



The Xeon family of processors have been around so long and are so vital to the fortunes of Intel that no one questions that there is a long-term roadmap for the devices. But thus far, Intel has only delivered one generation of Xeon Phi coprocessors, and the details have been scant on the second generation. Customers want to know there is a long-term plan for the Xeon Phi chips, which will be used both as processors and coprocessors starting with the “Knights Landing” variant of the multicore chips, and so at the SC14 supercomputing conference this week in New Orleans, Intel is lifting the veil a bit on its Knights roadmap.

The family already included “Knights Corner,” the current Xeon Phi chip used in coprocessors, and the “Knights Landing” chip expected in the second half of 2015. The roadmap now also includes a future chip called “Knights Hill,” which will pack even more floating point punch.

“Knights Hill is still a little out in time,” explains Charlie Wuischpard, vice president and general manager of workstation and high performance computing at Intel’s Data Center Group. “But we wanted to do this for a couple of reasons. For customers that are investing in Xeon Phi, they want to know that this is a multi-generational investment, not a one-shot deal. Two, we actually have to start development in overlapping ways, so we have already started development on Knights Hill to understand what kind of performance characteristics we want to achieve. So this is really to start paving a longer-term roadmap that our customers can rely on.”

Driving The Roadmap

“Knights Ferry,” you will recall, was a prototype multi-core chip with 32 X86 cores that debuted in the summer of 2010. It used Intel’s 45 nanometer processes and was an experimental coprocessor designed to test the idea of using multi-core coprocessors in conjunction with Xeon CPUs in hybrid systems. Intel had originally planned to launch itself into the external graphics chip business with a project called “Larrabee,” but decided to mothball the X86-based GPUs and kept the underlying technology as the Xeon Phi coprocessor. (Nvidia and AMD have come from the opposite direction, positioning their GPU chips as compute engines.)

“Knights Corner” is the current chip in the Xeon Phi family, and it was introduced to the world in the summer of 2011 and shipped in initial systems in the following year. The Knights Corner chip has 61 cores based on a heavily modified Pentium P54C core and is implemented in Intel’s 22 nanometer processes. Depending on the core count and their clock speed, this coprocessor delivers a maximum of around 1 teraflops of double precision number crunching performance. The Knights Corner Xeon Phi chip is only available on a PCI card form factor, and significantly, it only support PCI-Express 2.0 links. The last three generations of Intel Xeon processors can support much faster PCI-Express 3.0 links, which significantly cuts down on the latency of moving data back and forth between the CPU and the accelerator.

The second-generation of production Xeon Phi chips, Knights Landing, shifts the compute element to a modified version of the “Silvermont” Atom core, and there is speculation that Intel could add as many as 72 cores on a single chip – a number that Intel has not divulged and has not confirmed. (The official line is that the Knights Landing chip will have “at least 60 cores.”) The modified Silvermont core is expected to have four threads per core and importantly will offer full compatibility with the Xeon instruction set and support AVX-512 vector instructions. The Knights Landing chip will come with Intel’s Omni-Path networking fabric integrated on the die, which is neither Ethernet nor InfiniBand but rather a follow-on to InfiniBand that is compatible with the software stack of InfiniBand. (Intel used to call this future network fabric Omni Scale, but has just rebranded it Omni-Path. We are covering Omni-Path in a separate feature.) The Knights Landing chip will be etched using Intel’s current 14 nanometer processes, which is how the company can cram so many dies onto the die while at the same time presumably cranking up the clock speed. The company has said that Knights Landing will be available in systems starting in the second half of 2015 and will deliver around 3 teraflops of floating point processing at double precision. This is a big jump in theoretical peak performance.

