Second-Order Asymptotics for Erasure and List Decoding

Vincent Y. F. Tan

Department of Electrical and Computer Engineering,
Department of Mathematics,
National University of Singapore

Joint work with Pierre Moulin (UIUC)

July 3, 2014
Consider codes for discrete memoryless channels (DMCs) with fixed error $\epsilon \in (0, 1)$.
Consider codes for discrete memoryless channels (DMCs) with fixed error $\epsilon \in (0, 1)$

Strassen (1962) showed that for well-behaved DMCs that the maximum number of codewords for channel $W^n$ and error $\epsilon$ satisfies

$$\log M^*(W^n, \epsilon) = nC + \sqrt{nV} \Phi^{-1}(\epsilon) + O(\log n),$$

where $C$ and $V$ are the capacity and dispersion of the DMC $W$ resp.
Consider codes for discrete memoryless channels (DMCs) with fixed error \( \epsilon \in (0, 1) \).

Strassen (1962) showed that for well-behaved DMCs that the maximum number of codewords for channel \( W^n \) and error \( \epsilon \) satisfies

\[
\log M^*(W^n, \epsilon) = nC + \sqrt{nV}\Phi^{-1}(\epsilon) + O(\log n),
\]

where \( C \) and \( V \) are the capacity and dispersion of the DMC \( W \) resp.

What if we allow the decoder to (i) declare an erasure event or (ii) output a list of messages?
The decoding regions $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M$ are disjoint but do not fill up the whole space and

$$\mathcal{D}_0 := \mathcal{Y}^n \setminus \bigcup_{m=1}^{M} \mathcal{D}_m$$

is the subset of channel outputs leading to an erasure.
Forney (1968) derived exponential error bounds for erasure and list decoding using Gallager-type bounding techniques

Telatar (1994) considered decoders parametrized by an asymmetric relation $\succ$.
Forney (1968) derived exponential error bounds for erasure and list decoding using Gallager-type bounding techniques.

Telatar (1994) considered decoders parametrized by an asymmetric relation $\succ$.

Forney (1968) derived exponential error bounds for erasure and list decoding using Gallager-type bounding techniques.

Telatar (1994) considered decoders parametrized by an asymmetric relation $\succ$.


Merhav (2008) derived error exponents based on the type class enumerator method that are at least as good as Forney’s.

Somekh-Baruch and Merhav (2011) showed that the type class enumerator method is ensemble-tight.
Forney (1968) derived exponential error bounds for erasure and list decoding using Gallager-type bounding techniques.

Telatar (1994) considered decoders parametrized by an asymmetric relation $\succ$.


Merhav (2008) derived error exponents based on the type class enumerator method that are at least as good as Forney’s.

Somekh-Baruch and Merhav (2011) showed that the type class enumerator method is ensemble-tight.

Instead of exponents, we consider the fixed errors version of erasure decoding.
Basic Definitions: Code

An \textit{M-erasure code} for channel $W : \mathcal{X} \to \mathcal{Y}$ is a pair of maps $(f, \varphi)$ such that

$$f : [M] \to \mathcal{X}, \quad \text{and} \quad \varphi : \mathcal{Y} \to [M] \cup \{0\}$$
Basic Definitions: Code

- An $M$-erasure code for channel $W : \mathcal{X} \to \mathcal{Y}$ is a pair of maps $(f, \varphi)$ such that
  \[ f : [M] \to \mathcal{X}, \quad \text{and} \quad \varphi : \mathcal{Y} \to [M] \cup \{0\} \]

- Decoding regions $D_m = \varphi^{-1}(m)$ are disjoint
Basic Definitions: Code

- An $M$-erasure code for channel $W : \mathcal{X} \rightarrow \mathcal{Y}$ is a pair of maps $(f, \varphi)$ such that

$$f : [M] \rightarrow \mathcal{X}, \quad \text{and} \quad \varphi : \mathcal{Y} \rightarrow [M] \cup \{0\}$$

