PAPA - Packed Arithmetic on a Prefix Adder for Multimedia Applications

Neil Burgess
Division of Electronics,
School of Engineering,
Cardiff University,
CARDIFF CF24 3TF
U.K.
burgessn@cf.ac.uk

Abstract

This paper introduces PAPA: Packed Arithmetic on a Prefix Adder, a new approach to parallel prefix adder design that supports a wide variety of packed arithmetic computations, including packed add and subtract with saturation, packed rounded average, and packed absolute difference. The approach consists of altering the prefix adder cell logic equations to take advantage of a previously unused "don't care" state. Logical effort is employed to assess the delay of the new adder architecture by establishing the extra effort needed to select and drive the appropriate carry signal to the requisite sum sub-word. This adder will find applications in video processors and other multimedia-orientated processor chips that implement packed arithmetic operations.

1. Motivation

Multimedia processor chips (and others) make much use of "packed" arithmetic operations in order to accelerate a variety of digital signal processing algorithms for consumer applications. In such arithmetic units, long wordlength numbers are optionally treated as several independent shorter wordlength numbers — for example, a 32-bit word may be treated as 2 separate 16-bit words or as 4 8-bit words. The main motivation for this mode of operation is to support SIMD processing with its associated advantages in the context of a conventional pipelined load-store processor architecture [1]. Moreover, a common arithmetic operation used in video processing is "absolute difference", denoted \( |A-B| \), and used widely in video motion estimation and prediction algorithms. Hence, a most valuable operation is a "packed absolute difference" operation, which returns the absolute differences of a several independent pairs of 8-bit pixel values simultaneously.

Ordinarily, absolute differences are computed either by performing a subtraction operation followed by a separate "absolute value" operation, which returns the magnitude of a signed number, or by performing a comparison to order the operands followed by a subtraction in which the smaller operand is subtracted from the larger [2]. Instead of such two-step implementations, absolute differences can be obtained by computing both \( A-B \) and \( B-A \), and using the signs of the two results to select the positive result [3]. However, this is wasteful and a better technique is sought. Recently, some authors have described how absolute differences can be derived using a single prefix adder [4, 5], but have not extended this insight to packed arithmetic. Previously re-
ported implementations of packed arithmetic prefix adders include [6, 7], but neither of these proposals is able to support packed late increment operations, vital for computing absolute difference and rounded average instructions.

A second valuable arithmetic option for media applications is saturated arithmetic, in which overflows and underflows do not cause exceptions but rather return pre-defined "saturation constants" [8]. Such constants should ideally be incorporated with little or no performance overhead since the integer adder is typically on the critical path that defines a processor's clock rate. This paper describes how a prefix adder can be altered straightforwardly to support packed absolute difference and rounded average instructions as well as packed saturated arithmetic, and further describes how Logical Effort was employed to assess its performance potential.

2. Packed arithmetic on a parallel prefix adder

2.1 Prefix tree cell logic

The parallel prefix carry-lookahead adder is a popular VLSI design technique that accelerates an n-bit addition by means of a parallel prefix tree [9]. A block diagram of a prefix adder is illustrated in Figure 1, where the adder is seen to consist of three blocks: input bit propagate, generate, and not kill cells; the prefix tree; output sum cells. The input cells derive the bit propagate, generate, and not kill signals respectively according to:

\[ \begin{align*}
\bar{p}(i) & = a(i) \oplus b(i) \\
\bar{g}(i) & = a(i) \land b(i) \\
\bar{k}(i) & = a(i) \lor b(i)
\end{align*} \]  

(1a)  
(1b)  
(1c)

\[ \begin{align*}
\bar{p}(0:n-1) & \\
\bar{g}(0:n-1) & \\
\bar{k}(0:n-1) & \\
\bar{c}(1:n) & \\
\bar{a}(0:n)
\end{align*} \]

\[ \begin{align*}
A(0:n-1) & \\
B(0:n-1)
\end{align*} \]

