Coding for Efficient Communications (in 5G)
Workshop on 5G, CVR College of Engineering

Prasad Krishnan

International Institute of Information Technology, Hyderabad

June, 2016

5G Communications

- Higher data rates, Lower latencies
- mmWave communication, small cells.
- Massive MIMO
- Full Duplex Communications.
- D2D communications, IoT.
- Local Caching.
- and more..

A word of caution

"

..in many cases, the term 5G is bandied about as a panacea that already exists. Thats why **Seizo Onoe, CTO of NTT DOCOMO**, Japan's largest mobile carrier, is traveling around to conferences trying to keep everyone's expectations in check. "*In the early 2000s, there was a concrete 4G technology but no one called it 4G*," Onoe laments. "*Today, there are no contents of 5G but everyone talks about 5G, 5G, 5G.*"

" - **5 Myths about 5G, IEEE Spectrum, 25th May 2016**.

Outline

Source material

- IEEE Spectrum
- Standard books on LDPC Codes
- METIS 2020 (Mobile and wireless communications Enablers for the Twenty-twenty Information Society 5G) documents.
- 5G proposal documents from Samsung, Huawei, etc.
- Google...

Outline

The Channel

- Given to us by nature (can optimise, but fundamental nature cannot be changed).
- Modelling through mathematics (Models are not exact).
- Making appropriate assumptions are very important.
- Classic case - difference between Wired Point-to-Point (ex: telephone) and Wireless Channels.

Channel Models: AWGN and Binary Symmetric Channel

- AWGN : Typical Model for point to point (Noise signals are from a Gaussian Random Process)
- Output Y = Input X+ Noise Z, $Z \sim$N(0,N$_0$/2) (Sampling)
- Characterised by the conditional distribution $p(y|x)$ (For AWGN, $p(y|x) = \frac{1}{\sqrt{(\pi N_0)}} e^{-\frac{(y-x)^2}{N_0}}$)

Channel Models: AWGN and Binary Symmetric Channel

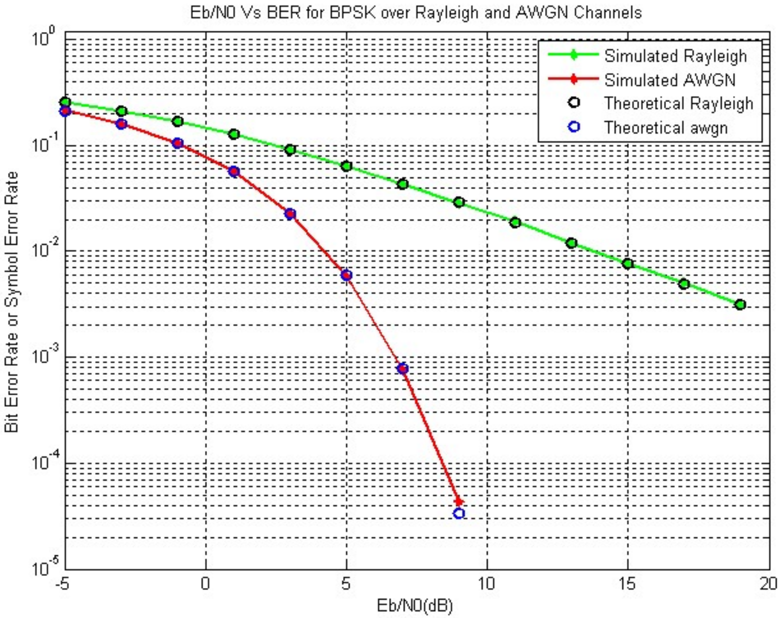
- AWGN : Typical Model for point to point (Noise signals are from a Gaussian Random Process)
- Output Y = Input X+ Noise Z, $Z \sim N(0,N_0/2)$ (Sampling)
- Characterised by the conditional distribution $p(y|x)$ (For AWGN, $p(y|x) = \frac{1}{\sqrt{(\pi N_0)}} e^{-\frac{(y-x)^2}{N_0}}$)
- Quantise X and Y to two values (for example using BPSK modulation) - Binary Symmetric Channel.

Channel Models: AWGN and Binary Symmetric Channel

- AWGN : Typical Model for point to point (Noise signals are from a Gaussian Random Process)
- Output Y = Input X+ Noise Z, $Z \sim$N(0,$N_0/2$) (Sampling)
- Characterised by the conditional distribution $p(y|x)$ (For AWGN, $p(y|x) = \frac{1}{\sqrt{(\pi N_0)}} e^{-\frac{(y-x)^2}{N_0}}$)
- Quantise X and Y to two values (for example using BPSK modulation) - Binary Symmetric Channel.

