The Implementation of a 2-core, Multi-threaded Itanium® Family Processor

Samuel Naffziger

Blaine Stackhouse

Tom Grutkowski

Intel Corp. Fort Collins, CO

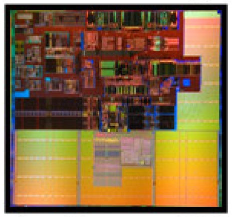
Outline

- Processor overview
- Core improvements
  - Cache hierarchy
  - Multi threading
  - Power reduction
- Handling 4 voltage and frequency domains
- Power management and reduction
- Conclusion

Itanium’s Progress

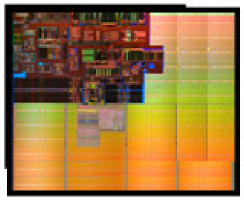
- 90nm implementation optimized for server workloads
  - Two dual-threaded high performance 64b EPIC cores on a single chip
  - (very) Large on-chip caches totaling 26.5MB
    - The same latencies as previous Itaniums (paper 26.8)
- 1.72 Billion transistors, 596mm^2
- Legacy system bus at higher frequency (666MT/s)
- High frequency operation representing at least a 20% increase over the 130nm implementation of the same core micro-architecture
- Power efficiency improvements that deliver a >2 fold compute capability increase at 23% *less* power consumption

McKinley



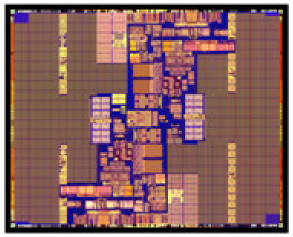
Single Core, 180nm
3MB L3
1 GHz, 130W, 1.5V

Madison



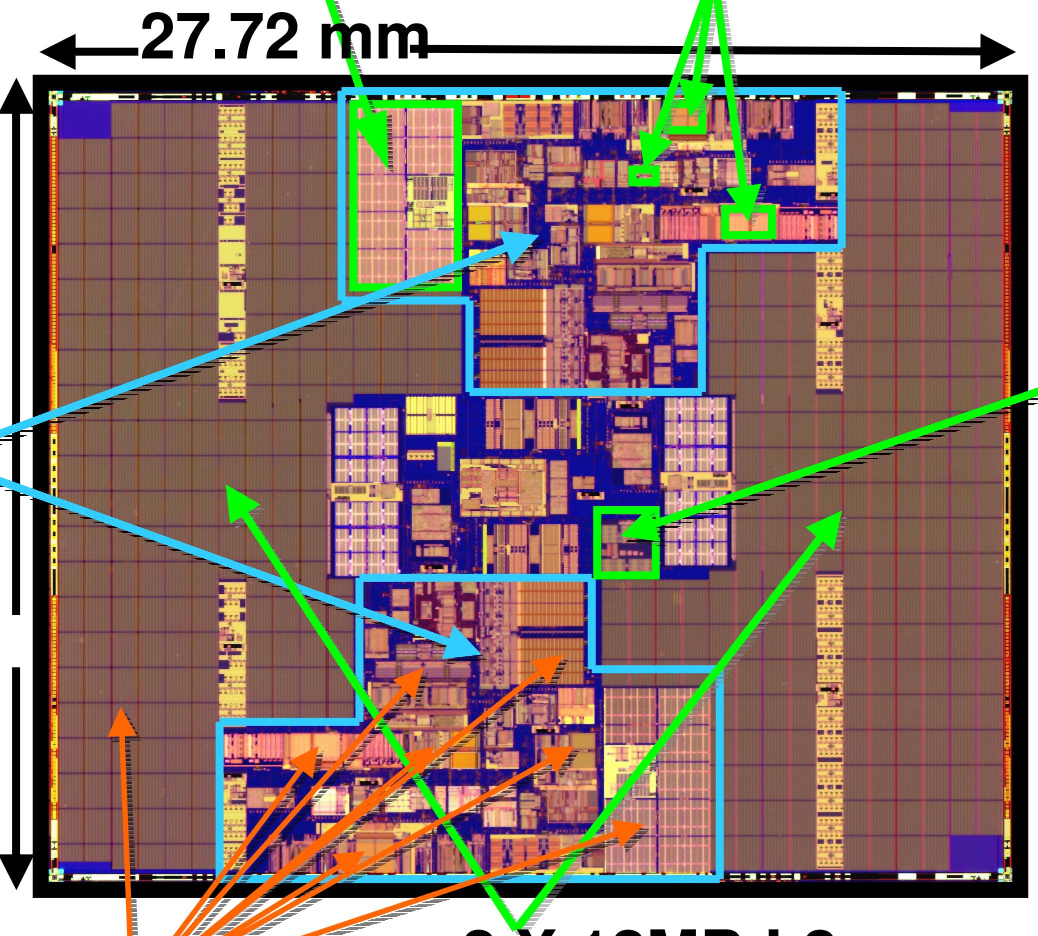
Single Core, 130nm
9MB L3
1.6 GHz, 130W, 1.35V