Input bit carry cells

Prefix carry tree

Sum cells

Figure 1 Block diagram of parallel prefix adder

The prefix carry tree expands the input bit generate and bit not kill signals, \(g(i)\) and \(\bar{k}(i)\), into "group generate" and "group not kill" signals through a number of levels of logic operations. \(G_g\) represents a group generate signal across the bits from significance \(w\) up to and including significance \(z\), and \(\bar{K_g}\) represents a group not kill signal across the same significances. Note that \(G_g^1 = g(i)\) and \(K_g^1 = k(i)\). Each level of logic in the tree widens the bit range of the groups at every significance until the lower value of the range covered by the group is 0, and the carry signals are obtained as:

\[ c(i) = G_g^0 \]  

(2)
Figure 2 shows one of the family of prefix trees proposed by Knowles for \( n = 32 \) [10]. The black squares in Figure 2 are prefix cells which implement the equation pair:

\[
G_i^r = G_i^l \lor \neg K_i^l \land G_i^r \\
\neg K_i^r = \neg K_i^l \land \neg K_i^r
\]  

and the grey squares are cut-down prefix cells implementing (3a) only. The output sum signals are derived according to:

\[
\text{Sum}_{A+B}: s(i) = G_{i,1} \oplus p(i)
\]  

![Diagram of prefix adder](image)

Legend:  
- XOR gate  
- PGK cell  
- black cell  
- grey cell  
- buffer

Figure 2 Knowles' [2,2,1,1,1] prefix adder

It has been shown that a trailing string of \((0, \neg k) = (0, 1)\) tuples ("CP conditions") identifies the trailing string of sum bits that must change from 1 to 0 if the sum is incremented [11]. Hence, the final group generate and group not kill signals can be combined with control signals, denoted "inc" and "abs", to derive alternative carry signals, \(c_{\text{inc}}(i)\) and \(c_{\text{abs}}(i)\) respectively, that yield results related to the original sum:

\[
\text{Incremented Sum, } A+\text{B+1}: c_{\text{inc}}(i) = G_{i,1}^0 \lor \neg K_{i,1}^0 \land \text{inc}
\]

\[
\text{Absolute Difference, } |A-B|: c_{\text{abs}}(i) = G_{i,1}^0 \lor \text{abs} \lor \neg K_{i,1}^0 \land \text{abs}
\]  

Packed versions of these two enhanced operations can be supported on a parallel prefix carry tree by defining a fourth symbol, denoted \(CB\) (B for "block"), that prevents carry information from traversing a column in the adder [12]. The \(CB\) condition can be represented by the "don't care" combination \((G_i^r, \neg K_i^r) = (1, 0)\), implying that the \(CG\) (carry generate) condition must be represented by \((G_i^r, \neg K_i^r) = (1, 1)\), and not \((G_i^r, \neg K_i^r) = (1, X)\). This adder is known as PAPA, and the logic equations for the PAPA black cell are [12]:
\[ G_i^w = G_i^f \lor -K_i^f \land G_i^w^* \]  \hspace{1cm} (7a)
\[ -K_i^w = -K_i^f \land (\neg K_i^w^* \lor G_i^f) \]  \hspace{1cm} (7b)
both of which are implementable as single CMOS logic gates.

A second cell for the packed arithmetic prefix adder is required that operates as a normal prefix cell if no CB conditions are received on the inputs, but which also converts CB’s to CP’s to enable the enhanced additions and subtractions described by (4) and (5). This cell is akin to the “grey” \( G_{i+1} \) (output only) cell in conventional prefix adders in that it replaces the standard “black” prefix cell at the foot of each column in the prefix carry tree. The logic equations of this second cell are [12]:
\[ G_i^x = \neg K_i^f \land (G_i^f \lor G_i^w^*) \]  \hspace{1cm} (8a)
\[ -K_i^x = G_i^f \lor -K_i^f \land -K_i^w \]  \hspace{1cm} (8b)

Again, both expressions are implementable as single CMOS logic gates and amount to no more than a simple rewiring of the PAPA black cell.

