$W \overset{d}{=} \text{Bin}\left(2^L - 1, \left(\frac{1}{2}\right)^{N(1-\varepsilon)}\right)$ "number of confusing code word".

(in distribution; $2^L - 1$: total # of available code words; $\left(\frac{1}{2}\right)^{N(1-\varepsilon)}$: proba 2 ≠ codewords agree on a bit)

$\Rightarrow \mathbb{E}(W) = 2^{N(R-1+\varepsilon)} \to 0$ if $\varepsilon < 1 - R$ $\Rightarrow$ $P_r[W \neq 0] \to 0$ - (↳ integer!)

First moment / Markov inequality for strictly positive codeword -

So that no confusion if $R < 1 - \varepsilon = C_{BEC}$ → Shannon theorem! ∎

Even the random codes achieve capacity - (Here we considered averages for codewords etc...)

HOMEWORK: proof for the BSC (text on website) -

Rk we proove that there exist such codes, but we did not show that we could not do better than this bound - if $R > C \Rightarrow P_{err} > 0$ coming from $H(\underline{X}|\underline{Y}) > 0$, Fano inequality -

⚠ In practice, encoding and decoding the Random Code ensemble takes exponential time and memory - $\mathcal{C}$ has to be described by $N 2^L$ bits. The encoding is easy by looking up the table - But the decoding is NP hard as one need to look for the corresponding match in the table - The randomness hinders any compression, we can only do exhaustive (es) search in the table -

## 4 LOW DENSITY PARITY CHECK CODES (LDPC). → putting some structure ;)

### 4A Linear codes

- $\{0,1\}^N$ is a linear space over $\mathbb{Z}_2 = \{0,1\}$ "scalars", addition mod 2, multiplication
- The codebook $\mathcal{C} \subset \{0,1\}^N$ is said to be a linear code if $\mathcal{C}$ is a linear subspace of $\{0,1\}^N$.
  - $\underline{x} + \underline{y} = (x_1 + y_1 \text{ mod } 2, x_2 + y_2, \dots x_N + y_N) \in \mathcal{C}$
  - $\underline{x} + \underline{x} = \underline{0} \in \mathcal{C}$ (the origin belongs to the linear subspace).
  - ↳ $\underline{0}$ is always a codeword of a linear code
  - All codewords are equivalent (can always make gauge transformation to change position of origin).

12/07/2017 RECALL THE IDEA:

$\#\mathcal{C} \ll \#\{0,1\}^N$ : if the codewords are sufficiently far from one another, we should be able to identify them despite of the noise corruption by the channel -
$2^L \qquad 2^N$

$R = \frac{L}{N} < 1$

↳ the codebook can be described by a set of $M$ linear equation, or equivalently as the Kernel of an operator: $\mathcal{C} = \ker H = \{x \in \{0,1\}^N : H\underline{x} = \underline{0}\}$. $H = \begin{bmatrix} 0\,0\,1\,0\,\dots \\ 0\,0\,0\,1\,\dots \\ \dots \end{bmatrix}$ ($M$ rows, $N$ columns)

↳ one line of $H\underline{x} = 0 \Rightarrow x_3 + x_{14} + x_{25} = 0 =$ nb of 1's among $\{x_3, x_{14}, x_{25}\}$ must be even → hence the name PARITY CHECKS

prop: $\dim \mathcal{C} =$ nb of linearly independent codewords

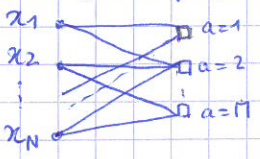
$|\mathcal{C}| = 2^{\dim \mathcal{C}}$

number of codewords = number of solutions

↳ suppose $M$ equations in $H\underline{x} = \underline{0}$ are independent, then $|\mathcal{C}| = 2^{N-M}$ -

general result of linear algebra: $N = \underbrace{\dim \ker H}_{=L} + \underbrace{\text{rk}(H)}_{=M \text{ if } H \text{ is full-rank}} \to R = \frac{N-M}{N} = 1 - \frac{M}{N}$ $M < N$.

