Title: CACHING INSTRUCTIONS FOR A MULTIPLE-STATE PROCESSOR

Abstract: A method and apparatus for caching instructions for a processor having multiple operating states. At least two of the operating states of the processor supporting different instruction sets. A block of instructions may be retrieved from memory while the processor is operating in one of the states. The instructions may be pre-decoded in accordance with said one of the states and loaded into cache. The processor, or another entity, may be used to determine whether the current state of the processor is the same as said one of the states used to pre-decode the instructions when one of the pre-decoded instructions in the cache is needed by the processor.
For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
CACHING INSTRUCTIONS FOR A MULTIPLE-STATE PROCESSOR

BACKGROUND

Field

[0001] The present disclosure relates generally to processing systems, and more specifically, to caching instructions for a multiple-state processor.

Background

[0002] Computers typically employ a processor supported by memory. Memory is a storage medium that holds the programs and data needed by the processor to perform its functions. Recently, with the advent of more powerful software programs, the demands on memory have been increasing at an astounding rate. The result is that modern processors require a large amount of memory, which is inherently slower than the smaller memories. Large memories with speeds capable of supporting today’s processors are simply too expensive for large scale commercial applications.

[0003] Computer designers have addressed this problem by organizing memory into several hierarchical components. The largest component, in terms of capacity, is commonly a hard drive. The hard drive provides large quantities of inexpensive permanent storage. The basic input/output system (BIOS) and the operating system are just a few examples of programs that are typically stored on the hard drive. These programs may be loaded into Random Access Memory (RAM) when the computer is operational. Software applications that are launched by a user may also be loaded into RAM from the hard drive. RAM is a temporary storage area that allows the processor to access the information more readily.

[0004] The computer’s RAM is still not fast enough to keep up with the processor. This means that processors may have to wait for program instructions and data to be written to and read from the RAM. Caches are used to increase the speed of memory access by making the information most often used by the processor readily available. This is accomplished by integrating a small amount of memory, known as a primary or Level 1 (L1) cache, into the processor. A secondary or Level 2 (L2) cache between the RAM and L1 cache may also be used in some computer applications.

[0005] The speed of the computer may be further improved by partially decoding the instructions prior to being placed into the cache. This process is often referred to as
“pre-decoding,” and entails generating some “pre-decode information” that can be stored along with the instruction in the cache. The pre-decode information indicates some basic aspects of the instruction such as whether the instruction is an arithmetic or storage instruction, whether the instruction is a branch instruction, whether the instruction will make a memory reference, or any other information that may be used by the processor to reduce the complexity of the decode logic. Pre-decoding instructions improves processor performance by reducing the length of the machine’s pipeline without reducing the frequency at which it operates.

[0006] Processors capable of operating in multiple states are becoming commonplace with today’s emerging technology. A “multiple state processor” means a processor that can support two or more different instruction sets. The ARM (Advance RISC Machine) processor is just one example. The ARM processor is an efficient, low-power RISC processor that is commonly used today in mobile applications such as mobile telephones, personal digital assistants (PDA), digital cameras, and game consoles just to name a few. ARM processors have historically supported two instructions sets: the ARM instruction set, in which all instructions are 32-bits long, and the Thumb instruction set which compresses the most commonly used instructions into a 16-bit format. A third instruction set that has recently been added to some ARM processors is “Thumb-2 Execution Environment” (T2EE). T2EE is an instruction set (similar to Thumb) that is optimized as a dynamic (JIT) compilation target for bytecode languages, such as Java and NET.

