Improved Layered Min-Sum Decoding Algorithm for Low Density Parity Check Codes

Jin Xie*, Liuguo Yin**, Ning Ge*, and Jianhua Lu*
*Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China
**School of Aerospace, Tsinghua University, Beijing, 100084, China
yinlg@tsinghua.edu.cn

Abstract: In this paper, an improved layered min-sum (ILMS) algorithm is proposed for the decoding of low-density parity-check (LDPC) codes. The proposed method speeds up the decoding by slightly amplifying the message update from check node to variable node. Simulations and analytical results show that compared with the standard layered min-sum decoding algorithm, the proposed scheme may reduce the number of iterations by 8.4%-62.6%, with hardware increase less than 2%.

Key–Words: LDPC, fast convergence decoding algorithm, layered decoding, min-sum algorithm, hardware implementation

1 Introduction

Due to their outstanding error-correcting capability and inherently parallelizable decoding scheme, low-density parity-check (LDPC) codes are adopted by many new generation communication standards, such as DVB-S2, WiMax (IEEE 802.16e), and wireless LAN (IEEE 802.11n). However, compared with turbo-product codes the decoding of an LDPC code needs much more iterations, which may be a big problem for high throughput decoder design. Hence, it is of great significance to improve the convergence speed of the LDPC decoding.

In recent years, many algorithms have been proposed for the decoding of an LDPC code. The sum-product decoding algorithm may obtain the best decoding performance while its hardware is also the most complex. The min-sum algorithm and its improved versions are good trade-off between decoding performance and hardware complexity. Moreover, these decoding algorithms may be implemented with two schedules. The first one is flooding scheme, which updates all check nodes and variable nodes in two successive steps. The second one is layered schedule [1], which offers much faster convergence speed by means of updating variable node messages sequentially rather than simultaneously [2][3]. Thereby, the min-sum algorithm integrated with layered decoding schedule has been shown to be the best choice for designing LDPC decoders with good decoding performance and low hardware complexity [4][5]. However, since the variable node message update of the layered min-sum algorithm is implemented sequentially rather than simultaneously, there is still a challenging problem for high throughput LDPC decoder design. Many research works have been conducted to further optimize the layered min-sum decoding algorithm to increase the convergence speed. In [6], a fast-convergence decoding method is proposed for LDPC codes used in the WiMax systems. This decoding method serially decodes block codes with identical parity-check matrix, and accelerates the iteration speed by breaking up large codes into small ones. However, this scheme is limited to QC-LDPC codes. In [7], another new decoding algorithm utilizing the zigzag connectivity of linearly encodable LDPC codes is introduced, which is also confined to IRA codes.

In this paper, an improved layered min-sum decoding algorithm is proposed for all kinds of LDPC codes. The idea of this algorithm comes from an observation that during the LDPC decoding, the check-node to variable-node messages computed in current iteration is generally more reliable than those computed in previous iteration. Thus the difference between the check-node to variable-node message from current iteration and that from previous iteration may denote the convergence direction of each variable-node. Based on this observation, we revised the iterative function of the layered min-sum algorithm by slightly amplifying the check-node to variable-node message update. Simulation results show that with the proposed decoding method, the number of iterations may reduce by 8.4% - 62.6% when compared with the decoding with the standard layered min-sum decoding algorithm, which is of great significance for high throughput LDPC decoder design.
The rest of this paper is organized as follows. In section 2, the standard layered min-sum decoding algorithm is described. Then, the improved layered min-sum algorithm is proposed in section 3, and simulation results for the proposed algorithm are presented in section 4. After that, the hardware implementation issue of the proposed method is discussed in section 5. Finally, conclusions are drawn in section 6.

2 LDPC Decoding Schedules

Consider an \((N, K)\) LDPC code defined by a parity check matrix \(H_{mn}\), where \(N\) is the codeword length, \(K\) is the number of information bits, and \(M = N - K\) is the number of parity check equations. A parity check code can be described by a bipartite graph (Tanner graph) with \(N\) variable-nodes corresponding bits in the codeword and \(M\) check-nodes representing parity check equations. The edges connecting variable-nodes and check-nodes map the ’1’s in \(H_{mn}\).

Furthermore, define

\[ N(m) = \{n : H_{mn} = 1\} \]

to be the set of variable-nodes connected to check-node \(m\), and \(K_m\) to be the number of elements in \(N(m)\).

