Low Latency Approximate Adder for Highly Correlated Input Streams

Xiaoliang Chen, Ahmed M. Eltawil, Fadi J. Kurdahi
Electrical Engineering and Computer Science Department
Univeristy of California Irvine
Irvine, CA, USA
{xiaoliac, aeltawil, kurdahi}@uci.edu

***Abstract*— Approximate computing helps achieve better performance or energy efficiency by trading accuracy. Most approximate adders are composed of multiple sub-adders and long carry chains are split to reduce latency, thus benefiting from the fact that carry propagation across long carry chains is rare for uniformly distributed inputs. One key tradeoff of these approximate adders is between latency and error rate. The more prediction bits are used, the lower is the error rate, but the latency is longer. In this paper, we present a Correlation Aware Predictor (CAP) which utilizes spatial-temporal correlation information of input streams to predict carry-in value for sub-adders. CAP uses less prediction bits which help reduce adder latency significantly. For highly correlated input streams, we found that CAP can reduce adder latency by about 23% at the same error rate compared to prior work. We implemented a CAP-based approximate adder in Verilog and synthesized with TSMC 16nm library. Synthesis results show that CAP-based adder can reduce latency by 25% and save 13% in silicon area compared to state-of-the-art.**

Keywords—approximate adders; speculative adders; carry predictor

I. INTRODUCTION

As a key arithmetic component in approximate computing [1-3], approximate addition has drawn lots of attentions recently and many approximate adders have been proposed [5-10]. Most of the approximate adders are split into *sub-adders,* and carry-in value for each sub-adder is predicted to reduce the latency. The main idea is for uniformly distributed inputs, carry propagation across long carry chains is rare. However, the assumption that input streams follow a uniform distribution is not always true. As pointed out in [4], multimedia signals like music have distributions that are closer to Gaussian shape, and therefore error probability can be very different compared with the uniform case which may incur over/under design for real applications.

In this paper, we propose new carry predictor to reduce error rate as well as latency for highly correlated input streams. Our major contributions as follows: (1) We show that correlation information in real application can be used to reduce the error rate and latency for carry-in prediction. (2) We proposed a Correlation Aware Predictor (CAP) utilizing correlation information of input streams to predict carry-in value for sub-adders. We show that a CAP based approximate adder can reduce latency significantly for highly correlated input streams compared with state-of-art approximate adders. (3) We experimented with a CAP based approximate adder processing 24-bit audio streams. Simulation results show that the proposed approximate adder can reduce latency and area for real applications.

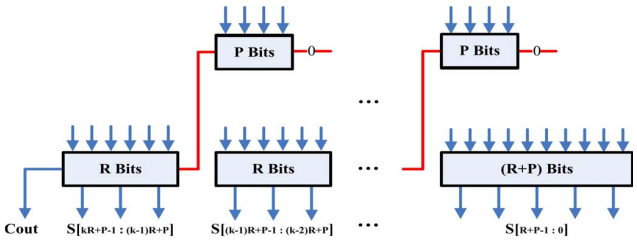

Figure 1. Block diagram of GeAr adder

II. BACKGROUND

Recently a general approximate adder model, Generic Accuracy Configurable (GeAr) adder [10], which supports variable approximation modes, is proposed. As shown in Fig. 1, any GeAr adder can be specified by three parameters (1) the adder width ***N*** (2) the number of prediction bits ***P*** of each sub-adder (3) the number of result bits ***R*** from each sub-adder. GeAr(*N*,*R*,*P*) adder uses multiple (*R*+*P*)-bit sub-adders in parallel to perform the approximate addition. Compared to an *N*-bit adder, the delay is significantly reduced since the longest carry propagation is no more than (*R*+*P*) bits.

The output of GeAr adder has errors when the carry-in value for any of sub-adders is mispredicted. An error occurs in an individual sub-adder when both of the following conditions occur: (i):The carry-in of sub-adder is 1, not 0; (ii):All *P* predictive bits of a sub-adder are propagating carry-in. Many research works [10-12] reported approximate adder error rate under uniformly distributed inputs. Generally, the more prediction bits are used, the lower is the error rate, but the longer is the adder delay.

Most recent works on approximate adders fall into the design space of the GeAr model. ACA-I in [5] uses multiple overlapping sub-adders with *R*=1. ETA-II in [6] made use of the carry predicted by one previous sub-adder, thus *R*=*P*.

