Digital Information Processing

“Information Theory”

Philippe Ciblat

Université Paris-Saclay & Télécom ParisTech

Outline

- Section 1: Information Theory tools
- Section 2: Optical nonlinear channels
- Section 3: Optical MIMO channels
  - Chromatic dispersion (CD) SISO channel
  - Polarization Mode Dispersion (PMD) MIMO channel
  - Polarization Dependent Loss (PDL) MIMO channel
  - Practical coding schemes

Section 1 : Information Theory tools

Goal for optical system

Transmitting

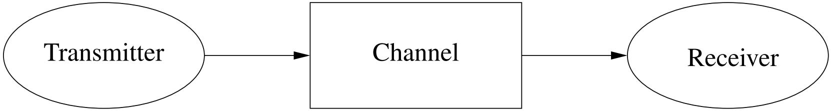
- over a long range (2000km)
- ultra-high data-rate (1Tbs per wavelength)
- with high fidelity (bit error rate at 10^{-15})

but constraints occur

- economical constraints (energy, complexity, cost)
- technological constraints (propagation, electronic devices)
- **theoretical constraints**

$\Rightarrow$ **Information Theory** **[Shannon1948]**

Notion of “information”



Idea

the received information is “something” which is transmitted but the receiver does not know it before the transmission

- Information thus deals with *uncertainty*
- Uncertainty is well modeled by *randomness*
- Examples :
  - A deterministic event does not contain information
  - A random event in communications: binary stream from a audio/video flow, ...

Information metric

Information corresponds to randomness

$\Updownarrow$

the more random the event is, the more information the event carries

Level (or quantity) of information carried by event $S = s$ has to satisfy following intuitive properties:

- $I(S = s)$ decreasing function wrt probability $Pr(S = s)$
- $I(S = s) = 0$ when $Pr(S = s) = 1$
- $I(S = s)$ is a continuous function wrt $Pr(S = s)$
- If two independent events $S_1 = s_1$ et $S_2 = s_2$ are observed, then the total received information is $I(S_1 = s_1) + I(S_2 = s_2)$

$$\Rightarrow \quad I(S = s) = -\log_2 \left(Pr(S = s) \right)$$

(Discrete) Entropy

- Let X be a discrete random variable (rv) whose values belong to $\mathcal{A} = \{x_1, \cdots, x_M\}$
- Let $p(x) = Pr(X = x)$
- Information in rv X, called *entropy*, is the average of information of each event. Therefore

$$H(X) = \mathbb{E}_x[I(X = x)] = -\sum_{m=1}^{M} p(x_m) \log_2(p(x_m))$$

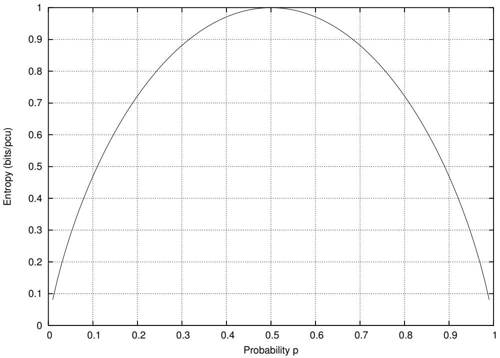
Remarks

- 2-base for logarithm since information bits per channel use (pcu)
- If X is N-variate, then $H(X) = \mathbb{E}_x[I(X = x)]/N$
- Entropy only depends on probability density function
- The semantic is not taken into account in Information Theory
  - In French, only the probability of occurrence of each character is taken into account, not the meaning of the words/sentences

Example: binary stream

- $Pr(X = 0) = p$
- $Pr(X = 1) = 1 - p$

$$H(X) = H(p) \stackrel{def.}{=} -p\log_2(p) - (1-p)\log_2(1-p)$$



(Other) Entropies

- Joint entropy: let $X_1, \cdots, X_n$ be random variables

$$H(X_1, \cdots, X_n) = -\sum_{x_1 \in \mathcal{A}_1} \cdots \sum_{x_n \in \mathcal{A}_n} p(x_1, \cdots, x_n) \log_2(p(x_1, \cdots, x_n))$$

- Conditional entropy: level of information in Y once X has been observed (and so known)

$$\begin{aligned} H(Y|X) &= -\mathbb{E}_X[\sum_{y \in \mathcal{A}_y} p(y|x) \log_2(p(y|x))] \\ &= -\sum_{x \in \mathcal{A}_x, y \in \mathcal{A}_y} p(x,y) \log_2(p(y|x)) \end{aligned}$$

