Design of Low-Density Parity Check Codes for 5G New Radio

Tom Richardson and Shrinivas Kudekar

The authors briefly review the requirements of the new channel code for 5G NR. They then describe the LDPC code design philosophy and how the broad requirements of 5G NR channel coding led to the introduction of novel structural features in the code design, culminating in an LDPC code that satisfies all the demands of 5G NR.

Abstract

Turbo codes, prevalent in most modern cellular devices, are set to be replaced by LDPC codes as the code for forward error correction. This transition was ushered in mainly because of the high throughput demands for 5G New Radio (NR). The new channel coding solution also needs to support incremental-redundancy hybrid ARQ, and a wide range of blocklengths and coding rates, with stringent performance guarantees and minimal description complexity. In this article, we first briefly review the requirements of the new channel code for 5G NR. We then describe the LDPC code design philosophy and how the broad requirements of 5G NR channel coding led to the introduction of novel structural features in the code design, culminating in an LDPC code that satisfies all the demands of 5G NR.

Introduction

5G NR Switches to LDPC Codes

The New Radio (NR) access technology will mark a transition in forward error correction (FEC) coding for Third Generation Partnership Project (3GPP) cellular technologies [1]. Turbo codes, which have been the primary coding scheme in the third and fourth generation (3G and 4G) will be replaced by low-density parity check (LDPC) codes. The primary driver for this transition is the high throughput requirement (5 Gb/s) for 5G systems. Turbo codes and LDPC codes are close cousins in that decoders for both are of the message passing type in which information is propagated inside a graphical structure representing the code. Compared to turbo code decoders, the computations for LDPC codes decompose into a larger number of smaller independent atomic units; hence, greater parallelism can be more effectively achieved in hardware. LDPC codes have already been adopted into other wireless standards including IEEE 802.11, digital video broadcast (DVB), and Advanced Television System Committee (ATSC). The broad requirements of 5G NR demand some innovation in the LDPC design. The need to support IR-hybrid automatic repeat request (HARQ) as well as a wide range of block sizes and code rates demands an adjustable design. In contrast, the 802.11 [10] codes consist of 12 codes, each distinguished by a distinct description. For 5G NR the number of possible codes measures in the thousands, so a unification of many codes into a single description was required.

The primary metrics of interest for the FEC in communication systems are complexity (which includes area and power) and performance (coding gain). State-of-the-art LDPC designs are very often the best option in both respects, which helps to explain their popularity. As shown in Fig. 1, LDPC codes can offer higher coding gains than turbo codes and have lower error floors, and, as, shown in Fig. 2, LDPC codes can simultaneously be computationally more efficient than turbo codes, that is, require fewer operations to achieve the same target block error rate (BLER) at a given energy per symbol (signal-to-noise ratio, SNR). Throughout the article, K and N denote the number of information and codeword bits, respectively.

Turbo codes (more specifically parallel concatenated turbo codes) are obtained by encoding an information bit sequence multiple times using a convolutional code, where the information bit sequence is permuted for the different encodings. LTE turbo codes use two such encodings with each encoding producing a parity bit sequence of essentially the same length as the information bit sequence. The main turbo decoding operation consists of soft decoding of the convolutional codes, which is performed on a one-dimensional graphical structure — a trellis — representing the convolutional code. That operation is inherently serial but can be parallelized by breaking the trellis into segments. Some computational overhead is incurred because results from the boundaries of those segments are typically discarded. Thus, there are practical barriers to making the segments very small and providing a very high degree of parallelism. LDPC decoders operate by repeatedly decoding single parity check codes involving a small number of bits. Decoding of these parity checks is operationally independent and hence easily parallelized.

The encoding of LTE turbo codes inherently yields rate 1/3 codes. Higher coding rates are achieved by puncturing (removing) parity bits, and IR-HARQ is achieved by subsequent transmission of the punctured bits. Decoding is performed for the rate 1/3 code with neutral values substituted for the received values of punctured bits. For a given information block length, the decoding complexity is therefore essentially the same as for the rate 1/3 case. The design of 5G

Digital Object Identifier:
10.1109/MCOM.2018.1700839

The authors are with Qualcomm.

NR LDPC codes consists of a high rate code, which we call the core, together with an extending sequence of parity bits that can be appended to lower the rate of the code. When these bits are not transmitted, they can be eliminated entirely from the decoding process. Thus, for the LDPC-based solution, the number of operations in decoding scales with the transmitted block length of the code rather than with the information block length. Consequently, the throughput of the LDPC decoder increases as the code rate increases, making it easier to reach high peak throughputs for a given low rate baseline throughput. This is summarized in Fig. 3.