- Wireless Channel: $Y = hX + Z$ (h is a random variable that models fading).

Fading Channel results in poor BER



Channel Capacity : Shannon's result

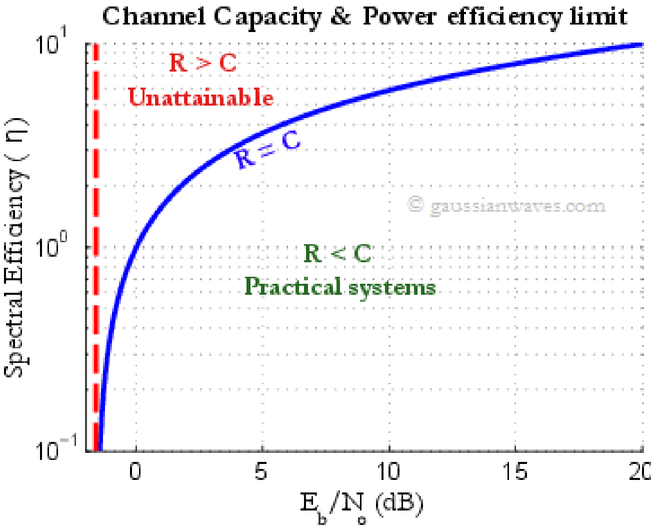
- Channels are noisy inherently. This limits the rate of communication.
- Capacity is maximum rate of transmission ($b/s/Hz$) - this is a function of $p(y|x)$ (and $p(x)$)
- Shannon's Theorem for point-to-point channels **Any rate of transmission (**$b/s/Hz$**) below Capacity is achievable. Any rate larger than C is always unachievable**

Capacity of AWGN

- For AWGN with bandwidth W:: Capacity $=$ $\frac{1}{2}log(1+SNR) = \frac{1}{2}log(1+\frac{P}{N_0 W})$.
- *Channel Coding* (Some appropriate function of the message bits should be transmitted, with appropriate decoding)
- **Only Existence of Good Codes is shown by Shannon.**
- Construction of 'good' codes has happened (for AWGN channels) over the last several decades since Shannon.

Channel Coding Block Diagram

Capacity curve for AWGN



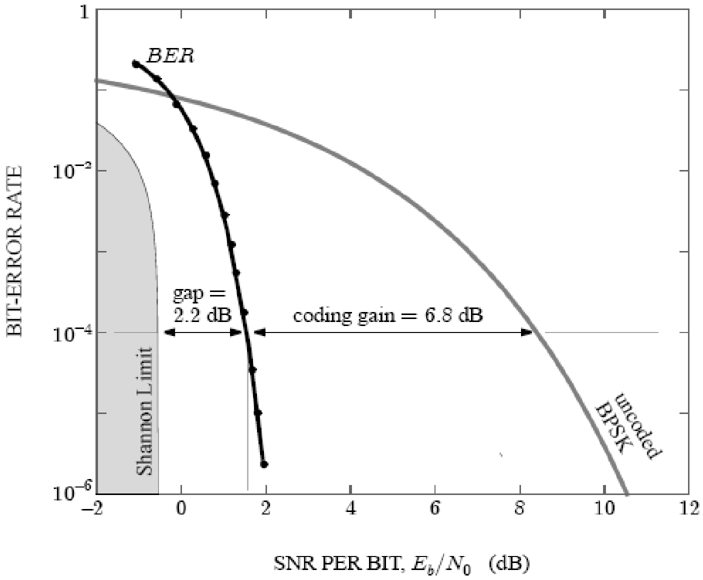
Channel Codes: Theory and Practice

- How to map the messages to the channel? What are the parameters of interest?
- Good error correcting performance.
- High rate of communication (close to Shannon limit) for a given *SNR*.
- For a given rate and given probability of error, should give best **coding gain** (gain in SNR(dB) over uncoded case).
- Low encoding complexity and decoding complexity (what “low” means changes with technology)

