Error Exponents in Asynchronous Communication

Da Wang, with
Venkat Chandar, Sae-Young Chung, Greg Wornell

Signals, Information
and Algorithms
Laboratory

ISIT 2011, St. Petersburg

August 2, 2011

Overview of Synchronization

Multiple levels of synchronization

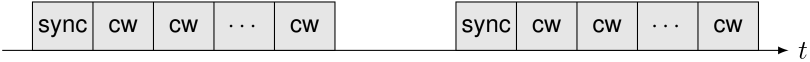
- Carrier synchronization
  - PLL, maybe pilot assisted
- Symbol synchronization/timing recovery
  - Match filter, DLL
- Frame synchronization
  - Initial/One-shot frame synchronization
  - Continuous frame synchronization

Scope of this talk

- Point-to-point communication
- Timing uncertainy in initial frame synchronization

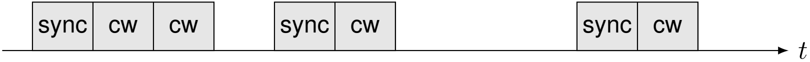
Two communication scenarios

Traditional



- Cost of synchronization: low, due to amortization

Emerging



- Examples: sensor networks, etc.
- Cost of synchronization: high, as amortization is insignificant

Separate vs.joint synchronization and coding

Separate synchronization and coding

$$\underbrace{\boxed{s_1}\boxed{s_2}\boxed{\dots}\boxed{s_k}}_{\text{symbols for synchronization}}\underbrace{\boxed{c_1}\boxed{c_2}\boxed{\dots}\boxed{\dots}\boxed{\dots}\boxed{c_{n-k}}}_{\text{symbols for information}}$$

Joint synchronization and coding

$$\underbrace{\boxed{c_1}\boxed{c_2}\boxed{\dots}\boxed{\dots}\boxed{\dots}\boxed{\dots}\boxed{\dots}\boxed{\dots}\boxed{\dots}\boxed{c_n}}_{\text{symbols for } \textit{both} \text{ synchronization and information}}$$

Question

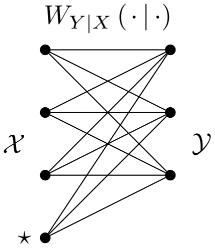
What is the performance loss of separate synchronization and coding, comparing to optimal joint synchronization and coding?

Asynchronous channel model

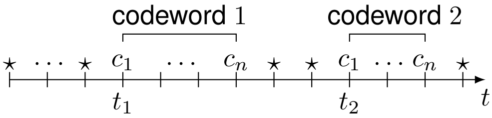
Asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$

- First proposed in [1]
- DMC with a special symbol $\star$
  - input symbol when nothing is sent.
  - models the effect of noise at channel output

Channel



- Tx sends codewords at time $t_1, t_2, \ldots$

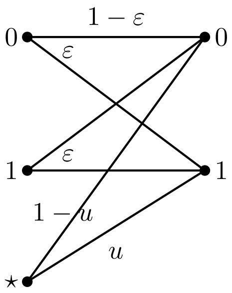


[1] Tchamkerten *et al.*, "Communication Under Strong Asynchronism", 2009

Asynchronous channel model
Examples

Asynchronous BSC

- crossover probability ε
- $W(\cdot \,|\, \star) = \text{Bernoulli}(u)$.



Asynchronous AWGN Channel

$$Y^n = \begin{cases} X^n + Z^n & \text{TX transmits} \\ Z^n & \text{TX silent} \end{cases}$$

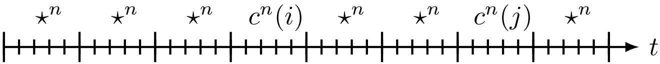
- X^n: average signal power P
- Z^n: average noise power 1
- $W(\cdot \,|\, \star) = \mathsf{N}(0,1)$

Asynchronous channel model

Slotted channel assumption

Each slot has length n, and contains

- either a noise sequence $\star^n$,
- or a codeword sequence
- example:



Receiver in asynchronous communication

In each time slot, received channel output y^n

$$\Downarrow$$

H_0: induced by noise sequence $\star^n$ vs. H_1: induced by some codeword $x^n(m)$