LDPC Codes

LDPC codes have two natural commonly used descriptions. One is based on a parity check matrix (PCM) where the code is understood to be the set of (right) null vectors of the PCM. The second is a graphical description known as a Tanner graph [2], which mirrors the parity check matrix. LDPC Tanner graphs are bipartite graphs where one set of nodes, the variable nodes, correspond to bits in the codeword, and the other set of nodes, the check nodes, correspond to the parity checks that the bits must satisfy. Thus, the variable nodes correspond to the columns of the parity check matrix, and the check nodes correspond to the rows. A variable node is connected to a check node if the corresponding entry in the parity check matrix is non-zero. Figure 4 provides an example of the two representations. The graphical representation is useful for visualization since decoding is best understood as a computational process that passes messages along the edges of graph, where the messages represent probability distributions on the associated bit, while processing and updating the messages at the nodes. The most powerful such algorithm is known as the sum-product (SP) algorithm [3], and performance results presented in this article are based on its use.

Practical LDPC designs use structured graphs wherein the structure is imposed to facilitate parallelism and to organize the message passing process. There is, however, another quite distinct notion of structure in the graph design, one which targets performance. This structure concerns large-scale statistics of the graphs and the messages passed on them and how those statistics affect performance. The LDPC design process typically consists of first describing the structure that governs the macroscopic statistics of the graph, and then producing an instance reflecting that structure. Typically, the desired macroscopic structure can be captured by a small graph, called a *base graph* or a *protograph* [4, 5]. Sometimes the base graph is already an expanded representation of the macroscopic structure. One of the first developments in the design of macroscopic structure to achieve capacity approaching LDPC codes was the use of irregularity [6]. The idea was to specify the degrees of the nodes in the Tanner graph but to connect the two sides essentially randomly. The structure, in this case, consists of specifying the degrees of the nodes, and irregularity amounts to having a variety of degrees. Most LDPC wireless standards adopted this idea with one additional piece of structure known as an accumulate chain (degree two variable nodes), which simplifies encoding and improves performance. For many years, it has been known that additional structural features can improve LDPC performance [4, 5, 15], and the 5G NR LDPC codes incorporate some of those features.

While the base graph encapsulates the desired macroscopic structure, a larger graph can be constructed by lifting the base graph. The lifted graph is formed by taking Z copies of the base

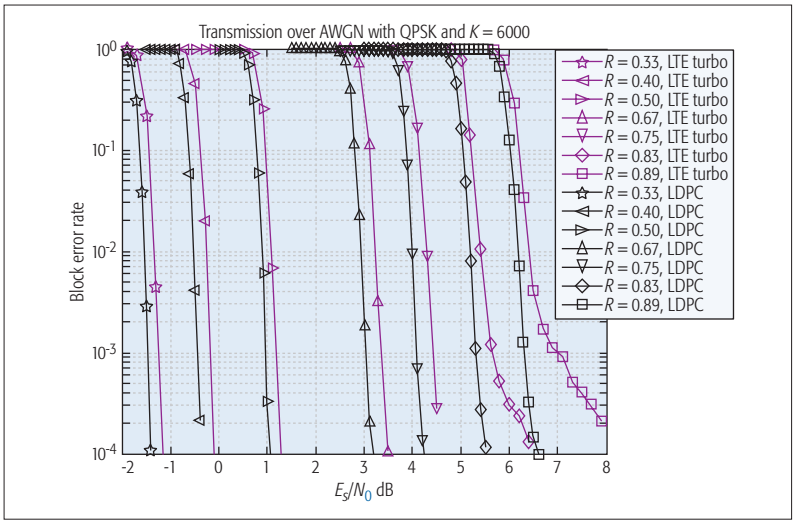

Figure 1. Performance comparison between LDPC with sum-product (SP) decoding and LTE turbo with log-MAP decoding. Both are decoded using their optimal decoders with sufficiently large number of iterations. The plot highlights the fundamental differences in the code structure itself.