- Decoding regions $\mathcal{D}_m = \varphi^{-1}(m)$ are disjoint

- The conditional undetected error is

$$\lambda_u(m) := \sum_{m' \in [M] \setminus \{m\}} W(\mathcal{D}_{m'} | f(m))$$
Basic Definitions: Code

- An $M$-erasure code for channel $W : \mathcal{X} \rightarrow \mathcal{Y}$ is a pair of maps $(f, \varphi)$ such that
  \[ f : [M] \rightarrow \mathcal{X}, \quad \text{and} \quad \varphi : \mathcal{Y} \rightarrow [M] \cup \{0\} \]

- Decoding regions $D_m = \varphi^{-1}(m)$ are disjoint

- The conditional undetected error is
  \[ \lambda_u(m) := \sum_{m' \in [M] \setminus \{m\}} W(D_{m'} | f(m)) \]

- The conditional erasure error is
  \[ \lambda_e(m) := W(D_0 | f(m)) \]
An \( M \)-erasure code for channel \( W : \mathcal{X} \to \mathcal{Y} \) is a pair of maps \((f, \varphi)\) such that
\[
f : [M] \to \mathcal{X}, \quad \text{and} \quad \varphi : \mathcal{Y} \to [M] \cup \{0\}
\]
Decoding regions \( D_m = \varphi^{-1}(m) \) are disjoint.

The conditional undetected error is
\[
\lambda_u(m) := \sum_{m' \in [M] \setminus \{m\}} W(D_{m'} | f(m))
\]

The conditional erasure error is
\[
\lambda_e(m) := W(D_0 | f(m))
\]

The conditional total error is
\[
\lambda_t(m) := \lambda_u(m) + \lambda_e(m).
\]
An \((M, \epsilon_u, \epsilon_t)\)-erasure code for \(W\) is an \(M\)-erasure code where

\[
\frac{1}{M} \sum_{m \in [M]} \lambda_u(m) \leq \epsilon_u, \quad \text{and} \quad \frac{1}{M} \sum_{m \in [M]} \lambda_t(m) \leq \epsilon_t.
\]
An \((M, \epsilon_u, \epsilon_t)\)-erasure code for \(W\) is an \(M\)-erasure code where

\[
\frac{1}{M} \sum_{m \in [M]} \lambda_u(m) \leq \epsilon_u, \quad \text{and} \quad \frac{1}{M} \sum_{m \in [M]} \lambda_t(m) \leq \epsilon_t.
\]

One can also make the above definition under the maximum error formalism.
**Basic Definitions: Code with Errors**

- An \((M, \epsilon_u, \epsilon_t)\)-erasure code for \(W\) is an \(M\)-erasure code where

\[
\frac{1}{M} \sum_{m \in [M]} \lambda_u(m) \leq \epsilon_u, \quad \text{and} \quad \frac{1}{M} \sum_{m \in [M]} \lambda_t(m) \leq \epsilon_t.
\]

- One can also make the above definition under the maximum error formalism.

- Results do not change.
An \((M, \epsilon_u, \epsilon_t)\)-erasure code for \(W\) is an \(M\)-erasure code where

\[
\frac{1}{M} \sum_{m \in [M]} \lambda_u(m) \leq \epsilon_u, \quad \text{and} \quad \frac{1}{M} \sum_{m \in [M]} \lambda_t(m) \leq \epsilon_t.
\]

One can also make the above definition under the maximum error formalism.

Results do not change.

Notice that we define the code in terms of undetected and total errors.
Basic Definitions: Code with Errors

- An \((M, \epsilon_u, \epsilon_t)\)-erasure code for \(W\) is an \(M\)-erasure code where

\[
\frac{1}{M} \sum_{m \in [M]} \lambda_u(m) \leq \epsilon_u, \quad \text{and} \quad \frac{1}{M} \sum_{m \in [M]} \lambda_t(m) \leq \epsilon_t.
\]

- One can also make the above definition under the maximum error formalism

- Results do not change

- Notice that we define the code in terms of undetected and total errors

- These correspond to error events \(\mathcal{E}_2\) and \(\mathcal{E}_1\) resp. in Forney and Merhav’s papers
Definition of Second-Order Coding Rate

