

Intel's 32nm Update: The Follow-on to Core i7 and More

Date: Feb 11, 2009
Type: CPU & Chipset
Manufacturer: Intel
Author: Anand Lal Shimpi
Page 1

Seven billion dollars.
That's the amount that Intel is going to spend in the US alone on bringing up its 32nm manufacturing process in 2009 and 2010.
These are the fabs Intel is converting to 32nm:

32nm Manufacturing Fabs



In Oregon Intel has the D1D fab which is already producing 32nm parts, and D1C which is scheduled to start 32nm production at the end of this year. Then two fabs in Arizona: Fab 32 and Fab 11X. Both of them come on line in 2010.
By the end of next year the total investment just to enable 32nm production in the US will be approximately eight billion dollars. In a time where all we hear about are bailouts, cutbacks and recession, this is welcome news.
If anything, Intel should have a renewed focus on competition given that its chief competitor finally woke up. That focus is there. The show must go on. 32nm will happen this year. Let's talk about how.

Page 2
The Manufacturing Roadmap

The tick-tock cadence may have come about at the microprocessor level, but its roots have always been in manufacturing. As long as I've been running AnandTech, Intel has introduced a new manufacturing process every two years. In fact, since 1989 Intel has kept up this two year cycle.

Process Name	P1264	P1266	P1268	P1270	P1272
Lithography	65nm	45nm	32nm	22nm	16nm
1 st Production	2005	2007	2009	2011	2013

Manufacturing

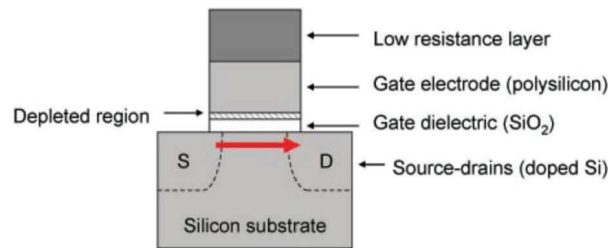
Development

We saw the first 45nm CPUs with the Penryn core back in late 2007. Penryn, released at the very high end, spent most of 2008 making its way mainstream. Now you can buy a 45nm Penryn CPU for less than \$100.

The next process technology, which Intel refers to internally as P1268, shrinks transistor feature size down to 32nm. The table above shows you that first production will be in 2009 and, after a brief pause to check your calendars, that means this year. More specifically, Q4 of this year.

I'll get to the products in a moment, but first let's talk about the manufacturing process itself.

Here we have our basic CMOS transistor:

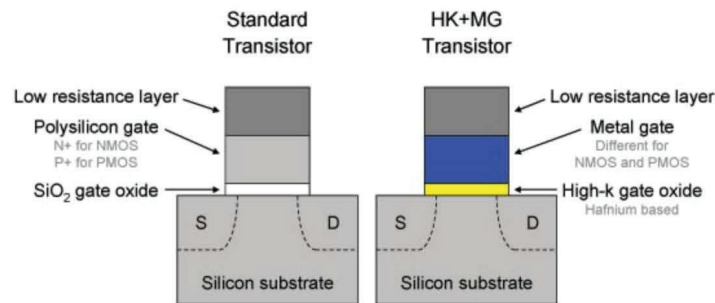


Current flows from source to drain when the transistor is on, and it isn't supposed to flow when it's off. Now as you shrink the transistor, all of its parts shrink. At 65nm Intel found that it couldn't shrink the gate dielectric any more without leaking too much current through the gate itself. Back then the gate dielectric was 1.2nm thick (about the thickness of 5 atoms), but at 45nm Intel's switched from a SiO₂ gate dielectric to a high-k one using Hafnium. That's where the high-k comes from.

The gate electrode also got replaced at 45nm with a metal to help increase drive current (more current flows when you want it to). That's where the metal gate comes from.

The combination of the two changes to the basic transistor gave us [Intel's high-k + metal gate transistors at 45nm](#), and at 32nm we have the second generation of those improvements.