- Differential entropy: level of information when X has values belonging to an uncountable set

$$H(X) = -\int p(x) \log_2(p(x)) dx$$

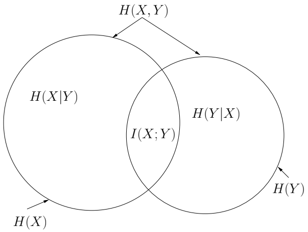
where $p(x)$ is the probability density function (pdf) of rv X

  – Gaussian case with $p(x) = (1/\sqrt{2\pi})e^{-x^2/2}$

Mutual information

- Let $I(X;Y)$ be the mutual information between X and Y
- It is the level of information shared by X and Y

$$I(X;Y) = H(X) - H(X|Y)$$



Examples

- If $Y = X$, then $I(X,Y) = H(X)$
- If Y and X independent, then $I(X,Y) = 0$

Channel capacity

- Let $X = [x_1, \cdots, x_N]$ be the transmitted signal
- Let T be the information rate: TN bits of information in X
- Let $Y = [y_1, \cdots, y_N]$ be the received signal
- Let C be the channel capacity. Then

$$C = \max_{p(x)} I(X;Y) \quad \text{(depends only on } p(y|x) \text{ and } p(x))$$

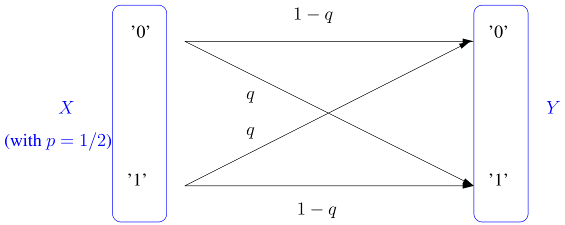
Theorem

There exists a coding scheme of length N with information rate T, such that,

$$T < C \quad \text{and} \quad \lim_{N\to\infty} P_e = 0$$

- Fundamental limit is the rate <u>not</u> the reliability
- Non-constructive theorem (large interleaver required)

Example: Binary Symmetric Channel (BSC)

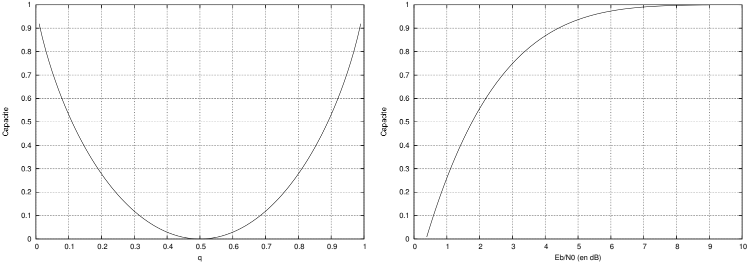


$$C = 1 - H(q)$$

- Channel corresponding with hard decision binary stream
- If physical channel is Gaussian, then

$$q = Q\left(\sqrt{\frac{2E_c}{N_0}}\right) = Q\left(\sqrt{\frac{2E_bC}{N_0}}\right) \quad \text{with} \quad E_c = E_x/2$$

Example: cont'd



- Capacity vanishes when $q = 0.5$
- Capacity vanishes when E_b/N_0 below 0.4dB

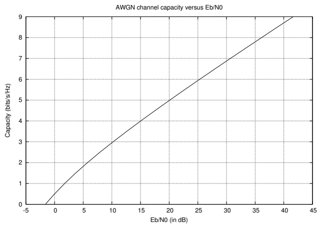
Example: Gaussian channel

$$Y = X + W$$

with W a white zero-mean circularly-symmetric Gaussian noise with variance N_0 per real dimension

The most famous expression of Information Theory

$$C = \log_2(1 + E_c/N_0)$$



- As $E_c = T \times E_b$ and $T < C$,
  $$\frac{E_b}{N_0} \geq \frac{2^T - 1}{T} \overset{T \to 0}{\geq} \underbrace{\log(2)}_{= -1.6\text{dB}}$$
- As for the rate (bits/s), $B \log_2(1 + P/BN_0)$ with bandwidth B (Hz)

Spectral efficiency (SE)

