Title

***Abstract*—**

***Index Terms*—Keywords**

I. Introduction

This paper is collocated into the reconfigurable computing context. An important problem into the reconfigurable computing is to analyse the error propagation when the circuit "under test" is setted with values of the frequency too high or in disagree with the values of the delay. The paper investigates the propagation of the error caused by a timing violation into a digital online multiplier.

.....

This work starts from the results obtained in a previus work on the Radix-2 Digit-parallel Online Multiplier (see [1]).
In [1] the autors have provided methodologies to implement a parallel version of the online multiplier and they have proposed a probabilistic model describing the propagations of the error caused by timing violation conditions. They also backed up the models with experimental results from an image processing application demostrating that this novel design methodology can lead to substantial performance benefits compared to the design method using conventional arithmetic. The purpose of this paper is to provide a numerical model estimating the error on the online multiplier and a model to estimate the probability of timing violation in order to provide an "a priori" estimation for the error expectation for different values of the circuit parameters (frequency, number of digits, online delay, etc).
In this paper we provide a formal model of the Overclocking Error for the radix-2 digit-parallel Online Multiplier. We also proved a mathematical '"support" to the experimental results obtained in [1] and, in particular, we estimate the expectation of the overclocking error for these results. We provide a model describing the probablity of timing violation as "chain start probability" and "chain rule probability" representing the probability that a chain starts during the operation and the probability that a chain starts & perpetuals for a fixed number of digits, respectively.

II. Background: Online Arithmetic

A. Key Features of Online Arithmetic

Online arithmetic has been widely used in numerous applications such as signal processing and control algorithms [xxx]. Online arithmetic was originally designed for digit serial operation, as illustrated in Fig xx. It can be seen that in order to generate the first output digit, δ digits of inputs are required, where δ is called the "online delay". Normally δ a small constant, which is independent of the precision. For ease of discussion, for the rest of this paper, the input data is assumed to be fixed point numbers in the range $(-1, 1)$. Based on this premise, the online representation of N-digit operands and result at iteration j are given by (xx), where $j \in [-\delta, N-1]$ and r denotes the radix [xxx].

$$X_{[j]} = \sum_{i=1}^{j+\delta} x_i r^{-i},\ Y_{[j]} = \sum_{i=1}^{j+\delta} y_i r^{-i},\ Z_{[j]} = \sum_{i=1}^{j} z_i r^{-i} \qquad (1)$$

MSD first operation is possible only if a redundant number system is used. Normally there are two most commonly used redundant number representations: carry-save (CS) [xxx] and signed-digit (SD) [xx]. With SD representation, each digit is represented using a redundant digit set $\{-a, \cdots, -1, 0, 1, \cdots, a\}$ where $a \in [r/2, r-1]$. In comparison, the standard non-redundant representation only uses a digit set $\{0, \cdots, r1\}$. Thus a standard number corresponds to several possible redundant representations. For example, the binary number 0.011 can be represented in SD form as 0.111, 0.101 or 0.011 among many other possible representations. Due to the redundancy, the MSDs of the result can be calculated using partial information from both inputs. Then the value of the number can be revised using the subsequent digits, because each number has multiple representations.

B. Binary Online Addition

Adders serve as a critical building block for arithmetic operations. To perform digit-parallel online addition, a redundant adder can be used directly. The structure of an online adder where all signals represented with SD numbers of digit set $\{1, 0, 1\}$ is shown in Fig. xxx. The module "3:2" denotes a 3:2 compressor, which takes three inputs and generates two outputs, and is logically equivalent to a full adder (FA). A major advantage of the redundant number system over the standard ripple-carry based arithmetic is that the propagation of carry is eliminated, resulting in a precision-independent computation time for addition. As labelled in Fig. xxx, ideally the computation delay of this adder is only two FA delays for any operand word-length, with the cost of one extra FA for each digit of operands. This makes the online adder suitable for building up more complex arithmetic operators such as multipliers to accelerate the sum of partial products [xx].