1063-6404/17 $31.00 © 2017 IEEE
DOI 10.1109/ICCD.2017.26

ACAA in [8], which can be configured to selectively turn on/off correction stage at runtime to achieve different accuracy levels, also has *R*=*P*. GDA is proposed in [9] to achieve better quality-effort tradeoff by changing the number of prediction bits. In [13] the author proposed correlation-aware speculative adder (CASA) based on the observation there is a high chance that carry-in to groups adder equals 1 when xor(msb)=1. In [14] the author proposed history aware adder by exploiting the temporal operand correlation of each instruction location. Since the GeAr model is general and covers cases including ACA-I, ETA-II, ACAA and GDA, in this paper, we will use GeAr adders as baseline design.

III. Study of Correlated Input Streams

In digital signal processing, random signals with uniform distribution are often considered as white noise. Useful signals usually have certain correlations. In this paper, correlations of input streams are classified into *spatial correlation* and *temporal correlation*. To better explain input stream correlations, we use a 20s pop song with 24-bit in 2's complement encoding as an example [16].

Spatial correlation measures the similarity of switching behavior between neighboring bits. Fig. 2(a) shows the switching probability of each bit. While least significant bits (LSB) have high switching activity and behave like random signals, the switching probabilities of most significant bits (MSB) is much smaller. Besides, neighboring bits have similar switching probability. In this case MSB is less "active" and has higher spatial correlation.

Temporal correlation measures the deviation of differences (or distance) between two successive inputs. Fig 2(b) is the histogram of differences between consecutive inputs. It follows zero-mean Gaussian distribution with a very small standard deviation (0.0013). Another observation is the signs of differences don't change frequently. This is due to the fact that low frequency components are dominant and signal value changes from increasing to decreasing (or vice versa) infrequently.

Given the observation that spatial and temporal correlations do exist in real input streams especially for most significant bits, we hypothesize that correlations exist between carry-in signals of consecutive operations on highly correlated input streams. Figure 2(c) and 2(d) show the switching probability of carry-in signals at different positions for unsigned addition and signed subtraction respectively. For most significant bits, there is a strong correlation between previous carry-in value and current value. This motivates us to enhance carry prediction by utilizing temporal correlation of carry signals.

IV. CAP Based Approximate Adder Design

In this section, we present the design of correlation aware predictor (CAP) based approximate adder and evaluate its performance. CAP is an extension to existing carry-in predictors such as GeAr [10]. As mentioned before, we use the GeAr as baseline design for comparison.

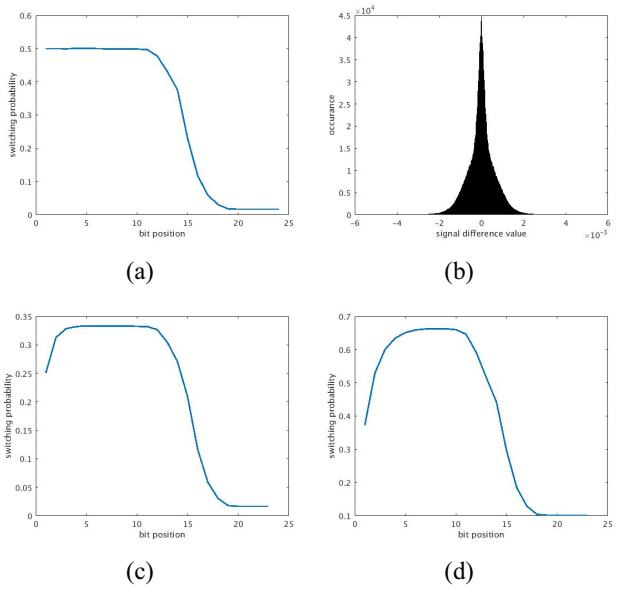

Figure 2 (a) histogram of audio signal (b) switching probability of audio signal (c) carry-in signal switching probability of unsigned addition (d) carry-in signal switching probability of signed subtraction

A. CAP Based Approximate Adder

We proposed correlation aware predictor (CAP) as shown in Fig. 3 which is a mux based predictor. Controlling signals are group generate (G) and kill (K) results of prediction bits. When G (or K) is 1, the output of mux would be 1 (or 0). In this case carry propagation chain is not activated and predicted carry-in is correct. When the value of both of G and K are zero, an estimated value must be provided since carry propagates across long carry chains and predictive bits are not sufficient. In most existing predictors such as GeAr, the estimated value is fixed at zero. To reduce error rate and latency, CAP has a flip-flop to store the carry-in result of the prior operation and use it to predict carry-in in this case. As discussed in section III, the prior result could be better guess for strongly correlated bits of highly correlated input streams.

