

Threshold-Based Fast Successive-Cancellation Decoding of Polar Codes

Citation for published version (APA):

Zheng, H., Hashemi, S. A., Balatsoukas-Stimming, A., Cao, Z., Koonen, T., Cioffi, J., & Goldsmith, A. (2021). Threshold-Based Fast Successive-Cancellation Decoding of Polar Codes. *IEEE Transactions on Communications*, 69(6), 3541-3555. Article 9373370. <https://doi.org/10.1109/TCOMM.2021.3064964>

Document license:

TAVERNE

DOI:

[10.1109/TCOMM.2021.3064964](https://doi.org/10.1109/TCOMM.2021.3064964)

Document status and date:

Published: 01/06/2021

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Threshold-Based Fast Successive-Cancellation Decoding of Polar Codes

Haotian Zheng^{ib}, *Student Member, IEEE*, Seyyed Ali Hashemi, *Member, IEEE*,
 Alexios Balatsoukas-Stimming^{ib}, *Member, IEEE*, Zizheng Cao^{ib}, *Member, IEEE*,
 Ton Koonen^{ib}, *Fellow, IEEE*, John M. Cioffi^{ib}, *Fellow, IEEE*,
 and Andrea Goldsmith^{ib}, *Fellow, IEEE*

Abstract—Fast SC decoding overcomes the latency caused by the serial nature of the SC decoding by identifying new nodes in the upper levels of the SC decoding tree and implementing their fast parallel decoders. In this work, we first present a novel sequence repetition node corresponding to a particular class of bit sequences. Most existing special node types are special cases of the proposed sequence repetition node. Then, a fast parallel decoder is proposed for this class of node. To further speed up the decoding process of general nodes outside this class, a threshold-based hard-decision-aided scheme is introduced. The threshold value that guarantees a given error-correction performance in the proposed scheme is derived theoretically. Analysis and hardware implementation results on a polar code of length 1024 with code rates $1/4$, $1/2$, and $3/4$ show that our proposed algorithm reduces the required clock cycles by up to 8%, and leads to a 10% improvement in the maximum operating frequency compared to state-of-the-art decoders without tangibly altering the error-correction performance. In addition, using the proposed threshold-based hard-decision-aided scheme, the decoding latency can be further reduced by 57% at $E_b/N_0 = 5.0$ dB.

Index Terms—Polar codes, fast successive-cancellation decoding, sequence repetition node, threshold-based hard-decision-aided scheme.

I. INTRODUCTION

POLAR codes represent a channel coding scheme that can provably achieve the capacity of binary-input memoryless channels [1]. The explicit coding structure and low-complexity

successive-cancellation (SC) decoding algorithm has generated significant interest in polar code research across both industry and academia. In particular, in the latest cellular standard of 5G [2], polar codes are adopted in the control channel of the enhanced mobile broadband (eMBB) use case. Although SC decoding provides a low-complexity capacity-achieving solution for polar codes with long block length, its sequential bit-by-bit decoding nature leads to high decoding latency, which constrains its application in low-latency communication scenarios such as the ultra-reliable low-latency communication (URLLC) [2] scheme of 5G. Therefore, the design of fast SC-based decoding algorithms for polar codes with low decoding latency has received a lot of attention [3].

A look-ahead technique was adopted to speed up SC decoding in [4]–[6] by pre-computing all the possible likelihoods of the bits that have not been decoded yet, and selecting the appropriate likelihood once the corresponding bit is estimated. Using the binary tree representation of SC decoding of polar codes, instead of working at the bit-level that corresponds to the leaf nodes of the SC decoding tree, parallel multi-bit decision is performed at the intermediate nodes of the SC decoding tree. An exhaustive-search decoding algorithm is used in [7]–[10] to make multi-bit decisions and to avoid the latency caused by the traversal of the SC decoding tree to compute the intermediate likelihoods. However, due to the high complexity of the exhaustive search, this method is generally only suitable for nodes that represent codes of very short lengths.

It was shown in [11] that a node in the SC decoding tree that represents a code of rate 0 (Rate-0 node) or a code of rate 1 (Rate-1 node) can be decoded efficiently without traversing the SC decoding tree. In [12], fast decoders of repetition (REP) and single parity-check (SPC) nodes were proposed for the SC decoding. Using these nodes, a fully-unrolled hardware architecture was proposed in [13]. Techniques were developed in [14]–[18] to adjust the codes that are represented by the nodes in the SC decoding tree to increase the number of nodes that can be decoded efficiently. However, these methods result in a degraded error-correction performance. On the basis of the works in [11], [12], five new nodes (Type-I, Type-II, Type-III, Type-IV, and Type-V) were identified and their fast decoders were designed in [19]. In [20], a generalized REP (G-REP) node and a generalized parity-check (G-PC) node were proposed to reduce the latency of the SC decoding even

Manuscript received March 20, 2020; revised August 29, 2020, December 13, 2020, and February 13, 2021; accepted February 14, 2021. Date of publication March 9, 2021; date of current version June 16, 2021. This work is supported by NWO Zwaartekracht program on Integrated Nanophotonics; ZJU-TU/e IDEAS project; Open Fund of the State Key Laboratory of Optoelectronic Materials and Technologies (Sun Yat-sen University); Sichuan Science and Technology Program(2020YFH0108). The work of Seyyed Ali Hashemi was supported by the Postdoctoral Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). This article was presented at the IEEE International Conference on Communications, June 2020. The associate editor coordinating the review of this article and approving it for publication was I. Tal. (*Corresponding author: Zizheng Cao.*)

Haotian Zheng, Alexios Balatsoukas-Stimming, Zizheng Cao, and Ton Koonen are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: h.zheng@tue.nl; a.k.balatsoukas.stimming@tue.nl; z.cao@tue.nl; a.m.j.koonen@tue.nl).

Seyyed Ali Hashemi and John M. Cioffi are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: ahashemi@stanford.edu; cioffi@stanford.edu).

Andrea Goldsmith is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: goldsmith@princeton.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCOMM.2021.3064964>.

Digital Object Identifier 10.1109/TCOMM.2021.3064964

0090-6778 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

further. In [21], seven of the most prevalent node patterns in short polar codes were analysed and efficient algorithms for processing these node patterns in parallel are proposed. However, the decoding of some of these node patterns leads to significant performance loss. Memory footprint optimization and operation merging were described in [22], [23] to improve the speed of SC decoding. The combination of fast SC and fast SC flip decoding algorithms was investigated in [24], [25]. Moreover, the identification and the utilization of the aforementioned special nodes were also extended to SC list (SCL) [26] decoding [27]–[31]. All these works require the design of a separate decoder for each class of node, which inevitably increases the implementation complexity. In addition, as shown in this work, the achievable parallelism in decoding can be further increased without degrading the error-correction performance.

For general nodes that do not fall in one of the above node categories, [32] proposed a hard-decision scheme based on node error probability. Specifically, in that work it was shown that extra latency reduction can be achieved when the communications channel has low noise. However, the hard-decision threshold is calculated empirically rather than for a desired error-correction performance. In [33], a hypothesis-testing-based strategy is designed to select reliable unstructured nodes for hard decision. However, additional operations are required to be performed to calculate the decision rule, thus, incurring extra decoding latency. For all the existing hard-decision schemes, a threshold comparison operation is required each time a general constituent code is encountered in the course of the SC decoding algorithm.

In this article, a fast SC decoding algorithm with a higher degree of parallelism than the state of the art is proposed. First, a class of *sequence repetition* (SR) nodes is proposed which provides a unified description of most of the existing special nodes. This class of nodes is typically found at a higher level and has a higher frequency of occurrence in the decoding tree than other existing special nodes. 5G polar codes with different code lengths and code rates can be represented using only SR nodes. Utilizing this class of nodes, a simple and efficient fast simplified SC decoding algorithm called the *SR node-based fast SC* (SRFSC) decoding algorithm is proposed, which achieves a high degree of parallelism without degrading the error-correction performance. Employing only SR nodes, the SRFSC decoder requires a lower number of time steps than most existing works. Moreover, SR nodes can be implemented together with other operation mergers, such as node-branch and branch mergers, to outperform the state-of-the-art decoders in terms of the required number of time steps. In addition, if a realistic hardware implementation is considered, the proposed SRFSC decoder provides up to 8% reduction in terms of the required number of clock cycles and 10% improvement in terms of the maximum operating frequency with respect to the state-of-the-art decoders on a polar codes of length 1024 changed with code rates 1/4, 1/2, and 3/4.

Second, a threshold-based hard-decision-aided (TA) scheme is proposed to speed up the decoding of the nodes that

are not SR nodes for a binary additive white Gaussian noise (BAWGN) channel. Consequently, a TA-SRFSC decoding algorithm is proposed that adopts a simpler threshold for hard-decision than that in [32]. The effect of the defined threshold on the error-correction performance of the proposed TA-SRFSC decoding algorithm is analyzed. Moreover, a systematic way to derive the threshold value for a desired upper bound for its block error rate (BLER) is determined. Performance results show that, with the help of the proposed TA scheme, the decoding latency of SRFSC decoding can be further reduced by 57% at $E_b/N_0 = 5$ dB on a polar code of length 1024 and rate 1/2. In addition, a multi-stage decoding strategy is introduced to mitigate the possible error-correction performance loss of TA-SRFSC decoding with respect to SC decoding, while achieving average decoding latency comparable to TA-SRFSC decoding.

The rest of this article is organized as follows. Section II gives a brief introduction to the basic concept of polar codes and fast SC decoding. In Section III, the SRFSC decoding algorithm is introduced. With the help of the proposed TA scheme, the TA-SRFSC decoding is presented in Section IV. Section V analyzes the decoding latency and simulation results are shown in Section VI. Finally, Section VII gives a summary of the article and concluding remarks.

II. PRELIMINARIES

A. Notation Conventions

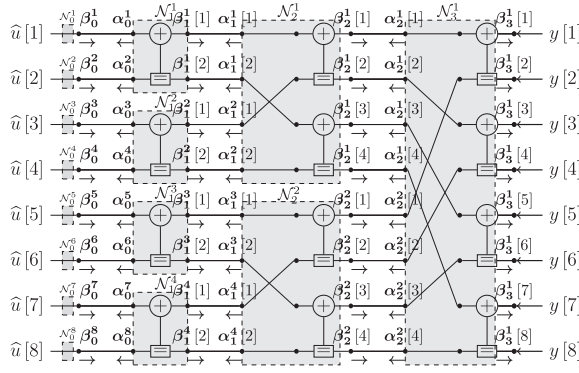
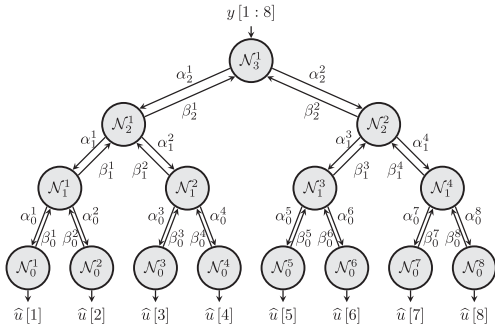
In this article, blackboard letters, such as \mathbb{X} , denote a set and $|\mathbb{X}|$ denotes the number of elements in \mathbb{X} . Bold letters, such as \mathbf{v} , denote a row vector, \mathbf{v}^T denotes the transpose of \mathbf{v} , and notation $\mathbf{v}[i:j]$, $1 \leq i < j$ represents a subvector $(\mathbf{v}[i], \mathbf{v}[i+1], \dots, \mathbf{v}[j])$. \oplus is used as the bitwise XOR operation and $\mathbf{v}[i:j] \oplus \mathbf{z} = (\mathbf{v}[i] \oplus \mathbf{z}, \mathbf{v}[i+1] \oplus \mathbf{z}, \dots, \mathbf{v}[j] \oplus \mathbf{z})$, $\mathbf{z} \in \{0, 1\}$. The Kronecker product of two matrices \mathbf{F} and \mathbf{G} is written as $\mathbf{F} \otimes \mathbf{G}$. \mathbf{F}_N represents an $N \times N$ square matrix and $\mathbf{F}_N^{\otimes n}$ denotes the n -th Kronecker power of \mathbf{F}_N . Throughout this article, $\ln(x)$ denotes the natural logarithm and $\log(x)$ indicates the base-2 logarithm of x , respectively.

