Morgan Stanley

IDF2009

INTEL DEVELOPER FORUM

PROCESSOR ARCHITECTURE AND HPC POINT OF VIEW

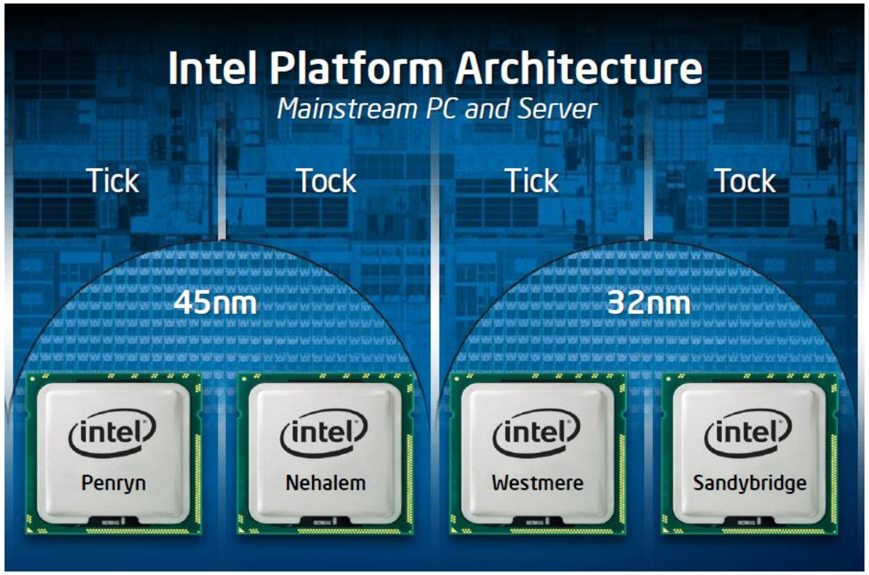
Zoltan Juhasz
zoltan.juhasz@morganstanley.com

Agenda

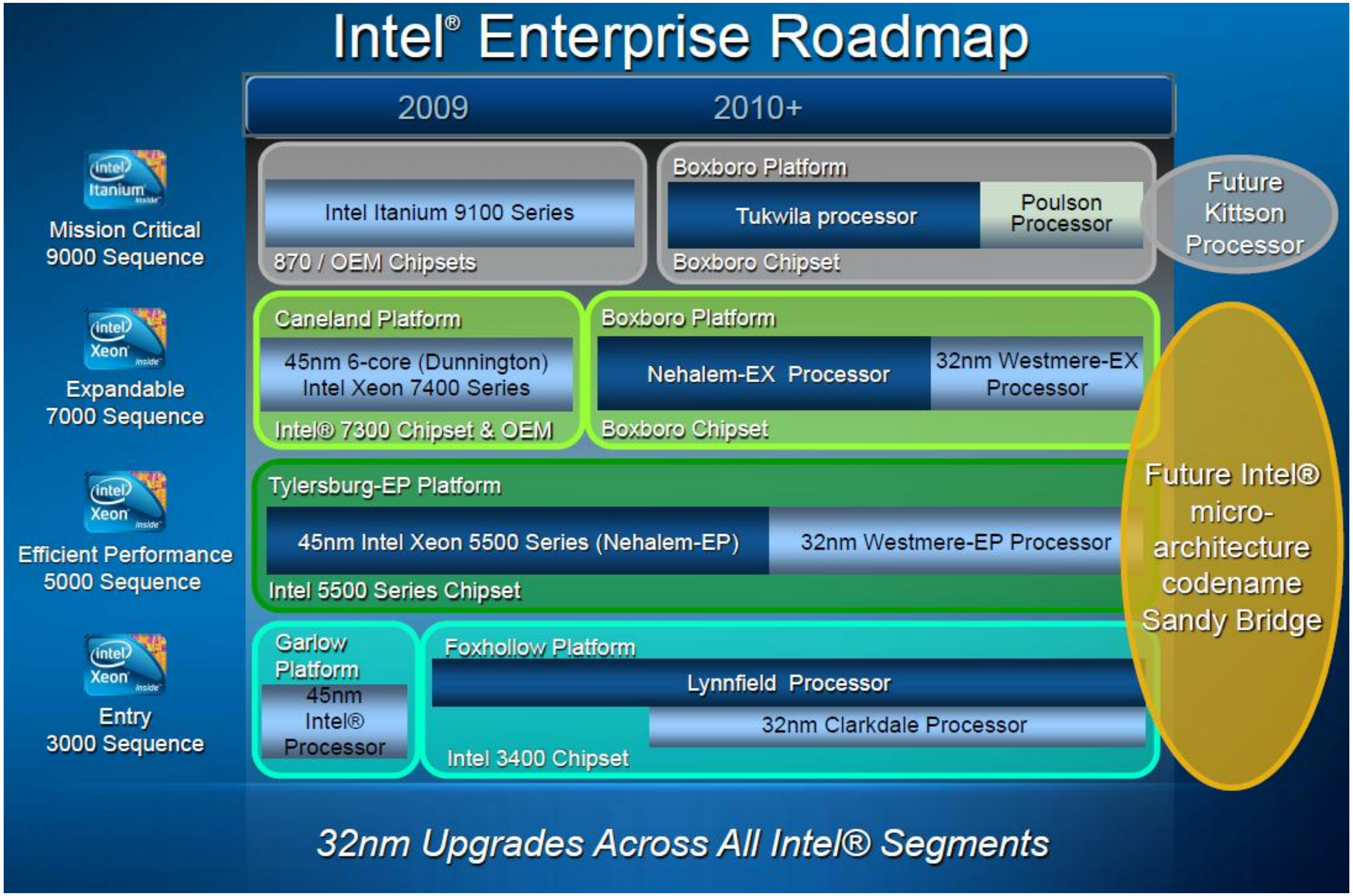
- Few quotes from the **Keynotes at IDF**
- Intel Server / Desktop **Roadmap**
- Key Points of **Intel Nehalem** Based Platforms
- Intel **Advanced Vector eXtension (AVX)** introduction
- Intel **Nehalem-EX** architecture: features and benefits
- **HP Extreme Scale Computing**: Server Futures (near term)

Keynotes from IDF

- Intel Platform Architecture: **Tick-Tock / next gen. silicon architecture - new microarchitecture**
- Intel is aggressively expanding to the **embedded world**
- Intel Platform Choice for Mission Critical Servers: **Nehalem-EX**
- **Improved energy efficiency**, and dissipation is another key area for Intel
- **New silicon technology**: 32nm (Q4 2009)