[0007] These multiple-state processors have significantly increased the capability of modern day computing systems, but can pose unique challenges to the computer designer. By way of example, if a block of instructions the size of one line in the L1 instruction cache contains instructions from multiple instruction sets, pre-decode information calculated assuming that the entire cache line contains instructions in one state cannot be used for those instructions that are actually in the other state. The solution described in this disclosure is not limited to ARM processors with Thumb and/or T2EE capability, but may be applied to any system that pre-decodes instructions for multiple instruction sets with overlapping instruction encodings prior to placing them into cache.
SUMMARY

[0008] One aspect of the present invention is directed to a method of operating a processor. The processor is capable of operating in different states, with each state supporting a different instruction set. The method includes retrieving a block of instructions from memory while the processor is operating in one of the states, pre-decoding the instructions in accordance with said one of the states, loading the pre-decoded instructions into cache, and determining whether the current state of the processor is the same as said one of the states used to pre-decode the instructions when one of the pre-decoded instructions in the cache is needed by the processor.

[0009] Another aspect of the present invention is directed to a processing system. The processing system includes memory, cache, a processor capable of operating in different states, each of the states supporting a different instruction set, the processor being further configured to retrieve a block of instructions from the memory while operating in one of the states, and a pre-decoder configured to pre-decode the instructions retrieved from the memory in accordance with said one of the states, wherein the processor is further configured to load the pre-decoded instructions into the cache, and, when one of the pre-decoded instructions in the cache is needed by the processor, determine whether its current state is the same as said one of the states used to pre-decode the instructions.

[0010] It is understood that other embodiments of the present invention will become readily apparent to those skilled in the art from the following detailed description, wherein various embodiments of the invention are shown and described by way of illustration. As will be realized, the invention is capable of other and different embodiments and its several details are capable of modification in various other respects, all without departing from the spirit and scope of the present invention. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Aspects of the present invention are illustrated by way of example, and not by way of limitation, in the accompanying drawings, wherein:

[0012] FIG. 1 is a functional block diagram illustrating an example of a processing system;
[0013] FIG. 2 is a functional block diagram illustrating an example of a pre-decoder;

[0014] FIG. 3A is a diagram illustrating the data structure of an instruction in cache;

[0015] FIG. 3B is a diagram illustrating the data structure of a cache line;

[0016] FIG. 3C is a diagram illustrating the data structure of a cache line with state information;

[0017] FIG. 4 is a flow diagram illustrating the operation of one embodiment of a multiple-state processor;

[0018] FIG. 5 is a diagram illustrating the data structure of a cache directory tag with state information; and

[0019] FIG. 6 is a flow diagram illustrating the operation of another embodiment of a multiple-state processor.

**Detailed Description**

[0020] The detailed description set forth below in connection with the appended drawings is intended as a description of various embodiments of the present invention and is not intended to represent the only embodiments in which the present invention may be practiced. The detailed description includes specific details for the purpose of providing a thorough understanding of the present invention. However, it will be apparent to those skilled in the art that the present invention may be practiced without these specific details. In some instances, well known structures and components are shown in block diagram form in order to avoid obscuring the concepts of the present invention.

[0021] FIG. 1 is a conceptual block diagram illustrating an example of a computer 100, or other processing system, capable of processing, retrieving and storing information. The computer 100 may be a stand-alone component, or alternatively, embedded in a device such as a mobile telephone, PDA, digital camera, game console, or any other suitable device. The computer 100 may include a processor 102, such as a microprocessor or other processing entity. In one embodiment of the computer 100, the processor 102 is a 32-bit embedded RISC microprocessor commonly referred to in the industry as an ARM processor. The ARM processor is commonly used as a platform to run complex operating systems for wireless, consumer, imaging, automotive, industrial and networking applications. Included in many ARM processors is a Thumb instruction set which compresses most of the commonly used 32-bit ARM instructions into 16-bit
forms. ARM has added 32-bit instructions to Thumb as well. The Thumb instruction set is shorthand for a subset of the 32-bit ARM instructions that can be used when code density is required. As indicated in the background portion of this disclosure, the ARM processor may also execute T2EE.

