A highly parallel Turbo Product Code decoder without interleaving resource
Cédric Leroux, Christophe Jego, Patrick Adde, Michel Jezequel, D. Gupta

To cite this version:

HAL Id: hal-00538606
https://hal.archives-ouvertes.fr/hal-00538606
Submitted on 22 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A HIGHLY PARALLEL TURBO PRODUCT CODE DECODER
WITHOUT INTERLEAVING RESOURCE

Camille Leroux, Christophe Jego, Patrick Adde, Michel Jezequel, Deepak Gupta

Institut TELECOM, TELECOM Bretagne, CNRS Lab-STICC
firstname.lastname@telecom-bretagne.eu

ABSTRACT
This article presents an innovative Turbo Product Code (TPC) decoder architecture without any interleaving resource. This architecture includes a full-parallel SISO decoder able to process n symbols in one clock period. Syntheses show the better efficiency of such an architecture compared with existing previous solutions. Considering a 6-iteration turbo decoder of a (32,26)\textsuperscript{2} BCH product code, synthesized in a 90nm CMOS technology, the resulting information throughput is 2.5Gb/s with an area of 233Kgates. Finally a second architecture enhancing parallelism rate is described. The information throughput is 33.7Gb/s while an area estimation gives \( A = 10\mu \text{m}^2 \).

Index Terms— Product codes, Parallel architectures, Ultra-high-speed integrated circuits

1. INTRODUCTION
Nowadays, high throughput telecommunication systems such as optical fiber transmission systems or passive optical networks require powerfull error correcting codes in order to increase their optical budget. Iterative decoding provides effective solutions for next generation optical systems. Recently, a (660,480) LDPC code decoder ASIC implementation was proposed. The information throughput is 2.4Gb/s while it could be enhanced to 10Gb/s with a (2048,1723) LDPC code [1]. Turbo product codes [2] also tend to be good candidates for emerging optical systems [3]. In [4], a BTC decoder is included in a 12.4Gb/s optical experimental setup. Since only a part of the transmitted data are actually coded, the information throughput of the BTC turbo decoder is 156Mb/s. The inherent parallel structure of the product code matrix confers to TPC good ability for parallel decoding. Nevertheless, enhancing parallelism rate rapidly induces the use of a prohibitive amount of memory. Some solutions were proposed in [5][6][7] to efficiently use the interleaving resources (IR), reaching high parallelism rates. However, a particular scheduling of the product code matrix enables to propose a turbo decoder architecture without any interleaving resource.

After a brief introduction of the TPC coding and decoding concept in section 2, section 3 reviews previous works and proposes an innovative TPC decoder architecture without any interleaving resource. This new TPC decoder includes a new full-parallel SISO decoder architecture which is described in section 4. Section 5 gives some synthesis results and demonstrates the better efficiency of the proposed BTC decoder. Finally, a solution is proposed to enhance parallelism rate while higthering the architecture efficiency. The interconnection issue is assessed and compared with an equivalent LDPC code decoder implementation.

2. TPC CODING AND DECODING PRINCIPLES

2.1. Product codes
The concept of product codes is a simple and efficient method to construct powerful codes with a large minimum Hamming distance \( d \) using cyclic linear block codes [8]. Let us consider two systematic cyclic linear block codes \( C\textsubscript{1} \) having parameters \((n\textsubscript{1},k\textsubscript{1},d\textsubscript{1})\) and \( C\textsubscript{2} \) having parameters \((n\textsubscript{2},k\textsubscript{2},d\textsubscript{2})\) where \( n\textsubscript{i}, k\textsubscript{i} \) and \( d\textsubscript{i} \) \((i = 1, 2)\) stand for code length, number of information symbols and minimum Hamming distance respectively. The product code \( P = C\textsubscript{1} \times C\textsubscript{2} \) is obtained by placing \((k\textsubscript{1} \times k\textsubscript{2})\) information bits in a matrix of \( k\textsubscript{1} \) rows and \( k\textsubscript{2} \) columns, coding the \( k\textsubscript{1} \) rows using code \( C\textsubscript{2} \) and coding the \( n\textsubscript{2} \) columns using code \( C\textsubscript{1} \).