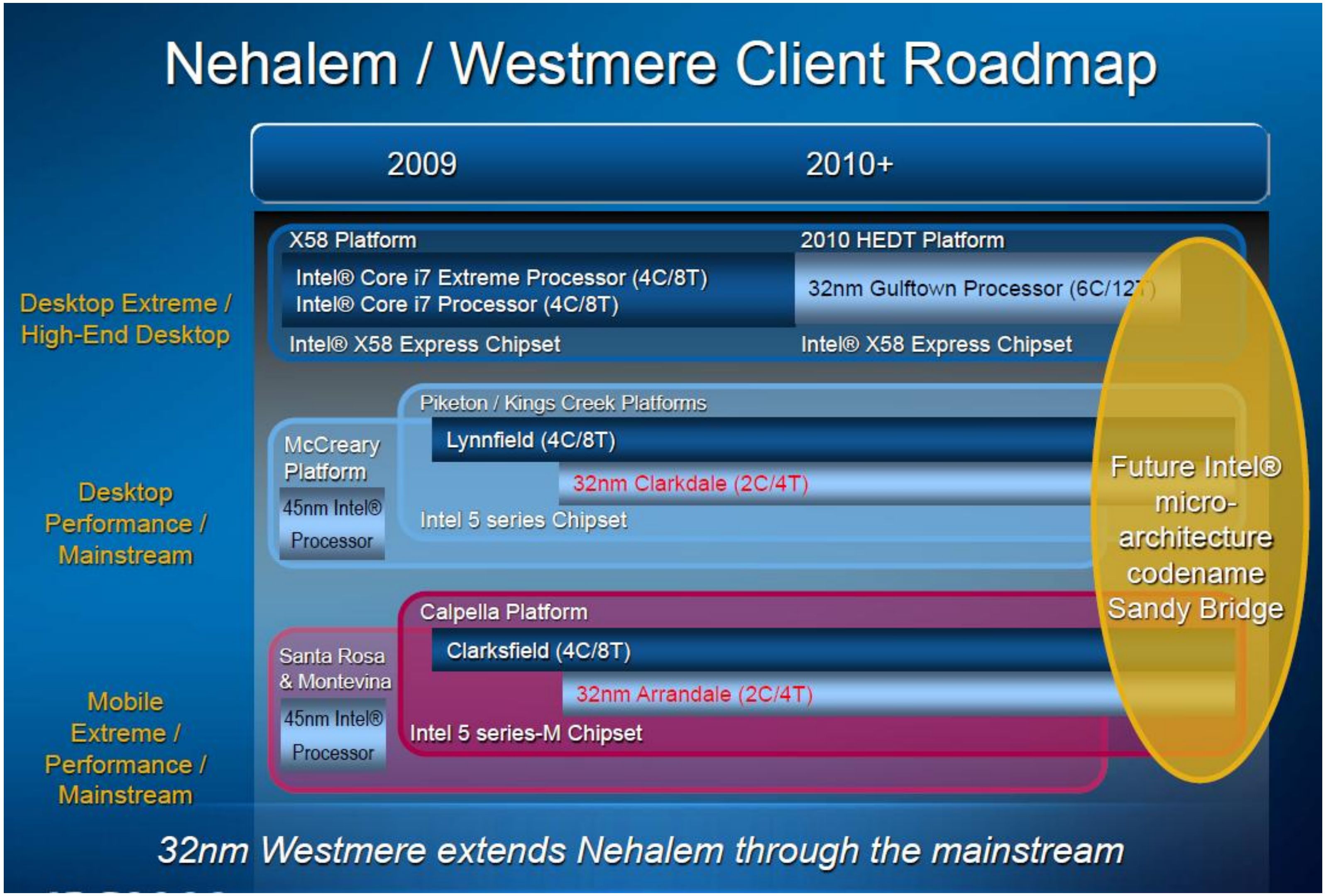


Keynotes: Server Roadmap


Intel® Enterprise Roadmap
2009
2010+
Mission Critical 9000 Sequence
Intel Itanium 9100 Series
870 / OEM Chipsets
Boxboro Platform
Tukwila processor
Poulson Processor
Boxboro Chipset
Future Kittson Processor
Expandable 7000 Sequence
Caneland Platform
45nm 6-core (Dunnington) Intel Xeon 7400 Series
Intel® 7300 Chipset & OEM
Boxboro Platform
Nehalem-EX Processor
32nm Westmere-EX Processor
Boxboro Chipset
Efficient Performance 5000 Sequence
Tylersburg-EP Platform
45nm Intel Xeon 5500 Series (Nehalem-EP)
32nm Westmere-EP Processor
Intel 5500 Series Chipset
Future Intel® micro-architecture codename Sandy Bridge
Entry 3000 Sequence
Garlow Platform
45nm Intel® Processor
Foxhollow Platform
Lynnfield Processor
32nm Clarkdale Processor
Intel 3400 Chipset
32nm Upgrades Across All Intel® Segments

Morgan Stanley

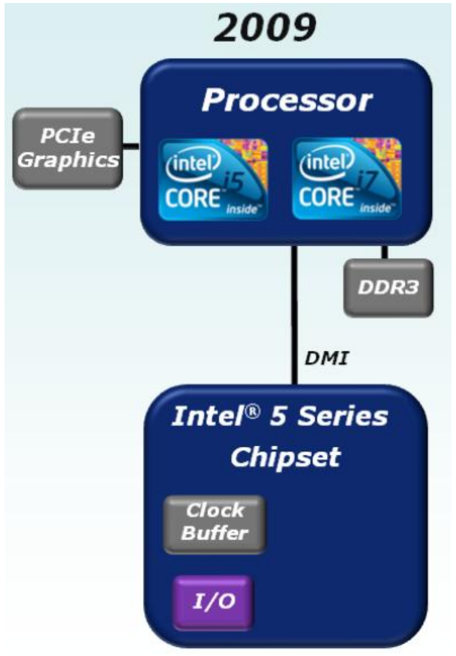
Keynotes: Client Roadmap

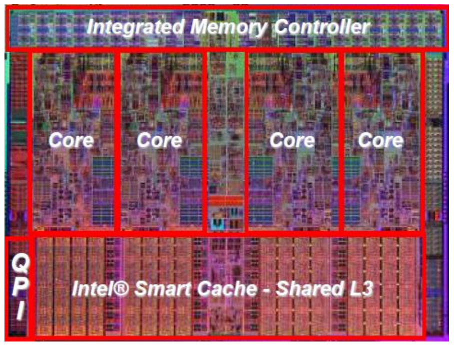

Nehalem / Westmere Client Roadmap
2009
2010+
Desktop Extreme / High-End Desktop
X58 Platform
Intel® Core i7 Extreme Processor (4C/8T)
Intel® Core i7 Processor (4C/8T)
Intel® X58 Express Chipset
2010 HEDT Platform
32nm Gulftown Processor (6C/12T)
Intel® X58 Express Chipset
Desktop Performance / Mainstream
McCreary Platform
45nm Intel® Processor
Piketon / Kings Creek Platforms
Lynnfield (4C/8T)
32nm Clarkdale (2C/4T)
Intel 5 series Chipset
Mobile Extreme / Performance / Mainstream
Santa Rosa & Montevina
45nm Intel® Processor
Calpella Platform
Clarksfield (4C/8T)
32nm Arrandale (2C/4T)
Intel 5 series-M Chipset
Future Intel® micro-architecture codename Sandy Bridge
32nm Westmere extends Nehalem through the mainstream

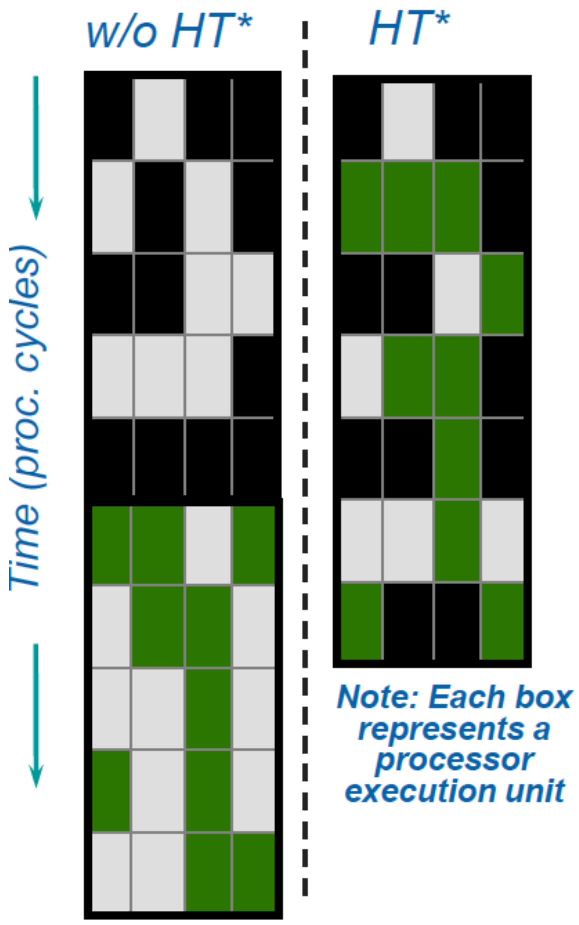
Morgan Stanley

Intel Nehalem Based Platforms

- Intel **Hyper-Threading** Technology
- Intel **Turbo Boost** Technology
- **PCI Express interface** moves to the processor





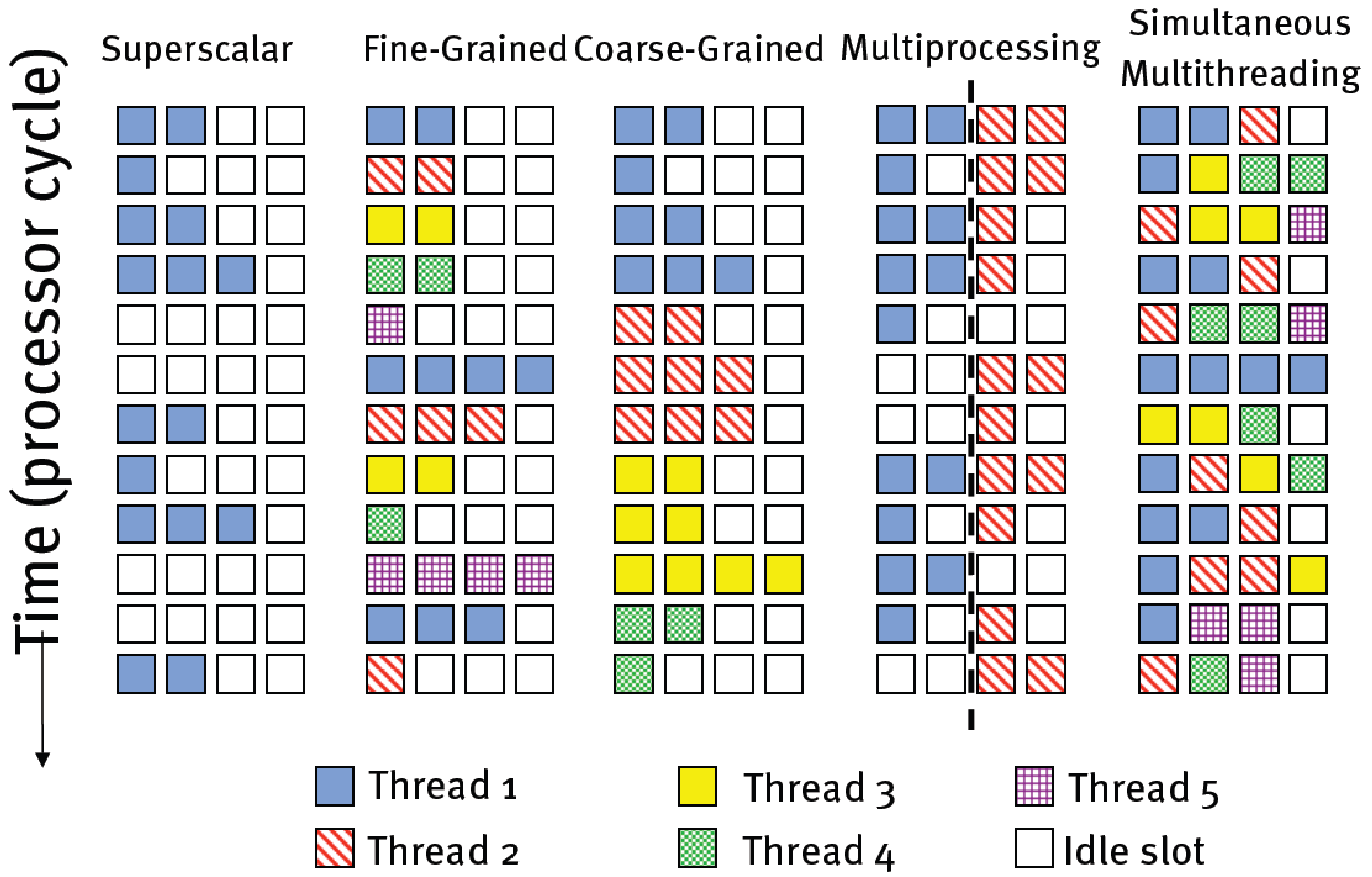


The third law

“Any sufficiently advanced technology is indistinguishable from magic.”

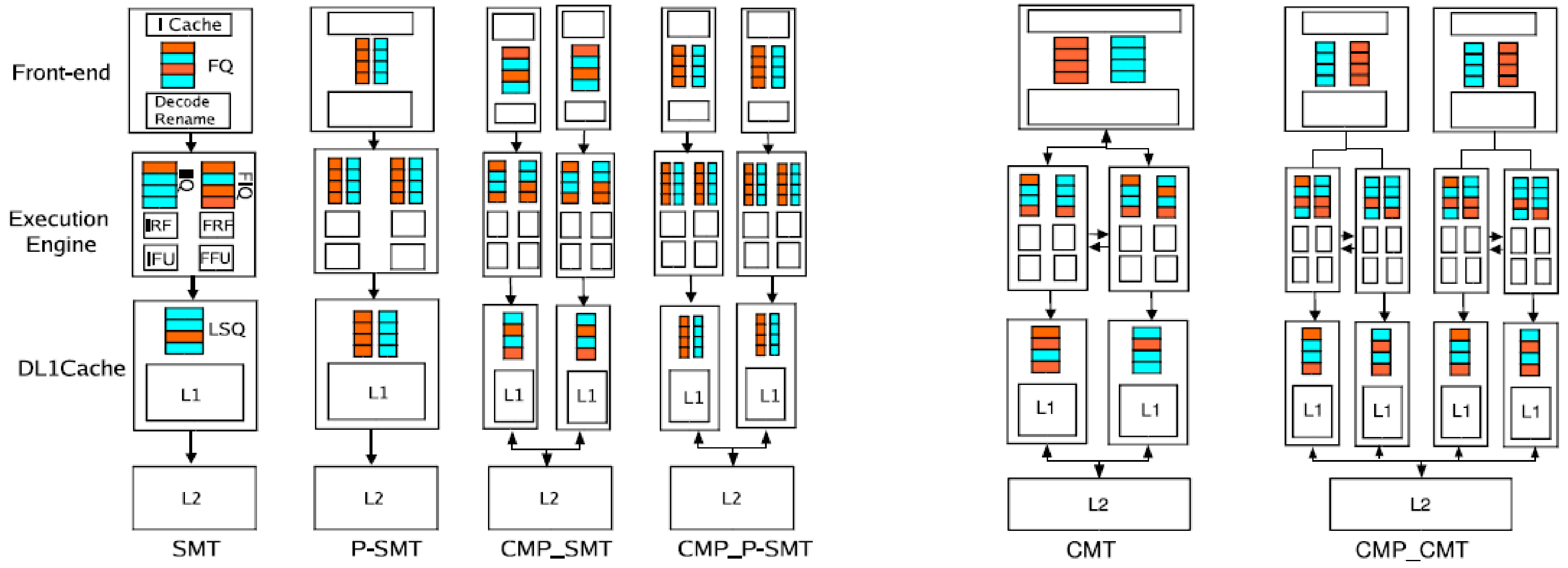
Arthur C. Clarke, “Profiles of The Future”, 1961 (Clarke’s third law)

Multithreaded Categories



Morgan Stanley

SMT – partitioned / clustered



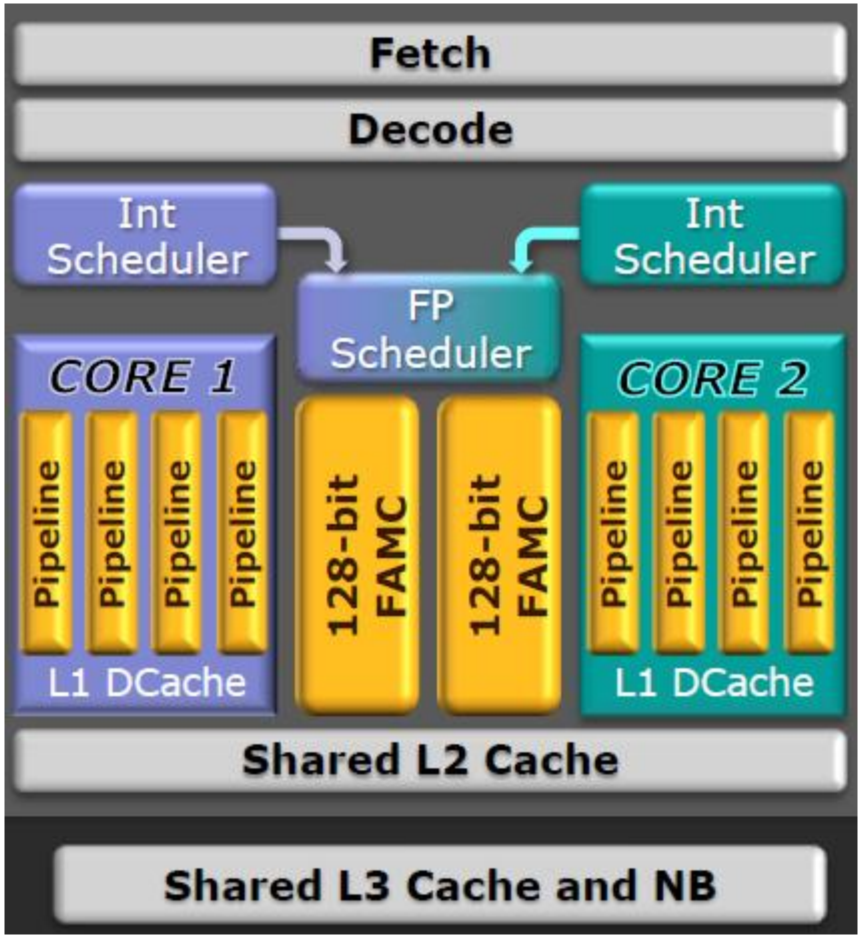
Difference lies is in the design decision what is
- **replicated** (register renaming logic, instruction pointer, ITLB, etc.)
- **partitioned** (reorder buffers, load/store buffers, various queues)
- **shared** (caches, micro architectural registers, execution units)

Will not help if one thread **monopolizes shared resources**, **cache effects** etc.

Morgan Stanley

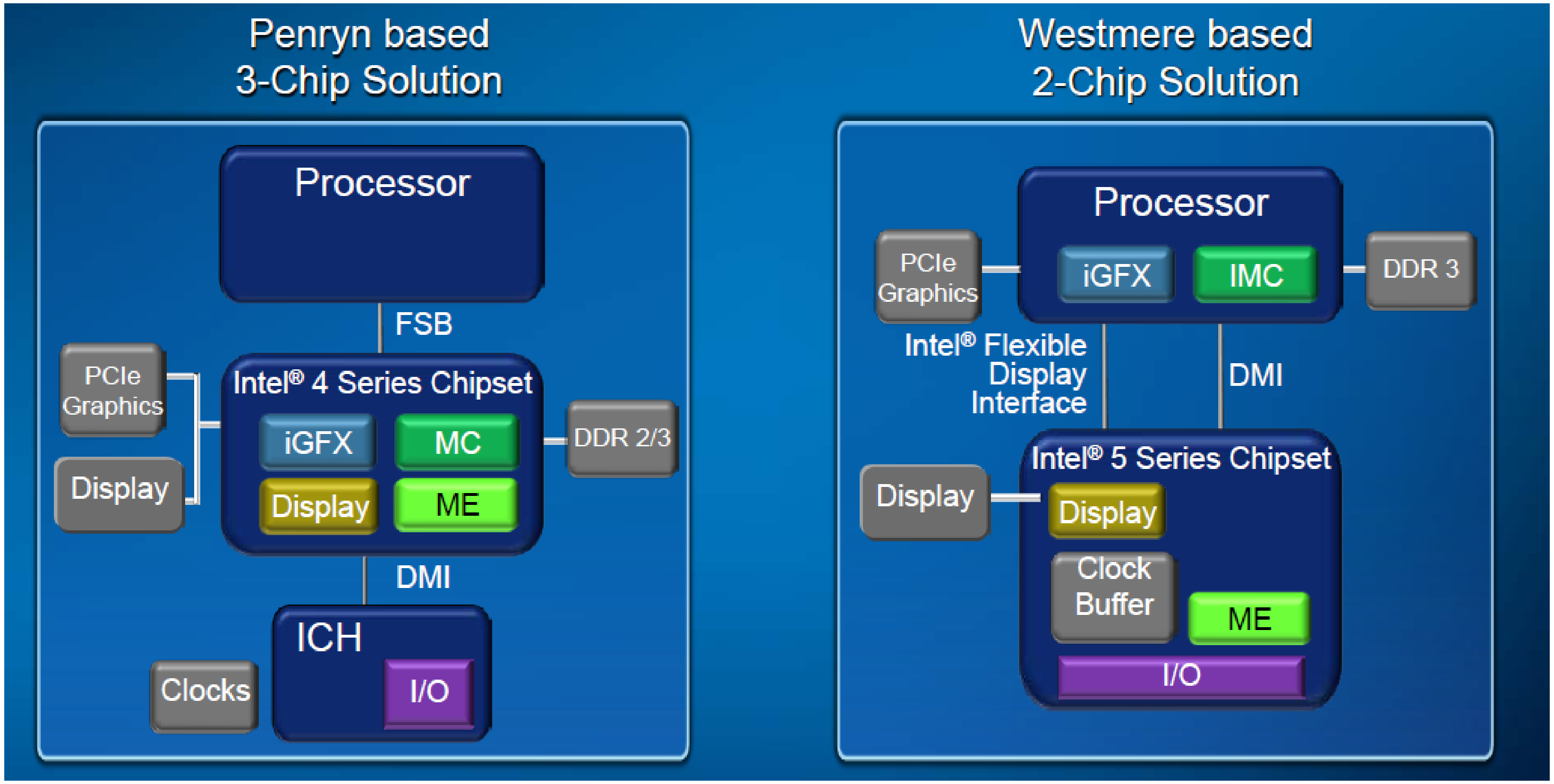
AMD Bulldozer

- Bulldozer implements a Cluster(-based) multiprocessing (CMT)
- Integrates two superscalar processors
- 2 Flexible Floating Point FAMC unit that can be dedicated or shared between the two core per cycle
- Independent integer and FP schedulers are provided



Morgan Stanley

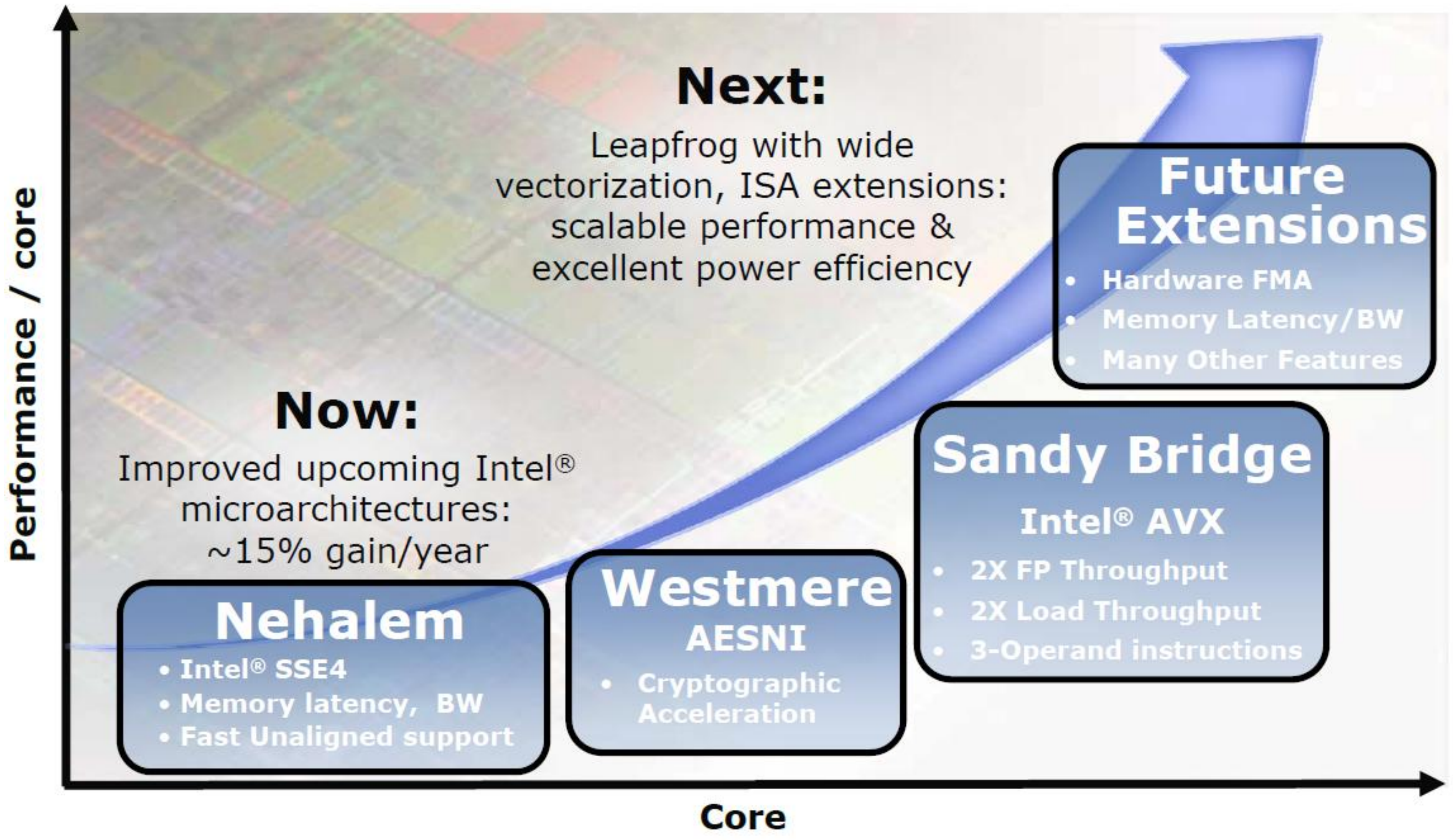
Keynotes: Intel Flexible Display Interface



Morgan Stanley

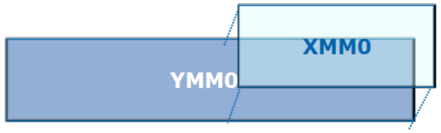
Intel Advanced Vector Extensions (AVX)

- Intel **Sandy Bridge Microarchitecture** features Intel **Advanced Vector Extension**

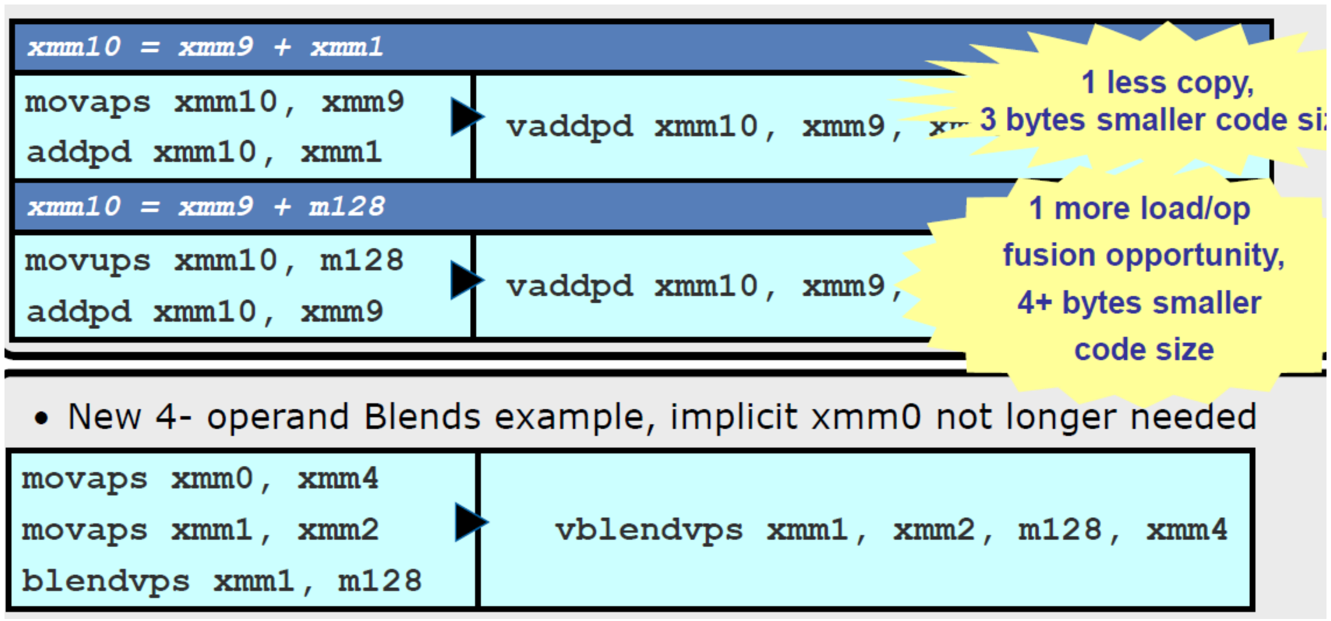


Intel AVX Key features and benefits

- **Wider Vectors** (128 to 256 bit with 2 128-bit load ports)