C. Binary Online Multiplication

Multiplication is another key arithmetic operator. Typically, online multiplication is performed in a recursive digit-serial manner, as illustrated in Algorithm II-C where both inputs and outputs are N-digit number as represented in (1). For a given iteration j, the product digit z_j is generated through a selection function $sel()$. For the radix r and a chosen digit set, there exits an appropriate selection method and a value of δ

which ensure convergence. As radix-2 is used most commonly in computer arithmetic, we keep $r = 2$ throughout this paper with the corresponding redundant digit set $\{\overline{1}, 0, 1\}$. In this case we have $\delta = 3$, and $sel()$ is given by (2), from which it can be seen that the decision is made only based on 1 integer bit and 1 fractional bit of $W_{[j]}$.

Online Multiplication

1: **Initialization:** $X_{[-\delta]} = Y_{[-\delta]} = P_{[-\delta]} = 0$
2: **Recurrence:** $for \;\; j = -\delta, \;\; -\delta + 1, \;\; \cdots, \;\; N - 1 \;\; do$
3: **Compute:**
4: $H_{[j]} = r^{-\delta} \left(x_{j+\delta+1} \cdot Y_{[j+1]} + y_{j+\delta+1} \cdot X_{[j]} \right)$
5: $W_{[j]} = P_{[j]} + H_{[j]}$
6: $z_j = sel(W_{[j]})$
7: $P_{[j+1]} = r \left(W_{[j]} - z_j \right)$

$$\begin{array}{ll} sel(W_{[j]}) = 1 & if \;\; W_{[j]} \geq \frac{1}{2} \\ 0 & if \;\; -\frac{1}{2} \leq W_{[j]} < \frac{1}{2} \\ \overline{1} & if \;\; W_{[j]} < -\frac{1}{2} \end{array} \tag{2}$$

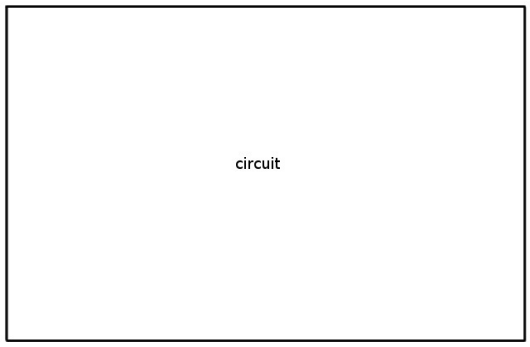

Fig. 1. . . . . .

.....

Let μ_{OM} denote the worst-case delay of an OM. It follows that if the clock period T_S is greater than μ_{OM}, correct results will be sampled. If, however, faster-than-rated sampling frequencies are applied such that $T_S < \mu_{OM}$, timing violations might happen and intermediate results will be sampled, potentially generating errors. For a given T_S, the maximum length of error-free propagation is described by (3) where f_S denotes the sampling frequency. We always ensure that the first digit of the product will be generated correctly, i.e. $b > \delta$.

$$b := \left\lceil \frac{T_S}{\mu} \right\rceil = \left\lceil \frac{1}{\mu \cdot f_S} \right\rceil \tag{3}$$

We now determine when this timing constraint is not met and the size of error in this case.

III. Probability of Timing Violations

For an N-digit OM, let a propagation chain be generated at stage S_τ with the length of $d(\tau)$ digits, then the stage number τ is bounded by (4). The presence of timing violation requires $d(\tau) > b$. Besides, the chain cannot propagate over S_{N-1}. Thus the bound of parameter $d(\tau)$ is given by (5).

$$-\delta \leq \tau \leq N - 1 - b = \tau_{max} \tag{4}$$

$$b < d(\tau) \leq N - 1 - \tau \tag{5}$$

However, the actual length of a "carry" chain is dependent upon input patterns. We then examine the relationship between $d(\tau)$ and the inputs that corresponds to the generation, propagation and annihilation of this carry chain. Let the specific input pattern of stage S_τ be represented by $C(\tau)$, which can be classified into four types, as listed in (6).

$$C(\tau) = \left\{ \begin{array}{llll} Case1 & (C_1(\tau)): & x_{\tau+\delta+1} = 0, & y_{\tau+\delta+1} = 0 \\ Case2 & (C_2(\tau)): & x_{\tau+\delta+1} \neq 0, & y_{\tau+\delta+1} \neq 0 \\ Case3 & (C_3(\tau)): & x_{\tau+\delta+1} \neq 0, & y_{\tau+\delta+1} = 0 \\ Case4 & (C_4(\tau)): & x_{\tau+\delta+1} = 0, & y_{\tau+\delta+1} \neq 0 \end{array} \right. \tag{6}$$

