

# LDPC codes and Message Passing decoders

– a brief review –

Valentin Savin

July 2013

# Outline

## **PART 1:**

LDPC codes and Message Passing (MP) decoders – brief review

## **PART 2:**

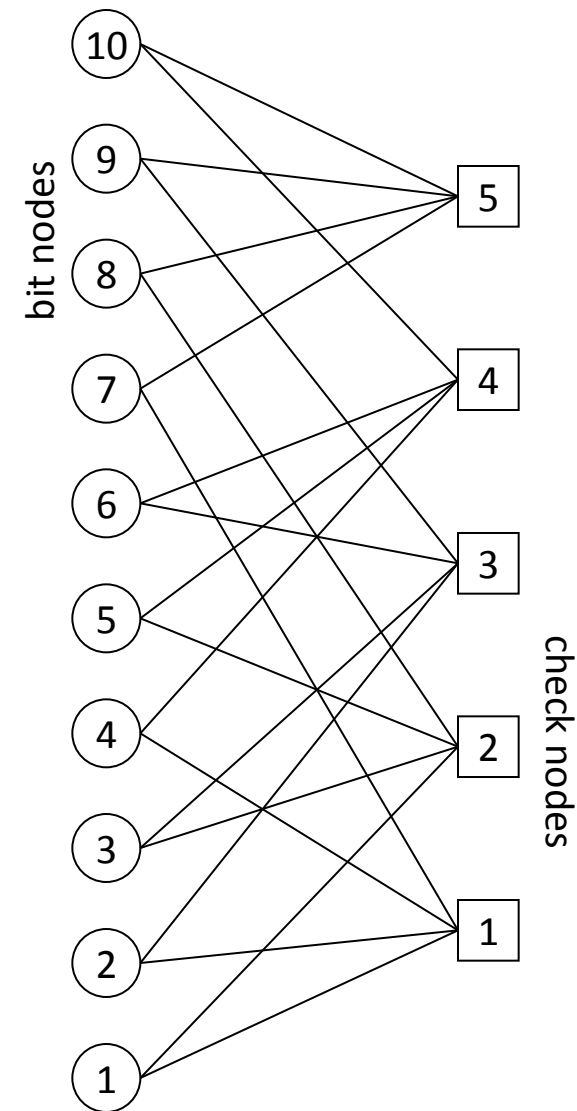
Fixed-point implementation and “noisy” decoders

# Graphical representation of LDPC codes

## LDPC code

- Linear block code defined by a sparse parity-check matrix  $H$
- $H$  is advantageously represented by a *bipartite (Tanner) graph*
  - circles: *variable-nodes* (or *bit-nodes*) representing coded bits (columns of  $H$ )
  - squares: *check-nodes* representing or parity-check equations (rows of  $H$ )
  - Graph *edges* correspond to the non-zero entries of  $H$
- Codeword: vector  $(x_1, \dots, x_N)$  such that  $H * (x_1, \dots, x_N)^T = 0$ 
  - The sum modulo 2 of bits connected to any check-node is zero

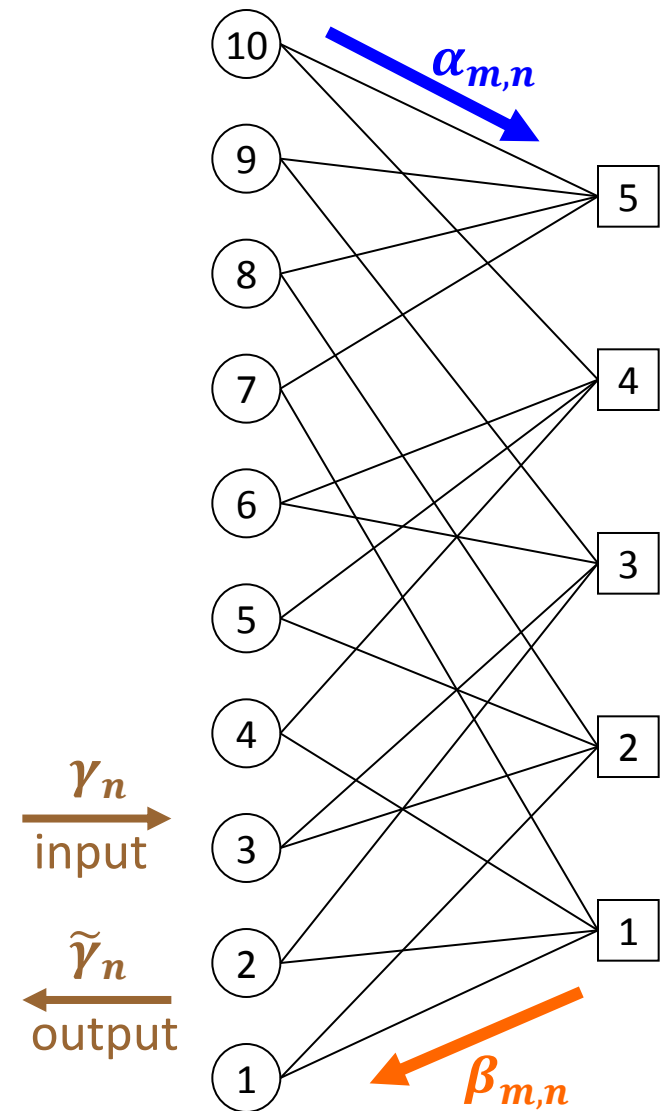
$$H = \begin{matrix} & \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \end{matrix} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \end{matrix}$$



# Message-Passing (MP) decoders

- Iterative exchange of messages between:
  - variable-nodes ( $n = 1, \dots, 10$ , on the left), and
  - check-nodes ( $m = 1, \dots, 5$ , on the right)
- Extrinsic-information exchange
  - $\beta_{m,n} = \text{funct}(\alpha_{m,n'}; n' \in H(m) \setminus n)$
  - $\alpha_{m,n} = \text{funct}(\gamma_n, \beta_{m',n}; m' \in H(n) \setminus m)$
- A posteriori information
  - $\tilde{\gamma}_n = \text{funct}(\gamma_n, \beta_{m,n}; m \in H(n))$
  - hard-decision is taken based on  $\tilde{\gamma}_n$  values

Remark:  $\gamma_n$  = input LLR;  $\tilde{\gamma}_n$  = output (AP-)LLR



# Message-Passing (MP) decoders

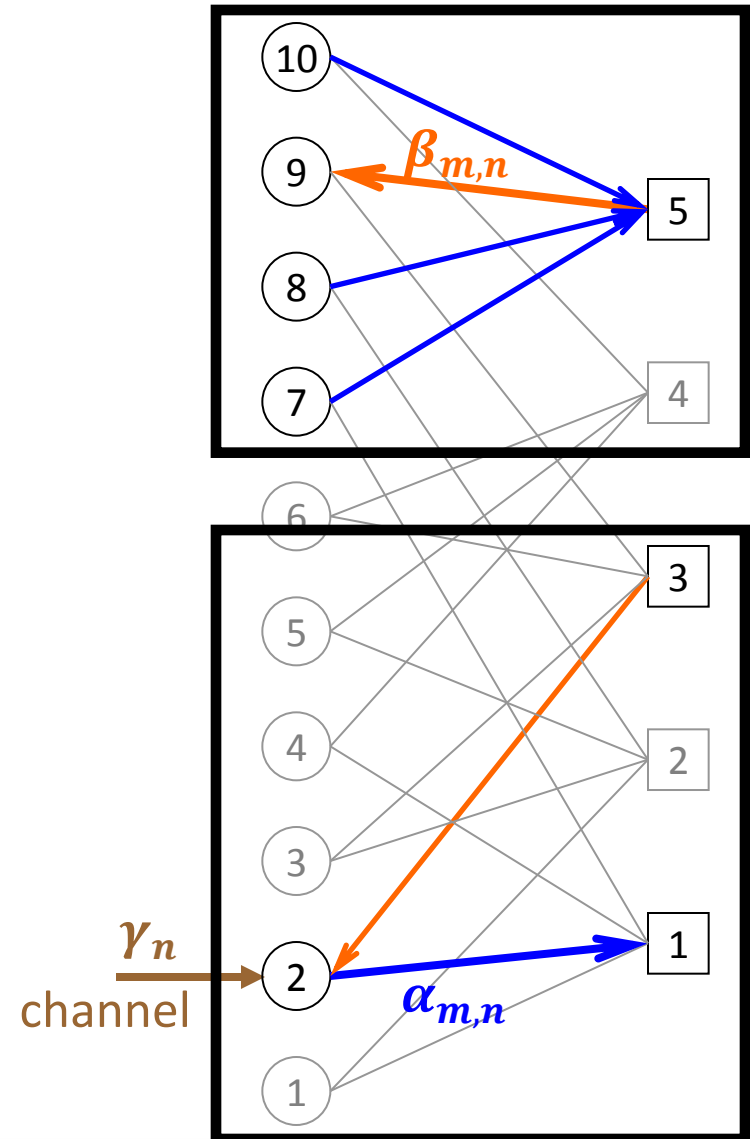
- Extrinsic-information exchange

CNU (Check-Node Unit)

- $\beta_{m,n} = \text{funct}(\alpha_{m,n'}; n' \in H(m) \setminus n)$

VNU (Variable-Node Unit)

- $\alpha_{m,n} = \text{funct}(\gamma_n, \beta_{m',n}; m' \in H(n) \setminus m)$



# Message-Passing (MP) decoders

- Several MP decoders, distinguished by the formulae used to compute check and variable-node messages
  - Gallager A, Gallager-B
  - Sum-Product (or Belief-Propagation)
  - Min-Sum

