

METHOD AND APPARATUS USING NOVEL OPERATIONS IN A PROCESSOR

BACKGROUND OF THE INVENTION

1. FIELD OF INVENTION

5 The present invention relates to the field of processor operation. In particular, the present invention describes an apparatus and method of performing a population count operation on packed data.

2. DESCRIPTION OF RELATED ART

10 Today, most personal computer systems operate with one instruction to produce one result. Performance increases are achieved by increasing execution speed of instructions and the processor instruction complexity; known as Complex Instruction Set Computer (CISC). Such processors as the Intel 80486™ microprocessor, available from Intel Corp. of Santa Clara, California, belong to the CISC category of processor.

15 Previous computer system architecture has been optimized to take advantage of the CISC concept. Such systems typically have data buses thirty-two bits wide. However, applications targeted at computer supported cooperation (CSC - the integration of teleconferencing with mixed media data manipulation), 2D/3D graphics, image processing, video compression/decompression, recognition
20 algorithms and audio manipulation increase the need for improved performance. But, increasing the execution speed and complexity of instructions is only one solution.

One common aspect of these applications is that they often manipulate large amounts of data where only a few bits are important. That is, data whose relevant bits are represented in much fewer bits than the size of the data bus. For example, processors execute many operations on eight bit and sixteen bit data (e.g., pixel
5 color components in a video image). Thus, a processor having a thirty-two bit data bus and registers, and executing one of these algorithms, can waste up to seventy-five percent of its data processing, carrying and storage capacity because only the first eight bits of data are important.

10 As such, what is desired is a processor that increases performance by more efficiently using the difference between the number of bits required to represent the data to be manipulated and the actual data carrying and storage capacity of the processor.

SUMMARY OF THE INVENTION

An apparatus and method of increasing processor performance by efficiently reducing the difference between data element length and actual carrying and storage capacity is described.

5 In one embodiment of the present invention, a processor manipulates packed data. The packed data includes a first data element D1 and a second data element D2. Each of said data elements has a predetermined number of bits. The processor comprises a decoder, a register, and a circuit. The decoder is for decoding a control signal responsive to receiving the control signal. The register is coupled to the
10 decoder. The register is for storing the packed data. The circuit is coupled to the decoder. The circuit is for generating a first result data element R1 and a second data element R2. The circuit is further for generating R1 to represent a total number bits set in D1, and the circuit is further for generating R2 to represent a total number bits set in D2.

15 In a second embodiment of the present invention, the above processor includes a second register, and the second register is for storing R1 and R2 as a result packed data.

 In another embodiment of the present invention, the above processor includes a first circuit that is coupled to the decoder and a second circuit that is coupled to the
20 decoder. The first circuit is for totaling a number of bits set in D1, and the second circuit is for totaling a number of bits set in D2.

In another embodiment of the present invention, the processor of the second embodiment wherein the packed data further includes a third data element D3 and a fourth data element D4. The circuit is further for generating a third result data element R3 and a fourth data element R4. The circuit is further for generating R3 to
5 represent a total number of bits set in D3, and for generating R4 to represent a total number of bits set in D4. The second register is for storing R1, R2, R3 and R4 as a result packed data.

In another embodiment of the present invention, the processor of the first embodiment wherein each of the data elements includes sixteen bits.

10 In another embodiment of the present invention, the processor of the first embodiment wherein a set bit equals one.

In another embodiment of the present invention, the processor of the first embodiment wherein the processor supports operations supported by an Intel 80486™ microprocessor.

15 In another embodiment of the present invention, the processor of the first embodiment wherein the processor supports operations supported by an Intel Pentium™ microprocessor.

In another embodiment of the present invention, the processor of the first embodiment wherein the processor supports operations defined by the IA™, Intel
20 Architecture.

In another embodiment of the present invention, the processor of the first embodiment wherein the processor includes a cache coupled to the decoder, and the cache is for storing data.

In another embodiment of the present invention, the processor of the first embodiment wherein the processor is further for storing integer data.

In another embodiment of the present invention, a computer system is for manipulating packed data. The packed data includes a plurality of data elements,
5 wherein each of the plurality of data elements has a predetermined number of bits. The computer system comprises a memory and a processor being coupled to the memory. The memory is for storing data and a control signal. The processor includes a decoder, a plurality of registers, and a circuit for performing an operation on the packed data. The decoder is for decoding the control signal. The plurality of
10 registers is for providing and storing the packed data. The plurality of registers is coupled to the decoder. The plurality of registers has a first register. The circuit is coupled to the decoder. Responsive to decoding the control signal, the decoder enables the first register to provide a first packed data. The first packed data has a first data element D1 and a second data element D2. The circuit generates a result
15 data element R1 equal to a total number of bits set in D1. The circuit generates a result data element R2 equal to a total number of bits set in D2. The circuit provides a result packed data to the plurality of registers. The result packed data includes R1 and R2. The plurality of registers stores the result packed data in one of the plurality of registers.

20 In another embodiment of the present invention, a method is for manipulating packed data in a computer system. The computer system includes a memory being coupled to a microprocessor. The microprocessor includes a decoder for decoding a control signal, a plurality of registers for storing the packed data, and a circuit for

processing the packed data. The packed data includes a first data element D1 and a second data element D2. Each of the data elements has a predetermined number of bits. The decoder is coupled to the plurality of registers and the circuit. The method the following steps. Responsive to receiving a control signal, decoding the control
5 signal. Providing the packed data to the circuit. Generating a first result data element representative of a number of bits set in D1. Generating a second result data element representative of a number of bits set in D2. Storing the first result data element and the second result data element as a result packed data.

