ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 397 – 401 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

LDPC ARCHITECTURE USING MEMORY BYPASSING SCHEME

Venu.I, Parameshwar. G

venuippalapalli@gmail.com
paramesh.gujjula@gmail.com
L I ETechnology,JNTUH,

Abstract:

Error correcting codes[2] insert redundancy into the transmitted data stream so that the receiver can detect and possibly correct errors that occur during transmission. In VLSI design, area of the chip, speed and power consumption are. Its performance is close to Shannon's capacity[1] limits. Inherent parallelism of the message passing decoding algorithm for LDPC codes makes them very suitable for hardware implementation. Inherent parallelism of the message passing decoding algorithm for LDPC codes makes them very suitable for hardware implementation. LDPC used in optical fiber communications[6], satellite, storage wireless, wired line. LDPC codes advantages are strong error control codes that operate efficiently at extremely low signal to noise ratios.

Sum-Product decoding algorithm has inherent parallelization. Decoding error is detectable event which results in a more reliable system. Min-Sum algorithm for low complexity design. The message passing algorithms used for its decoding. In the hardware architecture for implementation of Decoder for LDPC[6], all the internal modules such as check node and variable node memory organization, check node and variable node functional units, and switching network which can parallely establish the connections between memory banks and functional units. In order to reduce the energy consumption of the LDPC decoder[6], modified memory-bypassing scheme has been proposed. The amount of achievable memory bypassing depends on the decoding order of the layers

Keywords: LDPC, Shannon's capacity, Inherent parallesim, memory bypassing Scheme.

I. INTRODUCTION

LDPC codes are linear block codes . The set of valid code words C is defined as $H \bullet x^T = 0 \quad \forall x \in C$. The information $i = [a_1, a_2, ..., a_K]$ mapped into a codeword $c = [c_1, c_2, ..., c_K, c_{K+1}, ..., c_N]$, the mapping can be a linear mapping. The canonical form of a linear transformation is $c = i * G$ Where G is a $K \times N$ matrix and all the code words {c} are distinct when the rank of G is K are same. $i=[a_1,a_2,...a_K]$ is encoded uniquely as, $c=aG=[a_1,a_2,....a_K]G, \quad a_i \in GF(2)$ The dual space of a linear code C is denoted by C^T, which is a vector space of dimension (N-K). **Parity Check Matrix H.**

$$H = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N-K-1} \end{bmatrix} = \begin{bmatrix} h_{0,0} & h_{0,1} & \cdots & h_{0,N-1} \\ h_{1,0} & h_{1,1} & \cdots & h_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N-K-1,0} & h_{N-K-1,1} & \cdots & h_{N-K-1,N-1} \end{bmatrix}$$

i) The parity check theorem:

The parity check matrix[2] determine the minimum distance of the code. If c is the code word and G is the generator matrix then generator matrix can be obtained and the column rank of the matrix is equal to the row rank.

$$G = [P|I_K] = \left[\begin{array}{cccc|cccccc} p_{0,0} & p_{0,1} & \cdots & p_{0,N-K-1} & 1 & 0 & 0 & \cdots & 0 \\ p_{1,0} & p_{1,1} & \cdots & p_{1,N-K-1} & 0 & 1 & 0 & \cdots & 0 \\ p_{2,0} & p_{2,1} & \cdots & p_{2,N-K-1} & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{K-1,0} & p_{K-1,1} & \cdots & p_{K-1,N-K-1} & 0 & 0 & 0 & \cdots & 1 \end{array}\right]$$

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 397 – 401 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

After decoding, the last *K* symbols are removed from the selected codeword and passed along to the data sink. Transform the generator matrix into systematic form using Gaussian eliminations.

ii)LDPC Codes

Parity check matrix $H_{(N-K)\times N}$ consists of only 0's and 1's and sparse. If degrees per row or column are not constant, then the code is *irregular*. Otherwise it is regular. Irregular codes have better performance than regular codes. Code rate *R* is equal to *K=N*.

iii) Tanner Graph

Nodes separated into two classes, and edges are connecting two nodes that residing in the same class. They are VN or Bit Nodes" and CN $f_j, j = 1,.....N-K$ is connected to bit node $x_i, i = 1,.....N$; h_{ij} in *H* is a *one*". Each Bit node is connected to two check nodes and each check node has a degree of *four*.