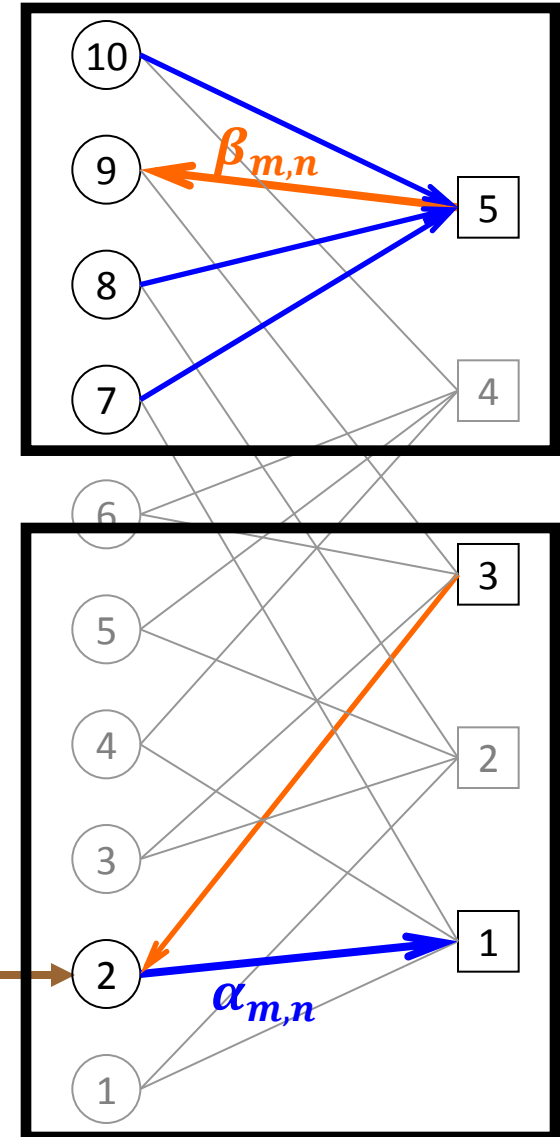
- Sum-Product decoding**

$$\beta_{m,n} = \left( \prod_{n' \in H(m) \setminus n} \text{sgn}(\alpha_{m,n'}) \right) \phi \left( \sum_{n' \in H(m) \setminus n} \phi(|\alpha_{m,n'}|) \right)$$

where  $\phi(x) = \log \left( \frac{e^x + 1}{e^x - 1} \right)$

$$\alpha_{m,n} = \gamma_n + \sum_{m' \in H(n) \setminus m} \beta_{m',n}$$

The outgoing  $\alpha_{m,n}$  message is the sum on channel LLR and incoming  $\beta_{m,n}$  messages



# Message-Passing (MP) decoders

- Several MP decoders, distinguished by the formulae used to compute check and variable-node messages
  - Gallager A, Gallager-B
  - Sum-Product (or Belief-Propagation)
  - Min-Sum

## Min-Sum decoding

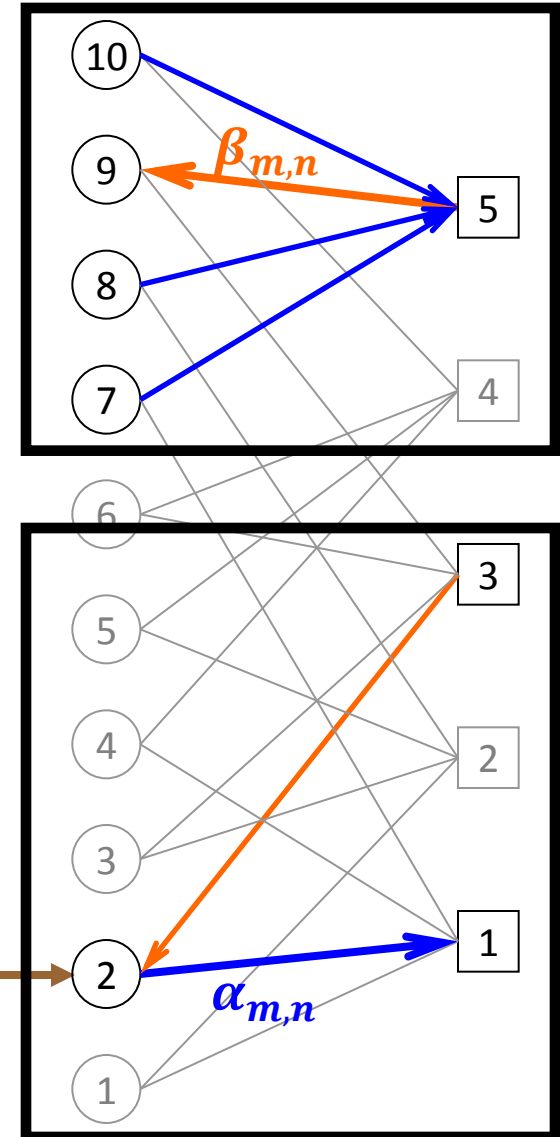
$$\beta_{m,n} = \left( \prod_{n' \in H(m) \setminus n} \text{sgn}(\alpha_{m,n'}) \right) \min_{n' \in H(m) \setminus n} (|\alpha_{m,n'}|)$$

The sign of the outgoing  $\beta_{m,n}$  message is the product of the signs of the incoming  $\alpha_{m,n'}$  messages

The absolute value of the outgoing  $\beta_{m,n}$  message is the minimum of the absolute values of the incoming  $\alpha_{m,n'}$  messages

$$\alpha_{m,n} = \gamma_n + \sum_{m' \in H(n) \setminus m} \beta_{m',n}$$

The outgoing  $\alpha_{m,n}$  message is the sum on channel LLR and incoming  $\beta_{m,n}$  messages



# Message-Passing (MP) Scheduling

- Order in which messages are passed between check and variable nodes
- Also determines the order in which CNUs and VNUs are processed

- **Flooding<sup>(\*)</sup> scheduling**

At each decoding iteration:

- all CNUs are processed and all check-node messages ( $\beta_{m,n}$ ) are sent from check to variable-nodes
- then all VNUs are processed, and all variable-node messages ( $\alpha_{m,n}$ ) are sent from variable to check-nodes

(\*) variable and check-nodes are “flooded” by incoming messages

- **Layered scheduling**

- parity check matrix is partitioned in horizontal layers

At each decoding iteration:

- all check nodes in one layer are processed and outgoing messages ( $\beta_{m,n}$ ) are sent to their neighbor variable nodes
- variable-nodes connected to check-nodes in the layer are “updated” (explained later on)
- move to next layer



# Min-Sum decoding with flooding scheduling

**Initialization:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\gamma_n = \log(\Pr(x_n = 0 | y_n) / \Pr(x_n = 1 | y_n))$$

$$\alpha_{m,n} = \gamma_n$$

variable-node messages are initialized acc. to channel LLRs

**Iterations:** for iter = 1,..., iter\_max

■ **CNU:**  $\forall m = 1, \dots, M; \forall n \in H(m)$

$$\beta_{m,n} = \left( \prod_{n' \in H(m) \setminus n} \text{sgn}(\alpha_{m,n'}) \right) \min_{n' \in H(m) \setminus n} (|\alpha_{m,n'}|)$$

all check-nodes are processed

new check-node messages are sent to neighbor var.-nodes

■ **VNU:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\alpha_{m,n} = \gamma_n + \sum_{m' \in H(n) \setminus m} \beta_{m',n}$$

all variable-nodes are processed

new var.-node messages are sent to neighbor check nodes

■ **AP-LLR:**  $\forall n = 1, \dots, N$

$$\tilde{\gamma}_n = \gamma_n + \sum_{m \in H(n)} \beta_{m,n}$$

a posteriori LLRs of variable-nodes are computed

N.B: these values are used to take a hard decision on each coded-bit (given by the sign of the AP-LLR)

Remark: at each iteration the decoder also computes the syndrome of the estimated word (determined by the signs of the AP-LLR values). Decoder stops if a codeword has been found (syndrome = 0), or if the maximum number of iterations (iter\_max) has been reached.

# Min-Sum decoding with flooding scheduling

**Initialization:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\gamma_n = \log(\Pr(x_n = 0 | y_n) / \Pr(x_n = 1 | y_n))$$

$$\alpha_{m,n} = \gamma_n$$

**Iterations:** for iter = 1,..., iter\_max

■ **CNU:**  $\forall m = 1, \dots, M; \forall n \in H(m)$

$$\beta_{m,n} = \left( \prod_{n' \in H(m) \setminus n} \text{sgn}(\alpha_{m,n'}) \right) \min_{n' \in H(m) \setminus n} (|\alpha_{m,n'}|)$$

■ **AP-LLR:**  $\forall n = 1, \dots, N$

$$\tilde{\gamma}_n = \gamma_n + \sum_{m \in H(n)} \beta_{m,n}$$

■ **VNU:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\alpha_{m,n} = \tilde{\gamma}_n - \beta_{m,n}$$

variable-node messages are initialized acc. to channel LLRs

all check-nodes are processed

new check-node messages are sent to neighbor var.-nodes

a posteriori LLRs of variable-node are computed

N.B: these values are used to take a hard decision on each coded-bit (given by the sign of the AP-LLR)

