3.2 A Family of 45nm IA Processors

Rajesh Kumar, Glenn Hinton

Intel, Hillsboro, OR

Nehalem is a family of next-generation IA processors for mobile, desktop and server segments implemented in 45nm high-κ metal-gate CMOS [1]. The family features a new system architecture, significantly enhanced Core architecture [2], innovations in power management and modular design. The 4-core 8MB-L3-cache die has 731M transistors. We introduce a coherent point–to-point link called QuickPath Interconnect that is the foundation for coherent communication between IA processors, chipsets, I/O hubs and coprocessors/accelerators for multiple generations. It features advanced power management, RAS capabilities and reduced hops/latency. At the physical level, it uses unidirectional variable-width differential signaling. The 45nm implementation features current-mode signaling with cascode driver, clock forwarding, source and sink termination, per-bit skew compensation and 2-tap driver equalization, as showin in Fig. 3.2.1. Signaling speed is up to 6.4GT/s in 45nm but the basic approach scales to 20GT/s using conventional copper interconnect [3].

An integrated memory controller supports 3 channels of DDR3 memory requiring a 1.4 to 1.6V supply, which would require a thick-oxide high-voltage transistor in the 1.1V 45nm process. The complexity and cost of developing two oxides on the high-κ metal-gate process is deemed excessive. Instead, we develop thin-gate circuit technology with a cascode topology, biased inputs and clamped outputs, as showin in Fig. 3.2.2. The controller performs *in situ* dynamic training per rank and per strobe group for higher performance.

This core increases instructions per cycle (IPC) significantly compared to the previous Merom architecture. It also improves floating-point performance, and virtualization and server performance without excessive pipelining and implements SSE4 for media and web acceleration. All architectural features are analyzed for power efficiency with key ideas generally targeted at improving more than 1% performance for 1% power increase. As an example, we evaluate alternative approaches for multi-threading, e.g., switch-on-event multi-threading (SoEMT) and choose 2-way SMT (simultaneous multi-threading), which allows for two concurrent logical processors in the core, exploits both parallelism under cache misses and higher functional unit utilization across threads with serial dependencies.

The memory system features 3 levels of on die caches: L1 32KB instruction and writeback data caches, 256KB private L2 with 10-cycle latency and shared L3 cache. L3 is modular from 2 to 24MB. For data integrity, L3 features double-bit error correction while the L2 and L1 have single-bit error correction. Most register files have parity.

The main modules: each core, LLC, memory controller and Quickpath link, have their own PLL/frequency and clock tree [4]. All cores have one VR while there is a separate VR for the shared memory system. A fine-grained synchronous low-latency FIFO that allows effectively arbitrary voltage/frequency combinations between communicating agents is implemented to avoid indeterminism problems with asynchronous FIFO in testing and system debug.

Idle power is becoming an important specification in all market segments. We integrate core-level power-gate transistors that shut off idle cores individually, completely removing idle power, as shown in Fig. 3.2.3. The architectural state is stored in chip memory and restored when that core again becomes active. Save/restore latency is minimized to make this scheme transparent to OS or software applications. Targeted innovations in circuit and process technology allowed aggressive design specifications to be met: <1% performance loss and >100× reduction in power. A total of 1.5m of total width per core of ultra-low-leakage PMOS transistors are used with the gate terminal adaptively switched to the highest available chip voltage to further reduce leakage. Traditional on-chip interconnect provides too high a resistance at the large currents involved so a new package-like very-thick (7μm) low resistance (~10× lower resistance than typical interconnect layers) on-die M9 tailored to this application is deposited on silicon, as shown in Fig. 3.2.4.

We integrate a dedicated microcontroller for power management with custom ISA and hardware optimized for power management arithmetic. The microcontroller monitors the operating environment (temperature, power consumption, etc.), the voltage/frequency characteristics of that particular die and the power-state requests of each of the cores, and resolves these independent requests to arrive at a unified power-management state and voltage for a specified frequency, thus reducing significant complexity and guard band. Implementing the power management algorithms via firmware enables flexible algorithms, and also provides the ability to enhance those algorithms both before and after shipping the product. The entire instruction space for this embedded microcontroller is patchable. The microcontroller monitors power in real time using integrated current sensor logic. If the total power consumed is less than safe limits (due to some cores turned off or low application power), it raises voltage/ frequency to convert this power headroom into performance.

