

Quantized Iterative Message Passing Decoders with Low Error Floor for LDPC Codes

Xiaojie Zhang and Paul H. Siegel
University of California, San Diego

1. Introduction

Low-density parity-check (LDPC) codes have been the focus of much research over the past decade as a consequence of their near Shannon-limit performance under iterative message-passing (MP) decoding [1]. However, the error floor phenomenon has hindered the adoption of LDPC codes and iterative decoders in some applications requiring very low error rates. Roughly speaking, an error floor is an abrupt change in the slope of the error-rate performance curve of an MP decoder in the high SNR region. Since many important applications, such as data storage and high-speed digital communication, often require extremely low error rates, the study of error floors in LDPC codes remains of considerable practical, as well as theoretical, interest.

LDPC codes are usually represented by a Tanner graph, as shown in Figure 1. Variable nodes represent codeword bits and check nodes represent parity-check constraints. The degree of a node refers to the number of edges adjacent to it. An LDPC code has the property that the node degrees are small relative to the codeword length.

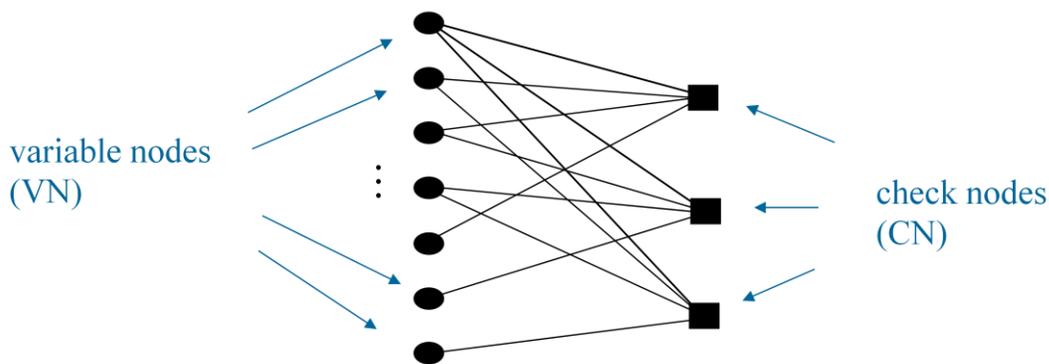


Figure 1. Tanner graph representation of LDPC codes

Optimal decoding of powerful LDPC codes is impractical, but suboptimal iterative message-passing (MP) decoding algorithms, such as the sum-product algorithm (SPA)

Research Highlight

and the min-sum algorithm (MSA), have been found to offer a useful tradeoff between decoder implementation complexity and error-rate performance.

The iterative decoder alternates between two phases, a “VN-to-CN” phase during which VNs send messages to CNs along their adjacent edges, and a “CN-to-VN” phase during which CNs send messages to their adjacent VNs. The message update rules, which we will now briefly describe for the SPA, are depicted schematically in Figures 2 and 3, respectively. In the figures, the set of neighboring CNs of VN i is denoted by $N(i)$, and the set of neighboring VNs of CN j is denoted by $N(j)$. In the initialization step of the decoding process, VN i forwards the same message to all of the CNs in $N(i)$, namely the log-likelihood ratio (LLR) L_i^{ch} derived from the corresponding channel output. In the CN-to-VN message update phase, CN j uses the incoming messages and the message update rule shown in Figure 2 to compute and forward, to VN i in $N(j)$, a new “CN-to-VN” message, $L_{j \rightarrow i}$. VN i then processes its incoming messages according to the update rule shown in Figure 3 and forwards to each adjacent CN an updated “VN-to-CN” message, $L_{i \rightarrow j}$. After a prespecified number of iterations, VN i sums all of the incoming LLR messages to produce an estimate of the corresponding code bit i . Note that all of the “CN-to-VN” message updates can be done in parallel, as can all of the “VN-to-CN” message updates. This enables efficient, high-speed software and hardware implementations of the decoding algorithm.

$$L_{j \rightarrow i} = 2 \tanh^{-1} \left(\prod_{i' \in N(j) \setminus i} \tanh \frac{L_{i' \rightarrow j}}{2} \right)$$