The high-k gate dielectric gets a little thinner (equivalent to a 0.9nm SiO₂ gate, but presumably thicker since it's Hafnium based, down from 1.0nm at 45nm) and we've still got a metal gate.



At 32nm the transistors are approximately 70% the size of Intel's 45nm hk + mg transistors, allowing Intel to pack more in a smaller area.

- 2nd generation high-k + metal gate transistors
- Immersion lithography on critical layers
- 9 copper + low-k interconnect layers
- ~70% dimension scaling from 45nm generation
- Pb-free and halogen-free packages

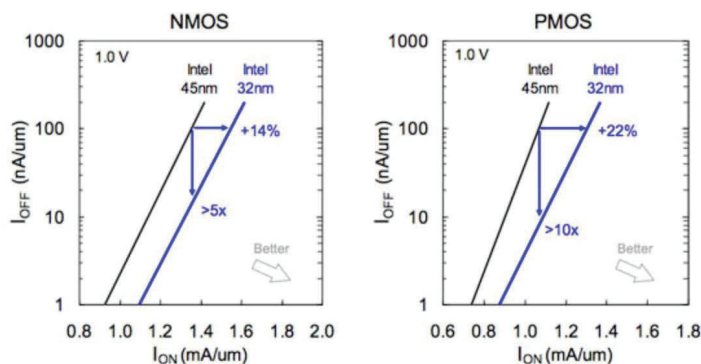
The big change here is that Intel is using immersion lithography on critical metal layers in order to continue to use existing 193nm lithography equipment. The smaller your transistors are, the higher resolution your equipment has to be in order to actually build them. Immersion lithography is used to increase the resolution of existing lithography equipment without requiring new technologies. It is a costlier approach, but one that becomes necessary as you scale below 45nm. Note that [AMD to made the switch to immersion lithography at 45nm](#).

- 2nd generation high-k + metal gate
 - 0.9nm equivalent oxide thickness high-k (scaled from 1.0 nm on 45nm)
 - Replacement Metal Gate process flow
 - 30nm gate length
 - 4th generation strained silicon
- >22% performance increase
- Tightest reported gate pitch

- Highest reported drive currents

Intel reported significant gains in transistor performance at 32nm; the graphs below help explain:

32nm Transistor Performance



32nm provides improved performance or reduced leakage
Highest drive current of all reported 32nm technologies

We're looking at the comparison of leakage current vs. drive current for both 32nm NMOS and PMOS transistors. The new transistors showcase a huge improvement in power efficiency. You can either run them faster or run them at the same speed and significantly reduce leakage current by a magnitude of greater than 5 - 10x compared to Intel's 45nm transistors. Intel claims that its 32nm transistors boast the highest drive current of all reported 32nm technologies at this point, which admittedly there aren't many.

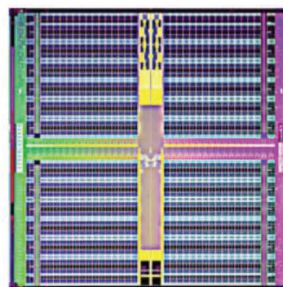
The power/performance characteristics of Intel's 32nm process make it particularly attractive for mobile applications. But more on that later.

Page 3

Fat Pockets, Dense Cache, Bad Pun

Whenever Intel introduces a new manufacturing process the first thing we see it used on is a big chip of cache. The good ol' SRAM test vehicle is a great way to iron out early bugs in the manufacturing process and at the end of 2007 Intel demonstrated its first 32nm SRAM test chip.

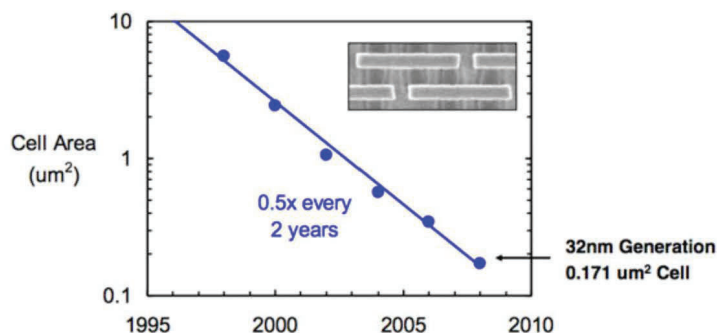
- 0.171 μm^2 cell
- 291 Mbit
- >1.9 billion transistors
- 4 GHz operation
- First demonstrated Sep '07