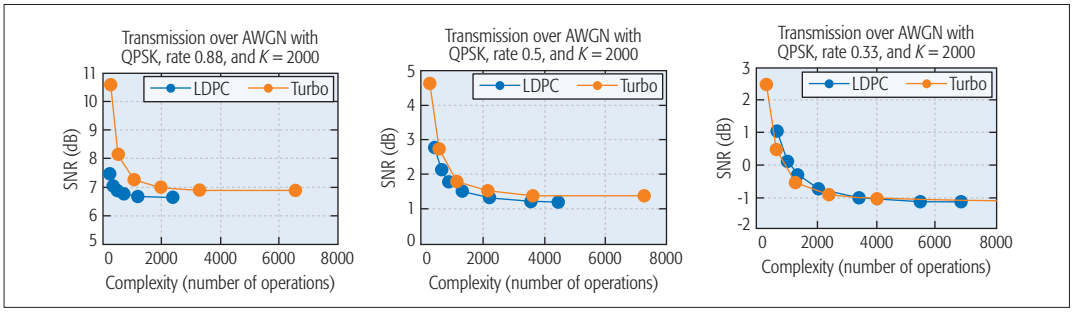

Figure 2. SNR vs. number of operations (add, multiply, min, etc.) needed by the decoder to achieve 1 percent of BLER. LDPC uses the "flooding" SP decoder, and LTE turbo codes use the max*-log-MAP decoder. Note that the performance of max*-log-MAP is very close to the optimal log-MAP decoder. If a "layered" SP decoder is used for LDPC codes, the number of iterations required for the same performance is halved, implying that the complexity is also halved. See [14] for more details.

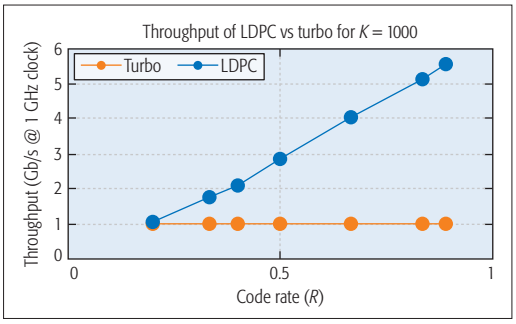

Figure 3. Throughput of LDPC and turbo decoders at different code rates for $K = 1000$.

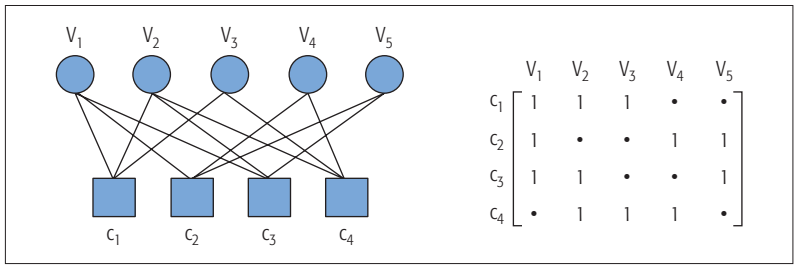

Figure 4. Tanner graph and parity check matrix for a length 5 rate 1/5 LDPC code.

graph and permuting the Z copies of the base edge between the Z copies. In practice, the set of allowed permutations is restricted to facilitate implementation, and a common choice is to restrict to cyclic permutations. The resulting LDPC codes are called quasi-cyclic because the code is invariant under the action of a cyclic group applied to the vectors of lifted bits. In contrast, a cyclic code is a code invariant to cyclic shifts of the entire code. The cyclic lifting structure is the structure that provides for a high degree of organized parallelism in implementation. The 802.11 codes, for example, have base graphs with 24 variable nodes and Z values of 27, 54, and 81. The DVB-S2 standard uses a Z of 360. The 5G NR codes use maximum Z of 384, but many smaller values are also supported.

For a base graph, we use the notation k_b and n_b to represent the corresponding base graph quantities so that if the lifting value is Z, we nominally have $K = Z\,k_b$ and $N = Z\,n_b$. We say "nominally" here because we may allow some sub-column shortening and puncturing so that more generally we have $Z\,(k_{b-1}) < K \le k_b$ and $Z\,(n_{b-2}) < N \le n_b$.

Meeting NR Block Length and Rate Requirements

The channel coding solution for NR needs to be rate-compatible to support IR-HARQ. The larger transmission frames needed for evolved mobile broadband (eMBB) warrant the use of large blocklengths for performance gains. On the other hand, single resource block allocations may lead to code blocklengths on the order of a few hundred coded bits. Thus, information blocklength sizes ranging from around 100 to 8000 need to be supported. A coding rate of 8/9 was chosen to meet the peak decoder throughput of 5 Gb/s, while a rate of 1/5 is needed to achieve extended coverage and meet the 100 Mb/s edge cell throughput goal.

Supporting a wide range of code rates naturally aligns with the need to support IR-HARQ. For example, a code with information blocklength $K = 800$ and rate 8/9, so $N = 900$, may need an IR-HARQ extension to $N = 2400$. That extension provides for a rate 1/3 code. Very often intermediate IR-HARQ extensions are also required, so the code should have good performance for rates from 8/9 down to 1/3. A goal in the LDPC design is to have good performance across IR-HARQ extensions so that the design of the highest rate codes together with their IR-HARQ extensions inherently provide good codes for all desired rates.

