Towards an Accurate High-Level Energy Model for LDPC Decoders

Jérémy Nadal*, Simon Brown*, Elsa Dupraz† and François Leduc-Primeau*

*Department of Electrical Engineering, École Polytechnique de Montréal, Montreal, CA.
†Department of Mathematical and Electrical Engineering, IMT Atlantique, Lab-STICC, UBL, Brest, France
Emails: *{jeremy.nadal, simon.brown, francois.leduc-primeau}@polymtl.ca, †elsa.dupraz@imt-atlantique.fr

***Abstract*—Estimating the energy consumption of LDPC decoders is a long and difficult task due to the large number of factors involved. Modern circuit synthesis tools can provide a satisfactory estimation of the power consumption, but this requires that the circuit be already implemented and it can take hours to provide the estimate. Currently, no accurate models are available to evaluate the decoding energy early in the design process. We propose a high-level energy model for flip-flop memory elements in LDPC architectures. The originality of the model is that it can analytically evaluate the variation of the energy due to the switching activity of the circuit gates, depending on the probability mass function (PMF) of the circuit inputs. Such PMFs are obtained through an adapted density evolution method that we propose. Therefore, the energy can be profiled for each decoding iteration and SNR value while considering several architecture choices. We illustrate the validity of the model by comparing the obtained energy estimates with measurements based on circuit simulations.**

***Index Terms*—LDPC, ASIC, density evolution, energy efficiency.**

I. Introduction

As communication speeds increase and data transmission channels become more complex, the use of error-correction codes becomes increasingly critical. Low-density parity-check (LDPC) codes [1] are widely used in modern communication systems such as 5G networks [2]. For future 6G networks, preliminary studies consider that peak data rate will be up to 50 times higher that of existing 5G networks [3]. Therefore, there is a need for massively parallel and pipelined hardware architectures, which raises the issue of their energy consumption. This impacts the battery life of user devices and the cost of electrical consumption in base stations for the operators. Particularly, energy efficiency now becomes an important metric for studies of 6G networks [3], and must be taken into account when designing LDPC codes, hardware architectures, and circuits.

Developing novel energy reduction techniques requires an accurate understanding of how the code construction, decoding algorithm, and hardware architecture affect energy consumption. Previous work studied the energy consumption of coding schemes in the asymptotic regime as the block error probability goes to zero [4], [5]. However, these results do not provide a way to optimize the energy consumption of finite-length codes and decoders. Reference [4] also proposed a non-asymptotic energy model for regular LDPC codes. Their model however is based on post-layout simulations of specific circuit architectures and does not take the actual switching activity into account, which prevents using it to optimize the decoder architecture. As such, there is a lack of accurate analytical methods that can predict the energy consumption of an LDPC decoder before it is fully implemented. Current methods are mainly empirical, where accurate estimation of the energy consumption requires time-consuming post-synthesis or post-layout simulations to generate signal activities [6].

Therefore, we propose in this paper a model to analytically estimate the energy consumption of LDPC decoders. As a first step, we focus on the energy consumption of flip-flop (FF)-based components such as registers and memories, since these components represent most of the energy consumption of the decoder [7]. The main originality of the approach is in the proposal of a clock-gated FF energy model where switching activity is taken into account. To predict the bit switching probabilities, a density evolution (DE)-based methodology is proposed. DE is a powerful analysis tool normally used for predicting the bit-error rate (BER) performance of LDPC decoders and designing good LDPC codes [8]. The proposed method aims to adapt DE to predict the probabilistic behaviour of signals in LDPC decoder circuits.

The rest of this paper is organized as follows. Section II describes the offset min-sum (OMS) decoding algorithm and its DE analysis. Section III proposes a novel energy model for clock-gated FF based on bit switching probabilities. In Section IV, these switching probabilities are evaluated through novel DE-based methodologies adapted for most LDPC decoder architectures. The proposed energy model is then compared and validated in Section V, with post-synthesis energy measurements from a highly parallel Δ-update LDPC decoder architecture. Finally, Section VI concludes this paper.

II. LDPC Decoder

A. Decoding algorithm

We consider an Additive White Gaussian Noise (AWGN) channel described by $y_i = x_i + w_i$, where y_i is the channel output, x_i is the i^{th} transmitted modulated symbol, and w_i is the additive Gaussian noise term of variance σ^2. For simplicity, BPSK modulation is considered. The received and quantized

This work was supported by grant ANR-17-CE40-0020 (project EF-FECtive) of the French National Research Agency ANR and by grant 2021-NC-286323 of Fonds de recherche du Québec – Nature et technologie.

channel log-likelihood ratio (LLR) L_i of the i^{th} coded bit is expressed as

$$L_i = \text{sat}_Q \left\lfloor \frac{2\alpha y_i}{\sigma^2} + \frac{1}{2} \right\rfloor, \quad (1)$$

where y_i is the received signal, σ^2 is the noise variance, α is a constant scaling factor, $\lfloor . \rfloor$ is the floor operator and sat_Q is a saturation operator which ensures that $L_i \in [-Q, Q]$, $Q \in \mathbb{N}^*$ being the maximum LLR magnitude.

