METHOD AND APPARATUS FOR TRANSMITTING MEMORY PRE-FETCH COMMANDS ON A BUS

A processing system and method is disclosed wherein a processor may be configured to predict an address of memory from which data will be needed, transmit to a memory controller a pre-fetch command for the data at the predicted address of the memory, and transmit to the memory controller a read request for the data at the predicted address of the memory if the data is needed.
Published:
— with international search report

GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR, OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
— as to the applicant’s entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

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.
METHOD AND APPARATUS FOR TRANSMITTING MEMORY PRE-FETCH COMMANDS ON A BUS

BACKGROUND

Field

[0001] The present disclosure relates generally to digital systems, and more specifically, to methods and apparatuses for transmitting memory pre-fetch commands over a bus in a processing system.

Background

[0002] Integrated circuits have revolutionized the electronics industry by enabling new applications which were difficult or impossible to realize with discrete devices. Integration allows complex circuits consisting of millions of electronic components to be packaged into a single chip of semiconductor material. As a result, powerful computation devices such as computers, cellular phones, personal digital assistants (PDAs), and the like may be reduced to hand-held devices.

[0003] Integrated circuits are widely used today to implement sophisticated circuitry such as general purpose and specific application processors. A typical integrated processor may include a central processing unit (CPU) with system memory. A high bandwidth system bus may be used to support communications between the two. A bus is typically a shared channel or path between components on an integrated circuit. In addition, there may also be an external bus which may be used to access low latency off-chip memory under control of an on-chip memory controller.

[0004] The off-chip memory is generally formatted into pages. A page is normally associated with a row of memory. In most applications, the memory controller is capable of only opening a limited number of pages at a time. An “open page” means that the memory is pointing to a row of memory and requires only a column access command from the memory controller to read the data. To access an unopened page of memory, the memory controller must present a row access command to the memory to move the pointer before presenting a column access command to read the data. As a result, there is a latency penalty for closing a page in memory and opening a new one.
Many CPU implementations include a pre-fetch mechanism to compensate for the high latency typically encountered with off-chip memory devices. These CPUs may be configured to evaluate instruction streams and make requests for pages from the off-chip memory devices that may be needed in the future in order to reduce the latency and improve CPU performance. However, these pre-fetch requests can often be incorrect due to branches in the instruction stream, as well as CPU interrupts. Moreover, once a pre-fetch request has been accepted on the bus, the data will be read from the off-chip memory and transferred on the bus back to the CPU even though the data may no longer be required. This often leads to lower bandwidth on the bus, delaying CPU access to subsequent data from the off-chip memory device as the speculative data is transferred across the bus to the CPU.

The use of pre-fetch commands to interface a CPU to an off-chip memory device has provided a workable solution for many years. However, as the operating speed of CPU continues to increase exponentially, it is becoming increasingly desirable to reduce the latency typically associated with off-chip memory devices.

**SUMMARY**

In one aspect of the present invention, a method of retrieving data from memory includes predicting an address of the memory from which data will be needed, transmitting to a memory controller a pre-fetch command for the data at the predicted address of the memory, transmitting to the memory controller a read request for the data at the predicted address of the memory, and receiving the data at the predicted address of the memory in response to the read request.

In another aspect of the present invention, a method of retrieving data from memory includes predicting an address from an unopened page of the memory from which data will be needed, transmitting to a memory controller a pre-fetch command for the data at the predicted address of the memory, reading the data at the predicted address into a pre-fetch buffer, and discarding the data from the pre-fetch buffer.

In yet another aspect of the present invention, a processing system includes memory, a memory controller, and a processor configured to predict an address of the memory from which data will be needed, transmit to the memory controller a pre-fetch command for the data at the predicted address of the memory, and transmit to the
memory controller a read request for the data at the predicted address of the memory if the data is needed, wherein the memory controller is further configured to deliver the data at the predicted address of the memory to the processor in response to the read request.