- Let us focus on practical linear modulations

$$x(t) = \sum_k s_k g(t - kT_s)$$

where

  - s_k belongs to 2^m-QAM
  - s_k is also the output of a N-length FEC with rate R
  - $g(t)$ a ρ-roll-off shaping filter ($\rho \approx 0$ for WDM Nyquist)
  - Information rate $T = mR$ pcu

- Data rate = T/T_s bits/s

- Maximal data rate = C/T_s bits/s

Maximal SE: C bits/s/Hz (zero BER and fixed SNR)

Practical SE: T bits/s/Hz (non-zero BER and fixed MCS)

Section 2 : Optical nonlinear channels

Nonlinear channels models

- To derive capacity, Input/Output closed-form expression required
- Hereafter, for sake of simplicity, one polarization

Nonlinearity $=$ Kerr effect (index of refraction depends on power)

Non-linear Schrödinger Equation (NSE)

$$i\frac{\partial X}{\partial z} + i\frac{\alpha}{2}X - \frac{\beta_2}{2}\frac{\partial^2 X}{\partial t^2} + \gamma_{NL}|X|^2 X = 0$$

with

- α the attenuation
- β_2 the dispersion (CD)
- γ_{NL} the nonlinear coefficient
- L the fiber/span length

No closed-form solution without further simplifications (α or β_2)

Volterra series

- Volterra series provide approximate solution to NSE
- So, in the frequency domain, we have [Brandt1997]

$$Y(f) \approx e^{2i\pi^2\beta_2 L f^2} X(f) + Y_{NL}(f) + W(f)$$

with

$$\begin{aligned} Y_{NL}(f) &= e^{2i\pi^2\beta_2 L f^2}(2i\pi)\gamma_{NL} \iiint \delta(f - f_1 + f_2 - f_3) X(f_1) X^*(f_2) X(f_3) \\ &\times \frac{e^{L\left(-\alpha + 2i\pi^2\beta_2(-f^2 + f_1^2 - f_2^2 + f_3^2)\right)} - 1}{-\alpha + 2i\pi^2\beta_2(-f^2 + f_1^2 - f_2^2 + f_3^2)} df_1 df_2 df_3 \end{aligned}$$

Special case: non-dispersion ($\beta_2 = 0$)

$$Y(t) = e^{i\gamma_{NL} L_{\text{eff}} |X(t)|^2} X(t) + W(t)$$

with $L_{\text{eff}} = (1 - e^{-\alpha L})/\alpha$

Literature on nonlinear capacity derivations

- **Numerical evaluation** [Essiambre2010]

$$I(X;Y) = \iint p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) dxdy \text{ with predefined } p(x)$$

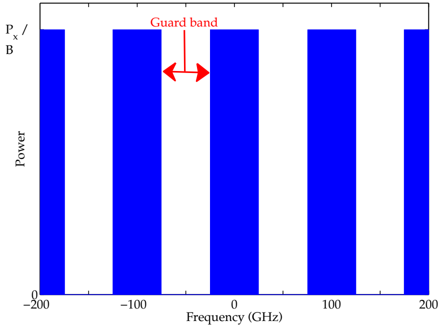
- **Lower-bound derivations** [Turitsyn2003]

- **Approximation derivations with $p(x)$ assumed Gaussian**
  - based on Volterra series [Tang2001,Tang2006]
  - based on additional multiplicative noise (valid only for specific set-ups) [Mitra2001,Kahn2004]
  - based on additional additive noise (valid only for specific set-ups) [Poggiolini2011]
  - based on perturbative solution of NSE [Narimanov2002]

In this lecture, we will focus on Tang's approach

Numerical results from the literature

- 1 span
- $B = 50$ GHz
- $B_g = 0$ GHz
- $N_c = 81$ WDM channels
- $L = 80$ km
- $N_0 = 10^{-5}\ \mathrm{mW}/\mathrm{GHz}$
- $\alpha = 0.2$ dB/km
- $\beta_2 = -21.6\ \mathrm{ps}^2/\mathrm{km}$
- $\gamma_{NL} = 1.22\ \mathrm{W}^{-1} \cdot \mathrm{km}^{-1}$