2.2 Prefix tree input and output cell logic

The inputs to the adder where CB conditions are injected to mark partition points of packed arithmetic operations require extra logic to force \((g, \neg k) = (1,0)\) instead of \((0,1)\). (Note that \((g,\neg k) = (1,1)\) or \((0,0)\) both already constitute partition points.) A simple way of accomplishing this is to use a pair of CMOS majority gates whose inputs are \(a(i), b(i)\), and \(\text{mode}\), a control signal indicating if an adder is to be partitioned at bit \(i\):
\[ -k = a(i) \land b(i) \lor \neg \text{mode} \land (a(i) \lor b(i)) \]  \hspace{1cm} (9a)
\[ g = a(i) \land b(i) \lor \text{mode} \land (a(i) \lor b(i)) \]  \hspace{1cm} (9b)

These expressions are both recognisable as CMOS minority gates, identical to those used in full adder implementations, and illustrated in Figure 3 [13].

![Figure 3 CMOS minority logic gate](image)

Once the carry signals (following either a full wordlength or packed arithmetic operation) emerge from the prefix tree, they must be combined with the bit propagate signals and other control signals so as to return the required results. However, a consequence of recoding the \(CG\) condition in order to accommodate the \(CB\) condition is that equations (5) and (6) are no longer correct. Instead, the \(G_{i+1}\) outputs give the carries for the sum, \(S\), while the \(\neg K_{i+1}\) outputs give the carries for the sum, \(S+1\), so that the output logic equations become:
Incremented Sum, $A+B+1$: $c_{inc}(t) = \neg inc \land G_{11}^0 \lor inc \land \neg K_{11}^0$  \hspace{1cm} (10)

Absolute Difference, $|A-B|$: $c_{abs}(t) = \neg abs \land \neg G_{11}^0 \lor abs \land \neg K_{11}^0$. \hspace{1cm} (11)

where $abs$ indicates that the difference of the original subtraction, $A-B$, is positive.

Now, the output logic must supply the correct control signals to the appropriate sub-adders so that the desired result is computed using some simple output logic. A block diagram of this approach is presented in Figure 4.

![Figure 4 Schematic diagram of PAPA output logic](image)

In Figure 4, there are two logic blocks per 8-bit sub-adder, labelled “control logic” and “sum set-up & select logic”. The signals “mode8” and “mode16” indicate whether packed 8-bit or packed 16-bit arithmetic is to be performed and conditionally select the appropriate $G_i^0$ and $\neg K_i^0$ bits to be fed to the 8-bit sum set-up blocks. The control logic runs largely in parallel with the operation of the prefix carry tree because under packed operation, the sub-adder carry outputs become available earlier than under full-wordlength operation. For example, under 8-bit operation (mode$8 = 1$), the sub-adder carry outputs are available after only three levels of prefix cells, and under 16-bit operation (mode$16 = 1$), they are available after only four levels of prefix cells. This can be exploited in the design of the control logic by careful scheduling of the logic, as shown in Figure 5, which illustrates the control logic for carry/ovf for bits (7:0) of the adder.

![Figure 5 Control logic for sub-adder computing s(7:0)](image)

The logic equations for the control logic blocks for the other sub-groups of output bits are all simpler to realise than Figure 5:

- $carry/ovf(15:8) = mode8 \land c(16) \lor mode16 \land c(16) \lor not16 \land c(32)$  \hspace{1cm} (12a)
- $carry/ovf(23:16) = mode8 \land c(24) \lor mode16 \land c(32) \lor not16 \land c(32)$  \hspace{1cm} (12b)
- $carry/ovf(31:24) = mode8 \land c(32) \lor mode16 \land c(32) \lor not16 \land c(32)$  \hspace{1cm} (12c)
The second logic block, labelled "sum set-up & select logic", prepares the prospective sum outputs based on the instruction being executed. It does such that when the carry/ovf signal does become available, it needs to select one of only two possible results. A representative set of multimedia-oriented instructions is presented in Table 1, and has been derived from recent multimedia instruction sets [14-16]. The Table also indicates, where appropriate, which of $G_{\kappa+1}$ and $-K_{\kappa+1}$ is to be combined with $\psi(n)$ to form the required sum, as discussed earlier. The rounded average function has been implemented as shown to balance the fan-out loads on the $G_{\kappa+1}$ and $-K_{\kappa+1}$ outputs. It introduces virtually no bias because the probability that an n-bit addition causes an overflow, $P(\alpha(n))$ is given by:

$$P(\alpha(n)) = 1/2 (1 - 2^{-n})$$

Thus there are almost equal probabilities that an average is truncated or incremented.