A binary (composite) hypothesis testing problem

Channel code $\mathcal{C} = \left\{x^n(m) | m = 1, 2, \cdots, M = e^{nR}\right\}$,

$$\begin{cases} H_0: & Y_i \overset{i.i.d.}{\sim} W\left(\cdot \,|\, \star\right) \\ H_1: & Y^n \sim W^n\left(\cdot | x^n(m)\right) \quad m \in \{1, 2, \cdots, M\} \end{cases}$$

Error events

- **Miss**: detect codeword as noise
- **False alarm**: detect noise as codeword
- **Decoding error**: decode into the wrong codeword

Performance metrics in asynchronous communication

Performance metrics

- Rate R
- Miss probability: P_{m}
- False alarm probability: P_{f}
- Decoding error probability: P_{d}
- Error exponents: E_{m}, E_{f}, E_{d}

$$P \approx \exp(-nE)$$

$$n \approx -\frac{1}{E} \log P$$

Central questions

For an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$,

- What are the fundamental tradeoffs between E_m and E_f, given $P_{\mathrm{m}}, P_{\mathrm{f}}, P_{\mathrm{d}} \to 0$?
  - Part I of the talk
- How do these tradeoffs compared with those obtained from the separate synchronization and coding approach?
  - Part II of the talk

Part I:

Fundamental Tradeoffs

Regimes of interest

General scenario

Given rate R, $P_{\mathrm{m}}, P_{\mathrm{f}}, P_{\mathrm{d}} \to 0$ as $n \to \infty$,

- Find the trade-off between E_{m} and E_{f}
- Similar to the Neyman-Pearson setup, but the objects of interest are error exponents.
- Results:
  - achievability schemes for DMC and AWGN channels: constant composition code
  - (multi-letter) outer bound for DMC

Special case

Given $P_{\mathrm{m}} \to 0$ ($E_{\mathrm{m}} = 0$), what is the optimal E_{f}?

- Motivation: communication really sparse
- ✓ Complete characterization $E_{\mathrm{f}}(R)$
- focus of this talk

Optimal E_{f} with $P_{\mathrm{m}} \to 0$
Main theorem

Define

$$P_Y(\cdot) \triangleq \sum_x W(\cdot \,|\, x)\, P_X(x)$$
$$W_\star \triangleq W(\cdot \,|\, \star),$$

then

$$E_{\mathrm{f}}(R) = \max_{P_X : I(P_X, W) \geq R} D\left(P_Y \,\|\, W_\star\right) + I(P_X, W) - R$$
$$= \max_{P_X : I(P_X, W) = R} D\left(P_Y \,\|\, W_\star\right)$$

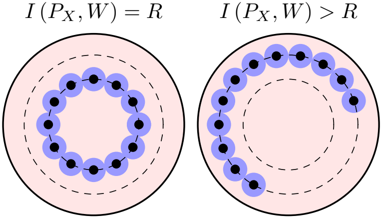
- Optimal E_{f} is the KL-divergence between
  - output distribution of the codebook P_Y
  - output distribution of the $\star$ symbol $W_\star$
- The two expressions correspond to two achievability schemes

Optimal E_{f} with $P_{\mathrm{m}} \to 0$

Two achievability schemes

Joint detection & decoding

- i.i.d.codebook P_X with $I\left(P_X, W\right) \geq R$
- Detect & decode simultaneously based on joint typicality
- Geometric view:



- More flexible codebook choice

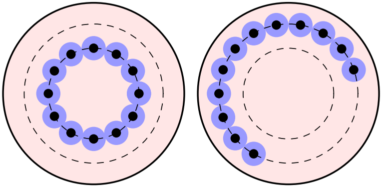
Optimal E_{f} with $P_{\mathrm{m}} \to 0$

Two achievability schemes

Joint detection & decoding

- i.i.d.codebook P_X with $I\left(P_X, W\right) \geq R$
- Detect & decode simultaneously based on joint typicality
- Geometric view:

$I\left(P_X, W\right) = R$ $\qquad$ $I\left(P_X, W\right) > R$

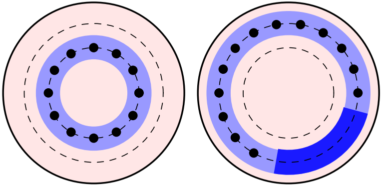


- More flexible codebook choice

Simpler detection

- i.i.d.codebook P_X with $I\left(P_X, W\right) = R$
- Detect based on output distribution
- Geometric view:

$I\left(P_X, W\right) = R$ $\qquad$ $I\left(P_X, W\right) > R$



- Simpler detection rule
- Regular decoding afterwards

Optimal E_f with $P_\mathrm{m} \to 0$

Example: Asynchronous BSC

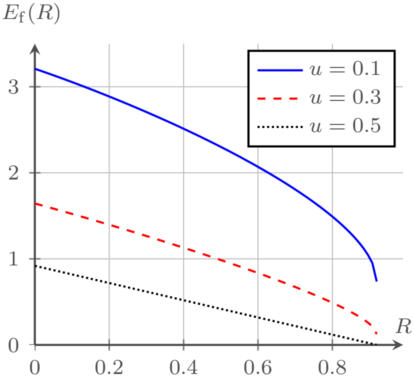
Optimal $E_\mathrm{f}(R)$

$$E_\mathrm{f}(R) = D\left(s \,\|\, u\right)$$

where $s \geq 0.5$ satisfies

$$H_b\left(s\right) - H_b\left(\varepsilon\right) = R.$$

BSC with $\varepsilon = 0.01$



Optimal E_f with $P_\mathrm{m} \to 0$
Example: Asynchronous AWGN

Optimal $E_\mathrm{f}(R)$

$$E_\mathrm{f}(R) = \mathsf{SNR}/2 - R$$

AWGN with SNR $= 10 = 20$dB

Part II:

Suboptimality of Separate Synchronization and Coding

Separate synchronization and coding (training)

Training-based scheme (separate synchronization and coding)

s_1	s_2	$\dots$	s_k	capacity achieving code

$k = (1 - R/C)n$ symbols $\qquad\qquad$ nR/C symbols

- detection algorithm operates on the k synchronization symbols only

- [2] shows training-based schemes achieve vanishing false alarm error exponent at capacity except for degenerate cases.
- With the slotted model, we quantify the performance loss due to training at any rate $R \in [0, C)$.

[2] Chandar *et al.*, "Training-based schemes are suboptimal for high rate asynchronous communication", 2009

Optimal E_{f} with $P_{\mathrm{m}} \to 0$

Performance of separate synchronization–coding

Training: best performance

$$E_{\mathrm{t}}(R) = (1 - R/C)\, D\left(W\left(\cdot \,|\, s^*\right) \| W_\star\right)$$

where $s^* = \arg\max_{s \in \mathcal{X}} D\left(W\left(\cdot \,|\, s\right) \| W_\star\right)$.

- standard large deviation argument

Optimal E_f with $P_\mathrm{m} \to 0$

Training-based scheme is suboptimal almost everywhere

Training-based scheme is suboptimal almost everywhere

For an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$ and $0 < R \leq C$,

$$E_\mathrm{t}(0) = E_\mathrm{f}(0) \text{ and } E_\mathrm{t}(R) \leq E_\mathrm{f}(R).$$

Furthermore, if the capacity achieving output distribution P_Y^* satisfies $D\left(P_Y^* \,\|\, W_\star\right) > 0$, then for all $R > 0$,

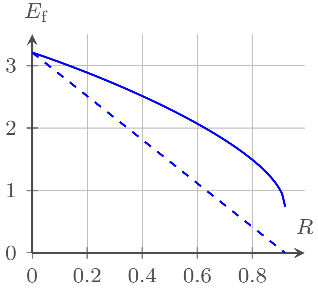
$$E_\mathrm{t}(R) < E_\mathrm{f}(R).$$

Proof: based on the concavity of $E_\mathrm{f}(R)$.

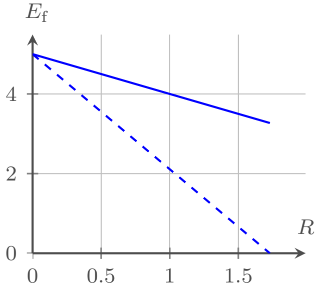
Training-based scheme is suboptimal almost everywhere