[0010] In a further aspect of the present invention, a processing system includes memory, a memory controller, means for predicting an address of the memory from which data will be needed, means for transmitting to the memory controller a pre-fetch command for the data at the predicted address of the memory, means for transmitting to the memory controller a read request for the data at the predicted address of the memory if the data is needed, and receiving means for receiving the data, wherein the memory controller is further configured to deliver the data at the predicted address of the memory to the receiving means in response to the read request.

[0011] 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**

[0012] Aspects of the present invention are illustrated by way of example, and not by way of limitation, in the accompanying drawings, wherein:

[0013] FIG. 1 is a conceptual block diagram illustrating an example of an integrated circuit with off-chip memory;

[0014] FIG. 2 is a conceptual block diagram illustrating another example of an integrated circuit with off-chip memory;

[0015] FIG. 3 is a timing diagram illustrating the use of a pre-fetch command to reduce the latency of a read operation from a SDRAM; and
[0016] FIG. 4 is a timing diagram illustrating the use of a pre-fetch command to reduce the latency of a read operation from a NAND Flash memory.

DETAILED DESCRIPTION

[0017] 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.

[0018] Several embodiments of an integrated circuit with off-chip memory will be described to illustrate various inventive concepts. In at least one embodiment described herein, a CPU integrated on a chip may interface with off-chip memory through a memory controller in a way that tends to reduce the latency that is typically encountered with such configurations. Those skilled in the art will readily appreciate that these inventive concepts are not limited to such configurations, and may be applied to any processing system in which a bus mastering device communicates with memory. By way of example, a CPU may use the various inventive concepts described throughout this disclosure to communicate with on-chip memory. Alternatively, a discrete processor, or other bus mastering device, may be used to communicate with discrete memory.

[0019] FIG. 1 is a conceptual block diagram illustrating an integrated circuit 102 with off-chip memory. The off-chip memory may be implemented with any storage device having a multiple page format, including by way of example, memory, registers, a bridge, or any other device capable of retrieving and storing information. In the embodiment shown in FIG. 1, the off-chip memory may include synchronous dynamic random access memory (SDRAM) 104 and NAND Flash memory 106. A CPU 108 may access the SDRAM 104 through a SDRAM memory controller 110, and access the NAND Flash memory 106 through a NAND Flash controller 112. The CPU 108 may
be any processing component, including by way of example, a microprocessor, a digital
signal processor (DSP), a programmable logic component, a combination of gates or
transistor logic, or any other processing component.

[0020] The integrated circuit 102 may include a system bus 114 to interface the
CPU 108 to the SDRAM controller 110 and the NAND Flash controller 112. The
system bus 114 may also be used to provide communications between various other
integrated components which have been omitted to avoid obscuring the various
inventive concepts. The system bus 114 may be configured in a variety of ways
depending on the specific application and the overall design constraints. By way of
example, the system bus 114 may be implemented as a shared bus or a point-to-point
switched connection.

[0021] In one embodiment of a processing system, a bus mastering device may be
used to generate pre-fetch commands to access unopened pages in memory. By way of
example, the CPU 108 shown in FIG. 1 may transmit a pre-fetch command over the
system bus 114 to the SDRAM memory controller 110. The pre-fetch command may
include an address that may be used by the SDRAM memory controller 110 to access a
page of memory in the SDRAM 104 in anticipation of a future request for the data by
the CPU 108. The SDRAM memory controller 110 may simply open the page of
memory in the SDRAM 104, or alternatively, read a portion or all of the data from the
page into a pre-fetch buffer. In the simplest configuration, the SDRAM memory
controller 110 simply opens the page of memory in the SDRAM 104, and the CPU 108
gets the benefit of lower latency due to the fact that the page is already open, and
therefore, does not require a row access command from the SDRAM memory controller
110. In an alternate configuration, the SDRAM memory controller 110 could
implement a pre-fetch buffer in which it will read data from the page and store that data
into its pre-fetch buffer. The amount of data read from the page may be predetermined,
or alternatively, broadcast with the pre-fetch command. When the CPU 108 is ready to
read the data, it simply reads the data from the pre-fetch buffer with a much lower
latency than reading from the SDRAM 104. If the CPU 108 does not need the data,
then the data may be discarded by the SDRAM memory controller 110, without having
been transmitted across the CPU to SDRAM bus 114. The SDRAM memory
controller 110 may determine that the CPU 108 does not need the data using any
number of algorithms. By way of example, the SDRAM memory controller 110 may discard the data if it does not receive a read request from the CPU 108 within a predetermined time or a predetermined number of read requests to a different page. Alternatively, the SDRAM memory controller 110 may discard the data if it receives a subsequent pre-fetch command before it receives a read request for the data in the pre-fetch buffer.