- A number \( r \in \mathbb{R} \) is \((\epsilon_u, \epsilon_t)\)-achievable for the channel \( W^n \) if there exists a sequence of \((M_n, \epsilon_{u,n}, \epsilon_{t,n})\) erasure codes such that

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n}}(\log M_n - nC) \geq r
\]

and

\[
\limsup_{n \to \infty} \epsilon_{u,n} \leq \epsilon_u, \quad \text{and} \quad \limsup_{n \to \infty} \epsilon_{t,n} \leq \epsilon_t
\]
A number \( r \in \mathbb{R} \) is \((\epsilon_u, \epsilon_t)\)-achievable for the channel \( W^n \) if there exists a sequence of \((M_n, \epsilon_{u,n}, \epsilon_{t,n})\) erasure codes such that

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n}} (\log M_n - nC) \geq r
\]

and

\[
\limsup_{n \to \infty} \epsilon_{u,n} \leq \epsilon_u, \quad \text{and} \quad \limsup_{n \to \infty} \epsilon_{t,n} \leq \epsilon_t
\]

The \((\epsilon_u, \epsilon_t)\)-optimum second-order coding rate \( r^*(\epsilon_u, \epsilon_t; W) \) is the supremum of all \( r \in \mathbb{R} \) that are \((\epsilon_u, \epsilon_t)\)-achievable for \( W^n \).
Definition of Second-Order Coding Rate

- A number $r \in \mathbb{R}$ is $(\epsilon_u, \epsilon_t)$-achievable for the channel $W^n$ if there exists a sequence of $(M_n, \epsilon_{u,n}, \epsilon_{t,n})$ erasure codes such that

$$\liminf_{n \to \infty} \frac{1}{\sqrt{n}} (\log M_n - nC) \geq r$$

and

$$\limsup_{n \to \infty} \epsilon_{u,n} \leq \epsilon_u, \quad \text{and} \quad \limsup_{n \to \infty} \epsilon_{t,n} \leq \epsilon_t$$

- The $(\epsilon_u, \epsilon_t)$-optimum second-order coding rate $r^*(\epsilon_u, \epsilon_t; W)$ is the supremum of all $r \in \mathbb{R}$ that are $(\epsilon_u, \epsilon_t)$-achievable for $W^n$

- Condition on code size means that there exists a sequence of $(M_n, \epsilon_{u,n}, \epsilon_{t,n})$-codes such that

$$\log M_n \approx nC + \sqrt{n} r^*(\epsilon_u, \epsilon_t; W), \quad \epsilon_{u,n} \leq \epsilon_u + o(1), \quad \epsilon_{t,n} \leq \epsilon_t + o(1)$$
Result for Second-Order Coding Rate

Recall the definition of the conditional information variance

\[ V(P, W) = \sum_x P(x) \sum_y W(y|x) \left[ \log \frac{W(y|x)}{PW(y)} - D(W(\cdot|x)\|PW) \right]^2 \]

Let \( \pi := \{P : I(P, W) = C\} \). Also recall

\[ V_\epsilon(W) = \begin{cases} V_{\min}(W) := \min_{P \in \pi} V(P, W) & \text{if } \epsilon < 0.5 \\ V_{\max}(W) := \max_{P \in \pi} V(P, W) & \text{if } \epsilon \geq 0.5 \end{cases} \]
Recall the definition of the conditional information variance

\[ V(P, W) = \sum_x P(x) \sum_y W(y|x) \left[ \log \frac{W(y|x)}{PW(y)} - D(W(\cdot|x)\|PW) \right]^2 \]

Let \( \Pi := \{ P : I(P, W) = C \} \). Also recall

\[ V_\epsilon(W) = \begin{cases} 
V_{\text{min}}(W) := \min_{P \in \Pi} V(P, W) & \text{if } \epsilon < 0.5 \\
V_{\text{max}}(W) := \max_{P \in \Pi} V(P, W) & \text{if } \epsilon \geq 0.5 
\end{cases} \]