The classification is based on whether a digit is zero, as all internal signals are reset to zero initially. Under the assumption that all digits are mutually independent and uniformly sampled from the digit set $\{\overline{1}, 0, 1\}$ as given in Section II, the probabilities of $C_1, ..., C_4$ are given in (7).

$$\begin{array}{ll} C_1(\tau) : \frac{1}{(2a+1)^2}, & C_2(\tau) : \frac{(2a+1)^2 - 2(a+1) - 1}{(2a+1)^2} \\ C_3(\tau) : \frac{a+1}{(2a+1)^2}, & C_4(\tau) : \frac{a+1}{(2a+1)^2} \end{array} \tag{7}$$

Notice that no carry chain will be generated under Case 1, because $P_{[\tau+1]} = 0$. In other words, a chain could be generated if one of the cases $C_2(\tau)$,...,$C_4(\tau)$ occur. In this case we provide the following propositions which describe the probabilities of timing violations.

Proposition 3.1: (Chain rule probability)
Let $Pr_{\tau,d(\tau)}$ denote the probability that a carry chain start at stage S_τ, and it have lenght $d(\tau)$, $\forall \tau \in [-\delta, N - 1 - b]$, $\forall d(\tau) \in [b, N - 1 - \tau]$, then it is given by (8)

$$Pr_{\tau,d(\tau)} = \left(\frac{1}{(2a+1)^2} \right)^{\tau} \left(\frac{(2a+1)^2 - 1}{(2a+1)^2} \right)^{d(\tau)} \tag{8}$$

Proof:

Let E_2 be the event "the carry chain is generated at stage S_τ and it have length $d(\tau)$", we have

$$Pr_{\tau,d(\tau)} = Pr(E_2) = \underbrace{Pr(C_1) \cdots Pr(C_1)}_{\tau - 1 \;\; times} \cdot \tag{9}$$

$$\cdot \underbrace{Pr(C_2 \;\; or \;\; C_3 \;\; or \;\; C_4) \cdots Pr(C_2 \;\; or \;\; C_3 \;\; or \;\; C_4)}_{d(\tau) \;\; times} Pr(C_1)$$

that is

$$Pr_{\tau,d(\tau)} = Pr(C_1)^{\tau - 1} Pr(C_2 \cup C_3 \cup C_4)^{d(\tau)} Pr(C_1) \tag{10}$$

Substituting (7) into 10 yields (11).

$$Pr_{\tau,d(\tau)} = \left(\frac{1}{(2a+1)^2}\right)^{\tau-1} \left(\frac{(2a+1)^2-1}{(2a+1)^2}\right)^{d(\tau)} \frac{1}{(2a+1)^2} \tag{11}$$

Simplifying (11) will give (8).

♣

For all possible values of τ and $d(\tau)$, the possibility that timing violaions happen is given by Pr in (12). According to (4) and (5), the ranges in which the parameters τ and $d(\tau)$ are defined depend upon the value of the swamping period T_S. Hence Pr is a function of T_S.

$$Pr = \sum_{\tau} \sum_{d(\tau)} Pr_{\tau,d(\tau)} \tag{12}$$

IV. Error Estimation

In the presence of timing violation, multiple chains might not be correctly propagated in the online multiplication, resulting in overclocking errors generated from LSDs. Let z_i and $z_i{}'$ denote the correct value and the actual value of the output digit at the stage S_i, respectively. Then we have $z_i{}' = z_i + \varepsilon_i$ where ε_i is referred to as the overclocking error. We now locate the first output digit z_λ that contains error. For a given T_S and a minimum value of τ such that $d(\tau)_{max} > b$, we have $\lambda = \tau + d(\tau)_{max} + 1$ for chain annihilation. As such, the overclocking error can be expressed by (13)

$$|\varepsilon| = \left| \sum_{j=\lambda}^{N} r^{-j} \varepsilon_j \right| \tag{13}$$

We would like to investigate the statistical characteristics of the ε. In general, the distribution of ε can be obtained from the convolution of variable ε_j where $j \in [\lambda, N]$. Each variable ε_j is sum of uniform iid variables with mean zero, therefore its mean is zero and the mean of ε is zero as well.