The decoding algorithm exchanges messages between variable-nodes and check-nodes. The messages are posteriori probabilities in the form of Log Likelihood Ratio (LLR), which is defined as:

\[ LLR(x_i) = \log\left( \frac{P_r(x_i = 1)}{P_r(x_i = 0)} \right) \]  

2.1 Flooding Schedule

Let \(\lambda_n^k\) be the soft output (SO) of bit \(n\), \(\lambda_{nm}^k\) and \(\Lambda_{nm}^{k-1}\) be the message sent from variable-node \(n\) to check-node \(m\) and the message from check-node \(m\) to variable-node \(n\) in the \(k\)th iteration, respectively. The classic flooding schedule updates all the variable-nodes and check-nodes in two successive steps.

During the \(k\)th iteration, the check-node \(m\) receives the message to its neighboring variable-node \(n\) as:

\[ \lambda_{nm}^k = \lambda_{ch} + \sum_{n' \in N(m) \setminus n} \Lambda_{nm}^{k-1} \]  

where \(\lambda_{ch}\) is the channel a-priori information on the current bit.

Then, combined with the min-sum algorithm, updated message from check-node \(m\) to its adjacent variable-node \(n\) is computed as:

\[ \Lambda_{nm}^k = \prod_{n' \in N(m) \setminus n} \text{sign}(\lambda_{n'm}^k) \times \max(\min_{n' \in N(m) \setminus n} |\lambda_{n'm}^k| - \beta, 0) \]  

The final decision on a bit is taken in the end of the iteration \(I\). The SO of bit \(n\) is calculated as:

\[ \lambda_n = \lambda_{ch} + \sum_{n' \in N(m)} \Lambda_{nm}^I \]

If \(\lambda_n\) is negative, the bit \(n\) is judged as a 0, otherwise it is considered as a 1.

2.2 Layered Schedule

Unlike the two-phase update in flooding schedule, layered schedule considers the parity check matrix as layers of check equations and updates the variable-node information right after updating check-node information of current layer. Vertical shuffle layer and horizontal shuffle layer have been considered in [8]. In this section, only horizontal shuffle layered schedule is introduced.

The horizontal shuffle decoding is check-node centric, which updates the SO values for each check-node \(m\):

- first, the \(\Lambda_{nm}^k\) are updated using equation (3).
- then, all the \(\lambda_n\) coming out of the variable-nodes connected to this check-node are updated immediately with equation (4).

It is noted that the flooding schedule, the updated messages in the \(k\)th iteration are based on the messages calculated in the previous iteration. However, in the layered schedule, this computation is done on messages already updated on the \(k\)th iteration by a different check-node. Based on the fact that layered schedule converges faster than the flooding schedule, we conclude that using messages already updated in the \(k\)th iteration, which are statistically more reliable, accelerates the iteration. Further analysis shows that the difference between the \(k\)th and \((k-1)\)th messages is \((\Lambda_{nm}^k - \Lambda_{nm}^{k-1})\). Adding such a term to the SO value might result in a faster speed than standard layered schedule. Inspired by this, we propose the improved layered decoding as follows.

3 Improved Layered Min-Sum Decoding Algorithm

Similar to the standard layered decoding algorithm, the proposed decoding method can be summarized in three steps:
1. Initialization: A-posteriori information $\lambda_n^0$ is initialized with the channel LLR $\lambda_{ch}$, and Each $\Lambda_{nm}^0$ is initialized to zero. Specifically for BPSK signals transmitted in the AWGN channel,

$$\lambda_{ch} = \frac{2y_n}{\sigma^2}$$

$y_n$ being the the received BPSK signal value, and $\sigma^2$ being the noise variance.

2. Iterative decoding: During the $k$th iteration, we observed that:

$$\lambda_n^k = \lambda_{nm}^{k-1} - \Lambda_{nm}^{k-1}$$

As a result of Equ.(5), messages from variable-node to check-node could be generated using the SO and the meomory for $\lambda_{nm}^k$ is saved. When $\lambda_{nm}^k$ is extracted, update messages from variable-node $n$ to check-node $m$ using Eqn.(3).

At the same time, the SO of bit $n$ is computed as:

$$\lambda_n^{int} = \lambda_{nm}^k + \Lambda_{nm}^k$$

$$\lambda_n^k = (1 + \omega) * \lambda_n^{int} - \omega * \lambda_n^{k-1}$$

where $\lambda_n^{int}$ is the interim LLR and $\omega$ is the accelerating parameter. If we set $\omega = 0$, $\lambda_n^{int}$ is equal to $\lambda_n^k$ and our proposed algorithm goes back to the standard layered decoding algorithm.