Going back to the parity check representation:
(Tanner) factor graph representation of H.




edge between $i$ and $a$ $\Leftrightarrow$ $H_{ai}=1$

$\forall a$ $(\underline{x}) = 1$ $\left(\sum_{i=1}^{N} H_{ai}x_i = 0\right)$

↳ set of neighbors: $\partial i = \{a \in \{1,\dots M\} : H_{ai}=1\}$
$\partial a = \{i \in \{1,\dots N\} : H_{ai}=1\}$

Rk: This is equivalent to the CSP: XORSAT

$$\begin{cases} 0+0=0 \\ 0+1=1+0=1 \\ 1+1=0 \end{cases} \Rightarrow \begin{cases} TT \to F \\ FF \to F \\ TF \to T \end{cases}$$

or Ising spins: $\sigma_i = (-1)^{x_i} = \begin{cases} 1 & x_i=0 \\ -1 & x_i=1 \end{cases}$ $\qquad x_i+x_j+x_k=0 \Leftrightarrow \sigma_i\sigma_j\sigma_k = 1$ 3-spin int.

## 4B Definition of the simplest ensemble (Gallager 62)

⇒ random matrix H, sparse or low density

↳ uniform at random under the condition $\begin{cases} |\partial i| = \ell & \forall i \text{ nb of 1s in the } i\text{th column} \\ |\partial a| = k & \forall a \text{ nb of 1s in the } a\text{th line} \end{cases}$

(implies $N\ell = Mk$, $R = 1 - \frac{M}{N} = 1 - \frac{\ell}{k}$ $\ell < k$).

Rk: how to proceed to generate H? $\ell=3$ $k=4$

→ find random matching between half-edges
(neglect case when 2 half edges connect same □ and ○, $p = 1/N$
as well as  $\left(p = \frac{1}{N^3}\right)$ )

↳ we are not imposing that rows are independent, but in the thermodynamic limit, there are only a sub-extensive number of not independent equations.

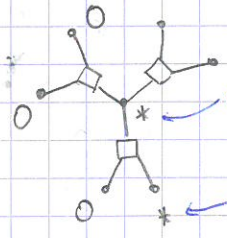
A crucial property of this construction: locally tree like graph = the neighborhood at distance $t$ of $i$ (arbitrarily chosen) is a tree with $p$.


↳ $t$ fixed, $N \to +\infty$ : choose finite number of neighbors among $N$ possible → proba $\left(\frac{1}{N}\right)$ → probability to pick twice same node in the process vanishing.

## 4C Analysis of BEC

Assume that we have sent $\underline{x} = (0,0,\dots 0)$ (w.l.o.g.) ⟶ received $\underline{y} = (0,0,*,*\dots *,0,0)$ ($\varepsilon$ fraction erased)

local picture:




the linear relation allow to guess for some non-received bits each time there exist a parity check with a single *. Can be repeated with newly guessed bits!

↳ when stopped: no more * left ⇒ PERFECT DECODING
remains □ with at least 2 erasures.

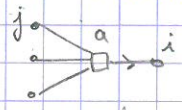
Now the question is, as function of the parameters of the model $(\varepsilon, \ell, k)$, when is perfect decoding possible?

↳ to answer: let's rephrase the problem as a message passing algorithm:

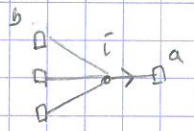
message from clause a to variable i $\quad a \xrightarrow[\mu_{a\to i}]{} i \quad$ $\mu_{a\to i} = \begin{cases} 0 & \text{"I'm sure you're a 0"} \\ * & \text{"I don't know"} \end{cases}$

RULES OF COMPUTATION OF THE MESSAGES: from initialization $\mu_{a\to i} = *$, $h_{i\to a} = y_i$.

$$\mu_{a\to i} = \begin{cases} 0 & \text{if } \forall j \in \partial a \setminus i \quad h_{j\to a} = 0 \\ * & \text{otherwise} \end{cases}$$




$$h_{i\to a} = \begin{cases} 0 & \text{if } y_i = 0 \text{ or } \forall b \in \partial i \setminus a \quad \mu_{b\to i} = 0 \\ * & \text{otherwise} \end{cases}$$




