©ISTOCKPHOTO.COM/KENTOH

CHANNEL CODING IN 5G NEW RADIO

Dennis Hui, Sara Sandberg, Yufei Blankenship, Mattias Andersson, and Leefke Grosjean

A Tutorial Overview and Performance Comparison with 4G LTE

Fifth-generation (5G) new radio (NR) holds promise in fulfilling new communication requirements that enable ubiquitous, low-latency, high-speed, and high-reliability connections among mobile devices. Compared to fourth-generation (4G) long-term evolution (LTE), new error-correcting codes have been introduced in 5G NR for both data and control channels. In this article, the specific low-density parity-check (LDPC) codes and polar codes adopted by the 5G NR standard are described. The purpose of each key component in these codes and the associated operations are explained. Performance and implementation advantages of these new codes are compared with those of 4G LTE.

Error-Correcting Codes in 5G NR

The advent of the 5G mobile communication networks augurs a new era of ubiquitous connections among new types of mobile devices including self-driving vehicles, unmanned aerial vehicles, and autonomous robots. At the heart of the 5G evolution over today's 4G networks, there is a set of new communication requirements beyond mere data-rate improvements that includes low latency, low power consumption, and high reliability. The foundation of 5G networks that enables the fulfillment of these new requirements is 5G NR.

The latest release (Release 15) of the cellular standard in the 3rd Generation Partnership Project [1] have formally announced the arrival of the 5G NR air interface. Compared to 4G LTE [2], a prominent feature of 5G NR is a pair of new error-correcting channel codes adopted, respectively, for data channels and control channels. Specifically, LDPC codes replaced turbo codes for data channels, and polar codes replaced tail-biting convolutional codes (TBCCs) for control channels. In this tutorial article, the specific LDPC codes and polar codes adopted by the 5G NR standard are described. The purpose of each key component in these codes and the associated operations are explained. The performance and implementation advantages of these new codes are compared with those of 4G LTE.

Digital Object Identifier 10.1109/MVT.2018.2867640
Date of publication: 1 October 2018

NR Data Channel

Introduction to NR LDPC Coding

LDPC codes were first introduced by Gallager in his doctoral thesis [3] but were then largely forgotten. They were rediscovered in the mid-1990s by several authors [4], [5]. LDPC codes have attracted a lot of interest since then, including the incorporation in several IEEE standards such as IEEE 802.16e, IEEE 802.11n, and so on. When 4G LTE was standardized approximately ten years ago, LDPC codes were also promising candidates, though ultimately turbo codes with new quadratic permutation polynomial interleavers were adopted instead.

The main advantages of 5G NR LDPC codes compared to turbo codes used in 4G LTE are

- better area throughput efficiency (e.g., measured in Gb/s/mm^2) and substantially higher achievable peak throughput
- reduced decoding complexity and improved decoding latency (especially when operating at high code rates) due to higher degree of parallelization
- improved performance, with error floors around or below the block error rate (BLER) 10^{-5} for all code sizes and code rates.

These advantages make NR LDPC codes suitable for the very high throughputs and ultrareliable low-latency communication targeted with 5G, where the targeted peak data rate is 20 Gb/s for downlink and 10 Gb/s for uplink.

Structure of NR LDPC Codes

The NR LDPC coding chain includes code block segmentation, cyclic-redundancy-check (CRC) attachment, LDPC encoding, rate matching, and systematic-bit-priority interleaving (see Figure 1). Specifically, code block segmentation allows very large transport blocks to be split into

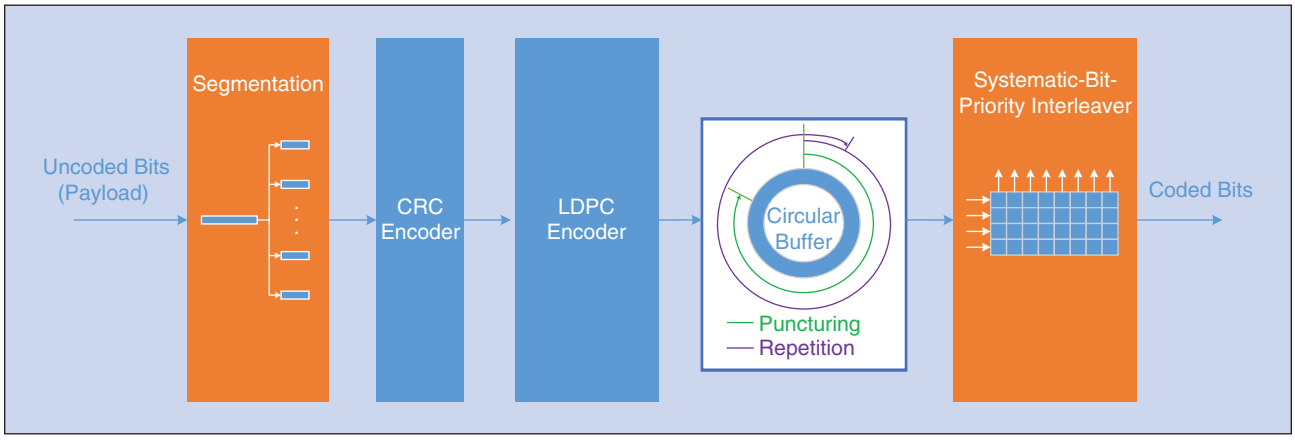


Figure 1 The NR LDPC coding chain.