[0022] In a similar manner, the CPU 108 shown in FIG. 1 may access the NAND Flash memory 106 by transmitting a pre-fetch command over the system bus 114 to the NAND Flash controller 112. The pre-fetch command may include an address that may be used by the NAND Flash controller 112 to access an unopened page of memory in the NAND Flash memory 106. The NAND Flash controller 112 may simply buffer the page of memory in the NAND Flash memory 106 to give the CPU 108 the benefit of lower latency for a subsequent read request. Alternatively, the NAND Flash controller 112 can read the data from the NAND Flash memory page buffer into a pre-fetch buffer in anticipation of a future request for the data by the CPU 108. If the CPU 108 does not need the data, then the data may be discarded by the NAND Flash controller 112. The NAND Flash controller 112 determines that the CPU 108 does not need the data in the same manner described earlier in connection with the SDRAM controller, or by any other means.

[0023] In the embodiments described thus far, the memory controller, whether it be the SDRAM memory controller 110, the NAND Flash controller 112, or any other memory controller, does not transmit the data to the bus mastering device in response to a pre-fetch command. The data is only transmitted across the system bus in response to a subsequent read request by the bus mastering device for the same address that was included in the pre-fetch command. Moreover, there is no requirement that the memory controller read the data into a pre-fetch buffer or even open the page in memory in response to the pre-fetch command. In certain embodiments, the pre-fetch command may provide nothing more than a notification from the bus mastering device that a certain address will be requested in the future and that if there is something that the memory controller can do to minimize the latency of the transfer it should do so, but is not required to do so.
In one embodiment of the integrated circuit 102, the system bus 114 may be implemented with an address channel, a write channel and a read channel. The address channel may be used by the CPU 108 to access a particular location in off-chip memory through the appropriate memory controller. Sideband signaling may be used to support pre-fetch commands, as well as other control signals that indicate, among other things, whether the CPU 108 is requesting a read or write operation. The write channel may be used to write data to off-chip memory through the appropriate memory controller and the read channel may be used to read data from off-chip memory through the appropriate memory controller. This bus structure is well known in the art.

In an alternative embodiment of the integrated circuit 102, the bus structure may be configured to eliminate the address channel. This concept is illustrated in FIG. 2. The address channel may be eliminated by redefining the write channel as a "transmit channel" 202. The transmit channel 202 may be used as a generic medium for transmitting information between the CPU 108 and the appropriate memory controller in a time division multiplexed fashion. This information may include address information, control signals, and data to be written to off-chip memory. An example of a control signal is a signal that indicates the number of bytes to be read from or written to off-chip memory in one read or write operation. Another example of a control signal is a signal that indicates which byte lanes of the transmit channel will be used to transmit the data to be written to off-chip memory. By way of example, 2-bytes of data transmitted on a 32-bit transmit channel 202 will need 2 of the 4 byte lanes. This control signal may be used to indicate to the appropriate memory controller which of the 2 byte lanes on the transmit channel 202 will be used to transmit the data.

The various concepts described in connection with FIG. 2 may be implemented using any number of protocols. In the detailed description to follow, an example of a bus protocol will be presented. This bus protocol is being presented to illustrate the inventive aspects of a processing system, with the understanding that such inventive aspects may be used with any suitable protocol. The basic signaling protocol for the transmit channel is shown below in Table 1. Those skilled in the art will readily be able to vary and/or add signals to this protocol in the actual implementation of the bus structure described herein.
### TABLE 1

