

High-Throughput Irregular LDPC Decoder Design with Parallel Layered Decoding Architecture



Kai Zhang, Xinming Huang
Department of Electrical and Computer Engineering
Worcester Polytechnic Institute
Email: kzhang@wpi.edu

Introduction

This poster presents a novel decoder architecture for the irregular Quasi-Cyclic (QC) Low-Density Parity-Check (LDPC) codes that can achieve over 2Gbps decoding throughput in a small chip area. Two new techniques are applied to improve overall throughput: parallel layered decoding algorithm (PLDA) and critical path splitting technique. Compared with conventional layered decoding algorithm (LDA), PPLDA assigns all of the message-passing destinations between different layers and allows parallel message updating and passing of all of the layers simultaneously. Moreover, because of the fixed message passing paths, the decoder contains no large multiplexer or demultiplexer based crossbars, thus can avoid the interconnection congestion problem. Critical path splitting technique is based on the artificial adjustments of starting points of different layers. Idle time intervals between different layers alleviate the delay of the decoder's critical path by splitting the long path into several register-coupled small sections. Min-sum algorithm and loosely coupled algorithm are both adopted to reduce interconnection complexity and chip area. To verify the proposed architecture, a rate-1/2 2304-bit irregular LDPC decoder is proposed using ASIC design in TSMC 90nm 1P9M CMOS process. The decoder achieves a maximum frequency of 950MHz, which enables almost 2.2Gbps decoding throughput at 10 iterations, while occupying a relative small area of 2.8mm² compared with previous architecture.

Code Structure

Here we use a rate-1/2 2304-bit irregular QC-LDPC code from WiMax standard, whose parity check matrix is shown below:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		94	73							22	79	9		55	83					7	0			
2		27																						
3				24	22	81				33														
4	61		47																					
5			39																					
6					46	40																		
7																								
8					95	53																		
9					11	73																		
10	12						81	24																
11																								
12	43																							

Parity-check matrix for rate-1/2 LDPC code in WiMax

Decoding Algorithm

- We first make some definitions as follows: Let c_n denote the n -th bit of a codeword and y_n denote the corresponding received value from the channel. Let r_{mn} (q_{mn}) be the CTV (VTC) message from check node m to variable node n at the k -th iteration. Let $N(m)$ denote the set of variables that participate in check m and $M(n)$ denote the set of checks that participate in variable n . The set $N(m)$ without variable n is denoted as $N(m)\setminus n$ and the set $M(n)$ without check m is denoted as $M(n)\setminus m$. Detailed steps of Min-Sum based decoding algorithm are described below:

Initialization:

Under the assumption of equal priori probability, compute the channel probability p_n (intrinsic information) of the variable node n , by:

$$p_n = \log \frac{P(y_n | c_n = 0)}{P(y_n | c_n = 1)}$$

The CTV message r_{mn} is set to be zero.

Iterative Decoding:

At the k -th iteration, for the variable node n , calculate VTC message q_{mn} by

$$q_{mn} = p_n + \sum_{m' \in M(n)\setminus m} r_{m'n}$$

Meanwhile, the decoder can make a hard decision by calculating the APP (a-posteriori probability) by

$$\Lambda_n = p_n + \sum_{m \in M(n)} r_{mn}$$

Decide the n -th bit of the decoded codeword $x_n=0$ if $\Lambda_n > 0$ and $x_n=1$ otherwise. The decoding process terminates when the entire codeword $x=[x_1, x_2, \dots, x_N]$ satisfy all of the M parity check equations: $Hx=0$, or the preset maximum number of iteration is consumed.