**Theorem (Tan-Moulin (2014))**

*For DMCs with* \( V_{\text{min}}(W) > 0 \), *and* \( 0 \leq \epsilon_u < \epsilon_t < 1 \), *we have*

\[ r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t). \]
Result for Second-Order Coding Rate

Recall the definition of the conditional information variance

\[ V(P, W) = \sum_x P(x) \sum_y W(y|x) \left[ \log \frac{W(y|x)}{PW(y)} - D(W(\cdot|x)\|PW) \right]^2 \]

Let \( \Pi := \{ P : I(P, W) = C \} \). Also recall

\[ V_\epsilon(W) = \left\{ \begin{array}{ll} V_{\min}(W) := \min_{P \in \Pi} V(P, W) & \text{if } \epsilon < 0.5 \\ V_{\max}(W) := \max_{P \in \Pi} V(P, W) & \text{if } \epsilon \geq 0.5 \end{array} \right. \]

Theorem (Tan-Moulin (2014))

For DMCs with \( V_{\min}(W) > 0 \), and \( 0 \leq \epsilon_u < \epsilon_t < 1 \), we have

\[ r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t). \]

Note that \( r^*(\epsilon_u, \epsilon_t; W) \) is independent of \( \epsilon_u \)

Vincent Y. F. Tan (NUS)
Remarks on Second-Order Coding Rate

- Direct part does not use Forney’s (1968) optimum decoding rule

\[ D_m^* := \left\{ y^n : \frac{W^n(y^n | f(m))}{\sum_{m' \neq m} W^n(y^n | f(m'))} \geq \xi \right\}, \quad \xi > 1, \]

which is based on the Neyman-Pearson lemma
Remarks on Second-Order Coding Rate

- Direct part does not use Forney's (1968) optimum decoding rule

$$D_m^* := \left\{ y^n : \frac{W^n(y^n | f(m))}{\sum_{m' \neq m} W^n(y^n | f(m'))} \geq \xi \right\}, \quad \xi > 1,$$

which is based on the Neyman-Pearson lemma.

- Forney (1968) also suggested the simpler but suboptimal rule

$$\tilde{D}_m := \left\{ y^n : W^n(y^n | f(m)) \geq \xi \max_{m' \neq m} W^n(y^n | f(m')) \right\}$$

- Instead we use thresholding of empirical mutual information (details to follow)
Remarks on Second-Order Coding Rate

Direct part does not use Forney’s (1968) optimum decoding rule

\[
\mathcal{D}_m^* := \left\{ y^n : \frac{W^n(y^n | f(m))}{\sum_{m' \neq m} W^n(y^n | f(m'))} \geq \xi \right\}, \quad \xi > 1,
\]

which is based on the Neyman-Pearson lemma

Forney (1968) also suggested the simpler but suboptimal rule

\[
\tilde{\mathcal{D}}_m := \left\{ y^n : W^n(y^n | f(m)) \geq \xi \max_{m' \neq m} W^n(y^n | f(m')) \right\}
\]

Instead we use thresholding of empirical mutual information (details to follow)

The rules \(\mathcal{D}_m^*\) and \(\tilde{\mathcal{D}}_m\) seem hard to analyze for fixed error asymptotics
Expected Rate

- Erasure probability is $\epsilon_e := \epsilon_t - \epsilon_u$
Erasure probability is $\epsilon_e := \epsilon_t - \epsilon_u$

Thus expected rate is approximately

$$\mathbb{E} \left[ R_e^{(n)} \right] = (1 - \epsilon_e) \left[ C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon_t) \right]$$

assuming Gaussian approximation is sufficiently accurate [Polyanskiy-Poor-Verdú (2010)]
Expected Rate

- Erasure probability is $\epsilon_e := \epsilon_t - \epsilon_u$

- Thus expected rate is approximately

$$
\mathbb{E} \left[ R_e^{(n)} \right] = (1 - \epsilon_e) \left[ C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon_t) \right]
$$

assuming Gaussian approximation is sufficiently accurate [Polyanskiy-Poor-Verdú (2010)]