In another embodiment of the present invention, a processor is for performing a
10 population count on a packed data. The packed data has four data elements, D1, D2, D3 and D4 respectively. Each of the data elements includes sixteen bits. The processor comprises a decoder, a plurality of registers, and an operation circuit. The decoder, responsive to receiving a control signal, decodes the control signal. The plurality of registers is for storing packed data. The plurality of registers couples to
15 the decoder. The plurality of registers has a first register. The operation circuit is for performing an operation on the packed data. The operation circuit is coupled to the decoder. Responsive to decoding the control signal, the decoder enables the first register to provide the first packed data to the operation circuit. The operation circuit includes a first, second, third, and fourth circuit. The first circuit is for
20 generating a result data element R1 equal to a total number of bits set in D1. The second circuit is for generating a second result data element R2 equal to a total number of bits set in D2. The third circuit is for generating a third result data element R3 equal to a total number of bits set in D3. The fourth circuit is for

generating a fourth result data element R4 equal to a total number of bits set in D4. The operation circuit provides the plurality of registers with a result packed data. The result packed data has the R1, R2, R3, and R4. The plurality of registers stores the result packed data in one of the plurality of registers.

- 5 In another embodiment of the present invention, method is for performing a population count operation on a packed data in a processor. The processor includes a decoder for decoding a control signal, a plurality of registers for storing the packed data, and a circuit for processing the packed data. The decoder is coupled to the plurality of registers and the circuit. The packed data includes four data elements
- 10 D1, D2, D3, and D4, respectively. The method comprises the following steps. Generating a first result data representative of a total number of bits equal to one in D1. Generating a second result data representative of a total number of bits equal to one in D2. Generating a third result data representative of a total number of bits equal to one in D3. Generating a fourth result data representative of a total number
- 15 of bits equal to one in D4. Storing D1, D2, D3, and D4, as a result packed data.

Thus, in a single operation, multiple data elements may be operated upon given a single packed data.

An object of the present invention is to produce a novel processor better suited to multimedia applications.

- 20 An object of the present invention is to improve the execution of multimedia algorithms by more efficiently using the difference between the length of data element manipulated and the actual data carrying and storage capacity of the processor.

An object of the present invention is to allow one operation to perform work on multiple data elements.

An object of the present invention is to improve the performance algorithms requiring the totaling of the number of bits set in a eight bit, sixteen bit, and thirty-
5 two bit operations in a processor having sixty-four bit registers.

An object of the present invention is to support population count operations on eight bit, sixteen bit, and thirty-two bit data elements in a processor having sixty-four bit registers.

Although a great deal of detail has been included in the description and figures,
10 our invention is defined by the scope of the claims. Only limitations found in those claims apply to our invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example, and not limitation, in the figures. Like references indicate similar elements.

Figure 1 illustrates an embodiment of the computer system using the methods
5 and apparatus of the present invention.

Figure 2 illustrates an embodiment of the processor of the present invention.

Figure 3 is a flow diagram illustrating the general steps used by the processor to manipulate data in the register file.

Figure 4a illustrates memory data types.

10 Figure 4b through Figure 4d illustrate in-register integer data representations.

Figure 5a illustrates packed data-types.

Figure 5b through Figure 5d illustrate in-register packed data representations.

Figure 6 illustrates the control signal format used in the computer system to indicate the use of packed data.

15 Figure 7 illustrates the steps followed by a processor when performing a population count operation on packed data.

Figure 8 illustrates the steps followed when performing a population count operation on one data element of a packed data and generating a single result data element for a result packed data.

20 Figure 9 illustrates one circuit for performing a population count operation on packed data having four word data elements.

Figure 10 illustrates a detailed circuit for performing a population count operation on one word data element of a packed data.

DESCRIPTION OF THE PREFERRED EMBODIMENT

DEFINITIONS

To provide a foundation for understanding the description of the embodiments of the present invention, the following definitions are provided.

- 5 Bit X through Bit Y:
 defines a subfield of binary number. For example, bit six
 through bit zero of the byte 00111010₂ (shown in base
 two) represent the subfield 111010₂. The '2' following a
10 binary number indicates base 2. Therefore, 1000₂ equals
 8₁₀, while F16 equals 15₁₀.
- R_x: is a register. A register is any device capable of storing
 and providing data. Further functionality of a register is
 described below.
- 15 DEST: is an address of a register.
 SRC1: is an address of a register.
 SRC2: is an address of a register.
- Result: is the data to be stored in the register addressed by DEST.
 Source1: is the data stored in the register addressed by SRC1.
20 Source2: is the data stored in the register addressed by SRC2.

COMPUTER SYSTEM

An apparatus and method of increasing processor performance by efficiently
reducing the difference between data element length and actual carrying and storage
capacity is described. In the following description, numerous specific details are set
25 forth such as packed data formats, control signal formats, etc., in order to provide a
thorough understanding of the present invention. It will be obvious, however, to one
skilled in the art that the present invention may be practiced without these specific

details. In other instances, well-known processor architecture, circuits, structures and techniques have not been shown in detail in order not to unnecessarily obscure the present invention.