Intel's 32nm SRAM test vehicle

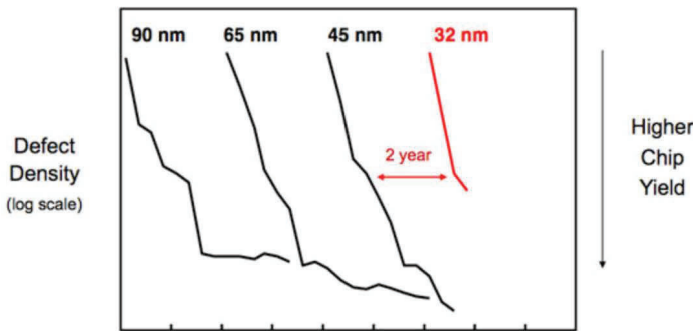
The 291Mbit chip was made up of over 1.9 billion transistors, switching at 4GHz, using Intel's 32nm process. The important number to look at is the cell size, which is the physical area a single bit of cache will occupy. At 45nm that cell size was 0.346 μm^2 (for desktop processors, Atom uses a slightly larger cell), compared to 0.370 μm^2 for AMD's 45nm SRAM cell size. At 32nm you can cut the area nearly in half down to 0.171 μm^2 for a 6T SRAM cell. This means that in the same die area Intel can fit twice the cache, or the same amount of cache in half the area. Given that Core i7 is a fairly large chip at 263 mm^2 I'd expect Intel to take the die size savings and run with them. Perhaps with a modest increase to L3 cache size.

SRAM Cell Size Scaling



A big reason we're even getting this disclosure today is because of how healthy the 32nm process is. Below we have a graph of defect density (number of defects

Big reason we're even getting this disclosure today is because of how healthy the 32nm process is. Below we have a graph of defect density (number of defects in silicon per area) vs time; manufacturing can't start until you're at the lowest part of that graph - the tail that starts to flatten out:



Intel's 45nm process ramped and matured very well as you can see from the chart. The 45nm process reached lower defect densities than both 65nm and 90nm and did it faster than either process. Intel's 32nm process is on track to outperform even that.

Two Different 32nm Processes?

With Intel now getting into the SoC business (System on a Chip), each process node will now have two derivatives - one for CPUs and one for SoCs. This started at 45nm with process P1266.8, used for Intel's consumer electronics and Moorestown CPUs and will continue at 32nm with the P1269 process.

	45 nm		32 nm		22 nm	
Process:	P1266	P1266.8	P1268	P1269	P1270	P1271
Products:	CPU	SoC	CPU	SoC	CPU	SoC

There are two major differences between the CPU and SoC versions of a given manufacturing process. One, the SoC version will be optimized for low leakage while the CPU version will be optimized for high drive current. Remember that graph comparing leakage vs. drive current of 45nm vs. 32nm? The P1268 process will exploit the arrows to the right, while P1269 will attempt to push leakage current down.

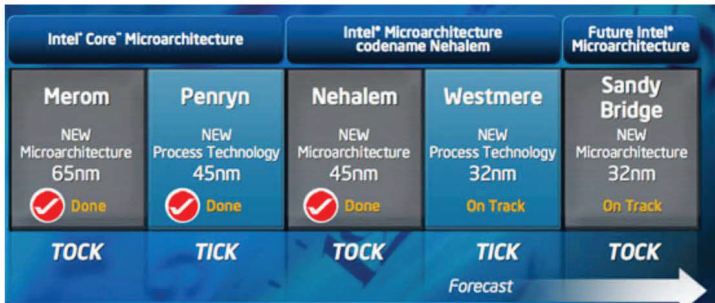
The second difference is that certain SoC circuits require higher than normal voltages and thus you need a process that can tolerate those voltages. Remember that with a SoC it's not always just Intel IP being used, there are many third parties that will contribute to the chips that eventually make their way into smartphones and other ultra portable devices.