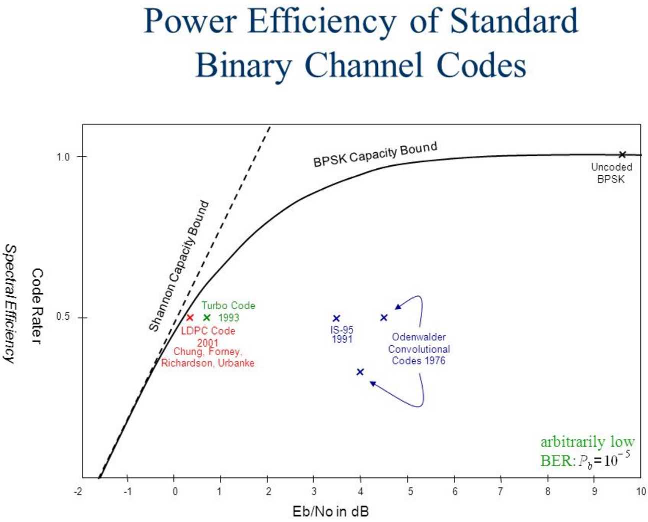
Good codes for AWGN Channels today and their characteristics

- Low Encoding Complexity Codes - Linear Codes.
- Two major classes of Linear Codes.
  - Block Codes (fixed block length)
  - Convolutional Codes (stream codes)
- Block Code variation : LDPC Codes
- Convolutional Code Variation : Turbo Codes.
- Both are 'long codes'. Transmission of the order of 1000s of bits are required before decoding.
- Require probabilistic decoding strategies to perform well.
- Performance - only a fraction of dB away from Shannon Capacity.

Coding gain illustration



Turbo Codes and LDPC Codes along Shannon Capacity Curve



Codes for 5G Communication

With the stringent demands of 5G communications, the best among the current candidate codes are considered.

- Turbo Codes (used in 4G already).
- LDPC Codes with Spatial Coupling (Better than Turbo - since 1960/1980s)
- Polar Codes (recent - 2008)
- Sparse Regression Codes (recent - 2010)

Outline

LDPC Codes Basics : Linear Block Codes

- Let $\boldsymbol{u}$ be the message vector and $\boldsymbol{x}$ is the corresponding codeword.
- The map connecting $\boldsymbol{u}$ and $\boldsymbol{x}$ is a linear map, i.e., connected by a *generator matrix* G.

$$\boldsymbol{x} = \boldsymbol{u}G$$

- G is a full-rank matrix of size $k \times n$ $(k \leq n)$.
- The code is called a linear code encoding k message symbols to n-length codeword.
- Corresponding to G, there is another full-rank matrix called the Parity Check matrix $H_{(n-k)\times n}$ such that $GH^T = \mathbf{0}$.
- Note that for any codeword $\boldsymbol{x}$, $H\boldsymbol{x}^T = \mathbf{0}$.

LDPC Codes

LDPC

If H is sparse (more 0s than 1s) then, the code is called a Low Density Parity Check Code (LDPC).

- **Regular LDPC Codes**: Rows of H have a constant weight (no. of 1s). Columns of H have a constant weight.
- Example construction : Take a single vector and shift it to get the rows.
- H matrix can be represented using a bipartite graph (Tanner graph)

LDPC Tanner Graph H matrix

- x_1 code bit is involved in two check bits - $\{m1, m3\}$.
- There are two code bits $\{x_1, x_2\}$ involved in the check bit m_1.

How to decode?

- Good codes use probabilistic decoding (algebraic decoding of certain codes exist, but such codes don't perform well).
- Basic rule : **Maximum Aposteriori Probability (MAP) Rule**
- MAP Rule: Choose that codeword $\boldsymbol{x}$ which is most probable given the output $\boldsymbol{y}$ received

$$\hat{\boldsymbol{x}} = argmax_{\boldsymbol{x} \in C} p(\boldsymbol{x}|\boldsymbol{y})$$

- Once $\hat{\boldsymbol{x}}$ is obtained, the receiver can get the corresponding message estimate $\hat{\boldsymbol{u}}$ (as it has the generator matrix).

Bit-wise decoding

- Estimate $\hat{x}_i$ one by one.

$$\begin{aligned}\hat{x}_i &= argmax_{x_i:\boldsymbol{x}\in C}\, p(x_i|\boldsymbol{y}) \\ &= argmax_{x_i:H\boldsymbol{x}^T=\boldsymbol{0}}\, p(x_i|\boldsymbol{y}) \\ &= argmax_{x_i\in\{0,1\}}\, p(x_i|\boldsymbol{y}), \text{under the condition that} \\ &\qquad \text{the check bits involving } x_i \text{ are zero.}\end{aligned}$$

- Choose $\hat{x}_i = 1$, if $p(x_i = 1|\boldsymbol{y}) > p(x_i = 0|\boldsymbol{y})$.
- Log-likelihood-ratio :

$$LLR(x_i|\boldsymbol{y}) = log\left(\frac{p(x_i = 1|\boldsymbol{y})}{p(x_i = 0|\boldsymbol{y})}\right) > 1.$$

- Similar for $\hat{x}_i = 0$.