The error detection and correction mechanism of the CAP based approximate adder is similar to GeAr [10]. Fig. 3 shows the structural diagram of CAP based approximate adder. For each carry predictor, a XOR gate is used to compare predicted carry-in with the carry-out of lower neighboring sub-adder. Once a mismatch is found in any of the sub-adders, the error detection circuit will flag an error and extra cycles are needed to correct the output. For CAP based approximate adders, flip-flops storing prior results also need update. The result is considered to be correct as long as no mismatch is detected.

B. Error Rate Analysis

To verify the effectiveness of CAP, we analyze error rate for the following cases, including 1) unsigned addition under uniform distributed inputs 2) unsigned addition under spatial correlated inputs and 3) signed subtraction under sine wave inputs.

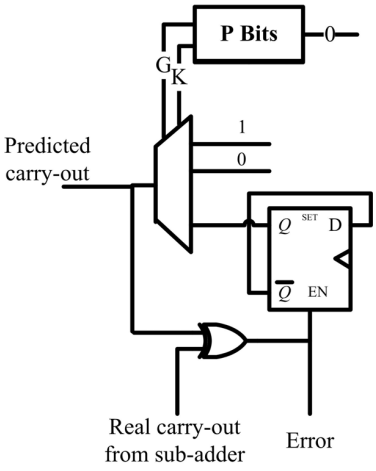

Figure 3. Implementation of correlation aware predictor

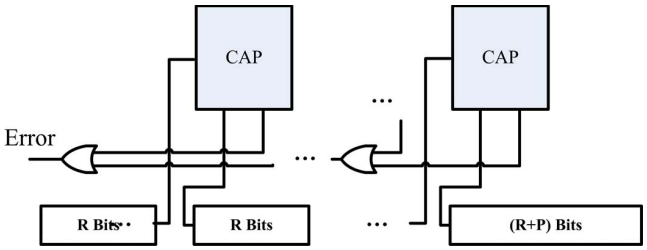

Figure 4. Structure of CAP based approximate adder

Case1: *Unsigned addition under uniformly distributed inputs*

GeAr(*N*,*R*,*P*) has *K*=(*N*-*P*)/*R* sub-adders. The first sub-adder is accurate. For sub-adder *k*, the probability of condition (i) is $0.5*(1-2^{-R})$ and the probability of condition (ii) is 2^{-P}. The error probability of *k*th sub-adder is approximately equal to $2^{-P+1}(1-2^{-R})$. The overall error probability is

$$ER_{uniform} = 1-(1-2^{-(P+1)}(1-2^{-R}))^{K-1} \quad (1)$$

For CAP based adders with the same configuration, the probabilities of condition (i) and (ii) are the same since correlation of successive carry values is 0 under uniformly distributed inputs. The overall error rate of CAP based on approximate adder is the same as GeAr under uniformly distributed inputs.

Case 2: *Unsigned addition under spatial correlated inputs*

For the *i*th bit, each input has four switching cases 00/11/01/10 and carry-out signal has 9 possible state transition cases *GG*/*GP*/*GK*/*PG*/*PP*/*PK*/*KG*/*KP*/*KK* (*G*/*P*/*K* represent carry *g*eneration/*p*ropagation/*k*ill respectively). For example, when inputs are a^{00} and b^{01}, carry-out state switches from *K* to *P* and the probability equals to prob(a^{00})*prob(b^{01}).

For CAP based adder, the probability of error condition (ii) is prob(*PP*)^P, while condition (i) XOR(*carry*(*t*-1),*carry*(*t*))=1 is met when the group generation signal of [i-P-R:i-P] under successive input vectors changes. The probability of error condition (i) for the *i*th bit can be represented as

