



Parallel Implementation of Fixed-Point FFTs on TigerSHARC® Processors

Contributed by Boris Lerner

Rev 1 – February 3, 2005

Introduction

The advance of modern highly paralleled processors, such as the Analog Devices TigerSHARC® family of processors, requires finding efficient ways to parallel implementations of many standard algorithms. This applications note not only explains how the fastest 16-bit FFT implementation on the TigerSHARC works, but also provides guidance about algorithm development, so you can apply similar techniques to other algorithms.

Generally, most algorithms have several levels of optimization, which are discussed in detail in this note. The first and most straightforward level of optimization is the paralleling of instructions, as permitted by the processor's architecture. This is simple and boring. The second level of optimization is loop unrolling and software pipelining to achieve maximum parallelism and to avoid pipeline stalls. Although more complex than the simple parallelism of level one, this can be done in prescribed steps without a firm understanding of the algorithm and, thus, requires little ingenuity. The third level is restructuring the math of the algorithm to still produce valid results, but fit the processor's architecture better. This requires a thorough understanding of the algorithm and, unlike software pipelining, there are no prescribed steps that lead to the optimal solution. This is where most of the fun in writing optimized code lies.

In practical applications, it is often unnecessary to go through all of these levels. When all of the

levels are required, it is best to perform these levels of optimization in reverse order. By the time the code is fully pipelined, it is too late to try to change the fundamental underlying algorithm. Thus, a programmer would have to think about the algorithm structure first and organize the code accordingly. Then, levels one and two (paralleling, unrolling, and pipelining) are usually done at the same time.

The code to which this note refers is supplied by Analog Devices. A 256-point FFT is used as the specific example, but the mathematics and ideas apply equally to other sizes (no smaller than 16 points).

As we will see, the restructured algorithm breaks down the FFT into much smaller parts that can then be paralleled. In the case of the 256-point FFT (its code listing is at the end of this applications note), the FFT is split into 16 FFTs of 16 points each and each, 16-point FFT is done in radix-4 fashion (i.e., each has only two stages). If we were to do a 512-point FFT, we would have to do 16 FFTs of 32 points each (and, also, 32 FFTs of 16 points each), each 32-point FFT would have the first two stages done in radix-4 and the last stage in radix-2. These differences imply that it would be difficult to write the code that is FFT size-generic. Although the implemented algorithm is generic and applies equally well to all sizes, the code is not, and it must be hand-tuned to each point size to be able to take full advantage of its optimization.

With all this in mind, let us dive into the fascinating world of fixed-point FFTs in the land of the TigerSHARC.

Standard Radix-2 FFT Algorithm

Figure 1 shows a standard 16-point radix-2 FFT implementation, after the input has been bit-reversed. Traditionally, in this algorithm, stages

1 and 2 are combined with the required bit reversing into a single optimized loop (since these two stages require no multiplies, only adds and subtracts). Each of the remaining stages is usually done by combining the butterflies that share the same twiddle factors together into groups (so the twiddles need only to be fetched once for each group).

