



DECLARATION OF GORDON MACPHERSON

I, Gordon MacPherson, am over twenty-one (21) years of age. I have never been convicted of a felony, and I am fully competent to make this declaration. I declare the following to be true to the best of my knowledge, information and belief:

1. I am Director Board Governance & IP Operations of The Institute of Electrical and Electronics Engineers, Incorporated ("IEEE").
2. IEEE is a neutral third party in this dispute.
3. I am not being compensated for this declaration and IEEE is only being reimbursed for the cost of the article I am certifying.
4. Among my responsibilities as Director Board Governance & IP Operations, I act as a custodian of certain records for IEEE.
5. I make this declaration based on my personal knowledge and information contained in the business records of IEEE.
6. As part of its ordinary course of business, IEEE publishes and makes available technical articles and standards. These publications are made available for public download through the IEEE digital library, IEEE Xplore.
7. It is the regular practice of IEEE to publish articles and other writings including article abstracts and make them available to the public through IEEE Xplore. IEEE maintains copies of publications in the ordinary course of its regularly conducted activities.
8. The article below has been attached as Exhibit A to this declaration:

A.	Blaine Stackhouse, et al., "A 65 nm 2-Billion Transistor Quad-Core Itanium Processor", published in IEEE Journal of Solid-State Circuits (Volume: 44, Issue: 1, page(s) 18 - 31), date of publication December 30, 2008.
----	--

9. I obtained a copy of Exhibit A through IEEE Xplore, where it is maintained in the ordinary course of IEEE's business. Exhibit A is a true and correct copy of the Exhibit, as it existed on or about December 19, 2022.

Google Exhibit 1128 Google v. Singular

10. The article and abstract from IEEE Xplore show the date of publication. IEEE Xplore populates this information using the metadata associated with the publication.
11. Blaine Stackhouse, et al., "A 65 nm 2-Billion Transistor Quad-Core Itanium Processor", published in IEEE Journal of Solid-State Circuits (Volume: 44, Issue: 1, page(s) 18 - 31), date of publication December 30, 2008. Copies of this publication was made available no later than the last day of the publication month. The article is currently available for public download from the IEEE digital library, IEEE Xplore.
12. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

I declare under penalty of perjury that the foregoing statements are true and correct.

Executed on: 12/19/2022

DocuSigned by:
Gordon Macpherson
E768DB210F4E4EF...

EXHIBIT A

A 65 nm 2-Billion Transistor Quad-Core Itanium Processor

Blaine Stackhouse, Sal Bhimji, Chris Bostak, Dave Bradley, Brian Cherkauer, Jayen Desai, Erin Francom, Mike Gowan, Paul Gronowski, Dan Krueger, Charles Morganti, and Steve Troyer

Abstract—This paper describes an Itanium processor implemented in 65 nm process with 8 layers of Cu interconnect. The 21.5 mm by 32.5 mm die has 2.05B transistors. The processor has four dual-threaded cores, 30 MB of cache, and a system interface that operates at 2.4 GHz at 105 °C. High speed serial interconnects allow for peak processor-to-processor bandwidth of 96 GB/s and peak memory bandwidth of 34 GB/s.

Index Terms—65-nm process technology, circuit design, clock distribution, computer architecture, microprocessor, on-die cache, voltage domains.

I. OVERVIEW

THE next generation in the Intel Itanium processor family code named Tukwila is described. The 21.5 mm by 32.5 mm die contains 2.05 billion transistors, making it the first two billion transistor microprocessor ever reported. Tukwila combines four ported Itanium cores with a new system interface and high speed serial interconnects to deliver greater than 2X performance relative to the Montecito and Montvale family of processors [1], [2]. Tukwila is manufactured in a 65 nm process with 8 layers of copper interconnect as shown in the die photo in Fig. 1. The Tukwila die is enclosed in a 66 mm × 66 mm FR4 laminate package with 1248 total landed pins as shown in Fig. 2.

A block diagram of the Tukwila processor is shown in Fig. 3. The die contains four multi-threaded high performance 64 bit cores. Associated with each core is 6 MB of level three cache implementing the Intel Cache Safe Technology [3]. A system interface is designed around a 12 port crossbar router that allows communication between the four cores, two home agents, and six IO channels. Associated with each home agent is a 1 MB directory cache in support of a directory-based cache coherence protocol. Dual integrated memory controllers allow communication to system memory through four full duplex FBD2 channels with a peak bandwidth of 34 GB/s. Four full width and two half width Intel QuickPath Interconnects (QPI) [4] allow processor to IO and processor-to-processor communication at a peak bandwidth of 96 GB/s. To connect the system interface to the core and IO physical layer, Tukwila implements a synchronizer and routing architecture that is distributed across the die. Finally, the charge rationing (QR) controller monitors chip activity factor, and together with the Tukwila clock system, allows

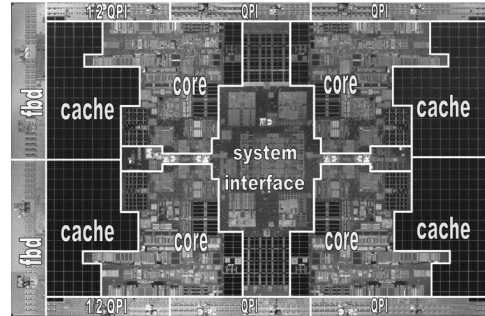


Fig. 1. Die photo.

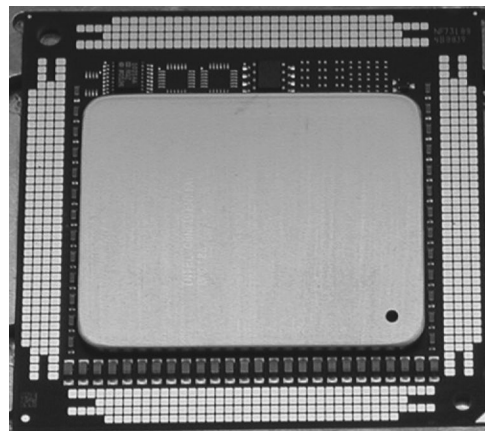


Fig. 2. Package photo.

dynamic modulation of the core voltage and frequency within a fixed power envelope.

Tukwila circuitry is partitioned as depicted in the table in Fig. 4. The cache, QR, and IO circuits are operated at high fixed voltages to ensure reliable circuit operation, while the core and system interface circuits are run at lower voltages to maximize power efficiency. The circuit design specific to each of these portions of the die will be covered in further detail in the remainder of this paper. Key emphasis areas will include low voltage circuit operation and circuit design in the presence of process variability.

II. CORE

This processor integrates 4 high speed dual-threaded cores onto a single die. Together with the Intel QPI links, FBD2 memory interfaces and system interface logic, Tukwila contains more than three times the logic circuits of its predecessor. To maximize performance in a given power envelope, the

Manuscript received March 31, 2008; revised August 13, 2008. Current version published December 24, 2008.

The authors are with Intel Corporation, Fort Collins, CO 80528 (e-mail: blaine.stackhouse@intel.com).

Digital Object Identifier 10.1109/JSSC.2008.2007150

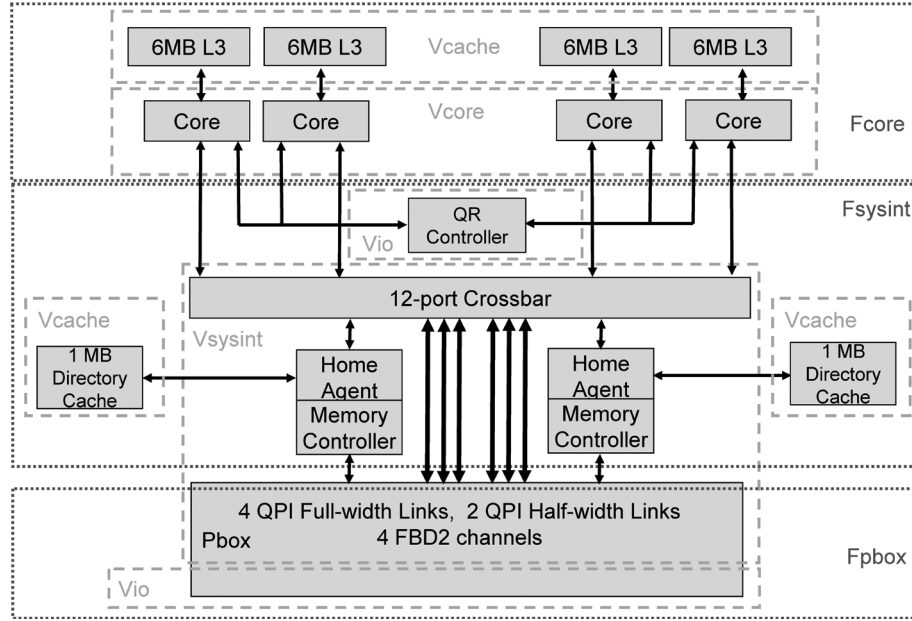


Fig. 3. Block diagram.

	Devices (million)	Area (mm ²)	Voltage (volts)	Frequency (GHz)
Core logic	430	276	0.9-1.2	Variable
L3 cache	1,420	191	1.1	Variable
SysInt logic	152	107	0.9-1.2	2.4
IO logic	39	123	1.1	2.4
QR logic	4	3	1.1	0.8

Fig. 4. Chip statistics.

voltage and frequency at which the cores operate can scale dynamically. Consequently, V_{CCmin} requirements for the cores are harder (lower V_{CCmin}) than previous generations which presents unique design challenges for the core's highly custom logic. A pulse based latching methodology and other challenges in moving the core from the 90 nm process to the 65 nm process generation, such as device variation, must also be solved.

Since the processor core is a port from 90 nm to 65 nm, substantial schedule savings are realized by using the many existing pulse-latch and entry-latch structures (Fig. 5) and placement from the previous generation. Pulse-latches are the dominant static state retention device on the die, excluding caches. Entry-latches retain state, produce a phase based monotonic output, and are used to drive dynamic logic circuits. Pulsed writes into these latches are self-timed, and can gain no margin as the clock frequency is reduced. Pulsed writes become even more difficult as V_{CC} is reduced with V_t consuming a larger portion of the supply and V_t increases further as temperature decreases. Consequently, only the very peak of the pulse is effective for writing (Fig. 6). Choosing

to make the pulse width wider can improve the write margin for a given pulsed structure, but also increases race exposure which requires extra effort to mitigate.