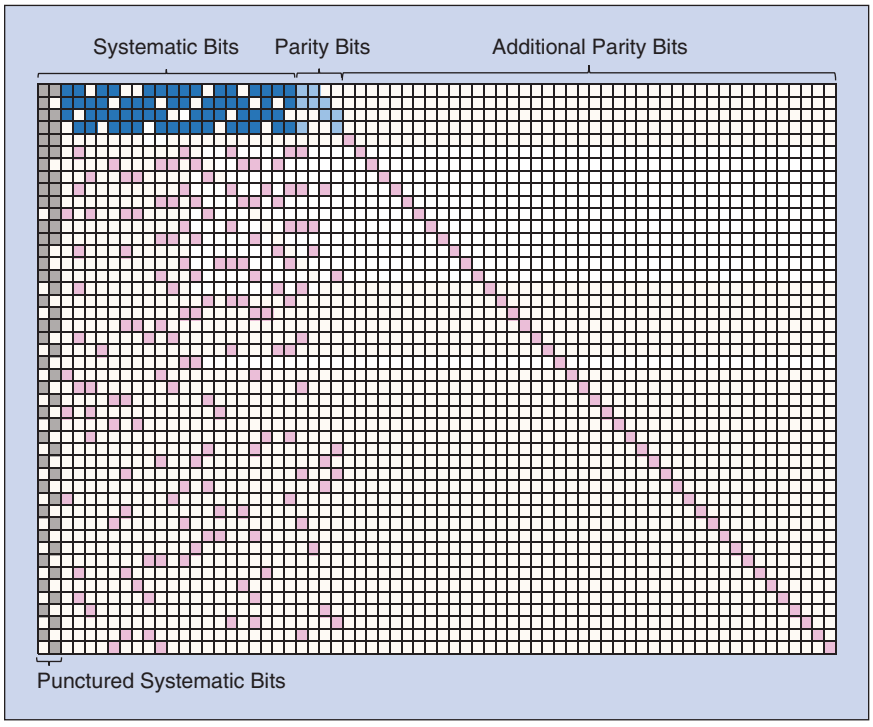


FIGURE 2 The structure of NR LDPC base matrix 1. Each square corresponds to one element in the base matrix or a $Z \times Z$ subblock in the PCM.

multiple smaller-sized code blocks that can be efficiently processed by the LDPC encoder/decoder. The CRC bits are then attached for error detection purposes. Combined with the built-in error detection of the LDPC codes through the parity-check (PC) equations, very low probability of undetected errors can be achieved. The rectangular interleaver with number of rows equal to the quadrature amplitude modulation (QAM) order improves performance by making systematic bits more reliable than parity bits for the initial transmission of the code blocks.

NR LDPC codes use a quasi-cyclic structure, where the parity-check matrix (PCM) is defined by a smaller base matrix [6]. Each entry of the base matrix represents either a $Z \times Z$ zero matrix or a shifted $Z \times Z$ identity matrix, where a cyclic shift (given by a shift coefficient) to the right of each row is applied.

TABLE 1 *NR LDPC base matrix parameters.*

Parameter	*Base Matrix 1*	*Base Matrix 2*
Minimum design code rate	1/3	1/5
Base matrix size	46 × 68	42 × 52
Number of systematic columns	22	10
Maximum information block size K	8,448 (= 22 × 384)	3,840 (= 10 × 384)
Number of nonzero elements	316	197

NR LDPC base matrix 1 is shown in Figure 2. The colored squares all correspond to shifted $Z \times Z$ identity matrices, while the white squares correspond to $Z \times Z$ zero matrices. The first two columns in the base matrix (gray), i.e., the $2 \times Z$ first columns in the PCM, correspond to punctured systematic bits that are never transmitted. Unlike the LDPC codes implemented in other wireless technologies, NR LDPC codes have a rate-compatible structure, similar to the codes proposed in [7]. Codewords of different rates can be generated by including a different number of parity bits, or equivalently, by using a smaller subset of the full PCM. This is especially useful for communication systems employing enhanced hybrid automatic repeat requests, allowing for the use of incremental redundancy instead of Chase combining for retransmissions. Another advantage of this structure is that for higher rates, the PCM, and, hence, the decoding complexity and latency, is smaller. This contrasts with the LTE turbo codes, which have constant decoding complexity and latency irrespective of the code rate.

Two Base Matrices

The NR data channel supports two base matrices to ensure that good performance and decoding latency is achieved for the full range of code rates and information block sizes in NR. The parameters of base matrix 1 and base matrix 2 are given in Table 1.

Base matrix 1, shown in Figure 2, is optimized for large information block sizes and high code rates. It is designed for a high maximum code rate of 8/9 and may be used for code rates up to $R = 0.95$. Base matrix 2 is optimized for small information block sizes and lower code rates than base matrix 1. The lowest code rate for base matrix 2 without using repetition is 1/5. This is significantly lower than for LTE turbo codes, which rely on repetition for code rates below 1/3. The NR LDPC codes can thereby also achieve an additional coding gain at low code rates, which makes them suitable for cases that require high reliability.

For each base matrix, 51 PCMs are defined, each corresponding to a subblock size Z between two and 384, thus resulting in different ranges of information block size. Specifying PCMs for many different block sizes allows the code performance to be optimized in each case. In total, there are 102 PCMs defined for the NR data

channel. For comparison, we note that IEEE 802.11n only specifies 12 PCMs with four different code rates and three different information block sizes.

The parameters for the two base matrices shown in Table 1 show a significant overlap both in information block size K (in bits) and in code rate R where both base matrix 1 and base matrix 2 may be used. For K and R in this region, the rules for selecting a base matrix based on information block size K and code rate R depend on the performance of the specific LDPC codes. In general, the base matrix with the best performance for a certain range of K and R is used.

From a decoding complexity point of view, for a given K value, it is beneficial to use base matrix 2, since it is more compact and utilizes a larger shift size Z (i.e., more parallelism) relative to base matrix 1. Typically, the decoding latency is proportional to the number of nonzero elements in the base matrix. Since base matrix 2 has much fewer nonzero elements than base matrix 1 for a given code rate (e.g., base matrix 2 contains 38% as many nonzero elements as base matrix 1 for code rate 1/3), its decoding latency is significantly lower.

Figure 3 shows the regions of code rate and information block size K, for which base matrix 1 and base matrix 2 are used. In general, base matrix 2 is used for low code rates, and base matrix 1 is used for high code rates. The exact switching point between the two base matrices also depends on the information block size, due to performance differences between the two base matrices in the different regions. For $K \leq 308$, only base matrix 2 is used since base matrix 2 performs better at all code rates in this range. For $308 \leq K \leq 3{,}840$, base matrix 2 is used for code rates up to 2/3 since base matrix 2 is optimized for $1/5 \leq R \leq 2/3$. Puncturing can be used to achieve higher code rates than 2/3 for base matrix 2, but base matrix 1 performs better in this range. For $K > 3{,}840$ and $R \leq 1/4$, base matrix 2 is used because of its superior performance over base matrix 1, which requires the use of repetition to reach any code rate lower than 1/3. The switch is not at a rate of 1/3 because code block segmentation results in fewer code blocks when base matrix 1 is used as opposed to base matrix 2, which offsets the additional coding gain from base matrix 2 for rates between 1/4 and 1/3 when $K > 3{,}840$.

