This paper presents a method for decoding high minimal distance ($d_{\text{min}}$) short codes, termed Cortex codes. These codes are systematic block codes of rate 1/2 and can have higher $d_{\text{min}}$ than turbo codes. Despite this characteristic, these codes have been impossible to decode with good performance because, to reach high $d_{\text{min}}$, several encoding stages are connected through interleavers. This generates a large number of hidden variables and increases the complexity of the scheduling and initialization. However, the structure of the encoder is well suited for analog decoding. A proof-of-concept Cortex decoder for the (8, 4, 4) Hamming code is implemented in subthreshold 0.25-µm CMOS. It outperforms an equivalent LDPC-like decoder by 1 dB at BER=$10^{-5}$ and is 44 percent smaller and consumes 28 percent less energy per decoded bit.

Keywords: Cortex codes, analog decoding, sub-threshold CMOS circuit.

I. Introduction

Over the last fifteen years, many improvements have been made in error correcting techniques due to the invention of turbo codes by Berrou in 1993 and the rediscovery of low-density parity check (LDPC) codes by Mackay in 1995. Besides the usual digital implementation of turbo and LDPC decoders, some research has shown the feasibility of implementing them using analog networks to fully exploit the information available from the channel [1]-[4]. These theoretical works were followed by fully operational prototypes using either sub-threshold-biased CMOS [5]-[8] or forward-biased bipolar transistors [9], [10] to build the computing cells. The main advantage of such analog implementations is that the analog network does not require any scheduling, meaning that the network converges to a stable state corresponding to the code word. Moreover, the analog decoder power consumption is mostly static; therefore, it remains the same as the throughput increases. That is not the case in a digital decoder for which the power consumption is proportional to the clock frequency, that is, to the data throughput. Another characteristic of an analog decoder is the parallel data processing, which helps increasing the data rate. This is also a disadvantage as it implies a direct relationship between the frame length and the size of decoder. Analog decoding has been thus confined to a frame length of at most a few hundred bits, despite some architectural solutions to cope with longer frame lengths [4], [11]. However, for these frame lengths, it is well known that turbo and LDPC codes are not at their best [12]. There are many applications requiring such short codes and for which analog decoding can be of interest if codes other than turbo and LDPC are used, such as sensor networks and volatile...
memories. A potential candidate is the Cortex code family. Cortex codes were invented by Carlach, from France Télécom, in 1999 [13] in an attempt to provide short codes with good minimal distances. The construction is multistage and uses a short base code to build the overall code. The base codes are interconnected through interleavers. Decoding has not been successful to date [14], [15], which has led to a drop in interest in such codes. This failure has been attributed partly to the fact that the resulting parity check matrix is not sparse and partly to the difficulty of correctly scheduling the complex digital network. Indeed, when decoding large frames, managing the large number of variables between decoding stages is not a simple matter.

An analog implementation seems a good choice because, once the input data is fed to the decoder, it converges naturally and continuously to a steady state corresponding to the most likely codeword.

This paper proposes a graph-based decoding structure which lends itself to an analog implementation. The background of the study is the (8, 4, 4) extended Hamming code because it is a simple and short code. A proof-of-concept Cortex analog decoder for this code is designed, tested, and compared to its equivalent LDPC-like decoder.

The remainder of this paper is organized as follows. Section II presents the construction of Cortex codes. Section III describes the decoder’s structure along with the decoding algorithm used. Section IV deals with the design of the decoders using PMOS Gilbert multipliers. In section V, experimental results of a proof-of-concept Cortex decoder are shown and compared to an LDPC-like decoder. Finally, section VI concludes the paper.

II. Cortex Codes

1. General Construction

Cortex codes are built using the divide and conquer principle. The idea is to split the whole frame of length \( k \) into \( m \) sub-frames of length \( b \), encode each sub-frame using a small base code, and then, after interleaving all of the resulting redundancy bits over the full length of the frame, to re-encode them using another encoding stage. This encoding process is repeated \( s \) times as shown in Fig. 1. The variables \( y_{ji} \) between each encoding stage are named hidden variables. The final encoding stage \( s \) outputs the redundancy bits \( r_{ij} \) which are concatenated to the initial frame to form the codeword. Hence, the code is of rate 1/2. The general form of the generator matrix is given as:

\[
G_{(s, k, d_{min})} = \begin{bmatrix} I_k & P_1 & P_2 & \cdots & P_s \end{bmatrix},
\]

where \( I_k \) is the identity matrix of length \( k \), \( P_i \) is the square matrix of size \( k \) representing the \( i \)-th encoding stage and is given by

\[
P_i = \begin{bmatrix} P_{i1} & 0_{kb} & \cdots & 0_{kb} & 0_{kb} \\
0_{kb} & P_{i2} & \cdots & 0_{kb} & 0_{kb} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0_{kb} & 0_{kb} & \cdots & P_{is} & 0_{kb} \\
0_{kb} & 0_{kb} & \cdots & 0_{kb} & P_{is} \end{bmatrix},
\]

where each \( P_{ij} \) is the matrix representing a base code, and \( 0_{kb} \) is a square null matrix of size \( b \).

Although it is not mandatory, it is simpler to use the same encoding stage built from the same base code and the same
interleaver throughout the encoding structure. Moreover, the base code is not necessarily a Cortex code itself. In [13], the same base code is used in the encoding stages, namely, the (8, 4, 4) Hamming code, but the interleavers are not all identical. The number of encoding stages used to build the code is directly related to the minimal distance. As shown in [16], when random interleavers are used and $s \to \infty$, the Cortex codes behave as random linear codes. Therefore, as $s$ increases, $d_{\text{min}}$ increases too.

2. Building the (8, 4, $x$) Hamming Code from the (4, 2, 2) Hadamard Code

The (8, 4) Hamming code has a Cortex construction using the (4, 2, 2) Hadamard code as a base code, which is denoted by $P_2$ in Fig. 2. Its matrix $P_2$ is given by (3). Using two encoding stages made of two base Hadamard encoders separated by one interleaver as shown in Fig. 2(a) yields a minimal distance of three. This is easy to verify from the generator matrix $G_{(8,4,3)}$ given by (4), which is obtained from Fig. 2(a) and (1), (2), and (3). Adding an encoding stage identical to the first two and using the same interleaver for simplicity as shown in Fig. 2(b) increases $d_{\text{min}}$ to four. This is the maximum $d_{\text{min}}$ that can be obtained for this block length. The generator matrix $G_{(8,4,4)}$ is given by (5). Thus, the (8, 4, 4) Hamming code is obtained.

This illustrates, by a simple example, how adding encoding stages increases the minimal distance. Thus, the modular construction of Cortex codes allows larger codes to be built with good minimal distance, using either the Cortex Hamming codes or the (4, 2, 2) Hadamard code as base bricks. This paper focuses, however, only on the Cortex (8, 4, 4) Hamming code to describe the proposed decoding method.

III. Cortex Decoder

The decoding structure resembles the encoding structure except that the edges are now bi-directional and the corresponding base code decoder replaces each base code encoder part. Due to the structure, digital decoding does not seem to be an appropriate method for this type of code. For instance, for the simple (8, 4, 4) Hamming digital decoder, considering that a decoding stage performs its task in a single clock cycle, three clock cycles are required to update the data on each side of the decoder. Because more than one iteration is required, usually six as in a digital turbo decoder, a total of eighteen ($6 \times 3$) clock cycles are needed to decode one frame.

On the contrary, in an analog implementation, once the data has been fed to the decoder, it is able to output the codeword after a time period at most equal to the frame duration, and usually much less than that [8]-[10], that is, eight clock cycles. Thus, analog decoding avoids complex message-passing scheduling and should definitively offer a reduced latency compared to a digital solution as the size of the decoder increases. This further motivates us to use an analog solution to attempt to successfully decode Cortex codes.

1. Decoding Algorithm

Note that, from (5), the resulting parity check matrix is dense, in the sense that it contains many ones. This is usually considered to yield unsatisfactory decoding performance when a belief propagation algorithm, such as the sum-product algorithm (SPA) for which operands are probabilities, is used. A non-sparse parity-check matrix implies a large number of cycles, which renders probability computation intractable [17]. Nevertheless, the SPA algorithm is very simple to implement in analog circuits as the previous works showed, hence the motivation for using it.
Fig. 3. (8, 4, 4) Hamming decoder using the SPA and based on the Tanner graph. One of the shortest cycles is shown with dashed lines.

Fig. 4. Cortex decoder construction: (a) bipartite graph of (4, 2, 2) Hadamard code and (b) Cortex graph of the (8, 4, 4) Hamming code. One of the shortest cycles is shown with dashed lines.