Unveiling Details of Knights Landing
(Next Generation Intel® Xeon Phi™ Products)

Platform Memory: DDR4 Bandwidth and Capacity Comparable to Intel® Xeon® Processors

Compute: Energy-efficient IA cores²

- Microarchitecture enhanced for HPC³
- 3X Single Thread Performance vs Knights Corner⁴
- Intel Xeon Processor Binary Compatible⁵

On-Package Memory:

- up to 16GB at launch
- 1/3X the Space⁶
- 5X Bandwidth vs DDR4⁷
- 5X Power Efficiency⁸

Jointly Developed with Micron Technology

2nd half '15
1st commercial systems

3+ TFLOPS¹
In One Package
Parallel Performance & Density

Helping turn that peak performance into actual performance are some important changes to the architecture of the Xeon Phi chips. The most important is the addition of between 8 GB and 16 GB of Hybrid Memory Cube (HMC) 3D stacked memory, developed in conjunction with Micron Technology, that will sit very close to the Knights Landing processor. This on-package memory will have five times the memory bandwidth of DDR4 main memory and take up one-third the space and be five times more power efficient than the GDDR5 buffer memory used in the current Knights Corner cards. For many memory-intensive applications, data sets will fit nicely in this “near memory” and performance relative to Xeon processors and prior generations of Xeon Phi coprocessors should be significantly higher. The Knights Landing version of the Xeon Phi will also support DDR4 main memory off the package, just like Xeon processors do. And importantly, it will come as both a coprocessor (support PCI-Express 3.0 slots, although Intel has not said this) and as a standalone processor in its own right and with its own socket.

“We just want to be balanced so we know we are covering the waterfront,” Wuischpard tells *EnterpriseTech*. “I think there are a lot of use cases for the card, and you will see that coming out with Knights Landing but a lot of that has to do with customers wanting to use the Intel architecture, but they want to pack it in in different ways and they want to re-use some of the existing systems they have. I would have to say that most of the interest we have is in the socketed part. That is the nirvana that our customers have been interesting in, getting that many threads and that much compute density. A number of these configurations are liquid cooled, even for Knights Landing, and there are both liquid and air cooled versions that will be coming out. The density and packaging will lend itself to liquid cooled solutions with Knights Landing and it will become an increasingly important aspect with Knights Hill.”

The infographic is titled "Industry and User Momentum" and is divided into three main sections: Knights Corner, Knights Landing, and Knights Hill. Knights Corner is labeled "1 TFLOPS¹" and shows a blue PCIe card. Knights Landing is labeled "3+ TFLOPS²" and shows a square chip and a blue PCIe card, with text listing features: "-Bootable Processor", "-On-Pkg, High BW Memory", and "-Integrated Fabric". It also states "2H'15 First Commercial Systems" and ">50 systems providers expected³". Knights Hill is labeled "Announcing" and lists "3rd Generation Intel® Xeon Phi™ Product Family", "2nd Generation Intel Omni-Path Architecture", and "10nm process technology". At the bottom, it says ">100 PFLOPS customer system compute commits to-date³".

Intel is not saying a whole lot publicly about Knights Hill just yet, but Wuischpard said that the chip would be manufactured using Intel's 10 nanometer process. This is the chip etching technology that Intel is just now perfecting for its "Broadwell" family of desktop and laptop processors. (The desktop, laptop, smartphone, and tablet chips get a new process first and then it is gradually moved to server processors after it is perfected and ramped up.) The only other thing we know is that the Knights Hill chip will have the second generation of the Omni-Path fabric integrated on the die as well.

So how far out is the Knights Hill version of the Xeon Phi? "If you look at the normal cadence of product releases, it is a couple of years out," Wuischpard says. The initial Knights Hill coprocessors and processors will be targeted for the pre-exascale supercomputing systems before the end of the decade, and then will trickle into enterprise systems as rapidly as software support, system designs, and budgets allow.