Performance of NR LDPC Codes

The performance of the NR LDPC codes over an additive white Gaussian noise channel has been evaluated using a normalized min-sum decoder, layered scheduling, and a maximum of 20 decoder iterations.

Figure 4 shows the signal-to-noise ratio (SNR) required to achieve certain BLER targets as a function of information block size K for code rate 1/2 and quaternary phase-shift keying (QPSK) modulation. The results show that the NR LDPC codes provide consistently good performance

THERE ARE THREE METHODS OF ADJUSTING THE CODE LENGTH IN NR POLAR CODING, SPECIFICALLY, PUNCTURING, SHORTENING, AND REPETITION.

over a large range of block sizes. For this code rate, base matrix 2 is used for all K for which it is defined, i.e., for $K \leq 3{,}840$, while base matrix 1 is used for larger values of K. As shown in the figure, there is a small gap in performance at the block size where the transition occurs between base matrix 1 and base matrix 2.

Figure 5 shows a comparison between the performance of the 5G NR LDPC codes and the 4G LTE turbo codes for $K = 6{,}144$ and typical combinations of code rate and modulation order. The LDPC encoder uses base matrix 2 for a code rate of 1/5, which means that the transport block must be segmented into two code blocks. Base matrix 1 is used for the other code rates. The performance of the LTE turbo codes has been evaluated using a max-log-map decoder with extrinsic log-likelihood ratio scaling by a factor of 0.75 and a maximum of eight

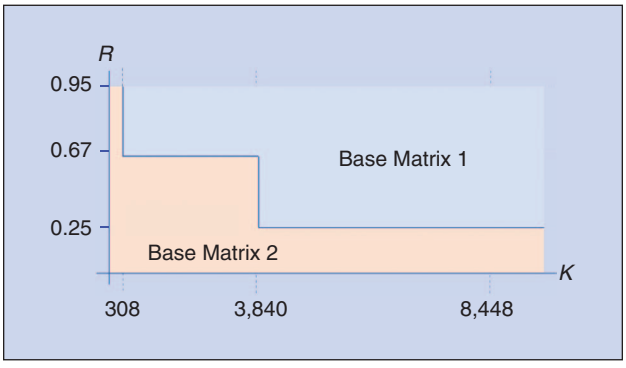


FIGURE 3 The usage of the two base matrices specified for the NR data channel. For K larger than the maximum information block size, code block segmentation is applied.

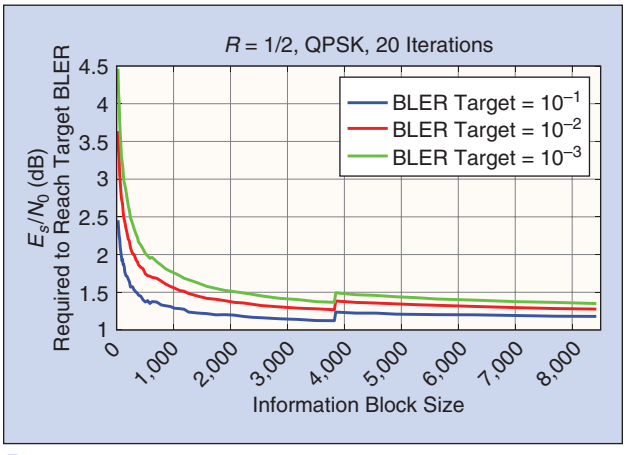


FIGURE 4 The performance of NR LDPC codes at code rate 1/2 for QPSK modulation.

decoder iterations. The NR LDPC codes and the LTE turbo codes have similar performance, except for high code rates where the LTE turbo codes have a tendency of an error floor. For example, in Figure 5, the LTE turbo code is 0.6 dB worse than the NR LDPC code at BLER $= 10^{-4}$ because of an error floor at code rate $R = 5/6$. The error floor becomes higher at higher code rates. See [12] for a thorough evaluation of the LTE turbo codes at different information block sizes and code rates.

NR Control Channel

Introduction to Polar Coding

Polar codes, originally proposed by Arıkan in [8], are the first class of constructive channel codes that was proven to achieve the symmetric (Shannon) capacity of a binary-input discrete memoryless channel using a low-complexity decoder, particularly, a successive cancelation (SC) decoder. The main idea of polar coding [8] is to transform a pair of identical binary-input channels into two distinct channels of different qualities: one better and one worse than the original binary-input channel. By repeating such a pairwise polarizing operation on a set of $N = 2^n$ independent uses of a binary-input channel, a set of 2^n bit-channels of varying qualities, in terms of the reliability in communicating a single bit, can be obtained for any given integer n. When n is large, some of these bit channels are nearly perfect (i.e., error-free) while the rest of them are nearly useless (i.e., totally noisy). The idea is to use the nearly perfect channels to transmit data to the receiver while setting the input to the useless channels to have fixed or frozen values (e.g., zero) that are known to the receiver. The input bits to the nearly useless and the nearly perfect bit channels are commonly referred to as *frozen* bits and *nonfrozen* (or *information*) bits, respectively. Only the nonfrozen bits are used to carry information. The choice of the set of nonfrozen bit locations, referred to as the *information set*, directly affects the performance of a polar code.