### Table 1 Representative set of multimedia-enriched add/subtract instructions

<table>
<thead>
<tr>
<th>OPERATION</th>
<th>$S = A + B$</th>
<th>$S = A + \neg B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>INST</td>
<td>$S = A + B$</td>
<td>$S = A + \neg B$</td>
</tr>
<tr>
<td>OVF</td>
<td>$S = A + B$</td>
<td>$S = A + \neg B$</td>
</tr>
<tr>
<td>no</td>
<td>$S = A + B$</td>
<td>$S = A + \neg B$</td>
</tr>
<tr>
<td>yes</td>
<td>$S = A + B$</td>
<td>$S = A + \neg B$</td>
</tr>
</tbody>
</table>

Figure 6 presents a possible implementation of Table 1. The multiplexer control signals, $invB$ and $media$ pick out sub-sets of results in the Table (i.e. columns of Table 10) and the logic has been organised so as to minimise the number of multiplexers needed. The saturation constant, $SatConst$, is also selected while the prefix tree is evaluating and fed to the sum output logic blocks.

![Figure 6 PAPA output logic to implement Table 1](image)

3. Logical Effort applied to a prefix adder

In this section, the principle of Logical Effort [17,18] is applied to a parallel prefix adder design so as to permit an assessment of the relative performance of PAPA. Logical Effort is a design methodology for estimating the number of CMOS stages (including buffers) required to implement a given logic function. The principle uses a small number of basic concepts, which are:
**logical effort** $g$ total FET gate capacitance of a CMOS logic gate relative to that of a minimum-sized inverter  

**electrical effort** $h$ ratio of output capacitance to input capacitance for each CMOS logic gate along a critical path  

**branching effort** $b$ ratio of total capacitative load on one CMOS logic gate's output along the critical path to the FET gate capacitance of the next CMOS gate  

**parasitic delay** $p$ total diffusion capacitance on the output node of a CMOS logic gate relative to that of a minimum-sized inverter

Logical Effort operates by calculating the total path effort along the critical path of a digital CMOS circuit as:

\[ F = GBH \]  

where $G = \Pi g$, $B = \Pi b$, and $H = \Pi h$. The last term reduces to the ratio of the output capacitance loading the last CMOS logic gate to the FET gate capacitance of the first CMOS logic gate along the critical path.

Once the path effort has been calculated, a near-optimum design for the CMOS circuit can be determined by deriving the number of CMOS stages (including buffers) required in the circuit as:

\[ N = \log_{3.3} F \]  

and sizing the FET's along the critical path such that the electrical effort of each logic gate (i.e. the ratio of the total output load capacitance to the input FET gate capacitance) $h = \frac{F^{1/n}}{g}$. Then the total delay of the CMOS circuit may be written as:

\[ D = NF^{1/n} + \sum p \]  

in arbitrary delay units. Dividing this expression by 5 will yield an approximation to the delay in terms of fan-out = 4 ("FO4") inverter delays.

By way of illustration, we shall now apply the tenets of Logical Effort to the prefix adder presented in Figure 2. The critical path of the adder runs vertically through the black cells down bit 15 to the grey cell in $4^{th}$ row, then out through the bit 30 grey cell, and is highlighted by a thick black line in Figure 2. This path has larger cell fan-outs than the path from $a(0)$ and $b(0)$ to $c(31)$. Also, the black cell $K$ outputs have higher fan-outs than the corresponding $G$ outputs so that the following analysis will consider the $K$ outputs of the black cells (grey cells have $G$ outputs only).

![Figure 7 Transistor diagrams of black prefix cell](image)

Figure 7 shows the transistor diagram of the two versions of the black prefix cell: one with true inputs, and the other with inverted inputs. The XOR gates are assumed to be implemented as
CMOS complex (2,2) AOI gates, with a logical effort on each input of 6/3. Table 2 gives the logical effort and parasitic delay of all the different logic gates used in the prefix adder.