As for performance, Intel is not saying much, but it looks like the leaps will continue to be large. "If you look at 1 teraflops for Knights Corner and 3 teraflops for Knights Landing, and you compare that to our Xeon line, which tends to increase performance at a certain rate and pace, then Xeon Phi is on a much more steep trajectory," says Wuischpard. "We know what we want to build, we know the characteristics and the performance it will have, and in our parlance we know what the 'landing zone' is and what the requirements are."

This is the real message that Intel wants to convey, that it has a multi-generational roadmap that can compete with the GPU roadmaps from Nvidia and AMD. (Nvidia is getting a lot more traction accelerating workloads, obviously.)

As for performance, it stands to reason that Intel will focus more on adding Atom cores to the Knights Hill chips rather than cranking up the clock speeds. These chips are designed explicitly for parallel applications, so more cores is better than higher clocks. The higher core counts will also provide a clean differentiation with the Xeon chips, which currently have a maximum of 18 cores with the "Haswell" Xeon E5-2600 v3 and which will no doubt keep adding more cores with each passing year. It is hard to figure how many more cores Intel can add with the Knights Hill chips. Intel has more than doubled the core count on the top-end Xeon E5s in the move from 32 nanometer to 22 nanometer processes, from eight-cores with the "Sandy Bridge" generation implemented in 32 nanometer to 18 cores with the Haswell chips implemented in 22 nanometers. If Intel expands the Knights Hill die size and also shrinks the transistors, it is perfectly feasible that it could double up the core count on Knights Hill compared to Knights Landing with the move from 14 nanometer to 10 nanometer, and therefore double up the performance at the same clock speed. Or, Intel could get microarchitecture improvements in the Atom cores used in the Knights chips and then lower the clock speed a bit to reduce the heat produced by the chip even more, allowing for denser packing of multiple units and possibly the addition of even more cores per die.

There are a lot of variables to play with, to be sure, and you can bet that Intel is working hard to engineer the Knights Hill chip such that it delivers the most improvement it can on performance per watt, since this is the metric that matters most. It seems unlikely that Intel could triple performance between the Knights Landing and Knights Hill generations, as it has done between Knights Corner and Knights Landing. But should that be possible, it would certainly be a welcome development for customers running parallel applications.

Knights Hill is a long way off. Right now Intel has to get Knights Landing out the door and ramping.

“We are still on track to deliver the first commercial systems in the second half of 2015,” Wuischpard says, adding that Intel has in excess of 50 design wins from system makers using the Knights Landing part and over 100 petaflops of aggregate performance from customers who are buying systems using the chip so far. Not all of these customers are national supercomputing labs, either. Xeon Phi is catching on in the commercial sector, too. But Wuischpard can’t name names.

“You can guess who they are in that they would be customers who would need that kind of performance for the problems they are trying to solve,” he says. “These are large Fortune 500 enterprises that we have really gone into exploration with for this kind of technology. Oil and gas, financial services, big pharma, and life sciences have been intrigued by the prospects. It is the usual suspects and it is not necessarily some brand new workload. But we are having those conversations, too.”

Wuischpard says that there are a lot more organizations who have asked for early release Knights Landing parts than Intel can supply, and that it is “trying to manage through the demand.” This is a good kind of problem to have ahead of a major product launch, which will probably come at the International Super Computing conference in Germany next June, if we had to guess.

Share this:



Categories: [Slider: Front Page](#), [Slider: Systems](#), [Systems](#)

Tags: [Intel](#), [Knights](#), [Knights Hill](#), [Knights Landing](#), [Omni-Path](#), [Xeon](#), [Xeon Phi](#)



About the author: Timothy Prickett Morgan

Editor in Chief, EnterpriseTech Prickett Morgan brings 25 years of experience as a publisher, IT industry analyst, editor, and journalist for some of the world’s most widely-read high-tech and business publications including The Register, BusinessWeek, Midrange Computing, IT Jungle, Unigram, The Four Hundred, ComputerWire, Computer Business Review, Computer System News and IBM Systems User.

Add a Comment

Name *

Email *

Website