A PCI Express system and a method of transitioning link state thereof are provided. The PCI Express system has an upstream device, a downstream device and a link. The upstream device and the downstream device transmit data packets to both via the link, but when the link is in a first link state, data packet transmission is forbidden. In the beginning, the link is in a second link state and data packet transmission is normal. The upstream device transmits a data packet via the link to the downstream device. A time period is counted when receiving the data packet. The downstream device asserts an acknowledge packet to the upstream device to response the data packet. After the timer is expired, the link is transited to the first link state.
FIG. 1
Start

Transit the link to a second link state

The upstream device asserts a data packet through the link to the downstream device

The downstream device counts a time period after receiving the data packet

The downstream device asserts an acknowledge packet to the upstream device

The downstream device stops counting and asserts a PM_Enter_L1 packet to the upstream device

The upstream device asserts a PM_Request_Ack packet

The link transits to a first link state

End

FIG. 2
PCI EXPRESS SYSTEM AND METHOD OF TRANSITIONING LINK STATE THEREOF

[0001] This application claims the benefit of U.S. provisional application Ser. No. 60/683,313, filed May 23, 2005, and the benefit of Taiwan application Serial No. 95107634, filed Mar. 7, 2006, the subject matters of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention
[0003] This invention relates to an apparatus and a method of transitioning state, more specifically to an apparatus and a method of transitioning link state for PCI Express.
[0004] 2. Description of the Related Art
[0005] As the wheel of time moves on endlessly, the Peripheral Device Interconnect (PCI) mainly for the use of Personal Computers faces a problem that the processors and input/output devices in the future generation need higher transmission bandwidth. Such transmission bandwidth gradually exceeds the working range of PCI. The industry thus develops a new generation PCI Express to be the standard regional input/output bus for all kinds of future processing platforms. The most significant feature is the improvement of transmission efficiency. Single direction transmission rate can reach up to 2.5 GHz. Furthermore, the transmission rate can be increased by expanding the number of lanes. For example, using 4 lanes can speed up the transmission rate to 4 times.

[0006] Advanced Configuration and Power Interface (ACPI) defines the device states (D-states) under every situation. And PCI Express further defines the link states (L-states) between devices. Each link state has a corresponding relation to a device state.

[0007] Device state D0 (Full-On) represents that the device is under normal operation. When the device is in the device state D0, the link between devices could be in the link states L0, L0s, or L1.

[0008] Device states D1 and D2 are not specifically defined in ACPI. In general, device state D2 is more power saving than device state D0 and D1 when the number of devices is few. Device state D1 is more power saving than device state D2 when the number of devices is relative large. Device states D1 and D2 could be corresponding to the link state L1.

[0009] Device state D3 (Off), including D3cold and D3hot states, represents the shut-down state. When the device is in D3cold state, the main power does not supply to the device. When the device is in D3hot state, the main power is supplied to the device. When the device is in D3cold state, the link between the devices could be under the link state L2 if there has an auxiliary power to supply power. If there is no auxiliary power, the link could be under the link power state L3. Device state D3hot corresponds to the link state L1.

[0010] Link state L0 is the state when the link between devices is in normal operation. Link state L0s can decrease the power consumption when the link has short idle periods during data transmission.

[0011] When the link is in the link state L1, the devices are in pause state with no request. This will decrease the demand of link power between devices. At the time there is no trigger of time pulse signal, and the Phase Lock Loop (PPL) will pause for any function.

[0012] Link state L2 and L3 are shut-down states. The difference between L2 and L3 is that the link state L2 is supplied by an auxiliary power, but the link state L3 has no auxiliary power.

[0013] However, when the link is in a power saving link state, such as the link state L1, the link has to transit to a normal link state so that data packet can be transmitted by the upstream device to the downstream device. After the transmission ends, the link will transit back to power saving link state L1. During the process, transmission error of data packet easily causes the link states transitioning repeatedly. More seriously it may cause the system to shut down.

SUMMARY OF THE INVENTION

[0014] To address the above-detailed deficiencies, the present invention provides an apparatus of PCI Express system and a method of transitioning link state thereof that avoids the transmission error of data packet.