Most common LDPC decoding algorithms are variants of the Sum-Product algorithm [9], in which messages are iteratively sent between variable nodes (VNs) and check nodes (CNs). Let us denote by $\gamma_{i\to j}^{(\ell)}$ the check-to-variable (C2V) message sent from CN i to VN j for decoding iteration $\ell \geq 0$, and by $\lambda_{j\to i}^{(\ell)}$ the V2C message sent from VN j to CN i at iteration ℓ. Initially, we set $\gamma_{i\to j}^{(0)} = 0 \ \forall (i,j)$, and $\lambda_{j\to i}^{(0)} = L_i \ \forall j$. In addition, $\mathcal{V}_i$ is the set of VN indices connected to CN i, and $\mathcal{C}_j$ is the set of CN indices connected to VN j. In this paper, we consider the widely used OMS decoder [10], which can be formulated through the following equations:

$$\gamma_{i\to j}^{(\ell)} = \prod_{j' \in \mathcal{V}_i \setminus j} \mathfrak{s}(\tilde{\lambda}_{j'\to i}^{(\ell)}) \times \max\Big(\min_{\forall j' \in \mathcal{V}_i \setminus j} |\tilde{\lambda}_{j'\to i}^{(\ell)}| - \beta, 0 \Big),$$
$$\lambda_{j\to i}^{(\ell)} = L_j + \sum_{i' \in \mathcal{C}_j \setminus i} \gamma_{i'\to j}^{(\ell)}, \quad (2)$$

where $\beta \in \mathbb{N}$ is an offset to be optimized, $\tilde{\lambda}_{j\to i}^{(\ell)} = \text{sat}_Q(\lambda_{j\to i}^{(\ell)})$, and $\mathfrak{s}(.)$ is the signum function. The decision on the j^{th} bit is taken by evaluating the sign of the a-posteriori (AP) belief $\Lambda_j^{(\ell)}$ assigned to VN j and defined as $\Lambda_j^{(\ell)} = \lambda_{j\to i}^{(\ell)} + \gamma_{i\to j}^{(\ell)}$.

B. Density evolution

Initially developed for Sum-Product decoding in [8], DE is an analysis tool which uses probabilistic properties of belief propagation to predict the behavior of a decoder at each iteration. That is, assuming that a given LDPC code is cycle-free, we can accurately predict the performance of a decoder at each iteration by expressing the successive probability distributions of messages exchanged in the decoder. The works presented in [11], [12] extend DE to quantized OMS decoding. Both the initial DE analysis [8] and the extensions provided in [11], [12] assume that the length of the LDPC code is infinite.

It is shown in [8] that the BER performance of the decoder is the same for any transmitted codeword. Therefore, to simplify equations, DE considers that the all-zero codeword is transmitted. In such case, the channel LLR cumulative distribution function (CDF), denoted $\Phi_L(k), k \in [-Q, Q]$, is

$$\Phi_L(k) = \frac{1}{2} + \frac{1}{\sqrt{4\sigma^2}} \text{erf}\left(\left(k + \frac{1}{2}\right) \frac{\sigma^2}{\alpha} - 1 \right), \quad (3)$$

where $\text{erf}(x) = 2/\sqrt{\pi} \int_0^x \exp[-t^2] dt$ is the Gauss error function. The probability-mass function (PMF) of the channel LLR, denoted $P_L(k)$, can then be expressed as $P_L(k) = \Phi_L(k) - \Phi_L(k-1), \ \forall k \in [-Q+1, Q-1]$, $P_L(-Q) = \Phi_L(-Q)$ and $P_L(Q) = 1 - \Phi_L(Q)$. The equation of the C2V messages PMF $P_{\gamma_{i\to j}}^{(\ell)}$, the V2C messages PMF $\lambda_{j\to i}^{(\ell)}$ and the AP $P_{\Lambda(k)}^{(\ell)}$ can be found in [11], [12] . The PMF of the saturated V2C message $\tilde{\lambda}_{j\to i}^{(\ell)}$ can then be obtained by respectively summing all $P_{\lambda_{j\to i}}^{(\ell)}(k), k > Q$ and $k < Q$, to $P_{\lambda_{j\to i}}^{(\ell)}(Q)$ and $P_{\lambda_{j\to i}}^{(\ell)}(-Q)$. The BER at iteration ℓ, denoted $B_{e,\text{inf}}^{(\ell)}$, can be calculated by evaluating the probability that Λ_j is negative.