The 5G NR LDPC codes are, strictly speaking, a concatenation of an LDPC code and a low-density generator matrix (LDGM) code. The structure begins with a relatively high rate "core"; this is the LDPC part. The base graph for the core has a small number (e.g., $m_{\text{core}} = 4$) of parity checks and some number (e.g., $k_{b\max} = 22$) of information variable nodes and m_{core} parity variable nodes. All additional variable nodes are extension degree one variable nodes each connected to a unique check node whose other variable node neighbors are taken from the core; this is the LDGM part. In general, the degree one variable nodes are the extension nodes used for IR-HARQ but the first degree one variable node is special for reasons discussed below, and is included in all code rates. For structural reasons that we describe shortly, the core portion of the graph does not perform very well without including at least some of the first degree one parity bits.

One aim of the 5G design was to have a compact description for many code blocklengths, and one important enabling idea is to have many Z liftings defined for a single base graph. Associated to each edge in the base graph is a descriptor so that given a target lifting value of Z, the associated lifting value for the edge is derivable from the descriptor. There are many ways to accomplish this, but we limit ourselves to a description of the method selected in 3GPP.

The set of lifting sizes supported for the graph are all values of the form $Z = A \times 2^j$ for $A = \{2, 3, 5, 7, 9, 11, 13, 15\}$ and $j = 0,1,\ldots$ and includes all such Z in the range from 2 to 384. This includes $Z = 2j \times \{2, 3, 5, 7, 9, 11, 13, 15\}$ for $j = 0, 1, 2, 3, 4, 5$ excluding those values exceeding $Z = 384$. Thus, the full lifting set Z is given by {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 40, 44, 48, 52, 56, 60, 64, 72, 80, 88, 96, 104, 112, 120, 128, 144, 160, 176, 192, 208, 224, 240, 256, 288, 320, 352, 384}. It is not clear that all lifting values, especially the small ones, will be used, but all are defined. The lifting values are organized into eight sets, one for each value of A. For each A, there is maximal $Z(A) = A2^j$ for some maximal $j = j(A)$. For each edge E in the base graph and each A, an integer value $L(A, E)$ in the range $[0, Z(A) - 1]$ is specified. Then, for a target lifting size $Z = A2^j$, the lifting value associated with the edge E is given by $L(A, E)$ modulo Z. Thus, the entire set of liftings can be represented by five 9-bit values and three 8-bit values (a 69-bit descriptor) per edge.

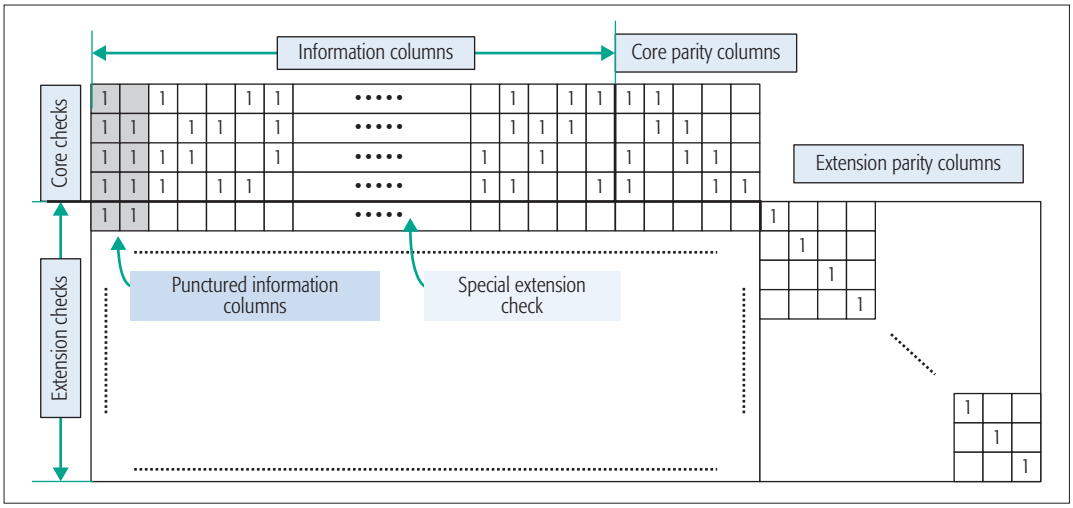

Figure 5. Sketch of base parity check structure for the 5G NR LDPC code.

A first step in LDPC design is to choose some macroscopic statistical properties of the Tanner graph structure. This part of the design process typically uses tools from asymptotic analysis under SP decoding such as density evolution (DE) and its approximations known as EXIT charts.

