Making the right links

RapidIO is being targeted at high performance computing applications as latency becomes a critical factor. By Graham Pitcher.

The extent to which ‘what goes around, comes around’ applies to technology. Despite the impression that companies are forging a future based on brand new ideas, that’s not always the case.

And so it is with RapidIO, the packet switched interconnect technology developed in the early 2000s for use in high performance computing (HPC) applications. Despite this ambition, RapidIO found most success in the wireless infrastructure sector. But the wheel is turning full circle as RapidIO begins to make inroads into today’s HPC applications – and data centres in particular.

Dev Paul, principal product manager with IDT, said the military and aerospace market was one of the first to actively embrace RapidIO. "Developers were looking for a technology to support processor to processor communications that met energy, space and latency constraints. Some of these companies were doing things like putting radar systems into the nose cones of aircraft and were concerned with things like power, space and cooling requirements."

At the time, developers were likely to have been using Ethernet or PCI Express (PCIe), "In fact," Paul claimed, "RapidIO was created to take business from both technologies. While Ethernet could scale to any topology, it was originally designed for LANs and needed a processor at each end of the link to terminate the protocol. PCIe, meanwhile, was all done in hardware. While it had low latency, it was limited." Since then, there have been several attempts to scale PCIe, but it hasn’t really been scalable to large multicore systems. If you needed to add boards, it was a complex task.

“RapidIO took these features and combined them into one protocol.”

Paul’s experience with RapidIO dates back to the mid 2000s, when he worked with Canadian company Tundra Semiconductor, which was acquired by IDT in 2009.

“Tundra started talking to communications companies who were using DSPs with memory interfaces. We identified wireless base stations as an application that would benefit from RapidIO and worked with Texas Instruments to put RapidIO into an IP core.”

A boost in RapidIO’s fortunes came in 2008, from two related developments. In one, Ericsson announced that it was using Tundra’s RapidIO technology in its LTE proof of concept system. “That started an acceleration in the use of RapidIO in the wireless market,” Paul noted. The other boost came when China rolled out mobile communications technology that took advantage of RapidIO’s capabilities.

“Since then,” Paul said, “IDT and Tundra captured all 3G and LTE development programmes. All of the top 10 vendors were applying the technology for connectivity on signal processing boards, as well as for extending systems.”

In the first generation of RapidIO, data was handled at 10Gbit/s. The second generation pushed that to 20Gbit/s, but the latest devices
handle 40Gbit/s. But work is now underway to increase the capacity to 100Gbit/s.

Looking to lay the foundations for future systems, IDT has launched a RapidIO interface product portfolio spanning data rates from 40 to 100Gbit/s. The products are intended to lower latency whilst improving bandwidth for HPC, wireless, analytics and embedded applications.

The first product on the road map is introducing a 40Gbit/s RapidIO 10xN IP core for application in ASICs, CPUs, DSPs, GPUs and FPGAs. The core is suitable for designs targeted at process nodes ranging from 45nm to 16nm.

According to IDT, the core is designed to serve as a basic building block and gateway between processing elements to create low latency, scalable, energy efficient computing. The lower latency and higher speed interface products will be essential, it claims, if the massive volumes of data generated by today’s electronics are to be handled efficiently.

“This initial 40Gbit/s product provides an exciting interconnect path for the Open Compute Project (OCP), where we are innovating from the system level down to the silicon level,” said OCP executive director Cole Crawford. “IDT’s RapidIO 10xN solutions support the OCP’s goal for open, interoperable, vendor agnostic solutions that take us from today’s custom vendor based HPC solutions towards open exascale computing.”

The core will also be used in a range of RapidIO based switching, bridging, processor and memory controller products currently on various drawing boards, including the switching collaboration between IDT and eSilicon to develop 100ns latency RapidIO switches operating at 40Gbit/s per port.

In Paul’s view, there will be no changes to the RapidIO standard as data rates move towards 100Gbit/s.

“With the exception,” he added, “that we will need to build in a faster SERDES.”

Having a core that can be used on a range of process technologies allows customers to optimise the SoC to the application. Paul said the choice of process technology will allow factors such as power consumption and clock rate to be tuned. “And the architecture scales to 100Gbit/s,” he continued, “which means it should be ‘future proof’.”

RapidIO is still finding plenty of business in the wireless sector, particularly as data rates increase. “The amount of time taken for all processing operations between the radio and backhaul has shrunk by a factor of five,” Paul claimed.

“Operators need to get data off the air at the base station, then off the analogue front end, digitise it using CPRI and so on. All that has to happen more quickly and with higher data rates.”

The growth in demand for HPC is also proving beneficial to RapidIO vendors. “In essence,” Paul observed, “RapidIO has always been about HPC, but applications were sometimes limited by latency. For example, when you’re performing a 2D FFT in real time with edge detection and filtering, there’s a lot of decision making involved.” That requires a lot of processing and, Paul added, there isn’t a single or a multicore processor that can handle such loads. “So it’s crunched by multiple devices,” he said, implying the need for high speed connectivity.

Data centres are also a major target (see fig 2). “Even as little as two years ago, search engine companies didn’t care about latency. Now, they are doing things like deep learning and real time analytics; all tasks where latency is critical – and Gigabit Ethernet has run out of steam,” said Paul.

“HPC and analytics customers are facing the same latency, power and space constrains as embedded customers, while trying to scale to large systems with thousands of nodes that stretch beyond petascale to exascale computing,” added Sean Fan, general manager of IDT’s Interface and Connectivity Division.

To support these applications, RapidIO silicon solutions have been taken from the wireless embedded sector and repackaged in a format which allows the necessary systems to be put together.