DE equations are only valid assuming a cycle-free code of infinite length, and provide asymptotic BER performance. The cycle-free assumption implies that messages are statistically independent, whereas the infinite-length assumption implies that the observed noise power within the received block averages to the channel noise power. This second assumption is what makes the DE unable to characterize the BER of finite-length codes [13]. However, the observed channel error rate can be modeled as a Gaussian random variable [14]. As shown in [13], [15], this model provides a good match to Monte-Carlo simulations in the waterfall region for moderately large code lengths.

III. Proposed energy model

This section proposes an energy model for clock-gated FFs, which serve to build registers and memory units. All data storage in an LDPC decoder is handled by one of these two components, and they represent most of the energy consumption of the decoder [7]. Therefore, an accurate energy model for an LDPC decoder can be developed based on an accurate model of the FF energy consumption. Specifically, most of the energy consumed by a FF is the result of the charge and discharge of the circuit's internal capacitance as a result of inputs switching between low and high voltage values. Therefore, it is critical to consider switching activity of the decoder signals when modeling energy consumption.

In a synchronous hardware architecture, each signal S, composed of $N_b = \lceil \log_2(2Q+1) \rceil$ bits, can be associated with a random variable of PMF $P_S^{(c)}(k)$ at a given clock event c. Equivalently, S can be represented in a bitwise format, with b_i being the i^{th} bit of S. Each bit b_i is random (but not independent or identically distributed) with probability $P_{b_i}^{(c)}$ of being equal to 1. Let us denote $\boldsymbol{P}_S^{(c)} = [P_S^{(c)}(-Q), ...P_S^{(c)}(Q)]$ the length-$2Q+1$ PMF vector of S, and $\boldsymbol{P_b^{(c)}} = [P_{b_0}^{(c)}, ..., P_{b_{N_b-1}}^{(c)}]$ the length-N_b probability bit vector. For sign-magnitude (SM) binary representation, the probability bit vector can be derived using $\boldsymbol{P_b^{(c)}} = \boldsymbol{T}\boldsymbol{P}_S^{(c)}$, where $\boldsymbol{T}$ is a $N_b \times (2Q+1)$ transformation matrix. Each element $T_{i,j}$ of $\boldsymbol{T}$ is defined as

$$T_{i,j} = \begin{cases} \lfloor \ |j - Q - 1| \times 2^{-i+1} \ \rfloor \mod 2 & \text{if } 1 \leq i < N_b, \\ \frac{1+\mathfrak{s}(Q+1-j)}{2} & \text{if } i = N_b. \end{cases}$$

When the two's complement (2C) representation is used, the transformation matrix is $\boldsymbol{C}$, with $C_{N_b,j} = T_{N_b,j}, \forall j$, $C_{i,j} = T_{i,j}, \forall i$ when $j \in [Q+1, 2Q+1]$ and $C_{i,j} = \lfloor \ (2^{N_b} + j - Q - 1) \times 2^{-i+1} \ \rfloor \mod 2$ when $j \in [1, Q]$ and $i \neq N_b$. Note that the SM and 2C representations are often used together in the same decoder architecture.

Using these transformation matrices, the bit probabilities can be derived for any signal if their associated PMFs $P_S^{(c)}$ are

known. It is then possible to estimate the energy consumption based on the bits switching activity. In this section, we propose to develop such a model for clock-gated flip-flop components.

A. Flip-flop and register energy model

The energy model we propose considers FFs whose input clock ports are connected to a clock gating unit (CGU). The CGU provides the ability to disable the FF clock when no new value needs to be stored, which greatly reduces its energy consumption. To derive the energy model, the following three assumptions are considered:

(A1): The CGU enable signal state, denoted $S_e(c) \in \{0,1\}$, is considered fully known. This assumption is quite realistic as enable signals are scheduled by a dedicated controller unit. We also define $S_{\overline{e}}(c) = 1 - S_e(c)$.

(A2): The input bit state of the FF is modeled as a random variable. We respectively define $P_+^{(c)}(r)$ and $P_-^{(c)}(r)$ the probabilities that the r^{th} input bit of a length-R flip flop array (register) is equal to 1 after ("+") and before ("−") the front edge of a given clock event c. Note that we have $P_-^{(c)}(r) = P_+^{(c-1)}(r)$. This assumption is perfectly realistic since the decoder input consists of noisy channel samples, which are random by nature.

(A3): It is considered that the input bit state before and after a clock event are statistically independent. Therefore, the probability $P_s^{(c)}(r)$ that the r^{th} input bit switches state at the clock event number c is $P_s^{(c)}(r) = P_+^{(c)}(r)(1-P_-^{(c)}(r)) + (1-P_+^{(c)}(r))P_-^{(c)}(r)$. The validity of this assumption rests on the fact that under the assumption of a cycle-free Tanner graph, any two messages associated with distinct VNs and distinct CNs are independent. In practice, LDPC codes contain cycles that introduce correlations between messages. As explained in Section II-B, these correlations usually do not have a significant effect on the decoding in the waterfall region, which is of interest for evaluating energy consumption. Therefore, (A3) is reasonable as long as processing elements are associated with distinct Tanner graph nodes in two successive clock cycles.