- Compare this rate to

$$
R_c^{(n)} = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon_e)
$$
Erasure probability is $\epsilon_e := \epsilon_t - \epsilon_u$

Thus expected rate is approximately

$$\mathbb{E} \left[ R_e^{(n)} \right] = (1 - \epsilon_e) \left[ C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon_t) \right]$$

assuming Gaussian approximation is sufficiently accurate [Polyanskiy-Poor-Verdú (2010)]

Compare this rate to

$$R_c^{(n)} = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon_e)$$

Obviously

$$\lim_{n \to \infty} \mathbb{E} \left[ R_e^{(n)} \right] = (1 - \epsilon_e)C < \lim_{n \to \infty} R_c^{(n)} = C$$

but what about at moderate blocklengths?
Numerical Example for BSC(0.11) and $\epsilon_u = 10^{-6}$

Finite blocklength bounds (DT, Meta-converse) included
Proof Idea: Converse

Recall that we want to prove that

\[ r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon t}} \Phi^{-1}(\epsilon_t). \]
Proof Idea: Converse

- Recall that we want to prove that
  \[ r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t). \]

- Converse is straightforward
Proof Idea: Converse

• Recall that we want to prove that
  \[ r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon_t} \Phi^{-1}(\epsilon_t)}. \]

• Converse is straightforward

• Every \((M, \epsilon_u, \epsilon_t)\)-code for \(W^n\) can be transformed into an \((M, \epsilon_t)\)-code for \(W^n\) (usual channel coding), i.e.,
  \[ M^*(W^n, \epsilon_u, \epsilon_t) \leq M^*(W^n, \epsilon_t) \]
Proof Idea: Converse

- Recall that we want to prove that
  \[ r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t). \]

- Converse is straightforward

- Every \((M, \epsilon_u, \epsilon_t)\)-code for \(W^n\) can be transformed into an \((M, \epsilon_t)\)-code for \(W^n\) (usual channel coding), i.e.,
  \[ M^*(W^n, \epsilon_u, \epsilon_t) \leq M^*(W^n, \epsilon_t) \]

- One can also observe that
  \[ \frac{1}{M} \sum_{m \in [M]} W^n(D_m | f(m)) \geq 1 - \epsilon_t \]
Proof Idea: Converse

- Recall that we want to prove that
  \[
  r^*(\epsilon_u, \epsilon_t; W) = \sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t).
  \]

- Converse is straightforward

- Every \((M, \epsilon_u, \epsilon_t)\)-code for \(W^n\) can be transformed into an \((M, \epsilon_t)\)-code for \(W^n\) (usual channel coding), i.e.,
  \[
  M^*(W^n, \epsilon_u, \epsilon_t) \leq M^*(W^n, \epsilon_t)
  \]

- One can also observe that
  \[
  \frac{1}{M} \sum_{m \in [M]} W^n(D_m | f(m)) \geq 1 - \epsilon_t
  \]

- By Strassen’s result, we are done since
  \[
  \log M^*(W^n, \epsilon_t) \leq nC + \sqrt{nV_{\epsilon_t}} \Phi^{-1}(\epsilon_t) + O(\log n)
  \]
Proof Idea: Achievability

Key realization: It suffices to prove that $\sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t)$ is a $(0, \epsilon_t)$-achievable second-order coding rate, i.e.,

$$\epsilon_u = 0, \quad \text{and} \quad \epsilon_e = \epsilon_t.$$
Proof Idea: Achievability

- Key realization: It suffices to prove that $\sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t)$ is a $(0, \epsilon_t)$-achievable second-order coding rate, i.e.,

$$\epsilon_u = 0, \quad \text{and} \quad \epsilon_e = \epsilon_t.$$

- Any achievable $(0, \epsilon_t)$-rate is also $(\epsilon_u, \epsilon_t)$-achievable for all $0 \leq \epsilon_u < \epsilon_t < 1$
Proof Idea: Achievability