Multiple Base Graphs

Although the above described eight PCMs would suffice to meet the requirements of 5G NR, there is room for improvement in latency and throughput at smaller blocklengths by introducing smaller base graphs with a smaller number of variable nodes. For a given blocklength, a relatively smaller base graph size implies a relatively larger Z and hence more parallelism in decoding operations. More parallelism in decoding implies lower latency and higher throughput. To achieve high code rate, however, the base graph cannot be too small. DE analysis predicts that smaller base graphs can perform well, with small gaps in capacity, although with reduced maximum code rate. Accordingly, 3GPP has agreed to consider two rate-compatible base graphs, BG1 and BG2, for the channel coding. BG1 is targeted for larger blocklengths ($500 \le K \le 8448$) and higher rates ($1/3 \le r \le 8/9$), whereas BG2 is targeted for smaller blocklengths ($40 \le K \le 2560$) and lower rates ($1/5 \le r \le 2/3$). The range of k_b is [19, 22] and [6, 10] for BG1 and BG2, respectively. The maximum Z for BG2 is limited to 256; extending this to 384 would make better use of available hardware and is under consideration. More generally, a wider range of supported k_b values would allow a wider range of code rates and block sizes to utilize full hardware parallelism, but this idea was not pursued in 3GPP.

Base Graph Shortening for Blocklength Granularity

Base graphs BG1 and BG2 have similar structure. We focus our description on BG1, which is the main 5G NR high-rate base graph. Base graph BG1 has 22 information bit columns. Using sub-column shortening, as discussed above, we see that any information block length in the range $21*Z$ to $22*Z$ for any of the above Z values can be supported. Even with the large set of Z indicated above, however, this does not cover the entire desired range. To complete the coverage, additional shortening at the base graph level is introduced. Note that for Z greater than 8, the largest ratio of a value of Z to the next smallest Z is 9/8. Thus, for a sufficiently large target information blocklength K, if Z is the smallest supported Z such that $K \le 22*Z$, the code with information blocklength $22*Z$ can be shortened to information blocklength K by shortening at most $(1/9)*22*Z = 2.44\ Z$ bits. Good performance can be expected if the base graph is designed to give good performance not just for $k_b = k_{bmax} = 22$ but also for $k_b = 19,20,21$. This kind of design flexibility distinguishes the 5G NR LDPC from previous designs.

The specific base graph structure adopted for 5G NR in 3GPP is illustrated in Fig. 5 as a sketch of the BG1 PCM where a 1 in the matrix indicates the presence of a base edge. The columns are broadly partitioned into three parts: information columns, core parity columns, and extension degree one parity-bit columns. In a lifted graph with lifting size Z, each column corresponds to Z variable nodes. In that case each 1 entry is replaced by an integer modulo Z to create a lifted PCM representation of the quasi-cyclic LDPC code. The rows are broadly partitioned into two parts: core check rows and extension check rows. The bits in the core parity columns are solved so that the core parity checks are satisfied. Each degree one parity bit is then simply a parity of a subset of core variable nodes.

As described above, the total number of base information columns is k_{bmax}. Shortening at the base graph level is achieved by eliminating the right-most systematic columns so that the k_b of the base graph in 5G NR is an adjustable parameter. The design of both the core and the IR-HARQ portion should be done so as to provide good performance. In the next section we describe some of the design details that enable this flexible design to yield very good code performance across the full range of block lengths and code rates.

5GNR LDPC Design Details

As mentioned above, a first step in LDPC design is to choose some macroscopic statistical properties of the Tanner graph structure. This part of the design process typically uses tools from asymptotic analysis under SP decoding such as density evolution (DE) and its approximations known as EXIT charts. The asymptotic analysis amounts to evaluating performance of an arbitrarily large graph instance that realizes the desired macroscopic statistics. It is a consequence of the law of large numbers that the asymptotic performance eval-

By introducing a punctured variable node into the base graph, an additional base parity check can be added without changing the code rate of the base structure. That increases the maximum available degree, but, it turns out, also significantly improves the performance that can be achieved by small degrees.

uation is essentially deterministic and, using the above-mentioned tools, relatively easy to compute. In this section, we discuss some features of the base graph design that have been incorporated into the agreed NR LDPC and emerged from this design process.

