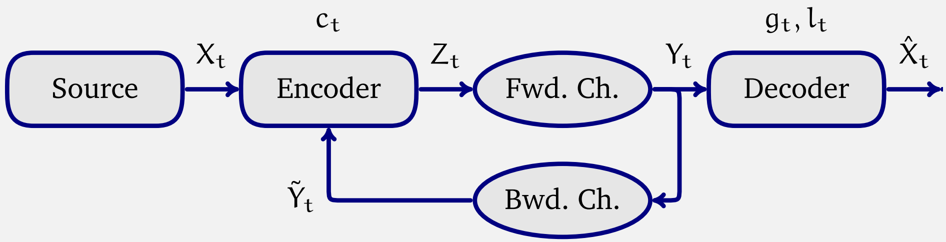
On real-time communication systems with noisy feedback

Aditya Mahajan and Demosthenis Teneketzis

DEPT. OF EECS,
UNIVERSITY OF MICHIGAN,
ANN ARBOR, MI. USA.

Presented at IEEE Information Theory Workshop, September 4, 2007.

Problem Formulation



Model

Discrete time, discrete valued

Source	: $\{X_t, t = 1, \dots, T\}$
Encoder	: $Z_t = c_t(X^t, \tilde{Y}^{t-1}, Z^{t-1})$
Forward ch.	: $Y_t = h(Z_t, N_t)$
Backward ch.	: $\tilde{Y}_t = \tilde{h}(Y_t, \tilde{N}_t)$
Decoder	: $\hat{X}_t = g_t(Y_t, M_{t-1})$
Memory Up.	: $M_t = l_t(Y_t, M_{t-1})$

Objective

Choose $(c_1, \dots, c_T)$, $(g_1, \dots, g_T)$, $(l_1, \dots, l_T)$ to minimize

$$\mathcal{J} := \mathbf{E}\left\{\sum_{t=1}^{T} \rho(X_t, \hat{X}_t)\right\}$$

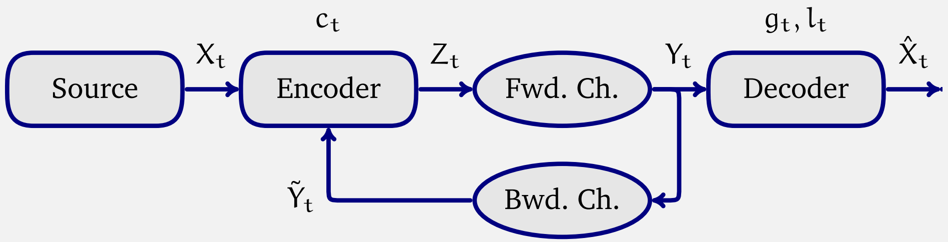
Outline

- Literature Overview
- Salient Features
- Solution Methodology

  (i) Structural Properties

  (ii) Global Optimization
- Conclusions

Literature Overview



real-time communication with **noisy feedback**

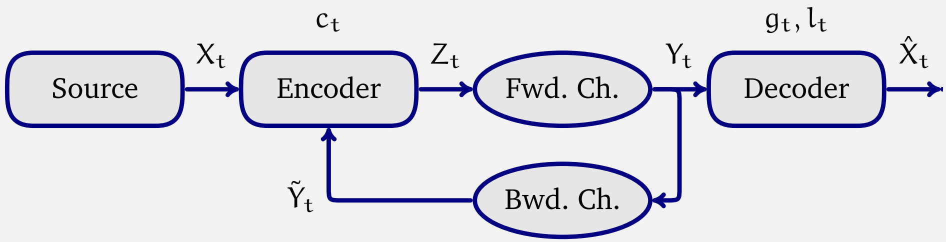
real-time communication

- *Types of problem*
  - (i) Source coding
  - (ii) Joint source-channel coding
- *Literature classification*
  - (i) Performance bounds
  - (ii) coding of individual sequences
  - (iii) coding of Markov sources

noisy feedback

- *Types of problem*
  - (i) Channel coding
- *Literature classification*
  - (i) Capacity of channels with memory and noisy feedback
  - (ii) Error exponents of memoryless channels with noisy feedback