B. Polar Codes

A polar code with code length $N = 2^n$ and information length K is denoted by $\mathcal{P}(N, K)$ and has rate $R = K/N$. The encoding can be expressed as $\mathbf{x} = \mathbf{u}\mathbf{G}_N$, where $\mathbf{u} = (u[1], u[2], \dots, u[N])$ is the input bit sequence and $\mathbf{x} = (x[1], x[2], \dots, x[N])$ is the encoded bit sequence. $\mathbf{G}_N = \mathbf{R}_N \mathbf{F}_2^{\otimes n}$ is the generator matrix, where \mathbf{R}_N is a bit-reversal permutation matrix and $\mathbf{F}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

The input bit sequence \mathbf{u} consists of K information bits and $N - K$ frozen bits. The information bits form set \mathbb{A} , transmitting information bits, while the frozen bits form set \mathbb{A}^c , transmitting fixed bits known to the receiver. For symmetric channels, without loss of generality, all frozen bits are set to zero [1]. To distinguish between frozen and information bits, a vector of flags $\mathbf{d} = (d[1], d[2], \dots, d[N])$ is used where each flag $d[k]$ is assigned as

$$d[k] = \begin{cases} 0, & \text{if } k \in \mathbb{A}^c, \\ 1, & \text{otherwise.} \end{cases} \quad (1)$$


 Fig. 1. SC decoding on the factor graph of a polar code $N = 8$.

 Fig. 2. Binary tree representation of a SC decoder for a polar code with $N = 8$.

The codeword \mathbf{x} is transmitted through a channel after modulation. In this article, non-systematic polar codes and binary phase-shift keying (BPSK) modulation that maps $\{0, 1\}$ to $\{+1, -1\}$ are considered. Transmission takes place over an additive white Gaussian noise (AWGN) channel.

C. SC Decoding and Binary Tree Representation

SC decoding can be illustrated on the factor graph of polar codes as shown in Fig. 1. The factor graph consists of $n + 1$ levels and, by grouping all the operations that can be performed in parallel, SC decoding can be represented as the traversal of a binary tree. This traversal is shown in Fig. 2, starting from the left side of the binary tree. At level j of the SC decoding tree with $n + 1$ levels, there are 2^{n-j} nodes ($0 \leq j \leq n$), and the i -th node at level j ($1 \leq i \leq 2^{n-j}$) of the SC decoding tree is denoted as \mathcal{N}_j^i . The left and the right child nodes of \mathcal{N}_j^i are \mathcal{N}_{j-1}^{2i-1} and \mathcal{N}_{j-1}^{2i} , respectively, as illustrated in Fig. 2. For \mathcal{N}_j^i , $\alpha_j^i[k]$, $1 \leq k \leq 2^j$, indicates the k -th input logarithmic likelihood ratio (LLR) value, and $\beta_j^i[k]$, $1 \leq k \leq 2^j$, denotes the k -th output binary hard-valued message. For the AWGN channel, the received vector $\mathbf{y} = (y[1], y[2], \dots, y[N])$ from the channel can be used to calculate the channel LLR as $2\mathbf{y}/\sigma^2$, where σ^2 is the variance of the Gaussian noise. SC decoding starts by setting $\alpha_n^1[1:N] = 2\mathbf{y}/\sigma^2$. A node will be activated once all its inputs are available. When LLR messages pass through a node in the factor graph which is indicated by the \oplus sign,

the f function over the LLR domain is executed as

$$\alpha_{j-1}^{2i-1}[k] = f(\alpha_j^i[2k-1], \alpha_j^i[2k]), \quad (2)$$

and when LLR messages pass through a node in the factor graph which is indicated by the \boxplus sign, the g function over the LLR domain is executed as

$$\alpha_{j-1}^{2i}[k] = g(\alpha_j^i[2k-1], \alpha_j^i[2k], \beta_{j-1}^{2i-1}[k]), \quad (3)$$

where

$$f(x, y) = 2 \operatorname{arctanh}\left(\tanh\left(\frac{x}{2}\right) \tanh\left(\frac{y}{2}\right)\right), \quad (4)$$

$$g(x, y, u) = (-1)^u x + y. \quad (5)$$

The f function can be approximated as [34]

$$f(x, y) = \operatorname{sign}(x) \operatorname{sign}(y) \min(|x|, |y|). \quad (6)$$

When the LLR value of the k -th bit at level zero α_0^k , $1 \leq k \leq N$, is calculated, the estimation of $u[k]$, denoted as $\hat{u}[k]$, can be obtained as

$$\hat{u}[k] = \hat{\beta}_0^k = \begin{cases} 0, & \text{if } k \in \mathbb{A}^c, \\ \frac{1 - \operatorname{sign}(\alpha_0^k)}{2}, & \text{otherwise.} \end{cases} \quad (7)$$

The hard-valued messages are propagated back to the parent node as

$$\begin{aligned} & \hat{\beta}_j^i[k] \\ &= \begin{cases} \hat{\beta}_{j-1}^{2i-1}\left[\frac{k+1}{2}\right] \oplus \hat{\beta}_{j-1}^{2i}\left[\frac{k+1}{2}\right], & \text{if } \operatorname{mod}(k, 2) = 1, \\ \hat{\beta}_{j-1}^{2i}\left[\frac{k}{2}\right], & \text{if } \operatorname{mod}(k, 2) = 0. \end{cases} \end{aligned} \quad (8)$$

After traversing all the nodes in the SC decoding tree, $\hat{\mathbf{u}}$ contains the decoding result. Thus the latency of SC decoding for a polar code of length N in terms of the number of time steps can be represented by the number of nodes in the SC decoding tree as [1]

$$\mathcal{T}_{\text{SC}} = 2N - 2. \quad (9)$$

D. Fast SC Decoding

The SC decoding has strong data dependencies that limit the amount of parallelism that can be exploited within the algorithm because the estimation of each bit depends on the estimation of all previous bits. This leads to a large latency in the SC decoding algorithm. It was shown in [9] that for a node \mathcal{N}_j^i , $\beta_j^i[1:2^j]$ can be estimated without traversing the decoding tree as

$$\hat{\beta}_j^i[1:2^j] = \arg \max_{\beta_j^i[1:2^j] \in \mathbb{C}_j^i} \sum_{k=1}^{2^j} (-1)^{\beta_j^i[k]} \alpha_j^i[k], \quad (10)$$

where \mathbb{C}_j^i is the set of all the codewords associated with node \mathcal{N}_j^i . Multibit decoding can be performed directly in an intermediate level instead of bit-by-bit sequential decoding at level 0, in order to traverse fewer nodes in the SC decoding tree and consequently, to reduce the latency caused by data computation and exchange. However, the evaluation of (10)

generally requires exhaustive search over all the codewords in the set \mathbb{C}_j^i which is computationally intensive in practice. In [12], a fast SC (FSC) decoding algorithm was proposed which performs fast parallel decoding when specific special node types are encountered. The sequences of information and frozen bits of these node types have special bit-patterns. Therefore, they can be decoded more efficiently without the need for exhaustive search. Using the vector \mathbf{d} , the five special node types proposed in [12] are described as:

- Rate-0 node: all bits are frozen bits, $\mathbf{d} = (0, 0, \dots, 0)$.
- Rate-1 node: all bits are non-frozen bits, $\mathbf{d} = (1, 1, \dots, 1)$.
- REP node: all bits are frozen bits except the last one, $\mathbf{d} = (0, \dots, 0, 1)$.
- SPC node: all bits are non-frozen bits except the first one, $\mathbf{d} = (0, 1, \dots, 1)$.
- ML node: a code of length 4 with $\mathbf{d} = (0, 1, 0, 1)$.

By merging the aforementioned special nodes, five additional nodes were proposed, namely, the REP-SPC node, which is a REP node followed by a SPC node; the P-01/P-0SPC node, which is generated by merging a Rate-0 node with a Rate-1/SPC node; and the P-R1/P-RSPC node, which is generated by merging a g function branch operation in (5) with the following Rate-1/SPC node. In [19], five additional special node types and their corresponding fast decoders were introduced. This enhanced FSC decoding algorithm can achieve a lower decoding latency than FSC decoding. The five special node types are:

- Type-I node: all bits are frozen bits except the last two, $\mathbf{d} = (0, \dots, 0, 1, 1)$.
- Type-II node: all bits are frozen bits except the last three, $\mathbf{d} = (0, \dots, 0, 1, 1, 1)$.
- Type-III node: all bits are non-frozen bits except the first two, $\mathbf{d} = (0, 0, 1, \dots, 1)$.
- Type-IV node: all bits are non-frozen bits except the first three, $\mathbf{d} = (0, 0, 0, 1, \dots, 1)$.
- Type-V node: all bits are frozen bits except the last three and the fifth to last, $\mathbf{d} = (0, \dots, 0, 1, 0, 1, 1, 1)$.

A generalized FSC (GFSC) decoding algorithm was proposed in [20] by introducing the G-PC node and the G-REP node (previously identified as Multiple-G0 nodes in [23]). The G-PC node is a node at level j having all its descendants as Rate-1 nodes except the leftmost one at a certain level $r < j$, that is a Rate-0 node. The G-REP node is a node at level j for which all its descendants are Rate-0 nodes, except the rightmost one at a certain level $r < j$, which is a generic node of rate C (Rate-C).

In [18], in addition to the nodes in [12], three special node types are added: the REP1 node, with the REP node on the left and the Rate-1 node on the right; the OREPSPC node, which is a concatenation of the Rate-0, REP, and SPC nodes; and the 001 node, whose left 3/4 of bits are frozen bits and right 1/4 of bits are unfrozen bits.

To further reduce the latency, the operation merging scenarios are generalized in [22] and [23] where, in addition to merging special nodes, branch operation ((4),(5) and (8)) merging is also considered. Consequently, REP-REPSPC, REP-Rate1, and Rate0-ML nodes are adopted as the merging of special

nodes, and F-REP node is adopted as a node-branch merger, which is generated by performing an f function operation followed by a REP node. The following branch operation mergers are further introduced:

- $F^{\times 2}$: two consecutive f function operations in (4).
- $G0^{\times 2}$: two consecutive G0 operations, where G0 is the g function in (5) assuming $u = 0$.
- $C^{\times 2}, C^{\times 3}$: up to three consecutive operations in (8).
- $C0^{\times 2}, C0^{\times 3}$: up to three consecutive operations in (8), assuming the estimations from the left branch are all zeros.
- G-F: g function operation followed by an f function operation.
- F-G0: f function operation followed by a G0 operation.

The key advantage of using specific parallel decoders for the aforementioned special nodes is that, since the SC decoding tree is not traversed when one of these nodes is encountered, significant latency saving can be achieved. For example, if \mathcal{N}_j^i is a Rate-1 node, hard decision decoding can be used to immediately obtain the decoding result as

$$\hat{\beta}_j^i[k] = h(\alpha_j^i[k]) = \begin{cases} 0, & \text{if } \alpha_j^i[k] \geq 0, \\ 1, & \text{otherwise.} \end{cases} \quad (11)$$

If \mathcal{N}_j^i is a SPC node, hard decision based on (11) is first derived followed by the calculation of the parity of the output using modulo-2 addition. The index of the least reliable bit is found as

$$k' = \arg \min_k |\alpha_j^i[k]|. \quad (12)$$

Finally, the bits in a SPC node are estimated as

$$\hat{\beta}_j^i[k] = \begin{cases} h(\alpha_j^i[k]) \oplus \text{parity}, & \text{if } k = k', \\ h(\alpha_j^i[k]), & \text{otherwise.} \end{cases} \quad (13)$$

This operation can be performed in a single time step [19]. Finally, if \mathcal{N}_j^i is a G-PC node, the decoding can be viewed as a parallel decoding of several separate SPC nodes. The decoding of a G-PC node can generally be performed in one time step considering parallel SPC decoders [20].

All of the aforementioned fast SC decoding algorithms perform decoding at an intermediate level of the decoding tree in order to reduce the number of traversed nodes. A decoding algorithm that can decode a node at a higher level of the decoding tree generally results in more savings in terms of latency than one that decodes a node at a lower level of the decoding tree.