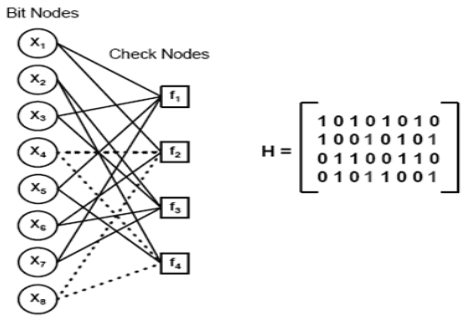


Fig.1Tannergraph representation

The code can also be described by bipartite graph, known as Tanner graph[3]. Variable nodes and Check nodes are connected to each other through a set of edges. An edge links the check node to the variable node if the element of the parity check matrix is non-null.

iv) LDPC Decoder

Information is sent along the edges of the Tanner graph[3]. Local computations are done in each node of the graph. To facilitate the subsequent iterative processing, one tries to keep the graph as low density as possible.

II. Hardware Implementation.

Architecture-aware codes are structured codes and based on specific pattern parity check matrix is build. It supports an efficient partial-parallel hardware VLSI implementation.

i. Designing the Parity Check Matrix

The Parity check matrix plays a major role in the performance the LDPC encoding/decoding. It can be random or structured depending on encoder/decoder. Random matrices are suitable for the decoders running on general purpose processors. It requires less memory. The generator matrix *G* derived by solving $GH^T = 0$. Perform the Gaussian elimination on the resulting matrix *G then* G = [I|P].

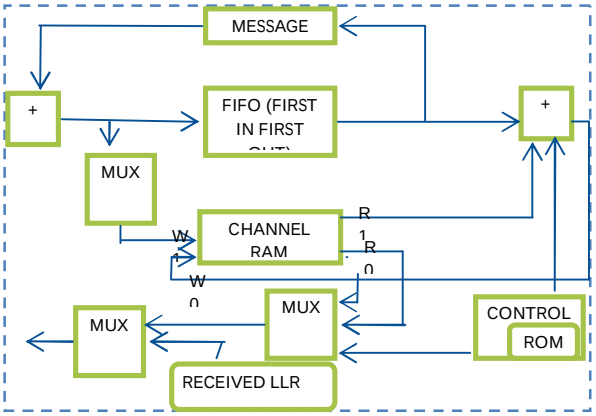


Fig. 2 Block Diagram of LDPC Architecture

H is composed of null sub-matrix or identity sub-matrix with different cyclic shifts. In the decoder, multiple SISO units work in parallel to calculate multiple check node process for a layer. Initially the Channel RAM used to store the input LLR of the received data. The shifter is used to perform the cyclic shift of the output messages so correct message sent to SISO for calculation. Sub array used to perform the subtraction and the results will be sent to the SISO unit and stored in the memory. For every CN, generate the signs for the outgoing messages, an index, magnitudes. The data generated by the SISO will be stored in the Message RAM and the Add-array performs the addition and the intermediate results stored in the FIFO. The add array results written back to the Channel RAM. Pipeline is used to increase the throughput.

ii. MEMORY BYPASSING SCHEME

De-couple the read and write order of the Channel RAM and the memory storing the intermediate messages and idling cycles are minimized. The messages read out from the Channel RAM will be stored in the intermediate data RAM

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 397 – 401 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

after the subtraction and read out from the intermediate data RAM for the check node update, the read order of the Channel RAM is the write order of the intermediate data RAM and the read order of the intermediate data RAM is the write order of the Channel RAM. The Limitations are

1.Consider overlapping between two consecutive layers for the memory bypassing. The no. of the latency cycles is 2. Sub-array, the SISO and the Add-array to finish the computation after the last incoming variable node is read in.