The buck doesn't stop here, in two more years we'll see the introduction of P1270, Intel's 22nm process. But before we get there, there's a little stop called Sandy Bridge. Let's talk about microprocessors for a bit now shall we?

Page 4

Tick-Tock: U R Doin it Right

Let's check the stats; Conroe in July 2006, Penryn in October 2007, Nehalem in November 2008. That's a tock, tick, and another tock, each about a year apart. Note that the cadence does appear to be slipping a bit, but we'll see exactly when in 2009 we get Westmere before making any accusations.



The next tick is, as I just mentioned, Westmere. It's a 32nm shrink of Nehalem, much like Penryn was a 45nm shrink of Conroe/Merom. And it's due out in the fourth quarter of this year.

Yesterday, Intel demonstrated working versions of its 32nm processors in both desktops and notebooks. The notebook aspect of the demonstration is very important, which I'll get to later. Both mobile and desktop versions of Westmere will be shipping from Intel in Q4.

Getting Complicated with Code Names

Nehalem is the overall name for Intel's 45nm desktop/mobile/server product family. At the high end we have Bloomfield, which is the quad-core, eight-thread, Core i7 processor we all long for. That's the only Nehalem derivative that's launched thus far.

Segment	Manufacturing Process	Socket	Processor	Cores	Threads	Release Date
High End Desktop	45nm	LGA-1366	Bloomfield	4	8	Q4 2008

Desktop		1366				2008
Mainstream Desktop	45nm	LGA-1156	Lynnfield	4	8	2H 2009
Mobile	45nm	mPGA-989	Clarksfield	4	8	2H 2009
4S Server	45nm	LGA-1567	Nehalem-EX	8	16	2H 2009
2S Server	45nm	LGA-1366	Nehalem-EP	4	8	1H 2009
1S Server	45nm	LGA-1156	Lynnfield	4	8	2H 2009

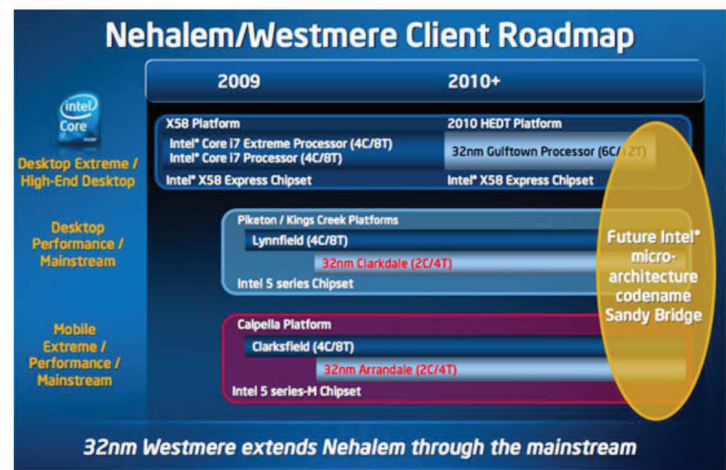
By the end of this year we'll see Lynnfield and Clarksfield. These are both quad-core, eight-thread Nehalem processors but at lower TDPs and price points. They will fit into Intel's unannounced LGA-1156 socket and only support two channels of DDR3 memory (compared to LGA-1366 and 3-channels with Core i7).

On the server side we'll have Nehalem-EX, an 8-core, 16-thread version. Nehalem EP a 4-core, 8-thread version. And Lynnfield again for the entry level servers. These are all 45nm parts and all due out by the end of this year.

Note that there's one name missing: Havendale. Havendale was supposed to be a 2-core Lynnfield + on-chip graphics, perfect for notebooks and low end desktops where quad-core isn't necessary. Unfortunately, Havendale got delayed until Q4 2009 with systems shipping in Q1 2010. That just happened to coincide with Intel's 32nm ramp so a very significant decision was made: Havendale got scrapped.