There are two circuits used to create pulse clocks (Fig. 7). Clock gaters typically drive a large number of latches along a short wire. A transfer gate is included in the internal delay chain that determines the clock pulse width. The slope of the transfer gate output is matched to the pulse latches, enabling the pulse width generated by the gaters to track latch write characteristics across PVT conditions. All gaters have programmable pulse widths. In new designs, a 20% wider pulse is software programmable. Ported designs have a metal option for an 8% wider pulse should it be needed. The other pulse generating circuit is a local pulse generator which can drive one to two latches when a larger gater is not practical. This structure requires the output pulse to reach the high V_{IH} of a feedback buffer before beginning to turn off the pulse. To further ensure a full-rail pulse, the output drive is $3\times$ that of the pulse gater, relative to the allowed output loading.

To achieve a wide PVT operating region, a new simulation methodology of all 10 million non-static circuits on the die (excluding L2 and L3 caches) was developed. This method requires specific functionality across 7 process corners, crossed with voltage from 0.7 V to 1.35 V, and temperature from -10°C to 125°C . In addition, multiple targeted device variation penalties are applied per circuit, which operate to make that device simulate worse than the design intent. To have less than 1% yield loss due to these circuits, a (root mean square) total of 6.4σ of transistor length, width, and V_t variation is applied to the FETs of each circuit for each robustness simulation measurement. The amount of variation applied to each FET is proportional to the effect it has on that particular measurement. Fig. 5 shows the variation applied to each transistor during a latch write-0 robustness simulation.

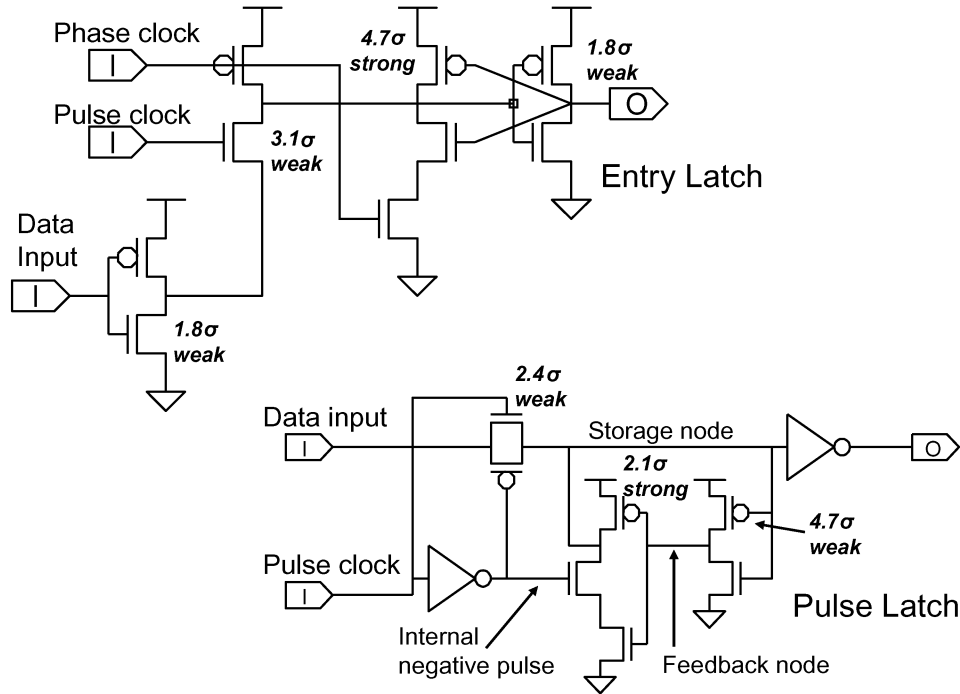


Fig. 5. Latches along with the applied variation used for write-0 robustness simulations.

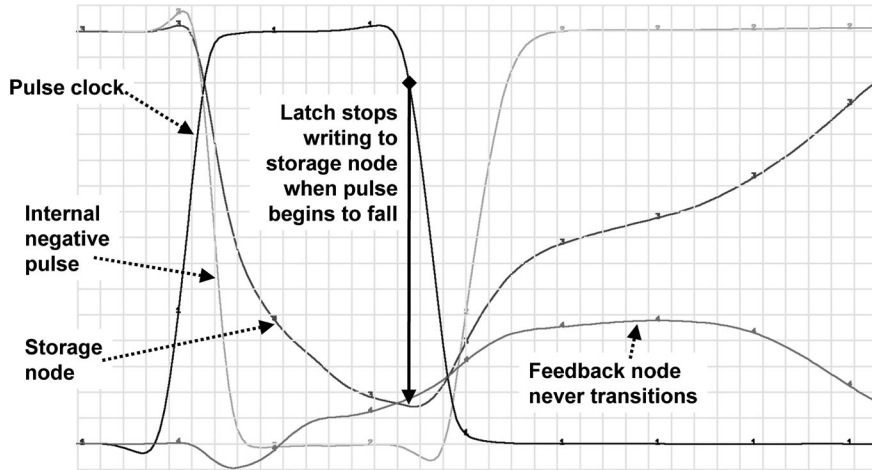


Fig. 6. Spice waveforms showing write-0 failure in pulse latch.

III. CACHES

The processor contains over 30 MB of on-die cache organized into three levels of hierarchy plus a directory cache for system coherency as shown in Fig. 8. This large quantity of on-chip memory presents significant challenges, especially in die area, power, yield, and error rates. The L2 and L3 caches contain redundant elements to allow for repair at manufacturing. ECC, parity, and Intel Cache Safe technology are used to alleviate the effects of VCC_{min} issues and soft errors. To address power and die area constraints, the L2 and L3 caches use the smallest available Intel 6 T SRAM cell [5]. While this enables the placement of over 30 MB of SRAM on the Tukwila die, it comes at the expense of VCC_{min} and performance. To offset these detrimental impacts, all caches using this memory cell are placed on a separate Vcache power supply that is maintained

at a fixed, higher voltage, regardless of any variations in the core voltage. All signals crossing the Vcache/Vcore boundary must be voltage translated with minimal area and timing overhead. Clock paths are maintained on the core supply to avoid any potential clock skew penalties, and data and control signals are only translated after they are combined with the clock. The voltage conversion incurs a 3% area penalty, but this is small compared to the 10–15% area penalty for a larger memory cell with better VCC_{min} characteristics.

Placing the L3 cache onto the Vcache power supply is relatively straightforward due to its physical separation from the core and its simple interface. The L3 cache follows the sub-array based design approach and the clockless cache design principles of previous Itanium products [6], [7] in which all clocks are confined to the datapath at the interface between the core and the cache. The entire L3 cache is placed on the Vcache supply, and all input and output

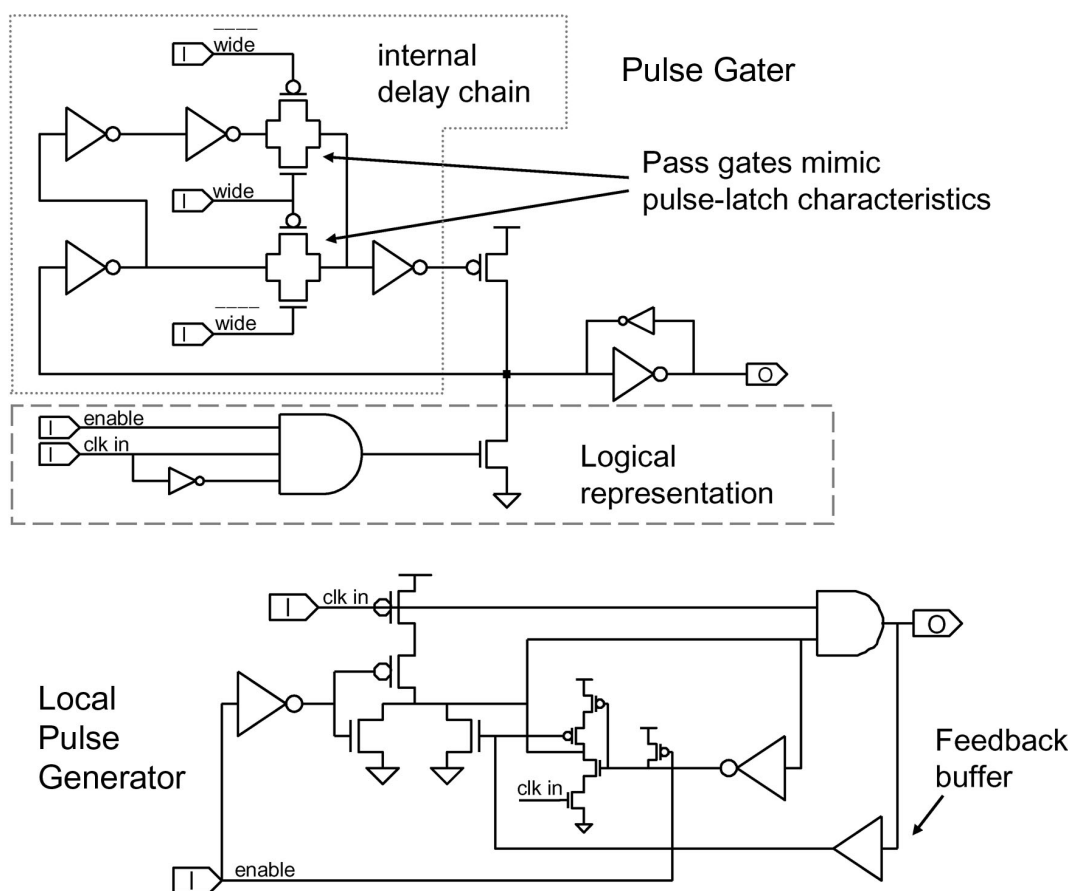


Fig. 7. Generating pulses.