Decoding LDPC Codes

- For general linear codes, this technique has exponential complexity with growing code length (n).
- How does LDPC codes get over the high complexity? - **Code structure, and 'belief propagation' (message passing) decoding**.
- $$\begin{aligned} LLR(x_i|\boldsymbol{y}) &= log\left(\frac{p(x_i=1|\boldsymbol{y})}{p(x_i=0|\boldsymbol{y})}\right) \\ &= log\frac{p(y_i|x_i=1)}{p(y_i|x_i=0)}) + log\left(\frac{p(x_i=1|y_j:j\neq i)}{p(x_i=0|y_j:j\neq i)}\right). \\ &= \text{ Intrinsic information} + \text{ Extrinsic information.} \end{aligned}$$

LDPC Codes - Belief Propagation Decoding

- It is easy to calculate the Intrinsic Information $(log\frac{p(y_i|x_i=1)}{p(y_i|x_i=0)})$ from the channel distribution.
- Extrinsic information : Very hard in general, but LDPC code structure makes it less complex.
- Extrinsic info: $log\left(\frac{p(x_i=1|y_j:j\neq i)}{p(x_i=0|y_j:j\neq i)}\right) = \sum_{m\in\mathcal{M}_i} D_{m,i}$, where
  - $\mathcal{M}_i$ is the set of check bits involving x_i.
  - $D_{m,i}$ is a function of Extrinsic information corresponding to all other code bits x_j $(j \neq i)$ which are involved in m^{th} check bit in $\mathcal{M}_i$.

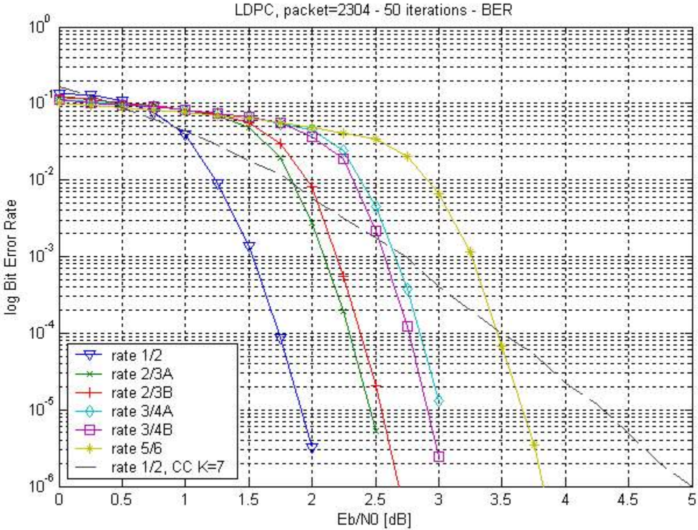
Belief propagation on the Tanner graph - From leaves to root

Belief Propagation on Tanner graph

The decoding is therefore recursive and iterative in nature.

- Algorithm: Unwrap the Tanner graph (analysis is easy if it is cycle-free).
- The LLR at each code bit x_i is initialised with the intrinsic information ($log \frac{p(y_i|x_i=1)}{p(y_i|x_i=0)}$).
- For a given number of iterations
  1. Process from Leaves to Root.
  2. At any check bit layer: Compute the D values at check bits using the LLR at the above layer and pass it to the below layer for calculating LLR values.
- After set number of iterations (around 10-20 is practical and gives good performance), declare the final LLRs for all the codebits.
- Choose the code bits according to the LLR values.

LDPC Codes performance



Outline

Some drawbacks

- LDPC codes (as well as the other candidates) are 'long codes' (10000 bits). This leads to latency (1000 bits or so) and higher power consumption.
- Low latency, low power, short block-length codes are very much in need.
- Improvements for short block-lengths are still open.

Didn't talk about

- LDPC Codes with Spatial Coupling.
- Polar Codes, Sparse Regression Codes.
- Space-time Codes for Large MIMO systems.
- Coded Caching for D2D communication.
- Video Coding
- Network Coding.