SCIENCE CHINA
Information Sciences

• **Supplementary File** •

A latency-reduced successive cancellation list decoder for polar codes

Yongrun YU[1], Zhiwen PAN[1*], Xiaosi TAN[1], Nan LIU[1], Xiaohu YOU[1] & Fei DING[2]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing, Jiangsu* 210096*, China;*
[2]*School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu* 210096*, China*

Appendix A Motivation of the definition of FCR0 and MSR1 node

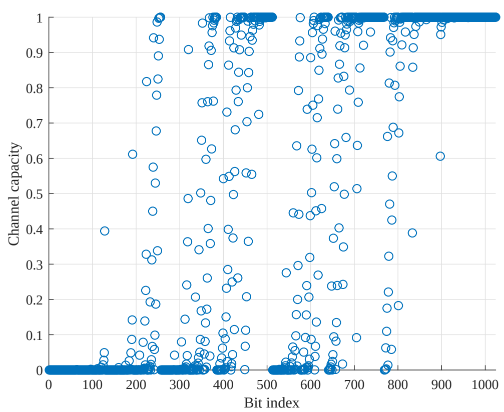

Figure A1 Capacity of synthesized bit channels assuming the underlying channels are BECs with erasure probability 0.5

Let $\mathbf{u}_1^N = (u_1, ..., u_k, u_{k+1}, ..., u_j, u_{j+1}, ..., u_N)$ denote the source bit sequence before encoding.

In terms of polar codes, it is known that source bits with smallest indices tend to be frozen bits as shown in Figure A1 [1] because the corresponding synthesized bit channels have low capacity (the underlying channels are binary erasure channels). Assuming that $(u_1, ..., u_k)$ is all frozen but u_{k+1} is an information bit, the bit cluster $(u_1, ..., u_k)$ is defined as first consecutive rate 0 (FCR0) node, i.e, FCR0 node refers to the consecutive frozen bits with smallest indices. An example of FCR0 node is shown in Figure A2 for polar codes with code length $N = 8$. The row at bottom represents the information bits (black circle) and the frozen bits (white circle) in the source bit sequence. The circles in higher levels indicate the property of the source bits that it contains. For example, the left-most circle at the second line from the bottom is white, which indicates that the two source bits under it (u_1 and u_2) are both frozen bits. Grey circle means that there are both information and frozen bits under it. When the erasure probability of underlying BECs becomes larger, the scale of FCR0 node increases as shown in Figure A3 because the capacities of synthesized bit channels with small indices approach zero. In this case, if the decoding of FCR0 node is simplified, then total decoding latency will be reduced.

According to the channel polarizing in Figure A1, the synthesized bit channels with large indices are likely to be the most reliable channels because the capacities approach 1 and thus a large portion of information index set $\mathcal{A}$ is located at the end of the source bit sequence. This phenomenon is also confirmed by [4], where a partial order of synthesized bit channels is proposed to estimate the reliability of the synthesized bit channels. A heuristic conclusion of [4] is that larger indices imply higher reliability. Assume that $(u_{j+1}, ..., u_N)$ is all information bit but u_j is frozen. Let integer $\gamma = N - 2^{\lfloor \log_2(N-j) \rfloor}$, where $\lfloor z \rfloor$ is the largest integer that does not exceed real number z. The bit cluster $(u_{\gamma+1}, ..., u_N)$ is defined as most significant

* Corresponding author (email: pzw@seu.edu.cn)

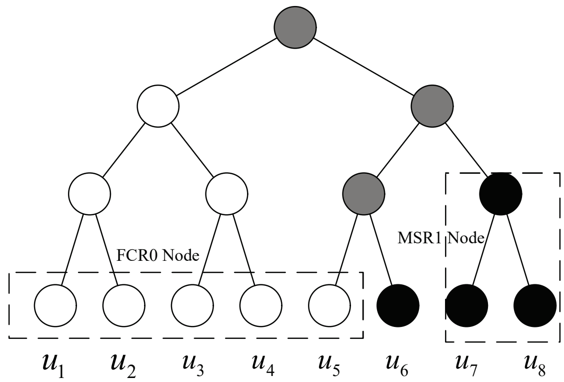

Figure A2 Polar code with length $N = 8$

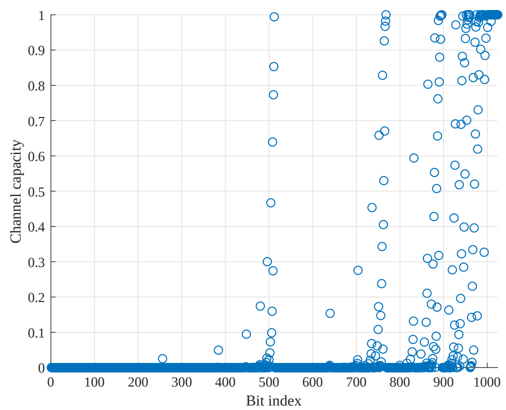

Figure A3 Capacity of synthesized bit channels assuming the underlying channels are BECs with erasure probability 0.9