2. Decoding Based on Tanner Graph

Codes are usually represented by means of graphs. Stemming directly from the generator matrix $G_{(8,4,4)}$, a decoder based on the Tanner graph when the SPA is used for the Cortex (8, 4, 4) Hamming code is represented in Fig. 3. First, note that the node’s degree is not constant and is at most four. In fact, the nodes degree increases with the code length (if $d_{\min}$ increases as well), which implies a higher complexity. Second, analyzing Fig. 3 shows that it contains 28 cycles and that the girth, that is, the length of the shortest cycle is four (one of them is represented with dashed lines in Fig. 3). Actually, there are four cycles at the girth. Because the girth is low and there are many short cycles, decoding is not optimal [17]. Because the parity check matrix is dense, decoding based on Tanner graph is termed LDPC-like, as it uses the same type of structure as that which is usually employed for LDPC decoders.

3. Decoding Based on Cortex Graph

Another representation is possible which takes into account the Cortex structure. Since the code is built based on the (4, 2, 2) Hadamard code, the (4, 2, 2) Hadamard decoder using the SPA is used as a base decoder. The bipartite graph representing the (4, 2, 2) Hadamard decoder is shown in Fig. 4(a). This version of the usual graph is simplified by noting that nodes of degree two perform no computation. Figure 4(b) shows the bipartite graph representing the Cortex (8, 4, 4) Hamming decoder built from the simplified (4, 2, 2) Hadamard graph. There are some interesting observations that can be made regarding the Cortex graph shown in that figure. First, all the nodes have the same degree. That is, the number of edges connecting them is always three. Any Cortex decoder built from the (4, 2, 2) Hadamard will have this property. Intuitively, this is interesting in terms of complexity because longer codes do not imply a complexity increase of the computing nodes. Second, the number of cycles is low (only six) and the girth is six (one of them is represented as a dashed line in Fig. 4(b)). This relatively large value for such a small code is obtained because the base code graph is cycle-free. Having cycle-free base code graphs helps keep the large girth once such graphs are connected through interleavers. Otherwise, the girth of the overall graph would be limited to that of the base code’s graph. Hence, decoding based on the proposed representation should yield better performance than decoding based on Tanner graph. Moreover, it should also have lower complexity. Both characteristics are summarized in Table 1.

4. Behavioral Simulation

Both (8, 4, 4) Hamming analog decoders are modeled using Simulink as described in [4]. The result of the behavioral simulation for an additive white Gaussian noise (AWGN) channel is shown in Fig. 5. For comparison, the maximum-likelihood decoding performance is also presented. The proposed Cortex decoding method is near optimal and outperforms LDPC-like decoding by 1 dB. This can be directly attributed to the comparatively high number of cycles and lower girth of the Tanner graph.

IV. Decoder Implementation

1. Design Guidelines

In [18], it was shown that an impaired transistor exponential

| Table 1. Cycle and complexity comparison. |
|------------------|------------------|------------------|
| Decoder          | Cortex           | LDPC-like        |
| Computing nodes  | 12               | 16               |
| Maximum node degree | 3              | 4               |
| Bi-directional connections | 8       | 12               |
| Number of cycles | 6                | 28               |
| Girth            | 6                | 4                |
| Number of girth cycles | 2       | 4                |
I-V characteristic, used to translate log-likelihood ratios (LLRs) represented by voltages into probability represented by currents, degrades the decoder’s performance. To avoid this, design guidelines were proposed in [19] for sub-threshold biased MOS. Only the main results are given here.

For weak-inversion, a better exponential MOS I-V characteristic is obtained if the inversion coefficient IC is much smaller than 1 and if the subthreshold slope factor $m_S$ [20] is constant. A constant slope factor is obtained if $V_{BS} = 0$. This also reduces the dependence of the drain current on the drain-source voltage $V_{DS}$. The reduction of the threshold voltage $V_T$ due to drain-induced barrier lowering and hence an increase of the drain current by electron injection from the source are limited if $V_{BS} = 0$. Finally, the exponential variations of the drain current are accentuated if either $V_{DS} < m_S \times U_T$ or $V_{BS} > 0$, where $U_T$ is the thermal potential equal to $26 mV$ at 300K. Fixing by design $V_{DS} \approx 4U_T$ and $V_{BS} = 0$ avoids such undesired effects. However, $V_{BS} = 0$ can be only guaranteed if PMOS transistors are used because they all have an individual substrate formed by the N-well. Because the source can be physically tied to the bulk, $V_{BS} = 0$.