Figure 1 illustrates a computer system 100 upon which one embodiment of the present invention can be implemented. Computer system 100 comprises a bus or other communication means 101 for communicating information, and a processing means 109 coupled with bus 101 for processing information. System 100 further comprises a random access memory (RAM) or other dynamic storage device 104 (referred to as main memory), coupled to bus 101 for storing information and instructions to be executed by processor 109. Main memory 104 also may be used for storing temporary variables or other intermediate information during execution of instructions by processor 109. Computer system 100 also comprises a read only memory (ROM) and/or other static storage device 106 coupled to bus 101 for storing static data and instructions for processor 109. Data storage device 107 is coupled to bus 101 for storing information and instructions.

Furthermore, a data storage device 107 such as a magnetic disk or optical disk and its corresponding disk drive can be coupled to computer system 100. Computer system 100 can also be coupled via bus 101 to a display device 121, such as a cathode ray tube (CRT), for displaying information to a computer user. An alphanumeric input device 122, including alphanumeric and other keys, is typically coupled to bus 101 for communicating information and command selections to processor 109. Another type of user input device is cursor control 123, such as a mouse, a trackball, or cursor direction keys for communicating direction information

and command selections to processor 109 and for controlling cursor movement on display 121. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), which allows the device to specify positions in a plane.

5 Another device that may be coupled to bus 101 is hard copy device 124 that may be used for printing instructions, data, or other information on a medium such as paper, film, or similar types of media. Additionally, computer system 100 can be coupled to a device for sound recording and/or playback 125 such as an audio digitizer coupled to a microphone for recording information. Further, the device
10 may include a speaker that is coupled to a digital to analog (D/A) converter for playing back the digitized sounds. Finally, computer system 100 can be a terminal in a computer network (e.g., a LAN).

Computer system 100 is useful for supporting computer supported conferencing (CSC - the integration of teleconferencing with mixed media data manipulation),
15 2D/3D graphics, image processing, video compression/decompression, recognition algorithms and audio manipulation.

PROCESSOR

Figure 2 illustrates a detailed diagram of processor 109. One skilled in the art would understand that processor 109 could be implemented on one or more
20 substrates using any of a number of process technologies, such as, BiCMOS, CMOS, and NMOS.

Processor 109 comprises a decoder 202 for decoding control signals and data used by processor 109. Data can then be stored in register file 204 via internal bus

205. As a matter of clarity, the registers of the preferred embodiment should not be limited in meaning to a particular type of circuit. Rather, a register of the preferred embodiment need only be capable of storing and providing data, and performing the functions described herein.

5 Depending on the type of data, the data may be stored in integer registers 209, status registers 208, or instruction pointer register 211. Integer registers 209 contains thirty-two integer registers, R_1 212a through R_n 212af. Each register is sixty-four bits in length. R_1 212a, R_2 212b and R_3 212c are examples of individual registers in integer registers 209. Status registers 208 indicate the status of processor
10 109. Instruction pointer 211 stores the address of the next instruction to be executed. Integer registers 209, status registers 208, and instruction pointer register 211 all connect to internal bus 205.

Arithmetic logic unit 203 (ALU) performs the arithmetic and logic operations carried out by processor 109. Such operations may include logical shifts, addition,
15 subtraction and multiplication, etc. ALU 203 connects to internal bus 205. Cache 206 is an optional element of processor 109 and can be used to cache data, including control signals, from, for example, main memory 104. Cache 206 is connected to decoder 202, and is connected to receive control signal 207.

Figure 3 illustrates the general operation of processor 109. That is, Figure 3
20 illustrates the steps followed by processor 109 while performing an operation on packed data, performing an operation on unpacked data, or performing some other operation. For example, such operations include a load operation to load a register in register file 204 with data from cache 206, main memory 104, static memory 106,

or mass storage device 107. In one embodiment of the present invention, processor 109 supports most of the instructions supported by the Intel 80486™, available from Intel Corporation of Santa Clara, California. In another embodiment of the present invention, processor 109 supports all the operations supported by the Intel 80486™, available from Intel Corporation of Santa Clara, California. In another embodiment of the present invention, processor 109 supports all the operations supported by the Pentium™, the Intel 80486™, the 80386™, the Intel 80286™, and the Intel 8086™, all available from Intel Corporation of Santa Clara, California. In another embodiment of the present invention, processor 109 supports all the operations supported in the IA™ - Intel Architecture, as defined by Intel Corporation of Santa Clara, California (see Microprocessors, Intel Data Books volume 1 and volume 2, 1992 and 1993, available from Intel of Santa Clara, California).

At block 301, the decoder 202 receives a control signal 207 from either the cache 206 or bus 101. Decoder 202 decodes the control signal to determine the operations to be performed.

Decoder 202 accesses the register file 204 at block 302. Registers in the register file 204 are accessed depending on the register address specified in the control signal 207. For example, for an operation on packed data, control signal 207 includes SRC1, SRC2 and DEST register addresses. SRC1 is the address of the first source register. SRC2 is the address of the second source register. In some cases, the SRC2 address is optional as not all operations require two source addresses. If the SRC2 address is not required for an operation, then only the SRC1 address is used. DEST is the address of the destination register where the result data is stored.