Enter the 32nm Lineup

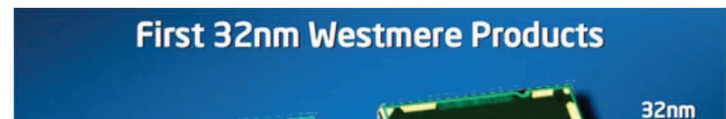
Instead of Havendale in Q4, we'll get Clarkdale and Arrandale. These are both dual-core, quad-thread processors, and both have on-package graphics. The CPU cores will be built on Intel's 32nm process and in fact, they will be the first Westmere CPUs shipping into the market.

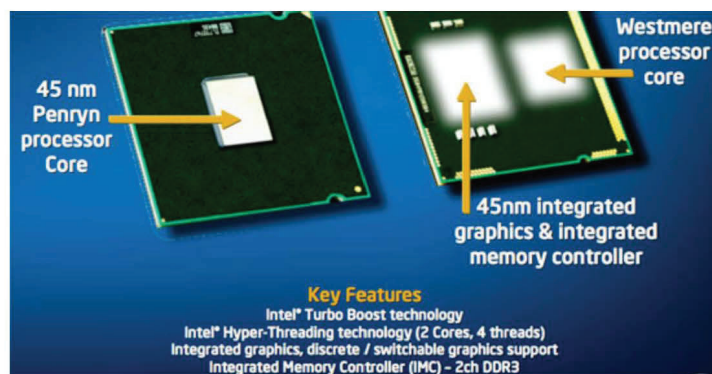


Now note that the dual-core market is the largest slice of the processor pie. Intel must be incredibly confident in its 32nm process to start shipping it into these demand markets first. Remember that both 65nm and 45nm initially launched on the high end desktop, but 32nm is making its debut in mainstream notebooks and desktops. The 32nm ramp is going to be a good one folks.

Segment	Manufacturing Process	Socket	Processor	Cores	Threads	Release Date
High End Desktop	32nm	LGA-1366	Gulftown	6	12	1H 2010
Mainstream Desktop	32nm	LGA-1156	Clarkdale	2	4	Q4 2009
Mobile	32nm	mPGA-989	Arrandale	2	4	Q4 2009
4S Server	32nm	LGA-1567	???	?	?	2010
2S Server	32nm	LGA-1366	???	?	?	2010
1S Server	32nm	LGA-1156	Clarkdale	2	4	2010

Clarkdale/Arrandale have 32nm CPUs but their on-package GPUs are still built on Intel's 45nm process; these are the GPUs that were supposed to be used for Havendale! It won't be until 2010 with Sandy Bridge that we see a 32nm CPU and 32nm GPU on the same package.





A side effect of the Clarkdale/Arrandale architecture is that the memory controller is now located on the GPU and not the CPU, although both are still on package and should still be quite low latency.

Westmere Family Mainstream Desktop / Mobile Processor

Key Features!:

- Intel microarchitecture codename Nehalem on 32nm
- Multi-Chip Package (MCP) processor with:
 - processor cores built on 32nm 2nd generation high-k metal gate process
 - Integrated graphics controller & memory controller built on 45nm high-k metal gate process
- Intel Turbo Boost technology
- Intel Hyper-Threading technology (2 Cores, 4 threads)
- Intel Smart Cache
- Integrated memory controller (IMC)
- Integrated, discrete / switchable graphics support
- Advanced Encryption Standard (AES) acceleration
- Compatible with the Intel 5 series chipset based platforms due to be released in 2H'09