Montecito



Dual Core, 90nm
24MB L3
1.8+ GHz,
100W, 1.0-1.2V

Business Critical Features



1MB L2I

2-Way Multi-Threading

1.72 Billion transistors

Dual Cores

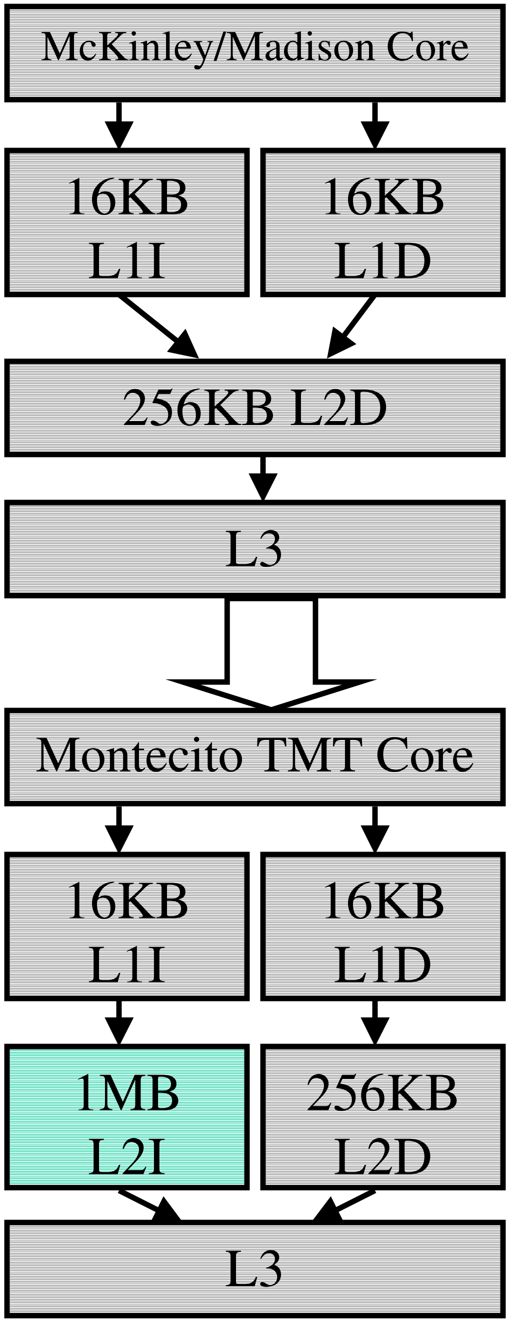
Foxton Power Controller

Soft Error Detection and Correction

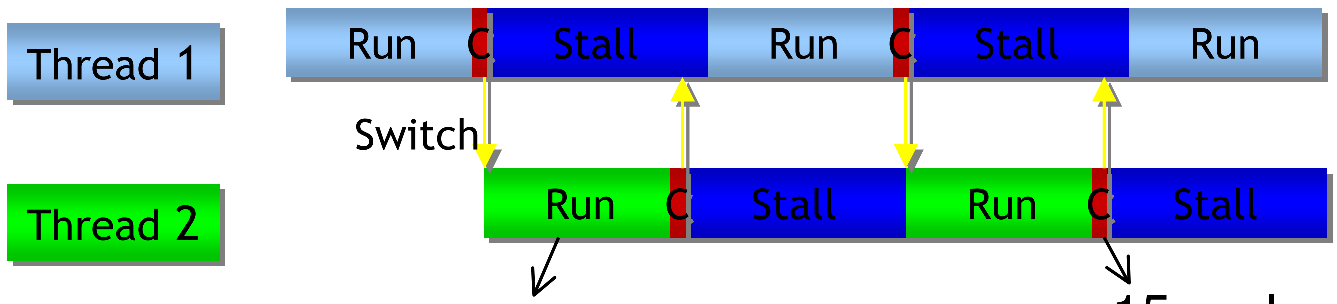
2 X 12MB L3 Caches with Pellston

Core Design Changes

- Temporal Multi-threading
  - Very low cost throughput performance boost
- New level of cache – 6 cycle 1MB L2I
  - Addresses largest CPI component for transaction processing (Instruction misses)
  - Frees 256KB L2D to be dedicated for data
- ECC add in L2T tags, and parity added to L1I TLB
- Parity in FP and Integer register files
- Second Shift/merge unit for encryption performance
- Power reduction with support for dynamic frequency and voltage