E. Threshold-Based Hard-Decision-Aided Scheme

For a binary AWGN channel with standard deviation σ_n , it was shown in [35] that, considering all the previous bits are decoded correctly, the LLR value $a_j^i[k]$, $1 \leq k \leq 2^j$, input into node \mathcal{N}_j^i can be approximated as a Gaussian variable using a Gaussian approximation as

$$\alpha_j^i[k] \sim \mathcal{N}(M_j^i[k], 2|M_j^i[k]|), \quad (14)$$

where $M_j^i[k]$ is the expectation of $\alpha_j^i[k]$ [36] such that

$$M_j^i[k] = \begin{cases} m_j^i, & \text{if } \beta_j^i[k] = 0, \\ -m_j^i, & \text{if } \beta_j^i[k] = 1, \end{cases} \quad (15)$$

and m_j^i can be calculated recursively offline assuming the all-zero codeword is transmitted as

$$m_n^1 = 2/\sigma_n^2, \quad m_{j-1}^{2i-1} = \varphi^{-1}(1 - [1 - \varphi(m_j^i)]^2), \\ m_{j-1}^{2i} = 2m_j^i, \quad (16)$$

where

$$\varphi(x) = \begin{cases} 1 - \frac{1}{\sqrt{4|x|}} \int_{-\infty}^{\infty} \tanh\left(\frac{u}{2}\right) e^{-\frac{(u-x)^2}{4|x|}} du, & x \neq 0, \\ 0, & x = 0. \end{cases} \quad (17)$$

It was shown in [32] that, when the magnitude of the LLR values at a certain node in the SC decoding tree is large enough, the node has enough reliability to perform hard decision directly at the node without tangibly altering the error-correction performance. To determine the reliability of the node, a threshold is defined in [32] as

$$T = c_t \log \frac{1 - \frac{1}{2} \operatorname{erfc}\left(0.5\sqrt{m_j^i}\right)}{\frac{1}{2} \operatorname{erfc}\left(0.5\sqrt{m_j^i}\right)}, \quad (18)$$

where $c_t \geq 1$ is a constant that is selected empirically. The hard-decision estimate of the received LLR values is calculated using

$$\text{HB}_j^i[k] = \begin{cases} 0, & \text{if } \alpha_j^i[k] > T, \\ 1, & \text{if } \alpha_j^i[k] < -T. \end{cases} \quad (19)$$

The issue with the method in [32] is that the threshold defined in (18) contains complex calculations of complementary error functions $\operatorname{erfc}(\cdot)$, making the corresponding calculation inefficient. Moreover, the hard-decision threshold is calculated empirically rather than for a desired error-correction performance and the threshold comparison in (19) is performed every time a node with no specific structure is encountered in the SC decoding process.

III. FAST SC DECODING WITH SEQUENCE REPETITION NODES

This section introduces a class of nodes that is at a higher level of the SC decoding tree and shows how parallel decoding at this class of nodes can be exploited to achieve significant latency savings in comparison with the state of the art.

A. Sequence Repetition (SR) Node

Let \mathcal{N}_j^i be a node at level j of the tree representation of SC decoding as shown in Fig. 2. An SR node is any node at stage j whose descendants are all either Rate-0 or REP nodes, except the rightmost one at a certain stage r , $0 \leq r \leq j$, that is a generic node of rate C . The structure of an SR node is depicted in Fig. 3. The rightmost node $\mathcal{N}_r^{i \times 2^{j-r}}$ at stage r is denoted as the source node of the SR node \mathcal{N}_j^i . Let $E = i \times 2^{j-r}$ so the source node can be denoted as \mathcal{N}_r^E .

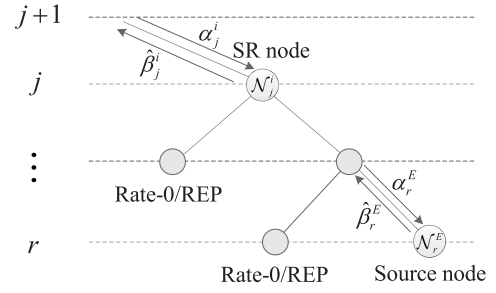


Fig. 3. General structure of SR Node.

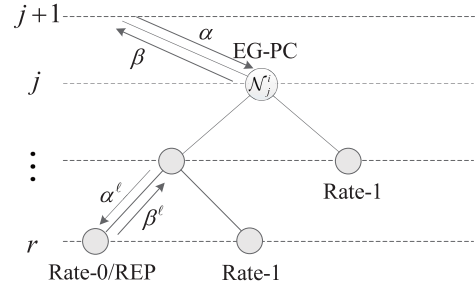


Fig. 4. General structure of Extended G-PC Node.

An SR node can be represented by three parameters as $\text{SR}(v, \text{SNT}, r)$, where r is the level of the SC decoding tree in which \mathcal{N}_r^E is located, SNT is the source node type, and $v = (v[j], v[j-1], \dots, v[r+1])$ is a vector of length $(j-r)$ such that for the left child node of the parent node of \mathcal{N}_r^E at level k , $r < k \leq j$, $v[k]$ is calculated as

$$v[k] = \begin{cases} 0, & \text{if the left child node is a Rate-0 node,} \\ 1, & \text{if the left child node is a REP node.} \end{cases} \quad (20)$$

Note that when $r = j$, \mathcal{N}_j^i is a source node and thus v is an empty vector denoted as $v = \emptyset$.

B. Source Node

To define the source node type, an extended class of G-PC (EG-PC) nodes is first introduced. The structure of the EG-PC node is depicted in Fig. 4. The EG-PC node is different from the G-PC node in its leftmost descendant node that can be either a Rate-0 or a REP node. The bits in an EG-PC node satisfy the following parity check constraint,

$$z = \bigoplus_{m=(k-1)2^{j-r}+1}^{k \times 2^{j-r}} \beta_j^i[m], \quad (21)$$

where $k \in \{1, \dots, 2^r\}$, and $z \in \{0, 1\}$ is the parity. Unlike G-PC nodes whose parity is always even ($z = 0$), the EG-PC node can have either even parity ($z = 0$) or odd parity ($z = 1$). The parity of the EG-PC node can be calculated as

$$z = \begin{cases} 0, & \text{if the leftmost node is Rate-0,} \\ h\left(\sum_{k=1}^{2^r} \alpha_k\right), & \text{otherwise,} \end{cases} \quad (22)$$

TABLE I
SR NODE REPRESENTATION OF COMMON NODE TYPES

Node Type	SR Node Representation	Length of v
Rate-0	$\text{SR}(\emptyset, \text{Rate-0}, j)$	0
REP	$\text{SR}(\emptyset, \dots, \emptyset, \text{Rate-1}, 0)$	j
SPC	$\text{SR}(\emptyset, \text{EG-PC}, j)$	0
Rate-1	$\text{SR}(\emptyset, \text{Rate-1}, j)$	0
P-01	$\text{SR}(\emptyset, \text{Rate-1}, j-1)$	1
P-0SPC	$\text{SR}(\emptyset, \text{EG-PC}, j-1)$	1
Type-I	$\text{SR}(\emptyset, \dots, \emptyset, \text{Rate-1}, 1)$	$j-1$
Type-II	$\text{SR}(\emptyset, \dots, \emptyset, \text{EG-PC}, 2)$	$j-2$
Type-III	$\text{SR}(\emptyset, \text{EG-PC}, j)$	0
Type-IV	$\text{SR}(\emptyset, \text{EG-PC}, j)$	0
Type-V	$\text{SR}(\emptyset, \dots, \emptyset, 1, \text{EG-PC}, 2)$	$j-2$
G-PC	$\text{SR}(\emptyset, \text{EG-PC}, j)$	0
G-REP	$\text{SR}(\emptyset, \dots, \emptyset, \text{Rate-C}, r)$	$j-r$
REP-SPC	$\text{SR}(1, \text{EG-PC}, 2)$	1
OREPSPC	$\text{SR}(\emptyset, 1, \text{EG-PC}, j-2)$	2
001	$\text{SR}(\emptyset, 0, \text{Rate-1}, j-2)$	2
REP-REPSPC	$\text{SR}(1, 1, \text{EG-PC}, j-2)$	2
Rate0-ML	$\text{SR}(\emptyset, 1, 0, \text{Rate-1}, 0)$	3
REP-Rate1(REP1)	$\text{SR}(1, \text{Rate-1}, j-1)$	1

where

$$\alpha_k = 2 \tanh^{-1} \left(\prod_{m=(k-1)2^{j-r}+1}^{k2^{j-r}} \tanh \left(\frac{\alpha_j^i[m]}{2} \right) \right).$$

After computing z , Wagner decoders [37] can be used to decode the 2^r SPC codes with either even or odd parity constraints. A Wagner decoder performs hard decisions on the bits, and if the parity does not hold, it flips the bit with the lowest absolute LLR value.

SPC, Type-III, Type-IV, and G-PC nodes can be represented as special cases of EG-PC nodes. As a result, most of the common special nodes can be represented as SR nodes, as shown in Table III-B. Note that the node-branch mergers like P-RSPC and F-REP, and branch operation mergers such as $F^{\times 2}$, G-F, $G0^{\times 2}$, F-G0, do not fit into the category of SR nodes. It is worth mentioning that other parameter choices for SR nodes may lead to the discovery of more types of special nodes that can be decoded efficiently, but this exploration is beyond the scope of this work.

C. Repetition Sequence

In this subsection, a set of sequences, called repetition sequences, is defined that can be used to calculate the output bit estimates of an SR node based on the output bit estimates of its source node. To derive the repetition sequences, v is used to generate all the possible sequences that have to be XORed with the output of the source node to generate the output bit estimates of the SR node. Let η_k denote the rightmost bit value of the left child node of the parent node of \mathcal{N}_r^E at level $k+1$. When $v[k+1] = 0$, the left child node is a Rate-0 node so $\eta_k = 0$. When $v[k+1] = 1$, the left child node is a REP node, thus η_k can take the value of either 0 or 1. The number of repetition sequences is dependent on the number of different values that η_k can take. Let W_v denote the number of '1's in v . The number of all possible repetition sequences is thus 2^{W_v} . Let $\mathbb{S} = \{s_1, \dots, s_{2^{W_v}}\}$ denote the set of all possible repetition sequences.

The output bits of SR node $\beta_j^i[1 : 2^j]$ have the property that their repetition sequence is repeated in blocks of length 2^{j-r} .

Let $\beta_r^E[1 : 2^r]$ denote the output bits of the source node of an SR node \mathcal{N}_j^i . The output bits for each block of length 2^{j-r} in \mathcal{N}_j^i with respect to the output bits of its source node can be written as

$$\beta_j^i[(k-1)2^{j-r}+1 : k2^{j-r}] = \beta_r^E[k] \oplus s_l, \quad (23)$$

where $k \in \{1, \dots, 2^r\}$ and $s_l = \{s_l[1], \dots, s_l[2^{j-r}]\}$ is the l -th repetition sequence in \mathbb{S} . To obtain the repetition sequence s_l and with a slight abuse of terminology and notation for convenience, the Kronecker sum operator \boxplus is used, which is equivalent to the Kronecker product operator, except that addition in GF(2) is used instead of multiplication. For each set of values that η_k 's can take, s_l can be calculated as

$$s_l = (\eta_r, 0) \boxplus (\eta_{r+1}, 0) \boxplus \dots \boxplus (\eta_{j-1}, 0). \quad (24)$$

Example 1 (Repetition sequences for $\text{SR}((1, 1), \text{EG-PC}, 2)$): Consider an example in which the SR node \mathcal{N}_4^1 is located at level 4 of the decoding tree and its source node \mathcal{N}_2^4 is an EG-PC node located at level 2. Since $v = (1, 1)$, $W_v = 2$ and $|\mathbb{S}| = 4$. For $\eta_1 \in \{0, 1\}$ and $\eta_2 \in \{0, 1\}$,

$$\begin{aligned} s_1 &= (0, 0) \boxplus (0, 0) = (0, 0, 0, 0), \\ s_2 &= (1, 0) \boxplus (0, 0) = (1, 1, 0, 0), \\ s_3 &= (0, 0) \boxplus (1, 0) = (1, 0, 1, 0), \\ s_4 &= (1, 0) \boxplus (1, 0) = (0, 1, 1, 0). \end{aligned}$$

For a polar code with a given d , the locations of SR nodes in the decoding tree are fixed and can be determined off-line. Therefore, the repetition sequences in \mathbb{S} of all of the SR nodes can be pre-computed and used in the course of decoding.