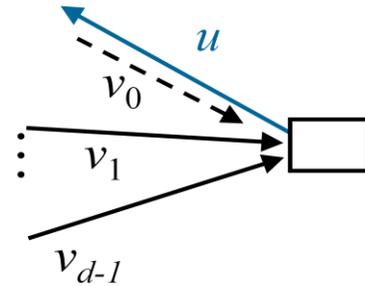
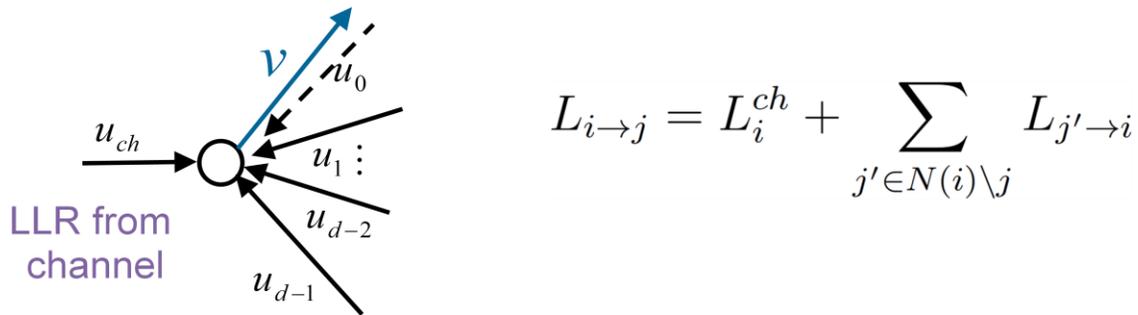


Figure 2. CN-to-VN message update: Each CN receives log-likelihood ratio (LLR) information from all of its neighboring VNs. For each such VN, it generates an updated “check-to-variable” message using the inputs from all other neighboring VNs.



$$L_{i \rightarrow j} = L_i^{ch} + \sum_{j' \in N(i) \setminus j} L_{j' \rightarrow i}$$

Figure 3. VN-to-CN message update: Each VN receives log-likelihood ratio (LLR) information from all of its neighboring CNs. For each such VN, it generates an updated “variable-to-check” message using the inputs from all other neighboring CNs.

2. Trapping Sets and Error Floors

Error patterns in the error floor region often correspond to sets of variable nodes that lie in subgraphs of the Tanner graph with special combinatorial structure. For the binary symmetric channel (BSC) and the AWGN channel, these sets and their induced subgraphs have been referred to as near-codewords, trapping sets, or absorbing sets. The term “trapping set” is often used generically in reference to such error-prone substructures. For example, Figure 4 shows the error floor of the rate-1/2, length-2640 Margulis code on the AWGN channel. Most SPA decoding failures in the error floor region for this code correspond to the two trapping sets shown in Figure 5. (See [4] [5] .)

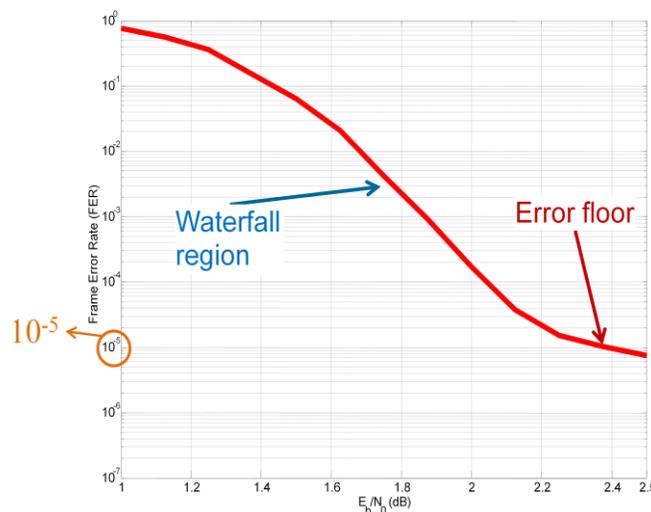
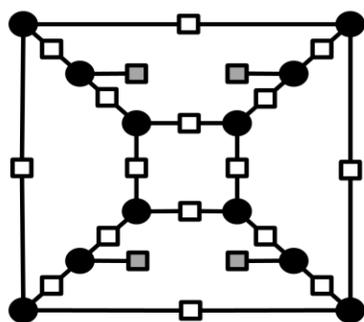
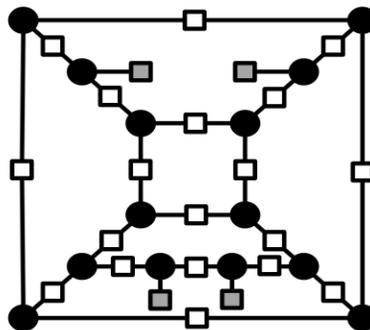


Figure 4. Error floor of Margulis code of length 2640, observed by Mackay-Postol in [4] and Richardson in [5] .