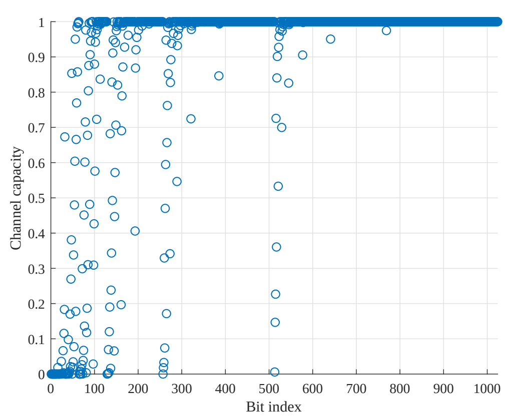

Figure A4 Capacity of synthesized bit channels assuming the underlying channels are BECs with erasure probability 0.1

rate 1 (MSR1) node, i.e., MSR1 node refers to the consecutive information bits with largest indices. An example of MSR1 node is shown in Figure A2. The number of information bits that are included in MSR1 is power of 2 and once the decoding of MSR1 is finished, the whole decoding process is also finished. When the erasure probability of underlying BECs becomes smaller, the scale of MSR1 node increases as shown in Figure A4 because the capacities of synthesized bit channels with large indices approach one. In this case, if the decoding of MSR1 node is simplified, then total decoding latency will be reduced.

When the underlying channels have moderate capacity (as shown in Figure A1), the scale of FCR0 and MSR1 node is also moderate. The joint simplified decoding of both FCR0 and MSR1 will produce lower decoding latency.

Appendix B Proof of lemma 1

The improved decoding of MSR1 node relies on the following lemma 1. In order to make the elaboration clear, some notations are defined as below.

Let $\mathcal{C}$ denote the code book of one binary linear block code and $\mathbf{c}_1^N = (c_1, ..., c_N)$ is a code word in $\mathcal{C}$ (every code word in $\mathcal{C}$ appears with equiprobability).The modulation scheme that is used to transmit $\mathbf{c}_1^N$ is binary-phase shift keying (BPSK). The labeling rule is: bit $0 \to +1$ and bit $1 \to -1$: The modulated sequence is represented by $\mathbf{s}_1^N = (s_1, ..., s_N), s_i \in \{-1, +1\}$. $\mathbf{s}_1^N$ is transmitted through binary-input additive white Gaussian noise (BI-AWGN) channel with noise variance σ^2. $\mathbf{y}_1^N = (y_1, ..., y_N)$ denotes the received signal, i.e., a noise version of $\mathbf{s}_1^N$. Let binary sequence $\hat{\mathbf{c}}_1^N$ denote the maximum likelihood decision (not decoding) result of $\mathbf{y}_1^N$, i.e., if $p(y_i|0) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(y_i-1)^2}{2\sigma^2}\right\} > p(y_i|1) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(y_i+1)^2}{2\sigma^2}\right\}$, then set $\hat{c}_i = 0$ and if $p(y_i|0) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(y_i-1)^2}{2\sigma^2}\right\} < p(y_i|1) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(y_i+1)^2}{2\sigma^2}\right\}$, then set $\hat{c}_i = 1$.

Lemma 1. If $\hat{\mathbf{c}}_1^N$ is exactly a code word in $\mathcal{C}$, then $\hat{\mathbf{c}}_1^N \in \mathcal{C}$ is the result of maximum likelihood decoding using received signal $\mathbf{y}_1^N = (y_1, ..., y_N)$.