[0022] The computer 100 may also include memory 104, which holds the program instructions and data needed by the processor 102 to perform its functions. The memory 104 may be implemented with RAM or other suitable memory, and may be comprised of the computer’s main memory and optionally a L2 cache. An instruction cache 106 may be used between the processor 102 and the memory 104. The instruction cache 106 is a relatively small, high-speed L1 cache used for the temporary storage of program instructions from the memory 104 to be executed by the processor 102. In one embodiment of the computer 100, the instruction cache 106 is a high-speed static RAM (SRAM) instead of the slower and cheaper dynamic RAM (DRAM) that may be used for the memory 104. The instruction cache 106 provides a mechanism for increasing processor access speed because most programs repeatedly access the same instructions. By keeping as much of this information as possible in the instruction cache 106, the processor 102 avoids having to access the slower memory 104. The computer 100 may also include a data cache (not shown) for the storage of data used in the execution of the instructions.

[0023] The instruction cache 106 provides storage for the most recently accessed instructions by the processor 102 from the memory 104. When the processor 102 needs instructions from the memory 104, it first checks the instruction cache 106 to see if the instruction is there. When an instruction required by the processor 102 is found in the instruction cache 106, the lookup is called a “cache hit.” On a cache hit, the instruction may be retrieved directly from the instruction cache 106, thus drastically increasing the rate at which instructions may be processed. An instruction required by the processor 102 that is not found in the instruction cache 106 results in a “cache miss.” On a cache miss, the processor 102 must fetch the required instruction from the memory 104, which takes considerably longer than fetching an instruction from the instruction cache 106. Usually the processor 102 fetches a “cache line” from memory 104. The cache line from the memory 104 may be stored in the instruction cache 106 for future access.

[0024] Computer performance may be further enhanced by pre-decoding the instructions from the memory 104 prior to being placed in the instruction cache 106. A
pre-decoder 108 takes the instructions as they are fetched from the memory 104, pre-decodes them in accordance with the operating state of the processor 102, and stores the pre-decode information in the instruction cache 106. Signaling from the processor 102 may be used to indicate the current operating state of the processor 102 for the pre-decoding operation.

[0025] FIG. 2 is a functional block diagram illustrating an example of a pre-decoder. The pre-decoder 108 includes an input register 202 for receiving the instructions in the cache line from the memory 104 (see FIG. 1). In the case of a multiple state processor, two pre-decoders may be used, one for each processor state 204a, 204b. The instructions in the input register 202 may be provided to the pre-decoders 204a, 204b to produce pre-decode information for each instruction. A multiplexer 206 may be used to select the appropriate pre-decoding information based on signaling from the processor 102 (see FIG. 1). In some embodiments of the pre-decoder 108, the input to the pre-decoders 204a, 204b may be gated to conserve power by enabling only the pre-decoder selected by the processor 102. Each instruction from the register 202 along with its corresponding pre-decode information from the multiplexer 206 may be placed in an output register 208. The contents of the output register may be read into both the processor 102 and the instruction cache 106 (see FIG. 1).

[0026] FIG. 3A illustrates how a single instruction 302 is stored with its pre-decode information 304 in the instruction cache 106. FIG. 3B illustrates how an entire cache line consisting of eight instructions is stored in the instruction cache 106. In this example, each instruction 302a-302h has its pre-decode information 304a-304h tacked onto the end. The pre-decoding information may be any number of bits depending on the specific application and the overall design constraints. Moreover, the use of eight instructions to illustrate a cache line is provided as an example only with the understanding that each cache line may hold any number of instructions.

[0027] The instruction cache 106 maintains a cache directory (not shown). The cache directory contains one entry or “tag” for each cache line. A one-to-one mapping exists between a cache directory tag and its associated cache line in the cache storage array. The cache directory tag contains the memory address for the first instruction in the cache line. The processor 102 (see FIG. 1) fetches an instruction by searching the cache directory for a tag having the memory address. A match signals a cache hit. The cache directory tag that matches dictates which cache line contains the required instruction.
The processor may then access the instruction and its pre-decode information. If the processor is unable to find a match (i.e., a cache miss), it will attempt to read a cache line from the memory 104 (see FIG. 1). The cache line retrieved from the memory is pre-decoded and placed into the instruction cache 106. One or more instructions may also be loaded directly into the processor 102 for immediate decoding and execution. The cache directory is updated so that processor 102 can determine where the cache line resides in the cache storage array on future accesses.