2. Computing Nodes

Unlike other CMOS analog decoders [5]-[8] which require a circuit to convert LLR from the channel into probability and a second circuit to multiply probabilities, the proposed decoder uses a PMOS-based Gilbert cell to perform both at the same time as in BJT-based decoders [9], [10]. This simplifies the overall circuit and thus reduces the power consumption. To further reduce the $V_{DS}$ variations, the cells have a symmetrical design. Dummy transistors are added so that there is the same number of MOSs between the power-rails in each branch. As an example, the soft-equality node is given in Fig. 6. Based on these design guidelines, the aspect ratio of the PMOS can be calculated as

$$\frac{W}{L} = \frac{I_{bias}}{2IC_0\mu_0C_{ox}U_T^2},$$

where $I_{bias}$ is the bias current equal to $1 \mu A$, $\mu_0$ is the carrier mobility, and $C_{ox}$ is the gate capacitance per unit area. We set the IC to 0.05. At less than 0.1, good exponential behavior is obtained, and above 0.01, the transistor aspect ratio of the Gilbert cell is reduced. The latter directly affects the maximum speed at which the decoder runs. Therefore, we chose to use minimum length transistors, which yield $W/L = 9/0.25$. The benefits of using PMOS rather than NMOS to build the computing cells come at the expense of a size increase. The aspect ratio of the transistors, 36, is larger than that used in other reported NMOS decoders, such as 10 in [8], due to a lower hole mobility (about three times) than that of an electron. The room taken by the N-well should be added, which further increases the area of the cell to about five times the area of an equivalent NMOS computing cell with the same IC and the same $I_{bias}$. The load NMOSs are $W/L = 5/0.25$. There are three such Gilbert multipliers per node of degree three and six per node of degree four.

3. Decoders Block Diagram

The block diagram of both the Cortex and the LDPC-like decoder are shown in Fig. 7. The only part that changes is the
Fig. 7. Decoder block diagram. The difference between the LDPC-like and Cortex decoders is the extrinsic block, which corresponds to Fig. 3 and Fig. 4(b), respectively.

extrinsic information block, which corresponds to the graphs shown in Fig. 3 and in Fig. 4(b) for the LDPC-like decoder and Cortex decoder, respectively. Considering the low number of analog inputs, we fed them in parallel to the decoders. Besides the clock, reset and write signals were added to synchronize the circuit with the outside test bench.

V. Experimental Results and Discussion

Both decoders were implemented on a chip. A microphotograph of the integrated circuit is given in Fig. 8 along with the floor plan. The total silicon area is 2.17 mm × 2.6 mm. Although they use the same computing cell, due to the higher node degree and larger number of interconnections, the LDPC-like decoder is almost twice as large as the Cortex decoder. To enable repeatable and reliable tests, random frames were generated, encoded, modulated, and altered by an AWGN channel and then demodulated on an FPGA. The resulting digital signals corresponding to the noisy LLRs were transferred onto a set of synchronized DACs to provide the eight analog inputs to the decoders.

The BER results are shown in Fig. 9 for the Cortex and LDPC-like decoders. As expected from the behavioral simulations, the Cortex decoder outperformed its LDPC-like counterpart by 1 dB at BER=10^-5. Table 2 summarizes the main features of some existing analog decoders and compares them to the two proposed decoders. Due to the lower complexity of the proposed decoders, their energy per decoded bit is two orders of magnitude lower than that of the sub-threshold analog decoders [6]-[8] using the SPA. The Tanner decoder for the (8, 4, 4) Hamming code (an LDPC-like decoder) presented in [5] performs better than the Cortex decoder in terms of energy consumption but is 1 dB worse than the Cortex decoder at a BER = 10^-6. Besides a smaller minimum size feature, 0.18-µm instead of 0.25-µm, the difference in terms of size is explained by the fact that a PMOS computing cell is five times larger than an NMOS one.