D. Decoding of SR Nodes

To decode SR nodes, the LLR values $\alpha_{r_l}^E[1 : 2^r]$ of the source node \mathcal{N}_r^E are calculated based on the LLR values $\alpha_j^i[1 : 2^j]$ of the SR node \mathcal{N}_j^i for every repetition sequence s_l as follows:

Proposition 1: Let $\alpha_j^i[1 : 2^j]$ be the LLR values of the SR node \mathcal{N}_j^i and $\alpha_{r_l}^E[1 : 2^r]$ be the LLR values of its source node \mathcal{N}_r^E associated with the l -th repetition sequence s_l . For $k \in \{1, \dots, 2^r\}$ and $l \in \{1, \dots, 2^{W_v}\}$,

$$\alpha_{r_l}^E[k] = \sum_{m=1}^{2^{j-r}} \alpha_j^i[(k-1)2^{j-r}+m] (-1)^{s_l[m]}. \quad (25)$$

Proof: See Appendix A. ■

Using (23) and (25), (10) can be written as

$$\begin{aligned} \hat{\beta}_j^i[1 : 2^j] &= \arg \max_{\beta_j^i[1:2^j] \in \mathbb{C}_j^i} \sum_{k=1}^{2^j} (-1)^{\beta_j^i[k]} \alpha_j^i[k] \\ &= \arg \max_{\substack{\beta_r^E[1:2^r] \in \mathbb{C}_r^E \\ s_l \in \mathbb{S}}} \sum_{k=1}^{2^r} (-1)^{\beta_r^E[k]} \sum_{m=1}^{2^{j-r}} \alpha_{r_l}^i[(k-1)2^{j-r}+m] (-1)^{s_l[m]} \\ &= \arg \max_{\substack{\beta_r^E[1:2^r] \in \mathbb{C}_r^E \\ l \in \{1, \dots, |\mathbb{S}|\}}} \sum_{k=1}^{2^r} (-1)^{\beta_r^E[k]} \alpha_{r_l}^E[k]. \end{aligned} \quad (26)$$

Thus, the bit estimates of an SR node $\hat{\beta}_j^i[1:2^j]$ can be calculated by finding the bit estimates of its source node $\beta_r^E[1:2^r]$ using (26) and the repetition sequences as shown in (23).

The decoding algorithm of an SR node \mathcal{N}_j^i is described in Algorithm 1. It first computes $\alpha_{r_l}^E$ for $l \in \{1, \dots, |\mathcal{S}|\}$ and generates $|\mathcal{S}|$ new paths by extending the decoding path at the $j-r$ rightmost bits corresponding to $\eta_r, \eta_{r+1}, \dots, \eta_{j-1}$. Note that the l -th path is generated when the repetition sequence is s_l and $\alpha_{r_l}^E$, $\hat{\beta}_{r_l}^E$, and $\hat{\beta}_{j_l}^i$ are its soft and hard messages. Then, the source node is decoded under the rule of the SC decoding. If the source node is a special node, a hard decision is made directly. Parity check and bit flipping will be performed further using Wagner decoder if the source node is an EG-PC node. Finally, the optimal decoding path index can be selected according to the comparison in (27) below and the decoding result is obtained using (23).

Algorithm 1 Decoding Algorithm of SR Node \mathcal{N}_j^i

Input: $\alpha_j^i[1:2^j], \mathcal{S}$

Output: $\hat{\beta}_j^i[1:2^j]$

1) Soft message computation

for $l \in \{1, \dots, |\mathcal{S}|\}$ **do**

 Calculate $\alpha_{r_l}^E$ according to (25).

2) Decoding of source node \mathcal{N}_r^E

for $l \in \{1, \dots, |\mathcal{S}|\}$ **do**

if $SNT=Rate-C$ **then**

 Decode source node \mathcal{N}_r^E using $\alpha_{r_l}^E$ and obtain $\hat{\beta}_{r_l}^E$;

else

if $SNT=Rate-0$ **then**

$\hat{\beta}_{r_l}^E[k] = 0, k \in \{1, \dots, 2^r\}$;

else // $SNT=Rate-1$ or $SNT=EG-PC$

$\hat{\beta}_{r_l}^E[k] = h(\alpha_{r_l}^E[k]), k \in \{1, \dots, 2^r\}$.

if $SNT=EG-PC$ **then**

 Perform parity check and bit flipping on $\hat{\beta}_{r_l}^E$ using $\alpha_{r_l}^E$.

3) Comparison and path selection

$$\hat{l} = \arg \max_{l \in \{1, \dots, |\mathcal{S}|\}} \sum_{k=1}^{2^r} |\alpha_{r_l}^E[k]|. \quad (27)$$

Return $\hat{\beta}_{j_l}^i$ to parent node according to (23).

Based on Algorithm 1, the SR node-based fast SC (SRFSC) decoding algorithm is proposed. It follows the SC decoding algorithm schedule until an SR node is encountered where Algorithm 1 is executed. Note that the $|\mathcal{S}|$ paths can be processed simultaneously and the path selection operation in step 3 of Algorithm 1 can be performed in parallel with the decoding of the source node in step 2 and the following g function operation. Once the selected index \hat{l} is obtained, only the l -th decoding path corresponding to s_l is retained and the remaining paths are deleted.

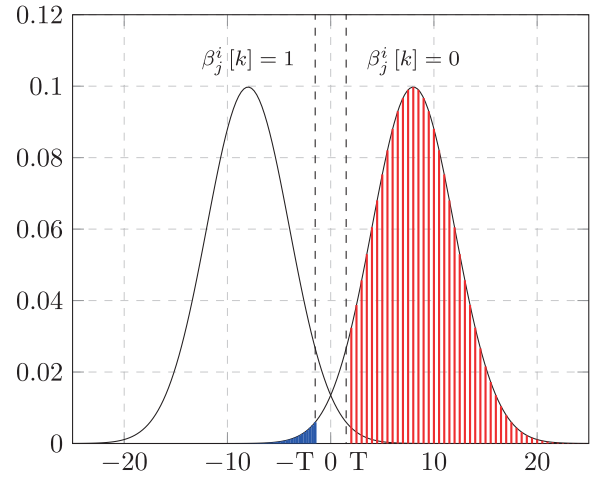


Fig. 5. Probability distribution of $\alpha_j^i[k]$ under Gaussian approximation for $m_j^i = 8$. The red dashed area represents the probability of correct hard decision and the blue solid area represents the probability of incorrect hard decision when $\beta_j^i[k] = 0$.

IV. HARD-DECISION-AIDED FAST SC DECODING WITH SEQUENCE REPETITION NODES

In this section, a novel threshold-based hard-decision-aided scheme is proposed, which speeds up the decoding of general nodes with no specific structure in the SRFSC decoding at high signal-to-noise ratios. Consequently, the *threshold-based hard-decision-aided SRFSC* (TA-SRFSC) decoding algorithm is proposed. In addition, a multi-stage decoding strategy is introduced to mitigate the possible error-correction performance loss of the proposed TA-SRFSC decoding.

A. Proposed Threshold-Based Hard-Decision-Aided Scheme

In a binary AWGN channel, the LLR value $\alpha_j^i[k]$ of any node \mathcal{N}_j^i can be approximated as a Gaussian variable as shown in Fig. 5. The red area in Fig. 5 represents the approximate probability of correct hard decision \tilde{P}_c and the blue area represents the approximate probability of incorrect hard decision \tilde{P}_e when $\beta_j^i[k] = 0$ such that

$$\tilde{P}_c = Q\left(\frac{T - m_j^i}{\sqrt{2m_j^i}}\right), \quad \tilde{P}_e = Q\left(\frac{T + m_j^i}{\sqrt{2m_j^i}}\right), \quad (28)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$. The area between the two dashed lines represents the approximate probability that a hard decision is not performed.

To simplify the calculation of the threshold, we take a different approach than [32] by using the Gaussian distribution of $\alpha_j^i[k]$ and constraining the approximate probability of error when $\beta_j^i[k] = 0$ to be

$$\tilde{P}_e = Q\left(\frac{T + m_j^i}{\sqrt{2m_j^i}}\right) < Q(c), \quad (29)$$

where c (and thus $Q(c)$) is a positive constant, whose selection method will be given in Observation 1. This is equivalent to

$T > -m_j^i + c\sqrt{2m_j^i}$. Therefore, the threshold is

$$T = \left| -m_j^i + c\sqrt{2m_j^i} \right|, \quad (30)$$

where the absolute value ensures T is positive for all values of $m_j^i > 0$ with any $c > 0$.

Under the Gaussian approximation in (14), for a node that undergoes hard decision in (19), the proposed threshold leads to a bounded probability of correct hard decision as shown in the following observation. Note that the analytical derivations are accurate under the assumption of Gaussian approximation.

Observation 1: Let $0.5 < \varepsilon < 1$ and $c > 0$ be real numbers such that

$$Q(c) \leq 1 - \sqrt[n]{\varepsilon}. \quad (31)$$

Performing hard decision in (19) with the threshold in (30) on nodes \mathcal{N}_j^i whose m_j^i satisfy

$$m_j^i \geq \frac{1}{2} \left[c - Q^{-1} \left(\frac{Q(c)}{\frac{1}{\sqrt[n]{\varepsilon}} - 1} \right) \right]^2, \quad (32)$$

has a probability of correct hard decision that is lower bounded by $\sqrt[n]{\varepsilon}$, under the assumption of Gaussian approximation (14) and assuming all prior bits are decoded correctly.

Explanation: See Appendix B.

The proposed TA scheme performs hard decision on a node \mathcal{N}_j^i only if (32) is satisfied. Furthermore, a hard decision is performed on a node if all of its input LLR values $\alpha_j^i[k]$ satisfy (19). Otherwise, standard SC decoding is applied on \mathcal{N}_j^i to obtain the decoding result. Compared to the TA scheme in [32], the threshold values of both methods can be computed off-line. However, the proposed TA scheme in this article has the following three advantages: 1) The proposed method can provide a higher latency reduction than the best case in [32] as shown in the simulation results; 2) The effect of the proposed threshold values on the error-correction performance is approximately predictable according to Observation 1; 3) According to (32), only a fraction of nodes undergo hard decision in the decoding process of the proposed TA scheme, which avoids unnecessary threshold comparisons. To speed up the decoding process, the proposed TA scheme is combined with SRFSC decoding that results in the TA-SRFSC decoding algorithm. In TA-SRFSC decoding, when one of the special nodes considered in SRFSC decoding is encountered, SRFSC decoding is performed and when a general node with no special structure is encountered, the proposed TA scheme is applied. The following observation provides an approximate upper bound on the BLER of the proposed TA-SRFSC decoding.

Observation 2: Let $\text{BLER}_{\text{TA-SRFSC}}$ and $\text{BLER}_{\text{SRFSC}}$ denote the BLER of the TA-SRFSC decoding and the SRFSC decoding respectively. We can have the following approximate inequality

$$\text{BLER}_{\text{TA-SRFSC}} \lesssim 1 - \varepsilon (1 - \text{BLER}_{\text{SRFSC}}). \quad (33)$$

Explanation: See Appendix C.

Observation 2 provides a method to derive the threshold value for a desired upper bound of the BLER for TA-SRFSC decoding approximately. In fact, a large threshold value results in a better error-correction performance than a small threshold value at the cost of lower decoding speed. Therefore, a trade-off between the error-correction performance and the decoding speed can be achieved with the proposed TA-SRFSC decoding algorithm.

B. Multi-Stage Decoding

To mitigate the possible error-correction performance loss of the proposed TA-SRFSC decoding, a multi-stage decoding strategy is adopted in which a maximum of two decoding attempts is conducted. In the first decoding attempt, TA-SRFSC decoding is used. If this decoding fails and if there existed a node that underwent hard decision in this first decoding attempt, then a second decoding attempt using SRFSC decoding is conducted. To determine if the TA-SRFSC decoding failed, a cyclic redundancy check (CRC) is concatenated to the polar code and it is verified after the TA-SRFSC decoding. Additionally, the CRC is verified after the second decoding attempt by the SRFSC decoding to determine if the overall decoding process succeeded.