(12,4) trapping set



(14,4) trapping set

Figure 5. Two trapping sets correspond to the dominant errors in the error floor region for the rate-1/2, length-2640 Margulis code. The VNs in the trapping set are represented as solid black circles, and the CNs are represented as squares. If the variable nodes are set to the value 1, the CNs where parity-checks are not satisfied are shown as shaded squares.

3. New Quantization Rule to Lower Error Floors

It is known that error floor characteristics also depend on system implementation issues, such as the quantization of channel LLR and message values, the decoding algorithm formulation, and the number of decoder iterations. In an idealized scenario, where all VNs outside the trapping set are assumed to have been correctly decoded, and where the VNs in the trapping set satisfy a certain separation assumption, we proved that decoders using the SPA and MSA could correct trapping set errors if the maximum magnitude of messages passed between nodes is not restricted [2] [3].

However, in practice, the messages must be represented by a limited number of bits. Typically, decoder implementations use uniform quantization, which permits fine resolution over a limited range (i.e., large values are “clipped”) or coarse resolution over larger range (i.e., message values are represented less accurately). If, as our idealized analysis suggests, a large range of values needs to be represented, the precision of small messages would have to be sacrificed, significantly degrading the error-rate performance. To resolve this dilemma, we propose a new quantization method which allows fine resolution for small messages and somewhat coarser representation over a range of larger messages. This method, which we call $(q+1)$ -bit quasi-uniform quantization, is portrayed graphically in Figure 6. The corresponding quantization table for $q=3$ is shown in Figure 7. We assume here that the Tanner graph is VN-degree

Research Highlight

regular, meaning that all variable nodes have the same degree, d_v ; however, similar results and quasi-uniform quantization techniques can be applied to LDPC codes with different VN-node degrees.

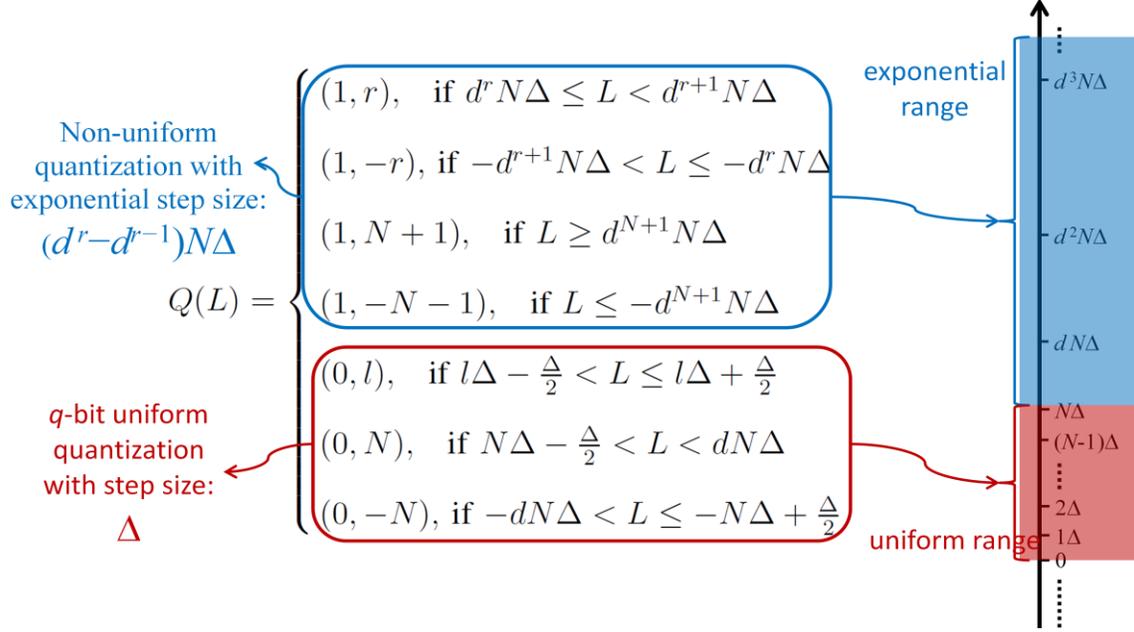


Figure 6. $(q+1)$ -bit quasi-uniform quantization, where $N = 2^{q-1} - 1$, $-N+1 \leq l \leq N-1$, $1 \leq r \leq N$, Δ is the nominal step size, and d is a quantization parameter within the range $(1, d_v - 1]$.

