On the Influence of Channel Uncertainty on the Shannon Capacity

Pablo Piantanida*
piantanida@lss.supelec.fr.

Jointly work with Gerald Matz†, Pierre Duhamel* and Samson Lasaulce*

*Laboratoire des Signaux et Systèmes, CNRS/Supélec,
Gif-sur-Yvette, France
†Institute of Communications and Radio-Frequency Engineering,
Vienna University of Technology, Wien, Austria

This research was partially supported by NEWCOM network of excellence.
Introduction and Motivation

- The **Shannon’s capacity** is the maximum amount of information that can pass through a channel without error, where the transmitter and the receiver are assumed to have full knowledge of the channel states.

- **Channel uncertainty**, caused by **time variations/fading**, interference, or channel **estimation errors**, can severely impair the performance of wireless systems.

- **Motivation**: A system designer must ensure reliable communication with a quality of service (QoS), no matter which degree of accuracy estimation arises during a transmission.

- **Our contribution**: A new notion: Estimation-induced outage capacity

- **Applications**: To compute information rates of realistic wireless systems under channel estimation errors, and in the context of cellular coverage.
Time-varying Discrete Memoryless Channels (DMCs)

DMC \( W(y|x, \theta) \in \mathcal{W}_\Theta = \{ W(\cdot|x, \theta) : \theta \in \Theta \} \) is a family of conditional probabilities on \( \mathcal{Y} \), parameterized by a vector \( \theta \in \Theta \), unknown channel state.

The channel state \( \theta \) is constant within blocks of duration \( T \) symbol periods coherence time, and these states are i.i.d. \( \theta \sim \psi(\theta) \).

CSIR: The transmitter sends training sequences \( x_T \) that allow the receiver to obtain an estimate \( \hat{\theta}_R \) of the channel state.

CSIT: The transmitter is provided of \( \hat{\theta}_T \) via a feedback channel.

Resulting a joint model \((\hat{\theta}_T, \hat{\theta}_R, \theta)\) i.i.d. \( \sim \psi(\hat{\theta}_T, \hat{\theta}_R, \theta) \).
Problem Statement

- For a given channel estimate $\hat{\theta} = (\hat{\theta}_T, \hat{\theta}_R)$
- We chose a message $m$ from the set $M = \{1, \ldots, [\exp(nR)]\}$
- A length-$n$ block code defined as a pair $(\varphi, \phi)$ of mappings:
  - Encoder $\varphi : M \times \Theta \mapsto X^n$ that utilizes $\hat{\theta}_T$
  - Decoder $\phi : Y^n \times \Theta \mapsto M \cup \{0\}$ that utilizes $\hat{\theta}_R$
- Maximum (over all messages) error probability

$$
\epsilon_{\text{max}}(\varphi, \phi, \hat{\theta}; \theta) = \max_{m \in M} \sum_{y \in Y^n : \phi(y, \hat{\theta}_R) \neq m} W^n(y | \varphi(m, \hat{\theta}_T), \theta)
$$

Note that actually this error probability is a random variable!

- **Capacity**: A rate $R \geq 0$ is called an $\epsilon$-achievable rate if there exists a sequence of length-$n$ block codes with error probability less than $\epsilon$. Then the $\epsilon$-capacity $C_\epsilon$ is defined as the largest $\epsilon$-achievable rate, and the capacity $C = \lim_{\epsilon \downarrow 0} C_\epsilon$. 
Performance Limits of Wireless Systems

- The pdf \((\psi(\theta), \mathcal{W}_\Theta)\) governing the channel variations are unknown

- Mismatched decoders: the decoder is restricted to be a metric, which is not necessarily matched to the channel (cf. Fischer 1971, Merhav et. al. 1994, Csiszar et. al. 1995, Lapidoth and Shamai 2002)

- The pdf \((\psi(\theta), \mathcal{W}_\Theta)\) governing the channel variations are perfectly known

- Perfect CSI at the receiver and imperfect at the transmitter: power allocation strategies can be employed (cf. Caire and Shamai 1999)


- Focus \(\implies\) Imperfect CSI at the transmitter and at the receiver: the resulting capacity will rely crucially on the error probability criteria adopted. Medard 2000 derives capacity bounds for AWGN channels with MMSE channel estimation and no CSIT
Can the capacity be defined in terms of the estimator performance?