Key realization: It suffices to prove that $\sqrt{V_{\epsilon_t}} \Phi^{-1}(\epsilon_t)$ is a $(0, \epsilon_t)$-achievable second-order coding rate, i.e.,

$$\epsilon_u = 0, \quad \text{and} \quad \epsilon_e = \epsilon_t.$$ 

Any achievable $(0, \epsilon_t)$-rate is also $(\epsilon_u, \epsilon_t)$-achievable for all $0 \leq \epsilon_u < \epsilon_t < 1$

Given a codebook $C_n = \{x^n(1), \ldots, x^n(M)\}$ and channel output $y^n$, decode to $\hat{m}$ if and only if it is the unique message such that

$$\hat{I}(x^n(\hat{m}) \land y^n) \geq \gamma$$

for some $\gamma > 0$, where $\hat{I}(x^n \land y^n)$ is the empirical mutual information of $x^n$ and $y^n$
Proof Idea: Achievability

- Key realization: It suffices to prove that $\sqrt{V_{\epsilon t}} \Phi^{-1}(\epsilon_t)$ is a $(0, \epsilon_t)$-achievable second-order coding rate, i.e.,

$$\epsilon_u = 0, \quad \text{and} \quad \epsilon_e = \epsilon_t.$$

- Any achievable $(0,\epsilon_t)$-rate is also $(\epsilon_u,\epsilon_t)$-achievable for all $0 \leq \epsilon_u < \epsilon_t < 1$

- Given a codebook $C_n = \{x^n(1), \ldots, x^n(M)\}$ and channel output $y^n$, decode to $\hat{m}$ if and only if it is the unique message such that

$$\hat{I}(x^n(\hat{m}) \wedge y^n) \geq \gamma$$

for some $\gamma > 0$, where $\hat{I}(x^n \wedge y^n)$ is the empirical mutual information of $x^n$ and $y^n$

- If there is no such message, or more than one, declare erasure
Proof Idea: Analysis of Undetected Error

- Pick $X^n(m), m \in [M]$ uniformly at random from a type class $\mathcal{T}_P$ with type $P \in \mathcal{P}_n(\mathcal{X})$
Proof Idea: Analysis of Undetected Error

- Pick $X^n(m), m \in [M]$ uniformly at random from a type class $\mathcal{T}_P$ with type $P \in \mathcal{P}_n(\mathcal{X})$

- Assuming message 1 sent, the undetected error is bounded as

$$ \mathbb{E} [\Pr(\mathcal{E}_u | C_n)] \leq \Pr \left[ \max_{m \in [M] \setminus \{1\}} \hat{I}(X^n(m) \wedge Y^n) \geq \gamma \right] $$
Proof Idea: Analysis of Undetected Error

- Pick $X^n(m), m \in [M]$ uniformly at random from a type class $\mathcal{T}_P$ with type $P \in \mathcal{P}_n(\mathcal{X})$

- Assuming message 1 sent, the undetected error is bounded as

$$\mathbb{E} \left[ \Pr(\mathcal{E}_u | C_n) \right] \leq \Pr \left[ \max_{m \in [M] \setminus \{1\}} \hat{I}(X^n(m) \land Y^n) \geq \gamma \right]$$

- Standard method of types result

$$\Pr \left[ \hat{I}(X^n(m) \land Y^n) \geq \gamma \right] \leq (n + 1)^{|\mathcal{X}|(1 + |\mathcal{Y}|)} 2^{-n\gamma}$$
Proof Idea: Analysis of Undetected Error

- Pick $X^n(m), m \in [M]$ uniformly at random from a type class $\mathcal{T}_P$ with type $P \in \mathcal{P}_n(\mathcal{X})$

- Assuming message 1 sent, the undetected error is bounded as

$$\mathbb{E} \left[ \Pr(\mathcal{E}_u | C_n) \right] \leq \Pr \left[ \max_{m \in [M] \setminus \{1\}} \hat{I}(X^n(m) \land Y^n) \geq \gamma \right]$$

- Standard method of types result

$$\Pr \left[ \hat{I}(X^n(m) \land Y^n) \geq \gamma \right] \leq (n + 1)^{|\mathcal{X}|(1+|\mathcal{Y}|)} 2^{-n\gamma}$$