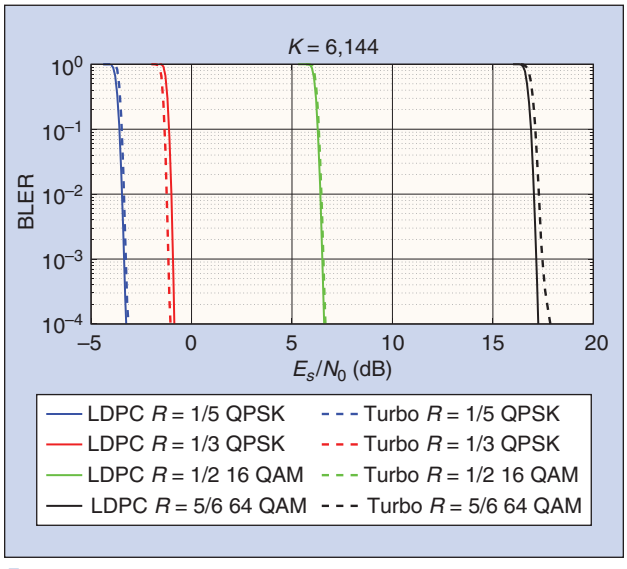


FIGURE 5 A performance comparison between NR LDPC codes (solid) and LTE turbo codes (dashed).

In practice, the relative reliabilities of the bit-channels may be sorted in an ascending order and be captured in a sequence of indices termed an *information sequence* that indicates the order with which the bit-channels should be used to carry data. For instance, the bit-channels corresponding to the last K indices in the information sequence should be used for carrying a given set of K data bits. Puncturing (or shortening) of coded bits, however, can make some bit-channels, whose locations depend on the puncturing (or shortening) pattern, highly unreliable (i.e., incapable of information carrying) and, thus, can alter the order with which the bit-channels should be used.

Despite their capacity-achieving capability, polar codes of finite size, when used alone, often exhibit a poor minimum distance property and do not perform well at a high SNR regime even with exhaustive-search maximum likelihood decoding [9]. Concatenation with an outer code, such as a CRC code, together with the use of a successive cancelation list (SCL) decoder, is needed to achieve competitive performance.

The main advantage of polar codes over the TBCCs and turbo codes used in 4G LTE, as well as the NR LDPC codes, is that polar codes with SCL and CRC outer codes typically yield better performance at moderate payload sizes on the order of $K = 250$ b or less, which is sufficient for typical control information. However, the NR polar code is significantly more complex than the TBCCs used in 4G LTE control channels.

Structure of NR Polar Codes

Figure 6 shows a complete coding chain for the NR polar codes. Some of the components in the chain applies only to either uplink or downlink. The core components that apply to both uplink and downlink are the CRC encoder, polar encoding kernel, and rate matcher.

CRC encoding is performed as is traditionally done. However, in polar coding, the purpose of the CRC bits is twofold. First, they are used for error detection to maintain a false-detection or false-alarm rate (FAR) below a certain threshold $P_{\text{FAR}} \approx 2^{-n_{\text{FAR}}}$. Second, they are also used for error correction at the end of SCL decoding with a target list size $L = 2^{n_L}$, to eliminate decoding paths in the list that fail CRC. The CRC length is set to $n_{\text{CRC}} = n_{\text{FAR}} + n_L$. For both uplink and downlink control channel, the target list size of eight (i.e., $n_L = 3$) is implicitly assumed in the specifications of 5G NR [1].

Basic polar encoding is performed in the polar encoding kernel for a given mother code with a size of $N = 2^n$ and a chosen information set in the same way as was

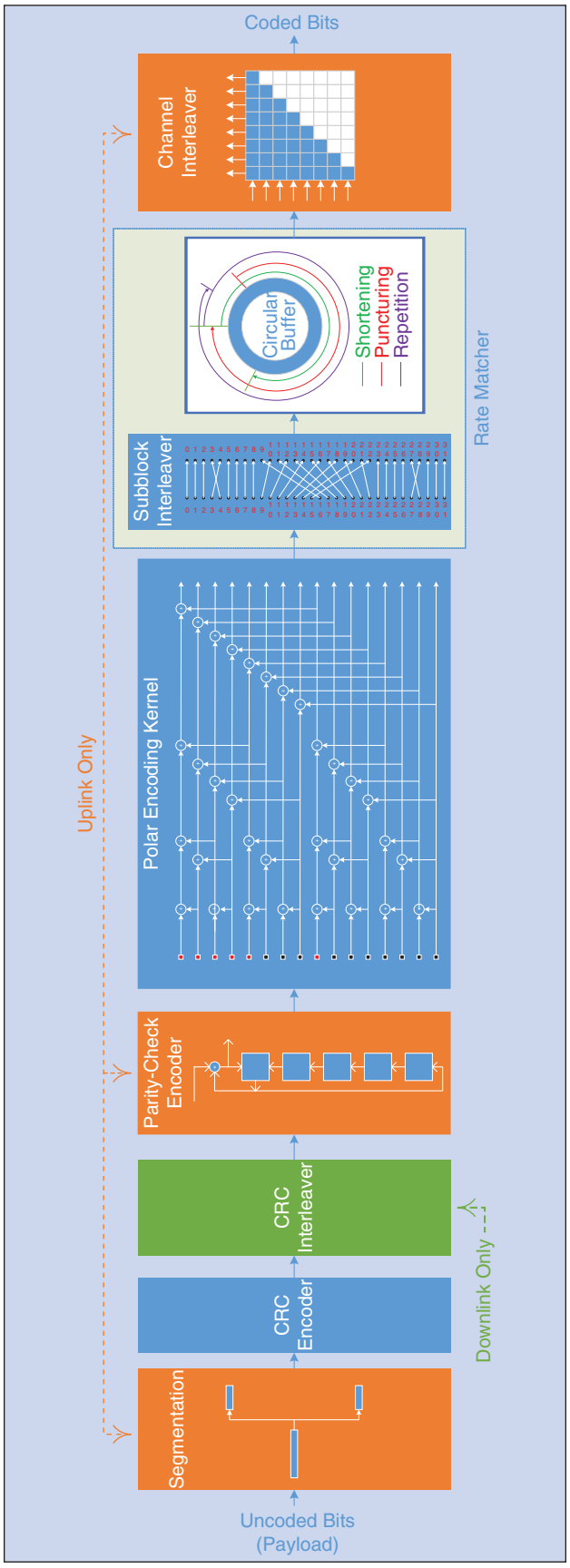


FIGURE 6 The NR polar coding chain.

proposed in [8], except without the bit-reversal permutation. In other words, the output of the NR polar encoding kernel is simply $\mathbf{x} = \mathbf{G}^{\otimes n}\mathbf{u}$, where $\mathbf{u}$ is the input to the polar encoding kernel, $\mathbf{G}$ is the 2×2 Arıkan kernel matrix, and $\otimes n$ denotes the n-time Kronecker power.