↳ CONVERGENCE: This case of belief propagation has a monotone dynamics as there can only be $* \to 0$ and no $0 \to *$. This implies that, whichever the update schedule, the algorithm reaches the same fixed point as the first decoder.

probabilistic analysis:

$$\begin{cases} \eta^{(t)} = \mathbb{P}\left[\mu_{a\to i}^{(t)} = 0\right] \\ \xi^{(t)} = \mathbb{P}\left[h_{i\to a}^{(t)} = 0\right] \end{cases}$$

randomness from: . construction of $\mathbb{H}$
. outcome of the channel (locations of erasure)
. edge $a \to i$ u.a.r. (which one considering)

From $\eta^{(0)} = 0$, updates: $\begin{cases} \eta^{(t+1)} = \left(\xi^{(t)}\right)^{k-1} \\ \xi^{(t)} = 1 - \varepsilon\left(1 - \eta^{(t)}\right)^{\ell-1} \end{cases}$

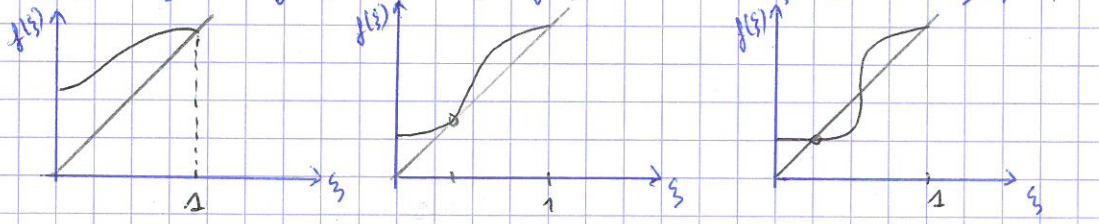
→ assume independence between $\xi_{i\to a}$'s is coming from the tree-like assumpt°

$\ell-1$ neighbors don't know about me

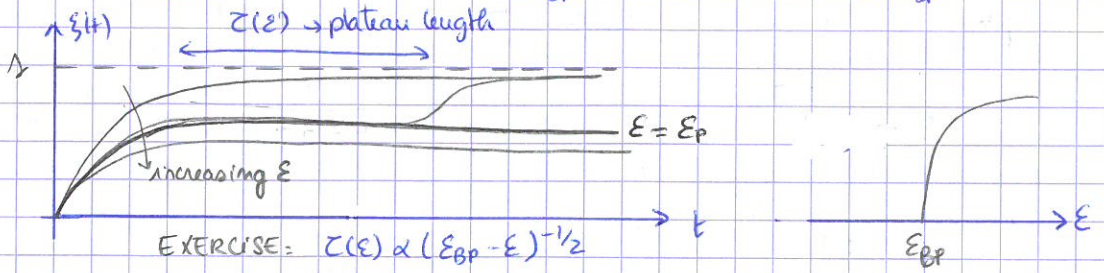
↑ proba of erasure at i

↳ closing on $\xi_s$: $\xi^{(t+1)} = f(\xi^{(t)})$ with $f(\xi) = 1 - \varepsilon\left(1 - \xi^{(t)_{k-1}}\right)^{\ell-1}$ assume $k \geq 3$, $\ell \geq 3$

fixed point analysis




evolution of $\xi^{(t)}$




↳ Very non trivial statement: $\varepsilon < \varepsilon_{BP}$ we have a perfect decoding in linear time.

How far is this algorithm from Shannon's theorem?

e.g. $\ell = 3$, $k = 4$ → $\varepsilon_{BP} = 0.647$ et $R = 1 - \frac{3}{4} = 0,25$ $\quad \varepsilon_{SPL} = 0,75$

↳ $\varepsilon_{BP} < \varepsilon_{SPL}$: - is the code bad?
- is the algorithm to decode bad?

$\varepsilon > \varepsilon_{BP}$: how many codewords are compatible with received $y$?

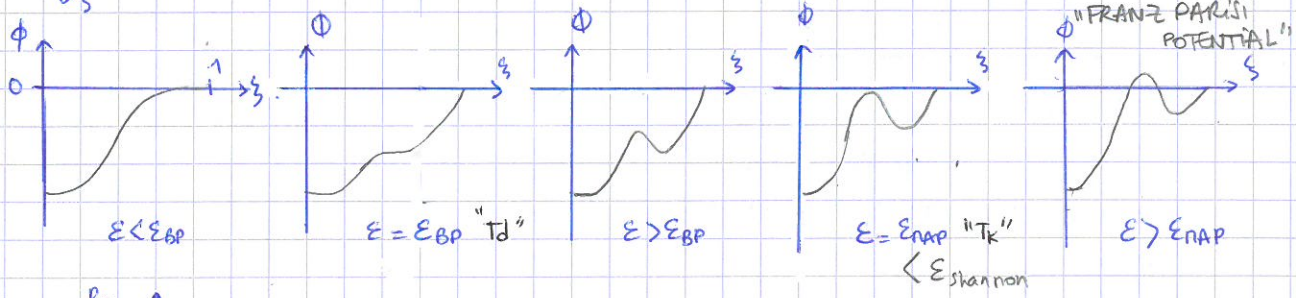
$N' = N \mathcal{E}(1-\xi^{k-1})^e$ $\qquad \xi = \xi^{(t=\infty)}$

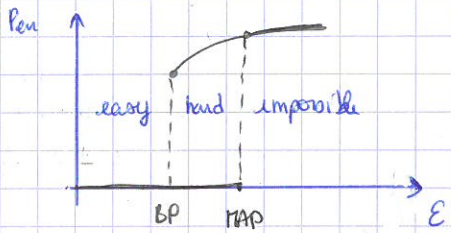
$M' = M\left(1-\xi^k - k\xi^{k-1}(1-\xi)\right)$ → at least 2 $\star$

$\Phi(\mathcal{E},\xi) = \frac{1}{N}(N'-M') = \mathcal{E}(1-\xi^{k-1})^e - \frac{e}{k}\left(1-\xi^k - k\xi^{k-1}(1-\xi)\right)$

↳ if independant, $2^{N'-M'}$ codewords compatible with received message.

$\frac{\partial \Phi}{\partial \xi} = 0$ when $\xi$ is a fixed point of $\xi = f(\xi)$.




to perfectly decode (≠ do better than a random guess)

Rk: Here, one can always do in the hard phase gauss elimination $O(N^3)$ - only for linear codes and BEC.

Rk: How close $\mathcal{E}_{MAP}$ to $\mathcal{E}_{shannon}$ ?
More generic random ensemble → $\{\lambda_e$ : fraction of variables of degree e
$\{\rho_k$ : checks k

↳ $\mathcal{E}_{BP}$, $\mathcal{E}_{MAP}$ depends on $\lambda$ and $\rho$ : their exist degree sequences such that $\mathcal{E}_{BP} = \mathcal{E}_{sh} - \delta \ \forall \delta > 0$.

## CONCLUSION AND PERSPECTIVE

Why information theory is relevant ?

↳ we have examined a set of variables $\underline{x} = (x_1, \dots x_N) \in \mathcal{X}^N$ { random
under a set of parity checks = quenched disorder $\underline{J}$ → graph of interactions { locally tree like

↳ from which we tried to study a probability distribution $\mu(\underline{x}, \underline{J}) = \frac{1}{Z(\underline{J})} \prod_{i=1}^{N} w_i(x_i, \underline{J}) \prod_a w_a(\underline{x}_{\partial a}, \underline{J})$.

↳ we have used the cavity method to compute "$Z$" $\quad \Phi = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}[\ln(\underline{J})]$
and marginals $\quad \mu(x_i, \underline{J})$.

↳ CAVITY METHOD = BP = MESSAGE PASSING:
- exact relation on a tree with message passing
- analysis on a random graph → RS
→ RSB in more complicated method.

Bikshadh