[0015] The present invention provides a method of PCI Express transitioning link state for a link between an upstream device and a downstream device. The upstream device and the downstream device transmit data to both through the link. Data transmission is forbidden when the link is in a first link state, and the downstream device is in the abnormal operation state. The method includes: transiting the link to a second link state in which data packet transmission is normal. Then the upstream device transmits a data packet to the downstream device through the link. Later, a time period is counted when the downstream device receives the data packet. Then the downstream device asserts an acknowledge packet to the upstream device for responding the data packet. When the time period is expired, and the downstream device asserts a power entry packet PM_Enter_L1 to the upstream device, and the link is then transited back to the first link state.

[0016] The present invention also provides a PCI Express system including an upstream device, a downstream device, and a link. The downstream device is in a first device state. The upstream device and the downstream device transmit data packets to both through the link. When the link is in a first link state, data transmission is forbidden. Thus the link transits to a second link state to normally transmit data packets. The upstream device transmits a data packet to the downstream device through the link. Then a time period is counted when the data packet is received, and the downstream device asserts an acknowledge packet to the upstream device for responding the data packet. When the time period is expired, the downstream device asserts a power entry packet PM_Enter_L1 to the upstream device, and the link is transited back to the first link state.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] These and other objects, features, and advantages of the present invention will become better understood with regard to the following description, and accompanying drawings where:

[0018] FIG. 1 is a schematic diagram of PCI Express link and layers.
FIG. 2 is a flowchart of the method of PCI Express transitioning link state.

FIG. 3 is a relative waveform of the link transition between first link state and second link state.

DETAILED DESCRIPTION OF THE INVENTION

Referring to FIG. 1, a schematic diagram of PCIe system 100 is shown. The PCIe system 100 of the present invention includes an upstream device 110, a downstream device 120, and a link 130 connected between the upstream device 110 and the downstream device 120.

The upstream device 110 includes: a Transaction Layer (TL) 111, a Data Link Layer (DLL) 112, and a Physical Layer (PHY) 113. And the downstream device 120 also includes: a Transaction Layer 121, a Data Link Layer 122, and a Physical Layer 123.

The upstream device can be, for example, a Root Complex (RC), and the downstream device as well can be an End Point (EP).

The Transaction Layers 111 and 121 respectively generate data packets to the Data Link Layers 112 and 122. The Transaction Layers 111 and 121 also respectively receive data packets from the Data Link Layers 112 and 122. Meanwhile the Transaction Layers 111 and 121 also manage the flow control between devices. Data packets generated by or received from the Transaction Layers 111 and 121 are regarded as Transaction Layer Packets (TLPs).

The Data Link Layers 112 is in charge of data packets transmission between the Physical Layers 113 and the Transaction Layer 111; similarly the Data Link Layers 122 is in charge of data packets transmission between the Physical Layers 123 and the Transaction Layer 121.

After receiving data packets, the Data Link Layers 112 and 122 respectively transmit TLPS to the corresponding Transaction Layers 112 and 121. The Data Link Layers 112 and 122 also respectively receive TLPS from the corresponding Transaction Layers 111 and 121, and then respectively output the data packets to the corresponding Physical Layer 113 and 123. When the Data Link Layers 112 and 122 transmit the data packets, error detection is performed for stably transmit the data packs.

Data packets transmitted between the Data Link Layer 112 and the Physical Layer 113 are or between the Data Link Layer 122 and the Physical Layer 123 are regarded as Data Link Layer Packets (DLLPs).

The Physical Layers 113 and 123 are in charge of data packet transmission via the link 130 between the upstream devices 110 and the downstream device 120.

The data packets from the downstream device 120 and received by Physical Layer 113 are transformed into DLLPs format and then transmitted to the Data Link Layer 112. The DLLPs from the Data Link Layer 112 are received by the Physical layer 113 and then transmitted to the downstream device 120 through the link 130. In the same manner, the data packets from the upstream device 110 and received by the Physical Layer 113 are transformed DLLPs format and then transmitted to the Data Link Layer 122. The DLLPs from the Data Link Layer 122 are received by the Physical Layer 123 and then transmitted to the upstream device 110 through the link 130.