- Enhances Data Rearrangement (broadcast, mask loads and permute data)
- **Flexible unaligned memory access** support (e.g. no penalty for unalgined loads on aligned memory)
- Mixing AVX / SSE code may incur penalty (AVX 256 dirties upper 128 bits)
- A **new** 3- and 4- operand **instruction format**:



xmm10 = xmm9 + xmm1

```
movaps xmm10, xmm9      ▶   vaddpd xmm10, xmm9, x
addpd xmm10, xmm1
```

1 less copy, 3 bytes smaller code si

xmm10 = xmm9 + m128

```
movups xmm10, m128      ▶   vaddpd xmm10, xmm9,
addpd xmm10, xmm9
```

1 more load/op fusion opportunity, 4+ bytes smaller code size

- New 4- operand Blends example, implicit xmm0 not longer needed

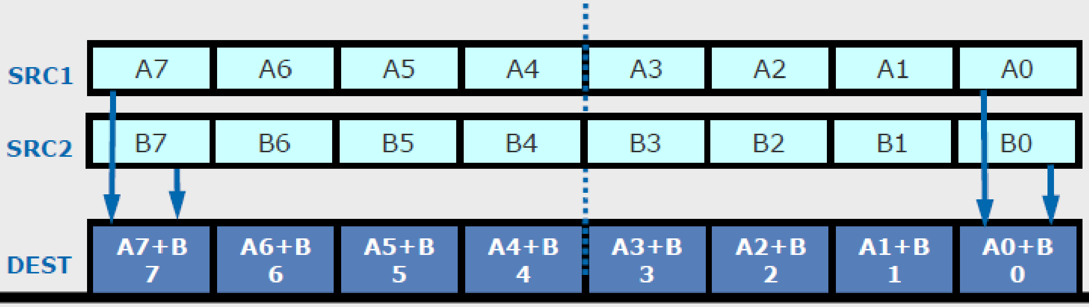
```
movaps xmm0, xmm4
movaps xmm1, xmm2       ▶   vblendvps xmm1, xmm2, m128, xmm4
blendvps xmm1, m128
```

Morgan Stanley

Intel AVX New Primitives

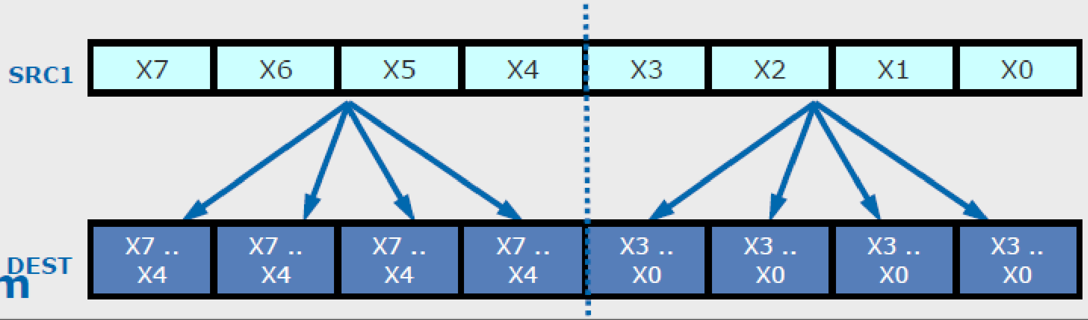
- **Simple in-lane instructions**
  - 2 lanes, 128 bit each

vAddPS dest, src1, src2



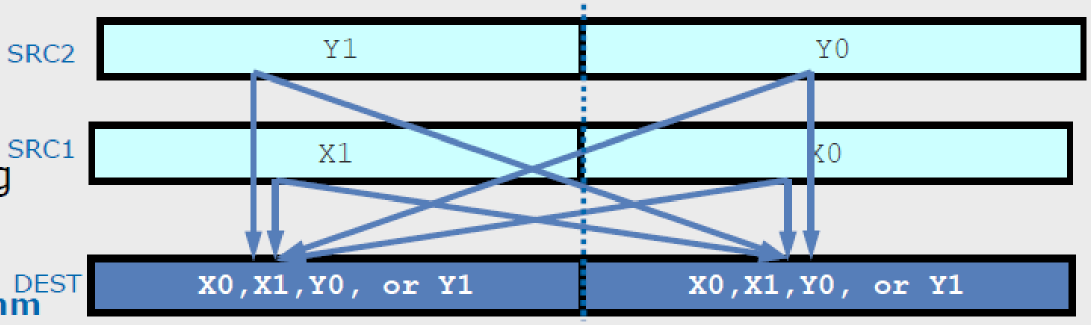
- **New in-lane PS and PD Permutes**
  - Permute controlled via immediate

vPermilPS dest, src, imm



- **New 128-bit permutes**
  - Useful for lane-crossing operations

vPerm2F128 dest, src1, src2, imm

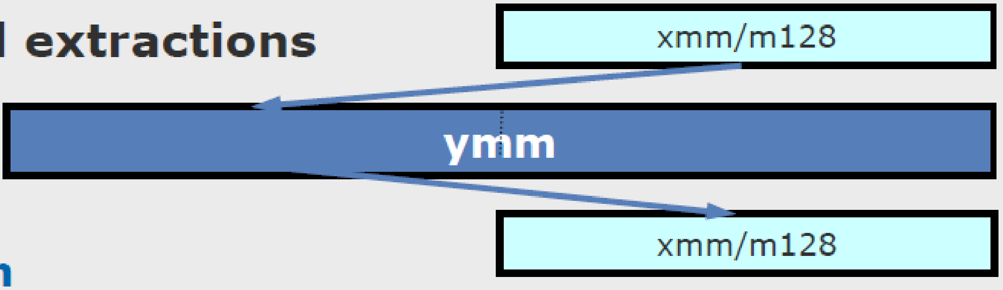


Intel AVX New Primitives

- **128-bit Insertions and extractions**
  - Useful for lane crossing operations

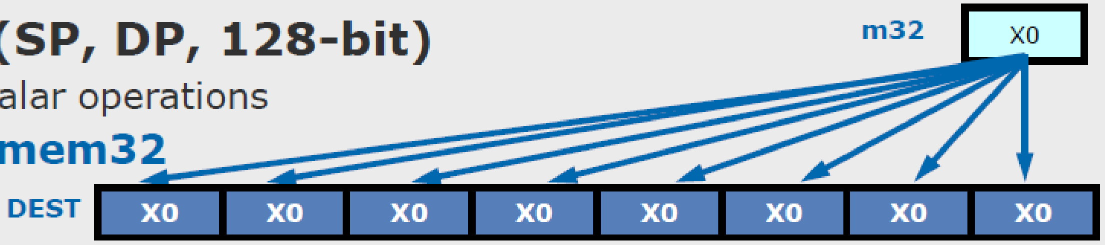
vInsertF128 dest, src, imm

vExtractF128 dest, src, imm



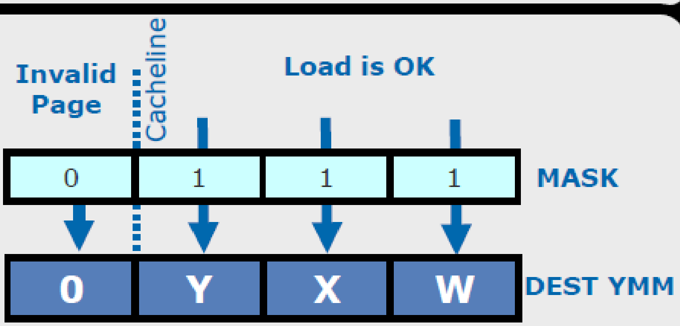
- **New Broadcast (SP, DP, 128-bit)**
  - Efficient Vector * Scalar operations