[0028] In the processing system of FIG. 1, where a multiple-state processor 102 is employed, multiple instruction sets may reside in the memory 104. In these multiple-state processing environments, a cache line retrieved from the memory 104 may include instruction for multiple states of the processor. Yet, each instruction in the cache line will be pre-decoded based on the current state of the processor 102, and the entire cache line is predecoded at once. If, for example, the processor 102 is operating in a particular state when retrieving a cache line from the memory 104, all instructions contained in the cache line will be pre-decoded as if they are instructions for that state before being placed into the instruction cache 106. A subsequent access to the instruction cache 106 when the processor 102 is in a different state cannot be properly executed because of the pre-decoding.

[0029] Several techniques may be employed by the processing system 100 to ensure that each instruction retrieved from the instruction cache 106 is not executed by the processor 102 with incorrect pre-decode information. One possible solution is to store “state information” with the pre-decoded instruction in each cache line. “State information” is defined as one or more bits that indicate which state the processor 102 was in when the associated cache line was pre-decoded. FIG. 3C illustrates how an entire cache line consisting of eight pre-decoded instructions is stored with state information. Similar to FIG. 3B, each instruction 302a-302h has its pre-decoding information 304a-304h tacked onto the end. However, in addition to what is shown in FIG. 3B, two bits of state information 306 is included at the beginning of the cache line. Two bits of state information provides up to four different operating states. Those skilled in the art will be readily able to determine the number of bits needed to support the requisite number of processor states in other applications.

[0030] FIG. 4 is a flow diagram illustrating the operation of a multiple-state processor using state information with the pre-decoded instruction to determine whether the pre-
decoding information for a particular instruction is correct. In step 402, the processor translates a virtual address for a required instruction into a memory address. The processor then searches the tags in the cache directory for the memory address in step 402. The search results in a cache hit or a cache miss. On a cache miss, the processor will attempt to retrieve the instruction from the memory in step 406. On a cache hit, the processor will retrieve the instruction from the instruction cache in step 408, and compare the corresponding state information with the current state of the processor in step 410. If the current state of the processor matches the state information for the instruction, the processor will decode and execute the instruction using the pre-decode information in step 412. Otherwise, the processor will pre-decode the instruction, in step 414, prior to decoding and executing the instruction in step 412.

[0031] Another possible solution is to include state information in each cache directory tag. FIG. 5 shows how the state information may be added to the tags in the cache directory. The cache directory 502 is shown with a number of tags 504, with each tag corresponding to a cache line in the cache storage array. The cache directory tag 504 contains an address field 506 and a state field 508. The address field 506 includes the memory address for the first instruction in the cache line, and the state field 508 includes state information identifying the state of the processor when the cache line was pre-decoded.

[0032] FIG. 6 is a flow diagram illustrating the operation of a multiple-state processor using state information in the cache directory tag to determine whether the pre-decoding information of a particular instruction is correct. In step 602, the processor performs the required calculations to generate an address with which to perform a cache lookup. In step 604, the processor adds state information to the memory address corresponding to the current state of the processor. The memory address in combination with the state information form a tag. The tag formed by the processor is then compared to the tags in the cache directory in step 606. In this case, a cache hit means not only that the instruction required by the processor is in the instruction cache, but that the pre-decoded instruction can be executed by the processor in its current state. On a cache hit, the processor retrieves the instruction in step 608, and decodes and executes the instruction using the pre-decoding information in step 610. On a cache miss, the processor retrieves the cache line with the required instruction from the memory in step 612, initiates a pre-decode of the cache line in step 614, and uses the pre-decode information
to decode and execute the instruction in step 616. In step 618, the processor stores the pre-decoded cache line into the instruction cache and updates its tag with the address and state information. The result is that multiple copies of a block of memory may exist in the instruction cache at different cache lines. The instructions residing in each of these copies are pre-decoded in different states, and a single lookup in the cache will never return multiple cache hits.