- Characterization of the estimator performance in terms of the conditional pdf \( \psi(\theta|\hat{\theta}) \) (estimator function, SNR, Rice K-factor, training sequence length-\( N \), etc.)

- The average of the error probability over all channel estimation errors:
  \[
  E_{\theta|\hat{\theta}} \{ e_{\max}(\varphi, \phi, \hat{\theta}; \theta) | \hat{\theta} = \hat{\theta}_0 \} \leq \epsilon
  \]

- The error probabilities larger than \( \epsilon \) occur with probability less than \( \gamma \)
  \[
  \Pr_{\theta|\hat{\theta}} \left( e_{\max}(\varphi, \phi, \hat{\theta}; \theta) \leq \epsilon | \hat{\theta} = \hat{\theta}_0 \right) \geq 1 - \gamma, \text{ robust to estimation errors!}
  \]
The capacity of an unknown DMC defined for the **average of error probability** criteria is given by

\[ C(\hat{\theta}) = \max_{P \in \mathcal{P}(\mathcal{X})} I(P, \tilde{W}(\cdot|\cdot, \hat{\theta})) \]

and

\[ \tilde{W}(y|x, \hat{\theta}) = \int_{\Theta} W(y|x, \theta)\psi(\theta|\hat{\theta})d\theta \]

This should be a **reasonably criteria** when the receiver can estimate the **channel accurately** (e.g. when the coherence time is sufficiently long \(\Rightarrow\) Medard 2000), **but it is not an adequate** criteria when a **reliable estimation of the fading coefficients is not available**

- The robust criteria \(\Rightarrow\) Estimation-induced outage capacity
Notion of Estimation-induced Outage Capacity

- For a given channel estimate $\hat{\theta} = \hat{\theta}_0$, and outage probability $\gamma$

- Any block code of random rate $\frac{1}{n} \log M_{\theta,\hat{\theta}}$ depends on the unknown channel realization $\theta$ through its probability of error.

- An outage rate $R \geq 0$ is $(\epsilon, \gamma)$-achievable if for every $\delta > 0$ and every sufficiently large $n$ there exists a sequence of length-$n$ block codes such that the rate satisfies

$$\Pr \left( \{ \theta \in \Lambda_\epsilon : n^{-1} \log M_{\theta,\hat{\theta}} \geq R - \delta \} \mid \hat{\theta} = \hat{\theta}_0 \right) \geq 1 - \gamma,$$

where $\Lambda_\epsilon = \{ \theta \in \Theta : e_{\max}(\varphi, \phi, \hat{\theta}; \theta) \leq \epsilon \}$ is the set of all channel states allowing for reliable decoding.

- This definition requires that maximum error probabilities larger than $\epsilon$ occur with probability less than $\gamma$. 
Coding Theorem

**Theorem:** Given a channel estimate \( \hat{\theta} \), and an outage probability \( 0 \leq \gamma < 1 \) the estimation-induced outage capacity is given by

\[
C(\gamma, \hat{\theta}) = \max_{P(\cdot | \hat{\theta}_T) \in \mathcal{P}(\mathcal{X})} \mathcal{C}(\gamma, \hat{\theta}, P),
\]

where

\[
\mathcal{C}(\gamma, \hat{\theta}, P) = \sup_{\Lambda \subseteq \Theta: \Pr(\Lambda | \hat{\theta}) \geq 1 - \gamma} \inf_{\theta \in \Lambda} I(P, W(\cdot | \cdot, \theta))
\]