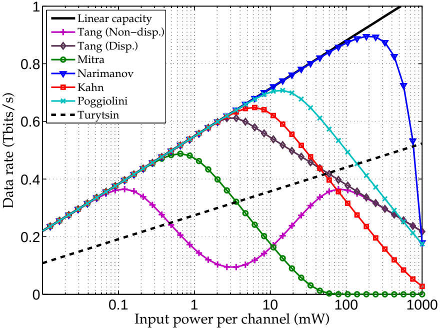


Remarks

- Results are scattered because of models
- Gaussian input derivations correct for practical input power
- Chromatic dispersion increases the nonlinear capacity

Numerical results from the literature

- 1 span
- $B = 50$ GHz
- $B_g = 0$ GHz
- $N_c = 81$ WDM channels
- $L = 80$ km
- $N_0 = 10^{-5}\ \mathrm{mW/GHz}$
- $\alpha = 0.2$ dB/km
- $\beta_2 = -21.6\ \mathrm{ps}^2/\mathrm{km}$
- $\gamma_{NL} = 1.22\ \mathrm{W}^{-1} \cdot \mathrm{km}^{-1}$



Remarks

- Results are scattered because of models
- Gaussian input derivations correct for practical input power
- Chromatic dispersion increases the nonlinear capacity

Derivations for non-dispersive case

$$\text{Maximal rate} = B\log_2\left(1+\frac{\frac{P_x}{1+(\gamma_{NL}L_{\text{eff}}N_cP_x)^2}}{N_0B+P_x\frac{3(\gamma_{NL}L_{\text{eff}}P_x)^2N_c^2}{2(1+(\gamma_{NL}L_{\text{eff}}N_cP_x)^2)^3}}\right) \quad \text{bits/s}$$

Remarks

- Derivations can be found in [Delesques2012]
- Red terms have been added and decrease the capacity
- Maximum value obtained for $P_x^* = (N_0B/3)^{1/3}/(\gamma_{NL}L_{\text{eff}}N_c)^{2/3}$ is

$$\text{Maximal rate}^* = \log_2\left(1+\frac{2/9^{2/3}}{(\gamma_{NL}L_{\text{eff}}N_cN_0B)^{2/3}}\right)$$

- At high P_x,

$$\text{Maximal rate} = B\log_2\left(1+\frac{1}{N_0B\gamma_{NL}^2L_{\text{eff}}^2N_c^2P_x}\right) \to 0$$

Derivations for dispersive case

$$\text{Maximal rate} = \int_{-B/2}^{B/2} \log_2\left(1 + \frac{P_x(1 + 16\pi^2\gamma_{NL}^2 L_{\text{eff}}^2 N_c^2 P_x^2)}{N_0 B + M_X(f)}\right) df$$

with

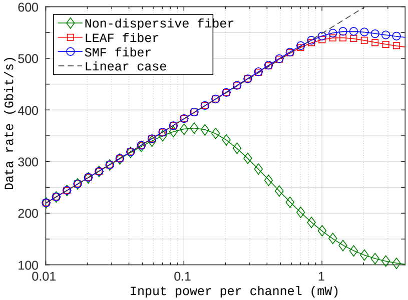
$$\begin{aligned} M_X(f) &= 8\pi^2\gamma_{NL}^2(P_x^3/B^2) \sum_{k_1,k_2,k_3=-N}^{N} \iiint \delta(f - f_1 + f_2 - f_3) \\ &\times \text{rect}\left(\frac{f_1 + k_1 B}{B}\right) \text{rect}\left(\frac{f_2 + k_2 B}{B}\right) \text{rect}\left(\frac{f_3 + k_3 B}{B}\right) \\ &\times \left|\frac{e^{L\left[-\alpha + 2i\beta_2\pi^2(-f^2 + f_1^2 - f_2^2 + f_3^2)\right]} - 1}{-\alpha + 2i\beta_2\pi^2(-f^2 + f_1^2 - f_2^2 + f_3^2)}\right|^2 df_1 df_2 df_3 \end{aligned}$$

Remarks

- Derivations can be found in `[Delesques2012]`
- Expressions difficult to manage
- But easy to evaluate numerically

Numerical illustrations

Same value as in slide 18 (except β_2: -5.1 for LEAF; -21.6 for SMF)



Remarks

- Dispersion increases the capacity
- Around 550 Gbits/s (11 bits/s/Hz) possible with Gaussian input!