2.Consider overlapping between three consecutive layers, the memory-bypassing operation can be divided into two cases:

1. Memory-bypassing between layers q+2 and q+1 and q+1 and q. The non-null entry that the layer q+2 are is common. It is limited by the latency cycles.

2.When the latency cycles < the no. of the non-null entry that the current layer q+2 are in common with the layers but not in common q+1 with layer. In order to overcome this, increase latency clock cycles.

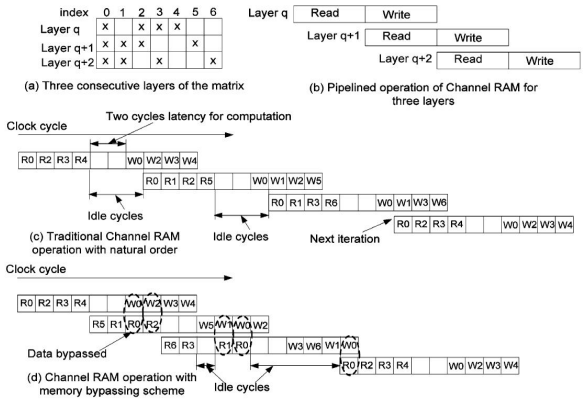


Fig 3 The Memory By Passing Operation For The Channel Ram In The Layered Ldpc Decoder

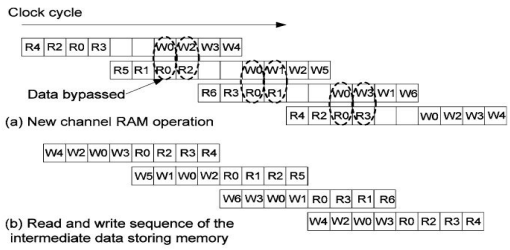
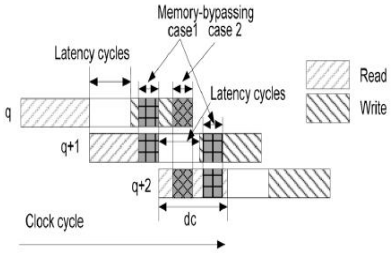


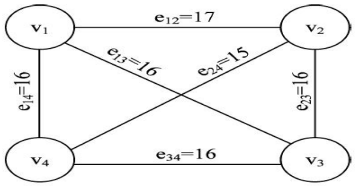
Fig 4 Memory Operations With Different Read And Write Order For The Matrix:-

The amount of the overlap directly depends on the order of the layer decoding. Determine the optimal decoding order and the non-null entries in every layer.



iii.Quick Searching Algorithm:-

Brute-force method to list out all the permutation orders of the layers and then compute the number of overlapped columns of each permutation. Disadvantage of bruteforce is that it is applicable for small no. of values only.



.

Fig.5 Searching Algorithm

Searching algorithm reduce the computational time and find the optimal decoding order for max. overlapping layers. In a graph problem. Let G=(V,E) be a complete graph where V and E are the set of the nodes and edges. Start from any of the node in G, visiting all the other nodes exactly once and returning back to the starting node. Brute force method is used to find the optimal solution for less no. of nodes. Heuristic algorithm used for large no. of nodes to find the best solution. Complexity reduced by using NP-hard and heuristic algorithms. Simulated annealing is used to implement the heuristic algorithm for two and three-layer-overlapping problems. For each move, a pair of layers is randomly selected and their orders in the decoding sequence are swapped.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 397 – 401 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

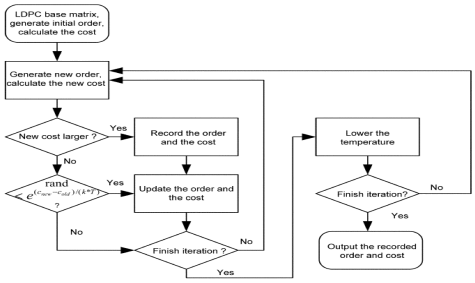


Fig. 6 The Flow Of The Simulated Annealing Algorithm For Finding The Order Of The Layer

Determine the optimal decoding order of the layers, the order of the column decoding of each layer has to be properly scheduled. Since the read and write order of the Channel RAM for a layer are decoupled, they can be scheduled differently and independently. The next layer q+1 are written first. The non-null entries that the layer q+2 are in common with the layer but not with the layer q+1 are written at the end. The other non-null entries are then scheduled in between randomly.