3. Check stop rules: compute the hard decoding output for current iteration:

$$\hat{c} = \begin{cases} 0 & \text{if } \lambda_n^k \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

If constraint function $H \cdot \hat{c} = 0$ is satisfied or the maximum number of iteration has performed, stop the iteration and output decoded results. Otherwise, continue the iteration.

Noted that in Eqn.(7) of the proposed algorithm the variable-node message update function has been revised by assigning more weight to current LLR . Merging Eqn.(5) and (6) into (7), we get that

$$\lambda_n^k = \lambda_{nm}^k + \Lambda_{nm}^k + \omega * (\Lambda_{nm}^k - \Lambda_{nm}^{k-1})$$

Compared with the standard layered decoding, a new term $(\Lambda_{nm}^k - \Lambda_{nm}^{k-1})$ is added. As the check-node to variable-node message becomes more reliable as the iteration proceeds, the term indicates the convergence direction for variable-node $n$. Therefore, adding such a term to the variable-node message update function will accelerate the convergence speed.

4 Performance Simulations and Numerical Results

To evaluate the performance of the proposed improved layered min-sum (ILMS) decoding and the standard layered min-sum (SLMS) decoding, we simulated with DVB-S2 LDPC codes. We choose a maximum of 500 iterations for each decoding. For the min-sum algorithm, $\beta = 0.5$ is used. The parameter $\omega$ in Eqn.(7) of the proposed algorithm is set to be 0.05. But further simulations demonstrate that for $0.02 < \omega < 0.06$ the proposed decoding algorithm has almost the same error-correcting performance and convergence speed. Thereby, the performance of the proposed algorithm is not sensitive to parameter $\omega$.

The average numbers of iterations for the two decoding algorithms are listed in Table 1, from which we can see that with the proposed decoding scheme the average number of iterations may be reduced by 8.4%-19.7% at high SNR; while the reduced number of iterations may be increased to be 30% - 62.6% when the channel SNR is low. Moreover, the BER performances with the proposed decoding algorithm and the standard layered min-sum decoding scheme are plotted in Figure 1. It is shown that the proposed decoding method achieves the same error correction performance as the standard layered min-sum decoding scheme. Besides, simulation results are better when the code rate is low, as Table 1 shows that at 1/4 code rate, the reduced iteration percentage changes slowly when the $E_s/N_0$ increases, which means our algorithm boosts the iteration significantly for a wider range of SNR.
### Table 1: Average Iteration Number of SLMS and ILMS

<table>
<thead>
<tr>
<th>Code Rate</th>
<th>$E_s/N_0$(dB)</th>
<th>Average Iteration Number</th>
<th>Average Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standard Layered Decoding</td>
<td>Improved Layered Decoding</td>
</tr>
<tr>
<td>1/4</td>
<td>-3.00</td>
<td>117.0</td>
<td>78.7</td>
</tr>
<tr>
<td></td>
<td>-2.95</td>
<td>77.7</td>
<td>56.5</td>
</tr>
<tr>
<td></td>
<td>-2.90</td>
<td>61.7</td>
<td>47.4</td>
</tr>
<tr>
<td></td>
<td>-2.85</td>
<td>51.9</td>
<td>41.7</td>
</tr>
<tr>
<td>1/2</td>
<td>0.85</td>
<td>84.7</td>
<td>63.8</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>34.9</td>
<td>28.5</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>25.0</td>
<td>21.8</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>21.6</td>
<td>19.4</td>
</tr>
<tr>
<td>2/3</td>
<td>2.95</td>
<td>94.9</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>33.2</td>
<td>23.3</td>
</tr>
<tr>
<td></td>
<td>3.05</td>
<td>21.3</td>
<td>18.8</td>
</tr>
<tr>
<td></td>
<td>3.10</td>
<td>17.8</td>
<td>16.3</td>
</tr>
</tbody>
</table>

### 5 Hardware Implementation Complexity of the Proposed Algorithm

Low-cost hardware architecture for standard layered algorithm has been proposed in many publications. Most of the effective techniques could be applied to our algorithm. First of all, the check-node message compression technique still works. As mentioned in [4], the magnitudes of all the $\Lambda_{nm}^k$ from the same check-node m have only two values: either the minimum, or the sub-minimum. So we choose to store the $K_m$ outgoing check-node message signs, 2 magnitudes and the index of the minimum. Second, the quantization and clipping results in [9] also stand for the proposed algorithm because the check-node outgoing messages are the same as the standard layered schedule.