<table>
<thead>
<tr>
<th>Signal</th>
<th>Definition</th>
<th>Driven By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock</td>
<td>The reference clock signal.</td>
<td>Processing System</td>
</tr>
<tr>
<td>Valid</td>
<td>Valid information is being transmitted on the transmit channel.</td>
<td>CPU</td>
</tr>
<tr>
<td>Type (2:0)</td>
<td>Indicates the type of information being transmitted on the transmit channel.</td>
<td>CPU</td>
</tr>
<tr>
<td>Transfer Ack</td>
<td>Indicates memory controller is capable of accepting information on the transmit channel.</td>
<td>Memory Controller</td>
</tr>
<tr>
<td>Transmit Channel</td>
<td>Channel driven by the CPU to transmit information.</td>
<td>CPU</td>
</tr>
</tbody>
</table>

[0027] The same signaling protocol may be used for the receive channel as shown below in Table 2.

### TABLE 2

<table>
<thead>
<tr>
<th>Signal</th>
<th>Definition</th>
<th>Driven By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock</td>
<td>The reference clock signal.</td>
<td>Processing System</td>
</tr>
<tr>
<td>Valid</td>
<td>Valid information is being transmitted on the receive channel.</td>
<td>Memory Controller</td>
</tr>
<tr>
<td>Type (2:0)</td>
<td>Indicates the type of information being transmitted on the receive channel.</td>
<td>Memory Controller</td>
</tr>
<tr>
<td>Transfer Ack</td>
<td>Indicates CPU is capable of accepting read data on the receive channel.</td>
<td>CPU</td>
</tr>
<tr>
<td>Receive Channel</td>
<td>Channel driven by the receiving component to broadcast information</td>
<td>Memory Controller</td>
</tr>
</tbody>
</table>

[0028] The definition of the Type field used in this signaling protocol is shown in Table 3.
### TABLE 3

<table>
<thead>
<tr>
<th>Type Value</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>Reserved</td>
</tr>
<tr>
<td>001</td>
<td>Valid Write Address</td>
</tr>
<tr>
<td>010</td>
<td>Valid Write Control Signals</td>
</tr>
<tr>
<td>011</td>
<td>Valid Write Data</td>
</tr>
<tr>
<td>100</td>
<td>Valid Read Pre-Fetch Command</td>
</tr>
<tr>
<td>101</td>
<td>Valid Read Address</td>
</tr>
<tr>
<td>110</td>
<td>Valid Read Control Signals</td>
</tr>
<tr>
<td>111</td>
<td>Valid Read Data</td>
</tr>
</tbody>
</table>

[0029] The definition of the Valid and Transfer Ack signals in this signaling protocol is shown in Table 4.

### TABLE 4

<table>
<thead>
<tr>
<th>Valid ; Transfer Ack</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ; 0</td>
<td>Valid information is not being transmitted, and the device at the other end of the bus is not ready to receive a transmission.</td>
</tr>
<tr>
<td>0 ; 1</td>
<td>Valid information is not being transmitted, but the device at the other end of the bus is ready to receive a transmission.</td>
</tr>
<tr>
<td>1 ; 0</td>
<td>Valid information is being transmitted, but the device at the other end of the bus is not ready to receive a transmission.</td>
</tr>
<tr>
<td>1 ; 1</td>
<td>Valid information is being transmitted, and the device at the other end of the bus is ready to receive a transmission.</td>
</tr>
</tbody>
</table>

[0030] FIG. 3 is a timing diagram illustrating an example of how the pre-fetch command can be used to reduce the latency of a read operation from the SDRAM. In order to avoid obscuring the concepts of the present invention, the example illustrated in FIG. 3 is limited to read operations only. Moreover, the control signals for these read operations will not be shown. It will be understood by those skilled in the art that the control signals may be transmitted over the transmit channel in parallel to the address information for any given read operation simply by increasing the width of the bus and providing signaling to identify the byte lanes carrying the control signals.
[0031] In this example, the latency for accessing data from an unopened page in the SDRAM is seven clock cycles, while the latency for accessing data from the same page once opened is four clock cycles. The first two read operations in this example are from the same page in the SDRAM and the third read operation is from a different page in the SDRAM. By utilizing the pre-fetch command, the latency of the third read operation may be reduced from seven clock cycles to four clock cycles.