Considering that \( C\textsubscript{1} \) and \( C\textsubscript{2} \) are linear codes, it is shown that all \( n\textsubscript{1} \) rows are codewords of \( C\textsubscript{2} \) exactly as all \( n\textsubscript{2} \) columns are codewords of \( C\textsubscript{1} \) by construction. Furthermore, the parameters of the resulting product code \( P \) are given by \( n\textsubscript{P} = n\textsubscript{1} \times n\textsubscript{2}, k\textsubscript{P} = k\textsubscript{1} \times k\textsubscript{2}, \) and \( d\textsubscript{P} = d\textsubscript{1} \times d\textsubscript{2} \) and the code rate \( R\textsubscript{P} \) is given by \( R\textsubscript{P} = R\textsubscript{1} \times R\textsubscript{2} \). Thus, it is possible to construct powerful product codes using linear block codes. In the following sections, we will consider a squared product code, meaning that \( n\textsubscript{1} = n\textsubscript{2} = n \). The most commonly used component codes are Bose Chaudhuri Hocquenghem (BCH) codes. These codes are an infinite class of linear cyclic block codes that have capabilities for multiple error detection and correction. Product codes were adopted in 2001 as an optional correcting code system for both the uplink and downlink of the IEEE 802.16 standard (WiMAX) [9].
2.2. Iterative decoding of product codes

TPC decoding involves successively decoding rows and columns using SISO decoders. Repeating this soft decoding during several iterations enables the decrease of the Bit Error Rate (BER). It is known as the TPC decoding process. Each decoder has to compute soft information \([R']_{it}\) from the channel received information \([R]\) and the previous half-iteration computed information \([R']_{it}\).

Despite the existence of other decoding algorithms, the Chase-Pyndiah SISO algorithm for a TPC decoding involves successively decoding rows and columns using SISO decoders each decoding \(C_w\) symbols in one clock period (denoted as \(m\)-decoders). A half-iteration structure includes \(m\) decoders each decoding \(m\) symbols in one clock period and an interleaving memory of size \(4qn^2\). The resulting throughput is \(O(m^2)\) while the overhead factor of the decoder complexity is \(\frac{m}{2}\).

In 2002, a new architecture was proposed \([5]\) in order to increase the parallelism degree without any extra storage between half-iteration. The idea was to store several product code matrix symbols at the same address and to use elementary decoders able to process \(m\) symbols during the same clock period (denoted as \(m\)-decoders). A half-iteration structure includes \(m\) decoders each decoding \(m\) symbols in one clock period and an interleaving memory of size \(4qn^2\). The resulting throughput is \(O(m^2)\) while the overhead factor of the decoder complexity is \(\frac{m}{2}\).

In [6], authors suggested to use a barrel shifter between decoding resources and the interleaving memory (IM) in order to avoid memory conflicts. This solution enables reaching the parallelism rate \(P = n\) by rotating the to be stored data. The extra-complexity only consists in a barrel shifter with a complexity in \(O(n \log(n))\). However, the IM requirement is still prohibitive.

In [7][13], an IM-less architecture is detailed and prototyped on an FPGA device. In this architecture, a particular scheduling of the product code matrix decoding enables the interleaving memory to be replaced by an interconnection network (omega network). The complexity of the interleaving resources is then drastically reduced and highly-parallel structure can be implemented onto low-cost targets.