Proof. Assume that $\hat{\mathbf{c}}_1^N$ is not the result of maximum likelihood decoding of received signal $\mathbf{y}_1^N$. Then there exists code word $\mathbf{b}_1^N = (b_1, ..., b_N) \in \mathcal{C}$ that is the result of maximum likelihood decoding of received signal $\mathbf{y}_1^N$. $\hat{\mathbf{c}}_1^N$ and $\mathbf{b}_1^N$ will differ at some bit indices. Denote the index set where $\hat{\mathbf{c}}_1^N$ and $\mathbf{b}_1^N$ have different bit values as $\mathcal{D}$, i.e., $\forall i \in \mathcal{D}$, $b_i = \hat{c}_i \oplus 1$, where $\oplus$ is bit-wise xor. Then according to the assumption, we have the following equation:

$$\frac{p(\mathbf{y}|\mathbf{b})}{p(\mathbf{y}|\hat{\mathbf{c}})} = \prod_{i=1}^{N} \frac{p(y_i|b_i)}{p(y_i|\hat{c}_i)} = \prod_{i\in\mathcal{D}} \frac{p(y_i|\hat{c}_i \oplus 1)}{p(y_i|\hat{c}_i)} > 1. \tag{B1}$$

For a given $i \in \mathcal{D}$, there exist two cases:

If $\hat{c}_i = 1$, then according to decision rule we have $p(y_i|0)/p(y_i|1) < 1$, which implies that:

$$\frac{p(y_i|\hat{c}_i \oplus 1)}{p(y_i|\hat{c}_i)} < 1. \tag{B2}$$

If $\hat{c}_i = 0$, then according to decision rule we have $p(y_i|1)/p(y_i|0) < 1$, which implies that :

$$\frac{p(y_i|\hat{c}_i \oplus 1)}{p(y_i|\hat{c}_i)} < 1. \tag{B3}$$

Since every factor $p(y_i|\hat{c}_i \oplus 1)/p(y_i|\hat{c}_i)$ in (8) is less than 1, it is impossible to have $p(\mathbf{y}|\mathbf{b})/p(\mathbf{y}|\hat{\mathbf{c}}) > 1$. Thus the assumption is wrong. $\hat{\mathbf{c}}$ is the result of maximum likelihood decoding of received signal $\mathbf{y}_1^N$. ■

Appendix C Observation of SCL decoding process

According to the Section III.B of [2], in sequential decoder, metric of surviving path is calculated to judge whether the current decoding result is correct. If the current decoding result is incorrect, the associated path metric will become more and more unreliable as the sequential decoding proceeds and thus the decoding error will be detected with probability one. Similarly, it can be inferred that in SCL decoding, the incorrect surviving paths with large LLR-based metric tend to be much more unreliable than the correct path, which indicates that the correct probability of the paths with large LLR-based metric is very small. Even if these unreliable paths are not accurately sorted and selected, the BLER performance loss will be negligible as long as most reliable paths, i.e., paths with small metric, are preserved. This gives inspiration to a suboptimal sorting scheme that has lower latency.

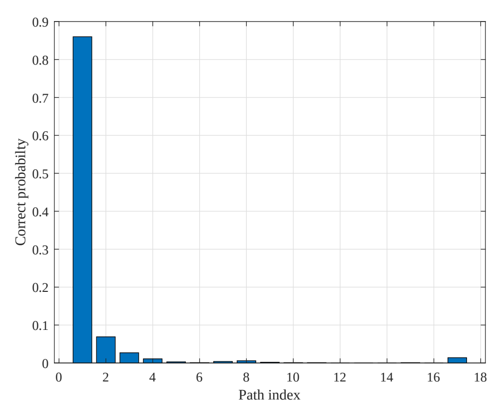

Figure C1 Correct probability of candidate paths, (128,64) polar codes, $L = 16$

Figure C1 shows the correct probability of candidate paths in SCL decoder in the case of BI-AWGN channel with (128, 64) polar codes, i.e., polar codes with length 128 and rate 0.5. The decoder is SCL decoder with list size $L = 16$ and signal-to-noise-ratio (SNR) is 1.75dB. The first column from the left means that the correct probability of the decoding result

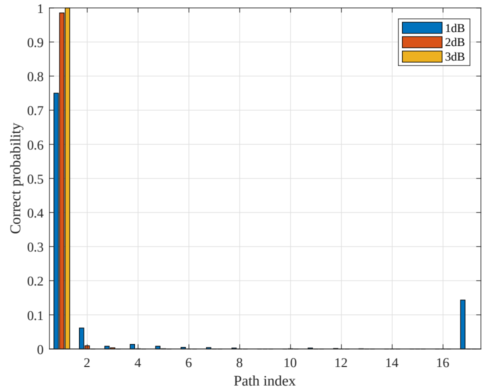

Figure C2 Correct probability of candidate paths, (1024,512) polar codes, $L = 16$