Table 2 Logical effort and parasitic delays of CMOS cells in prefix adder

<table>
<thead>
<tr>
<th>Cell type</th>
<th>CMOS gate</th>
<th>logical effort, (g)</th>
<th>parasitic delay, (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGK cell</td>
<td>NOR</td>
<td>5/3</td>
<td>6/3</td>
</tr>
<tr>
<td>Black cell (inv(2) i/p)</td>
<td>NOR</td>
<td>5/3</td>
<td>6/3</td>
</tr>
<tr>
<td>Black cell (true i/p)</td>
<td>NAND</td>
<td>4/3</td>
<td>6/3</td>
</tr>
<tr>
<td>Grey cell (inv(2) i/p)</td>
<td>(2,1) AOI</td>
<td>6/3</td>
<td>8/3</td>
</tr>
<tr>
<td>Grey cell (true i/p)</td>
<td>(2,1) AOI</td>
<td>6/3</td>
<td>7/3</td>
</tr>
<tr>
<td>XOR cell</td>
<td>(2,2) AOI</td>
<td>6/3</td>
<td>12/3</td>
</tr>
</tbody>
</table>

The path effort of the adder can now be calculated by taking into account the fan-out loads at each node along the critical path in order to derive the branching effort. In computing the branching effort, track capacitance has been calculated as one minimum-geometry \(n\)-FET per lateral cell traversed, equivalent to 1/3 of the gate capacitance of a minimum-size inverter. Hence, a track that travels a lateral distance of four cells has been allocated a track fan-out of 4/3 minimum size inverters. The full calculation of the path effort, \(F\), for the adder is presented in Table 3.

Table 3 Calculation of Path Effort for standard prefix adder

<table>
<thead>
<tr>
<th>CMOS cell</th>
<th>logical effort, (g)</th>
<th>fanout load</th>
<th>branching effort, (b)</th>
<th>parasitic delay, (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGK block</td>
<td>5/3</td>
<td>(track+2NOR+OAI)</td>
<td>(1+2\times5+6\times3 / (5/3)) = 17/5</td>
<td>6/3</td>
</tr>
<tr>
<td>Prefix 1</td>
<td>5/3</td>
<td>(track+2NAND+AOI)</td>
<td>(2+2\times4+6\times3 / (4/3)) = 16/4</td>
<td>6/3</td>
</tr>
<tr>
<td>Prefix 2</td>
<td>4/3</td>
<td>(track+2NOR+OAI)</td>
<td>(4+2\times5+6\times3 / (5/3)) = 20/5</td>
<td>6/3</td>
</tr>
<tr>
<td>Prefix 3</td>
<td>5/3</td>
<td>(track+2NAND+AOI)</td>
<td>(8+2\times4+6\times3 / (6/3)) = 22/6</td>
<td>6/3</td>
</tr>
<tr>
<td>Prefix 4</td>
<td>6/3</td>
<td>(track+2OAI+NOT)</td>
<td>(16+2\times6+3\times3 / (6/3)) = 31/6</td>
<td>8/3</td>
</tr>
<tr>
<td>Prefix 5</td>
<td>6/3</td>
<td>NOT+(2,2) AOI</td>
<td>(3+6\times3 / 6 = 9/6)</td>
<td>7/3</td>
</tr>
<tr>
<td>XOR (2,2)</td>
<td>6/3</td>
<td>assume 3 NOT's</td>
<td>1</td>
<td>12/3</td>
</tr>
</tbody>
</table>

Delay estimate: \(G = \Pi g = 49.4\); \(B = \Pi b = 1545.9\); \(H = 1\); \(\Sigma p = 17\)

\(F = GBH = 76367 \rightarrow N = \log_{2}F = 8.8\); \(\therefore F^N = 3.6\)

\(D = 8.8 \times 3.6 + 17 = 31.7 + 17 = 48.5\) delay units = 10 FO4 delays

Hence, according to the logical effort method, the prefix adder of Figure 2 should have a delay of around 10 FO4 inverter delays.