Example: BSC & AWGN

BSC with $\varepsilon = 0.01, u = 0.1$



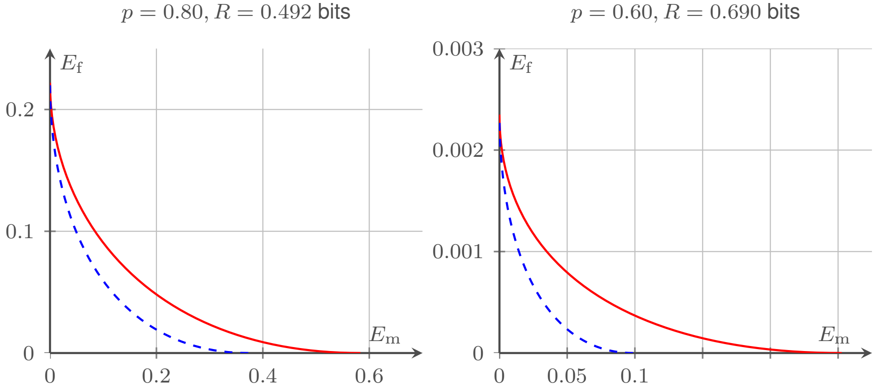
AWGN with SNR=20dB



- Larger gap at higher rate
- Smaller difference at lower rate

Trade-off between E_{m} and E_{f}

Special case: BSC with $u = 0.5$

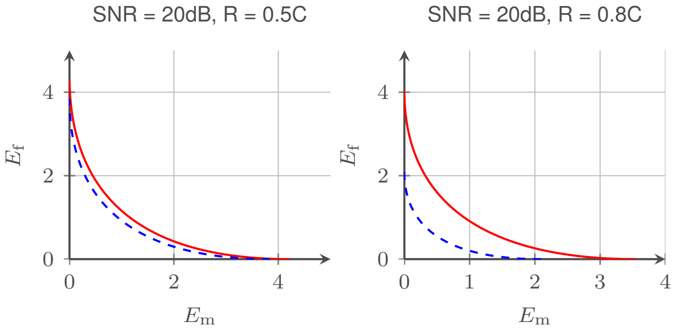


achievable performance: **joint** sync–coding

optimal detection: **separate** sync–coding

Trade-off between E_m and E_f

Special case: AWGN with SNR$=20$dB



——	**achievable** performance: **joint** sync–coding
- - -	**optimal** detection: **separate** sync–coding

Concluding Remarks

Insights

- We quantify the suboptimality of training-based schemes
  - The performance loss is significant in the high rate regime.
  - Advanced the insights from [3] .
- Coding schemes for joint synchronization and coding
  - To maximize $E_{\mathrm{f}}(R)$, i.i.d.codebook is sufficient
    - Typicality decoding, or
    - Detection based on empirical distribution
  - For tradeoff between E_{f} and E_{m}, constant composition codebook is better than the i.i.d.codebook.

[3] Tchamkerten *et al.*, On the capacity region of asynchronous channels, 2008

Backup slides

Discussions

Discussion: slotted vs.unslotted model

$$\text{synchronize} = \underbrace{\overbrace{\text{detect the presence}}^{\text{slotted model}} + \text{locate the position}}_{\text{unslotted model}}$$

Optimal E_{f} with $P_{\mathrm{m}} \to 0$

- [4] shows that, in the unslotted model, after detection, using a prefix with sub-linear length is sufficient to locate the codeword with $P_{\mathrm{m}} \to 0$
  - sub-linear length: does not affect the error exponents E_{f}
  - prefix design: maximum length shift register sequence
- Results of slotted model also hold for the unslotted model!

Positive E_{m} requirement

- Stronger requirement for "position location"
- Results for slotted model do not hold for the unslotted model.
  - Only serve as "upper bounds"

[4] Tchamkerten *et al.*, On the capacity region of asynchronous channels, 2008

Connection to the Single Message Unequal Error Protection

Similarities

Single message UEP	**Asynchronous Communication**
Special codeword	Noise sequence $\star^n$
Regular codewords	Codewords
Miss (special codeword)	False alarm (of noise sequence)
False alarm (of regular codeword)	Miss (a codeword)

Differences

- In UEP, one can design the special codeword
- In asynchronous communication, it is constrained to be repetition (of $\star$).