The function of the rate matcher is to adjust the code length M (i.e., the number of coded bits at the output of the rate matcher) to match the amount of available radio resources. There are three methods of adjusting the code length in NR polar coding, specifically, puncturing, shortening, and repetition. Shortening typically performs better at high code rates while puncturing performs better at low code rates. All three methods of code length adjustment are implemented by a size-N subblock interleaver and a circular buffer, as illustrated in Figure 6.

The purpose of the subblock interleaver is to arrange the coded bits in the order with which they are discarded or admitted before placing them into the circular buffer. The interleaver simply permutes the five most significant bits of the binary representation of the indices of the polar encoded bits according to a predetermined integer sequence of length 32, as illustrated in Figure 6. Such a permutation is equivalent to dividing the N coded bits into 32 subblocks, each of size $N/32$, and to permute these subblocks according to the predetermined integer sequence.

Depending on M and N, the coded bits are selected from the circular buffer as described in Table 2, where n_{CRC} is the number of CRC bits and $R_{\text{ps}} = 7/16$ is a threshold for determining whether puncturing or shortening should be used.

Depending on the method of rate matching and the number of discarded coded bits, a corresponding set of bit-channel indices needs to be prefrozen or skipped from the information sequence when selecting the information set. In particular, the indices of those coded bits that are discarded through puncturing or shortening are

TABLE 2 *Bit selection from circular buffer for different rate matching methods.*

Method	*Condition*	*Operation*
None	$M = N$	Extracting all N bits from the circular buffer starting from the first position.
Puncturing	$M < N$ and $\frac{K + n_{\text{CRC}}}{M} \leq R_{\text{ps}}$	Extracting M consecutive bits starting from the $(N - M + 1)$th position of the circular buffer.
Shortening	$M < N$ and $\frac{K + n_{\text{CRC}}}{M} > R_{\text{ps}}$	Extracting M consecutive bits starting from the first position of the circular buffer.
Repetition	$M > N$	Extracting M consecutive bits starting from the first position of the circular buffer and wrap around.

BESIDES PERFORMANCE, THE IMPLEMENTATION COMPLEXITY OF THE DECODER THAT IS REQUIRED TO SUPPORT A DESIRED DATA THROUGHPUT ALSO PLAYED AN IMPORTANT ROLE IN THE SELECTION OF NR CODING SCHEMES.

skipped from the information sequence when determining the information set.

Coding for Downlink Control Information

CRC interleaving is the unique component in the coding chain that is performed in the downlink only. The purpose of the CRC interleaver is to distribute the CRC bits more evenly within the block of information and frozen bits. This contrasts with the conventional approach of clustering all the CRC bits at the end of the information block [9]. The CRC interleaver is designed in such a way that each CRC bit only depends on the previous information bits during SC/SCL decoding so that the CRC bit value can be calculated without waiting for information bits that come after the given CRC bit. Hence, the CRC bits can be used as dynamic frozen bits [10] to assist surviving path selection during SC/SCL decoding. These distributed CRC bits also allow CRC checks to be performed earlier in the decoding processing so that the list decoding process can terminate when all candidate paths in the list fail a CRC check. Consequently, some early termination benefits at the receiver, albeit limited, can be obtained by using the distributed CRC bits. The size of the CRC interleaver is limited to support a maximum downlink control information (DCI) size of 140 b.

Because of the increase in the number of blind decoding attempts of DCI in NR compared to LTE, the CRC for DCI has a total length of $n_{\mathrm{CRC}} = 24$. Discounting the additional CRC length required for a target list size of eight (i.e., $n_L = 3$), this gives an error detection capability of 2^{-21} for DCI in NR, as compared to 2^{-16} of LTE. Among the 24 CRC bits, seven are distributed among the information bits, and 17 are clustered at the end.

Coding for Uplink Control Information

The maximum mother code block size supported by the NR polar code is $N_{\max} = 1,024$. To avoid an excessive use of repetition and to support larger information block size K, segmentation of coded bits into two blocks of equal size is performed when $K \geq 360$ and $M \geq 1,088$ or when $K > 1,013$. The maximum uplink control information (UCI) payload size K supported by polar code is $K_{\max} = (5/6) \times 2,048 = 1,706$, which can only be achieved when segmentation is applied. The minimum UCI size supported by the polar code is $K_{\min} = 12$. For $K < 12$, short block codes in LTE, i.e., repetition, simplex, and Reed–Muller codes, are used instead.

Different numbers of CRC bits for UCI are used as a function of K. For $12 \leq K \leq 19$, only $n_{\mathrm{CRC}} = 6$ CRC bits are used, while for $K > 19$, $n_{\mathrm{CRC}} = 11$ CRC bits are used.

Conventional frozen bits of polar codes typically carry a fixed value (commonly, zero) that is known to the decoder. However, it is possible to improve, albeit slightly in most cases, the error-correcting performance of a polar code by making the frozen bits dependent on the values of the information (i.e., nonfrozen) bits [10]. This kind of data-dependent frozen bits is commonly referred to as dynamic frozen bits, PC frozen bits, or simply PC bits. The addition of these PC bits is equivalent to the addition of another layer of outer code between the CRC code and the polar code.

To facilitate SCL decoding, the value of each PC bit is restricted to depend only on those previously decoded data bits in the successive decoding order, so that the values of these PC bits can be readily determined during successive decoding from previously decoded bits for each hypothesized decoding path in the list.