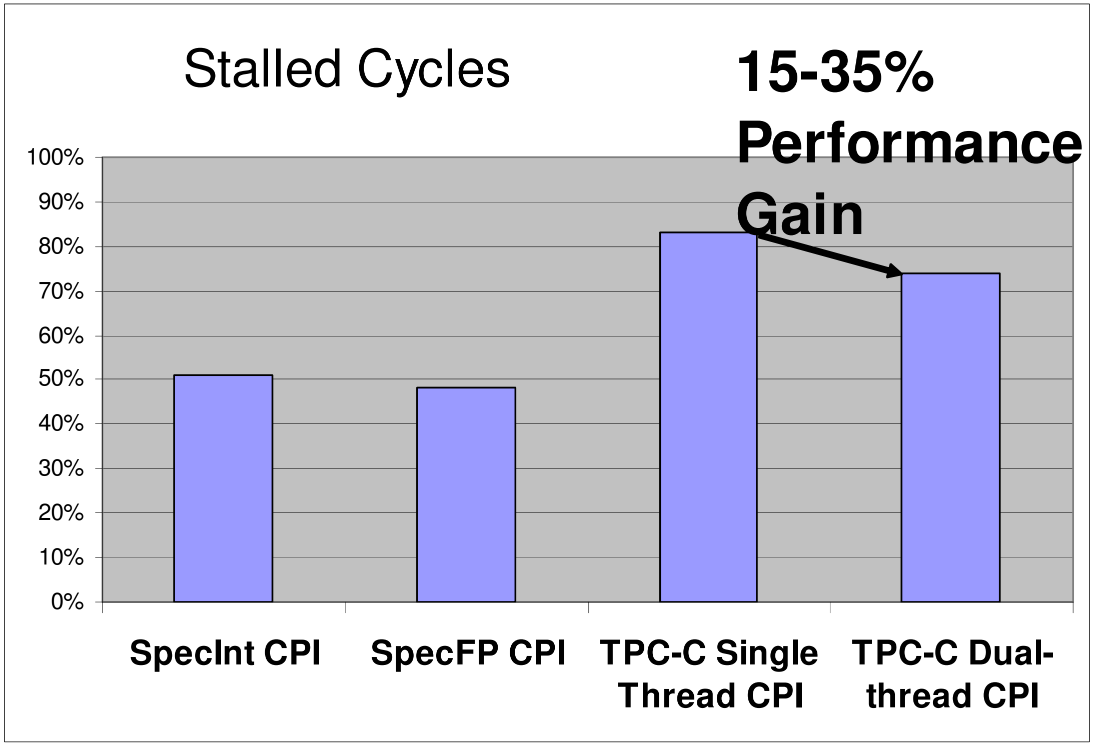


Temporal Multi-Threading



Overlap memory latency stall of
Thread 1 with execution of Thread 2

15 cycle penalty



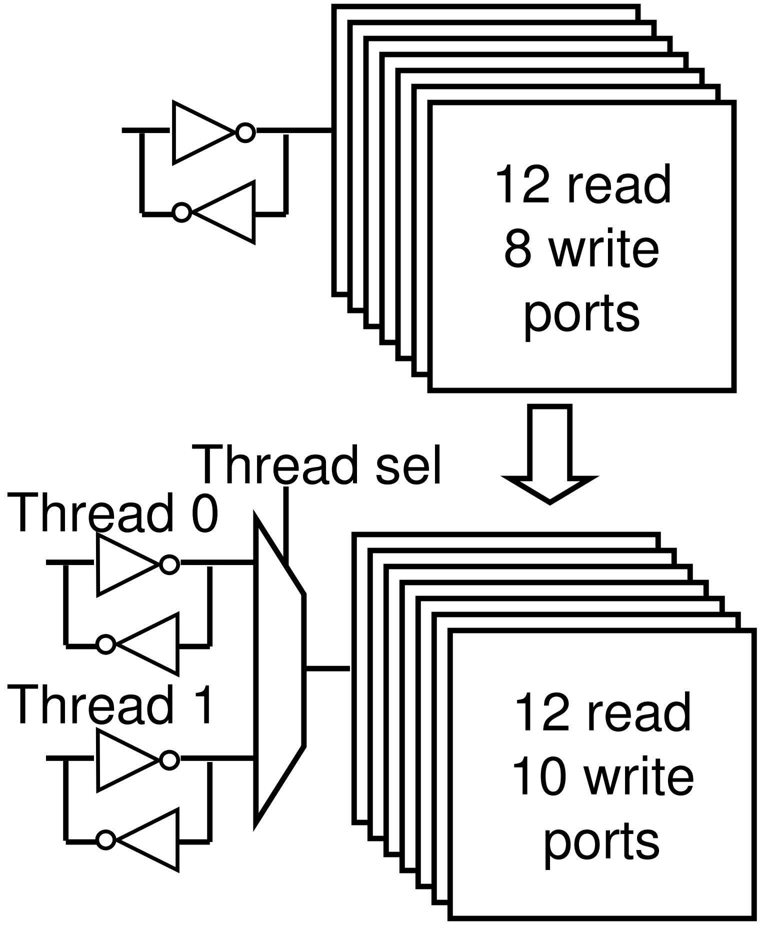
15-35% Performance Gain

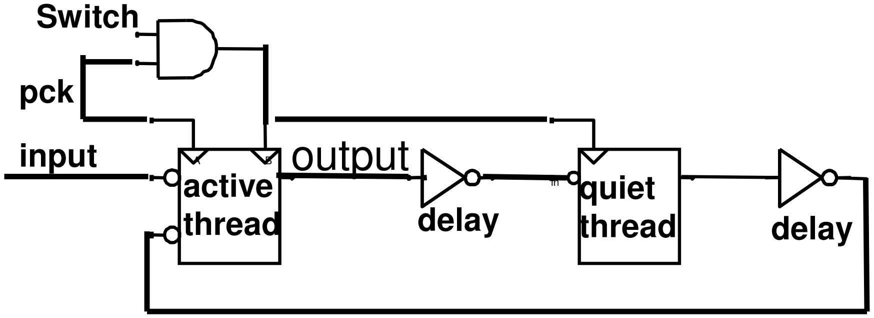
2% core area impact and 0% cycle time impact

Multi-Threading Implementation

- 15% register file growth with TMT using mux scheme
  - See Paper 20.5
- Key architectural state threaded, rest is flushed on a switch
  - 2X latch area for 1.3% of total latches