In the absence of clock gating, we consider that the energy consumption $E_{FF}^{(c)}(r)$ of the r^{th} FF in a register for clock cycle c comes from the switching activity of the clock and input signals, regardless of the FF output state. Therefore, we have:

$$E_{FF}^{(c)}(r) = E_C + E_{sc} P_s^{(c)}(r)\,, \tag{4}$$

where E_C is the energy consumed by the clock activity and E_{sc} is the energy consumed when the FF input switches state. When adding clock gating, the energy added by the CGU is

$$E_{CG}^{(c)} = E_{ce} S_e^{(c)} + E_{\overline{ce}} S_{\overline{e}}^{(c)}, \tag{5}$$

where E_{ce} and $E_{\overline{ce}}$ correspond to the CGU energy when its input-enable signal is set to 1 and 0, respectively. The parameter E_{ce} depends on the fanout of the CGU sub-clock port. For a clock-gated register composed of W FFs, E_{ce} depends on W since the CGU sub-clock drives W FFs. The

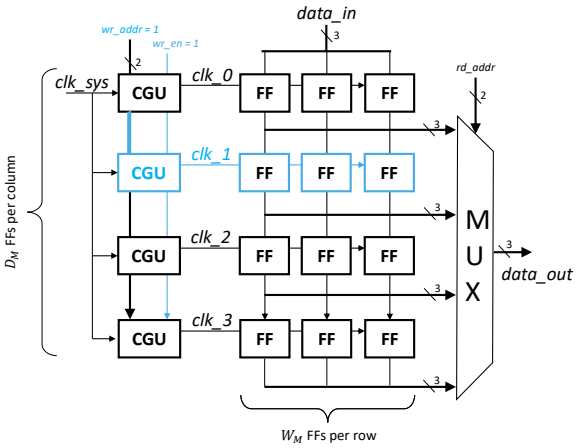

Fig. 1. Reference clock-gated FF-based memory (example with $D_M = 4$ and $W_M = 3$).

energy consumption of such register, denoted $E_W^{(c)}$, is simply the sum of the energy of the W FFs and CGUs:

$$E_W^{(c)} = E_{CG}^{(c)} + S_e^{(c)} \sum_{r=1}^{W} E_{FF}^{(c)}(r) + S_{\overline{e}}^{(c)} E_{s\overline{c}} \sum_{r=1}^{W} P_s^{(c)}(r)$$

$$E_W^{(c)} = \underbrace{E_{CG}^{(c)} + W E_C S_e^{(c)}}_{E_0^{(c)}(W)} + P_S^{(c)} \underbrace{\left(S_e^{(c)} E_{sc} + S_{\overline{e}}^{(c)} E_{s\overline{c}}\right)}_{E_S^{(c)}}, \tag{6}$$

where $P_S^{(c)} = \sum_{r=1}^{W} P_s^{(c)}(r)$, and $E_{s\overline{c}}$ is the energy consumed by the FF when the sub-clock is disabled and input bits are switching. From the above equation, it can be inferred that the total energy consumption of the clock-gated length-W register is composed of a deterministic part ($E_0^{(c)}(W)$) and a probabilistic part ($P_S^{(c)} E_S^{(c)}$). Furthermore, $E_0^{(c)}$ provides a lower bound of the energy consumption, while $E_0^{(c)}(W) + W E_S^{(c)}$ corresponds to an upper bound.

B. Energy model of memory units

We now use the FF energy model presented previously to derive the energy consumption of memories built out of standard-cell FFs. The standard-cell-based memory (SCM) reference architecture is depicted in Figure 1. Although SCMs are not area-efficient, they can be more energy-efficient than static random-access memory (SRAM) macros when operating at low voltages [16]. They are thus a good choice when energy efficiency is the primary design objective.

The memory storage matrix is composed of D_M rows of W_M FFs, with D_M being the depth of the memory and W_M the data width. The i^{th} input data bit is connected to all the FF inputs located at the i^{th} column. All the FF outputs at a given column i are fed to a D-to-1 multiplexer. When considering clock gating, D_M CGUs are instantiated, and each CGU outputs a dedicated sub-clock, generated from the main system clock, for each FF row. These sub-clocks can be enabled or disabled by the CGU depending on the write address and the write-enable signals. If the write-enable signal