Research Highlight

| Messages | Ranges | Values | |
|----------|------------------|--------|---|
| (0,000) | (-0.5, 0.5] | 0 | 3-bit uniform quantization with step size: 1 |
| (0,001) | (0.5, 1.5] | 1 | |
| (0,010) | (1.5, 2.5] | 2 | |
| (0,011) | (2.5, 9) | 3 | |
| (1,000) | [9, 27) | 9 | 3-bit exponential quantization step size: $3 \cdot (3^r - 3^{r-1})$ |
| (1,001) | [27, 81) | 27 | |
| (1,010) | [81, 243) | 81 | |
| (1,011) | [243, ∞) | 243 | |

Max magnitude in 4-bit uniform quantization with step size 1 is only 7

Figure 7. An example of (3+1)-bit quasi-uniform quantization, where $\Delta = 1$, $d = 3$, for the positive range of message values.

4. Numerical Results

Figure 8 shows SPA performance results for a rate-0.3, length-640, quasi-cyclic LDPC code on the AWGN channel, with a maximum of 200 decoder iterations. We can see that the proposed quantization performs even better than floating-point SPA due to its faster convergence. Note that the slope of the error rate curve of the (5+1)-bit quantized SPA is the same as the LP decoder, which is steeper than the floating-point SPA curve.

In Figure 9, we show results for the rate-0.5, length-2640 Margulis code. In this decoder performance comparison, we also considered the “dual quantization” SPA decoding technique proposed in [6]. In dual quantization, two uniform quantization rules with different step sizes are used in CN-to-VN message update. Specifically, using the notation of [6], Q_m/f quantization uses a signed fixed-point number with m bits to the left of the radix point to represent integer values, and f bits to the right of the radix point to represent fractional values. For example, a Q4.2 quantizer has uniform quantization step size of 0.25 and a range of $[-7.75, 7.75]$.

In the figure, we see that the proposed (5+1)-bit quasi-uniform quantizer has the best error-floor performance; it even improves upon the 64-bit double-precision floating-point SPA decoder in low error-rate region as a result of its faster convergence.

Research Highlight

Extensive computer simulation results, not shown here, show that the error-rate performance in the water-fall region can also be improved by using quasi-uniform quantization with carefully chosen quantization parameters.

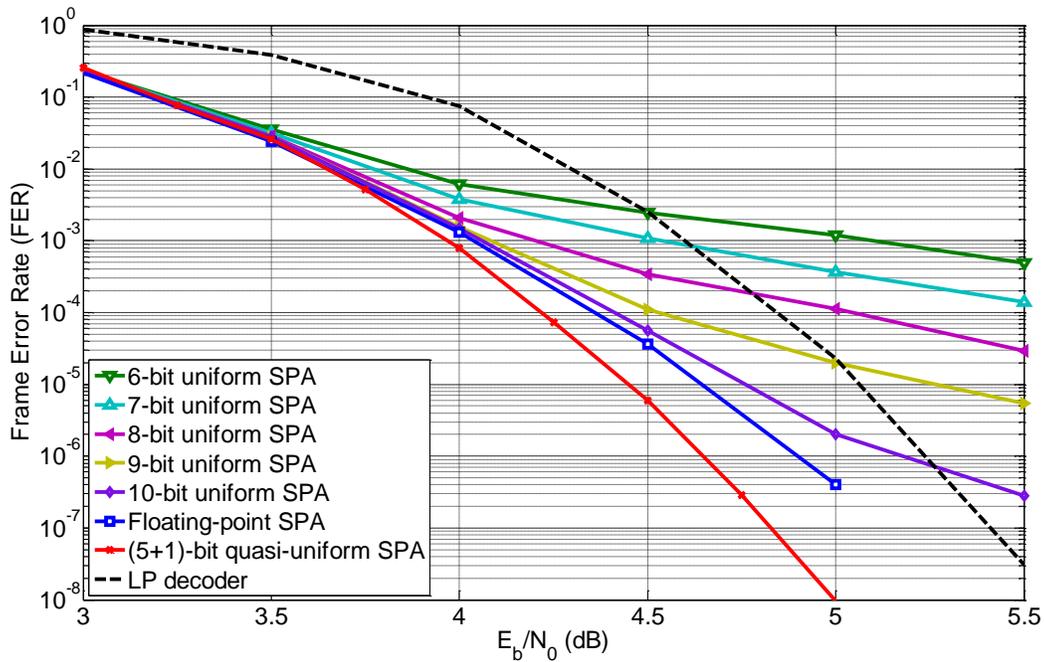


Figure 8. FER results of SPA decoder on the (640,192) QC-LDPC code on AWGNC. The uniform quantization step $\Delta = 0.25$, and $d = 1.5$ in (5+1)-bit quasi-uniform quantization, maximum number of iterations is 200.