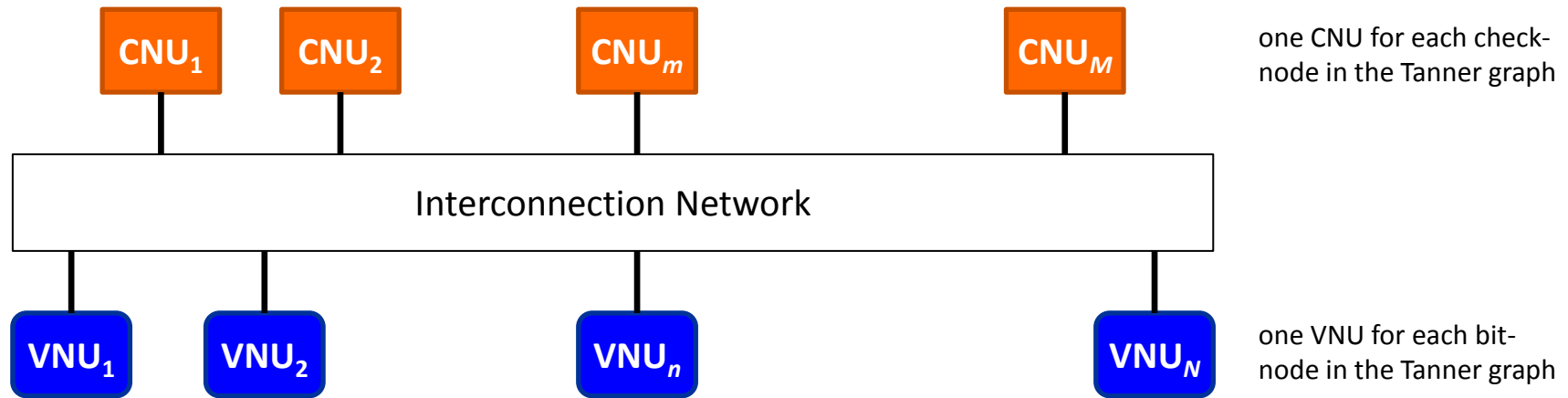
all variable-nodes are processed

new var.-node messages are sent to neighbor check nodes

**AP-LLR and VNU steps can be advantageously switched**  
(they can actually be merged in a single processing unit)

Remark: at each iteration the decoder also computes the syndrome of the estimated word (determined by the signs of the AP-LLR values). Decoder stops if a codeword has been found (syndrome = 0), or if the maximum number of iterations (iter\_max) has been reached.

# MP decoding with flooding scheduling



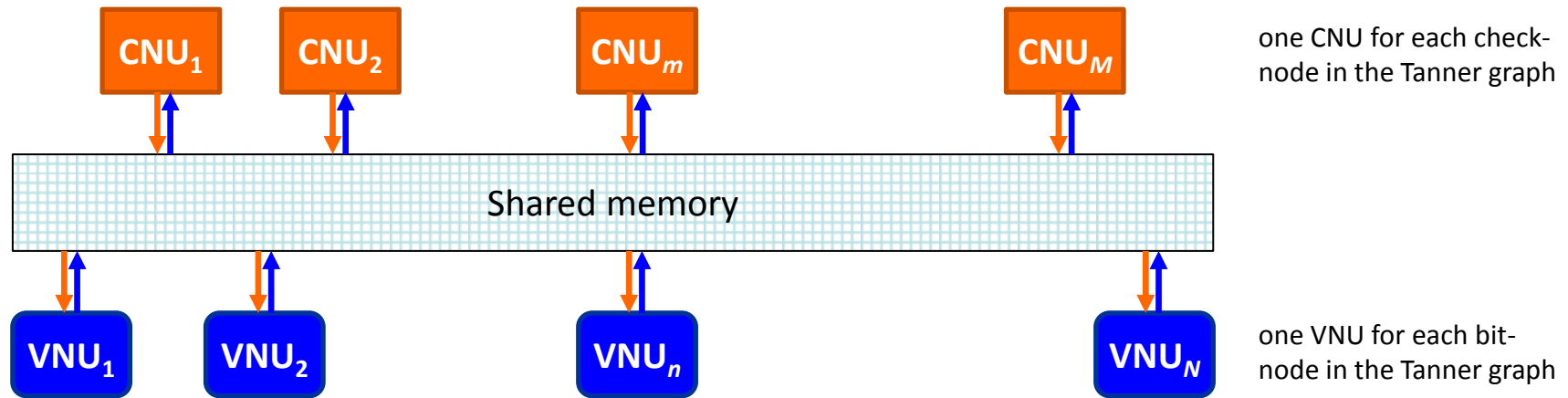
- First half iteration: all CNUs are processed
  - **processed in parallel** if the system allows parallel computation
  - otherwise, **processed sequentially**
- Second half iteration: all VNUs are processed
  - **processed in parallel** if the system allows parallel computation
  - otherwise, **processed sequentially**
- **Main issue: complexity of the interconnection network**

**fully parallel decoder**

**sequential decoder**

but the same  
flooding scheduling!

# MP decoding with flooding scheduling



- First half iteration: all CNUs are processed
  - **processed in parallel** if the system allows parallel computation
  - otherwise, **processed sequentially**
- Second half iteration: all VNUs are processed
  - **processed in parallel** if the system allows parallel computation
  - otherwise, **processed sequentially**
- **Main issue: dealing with memory conflicts**

**fully parallel decoder**

**sequential decoder**

but the same  
flooding scheduling!

# MP decoding with layered scheduling

- The parity check matrix  $H$  is partitioned in  $L$  horizontal layers
  - Each layer contains  $\mu = M/L$  rows
    - Layer  $l$  is determined by  $\mu$  consecutive rows:  $\mathcal{M}_l = (l-1)\mu + 1, \dots, l\mu$
  - Matrix  $H$  is usually designed such that any variable-node is connected at most once to any layer (e.g. Quasi-Cyclic LDPC codes).

$$H = \left( \begin{array}{ccc|cc} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ \hline & & & & \\ & & & & \\ & & & & \\ \hline & & & & \\ & & & & \\ & & & & \\ \hline 0 & 1 & 0 & 1 & 1 \end{array} \right) \begin{array}{l} \mathcal{M}_1 (\mu \text{ rows}) \\ \\ \\ \mathcal{M}_2 (\mu \text{ rows}) \\ \\ \\ \mathcal{M}_3 (\mu \text{ rows}) \end{array}$$

## Decoding scheduling (how it works):

Process check nodes in the first layer, then update all bit-nodes connected to the layer

Move to the next layer: process check nodes in the layer, then update all bit-nodes connected to this layer

Continue with following layers...

then start a new decoding iteration

# Min-Sum decoding with layered scheduling

**Initialization:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\tilde{\gamma}_n = \gamma_n = \log(\Pr(x_n = 0 | y_n) / \Pr(x_n = 1 | y_n))$$

$$\beta_{m,n} = 0$$

**Iterations:** for iter = 1,..., iter\_max

▪ **Loop over horizontal layers:**  $\forall l = 1, \dots, L$

▪ **VNU:**  $\forall m \in \mathcal{M}_l; \forall n \in H(m)$

$$\alpha_{m,n} = \tilde{\gamma}_n - \beta_{m,n}$$

▪ **CNU:**  $\forall m \in \mathcal{M}_l; \forall n \in H(m)$

$$\beta_{m,n} = \left( \prod_{n' \in H(m) \setminus n} \text{sgn}(\alpha_{m,n'}) \right) \min_{n' \in H(m) \setminus n} (|\alpha_{m,n'}|)$$

▪ **AP-LLR:**  $\forall m \in \mathcal{M}_l; \forall n \in H(m)$

$$\tilde{\gamma}_n = \alpha_{m,n} + \beta_{m,n}$$

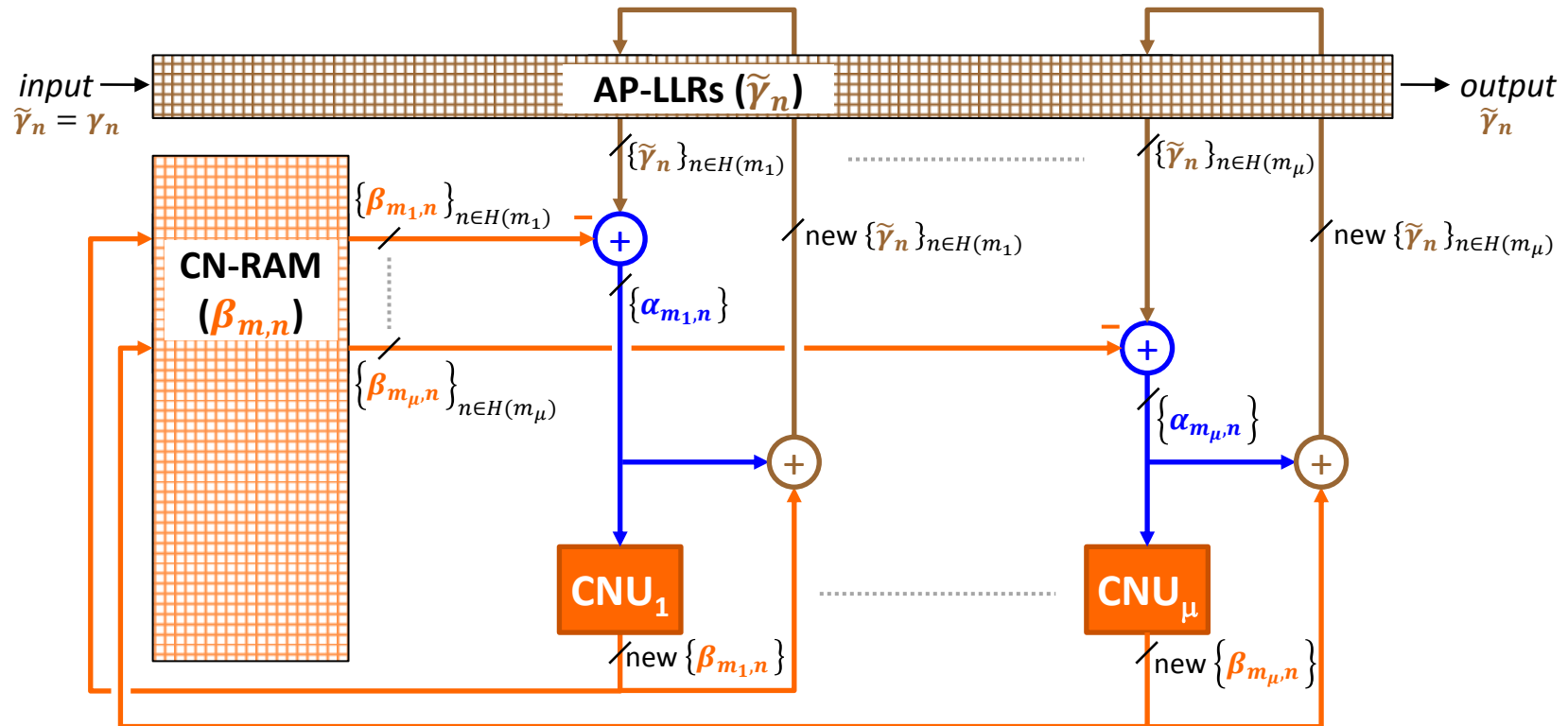
AP-LLR values are initialized from channel LLR values  
check-node messages are initialized to 0

for each check-node in the current layer, incoming  
variable-to-check ( $\alpha_{m,n}$ ) messages are computed by  
subtracting the  $\beta_{m,n}$  message from the AP-LLR value

check-nodes in the current layer are processed  
new check-to-variable messages  $\beta_{m,n}$  are computed

a posteriori LLR values are updated according to the new  
check-to-variable ( $\beta_{m,n}$ ) messages

# MP decoding with layered scheduling



- Number of CNU = number of for check-node per layer
  - processed in parallel** if the system allows parallel computation
  - otherwise, **processed sequentially**
- No memory conflicts, provided that any variable-node is connected to at most one check-node in each layer

**partially parallel decoder**

**sequential decoder**

but the same  
layered scheduling!

# MP decoding with layered scheduling

## Interest of layered scheduling:

- Simple architecture, with a flexible degree of parallelism
- Converges (twice) faster than flooding scheduling
  - better performance, if the maximum number of decoding iterations is small

## Faster convergence:

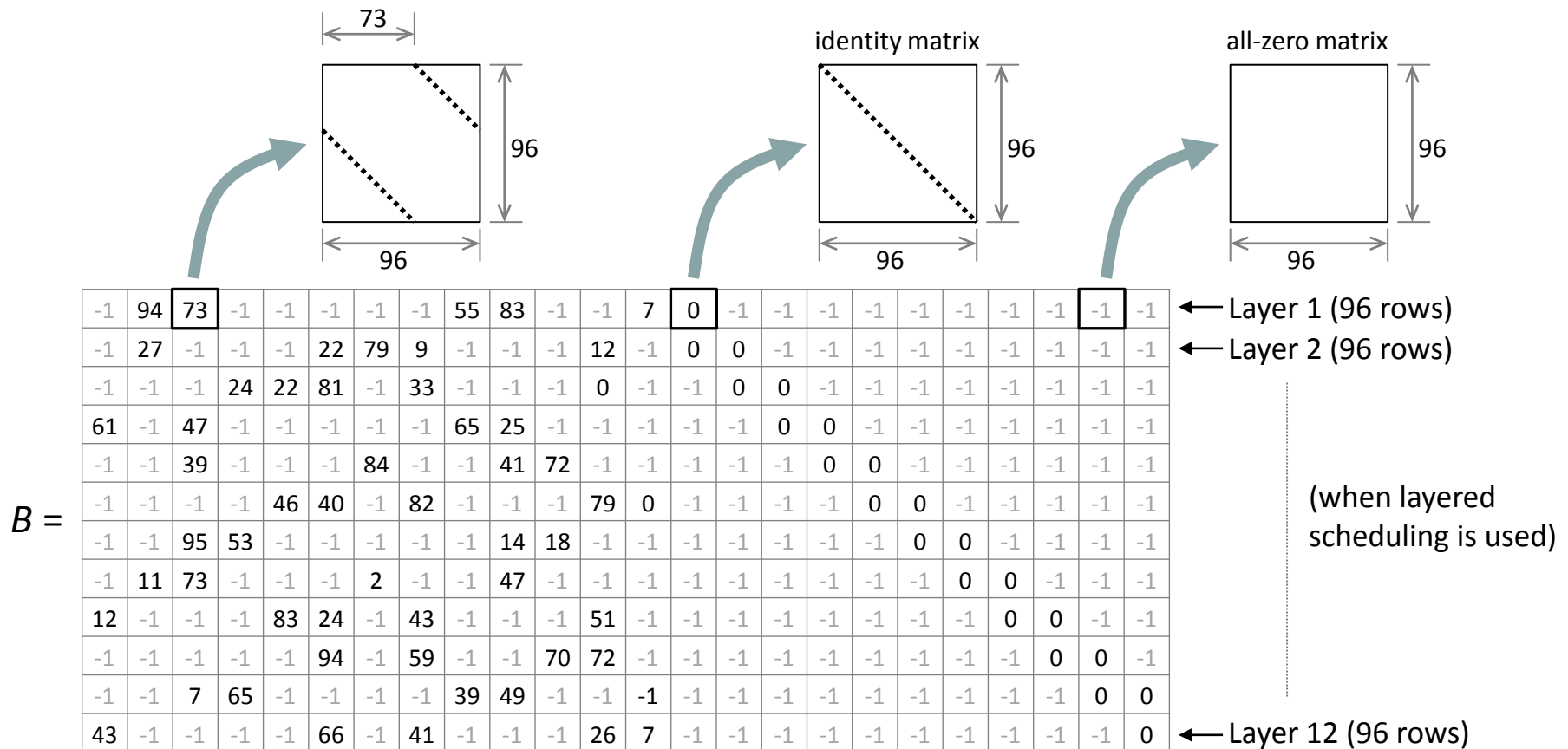
- AP-LLR values of variable-nodes connected to a layer are updated after the check-nodes in that layer have been processed
- When the next layer is processed, the incoming check-node messages ( $\alpha_{m,n}$ ) are derived from the AP-LLR values of the corresponding variable-nodes .
- Hence, incoming messages incorporate the contribution of the check-nodes from the previous layers (even those processed in the same iteration)
- This yield a faster propagation (\*) of messages through the decoding graph, thus speeding up the decoder convergence

(\*) It can be proved that messages propagate twice faster through cycle-free graphs  
In practice, the propagation speed up factor is less than twice...

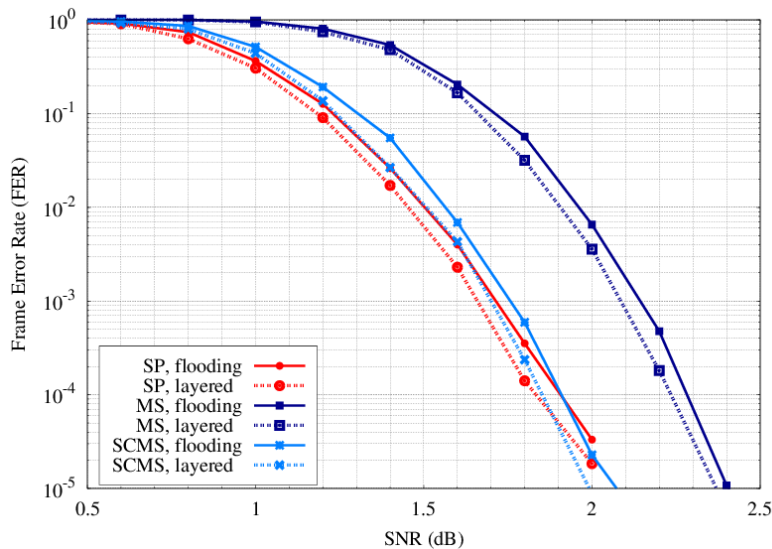


# Example-1: WiMAX LDPC code with rate 1/2

- Parity-check matrix  $H$  of size 1152 x 2304
- **Quasi-Cyclic** code:  $H$  is obtained by expanding a *base matrix*  $B$  of size 12 x 24
  - each entry of  $B$  is expanded to a square matrix of size 96 x 96
  - a non-negative entry  $b$  is replaced by a cyclic permutation matrix = identity right-shifted by  $b$  columns



# Example-1: WiMAX LDPC code with rate 1/2



AWGN, QPSK

Maximum number of decoding iterations = 50

Color code:

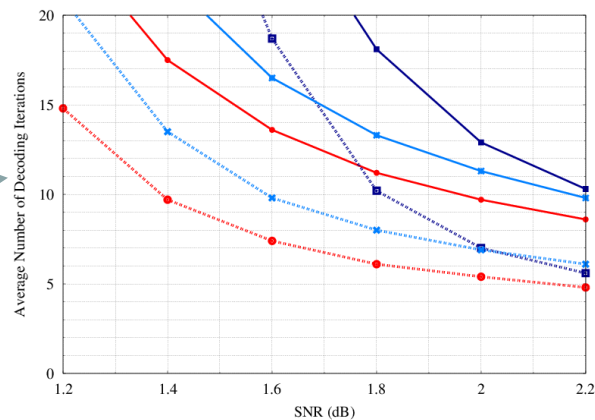
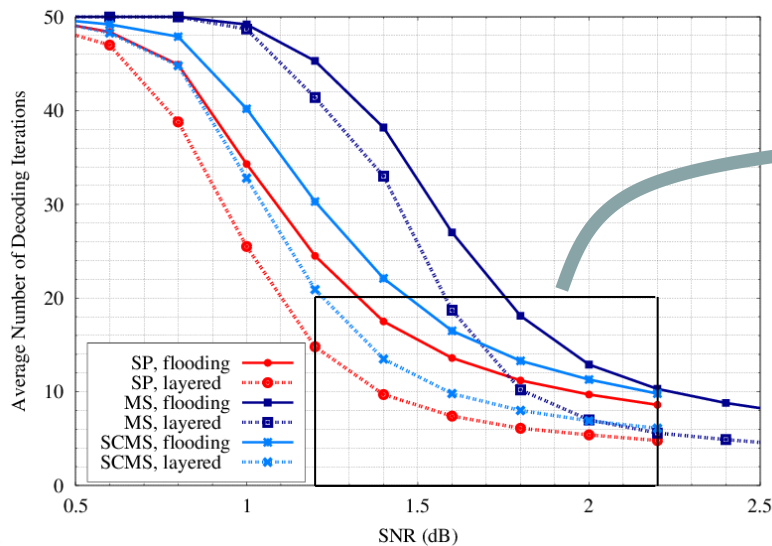
Sum-Product (SP)

Min-Sum (MS)

Self-Corrected Min-Sum (SCMS)

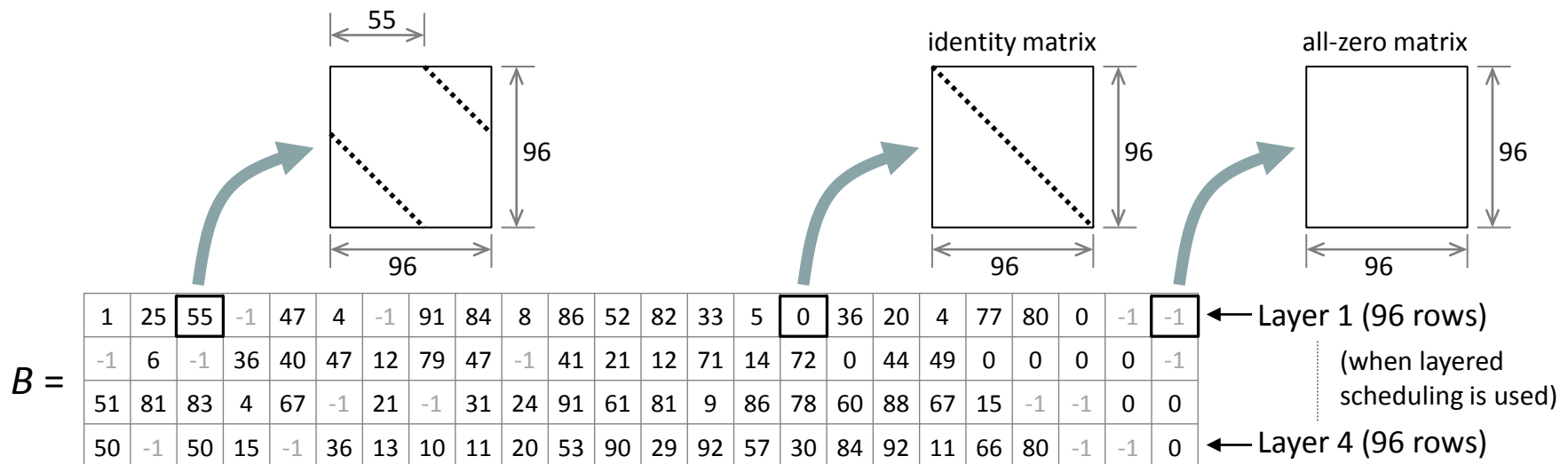
Solid curves : flooding scheduling

Dashed curves: layered scheduling

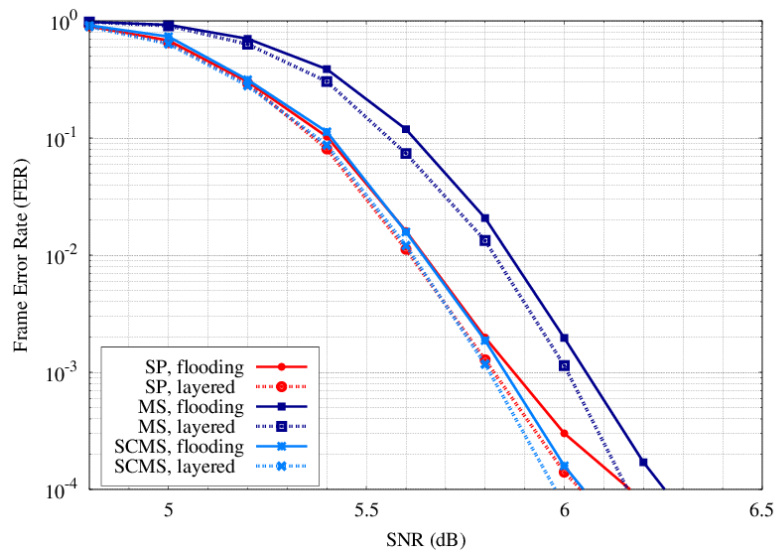


# Example-2: WiMAX LDPC code with rate 5/6

- Parity-check matrix  $H$  of size 384 x 2304
- **Quasi-Cyclic** code:  $H$  is obtained by expanding a *base matrix*  $B$  of size 4 x 24
  - each entry of  $B$  is expanded to a square matrix of size 96 x 96
  - a non-negative entry  $b$  is replaced by a cyclic permutation matrix = identity right-shifted by  $b$  columns



# Example-2: WiMAX LDPC code with rate 5/6



AWGN, QPSK

Maximum number of decoding iterations = 50

Color code:

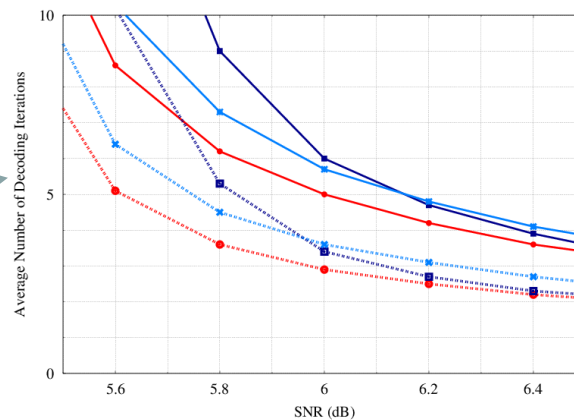
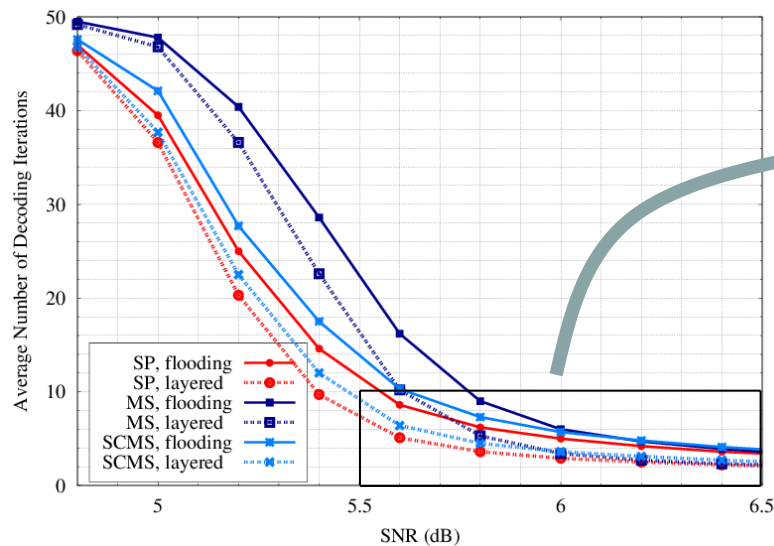
Sum-Product (SP)

Min-Sum (MS)

Self-Corrected Min-Sum (SCMS)

Solid curves : flooding scheduling

Dashed curves: layered scheduling



# Outline

## PART 1:

LDPC codes and Message Passing (MP) decoders – brief review

## PART 2:

Fixed-point implementation and “noisy” arithmetic units

# Impact of the scheduling and of the quantization of the decoder performance

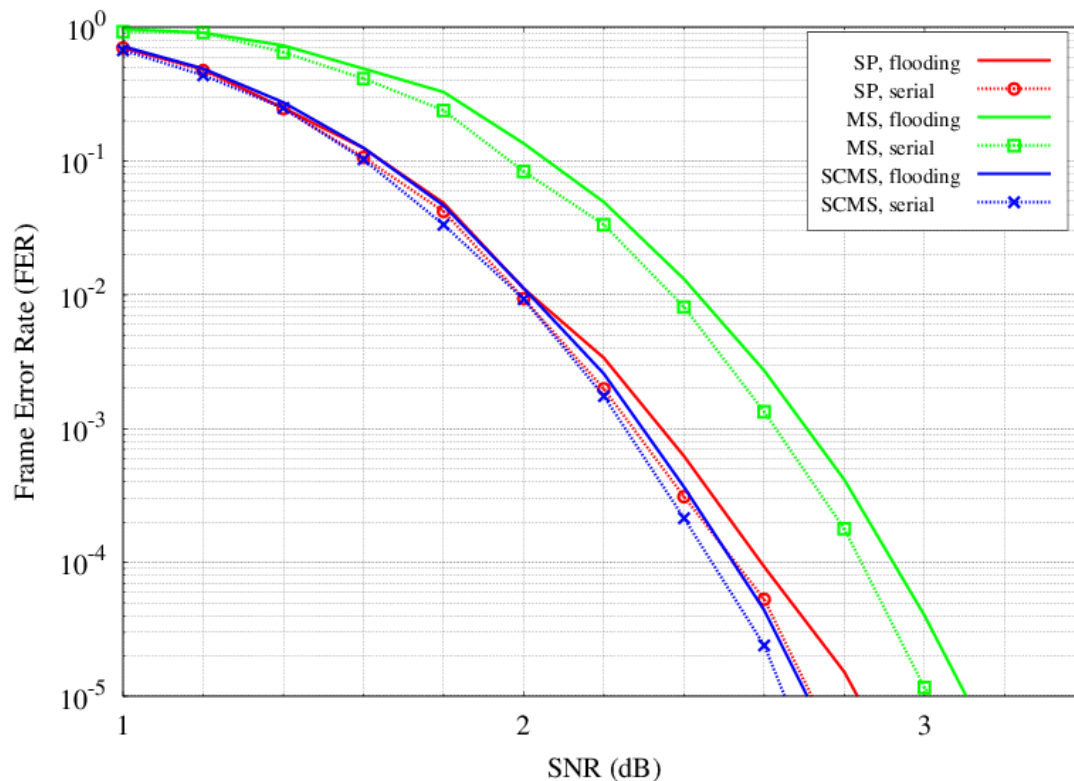
## 1. Noiseless devices

- Fixed-point decoders
  - AP-LLR ( $\tilde{\gamma}_n$ ) quantized on **6 bits**
  - IN-LLR ( $\gamma_n$ ) and exchanged messages ( $\alpha_{m,n}$ ,  $\beta_{m,n}$ ) quantized on **4 or 5 bits**
- Fixed-point arithmetic
  - **6-bit adders**
    - exchanged messages and IN/AP-LLR values are saturated if out of corresponding range
  - **noiseless** (“exact”) **arithmetic**

- all simulations: Mackay’s regular (3,6)-LDPC code, [K = 504, N = 1008]

# Floating-point simulations

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code,  $[K = 504, N = 1008]$
- Floating-point decoders



## Decoder

Sum-Product

Min-Sum

Self-Corrected Min-Sum

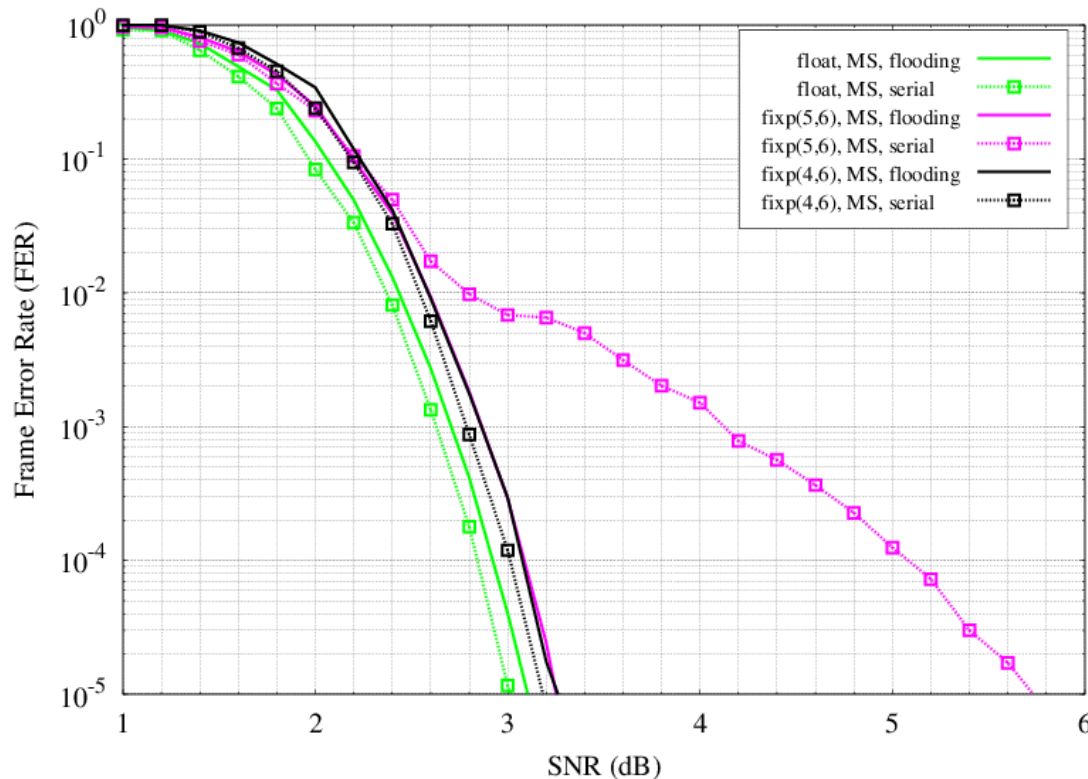
## Scheduling

Solid curves: **flooding**

Dashed curves: **serial**

# Fixed-point simulations / Min-Sum decoder

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point quantization: ( $x$ ,  $y$ )
  - $x$  bits for IN-LLR & exchanged messages,  $y$  bits for AP-LLR



## Color code

floating-point

fixed-point(5, 6)

fixed-point(4, 6)

## Scheduling

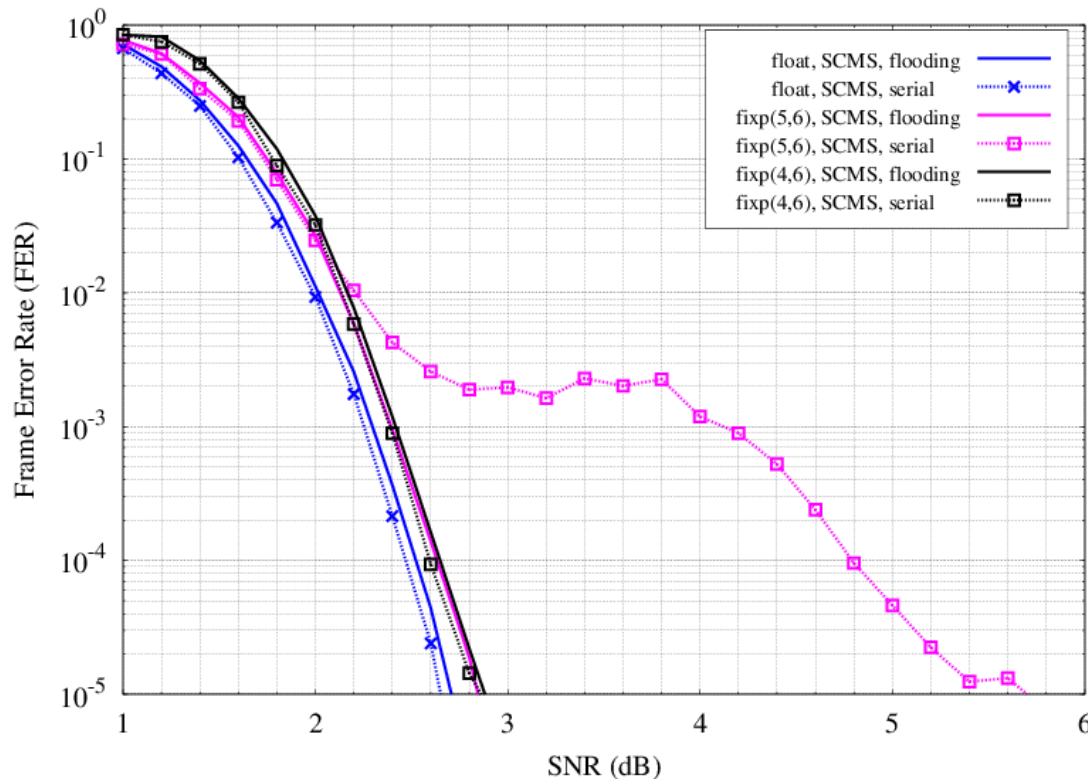
Solid curves: **flooding**

Dashed curves: **serial**



# Fixed-point simulations / SC-Min-Sum decoder

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point quantization: ( $x$ ,  $y$ )
  - $x$  bits for IN-LLR & exchanged messages,  $y$  bits for AP-LLR



## Color code

floating-point

fixed-point(5, 6)

fixed-point(4, 6)

## Scheduling

Solid curves: **flooding**

Dashed curves: **serial**

# Impact of the scheduling and of the quantization of the decoder performance

## 2. Noisy devices

- Fixed-point decoders
  - AP-LLR ( $\tilde{\gamma}_n$ ) quantized on **6 bits**
  - IN-LLR ( $\gamma_n$ ) and exchanged messages ( $\alpha_{m,n}$ ,  $\beta_{m,n}$ ) quantized on **4 bits**
- Fixed-point arithmetic
  - **6-bit adders**
    - exchanged messages and IN/AP-LLR values are saturated if out of corresponding range
  - **noisy** (probabilistic) **arithmetic**

- all simulations: Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]

# Min-Sum decoder on **faulty** devices

**Initialization:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\gamma_n = \log(\Pr(x_n = 0 | y_n) / \Pr(x_n = 1 | y_n))$$

$$\alpha_{m,n} = \gamma_n$$

## Iterations

- **CNU:**  $\forall m = 1, \dots, M; \forall n \in H(m)$

$$\beta_{m,n} = \left( \prod_{n' \in H(m) \setminus n} \text{sgn}(\alpha_{m,n'}) \right) \min_{n' \in H(m) \setminus n} (|\alpha_{m,n'}|)$$

- **AP-LLR:**  $\forall n = 1, \dots, N$

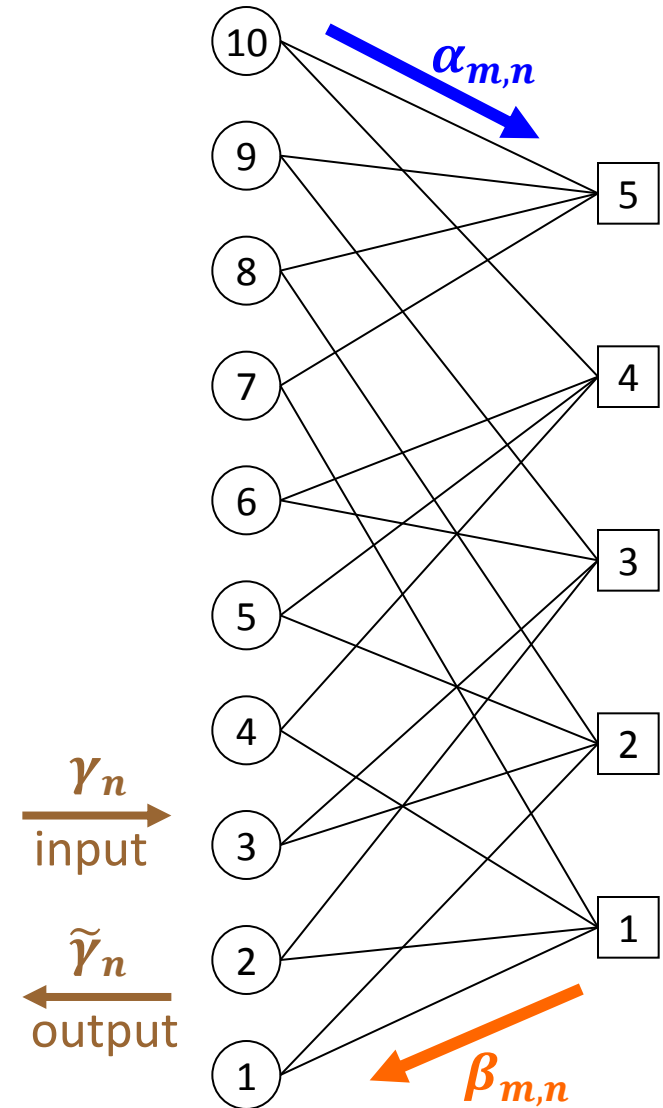
$$\tilde{\gamma}_n = \gamma_n + \sum_{m \in H(n)} \beta_{m,n}$$