Referring to FIG. 2, a flowchart of PCI Express transitioning state is shown. The method is applied to the link 130 between the upstream device 110 and the downstream device 120.

The present invention provides an apparatus and a method for transiting link state and transmitting data when under abnormal working state. That is to say, the downstream device 120 is in a non-first device state (the first device state for example is D0 state). Assume the initial link state of the link 130 is in a first link state (ex. L1 state), data transmission is forbidden.

The Data packets can not be transmitted through the link 130 in the first link state. As a result the link state of the link 130 has to transit to a second link state (ex. L0 state) so that data transmission can be normal (step 21). Later as shown in step 22, the upstream device 110 asserts a data packet (ex. TLP) to the downstream device 120 through the link 130. The data packet is a command for changing or reading the device state of the downstream device 120. In step 23, a time period is counted when the downstream device 220 receives the data packet. In step 24, the downstream device 120 asserts an acknowledge packet to the upstream device 10 for responding the data packet. In step 25, when the time period is expired, the downstream device 120 asserts a power entry packet, PM_Enter_L1 (ex. DLLP), to the upstream device 110. In step 26, the upstream device 110 asserts a power request acknowledge packet, PM_Request_Ack, to the downstream device 120. In step 27, after receiving the PM_Request_Ack, the link 130 is transited to the first link state (ex. L0 state).

In step 23 to step 25, the time period could be: immediate time-out, (1 ClgW+10 cycles), (32 QW TLP+1 ClgW+10 cycles), (2*32 QW TLP+4 ClgW+10 cycles).

Wherein, the ClgW is one data packet transmission period.

The CfgW is, for example a transmission period of a TLP transmitted from the Transaction Layer 111 of the upstream device 110 to the downstream device 120. 10 cycles represents a time period for the downstream device to process a TLP. QW TLP represents the QW length of a TLP (ex: 1 QW TLP means that the TLP length is 1 QW, and 1 QW is 8 bytes. Consequently 32 QW TLP is a TLP of 256 Bytes).

The time period counted to allow the acknowledge packet is received earlier than the power entry packet, PM_Enter_L1. This ensures that the acknowledge packet is received and the link 130 is later transited to the first link state (ex. L1 state) by receiving the power entry packet, PM_Enter_L1.

FIG. 3 is the relative waveform of the link 130 transitioning between the first link state (ex. L1 state) and the second link state (ex. L0 state).

Assume that the initial state of the downstream device 120 is in the first device state (ex. D0 state), and the link state is in L0 state. After idle for a while, the downstream device 120 is transited to a second device state (ex. D1 state), which is the non-first device state, at time point 11. At t1, the downstream device 120 also asserts a power entry...
packet PM_Enter_L1 to the upstream device 110. At time point t2, the downstream device 220 receives the PM_Request_Ack and then the link 130 is transitioned to the link state L1.

[0039] In the following paragraphs, the process described below is referred to the flowchart in FIG. 2. After time point t2, the downstream device 120 maintains in the second device state D1, the link 130 is in the link state L1 in which data packet transmission is forbidden.

[0040] If a data packet transmission is needed, the link 130 is transitioned from link state L1 to the link state L0 for allowing data packet transmission. At time point t3, the data packet is transmitted. When the downstream device 120 receives the data packet, a time period t4 is counted and an acknowledge packet is asserted after processing the data packet. Later at time point t4, when the time period is expired, the downstream device 120 asserts a power entry packet PM_Enter_L1 data packet. At time point t5, after the downstream device 120 receives a power request acknowledge packet, PM_Request_Ack, the link 130 is transitioned to the link state L1.

[0041] The PCI Express system and method of transitioning link state thereof revealed in the present invention has the advantage of transitioning the link state from which data packet transmission is forbidden to which is allowed. And data transmission error during the transition of the link state can be avoided. Furthermore, the present invention avoids system shut-down causing by repeated link state transitioning, and satisfies the power saving purpose of the prior art.