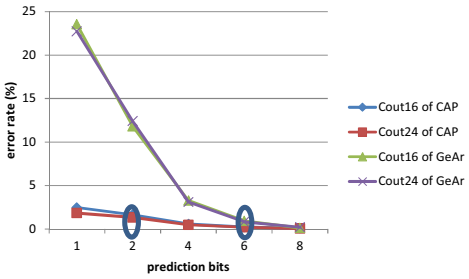

Figure 5. Error rate of GeAr VS. CAP under spatially correlated inputs

$$\begin{aligned}&\text{prob}(\text{XORG}(i-P-R{:}i-P))\\&= \text{prob}(GP_{i-P}) + \text{prob}(PP_{i-P}*GP_{i-P-1})\\&+ \text{prob}(PP_{i-P}* PP_{i-P-1} *GP_{i-P-2}) + \ldots\\&+ \text{prob}(PP_{i-P}* PP_{i-P-1}*\ldots * PP_{i-P-R-1}*GP_{i-P-R})\\&\approx \text{prob}(GP)*(1-\text{prob}(PP)^R)/(1-\text{prob}(PP)) \quad (2)\end{aligned}$$

So probability that CAP has an error at the *i*th bit is

$$\text{prob}(PP)^P * \text{prob}(GP) * (1-\text{prob}(PP)^R)/(1-\text{prob}(PP)) \quad (3)$$

Equation (3) shows that error rate of CAP is related to carry state transition probability (prob(*GP*) and prob(*KP*)) of each bit. For input streams with high spatial correlation, state transition probabilities are small, thus the error rate of CAP can be greatly reduced.

To verify the analysis result, we generate 32-bit highly correlated input signals with switching probability of the *i*th bit equals to 0.5^i. Fig. 5 shows the error rate of CAP predicted carry-out and GeAr predicted carry-out at bit 16 and bit 24. Simulation results show the error rate of CAP is much smaller than GeAr with the same number of prediction bits.

Case 3: *Signed subtraction under temporal correlated inputs*

As mentioned in the previous section, the signs of differences between consecutive inputs that has strong temporal correlation change infrequently. Intuitively, signs of differences between consecutive inputs reflect the tendency of signal changing, either increasing or decreasing. CAP can benefit a lot if signal frequency is low. For example, for a sine wave input with signal frequency *K1*, sampling frequency *K2*, signs of signal difference changes twice in *K2*/*K1* operations so the error rate of CAP is 2**K1*/*K2*.

C. Synthesis Result

We implemented two 32-bit approximate adders (GeAr-based and CAP-based) in Verilog and synthesized them using Synopsys Design Compiler with TSMC 16nm library. Both adders have four 8-bit sub-adders but CAP uses much fewer predictive bits to achieve the same error rate for highly correlated input streams. As shown in Figure 5, CAP with 2 predictive bits achieves the same error rate as GeAr with 6 predictive bits. Table I shows the latency, area and power results of GeAr(32,8,6) and CAP(32,8,2). CAP can reduce latency by 23.33% and save as much as 15.9% in silicon area.

TABLE I. SYNTHESIS RESULTS FOR GEAR(32,8,6) AND CAP(32,8,2)

	GeAr(32,8,6)	*CAP(32,8,2)*	*Reduction*
Delay(ns)	0.30	0.23	23.33%
Area(μm^2)	39.35	33.074	15.9%

V. EXPERIMENTAL RESULTS

In this section, we experiment a 24-bit CAP based approximate adder and a GeAr adder for interchannel decorrelation used in audio encoding standards FLAC [17] to reduce redundancy and save bandwidth. Comparison results on delay, area and power are also reported.

Most of the existing approximate adders [6,7,8,10] use fixed result bits *R* and predictive bits *P* for all sub-adders since for uniform distributed inputs, the error rate is only decided by the number of prediction bits. This may produce non-optimal design for non-uniformly distributed inputs. To explore design space more efficiently, we implemented a procedure which assigns just enough prediction bits for each sub-adder thus produce better design in terms of latency and area. The procedure produced CAP based adder which can be represented as

$$(N,P,R) = (24, [8,6,4,6],[0,2,4,1])$$

Note that previously we use (*N*,*R*,*P*) to describe adder configuration which implies all sub-adders have the same result bit *R* and predictive bits *P*. They become array with *K* elements, where *K* is the number of sub-adders. The adder employs a hybrid prediction scheme where the last predictor (where signal bits are highly correlated) is CAP based while the other predictors (where signal bits are less correlated) are GeAr based. GeAr based predictor suffers from high error rate when predicting for the last sub-adder. According to [15], 99% of predicted errors can be fixed in two cycles and therefore effectiveness delay can be estimated as (1+ER)*delay. To achieve the best effective delay, GeAr(24,8,5) is selected for comparison.