4. Logical Effort applied to the PAPA architecture

Next, Logical Effort is applied to the same prefix adder topology, but now modified to implement PAPA, in order to assess the performance overhead in modifying the cells of the standard
prefix tree. The transistor diagrams for the PAPA black cells are given in Figure 8 and the calculation of the new path effort in Table 4.

![Transistor Diagrams](image)

Figure 8 Transistor diagrams of PAPA black and grey cells

<table>
<thead>
<tr>
<th>CMOS cell</th>
<th>logical effort, $g$</th>
<th>fanout load</th>
<th>branching effort, $b$</th>
<th>parasitic delay, $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGKB</td>
<td>6/3</td>
<td>(trk+2AOI+IOI)</td>
<td>(1+2×6+5)/3 / (5/3) = 18/5</td>
<td>12/3</td>
</tr>
<tr>
<td>Prefix 1</td>
<td>5/3</td>
<td>(trk+2AOI+IOI)</td>
<td>(2+2×6+4)/3 / (6/3) = 18/6</td>
<td>7/3</td>
</tr>
<tr>
<td>Prefix 2</td>
<td>6/3</td>
<td>(trk+2AOI+IOI)</td>
<td>(2×6+6+5)/3 / (5/3) = 21/5</td>
<td>8/3</td>
</tr>
<tr>
<td>Prefix 3</td>
<td>5/3</td>
<td>(trk+2AOI+IOI)</td>
<td>(8+3×6+4)/3 / (6/3) = 30/6</td>
<td>7/3</td>
</tr>
<tr>
<td>Prefix 4</td>
<td>6/3</td>
<td>(trk+2AOI+NOT)</td>
<td>(16+2×6+3)/3 / (6/3) = 31/6</td>
<td>8/3</td>
</tr>
<tr>
<td>Prefix 5</td>
<td>6/3</td>
<td>NOT+(2,2)AOI</td>
<td>(3+6+3) / (6/3) = 9/6</td>
<td>7/3</td>
</tr>
<tr>
<td>XOR (2,2)</td>
<td>6/3</td>
<td>assume 3 NOT’s</td>
<td>1</td>
<td>12/3</td>
</tr>
</tbody>
</table>

Delay estimate: $G = \prod g = 88.9; \hspace{1em} B = \prod b = 1757.7; \hspace{1em} H = 3 / (6+6)/3 = 0.75; \hspace{1em} 2p = 20.3$

$F = GBH = 117795 \rightarrow N = \log_{10} F = 9.1; \hspace{1em} \therefore F^{\text{perl}} = 3.6$

$D = 9.1 \times 3.6 + 20.3 = 32.8 + 20.3 = 53.1$ delay units = 10.5 FO4 delays

PAPA (excluding the output logic blocks) runs approximately 5% slower than the standard prefix adder. This increase in delay is mostly due to the increase in parasitic delays, $p$, of the more complex gates employed in the prefix adder. Harris has shown that this definition of parasitic delay is pessimistic [18] so that the delay increase incurred in converting a prefix adder to a PAPA adder is likely to be even less than the figure of 5% arrived at here.

Finally, Logical Effort is applied to the output logic blocks to determine how closely the delays of the logic blocks match one another and how much delay penalty is introduced. There are two distinct paths through the PAPA output logic, corresponding to the two output logic blocks drawn in Figures 5 and 6. In both cases, the last carry signal to become available is $c(32)$ and so the logical effort analyses both start from the last row of output grey cells of the prefix tree. In the control logic blocks, the sub-adder m.s.b. carry outputs traverse some logic stages before being buffered and broadcast to the output multiplexers in the sum logic blocks. The analysis of the control logic block is presented in Table 5 and includes a buffer to drive $c(32)$ across the whole width of the adder. $H$ is set to 3/4 in common with Table 4.
Table 5 Calculation of Path Effort for control logic