Literature Overview



real-time communication with **noisy feedback**

real-time communication

- *Types of problem*
  - (i) Source coding
  - (ii) Joint source-channel coding
- *Literature classification*
  - (i) Performance bounds
  - (ii) coding of individual sequences
  - (iii) coding of Markov sources

noisy feedback

- *Types of problem*
  - (i) Channel coding
- *Literature classification*
  - (i) Capacity of channels with memory and noisy feedback
  - (ii) Error exponents of memoryless channels with noisy feedback

Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

$\Longrightarrow$

Do not know
how to apply
Information Theory

Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

$\Longrightarrow$

Do not know
how to apply
Information Theory

- *Use Stochastic Optimization*

Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

$\Longrightarrow$

Do not know
how to apply
Information Theory

- *Use Stochastic Optimization*

  Which solution methodology?

  - Markov Decision Theory
  - Orthogonal Search
  - Standard Form
  - ??

Stochastic Optimization

- *Markov Decision Theory* *(not applicable)*
  - ▷ works only for one decision maker with perfect recall

Stochastic Optimization

- Markov Decision Theory (not applicable)
  - works only for one decision maker with perfect recall
- *Orthogonal Search* *(not appropriate)*
  - May not converge
  - only guarantees local optima

Stochastic Optimization

- Markov Decision Theory (not applicable)
  - works only for one decision maker with perfect recall
- Orthogonal Search (not appropriate)
  - May not converge
  - only guarantees local optima
- *Standard Form — Witsenhausen, 1973* *(not appropriate)*
  - Does not extend to infinite horizon

Stochastic Optimization

- Markov Decision Theory (not applicable)
  - ▷ works only for one decision maker with perfect recall
- Orthogonal Search (not appropriate)
  - ▷ May not converge
  - ▷ only guarantees local optima
- Standard Form — Witsenhausen, 1973 (not appropriate)
  - ▷ Does not extend to infinite horizon
- *Our Methodology*

  (i) Identify structural properties

  (ii) Use structural properties to solve the global optimization problem

Our Methodology
(i) Structural Properties

Structural Properties

Need

- Encoder: $Z_t = c_t(X^t, Z^{t-1}, \tilde{Y}^{t-1}), \qquad c_t \in \mathcal{C}_t : \mathcal{X}^t \times \mathcal{Z}^{t-1} \times \tilde{\mathcal{Y}}^{t-1} \to \mathcal{Z}$
- Domain changing with time.
- *Makes the infinite horizon design hard*
- *Can we simplify implementation?*

Preliminaries

- Notion of Information
- Notion of Beliefs

Preliminaries

- Let $(X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)$ be defined on $(\Omega, \mathfrak{F}, P)$.

Preliminaries

- Let $(X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)$ be defined on $(\Omega, \mathfrak{F}, P)$.

Information	Everything that can be inferred from the data

Preliminaries

- Let $(X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)$ be defined on $(\Omega, \mathfrak{F}, P)$.

Information	Everything that can be inferred from the data

Information at Encoder

$$^1E_t := (X^t, Z^{t-1}, \tilde{Y}^{t-1}), \qquad ^1\mathfrak{E}_t := \sigma(^1E_t; {}^1\phi^{t-1}),$$
$$^2E_t := (X^t, Z^t, \tilde{Y}^{t-1}), \qquad ^2\mathfrak{E}_t := \sigma(^2E_t; {}^2\phi^{t-1}),$$
$$^3E_t := (X^t, Z^t, \tilde{Y}^t), \qquad ^3\mathfrak{E}_t := \sigma(^3E_t; {}^3\phi^{t-1}).$$

Information at Decoder

$$^1R_t := (M_{t-1}), \qquad ^1\mathfrak{R}_t := \sigma(^1R_t; {}^1\phi^{t-1}),$$
$$^2R_t := (Y_t, M_{t-1}), \qquad ^2\mathfrak{R}_t := \sigma(^2R_t; {}^2\phi^{t-1}),$$
$$^3R_t := (Y_t, M_{t-1}), \qquad ^3\mathfrak{R}_t := \sigma(^3R_t; {}^3\phi^{t-1}).$$

Preliminaries

- Let $(X_1, \dots, X_T, N_1, \dots, N_T, \tilde{N}_1, \dots, \tilde{N}_T)$ be defined on $(\Omega, \mathfrak{F}, P)$.