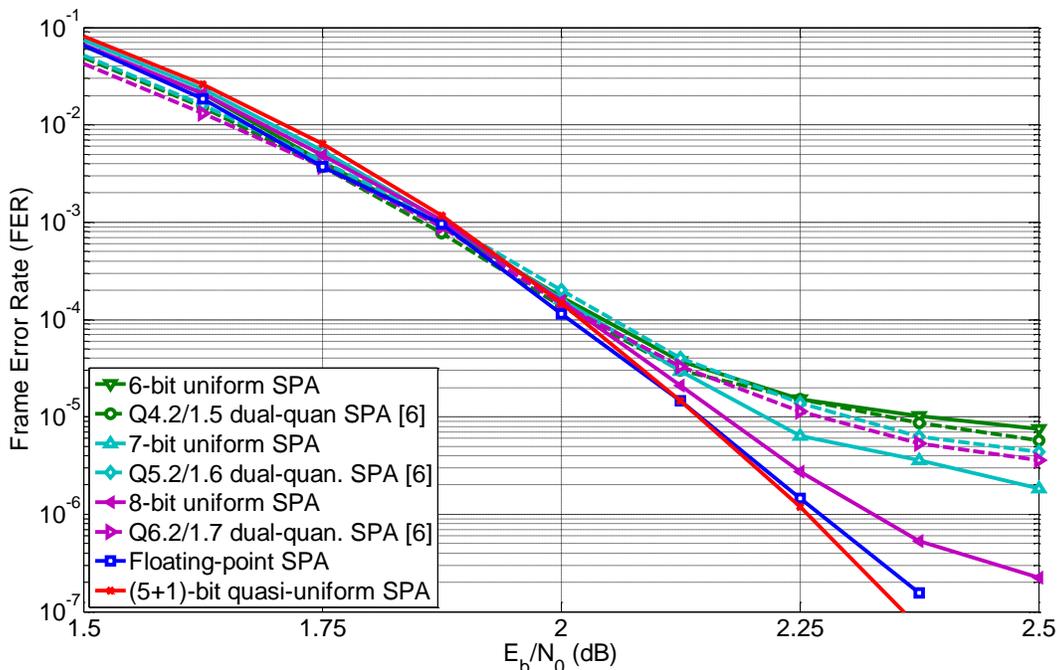


Figure 9. FER results of approximate-SPA decoder on the Margulis code of length 2640 on AWGNC. Uniform quantization step $\Delta = 0.25$, and $d = 1.3$ in (5+1)-bit quasi-uniform quantization, maximum number of iterations is 200.

5. Summary and Conclusions

In this research, we have shown that the use of uniform quantization in iterative message-passing decoding can be a significant factor in the appearance of error floors in LDPC code performance. To address this problem, we have proposed a new quasi-uniform quantization method that effectively extends the dynamic range of the quantized message values. Without modifying the CN-to-VN and VN-to-CN message update rules or adding extra stages to standard iterative decoding algorithms, the use of quasi-uniform quantization was shown to significantly lower the error floors of two well-studied LDPC codes.

References

- [1] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. 8, pp. 21-28, Jan. 1962.
- [2] X. Zhang and P. Siegel, "Quantized Min-Sum Decoders with Low Error Floor for LDPC codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Cambridge, MA, July 2-5, 2012.
- [3] X. Zhang and P. Siegel, "Efficient Algorithms to Find All Small Error-Prone Substructures in LDPC Codes," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Houston, TX, Dec. 5-9, 2011.

Research Highlight

- [4] D. MacKay and M. Postol, "Weakness of Margulis and Ramanujan-Margulis low-density parity check codes," *Electron. Notes Theor. Comp. Sci.*, vol. 74, 2003.
- [5] T. Richardson, "Error-floors of LDPC codes," in *Proc. of the 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 1-3, 2003, pp. 1426-1435.
- [6] Z. Zhang, "Design of LDPC decoders for improved low error rate performance," Ph.D. dissertation, Univ. of California at Berkeley, 2009.