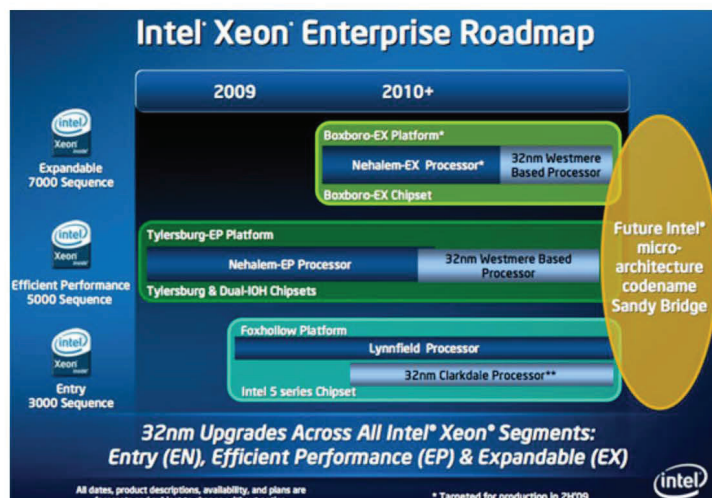
Worlds First 32nm Based Processor

Keep following; if you want a quad-core Westmere, your only option will be in the LGA-1366 socket with Gulftown. **Core i7 will get replaced with a six-core, twelve-thread processor in early 2010.** There won't be a 32nm quad-core part on the desktop until the end of 2010 with Sandy Bridge.

Page 6

The Server Roadmap

Intel's 32nm server roadmap is notably different from the desktop roadmap. Nehalem-EX will ship into the Xeon 7000 series as an 8-core, 16-thread part. It will eventually get replaced sometime in 2010 with a 32nm Westmere derivative.



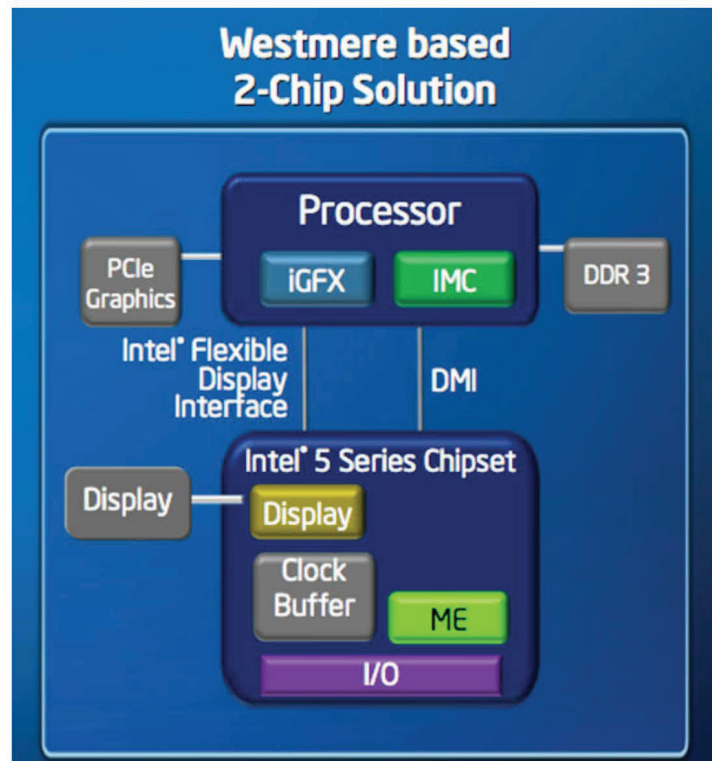
We'll see a 32nm six-core Westmere based processor in the Xeon 5000 series in 2010.

Finally Lynnfield and Clarkdale will be carried over to the entry level Xeon platforms at the end of this year and into 2010.

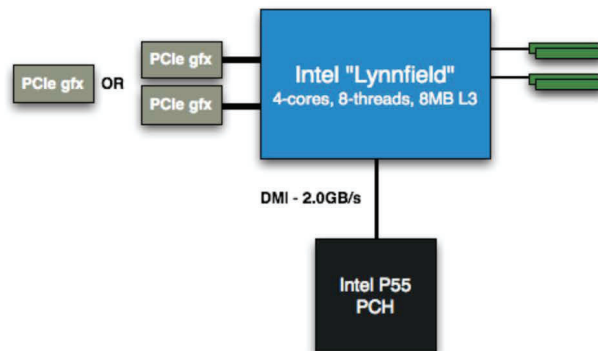
What About Chipsets?

Intel's X58 chipset will remain the top dog through 2010. Chances are that we won't see it replaced until the next tock with Sandy Bridge. Now that isn't to say that the six-core 32nm Gulftown will work in existing X58 motherboards; while that would be nice, Intel does have a habit of forcing motherboard upgrades, we'll have to wait and see.