Information — Everything that can be inferred from the data

Information at Encoder

$$^1E_t := (X^t, Z^{t-1}, \tilde{Y}^{t-1}), \qquad {}^1\mathfrak{E}_t := \sigma({}^1E_t; {}^1\phi^{t-1}),$$
$$^2E_t := (X^t, Z^t, \tilde{Y}^{t-1}), \qquad {}^2\mathfrak{E}_t := \sigma({}^2E_t; {}^2\phi^{t-1}),$$
$$^3E_t := (X^t, Z^t, \tilde{Y}^t), \qquad {}^3\mathfrak{E}_t := \sigma({}^3E_t; {}^3\phi^{t-1}).$$

Information at Decoder

$$^1R_t := (M_{t-1}), \qquad {}^1\mathfrak{R}_t := \sigma({}^1R_t; {}^1\phi^{t-1}),$$
$$^2R_t := (Y_t, M_{t-1}), \qquad {}^2\mathfrak{R}_t := \sigma({}^2R_t; {}^2\phi^{t-1}),$$
$$^3R_t := (Y_t, M_{t-1}), \qquad {}^3\mathfrak{R}_t := \sigma({}^3R_t; {}^3\phi^{t-1}).$$

Beliefs — What one decision maker thinks about the data at other nodes

Preliminaries

- Let $(X_1, \dots, X_T, N_1, \dots, N_T, \tilde{N}_1, \dots, \tilde{N}_T)$ be defined on $(\Omega, \mathfrak{F}, P)$.

Information — Everything that can be inferred from the data

Information at Encoder

$$^1E_t := (X^t, Z^{t-1}, \tilde{Y}^{t-1}), \qquad ^1\mathfrak{E}_t := \sigma(^1E_t; {}^1\phi^{t-1}),$$
$$^2E_t := (X^t, Z^t, \tilde{Y}^{t-1}), \qquad ^2\mathfrak{E}_t := \sigma(^2E_t; {}^2\phi^{t-1}),$$
$$^3E_t := (X^t, Z^t, \tilde{Y}^t), \qquad ^3\mathfrak{E}_t := \sigma(^3E_t; {}^3\phi^{t-1}).$$

Information at Decoder

$$^1R_t := (M_{t-1}), \qquad ^1\mathfrak{R}_t := \sigma(^1R_t; {}^1\phi^{t-1}),$$
$$^2R_t := (Y_t, M_{t-1}), \qquad ^2\mathfrak{R}_t := \sigma(^2R_t; {}^2\phi^{t-1}),$$
$$^3R_t := (Y_t, M_{t-1}), \qquad ^3\mathfrak{R}_t := \sigma(^3R_t; {}^3\phi^{t-1}).$$

Beliefs — What one decision maker thinks about the data at other nodes

Belief of the encoder

$$^iB_t(^ir) := \Pr\left(^iR_t = {}^ir \,\middle|\, {}^i\mathfrak{E}_t\right).$$

Belief of the decoder

$$^iA_t(^ie) := \Pr\left(^iE_t = {}^ie \,\middle|\, {}^i\mathfrak{R}_t\right),$$

Structure of Opt Encoders

- *The main idea*
  - Fix the decoding and memory update function
  - Look at the problem from the encoder's point of view
  - Derive qualitative properties of optimal encoders

Structure of Opt Encoders

- *The main idea*
  - Fix the decoding and memory update function
  - Look at the problem from the encoder's point of view
  - Derive qualitative properties of optimal encoders

Structure of Optimal Encoders	$Z_t = c_t(X_t, {}^1B_t), \quad t = 2, \dots, T$

Structure of Opt Encoders

- *The main idea*
  - Fix the decoding and memory update function
  - Look at the problem from the encoder's point of view
  - Derive qualitative properties of optimal encoders

Structure of Optimal Encoders	$Z_t = c_t(X_t, {}^1B_t), \quad t = 2, \dots, T$

- We recover the structural results of previous models considered in literature
- *Systems where the encoder knows the decoder's information*
  - Real-time source coding (Witsenhausen, 1979)
  - Real-time joint source-channel coding with noiseless feedback (Walrand and Varaiya, 1982)
- *Systems where the encoder does not know the decoder's information*
  - Real-time joint source-channel coding with no feedback (Teneketzis, 2006)