If the decoding process does not stop, then, calculate the CTV message r_{mn} for the check node m , by

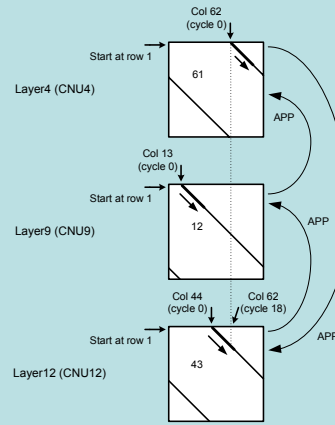
$$r_{mn} = \prod_{n' \in N(m)\setminus n} \text{sign}(q_{mn'}) \times \left(\min_{n' \in N(m)\setminus n} |q_{mn'}| \times \alpha \right)$$

Here, a normalized factor α is introduced to compensate for the performance loss exiting in the min-sum algorithm without compensation compared to BP algorithm. In this poster, α is set to be 0.75.

Parallel Layered Decoding Architecture

In conventional layered decoding algorithm (LDA), data processing of lower layer requires values of messages coming from upper layer, making the LDA totally serial. The decoder has to process data from the first layer until the last layer, which causes serious throughput limitation due to large decoding latency.

Parallel layered decoding architecture (PLDA) is able to maintain high throughput using much less hardware resources. It is similar to the classic partly parallel decoding architecture, but with message passing between different layers and therefore can also accelerate the convergence speed. PLDA allows parallel processing among different layers while the updating of messages is in sequential. CTV messages r_{mn} are calculated during the horizontal processing and then transmitted to overlapped positions in other layers. Specific network precisely regulates the transmission directions of the messages to guarantee that at another layer, messages from the previous layer must arrive before they are used.



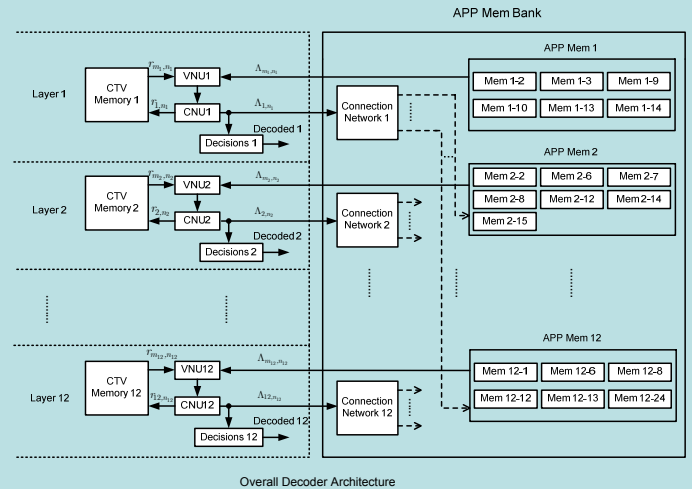
Let us pick up the first column of the H base matrix for example. Non-zero entries are at the 4th, 9th and 12th layer whose permutation numbers s are 61, 12 and 43, respectively. The newly updated CTV message r_{mn} from the 4th layer will be used by the 9th and 12th layer 49 and 18 clock cycles later. As a result, this message from the 4th layer only needs to be transmitted to the layer that demands it most urgently, that is the 12th layer. When the processing unit of the 12th layer moves to column 64, the CTV message from the 4th layer is used to update the VTC message and then again get a new CTV message which will be passed to the 9th layer. Message passing path for the first column should be a cyclic loop which assigns the direction as 4th-12th-9th-4th layer

Message Passing Directions among different layers

Therefore, we can conclude the message passing rule between different layers at the vertical block j as follows: summation message at layer i' will be transmitted to the layer that has the maximum permutation value cyclically smaller than the permutation value at layer i' .

Overall Decoder Architecture

The updated APP messages obtained by one layer are stored in the APP Mem Bank and used by another layer. Hence, convergence speed can be doubled and decoding latency per iteration can also be reduced significantly. Besides, a VNU and a CNU of the same layer can be merged and split into several pipelined stages by inserting registers. As a result, overall decoding frequency is expected to be enhanced. All of these methods result in great improvement in decoding throughput.



Overall Decoder Architecture

Result Highlights

- Decoding throughput: 2.2 Gbps
- Operating frequency: 950MHz
- Core area: 2.8mm²