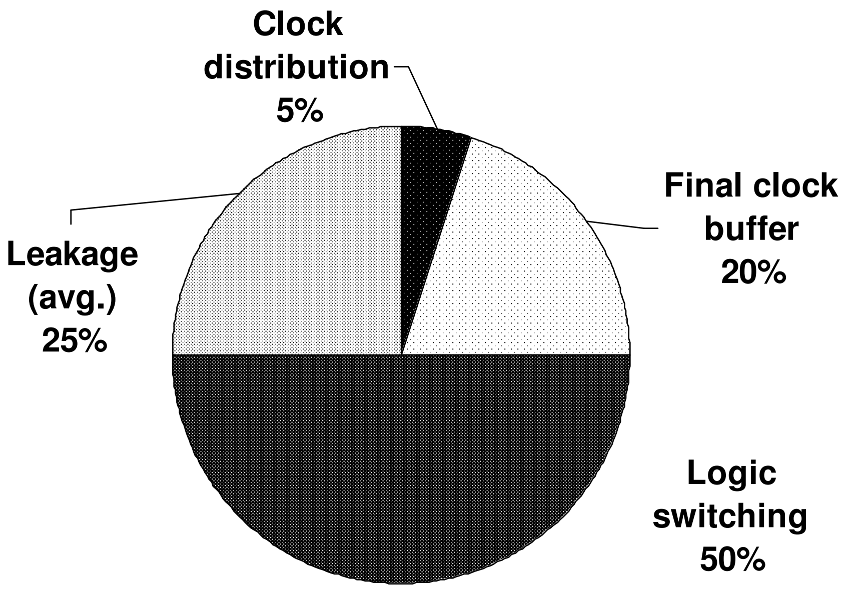
Register file impact

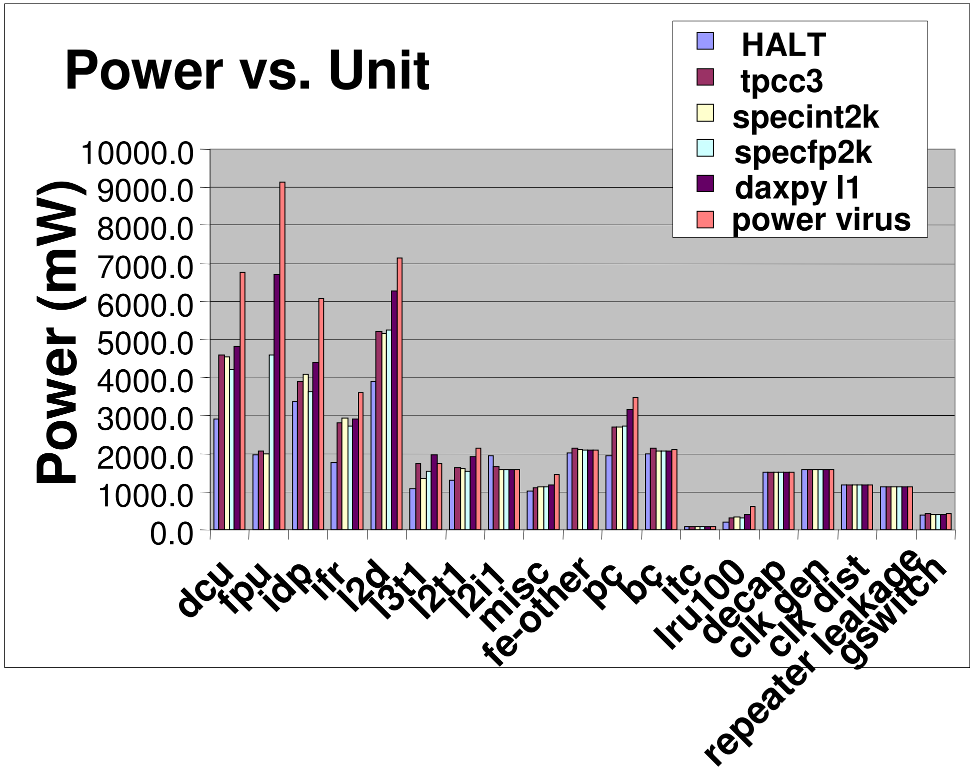




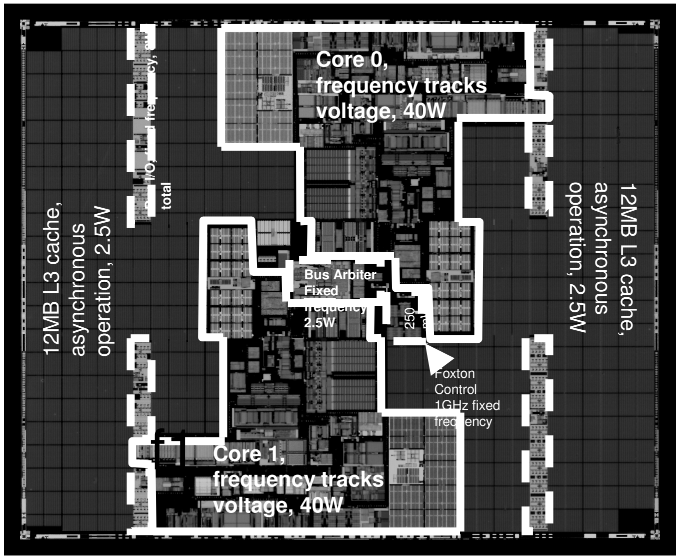
Core Power Reduction

- To improve power efficiency, we focused the team on reducing power consumption for *typical applications*
  - Developed a vector based tool to accurately measure switching and leakage power
  - Achieved a 28% overall reduction from the ported core even with the new features.