When K of UCI is small, in the range of $12 \leq K \leq 19$, $n_{\mathrm{PC}} = 3$ PC bits are computed and placed at the input of the polar encoder. Outside this range of K, no PC bit is added to the input of the polar encoder. In a clear majority of cases, the PC bits are placed in the least reliable positions among the $(K + n_{\mathrm{CRC}} + n_{\mathrm{PC}})$ most reliable positions at the input of the polar encoder. The exception is when $M > 195 + K$, in which case, one of the PC bits is moved to the most reliable position among those positions with indices whose binary representations have the smallest weight (i.e., number of ones) among the $(K + n_{\mathrm{CRC}})$ most reliable positions. The value of each PC bit is simply the modulo-2 cumulative addition of every fifth bit value in front of the PC bit, except for the values of other PC bits. The computation of PC bits can be carried out by a length-five cyclic shift register, as illustrated in Figure 6.

Both the channel fading and the difference in reliabilities among individual bits carried by a high-order modulation symbol can cause different coded bits to experience significantly different channel qualities in the transmission of UCI. To mitigate their adverse impact on performance, a channel interleaver is used to attain an averaging effect. In NR polar coding, a triangular interleaver, as illustrated in Figure 6, with a linearly decreasing number of columns in each row is used. Coded bits are written into the memory of this interleaver rowwise and are then read out columnwise. Such an interleaver can significantly improve the performance of the NR polar code when it is used with a high-order modulation, such as 16 QAM.

Performance of NR Polar Codes

In Figure 7, the performance of the NR polar codes in terms of BLER versus E_s/N_0 for 72 payload bits in the

downlink is shown. For comparison, the performance of LTE TBCCs for the same values of K and M is shown, where the LTE TBCC is updated to use 21 CRC bits instead of the 16-b CRC in the LTE DCI. As shown in this figure, the NR polar code outperforms the LTE TBCC at low code rates, and the gap increases as the code rate decreases.

Figure 8 compares the performance of the NR polar codes with that of the LTE turbo codes in the uplink for different rates, modulations, and list sizes of SCL decoders. For a list size of $L = 16$, only the best eight candidate paths from the list are tested by CRC to avoid an increase in P_{FAR}.

As shown in Figure 8, the performance of the NR polar codes with a list size of eight (or more) and the LTE turbo codes are comparable except at high and low code rates. For high code rates, the LTE turbo codes exhibit an error-floor phenomenon with significant performance degradation at low BLER (e.g., around 1% for $R = 5/6$) because of heavy puncturing of the mother code at a rate of 1/3. For low code rates ($R < 1/3$), the NR polar codes also outperform the LTE turbo codes due to the considerable repetition required in the LTE turbo codes with a single mother-code rate. Note that only a small performance improvement is observed with the NR polar codes when the list size is increased beyond the target list size of eight.

Figure 9 shows the SNR required to achieve a target BLER of 0.1%, 1%, and 10% as a function of payload size K at code rate $R = 1/2$ with QPSK for SC and SCL decoder with list size of eight. The impact of different

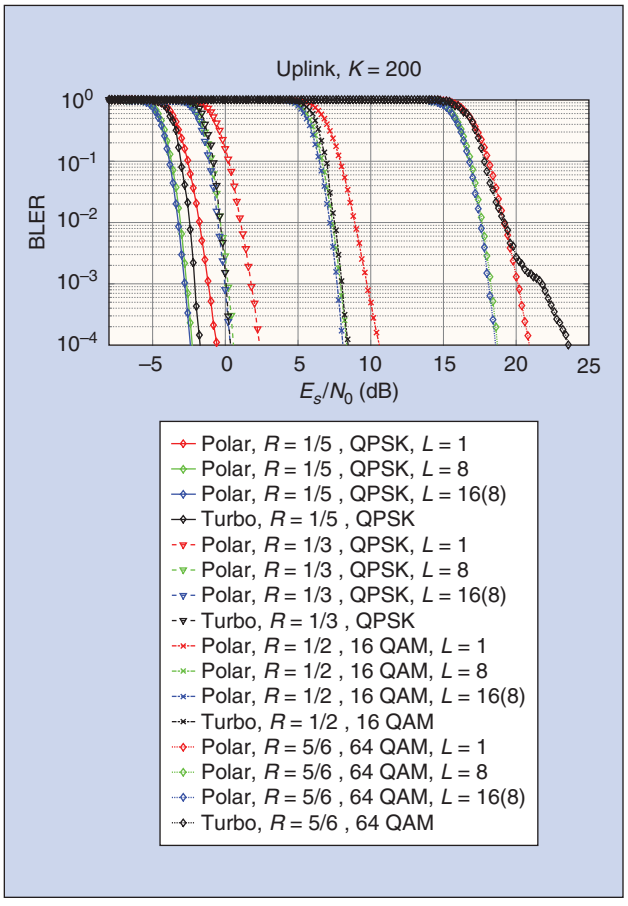


FIGURE 8 A performance comparison between the NR polar codes in uplink and the LTE turbo codes for 200 payload bits.

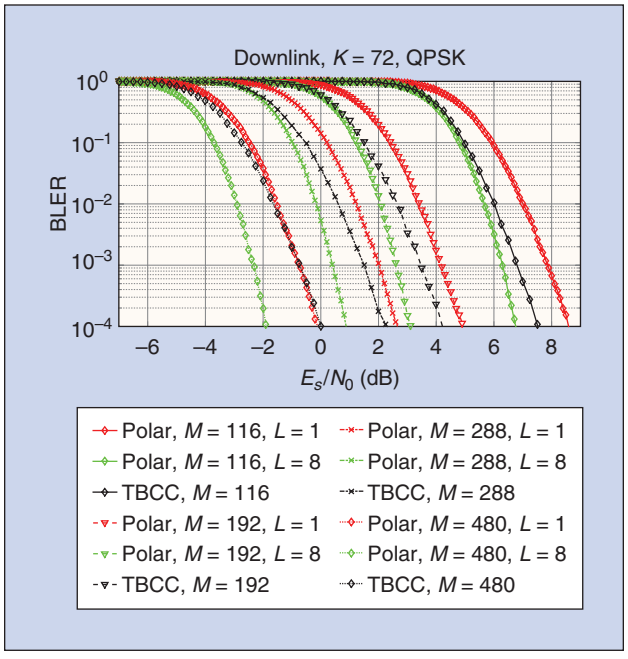


FIGURE 7 A performance comparison between the NR polar code in the downlink and the LTE TBCC with 21 CRC bits for 72 payload bits.

