**3GPP TSG RAN WG1 Meeting #86bis R1-1608865**

**Lisbon, Portugal, Oct 10 - 14, 2016**

**Agenda Item: 8.1.4.1**

**Source: Huawei, HiSilicon**

**Title: Design aspects of Polar Code and LDPC for NR**

**Document for: Discussion and Decision**

# Introduction

When constructing a Polar Code, [1] increases its minimum coding distance by adding a self-parity-check function over some selected PC (Parity-Check) frozen bits so that a List 8 Parity Check Successive Cancelation List Polar Decoder (PC-SCL) outperforms a 8-iter LTE-Turbo MLM (maximum-log-map) decoder and a 20-iter LDPC LOMS (layered-offset-min-sum) decoder in most eMBB cases as illustrated [1]. A PC-SCL decoder avoids using CRC bits to select the decoding final path, thereby ensuring as good detection capability as TBCC Viterbi decoder.

To reduce the decoding latency of a SCL (Successive Cancelation list) decoder for the eMBB data and URLLC data channels, [1] divides one block into equal-sized segments and chains them by a cross-parity-check function. The resulting multiple chained PC-based Polar Code segments are simultaneously decoded by multiple PC-SCL decoders in a parallel.

In this contribution, we focus on the design implementation aspects such as area and latency of the polar code for a NR terminal. We use the ME-LDPC proposed in [2] for the comparison purpose. We select the following typical study cases to cover the design aspect exploration.

|  |  |  |
| --- | --- | --- |
| **Decoding Scenarios** | **Huawei’s Chained Polar Code** | **Qualcomm’s (QC) LDPC** |
| Decoding Latency < 16us for eMBB data channel (peak code rate) | PC-SCL List 8 decoder  N = 2048 for K < 2000 bits (2K PC-basedcodeword)  N = 4096 or 2000 < K < 4000 bits (2 chained 2K PCC segments)  N = 8192 for 4000 < K < 8000 bits (2 chained 4K PCC segments) | 20-iteration LOMS decoder |
| Decoding Latency < 16us for URLLC (low code rate < 1/3) | PC-SCL List 16 decoder  N = 16384 for K <= 1600 bits (2 chained 8K PCC segments) | 50-iteration LOMS decoder for code rate lower than 1/3 |

Table 1 Typical Decoding Scenarios for Comparison between Polar Code and ME-LDPC

Section 5 discusses a flexible polar decoding architecture that demonstrates the hardware efficiency. The same PC-SCL decoder can be used to decode any codeword sizes (N) from 1K to 16K with list size from 2 to 32.

The design parameters through this contribution are synthesized with 14nm ASIC technology at a pre-layout stage. Note that the area and latency are expected to improve along with the development of the ASIC technology toward to favouring the Polar decoding implementation.

The following denotations are used through this contribution:

K: information block length

N: mother Polar Code length (power of two)

L: List size of polar PC-SCL or CA-SCL decoder

R: Code Rate

# Polar Code for eMBB Data Channel

[5] specifies the shortest decoding latency Δt for eMBB data channel to be 16us (two OFDM symbols duration, = 66us/4) or even 8us (one OFDM symbols duration, =66us/8) with a subcarrier spacing of 60Hz in NR. To meet this latency with an outstanding BLER performance, a PC-SCL List 8 decoding implementation is studied with the reasonable hardware resource consumption.

A PC-SCL List 8 decoder reaches similar performance of a CA-SCL List 32 decoder for a conventional polar at high code rates. When both decoding algorithm use List 32 (PC-SCL-32 and CA-SCL-32), the PC-based polar code has better performance, except at very low coding rates. The gain of a PC-based polar code becomes larger as the coding rate increases, as demonstrated in [1].

This PC-SCL decoder implementation employs a SSC (4-bit aggregation) decoding algorithm to reduce the latency. An effortless AND and OR circuit can easily determines the “Rate-0” and “Rate-1” of 4-bit group on the fly from the frozen-bit set generated from the PW-based method defined in [7]. A Selective-Path-Extension technique disclosed in [6] with a ratio of 70% is applied to reduce the latency too. In case of the information blocks longer than 2000 bits, a chained PC-based polar code is constructed as proposed in [1].

Figure 1 illustrates a general diagram with a dual-segment PC-based polar code and its parallelized decoding.