Asynchronous communication

- can be viewed as UEP with constraint on special message design
- can be extended to obtain results on single message UEP that is more general than [5]

[5] Borade *et al.*, "Unequal Error Protection: An Information-Theoretic Perspective", 2009

Future directions

- More complete characterization of the error exponents
- Sequential detection
- Unslotted model

More details

Detailed analysis

Acceptance region for codewords: $\mathcal{A}_n$

- If $y^n \in \mathcal{A}_n$, we consider the channel input to be a certain codeword $x^n(m)$
- otherwise we consider the channel input as $\star^n$.
- $\mathcal{A}_n$: the acceptance region for codewords
- $\mathcal{B}_n \triangleq \mathcal{A}_n^c$ to be the rejection region for codewords

Error probabilities analysis

$$
\begin{aligned}
P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) &\triangleq \max_m P_{\mathrm{m}}(m) && \triangleq \max_m W^n\left(\mathcal{A}_n^c \,|\, x^n(m)\right) \\
P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) &\triangleq W^n\left(\mathcal{A}_n \,|\, \star^n\right) && \triangleq W_\star^n(\mathcal{A}_n) \\
P_{\mathrm{d}}\left(\mathcal{C}^{(n)}\right) &\triangleq \max_m P_{\mathrm{d}}(m) && \triangleq \max_m \sum_{\hat{m}\neq m} W^n\left(g_n^{-1}(\hat{m}) \middle|\, f_n(m)\right)
\end{aligned}
$$

Scenario 1

Optimal E_f with $P_\mathrm{m} \to 0$

Optimal E_{f} with $P_{\mathrm{m}} \to 0$

Converse

Main idea

- $P_{\mathrm{m}} \to 0$ and $P_{\mathrm{d}} \to 0$
  - The acceptance region $\mathcal{A}_n$ for codewords cannot be too small
  - Essentially no smaller than the union of each codeword's typical shell
- P_{f} cannot be too small
  - as the chance that $\star^n$ falls into the acceptance region cannot be too small

Technique

- First prove it for constant composition code
- Then extends to general code
- Every code has a constant composition subcode with essentially the same rate

Scenario 2

Optimal E_{m} with $P_{\mathrm{f}} \to 0$

Main results

$$\underline{E_{\mathrm{m}}}(R) \leq E_{\mathrm{m}}(R) \leq \overline{E_{\mathrm{m}}}(R)$$

Lower bound

Let $\mathcal{V}_\star = \{V : \sum_x P_X(x) V(\cdot \,|\, x) = W_\star\}$,

$$\underline{E_{\mathrm{m}}}(R) = \max_{P_X : I(P_X, W) \geq R} \; \min_{V \in \mathcal{V}_\star} D\left(V \,\|\, W | P_X\right)$$

where $D\left(V \,\|\, W | P_X\right) \triangleq \mathbb{E}_{P_X}\left[D\left(V\left(\cdot \,|\, X\right) \,\|\, W\left(\cdot \,|\, X\right)\right)\right]$.

Upper bound

$$\overline{E_{\mathrm{m}}}(R) = \max_{P_X : I(P_X, W) \geq R} \mathbb{E}_{P_X}\left[D\left(W_\star \,\|\, W\left(\cdot \,|\, X\right)\right)\right],$$

Optimal E_{m} with $P_{\mathrm{f}} \to 0$

Achievability scheme

- Constant composition codebook with type P_X
- Rejection region for codewords: typical shell of the noise sequence $\star^n$
- $V \in \mathcal{V}_\star$: "confusing" channel realizations
  - Makes the output type of a codeword same as $\star^n$

Upper bound proof idea

- Rate R: a constraint on the codebook type P_X
- Consider a single codeword x^n with type P_X and $\star^n$
- Swap the role of the two sequences
  - x^n: "noise sequence"
  - $\star^n$: "codeword sequence"
- Apply the result on $E_{\mathrm{f}}(R)$, and average over the type P_X

Scenario 3

Trade-off between

E_{m} and E_{f}

For an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$, given a rate R and a miss error exponent constraint e_{m},

$$\underline{E_{\mathrm{f}}}(R, e_{\mathrm{m}}) = \max_{P_X : I(P_X, W) \geq R} \; \min_{V : D(V \,\|\, W | P_X) \leq e_{\mathrm{m}}} \left[D\left(Q_V \,\|\, W_\star\right) + \left| I\left(P_X, V\right) - R \right|^+ \right]$$

where $Q_V(\cdot) = \sum_x P_X(x) V(\cdot \,|\, x)$.