<table>
<thead>
<tr>
<th>Throughput</th>
<th>IM</th>
<th>Internal</th>
</tr>
</thead>
<tbody>
<tr>
<td>[5] (O(m^2), m &lt; n)</td>
<td>(O(4qn^2))</td>
<td>(O(2sqnm))</td>
</tr>
<tr>
<td>[6] (O(n))</td>
<td>(O(4qn^2)) + (O(n \log n))</td>
<td>(O(2sqn^2))</td>
</tr>
<tr>
<td>[7][13] (O(n))</td>
<td>(O(n \log n))</td>
<td>(O(2sqn^2))</td>
</tr>
</tbody>
</table>

Table 1. Throughput and complexity order of previous high-speed architectures
These three architectures can reach high parallelism degrees (i.e., high throughput). However, as illustrated in Table 1, the internal memory complexity remains an important issue in the design of high-throughput TPC decoders. \( s \) is the number of pipeline stages inside the SISO decoder. Moreover, in these architectures, decoding resources consist in a duplication of sequential decoders. Increasing the parallelism rate by duplicating computation resources is inefficient since the reuse of available resources is not optimized. In the next section, we propose to merge the duplicated sequential SISO decoders into one full-parallel SISO decoder.

### 3.2. Proposed IM-free architecture using full-parallel SISO decoder

Considering that one can design a SISO decoder able to process \( n \) symbols in parallel in one clock period, a product code matrix can be decoded without any interleaving resource as shown in Figure 1.

![Proposed parallel decoding scheduling of a product code matrix](image)

At \( t = 0 \), the full-parallel SISO decoder processes the column 1. During the next clock period, \( n \) sequential SISO decoders start decoding the first symbol of each row while the parallel decoder process the column 2. During the \( n \)th clock period, sequential decoders complete matrix decoding while the parallel decoder is already decoding the next matrix.

Data generated by the parallel decoder are immediately used by a sequential decoder. Consequently, no IM or data routing resources are required between the full-parallel decoder and sequential decoders. The resulting proposed architecture and the typical previous architecture for one iteration is depicted on Figure 2.

![Previous and proposed TPC decoder architectures](image)

**Fig. 2.** Previous TPC decoder architecture (a) and proposed full-parallel SISO based TPC decoder architecture (b)

### 4. ARCHITECTURE OF A FULL-PARALLEL COMBINATORY SISO DECODER

#### 4.1. Algorithmic parameter reduction

As explained in section 2, the Chase-Pyndiah algorithm includes parameters \((p, T, Cw, q)\) which impact on both the performance and the complexity of the turbo decoding. Performing 8 iterations, the parameter set \( p_0 = \{p = 5, T = 16, Cw = 3, q = 5\} \) gives the best performance for a reasonable complexity [11]. However, algorithmic simulations showed that the reduced parameter set \( p_1 = \{p = 3, T = 8, Cw = 0, q = 5\} \) only induce a performance loss of 0.25dB at BER = \( 10^{-6} \) while it becomes null below BER = \( 10^{-9} \). Consequently, using \( p_1 \) enables the architecture to be simplified for a limited performance lost.

#### 4.2. Full-parallel SISO decoder architecture

Figure 3 depicts the architecture of the full-parallel SISO decoder. It was firstly designed totally combinatorial, then, a critical path study enabled the insertion of pipeline stages within the structure.

![Combinatorial version of the full-parallel SISO decoder](image)

**Fig. 3.** Combinatorial version of the full-parallel SISO decoder
4.2.1. Reception stage

The reception stage corresponds to steps (1-3) of the Chase-Pyndiah algorithm detailed in section 2.

The syndrome of the incoming vector $R_{tl}'$ can be derived as $S(R_{tl}') = H \times sign(R_{tl}')$ where $H$ is the parity check matrix of the BCH code. A straightforward implementation of such a matrix multiplication is depicted on figure 4. The $H$ matrix, the corresponding parity check equations and the syndrome $S(t_0) = [s_2, s_1, s_0]$ implementation of a (7,4) BCH code are detailed.

$$H = \begin{bmatrix}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}$$

**Fig. 4.** BCH(7,4) code: (a) Parity check matrix (b) Parity check equations (c) Syndrome parallel computation implementation

It can be noticed that some parity check equations have similar terms. For instance, the term $(x_1 \oplus x_0)$ is used in both $s_1$ and $s_2$ computation. This enables a reuse of computation resources for an even more efficient implementation. The parity of the incoming vector $R_{tl}'$ is computed with a similar structure by “xoring” $(n - 1)$ incoming bits.