Dual-segmented PC-basedPolar Code using a cross-parity-check Function



Figure 1 PC-SCL parallel decoding of a dual-segmented PC-basedPolar Code

|  |  |  |  |
| --- | --- | --- | --- |
| **Implementation** | **N = 2K** | **N = 4K** | **N = 8K** |
| Code Construction | Polar Code with self-parity check (PC) | Chained Polar Code with cross-parity check (PCC) | Chained Polar Code with cross-parity check (PCC) |
| Decoding Algorithm | PC-SCL List 8 | PC-SCL List 8 | PC-SCL List 8 |
| Decoding resources | N = 2K L=8 1x 2K PC-basedsegment | N = 4K L=8 2x 2K PCC-enabled segments | N = 8K L=8 2x 4K PCC-enabled segments |
| Internal decoding memory | 45 kbits | 87 kbits | 173 kbits |
| Estimated Area | 0.25 mm2 | 0.28 mm2 | 0.30 mm2 |
| Latency (ns) @ 1GHz with highest code rate (R=8/9) | 3574 | 3663 | 7645 |
| Normalized T/P (bit/cycle) | 0.51 | 0.99 | 0.95 |
| Hardware efficiency (Gpbs/mm2) | 2.07 | 3.60 | 3.14 |

Table 2 Key Parameters of PC-SCL List 8 Decoder for eMBB data channel (R=8/9)

The decoding latency for the case K ~= 8000 bits (with 2 chained 4K Polar Code segments) is illustrated as follow:



Figure 2 Latency Reduction attributed to dual Chained Polar Code Construction with R=8/9

Although it would reduce the latency to divide a block shorter than 2000 bits into two segments, the priority in a flexible decoding architecture would be to enhance the performance by increasing the list size if Δt is reached.

***Observation-1****: Using the chained polar code construction, the decoding latency of a PC-SCL List 8 decoder is less than 8 us in the case of the highest code rate and longest block length in the NR data channel.*

# Polar Code for URLLC Channel

[10] agrees that the code rate (R) of URLLC is from 1/3 down to 1/12 and the information block length (K) is up to 200 bytes. [3]’s simulation results show that a PC-SCL List 16 decoder outperforms both 50-iteration LOMS LDPC decoder and 8-iter MLM Turbo decoder. To match the same performance of PC-SCL List-16 decoder (especially for R<1/5), a LDPC decoder tends to iterate over 50 times.

For URLLC, the same PC-SCL decoding algorithm is used as that for eMBB data channel, except that the only it employs a 2-bit aggregation decoding algorithm and a list size of 16 to provide a more reliable decoding algorithm. To meet the decoding latency Δt = 16us, we present the results of a dual-chained polar code for 200 information bytes at rate of 1/8. The BLER performance of a PC-SCL List 16 is comparable to a CA-SCL List 32 decoder for low rate, as demonstrated in [1].

|  |  |
| --- | --- |
| **Implementation** | **N = 16K (K ~= 200 bytes)** |
| Code Construction | Chained Polar Code with cross-parity check (PCC) |
| Decoding Algorithm | PC-SCL List 16 |
| Decoding resources | N = 16K L=16 2x 8K PCC-enabled segments |
| Internal decoding memory | 693 kbits |
| Estimated Area | 0.45 mm2 |
| Latency (ns) @ 1GHz with lowest code rate (R=1/8) | 9000 |
| Normalized T/P (bit/cycle) | 0.23 |
| Hardware efficiency (Gpbs/mm2) | 0.55 |

Table 3 Key Parameters of Huawei’s PC-SCL List 16 Decoder for uRLLC channel (R=1/8)

***Observation-2:*** *Using the chained polar code construction, the decoding latency of a PC-SCL List 16 decoder is less than 16 us in the case of the lowest code rate and the longest block length in the URLLC channel.*

# Polar Code for Control Channel

As a PC-basedpolar is applied to NR control channel [13], the same PC-SCL decoder architecture is also leveraged for the control channel decoding. However, the internal decoder resources are optimized to provide ultra short decoding latency.

|  |  |
| --- | --- |
| **Implementation** | **N = 1024 (K <= 200 bits)** |
| Code Construction | Polar Code with self-parity check (PC) |
| Decoding Algorithm | PC-SCL List 8 |
| Decoding resources | N=1K, L=8 1K PC-based segment |
| Internal decoding memory | 23 kbits |
| Estimated Area | 0.24 mm2 |
| Latency (ns) @ 1GHz with lowest code rate (R=1/5) | 470 |
| Normalized T/P (bit/cycle) | 0.44 |
| Hardware efficiency (Gpbs/mm2) | 1.82 |

Table 4 Key Parameters of a PC-SCL List 8 decoder for control channel (R=1/5)

***Observation-3:*** *The decoding latency of a PC-SCL List 8 decoder meets the latency requirement of the NR control channel.*

# Flexible SCL Polar Decoder

A real NR terminal decoder chip needs to support diverse codeword lengths and rates given a hardware die area constraint. It is well known that a SCL decoding implementation is bounded by the memory more than computational logics and that its memory size and organization are in term of list size. Moreover, the polarization theory shows that a longer mother code length (N) possesses more polarization and requires a smaller list size (L) to meet the BLER performance target. Both of these characteristics allow the construction of a flexible decoding architecture based on a simple hedge rule: when N goes down to N/2, L is doubled to 2\*L.

In our test, we built a flexible quad-codeword parallel SCL decoding architecture to support a mother codeword length (N) from 16K with List 2 down to N=1K with List 32. The decoder uses a total of 380 kbits and its area is estimated to 0.51 mm2.