- **VNU:**  $\forall n = 1, \dots, N; \forall m \in H(n)$

$$\alpha_{m,n} = \tilde{\gamma}_n - \beta_{m,n}$$

comparators

adders



# Error models for faulty components

## ■ Probabilistic adder ( $Q=6$ bits)

- Two parameters: the **depth**  $D$  and the **error probability**  $P_a$
- $P_a$  is the probability that an error occurs on at least one of the  $D$  LSBs

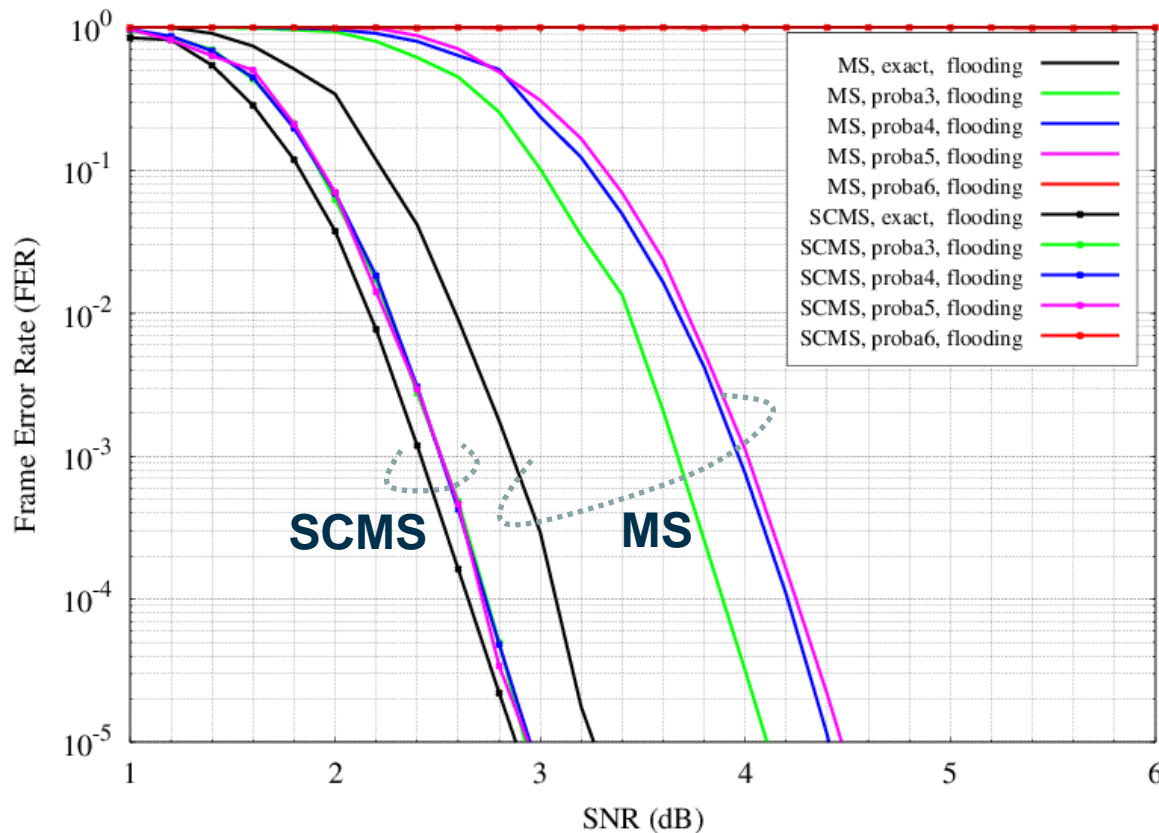
	$Q$		$D$		2	1	
correct output	1	0	1	1	0	1	two's complement
error pattern			0	1	1	0	rand integer in $[1, 2^D-1]$
erroneous output	1	0	1	0	1	1	

## ■ Probabilistic comparator

- $P_c$  is the probability that the output is in error

# Flooding implementation / SCMS vs. MS

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point decoders: 4 / 6 bits (IN-LLR & exchanged messages / AP-LLR)



Comp.err. prob:  $P_c = 0.01$   
Adder err. prprob:  $P_a = 0.01$

Color code:

**Noiseless**

**Depth = 3**

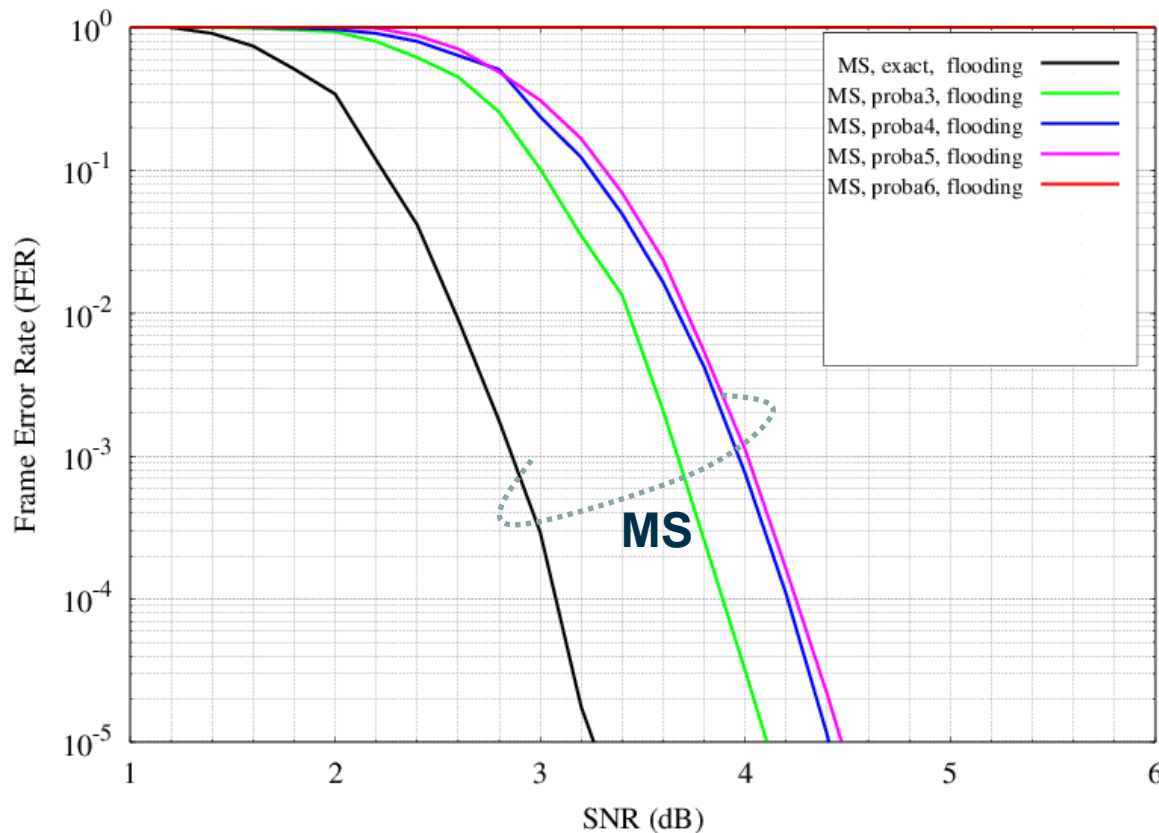
**Depth = 4**

**Depth = 5**

**Depth = 6**

# MS decoder / flooding implementation

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point decoders: 4 / 6 bits (IN-LLR & exchanged messages / AP-LLR)



Comp.err. prob:  $P_c = 0.01$

Adder err. prprob:  $P_a = 0.01$

Color code:

Noiseless

Depth = 3

Depth = 4

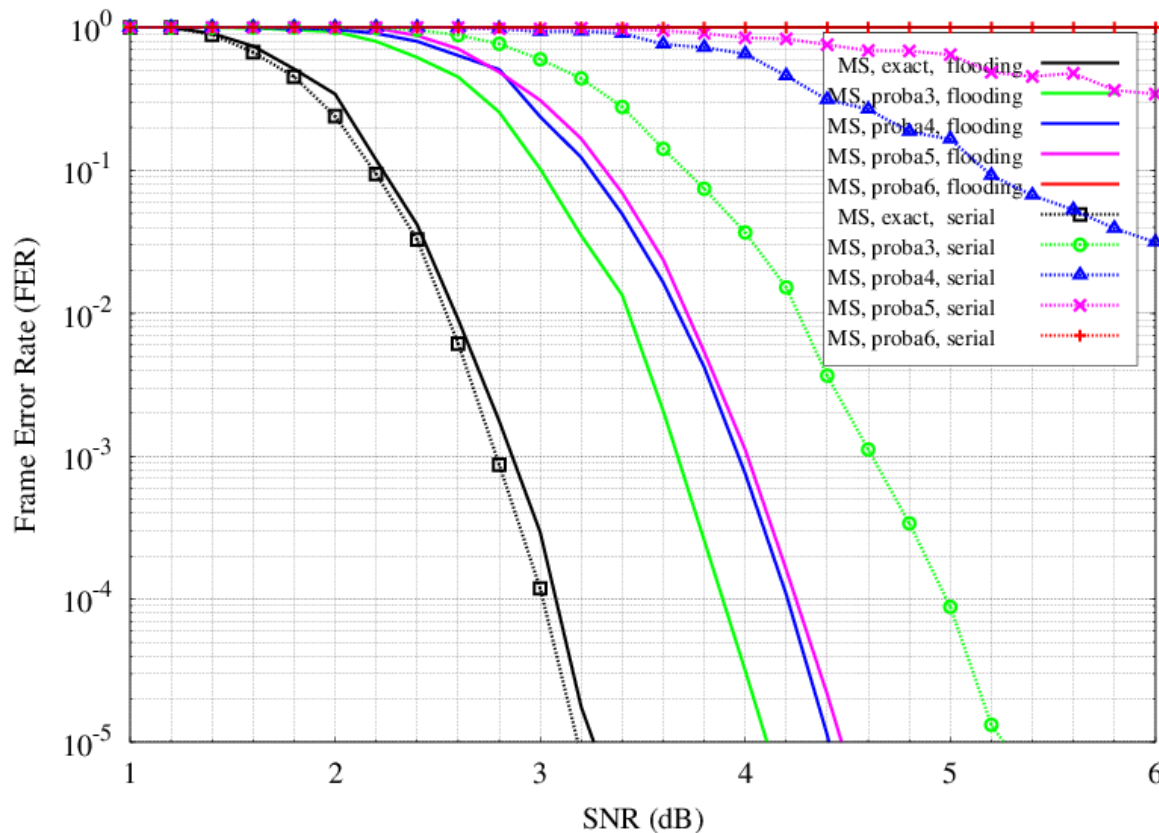
Depth = 5

Depth = 6

Solid curves: flooding

# MS decoder / flooding vs. serial implementation

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point decoders: 4 / 6 bits (IN-LLR & exchanged messages / AP-LLR)



Comp.err. prob:  $P_c = 0.01$   
Adder err. prrob:  $P_a = 0.01$

Color code:

Noiseless

Depth = 3

Depth = 4

Depth = 5

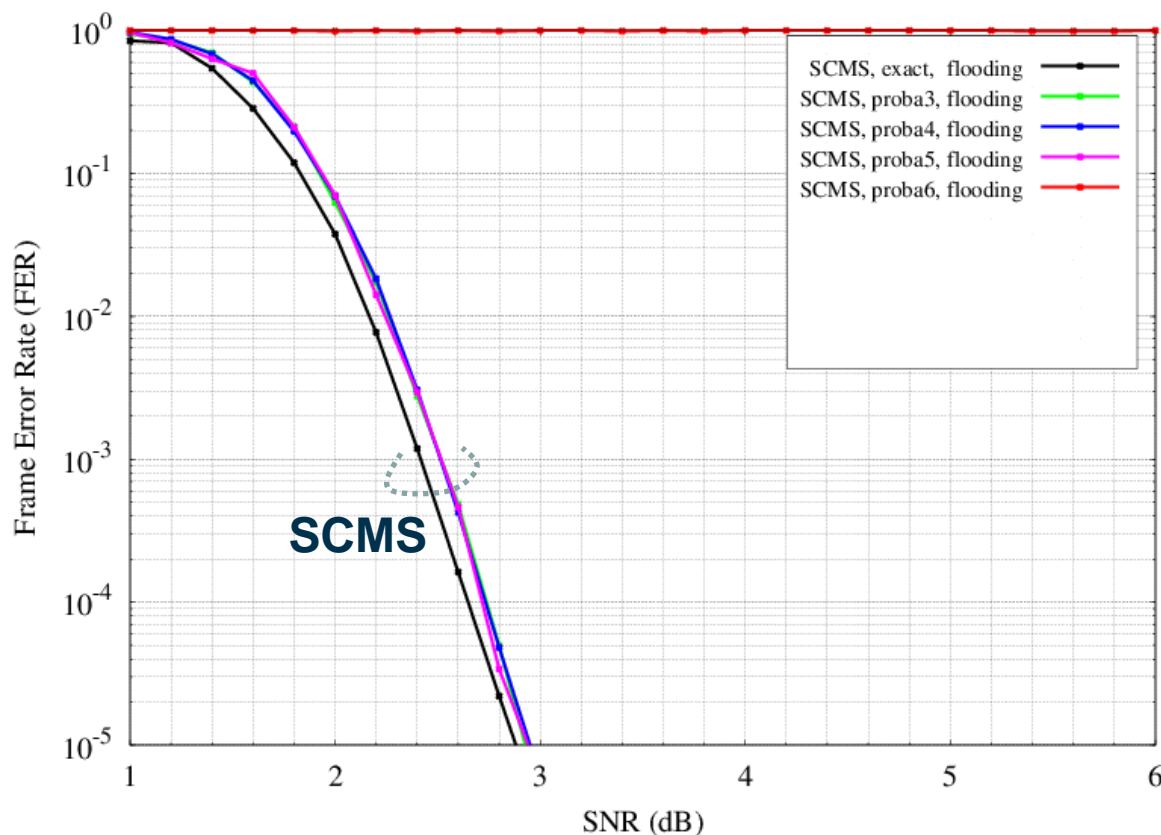
Depth = 6

Solid curves: **flooding**

Dashed curves: **serial**

# SCMS decoder / flooding implementation

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point decoders: 4 / 6 bits (IN-LLR & exchanged messages / AP-LLR)



Comp.err. prob:  $P_c = 0.01$   
Adder err. prrob:  $P_a = 0.01$

Color code:

Noiseless

Depth = 3

Depth = 4

Depth = 5

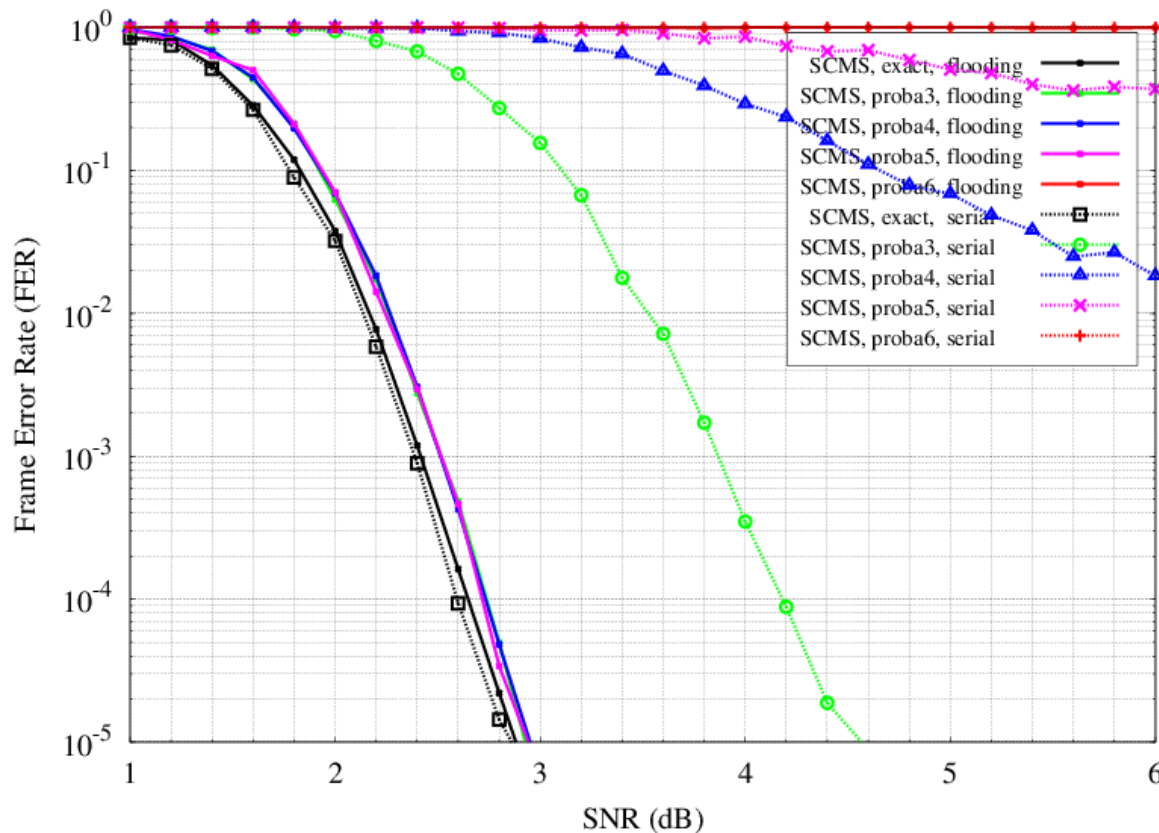
Depth = 6

Solid curves: **flooding**



# SCMS decoder / flooding vs. serial implementation

- AWGN, QPSK
- Mackay's regular (3,6)-LDPC code, [K = 504, N = 1008]
- Fixed-point decoders: 4 / 6 bits (IN-LLR & exchanged messages / AP-LLR)



Comp.err. prob:  $P_c = 0.01$   
Adder err. prrob:  $P_a = 0.01$

Color code:

Noiseless

Depth = 3

Depth = 4

Depth = 5

Depth = 6

Solid curves: **flooding**

Dashed curves: **serial**



# leti

LABORATOIRE D'ÉLECTRONIQUE  
ET DE TECHNOLOGIES  
DE L'INFORMATION

CEA-Leti  
MINATEC Campus, 17 rue des Martyrs  
38054 GRENOBLE Cedex 9  
Tel. +33 4 38 78 36 25

[www.leti.fr](http://www.leti.fr)



## Merci de votre attention



énergie atomique • énergies alternatives