[0032] A System Clock 302 may be used to synchronize communications between the CPU and the SDRAM. The System Clock 302 is shown with twenty-one clock cycles, with each cycle numbered sequentially for ease of explanation.

[0033] The first read operation from an address in a closed page of memory may be initiated by the CPU during the first clock cycle 301. This may be achieved by asserting the Valid signal 304, setting the Type field 306 appropriately, and transmitting the address \( A_0 \) for the first read operation to the SDRAM controller on the Transmit Channel 308. The Type field 306 may be used by the SDRAM controller to determine that the CPU is requesting a read operation from the SDRAM at the address \( A_0 \). At the end of the first clock cycle 301, the CPU detects the assertion of the Transfer Ack signal 310 by the SDRAM controller, and determines that the transmission has been received and accepted.

[0034] In response to the first read operation request, the SDRAM controller may open the appropriate page in the SDRAM and begin to read the data from the address \( A_0 \). Due to the latency of the SDRAM, a seven clock cycle delay may be experienced before the data is available. Once the data is available, the SDRAM controller may assert the Valid signal 312 and set the Type Field 314 to signal a data transmission on the Receive Channel 316. The data transmission in this example requires 4 clock cycles to accommodate the amount of data requested by the CPU and the width of the Receive Channel 316. Consequently, the data may be transmitted over the Receive Channel 316 from the SDRAM controller to the CPU during the eighth, ninth, tenth and eleventh clock cycles 315, 317, 319, and 321. At the end of each of these clock cycles, the SDRAM controller detects the assertion of the Transfer Ack signal 318 by the CPU, and determines that the data transmission has been received and accepted.
During the eighth clock cycle 315, a second read operation may be requested by the CPU from the same page of the SDRAM containing the address A₀ for the first read operation. This may be achieved by asserting the Valid signal 304, setting the Type field 306 appropriately, and transmitting the address A₁ for the second read operation to the SDRAM controller on the Transmit Channel 308. The Type field 306 may be used by the SDRAM controller to determine that the CPU is requesting a read operation from the SDRAM at the address A₁. At the end of the eighth clock cycle 315, the CPU detects the assertion of the Transfer Ack signal 310 by the SDRAM controller, and determines that the transmission has been received and accepted.

Since the page in the SDRAM is already open, the latency of the SDRAM is reduced to four cycles, becoming available for transmission at the beginning of the twelfth clock cycle 323. Once the data is available, the SDRAM controller may assert the Valid signal 312 and set the Type Field 314 to signal a data transmission on the Receive Channel 316. The data transmission in this example also requires 4 clock cycles because of the amount of data requested and the width of the Receive Channel 316, and therefore, may be transmitted over the Receive Channel 316 during the twelfth, thirteenth, fourteenth, and fifteenth clock cycles 323, 325, 327 and 329. At the end of each of these clock cycles, the SDRAM controller detects the assertion of the Transfer Ack signal 318 by the CPU, and determines that the data transmission has been received and accepted.

During the eighth clock cycle 315, the CPU also predicts that it will need data from an address A₂ from a different page in the SDRAM, and as a result, may transmit a pre-fetch command over the Transmit Channel 308 in the ninth clock cycle 317. This may be achieved by asserting the Valid signal 304, setting the Type field 306 to indicate a pre-fetch command, and transmitting the address A₂ to the SDRAM controller on the Transmit Channel 308. In response to the pre-fetch command, the SDRAM controller may begin to open the new page of memory in the SDRAM in parallel with the transfer of data to the CPU over the Receive Channel 316 for the first read operation. At the end of the ninth clock cycle 317, the CPU detects the assertion of the Transfer Ack signal 310 by the SDRAM controller, and determines that the pre-fetch command has been received and accepted.
During the fourteenth clock cycle 327, the CPU determines that it definitely needs the data from the address A₂ in the SDRAM, and thus transmits a read request for this data. This may be achieved by asserting the Valid signal 304, setting the Type field 306 appropriately, and transmitting the address A₂ for the third read operation to the SDRAM controller on the Transmit Channel 308. The Type field 306 may be used by the SDRAM controller to determine that the CPU is requesting a read operation from the SDRAM at the address A₂. At the end of the fourteenth clock cycle 327, the CPU detects the assertion of the Transfer Ack signal 310 by the SDRAM controller, and determines that the transmission has been received and accepted.

