PCI

By Richard Solomon

PCI Express (PCIe) has been known for its high performance since the early 2000s, but many designers may not be aware of the power efficiencies that its current specifications can attain. To understand PCIe's low-power modes, it is helpful to consider history—including PCIe's predecessor parallel bus, PCI.

HISTORY OF POWER MANAGEMENT

In the 1990s, power management mostly meant, "turn off devices that aren't being used," and many servers and data center applications didn't even bother with that. To improve the granularity of device power management, PCI added the concept of device power states D0, D1, D2, and D3 in the late 1990s. Devices in state D0 were operating normally, while the ones in D3 actually were off, or at least in a state in which their power could be removed. Devices in D1 or D2 were specified as not in use, but had only vague indications of what power-saving methods they might be using.

This model was largely consistent with the idea of laptop-like usage. In such cases, a user might close the laptop and walk away. The system would then turn off as many devices as possible but leave them active enough that the operating system would only need to do minimal work to resume. After a period of inactivity, the laptop would turn the remaining devices completely off and switch to the save state to later resume normal operations.

PCI Express has been known for its high performance since the early 2000s, but many designers may not be aware of the power efficiencies.

EXTENSION TO PCIe LINK STATES

The PCIe architecture was built based on the traditional concepts with link power states L0, L0s, L1, L2, and L3, which roughly correspond to the D states. L0 is a link operating normally, and L3 is one that is off. The intermediate states were better defined in PCIe: L2 is a link state identical to L3, but its power has not (yet) been removed, and L1 is a link state in which no data are being transferred, so key portions of the PCIe transceiver logic can be turned off. The "L0s" state is one in which data may be in the process of being transferred in one direction but not the other, so each of the two devices on a link can idle its transmitter independently.

LIMITATIONS AND TRADEOFFS

The L3 state provides significant power savings because all logic associated with the PCIe link can be turned off. L2 allows for nearly as much power savings as L3, although it requires some logic to remain alive to process the exit and reset the sequence required to resume L0 and normal operations if the power was not removed. Exiting L2 is a lengthy process that requires reconfiguring the devices and allowing an amount of time that is comparable with what is needed to power on tasks such as relocking phase-locked loops (PLLs). The L1 and L0s states provide returns that revert to normal operations rapidly and use in-band signaling to do so. In-band signaling means the high-speed circuits in the PCIe transceivers need to stay alive, therefore reducing the possible power savings.

MOBILE NEEDS DICTATE GREATER SAVINGS

As the industry moved to tablets and other handheld/mobile devices, the focus of power management shifted from gross "on" versus "off" to finer-grained moment-by-moment switching. For these applications, L2 resume latencies were too high to for this rapid and frequent state switching, while L1 power savings were too low to meet the device power-consumption goals. An innovative solution to this conundrum came in the form of the L1 PM substates with clock request signal (CLKREQ#) engineering change notice, commonly referred to as *L1 substates*.

PCIe L1 SUBSTATES EXPLAINED

The fundamental idea behind L1 substates is to use something other than the high-speed logic inside the PCIe transceivers to wake the devices. This is done by adding additional functionality to an existing PCIe pin (CLKREQ#) to provide a very simple signaling protocol, allowing the PCIe

Digital Object Identifier 10.1109/MCE.2019.2892267
Date of publication: 8 April 2019

Table 1. The port circuit power states in L1 substates.

Port Circuit Power On/Off			
Substate	**PLL**	**Receive/Transmit**	**Common-Mode Keeper**
L1.0	On	Off/Idle	On
L1.0 + CLKREQ	Off	Off/Idle	On
L1.1	Off	Off	On
L1.2	Off	Off	Off

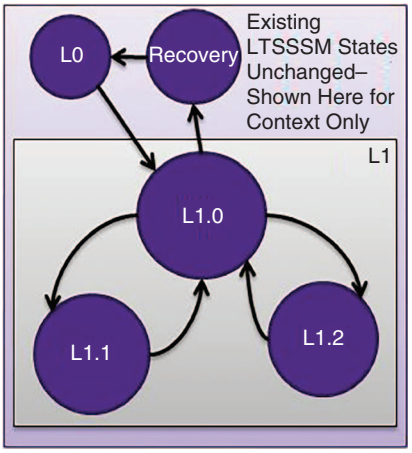

FIGURE 1. The L1 substates state machine. LTSSM: link training and status state machine.

transceivers to turn off their high-speed circuits and rely on the new signaling to wake them again. In fact, two of these new substates, L1.1 and L1.2, were defined, providing their own power versus exit latency tradeoff choices (see Figure 1).

L1 SUBSTATE RESULTS

Both L1.1 and L1.2 permit the PCIe transceivers to turn off their PLLs along with their receivers and transmitters, while L1.2 even allows turning off the common mode keeper circuits (see Table 1). The results are very significant; with modern silicon processes, a representative PCIe4.0 ×4 PHY (four transceivers plus the related digital logic for four lanes) running at the full 16 GT/s data rate in L0 consumes in the range of 400–500 mW. Utilizing L1.1, the same PHY's power consumption drops by a factor of around 20× to consume only 20–30 mW. Accepting the slightly longer exit latency of L1.2 permits power consumption to fall by another 10× to 2–3 mW.

ADDITIONAL ARCHITECTURAL POWER SAVINGS

While not as easily quantified, PCIe's load/store architecture with bus mastering allows overall system power optimization. Intelligent bus-mastering PCIe-based I/O adapters allow the system's CPU to enter low-power states while the adapter moves data to and from main memory. Emerging PCIe-based computing accelerators for tasks such as artificial intelligence and machine learning allow system architects to better optimize the system's power by leveraging optimized task-specific hardware rather than overworking general-purpose CPUs.

Additionally, the move to push/push software interfaces such as Nonvolatile Memory Express is helping with the optimization. By reducing or eliminating power-consuming CPU polling loops, these combined systems can allow the main CPU to go to sleep for the most part, while intelligent PCIe-based adapters access system memory via bus mastering and even communicate on a peer-to-peer basis with PCIe-based accelerators.

CONCLUSION

The low exit latencies and increased power savings of the L1 substates feature, combined with PCIe's bus-mastering load/store architecture and upcoming 32 GT/s speed, can provide both the performance and low power needed for storage, accelerated computing, networking, and other high-speed devices, both current and futuristic.

ABOUT THE AUTHOR

Richard Solomon (rsolomon@synopsys.com) serves as the vice chair of the PCI-SIG and is the technical marketing manager for the Synopsys DesignWare PCI Express Controller IP.