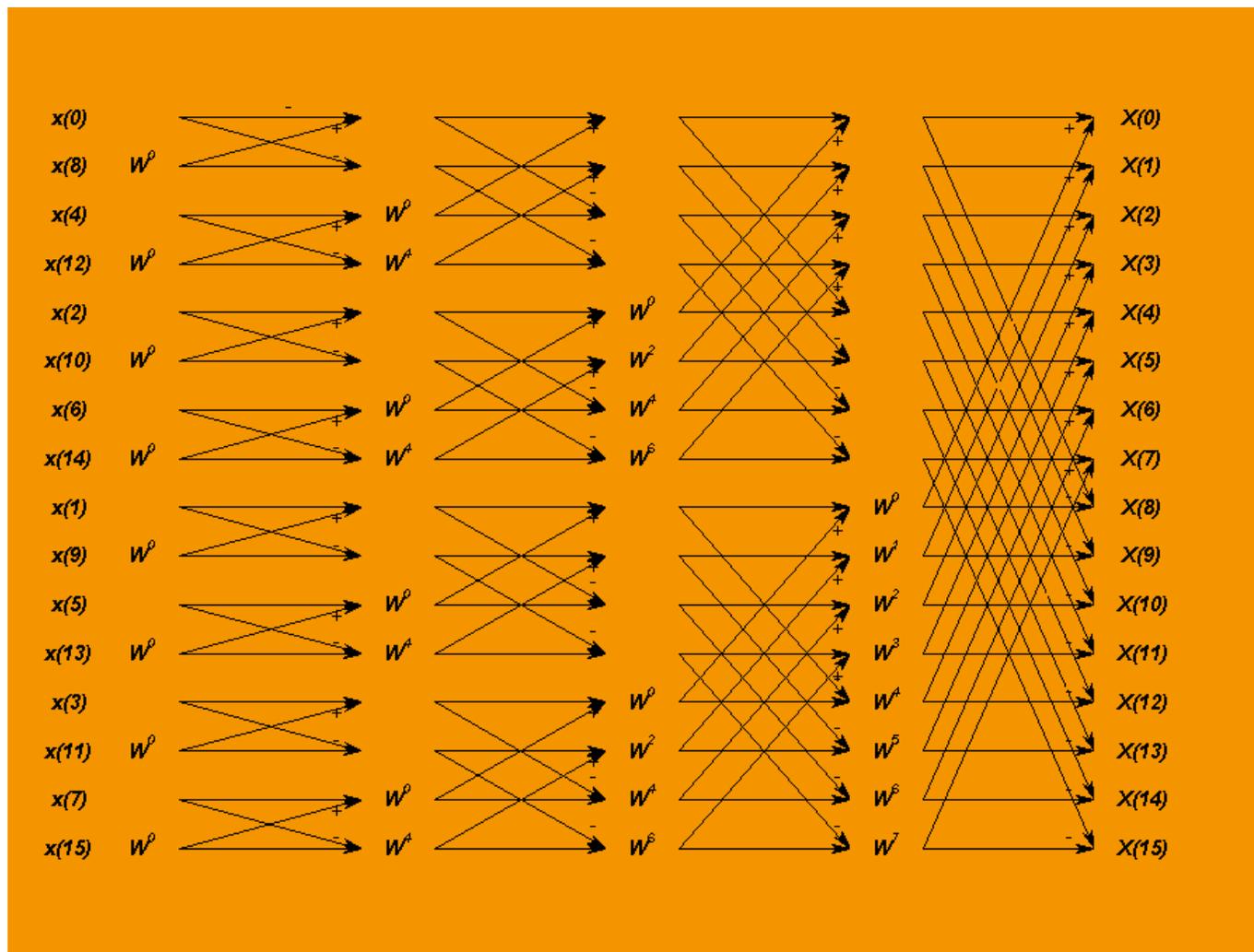


Figure 1. Standard Structure of the 16-Point FFT

Since TigerSHARC processors offer vectorized 16-bit processing on packed data, we would like to parallel this algorithm into at least as many parallel processes as the TigerSHARC can handle. An add/subtract instruction of the TigerSHARC (which is instrumental in

computing a fundamental butterfly) can be paralleled to be performed on eight 16-bit values per cycle (four in each compute block of the TigerSHARC processor). Since data is complex, this equates to four add/subtracts of data per cycle. Thus, we would like to break the FFT into

at least four parallel processes. Looking at the diagram of Figure 1, it is clear that we can do this by simply combining the data into blocks, four points at a time, i.e.:

$$1st\ block = \{x(0), x(8), x(4), x(12)\}$$

$$2nd\ block = \{x(2), x(10), x(6), x(14)\}$$

$$3rd\ block = \{x(1), x(9), x(5), x(13)\}$$

$$4th\ block = \{x(3), x(11), x(7), x(15)\}$$

These groups have no interdependencies and will parallel very nicely for the first two stages of the FFT. After that we are in trouble and the parallelism is gone. At this point, however, we could re-arrange the data into different blocks to ensure that the rest of the way the new blocks do not crosstalk to each other and, thus, can be paralleled. A careful examination shows that the required re-arrangement is an operation of interleaving (or de-interleaving), with new blocks given by:

$$1st\ block = \{x(0), x(2), x(1), x(3)\}$$

$$2nd\ block = \{x(8), x(10), x(9), x(11)\}$$

$$3rd\ block = \{x(4), x(6), x(5), x(7)\}$$

$$4th\ block = \{x(12), x(14), x(13), x(15)\}$$

Another way to look at these new blocks is as the 4x4 matrix transpose (where blocks define the matrix rows). Of course, there is a significant side effect—after the data re-arrangement the last stages will parallel, but will not produce data in the correct order. Maybe we can compensate for this by starting with an order other than the bit-reverse we had started with, but let us leave this detail for the rigorous mathematical analysis that comes later.

At this time, the analysis of the 16 point FFT seems to suggest that, in general, given an N point FFT, we would like to view it in two dimensions as a $\sqrt{N} \times \sqrt{N}$ matrix of data and parallel-process the rows or columns, then transpose the matrix and parallel-process the rows or columns again. Another requirement that

comes from this is that N must be a perfect square. As it turns out, we can dispose of this requirement, but that will be discussed later. At this time we are concerned with the 256-point FFT and, as luck would have it, $256 = 16^2$.

So, which shall we parallel, the rows or the columns? The answer lies in the TigerSHARC processor's vector architecture. When the TigerSHARC processor fetches data from memory, it fetches it in chunks of 128 bits at a time (i.e., four 16-bit complex data points) and packs it into quad or paired (for SIMD fetches) registers. Then it vectorizes processing across the register quad or pair. Thus, it is the columns of the matrix that we want to parallel (i.e., we would like to structure our math so that all the columns of the matrix are independent from one another).

Now that we know what we would like the math to give us, it is time to do this rigorously in the language of mathematics.

Mathematics of the Algorithm

The following notation will be used:

N = Number of points in the original FFT (256 in our example),

$$M = \sqrt{N},$$

\hat{x} will stand for the Discrete Fourier Transform (henceforth abbreviated as DFT) of x .

Now, given signal x ,

$$\hat{x}(n) = \sum_{k=0}^{N-1} x(k) e^{-\frac{2\pi i n k}{N}} = \sum_{m=0}^{M-1} \sum_{l=0}^{M-1} x(Ml + m) e^{-\frac{2\pi i n (Ml + m)}{N}} =$$

$$\sum_{m=0}^{M-1} e^{-\frac{2\pi i n m}{N}} \sum_{l=0}^{M-1} x(Ml + m) e^{-\frac{2\pi i n l}{M}} = \sum_{m=0}^{M-1} e^{-\frac{2\pi i n m}{N}} \hat{x}_m(n)$$

where:

$$x_m(l) := x(Ml + m) \tag{1}$$

and \hat{x}_m is this function's M -point DFT. Now, we view the output index n as arranged in an $M \times M$ matrix (i.e., $n = Ms + t$, $0 \leq s, t < M - 1$) Thus,

$$\hat{x}(Ms + t) = \sum_{m=0}^{M-1} e^{\frac{-2\pi i(Ms+t)m}{N}} \hat{x}_m(Ms + t) = \sum_{m=0}^{M-1} e^{\frac{-2\piism}{M}} e^{\frac{-2\piitm}{N}} \hat{x}_m(t)$$

since \hat{x}_m , being an M -point DFT, is periodic with period = M . Thus,

$$\hat{x}(Ms + t) = \sum_{m=0}^{M-1} e^{\frac{-2\piism}{M}} x_t^*(m) = x_t^*(s), \quad (2)$$

where:

$$x_t^*(m) := e^{\frac{-2\piim}{N}} \hat{x}_m(t) \quad (3)$$

and x_t^* is this function's M -point DFT.