SRC1, SRC2 and DEST are described more fully in relation to Figure 6. The data stored in these registers is referred to as Source1, Source2, and Result respectively. Each of these data is sixty-four bits in length.

Where the control signal requires an arithmetic or logic operation, at step 303,
5 ALU 203 will be enabled to perform this operation on accessed data from register file 204. Once the arithmetic or logic operation has been performed in ALU 203, at step 304, the result is stored back into register file 204 according to requirements of control signal 207.

DATA AND STORAGE FORMATS

10 Figure 4a illustrates some of the data formats as may be used in the computer system of figure 1. Processor 109 can manipulate these data formats. Multimedia algorithms often use these data formats. A byte 401 contains eight bits of information. A word 402 contains sixteen bits of information, or two bytes. A doubleword 403 contains thirty-two bits of information, or four bytes. Thus,
15 processor 109 executes control signals that may operate on any one of these memory data formats.

In the following description, references to bit, byte, word, and doubleword subfields are made. For example, bit six through bit zero of the byte 00111010₂ (shown in base 2) represent the subfield 111010₂.

20 Figure 4b through Figure 4d illustrate in-register integer data representations used in one embodiment of the present invention. For example, unsigned byte 410 can represent data stored in integer register R₁ 212a. A register, in integer registers 209, is sixty-four bits in length.

In-register unsigned byte representation 410 illustrates processor 109 storing a byte 401 in integer registers 209, the first eight bits, bit seven through bit zero, in that integer register are dedicated to the data byte 401. These bits are shown as {b}. To properly represent this byte, the remaining 56 bits must be zero.. For a signed
5 byte 411, integer registers 209 store the data in the first seven bits, bit six through bit zero, to be data. The seventh bit represents the sign bit, shown as an {s}. The remaining bit sixty-three through bit eight are the continuation of the sign for the byte.

Unsigned word in-register data representation 412 is stored in one register of
10 integer registers 209. Bit fifteen through bit zero contain the unsigned word 402. These bits are shown as {w}. To properly represent this word, the remaining bit sixty-three through bit sixteen must be zero. Signed word 402 is stored in bit fourteen through bit zero as shown in the signed word in-register data representation 413. The remaining bit sixty-three through bit fifteen is the sign field.

15 A doubleword 403 can be stored as an unsigned doubleword or a signed doubled word 414 or 415. Bit thirty-one through bit zero of an unsigned doubleword 414 are the data. These bits are shown as {d}. To properly represent this unsigned doubleword, the remaining bit sixty-three through bit thirty-two must be zero. Integer register 209 stores a signed doubleword 415 in its bit thirty through
20 bit zero; the remaining bit sixty-three through bit thirty-one are the sign field.

As indicated by the above Figure 4b through Figure 4d, storage of some data types in a sixty-four bit wide register is an inefficient method of storage. For example, for storage of a byte in-register data representation 410 bit sixty-three

through bit eight must be zero, while only bit seven through bit zero may contain non-zero bits. Thus, a processor storing a byte in an integer register uses only 12.5% of its capacity. Similarly, only the first few bits of operations performed by ALU 203 will be important.

5 Figure 5a illustrates the data formats for packed data. Three packed data formats are illustrated; packed byte 501, packed word 502, and packed doubleword 503. Packed byte, in the preferred embodiment of the present invention, is sixty-four bits long containing eight data elements. Each data element is one byte long. Generally, a data element is an individual piece of data that is stored in a single
10 integer register with other data elements of the same length. In the preferred embodiment of the present invention, the number of data elements stored in an integer register is sixty-four bits divided by the length in bits of a data element.

 Packed word data 502 is sixty-four bits long and contains four word data elements 402. Each word data element 402 contains sixteen bits of information.

15 Packed doubleword data 503 is sixty-four bits long and contains two doubleword 403 data elements. Each doubleword data element 403 contains thirty-two bits of information.

 Figure 5b through Figure 5d illustrate the in-register packed data storage representation. Unsigned packed byte in-register representation 510 illustrates the
20 storage of packed data type 501 in one of the integer registers R₀ 212a through R_N 212af. Information for each byte data element is stored in bit seven through bit zero for byte zero, bit fifteen through bit eight for byte one, bit twenty-three through bit sixteen for byte two, bit thirty-one through bit twenty-four for byte three, bit thirty-

nine through bit thirty-two for byte four, bit forty-seven through bit forty for byte five, bit fifty-five through bit forty-eight for byte six and bit sixty-three through bit fifty-six for byte seven. Thus, all available bits are used in the integer register. This storage arrangement increases the storage efficiency of the processor. As well, with
5 eight data elements accessed, one operation can now be performed on eight data elements simultaneously. Signed packed byte in-register representation 511 is similarly stored in an integer register in integer registers 209. Note that only the eighth bit of every byte data element is the sign bit; no other bits are used to indicate sign.

10 Unsigned packed word in-register representation 512 illustrates how word three through word zero are stored in one integer register of integer registers 209. Bit fifteen through bit zero contain the data element information for word zero, bit thirty-one through bit sixteen contain the information for data element word one, bit forty-seven through bit thirty-two contain the information for data element word two
15 and bit sixty-three through bit forty-eight contain the information for data element word three. Signed packed word in-register representation 513 is similar to the unsigned packed word in-register representation 512. Note that only the sixteenth bit of each word data element contains the sign indicator.