As shown in the next section, most of the received frames are decoded correctly by the proposed TA-SRFSC decoding in the first decoding attempt. As a result, the average decoding latency of the proposed multi-stage SRFSC decoding is very close to that of the TA-SRFSC decoding with CRC bits. However, the error-correction performance of the proposed multi-stage SRFSC decoding algorithm is slightly worse than that of SRFSC decoding due to the addition of CRC bits. As such, multi-stage SRFSC decoding trades off a slight degradation in error-correction performance to obtain a significant reduction in the average decoding latency. It is worth noting that, in order to improve the error-correction performance of the proposed scheme, the proposed multi-stage decoding strategy can be generalized to have more than two decoding attempts. In this scenario the first two attempts are the same as described above, i.e., TA-SRFSC decoding is used first followed by SRFSC decoding.¹ The third and any subsequent attempts would use increasingly more powerful decoding techniques than the first two, such as SCL decoding.

V. DECODING LATENCY

In this section, the decoding latency of the proposed fast decoders is analyzed under two cases: 1) with no resource limitation and 2) with hardware resource constraints. The former is to facilitate the comparisons with the works that do not consider hardware implementation constraints [19], [20], and the latter is for comparison with those that do [12], [18], [23].

¹If the TA-SRFSC decoder does not satisfy the CRC, the SRFSC decoding either should start from the beginning or from the first node which underwent the hard decision. In this article, the SRFSC decoder starts decoding from the first bit to avoid storing intermediate LLRs.

A. No Resource Limitation

In this case, the same assumptions as in [1], [19] are used and the decoding latency is measured in terms of the number of required time steps. More specifically: 1) there is no resource limitation so that all the parallelizable instructions are performed in one time step, 2) bit operations are carried out instantaneously, 3) addition/subtraction of real numbers and check-node operation consume one time step, and 4) Wagner decoding can be performed in one time step.

1) *SRFSC and TA-SRFSC*: For any node \mathcal{N}_j^i that has no special structure, e.g. a parent node of an SR node, and that satisfies (32), the threshold comparison in (19) is performed in parallel with the calculation of the LLR values of its left child node. Therefore, the proposed hard decision scheme does not affect the latency requirements for the nodes that undergo hard decision.

The number of time steps required for the decoding of the SR node is calculated according to Algorithm 1. In Step 1, the calculation of the LLR values for the source node requires one time step if $\mathbf{v} \neq \emptyset$. If $\mathbf{v} = \emptyset$, then the LLR values of the source node are available immediately. Thus, the required number of time steps for Step 1 is

$$\mathcal{T}_1 = \begin{cases} 0, & \text{if } \mathbf{v} = \emptyset, \\ 1, & \text{if } \mathbf{v} \neq \emptyset. \end{cases} \quad (34)$$

The time step requirement of Step 2 depends on the source node type. If SNT = Rate-C, the time step requirement of Step 2 is the time step requirement of the Rate-C node. If SNT = Rate-0 or Rate-1, then there is no latency overhead in Step 2. If SNT = EG-PC and in accordance with Section III-B, z can be estimated in two time steps if the leftmost node is a REP node (one time step for performing the check-node operation and one time step for adding the LLR values). Also, Wagner decoding can be performed in parallel with the estimation of z assuming $z = 0$ or $z = 1$. As such, at most two time steps are required for parity check and bit flipping of the EG-PC node. The required number of time steps for Step 2 is

$$\mathcal{T}_2 = \begin{cases} 0, & \text{if SNT = Rate-0 or Rate-1,} \\ 1 \text{ or } 2, & \text{if SNT = EG-PC,} \\ 2^{r+1} - 2, & \text{if SNT = Rate-C.} \end{cases} \quad (35)$$

Step 3 consumes two time steps using an adder tree and a comparison tree if $|\mathbb{S}| > 1$ and no time steps otherwise. Thus, the number of required time steps for Step 3 is

$$\mathcal{T}_3 = \begin{cases} 0, & \text{if } |\mathbb{S}| = 1, \\ 2, & \text{if } |\mathbb{S}| > 1. \end{cases} \quad (36)$$

Since path selection in Step 3 can be executed in parallel with the decoding of source node in Step 2 and the following g function calculation, the total number of time steps required to decode an SR node can be expressed as

$$\mathcal{T}_{\text{SR}} = \mathcal{T}_1 + \max(\mathcal{T}_2, \mathcal{T}_3 - 1), \quad (37)$$

where $\mathcal{T}_3 - 1$ indicates that at least one time step in \mathcal{T}_3 can be reduced by parallelizing Step 3 and the g function

calculation. Therefore, \mathcal{T}_{SR} is a variable that is dependent on its parameters. However, with a given polar code, the total number of time steps required for the decoding of the polar code using SRFSC decoding is fixed, regardless of the channel conditions.

2) *Multi-Stage SRFSC*: Let $\mathcal{T}_{\text{TA-SRFSC}_{\text{crc}}}$ and $\mathcal{T}_{\text{SRFSC}_{\text{crc}}}$ denote the average decoding latency of the proposed TA-SRFSC decoding and the SRFSC decoding using CRC bits, respectively. The average decoding latency of multi-stage SRFSC decoding, $\mathcal{T}_{\text{Multi-stage SRFSC}}$, is

$$\mathcal{T}_{\text{Multi-stage SRFSC}} = \mathcal{T}_{\text{TA-SRFSC}_{\text{crc}}} + P_{\text{Re-decoding}} \mathcal{T}_{\text{SRFSC}_{\text{crc}}}, \quad (38)$$

where $P_{\text{Re-decoding}}$ indicates the probability that TA-SRFSC decoding fails and there is at least one node that undergoes hard decision. Note that $P_{\text{Re-decoding}}$ is less than or equal to the probability that the output of TA-SRFSC decoding fails the CRC verification, which can be approximated by $\text{BLER}_{\text{TA-SRFSC}_{\text{crc}}}$. The approximation is due to the fact that the undetected error probability of CRC is negligible when its length is long enough [38]. In accordance with Observation 2, the average decoding latency requirement for the proposed multi-stage SRFSC decoding is

$$\begin{aligned} \mathcal{T}_{\text{Multi-stage SRFSC}} &\lesssim \mathcal{T}_{\text{TA-SRFSC}_{\text{crc}}} + (1 - \varepsilon(1 - \text{BLER}_{\text{SRFSC}_{\text{crc}}})) \mathcal{T}_{\text{SRFSC}_{\text{crc}}}. \end{aligned} \quad (39)$$

Since the decoding latency of SRFSC decoding is fixed, the average decoding latency and the worst case decoding latency of SRFSC decoding are equivalent. The worst case decoding latency of the proposed TA-SRFSC decoding can be calculated when none of the nodes in the decoding tree undergo hard decision. This occurs when the channel has a high level of noise. Thus, the worst case decoding latency of the proposed TA-SRFSC decoding is equivalent to the decoding latency of the SRFSC decoding. Moreover, when the channel is too noisy, $P_{\text{Re-decoding}} \approx 0$, because almost none of the nodes undergo hard decision. Thus, the worst case decoding latency of the proposed multi-stage SRFSC decoding is equivalent to the worst case decoding latency of TA-SRFSC decoding using a CRC, which is the latency of SRFSC decoding with a CRC.

B. With Hardware Resource Constraints

In this case, the decoding latency is measured in terms of the number of clock cycles in the actual hardware implementation. Contrary to Section V-A, a time step may take more than one clock cycle in a realistic hardware implementation. First, the idea of a semi-parallel decoder in [39] is adopted in which at most P processing elements can run at the same time. Thus, parallel f function or g function operations on more than P processing elements require more than one clock cycle. Second, the adder tree used for the calculation of LLR values in step 1 of Algorithm 1, and the compare-and-select (CS) tree used by the Wagner decoder to find the index of the least reliable input bit in step 2 of Algorithm 1, both

TABLE II
THE NUMBER OF SR NODES WITH DIFFERENT $|\mathcal{S}|$ AND THE
NUMBER OF GENERAL NODES IN 5G POLAR CODES OF
LENGTHS $N \in \{128, 512, 1024\}$ AND
RATES $R \in \{1/4, 1/2, 3/4\}$

N	R	SR					General	
		$ \mathcal{S} $					Total	Total
		1	2	4	8	16		
128	1/4	1	0	2	1	0	4	3
	1/2	4	3	1	0	0	8	7
	3/4	8	2	0	0	0	10	9
512	1/4	12	2	2	1	0	17	16
	1/2	15	5	2	0	1	23	22
	3/4	13	5	1	0	1	20	19
1024	1/4	17	6	2	2	1	28	27
	1/2	25	8	2	3	1	39	38
	3/4	29	8	2	1	0	40	39

require pipeline registers to reduce the critical path delay and improve the operating frequency and the overall throughput. The addition of these pipeline registers increases the number of decoding clock cycles. The exact number depends on the implementation and the given polar code.

VI. RESULTS AND COMPARISON

In this section, the average decoding latency and the error-correction performance of the proposed decoding algorithms are analyzed and compared with state-of-the-art fast SC decoding algorithms. To derive the results, polar codes of length $N \in \{128, 512, 1024\}$, which are adopted in the 5G standard [40], are used. To do a fair comparison, for baseline decoding algorithms without hardware implementation, the latency is calculated under the assumptions in Section V-A. Otherwise, the hardware implementation results are reported.

To simulate the effect of ε on the error-correction performance and the latency of the proposed decoding algorithms, the assumption in [32] is adopted and three values of $\varepsilon \in \{0.9, 0.99, 0.999\}$ are selected. In accordance with (31), $c \geq 3.8$ for $\varepsilon = 0.9$, $c \geq 4.3$ for $\varepsilon = 0.99$, and $c \geq 4.8$ for $\varepsilon = 0.999$. According to (29), with the increasing of c , \tilde{P}_e decreases. At the same time, fewer nodes undergo hard decision and the decoding latency increases. To get a good trade-off between error-correction performance and latency, we set $c = 3.8$ for $\varepsilon = 0.9$, $c = 4.3$ for $\varepsilon = 0.99$, and $c = 4.8$ for $\varepsilon = 0.999$. Consequently, $m_j^i \geq 9.3891$ for $\varepsilon = 0.9$, $m_j^i \geq 14.7255$ for $\varepsilon = 0.99$, and $m_j^i \geq 16.1604$ for $\varepsilon = 0.999$ in accordance with (32). Using these values, the threshold T defined in (30) and the BLER upper bound for the TA-SRFSC decoding in (33) can be calculated for different values of ε .

Table II reports the number of SR nodes with different $|\mathcal{S}|$, the total number of SR nodes, and the total number of general nodes with no special structure, at different code lengths and rates. It can be seen that when the code length is 128, 512, and 1024, the codes with rate 1/2 have the largest proportion of nodes with $|\mathcal{S}| > 1$, respectively. This in turn results in more latency savings because a higher degree of parallelism can be exploited with these nodes. Thanks to the binary tree structure and the fact that SR nodes are always present at an

TABLE III
NODE LENGTH OF SR NODES WITH DIFFERENT $|\mathcal{S}|$ IN 5G POLAR CODES
OF LENGTHS $N \in \{128, 512, 1024\}$ AND RATES $R = 1/2$

N	Length	$ \mathcal{S} $					
		1	2	4	8	16	
128	8	2	2	0	0	0	
	16	0	1	1	0	0	
	32	2	0	0	0	0	
512	8	7	3	0	0	0	
	16	4	1	2	0	0	
	32	3	1	0	0	0	
	64	1	0	0	0	0	
	128	0	0	0	0	1	
1024	8	10	6	0	0	0	
	16	7	1	2	0	0	
	32	4	0	0	3	0	
	64	2	1	0	0	1	
	128	2	0	0	0	0	

intermediate level of the decoding tree, the number of traversed general nodes with no special structure equals the total number of SR nodes minus one, since they are the parent nodes of SR nodes. Table III shows the length of SR nodes with different $|\mathcal{S}|$ at different code lengths when $R = 1/2$. The length of the SR nodes in the decoding tree corresponds to the level in the decoding tree that they are located. SR nodes with larger $|\mathcal{S}|$ that are located on a higher level of the decoding tree contribute more in the overall latency reduction.