Structure of the Core Graph

One feature present in the 5G NR LDPC codes not found in other wireless standards (although it has been used in proprietary solutions and proposed in other standards) is punctured variable nodes, also known as state variable nodes. It has been found that the introduction of relatively high degree punctured variable nodes can produce improved performance at lower complexity. There are several theoretical results, some of which we discuss below, that help to explain this improvement. Another advantage offered by punctured nodes in practical settings is their effect on the number of base check nodes. Traditional irregular LDPC designs, such as those used in 802.11, require relatively high degree variable nodes to achieve good performance. To facilitate implementation, however, it is desirable that base graphs should not have multiple edges between a base variable and check node. With very small base graphs having very few base parity checks, this imposes a strong limit on the maximum variable node degree and hence strongly limits performance. By introducing a punctured variable node into the base graph, an additional base parity check can be added without changing the code rate of the base structure. That increases the maximum available degree but, it turns out, also significantly improves the performance that can be achieved by small degrees.

LDPC was invented by Robert Gallager in his 1961 Ph.D. thesis [8]. One result in Gallager's thesis was a proof that LDPC codes cannot reach Shannon capacity without the degrees of the check nodes becoming arbitrarily large. The entropy of the syndrome (the parity checks) needs to be maximal to reach capacity, and it cannot be maximal with bounded degree. Recently [7], it was shown that capacity could be achieved on the binary erasure channel using a Tanner graph with bounded degree. This is not a contradiction because Gallager assumed that the PCM was for the transmitted code. If one punctures (does not transmit) some of the bits associated with the variable nodes, the Tanner graph no longer corresponds to a PCM of the transmitted code. To obtain a PCM for the transmitted code, one needs to eliminate the punctured variables from the punctured node, including PCM representation. The elimination process increases the density (the number of 1s) in the matrix, and Gallager's obstacle to capacity is thereby circumvented.

The core portion of the 5G NR code uses variable nodes of small degree, the largest being four, thus strongly limiting the degree of the punctured node if no double edges are allowed. It was found that by adding additional punctured nodes, significant further improvements could be obtained. The additional punctured node did not, however, truly result in an additional parity check because to achieve good performance it was also found that a low degree parity bit involving the two punctured nodes was needed. Thus, the 5G NR LDPC design incorporates two punctured information columns, and the very first extension degree one parity bit is a special low degree parity connected to the punctured nodes that should not be punctured for any code rate. The UMB design [15] has three state nodes and does not have this special parity bit. In some cases, to achieve high code rate it may be preferable to puncture the core parity bits rather than puncture this special degree one parity bit.

Core Information Columns: The base core information bits other than the two puncture variable nodes discussed above will typically have low core degree. In BG1 they are all degree 3, and in BG2 some degree 2s have been included. Generally, with all such nodes set to degree 3, good performance can be achieved for a large range of the number of information columns. Thus, this structure is good for designing an extensible and flexible LDPC coding system such as required for 5G NR.

Extension Structure

Given a core graph constructed with the features mentioned above, the next task in code design is to optimize the connectivity of the extension HARQ parity checks. Ultimately, the design is a joint problem since good performance across all rates and block sizes is desired. The core graph design structure was chosen because of its ability to support good and flexible extensions, and good performance is obtained with subsequent extension optimization.

Since each extension bit can potentially be used as the last bit in the base code, the extension bits are optimized incrementally, each one conditioned on the design selection of the preceding extension bits. The main tools used for making the selection are the asymptotic performance characterizations described earlier. One challenge in the 5G NR design is to ensure good performance not only for the full base graph but also for its shortened versions.

Quasi and Full Row Orthogonality: In a typical hardware implementation of a quasi-cyclic LDPC decoder [9], there are Z processors, where Z corresponds to the maximum lift value. These Z processors perform variable node and check node update operations. More precisely, in each clock cycle, the Z processors are working on one layer (row in the base graph) and processing Z edges in parallel. The decoder performs Z variable node updates and at the same time performs Z check node updates of the next layer. If the edge connectivity of the base graph is such that variable nodes are connected to two consecutive layers, there is potential for a negative impact to decoder performance. Indeed, when the second such layer is processed, the updated variable node operation may not yet be available, so the potential gain from the previous layer processing is not available to and does not benefit the performance of the current layer.