 Unsigned packed doubleword in-register representation 514 shows how integer
20 registers 209 store two doubleword data elements. Doubleword zero is stored in bit thirty-one through bit zero of the integer register. Doubleword one is stored in bit sixty-three through bit thirty-two of the integer register. Signed packed doubleword in-register representation 515 is similar to unsigned packed doubleword in-register

representation 514. Note that the signed bit is the thirty-second bit of the doubleword data element.

In one embodiment of the present invention, the individual programming processor 109 must track whether an addressed register, R₁ 212a for example, is
5 storing packed data or simple integer data. One skilled in the art would understand that in an alternative embodiment, processor 109 could track the type of data stored in individual registers of integer registers 209. This alternative embodiment could then generate errors if, for example, a packed addition operation were attempted on simple integer data.

10

CONTROL SIGNAL FORMATS

The following describes the control signal formats used by processor 109 to manipulate packed data. In one embodiment of the present invention, control signals are represented as thirty-two bits. Decoder 202 may receive control signal 207 from bus 101. However, it is obvious to one skilled in the art that decoder 202
15 can also receive such control signals from cache 206.

Figure 6 illustrates the general format for a control signal operating on packed data. Operation field OP 601, bit thirty-one through bit twenty-six, provides information about the operation to be performed by processor 109; for example, packed addition, packed subtraction, etc.. SRC1 602, bit twenty-five through
20 twenty, provides the source register address of a register in integer registers 209. This source register contains the first packed data, Source1, to be used in the execution of the control signal. Similarly, SRC2 603, bit nineteen through bit fourteen, contains the address of a register in integer registers 209. This second

source register contains the packed data, Source2, to be used during execution of the operation. DEST 605, bit five through bit zero, contains the address of a register in integer registers 209. This destination register will store the result packed data, Result, of the packed data operation. Although this general format for a control
5 signal uses source and destination registers in integer registers 209, one skilled in the art would understand that an alternative embodiment can use the source and destination addresses of memory locations in cache 206 or main memory 104.

Control bits SZ 610, bit twelve and bit thirteen, indicates the length of the data elements in the first and second packed data source registers. If SZ 610 equals 012,
10 then the packed data is formatted as packed byte data type 501. If SZ 610 equals 102, then the packed data is formatted as packed word data type 502. SZ 610 equaling 002 or 112 is reserved, however, given the above description, one of ordinary skill in the art would understand that one of these values could be used for doubleword size data type 503.

15 Control bit T 611, bit eleven, indicates whether the operation is to be carried out with saturate mode. If T 611 equals one, then a saturating operation is performed. If T 611 equals zero, then a nonsaturating operation is performed. Saturating operations will be described later.

Control bit S 612, bit ten, indicates the use of a signed operation. If S 612
20 equals one, then a signed operation is performed. If S 612 equals zero, then an unsigned operation is performed.

DESCRIPTION OF SATURATE/UNSATURATE

As mentioned previously, T 611 indicates whether operations optionally saturate. Where the result of an operation, with saturate enabled, overflows or underflows the range of the data, the result will be clamped. Clamping means
5 setting the result to a maximum or minimum value should a result exceed the range's maximum or minimum value. In the case of underflow, saturation clamps the result to the lowest value in the range and in the case of overflow, to the highest value.
The allowable range for each data format is shown in Table 1.

Data Format	Minimum Value	Maximum Value
Unsigned Byte	0	255
Signed Byte	-128	127
Unsigned Word	0	65535
Signed Word	-32768	32767
UnSigned Doubleword	0	$2^{64}-1$
Signed Doubleword	-2^{63}	$2^{63}-1$

Table 1

10

As mentioned above, T 611 indicates whether saturating operations are being performed. Therefore, using the unsigned byte data format, if an operation's result = 258 and saturation was enabled, then the result would be clamped to 255 before being stored into the operation's destination register. Similarly, if an operation's
15 result = -32999 and processor 209 used signed word data format with saturation enabled, then the result would be clamped to -32768 before being stored into the operation's destination register.

POPULATION COUNT

One embodiment of the present invention enables population count operations to be performed on packed data. That is, the present invention generates a result data element for each data element of a first packed data. Each result data element represents the number of bits set in each corresponding data element of the first packed data. In one embodiment, the total number of bits set to one is counted.

Table 2a illustrates an in-register representation of a population count operation on a packed data. The first row of bits is the packed data representation of a Source1 packed data. The second row of bits is the packed data representation of the Result packed data. The number below each data element bit is the data element number. For example, Source1 data element 0 is 1000111110001000₂. Therefore, if the data elements are sixteen bits in length (word data), and a population count operation is performed, ALU 203 produces the Result packed data as shown.

01110010 00000101	11111111 11111111	01111111 11111111	10001111 10001000
<u> </u> = <u> </u> ²	<u> </u> = <u> </u> ²	<u> </u> = <u> </u> ¹	<u> </u> = <u> </u> ⁰
00000000 00000110	00000000 00010000	00000000 00001111	00000000 00000111
₃	₂	₁	₀

Table 2a

In another embodiment, population counts are performed on eight bit data elements. Table 2b illustrates an in-register representation of a population count on a packed data having eight eight-bit packed data elements.