Since the SDRAM controller has already opened the page of memory in the SDRAM containing the address A₂, the latency of the SDRAM may be reduced to four cycles, with the data becoming available for transmission at the beginning of the eighteenth clock cycle 335. Once the data is available, the SDRAM controller may assert the Valid signal 312 and set the Type Field 314 to signal a data transmission on the Receive Channel 316. The data transmission in this example requires 4 clock cycles because of the amount of data requested and the width of the Receive Channel 316, and therefore, may be transmitted over the Receive Channel 316 during the eighteenth, nineteenth, twentieth, and twenty-first clock cycles 335, 337, 339 and 341. At the end of each of these clock cycles, the SDRAM controller detects the assertion of the Transfer Ack signal 318 by the CPU, and determines that the data transmission has been received and accepted.

FIG. 4 is a timing diagram illustrating an example of how the pre-fetch command can be used to reduce the latency of a read operation from the NAND Flash memory. In a manner similar to that used to describe pre-fetch commands in connection with the SDRAM, the example illustrated in FIG. 4 is limited to read operations only, and the control signals for those read operations are not shown. It will be understood by those skilled in the art that the control signals may be transmitted over the transmit channel in parallel to the address information for any given read operation simply by increasing the width of the bus and providing signaling to identify the byte lanes carrying the control signals.
In this example, the CPU will read data from the NAND Flash memory, which has a latency of n cycles. This latency is equal to the time required to read a full page of data, typically 512 bytes, from the NAND Flash memory into a large read buffer contained within the NAND Flash controller. The initial latency in most cases can be hundreds of clock cycles. If the CPU were to simply request a read operation to the NAND Flash controller for data from a closed page, the CPU may be stalled while it waited for the new page of data to be loaded into the NAND Flash controller. By issuing the pre-fetch command, the NAND Flash controller can read the data into the read buffer while the CPU continues to execute other instructions. In addition, the receive channel is available for data transmission to the CPU for other read operations. This tends to significantly improve the utilization of the receive channel.

During the first clock cycle 301, the CPU predicts that it will need data from an address A₀ in the NAND Flash memory, and as a result, may transmit a pre-fetch command over the Transmit Channel 308. This may be achieved by asserting the Valid signal 304, setting the Type field 306 to indicate a pre-fetch command, and transmitting the address A₀ for the pre-fetch command to the NAND Flash controller on the Transmit Channel 308. In response to the pre-fetch command, the NAND Flash controller may begin loading the new page of memory from the NAND Flash memory into the read buffer in the NAND Flash controller. At the end of the first clock cycle 301, the CPU detects the assertion of the Transfer Ack signal 310 by the NAND Flash controller, and determines that the pre-fetch command has been received and accepted.

In most cases, the process of loading the new page from the NAND Flash memory into the NAND Flash controller takes many cycles to complete. During this period, the CPU may continue to fetch and execute instructions from its external cache or external memories while it waits for the NAND Flash memory to load the new page of memory into the NAND Flash controller. In clock cycle n 401, where n is greater than or equal to the time required to load the read buffer in the NAND Flash controller, the CPU may request a read operation from the address A₀ by asserting the Valid signal 304, setting the Type field 306 appropriately, and transmitting the address A₀ to the NAND Flash controller on the Transmit Channel 308. At the end of the n clock cycle 401, the CPU detects the assertion of the Transfer Ack signal 310 by the NAND Flash controller, and determines that the transmission has been received and accepted.
Since the read buffer in the NAND Flash memory has already been loading with the page of memory from the NAND Flash controller, the latency may be reduced to the time necessary for the CPU to read the buffer in the NAND Flash controller across the Receive Channel. In this example, the read latency is reduced to four clock cycles. As a result, the data may become available four clock cycles after the CPU requests the read operation. Once the data is available, the NAND Flash controller may assert the Valid signal 312 and set the Type Field 314 to signal a data transmission on the Receive Channel 316. The data transmission in this example is transmitted over the Receive Channel 316 in four clock cycles. At the end of each of these clock cycles, the NAND Flash controller detects the assertion of the Transfer Ack signal 318 by the CPU, and determines that the data transmission has been received and accepted.