Cache Array	Size	ECC/Parity	Phase/Cycle Access	Voltage Domain	Ramcell Sleep	Redun-dancy	CacheSafe Technology
L3	6MB/ Core	ECC	Cycle	Vcache	Yes	Yes (Block)	Yes
L3 Tag	250KB/ Core	ECC	Phase	Vcache	No	Yes (Entry)	No
L2 Instr	512KB/ Core	Parity	Cycle	Vcache	No	Yes (Block)	Yes
L2 Instr Tag	20KB/ Core	Parity	Phase	Vcache	No	No	No
L2 Data	256KB/Core	ECC	Phase	Vcache	No	Yes (Row)	Yes
L2 Data Tag & L1	Various	Various	Various	Vcore	No	No	No
Directory Cache	1.9MB/Die	Parity	Cycle	Vcache	No	Yes (Column)	Yes
Changes from Montvale in gray							

Fig. 8. Cache organization and feature summary.

signals are voltage converted in the datapath using a standard cross-coupled PFET topology [8]. The four cycle access time easily absorbs the additional delay of the voltage conversion, and the datapath area growth is minimal.

For the caches embedded in the CPU core, such as the L2 instruction and data arrays, the power supply separation is more complicated. Vcache supplies the memory cell and all circuits in contact with it, including the wordlines, bitlines, and column

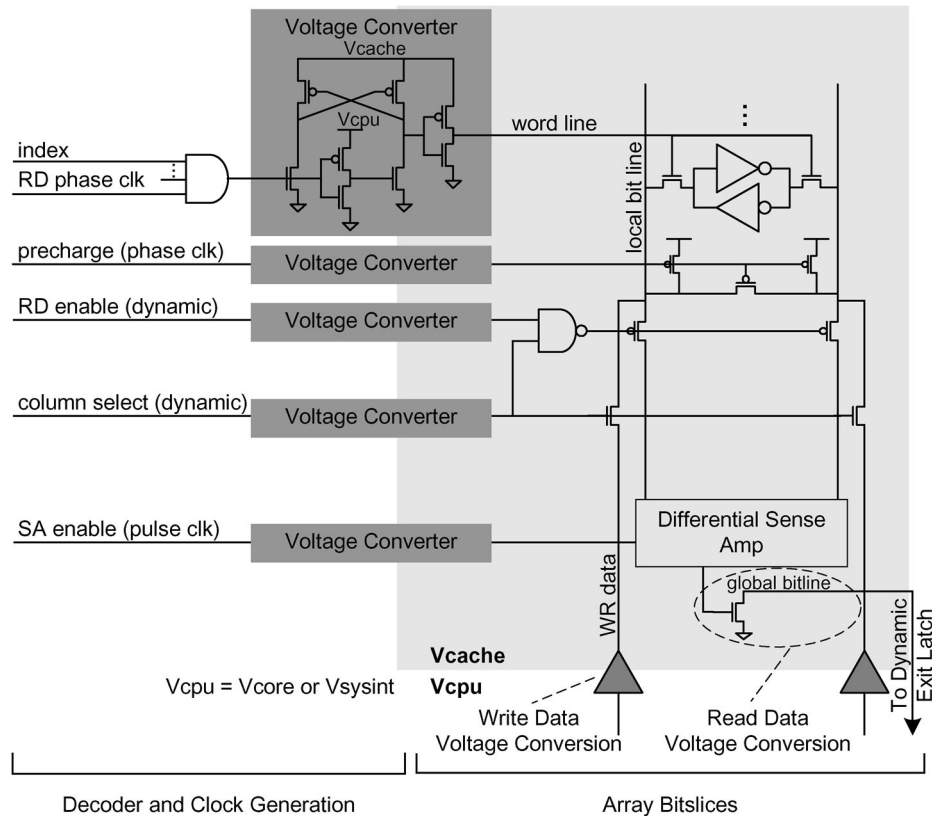


Fig. 9. Cache voltage conversion overview.

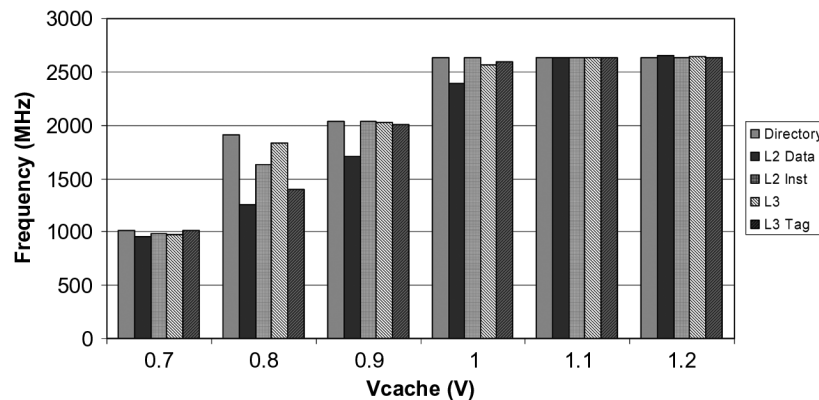


Fig. 10. Measured cache frequency versus cache power supply level.

I/O circuitry. The surrounding logic and datapath are in the core voltage domain. As shown in Fig. 9, voltage conversion between the two power supply domains is necessary for any signal entering or leaving the array. To avoid any clock skew penalties, voltage conversion is performed after clock and address/control/data are logically combined. Cross-coupled PFET voltage converters similar to the ones used in the L3 cache are embedded in the output stages of all decoder signals, including wordlines, column mux selects, and precharge signals. Read data coming from the sense-amp output uses a dynamic global bitline as an implicit voltage converter; it drives the gate of an NMOS pull-down device on a dynamic global bitline, whose precharger and keeper are tied to the core power. Write data is static and uses explicit voltage converters within the bitslices

that are based on a different topology that only requires the Vcache supply. While this topology has a potentially high current draw during a write operation, it was deemed a reasonable tradeoff since writes are relatively infrequent, and it eliminated the need for routing both power supplies to the tight bitslice I/O section.

Post-silicon tests on the caches have shown robust operation and frequency across the Vcache range. The L2, L3 and directory caches have been tested with functional content and programmable built in self test (BIST) algorithms. No functional, electrical, or voltage conversion issues have been found. Fig. 10 shows Vcache voltage versus frequency for typical BIST content averaged across 32 instances of each array at 25 °C. The data from these parts are representative of the entire popula-

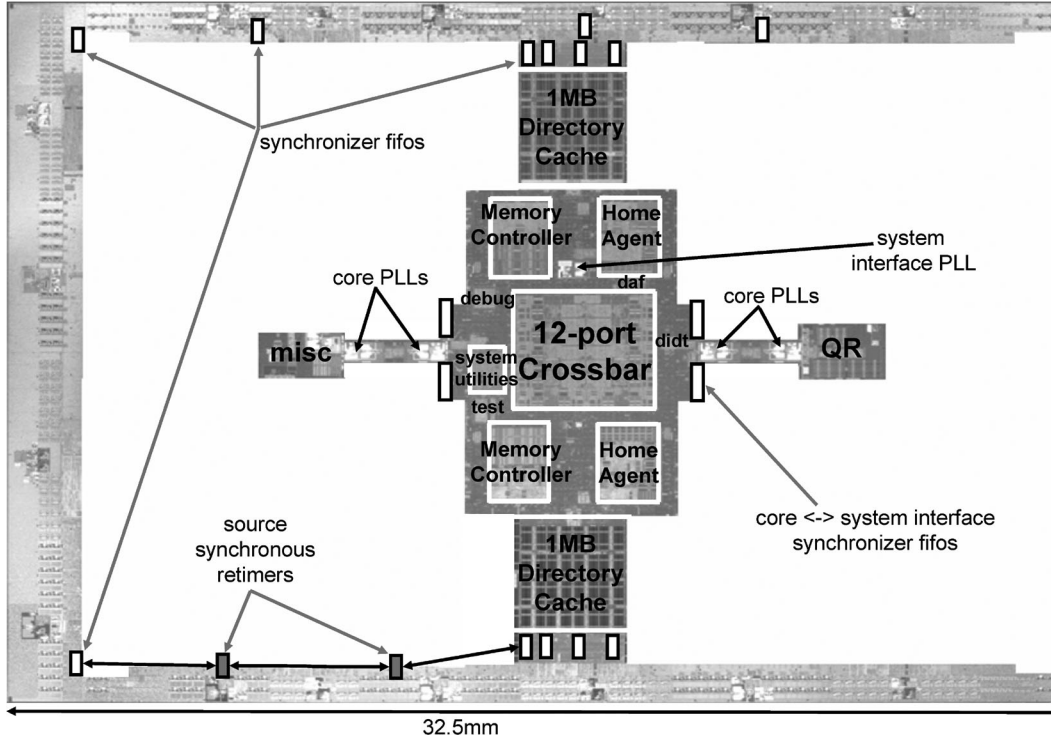


Fig. 11. System interface design and floorplan.

tion. The average array frequency at nominal voltage (1.1 V) is well beyond the design target and is not a limiter. The maximum frequency is limited by measurement and test circuitry restrictions and not the actual array read/write. Yield results for the caches have been good, and cache repair has successfully salvaged much of the fallout. The distribution of failures from stability at low voltage has been in line with expectations for the process generation.

IV. SYSTEM INTERFACE

The Tukwila system interface (Fig. 11) features dual integrated memory controllers that interface the microprocessor to four FBD2 memory channels. Each memory controller is paired with a home agent that implements the QuickPath Interconnect directory-based cache coherence protocol. Together the memory controller and the home agent implement a rich set of memory RAS features [9]. Directory coherency enables system performance to better scale to large numbers of sockets. Each home agent uses a 1 MB directory cache to reduce the latency of cache to cache transfers. In the center of the system interface is a 12-port crossbar router for concurrent full-duplex intra- and inter-processor communications. Each of the 2 home agent and 4 core crossbar ports have twice the bandwidth, 19.2 GB/s per direction, of the QuickPath Interconnect ports. A system utilities controller performs functions such as processor configuration access and implements SMBus master and off-chip Flash ROM interfaces. The processor's testability and initialization controllers are also contained within the system interface. The QR (charge rationing) controller, a system interface digital activity estimator, and the central di/dt controller also reside in the system interface.