01111111	01010101	10101010	10000001	10000000	11111111	11001111	00000000
<u> </u> = <u> </u> ⁷	<u> </u> = <u> </u> ⁶	<u> </u> = <u> </u> ⁵	<u> </u> = <u> </u> ⁴	<u> </u> = <u> </u> ³	<u> </u> = <u> </u> ²	<u> </u> = <u> </u> ¹	<u> </u> = <u> </u> ⁰
00000111	00000100	00000100	00000010	00000001	00001000	00000110	00000000
₇	₆	₅	₄	₃	₂	₁	₀

Table 2b

In another embodiment, population counts are performed on thirty-two bit data elements. Table 2c illustrates an in-register representation of a population count on a packed data having two, thirty-two bit, packed data elements.

11111111 11111111 11111111 11111111	10000000 11110000 11001111 10001000
=	=
00000000 00000000 00000000 00100000	00000000 00000000 00300000 00001101

Table 2c

5 Population counts can also be performed on sixty-four bit integer data. That is, the number of bits set to one, in sixty-four bits of data, is totalled. Table 2d illustrates an in-register representation of a population count on sixty-four bit integer data.

11111111 11111111 11111111 11111111 10000000 11110000 11001111 10001000
=
00000000 00000000 00000000 00100000 00000000 00000000 00000000 00101101

Table 2d

10

A METHOD OF PERFORMING A POPULATION COUNT

Figure 7 illustrates one method of performing a population count operation on packed data. At step 701, responsive to receiving a control signal 207, decoder 202 decodes that control signal 207. In one embodiment, control signal 207 is supplied via bus 101. In another embodiment, control signal 207 is supplied by cache 206. Thus, decoder 202 decodes: the operation code for population count, and SRC1 602 and DEST 605 addresses in integer registers 209. Note that SRC2 603 is not used in this present embodiment of the present invention. As well, saturate/unsaturate,

signed/unsigned, and length of the data elements in the packed data are not used in this embodiment. In the present embodiment of the invention, only sixteen bit data element length packed addition is supported. However, one skilled in the art would understand that population counts can be performed on packed data having eight
5 packed byte data elements or two packed doubleword data elements.

At step 702, via internal bus 205, decoder 202 accesses integer registers 209 in register file 204 given the SRC1 602 address. Integer registers 209 provides ALU 203 with the packed data, Source1, stored in the register at this address. That is, integer registers 209 communicate the packed data to ALU 203 via internal bus 205.

10 At step 703, decoder 202 enables ALU 203 to perform a population count operation. In an alternative embodiment, decoder 202 further communicates, via internal bus 205, the length of packed data elements.

At step 705, assuming the length of the data elements is sixteen bits, then ALU 203 totals the number of bits set in bit fifteen through bit zero of Source1, producing
15 bit fifteen through bit zero of Result packed data. In parallel with this totaling, ALU 203 adds totals thirty-one through bit sixteen of Source1, producing bit thirty-one through bit sixteen of Result packed data. In parallel with the generation of these totals, ALU 203 totals bit forty-seven through bit thirty-two of Source1, producing bit forty-seven through bit thirty-two of Result packed data. In parallel with the
20 generation of these totals, ALU 203 totals bit sixty-three through bit forty-eight of Source1, producing bit sixty-three through bit forty-eight of Result packed data.

At step 706, decoder 202 enables a register in integer registers 209 with DEST 605 address of the destination register. Thus, the Result packed data is stored in the register addressed by DEST 605.

A METHOD OF PERFORMING A POPULATION COUNT ON ONE DATA ELEMENT

5 Figure 8 illustrates the steps followed when performing a population count operation on one data element of a packed data and generating a single result data element for a result packed data. At step 810a, a column sum, CSum1a, and a column carry, CCarry 1a, are generated from Source1 bits fifteen, fourteen, thirteen and twelve. At step 810b, a column sum, CSum1b, and a column carry, CCarry 1b, are generated from Source1 bits eleven, ten, nine and eight. At step 810c, a column sum, CSum1c, and a column carry, CCarry 1c, are generated from Source1 bits seven, six, five and four. At step 810d, a column sum, CSum1d, and a column carry, CCarry 1d, are generated from Source1 bits three, two, one and zero. In one embodiment of the present invention, steps 810a-d are performed in parallel. At step 15 820a, a column sum, CSum2a, and a column carry, CCarry 2b, are generated from CSum1a, CCarry1a, CSum1b, and CCarry1b. At step 820b, a column sum, CSum2b, and a column carry, CCarry 2b, are generated from CSum1c, CCarry1c, CSum1d, and CCarry1d. In one embodiment of the present invention, steps 820a-b are performed in parallel. At step 830, a column sum, CSum3, and a column carry, CCarry 3, are generated from CSum2a, CCarry2a, CSum2b, and CCarry2b. At step 20 840, a Result is generated from CSum3 and CCarry3. In one embodiment, the Result is represented in sixteen bits. In this embodiment, as only bit four through bit zero are need to represent the maximum number of bits set in a Source1, bits fifteen

through five are set to zero. The maximum number of bits for Source1 is sixteen. This occurs when Source1 equals 1111111111111112. The Result would be sixteen and would be represented by 000000000000100002.