Selecting the least reliable bits among the incoming vector in parallel requires a sorting network. Such structures are composed of interconnected Compare and Select operators (CS). The interconnection scheme depends on the considered sorting algorithm. Many parallel sorting algorithm are conceivable [14]. However, most of them are optimized for a complete sorting, while the Chase-Pyndiah algorithm only requires a partial sorting (i.e. extracting $p$ minima). Consequently we devised a network optimized, in terms of area and critical path, for the partial sorting of $p=3$ values among $n=32$, as depicted on figure 5. The structure is based on shuffle networks coupled with local minima computation blocks. After the first shuffle stage, $min_1$ is in the lower section while the upper section can either contain $min_2$ or $min_3$ or no minimum. The same reasoning is applied recursively. After 5 shuffle stages, the minimum is determined while 5 values can still be $min_2$ and $min_3$. A local sorting of this 5 values enables the determination of $min_2$ and $min_3$ value. This partial sorting network requires 35 CS and 29 minimum elements. The critical path consists in 9 comparisons stages.

**Fig. 5.** Optimized sorting network for least reliable bits selection

4.2.2. Test pattern processing stage

The test pattern processing stage corresponds to steps (4-5) in the Chase-Pyndiah algorithm detailed in section 2. Instead of being processed sequentially, test patterns are processed in parallel. The syndrome of each test pattern is computed by adding $S(t_0)$ with the position of the inverted reliable bits. The parity management block computes the parity of $R_{tl}'$ considering the parity of $R_{tl}'$ and the detection of an error which is the case when $S(t_0) \neq 0$. Metrics of each test pattern is then computed by adding the contribution of each inverted bit in the current test pattern (least reliable bits, syndrome corrected bits and, the new parity bit). The minimum metric is determined in the DW selection block. The structure is a simple minimum selection tree. The multiplexer selects $R_{tl}' (S(t_i))$ in order to compute test pattern metrics.

4.2.3. Soft output computation stage

The last stage is a duplication of $n$ soft output computation blocks. As shown on figure 6, this block first computes the new reliability $F_{it}$ of each symbol. Since, no competitor is considered, the $\beta$ value is automatically assigned. The $\beta$ value is based on an estimation of the competitor word metric value. It is calculated with the reliability of the corrected bit and the least reliable bits. Then, the extrinsic information is computed and damped by the coefficient $\alpha_{it}$ which is devised to be a power of 2 making the multiplication a simple bit shifting. Finally, the channel information is added to generate the soft output $R_{tl}'(S(t_i))$. Within this block, all computation are performed in sign and magnitude format. Other arithmetic format were explored but the chosen one requires less computation resources than others.

5. AREA VS THROUGHPUT TRADEOFF

In this section, we compare the parallel and the sequential SISO decoders in terms of throughput, complexity and ef-
Fig. 6. Soft output computation stage

ficiency. Then a pipelined version of the parallel SISO decoder is proposed. Logic syntheses were performed using Synopsys Design Compiler with a ST-microelectronics 90nm CMOS process. The area is transposed in logic gate count. One equivalent logic gate corresponds to the area of a 2-input NAND gate. It enables a more technology-independent measure of the complexity.

5.1. Logic synthesis results

We designed 5 versions of the (32,26) BCH parallel SISO decoder having from 1 to 5 pipeline stages. The 1-pipeline stage version is a full-combinational architecture with register banks only in input and output. Table 2 summarizes synthesis results of the 5 different full-parallel SISO decoder and compare them with duplicated sequential SISO.