Table IV reports the number of time steps required to decode polar codes of lengths $N \in \{128, 512, 1024\}$ and rates $R \in \{1/4, 1/2, 3/4\}$ with the proposed SRFSC decoding algorithm, and compares it with the required number of time steps of the decoders in [12], [18]–[20], [23]. Note that the decoder in [23] has the minimum required time steps amongst all the previous works considered in Table IV. Three versions of the proposed SRFSC decoder are considered in Table IV: the SRFSC decoder that only utilizes SR nodes; the SRFSC' decoder that considers SR nodes and P-OSPC nodes that is adopted in all other baseline decoders in Table IV; and the SRFSC'' decoder that uses SR nodes, P-OSPC nodes, and the operation mergers $F^{\times 2}$ and G-F. It can be seen that the SRFSC'' decoder requires fewer time steps with respect to other decoders, except for the case of $N = 128$ and $R = 3/4$, which has a frequent occurrence of REP-SPC nodes that provides an advantage for [23]. This is due to the fact that the REP-SPC decoder in other works decodes REP and SPC nodes in parallel, consuming only a single time step, while two time steps are needed to decode the REP-SPC node in the proposed SRFSC decoder.

Table V compares the required number of clock cycles and the maximum operating frequency in the hardware implementation of different decoders for a polar code with $N = 1024$, $R \in \{1/2, 1/4, 3/4\}$, and $P = 64$ for the proposed SRFSC decoding algorithm (taken from [41]) and the decoders in [12], [18], [22], [23]. All FPGA results are for an Altera Stratix IV EP4SGX530KH40C2 FPGA device and all ASIC results use TSMC 65nm CMOS technology. The required number of clock cycles for the proposed SRFSC decoder can be further reduced if we consider node-branch merging and branch

TABLE IV
NUMBER OF TIME STEPS FOR DIFFERENT FAST SC DECODING ALGORITHMS OF POLAR CODES
OF LENGTHS $N \in \{128, 512, 1024\}$ AND RATES $R \in \{1/4, 1/2, 3/4\}$

N	R	[12]	[18]	[19]	[20]	[23]	SRFSC	SRFSC'	SRFSC''
128	1/4	25	23	24	23	10	13	12	10
	1/2	33	31	24	24	21	25	24	20
	3/4	22	22	22	22	16	29	25	20
512	1/4	73	70	63	63	42	57	52	40
	1/2	87	83	77	77	56	72	66	52
	3/4	79	73	64	64	50	63	56	45
1024	1/4	122	115	110	108	68	92	85	66
	1/2	156	144	138	138	94	127	115	90
	3/4	138	132	116	116	89	123	111	86

TABLE V
REQUIRED NUMBER OF CLOCK CYCLES AND MAXIMUM OPERATING
FREQUENCY FOR DIFFERENT DECODING ALGORITHMS OF
POLAR CODES OF LENGTH $N = 1024$ WITH RATE
 $R = \{1/4, 1/2, 3/4\}$ AND $P = 64$

	clock cycles			f_{\max} (MHz)	
	R				
	1/4	1/2	3/4	FPGA	ASIC
SRFSC	186	222	200	109.6	—
SRFSC*	155	191	166	109.6	—
[23]	168	198	181	—	430
[22]	189	214	185	89.6	420
[18]	221	252	216	—	450
[12]	225	270	225	99.8	450

operation merging similar to [23]. We have verified that the introduction of P-RSPC node to the proposed SRFSC decoder lengthens the critical path of the decoder, but the merging of branch operations, $F^{\times 2}$ and G-F, does not have an effect on the operating frequency. Thus, an enhanced SRFSC decoder with merging branch operations $F^{\times 2}$ and G-F is implemented and denoted as SRFSC* in Table V. It can be seen that the SRFSC* decoder requires a smaller number of clock cycles compared to other decoders and the reduction with respect to [23] is 8%, 4%, and 8% at $R \in \{1/4, 1/2, 3/4\}$, respectively. Moreover, the SRFSC decoder and the SRFSC* decoder both have the highest maximum operating frequency f_{\max} when implemented on the FPGA. Compared with the works that reported f_{\max} results for FPGA implementations, our decoders achieve an improvement of at least 10%.

Fig. 6 shows the BLER performance of different decoding algorithms when $N = 1024$ and $R = 1/2$, for different values of energy per bit to noise power spectral density ratio (E_b/N_0). For each value of ε , the BLER of TA-SRFSC decoding is depicted together with the upper bound calculated in Observation 2. It can be seen that the introduction of the TA scheme results in BLER performance loss for the proposed TA-SRFSC decoding with respect to SC and SRFSC decoding, especially at higher values of E_b/N_0 . The simulations confirm that the BLER curves of TA-SRFSC decoding fall below their respective upper bounds. It can also be seen that, as the E_b/N_0 value increases beyond a specific point, the BLER performance of the TA-SRFSC decoding degrades. This is because of the difference in the performance of hard-decision and soft-decision decoding. In accordance with (32), more

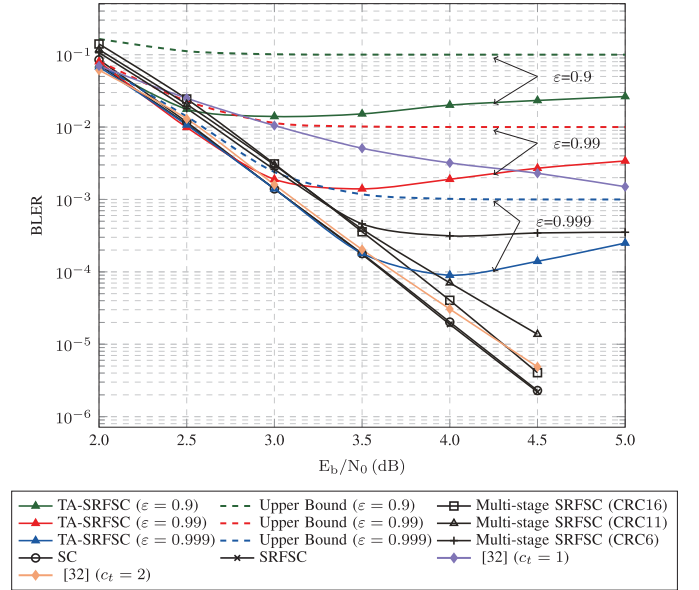


Fig. 6. BLER performance of different decoding algorithms for the 5G polar code of length $N = 1024$ and rate $R = 1/2$.

nodes undergo hard decision decoding for larger values of E_b/N_0 . Therefore, while the channel conditions improve, the hard decision decoding introduces errors that reduce the error-correction performance gain associated with these large E_b/N_0 values. As a result, the BLER performance degrades after a certain value of E_b/N_0 . This phenomenon exists as long as there are nodes that can undergo hard decision decoding. After all the nodes are decoded using hard decision, the BLER performance improves again as E_b/N_0 increases.

The proposed multi-stage SRFSC decoder is implemented using three different CRC lengths that are adopted in 5G to identify whether the decoding succeeded or failed: the CRC of length 6 (CRC6) with generator polynomial $D^6 + D^5 + 1$, the CRC of length 11 (CRC11) with generator polynomial $D^{11} + D^{10} + D^9 + D^5 + 1$, and the CRC of length 16 (CRC16) with generator polynomial $D^{16} + D^{12} + D^5 + 1$. For all values of ε , the proposed multi-stage SRFSC decoding results in almost the same BLER performance. Therefore, only the curve with $\varepsilon = 0.9$ is plotted in Fig. 6 for the multi-stage SRFSC decoding. CRC16 provides a better error-correction performance compared to CRC6 and CRC11, especially at high values of E_b/N_0 , due to high undetected probability

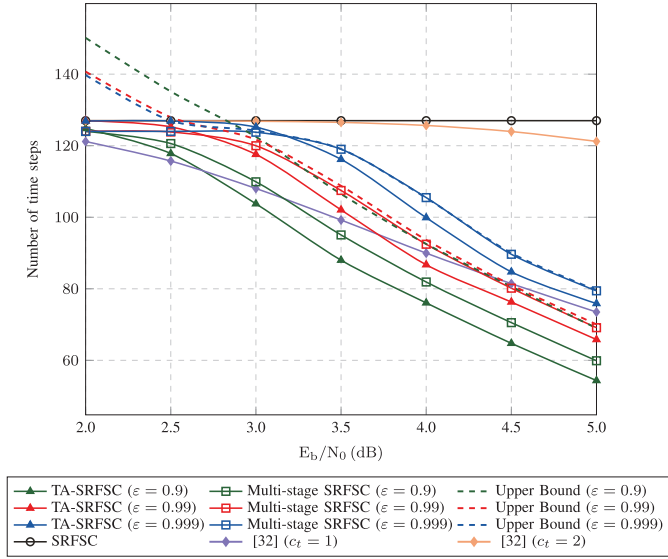


Fig. 7. Average decoding latency of different decoding algorithms for the 5G polar code of length $N = 1024$ and rate $R = 1/2$.

of error for short CRC lengths. Hence, CRC16 is selected for the polar code of length $N = 1024$ in this article. It can be seen that the multi-stage SRFSC decoding with CRC16 demonstrates a slight performance loss compared to the conventional SC and SRFSC decoders as a result of adding extra CRC bits. For polar codes of other lengths and rates, a similar trend can also be observed when comparing the BLER performance of different schemes.

Fig. 7 presents the average decoding latency in terms of the required number of time steps for the proposed TA-SRFSC decoding and multi-stage decoding with CRC16. In particular, it compares them with the latency of SRFSC decoding and the decoder in [32] at different values of E_b/N_0 when $N = 1024$ and $R = 1/2$. It can be seen that the required number of time steps for the proposed TA-SRFSC decoding decreases as E_b/N_0 increases and is reduced by 40% for $\epsilon = 0.999$, by 48% for $\epsilon = 0.99$, and by 57% for $\epsilon = 0.9$, compared to SRFSC decoding at $E_b/N_0 = 5$ dB. In addition, the proposed multi-stage SRFSC decoding reduces the required number of time steps by 37% for $\epsilon = 0.999$, by 46% for $\epsilon = 0.99$, and by 53% for $\epsilon = 0.9$, with respect to SRFSC decoding at $E_b/N_0 = 5$ dB. The required number of time steps for the proposed multi-stage SRFSC decoding outperforms the method in [32] with $c_t = 1$ by 19% for $\epsilon = 0.9$ at $E_b/N_0 = 5$ dB while providing a significantly better BLER performance. Fig. 7 also presents the approximate upper bound derived in (39). It can be seen in the figure that the upper bound in (39) becomes tighter as ϵ increases.

Fig. 8 compares the average number of threshold comparisons in (19) for the proposed TA-SRFSC decoder with $\epsilon \in \{0.9, 0.99, 0.999\}$, and the decoder in [32] with $c_t \in \{1, 2\}$ for the 5G polar code of length $N = 1024$ and rate $R = 1/2$. It can be seen that the proposed TA-SRFSC decoder shows significant benefits with respect to the decoder of [32] in terms of the average number of threshold comparisons. The TA-SRFSC decoder with $\epsilon = 0.9$ provides at least 36% reduction with respect to [32] with $c_t = 1$ while having

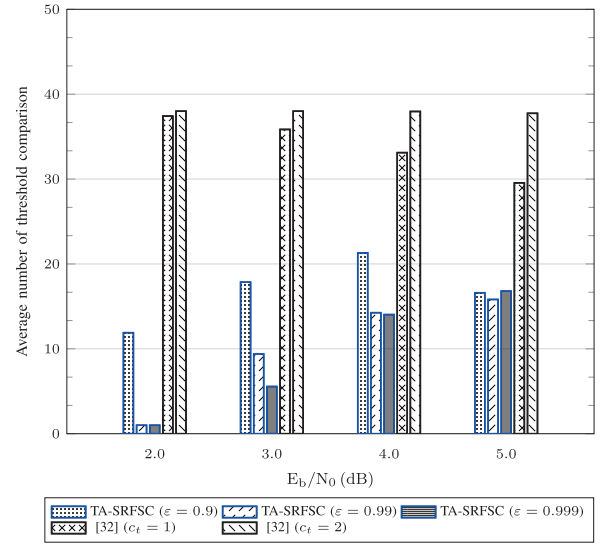


Fig. 8. Average number of threshold comparisons of the proposed TA-SRFSC decoding in comparison with the hard-decision scheme in [32] for the 5G polar code of length $N = 1024$ and rate $R = 1/2$.

a lower decoding latency. This means the decoder in [32] executes many unnecessary threshold comparison operations, while TA-SRFSC decoding only makes hard decisions when a node satisfies the condition in Equation (32).