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.

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.
[0047] 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:
CLAMS

1. A method of retrieving data from memory, comprising:
   predicting an address of the memory from which data will be needed;
   transmitting to a memory controller a pre-fetch command for the data at
   the predicted address of the memory;
   transmitting to the memory controller a read request for the data at the
   predicted address of the memory; and
   receiving the data at the predicted address of the memory in response to
   the read request.

2. The method of claim 1 wherein the memory controller opens a page in
   the memory in response to the pre-fetch command, the page including the data at the
   predicted address.

3. The method of claim 2 wherein the memory controller reads into a pre-
   fetch buffer at least a portion of the data from the page including the data at the
   predicted address, and wherein the data at the predicted address is received from the
   pre-fetch buffer in response to the read request.

4. The method of claim 3 wherein the size of said at least a portion of the
   data from the page read into the pre-fetch buffer is a predetermined size.

5. The method of claim 3 wherein the size of said at least a portion of the
   data from the page read into the pre-fetch buffer is transmitted to the memory controller
   with the pre-fetch command.

6. The method of claim 1 wherein the memory comprises SDRAM.

7. The method of claim 1 wherein the memory comprises NAND Flash
   memory.
8. The method of claim 1 further comprising transmitting the predicted address to the memory controller with the pre-fetch command and the read request.

9. The method of claim 8 wherein the predicted address is transmitted to the memory controller on a first channel of a bus, and wherein the data is received in response to the read request on a second channel of the bus.

10. The method of claim 9 wherein the pre-fetch command and the read request are transmitted to the memory controller on a common channel.

11. The method of claim 10 wherein the pre-fetch command comprises a first code and the read request comprises a second code different from the first code.

12. The method of claim 9 further comprising transmitting transfer qualifiers associated with the pre-fetch command and the read request to the memory controller on the first channel of the bus.

13. The method of claim 9 further comprising transmitting data to be written to the memory on the first channel of the bus, and transmitting an address to write the data to the memory on the first channel of the bus.

14. A method of retrieving data from memory, comprising:
   predicting an address from an unopened page of the memory from which data will be needed;
   transmitting to a memory controller a pre-fetch command for the data at the predicted address of the memory;
   reading the data at the predicted address into a pre-fetch buffer; and
   discarding the data from the pre-fetch buffer.

15. The method of claim 14 wherein the data is discarded from the pre-fetch buffer in response to the transmission to the memory controller of a predetermined number of the read requests for data from one or more different pages of the memory without a read request for the data at the predicted address in the memory being transmitted to the memory controller.
16. The method of claim 14 wherein the data is discarded from the pre-fetch buffer after a predetermined time without a read request for the data at the predicted address in the memory being transmitted to the memory controller.

17. The method of claim 14 wherein the data is discarded by the pre-fetch buffer in response to a subsequent pre-fetch command transmitted to the memory controller before a read request for the data at the predicted address in the memory is transmitted to the memory controller.

18. A processing system, comprising:
   memory;
   a memory controller; and
   a processor configured to predict an address of the memory from which data will be needed, transmit to the memory controller a pre-fetch command for the data at the predicted address of the memory; and transmit to the memory controller a read request for the data at the predicted address of the memory if the data is needed;
   wherein the memory controller is further configured to deliver the data at the predicted address of the memory to the processor in response to the read request.

19. The processing system of claim 18 wherein the memory controller is further configured to open a page in the memory in response to the pre-fetch command, the page including the data at the predicted address.

20. The processing system of claim 19 further comprising a pre-fetch buffer; and wherein the memory controller is further configured to read into the pre-fetch buffer at least a portion of the data from the page including the data at the predicted address, and to deliver the data at the predicted address to the processor from the pre-fetch buffer in response to the read request.