Power efficiency in design is addressed by 1) static CMOS 2) low-voltage operation and 3) smaller layout. In the quest for performance at all cost, high-performance-microprocessor designers have developed advanced circuit families: domino, self-resetting, low-voltage swing (LVS) [5], with each new family being slighter faster than the previous, but at the expense of increasing design complexity and power consumption. For example, domino circuits consume 2 to 5× more power for same functionality as static CMOS. This processor uses static CMOS for all datapath and control, eliminating domino circuits, tri-state buses and similar power-hungry techniques. To do so without increasing latency or pipelining, innovations in fundamental algorithms, such as instruction decode, are made, decreasing the total logical work. Low-voltage operation, down to 0.7V is enabled by several circuit and power-plane changes. The most critical analog circuitry like PLL is on a separate 1.4 to 1.6V supply. Generic analog and L3 SRAM are on a 0.9 to 1.1 V supply. All the 6T SRAM cells in the core are replaced by more robust 8T SRAM. This chip also improves chip density by about 20% (iso-process), by innovations in memory arrays and layout techniques. Instead of the usual memory-array layout that embeds read ports with each storage element, read ports were clustered into a separate cell, as shown in Fig. 3.2.5, for all highly ported register file blocks. At iso-delay this reduced local read-bitline length 60%, read pull-down size 50%, global bitline length 15%, wordline driver size 10%, and layout area 10% (7 or 8 ports) to 30% (12+ ports). Figure 3.2.6 shows the die micrograph.

Acknowledgment:
We thank the entire Nehalem team for the impeccable execution.

References:

[1] K. Mistry, C. Allen, C. Auth, et al., “A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging”, *IEDM Dig. Tech. Papers*, Dec. 2007.

[2] N. Sakran, M. Yuffe, M. Mehalel, et al., “The Implementation of the 65nm Dual-Core 64b Merom Processor”, *ISSCC Dig. Tech. Papers*, Feb. 2006.

[3] B. Casper, J. Jaussi, F. O’Mahony, et al., “A 20Gb/s Forwarded Clock Transceiver in 90nm CMOS”, *ISSCC Dig. Tech. Papers*, Feb. 2006.

[4] N. Kurd, J. Douglas, P. Mosalikanti and R. Kumar, “Next Generation Intel® Micro-Architecture (Nehalem) Clocking Architecture”, *IEEE Symp. VLSI Circuits*, Jun. 2008.

[5] D.J. Deleganes, M. Barany, G. Geannopoulos, et al.,”Low-Voltage-Swing Logic Circuits for a 7Ghz X86 Integer Core”, *ISSCC Dig. Tech. Papers*, Feb. 2004.

978-1-4244-3457-2/09/$25.00 ©2009 IEEE

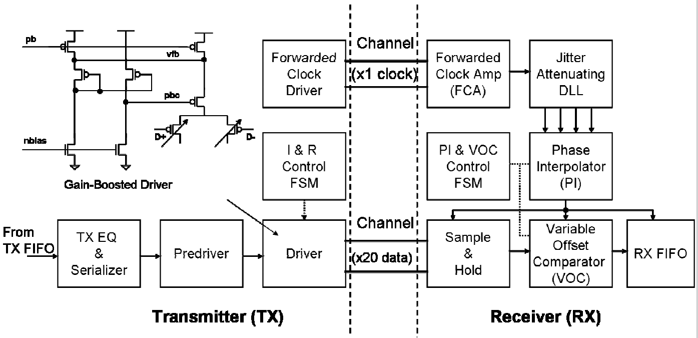

Figure 3.2.1: QuickPath physical layer showing differential signaling, clock forwarding and receiver block.

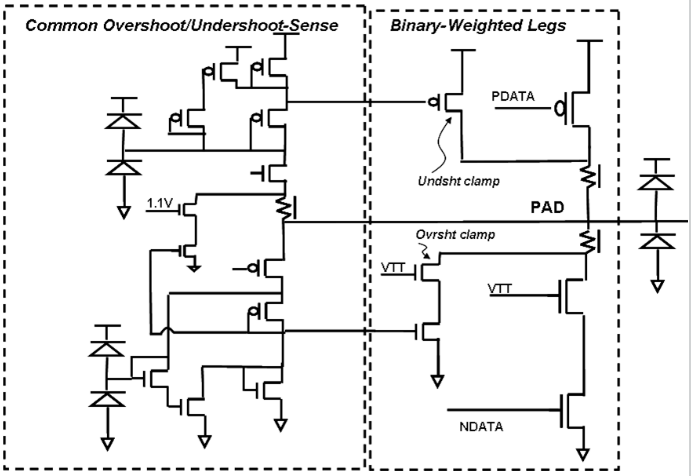

Figure 3.2.2: DDR3 driver using thin gate oxide process showing driver, clamp and bias circuitry.