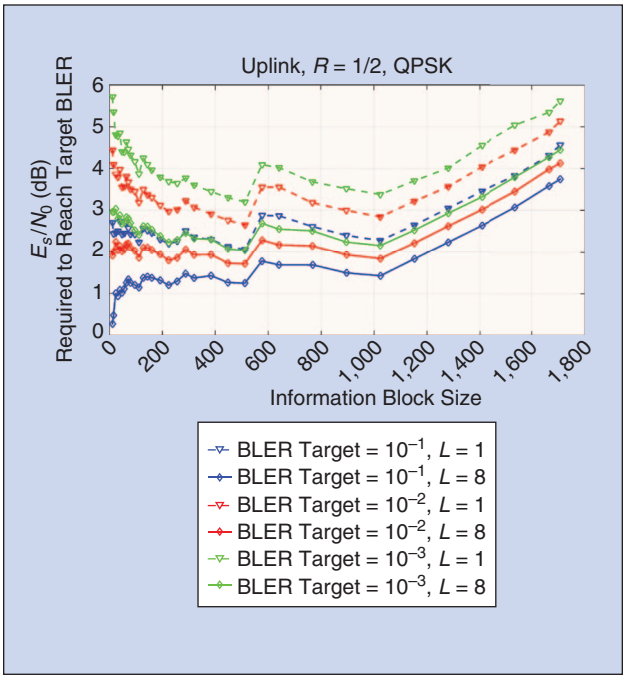


FIGURE 9 The required SNR to achieve 0.1%, 1%, and 10% BLER at rate 1/2 with the NR polar codes in uplink.

TABLE 3 *The performance comparison of the NR LDPC codes, NR polar codes, and LTE turbo codes. The performance numbers state the SNR in decibels required to achieve the desired BLER.*

			Performance (dB)		
Info Block Length (K)	*Code Rate*	*BLER Target*	*NR LDPC*	*NR Polar*	*LTE Turbo*
40 b	1/3 QPSK	0.1	0.12	–1.37	–0.09
		0.0001	2.84	1.15	2.56
	5/6 64 QAM	0.1	18.60	15.72	18.07
		0.0001	22.56	19.00	22.38
400 b	1/3 QPSK	0.1	–0.77	–0.51	–0.99
		0.0001	0.25	0.94	–0.05
	5/6 64 QAM	0.1	17.61	17.05	17.65
		0.0001	19.42	18.55	19.51
6144 b	1/3 QPSK	0.1	–1.12	n/a	–1.34
		0.0001	–0.86	n/a	–1.05
	5/6 64 QAM	0.1	16.90	n/a	17.11
		0.0001	17.26	n/a	17.86

n/a: not applicable.

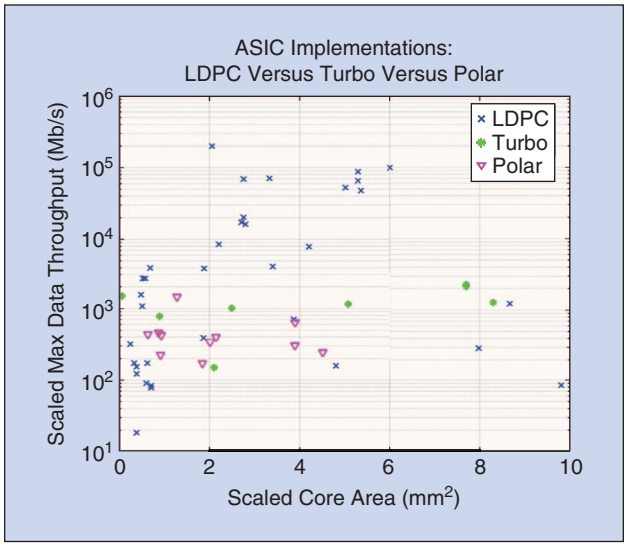


FIGURE 10 Examples of scaled core chip area of application-specified integrated circuit implementation versus scaled maximum achievable data throughput of LDPC, turbo and polar codes. ASIC: application-specified integrated circuit.

rate-matching methods and segmentation is displayed. The required SNR tends to decrease as K increases to a point of around $K = K_{\max}R = 1,024R$ when repetition replaces puncturing or shortening during rate-matching. From that point on, the required SNR keeps increasing until $K = 1,088R$ where the required SNR starts to decrease again due to segmentation that allows the rate-matching to switch back to puncturing/shortening. From around $K = 2K_{\max}R = 2,048R$, the required SNR keeps increasing as K increases because there is no additional segmentation beyond two segments, and repetition of the $N = 1,024$ polar code core is used exclusively from that point on.

Performance and Complexity Comparison

To highlight the reasoning behind the selection of LDPC and polar codes for NR, the performance and complexity of the NR LDPC codes and the NR polar codes are compared with those of the LTE turbo codes that are presented.

Table 3 summarizes the performance comparison among different coding schemes at various values of K and R. As shown, the NR polar codes outperform both the NR LDPC and the LTE turbo codes, by around 1 to 3 dB, for small K (e.g., 40 b). For moderate K (e.g., 400 b), the NR polar codes also have superior performance at high code rates but may suffer from segmentation loss at low code rates. The NR polar codes are not defined for $K > 1,706$. For large K (e.g., 6,144 b), the LDPC codes typically outperform the turbo codes at high code rates while the turbo codes perform slightly better than the LDPC codes at low code rates. Note that $R = 1/3$ is the native code rate of the LTE turbo codes where their performance is most optimized.

Besides performance, the implementation complexity of the decoder that is required to support a desired data throughput also played an important role in the selection of NR coding schemes. Since an increase in the silicon footprint can often be traded for an increase in data throughput, the area efficiency in terms of the tradeoff between the maximum achievable throughput and the required core area of hardware is most relevant in code selection. Figure 10 shows a scatterplot of the achievable application-specified integrated circuit area efficiency of the LDPC, turbo, and polar decoders, based on published implementations (see [13]–[15] and references therein for details), all normalized to a technology scale of 65 nm (see [15] for details on normalization). As the plot shows, the LDPC codes can be implemented much more efficiently than both the turbo codes and polar codes.