If the latency of the decoding data-path is zero, i.e., the SISO can start writing to the Channel RAM for a certain layer. If the no. of overlapping between layers q and q+2 is larger than the latency of the data-path, memory bypassing cannot be realized. To achieve the maximum memory bypassing operation, add idle clock cycles in the pipeline. When the latency > the no. of overlapping between layers q and q+2 , memory bypassing can always be achieved for all the overlapped columns without the need of inserting idle clock cycles. SISO units working in parallel to calculate multiple check nodes processing for a layer at the same time. The received messages are quantized into 5 bit and the bit-width of the soft output for the variable nodes is 6. The two-output approximation approach [15] is used to implement the SISO. For every check node, two magnitudes, an index, and signs from all variable nodes are generated and stored in the message RAM [23]. The decoder will stop decoding if the signs of the messages An q+1 in one iteration during the decoding satisfy all the parity checks or the number of iteration equals to a pre-defined value i.e.15. The read & write order of the Channel RAM for the non-zero entries within a layer is scheduled to the overlapped columns and minimize the idle cycles. The read and write order of the intermediate data RAM is then fixed, as they are the same as the write and read order of the Channel RAM. By de-coupling the read and write order of the intermediate data RAM. By combining the channel RAM and intermediate data RAM. The messages stored in the same address in the Channel RAM at the same time. Single four port memory is used to implement both the Channel RAM and the intermediate data RAM and it reduces the area. R0 and W0 are used for the read and write access of the original Channel RAM and R1 and W1 are used for the read and write access of the original intermediate data RAM. If the updated results are used directly in the decoding, they will be sent to the shifter directly through a mux-array. The write port W0 and the read port R0 are disabled. Otherwise, the updated messages will be written into the new Channel RAM through the write port W0. The read port R0 used to decode the message in the next layer. Muxes is added to select the output of the Add array and that of the Channel RAM and pipeline registers are added after the Add-array, to implement the memory bypassing scheme. A ROM containing the order of the updated process is added and it is used together with the index generated in the SISO to select the two magnitudes for the update process. By de-coupling the read and write order of the Channel RAM and using the memory bypassing scheme, the no. of read and write access of the Channel RAM is reduced. At the same time, the idle cycles also reduced.

III. Decoding Algorithms for LDPC Codes

The algorithm iteratively computes the distributions of variables in graph-based models such as Message passing algorithm, Sum-Product algorithm or BP algorithm[4].

i. Regular BP algorithm:

The regular BP algorithm[4] with sign-magnitude processing is as follows:

Initialization: $E_{n,m}^{(0)} = 0$

Variable node update rule:

$$T_{n,m}^{(i)} = I_n + \sum_{m' \in \mathrm{M}(n) \backslash m} E_{n,m'}^{(i-1)}$$

Check node update rule:

$$E_{n,m}^{(i)} = \prod_{n' \in \mathrm{N}(m) \backslash n} \mathrm{sgn}\left(T_{n',m}^{(i)}\right) \times \Phi\left(\sum_{n' \in \mathrm{N}(m) \backslash n} \Phi\left(\left|T_{n',m}^{(i)}\right|\right)\right)$$

Last variable node update rule:

$$T_n = I_n + \sum_{m \in \mathrm{M}(n)} E_{n,m'}^{(i-1)}$$

Where, $T_{n,m}$ is the information sent by a VN *n* to its connected check node *m*; $E_{n,m}$ is the message passed from check node *m* to the connected VN *n* ;

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 397 – 401 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

$M(n)$ is the set of check nodes connected to variable node *n*; $N(m)$ is the set of variable nodes connected to CN *m*;

ii. The Min-Sum algorithm:

The difference between BP algorithm and Min-Sum[5] algorithm is that the magnitude part of the check node update rule.