A. Design Methodologies

The system interface RTL model and circuit design were developed concurrently. To effectively mitigate the effects of a high RTL change rate late in the design cycle and to reduce design time, fully static cell-based design and automation were utilized.

Control blocks with embedded register files are fully synthesized. A technique called hardcoding enabled designers to directly control the synthesis flow to fix critical timing and routing issues. Hardcoding was used to replace logic cones with custom designed cell-based circuits to solve critical timing paths and to create embedded datapath and repeater structures in synthesized control blocks. Larger datapaths were implemented with an automated datapath compiler which derives layout specifications from specially annotated schematics. A register file generator produced tape release quality layout for 36 unique register files in the system interface. The register file generator could not create completed implementations of a limited number of CAMs and special topologies; however the tool did provide baseline designs that were hand customized for each application.

B. Timing and Power

To achieve high system interface yield at 2.4 GHz the system interface pre-silicon timing target was significantly higher than 2.4 GHz. To achieve this cycle time requirement the design team enabled synthesis to automatically shift as many as 5 gate delays across latch boundaries. This was done by enabling synthesis to understand the transparency of the pulse latches, and providing latch variants with either fast clock to out delays or relaxed data setup times. Furthermore, engineer driven clock

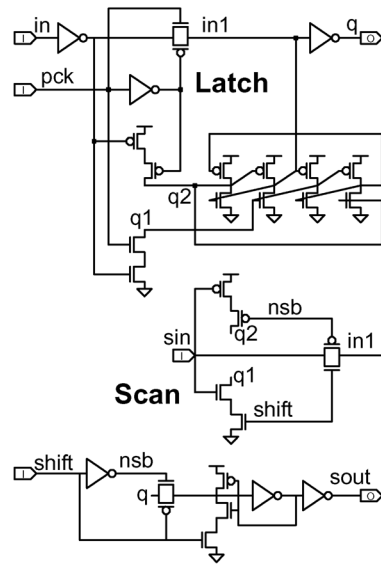
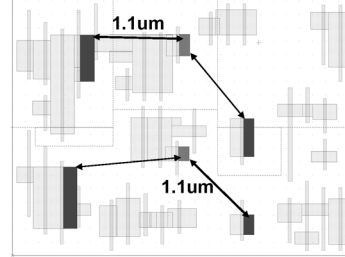


Fig. 12. SE-hardened pulse latch.

parameter	% of unprotected
Area	134%
pck load	136%
flowthru (in to q)	98%
pck → out	96%
Setup time	106%
SER (FIT)	100x better
Standby power	127%
Active power	125%



manipulations could shift as many as 5 additional gate delays across a clock boundary. These capabilities are in addition to the post silicon clock programmability that was developed for the 90 nm product [2].

Power dissipation is minimized with the usual techniques: clock gating at various levels of granularity, 85% low leakage FET usage, and FET downsizing on non-critical paths. More importantly, the timing margin described above enables the typical die to operate at reduced voltage, reducing power by more than 20%.

C. Soft Error Rate

Enterprise-class customers demand that per-socket SER (Soft Error Rate) not increase from one processor generation to the next. With triple the logic states per socket of the 90 nm processor, we maintain acceptable SER by extensive use of SE-hardened circuits.

Since a majority of the core is ported from 90 nm, we could not make a drastic reduction in per-core SER. This forced the design team to be very aggressive in addressing SER in the system interface. More than 99% of the latches in the system interface are SE-hardened, and 33% of the core latches have also been converted. Furthermore, the system interface protects 34 unique small register files, including all CAMs, with SE-hardened storage.

Interlocking feedback structures [10] are chosen for both types of storage. Fig. 12 compares the hardened latch to an unprotected latch. The SE-hardened pulse latch uses a write structure that achieves timing characteristics comparable to an unprotected pulse latch. The write mechanism presented here pulls the feedback out of the way, resulting in excellent clock-to-output delays, which are 25 to 30% faster than writing two same sense nodes.

For the SE-hardened register file cell, the redundant write construct that is required makes the cell harder to write, but write timing is generally not critical. The load on the read bitlines and

wordlines increase by 3% and 10%, respectively, making timing only slightly worse.

The SER benefits of a redundant storage structure depend on layout that physically separates the storage node diffusions. We use a layout that keeps the latch nodes that are sensitive to multi-bit strikes a minimum of 1.1 μm apart. This results in a projected 100 $\times$ reduction in SER over the unprotected latch (Fig. 5). The more compact register file cell is projected to achieve an 80 $\times$ improvement. This spacing dependent strategy is expected to work for at least 2 more process generations.

The cost of the SE-hardened latch is a 34% area increase and 25% higher power consumption. Including the entire overhead, the area cost for hardening a 32 b $\times$ 32 entry, 1-read, 1-write register file is 25%. In comparison, ECC costs the same 25% area for small amounts of storage, and presents difficult pipelining and timing problems that are avoided with this SE-hardened design.

D. Cross Clock-Domain Synchronization

The system interface has a central communication role, containing most of the processor's voltage and frequency domain crossings. There are 12 independent frequency domains (2 core pairs, 1 system interface, 1 QR controller, 2 FBD2 domains, and 6 Intel QuickPath Interconnect domains.) To ensure deterministic data transfers between these domains, synchronizers are employed, consisting of FIFO's with deterministic read and write patterns based on the clock ratio between the two domains. Depending on the local clock distribution near the synchronizer, it will be one of two types.

For core to system interface communication, the two clock domains are physically adjacent. This allows a traditional FIFO structure, with writes happening on a local send clock, and reads happening on a local receive clock. Timing and verification for this type of FIFO is handled easily by the standard tools. Because the cores and the system interface are in separate voltage domains, a voltage conversion is necessary. Each entry

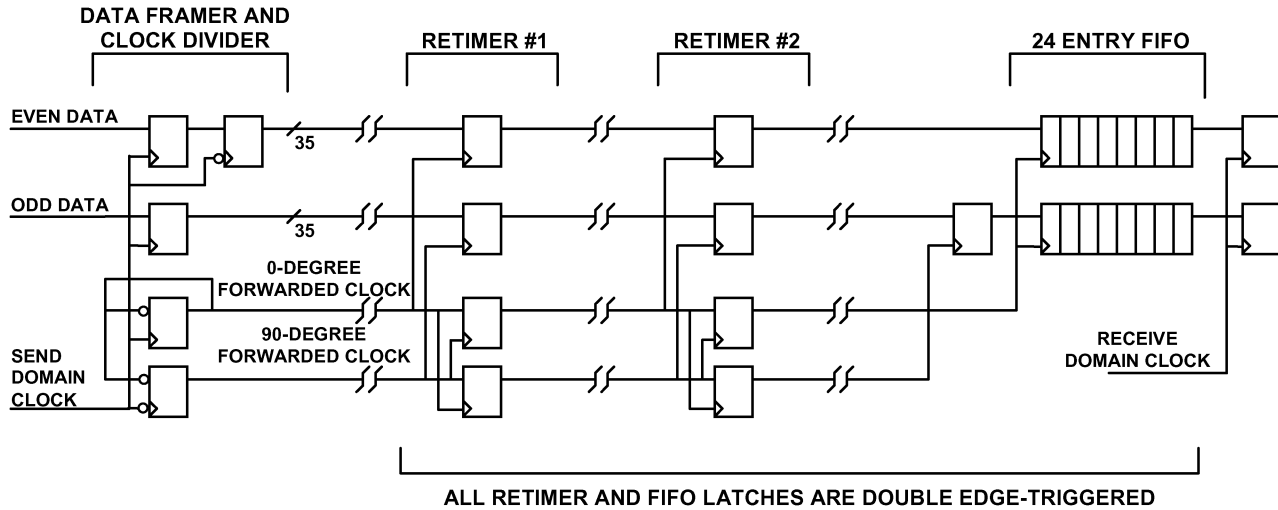


Fig. 13. Single synchronizer link, showing data framer, clock divider, retiming stations, and FIFO.

of the FIFO is independently level-shifted, hiding the delay in the FIFO latency rather than in a timing path.

For the other domain transitions, such as from the system interface to the FBD2 and Intel QuickPath Interconnect links, the clock distributions are separated by as much as 15 mm. In these cases, it is impractical to extend the clock distribution to a point where both meet, so a source-synchronous, forwarded clock scheme is employed. Two half-rate, quadrature clocks are sent along with up to 70 bits of data, such that the clock delay is matched to the data delay. At the destination, both edges of the forwarded clocks are used to write data into the FIFO, and the local clock reads the data after a programmable latency.

For ideal matching, all data and clock wires would be fully shielded, but this uses too many wires. Instead, wires are half-shielded (25% fewer tracks than full-shielding), coupled on one side to a shield, and on the other side to a data wire transitioning on the opposite clock phase. Because the noise event is separated from the transition by a clock phase, the victim has enough time to recover and minimal skew is introduced.

At two points along the route, data bits are retimed by the forwarded clocks, to realign data and clocks, and to reduce the chance of a skew induced failure. The clocks also retime each other, producing new forwarded clocks. A retimer is simply a station where all data and clock bits are latched to remove accumulated skew and line them up to a common clock again (Fig. 13).

Silicon results indicate that the synchronizer is robust, with margin beyond the system interface operating point, allowing for normal circuit variations without impacting chip yield or performance.

V. PBOX

This processor marks the transition of the Itanium architecture from a front-side bus based I/O design to a high-speed serial links based design. This processor includes 400+ QuickPath interconnect pins organized into 4 full-width ports and 2 half-width ports, each of which can be used for processor-to-processor communication or processor-to-I/O communication. In addition to the QuickPath interconnect pins, there are 184+ pins

devoted to FBD2 memory, organized into 4 channels. In each case, each differential pair is capable of achieving 4.8 Gb/s.