Unlike all the other decoders in Table 2, the decoder implemented in [21] does not use the SPA but the min-sum (MS) algorithm. NMOS and PMOS current mirrors are used to implement the algorithm instead of multipliers in the other decoders. Although it works in strong inversion, it is still slower than BJT decoders. Compared to the Cortex decoder, the MS LDPC is larger, given that it uses a smaller minimum length technology (0.18-µm) than the one used for the Cortex decoder, while consuming five times more energy per decoded bit. However, its data throughput is three times higher thanks to the MOS operating in the saturation region.

It is striking that the decoder with the second lowest energy consumption per bit is the BTJ-based decoder from [9]. This is due to a high throughput as explained in the introduction. This
shows that the BJT-based analog decoder should not be discarded when low energy consumption is required. In that respect, the decoder presented in [10] also achieves good performance when compared to most of the subthreshold CMOS decoders.

Comparing the two proposed decoders shows that, as expected, the Cortex decoder consumes 28 percent less energy per decoded bit than the LDPC-like decoder. The discrepancies between behavioral simulations and actual measurements are about 0.3 dB, which is consistent with the mismatch described in [8].

VI. Conclusion

A modular method to easily decode Cortex codes has been proposed. The decoding algorithm applied is the sum-product algorithm. An LDPC-like decoding method based on the Tanner graph, is outperformed by the Cortex decoder in terms of complexity (44 percent smaller), power consumption (28 percent less), and decoding performance (1 dB at BER=10⁻⁵). The first working prototype confirmed all these. Fabricated in a CMOS 0.25µm technology, these decoders are also the first reported decoders that use PMOS-based computing cells. Further works are required to build useful codes using the proposed method.

References


Jorge Pérez-Chamorro received the degree of electronics engineer from the Universidad de los Andes, Bogota, Colombia, in 2005. During his doctoral studies, he focused on channel coding and decoding and on analog circuits design. He was an invited researcher at the High Capacity Digital Communications Laboratory of the ECE Department of the University of Alberta, Edmonton, Canada, in 2007. He received his PhD (with highest honours) from TELECOM Bretagne, Brest, France, in 2009. Afterwards, he joined INVIA, a start-up company based in Aix en Provence, France, specializing in the design of IPs for secure platforms. His research interests are error correcting codes, bipartite graphs, decoding algorithms, and code compression.

Cyril Lahuec received the BSc (Hons) degree from the University of Central Lancashire, UK, in 1993, and his MEng and PhD degrees from Cork Institute of Technology, Ireland, in 1999 and 2002, respectively. He was with Parthus Technologies (now Ceva), Cork for his PhD work and then as a consultant. He joined the Electronics Department of TELECOM Bretagne as a full-time lecturer in 2002. His research interests are in frequency synthesis, analog IC design, channel decoding, and biomedical applications.

Jorge Pérez-Chamorro

Cyril Lahuec

Jorge Pérez-Chamorro received the degree of electronics engineer from the Universidad de los Andes, Bogota, Colombia, in 2005. During his doctoral studies, he focused on channel coding and decoding and on analog circuits design. He was an invited researcher at the High Capacity Digital Communications Laboratory of the ECE Department of the University of Alberta, Edmonton, Canada, in 2007. He received his PhD (with highest honours) from TELECOM Bretagne, Brest, France, in 2009. Afterwards, he joined INVIA, a start-up company based in Aix en Provence, France, specializing in the design of IPs for secure platforms. His research interests are error correcting codes, bipartite graphs, decoding algorithms, and code compression.

Cyril Lahuec received the BSc (Hons) degree from the University of Central Lancashire, UK, in 1993, and his MEng and PhD degrees from Cork Institute of Technology, Ireland, in 1999 and 2002, respectively. He was with Parthus Technologies (now Ceva), Cork for his PhD work and then as a consultant. He joined the Electronics Department of TELECOM Bretagne as a full-time lecturer in 2002. His research interests are in frequency synthesis, analog IC design, channel decoding, and biomedical applications.

Michel Jézéquel

Michel Jézéquel received the degree of “Ingénieur” in electronics from the École Nationale Supérieure de l’Électronique et de ses Applications, Paris, France, in 1982. From 1983 to 1986 he was a design engineer at CIT ALCATEL in Lannion, France. Then, after experience with a small company, he took a one-year course on software design. In 1988, he joined the École Nationale Supérieure des Télécommunications de Bretagne where he is currently involved with design issues of analog channel decoders and related topics.

Jorge Pérez-Chamorro

Cyril Lahuec

Michel Jézéquel