vBroadcastPS dest, mem32



- **New Conditional SIMD Loads and Stores**
  - Avoid page faults, segment violations, memory transaction if the mask is 0
  - Allow more automatic compiler vectorization

vMaskMovPD dest, mask, mem256



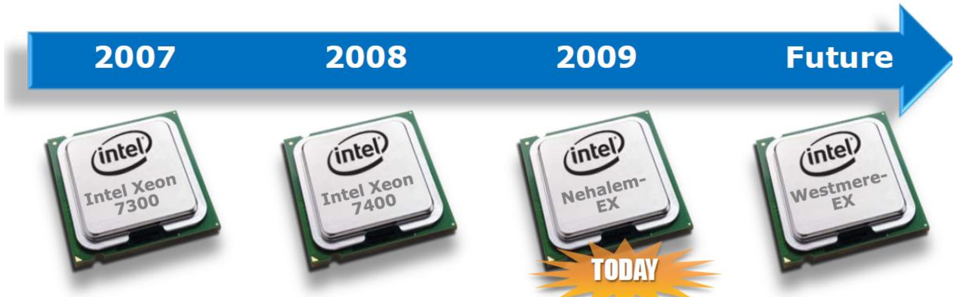
Intel AVX software support

- Intel Compiler: 11.1
- Intel IPP 6.1
- Intel MKL 10.2
- Intel TBB 2.2
- GNU GCC 4.4.1
- Microsoft Visual Studio 2010 Beta 2

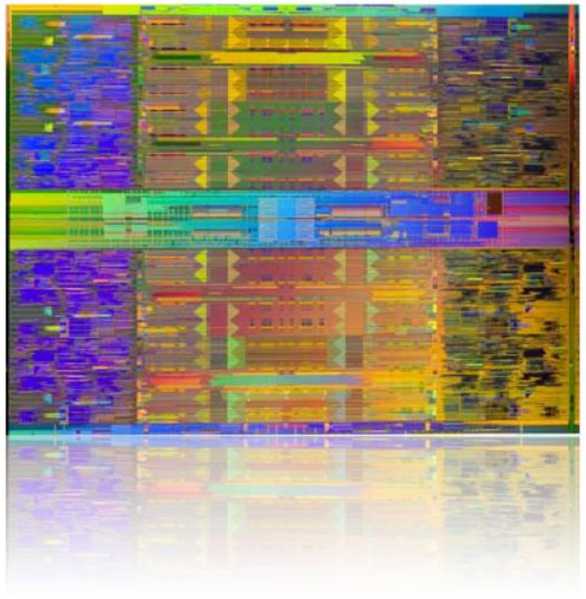
- **Update on Fused Multiply Add**: Not supported on Sandy Bridge Microarchitecture

Intel Nehalem-EX Architecture

- **Scalable** Nehalem-EX Architecture
- Reliability, Availability and Serviceability **(RAS)**
- The platform for **x86-64 High-End Computing**
- Several **uncore improvements**



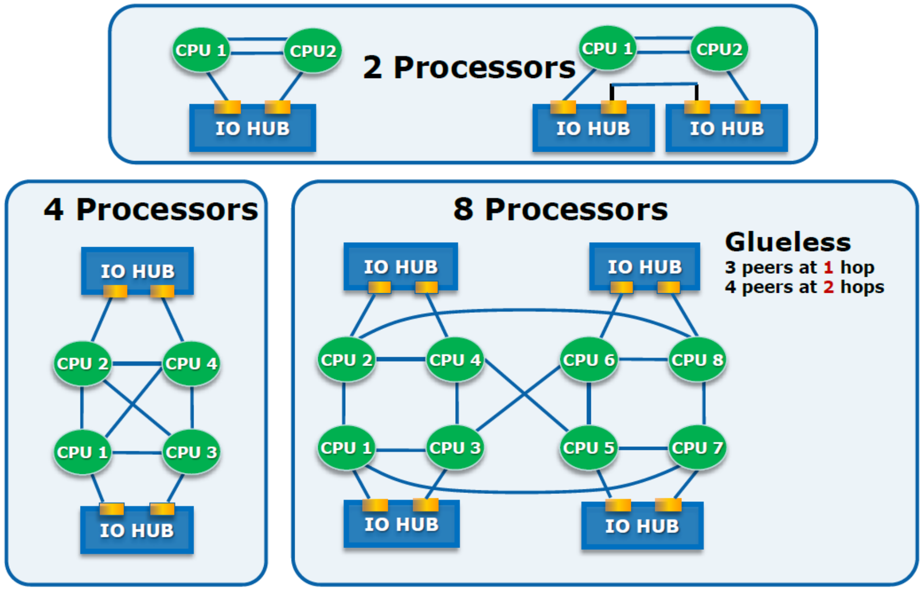
Expandability and Scalability
Reliability, Availability, and Serviceability (RAS)

