"SATA-IV" Proposal: A Flexible Topology for PCIe 3.0

http://thessdreview.com/Forums/ssd-discussion/1555.htm

17-Dec-11
10:45 AM

This archives our recent correspondence with the PCI Special Interest Group concerning our "SATA-IV" Proposal:

Sent: Thursday, November 17, 2011 9:30 AM
To: pr@pcisig.com
Subject: PCI-SIG Contact Request

We push the envelope onwards and upwards with our "SATA-IV" Proposal: A Flexible Topology for PCIe 3.0

12Gbps Is Here!!!! So is that SATA 6.0 or 4.0?
http://thessdreview.com/Forums/raid-enterprise/1377-2.htm#post13318

I personally believe this [latter] approach is going down the wrong path, after taking a wrong turn at a junction several months past.

Here's why:

PCIe 3.0 specs introduce 2 very significant changes:

8 GHz transmission clocks
-and-
128b/130b "jumbo frames"

This permits each bus channel to transmit 8G/8b = 1.0 GB per second in each direction, per x1 PCIe lane -- at the bus level.

I see no reason why this PCIe 3.0 "topology" cannot be extended outwards over flexible SATA & SAS cables:

I haven't done the engineering analysis, but I don't expect that too much change will be needed (if any) for flexible SATA and SAS cables to handle 8G transmission rates.

Therefore, I believe that the solid-state industry should be "thinking forward" to SSDs with an 8G "option" and a corresponding "option" in add-on controllers, like those being discussed at this excellent website -- *i.e.* either a setting in the card's Option ROM, and/or a physical jumper.

This combination of features will be particularly interesting when RAID arrays of SSDs with very large DRAM caches are moving data in both directions at 1.0 GB/second *e.g.* Plextor's SSDs w/ 256MB DDR3 cache!

Let's call it "SATA-IV" (in logical progression), but that nomenclature also needs to accommodate the "option" to change the transmission protocol so as to exploit the obvious advantages that result from 16 bytes per "frame" with only 1 extra start bit and only 1 extra stop bit aka "128b/130b".

This is already a feature of the PCIe 3.0 spec published a year ago, and should be available in all PCIe 3.0 chipsets now coming to market *e.g.* cf. motherboards w/ LGA 2011 sockets.

I welcome your thoughts, reactions, criticisms, etc.

MRFS
(User I.D. abbreviating “Memory Resident File Systems”)

p.s. See also some of my early writing on this point, attached and here:

SSD Technology Improvement Could Match DDR2 Performance
http://benchmarkreviews.com/index.php?option=com_content&task=view&id=10291&Itemid=22

to: "mrfsys@gmail.com" <mrfsys@gmail.com>
cc: "pr@pcisig.com" <pr@pcisig.com>
date: Thu, Nov 17, 2011 at 3:52 PM

Hi Paul,

Thank you for reaching out to the PCI-SIG and for passing this along.

We will share this with PCI-SIG representatives.

Best,
Claire Castellanos
Nereus for PCI-SIG

You're very welcome, Claire.

p.s. One truly fabulous feature of the PCIe 3.0 specification is the simplicity that obtains for capacity planning, when raw bandwidth is an even 1.0 Gigabyte per second for each x1 PCIe 3.0 lane.

What I have been seeing, *in my mind's eye*, is a logical extension of this elegant topology, out beyond the "fixed" traces of motherboard chipsets, to any number of flexible SATA and/or SAS cables that also support this very same raw bandwidth: 1.0 GB/second, using the same performance characteristics: 8G/8b

Another neat advantage of this "flexible topology" is the fact that it can be implemented with existing PCIe 3.0 chipsets, *even though* their native SATA protocols are presently limited to the old 8b/10b frame (8 data bits + 1 start bit + 1 stop bit = 10 bits per byte):

A builder need only opt for a compatible add-on controller that supports both the 8G transmission rate **-and-** the 128b/130b "jumbo frame" -- either in an Option ROM on-board the add-on card, and/or via hardware jumpers.

Now we are looking at a very extensible and very scalable property of PCIe 3.0 systems: given the enormous raw bandwidth that is offered by PCIe 3.0 x16 slots -- 16 GB/second in each direction (1GB/lane) -- then the builder need only select the appropriate add-on controller, and then decide on the appropriate number of those add-on controllers, to create a very fast

storage subsystem, particularly when it is populated with compatible "8G" SSDs that also understand the new 128b/130b "jumbo frame".

Lastly, if this proposal is timely implemented by chipset vendors, then it is only logical that both the 8G and the 128b/130b jumbo frame should be made options with native support in upcoming PCIe chipsets, perhaps the very next generation aka PCIe 4.0.

Clearly, these are features that should be standardized worldwide, so that the entire IT industry can share in the obvious performance benefits.

Thank you very much for your interest in this Proposal.

See also:

SSD Technology Improvement Could Match DDR2 Performance
http://benchmarkreviews.com/index.php?option=com_content&task=view&id=10291&Itemid=22

Visible Computer Futures
http://benchmarkreviews.com/index.php?option=com_content&task=view&id=12508&Itemid=21

Overclocking Data Storage Subsystems: Variable Channel Bandwidth
(hard copy attached)

--
Sincerely yours,
/s/ Paul A. Mitchell, B.A., M.S., Instructor,
Inventor and Systems Development Consultant

p.s. The version of "Overclocking Data Storage Subsystems" re-published at Benchmark Reviews: Performance Computer Hardware Tests
did NOT include any of the Figures. Those Figures are available here:

Overclocking Data Storage Subsystems (hard copy attached)