MOST SIGNIFICANT BITS

INTEL ADAPTS LARRABEE FOR HPC

By Tom Halfhill {6/14/10-02}

When you've got a lemon, make lemonade. Intel's troubled manycore-processor project is steering away from discrete 3D graphics in favor of high-performance computing (HPC), mainly for scientific and engineering applications. It's a wise maneuver that will salvage Intel's investment in the Larrabee project, and the new direction is better suited to Intel's experience and expertise. But it won't avoid a collision with Nvidia, which is surging into the same market.

Intel revealed new details about its HPC strategy at the recent Super Computing Conference in Hamburg, Germany. The x86-based family of GPUs code-named Larrabee will spawn a new family of manycore processors code-named Knights. Both Larrabee and Knights can integrate dozens of x86 processor cores on a single chip. Intel now refers to this technology as the Many Integrated Core (MIC) architecture. The first implementation of the MIC architecture will integrate numerous 64-bit x86 cores with coherent caches, 512-bit-wide vector-processing units, and an on-chip ring network.

In other words, the first implementation of MIC draws heavily on the Larrabee design we described in 2008. (See *MPR 9/29/08-01*, "Intel's Larrabee Redefines GPUs.") Since then, Intel has decided that the initial Larrabee-family chips can't compete with established GPUs from market leaders AMD and Nvidia. Last December, Intel announced that the first Larrabee chip was demoted to a "software-development platform," although the company said it remains committed to making discrete GPUs in the future. (See *MPR 12/28/09-02*, "Editorial: Augmented Reality—and Larrabee.")

Knights is the heir apparent. Although the first Knights design is obviously derived from Larrabee, Intel says future designs could implement the MIC architecture differently. Knights chips will act as coprocessors for x86 host CPUs or perhaps as the main processors in custom-built supercomputers from other companies. Intel has no plans to resume making its own supercomputers, as the company did years ago. As a component supplier, Intel is in more than 80% of the world's Top 500 supercomputers. (See *www.top500.org*.)

The first MIC processor is a chip code-named Aubrey Isle, a 32-core design clocked at 1.2GHz. Four-way Hyper-Threading enables a total of 128 simultaneous hardware-managed threads. Each core has a 256KB L2 cache, which they can share coherently with other cores. All cores share 1–2GB of external GDDR5 "graphics" memory, which isn't limited to graphics data. For HPC applications, this memory is a general-purpose repository for any kind of data—essentially, it's an off-chip L3 cache.

Intel showed Aubrey Isle on a PCI Express card code-named Knights Ferry. The card is entering production this year, but apparently, it's intended for developers, not customers. The first actual product will be another card code-named Knights Corner, which will have a different MIC processor. That chip (code-name unknown) will integrate at least 50 cores and will be manufactured in 22nm technology, which Intel plans to debut in 2011.

Don't confuse MIC or Knights with a 48-core chip that Intel showed earlier this year. Called the Single-chip Cloud Computer (SCC), this processor is an experimental design using simpler x86 cores. SCC uses Pentium-class 32-bit cores, whereas Larrabee and Knights use 64-bit x86 cores with much beefier FPUs and vector-processing units. Larrabee and Knights draw more heavily on Intel's Tera-Scale project. (See *MPR 4/9/07-01*, "Low-Key Intel 80-Core Intro: The Tip of the Iceberg.") SCC is a different research project for a different market. (See *MPR 4/26/10-01*, "The Single-chip Cloud Computer.")

Meanwhile, Nvidia keeps pushing its new Fermi architecture, a GPU-first design with several new features for HPC. Fermi chips are in production now, and dozens of HPC customers are using Nvidia's Cuda software-development tools. (See *MPR 10/5/09-01*, "Looking Beyond Graphics.")

Intel struck back last year by acquiring Cilk Arts and RapidMind to bolster its own software-development tools for parallel processing. These tools will allow developers to compile programs written in C, C++, and FORTRAN for mainstream x86 processors (e.g., Xeon) or MIC processors. (See *MPR 9/14/09-01*, "Summer Shopping Spree.") AMD is pursuing the same market with its ATI Stream platform. (See *MPR 12/22/08-01*, "AMD's Stream Becomes a River.")

Even if Intel never ships a discrete GPU into the graphics market, the company will have to overcome the momentum of AMD and Nvidia GPUs in the HPC market. But for HPC applications, the general-purpose x86 architecture looks better when pitted against processors designed primarily to accelerate 3D graphics. ♦