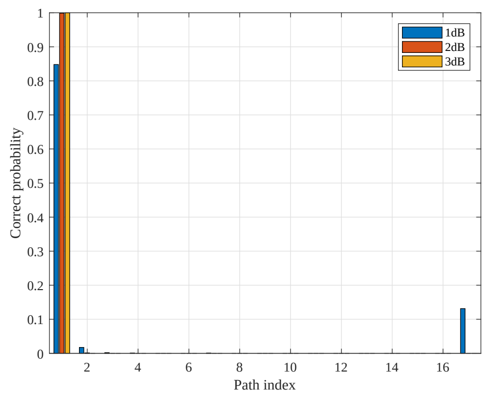

Figure C3 Correct probability of candidate paths, (2048,1024) polar codes, $L = 16$

stored in the path with the minimum metric is 85%. The second column from the left means that the correct probability of the decoding result stored in the path with the second minimum metric is about 8%. The rest columns follow the same meaning except the right-most one (the 17th column) which means that none of the decoding results stored in the 16 lists is correct. Results show that the correct probability of the paths with minimum and second minimum metric is dominant and the correct probability of the worst paths is negligible. The results of different simulation configurations are presented in Figure C2 and Figure C3, where the columns have the same meaning as Figure C1. Figure C2 and C3 show the correct path probability of (1024,512) and (2048, 1024) polar codes, respectively. The SNR operating points are 1,2 and 3dB. The list size of SCL decoder is 16, too. Both Figure C1, C2 and C3 show that the correct probability of the paths with minimum and second minimum metric is dominant, especially in high SNR region.

The results in above figures imply that if the paths with minimum and second minimum metrics are preserved, then it is very likely that the correct path is survived. Therefore, it is not necessary to exactly select paths with L smallest metrics as long as the two paths with minimum and second minimum metric are preserved. The proposed suboptimal path sorting algorithm is described below. Note that the proposed sorting still preserves L paths in every sorting procedure.

Appendix D Proof of lemma 2

When decoding the information bit u_i , L surviving paths split into $2L$ candidate paths. The matrix that saves the $2L$ values of path metric is:

$$\mathbf{P} = \begin{pmatrix} PM_1 & ... & PM_i & ... & PM_j & ... & PM_L \\ PM_1^+ & ... & PM_i^+ & ... & PM_j^+ & ... & PM_L^+ \end{pmatrix}, \tag{D1}$$

where PM_i^+ represents that $PM_i^+ \geqslant PM_i$ holds. The partial order between PM_i^+ and PM_i is naturally obtained [3].

The proposed sorting scheme compares PM_i with PM_{i+1}^+ and select the path corresponding to the smaller metric as surviving path, i.e., if $PM_i \leqslant PM_{i+1}^+$, then the candidate path with metric PM_i is preserved and the candidate path with metric PM_{i+1}^+ is discarded. PM_L should be compared with PM_1^+. This sorting procedure needs L times of comparisons and as a result, L survival paths are selected.

Lemma 2. Through the proposed sorting algorithm, the two paths corresponding to the minimum and second minimum metrics are preserved.

Proof. The path with minimum metric is obviously preserved because none of other paths owns smaller metrics. The minimum metric is always in the first row according to (D1). Assume that the path with second minimum metric is eliminated. Under this assumption, the second minimum metric must be eliminated by the minimum metric, i.e., if PM_k is the minimum metric, then the second minimum metric is PM_{k+1}^+ and thereby PM_{k+1}^+ is eliminated. This is impossible because PM_{k+1} is smaller than PM_{k+1}^+ and thus PM_{k+1} is the second minimum metric, which contradicts that PM_{k+1}^+ is the second minimum metric. Therefore, the path with second minimum metric is preserved. ■

Appendix E Analysis of latency reduction

Here we assume that the computation of LLR of one synthesized bit channel takes unit time.

It is easy to analyze the reduction of decoding latency due to the improved decoding of $\mathcal{F}$. Assume that there are $N-K$ frozen bits in the source bit sequence and $\mathcal{F}$ includes N_0 frozen bits, where N is code word length and K is the number of information bits. In standard SCL decoder, the number of LLRs that correspond to frozen bits is $N-K$. With the improved decoding of $\mathcal{F}$, the number of LLRs that correspond to frozen bits is $(N-K)-N_0$. The ratio is:

$$r_0 = \frac{(N-K)-N_0}{N-K} = 1 - \frac{N_0}{N-K}. \tag{E1}$$

Besides reduced calculations of LLR of frozen bit, the update of path metric in $\mathcal{F}$ is also omitted.