The mutual information,

\[
I(P, W(\cdot | \cdot, \theta)) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x)W(y|x, \theta) \log \frac{W(y|x, \theta)}{Q(y|\theta)}
\]

with \( Q(y|\theta) = \sum_{x \in \mathcal{X}} P(x)W(y|x, \theta) \)
Remarks:

- The supremum in the capacity expression is taken over all subsets $\Lambda$ of $\Theta$ that have (conditional) probability at least $1 - \gamma$

- Explicit way to evaluate the tradeoff of maximal outage rate vs. outage probability $\gamma$, for an unknown but estimated channel for arbitrary estimation accuracies without additional assumptions

- This capacity is seen to equal the maximum capacity of all compound channels that are contained in $\mathcal{W}_\Theta$ and, conditioned on $\hat{\theta}$, have sufficiently high probability

- We will consider the average over a large number of channel estimates $\hat{\theta}$ (coherence intervals)

$$\bar{C}(\gamma) = E_{\hat{\theta}} \{ C(\gamma, \hat{\theta}) \}$$

which describes average outage rates respect to the joint distribution $\psi(\hat{\theta}) = \psi(\hat{\theta}_T, \hat{\theta}_R)$
Sketch of Proof

This proof is based on an extension of the maximal code lemma to bound the minimum random size of the images for the considered channels, according to the notion of estimation-induced outage capacity.

**Generalized Maximal Code Lemma:** Let $\mathcal{I}_\gamma$ denote the set of all common $\eta$-images $\mathcal{B}^n \subseteq \mathcal{Y}^n$ from the collection of DMCs having high probability,

$$\mathcal{I}_\gamma(\mathcal{A}^n, \hat{\theta}, \eta) = \{ \mathcal{B}^n : \Pr(W^n(\mathcal{B}^n|x, \theta) \geq \eta|\hat{\theta}) \geq 1 - \gamma \}$$

for all $x \in \mathcal{A}^n$. The minimum of the cardinalities of all common $\eta$-images $\mathcal{B}^n$ is

$$g_\gamma(\mathcal{A}^n, \hat{\theta}, \eta) = \min_{\mathcal{B}^n \in \mathcal{I}_\gamma(\mathcal{A}^n, \hat{\theta}, \eta)} \|\mathcal{B}^n\|$$

This extension is based on the notion of robust $l$-typical sets.
A code \((x_1(\hat{\theta}_T), \ldots, x_M(\hat{\theta}_T); D_1^n(\hat{\theta}), \ldots, D_M^n(\hat{\theta}))\) according to our definition consists of a set of codewords \(x_m(\hat{\theta}_T)\) and associated decoding sets \(D_m^n(\hat{\theta})\) (i.e., the decoder reads \(\phi(y, \hat{\theta}) = m\) iff \(y \in D_m^n(\hat{\theta})\)).

For any set \(\mathcal{A}^n\), we call a code admissible if \(x_m(\hat{\theta}_T) \in \mathcal{A}^n\), all decoding sets \(D_m^n(\hat{\theta}) \subseteq \mathcal{Y}^n\) are mutually disjoint, and the set

\[
\Lambda_\varepsilon = \left\{ \theta \in \Theta : \max_{m \in M} W^n((D_m^n(\hat{\theta}))^c | x_m(\hat{\theta}_T), \theta) \leq \varepsilon \right\},
\]

satisfies that \(\Pr(\Lambda_\varepsilon | \hat{\theta}) \geq 1 - \gamma\).

The proof of the Theorem is obtained by bounding the codebook size \(M_{\theta, \hat{\theta}}\) and the cardinalities \(g_{\gamma}(\mathcal{A}^n, \hat{\theta}, \eta)\) from above and below.
An Application Example: Ricean Channels