This flexible polar decoding architecture is built on the exiting implementation introduced in [12] at the #86 meeting. We design the internal LLR memory for the stages 3 to 7 for the longest mother code (Nmax) and the sort unit for the largest list size (Lmax) so that they can be re-used by any different valid combinations of N and L. The chained PC-SCL decoder can be naturally integrated into this architecture.

The flexible polar decoding architecture is illustrated as below:



Figure 3 Quad-Codeword Flexible SCL Polar Decoder

We measure the hardware efficiency (information bits Gbps/mm2) of SCL decoder at 1GHz frequency for R=1/8 and R=8/9 respectively. Note that our flexible decoder supports the decoding of 4 codewords simultaneously within 0.51mm2 (0.21mm2 of logic and 0.30mm2 of memory) in 14nm ASIC technology. The throughput numbers are normalized for a 1mm2 die area.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N=1K | N=2K | N=4K | N=8K | N=16K |
| L=2 | 4.46 | 4.53 | 5.20 | 5.96 | 6.22 |
| L=4 | 2.76 | 2.99 | 3.51 | 4.03 |  |
| L=8 | 1.66 | 1.89 | 2.26 |  |  |
| L=16 | 0.90 | 0.68 |  |  |  |
| L=32 | 0.52 |  |  |  |  |

Table 5 Hardware Efficiency (Info Gbps/mm2) at R=1/8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N=1K | N=2K | N=4K | N=8K | N=16K |
| L=2 | 9.44 | 10.40 | 11.45 | 13.31 | 13.43 |
| L=4 | 6.25 | 7.62 | 8.69 | 10.45 |  |
| L=8 | 3.96 | 5.31 | 6.30 |  |  |
| L=16 | 2.03 | 1.50 |  |  |  |
| L=32 | 1.25 |  |  |  |  |

Table 6 Hardware Efficiency (Info Gbps/mm2) at R=8/9

In term of the mother codeword length N and code rate R, this flexible decoding architecture selects the right list size in such a way to leverage both the BLER performance and decoding throughput.

As an example of an over 10Gbps peak information throughput, [4] proposes to use a 7-iteration LOMS LDPC decoder to reach 12.2\*f (f is the frequency) in case of Kmax = 8000 bit and Rmax = 8/9. However, its BLER performance can be matched by List = 2 or List = 4 PC-SCL decoder ([3]). In table 5, the peak throughput is over 13 Gbps given 1-mm2.

***Observation-4****: With a flexible decoding architecture, CA-SCL or PC-SCL decoders can reach a very high throughput for NR.*

In [11], over 1Tbps throughput has been reported at 65nm at 500MHz. However, such a high throughput is reached with very less granularity. Optionally, a hardware accelerator with fixed code rate and block length can be integrated into the decoding chip to come up with over hundreds of Gbps throughput if the granularity is compromised. It can be integrated into the decoder as an accelerator.



Figure 4 Functional Specification of a Flexible Decoding Architecture

# Conclusion

***Observation-1****: Using the chained polar code construction, the decoding latency of a PC-SCL List 8 decoder is less than 8 us in the case of the highest code rate and longest block length in the NR data channel.*

***Observation-2:*** *Using the chained polar code construction, the decoding latency of a PC-SCL List 16 decoder is less than 16 us in the case of the lowest code rate and the longest block length in the URLLC channel.*

***Observation-3:*** *The decoding latency of a PC-SCL List 8 decoder meets the latency requirement of the NR control channel.*

***Observation-4****: With a flexible decoding architecture, CA-SCL or PC-SCL decoders can reach a very high throughput for NR.*

# References

1. R1-1608862: “Polar Code Construction for NR”, Huawei, HiSilicon
2. R1-166370: “R1-166370\_LDPC\_rate\_compatible\_design”, Qualcomm
3. R1-1608864: “Performance Evaluation for NR Channel Coding”, Huawei, HiSilicon
4. R1- 166372, “EMBB\_ChannelCoding\_Implementation”, Qualcomm
5. R1-166933, “Latency Consideration of Channel Coding Candidates“, Ericsson
6. R1-164040, “On latency and complexity”, Huawei, HiSilicon
7. R1-167209, “Polar code design and rate-matching Huawei, HiSilicon
8. R1-166369, “Polar HARQ”, Qualcomm
9. C.Xiong, J.Lin, Z.Yan, “A multi-mode area-efficient SCL polar decoder”
10. R1-168371, “Way-Forward on URLLC evaluation parameter and LLS method”
11. P. Giard, G. Sarkis, C. Thibeault, W. Gross, “Multi-mode Unrolled Architectures for Polar Decoders”
12. R1-167213 Computational and implementation complexity of channel coding schemes
13. R1-1608863: “Evaluation of channel coding schemes for control channel”, Huawei, HiSilicon