[0033] The various illustrative logical blocks, modules, circuits, elements, and/or components described in connection with the embodiments disclosed herein may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic component, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general-purpose processor may be a microprocessor, but in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing components, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[0034] The methods or algorithms described in connection with the embodiments disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may reside in RAM memory, flash memory, ROM memory, EPROM memory, EEPROM memory, registers, hard disk, a removable disk, a CD-ROM, or any other form of storage medium known in the art. A storage medium may be coupled to the processor such that the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor.

[0035] The previous description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the present invention. Various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without departing from the spirit or scope of the invention. Thus, the present invention is not intended to be limited to the embodiments shown herein, but is to be accorded the full scope consistent with the claims, wherein reference to an element in the singular is not intended to mean
“one and only one” unless specifically so stated, but rather “one or more.” All structural and functional equivalents to the elements of the various embodiments described throughout this disclosure that are known or later come to be known to those of ordinary skill in the art are expressly incorporated herein by reference and are intended to be encompassed by the claims. Moreover, nothing disclosed herein is intended to be dedicated to the public regardless of whether such disclosure is explicitly recited in the claims. No claim element is to be construed under the provisions of 35 U.S.C. §112, sixth paragraph, unless the element is expressly recited using the phrase “means for” or, in the case of a method claim, the element is recited using the phrase “step for.”

WHAT IS CLAIMED IS:
CLAIMS

1. A method of operating a processor, the processor being capable of operating in different states, at least two of the states supporting different instruction sets, the method comprising:

   retrieving a block of instructions from memory while the processor is operating in one of the states;
   
   pre-decoding the instructions in accordance with said one of the states;
   
   loading the pre-decoded instructions into cache; and
   
   determining whether the current state of the processor is the same as said one of the states used to pre-decode the instructions when one of the pre-decoded instructions in the cache is needed by the processor.

2. The method of claim 1 further comprising executing said one of the pre-decoded instructions in the cache if it is determined that the current state of the processor is the same as said one of the states used to pre-decode the instructions.

3. The method of claim 1 further comprising loading state information in the cache with the pre-decoded instructions, the state information indicating that the instructions were pre-decoded in accordance with said one of the states.

4. The method of claim 3 wherein the determination as to whether the current state of the processor is the same as said one of the states used to pre-decode the instructions is made by comparing the current state of the processor to the state information in the cache.

5. The method of claim 4 further comprising executing said one of the pre-decoded instructions in the cache if the comparison between the current state of the processor and the state information in the cache indicates that the current state of the processor is the same as said one of the states used to pre-decode the instructions.

6. The method of claim 4 wherein said one of the pre-decoded instructions includes one of the instructions retrieved from the memory and pre-decode information,
the method further comprising ignoring the pre-decode information if the comparison between the current state of the processor and the state information in the cache indicates that the current state of the processor is different from said one of the states used to pre-decode the instructions.

7. The method of claim 6 further comprising processing said one of the instructions without the pre-decode information before execution.

8. The method of claim 1 wherein the cache includes a tag associated with the pre-decoded instructions, the tag being used by the processor to determine whether said one of the pre-decoded instructions needed by the processor is in the cache, method further comprising adding state information to the tag, the state information indicating that the instructions were pre-decoded in accordance with said one of the states.

9. The method of claim 8 wherein the determination as to whether the current state of the processors is the same as said one of the states used to pre-decode the instructions is made by checking the tag.

10. The method of claim 9 further comprising retrieving said one of the pre-decoded instructions from the cache if the tag indicates that the current state of the processor is the same as said one of the states used to pre-decode the instructions.

11. The method of claim 9 further comprising retrieving another copy of the block of instructions from the memory if the tag indicates that the current state of the processor is different from said one of the states used to pre-decode the instructions.