Initialization: $E_{n,m}^{(0)} = 0$

Variable node update rule:

$$T_{n,m}^{(i)} = I_n + \sum_{m' \in M(n) \setminus m} E_{n,m'}^{(i-1)}$$

Check node update rule:

$$E_{n,m}^{(i)} = \prod_{n' \in N(m) \setminus n} \operatorname{sgn}\left(T_{n',m}^{(i)}\right) \times \min_{n' \in N(m) \setminus n} \left|T_{n',m}^{(i)}\right|$$

$$T_n = I_n + \sum_{m \in M(n)} E_{n,m'}^{(i-1)}$$

It is not suited for irregular codes. Check node update by the selection of min. input value. Two messages are saved for each parity check equation. Compensation in terms of subtraction or multiplication. Compensation factor is code specific and unchanged for all iterations and all SNR values.

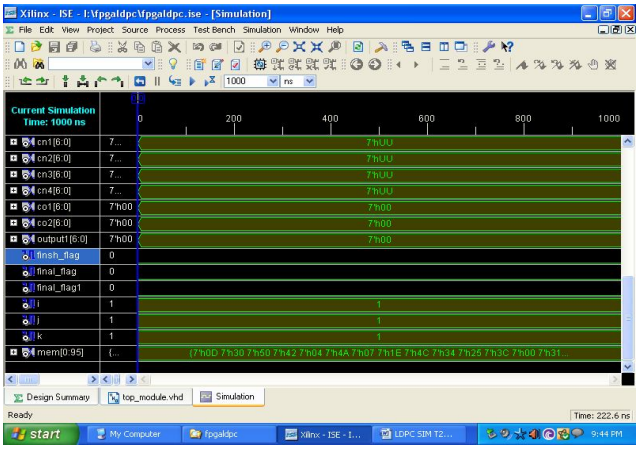


Fig. 7 Output of the LDPC decoder

IV.CONCLUSION

The improved memory-bypassing scheme will reduce the memory access and the energy consumption of the LDPC decoder[6]. It uses the characteristic of the LDPC parity check matrix and searching algorithm. The main advantage of searching algorithm is to find the optimal decoding order that results in maximum number of memory bypassing. The optimum reduction in the memory access can be achieved for LDPC decoder and the memory access is reduced. The decoders were implemented and synthesized with Synopsys. The power consumption of the decoder was simulated using Synopsys VCS-MX and Prime-Time. It compares the clock cycle, and the idle cycles and throughput of the decoder. It reduces the idle cycles by 21.2%–41.3% and the number of idle cycles is reduced by 1.0%–13.2% in view of design. The power consumption of different decoders reduces the energy consumption of the channel RAM and FIFO by 37.6%–47.5%. The reduction in energy is 27.2% to 38.1%. The implemented Decoder for LDPC used for cell-phones to inter-planetary communications[6] systems and observe the exact original data.

V. REFERENCES:

[1] David J.C. MacKay and Radford M. Neal, "Near Shannon Limit Performance of Low Density Parity Check Codes," Electronics Letters, July 1996

[2] Pless, Vera (1982). *Introduction to the theory of error-correcting codes*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons. pp. 8. ISBN 0-471-08684-3.

[3] T. Etzion, A. Trachtenberg, and A. Vardy, Which Codes have Cycle-Free Tanner Graphs?, IEEE Trans. Inf. Theory, 45:6.

[4] , J.S.; Freeman, W.T.; Weiss, Y.; Y. (July 2005). "Constructing free-energy approximations and generalized belief propagation algorithms". *IEEE Transactions on Information Theory* 51 (7): 2282–2312. doi:10.1109/TIT.2005.850085.
Retrieved 2009-03-28.

[5] David J.C. MacKay (2003). Exact Marginalization in Graphs. In David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*, pp. 334–340. Cambridge: Cambridge University Press.

[6] Schramm, W. (1954). How communication works. In W. Schramm (Ed.), *The process and effects of communication* (pp. 3-26). Urbana, Illinois: University of Illinois Press