The chip implements a common transmitter design which maintains a balance between power efficiency and high-speed functionality (Fig. 14). Each transmitter lane (i.e. the logic and analog circuits associated with one differential pair output) receives 2 bits of data per 2.4 GHz clock. These bits then enter a 4-tap, 6-bit FIR filter which performs digital de-emphasis. The implementation of this power-efficient filter uses a set of multiplexers for each pair of taps to select between four pre-computed static coefficient sums and then uses a single adder to combine the results from each pair of taps. This is significantly more power efficient than a more traditional FIR implementation which might use 3 adders in the implementation of a 4 tap FIR. The output of this de-emphasis filter then enters a FIFO which is used both to cross from the general-purpose digital clock domain to a low-skew, shallow-distribution transmit clock domain and to cross from V_{sysint} to V_{io} . The output of this FIFO traverses some latches, enters a serializing mux which selects between the two phases of data, and ultimately feeds the analog driver which is basically a current-steering binary DAC. Each of the binary weighted current sources is implemented with two cascode-connected PMOS devices, and this current is switched between the two differential outputs via two PMOS digital switches.

The chip also implements a common receiver design which can operate both in a forwarded clock mode and in a CDR mode (Fig. 15). The receiver's clock generation path begins with a multiplexer which selects between the transmit clock (for CDR and test applications) and an external clock received through an AC-coupled clock receiver. The output of this multiplexer then enters a duty cycle corrector (DCC) and a DLL which generates 8 evenly spaced clocks. The input clock path through the DLL is powered by an on-die voltage regulator which takes in 1.8 V and generates a clean 1.1 V. These full-swing clocks are then distributed (with repeaters) to each of the receiver lanes. Each lane contains three phase-interpolators (PIs), each of which further subdivides the DLL clocks to generate a clock with a 1/64 cycle granularity. These three clocks are used to clock three latching

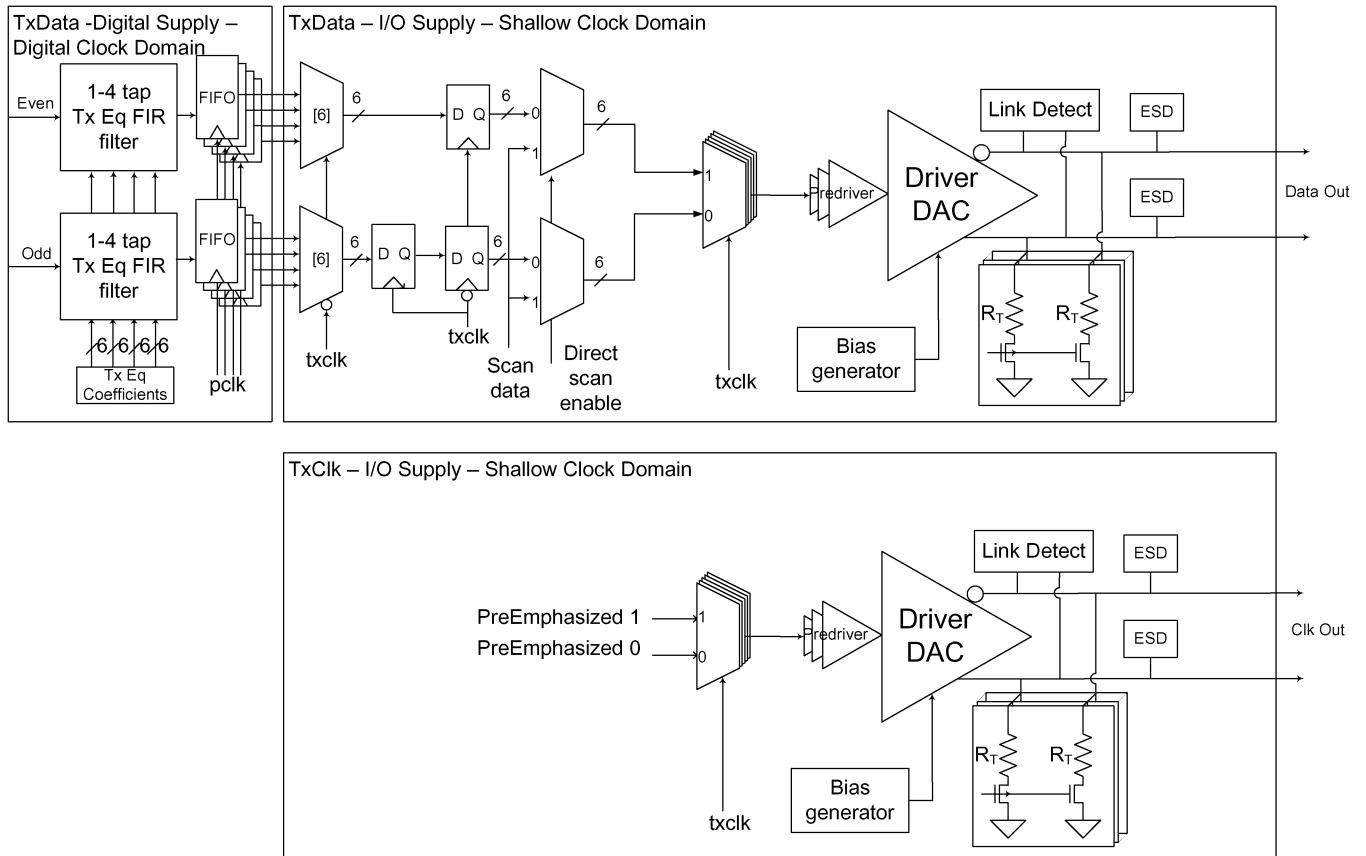


Fig. 14. P-box transmitter diagram.

differential amplifiers which sample the differential data input pins. Each of these amplifiers is offset compensated by mixing in a DC differential error term generated from a 7-bit DAC with ~ 1.5 mV step size. The correct offset for each DAC is trained at reset time. Two of these amplifiers correspond to data samples taken in the center of each phase of the data eye, while the third sample is taken between the two eyes and is used in CDR mode. This chip uses private PIs for each of the two data samples (rather than using both edges of one PI's clock) to better center the samples in the case of duty cycle error from both the far-end transmitter as well as the on die Rx clock generation path. The two data samples are then deterministically retimed to a common clock and fed into a 32-sample-deep drift FIFO.

The I/O blocks also contain support circuits which provide a variety of functions. An on-die bandgap reference (tunable in the manufacturing flow) is used to generate bias currents which are distributed to each lane and other assorted analog circuits. These bias currents are generated with high-impedance cascode PMOS-mirrors off of the 1.8 V supply (for headroom concerns). To ameliorate noise concerns, care was taken in the layout to both laterally shield these ~ 100 μ A bias currents with the appropriate supply and to ensure only quiet signals orthogonally crossed these signals on adjacent metal layers (e.g. power supplies, static configuration signals, etc.). Pad termination resistance compensation (R_{comp}) is achieved by using an off-die precision resistor to periodically tune a replica variable termination resistor. The digital codes associated with this tuning are then periodically broadcast to the termination resistors for the nearest

40–52 pads. Transmitter current compensation is achieved by constantly modulating a DAC-controlled bias current driving a replica transmitter and comparing the output voltage of that replica transmitter to one generated by the bandgap reference. The DAC-controlled bias current is then mirrored and sent out to every Tx lane in the channel/port. This mirror contains some small amount of per-lane current adjustments that can be programmed at manufacturing test time so as to increase the current to “weak” lanes and hence improve yield.

The chip's post silicon I/O performance has largely matched pre-silicon expectations. An eye with a height of 571 mV and width of 126 ps has been demonstrated on a 6.4 Gb/s link with a 0'' channel. At 4.8 Gb/s on a 38'', 2 connector channel, the eye height is 182 mV and width is 165 ps. Empirically it has been determined that utilizing only 2 of the 4 taps in the de-emphasis filter (i.e. setting the coefficients for the 3rd and 4th tap to 0) produces the optimal eye height. Additionally, the offset compensation for the latching amplifiers in the receiver has proven to be essential for maximizing performance. Trained offset compensations as high as ± 80 mV have been observed.

VI. CLOCKS

Fig. 16 shows a schematic of the connectivity of the clock tree. The clock tree generates all internally used clocks from a 133 MHz system clock reference. The maiden PLL is intended to filter worst case jitter from the external clock source and generate a 133 MHz clock, called the L0 clock. The L0 clock is distributed so that its arrival times are matched at the core and