Our Methodology

(ii) Global Optimization

Global Optimization

Information at Encoder

$$^{i}\mathfrak{E}_t = \sigma(^{i}E_t; {}^{i}\phi^{t-1})$$

Information at Decoder

$$^{i}\mathfrak{R}_t = \sigma(^{i}R_t; {}^{i}\phi^{t-1})$$

Global Optimization

Information at Encoder
${}^{i}\mathfrak{E}_t = \sigma({}^{i}E_t; {}^{i}\phi^{t-1})$

Information at Decoder
${}^{i}\mathfrak{R}_t = \sigma({}^{i}R_t; {}^{i}\phi^{t-1})$

- Information is non nested. ${}^{i}\mathfrak{E}_t \not\subseteq {}^{i}\mathfrak{R}_t \quad {}^{i}\mathfrak{E}_t \not\supseteq {}^{i}\mathfrak{R}_t$

Global Optimization

Information at Encoder
$^{i}\mathfrak{E}_t = \sigma(^{i}E_t; {^{i}\phi^{t-1}})$

Information at Decoder
$^{i}\mathfrak{R}_t = \sigma(^{i}R_t; {^{i}\phi^{t-1}})$

- Information is non nested. $^{i}\mathfrak{E}_t \not\subseteq {^{i}\mathfrak{R}_t}$ $\quad ^{i}\mathfrak{E}_t \not\supseteq {^{i}\mathfrak{R}_t}$

Aumann's notion on Common Knowledge	$^{i}\mathfrak{K}_t := {^{i}\mathfrak{E}_t} \cap {^{i}\mathfrak{R}_t}$

Global Optimization

Information at Encoder
${}^{i}\mathfrak{E}_t = \sigma({}^{i}E_t; {}^{i}\phi^{t-1})$

Information at Decoder
${}^{i}\mathfrak{R}_t = \sigma({}^{i}R_t; {}^{i}\phi^{t-1})$

- Information is non nested. ${}^{i}\mathfrak{E}_t \not\subseteq {}^{i}\mathfrak{R}_t$ ${}^{i}\mathfrak{E}_t \not\supseteq {}^{i}\mathfrak{R}_t$

Aumann's notion on Common Knowledge

$${}^{i}\mathfrak{K}_t \coloneqq {}^{i}\mathfrak{E}_t \cap {}^{i}\mathfrak{R}_t$$

- Choose future decision rules based on common knowledge.
- Or, since we are only interested in performance, choose future decision rules based on common belief: $P\big|_{{}^{i}\mathfrak{K}_t}$

Global Optimization

Information at Encoder
${}^i\mathfrak{E}_t = \sigma({}^iE_t; {}^i\phi^{t-1})$

Information at Decoder
${}^i\mathfrak{R}_t = \sigma({}^iR_t; {}^i\phi^{t-1})$

- Information is non nested. ${}^i\mathfrak{E}_t \not\subseteq {}^i\mathfrak{R}_t \quad {}^i\mathfrak{E}_t \not\supseteq {}^i\mathfrak{R}_t$

Aumann's notion on Common Knowledge

$${}^i\mathfrak{K}_t := {}^i\mathfrak{E}_t \cap {}^i\mathfrak{R}_t$$

- Choose future decision rules based on common knowledge.
- Or, since we are only interested in performance, choose future decision rules based on common belief: $P\Big|_{{}^i\mathfrak{K}_t}$

- Need to find the set of all feasible realizations of common information (or common belief)
- ${}^i\mathfrak{E}_t$ and ${}^i\mathfrak{R}_t$ depend on past decision rules — so does ${}^i\mathfrak{K}_t$.