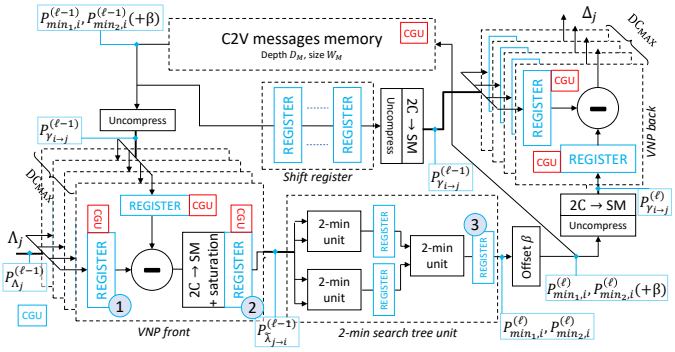

Fig. 2. Internal signal PMFs of a check-node processing unit

is disabled, then the total energy consumption $E^{(c)}_{M,\overline{e}}(D_M)$ of the CGUs and FFs in the memory unit is

$$E^{(c)}_{M,\overline{e}}(D_M) = D_M\Big(E_{\overline{ce}} + E_{s\overline{c}}\sum_{r=1}^{W_M} P^{(c)}_s(r)\Big). \tag{7}$$

It is worth mentioning that the input signal activities have a non-negligible impact on the energy consumption when the memory is not solicited ($D_M W_M E_{s\overline{c}}$ in the worst case). When the write-enable signal is activated, only one FF row is activated, and the energy of the memory unit becomes

$$E^{(c)}_{M,e}(D_M) = E_{ce} + \sum_{r=1}^{W_M} E^{(c)}_{FF}(r) + E^{(c)}_{M,\overline{e}}(D_M - 1). \tag{8}$$

A non clock-gated register can optionally be added at the output of the multiplexer. This adds $E^{(c)}_{W_M}$ to the energy obtained with the above equations.

IV. Derivation of Signal Activities

The previous section described a clock-gated FF energy model based on signal activities, assuming the PMFs of the decoder signals $P^{(c)}_S$ are known. An analytical evaluation of these PMFs is however not straightforward if typical DE methods are employed. In particular, some signals in LDPC decoders are not well modeled with typical DE, such as signals associated with finding the first and second minima in the OMS decoder. Furthermore, the all-zero transmitted codeword assumption cannot be used when evaluating switching activity. This section aims to adapt the DE equations to solve these issues.

A. Signal PMFs in the processing core

Figure 2 shows the architecture of the main processing unit used in a Δ-update LDPC decoder [6]. In this figure, the PMFs for the main signals of interest are labelled. Most of these PMFs can be derived using the equations presented in Section II-B. One exception concerns the signal output of the "2-min search tree" unit, whose function is to find the first and second minimum values of $d_c(i) \triangleq |\mathcal{V}_i|$ input V2C message magnitudes. This output is connected to several registers and to the "C2V message memory." Therefore, the associated PMFs, denoted $P^{(\ell)}_{\min_1,i}$ and $P^{(\ell)}_{\min_2,i}$, are critical to assess the energy consumption of the decoder.

Finding the two smallest values among the V2C message magnitudes $\{|\tilde{\lambda}^{(\ell)}_{j\to i}|\}_{\forall j\in\mathcal{V}_i}$ is equivalent to finding the two largest values among the $\{Y^{(\ell)}_{j\to i}\}_{\forall j\in\mathcal{V}_i}$ variables, with $Y^{(\ell)}_{j\to i} \triangleq Q - |\tilde{\lambda}^{(\ell)}_{j\to i}|$. In such case, $P^{(\ell)}_{\max_u,i}(k) \triangleq P^{(\ell)}_{\min_u,i}(Q-k)$, $u\in\{1,2\}$, corresponds to the $(d_c(i)-r+1)^{th}$-order statistic of the $\{Y^{(\ell)}_{j\to i}\}_{\forall j\in\mathcal{V}_i}$ random variables with CDF $\Phi^{(\ell)}_{Y_{j\to i}}$. The r^{th}-order statistic CDF is equal to the probability that at most $d_c(i)-r$ variable values are strictly superior to k [17]. Therefore, we have

$$\Phi^{(\ell)}_{\max_1,i}(k) = \prod_{j\in\mathcal{V}_i} \Phi^{(\ell)}_{Y_{j\to i}}(k) \tag{9}$$

$$\Phi^{(\ell)}_{\max_2,i}(k) = \Phi^{(\ell)}_{\max_1,i}(k) + \sum_{j\in\mathcal{V}_i} \overline{\Phi}^{(\ell)}_{Y_{j\to i}}(k) \prod_{j'\in\mathcal{V}_i\setminus j} \Phi^{(\ell)}_{Y_{j\to i}}(k), \tag{10}$$

where $\overline{\Phi}^{(\ell)}_{Y_{j\to i}}(k) = 1 - \Phi^{(\ell)}_{Y_{j\to i}}(k)$ and $\Phi^{(\ell)}_{\max_u,i}(k) = \sum_{l=0}^{k} P^{(\ell)}_{\max_u,i}(l)$ corresponds to the CDF of the first ($u=1$) and second ($u=2$) maximum values of the $\{Y^{(\ell)}_{j\to i}\}_{\forall j\in\mathcal{V}_i}$ random variables.