Achievability

- a code achieves miss error exponent e_{m}
  - $\Rightarrow$ it also achieves $P_{\mathrm{m}} \to 0$ for any V s.t. $D\left(V \,\|\, W | P_X\right) \leq e_{\mathrm{m}}$
    - use the typical shell of all these Vs for the detection region of the noise sequence $\star^n$
- Check: $e_{\mathrm{m}} = 0 \quad \Leftrightarrow \quad V = W \quad \Rightarrow \quad \underline{E_{\mathrm{f}}}(R, e_{\mathrm{m}} = 0) = E_{\mathrm{f}}(R)$

Similarly, given a rate R and a false alarm error exponent constraint e_{f}, the following lower bound for the miss reliability function is achievable via a sequence of constant composition codebooks

$$\underline{E_{\mathrm{m}}}(R, e_{\mathrm{f}}) = \max_{P_X : I(P_X, W) \geq R} \; \min_{V : D(Q_V \,\|\, Q_Y) \leq e_{\mathrm{f}}} D\left(V \,\|\, W | P_X\right).$$

Trade-off between E_{m} and E_{f}

Special case: BSC with $u = 0.5$

Joint Synchronization & Coding

$$e_{\mathrm{f}}(\delta) \leq D\left(\delta \| u\right)$$

$$e_{\mathrm{m}}(\delta) \leq \min_{\kappa \in [\delta - \bar{p}\varepsilon, \kappa^*]} \left[\bar{p} D\left(\frac{\delta - \kappa}{\bar{p}} \middle\| \varepsilon\right) + p D\left(\frac{\kappa}{p} \middle\| \bar{\varepsilon}\right)\right]$$

where $\bar{x} \triangleq 1 - x$ and $\kappa^* = \min\{\delta, p(1-\varepsilon)\}$.

Training

$$e_{\mathrm{m}}(\lambda) \leq \left(1 - \frac{R}{C}\right) D\left(q_\lambda \| \varepsilon\right)$$

$$e_{\mathrm{f}}(\lambda) \leq \left(1 - \frac{R}{C}\right) D\left(q_\lambda \| u\right)$$

where $q_\lambda \propto \varepsilon^\lambda u^{1-\lambda}$, $\lambda \in [0, 1]$.

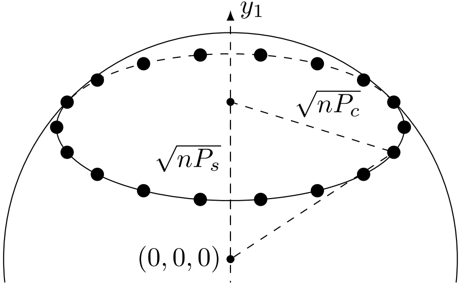
Trade-off between E_{m} and E_{f}

Achievability result for AWGN

Code design

Find P_c and P_s satisfy

- $R = \log(1 + P_c)/2$
- $P_s = P - P_c$
- $X^n = (\sqrt{nP_s}, \hat{X}_1, \cdots, \hat{X}_{n-1})$



Decision rule

$$\begin{cases} H_0: & \text{noise} \\ H_1: & \text{some codeword} \end{cases} \quad \Longrightarrow \quad a y_1 + b\|y_2^n\| \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\gtrless}} \sqrt{n}\eta$$

Trade-off between E_{m} and E_{f}

Special case: AWGN with SNR$=20$dB

Joint Synchronization & Coding	Training
$e_{\mathrm{f}}(\eta) \le \max_{(a,b)\in[0,1]^2} \min_{0\le r\le \eta-b} \left[\frac{r^2}{2a^2} + I_{\chi_1^2}\left(\frac{(\eta-r)^2}{b^2}\right)\right]$ $e_{\mathrm{m}}(\eta) \le \max_{(a,b)\in[0,1]^2} \min_{\eta-b\sqrt{P_c+1}\le r\le \eta} \left[\frac{(r-a\sqrt{P_s})^2}{2a^2} + I_{\mathsf{SG}}\left(P_c, \frac{(\eta-r)^2}{b^2}\right)\right]$	$e_{\mathrm{m}}(\eta) \le (\sqrt{P_s}-\eta)^2/2$ $e_{\mathrm{f}}(\eta) \le \eta^2/2$

where

$$I_{\chi_1^2}(x) \triangleq \frac{1}{2}(x - \ln x - 1)$$

$$I_{\mathsf{SG}}(P,\eta) \triangleq \frac{1}{2}\left(P + \eta - \sqrt{1+4P\eta} - \log\left[\frac{\sqrt{1+4P\eta}-1}{2P}\right]\right).$$

More AWGN results