With both complexity and performance taken into consideration, it was apparent during 5G NR standardization that LDPC codes are most suitable for long block lengths typically used in data channels, while polar codes are well suited for short block lengths commonly used in control channels.

Conclusions and Future Directions

This article provides a tutorial overview of the two new channel coding schemes, LDPC and polar codes, used in the data and control channels of the 5G NR air interface. A complete coding chain that includes the key components is shown for each of these codes. The purpose and

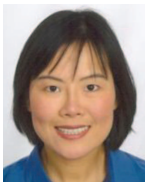
the associated operations of each key component in these codes were explained. We compared these new codes with those used in 4G LTE, i.e., turbo codes and TBCCs, and highlighted the performance and implementation advantages of the new codes over those being replaced. We show that the performance of the NR LDPC codes is similar to that of the LTE turbo codes except at high code rates where the turbo codes tend to exhibit an error floor, or at low code rates where repetition of coded bits is needed for the LTE turbo codes. The advantage of the NR LDPC codes lies in their capability to achieve higher throughput and lower latency than LTE turbo codes since the LDPC decoder structure allows much higher degree of parallel processing. On the other hand, the NR polar codes generally outperform the LTE TBCCs for those payload sizes of interest for control information, especially at low code rates. However, the complexity of the NR polar codes is significantly higher than that of the LTE TBCCs.

We conclude this article with some future areas of study in channel coding for 5G NR or beyond. Although some proposed solutions [11] exist in the literature, the problem of how to effectively support incremental redundancy in retransmissions with polar codes has not been resolved in practice. In fact, this is a main reason why polar codes cannot be used for data channels where retransmission is crucial for system performance. For LDPC codes, it may be necessary to introduce new PCMs if ultrareliable link performance is required. For example, if the BLER target is lowered to 10^{-7} for infrastructure backhaul of intelligent transport systems, it is unknown if the existing PCMs can provide sufficient reliability since their performances are verified only down to $10^{-4} \sim 10^{-5}$ during their design.

Author Information

Dennis Hui (dennis.hui@ericsson.com) is with Ericsson Research in Santa Clara, California. His research interests encompass both theoretical and practical areas, including lossy source coding; channel coding; non-Gaussian interference mitigation; multiple-input, multiple-output/beamforming; coordinated multipoint communications; and distributed coordination.

Sara Sandberg (sara.sandberg@ericsson.com) was previously an assistant professor at the Luleå University of Technology, Sweden. She joined Ericsson Research in Luleå, Sweden, in 2011. Her research focuses on fifth-generation technologies, including channel coding, multiple-input, multiple-output, and the Internet of Things, and she holds many patents and several publications in this field.



Yufei Blankenship (yufei.blankenship@ericsson.com) is currently a standards researcher with Ericsson in Chicago, Illinois. For 3rd Generation Partnership Project New Radio Release 15, her focuses were channel coding and ultrareliable and low-latency communications. She holds numerous patents and publications in the field of wireless communications.

Mattias Andersson (mattias.w.andersson@ericsson.com) is with Ericsson Research in Stockholm, Sweden, focusing on ultrareliable and low-latency communications, carrier aggregation, and channel coding for long-term evolution and new radio.

Leefke Grosjean (leefke.grosjean@ericsson.com) is with Ericsson Research in Stockholm, Sweden. She is an experienced researcher working on standardization and future trends in channel coding.

References

[1] 3GPP. (2018). Multiplexing and channel coding. 3rd Generation Partnership Project. Sophia Antipoli, France. TS 38.212, v15.0.0, Release 15. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.212/38212-f20.zip

[2] ETSI. (2014). LTE; evolved universal terrestrial radio access (E-UTRA); multiplexing and channel coding. European Telecommunications Standards Inst. Sophia-Antipolis, France. TS 136 212 v12.2.0, Release 12. [Online]. Available: http://www.etsi.org/deliver/etsi_ts/136200_136299/136212/12.02.00_60/ts_136212v120200p.pdf

[3] R. G. Gallager, *Low Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[4] D. MacKay and R. Neal, "Good codes based on very sparse matrices," in *Proc. 5th IMA Conf. Cryptography and Coding*, Oct. 1995, pp. 100–111.

[5] N. Alon and M. Luby, "A linear time erasure-resilient code with nearly optimal recovery," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1732–1736, Nov. 1996.

[6] G. Liva, W. E. Ryan, and M. Chiani, "Quasi-cyclic generalized LDPC codes with low error floors," *IEEE Trans. Commun.*, vol. 56, no. 1, pp. 49–57, Jan. 2008.

[7] T. Y. Chen, K. Vakilinia, D. Divsalar, and R. D. Wesel, "Protograph-based raptor-like LDPC codes," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1522–1532, May 2015.

[8] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3051–3073, July 2009.

[9] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Symp. Information Theory*, 2011, pp. 1–5.

[10] P. Trifonov and V. Miloslavskaya, "Polar codes with dynamic frozen symbols and their decoding by directed search," in *Proc. 2013 IEEE Information Theory Workshop*, Sept. 2013, pp. 1–5.

[11] S. Hong, D. Hui, and I. Marić, "Capacity-achieving rate-compatible polar codes," *IEEE Trans. Inf. Theory*, vol. 63, pp. 7620–7632, Dec. 2017.

[12] Ericsson. (2016, May 23–27). Performance evaluation of turbo codes and LDPC codes at higher code rates. Ericsson. Nanjing, China. Tech. Rep. R1-164359, 3GPP TSG RAN WG1 Meeting 85. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_85/Docs/R1-164359.zip

[13] Ericsson. (2016, Aug. 22–26). Survey of channel coding implementations. Ericsson. Gothenburg, Sweden. Tech. Rep. R1-166936, 3GPP TSG RAN WG1 Meeting 86. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_86/Docs/R1-166936.zip

[14] C. Chavet and P. Coussy, Eds., *Advanced Hardware Design for Error Correcting Codes*, New York, NY: Springer, 2015.

[15] R. G. Maunder. (Aug. 2016). Survey of ASIC implementations of LDPC decoders. University of Southampton Institutional Repository. [Online]. Available: http://eprints.soton.ac.uk/399259/

VT