VII. CONCLUSION

In this work, a new sequence repetition (SR) node is identified in the successive-cancellation (SC) decoding tree of polar codes and an SR node-based fast SC (SRFSC) decoder is proposed. In addition, to speed up the decoding of nodes with no specific structure, the SRFSC decoder is combined with a threshold-based hard-decision-aided (TA) scheme and a multi-stage decoding strategy. We show that this method further reduces the decoding latency at high SNRs. In particular, hardware implementation results on an FPGA for a polar code of length 1024 with code rates 1/4, 1/2, and 3/4 show that the proposed SRFSC decoder with merging branch operations requires up to 8% fewer clock cycles and achieves 10% higher maximum operating frequency compared to state-of-the-art decoders. In addition, the proposed TA-SRFSC decoding reduces the average decoding latency by 57% with respect to SRFSC decoding at $E_b/N_0 = 5$ dB on a polar code of length 1024 and rate 1/2. This average latency saving is particularly important in real-time applications such as video. Future work includes the design of a fast SC list decoder using SR nodes.

APPENDIX A PROOF OF PROPOSITION 1

Let I_k denote the $k \times k$ identity matrix for $k \geq 1$. Since the source node is the rightmost node in an SR node, the g function calculation in (3) can be used as

$$\begin{aligned} \alpha_r^E [1 : 2^r] \\ = \alpha_j^i [1 : 2^j] \times \left(I_{2^{j-1}} \otimes ((-1)^{\eta_{j-1}}, 1)^T \right) \end{aligned}$$

$$\begin{aligned} & \times \left(I_{2^{j-2}} \otimes ((-1)^{\eta_{j-2}}, 1)^T \right) \\ & \times \cdots \times \left(I_{2^r} \otimes ((-1)^{\eta_r}, 1)^T \right). \end{aligned}$$

Using the identity $(A \otimes B) \times (C \otimes D) = (A \times C) \otimes (B \times D)$ with $A = I_{2^{j-2}}$, $B = I_2 \otimes ((-1)^{\eta_{j-1}}, 1)^T$, $C = I_{2^{j-2}}$, and $D = ((-1)^{\eta_{j-2}}, 1)^T$ results in

$$\begin{aligned} \alpha_r^E [1 : 2^r] &= \alpha_j^i [1 : 2^j] \times \cdots \left[(I_{2^{j-2}} \times I_{2^{j-2}}) \right. \\ & \quad \left. \otimes \left(I_2 \otimes ((-1)^{\eta_{j-1}}, 1)^T \times ((-1)^{\eta_{j-2}}, 1)^T \right) \right] \\ & \quad \times \left(I_{2^{j-3}} \otimes ((-1)^{\eta_{j-3}}, 1)^T \right) \\ & \quad \times \cdots \times \left(I_{2^r} \otimes ((-1)^{\eta_r}, 1)^T \right), \end{aligned}$$

which can be written as

$$\begin{aligned} \alpha_r^E [1 : 2^r] &= \alpha_j^i [1 : 2^j] \times I_{2^{j-2}} \otimes \left(((-1)^{\eta_{j-2}}, 1)^T \otimes ((-1)^{\eta_{j-1}}, 1)^T \right) \\ & \quad \times \left(I_{2^{j-3}} \otimes ((-1)^{\eta_{j-3}}, 1)^T \right) \times \cdots \times \left(I_{2^r} \otimes ((-1)^{\eta_r}, 1)^T \right), \end{aligned}$$

where the identity $I_2 \otimes (a_1, \dots, a_k)^T \times (b_1, b_2)^T = (b_1, b_2)^T \otimes (a_1, \dots, a_k)^T$ is used. Repeating the above procedures results in

$$\begin{aligned} \alpha_r^E [1 : 2^r] &= \alpha_j^i [1 : 2^j] \times I_{2^r} \otimes \left(((-1)^{\eta_r}, 1)^T \otimes \cdots \otimes ((-1)^{\eta_{j-1}}, 1)^T \right) \\ &= \alpha_j^i [1 : 2^j] \times I_{2^r} \otimes \left((-1)^{s_l[1]}, (-1)^{s_l[2]}, \dots, (-1)^{s_l[2^{j-r}]} \right)^T. \end{aligned}$$

Thus for $k \in \{1, \dots, 2^r\}$,

$$\alpha_{r_l}^E [k] = \sum_{m=1}^{2^{j-r}} \alpha_j^i [(k-1)2^{j-r} + m] (-1)^{s_l[m]}.$$

APPENDIX B

EXPLANATION OF OBSERVATION 1

To explain Observation 1, Observation 3 is first introduced as follows.

Observation 3: Under the Gaussian approximation assumption in (14), for any node \mathcal{N}_j^i whose 2^j bits undergo a hard decision in (19), assuming all prior bits are decoded correctly, the probability of correct decoding can be calculated as $\left(\frac{\tilde{P}_c}{\tilde{P}_c + \tilde{P}_e} \right)^{2^j}$.

Explanation: In accordance with Fig. 5, for any node \mathcal{N}_j^i , considering all the previous bits are decoded correctly, the probability that the k -th bit ($1 \leq k \leq 2^j$) in the node undergoes a hard decision is $\tilde{P}_c + \tilde{P}_e$. Moreover, The probability of a correct hard decision for the k -th bit in the node is \tilde{P}_c , regardless of the value of $\beta_j^i[k]$. Thus, the conditional probability that a hard decision on the k -th bit is correct given that the k -th bit undergoes a hard decision is $\frac{\tilde{P}_c}{\tilde{P}_c + \tilde{P}_e}$. Since the LLR values of bits in a node are independent of each other, the conditional probability that hard decisions on all the 2^j

bits of node \mathcal{N}_j^i are correct given that all its 2^j bits undergo hard decisions can be calculated as $\left(\frac{\tilde{P}_c}{\tilde{P}_c + \tilde{P}_e} \right)^{2^j}$.

To have a probability of correct decoding of at least ε for all the nodes that undergo hard decision in a polar code of length 2^n , any such node \mathcal{N}_j^i is assumed to have the probability of correct decoding of at least $2^{(n-j)}\sqrt{\varepsilon}$. Therefore and by using the result in Observation 3, we have

$$\left(\frac{\tilde{P}_c}{\tilde{P}_c + \tilde{P}_e} \right)^{2^j} \geq 2^{(n-j)}\sqrt{\varepsilon}, \quad (40)$$

which is equivalent to

$$\frac{\tilde{P}_c}{\tilde{P}_c + \tilde{P}_e} \geq 2^{n\sqrt{\varepsilon}}. \quad (41)$$

If $m_j^i \leq 2c^2$, then $T = -m_j^i + c\sqrt{2m_j^i}$, $\tilde{P}_c = Q\left(c - \sqrt{2m_j^i}\right)$, and $\tilde{P}_e = Q(c)$. When $0.5 < \varepsilon < 1$, (41) can be written as

$$\frac{1}{2} \left[c - Q^{-1} \left(\frac{Q(c)}{\frac{1}{2^{n\sqrt{\varepsilon}}} - 1} \right) \right]^2 \leq m_j^i \leq 2c^2, \quad (42)$$

which requires

$$Q(c) \leq 1 - 2^{n\sqrt{\varepsilon}}. \quad (43)$$

If $m_j^i \geq 2c^2$, then $T = m_j^i - c\sqrt{2m_j^i}$, $\tilde{P}_c = Q(-c)$, and $\tilde{P}_e = Q\left(\sqrt{2m_j^i} - c\right)$. Thus (41) can be written as

$$m_j^i \geq \max \left\{ 2c^2, \frac{1}{2} \left[c + Q^{-1} \left(\left(\frac{1}{2^{n\sqrt{\varepsilon}}} - 1 \right) Q(-c) \right) \right]^2 \right\}. \quad (44)$$

If (43) holds and by using the fact that $Q(-c) = 1 - Q(c)$, then

$$2c^2 \geq \frac{1}{2} \left[c + Q^{-1} \left(\left(\frac{1}{2^{n\sqrt{\varepsilon}}} - 1 \right) Q(-c) \right) \right]^2. \quad (45)$$

Thus $m_j^i \geq 2c^2$, which always holds based on the initial assumption. Therefore, it suffices that

$$m_j^i \geq \frac{1}{2} \left[c - Q^{-1} \left(\frac{Q(c)}{\frac{1}{2^{n\sqrt{\varepsilon}}} - 1} \right) \right]^2, \quad (46)$$

and (43) to ensure (41). In other words, assuming all previous bits are decoded correctly, the probability that any node \mathcal{N}_j^i that undergo hard decision (19) in the decoding process is decoded correctly is lower bounded by $2^{(n-j)}\sqrt{\varepsilon}$ if (43) and (46) are satisfied.

APPENDIX C

EXPLANATION OF OBSERVATION 2

Note that based on Observation 1, any node \mathcal{N}_j^i that is decoded using (19) has a probability of correct hard decision of approximately greater than or equal to $2^{(n-j)}\sqrt{\varepsilon}$. For any node that undergoes the SRFSC decoding, the probability of correct decoding is determined by the error rate of

SRFSC decoding. Thus, the probability of correct decoding for TA-SRFSC decoding is approximately greater than or equal to $\varepsilon(1 - \text{BLER}_{\text{SRFSC}})$. Consequently, $\text{BLER}_{\text{TA-SRFSC}} \lesssim 1 - \varepsilon(1 - \text{BLER}_{\text{SRFSC}})$.