Equations (9) and (10) can be used to derive the signal activities of all input registers storing the first and second minimum message value, with the exception of the internal registers in the "2-min search tree." This tree of elementary 2-min units is composed of $\lceil \log_2 \mathrm{DC}_{\mathrm{MAX}} \rceil$ levels. Each level can optionally be terminated by a pipeline register in order to increase the processing throughput of the decoder. To derive the activity of these registers, (9) and (10) must be adapted by replacing $\mathcal{V}_i$ with $\mathcal{V}'_i(s,u)$, where $\mathcal{V}'_i(s,u) \subseteq \mathcal{V}_i$ is the subset of VN indices that are part of the computational tree for the u-th 2-min unit in level s of the tree.

It is worth mentioning that the proposed model can be easily adapted for other Min-Sum decoder architectures compatible with assumption (A3) introduced in Section III-A. In such case, the registers and memories are used to store the same variables, so the signal activities are similar. Therefore, adapting the model for different architectures only requires to identify where the registers are located.

B. Density evolution with random codewords

Standard DE [8] provides the message PMFs through successive decoding iterations, under the assumption that the all-zero codeword is transmitted. However, this assumption does not allow to describe the message PMFs in a real decoder, since random codewords are transmitted in practice. Therefore, standard DE cannot be straightforwardly applied to derive the signal switching activities. Instead, we consider that the bits b_j affected to the VN indexes $j\in\mathcal{V}_i$ are random with probability $1/2$ of being 1 or 0, under the condition that the transmit bits fulfill parity check equation at the i^{th} row of the PCM: $\bigoplus_{j'\in\mathcal{V}_i} b_{j'} = 0$, with $\oplus$ being the logic *exclusive-or* operator. Under this condition, it can be shown that the PMF of the C2V messages, denoted $P^{(\ell)}_{\gamma_{i\to j}|\overline{\oplus}}(k)$, becomes

$$P^{(\ell)}_{\gamma_{i\to j}|\overline{\oplus}}(k) = \frac{1}{2}\left(P\Big(\gamma^{(\ell)}_{i\to j} = k \mid \bigoplus_{j'\in\mathcal{V}_i\setminus j} b_{j'} = 0\Big)\right.$$

TABLE I
ESTIMATED MODEL PARAMETERS FOR THE 65NM TSMC TECHNOLOGY

Notation	Energy Value (fJ)	Description
$E_{\overline{ce}}$	5.16	Clock activity of CGU when disabled
$E_{ce}(W)$	$E_\alpha W + E_\beta$	Clock activity of CGU when enabled
E_α	0.57 /bit	Energy per sub-clock
E_β	7.29	CGU base energy when enabled
E_C	5	Clock activity in FF
$E_{s\,\overline{c}}$	2.3	FF input activities when clock disabled
E_{sc}	5.24	FF input activities when clock enabled

$$+ P\Big(\gamma_{i\to j}^{(\ell)} = k \mid \bigoplus_{j'\in\mathcal{V}_i\setminus j} b_{j'} = 1\Big)\Bigg)$$
$$= \frac{1}{2}\Big(P_{\gamma_{i\to j}}^{(\ell)}(k) + P_{\gamma_{i\to j}}^{(\ell)}(-k)\Big), \qquad (11)$$

where $P_{\gamma_{i\to j}}^{(\ell)}(k)$ is the PMF of the C2V messages when the all-zero codeword is transmitted.

Similarly, the V2C message PMF given the transmitted bits fulfill the $|\mathcal{C}_j|$ local parity-check equations is expressed as

$$P_{\lambda_{j\to i}|\,\overline{\oplus}}^{(\ell)}(k) = \frac{1}{2}\Big(P_{\lambda_{j\to i}}^{(\ell)}(k) + P_{\lambda_{j\to i}}^{(\ell)}(-k)\Big). \qquad (12)$$

V. VALIDATION

The aim of this section is to estimate the parameters of the clock-gated FF model proposed in Section III, and to validate the model by comparing the predicted energy with the one estimated on post-synthesis circuits.

A. Estimation of the energy model parameters

The energy parameters of the FF model are presented in Table I. They are obtained through power estimation performed using the `Cadence Genus` tool on simple test circuits composed of registers and memory units. Estimations are based on switching activity files generated through post-synthesis simulations. The architectures are synthesized using TSMC CMOS 65nm GP technology. The E_{ce} and $E_{\overline{ce}}$ energy parameters are directly obtained through the power estimation tools. The parameter E_C is then deduced by setting all FF input bits to 0 so that no activities are generated. Then, E_{sc} and $E_{s\,\overline{c}}$ are obtained by estimating the energy consumption when the input bits switch states at each clock cycle, for both cases where the clock-gating is enabled and disabled. The CGU energy when the sub-clock is enabled E_{ce} depends on W. All W FFs being identical, E_{ce} is expected to be linear with the fanout of the CGU. Thus it is modeled as $E_{ce}(W) = E_\alpha W + E_\beta$, which was confirmed to closely match empirical results.