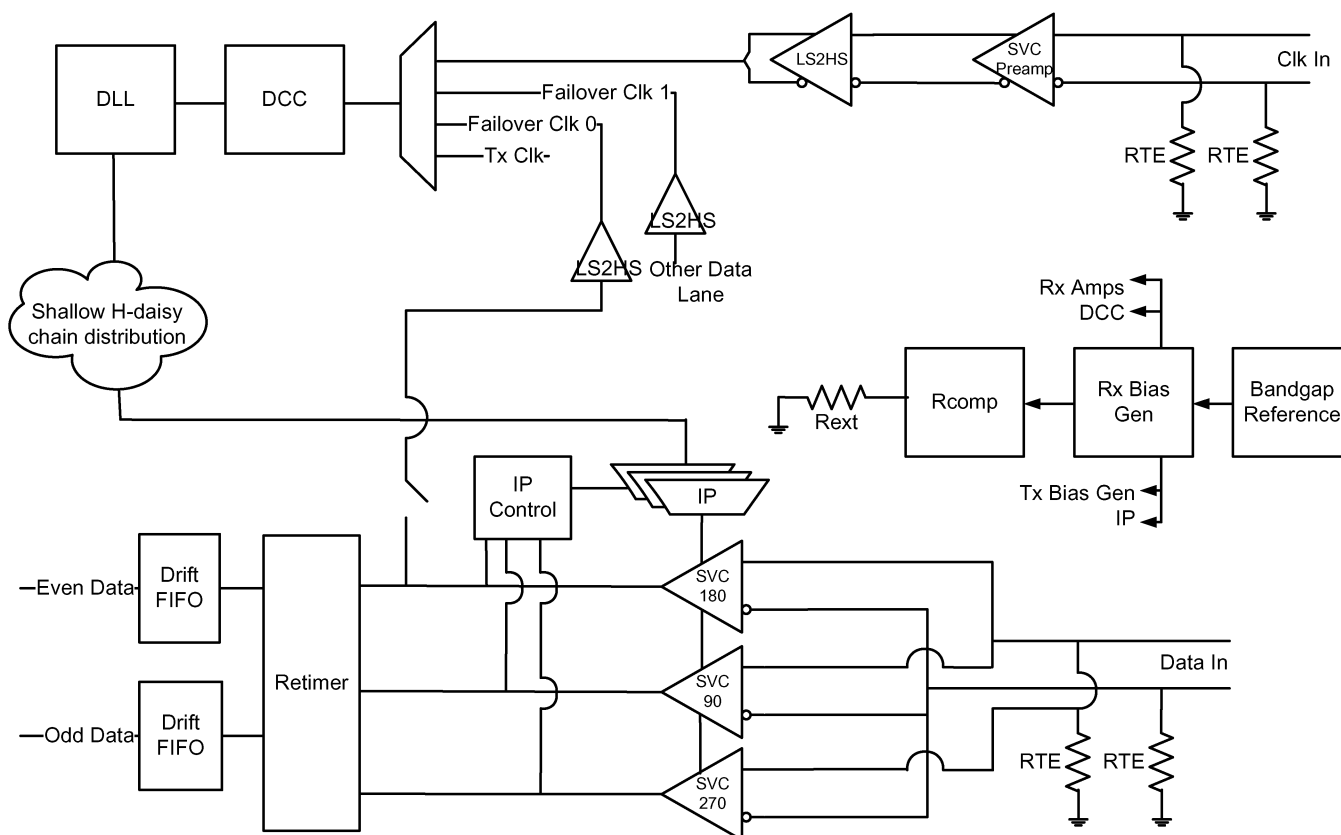


Fig. 15. P-box receiver diagram.

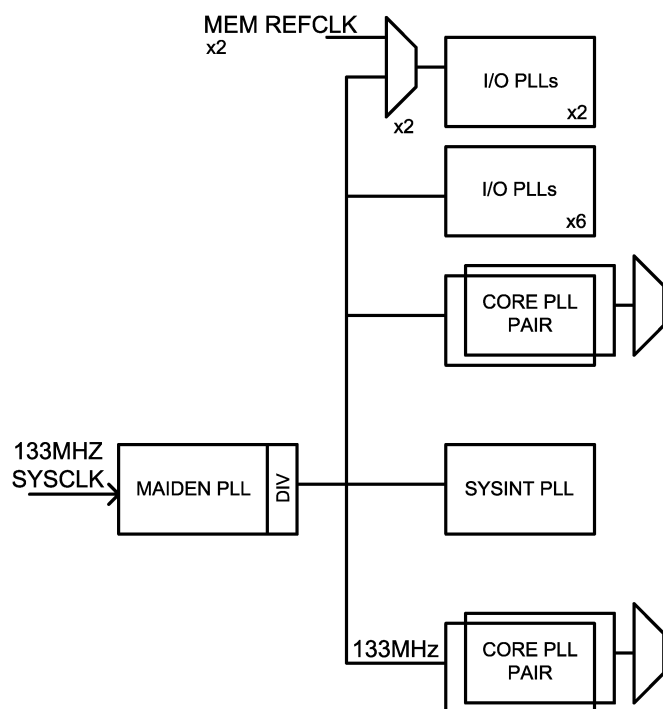


Fig. 16. Clock PLL diagram.

system interface PLLs; due to floorplanning and power domain issues, the L0 clock is serially routed to the I/O PLLs. The rising

edge of the L0 clock at the core and system interface PLLs is the common “heartbeat” used for deterministic data transfer between clock domains. Each of the core, I/O, and system interface PLLs are followed by a global clock distribution network for that specific region of the chip. This global clock distribution includes a pseudo-differential route, a final global clock buffer (SLCB), and local gates for driving latches. Fig. 17 shows the physical locations of the PLL blocks on the die and the locations of the various clock domains.

Each PLL macro block contains two differently tuned, fuse-selectable self-biased analog PLL cores, a bandgap reference, a voltage regulator, and a duty cycle corrector. The bandgap reference is used in concert with the voltage regulator to provide a regulated supply voltage to the PLL. For the core, system interface, and maiden PLLs the analog PLL cores differ in how their F-V curves are tuned depending on frequency synthesis requirements. For the I/O PLL, one of the two analog PLL cores has the capability to optimize capacitive loading to ensure best signal swing, and lowest jitter, for a given desired PLL output frequency. All PLL macros have manufacturing hooks to enable F-V curve data collection and to enable optimized gear-shifting to keep the VCOs in the sweet spot of their operating range. There are two different types of duty cycle correctors used in the PLL macros. For the I/O PLL a continuously operating analog DCC is reused from the I/O receiver design. For the core and system interface PLLs, a digitally controlled DCC is used to measure and optimize duty cycle trim in the manufacturing process.

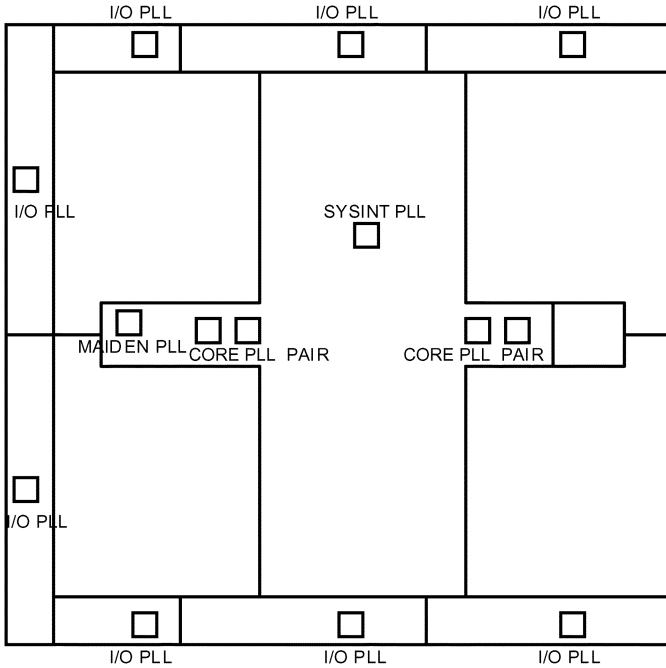


Fig. 17. PLL locations.

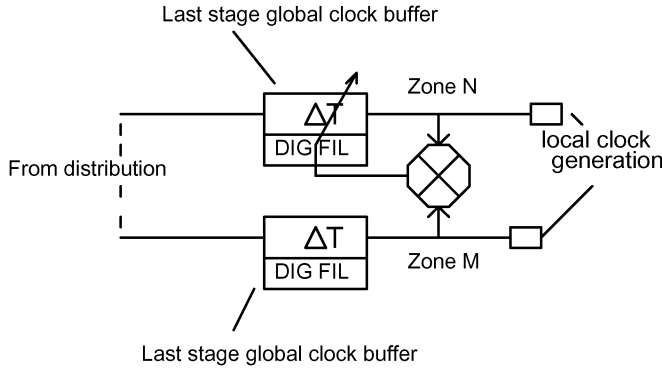


Fig. 18. RAD diagram.

Within each clock domain, the PLL route is distributed in a pseudo-differential manner with route branch balancing. The pseudo-differential route is converted to a single-ended signal in the final global clock buffer, called the SLCB. The SLCBs also implement the control and delay lines for an active deskew system which has been leveraged from previous Itanium designs [11]. This active deskew system is intended to compensate for random and systematic sources of skew within a given PLL's clock zone. The schematic for the deskew system is shown in Fig. 18.

Deterministic data transfers between clock domains require a set of FIFOs between the domains. Fig. 3 shows the locations of various clock domains on the die. Note that some I/O and system interface clock domains must communicate deterministically even though they are widely separated. Such FIFOs exist for core $\leftrightarrow$ system interface transfers and system interface $\leftrightarrow$ I/O transfers. These domains use the heartbeat timing reference as a common timestamp to control read and write FIFO operation. As mentioned previously, this heartbeat timing corresponds to L0 clock arrival at the core and system interface PLLs.

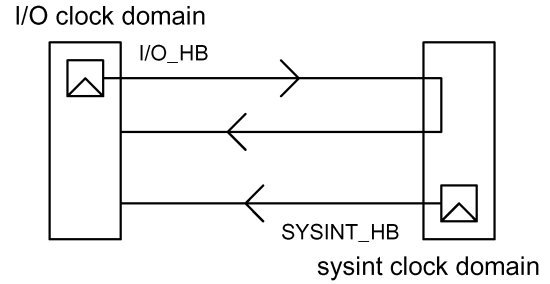


Fig. 19. P-box heartbeat.

Through the standard technique of matching the clock delay to the latches in the PLL's feedback in the core and sysint PLLs, we identify the core and system interface clock edges which are aligned to the heartbeat timing. Generating I/O clocks which are aligned to heartbeat cannot leverage this same standard use of replica feedback in the I/O PLLs for two reasons. First, the L0 clock arrival time at the I/O PLLs do not match the arrival at the core and sysint PLLs. Second, because of a desire to minimize I/O jitter, a minimal delay "tight" feedback is used for the I/O PLLs. To address these issues with the generation of heartbeats in the I/O clock domain, we implemented a digital DLL which is split between the two clock domains of interest. See Fig. 19 for a description of this DLL, which tries to optimally place candidate I/O clock heartbeats to align with the transmitted sysint heartbeat under the assumption that the routes in the figure are well matched. Note that the separation between the I/O clock and system interface clock domains is more than 1 cm in the worst case. Due to this indirect heartbeat generation method, the system interface $\leftrightarrow$ I/O FIFOs must tolerate extra skew and latency. (Please note that these system interface $\leftrightarrow$ I/O/FIFOs also include the source synchronous transfer scheme which is explained elsewhere.)

The core clocks are generated by muxing between one of two PLLs. This muxing is designed to operate without limiting the full core frequency operating range. Furthermore, these frequency switches must occur in a way which allows deterministic operation of the FIFOs which transfer data between the core and sysint clock domains. To enable frequency switches to meet these requirements, frequency switches occur at a heartbeat boundary only because the two PLLs work to align their output clock heartbeat edges; to allow for drift between the two PLLs heartbeats, the core clocks will drop one clock cycle during the frequency transition. The FIFO which controls core $\leftrightarrow$ sysint communication is told in advance that an edge will be dropped during a frequency switch. This allows the FIFO to maintain deterministic operation across switches. Please see waveforms in Fig. 20.