Let $v\varepsilon$ and $v\varepsilon_j$ denote the variances of variable ε and ε_j, respectively. We can form the relationship between them as given in (14) based on (13).

$$v\varepsilon = \sum_{j=\lambda}^{N} r^{-2j} v\varepsilon_j \tag{14}$$

Normally for a given algorithm, in this case the Algorithm II-C, the value of ε_j can be determined based on the inputs, which are uniformly distributed with the digit set $\{\overline{a}, \cdots, 0, \cdots, a\}$. Hence for a single digit of inputs $X_{[j]}$ and $Y_{[j]}$, its variance can be denoted by $v\varepsilon^{in}$, which is a fixed value with respect to a given value of a. Let the variances of the input signals $X_{[j]}$ and $Y_{[j]}$ be denoted by $v\varepsilon_j^y$ and $v\varepsilon_j^x$, respectively. Then we have (15) according to (1).

$$v\varepsilon_j^y = v\varepsilon_j^x = \sum_{i=1}^{j+\delta} r^{-2i} v\varepsilon^{in} \tag{15}$$

In Algorithm II-C, the variances of the input data propagate through for all iterations. For a single iteration j, we can calculate $v\lambda_j$ based on $v\varepsilon_j^y$ and $v\varepsilon_j^x$. For example initially the variance of $H_{[j]}$ is given by (16).

$$v\varepsilon_j^H = r^{-2\delta}(v\varepsilon^{in} v\varepsilon_{j+1}^y + v\varepsilon^{in} v\varepsilon_j^x) \tag{16}$$

This value propagates throughout the entire iteration and finally we have the value of $v\varepsilon_j$ in (17).

$$v\varepsilon_j = r^2 \left(r^{-2\delta} \cdot v\varepsilon_j^H\right) \tag{17}$$

Combining (16) and (17) yields

$$v\varepsilon_j = r^2 \left(r^{-2\delta} \left(v\varepsilon^{in} v\varepsilon_{j+1}^y + v\varepsilon^{in} v\varepsilon_j^x\right)\right). \tag{18}$$

Finally, (18) can be used to derive the expression of $v\varepsilon$ in (19) according to (14). Notice that this expression forms the relationship between the variance of overclocking error and the given input distribution.

$$\begin{aligned} v\varepsilon &= \mathcal{F}(T_S, \mu, v\varepsilon^{in}, \delta) \\ &= \textstyle\sum_{k=\lambda}^{N} r^{-2k} \left(r^{2(1-\delta)} \left(v\varepsilon^{in} v\varepsilon_{k+1}^y + v\varepsilon^{in} v\varepsilon_k^x\right)\right) \end{aligned} \tag{19}$$

V. Model Verification with FPGA results

In this Section, we compare the proposed models for both the probability and the magnitude of overclocking errors with the experimental results from FPGAs. In the verification, we choose the redundant digit set $\{\overline{1}, 0, 1\}$, which is most commonly used for binary redundant representations.

A. Verification of Models for Error Probability

We initially verify the models for the probability of timing violations. With the usage of the redundant digit set $\{\overline{1}, 0, 1\}$, we have $a = 1$. Substituting this in (8) yields (20).

$$Pr_{\tau,d(\tau)} = \left(\frac{1}{9}\right)^{\tau} \left(\frac{8}{9}\right)^{d(\tau)} \tag{20}$$

Combining (20) and (12) we obtain the values of the probability of timing violations with respect to a variety of sampling period T_S. The results of an 8-digit OM are shown in Fig. 2. For comparison, we also plot the results obtained from a Virtex-6 FPGA, and the data are obtained from post place-and-route simulations. It can be seen that the modelled probability values match well with the experimental results.

B. Verification of Models for Error Magnitude

We then verify the proposed models for the estimation of timing errors. Fig. 3 shows the distribution of the error ε on the FPGA results of the Algorithm 1 with input data uniformly distributed into the digit set $\{\overline{1}, 0, 1\}$.