Section 3 : Optical MIMO channels

MIMO models

- Two polarizations (2×2 MIMO)
  - Chromatic dispersion (CD) + Polarization Dispersion Mode (PMD)

$$Y(f) = e^{2i\pi^2\beta_2 L f^2}\mathbf{H}_{\mathrm{PMD}}X(f)+W(f) \quad \text{with} \quad \mathbf{H}_{\mathrm{PMD}} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

  - Polarization Dependent Loss (PDL)

$$Y(f) = \mathbf{H}_{\mathrm{PDL}}X(f)+W(f) \quad \text{with} \quad \mathbf{H}_{\mathrm{PDL}} = \mathbf{U}^H \begin{bmatrix} \sqrt{1+\gamma} & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \mathbf{V}$$

with **U** and **V** two unitary matrices and $\gamma \in]-1,1[$ the PDL loss

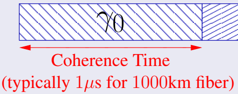
PDL coefficient: $\Gamma = 10\log_{10}(\frac{1+\gamma}{1-\gamma})$

- Extension : multi-modes fiber (not addressed in these slides)
- Nonlinear effect neglected hereafter

Random channel

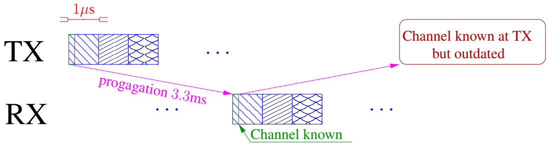
Block fading channel

θ and γ are time-varying!



Assuming transmission over $B = 50$GHz, so $T_s = 20$ps

- Channel static over 50,000 symbols!
- Channel known at RX (at very low cost)
- In contrast, channel unknown at TX due to propagation duration



  – outdated Channel State Information at the Transmitter (CSIT)
  – **TX has to view channel as <u>random</u>**

Capacity for random channels

- How defining capacity (which is channel-dependent) when channel is time-varying !
- Solution provided for wireless channel in `[Shamai1994]`

First case: when a codeword encounters every channel realization

Ergodic capacity

$$C = \mathbb{E}_{channel}\left[C(channel)\right]$$

Example: codeword of length 4320 with QPSK $\Rightarrow$ 43.2ns

- without interleaver, one codeword views one channel realization: ergodic capacity not suitable
- with interleaver of size 200Mbits, each bit within one codeword views one channel realization:
  - latency: 2.16 ms $\Rightarrow$ yes
  - interleaver size $\Rightarrow$ no

Capacity for random channels

- How defining capacity (which is channel-dependent) when channel is time-varying !
- Solution provided for wireless channel in [Shamai1994]

First case: when a codeword encounters every channel realization

Ergodic capacity

$$C = \mathbb{E}_{channel}\left[C(channel)\right]$$

Example: codeword of length 4320 with QPSK $\Rightarrow$ 43.2ns

- without interleaver, one codeword views one channel realization: ergodic capacity not suitable
- with interleaver of size 200Mbits, each bit within one codeword views one channel realization: ergodic capacity not suitable
  - latency: 2.16 ms $\Rightarrow$ yes
  - interleaver size $\Rightarrow$ no

Definition of outage probability

Second case: when a codeword encounters one channel realization

Outage probability

Codeword for which $C(channel)$ is less than the spectral efficiency R is in "outage"

$$P_{\mathrm{out.}} = Pr_{channel}(C(channel) < R)$$

- Realistic scheme when channel is slow block fading $\Rightarrow$ yes
- Realistic scheme according to our interleaver size $\Rightarrow$ yes

outage capacity suitable

What do we need for deriving outage capacity:

- Closed-form expression for $C(channel)$
- Statistical model for the channel (through either θ or γ)

MIMO channel capacity

Let us assume

$$Y(f) = \mathbf{H}(f)X(f) + W(f)$$

with

- $X(f)$ white Gaussian process with energy E_c
- $W(f)$ white Gaussian noise with variance N_0 per real dimension

Channel capacity [Telatar1995]

$$C(channel) = \int_{-B/2}^{B/2} \log_2 \left(\det \left(\mathbf{I}_2 + \frac{E_c}{2N_0} \mathbf{H}(f)\mathbf{H}(f)^{\mathrm{H}} \right) \right) df$$