The reduction of decoding latency due to the improved decoding of $\mathcal{M}$ is analyzed as follows. When the SCL decoder is initialized, there is only one path. After the decoding of the first information bit, one path splits into two paths. This process continues until the number of candidate path approaches $2L$, which indicates that the path sorting operation is needed. When decoding each information bit after the first sorting operation, there are $2L$ candidate paths in the decoder and thus two LLRs of synthesized bit channel should be computed. In standard SCL decoder, the number of LLRs that correspond to information bit is:

$$1+2+4+...+L+2L(K-1-log_2L) = 2L(K-log_2L)-1. \tag{E2}$$

With the improved decoding of $\mathcal{M}$, the number of LLRs that correspond to information bits is:

$$1+2+4+...+L+2L(K-1-N_1-log_2L) = 2L(K-N_1-log_2L)-1, \tag{E3}$$

where N_1 is the number of information bits that $\mathcal{M}$ includes. The ratio of (16) and (17) is:

$$r_1 = \frac{2L(K-N_1-log_2L)-1}{2L(K-log_2L)-1} = 1 - \frac{N_1}{K-log_2L-\frac{1}{2L}}. \tag{E4}$$

For example, in the case of (1024, 512) polar codes under SCL decoder with list size $L=32$ and $\mathcal{M}$ includes $N_1=128$ bits, $r_1 \approx 74.8\%$, i.e., 25.2 % calculations for decoding information bits are omitted.

The decoding latency improved by the suboptimal sorting algorithm is significant. There are only L times of comparisons during each path sorting procedure and only L comparators are required, which is much simpler than the existing sorting algorithms in [5–10]. The detailed complexity (numbers of required compare and select units) comparison is given in Table E1. Number M in bit-wise sorter [9] is the number of bits of metric value stored in the hardware memory. c_{L-1} in simplified bitonic sorter [10] denotes the size of $(L-1)$-wire sorting network. L deontes the list size of SCL decoder.

Table E1 Sorting complexity comparison with existing schemes

Sorting scheme	Complexity
Bitonic sorter [5]	$\frac{L}{2}(1+\log L)(2+\log L)$
Full bitonic extractor [6]	$L((1+\log_2 L)^2/2-\log_2 L+1)$
Simplified bubble sorter [7]	$\frac{L}{2}(L-1)$
Pruned bitonic sorter [7]	$(\frac{L}{2}-1)(\log L)(2+\log L)+1$
Odd-even sorter [8]	$(\log_2 L)(\frac{L}{4}\log_2 L-1)+\frac{7}{4}L-2$
Bit-wise sorter [9]	ML
Simplified bitonic sorter [10]	$c_{L-1}+(L-1)+(\frac{L}{2}-1)\log_2 L$
Proposed suboptimal sorter	L

Appendix F More simulation results

Here we demonstrate the BLER performance of the proposed decoder under list size 4, 8 and 16.

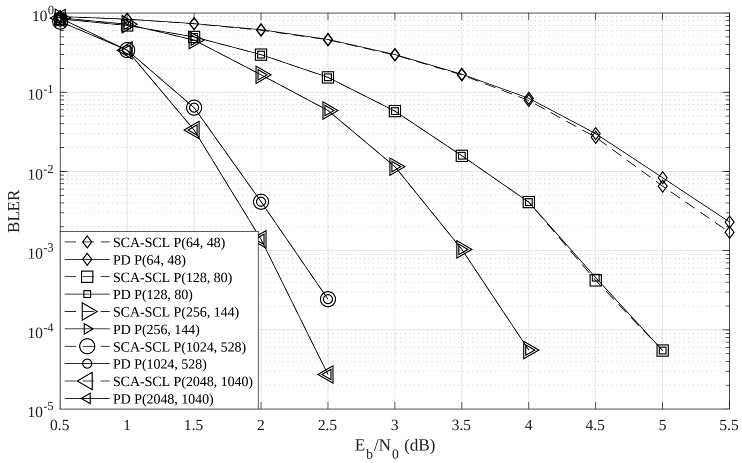

Figure F1 BLER of list size $L = 4$.