Implementation of the Algorithm

Equations (1), (2), and (3) show how to compute the DFT of x using the following steps (here we go back to our specific example of $N=256$, $M=16$):

1. Arrange the 256 points of the input data $x(n)$ linearly, but think of it as a 16×16 matrix:

$$\begin{bmatrix} x(0) & x(1) & x(2) & \cdots & x(15) \\ x(16) & x(17) & x(18) & \cdots & x(31) \\ x(32) & x(33) & x(34) & \cdots & x(47) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(240) & x(241) & x(242) & \cdots & x(255) \end{bmatrix}$$

2. Using equation (1), re-write as:

$$\begin{bmatrix} x_0(0) & x_1(0) & x_2(0) & \cdots & x_{15}(0) \\ x_0(1) & x_1(1) & x_2(1) & \cdots & x_{15}(1) \\ x_0(2) & x_1(2) & x_2(2) & \cdots & x_{15}(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_0(15) & x_1(15) & x_2(15) & \cdots & x_{15}(15) \end{bmatrix}$$

3. We now compute parallel FFTs on columns (as mentioned before, TigerSHARC processors do this very efficiently) obtaining:

$$\begin{bmatrix} \hat{x}_0(0) & \hat{x}_1(0) & \hat{x}_2(0) & \cdots & \hat{x}_{15}(0) \\ \hat{x}_0(1) & \hat{x}_1(1) & \hat{x}_2(1) & \cdots & \hat{x}_{15}(1) \\ \hat{x}_0(2) & \hat{x}_1(2) & \hat{x}_2(2) & \cdots & \hat{x}_{15}(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{x}_0(15) & \hat{x}_1(15) & \hat{x}_2(15) & \cdots & \hat{x}_{15}(15) \end{bmatrix}$$

4. We point-wise multiply this by matrix

$$\left[e^{\frac{-2\piim}{256}} \right]_{0 \leq t, m \leq 15}$$

to obtain

$$\begin{bmatrix} \hat{x}_0(0)e^{\frac{-2\pi i0}{256}} & \hat{x}_1(0)e^{\frac{-2\pi i0}{256}} & \hat{x}_2(0)e^{\frac{-2\pi i0}{256}} & \cdots & \hat{x}_{15}(0)e^{\frac{-2\pi i0}{256}} \\ \hat{x}_0(1)e^{\frac{-2\pi i0}{256}} & \hat{x}_1(1)e^{\frac{-2\pi i1}{256}} & \hat{x}_2(1)e^{\frac{-2\pi i2}{256}} & \cdots & \hat{x}_{15}(1)e^{\frac{-2\pi i15}{256}} \\ \hat{x}_0(2)e^{\frac{-2\pi i0}{256}} & \hat{x}_1(2)e^{\frac{-2\pi i2}{256}} & \hat{x}_2(2)e^{\frac{-2\pi i4}{256}} & \cdots & \hat{x}_{15}(2)e^{\frac{-2\pi i30}{256}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{x}_0(15)e^{\frac{-2\pi i0}{256}} & \hat{x}_1(15)e^{\frac{-2\pi i15}{256}} & \hat{x}_2(15)e^{\frac{-2\pi i30}{256}} & \cdots & \hat{x}_{15}(15)e^{\frac{-2\pi i225}{256}} \end{bmatrix}$$

which, according to equation (3) is precisely

$$\begin{bmatrix} x_0^*(0) & x_0^*(1) & x_0^*(2) & \cdots & x_0^*(15) \\ x_1^*(0) & x_1^*(1) & x_1^*(2) & \cdots & x_1^*(15) \\ x_2^*(0) & x_2^*(1) & x_2^*(2) & \cdots & x_2^*(15) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{15}^*(0) & x_{15}^*(1) & x_{15}^*(2) & \cdots & x_{15}^*(15) \end{bmatrix}$$

5. Now we would like to compute the 16-point FFTs of $x_t^*(m)$, but these are arranged to be paralleled in rows instead of columns. Thus, we have to transpose to obtain

$$\begin{bmatrix} x_0^*(0) & x_1^*(0) & x_2^*(0) & \cdots & x_{15}^*(0) \\ x_0^*(1) & x_1^*(1) & x_2^*(1) & \cdots & x_{15}^*(1) \\ x_0^*(2) & x_1^*(2) & x_2^*(2) & \cdots & x_{15}^*(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_0^*(15) & x_1^*(15) & x_2^*(15) & \cdots & x_{15}^*(15) \end{bmatrix}$$

6. We compute parallel FFTs on the columns and use equation (2) to obtain

$$\begin{bmatrix} \hat{x}(0) & \hat{x}(1) & \hat{x}(2) & \cdots & \hat{x}(15) \\ \hat{x}(16) & \hat{x}(17) & \hat{x}(18) & \cdots & \hat{x}(31) \\ \hat{x}(32) & \hat{x}(33) & \hat{x}(34) & \cdots & \hat{x}(47) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{x}(240) & \hat{x}(241) & \hat{x}(242) & \cdots & \hat{x}(255) \end{bmatrix}$$

This is the FFT result that we want, and it is in the correct order! The math is done, and we are ready to consider the programming implementation. In the following discussion, we will refer to the steps outlined above as Steps 1 through 6.

Programming Implementation

We will go through the steps of the previous section, one step at a time.