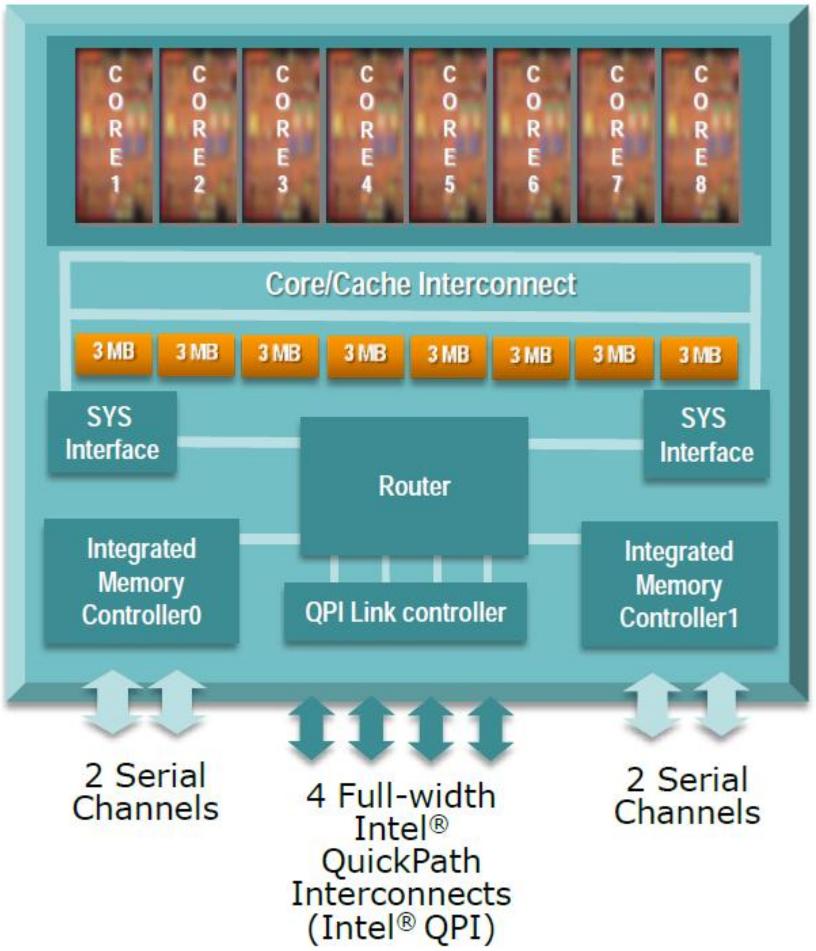


Morgan Stanley

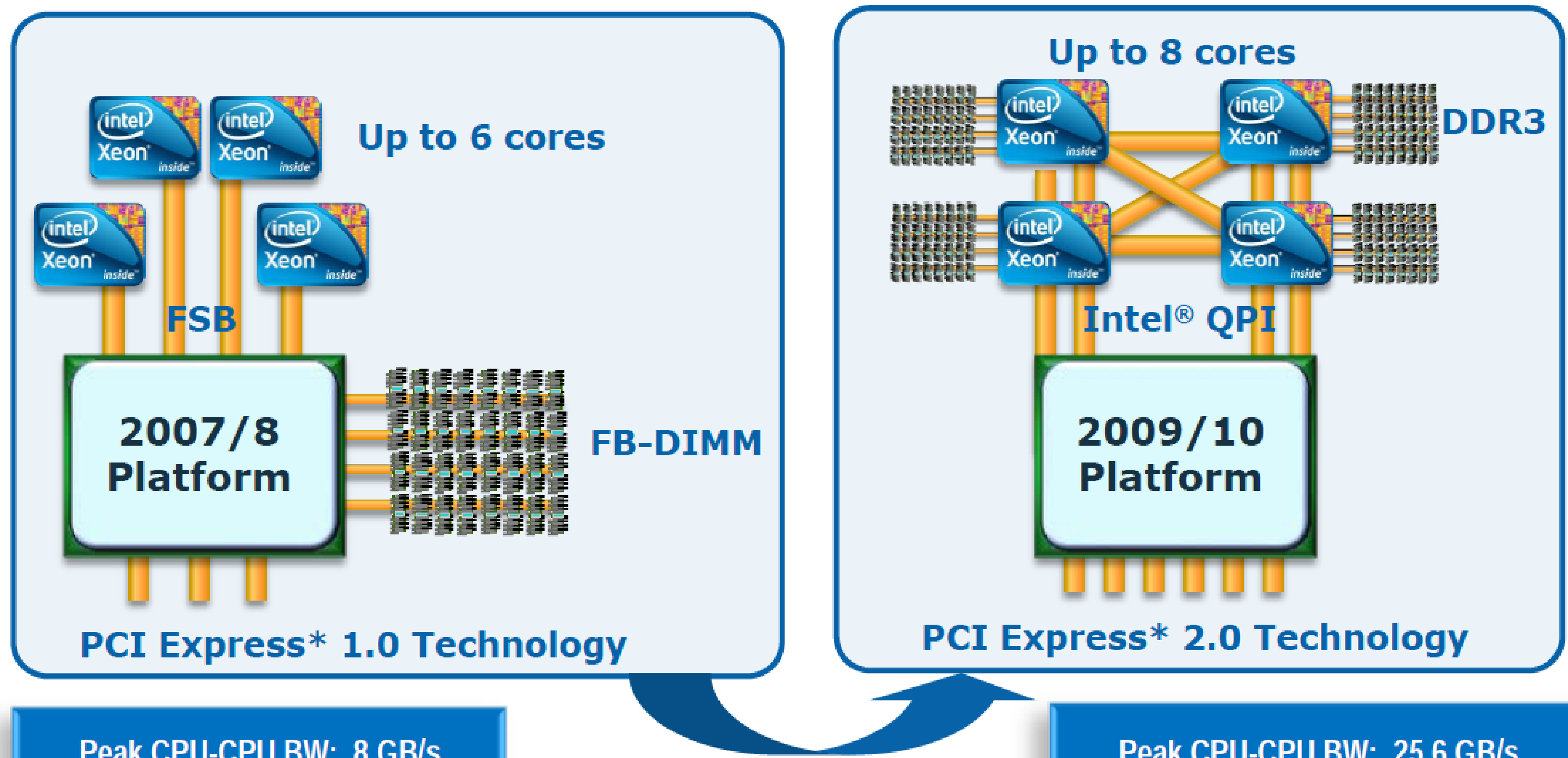
Scalable Processor Cores

- Up to **8 cores** per socket
- **24 MB shared last level cache**
- 2,4 and **up to 8+ processors**





Platform Capability Enhancement



Peak CPU-CPU BW: 8 GB/s

Peak CPU-Chipset: 34 GB/s

Peak Memory BW: 32 (FBD-667)

Peak I/O BW: 2.5GT/s , PCIe 1.0

Architectural Innovations:

- Front Side Bus (FSB) to Intel® QuickPath Interconnect with integrated memory controller
- FB-DIMM to DDR3
- PCI Express* Technology (PCIe) 1.0 to 2.0...

Peak CPU-CPU BW: 25.6 GB/s

Peak CPU-Chipset: 102.4 GB/s

Peak Memory BW: 200 (DDR3-1067)

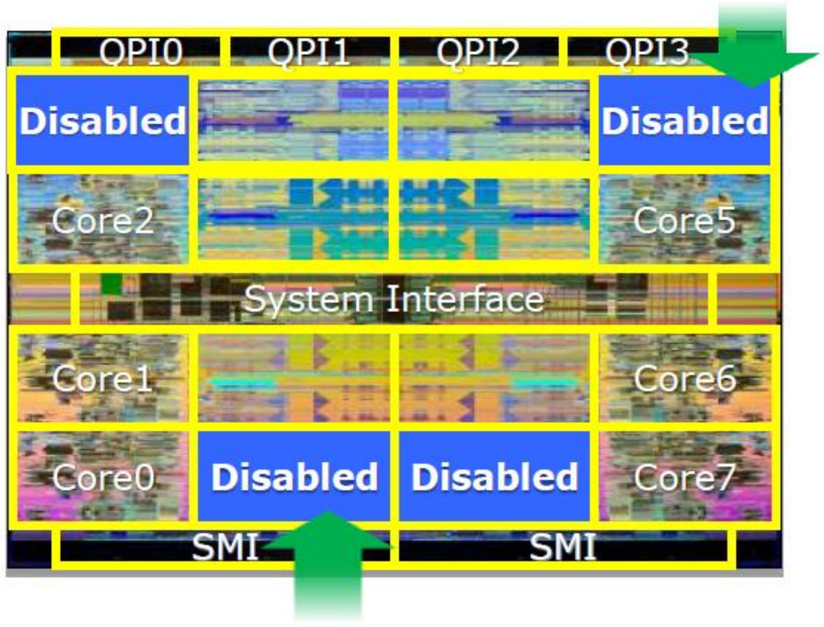
Peak I/O BW: 5GT/s, PCIe 2.0

Morgan Stanley

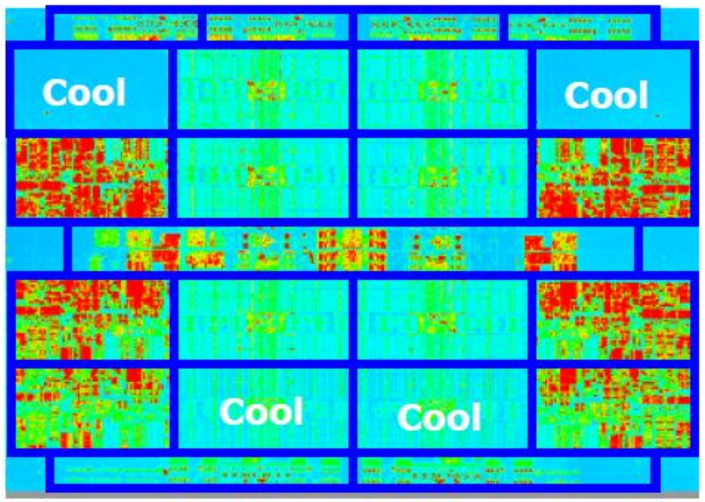
Power Optimization / Energy efficiency

Leakage Control at Core and Cache Level

New to NHM-EX: Power gating disabled cores



Power gating disabled cache

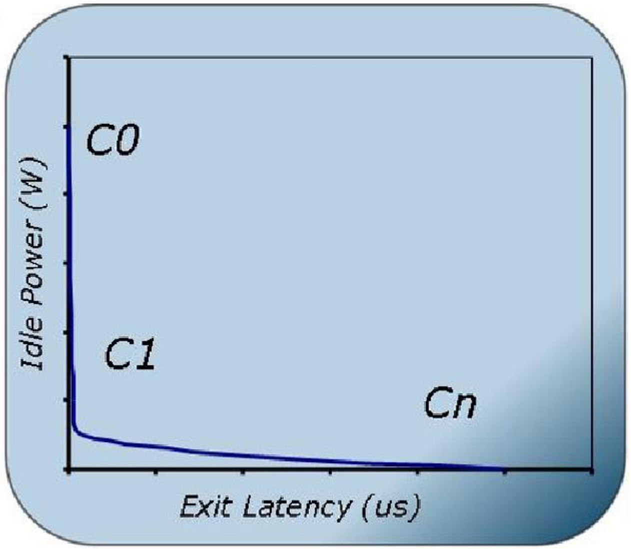


Infrared back-side emission shows core leakage suppressed and cache leakage reduced

Core Level: 40x leakage reduction when shut off
Cache Level: 35% during sleep & 83% when shut off

Reducing Idle Power Consumption

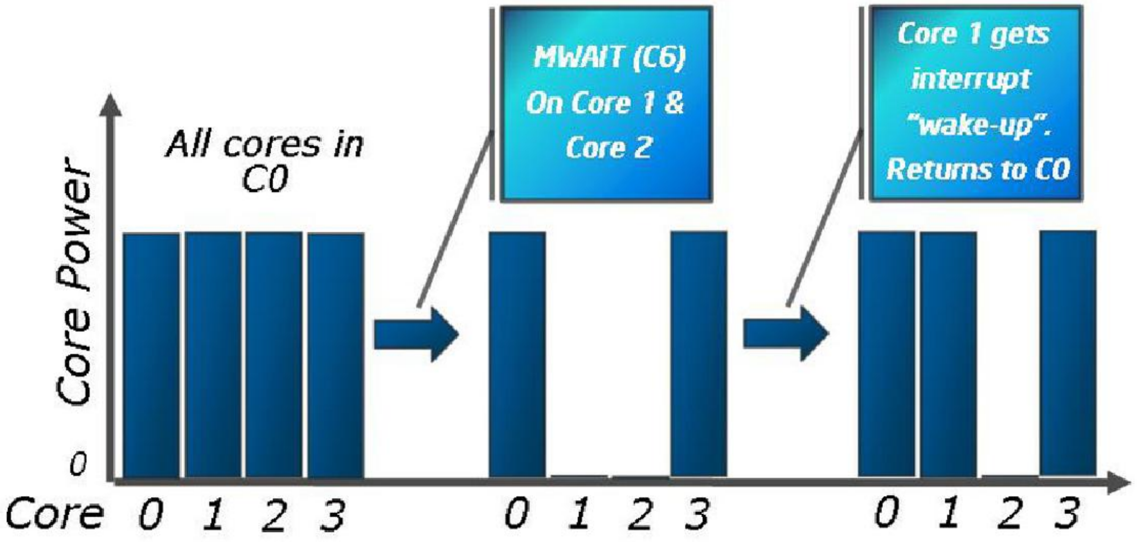
- OS initiates C-state en entry by Mwait instruction
  - **C-states** – “idle” states
    - Different C states means different exit latency
    - Windows CPU Power Management Framework (Windows 7)
  - **P-states** – “active” states (C0)