- Thus, by choosing $\gamma = \frac{1}{n} \log M + O\left(\frac{\log n}{n}\right)$, we have

$$\mathbb{E} \left[ \Pr(\mathcal{E}_u | C_n) \right] \leq O\left(\frac{1}{\sqrt{n}}\right)$$
The erasure event $\mathcal{E}_e$ can be decomposed into two events

\begin{align*}
\mathcal{E}_{e}^{(1)} &:= \{ \hat{I}(X^n(m) \land Y^n) < \gamma, \ \forall m \in [M] \} \\
\mathcal{E}_{e}^{(2)} &:= \{ \hat{I}(X^n(m) \land Y^n) \geq \gamma, \ \text{for at least two } m \in [M] \}
\end{align*}
Proof Idea: Analysis of Erasure Error I

- The erasure event $\mathcal{E}_e$ can be decomposed into two events

  
  \[ \mathcal{E}_e^{(1)} := \left\{ \hat{I}(X^n(m) \wedge Y^n) < \gamma, \ \forall m \in [M] \right\} \]
  \[ \mathcal{E}_e^{(2)} := \left\{ \hat{I}(X^n(m) \wedge Y^n) \geq \gamma, \ \text{for at least two } m \in [M] \right\} \]

- Furthermore,

  \[ \mathcal{E}_e^{(2)} \subseteq \mathcal{F}_e^{(2)} := \left\{ \max_{m \in [M] \setminus \{1\}} \hat{I}(X^n(m) \wedge Y^n) \geq \gamma \right\} \]

  But then $\mathcal{F}_e^{(2)}$ is exactly the undetected error event, thus

  \[ \mathbb{E} \left[ \Pr(\mathcal{F}_e^{(2)} \mid \mathcal{C}_n) \right] \leq O \left( \frac{1}{\sqrt{n}} \right) \]
Proof Idea: Analysis of Erasure Error II

We are left with

$$E_e^{(1)} := \{ \hat{I}(X^n(m) \land Y^n) < \gamma, \ \forall m \in [M] \}$$
Proof Idea: Analysis of Erasure Error II

- We are left with
  \[ \mathcal{E}_e^{(1)} := \{ \hat{I}(X^n(m) \land Y^n) < \gamma, \ \forall m \in [M] \} \]

- But this can be upper bounded as
  \[ \mathcal{E}_e^{(1)} \subset \mathcal{F}_e^{(1)} := \{ \hat{I}(X^n(1) \land Y^n) < \gamma \} \]
Proof Idea: Analysis of Erasure Error II

- We are left with
  \[ E^{(1)}_e := \{ \hat{I}(X^n(m) \wedge Y^n) < \gamma, \quad \forall m \in [M] \} \]

- But this can be upper bounded as
  \[ E^{(1)}_e \subset F^{(1)}_e := \{ \hat{I}(X^n(1) \wedge Y^n) < \gamma \} \]

- But then \( E[Pr(F^{(1)}_e | C_n)] \) can be bounded using standard dispersion techniques [Wang-Ingber-Kochman (2011)]
Proof Idea: Analysis of Erasure Error II

- We are left with
  \[ \mathcal{E}_e^{(1)} := \{ \hat{I}(X^n(m) \land Y^n) < \gamma, \ \forall m \in [M] \} \]

- But this can be upper bounded as
  \[ \mathcal{E}_e^{(1)} \subset \mathcal{F}_e^{(1)} := \{ \hat{I}(X^n(1) \land Y^n) < \gamma \} \]

- But then \( \mathbb{E}[\Pr(\mathcal{F}_e^{(1)} | C_n)] \) can be bounded using standard dispersion techniques [Wang-Ingber-Kochman (2011)]

- Thus, we may choose
  \[ \log M \approx nI(P, W) + \sqrt{nV(P, W)} \Phi^{-1}(\epsilon_t) \]
  resulting in
  \[ \mathbb{E} \left[ \Pr(\mathcal{F}_e^{(1)} | C_n) \right] \approx \epsilon_t \]
Proof Idea: Derandomization

Finally we need to show that a sequence of deterministic codes exists.