Steps 1 and 2 do not need to be programmed. The input data is already arranged in the proper order.

Step 3 requires us to compute 16 parallel 16-point FFTs on the columns of the input matrix. As mentioned before, the TigerSHARC can easily parallel four FFTs at a time, thanks to its vector processing, so we can do four FFTs at a time and repeat this process four times to compute all 16 FFTs. 16-point FFTs can be done very efficiently in radix-4, resulting in two stages, at four butterflies per stage. To minimize overhead, it is more efficient to compute only the first stage for each of the four sets of the four FFTs, followed by computing the second stage

for each of the four sets of the four FFTs (instead of doing both stages for each set).

Step 4 is a point-wise complex multiply (256 multiplies in total), and Step 5 is a matrix transpose. These two steps can be combined—while multiplying the data points we can store them in a transposed fashion.

Step 6 is identical to Step 3—we have to compute 16 parallel 16-point FFTs on the columns of our new input matrix. This part need not need be written. We can simply branch to the code of Step 3, remembering to exit the routine once the FFTs are finished (instead of going on to the Step 4, as before).

Figure 2 represent buffers containing the data, and arrows correspond to transformations of data between buffers.

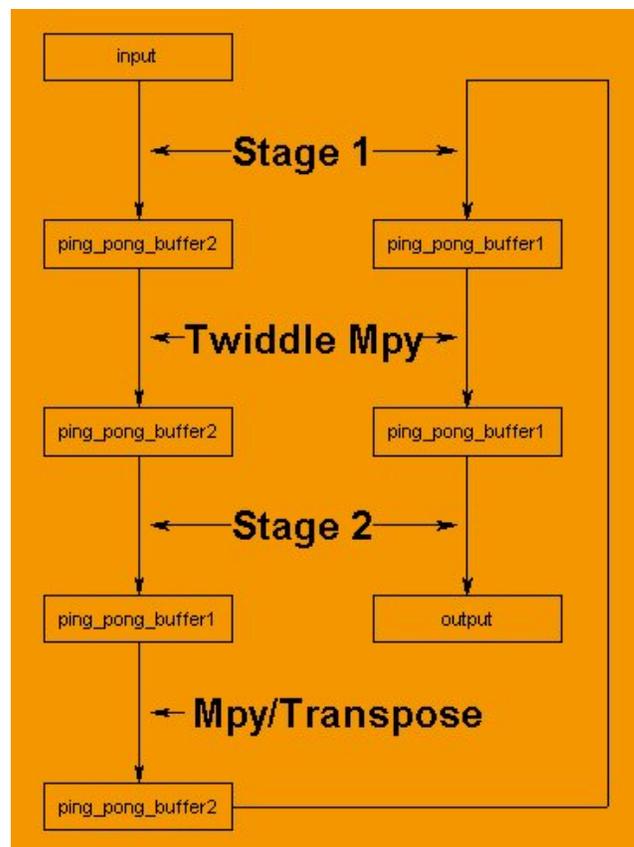


Figure 2. Block Diagram of the Code

Pipelining the Algorithm - Stage 1

Let us concentrate on Step 3 first.

<i>Mnemonic</i>	<i>Operation</i>
<i>F1</i>	<i>Fetch 4 complex Input1 of the 4 butterflies=4 32-bit values</i>
<i>F2</i>	<i>Fetch 4 complex Input2 of the 4 butterflies=4 32-bit values</i>
<i>F3</i>	<i>Fetch 4 complex Input3 of the 4 butterflies=4 32-bit values</i>
<i>F4</i>	<i>Fetch 4 complex Input4 of the 4 butterflies=4 32-bit values</i>
<i>AS1</i>	<i>F1+/-F2</i>
<i>AS2</i>	<i>F3+/-F4</i>
<i>MPY1</i>	<i>1st half of (F1-F2)*(-i) Note that we can do only 2 complex mpy's per cycle</i>
<i>M1</i>	<i>Move MPY1 into compute block register</i>
<i>MPY2</i>	<i>2nd half of (F1-F2)*(-i) Note that we can do only 2 complex mpy's per cycle</i>
<i>M2</i>	<i>Move MPY2 into compute block register</i>
<i>AS3</i>	<i>(F3 + F4)+/(F1+F2) =Output1 and Output3 of the 4 butterflies</i>
<i>AS4</i>	<i>(F3 - F4)+/(F1-F2)*(-i) =Output2 and Output 4 of the 4 butterflies</i>
<i>S1</i>	<i>Store (Output1) of the 4 butterflies=4 32-bit values</i>
<i>S2</i>	<i>Store (Output2) of the 4 butterflies=4 32-bit values</i>
<i>S3</i>	<i>Store (Output3) of the 4 butterflies=4 32-bit values</i>
<i>S4</i>	<i>Store (Output4) of the 4 butterflies=4 32-bit values</i>

Table 1. Single Butterfly of Stage 1 Done Linearly – Logical Implementation

Table 1 lists the operations necessary to perform four parallel radix-4 complex butterflies of stage 1 in vector fashion of the TigerSHARC processor. Actually, this portion is the same for stages other than first, except that the other stages also require a complex twiddle multiply at the beginning of the butterfly. This makes other stages more complicated, and they will be dealt with in the next section.

