# Block Markov coding scheme

Let

$N$ : Block length
$B$ : Number of blocks
$M^i$ : Message sizes
$d$ : Decoding delay (in blocks)

A $(N, B, M^1, M^2, d)$ block Markov (BM) coding scheme is a tuple $(f^1, f^2, g)$ where $f^i = (f^i_1, \ldots, f^i_B)$, $g = (g_1, \ldots, g_B)$, and

$$f^i_b : (\mathcal{M}^i)^b \times \mathcal{Y}^{(b-1)N} \mapsto (\mathcal{X}^i)^N$$

$$g_b : \mathcal{Y}_{1:b} \mapsto (\mathcal{M}^i, \mathcal{M}^2)$$

# Block Markov coding scheme

## Cost function

For any scheme $S = (f^1, f^2, g)$ and all $b > d$

$$J_b(S) = \mathbb{P}(W_{1:b-d} \neq \hat{W}_{d+1:b} \mid S)$$
$$+ \mathbb{P}(W_{b-d+1:B-d} \neq \hat{W}_{b+1:B} \mid W_{1:b-d} = \hat{W}_{d+1:b}, S)$$

## $\varepsilon$-strong achievability

For any $\varepsilon > 0$, the rate pair $(R^1, R^2)$ is $\varepsilon$-strongly achievable if for all sufficiently large block length and number of blocks, $\exists$ coding schemes $S = (f^1, f^2, g)$ such that

1. $(B-d)\dfrac{\log M^i}{BN} \geq R^i - \varepsilon$
2. $J_b(S) \leq \varepsilon$ for all $b > d$.

# Problem formulation

Can we restrict attention to a subclass $\mathcal{S}^*$ of all coding schemes $\mathcal{S}$ such that $\exists S^* \in \mathcal{S}^*$ and

$$J_b(S^*) \leq J_b(S), \quad \forall S \in \mathcal{S} \text{ and } \forall b > d.$$

# Structure of optimal decoder

## Optimistic MAP decoder

$$g_b(y_{1:b}) = \arg\min_{w} \mathbb{P}(W_{b-d} = w \mid Y_{1:b} = y_{1:b}, \hat{W}_{d+1:b-1} = W_{1:b-1}, f^1, f^2)$$

# Structure of optimal decoder

## Optimistic MAP decoder

$$g_b(y_{1:b}) = \arg\min_w \mathbb{P}(W_{b-d} = w \mid Y_{1:b} = y_{1:b}, \hat{W}_{d+1:b-1} = W_{1:b-1}, f^1, f^2)$$

## Lemma (Optimal decoder)

For any coding scheme $S$, changing the decoder to be an optimistic MAP decoder improves $J_b$ universally (i.e., for all $b > d$).

Recall

$$\begin{aligned} J_b(S) &= \mathbb{P}(W_{1:b-d} \neq \hat{W}_{d+1:b} \mid S) \\ &\quad + \mathbb{P}(W_{b-d+1:B-d} \neq \hat{W}_{b+1:B} \mid W_{1:b-d} = \hat{W}_{d+1:b}, S) \end{aligned}$$

# Dependency of MAP decoder on the encoder

$W_{b-d}$ depends on $Y_{1:b}$ through the channel inputs $X_{b-d+1:b}$.

# Dependency of MAP decoder on the encoder

$W_{b-d}$ depends on $Y_{1:b}$ through the channel inputs $X_{b-d+1:b}$.

Suppose $\hat{W}_{d+1:b-1} = W_{1:b-d-1}$.

$$x^i_{b-d} = f^i_b(w_{1:b-d}, y_{1:b-d-1}) = \varphi^{i,(0)}_{b-d}(w_{b-d})$$

# Dependency of MAP decoder on the encoder

$W_{b-d}$ depends on $Y_{1:b}$ through the channel inputs $X_{b-d+1:b}$.