REFERENCES

- [1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- [2] Study on New Radio Access Technology, document Final Report of 3GPP TSG RAN WG1 #87 v1.0.0, Reno, NV, USA, Nov. 2016. [Online]. Available: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_87
- [3] K. Niu, K. Chen, J. Lin, and Q. Zhang, "Polar codes: Primary concepts and practical decoding algorithms," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 192–203, Jul. 2014.
- [4] C. Zhang, B. Yuan, and K. K. Parhi, "Reduced-latency SC polar decoder architectures," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3471–3475.
- [5] B. Yuan and K. K. Parhi, "Low-latency successive-cancellation polar decoder architectures using 2-bit decoding," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 4, pp. 1241–1254, Apr. 2014.
- [6] A. Mishra et al., "A successive cancellation decoder ASIC for a 1024-bit polar code in 180 nm CMOS," in *Proc. IEEE Asian Solid State Circuits Conf. (A-SSCC)*, Dec. 2012, pp. 205–208.
- [7] B. Yuan and K. K. Parhi, "Reduced-latency LLR-based SC list decoder for polar codes," in *Proc. 25th Ed. Great Lakes Symp. VLSI*, May 2015, pp. 107–110.
- [8] B. Yuan and K. K. Parhi, "Low-latency successive-cancellation list decoders for polar codes with multibit decision," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 10, pp. 2268–2280, Oct. 2015.
- [9] G. Sarkis and W. J. Gross, "Increasing the throughput of polar decoders," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 725–728, Apr. 2013.
- [10] C. Husmann, P. C. Nikolaou, and K. Nikitopoulos, "Reduced latency ML polar decoding via multiple sphere-decoding tree searches," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1835–1839, Feb. 2018.
- [11] A. Alamdar-Yazdi and F. R. Kschischang, "A simplified successive-cancellation decoder for polar codes," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1378–1380, Dec. 2011.
- [12] G. Sarkis, P. Giard, A. Vardy, C. Thibault, and W. J. Gross, "Fast polar decoders: Algorithm and implementation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 946–957, May 2014.
- [13] P. Giard, G. Sarkis, C. Thibault, and W. J. Gross, "Multi-mode unrolled architectures for polar decoders," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 9, pp. 1443–1453, Sep. 2016.
- [14] Z. Huang, C. Diao, and M. Chen, "Latency reduced method for modified successive cancellation decoding of polar codes," *Electron. Lett.*, vol. 48, no. 23, pp. 1505–1506, Nov. 2012.
- [15] A. Balatsoukas-Stimming, G. Karakostas, and A. Burg, "Enabling complexity-performance trade-offs for successive cancellation decoding of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2977–2981.
- [16] L. Zhang, Z. Zhang, X. Wang, C. Zhong, and L. Ping, "Simplified successive-cancellation decoding using information set reselection for polar codes with arbitrary blocklength," *IET Commun.*, vol. 9, no. 11, pp. 1380–1387, Jul. 2015.
- [17] P. Giard, G. Sarkis, C. Thibault, and W. J. Gross, "A 638 Mbps low-complexity rate 1/2 polar decoder on FPGAs," in *Proc. IEEE Workshop Signal Process. Syst. (SiPS)*, Oct. 2015, pp. 1–6.
- [18] P. Giard, A. Balatsoukas-Stimming, G. Sarkis, C. Thibault, and W. J. Gross, "Fast low-complexity decoders for low-rate polar codes," *J. Signal Process. Syst.*, vol. 90, no. 5, pp. 675–685, May 2018.
- [19] M. Hanif and M. Ardakani, "Fast successive-cancellation decoding of polar codes: Identification and decoding of new nodes," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2360–2363, Nov. 2017.
- [20] C. Condo, V. Bioglio, and I. Land, "Generalized fast decoding of polar codes," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
- [21] H. Gamage, V. Ranasinghe, N. Rajatheva, and M. Latva-aho, "Low latency decoder for short blocklength polar codes," 2019, *arXiv:1911.03201*. [Online]. Available: <http://arxiv.org/abs/1911.03201>
- [22] F. Ercan, C. Condo, and W. J. Gross, "Reduced-memory high-throughput fast-SSC polar code decoder architecture," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2017, pp. 1–6.
- [23] F. Ercan, T. Tonnellier, C. Condo, and W. J. Gross, "Operation merging for hardware implementations of fast polar decoders," *J. Signal Process. Syst.*, vol. 91, no. 9, pp. 995–1007, Sep. 2019.
- [24] P. Giard and A. Burg, "Fast-SSC-flip decoding of polar codes," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2018, pp. 73–77.
- [25] F. Ercan, T. Tonnellier, and W. J. Gross, "Energy-efficient hardware architectures for fast polar decoders," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 1, pp. 322–335, Jan. 2020.
- [26] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.
- [27] G. Sarkis, P. Giard, A. Vardy, C. Thibault, and W. J. Gross, "Fast list decoders for polar codes," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 2, pp. 318–328, Feb. 2016.
- [28] S. A. Hashemi, C. Condo, and W. J. Gross, "A fast polar code list decoder architecture based on sphere decoding," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 12, pp. 2368–2380, Dec. 2016.
- [29] S. A. Hashemi, C. Condo, and W. J. Gross, "Fast and flexible successive-cancellation list decoders for polar codes," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5756–5769, Nov. 2017.
- [30] S. A. Hashemi, C. Condo, M. Mondelli, and W. J. Gross, "Rate-flexible fast polar decoders," *IEEE Trans. Signal Process.*, vol. 67, no. 22, pp. 5689–5701, Nov. 2019.
- [31] M. H. Ardakani, M. Hanif, M. Ardakani, and C. Tellambura, "Fast successive-cancellation-based decoders of polar codes," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4562–4574, Jul. 2019.
- [32] S. Li, Y. Deng, L. Lu, J. Liu, and T. Huang, "A low-latency simplified successive cancellation decoder for polar codes based on node error probability," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2439–2442, Dec. 2018.
- [33] H. Sun, R. Liu, and C. Gao, "A simplified decoding method of polar codes based on hypothesis testing," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 530–533, Mar. 2020.
- [34] C. Leroux, I. Tal, A. Vardy, and W. J. Gross, "Hardware architectures for successive cancellation decoding of polar codes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1665–1668.
- [35] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3221–3227, Nov. 2012.
- [36] Z. Zhang, L. Zhang, X. Wang, C. Zhong, and H. V. Poor, "A split-reduced successive cancellation list decoder for polar codes," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 2, pp. 292–302, Feb. 2016.
- [37] R. Silverman and M. Balser, "Coding for constant-data-rate systems," *Trans. IRE Prof. Group Inf. Theory*, vol. 4, no. 4, pp. 50–63, Sep. 1954.
- [38] M. El-Khamy, J. Lee, and I. Kang, "Detection analysis of CRC-assisted decoding," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 483–486, Mar. 2015.
- [39] C. Leroux, A. J. Raymond, G. Sarkis, and W. J. Gross, "A semi-parallel successive-cancellation decoder for polar codes," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 289–299, Jan. 2013.
- [40] Multiplexing and Channel Coding, document 3GPP TS RAN 38.212 v1.2.1, Dec. 2017. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/38_series/38.212/38212-f30.zip
- [41] H. Zheng, A. Balatsoukas-Stimming, Z. Cao, and A. M. J. Koonen, "Implementation of a high-throughput fast-SSC polar decoder with sequence repetition node," in *Proc. Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2020, pp. 1–6. [Online]. Available: <http://arxiv.org/abs/2007.11394>



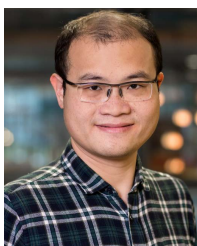
Haotian Zheng (Student Member, IEEE) was born in Xiangyang, China. He received the B.S. degree in information and communication engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2012, and the M.E. degree in communication and information systems from the National Digital Switching System Engineering and Technological Research Center (NDSC), Zhengzhou, China, in 2015. He is currently pursuing the Ph.D. degree with the Electro-Optical Communications (ECO) Group, Department of Electrical and Engineering, Eindhoven University of Technology. His research interests include error-correcting codes and the VLSI implementation of digital signal processing systems.



Seyyed Ali Hashemi (Member, IEEE) was born in Qaemshahr, Iran. He received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2009, the M.Sc. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2011, and the Ph.D. degree in electrical and computer engineering from McGill University, Montréal, QC, Canada, in 2019. He is currently a Post-Doctoral Fellow with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. His research interests include machine learning for communications, error-correcting codes, and VLSI implementation of digital signal processing systems. He was a recipient of the Best Student Paper Award at the 2016 IEEE International Symposium on Circuits and Systems (ISCAS 2016) and the Postdoctoral Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) in 2018.



Alexios Balatsoukas-Stimming (Member, IEEE) received the Diploma and M.Sc. degrees in electronics and computer engineering from the Technical University of Crete, Chania, Greece, in 2010 and 2012, respectively, and the Ph.D. degree in computer and communications sciences from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2016. He then spent one year at the European Laboratory for Particle Physics (CERN) as a Marie Skłodowska-Curie Post-Doctoral Fellow, after which he was a Post-Doctoral Researcher with the Telecommunications Circuits Laboratory, EPFL, from 2018 to 2019. He is currently an Assistant Professor with the Eindhoven University of Technology, The Netherlands. His research interests include VLSI circuits for signal processing and communications, error-correction coding theory and practice, and the applications of machine learning to signal processing for communications. He has served as a track chair and technical program committee member for multiple IEEE conferences. He is an Associate Editor of IEEE COMMUNICATIONS LETTERS and for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Series on Machine Learning in Communications and Networks.



Zizheng Cao (Member, IEEE) is currently a tenured Assistant Professor in beam-steered optoelectronic system with the Electro-Optical Communications (ECO) Group, Department of Electrical Engineering, Eindhoven University of Technology (TU/e). His research interests include from his curiosity on how to efficiently deliver and detect wireless signals. It is a key for wireless communication and remote/non-invasive sensing. Next to the existing technologies in time and frequency domain, the approaches of spatial focusing and isolation are widely recognized as promising but not fully established solutions. His current research interests include enable efficient delivery and detection of weak signals by means of beam-steering/forming. He mainly works at two areas new mechanisms, devices, algorithms and architectures to enable lightwave/millimeter wave various kinds of beam steering systems, including scenarios of line-of-sight (LoS), non-line-of-sight (NLoS) and complex wavefront; and the applications of beam steering systems to indoor communication (e.g., optical wireless communication, OWC), metrology (e.g., LiDAR) and healthcare (e.g., blood sensing). His pioneering work results in a complete optically controlled mm-wave beam steering communication system, a new type of OWC transmitters for OWC system with NLoS beam steering, and a new design approach to simultaneously optimize optical and electrical properties of OWC receivers. These results lead to several publications in Nature LSA, Wiley LPR (cover story), and invited papers in IEEE journals. He has established a complete set of hardware and software to carry out aforementioned research, including state-of-art OWC testbeds, advanced IM-DD and coherent DSP algorithms (offline and real time), self-developed design kit of functional photonic system-on-chip (photonic integrated circuit, PIC), and automatic measurement setups of electrooptical chips.



Ton Koonen (Fellow, IEEE) is currently a Full Professor of electro-optical communications and also a Chair of broadband communication networks with the Department Telecommunication Technology and Electromagnetics. He has initiated and led several European and national R&D projects in this area on dynamically reconfigurable hybrid fiber access networks, fiber-wireless, packet-switched access, and short-range multimode (polymer) optical fiber networks, and label-controlled optical packet routed networks. His research interests include computer systems, architectures and networks, telecommunications, broadband and optical fiber-to/in-the-home, optical fiber-supported in-building networks (including optical wireless communication techniques, radio-over-fiber techniques, and high-capacity plastic optical fiber (POF) techniques), optical access networks, and spatial division multiplexed systems. His group has, for example, developed a Wi-Fi network that transmits signals via infrared light, achieving a speed of 42.8 Gb/s, 100 times faster than current networks generally achieve. At this speed, an entire film could be transferred in one second.



John M. Cioffi (Fellow, IEEE) received the B.S.E.E. degree from IL, USA, in 1978, and the Ph.D.E.E. degree in Stanford in 1984. He was a Founder of Amati in 1991 (1997 purchased by TI); a Chairman and a CEO ASSIA, Inc. Since 1986, he was a Professor of the Department of EE, Stanford University, where he is currently an Emeritus. He has more than 800 articles and more than 150 heavily licensed patents. He has served over a dozen boards of directors, currently a Chairman with PhyTunes, a Vice-Chairman with Marconi Society, Tinoq, and AltoBeam. His current interest includes high-performance digital transmission. His awards include IEEE AG Bell in 2010, Kirchmayer in 2014 and Millennium Medals; Member Internet in 2014 and Consumer-Electronics in 2018 Halls of Fame; Marconi Fellow in 2006; Member, U.S. National in 2001, and U.K. Royal in 2009 Engineering Academies.



Andrea Goldsmith (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from U.C. Berkeley. She was the Stephen Harris Professor of engineering and a Professor of electrical engineering with Stanford University, where she is currently a Harris Professor Emerita. She is also the Dean of engineering and applied science and the Arthur LeGrand Doty Professor of electrical engineering with Princeton University. Her research interests include information theory, communication theory, and signal processing, and their application to wireless communications, interconnected systems, and neuroscience. She founded and served as a Chief Technical Officer for Plume WiFi (formerly Accelera, Inc.) and of Quantenna (QTNA), Inc. She currently serves on the Board of Directors for Medtronic (MDT) and Crown Castle Inc. (CCI). She has authored the book *Wireless Communications* and coauthored the books *MIMO Wireless Communications* and *Principles of Cognitive Radio* (Cambridge University Press) and an inventor on 29 patents.

Dr. Goldsmith is a member of the National Academy of Engineering and the American Academy of Arts and Sciences. She is a fellow of the Stanford, and has received several awards for her work, including the Marconi Prize, the IEEE Sumner Technical Field Award, the ACM Athena Lecturer Award, the ComSoc Armstrong Technical Achievement Award, the Kirchmayer Graduate Teaching Award, the WICE Mentoring Award, and the Silicon Valley/San Jose Business Journal's Women of Influence Award. She is currently the Founding Chair of the IEEE Board of Directors Committee on Diversity, Inclusion, and Ethics. She served as the President for the IEEE Information Theory Society in 2009, as a Founding Chair for its Student Committee, and as a Founding Editor-in-Chief for IEEE JOURNAL ON SELECTED AREAS OF INFORMATION THEORY. She has also served on the Board of Governors for both the IEEE Information Theory and Communications Societies. At Stanford, she has served as the Chair of Stanford's Faculty Senate and for multiple terms as a Senator, and on its Academic Council Advisory Board, Budget Group, Committee on Research, Planning and Policy Board, Commissions on Graduate and on Undergraduate Education, Faculty Women's Forum Steering Committee, and Task Force on Women and Leadership.