The rest of the Nehalem/Westmere family will rely on Intel's upcoming P55 chipset:



Originally both Lynnfield and Havendale were to have an on-package PCIe controller, I'm not sure if that has changed with the Havendale cancellation but I see no reason for it to have. In which case a Lynnfield system will still look like this:



Page 7

Westmere's New Instructions

Much like Penryn and its new SSE4.1 instructions, Westmere comes with 7 new instructions added to those already in Core i7. These instructions are specifically focused on accelerating encryption/decryption algorithms. There's a single carryless multiply instruction (PCLMULQDQ...I love typing that) and 6 instructions of AES.

Westmere: AES New Instructions

- Use additional transistor budget to add new capabilities
 - Similar to adding SSE4.1 in Penryn (45nm tick)
- 7 new instructions for accelerating encryption/decryption algorithms
 - Carryless multiply (PCLMULQDQ)
 - 6 instructions for AES
- Example client usage
 - Enables full disk encryption

Intel gives the example of hardware accelerated full disk encryption as a need for these instructions. With the new instructions being driven into the mainstream

first, we'll probably see quicker than usual software adoption.

Final Words

What is there to say other than: it's a healthy roadmap. The only casualty I've seen is Havendale but I'd gladly trade Havendale for a 32nm version. But let's get down to what this means for what you should buy and when.

At the very high end, Core i7 users have little reason to worry. While Intel is expected to bump i7 up to 3.33GHz in the near future, nothing below i7 looks threatening in 2009. Moving into 2010, the 6-core 32nm i7 successor should be extremely powerful. Intel's strategy with LGA-1366 makes a lot of sense: if you want more cores, this is the platform you're going to have to be on.

Now although I said that nothing will threaten Core i7 this year, you may be able to get i7-like performance out of Lynnfield in the second half. A quad-core Lynnfield running near 3GHz, should offer much of the performance of an i7 with a lower platform cost. Remember back to our original i7 review; we didn't find a big performance benefit from three channels of DDR3 versus two.

Lynnfield is on track for a 2H 2009 introduction and if you're unable to make the jump to i7 now, you'll probably be able to get i7-like performance out of Lynnfield in about 6 months. Intel did mention that the most overclockable processors would go into the LGA-1366 socket. Combine better overclockability with the promise of 6 cores in the future and it seems like LGA-1366 is shaping up to be a platform that's going to stick around despite cheaper alternatives.

The 32nm Clarkdale/Arrandale parts arriving by the end of this year really means one very important thing: the time to buy a new notebook will be either in Q4 2009 or Q1 2010. A 2-core, 4-thread 32nm Westmere derivative is not only going to put current Penryn cores to shame, it's going to be extremely power efficient. In its briefing yesterday, Intel mentioned that while Clarkdale/Arrandale clock speeds and TDPs would be similar to what we have today, you'll be getting much more performance. Seeing what we've seen thus far with Nehalem, I'd say a 2-core, 32nm version in a notebook is going to be reason enough for you to want to upgrade.

If I had to build a new desktop today I'd go Core i7 and think about upgrading to a 6-core version sometime next year. If I couldn't or didn't need to build today, then the thing to wait for is Lynnfield. Four cores that should deliver i7-like performance just can't be beat, and platform costs will be much cheaper by then (expect ~\$100 motherboards and near price parity between DDR3 and DDR2).

On the mainstream quad-core side, it may not make sense to try to upgrade to 32nm quad-core until Sandy Bridge at the end of 2010. If you buy Lynnfield this year, chances are that you won't feel a need to upgrade until late 2010/2011.

On the notebook side, if I needed one today I'd buy whatever I could keeping in mind that within a year I'm going to want to upgrade. I mentioned this in much of my recent Mac coverage; if you bought a new MacBook, it looks great, but the one you're going to really want will be here in about a year.

We owe Intel a huge thanks for being so forthcoming with its roadmaps. It's going to be a good couple of years for performance.