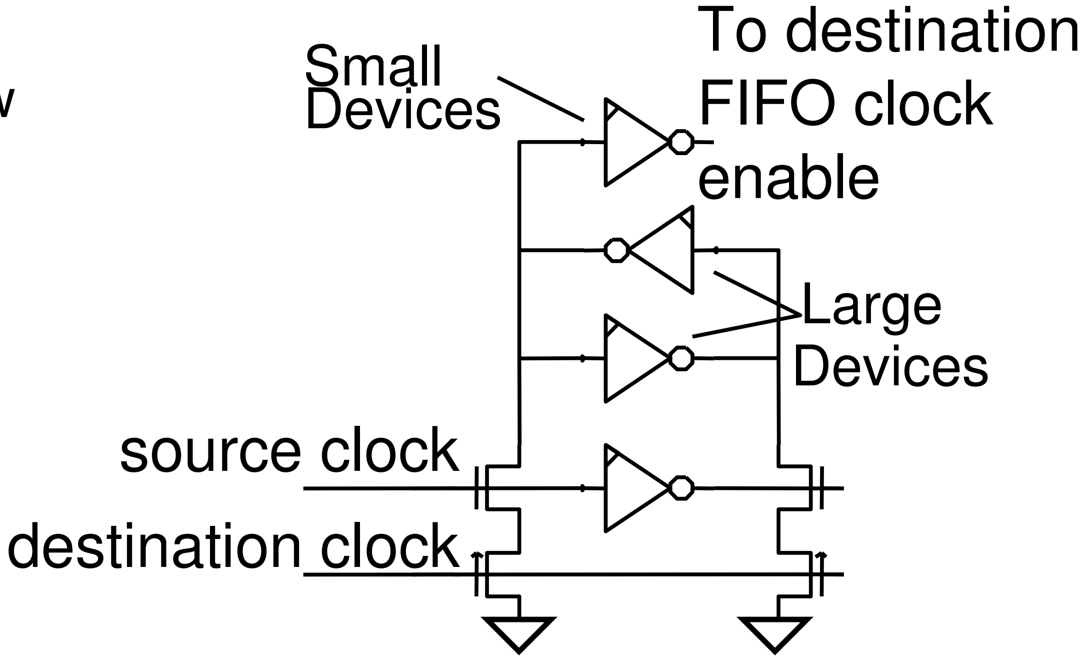


Multiple Voltage and Frequency Domains



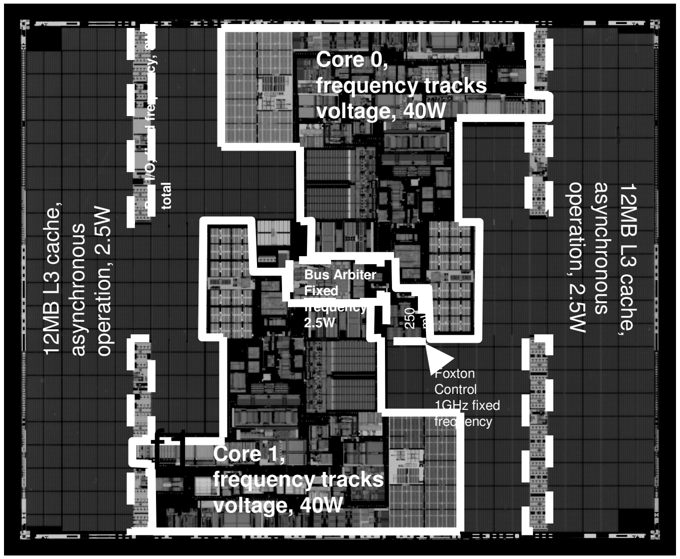
- To ensure minimal latency impact, a high gain resolver is needed to cross clock domains

Failure probability Proportional to $\text{Exp}\left(\frac{-T_w}{Tr}\right)$
Where Tw is the time window for resolution, *Tr* is the resolver time constant ≈ 5ps