Global Optimization

Aumann's notion on Common Knowledge	${}^{i}\mathfrak{K}_t := {}^{i}\mathfrak{E}_t \cap {}^{i}\mathfrak{R}_t$

- Can work with a super-set of the set of all feasible realizations of common knowledge

Global Optimization

Aumann's notion on Common Knowledge

$$^{i}\mathfrak{K}_t := {}^{i}\mathfrak{E}_t \cap {}^{i}\mathfrak{R}_t$$

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose Total Information: $^{i}\mathfrak{T}_t := \sigma(X_t, {}^{i}B_t, {}^{i}R_t; {}^{i}\phi_t) \supseteq {}^{i}\mathfrak{K}_t$
- Set of all realizations depends on the past decision rules

Global Optimization

Aumann's notion on Common Knowledge

$$ {}^{i}\mathfrak{K}_t := {}^{i}\mathfrak{E}_t \cap {}^{i}\mathfrak{R}_t $$

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose Total Information: ${}^{i}\mathfrak{T}_t := \sigma(X_t, {}^{i}B_t, {}^{i}R_t; {}^{i}\phi_t) \supseteq {}^{i}\mathfrak{K}_t$
- Set of all realizations depends on the past decision rules

The Image Space

- $(X_t, {}^{i}B_t, {}^{i}R_t) : (\Omega, \mathfrak{F}, P) \to (\mathcal{X} \times {}^{i}\mathcal{B} \times {}^{i}\mathcal{R},\ \mathbb{B}\left(\mathcal{X} \times {}^{i}\mathcal{B} \times {}^{i}\mathcal{R}\right),\ P')$

Global Optimization

Aumann's notion on Common Knowledge $\quad {}^i\mathfrak{K}_t \coloneqq {}^i\mathfrak{E}_t \cap {}^i\mathfrak{R}_t$

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose Total Information: $\quad {}^i\mathfrak{T}_t \coloneqq \sigma(X_t, {}^iB_t, {}^iR_t; {}^i\phi_t) \supseteq {}^i\mathfrak{K}_t$
- Set of all realizations depends on the past decision rules

The Image Space

- $(X_t, {}^iB_t, {}^iR_t) : (\Omega, \mathfrak{F}, P) \to (\mathcal{X} \times {}^i\mathcal{B} \times {}^i\mathcal{R},\ \mathbb{B}\left(\mathcal{X} \times {}^i\mathcal{B} \times {}^i\mathcal{R}\right),\ P')$

- Only the measure P' depends on ${}^i\phi^{t-1}$
- Hard to determine set of feasible realizations.

Global Optimization

Aumann's notion on Common Knowledge $\quad {}^{i}\mathfrak{K}_t := {}^{i}\mathfrak{E}_t \cap {}^{i}\mathfrak{R}_t$

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose Total Information: $\quad {}^{i}\mathfrak{T}_t := \sigma(X_t, {}^{i}B_t, {}^{i}R_t; {}^{i}\phi_t) \supseteq {}^{i}\mathfrak{K}_t$
- Set of all realizations depends on the past decision rules

The Image Space

- $(X_t, {}^{i}B_t, {}^{i}R_t) : (\Omega, \mathfrak{F}, P) \to (\mathcal{X} \times {}^{i}\mathcal{B} \times {}^{i}\mathcal{R},\ \mathbb{B}\left(\mathcal{X} \times {}^{i}\mathcal{B} \times {}^{i}\mathcal{R}\right),\ P')$

- Only the measure P' depends on ${}^{i}\phi^{t-1}$
- Hard to determine set of feasible realizations.

- Work with the set of all probability measures on

$$(\mathcal{X} \times {}^{i}\mathcal{B} \times {}^{i}\mathcal{R},\ \mathbb{B}\left(\mathcal{X} \times {}^{i}\mathcal{B} \times {}^{i}\mathcal{R}\right))$$

Global Optimization

Information States

$$^{1}\pi_t = \Pr\left(X_t, M_{t-1}, {}^{1}B_t\right),$$

$$^{2}\pi_t = \Pr\left(X_t, Y_t, M_{t-1}, {}^{2}B_t\right),$$

$$^{3}\pi_t = \Pr\left(X_t, Y_t, M_{t-1}, {}^{3}B_t\right).$$

Global Optimization

Information States

$$^{1}\pi_t = \Pr\left(X_t, M_{t-1}, {}^{1}B_t\right),$$
$$^{2}\pi_t = \Pr\left(X_t, Y_t, M_{t-1}, {}^{2}B_t\right),$$
$$^{3}\pi_t = \Pr\left(X_t, Y_t, M_{t-1}, {}^{3}B_t\right).$$