VII. VFM

This processor includes a new design for power management that is based upon digital activity sensing and discrete voltage-frequency pairs. The digital activity sensing is accomplished by monitoring approximately 120 architectural events in each of the four cores. Each event is weighted to represent the relative amount of capacitance switched by that event. A central control unit accumulates these and makes a decision every $\sim 8 \mu\text{s}$ what

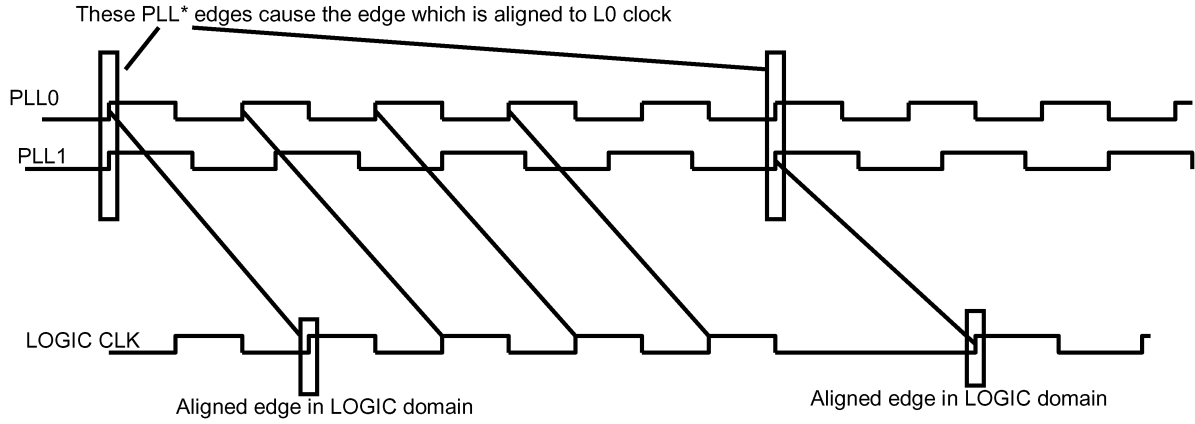


Fig. 20. Dual PLL switch.

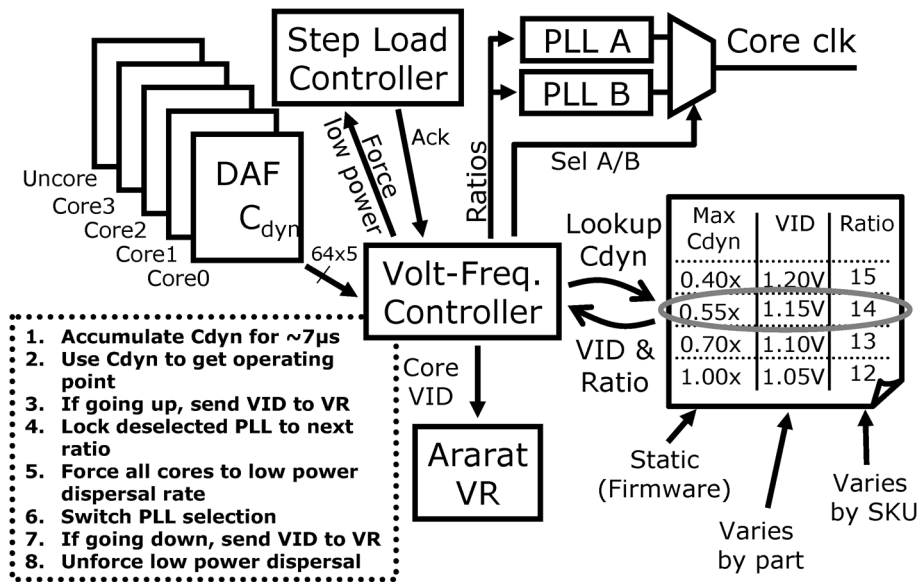


Fig. 21. Adaptive voltage and frequency overview.

core voltage-frequency pair will give the best frequency without exceeding the power envelope for that level of activity.

The maximum sustainable capacitance thresholds are stored along with voltage-frequency pairs in a look-up table fused at production. As capacitance varies little from part-to-part, weights and switching thresholds are kept identical among all parts. This assures that the frequency decisions of every part are identical, eliminating the part-to-part and environmental variability [2].

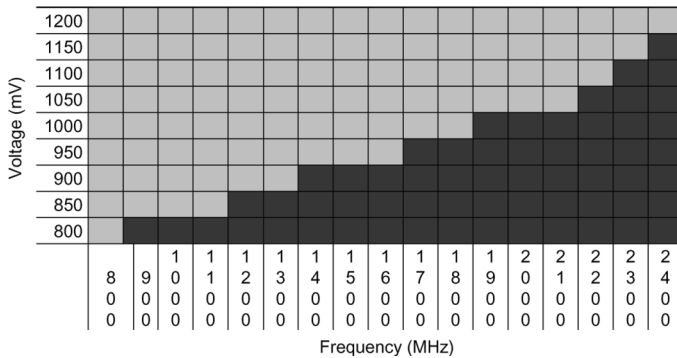
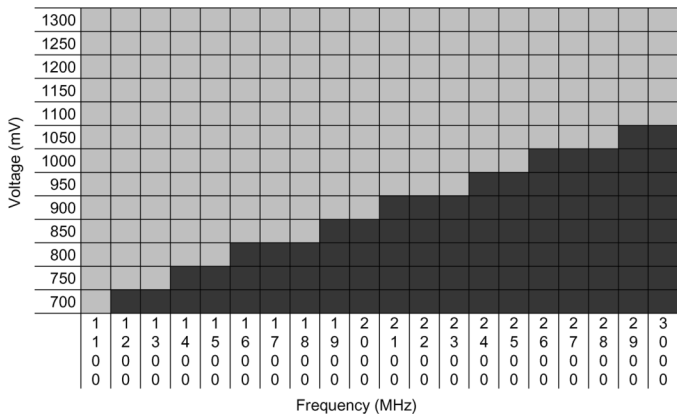
A block diagram of the adaptive voltage and frequency subsystem is shown in Fig. 21. The Digital Activity Factor (DAF) accumulators in the cores and uncore sum the dynamic capacitance (Cdyn) switched over a time period. The Voltage-Frequency Controller periodically tallies these accumulated sums and, with the aid of a lookup table, determines the appropriate voltage ID (VID) and clock ratio. The Voltage-Frequency Controller drives the voltage and frequency changes to the Ararat voltage regulator (VR) and the core PLLs. Frequency changes are accomplished by two parallel PLLs that are mux-selected. One drives the core clock while the other locks to the next requested ratio. When the ratio change is executed, the core clock

is halted momentarily, the PLL mux select is switched, and the core clock is restarted at the new ratio.

This processor also includes a tunable di/dt controller [12], [13]. This controller prevents sudden changes in current by limiting how quickly the issue rate of certain high-power instructions can increase and by injecting artificial activity to smooth out sudden drops in issue rate of these instructions. This controller has been simulated to reduce first droop in the cores by 30%–50%. The adaptive voltage and frequency changes utilize the di/dt controller to restrict the issue rate of instructions momentarily before and after the frequency change in order to reduce the di/dt associated with stopping and restarting the core clock.

VIII. RESULTS

The circuit design methodology described herein has delivered good silicon results in the 65 nm process technology. Fig. 22 shows the core frequency versus voltage shmoo on a typical part at a junction temperature of 80 °C. The processor exhibits a wide range of operation from 800 mV to 1.2 V. Fig. 23 shows the system interface frequency versus voltage

Fig. 22. Core shmoo at $T_j = 80^\circ\text{C}$.Fig. 23. System interface shmoo at $T_j = 80^\circ\text{C}$.

shmoo on the same part. The design demonstrates an operating range from 700 mV to 1.20 V with good operating margin above the system interface frequency of 2.4 GHz.

ACKNOWLEDGMENT

The authors acknowledge the effort and contributions of the entire design team from Fort Collins, Colorado and Hudson, Massachusetts.

REFERENCES

- [1] S. Naffziger *et al.*, "The implementation of the Itanium 2 microprocessor," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1448–1460, Nov. 2002.
- [2] S. Naffziger *et al.*, "The implementation of a 2-core, multi-threaded Itanium family microprocessor," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 197–209, Jan. 2006.
- [3] Intel Corp., "Mainframe reliability on industry-standard servers: Intel Itanium-based servers are changing the economics of mission-critical computing," white paper, 2007 [Online]. Available: http://download.intel.com/products/processor/itanium/RAS_WPaper_Final_1207.pdf, white paper
- [4] Intel Corp., "Intel QuickPath Architecture: A new system architecture for unleashing the performance of future generations of Intel multi-core microprocessors," white paper, 2008 [Online]. Available: <http://www.intel.com/technology/quickpath/whitepaper.pdf>
- [5] K. Zhang *et al.*, "A 3-GHz 70-Mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2005.
- [6] D. Weiss *et al.*, "The on-chip 3 MB subarray based 3rd level cache on an Itanium microprocessor," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2002.

- [7] J. Wu *et al.*, "The asynchronous 24 MB on-chip level-3 cache for a dual-core Itanium-family processor," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2005.
- [8] R. Krishnamurthy, *et al.*, "Hybrid CVSL pass-gate level converting sequential circuit for multi-VCC microprocessor," US Patent 7,132,856.
- [9] E. Delano, "Tukwila: A quad-core Intel Itanium processor," in *Proc. Hot Chips 20*, Aug. 2008.
- [10] P. Hazucha *et al.*, "Measurements and analysis of SER-tolerant latch in a 90-nm dual-Vt CMOS process," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1536–1543, Sep. 2004.
- [11] T. Fischer *et al.*, "A 90 nm variable-frequency clock system for a power-managed Itanium-family processor," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2005, Paper 16.2.
- [12] M. D. Powell and T. N. Vijaykumar, "Pipeline muffling and a priori current ramping: Architectural techniques to reduce high-frequency inductive noise," in *Proc. Int. Symp. Low Power Electronics and Design (ISLPED)*, Aug. 2003.
- [13] M. D. Powell and T. N. Vijaykumar, "Pipeline damping: A micro-architectural technique to reduce inductive noise in supply voltage," in *Proc. 30th Int. Symp. Computer Architecture (ISCA 30)*, Jun. 2003.