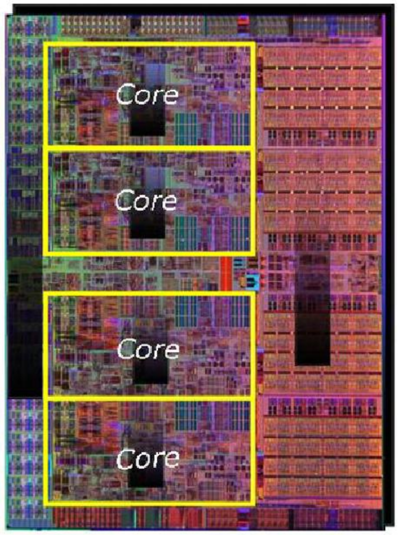


Morgan Stanley

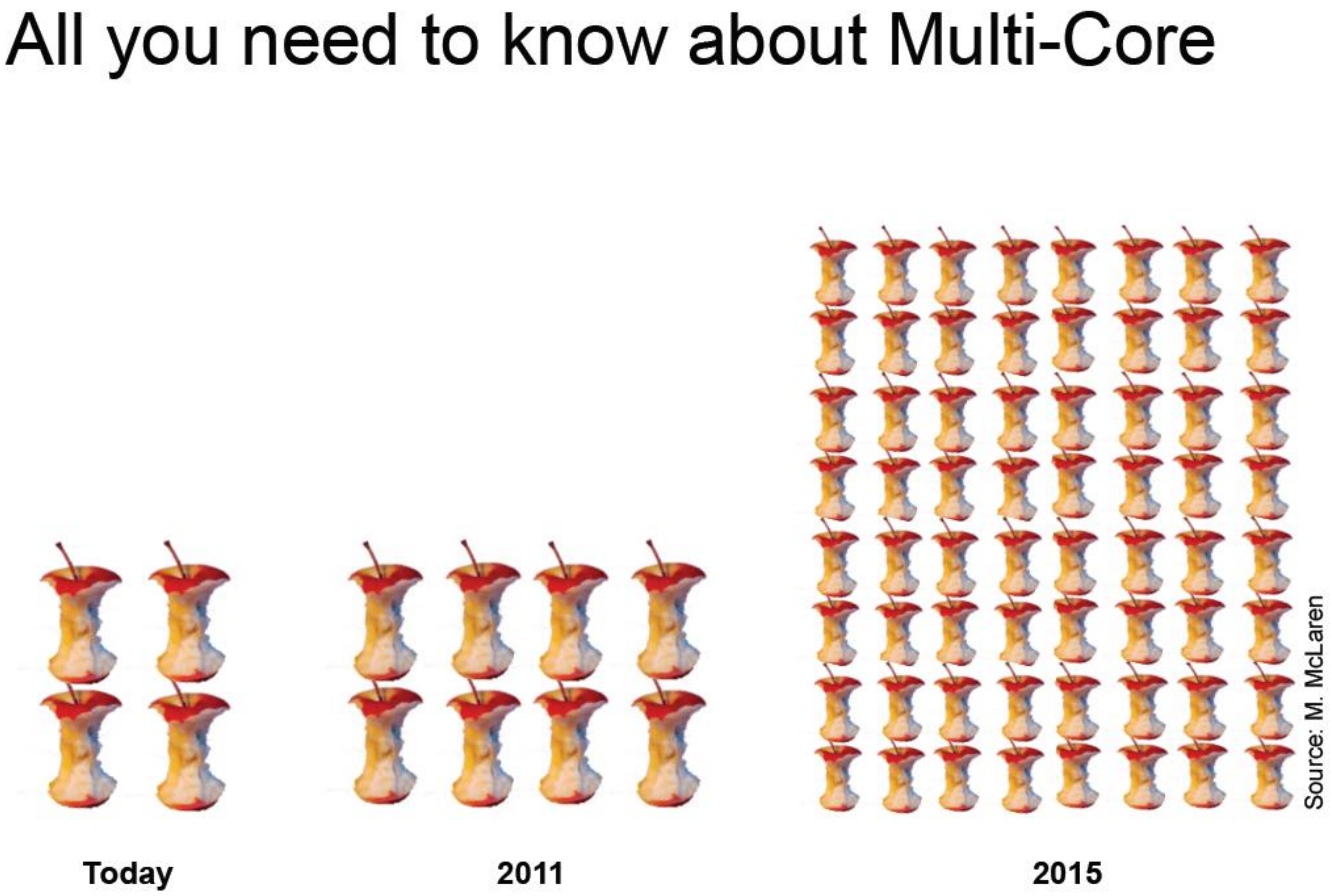
Deep Power Down Technology (C6)

- Integrated Power Gate enables **a per core C6 state** and individual cores transition to a **~0W Power State**
- Uncore logic? When all cores in C6, **package can transition to C6**





Near-Term Server Futures



Company server architecture

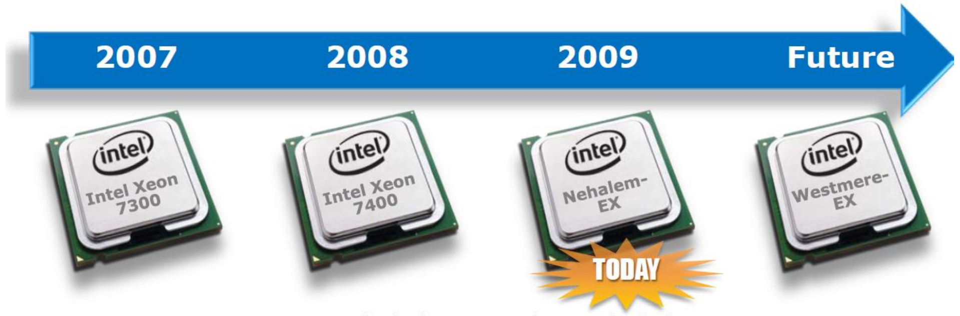
- HPC segment is moving towards high density commodity hardware
- **Past:** 2-socket server (2+2 / 4+4 cores), P4 / Opteron processors
- **Now**: 2-socket servers (4+4 cores), Core2 architecture
- **Future**: 2-socket servers (4+4 / 6+6 cores), Nehalem / Westmere architecture
- **Focus:** best performance per unit price, not absolute performance

What Are The New Cores Like?

- Just like old ones
- **Clock rates NOT going up**
  - **Incore improvements**: more FLOPS/tick, better virtualization support, HTT
  - **Uncore improvements**: integrated memory controllers, etc.
- **Takes a little pressure off of memory latency**
  - Memory used to grow twice as distant (in clocks) every eighteen months
- **Does not take pressure off of memory bandwidth**
- **Bandwidth**: 0.5B/FLOP is ‘good’ for HPC
  - Nehalem gives us 40 GB/s : (3Ghz * 4 F/tick * 4 cores * 2 sockets) = 0.4
- **Capacity**: up to 1.5, 3, 6GB/core (48GB in a 2s-4c system)

Prognostication

- **Nealem: wonderful gift of memory bandwidth**
  - 1.3x – 3x performance boost
- **Core count ++ means lower bandwidth/core**
  - Westmare @ 6 cores
  - Nehalem-EX @ 8 cores, but more bandwidth



Expandability and Scalability
Reliability, Availability, and Serviceability (RAS)

Evolutionary Improvements Running Out of Gas

- **S/W has been relying on H/W improvements**
  - Moore’s Law hid inefficiencies
    - “Every programmer’s job is to consume twice as much compute power every eighteen moths!”
  - Get three programmers together, and they’ll first have to define nine levels of abstraction to ensure optimal job security.
- **S/W needs to be come more efficient**
  - Requires effort and expertise from S/W part
  - Olukotun’s view 3 real dimensions of processor performance:
    - clock frequency
    - superscalar instruction issue
    - multiprocessing
  - First two have been pretty much exhausted, programmers need to switch.

The Memory Wall

- **2014: assume 400GF mainstream processor**
  - Aggressively multi-core
  - 0.5 B/FLOP means 200 GB/s, ~10x today's chips
    - **Opinion: won't happen in copper**
      - Optical connection
      - Stacked memory (more levels)
      - ?

    - ... maybe the hardware folks can save your sorry selves once again!

Q/A