By Markov's inequality, there exists a sequence of deterministic codes $C_n$ such that for every $\{\theta_n\}_{n \in \mathbb{N}} \subset (0,1)$,

$$\epsilon_u,n(C_n) \leq 1 - \theta_n \mathbb{E}[(\epsilon_u,n(C_n))]$$

But in our case $\mathbb{E}[\epsilon_u,n(C_n)] = O(\sqrt{1/n})$ so we may choose $\theta_n = n^{-\frac{1}{4}}$ (say) and we're done...

Notice that the undetected error being 0 asymptotically is important for derandomization.
Proof Idea: Derandomization

- Finally we need to show that a sequence of deterministic codes exists.
- By Markov’s inequality, there exists a sequence of deterministic codes $C_n$ such that for every $\{\theta_n\}_{n \in \mathbb{N}} \subset (0, 1)$,
  \[
  \epsilon_{u,n}(C_n) \leq \frac{1}{\theta_n} \mathbb{E}[\epsilon_{u,n}(C_n)], \quad \text{and} \quad \epsilon_{t,n}(C_n) \leq \frac{1}{1 - \theta_n} \mathbb{E}[\epsilon_{t,n}(C_n)]
  \]
Proof Idea: Derandomization

- Finally we need to show that a sequence of deterministic codes exists.

- By Markov’s inequality, there exists a sequence of deterministic codes $C_n$ such that for every $\{\theta_n\}_{n \in \mathbb{N}} \subset (0, 1)$,

\[
\epsilon_{u,n}(C_n) \leq \frac{1}{\theta_n} \mathbb{E}[\epsilon_{u,n}(C_n)], \quad \text{and} \quad \epsilon_{t,n}(C_n) \leq \frac{1}{1 - \theta_n} \mathbb{E}[\epsilon_{t,n}(C_n)]
\]

- But in our case

\[
\mathbb{E}[\epsilon_{u,n}(C_n)] = O \left( \frac{1}{\sqrt{n}} \right)
\]

so we may choose $\theta_n = n^{-1/4}$ (say) and we’re done...
Proof Idea: Derandomization

- Finally we need to show that a sequence of deterministic codes exists

- By Markov’s inequality, there exists a sequence of deterministic codes $C_n$ such that for every $\{\theta_n\}_{n \in \mathbb{N}} \subset (0, 1)$,
  \[\epsilon_{u,n}(C_n) \leq \frac{1}{\theta_n} \mathbb{E}[\epsilon_{u,n}(C_n)], \quad \text{and} \quad \epsilon_{t,n}(C_n) \leq \frac{1}{1 - \theta_n} \mathbb{E}[\epsilon_{t,n}(C_n)]\]

- But in our case
  \[\mathbb{E}[\epsilon_{u,n}(C_n)] = O\left(\frac{1}{\sqrt{n}}\right)\]
  so we may choose $\theta_n = n^{-1/4}$ (say) and we’re done...

- Notice that the undetected error being 0 asymptotically is important for derandomization
Conclusion

- Provided fixed error asymptotic analysis of erasure decoding
Conclusion

- Provided fixed error asymptotic analysis of erasure decoding
- Paper contains analysis for list decoding: Extension of hypothesis testing converse to list decoding [cf. Kostina-Verdú (2012-2013)]
Conclusion

- Provided fixed error asymptotic analysis of erasure decoding
- Paper contains analysis for list decoding: Extension of hypothesis testing converse to list decoding [cf. Kostina-Verdú (2012-2013)]
- Extended version contains analysis for erasure decoding for the Slepian-Wolf problem
- arxiv.org/pdf/1402.4881
Conclusion

- Provided fixed error asymptotic analysis of **erasure decoding**

- Paper contains analysis for **list decoding**: Extension of hypothesis testing converse to list decoding [cf. Kostina-Verdú (2012-2013)]

- Extended version contains analysis for erasure decoding for the **Slepian-Wolf problem**

- arxiv.org/pdf/1402.4881

- More refined tradeoff between total and undetected errors (joint work with M. Hayashi)