Blaine Stackhouse received the B.S.E.E. degree from Virginia Tech, Blacksburg, VA, in 1987 and the M.S.E.E. degree from North Carolina State University, Raleigh, NC, in 1989.

He joined Digital Equipment Corporation and designed CPU peripheral chips including the Giga Port Interface router chip and the first PCI-to-PCI bridge chip. He moved to Hewlett Packard in 1994 and designed caches and IO on the PCXL-2 processor. In 1996, he transitioned to the Intel McKinley processor development team and worked as the DCU unit lead

for both the pre-silicon design and post-silicon debug phases of the project. Next he worked on the Montecito microprocessor as the L1D unit lead and the project memory array lead. In 2004, he joined Intel Corporation as a Principal Engineer leading the circuit design for Tukwila microprocessor. He is currently responsible for the electrical debug effort on Tukwila. He holds five US patents with several pending and has been an author on 15 internal and external conference papers.



Sal Bhimji received the B.S.E.E. degree from Cornell University, Ithaca, NY, in 2002.

He is an engineer in the Enterprise Microprocessor Group at Intel Corporation and is currently the technical leader for cache physical design and debug on Tukwila. He joined Intel in 2002 where he has worked on Itanium and Xeon microprocessors.



Chris Bostak received the B.S. and M.S. degrees in electrical engineering from the University of Virginia, Charlottesville, VA, in 1989, and from Stanford University, Stanford, CA, in 1993, respectively.

His professional work started at HP in 1989, as a manufacturing test development engineer for microwave communications test equipment. Subsequently, he designed components of digital video file servers at HP. Since 1997, he has designed circuits for Itanium processors including cache control and

power management circuits first at HP and then at Intel starting in 2005. On Tukwila, he led the analog IO design team.



methodology.

Dave Bradley received the B.S.E.E. degree in 1996 from Virginia Tech, Blacksburg, VA, and has since been a VLSI designer for HP and now Intel. He has served as a lead physical designer on L1 and L2 caches on previous Itanium processors. He has also served as a pre-silicon timing convergence lead and power convergence lead, and he has led efforts in post-silicon clock system debug. His current role is the physical design lead for the Tukwila I/O, and his areas of expertise include high speed circuit design, power reduction techniques, and electrical



U.S. patents and has authored or coauthored more than 20 papers on microprocessor design and high-speed circuits.

Paul Gronowski received the B.S.E.E. degree from the University of Cincinnati, Cincinnati, OH, in 1984.

He is a Principal Engineer in the Enterprise Microprocessor Group at Intel Corporation, and is currently leading the full-chip integration on a next generation Itanium processor. He joined Intel in 2001 and has worked on both Itanium and Xeon processors. Prior to that, he worked for Digital Equipment Corporation and Compaq Computer Corporation in Hudson, MA, where he contributed to six new and two derivative Alpha and VAX processor designs. He holds two



Throughout his career, he has served in various circuit design and debug capacities on mobile, desktop, Xeon, and Itanium processor families. He is currently a Senior Staff Engineer leading post-silicon power validation, VFM, and frequency projection for Tukwila.

Brian Cherkauer received the B.S.E.E. degree from the State University of New York, Buffalo, in 1990, and the M.S.E.E. degree and the Ph.D. degree in electrical engineering from the University of Rochester, Rochester, NY, in 1991 and 1995, respectively. His doctoral work concerned the field of high-speed, low power integrated circuit design.

He joined Intel Corporation, Santa Clara, CA, in 1995 in the Mobile and Handheld Products Group where he worked on design, power projections, and power-saving features for laptop processors.



ology, and SER immunity.

Dan Krueger received the B.S.E.E. degree in 1989 and the M.S.E.E. degree in 1991 from Washington University in Saint Louis.

He has been a VLSI designer for IBM and HP, and is now with Intel. He has been the physical design lead for the integer datapath and clocking systems on multiple PA-RISC CPUs. He also led production test development and yield analysis for these same CPUs. His current role is the physical design lead for the Tukwila uncore. His areas of expertise include high speed circuit design, electrical method-



Jayen Desai received the B.S.E.E. and M.S.E.E. degrees from Stanford University, Stanford, CA, in 1994.

Since then, he has worked at HP and Intel on PA-RISC and Itanium processors in the areas of timing convergence, clock generation, synchronizer design, analog design, and I/O design. He is a principal engineer at Intel's Fort Collins Design Center, Fort Collins, CO.



Charles Morganti received the B.S. degree in engineering in 1993 from Harvey Mudd College and the M.S.E.E. degree in 1994 from Colorado State University.

He joined Intel in 1995 and has worked in the area of cache design for multiple generations of Intel Itanium processors. He was one of the lead cache designers on the Tukwila project and is currently developing the cache design methodology for a next generation Itanium processor.



Erin Francom received the B.S.E.E. degree from Brigham Young University, Provo, UT, in 1996.

He joined Hewlett-Packard's microprocessor design lab in Fort Collins, CO, in 1996 and worked on PA-RISC and Itanium processors. This work included high-speed custom circuit and clocking designs and post-Si validation. In 2005 he joined Intel Corporation in Fort Collins. He is currently a Technical Lead in the Enterprise Microprocessor Group working on the next generation Itanium microprocessor. He holds two U.S. patents.



Steve Troyer received the B.S.E.E. degree from Case Western Reserve University, Cleveland, OH, and the B.A. degree in physics from Goshen College, Goshen, IN, in 2000.

In 2005, he joined the Intel Corporation. Currently, he is the physical design lead on the Tukwila synchronizer. Before joining Intel, he worked for Agilent Technologies, where he contributed to several Serdes transmitter and receiver designs.



Mike Gowan received the B.S.E.E. degree from the Purdue University, West Lafayette, IN, and the M.S.E.E. degree from North Carolina State University, Raleigh, NC.

He is a Senior Engineering Manager in the Enterprise Microprocessor Group at Intel Corporation. He joined Intel in 2001 where he has worked on several generations of Itanium Processors. Previously, he worked for Digital Equipment Corporation and Compaq Computer Corporation in Hudson, MA, where he contributed to Alpha microprocessor and

communications chip designs. He holds two patents and has coauthored more than 12 papers on microprocessors and communications designs.

Journals & Magazines > IEEE Journal of Solid-State C... > Volume: 44 Issue: 1

A 65 nm 2-Billion Transistor Quad-Core Itanium Processor

Publisher: IEEE

Cite This

PDF

Blaine Stackhouse ; Sal Bhimji ; Chris Bostak ; Dave Bradley ; Brian Cherkauer ; Jaye... All Authors

65
Paper
Citations

66
Patent
Citations

1081
Full
Text Views

Alerts

Manage Content Alerts

Add to Citation Alerts

Abstract

Document Sections

I. Overview

II. Core

III. Caches

IV. System Interface

V. Pbox

Show Full Outline

Authors

Figures

References

Citations

Keywords

Download PDF

Abstract:This paper describes an Itanium processor implemented in 65 nm process with 8 layers of Cu interconnect. The 21.5 mm by 32.5 mm die has 2.05B transistors. The processor h... **View more**

Metadata

Abstract: This paper describes an Itanium processor implemented in 65 nm process with 8 layers of Cu interconnect. The 21.5 mm by 32.5 mm die has 2.05B transistors. The processor has four dual-threaded cores, 30 MB of cache, and a system interface that operates at 2.4 GHz at 105degC . High speed serial interconnects allow for peak processor-to-processor bandwidth of 96 GB/s and peak memory bandwidth of 34 GB/s.

Published in: IEEE Journal of Solid-State Circuits (Volume: 44 , Issue: 1, January 2009)

Page(s): 18 - 31

Date of Publication: 30 December 2008

INSPEC Accession Number: 10370957

DOI: 10.1109/JSSC.2008.2007150

More Like This

A new multiprocessor interconnection network with wide sense nonblocking capabilities

Proceedings of the 32nd Midwest Symposium on Circuits and Systems, Published: 1989

Selective Filling and Sintering of Copper Nanoclusters for Interconnect

IEEE Transactions on Nanotechnology Published: 2007

Show More

☰ Contents

I. Overview

The next generation in the Intel Itanium processor family code named Tukwila is described. The 21.5 mm by 32.5mm die contains 2.05 billion transistors, making it the first two billion transistor microprocessor ever reported. Tukwila combines four ported Itanium cores with a new system interface and high speed serial interconnects to deliver greater than 2X performance relative to the Montecito and Montvale family of processors [1], [2]. Tukwila is manufactured in a 65 nm process with 8 layers of copper interconnect as shown in the die photo in Fig. 1. The Tukwila die is enclosed in a 66 mm × 66 mm FR4 laminate package with 1248 total landed pins as shown in Fig. 2. Fig. 1.

Die photo.

Fig. 2.
Package photo.

Authors	▼
Figures	▼
References	▼
Citations	▼
Keywords	▼
Metrics	▼

IEEE Personal Account

CHANGE
USERNAME/PASSWORD

Purchase Details

PAYMENT OPTIONS
VIEW PURCHASED
DOCUMENTS

Profile Information

COMMUNICATIONS
PREFERENCES
PROFESSION AND
EDUCATION
TECHNICAL INTERESTS

Need Help?

US & CANADA: +1 800 678
4333
WORLDWIDE: +1 732 981
0060
CONTACT & SUPPORT

Follow



IEEE Account	Purchase Details	Profile Information	Need Help?
» Change Username/Password	» Payment Options	» Communications Preferences	» US & Canada: +1 800 678 4333
» Update Address	» Order History	» Profession and Education	» Worldwide: +1 732 981 0060
	» View Purchased Documents	» Technical Interests	» Contact & Support

[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [Sitemap](#) | [Privacy & Opting Out of Cookies](#)

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.
© Copyright 2022 IEEE - All rights reserved. Use of this web site signifies your agreement to the terms and conditions.