Thus, to calculate four result data elements for a population count operation on
5 a sixty-four bit packed data, the steps of Figure 8 would be performed for each data element in the packed data. In one embodiment, the four sixteen bit result data elements would be calculated in parallel.

A CIRCUIT FOR PERFORMING A POPULATION COUNT

The preferred embodiment of the present invention employs numerous circuits
10 to implement the invention. Figure 9 illustrates one circuit for performing a population count operation on a packed data having four word data elements. Figure 10 illustrates a detailed circuit for performing a population count operation on one word data element of a packed data.

Figure 9 illustrates a circuit wherein Source1 bus 901 carries information
15 signals to the popcnt circuits 908a-d via Source1_N 906a-d. Thus, popcnt circuit 908a totals the number of bits set in bit fifteen through bit zero of Source1, producing bit fifteen through bit zero of Result. Popcnt circuit 908b totals the number of bits set in bit thirty-one through bit sixteen of Source1, producing bit thirty-one through bit sixteen of Result. Popcnt circuit 908c totals the number of
20 bits set in bit forty-seven through bit thirty-two of Source1, producing bit forty-seven through bit thirty-two of Result. Popcnt circuit 908d totals the number of bits set in bit sixty-three through bit forty-eight of Source1, producing bit sixty-three through bit forty-eight of Result. Enable 904a-d receives, from ALU 910, via

control 903, control signals enabling popcnt circuits 908a-d to perform population count operations, and to place a Result on the Result Bus 909. One skilled in the art would be able to create such a circuit given the illustrations in Figures 1-9 and the above description.

- 5 Popcnt circuits 908a-d communicate result information of a packed population count operation onto Result bus 909, via result out 907a-d. This result information is then stored in the integer register specified by the DEST 605 register address.

A CIRCUIT FOR PERFORMING A POPULATION COUNT ON ONE DATA ELEMENT

- Figure 10 illustrates a detailed circuit for performing a population count
10 operation on one, word, data element of a packed data. In particular, Figure 10 illustrates a portion of popcnt circuit 908a. To achieve the maximum performance for applications employing a population count operation, the operation should be complete within one clock cycle. Therefore, given that accessing a register and storing a result requires a certain percentage of the clock cycle, the circuit of Figure
15 9 completes its operation within approximately 80% of one clock period. This circuit has the advantage of allowing processor 109 to execute a population count operation on four sixteen bit data elements in one clock cycle.

- Popcnt circuit 908a employs 4->2 carry-save adders (unless otherwise specified, CSA will refer to a 4->2 carry-save adder). 4->2 carry-save adders, as may be
20 employed in the popcnt circuit 908a-d, are well known in the art. A 4->2 carry-save adder is an adder that adds four operands, resulting in two sums. Since the population count operation in popcnt circuit 908a involves sixteen bits, the first level includes four 4->2 carry-save adders. These four 4->2 carry-save adders transform

the sixteen one-bit operands into eight two-bit sums. The second level transforms the eight two-bit sums into four three-bit sums, and the third level transforms the four three-bit sums into two four-bit sums. Then a four-bit full adder, adds the two four-bit sums to generate a final result.

- 5 Although 4->2 carry-save adders are used, an alternative embodiments could employ 3->2 carry-save adders. Alternatively, a number of full adders could be used; however, this configuration would not provide a result as quickly as the embodiment shown in Figure 10.

- 10 Source1_{IN 15-0} 906a carries bit fifteen through bit zero of Source1. The first four bits are coupled to the inputs of a 4->2 carry-save adder (CSA 1010a). The next four bits are coupled to the inputs of CSA 1010b. The next four bits are coupled to the inputs of CSA 1010c. The final four bits are coupled to the inputs of CSA 1010d. Each CSA 1010a-d generates two, two-bit, outputs. The two, two bit, outputs of CSA 1010a are coupled to two inputs of CSA 1020a. The two, two bit, outputs of CSA 1010b are coupled to the other two inputs of CSA 1020a. The two, two bit outputs of CSA 1010c are coupled to two inputs of CSA 1020b. The two, two bit outputs of CSA 1010d are coupled to the other two inputs of CSA 1020b. Each CSA 1020a-b generates two, three bit, outputs. The two, three bit, outputs of 1020a are coupled to two inputs of CSA 1030. The two, three bit, outputs of 1020b are coupled to the other two inputs of CSA 1030. CSA 1030 generates two, four bit, outputs.

20 These two four bit outputs are coupled to two inputs of a full adder (FA 1050). FA 1050 adds the two four bit inputs and communicates bit three through bit zero of

Result Out 907a as a total of the addition of the two, four bit, inputs. FA 1050 generates bit four of Result Out 907a through carry out (CO 1052). In an alternative embodiment, a five bit full adder is used to generate bit four through bit zero of Result Out 907a. In either case, bit fifteen through bit five of Result Out 907a are
5 tied to zero. As well, any carry inputs to the full adder are tied to zero.

Although not shown in Figure 10, one skilled in the art would understand that Result Out 907a could be multiplexed or buffered onto Result bus 909. The multiplexor would be controlled by Enable 904a. This would allow other ALU circuits to write data onto Result bus 909.

10

Thus, an apparatus and method of increasing processor performance by efficiently reducing the difference between data element length and actual carrying and storage capacity has been described. In particular, a processor capable of performing population count operations on packed data has been described.