Table II is the comparison results between GeAr(24,8,5) and CAP(24,[8,6,4,6],[0,2,4,1]) on delay, area, error rate and effective delay. CAP based approximate adder consumes 13.17% less in silicon area and reduces latency by 25.92% compared to GeAr. It's interesting to note though CAP has higher error rate, its effective delay is still lower than GeAr due to latency reduction. By scaling supply voltage from 0.9V down to 0.71V so that CAP has the same effective delay as GeAr, our simulation results show that 35% power saving can be achieved.

VI. CONCLUSIONS

In this paper, we first studied the spatial/temporal correlations existing in real input streams as well as carry-in signals. For highly correlated input streams, we proposed correlation aware predictor which unitizes correlation information to predict carry-in value to reduce latency. Analysis and simulation results confirmed its effectiveness. Finally we implemented a 24-bit CAP based approximate adder and demonstrated its usefulness for real application.

TABLE II. SYNTHESIS RESULTS FOR GEAR(24,8,5) AND CAP(24,[8,6,4,6],[0,2,4,1])

	GeAr(24,8,5)	*CAP(24, [8,6,4,6],[0,2,4,1])*	*Reduction*
Delay(ns)	0.27	0.2	25.92%
Area(μm^2)	26.5	23.01	13.17%
Error Rate	0.0757	0.159	-110.03%
Effective Delay(ns)	0.2904	0.2318	20.18%

REFERENCES

[1] J. Han and M. Orshansky, "Approximate computing: An merging paradigm for energy-efficient design", in European Test Symposium, 2013, pp. 1-6

[2] V. Chippa, S. Chakradhar, K. Roy, and A. Raghunatha, "Analysis and characterization of inherenet application resilience for approximate computing", Design Automaton Conference (DAC), 2013

[3] K. Roy and A. Raghunathan, "Approximate Computing: An EnergyEfficient Computing Technique for Error Resilient Applications," in Proc. IEEE Comput. Soc. Annual Symp. VLSI, 2015, pp. 473–475.

[4] D. Esposito, G . Cas etl, "Approximate Adder With Output Correction for Error Tolerant Applications and Gaussian Distributed Inputs", ISCAS 2016

[5] A. K. Verma, P. Brisk, P. Ienne, "Variable Latency Speculative Addition: A New Paradigm for Arithmetic Circuit Design," proc. of Design, Automation and Test in Europe, DATE '08, pp.1250-1255, March 2008

[6] N. Zhu, W. L. Goh, G. Wang, K. S. Yeo, "An enhanced low-power high-speed adder for error tolerant applications", proc. of International Symp. on Integrated Circuits, (ISIC), pp. 67-69, 2009.

[7] K. Du, "High performance reliable variable latency carry select addition", DATE 2012

[8] A. B. Kahng, S. Kang, "Accuracy configurable adder for approximate arithmetic design", proc. of Design Automation Conf., DAC'12, pp.820-825, June 2012, San Francisco.

[9] R. Ye, T. Wang, F. Yuan, R. Kumar, and Q. Xu, "On reconfiguration-oriented approximate adder design and its application", ICCAD 2013

[10] M. Shafique, W. Ahmad, R. Hafiz, J. Henkel, "A low-latency generic accuracy configurable adder", proc. of Design Automation Conf., DAC'15, pp. 86:1-86:6, June 2015, San Francisco.

[11] C. Liu, J. Han, and F. Lombardi, "An analytical framework for evaluating the error characteristics of approximate adders," IEEE Trans. Comput., vol. 64, no. 5, pp. 1268–1281, May 2015.

[12] S. Mazahir, O. Hasan, R. Hafiz, M. Shafique, and J. Henkel. Probabilistic Error Modeling for Approximate Adders. IEEE Transactions on Computers, pages 515–530, 2016.

[13] G. Liu, Y. Tao, M. Tan, and Z. Zhang. "CASA: Correlation-aware speculative adders". In Proceedings of the International Symposium on Low Power Electronics and Design, 2014.

[14] Ali Mural Gok, Nikos Hardavellas, "VaLHALLA: variable Latency History Aware Local-carry Lazy Adder", proc. of the on Creat Lakes Synposium on VLSI, GLSVLSI 2017, pp.17-22.

[15] A. A. Del Barrio, R. Hermida, S. O. Memik, J. M. Mendıas, and M. C. Molina, "Multispeculative addition applied to datapath synthesis," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 31, no. 12, pp. 1817–1830, Dec. 2012.

[16] http://helpguide.sony.net/high-res/sample1/v1/en/index.html

[17] https://xiph.org/flac/format.html#interchannel