12. The method of claim 11 further comprising executing one of the instructions from the second copy of instructions.

13. The method of claim 11 further comprising pre-decoding said another copy of the instructions in accordance with the current state of the processor, and loading the pre-decoded second copy of the instructions into the cache.
14. The method of claim 13 further comprising associating a second tag with the pre-decoded instructions from the second copy of instructions, and adding state information to the second tag, the state information added to the second tag indicating that the second copy of the instructions were pre-decoded in accordance with the current state of the processor.

15. The method of claim 1 wherein the operating states of the processor includes at least two of ARM, Thumb and T2EE.

16. A processing system, comprising
memory;
cache;

a processor capable of operating in different states, each of the states supporting a different instruction set, the processor being further configured to retrieve a block of instructions from the memory while operating in one of the states; and

a pre-decoder configured to pre-decode the instructions retrieved from the memory in accordance with said one of the states;

wherein the processor is further configured to load the pre-decoded instructions into the cache, and, when one of the pre-decoded instructions in the cache is needed by the processor, determine whether its current state is the same as said one of the states used to pre-decode the instructions.

17. The processing system of claim 16 wherein the processor is further configured to execute said one of the pre-decoded instructions in the cache if the processor determines that its current state is the same as said one of the states used to pre-decode the instructions.

18. The processing system of claim 16 wherein the processor is further configured to load state information in the cache with the pre-decoded instructions, the state information indicating that the instructions were pre-decoded in accordance with said one of the states.
19. The processing system of claim 18 wherein the processor is further configured to determine whether its current state is the same as said one of the states used to pre-decode the instructions by comparing the current state of the processor to the state information in the cache.

20. The processing system of claim 19 wherein the processor is further configured to execute said one of the pre-decoded instructions in the cache if the comparison between the current state of the processor and the state information in the cache indicates that the current state of the processor is the same as said one of the states used to pre-decode the instructions.

21. The processing system of claim 19 wherein said one of the pre-decoded instructions includes one of the instructions retrieved from the memory and pre-decoding information, and wherein the processor is further configured to ignore the pre-decoding information if the comparison between the current state of the processor and the state information in the cache indicates that the current state of the processor is different from said one of the states used to pre-decode the instructions.

22. The processing system of claim 21 wherein the processor is further configured to process said one of the instructions without the pre-decode information before execution if the comparison between the current state of the processor and the state information in the cache indicates that the current state of the processor is different from said one of the states used to pre-decode the instructions.

23. The processing system of claim 16 wherein the cache includes a tag associated with the pre-decoded instructions, the tag being used by the processor to determine whether said one of the pre-decoded instructions needed by the processor is in the cache, and wherein the processor is further configured to add state information to the tag, the state information indicating that the instructions were pre-decoded in accordance with said one of the states.

24. The processing system of claim 23 wherein the processor is further configured to determine whether its current state is the same as said one of the states used to pre-decode the instructions by checking the tag.
FIG. 1

FIG. 5
FROM MEMORY

INPUT REGISTER

ARM PRE-DECODER

THUMB PRE-DECODER

FROM PROCESSOR

TO INSTRUCTION CACHE

FIG. 2
4/5

402
GENERATE MEMORY ADDRESS

406
RETRIEVE INSTRUCTION FROM MEMORY

404
ACCESS CACHE DIRECTORY

408
CACHE "HIT"

410
COMPARE STATE INFORMATION TO CURRENT PROCESSOR STATE

412
DECODE AND EXECUTE INSTRUCTION

414
PRE-DECODE INSTRUCTION

FIG. 4
GENERATE MEMORY ADDRESS

ADD STATE INFORMATION

CACHE "MISS" ACCESS CACHE DIRECTORY CACHE "HIT"

RETRIEVE CACHE LINE FROM MEMORY

PRE-DECODE CACHE LINE

STORE CACHE LINE IN CACHE

RETRIEVE INSTRUCTION FROM CACHE

DECODE AND EXECUTE INSTRUCTION

FIG. 6