This can be circumvented by introducing additional delay in the update process, but that results in a slowing down of the decoder, also potentially degrading performance. This problem can be avoided if the consecutive layers are orthogonal, meaning they have no variable node neighbors in common. Typically, this is not feasi-

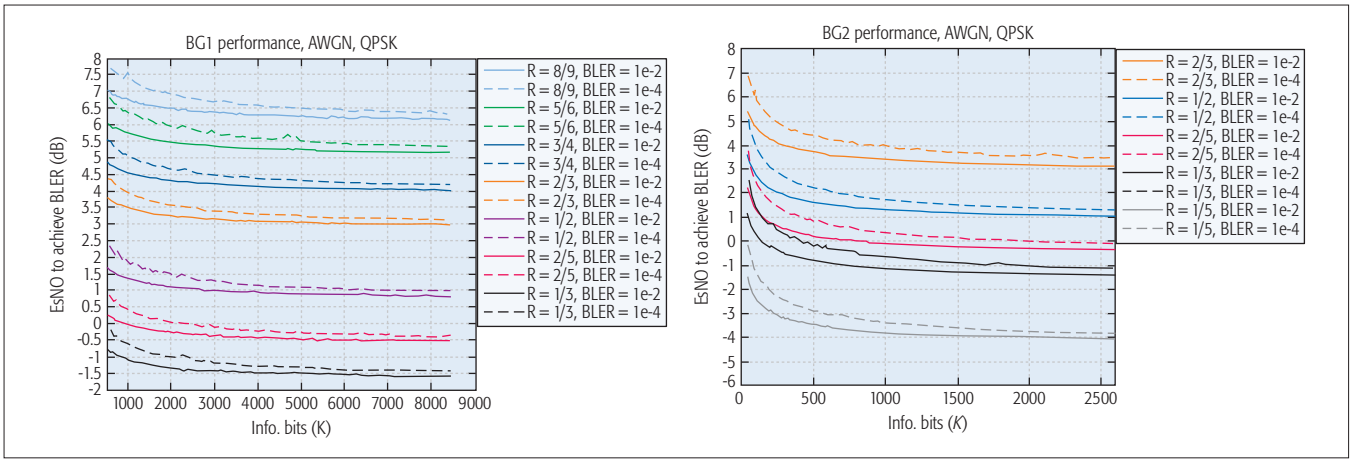

Figure 6. Robust performance of BG1 (left) and BG2 (right) with no error-floors till BLER of 1e-4 across finely granular blocklength across all rates.

ble for the core graph since there are very few layers and any further restriction on their connectivity would lead to performance degradation. Also, typically for the first few HARQ parity layers, corresponding to relatively high coding rates, the DE optimization dictates that both the punctured nodes are connected to each such layer. However, for these layers, imposing orthogonality on the remaining variable nodes results in negligible ideal performance degradation while enhancing practical performance. This construction is termed *quasi-row orthogonality*. When more layers are added, resulting in mid-to-low rates, the density of added edges gradually decreases and the orthogonality constraint does not hamper the optimization. We further observe that deeper layers are typically not connected to both the punctured nodes, and the design has nearly balanced connectivity of the two punctured nodes, meaning that they each connect to approximately the same number of layers. Thus, in this case it is possible to achieve *full orthogonality* for such layers, and the 5G NR codes have adopted this feature.

Lifting Design

Given the base graph, the next step in the design process is to determine the lift values. The structure for lift representation described above creates dependencies where large lifts contain the smaller ones. These dependencies present some challenges to designing a good set of lift values.

In general, the lift values are designed to avoid certain structures in the full Tanner graph that lead to bad performance. Loops in the graph are the fundamental reason for the sub-optimality of SP decoding, and the elimination of certain loops in the graph is often used as a heuristic for graph lifting optimization. The effects of loops in the graph and their interaction with graph structure are a complicated topic [11] that is beyond the scope of this article.

To give a sense of the dependencies in the lifting design, consider liftings for two sizes, Z and $2Z$. Given the 5G NR lifting definition, the liftings for the Z-lifted graph are those for the $2Z$-lifted graph taken modulo Z. Equivalently, if L is a lift value for the Z-lifted graph, the lifted value for the $2Z$-lifted graph is either L or $L + Z$. If a good design has been achieved for the Z-lifted graph, it then follows that loops in the $2Z$-lifted graph are a subset (with suitable matching interpretation) of those in the Z-lifted graph. This property ensures that good design for the Z-lifted graph is inherited in the $2Z$-lifted graph, and the decision whether to add the additional Z to L can be made to optimize the $2Z$-lifted graph without affecting the Z-lifted design.

Compared with the UMB design [15], the requirements of peak rate of 8/9 as opposed to 3/4, finely granular blocklength, and compact description led to a different design in 5G. More specifically, [15] has six different base graphs, one for each k_b = 6, 7, 8, 9, 10, 11, and the lifting values used were Z = {16, 32, 64, 128, 256}. HARQ extensions in [15] used degree two variable node splitting along with degree one extension.