B. Validation of the proposed energy model

To validate the energy model, the check node processing element (CNPE) architecture proposed in [6] and shown in Figure 2 is considered. A CNPE can process up to $\text{DC}_{\text{MAX}} = 26$ messages in parallel at each clock cycle. The CNPE input corresponds to DC_{MAX} APs of 8 bits and outputs DC_{MAX} $\Lambda_j^{(\ell)}$ variables on $N_b + 1$ bits, with $N_b = 5$ being the number of bits used to represent C2V messages (including the sign).

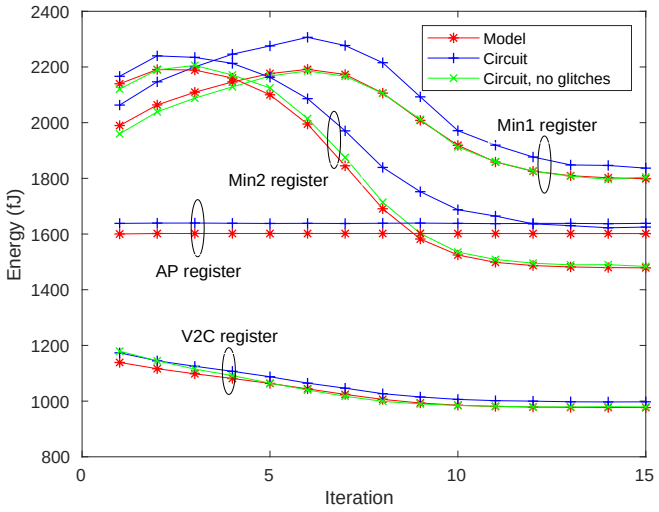

Fig. 3. Energy per iteration of the CNPE registers.

The memory storing the two minimum C2V message values of $N_b - 1$ bits, the index of the minimum (5 bits) and the DC_{MAX} message signs is of depth $D_M = 96$ and width $D_W = 41$. The offset is fixed to $\beta = 1$.

The energy is estimated through post-synthesis simulations using the same tools presented in Section V-A. To evaluate the impact of the signal activities on the energy consumption, the energy is estimated for each decoding iteration independently. To do so, the AP samples are generated based on their expected PMFs at a given iteration and clock cycle index for 100 codewords. This method artificially removes the cycles in the code, and has negligible impact on the decoder performance evaluation as explained in Section II-B. The PMFs of the received LLRs are set according to (3) with $\sigma^2 = 0.7079$ and $\alpha = 3$. Since the aims of this section is to validate the accuracy of the energy model with respect to measurements from post-synthesis simulations, the choice of the code can be arbitrary. The model must be able to predict the energy consumption independently of the type of input PMFs being considered in the CNPE. Therefore, for simplicity, we consider that the PMFs of the CNPE input signals are obtained assuming an infinite-length code based on the following protograph S_P [18]:

$$S_P = \begin{bmatrix} 0 & 2 & 3 & 1 \\ 2 & 0 & 3 & 2 \end{bmatrix}. \qquad (13)$$

The validation and study of the energy model for the complete LDPC decoder and several codes is left for future work.

Figure 3 shows the predicted energy obtained with the proposed model and the energy estimated on the CNPE circuit for registers storing the APs, identified ① in Figure 2, the registers storing the V2C messages ② and the registers storing the first/second minimum variables $\text{min}_1/\text{min}_2$ ③. All energies are normalized per register and per iteration. For all registers, the predicted energy at each iteration follows the same trend as the circuit simulation. The energy estimated from the circuit is only between 2% (V2C register) and 8% (min2 register) higher than the predicted energy, which is acceptable for a high-level

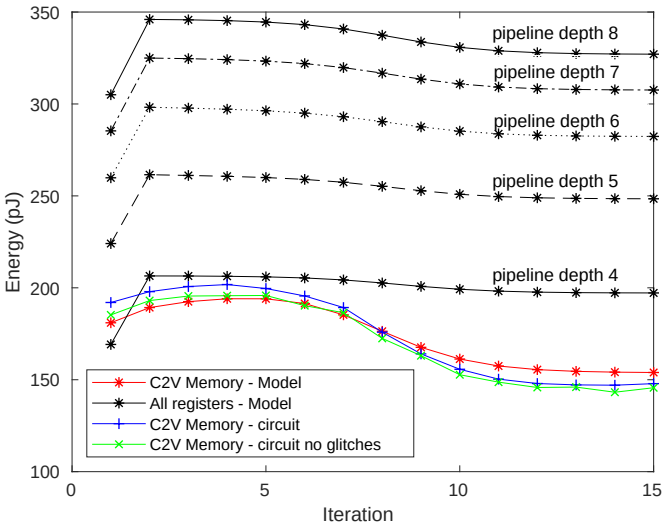

Fig. 4. Energy per iteration of the C2V memory unit and all the registers.

model. This slight difference can be explained by the presence of additional switching due to the signals propagating through the logic units with different delays. This additional activity, known as *glitches*, is not taken into account in the proposed model, since this would require information on propagation delays, which is only available after circuit synthesis. Figure 3 also shows the register energy when glitches are removed. In this case, the predicted energy is within 1% of the simulation result. This demonstrates the accuracy of the model and the validity of the proposed approach.