Several slice through Fig. 3 are presented in Fig. 4 for different sampling period values and the corresponding variances of overclocking errors. We observe that the profile of the error distribution changes as the value of the sampling period changes, and the width of the bell changes as well also (see Fig. 4). The value of the mean is zero for all the values of the sampling period, while the variance of

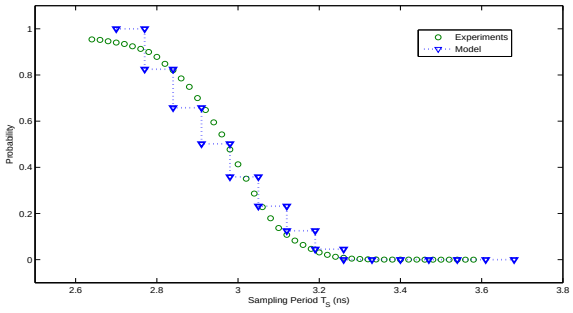

Fig. 2. Probability of timing violation for $N = 8$ and $\delta = 3$

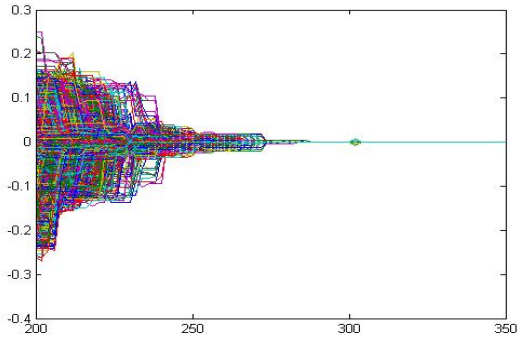

Fig. 3. Distribution of overclocking errors with respect to a variety of sampling period. The Xlable should be "Sampling period", which varies from 2ns to 3.5ns, and the Ylabel is "Error Values".

the distribution changes as the sampling period changes. In particular, the values of the variance decrease as the sampling period increases in agreement with the information provided by the values of the probability of timing violation.

We have compared these experimental results with the results provided by the model defined in (19). The Figure

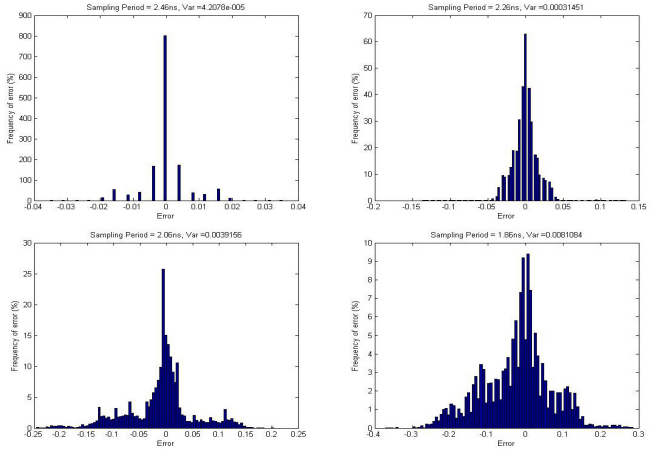

Fig. 4. Several slice through Fig. 3 that shows error distribution with respect to a variety of sampling periods and the corresponding variances of overclocking error. The results are obtained from experiments.

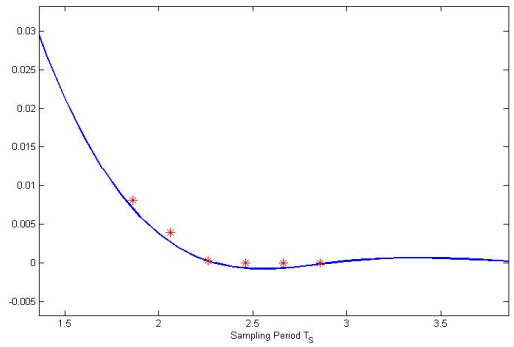

Fig. 5. Variances of overclocking errors. The results are obtained from the proposed models (blue line) and the experiments (red stars). (Legend and xylabels to be added.)

5 shows the curve provided by the model (blue line) and the experimental results (red stars). It can be seen that the modelled results are close to the experimental results.

REFERENCES

[1] Datapath synthesis for overclocking: Online Arithmetic for Latency-Accuracy Trade-offs.
[2] J. B. Uspensky, Introduction to Mathematical Probability (New York: McGraw-Hill, 1937)
[3] Ross, Introduction to Probability Models, University of Southern California - Los Angeles (CA), 2010 ELSEVIER
[4] C. de Boor, A Practical Guide to Splines, Springer-Verlag, 1978.
[5] Kendall E. Atkinson, On the order of convergence of natural cubic spline interpolation, SIAM Journal on Numerical Analysis, Vol 5 No 1, 1968
[6] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.