21. The processing system of claim 20 wherein the size of said at least a portion of the data from the page read into the pre-fetch buffer is a predetermined size.
22. The processing system of claim 20 wherein the processor is further configured to transmit to the pre-fetch buffer the size of said at least a portion of the data from the page to be read into the pre-fetch buffer.

23. The processing system of claim 20 wherein the pre-fetch buffer is located in the memory controller.

24. The processing system of claim 20 wherein the memory controller is further configured to discard the data in the pre-fetch buffer if the memory controller receives from the processor a predetermined number of read requests for data from one or more different pages of the memory after the pre-fetch command without receiving the read request for the data at the predicted address of the memory.

25. The processing system of claim 18 wherein the memory comprises SDRAM.

26. The processing system of claim 18 wherein the memory comprises NAND Flash memory.

27. The processing system of claim 18 wherein the processor is further configured to transmit the predicted address to the memory controller with the pre-fetch command and the read request.

28. The processing system of claim 27 further comprising a bus coupling the processor to the memory controller, the bus having first and second channels, wherein the processor is further configured to transmit the predicted address to the memory controller on a first channel of a bus, and wherein the memory controller is further configured to transmit the data to the processor on the second channel of the bus.

29. The processing system of claim 28 further comprising a common channel between the processor and the memory controller, and wherein processor is further configured to transmit the pre-fetch command and the read request to the memory controller on the common channel.
30. The processing system of claim 29 wherein the processor is further configured to generate a first code comprising the pre-fetch command and a second code comprising the read request, the first code being different from the second code.

31. The processing system of claim 28 wherein the processor is further configured to transmit transfer qualifiers associated with the pre-fetch command and the read request to the memory controller on the first channel of the bus.

32. The processing system of claim 28 wherein the processor is further configured to transmit data to be written to the memory on the first channel of the bus, and transmit an address to write the data to the memory on the first channel of the bus.

33. A processing system, comprising:
   memory;
   a memory controller;
   means for predicting an address of the memory from which data will be needed;
   means for transmitting to the memory controller a pre-fetch command for the data at the predicted address of the memory;
   means for transmitting to the memory controller a read request for the data at the predicted address of the memory if the data is needed; and
   receiving means for receiving the data;
   wherein the memory controller is further configured to deliver the data at the predicted address of the memory to the receiving means in response to the read request.
# International Search Report

## A. Classification of Subject Matter

G06F13/02  G06F9/38

According to International Patent Classification (IPC) or to both national classification and IPC

## B. Fields Searched

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

## C. Documents Considered to be Relevant

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>figures 1,3 paragraph '0012'! - paragraph '0013'! paragraph '0017'! paragraph '0020'! - paragraph '0021'! paragraph '0026'!</td>
<td>3-5, 14-17, 20-24</td>
</tr>
<tr>
<td></td>
<td>figures 2,3 paragraph '0023'! - paragraph '0025'!</td>
<td></td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C

See patent family annex

- * Special categories of cited documents
  - **A** document defining the general state of the art which is not considered to be of particular relevance
  - **E** earlier document but published on or after the international filing date
  - **L** document which may throw doubts on priority claims(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
  - **O** document referring to an oral disclosure, use, exhibition or other means
  - **P** document published prior to the international filing date but later than the priority date claimed

- **T** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

Date of the actual completion of the international search: 3 February 2006

Date of mailing of the international search report: 10/02/2006

Name and mailing address of the ISA/ European Patent Office, P. B. 5618 Patentlaan 2 NL – 2280 HV RIJKWIJK Tel. (+31-70) 340-2940, Tx. 31 651 epo nl, Fax (+31-70) 340-3016

Authorized Officer: Filip, L
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>DE 10393803 T5</td>
<td>20-10-2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2004061685 A2</td>
<td>22-07-2004</td>
</tr>
<tr>
<td>US 2004123042 A1</td>
<td>24-06-2004</td>
<td>CN 1508694 A</td>
<td>30-06-2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2004199684 A</td>
<td>15-07-2004</td>
</tr>
</tbody>
</table>