Suppose $\hat{W}_{d+1:b-1} = W_{1:b-d-1}$.

$$x^i_{b-d} = f^i_b(w_{1:b-d}, y_{1:b-d-1}) = \varphi^{i,(0)}_{b-d}(w_{b-d})$$

$$x^i_{b-d+1} = f^i_b(w_{1:b-d+1}, y_{1:b-d}) = \varphi^{i,(1)}_{b-d+1}(w_{b-d}, w_{b-d+1})$$

# Dependency of MAP decoder on the encoder

$W_{b-d}$ depends on $Y_{1:b}$ through the channel inputs $X_{b-d+1:b}$.

Suppose $\hat{W}_{d+1:b-1} = W_{1:b-d-1}$.

$$\begin{aligned}
x^i_{b-d} &= f^i_b(w_{1:b-d}, y_{1:b-d-1}) = \varphi^{i,(0)}_{b-d}(w_{b-d}) \\
x^i_{b-d+1} &= f^i_b(w_{1:b-d+1}, y_{1:b-d}) = \varphi^{i,(1)}_{b-d+1}(w_{b-d}, w_{b-d+1}) \\
\dots &= \dots \\
x^i_{b-d+k} &= f^i_b(w_{1:b-d+k}, y_{1:b-d+k}) = \varphi^{i,(k)}_{b-d+k}(w_{b-d:b-d+k}) \\
\dots &= \dots
\end{aligned}$$

# Dependency of MAP decoder on the encoder

$W_{b-d}$ depends on $Y_{1:b}$ through the channel inputs $X_{b-d+1:b}$.

Suppose $\hat{W}_{d+1:b-1} = W_{1:b-d-1}$.

$$
\begin{aligned}
x^i_{b-d} &= f^i_b(w_{1:b-d}, y_{1:b-d-1}) = \varphi^{i,(0)}_{b-d}(w_{b-d}) \\
x^i_{b-d+1} &= f^i_b(w_{1:b-d+1}, y_{1:b-d}) = \varphi^{i,(1)}_{b-d+1}(w_{b-d}, w_{b-d+1}) \\
\dots &= \dots \\
x^i_{b-d+k} &= f^i_b(w_{1:b-d+k}, y_{1:b-d+k}) = \varphi^{i,(k)}_{b-d+k}(w_{b-d:b-d+k}) \\
\dots &= \dots
\end{aligned}
$$

$\mathbb{P}(W_{b-d} \mid Y_{1:b}, \hat{W}_{d+1:b-1} = W_{1:b-d-1})$ can be computed using the partially evaluated functions $\varphi^{i,(0)}_{b-d}(\cdot), \dots, \varphi^{i,(d)}_{b}(\cdot)$.

# Structure of optimal decoder

$$
\begin{aligned}
g_b(y_{1:b}) &= \arg\min_{w} \mathbb{P}(W_{b-d} = w \mid Y_{1:b} = y_{1:b}, \hat{W}_{d+1:b-1} = W_{1:b-1}, f^1, f^2) \\
&= \tilde{g}_b(\varphi^{(0)}_{b-d}, \dots, \varphi^{(d)}_b)
\end{aligned}
$$

where for $k \le d$

$$
\varphi^{i,(k)}_b(w^i_{b-k:b}) = f_i(w^i_{1:b-k-1}, w^i_{b-k:b}, y_{1:b-1})
$$

and $\varphi^{(d)}_b = \left(\varphi^{1,(d)}_b, \varphi^{2,(d)}_b\right)$

Note that the decoder $\tilde{g}_b$ is a **fixed memory** decoder.

# Structure of optimal decoder

$$g_b(y_{1:b}) = \arg\min_{w} \mathbb{P}(W_{b-d} = w \mid Y_{1:b} = y_{1:b}, \hat{W}_{d+1:b-1} = W_{1:b-1}, f^1, f^2)$$
$$= \tilde{g}_b(\varphi_{b-d}^{(0)}, \dots, \varphi_b^{(d)})$$

where for $k \le d$

$$\varphi_b^{i,(k)}(w_{b-k:b}^i) = f_i(w_{1:b-k-1}^i, w_{b-k:b}^i, y_{1:b-1})$$

and $\varphi_b^{(d)} = \left(\varphi_b^{1,(d)}, \varphi_b^{2,(d)}\right)$

Note that the decoder $\tilde{g}_b$ is a **fixed memory** decoder.