<i>Cycle/Operation</i>	<i>JALU</i>	<i>KALU</i>	<i>MAC</i>	<i>ALU</i>
<i>1</i>	<i>F1</i>	<i>S4---</i>		<i>AS3--</i>
<i>2</i>	<i>F2</i>	<i>S1--</i>	<i>MPY1-</i>	<i>AS2-</i>
<i>3</i>	<i>F3</i>	<i>S2--</i>	<i>M1-, MPY2-</i>	<i>AS4--</i>
<i>4</i>	<i>F4</i>	<i>S3--</i>	<i>M2-</i>	<i>AS1</i>
<i>5</i>	<i>F1+</i>	<i>S4--</i>		<i>AS3-</i>
<i>6</i>	<i>F2+</i>	<i>S1-</i>	<i>MPY1</i>	<i>AS2</i>
<i>7</i>	<i>F3+</i>	<i>S2-</i>	<i>M1, MPY2</i>	<i>AS4-</i>
<i>8</i>	<i>F4+</i>	<i>S3-</i>	<i>M2</i>	<i>AS1+</i>
<i>9</i>	<i>F1++</i>	<i>S4-</i>		<i>AS3</i>
<i>10</i>	<i>F2++</i>	<i>S1</i>	<i>MPY1+</i>	<i>AS2+</i>
<i>11</i>	<i>F3++</i>	<i>S2</i>	<i>M1+, MPY2+</i>	<i>AS4</i>
<i>12</i>	<i>F4++</i>	<i>S3</i>	<i>M2+</i>	<i>AS1++</i>
<i>13</i>	<i>F1+++</i>	<i>S4</i>		<i>AS3+</i>
<i>14</i>	<i>F2+++</i>	<i>S1+</i>	<i>MPY1++</i>	<i>AS2++</i>
<i>15</i>	<i>F3+++</i>	<i>S2+</i>	<i>M1++, MPY2++</i>	<i>AS4+</i>
<i>16</i>	<i>F4+++</i>	<i>S3+</i>	<i>M2++</i>	<i>AS1++ +</i>

Table 2. Pipelined Butterflies – Stage 1

Table 2 shows the butterflies pipelined. A “+” in the operation indicates the operation that

corresponds to the next set of the butterflies and a “-” corresponds to the operation in the previous set of butterflies. All instructions are paralleled, and there are no stalls.

Pipelining the Algorithm - Stage 2

To do a butterfly for stage 2, one must perform the same computations as for stage 1, except for an additional complex multiply of each of the 16 (4 paralleled x 4 points) inputs at the beginning. This creates a problem. The original butterfly has two SIMD complex multiplies in it already. Adding 8 more makes 10 complex multiplies, while ALU, fetch, and store remain at 4. This would make the algorithm unbalanced—too many multiplies and too few other compute units will not parallel well. The best that can be done this way is 10 cycles per vector of butterflies.

It turns out that each of the two original multiplies (being multiplies by $-i$) can be replaced by one short ALU (negate) and one long rotate. In addition, two register moves are required to ensure that the data is back to being packed in long registers for the parallel add/subtracts that follow. For the vectored butterfly, this would leave a total of 8 multiplies, 6 ALUs (4 add/subtracts and 2 negates), and 2 shifts (rotates)—perfectly balanced for an 8-cycle execution. Also, 4 fetches and 4 stores leave plenty of room for register moves.

Table 3 lists the operations necessary to perform vectored (i.e., four parallel) radix-4 complex butterflies of stage 2 on the TigerSHARC processor. Things have gotten significantly more complex!

Table 4 shows the butterflies pipelined. A “+” in the operation indicates the operation that corresponds to the next set of the butterflies, and a “-” corresponds to the operation in the previous set of the butterflies.

<i>Mnemonic</i>	<i>Operation</i>
<i>F1</i>	<i>Fetch 4 complex Input1 of the 4 butterflies=4 32-bit values</i>
<i>F2</i>	<i>Fetch 4 complex Input2 of the 4 butterflies=4 32-bit values</i>
<i>F3</i>	<i>Fetch 4 complex Input3 of the 4 butterflies=4 32-bit values</i>
<i>F4</i>	<i>Fetch 4 complex Input4 of the 4 butterflies=4 32-bit values</i>
<i>MPY1</i>	<i>1st half of F1*twiddle</i>
<i>M1</i>	<i>Move MPY1 into compute block register</i>
<i>MPY2</i>	<i>2nd half of F1*twiddle</i>
<i>M2</i>	<i>Move MPY2 into compute block register</i>
<i>MPY3</i>	<i>1st half of F2*twiddle</i>
<i>M3</i>	<i>Move MPY3 into compute block register</i>
<i>MPY4</i>	<i>2nd half of F2*twiddle</i>
<i>M4</i>	<i>Move MPY4 into compute block register</i>
<i>MPY5</i>	<i>1st half of F3*twiddle</i>
<i>M5</i>	<i>Move MPY5 into compute block register</i>
<i>MPY6</i>	<i>2nd half of F3*twiddle</i>
<i>M6</i>	<i>Move MPY6 into compute block register</i>
<i>MPY7</i>	<i>1st half of F4*twiddle</i>
<i>M7</i>	<i>Move MPY7 into compute block register</i>
<i>MPY8</i>	<i>2nd half of F4*twiddle</i>
<i>M8</i>	<i>Move MPY8 into compute block register</i>
<i>AS1</i>	<i>M1,M2+/-M3,M4</i>
<i>AS2</i>	<i>M5,M6+/-M7,M8</i>
<i>A1</i>	<i>Negate (M1-M3)</i>
<i>MV1</i>	<i>Move (M1-M3) into a pair of the register that contains A1</i>

<i>R1</i>	<i>Rotate the long result of A1,MV1 – now the low register contains (M1-M3)*(-i)</i>
<i>A2</i>	<i>Negate (M2-M4)</i>
<i>MV2</i>	<i>Move (M2-M4) into a pair of the register that contains A2</i>
<i>R2</i>	<i>Rotate the long result of A2,MV2 – now the low register contains (M2-M4)*(-i)</i>
<i>AS3</i>	$(F3 + F4) + /-(F1+F2)$
<i>AS4</i>	$(F3 - F4) + /-(F1-F2)*(-i)$ <i>Here (F1-F2)*(-i) was obtained by R1 and R2</i>
<i>S1</i>	<i>Store 1 of the 4 butterflies=4 32-bit values</i>
<i>S2</i>	<i>Store 2 of the 4 butterflies=4 32-bit values</i>
<i>S3</i>	<i>Store 3 of the 4 butterflies=4 32-bit values</i>
<i>S4</i>	<i>Store 4 of the 4 butterflies=4 32-bit values</i>

Table 3. Single Butterfly of Stage 2 Done Linearly – Logical Implementation

All instructions are paralleled, and there are no stalls. There is still a question of twiddle fetch which was not addressed, but there are so many JALU and KALU instruction slots still available that scheduling twiddle fetches will not cause any problems (they are actually the same value per vector, so broadcast reads will bring them in efficiently).

<i>Cycle/Operation</i>	<i>JALU</i>	<i>KALU</i>	<i>MAC</i>	<i>ALU</i>
<i>1</i>	<i>F1</i>	<i>MV1--</i>	<i>M3-, MPY4-</i>	<i>A1--</i>
<i>2</i>	<i>F2</i>		<i>M4-, MPY5-</i>	<i>A2--</i>
<i>3</i>		<i>S3---</i>	<i>M5-, MPY6-</i>	<i>R1--</i>
<i>4</i>		<i>S4---</i>	<i>M6-, MPY7-</i>	<i>R2--</i>
<i>5</i>	<i>F3</i>	<i>MV2--</i>	<i>M7-, MPY8-</i>	<i>AS3--</i>
<i>6</i>	<i>F4</i>		<i>M8-, MPY1</i>	<i>AS1-</i>

<i>7</i>		<i>S1--</i>	<i>M1, MPY2</i>	<i>AS4--</i>
<i>8</i>		<i>S2--</i>	<i>M2, MPY3</i>	<i>AS2-</i>
<i>9</i>	<i>F1+</i>	<i>MV1-</i>	<i>M3, MPY4</i>	<i>A1-</i>
<i>10</i>	<i>F2+</i>		<i>M4, MPY5</i>	<i>A2-</i>
<i>11</i>		<i>S3--</i>	<i>M5, MPY6</i>	<i>R1-</i>
<i>12</i>		<i>S4--</i>	<i>M6, MPY7</i>	<i>R2-</i>
<i>13</i>	<i>F3+</i>	<i>MV2-</i>	<i>M7, MPY8</i>	<i>AS3-</i>
<i>14</i>	<i>F4+</i>		<i>M8, MPY1+</i>	<i>AS1</i>
<i>15</i>		<i>S1-</i>	<i>M1+, MPY2+</i>	<i>AS4-</i>
<i>16</i>		<i>S2-</i>	<i>M2+, MPY3+</i>	<i>AS2</i>
<i>17</i>	<i>F1++</i>	<i>MV1</i>	<i>M3+, MPY4+</i>	<i>A1</i>
<i>18</i>	<i>F2++</i>		<i>M4+, MPY5+</i>	<i>A2</i>
<i>19</i>		<i>S3-</i>	<i>M5+, MPY6+</i>	<i>R1</i>
<i>20</i>		<i>S4-</i>	<i>M6+, MPY7+</i>	<i>R2</i>
<i>21</i>	<i>F3++</i>	<i>MV2</i>	<i>M7+, MPY8+</i>	<i>AS3</i>
<i>22</i>	<i>F4++</i>		<i>M8+, MPY1++</i>	<i>AS1+</i>
<i>23</i>		<i>S1</i>	<i>M1++, MPY2++</i>	<i>AS4</i>
<i>24</i>		<i>S2</i>	<i>M2++, MPY3++</i>	<i>AS2+</i>
<i>25</i>	<i>F1+++</i>	<i>MV1+</i>	<i>M3++, MPY4++</i>	<i>A1+</i>
<i>26</i>	<i>F2+++</i>		<i>M4++, MPY5++</i>	<i>A2+</i>
<i>27</i>		<i>S3</i>	<i>M5++, MPY6++</i>	<i>R1+</i>
<i>28</i>		<i>S4</i>	<i>M6++, MPY7++</i>	<i>R2+</i>

Table 4. Pipelined Butterflies – Stage 2

The multiplies and transpose of Steps 4 and 5 are very simple to pipeline. They involve only fetches, multiplies, and stores, so the pipelining of these parts of the algorithm is not discussed in detail here.

The Code

Now, writing the code is trivial. The ADSP-TS201 TigerSHARC processor is so flexible that it takes all the challenge right out of it. Just follow the pipelines of [Table 2](#) and [Table 4](#) and the code is done. The resulting code is shown in [Listing 1](#).

Now, for the bottom line—how much did the cycle count improve? [Table 5](#) lists cycle counts for the old and new implementations of the 16-bit complex input FFTs. As shown, the cycle counts have improved considerably.

<i>Points N</i>	<i>Old Implementation</i>	<i>New Implementation</i>
<i>64</i>	<i>302</i>	<i>147</i>
<i>256</i>	<i>886</i>	<i>585</i>
<i>1024</i>	<i>3758</i>	<i>2725</i>
<i>2048</i>	<i>7839</i>	<i>5776</i>
<i>4096</i>	<i>16600</i>	<i>12546</i>

Table 5. Core Clock Cycles for N-Point 16-bit Complex FFT

Usage Rules

The C-callable complex FFT routine is called as

```
fft256pt(&(input), &(ping_pong_buffer1),
        &(ping_pong_buffer2), &(output));
```

where:

input -> FFT input buffer,

output -> FFT output buffer,

ping_pong_bufferx are the ping pong buffers.

All buffers are packed complex values in normal (not bit-reversed) order.

ping_pong_buffer1 and *ping_pong_buffer2* must be two distinct buffers. However, depending on the routine's user requirements, some memory

optimization is possible. *ping_pong_buffer1* can be made the same as *input* if *input* does not need to be preserved. Also, *output* can be made the same as *ping_pong_buffer2*. Below is an example of the routine usage with minimal use of memory:

```
fft256pt (&(input), &( input),
          &( output), &(output));
```

To eliminate memory block access conflicts, *input* and *ping_pong_buffer1* must reside in a different memory block than *ping_pong_buffer2* and *output*, and the twiddle factors must reside in a different memory block than the ping-pong buffers. Of course, all code must reside in a block that is different from all the data buffers, as well.

Remarks

The example examined here is that of a 256-point FFT. At the time of writing this note, 64-point, 1024-point, 2048-point and 4096-point FFT examples using the algorithm described above have also been written. In those cases, the FFTs were viewed as 8x8, 32x32, 32x64, and 64x64 matrices, respectively. The 32-point FFTs were done in radix-4 (all the way to the last stage) and the last stage was done in the traditional radix-2.

The 2048-point FFT was arranged in a matrix of 32 columns and 64 rows. 32 FFTs of 64 points each are done in parallel on the columns. Applying a point-wise multiply and transpose gives a matrix of 64 columns and 32 rows. Doing 64 FFTs of 32 points each in parallel on the columns completes the algorithm. The only side effect is that the parallel FFT portion of the code cannot be re-used (remember, the algorithm needs it twice) because the number of rows and columns is no longer the same. This results in longer source code, but the cycle count efficiency is just as good.

Appendix

Complete Source Code of the Optimized FFT

```

/*****
fft256pt.asm
Prelim rev.      August 10, 2004   BL
This is assembly routine for the complex C-callable 256-point 16-bit FFT on
TigerSHARC family of DSPs.
I. Description of Calling.
1. Inputs:
   j4 -> input
   j5 -> ping_pong_buffer1
   j6 -> ping_pong_buffer2
   j7 -> output
2. C-Calling Example:
   Fft256pt(&(input), &(ping_pong_buffer1), &(ping_pong_buffer2), &(output));
3. Limitations:
   a. All buffers must be aligned on memory boundary which is a multiple of 4.
   b. Buffers input and ping_pong_buffer2 must be aligned on memory boundary
      which is a multiple of 256.
   c. If memory space savings are required and input does not have to be
      preserved, ping_pong_buffer1 can be the same buffer as input with no
      degradation in performance.
   d. If memory space savings are required, output can be the same buffer
      as ping_pong_buffer2 with no degradation in performance.
4. For the code to yield optimal performance, the following must be observed:
   a. Buffer input must have been cached previously. This is reasonable to
      assume since any engine that would have brought the data into internal
      memory, such as a DMA, would also have cached it.
   b. input and ping_pong_buffer2 must be located in different memory blocks.
   c. ping_pong_buffer1 and ping_pong_buffer2 must be located in different
      memory blocks.
   d. ping_pong_buffer1 and output must be located in different memory blocks.
   e. twiddles and input must be located in different memory blocks.
   f. AdjustMatrix and ping_pong_buffer1 must be located in different memory
      blocks.
II. Description of the FFT algorithm.
1. All data is treated as complex packed data.
2. An application note will be provided for the description of the math of
   the algorithm.
*****/

*****
//***** Includes *****/
#include <defTS201.h>

*****
.section data6a;
.align 4;                                     // align to quad
.var _AdjustMatrix[256] = "MatrixCoeffs.dat";

.align 4;                                     // align to quad
.var _twiddles16[32] = "Twiddles16.dat";

*****
.section program;
.global _fft256pt;

/***** Start of code *****/
_fft256pt:
j2=j4+64;;
j0=j4+0;          k1=j6;;
j3=j4+(128+64);  LC1=2;;
r5:4 =br Q[j2+=32];  k3=k31+(_twiddles16+2);;
j1=j4+128;;
                                           // -----
                                           // F1-- | | | | |
                                           // -----
.align_code 4;
_VeriticalLoop:

/***** Stage 1 *****/
// 1st time: From j0,j1,j2,j3-> input to k1-> ping_pong_buffer2
// 2nd time: From _ping_pong_buffer2 to _ping_pong_buffer1
r7:6 =br Q[j3+=32];  kL1=k31+252;;
r1:0 =br Q[j0+=32];  r31=0x80000000;;
r3:2 =br Q[j1+=32];  kB3=k31+_twiddles16;
r5:4 =br Q[j2+=32];  kB1=k1+4;;
                                           // -----
                                           // F2-- | | | | |
                                           // F3-- | | | | |
                                           // F4-- | | | | |
                                           // F1-- | | | | |
                                           // -----
r7:6 =br Q[j3+=32];  jL0=252;
r1:0 =br Q[j0+=32];  kL3=k31+32;
r3:2 =br Q[j1+=32];  LC0=4;
r5:4 =br Q[j2+=32];  jB0=kB1;
r24=mr1:0;
r25=mr1:0;
mr1:0+=r14**r31(C); sr9:8=r1:0+r3:2;
mr1:0+=r15**r31(C); sr19:18=r11:10+r25:24;
sr17:16=r9:8+r13:12; sr21:20=r9:8-r13:12;
sr11:10=r1:0-r3:2;
sr15:14=r5:4-r7:6;;
sr29:28=r5:4+r7:6;
sr17:16=r9:8+r13:12;
sr21:20=r9:8-r13:12;
                                           // F2- | MPY1-- | AS2-- |
                                           // F3- | MPY2-- | M1-- |
                                           // F4- | | | | |
                                           // F1- | | | | |
                                           // -----
.align_code 4;
_VerFFTStage1:
r7:6 =br Q[j3+=32];  cb Q[k1+=32]=r17:16;
r1:0 =br Q[j0+=32];  cb Q[k1+=-16]=r21:20;
r3:2 =br Q[j1+=32];  cb Q[k1+=32]=r19:18;
r24=mr1:0;
r25=mr1:0;
mr1:0+=r14**r31(C); sr9:8=r1:0+r3:2;
mr1:0+=r15**r31(C); sr19:18=r11:10+r25:24;
sr27:26=r1:0-r3:2;
sr23:22=r11:10-r25:24;;
sr17:16=r9:8+r13:12;
sr21:20=r9:8-r13:12;
sr11:10=r1:0-r3:2;
sr15:14=r5:4-r7:6;;
sr29:28=r5:4+r7:6;
sr17:16=r9:8+r13:12;
sr21:20=r9:8-r13:12;
                                           // F2 | MPY1- | AS2- | S1-- |
                                           // F3 | MPY2- | M1-- | AS4-- |
                                           // F4 | | | | |
                                           // F1 | | | | |
                                           // -----

```



```

r3:2=  Q[j9+=16];  r19:18=Q[k3+=16];  r27=mr1:0,  mr1:0+=r7**r23 (C);  // F2  MPY8-  M7-  TF2
r5:4=  Q[j9+=16];  r21:20=Q[k3+=16];  r31=mr1:0,  mr1:0+=r0**r16 (C);  // F3  MPY1  M8-  TF3
r7:6=  Q[j9+=16];  r23:22=Q[k3+=16];  r8=mr1:0,  mr1:0+=r1**r17 (C);  // F4  MPY2  M1  TF4
Q[j10+=16]=yr27:24;  k9=k8+64;  r12=mr1:0,  mr1:0+=r2**r18 (C);  // S1-  MPY3  M2
Q[j10+=16]=yr31:28;  j2=k9;  r9=mr1:0,  mr1:0+=r3**r19 (C);  // S2-  MPY4  M3
Q[j10+=16]=xr27:24;  k9=k8+(128+64);  r13=mr1:0,  mr1:0+=r4**r20 (C);  // S3-  MPY5  M4
Q[j10+=-44]=xr31:28;  j3=k9;  r10=mr1:0,  mr1:0+=r5**r21 (C);  // S4-  MPY6  M5
// -----
r1:0=  Q[j9+=16];  r17:16=Q[k3+=16];  r14=mr1:0,  mr1:0+=r6**r22 (C);  // F1+  MPY7  M6  TF1+
r3:2=  Q[j9+=16];  r19:18=Q[k3+=16];  r11=mr1:0,  mr1:0+=r7**r23 (C);  // F2+  MPY8  M7  TF2+
r5:4=  Q[j9+=16];  r21:20=Q[k3+=16];  r15=mr1:0,  mr1:0+=r0**r16 (C);  // F3+  MPY1+  M8  TF3+
r7:6=  Q[j9+=-236];  r23:22=Q[k3+=k2];  r24=mr1:0,  mr1:0+=r1**r17 (C);  // F4+  MPY2+  M1+  TF4+
Q[j10+=16]=yr11:8;  r28=mr1:0,  mr1:0+=r2**r18 (C);  // S1  MPY3+  M2+
Q[j10+=16]=yr15:12;  r25=mr1:0,  mr1:0+=r3**r19 (C);  // S2  MPY4+  M3+
Q[j10+=16]=xr11:8;  r29=mr1:0,  mr1:0+=r4**r20 (C);  // S3  MPY5+  M4+
Q[j10+=-44]=xr15:12;  r26=mr1:0,  mr1:0+=r5**r21 (C);  // S4  MPY6+  M5+
// -----
r1:0=  Q[j9+=16];  r17:16=Q[k3+=16];  r30=mr1:0,  mr1:0+=r6**r22 (C);  // F1++  MPY7+  M6+  TF1++
r3:2=  Q[j9+=16];  r19:18=Q[k3+=16];  r27=mr1:0,  mr1:0+=r7**r23 (C);  // F2++  MPY8+  M7+  TF2++
r5:4=  Q[j9+=16];  r21:20=Q[k3+=16];  r31=mr1:0,  mr1:0+=r0**r16 (C);  // F3++  MPY1++  M8+  TF3++
r7:6=  Q[j9+=16];  r23:22=Q[k3+=16];  r8=mr1:0,  mr1:0+=r1**r17 (C);  // F4++  MPY2++  M1++  TF4++
Q[j10+=16]=yr27:24;  r12=mr1:0,  mr1:0+=r2**r18 (C);  // S1+  MPY3++  M2++
Q[j10+=16]=yr31:28;  r9=mr1:0,  mr1:0+=r3**r19 (C);  // S2+  MPY4++  M3++
Q[j10+=16]=xr27:24;  r13=mr1:0,  mr1:0+=r4**r20 (C);  // S3+  MPY5++  M4++
// -----
.align_code 4;
if NLC0E, jump _MultXposeLoop;
Q[j10+=4]=xr31:28;  r10=mr1:0,  mr1:0+=r5**r21 (C);  // S4+  MPY6++  M5++
// -----
.align_code 4;
jump _VerticalLoop;
r5:4=br Q[j2+=32];  k3=k31+(_twiddles16+2);  // Repeat the vertical loop
// with swapped pointers

//***** Done *****
_fft256pt.end:

```

Listing 1. *fft256pt.asm*

References

[1] *ADSP-TS201 TigerSHARC Processor Programming Reference*. Revision 1.0, August 2004. Analog Devices, Inc.

Document History

Revision	Description
<i>Rev 1 – February 03, 2005 by Boris Lerner</i>	Initial Release