Figure 4 shows the predicted energy obtained with the proposed model and the energy estimated on the C2V memory unit of the CPNE circuit. The predicted energy follows similar variations as the estimation, particularly during the first iterations. This shows the importance of considering the signal activity for predicting the energy consumption. After the 7$^{\text{th}}$ iteration, the model slightly deviates, where the energy is up to 4% higher than the estimated ones, which is acceptable for a high-level model. The total energy of all the registers, predicted by the model, is also shown in this figure for different CNPE pipeline depths $D_P(\text{CNPE})$. Note that increasing pipeline depth allows achieving a higher processing throughput. It can be seen that the energy consumed by the registers is of the same order of magnitude as the memory unit. Furthermore, the register energy greatly depends on the number of pipeline stages, with up to 32% energy increase between $D_P(\text{CNPE}) = 4$ and $D_P(\text{CNPE}) = 8$. This demonstrates the importance of considering the register energy and the number of pipeline stages when studying the energy consumption of LDPC decoders.

VI. Conclusion

This paper proposes a novel high-level energy model for LDPC decoders. The model predicts the energy consumption of clock-gated FFs and standard-cell-based memory units based on signal activities. The switching activity of each bit is derived through a novel DE method adapted for this energy model. The model is validated by comparing the predicted energy with the energy estimated by a power analysis tool after circuit synthesis. Future work will study how this energy model can be used to optimize the energy efficiency of LDPC decoders, for instance in terms of quantization bits, parallelism level and number of pipeline registers, as well as to guide the construction of energy-efficient codes.

References

[1] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[2] J. H. Bae, A. Abotabl, H.-P. Lin, K.-B. Song, and J. Lee, "An overview of channel coding for 5G NR cellular communications," *APSIPA Transactions on Signal and Information Processing*, vol. 8, p. e17, 2019.

[3] I. F. Akyildiz, A. Kak, and S. Nie, "6g and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[4] K. Ganesan, P. Grover, J. Rabaey, and A. Goldsmith, "On the total power capacity of regular-ldpc codes with iterative message-passing decoders," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 2, pp. 375–396, 2016.

[5] C. G. Blake and F. R. Kschischang, "Energy, latency, and reliability tradeoffs in coding circuits," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 935–946, 2019.

[6] J. Nadal, M. Fiorentino, E. Dupraz, and F. Leduc-Primeau, "A deeply pipelined, highly parallel and flexible LDPC decoder," in *2020 18th IEEE International New Circuits and Systems Conference (NEWCAS)*, 2020, pp. 263–266.

[7] E. Amador, R. Pacalet, and V. Rezard, "Optimum LDPC decoder: A memory architecture problem," in *2009 46th ACM/IEEE Design Automation Conference*, 2009, pp. 891–896.

[8] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 599–618, 2001.

[9] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[10] M. P. C. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Transactions on Communications*, vol. 47, no. 5, pp. 673–680, May 1999.

[11] A. Balatsoukas-Stimming and A. Burg, "Density evolution for min-sum decoding of LDPC codes under unreliable message storage," *IEEE Communications Letters*, vol. 18, no. 5, pp. 849–852, 2014.

[12] C. Kameni Ngassa, V. Savin, E. Dupraz, and D. Declercq, "Density evolution and functional threshold for the noisy min-sum decoder," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1497–1509, 2015.

[13] F. Leduc-Primeau and W. J. Gross, "Finite-length quasi-synchronous LDPC decoders," in *2016 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, 2016, pp. 325–329.

[14] R. Yazdani and M. Ardakani, "Waterfall performance analysis of finite-length LDPC codes on symmetric channels," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3183–3187, 2009.

[15] E. Dupraz and F. Leduc-Primeau, "Noisy density evolution with asymmetric deviation models," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1403–1416, 2020.

[16] P. Meinerzhagen, S. M. Y. Sherazi, A. Burg, and J. N. Rodrigues, "Benchmarking of standard-cell based memories in the sub-v_{t} domain in 65-nm CMOS technology," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 2, pp. 173–182, 2011.

[17] H. A. David and H. N. Nagaraja, *Order Statistics*. American Cancer Society, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471667196.ess6023.pub2

[18] E. Dupraz, F. Leduc-Primeau, and F. Gagnon, "Low-latency LDPC decoding achieved by code and architecture co-design," in *2018 IEEE 10th International Symposium on Turbo Codes Iterative Information Processing (ISTC)*, 2018, pp. 1–5.