**Proposition (Optimal decoder)**

Restricting attention to finite memory decoders of the above form does not affect the $\varepsilon$-strongly-achievable region.

# Structure of optimal encoders

- Finite memory decoder
- $d$-block decoding delay

**The argument of Witsenhausen-79 for the structure of optimal real-time encoder applies**, although we are looking at multi-terminal setup with a different objective function.

# Structure of optimal encoders

- Finite memory decoder
- $d$-block decoding delay

**The argument of Witsenhausen-79 for the structure of optimal real-time encoder applies**, although we are looking at multi-terminal setup with a different objective function.

### Proposition (Optimal Encoder)

Restricting attention to encoders of the form

$$x_b^i = f_b^i(w_{b-d:b}^i, \varphi_{b-d}^{(0)}(\cdot), \ldots, \varphi_{b-1}^{(d-1)}(\cdot))$$

does not affect the $\varepsilon$-strongly-achievable region.

# Proof idea

Proof proceeds by backward induction. Assume both encoders from block $b+1 : B$ have the desired structure.

Pick any encoder and look at $f_b^i$. The costs of interests are $J_b$, $J_{b+1}$, …, $J_B$.

# Proof idea

Proof proceeds by backward induction. Assume both encoders from block $b+1:B$ have the desired structure.

Pick any encoder and look at $f_b^i$. The costs of interests are $J_b$, $J_{b+1}$, …, $J_B$.

Assume $\hat{W}_{1:b-1} = W_{d+1:b-d-1}$.

Can pick a randomized encoder $\tilde{f}_b^i$ that induces the same distribution on the future realizations of the partially evaluated functions $\{(\varphi_{c-d}^{(0)}(\cdot), \dots, \varphi_c^{(d)}(\cdot))\}_{c=b}^{B}$ as given by $f_b^i$.

# Summary

- Structure of Encoder

$$x_b^i = f_b^i(w_{b-d:b}^i, \varphi_{b-d}^{(0)}(\cdot), \dots, \varphi_{b-1}^{(d-1)}(\cdot))$$

- Structure of Decoder

$$\hat{w}_b = \tilde{g}_b(\varphi_{b-d}^{(0)}, \dots, \varphi_b^{(d)})$$

# Comparison with achievable schemes

## Decoding with one block delay

- Our result

$$x_b^i = f_b^i(w_{b-1}^i, w_b^i, \varphi_{b-1}^{(0)}(\cdot))$$

$$\hat{w}_b = \tilde{g}_b(\varphi_{b-1}^{(0)}, \varphi_b^{(1)})$$

- Cover-Leung

$$x_b^i = f_b^i(w^i, w_{b-1}^i, u)$$

The decoder is a typicality decoder.

# Comparison with achievable schemes

- At the encoder, $\varphi_{b-1}^{(0}$ plays the role of the auxiliary random variable $u$.

- Compare from coding with causal side information (code function = auxiliary random variable). This case is different because the code function is at a block level, while the auxiliary random variable is at the symbol level.

# Comparison with achievable schemes

## Coding with two blocks of delay

- Our result

$$x_b^i = f_b^i(w_{b-2}^i, w_{b-1}^i, w_b^i, \varphi_{b-2}^{(0)}(\cdot), \varphi_{b-1}^{(1)})$$
$$\hat{w}_b = \tilde{g}_b(\varphi_{b-2}^{(0)}, \varphi_{b-1}^{(1)}, \varphi_b^{(2)})$$

- Bross-Lapidoth and Venkataramanan-Pradhan

Both have multiple auxiliary random variables and many other conditions. Difficult to make a direct comparison.

# Conclusion

- At the block level, block Markov coding schemes are similar to real-time communication.
- Stochastic control provides some insights into design of block Markov coding schemes and some intuition on the choice of auxiliary variables.
- MAC with multiple encoders and feedback?
- Other multi-terminal communication systems?

Thank you