We consider a single user, narrowband block flat-fading communication model

\[ Y[i] = H X[i] + Z[i], \]

\[ Y[i] \] is the discrete-time received signal,
\[ X[i] \] denotes the transmit signal with \( E\{ |X[i]|^2 \} \leq P(\hat{\theta}_T) \),
\[ H = \theta \sim \psi(\theta) = \mathcal{CN}(\mu_h, 2\sigma^2_h) \] is the fading coefficient, Rice factor \( K_h = \frac{|\mu_h|^2}{2\sigma^2_h} \),
\[ Z[i] \sim \mathcal{CN}(0, \sigma_Z^2) \] is the iid additive noise, (noise and fading coefficient are statistically independent), and the average power constraint \( E_{\hat{\theta}_T}\{\mathcal{P}(\hat{\theta}_T)\} \leq \bar{P} \)

The actual codeword is preceded by a length-\( N \) training sequence \( x_T = [x_0, \ldots, x_{N-1}] \) known by the receiver. This enables maximum-likelihood (ML) estimation of the fading coefficient \( \theta \) at the receiver \( \hat{\theta}_R = \hat{H} \)
Estimation-induced Outage Capacity of Ricean Channels

The ML estimate of $\theta = H$ is given by $\hat{\theta} = H + W$, where $W \sim \mathcal{CN}(0, \sigma_W^2)$ and the estimation error $\sigma_W^2 = \sigma_Z^2/(NP_T)$.

Using this relation and the channel’s a priori distribution $\psi(\theta)$ we compute the a posteriori distribution $\psi(\theta|\hat{\theta}) \sim \mathcal{CN}(\tilde{\mu}(\hat{\theta}), 2\tilde{\sigma}^2)$, where

$$\tilde{\mu}(\hat{\theta}) = \rho \mu_h + (1 - \rho)\hat{\theta}, \quad \text{with} \quad \rho = \frac{\sigma_W^2}{\sigma_W^2 + 2\sigma_h^2} \quad \text{and} \quad \tilde{\sigma}^2 = \rho \sigma_h^2$$

To evaluate the capacity we have to determine the optimum set $\Lambda^*$, and the associated channel state $\theta^* \in \Lambda^*$ minimizing mutual information.

Mutual information is a monotone and increasing function in $r = |\theta|$. Then the problem now reduces to finding the optimum positive real interval $\tilde{\Lambda}^* = [r^*, \infty[$ having probability $1 - \gamma$. Resulting

$$C(\gamma, \hat{\theta}_0) = \log_2 \left(1 + \frac{(r^*(\gamma, \hat{\theta}_0))^2 P(\hat{\theta}_T, 0)}{\sigma_Z^2}\right)$$
Considered Scenario

- We consider the following scenario:
  - No feedback channel is available (no CSIT), we compare our results with the capacity of a system where the receiver uses a mismatched ML decoder based on $\hat{\theta}_R$
  - Instantaneous and error-free feedback channel is available ($\hat{\theta}_T = \hat{\theta}_R$)

- We derive optimal transmitter power allocation strategies that achieve the mean outage capacity

- By simulating this model, we evaluate the impact of the channel estimate and the channel characteristics (SNR, Ricean K-factor, training sequence length, feedback, etc.) on the mean outage capacity
Simulation Results

Average of estimation-induced outage capacity and the mismatched ML decoding capacity vs SNR, for various outage probabilities
Average estimation-induced outage capacity for different amounts of training, without feedback (no CSIT) and with perfect feedback (CSIT=CSIR) vs. SNR
Summary

- We have studied the problem of reliable communications over unknown DMCs when the receiver and transmitter only know an estimate of the channel state.

- We proposed to characterize the information theoretic limits of such scenarios in terms of the novel notion of estimation-induced outage capacity.

- Our model is relevant e.g. for communication systems where a quality of service (QoS) in terms of error performance must be ensured although significant channel variations occur due to user mobility.

- In the context of cellular coverage, where the average rate would characterize performance over multiple communication sessions in a large number of geographic locations. A system designer must ensure a quality service, i.e. reliable communication over \( (1 - \gamma) \)-percent of users, no matter which degree of accuracy estimation arises during the connection session.

- We used a Ricean fading channel model to illustrate our approach. Results indicate that the mismatched ML decoder can be largely suboptimal for the considered class of channels.