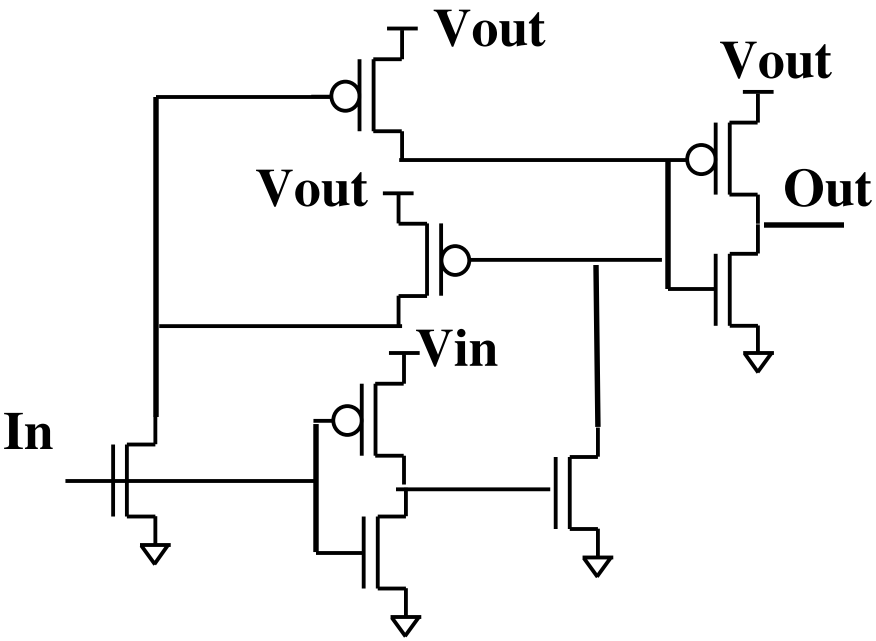


- Determinism for debug / test provided with a fixed frequency mode (FFM) enabled with dynamic deskew

Multiple Voltage and Frequency Domains



- To ensure reliable crossing of voltage domains, with up to 400mV offsets, a robust level shifter is used
  - Vcache ←→ Vcore
  - Vcore ←→Vfixed
  - Vfixed ←→Vtt
- A tool was developed to identify all domain crossings and check for the presence of level shifters

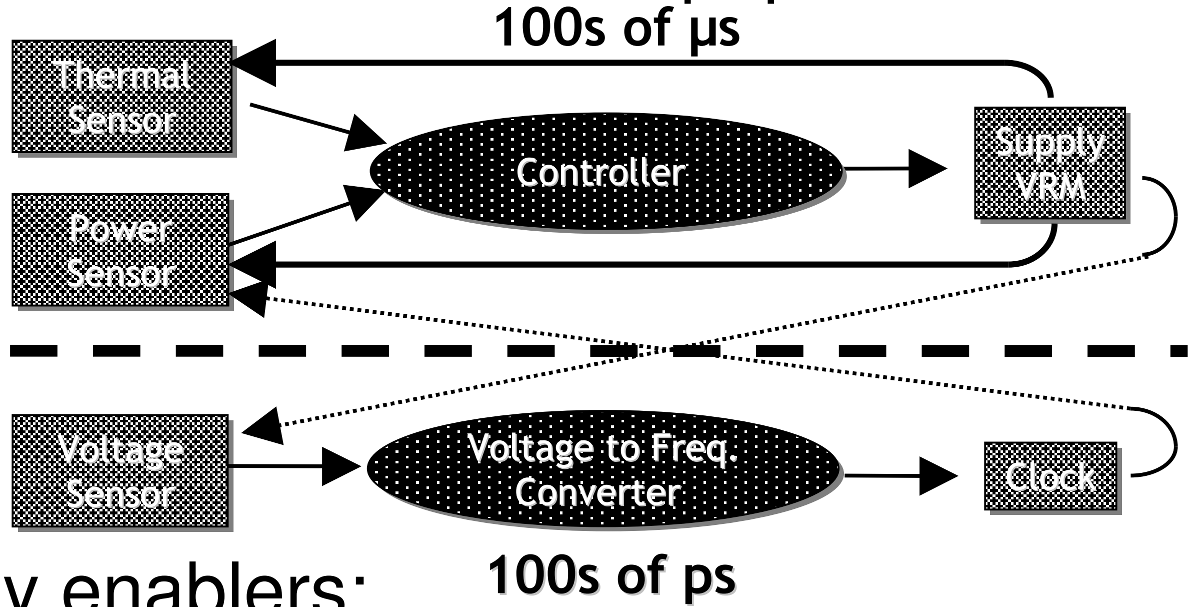


Power Efficiency Improvements

- A primary design goal of Montecito was the improvement of power efficiency
  - *Performance / Watt*
  - Lower acquisition and operating costs, reduced form factor, better compute density
- One must measure what is being optimized
  - Temperature measurements do not suffice – an environmentally dependent proxy for power
  - ➔ An ammeter is required
- To avoid wasted watts, have the chip instantaneously optimize operating point for variations in {voltage, temperature, workload, silicon processing}
  - ➔Requires self selection of voltage with dynamic frequency in conjunction with temperature and current measurement