[0042] Although the present invention has been described in considerable detail. Those skilled in the art should appreciate that they can readily use the disclosed conception and specific embodiments as a basis for designing or modifying other structures for carrying out the same purposes of the present invention without departing from the spirit and scope of the invention as defined by the appended claims.

What is claimed is:

1. A method of transitioning link state, for a link connected between an upstream device and a downstream device, wherein an initial link state of the link is in a first link state, and the downstream device is in an abnormal operation state, the method comprising:

   transitioning the link state from the first link state to a second link state;

   asserting a data packet by the upstream device;

   counting a time period when the downstream device receives the data packet;

   asserting an acknowledge packet by the downstream device to the upstream device as a response to the data packet;

   asserting a power entering packet by the downstream device to the upstream device when the time period is expired; and

   asserting a power request acknowledge packet by the upstream device to respond to the power entering packet so as to transit the link from the second link state to the first link state.

2. The method of claim 1 wherein the data packet is a Transaction Layer Packet (TLP); the power entering packet and the power request acknowledge packet are respectively Data Link Layer Packets (DLPs).

3. The method of claim 1 wherein the time period is not less than one data packet transmission period plus 10 cycles.

4. The method of claim 3 wherein the time period further comprises 32 QW TLPs transmission period.

5. The method of claim 1 wherein the time period ensures the upstream device receives the acknowledge packet earlier than the power entering packet.

6. The method of claim 1 wherein the first link state is the link state L0.

7. The method of claim 1 wherein the second link state is the link state L1.

8. The method of claim 1 wherein the upstream device is a Root Complex and the downstream device is an End Point.

9. The method of claim 1 wherein the method is applied to a PCI Express link.

10. A data transmission system comprising:

     an upstream device, for asserting a data packet;

     a downstream device, a time period is counted when the downstream device receives the data packet from the upstream device, at the same time, the downstream device asserts an acknowledge packet to the upstream device; and

     a link, connected between the upstream device and the downstream device for data transmission;

wherein the downstream device asserts a power entering packet to the upstream device when the time period is expired and the time period ensures the acknowledge packet is received earlier than power entering packet.

11. The system of claim 10 wherein an initial link state of the link is at L1 state.

12. The system of claim 11 wherein the link is transitioned to the link state L0 before transmitting the data packet.

13. The system of claim 12 wherein the upstream device asserts a power request acknowledge packet to the downstream device as a response to the power entering packet, and the link is transitioned to L1 state.

14. The system of claim 10 wherein the initial link state of the downstream device is in an abnormal operation state.

15. The system of claim 10 wherein the data packet is utilized to change or read the state of the downstream device.

16. The system of claim 10 wherein the time period is not less than one data packet transmission period plus 10 cycles.

17. The system of claim 16 wherein the time period further comprises 32 QW TLPs transmission period.

18. The system of claim 10 wherein the data transmission system is a PCI Express system.

19. A method of link data transmission, for a link connected between an upstream device and a downstream device; wherein a initial link state of the link is unable to transmit data packets; the downstream device is in an abnormal operation state, the method comprising:

   asserting a data packet by the upstream device;

   counting a time period when the downstream device receives the data packet;
asserting an acknowledge packet by the downstream device to the upstream device as a response to the data packet; and

asserting a power entry packet by the downstream device to the upstream device when the time period is expired;

wherein the time period ensures the acknowledge packet is received earlier than the power entry packet.

20. The method of claim 19 wherein the method further comprising:

transiting the link state to a state able to transmit data packet.

21. The method of claim 20 wherein the method further comprising:

asserting a power request acknowledge packet responding to the power entering packet by the upstream device so as to transit the link state to a state unable to transmit data packet.

22. The method of claim 19 wherein the time period is not less than one data packet transmission period plus 10 cycles.

23. The method of claim 22 wherein the time period further comprises 32 QW TLPS transmission period.

24. The method of claim 19 wherein the method is applied to a PCI Express system.

* * * * *