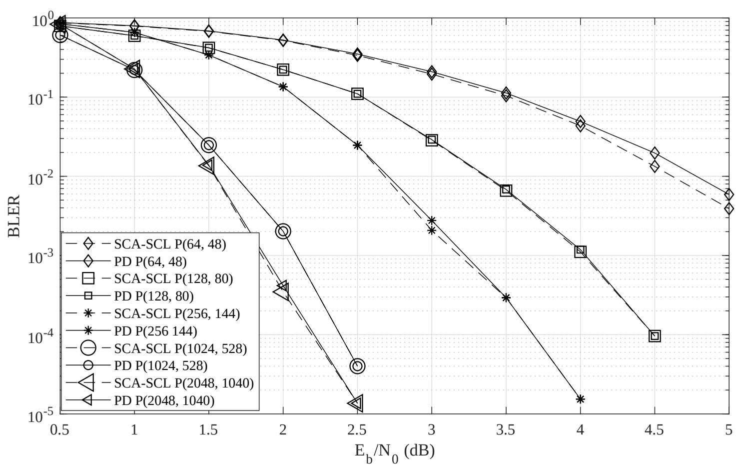

Figure F2 BLER of list size $L = 8$.

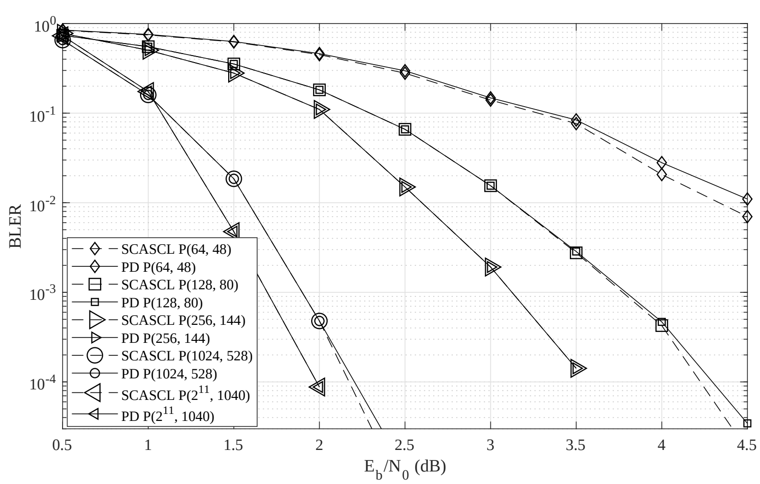

Figure F3 BLER of list size $L = 16$.

In all above figures, except $P(64,48)$ and $P(128,64)$ suffering a little BLER loss when SNR is high (more than 4dB), under all remaining polar code configurations, the BLER between standard CA-SCL and the proposed decoder are almost overlapped for various configurations. For $P(64,48)$ and $P(128,80)$, the BLER gap in high SNR region may results from that the rate of short and moderate polar code tend to be influenced by the length of outer CRC code. For example, in $P(64,48)$, $R = 48/64 = 0.75$ is relatively higher. This implies that there are relatively more suboptimal sorting process, which may introduce some BLER degradation.

References

1 Arikan E. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans Inf Theory, 2009, 55: 3051-3073

2 Arikan E. On the origin of polar coding. IEEE J Sel Area Comm, 2016, 34: 209-223

3 Balatsoukas-Stimming A, Parizi M B, Burg A. LLR-based successive cancellation list decoding of polar codes. IEEE Trans Signal Process, 2015, 63: 5165-5179

4 Schurch C. A partial order for the synthesized channels of a polar code. In: Proceedings of the 2016 IEEE International Symposium on Information Theory, Barcelona, 2016. 220-224

5 Lin J, Yan Z. Efficient list decoder architecture for polar codes. In: Proceedings of the 2014 IEEE International Symposium on Circuits and Systems, Melbourne VIC, 2014. 1022-1025

6 Lin J, Yan Z. An efficient list decoder architecture for polar codes. IEEE Trans VLSI Syst, 2015, 23: 2508-2518

7 Balatsoukas-Stimming A, Parizi M B, Burg A. On metric sorting for successive cancellation list decoding of polar codes. In: Proceedings of the 2015 IEEE International Symposium on Circuits and Systems, Lisbon, 2015. 1993-1996

8 Kong B Y, Yoo H, Par I C. Efficient sorting architecture for successive-cancellation-list decoding of polar codes. IEEE Trans Circuits Syst II Express Briefs, 2016, 63: 673-677

9 Shi B, Han B, Gao J L, et al. Enhanced successive cancellation list decoding of polar codes. IEEE Commun Lett, 2017, 21: 1233-1236

10 Song H, Zhang S, You X, et al. Efficient metric sorting schemes for successive cancellation list decoding of polar codes. In: Proceedings of the 2017 IEEE International Symposium on Circuits and Systems, Baltimore, 2017. 1-4