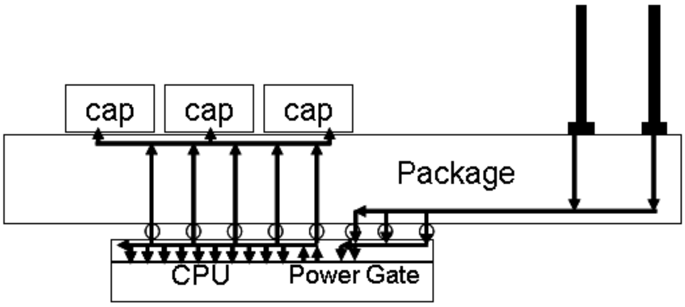

Figure 3.2.3: Power gate transistor location in the power delivery flow.

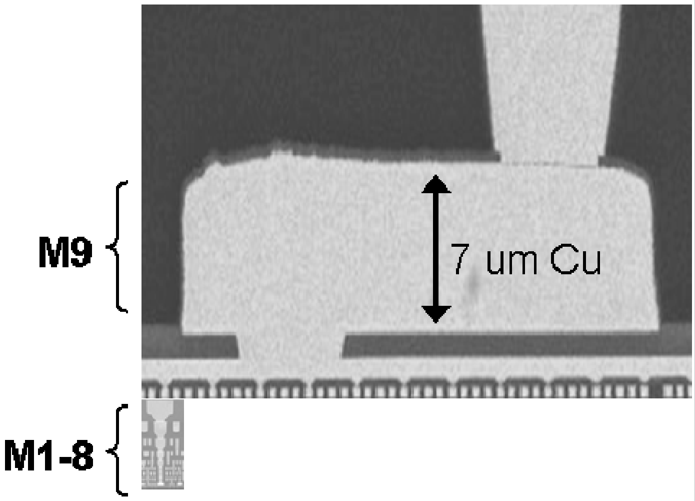

Figure 3.2.4: 45nm interconnect stack showing low-resistance M9 used for power gate.

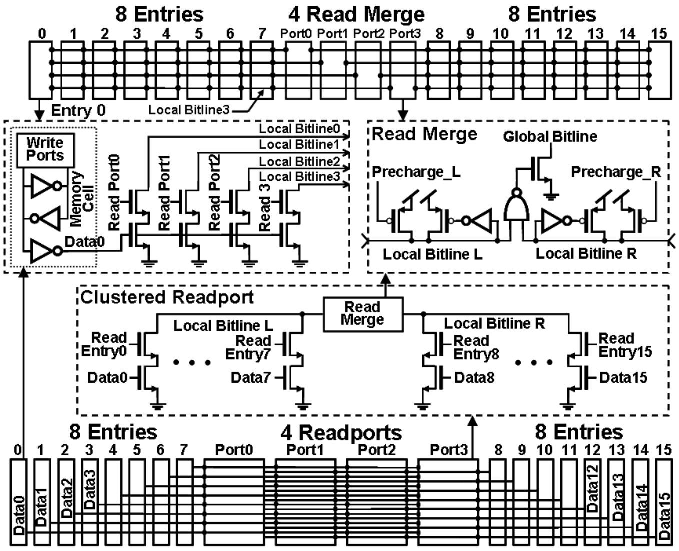

Figure 3.2.5: Clustered layout technique for memory arrays.

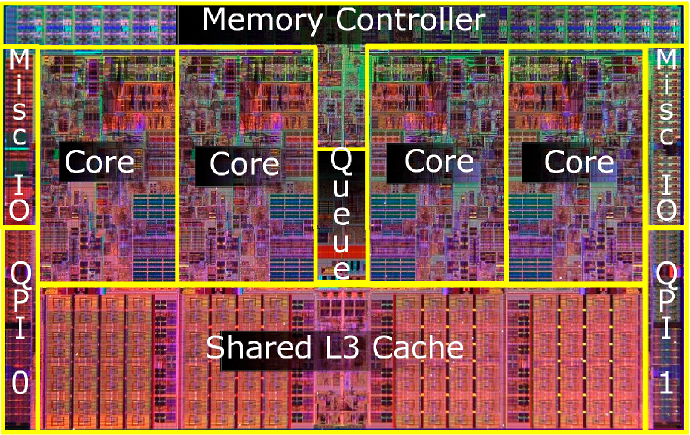

Figure 3.2.6: Nehalem 4-core die micrograph.