<table>
<thead>
<tr>
<th>Cell</th>
<th>logical effort, $g$</th>
<th>fanout load</th>
<th>branching effort, $b$</th>
<th>parasitic delay, $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>c(w) buffer</td>
<td>3/3</td>
<td>trk + 4AOI</td>
<td>(32+4×6)/3 / (6/3) = 56/6</td>
<td>3/3</td>
</tr>
<tr>
<td>(2,1) AOI</td>
<td>6/3</td>
<td>trk+8MUX+NOT</td>
<td>(8+8×6+3)/3 / (6/3) = 59/6</td>
<td>7/3</td>
</tr>
<tr>
<td>MUX</td>
<td>6/3</td>
<td>assume 3 NOT's</td>
<td>1</td>
<td>12/3</td>
</tr>
</tbody>
</table>

Delay estimate: $G = 11g = 4$; $B = 11b = 91$; $H = 0.75$; $\Sigma p = 7.3$

$F = GBH = 273 \rightarrow N = \log_{2} F = 4.4$; $F^{1/v} = 3.6$

$D = 4.4 \times 3.6 + 7.3 = 23.1$ delay units $= 4.5$ FO4 delays

In the sum logic block, sum signals are derived from the group generate and group not kill outputs of the prefix carry tree, and then traverse through multiplexers as shown in Figure 6. The Logical Effort analysis is presented in Table 6, where, as before, the MUX’s are assumed to be constructed from (2,2) AOI gates and that $H = 0.75$.

Tables 5 and 6 show that both output logic blocks each add around 5 FO4 inverter delays to the delay of the prefix carry tree and are thus well matched for speed, given the pessimistic values for parasitic delay, $p$. The delay due to the output XOR gates is around 1 FO4 inverter delay so that the output logic adds 4 FO4 delays to the 10.5 FO4 delays of the PAPA adder, for a total of 14.5 FO4 delays.

Table 6 Calculation of Path Effort for control logic

<table>
<thead>
<tr>
<th>CMOS cell</th>
<th>logical effort, $g$</th>
<th>fanout load</th>
<th>branching effort, $b$</th>
<th>parasitic delay, $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>XOR</td>
<td>6/3</td>
<td>2 MUX</td>
<td>(2 × 6/3) / (6/3) = 12/3</td>
<td>12/3</td>
</tr>
<tr>
<td>MUX</td>
<td>6/3</td>
<td>1 MUX</td>
<td>1</td>
<td>12/3</td>
</tr>
<tr>
<td>MUX</td>
<td>6/3</td>
<td>1 MUX</td>
<td>1</td>
<td>12/3</td>
</tr>
<tr>
<td>MUX</td>
<td>6/3</td>
<td>3 NOT's</td>
<td>1</td>
<td>12/3</td>
</tr>
</tbody>
</table>

Delay estimate: $G = 11g = 16$; $B = 11b = 2$; $H = 0.75$; $\Sigma p = 16$

$F = GBH = 24 \rightarrow N = \log_{2} F = 2.5$; $F^{1/v} = 3.6$

$D = 2.5 \times 3.6 + 16 = 25$ delay units $= 5$ FO4 delays

5. Summary and Future Work

This paper has introduced PAPA - a means of implementing Packed Arithmetic on a Prefix Adder [19]. The key insight has been a new definition of the prefix tree’s function, taking advantage of a previously unused ‘don’t care state’. This adder should be of use in the efficient design of VLSI processor chips implementing media enhanced instruction sets because of its support of SIMD operation of instructions such as packed addition and subtraction with saturation, and packed absolute difference and rounded average. The PAPA concept can be extended to cover other media-oriented instructions not considered in this paper; for example, Galois field arithmetic for Reed-Solomon coding can be supported by using the sub-word carry signals to select between $S$ and $S^+1$. Similarly, min/max instructions should be straightforward to implement. The wide variety of possible overflow conditions that depend on which, if either, of the operands were signed require further study to assess the impact on the adder’s performance of supporting all
mechanisms. Finally, detailed VLSI layout and simulation would give a more accurate assessment of PAPA's performance relative to a standard prefix adder.

6. Acknowledgement

The author wishes to thank Professors Ruby Lee (Princeton University) and David Harris (Harvey Mudd College, Claremont) for their time, hospitality, and friendship that was given during the inception and writing of this paper.

7. References