$s$ is the number of pipeline stages, $f_{\text{max}}$ is the maximum working frequency reached during synthesis, $A$ represents the area of the design in equivalent gate count, the parallelism rate $P$ corresponds to the number of processed symbols per clock period. $T_{\text{in}}$ is the input throughput such as $T_{\text{in}} = P \times f$ and $T_{\text{out}}$ is the information throughput $T_{\text{out}} = R_p \times P \times f$. Increasing throughput regardless of turbo decoder complexity is not relevant. In order to compare the throughput and complexity of SISO decoders, the efficiency of each decoder is defined as: 

$$\eta = \frac{T_{\text{out}}}{A}.$$

Table 2 shows that the parallel SISO decoder can reach high throughput with low complexity. This clearly shows the higher efficiency of the parallel SISO decoder compared with a duplication of $n$ sequential decoders. Actually a fully parallel structure enables a better reuse of computation and memory resources. For instance, in a parallel SISO decoder, the memory requirement for $R$ and $R'_t$ are $O(2^n q n)$ while in a sequential-based parallel decoder, it is $O(2^n q n^2)$.

Since the resulting throughput is limited by the slower stage, the maximum throughput of a full-iteration module is 14.8Gb/s. In this case, a full-iteration of the proposed architecture (see figure 2b) takes 236Kgates while the previous architecture (see figure 2a) takes 400Kgates.

5.2. Towards a maximal parallelism rate

Previous synthesis results showed the better efficiency of parallel architecture compared with duplicated architectures.

![Diagram](image)

Alternative turbo decoding scheduling for enhanced parallelism rate

Figure 5.2 proposes an alternate product code matrix parallel decoding scheme where $m$ parallel decoders are used for column decoding while row decoding is performed by $n$ $m$-decoders. A $m$-dec decoder can decode $m$ symbols in one clock period and $1 < m < n$ [5]. In such an architecture, the maximum reachable parallelism rate $P = n^2$ can be achieved by using $n$ full-parallel SISO decoder for column decoding and $n$ full-parallel SISO decoder for row decoding. Considering that the TPC decoder includes $2n$ parallel SISO decoders (with $s = 5$ pipeline stages) for $i$ decoding iterations, the turbo decoder output throughput is

$$T_{\text{out}} = \frac{P \times f \times R_p}{i \times t}$$

Placing and routing $2n$ parallel SISO decoders would reduce the maximum working frequency of the parallel SISO decoder. Nevertheless, using the previous equation, an information throughput $T_{\text{out}} = 10$Gb/s is reached when $f > 88$MHz which is a most probably achievable frequency. Furthermore,
a reasonable working frequency of $f=300$MHz leads to a (32,26)$^2$ BCH product code turbo decoder with an information throughput $T_{out}=33.7$Gb/s. The total area of such a parallel turbo decoder is $A = 10\mu$m$^2$. The achieved parallelism rate ($P = n^2 = 1024$) is similar to a full parallel $n = 1024$ LDPC decoder. The number of interconnection among the product code turbo decoder is $I_n(TPCD) = 2 \times n^2_{BCH} \times q$ while an equivalent full parallel 1024-LDPC decoder would have $I_n(1024-LDPC) = n_{LDPC} \times n \times d_v$ where $d_v$ represents the variable node degree. Consequently, as long as $d_v > 2$, the following inequation is verified: $I_n(TPCD) < I_n(1024 - LDPC)$.

6. CONCLUSION

High throughput TPC decoder architecture complexity is made prohibitive by the amount of memory usually required for data interleaving and pipelining. In this article, we proposed an innovative product code matrix decoding scheduling which enables any interleaving resource to be removed. The resulting architecture requires a full-parallel SISO decoder able to decode $n$ symbols in one clock period. Such a SISO decoder architecture is described and includes a new optimized parallel sorting network. ASIC based-logic synthesis showed the better efficiency of the proposed architecture. Actually, compared to a previous architecture, the area is reduced while the throughput is the same. Finally, in order to enhance parallelism rate and throughput, a slightly different scheduling is proposed. With such a scheduling, the maximum parallelism rate $O(n^2)$ can be achieved. An area estimation showed that a (32,26)$^2$ BCH code can be decoded at 33.7Gb/s with an estimated silicon area of 10$\mu$m$^2$.

7. REFERENCES