Performance Results

The performance of LDPC codes generated using the design philosophy described in this article and finally adopted in 3GPP [12, 13] is shown in Fig. 6. For both BG1 and BG2, the information blocklength K is sampled in steps of 16 bits from 500 to 8192 bits and from 100 to 2560 bits, respectively. The decoder used is a floating-point SP decoder with 50 flooding iterations. It is observed that the solution has robust, smooth, and consistent performance without error floor down to at least BLER of 1e-4 across all blocklengths.

Conclusions

We describe in this article LDPC codes as a new coding scheme for NR that can address all the requirements of NR including high performance, IR HARQ capability, high parallelism, low decoding latency, large range of rate and blocklengths, fine blocklength granularity, and compact description.

References

[1] 3GPP TSG RAN Meeting #71, RP-160671, "New SID Proposal: Study on New Radio Access Technology," NTT DOCOMO Inc., Göteborg, Sweden, 7–10 Mar. 2016.
[2] R. Tanner, "A Recursive Approach to Low Complexity Codes," *IEEE Trans. Info. Theory*, vol. 27, no. 5, Sept. 1981, pp. 533–47.
[3] F. Kschischang, B. Frey, and H-A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," *IEEE Trans. Info. Theory*, vol. 47, no. 2, Feb. 2001, pp. 498–519.

[4] T. Richardson and R. Urbanke, "Multi-Edge Type LDPC Codes," Wksp. Honoring Prof. Bob McEliece on His 60th Birthday, 2002.
[5] D. Divsalar *et al.*, "Capacity-Approaching Protograph Codes," *IEEE JSAC*, vol. 27, no. 6, Aug. 2009, pp. 876–88.
[6] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of Capacity-Achieving Irregular Low-Density Parity-Check Codes," *IEEE Trans. Info. Theory*, vol. 47, no. 2, Feb. 2001, pp. 619–37.
[7] H. Pfister, I. Sason, and R. Urbanke, "Capacity-Achieving Ensembles for the Binary Erasure Channel with Bounded Complexity," *IEEE Trans. Info. Theory*, vol. 51, no. 7, July 2005, pp. 2352–79.
[8] R. Gallager, *Low-Density Parity-Check Codes*, thesis, MIT Press, 1963.
[9] D. Hocevar, "A Reduced Complexity Decoder Architecture via Layered Decoding of LDPC Codes," *Proc. IEEE Wksp. Sig. Processing Sys.*, Austin, TX, Oct. 2004, pp. 107–12.
[10] IEEE 802.11n-D2.0, "IEEE 802.11n Wireless LAN Medium Access Control MAC and Physical Layer PHY specifications," 2007.
[11] T. Richardson, "Error Floor of LDPC Codes," *Proc. 41st Annual Allerton Conference on Communication, Control, Computation*, Urbana-Champaign, IL, 2003, pp. 1426–35.
[12] 3GPP TSG RAN WG1 NR Ad-Hoc # 2, R1-1711545, "Summary of [89-24] Email Discussion on LDPC Code Base Graph #1 for NR," Nokia, Alcatel-Lucent Shanghai Bell, Qingdao, China, 27–30 June 2017.
[13] 3GPP TSG RAN WG1 NR Ad-Hoc # 2, R1-1711546, "Summary of [89-25] Email Discussion on LDPC code Base Graph #2 for NR," Nokia, Alcatel-Lucent Shanghai Bell, Qingdao, China, 27–30 June 2017.
[14] 3GPP TSG RAN WG1 NR #85, R1-164704, "Channel Coding Evaluations Assumptions — Performance and Complexity," Qualcomm Inc., Nanjing, China, 23–27 May 2016.
[15] 3GPP2, "C.S0084-001-0 v3.0, Physical Layer for Ultra Mobile Broadband (UMB) Air Interface Specification," Aug. 2008.

Biographies

Tom Richardson [S'01, SM'02, F'07] is VP, Engineering at Qualcomm, New Jersey. Receiving his Ph.D. in 1990 from MIT, he was a member of the Bell Labs Mathematical Sciences Research Center until 2000. From 2000 to 2006 he was vice president and chief scientist at Flarion Technologies. He was a co-recipient of the 2002 and 2013 Information Theory Society Paper Award, the 2011 Kobayashi award, and the 2014 Hamming Medal and was elected in 2011 to the National Academy of Engineering.

Shrinivas Kudekar obtained his Bachelor's degree in electrical engineering from Indian Institute of Technology Bombay in 2003 and his Ph.D. from the School of Computer and Communication Sciences, EPFL, Switzerland, in 2009. He is presently at Qualcomm Research, New Jersey, and is working on next-generation wireless communication technologies. He was a co-recipient of the 2013 IEEE Information Theory Society Best Paper Award and the 2016 ACM Symposium on Theory of Computing Best Paper Award.