Application to PMD

- As $\mathbf{H}_{\mathrm{PMD}}$ is a rotation, we have $\mathbf{H}_{\mathrm{PMD}}\mathbf{H}_{\mathrm{PMD}}^{\mathrm{H}} = \mathbf{I}_2$
- Consequently

$$C = 2B \log_2(1 + E_c/2N_0) \approx 2C_{SISO}$$

- Capacity is PMD insensitive and is twice due to PolMux

Capacity in PDL case

We have

$$C(\gamma) = B\log_2\left((1+\rho)^2 - \rho^2\gamma^2\right)$$

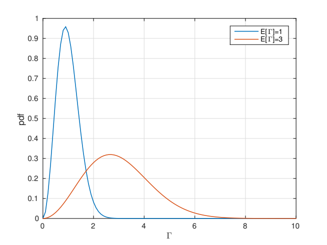
with $\rho = E_c/2N_0$

- Channel capacity depends on PDL value
- Outage probability is needed
- A statistical model of PDL value is required!

Γ Maxwellian model

$$p(x) = \begin{cases} \sqrt{\frac{2}{3}}\frac{x^2}{\sigma^3}e^{-\frac{x^2}{2\sigma^2}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

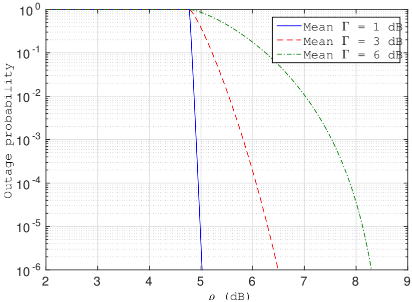
with $\sigma = \sqrt{\pi/8}\mathbb{E}[\Gamma]$



Outage probability derivations [Delesques2012]

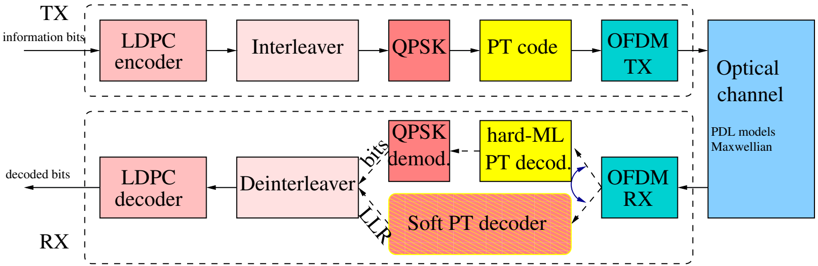
$$P_{\text{out.}} = \begin{cases} 1 & \text{if } \rho < \sqrt{2^{R/B}} - 1 \\ 2Q\left(\frac{T}{\sigma}\right) + \sqrt{\frac{2}{\pi\sigma^2}}\, T e^{\frac{-T^2}{2\sigma^2}} & \text{if } \sqrt{2^{R/B}} - 1 < \rho < \frac{1}{2}(2^{R/B} - 1) \\ 0 & \text{otherwise} \end{cases}$$

with $T = (20/\log(10))\mathrm{atanh}\left(\sqrt{1 - (2^{R/B} - 1 - 2\rho)/\rho^2}\right)$



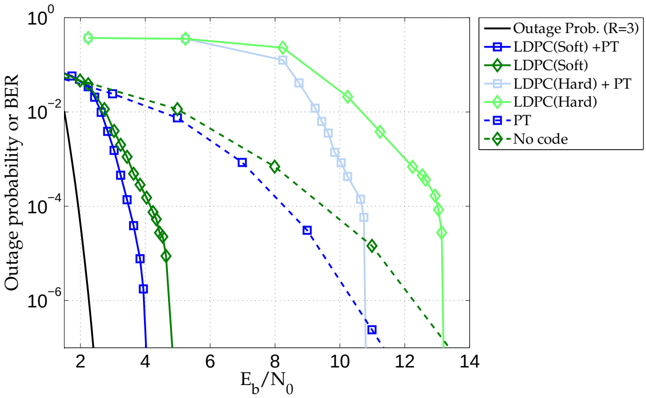
- far away from constant model
- beyond a SNR threshold, no outage
- fit well with simulations

Numerical illustrations (I)



- FEC Code: LDPC Code
- Modulation scheme: QPSK
- Polarization-Time code: Silver Code

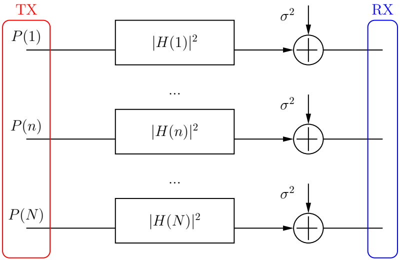
Numerical illustrations (II)