The data path of our proposed algorithm is shown in Figure 2, which is similar to the data path in [4]. It operates in a serial-in, serial-out way, which is convenient for check-node message update and programmable for different codes. In Eqn.(7), $\Lambda_{nm}^{k-1}$ is required to generate $\Lambda_{n}^{k}$. As a result, $\Lambda_{n}^{k-1}$ is stored into delay registers. The data path decompresses the check-node message on-the-fly, thus effectively reduces the memory especially for high-degree check nodes. In order to implement Eqn.(7) in an economical way, multiplication could be replaced by shift and addition. As mentioned above, our algorithm is insensitive to accelerating parameter $\omega$. For example, $\omega$ is set as 1/32 and the multiplication equals to shift arithmetic right (SAR) 5 bits. As a result, only 3 bits are required to store $\omega \times \Lambda_{n}^{k-1}$ if we use 8 bits to quantize $\Lambda_{n}^{k}$. Similarly, $(1 + \omega) \times \Lambda_{n}^{int}$ is the sum of $\Lambda_{n}^{int}$ and its SAR.
In many applications, a fully serial solution for LDPC decoder will lead to a low throughput, while a fully parallel solution probably results in congestion in layout due to wire connections between all the check-nodes and variable-nodes. Therefore, for structured LDPC codes like DVB-S2 or WiMax, semi-parallel layered decoding is a useful way to increase the throughput of the whole system as shown in Figure 3. Take DVB-S2 for example, the input parity-check LLRs are stored in an interleaved way so that the corresponding parity check equations are quasi cyclic structured. Blocks of 360*360 submatrix are combination of shifted identity matrix, so the SO values could be read as vectors from the memory and rotated by a barrel shifter to match the order of check equations. The above data path is duplicated to process the incoming SO values simultaneously. After updated by the data paths, the SO values are rotated back and returned to the same position in the SO memory. The addresses of SO values and shift magnitudes are extracted from parity check matrix. The addresses of check-node messages go ahead one step as the data paths have processed current block of parity equations.

The decoders in [2] and [4] use 30 iterations to achieve the BER performance. The proposed architecture, however, reduces 5 iterations, e.g. 25 iterations is enough for the same performance. Thus the throughput increases about 17% with the same clock frequency. At the same time, the cost increases slightly. A 1/2 code rate DVB-S2 decoder is implemented on Xilinx FPGA Virtex 4 LX100. There are 90 parallel data paths in our decoder, and 8 bit inputs and 6 bit message value are used. Consumed resources are listed in Table 2, which depend on the parallel level and quantization bits. Comparing our algorithm with the standard one, only 2% more 4-input LUTs are required and no extra RAMB16s are instantiated. In the case of 45 parallel data paths, synthesis report shows that about 1% more LUTs are required. So approximately the overhead is 1% more LUTs for every 45 data path.

The static timing report shows that a maximum clock frequency of 115.7 MHz is achieved. The 1/2 code rate decoder throughput is 119 Mbps, which is calculated by:

\[
\text{Throughput} = \frac{f_{\text{clk}} \cdot \text{frame length} \cdot Ndp}{Niter \cdot K_m \cdot Ncheck}\]  

(10)

where \(f_{\text{clk}}\) is the operating clock frequency, \(Ndp\) is the number of data path, \(Niter\) is the number of iteration, and \(K_m\) is a constant for each specific code rate in DVB-S2, \(Ncheck\) is the number of check equations.

### Table 2: Synthesis Results On XC4VLX100 FPGA

<table>
<thead>
<tr>
<th>Adders/Subtractors</th>
<th>XST Synthesis Report</th>
<th>LS</th>
<th>ILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counters</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Registers</td>
<td>9557</td>
<td>7577</td>
<td></td>
</tr>
<tr>
<td>Comparators</td>
<td>551</td>
<td>551</td>
<td></td>
</tr>
<tr>
<td>Multiplexers</td>
<td>450</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>Xors</td>
<td>720</td>
<td>720</td>
<td></td>
</tr>
<tr>
<td>Total Number of 4-LUT</td>
<td>33%</td>
<td>31%</td>
<td></td>
</tr>
<tr>
<td>Number of RAMB16s</td>
<td>55%</td>
<td>55%</td>
<td></td>
</tr>
</tbody>
</table>

### 6 Conclusions

In this paper, we have proposed an improved layered min-sum algorithm for the decoding of low-density parity-check (LDPC) codes. The proposed method speeds up the decoding by slightly amplifying the message update from check node to variable node. It is shown that with our improved scheme the number of iterations may be reduced by 8.4%-62.6%, while the hardware complexity of our method is almost the same as that of the standard min-sum algorithm.

Acknowledgements: This work was supported in part by the National Science Foundation of China (60525107, 60532070) and the National Fundamental Research Program of China (2007CB310600).

### References:


