



\*159159\*

ARRAY SEARCHING OPERATIONSBACKGROUND

This invention relates to array searching operations  
5 for a computer.

Many conventional programmable processors, such as  
digital signal processors (DSP), support a rich instruction  
set that includes numerous instructions for manipulating  
arrays of data. These operations are typically  
10 computationally intensive and can require significant  
computing time, depending upon the number of execution  
units, such as multiply-accumulate units (MACs), within the  
processor.

15 DESCRIPTION OF DRAWINGS

Figure 1 is a block diagram illustrating an example of  
a pipelined programmable processor according to the  
invention.

Figure 2 is a block diagram illustrating an example  
20 execution pipeline for the programmable processor.

Figure 3 is a flowchart for implementing an example  
array manipulation machine instruction according to the  
invention.

Figure 4 is a flowchart of an example routine for  
25 invoking the machine instruction.

Figure 5 shows a search instruction; and

\*000002\*



Figure 6 shows N/M search instruction.

#### DESCRIPTION

Figure 1 is a block diagram illustrating a programmable processor 2 having an execution pipeline 4 and a control unit 6. Processor 2, as explained in detail below, reduces the computational time required by array manipulation operations. In particular, processor 2 may support a machine instruction, referred to herein as the SEARCH instruction, that reduces the computational time to search an array of numbers in a pipelined processing environment.

Pipeline 4 has a number of stages for processing instructions. Each stage processes concurrently with the other stages and passes results to the next stage in pipeline 4 at each clock cycle. The final results of each instruction emerge at the end of the pipeline in rapid succession.

Control unit 6 controls the flow of instructions and data through the various stages of pipeline 4. During the processing of an instruction, for example, control unit 6 directs the various components of the pipelined to fetch and decode the instruction, perform the corresponding operation and write the results back to memory or local registers.

Figure 2 illustrates an example pipeline 4 configured according to the invention. Pipeline 4, for example, has five stages: instruction fetch (IF), decode (DEC), address

calculation (AC), execute (EX) and write back (WB).

Instructions are fetched from memory, or from an instruction cache, during the IF stage by fetch unit 21 and decoded within address registers 22 during the DEC stage. At the  
5 next clock cycle, the results pass to the AC stage, where data address generators 23 calculate any memory addresses that are necessary to perform the operation.

During the EX stage, execution units 25A through 25M perform the specified operation such as, for example, adding  
10 or multiplying numbers, in parallel. Execution units 25 may contain specialized hardware for performing the operations including, for example, one or more arithmetic logic units (ALU's), floating-point units (FPU) and barrel shifters. A variety of data can be applied to execution units 25 such as  
15 the addresses generated by data address generator 23, data retrieved from data memory 18 or data retrieved from data registers 24. During the final stage (WB), the results are written back to data memory or to data registers 24.

The SEARCH instruction supported by processor 2, may  
20 allow software applications to search an array of N data elements by issuing N/M search instructions, where M is the number of data elements that can be processed in parallel by execution units 25 of pipeline 4. Note, however, that a single execution unit may be capable of executing two or  
25 more operations in parallel. For example, an execution unit

may include a 32-bit ALU capable of concurrently comparing two 16-bit numbers.

Generally, the sequence of SEARCH instructions allow the processor to process M sets of elements in parallel to identify an ``extreme value'', such as a maximum or a minimum, for each set. During the execution of the search instructions, processor 2 stores references to the location of the extreme value of each of the M sets of elements. Upon completion of the N/M instructions, as described in detail below, the software application analyzes the references to the extreme values for each set to quickly identify an extreme value for the array. For example, the instruction allows the software applications to quickly identify either the first or last occurrence of a maximum or minimum value. Furthermore, as explained in detail below, processor 2 implements the operation in a fashion suitable for vectorizing in a pipelined processor across the M execution units 25.

As described above, a software application searches an array of data by issuing N/M SEARCH machine instructions to processor 2. Figure 3 is a flowchart illustrating an example mode of operation 20 for processor 2 when it receives a single SEARCH machine instruction. Process 20 is described with reference to identifying the last occurrence of a minimum value within the array of elements; however, process 20 can be easily modified to perform other functions

such as identifying the first occurrence of a minimum value, the first occurrence of a maximum value or a last occurrence of a maximum value.

For exemplary purposes, process 20 is described in assuming M equals 2, i.e., processor 2 concurrently processes two sets of elements, each set having  $N/2$  elements. However, the process is not limited as such and is readily extensible to concurrently process more than two sets of elements. In general, process 20 facilitates vectorization of the search process by fetching pairs of elements as a single data quantity and processing the element pairs through pipeline 4 in parallel, thereby reducing the total number of clock cycles necessary to identify the minimum value within the array. Although applicable to other architectures, process 20 is well suited for a pipelined processor 2 having multiple execution units in the EX stage. For each set of elements, process 20 maintains two pointer registers,  $P_{\text{Even}}$  and  $P_{\text{Odd}}$ , that store locations for the current extreme value within the corresponding set. In addition, process 20 maintains two accumulators, A0 and A1, that hold the current extreme values for the sets. The pointer registers and the accumulators, however, may readily be implemented as general-purpose data registers without departing from process 30.

Referring to Figure 3, in response to each SEARCH instruction, processor 2 fetches a pair of elements in one

clock cycle as a single data quantity (21). For example, processor 2 may fetch two adjacent 16-bit values as one 32-bit quantity. Next, processor 2 compares the even element of the pair to a current minimum value for the even elements (22) and the odd element of the pair to a current minimum value for the odd elements (24).

When a new minimum value for the even elements is detected, processor 2 updates accumulator A0 to hold the new minimum value and updates a pointer register  $P_{\text{Even}}$  to hold a pointer to corresponding data quantity within the array (23). Similarly, when a new minimum value for the odd elements has been detected, processor 2 updates accumulator A1 and a pointer register  $P_{\text{Odd}}$  (25). In this example, each pointer register  $P_{\text{Even}}$  and  $P_{\text{Odd}}$  points to the data quantity and not the individual elements, although the process is not limited as such. Processor 2 repeats the process until all of the elements within the array have been processed (26). Because processor 2 is pipelined, element pairs may be fetched until the array is processed.

The following illustrates exemplary syntax for invoking the machine instruction:

$$(P_{\text{Odd}}, P_{\text{Even}}) = \text{SEARCH } R_{\text{Data}} \text{ LE}, R_{\text{Data}} = [P_{\text{fetch\_addr}}++]$$

Data register  $R_{\text{Data}}$  is used as a scratch register to store each newly fetched data element pair, with the least significant word of  $R_{\text{Data}}$  holding the odd element and the

most significant word of  $R_{Data}$  holding the even element. Two accumulators, A0 and A1, are implicitly used to store the actual values of the results. An additional register,  $P_{fetch\_addr}$ , is incremented when the SEARCH instruction is  
5 issued and is used as a pointer to iterate over the  $N/2$  data quantities within the array. The defined condition, such as ``less than or equal'' (LE) in the above example, controls which comparison is executed and when the pointer registers  $P_{Even}$  and  $P_{Odd}$ , as well as the accumulators A0 and A1, are  
10 updated. The ``LE'', for example, directs processor 2 to identify the last occurrence of the minimum value.

In a typical application, a programmer develops a software application or subroutine that issues the  $N/M$  search instructions, probably from within a loop construct.  
15 The programmer may write the software application in assembly language or in a high-level software language. A compiler is typically invoked to process the high-level software application and generate the appropriate machine instructions for processor 2, including the SEARCH machine  
20 instructions for searching the array of data.

Figure 4 is a flowchart of an example software routine  
30 for invoking the example machine instructions illustrated above. First, the software routine 30 initializes the registers including initializing A0 and A1 and pointing  $P_{Even}$   
25 and  $P_{Odd}$  to the first data quantity within the array (31). In one embodiment, software routine 30 initializes a loop

count register with the number of SEARCH instructions to issue (N/M). Next, routine 30 issues the SEARCH machine instruction N/M times. This can be accomplished a number of ways, such as by invoking a hardware loop construct supported by processor 2. Often, however, a compiler may unroll a software loop into a sequence of identical SEARCH instructions (32).

After issuing N/M search instructions, A0 and A1 hold the last occurrence of the minimum even value and the last occurrence of the minimum odd value, respectively. Furthermore,  $P_{\text{Even}}$  and  $P_{\text{Odd}}$  store the locations of the two data quantities that hold the last occurrence of the minimum even value and the last occurrence of the minimum odd value.

Next, in order to identify the last occurrence of the minimum value for the entire array, routine 30 first increments  $P_{\text{Odd}}$  by a single element, such that  $P_{\text{Odd}}$  points directly at the minimum odd element (33). Routine 30 compares the accumulators A0 and A1 to determine whether the accumulators contain the same value, i.e., whether the minimum of the odd elements equals the minimum of the even elements (34). If so, the routine 30 compares the pointers to determine whether  $P_{\text{Odd}}$  is less than  $P_{\text{Even}}$  and, therefore,  $P_{\text{Odd}}$  and  $P_{\text{Even}}$  whether the minimum even value occurred earlier in the array (35). Based on the comparison, the routine determines whether to copy  $P_{\text{Odd}}$  into  $P_{\text{Even}}$  (37).

When the accumulators A0 and A1 are not the same, the routine compares A0 to A1 in order to determine which holds the minimum value (36). If A1 is less than A0 then routine 30 sets  $P_{\text{Even}}$  equal to  $P_{\text{Odd}}$ , thereby copying the pointer to the  
5 minimum value from  $P_{\text{Odd}}$  into  $P_{\text{Even}}$  (37).

At this point,  $P_{\text{Even}}$  points to the last occurrence of the minimum value for the entire array. Next, routine 30 adjusts  $P_{\text{Even}}$  to compensate for errors introduced to the pipelined architecture of processor 2 (38). For example,  
10 the comparisons described above are typically performed in the EX stage of pipeline 4 while incrementing the pointer register  $P_{\text{fetch\_addr}}$  typically occurs during the AC stage, thereby causing the  $P_{\text{Odd}}$  and  $P_{\text{Even}}$  to be incorrect by a known quantity. After adjusting  $P_{\text{Even}}$ , routine 30 returns  $P_{\text{Even}}$  as a  
15 pointer to the last occurrence of the minimum value within the array(39).

Figure 5 illustrates the operation for a single SEARCH instruction as generalized to the case where processor 2 is capable of processing M elements of the array in parallel,  
20 such as when processor 2 includes M execution units. The SEARCH instruction causes processor 2 to fetch M elements in a single fetch cycle (51). Furthermore, in this example, processor 2 maintains M pointer registers to store addresses (locations) of a corresponding extreme value for each of the  
25 M sets of elements. After fetching the M elements, processor 2 concurrently compares the M elements to a

current extreme value for the respective element set, as stored in M accumulators (52). Based on the comparisons, processor 2 updates the M accumulators and the M pointer registers (53).

5        Figure 6 illustrates the general case where a software application issues N/M SEARCH instructions and, upon completion of the instructions, determines the extreme value for the entire array. First, the software application initializes a loop counter, the M accumulators used to store  
10 the current extreme values for the M element sets and the M pointers used to store the locations of the extreme values (61). Next, the software application issues N/M SEARCH instructions (62). After completion of the instructions, the software application may adjust the M pointer registers  
15 to correctly reference its respective extreme value, instead of the data quantity holding the extreme value (63). After adjusting the pointer registers, the software application compares the M extreme values for the M element sets to identify an extreme value for the entire array, i.e., a  
20 maximum value or a minimum value (64). Then, the software application may use the pointer registers to determine whether more than one of the element sets have an extreme value equal to the array extreme value and, if so, determine which extreme value occurred first, or last, depending upon  
25 the desired search function (65).

Various embodiments of the invention have been described. For example, a single machine instruction has been described that searches an array of data in a manner that facilitates vectorization of the search process within  
5 a pipelined processor. The processor may be implemented in a variety of systems including general purpose computing systems, digital processing systems, laptop computers, personal digital assistants (PDA's) and cellular phones. For example, cellular phones often maintain an array of  
10 values representing signal strength for services available 360° around the phone. In this context, the process discussed above can be readily used upon initialization of the cellular phone to scan the available services and quickly select the best service. In such a system, the  
15 processor may be coupled to a memory device, such as a FLASH memory device or a static random access memory (SRAM), that stores an operating system and other software applications. These and other embodiments are within the scope of the following claims.

What is claimed is:

- 1 1. A method comprising:
  - 2 receiving a machine instruction directing a processor
  - 3 to search a plurality of data elements; and
  - 4 executing the machine instruction by:
    - 5 retrieving M data elements in a single fetch
    - 6 cycle;
    - 7 concurrently comparing the M data elements to a
    - 8 corresponding current extreme value; and
    - 9 updating a set of references based on the
    - 10 comparisons.
- 1 2. The method of claim 1, wherein retrieving the M data
  - 2 elements comprises retrieving the M data elements as a
  - 3 single data quantity containing the M data elements.
- 1 3. The method of claim 2, wherein the set of references
  - 2 comprise pointer registers to store addresses for data
  - 3 quantities.
- 1 4. The method of claim 1, wherein  $M = 1$ .
- 1 5. The method of claim 1, wherein  $M = 2$ .
- 1 6. The method of claim 1, wherein executing the machine 1
  - 2 instruction further includes:

3       storing the current extreme values in M accumulators;  
4    and  
5       copying the M data elements to the accumulators based  
6    on the comparisons.

1    7.    The method of claim 5, wherein concurrently comparing  
2    the data elements comprises processing a first data element  
3    with a first execution unit of a pipelined processor and  
4    processing a second data element with a second execution  
5    unit of the pipelined processor.

1    8.    The method of claim 5, wherein concurrently comparing  
2    the data elements comprises concurrently processing a first  
3    data element and a second data element within a single  
4    execution unit of a pipelined processor.

1    9.    The method of claim 1, wherein concurrently comparing  
2    each of the data elements to a current extreme value  
3    includes determining whether each of the data elements is  
4    less than the corresponding current extreme value.

1    10.   The method of claim 1, wherein concurrently comparing  
2    each of the data elements to a current extreme value  
3    includes determining whether each of the data elements is  
4    greater than the corresponding current extreme value.

1    11.   A method for searching an array of N data elements for  
2    a value comprising:

3       issuing N/M machine instructions to a processor,  
4       wherein the processor is adapted to process M data elements  
5       in parallel; and

6       analyzing results of the machine instructions to  
7       identify a value for the array.

1   12.   The method of claim 11, further comprising:

2       executing each machine instruction by:

3       retrieving M data elements in a single fetch cycle,

4       concurrently comparing each of the M data elements to a  
5       corresponding current extreme value, and

6       updating the references based on the comparisons.

1   13.   A method comprising:

2       retrieving the pair of data elements from an array of  
3       elements in a single fetch operation, wherein the pair of  
4       data elements includes an even data element and an odd data  
5       element;

6       substantially comparing the even element of the pair  
7       and the odd element of the pair; and

8       substantially fetching and comparing the remaining  
9       pairs of data elements of the array until all of the data  
10       elements of the array have been processed.

1   14.   The method of claim 13, wherein substantially comparing

2       the pair of data elements includes setting an even minimum

3       value as function of the even element of the element pair

4 and setting an odd minimum value as function of the odd  
5 element of the element pair.

1 15. The method of claim 13, wherein substantially comparing  
2 the pair of data elements includes maintaining a first  
3 accumulator to store a minimum value for the even elements  
4 and a second accumulator to store a minimum value for the  
5 odd elements.

1 16. The method of claim 13, further including maintaining a  
2 first pointer register to store an address for the minimum  
3 value of the even data elements and maintaining a second  
4 pointer register to store an address for the minimum value  
5 of the odd data elements.

1 17. The method of claim 16, further including adjusting at  
2 least one of the pointer registers after processing all of  
3 the pairs of data elements to account for a number of stages  
4 in the pipeline.

1 18. The method of claim 13, wherein the method is invoked  
2 by issuing N/M machine instructions to a programmable  
3 processor, wherein N equals the number of elements in the  
4 array and M equals the number of data elements that the  
5 processor can concurrently compare.

- 1 19. An apparatus comprising:  
2 a pipeline adapted to process M data elements in  
3 parallel; and  
4 a control unit adapted to direct the execution pipeline  
5 to search an array of N data elements in response to N/M  
6 machine instructions.
- 1 20. The apparatus of claim 19, wherein in response to the  
2 machine instructions, the control unit directs the pipeline  
3 to retrieve M data elements from the array of elements in a  
4 single fetch operation and concurrently compare the data  
5 elements to a corresponding current extreme value.
- 1 21. The apparatus of claim 19, wherein the pipeline  
2 includes M registers adapted to store references to the  
3 extreme values.
- 1 22. The apparatus of claim 21, wherein the registers are  
2 pointer registers.
- 1 23. The apparatus of claim 21, wherein the registers are  
2 general-purpose data registers.
- 1 24. The apparatus of claim 18, wherein the pipeline  
2 includes M accumulators to store M current extreme values.
- 1 25. The apparatus of claim 18, wherein the pipeline  
2 includes M general-purpose registers to store M current  
3 extreme values.

1 26. An article comprising a medium having computer-  
2 executable instructions stored thereon for compiling a  
3 software program, wherein the computer-executable  
4 instructions are adapted to generate N/M machine  
5 instructions to search an array of N data elements, each  
6 machine instruction causing a programmable processor to:  
7       retrieve M data elements from an array of N elements in  
8 a single fetch operation; and  
9       substantially compare each of the M data elements to a  
10 corresponding current extreme value.

1 27. The article of claim 26, wherein each machine  
2 instruction causes the processor to update a set of  
3 references based on the comparisons.

1 28. The article of claim 26, wherein each machine  
2 instruction causes the processor to concurrently process a  
3 first data element and a second data element within a single  
4 execution unit of a pipelined processor.

1 29. A system comprising:  
2       a memory device; and  
3       a processor coupled to the memory device, wherein the  
4 processor includes a pipeline configured to process M data  
5 elements in parallel and a control unit configured to

6 direct the pipeline to search an array of N data elements  
7 in response to N/M machine instructions.

1

1 30. The system of claim 29, wherein in response to each  
2 machine instructions, the control unit directs the pipeline  
3 to retrieve M data elements from the array of elements in a  
4 single fetch operation and concurrently compare the data  
5 elements to a corresponding current extreme value.

1

1 31. The system of claim 29, wherein the pipeline includes  
2 M registers configured to store references to the extreme  
3 values.

1

1 32. The system of claim 31, wherein the registers are  
2 pointer registers.

1

1 33. The system of claim 31, wherein the registers are  
2 general-purpose data registers.

1

1 34. The system of claim 29, wherein the memory device  
2 comprises static random access memory.

1

2 35. The system of claim 29, wherein the memory device  
3 comprises FLASH memory.