$\Longrightarrow$

A Dynamic System

$$^{2}\pi_t = {}^{1}Q(c_t)\, {}^{1}\pi_t,$$
$$^{3}\pi_t = {}^{2}Q\, {}^{2}\pi_t,$$
$$^{1}\pi_{t+1} = {}^{3}Q(l_t)\, {}^{3}\pi_t.$$

$$\mathbf{E}\left\{\rho(X_t, \hat{X}_t) \,\middle|\, c^t, g^t, l^{t-1}\right\} = {}^{2}\rho({}^{2}\pi_t, g_t)$$

Global Optimization

Information States

$$^1\pi_t = \Pr\left(X_t, M_{t-1}, {}^1B_t\right),$$
$$^2\pi_t = \Pr\left(X_t, Y_t, M_{t-1}, {}^2B_t\right),$$
$$^3\pi_t = \Pr\left(X_t, Y_t, M_{t-1}, {}^3B_t\right).$$

$\Longrightarrow$

A Dynamic System

$$^2\pi_t = {}^1Q(c_t)\,{}^1\pi_t,$$
$$^3\pi_t = {}^2Q\,{}^2\pi_t,$$
$$^1\pi_{t+1} = {}^3Q(l_t)\,{}^3\pi_t.$$

$$\mathbf{E}\left\{\rho(X_t, \hat{X}_t)\,\middle|\, c^t, g^t, l^{t-1}\right\} = {}^2\rho({}^2\pi_t, g_t)$$

An equivalent problem

Determine *meta-functions* $^1\Delta_t$, $^2\Delta_t$, and $^3\Delta_t$ such that:

$$c_t = {}^1\Delta_t({}^1\pi_t), \qquad {}^2\pi_t = {}^1Q(c_t)\,{}^1\pi_t,$$
$$g_t = {}^2\Delta_t({}^2\pi_t), \qquad {}^3\pi_t = {}^2Q\,{}^2\pi_t,$$
$$l_t = {}^3\Delta_t({}^3\pi_t), \qquad {}^1\pi_{t+1} = {}^3Q(l_t)\,{}^3\pi_t.$$

to minimize a total cost

$$\mathcal{J}_T(\Delta^T, {}^1\pi_1) = \sum_{t=1}^{T} {}^2\rho({}^2\pi_t, g_t).$$

Sequential Decomposition

$$
\begin{aligned}
{}^1V_t({}^1\pi) &= \inf_{c \in \mathcal{C}} {}^2V_t\left({}^1Q(c)\,{}^1\pi\right), \\
{}^2V_t({}^2\pi) &= \min_{g \in \mathcal{G}} {}^2\rho({}^2\pi, g) + {}^3V_t({}^2Q\,{}^2\pi), \\
{}^3V_t({}^3\pi) &= \min_{l \in \mathcal{L}} {}^1V_{t+1}\left({}^3Q(l)\,{}^3\pi\right).
\end{aligned}
$$

The arg min at each step determines an optimal meta-function

Sequential Decomposition

$$
\begin{aligned}
{}^1V_t({}^1\pi) &= \inf_{c \in \mathcal{C}} {}^2V_t\big({}^1Q(c)\,{}^1\pi\big), \\
{}^2V_t({}^2\pi) &= \min_{g \in \mathcal{G}} {}^2\rho({}^2\pi, g) + {}^3V_t({}^2Q\,{}^2\pi), \\
{}^3V_t({}^3\pi) &= \min_{l \in \mathcal{L}} {}^1V_{t+1}\big({}^3Q(l)\,{}^3\pi\big).
\end{aligned}
$$

The arg min at each step determines an optimal meta-function

- *Complexity*
  - Search complexity is linear in time (versus exponential in time for a brute force search)
  - Search complexity is exponential in the size of the alphabets

Extensions

- Infinite horizon
  - Expected discounted distortion
  - Average distortion per unit time
- Active feedback
- Channels with memory
- k-th order Markov source
- distortion with d-step delay

Conclusion

- An alternative approach to real-time communication
- Use stochastic optimization
  - (i) Derive structural results
  - (ii) Use structural results for global optimization
- A systematic search algorithm to determine optimal design

Future Directions

- Performance bounds
- Numerical algorithms
- Multi-terminal systems

Thank You