Remarks

- Soft decoding is significantly better than hard
- Gain offered by PT Codes and FEC are cumulative
- Best practical system is at 1.5 dB from the outage

What can we do with CSIT: an example (I)



- Data rate maximization
- Power constraint:

$$\sum_{n=1}^{N} P(n) = P_{\max}$$

  with maximum power $P_{\max}$
- Perfect CSIT

Problem: capacity optimization

$$[P(1)^*, \cdots, P(N)^*] = \arg \max_{P(1),\cdots,P(N)} \sum_{n=1}^{N} \log_2 \left(1 + |H(n)|^2 \frac{P(n)}{\sigma^2}\right)$$

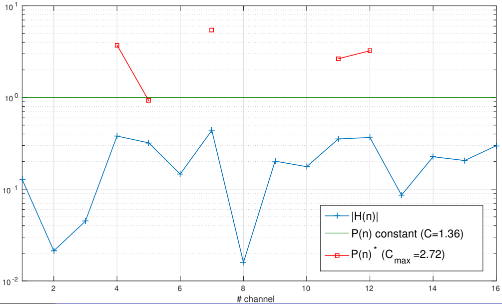
What can we do with CSIT: an example (II)

Solution: waterfilling [Shannon1948]

$$P(n)^* = \left(\nu - \frac{\sigma^2}{|H(n)|^2}\right)^+$$

with

- ν chosen s.t. $\sum_{n=1}^{N} P(n)^* = P_{\max}$
- $(\bullet)^+ = \max(0, \bullet)$



Conclusion

- On nonlinear impairments
  - Strong degradation of the performance if Gaussian-distributed
- On MIMO impairments
  - Only PDL leads to an issue
  - Powerful practical techniques almost achieve the fundamental limit
- Future works
  - What's really happened for capacity when QAM employed
  - Best waveform (and so probability distribution) for nonlinearity
  - Other MIMO issues : multi-modes fibers [Awwad2015]

References

[Shannon1948] C. Shannon, A mathematical theory of communication, Bell System Technical Journal, 1948

[Shamai1994] L. Ozarow, S. Shamai, A. Wyner, Information theoretic consideration for cellular mobile radio, IEEE Trans. Vehicular Technology, 1994

[Brandt1997] K. Peddanarappagari, M. Brandt-Pearce, Volterra series transfer function of single-mode fibers, IEEE Jnl of Lightwave Technology, 1997

[Essiambre2010] R.J. Essiambre, G Kramer, P. Winzer, G. Foschini, B. Goebel, Capacity limits of optical fiber networks, IEEE Jnl of Lightwave Technology, 2010

[Turitsyn2003] K. Turitsyn, S. Derevyanko, I. Yurkevich, Information capacity of optical fiber channels with zero average dispersion, Physical Review Letters, 2003

[Tang2001] J. Tang, The shannon channel capacity of dispersion-free nonlinear optical fiber transmission, IEEE Jnl of Lightwave Technology, 2001

[Tang2006] J. Tang, A comparison study of the Shannon channel capacity of various nonlinear optical fibers, IEEE Jnl of Lightwave Technology, 2006

[Mitra2001] P. Mitra, J. Stark, Nonlinear limits to the information capacity of optical fiber communications, Nature, 2001

[Kahn2004] J. Kahn, P. Ho, Spectral efficiency limits and modulation/detection techniques for DWDM systems, IEEE Jnl Selected topics in Quantum Electronics, 2004

[Poggiolini2011] P. Poggiolini, A. Carena, V. Curri, G. Bosco, F. Forghieri, Analytical Modeling of Nonlinear Propagation in Uncompensated Optical Transmission Links, IEEE Photonics Letters, 2011

[Narimanov2002] E. Narimanov, P. Mitra, The channel capacity of a fiber optics communication system: Perturbation theory, IEEE Jnl of Lightwave Technology, 2002

[Delesques2012] P. Delesques, Analyses of transmission and switching capacities in optical networks, PhD thesis, Telecom ParisTech, 2012

[Telatar1995] E. Telatar, Capacity of Multi-antenna Gaussian Channels, Europena Transactions on Telecommunications, 1995

[Awwad2015] E. Awwad, Emerging space-time coding techniques for optical fiber transmission systems, PhD thesis, Telecom ParisTech, 2015