CLAIMS



1. A processor for manipulating packed data, said packed data including a first data element D1 and a second data element D2, each of said data elements having a predetermined number of bits, said processor comprising:
 - a) a decoder for decoding a control signal responsive to receiving said control signal;
 - b) a register being coupled to said decoder, said register for storing said packed data;
 - c) a circuit being coupled to said decoder, said circuit for generating a first result data element R1 and a second data element R2, said circuit for generating R1 to represent a total number bits set in D1, and said circuit for generating R2 to represent a total number bits set in D2.
2. The processor of Claim 1 wherein said processor includes a second register, and said second register for storing R1 and R2 as a result packed data.
3. The processor of Claim 2 wherein said circuit includes a first circuit being coupled to said decoder and a second circuit being coupled to said decoder, said first



circuit for totaling a number of bits set in D1, and said second circuit for totaling a number of bits set in D2.

4. The processor of Claim 2 wherein said packed data further includes a third data element D3 and a fourth data element D4, said circuit further for generating a third result data element R3 and a fourth data element R4, said circuit for generating R3 to represent a total number of bits set in D3, said circuit for generating R4 to represent a total number of bits set in D4, and said second register for storing R1, R2, R3 and R4 as a result packed data.

5. The processor of Claim 1 wherein each of said data elements includes sixteen bits.

6. The processor of Claim 1 wherein a set bit equals one.

7. The processor of Claim 1 wherein said processor supports operations supported by an Intel 80486™ microprocessor.

8. The processor of Claim 1 wherein said processor supports operations supported by an Intel Pentium™ microprocessor.

9. The processor of Claim 1 wherein said processor supports operations defined by the IA™, Intel Architecture.



10. The processor of Claim 1 wherein said processor includes a cache coupled to said decoder, and said cache for storing data.

11. The processor of Claim 1 wherein said processor is further for storing integer data.

12. A computer system for manipulating packed data, said packed data including a plurality of data elements, each of said plurality of data elements having a predetermined number of bits, said computer system comprising:

- a) a memory for storing data and a control signal, and
- b) a processor coupling to said memory, said processor having:
 - i) a decoder for decoding said control signal;
 - ii) a plurality of registers for providing and storing said packed data, said plurality of registers being coupled to said decoder, said plurality of registers having a first register, and
 - iii) a circuit for performing an operation on said packed data, said circuit being coupled to said decoder, responsive to decoding said control signal said decoder enabling said first register to provide a first packed data, said first packed data having a first data element D1 and a second data element D2, said circuit generating a result data element R1 equal to a total number of bits set in D1, said circuit generating a result data element R2



equal to a total number of bits set in D2, said circuit providing a result packed data to said plurality of registers, said result packed data including R1 and R2, said plurality of registers storing said result packed data in one of said plurality of registers.

13. A method of manipulating packed data in a computer system, said computer system including a memory being coupled to a microprocessor, said microprocessor including a decoder for decoding a control signal, a plurality of registers for storing said packed data, and a circuit for processing said packed data, said packed data including a first data element D1 and a second data element D2, each of said data elements having a predetermined number of bits, said decoder being coupled to said plurality of registers and said circuit, said method comprising the steps of:

- a) responsive to receiving a control signal, decoding said control signal;
- b) providing said packed data to said circuit;
- c) generating a first result data element representative of a number of bits set in D1;
- d) generating a second result data element representative of a number of bits set in D2, and
- e) storing said first result data element and said second result data element as a result packed data.



14. A processor for performing a population count on a packed data, said packed data having four data elements, D1, D2, D3 and D4 respectively, each of said data elements including sixteen bits, said processor comprising:

- a) a decoder responsive to receiving a control signal for decoding said control signal;
- b) a plurality of registers for storing packed data, said plurality of registers coupling to said decoder, said plurality of registers having a first register, and
- c) an operation circuit for performing an operation on said packed data, said operation circuit being coupled to said decoder, responsive to decoding said control signal said decoder enabling said first register to provide said first packed data to said operation circuit, said operation circuit including:
 - i) a first circuit for generating a result data element R1 equal to a total number of bits set in D1,
 - ii) a second circuit for generating a second result data element R2 equal to a total number of bits set in D2,
 - iii) a third circuit for generating a third result data element R3 equal to a total number of bits set in D3, and
 - iv) a fourth circuit for generating a fourth result data element R4 equal to a total number of bits set in D4, said operation circuit providing said plurality of registers with a result packed data,



said result packed data having said R1, R2, R3, and R4, said plurality of registers storing said result packed data in one of said plurality of registers.

15. A method of performing a population count operation on a packed data in a processor, said processor including a decoder for decoding a control signal, a plurality of registers for storing said packed data, and a circuit for processing said packed data, said decoder being coupled to said plurality of registers and said circuit, said packed data including four data elements D1, D2, D3, and D4, respectively, said method comprising the steps of:

- a) generating a first result data representative of a total number of bits equal to one in D1;
- b) generating a second result data representative of a total number of bits equal to one in D2;
- c) generating a third result data representative of a total number of bits equal to one in D3;
- d) generating a fourth result data representative of a total number of bits equal to one in D4, and
- e) storing D1, D2, D3, and D4, as a result packed data.