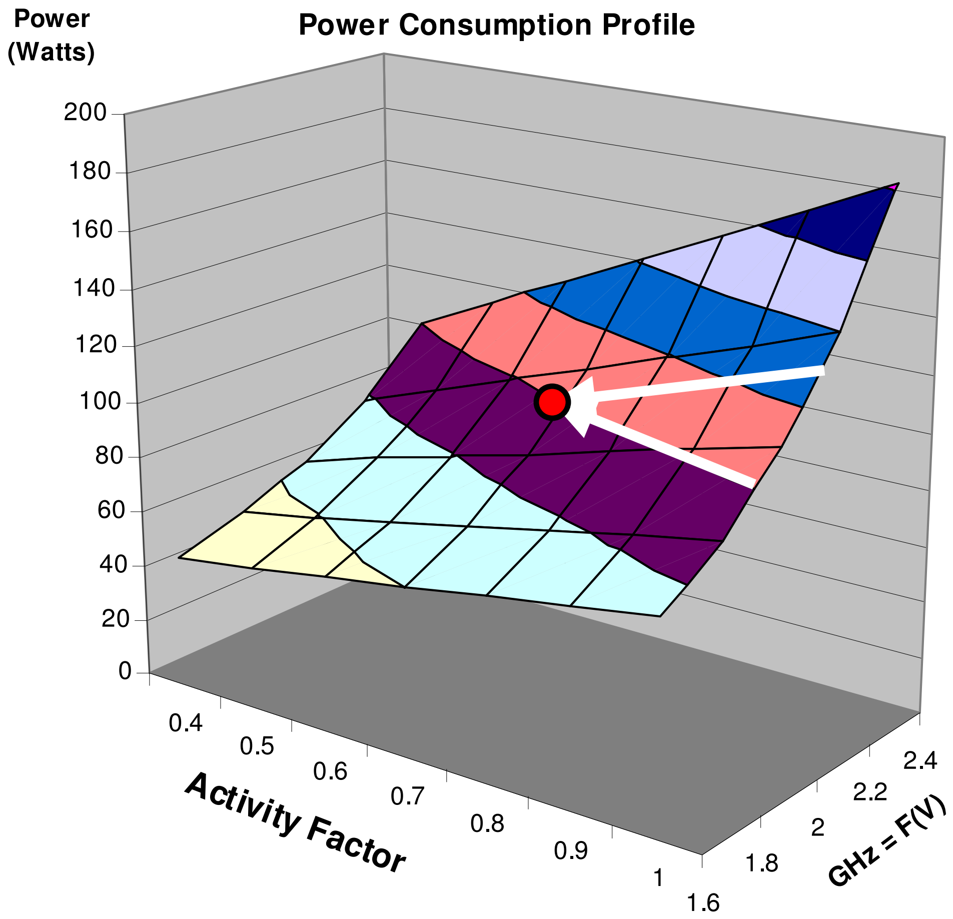
Power Management Loop

- For further details, see paper 16.7



- Key enablers:
  - High accuracy ammeter
  - Dynamic voltage control
  - Fast frequency synthesis as a function of Vcc
  - Accurate thermal measurement

Power Consumption Contour

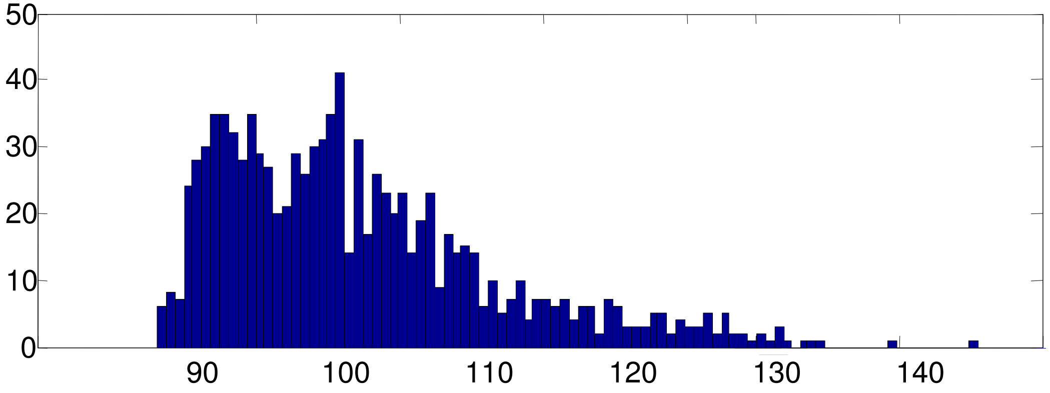


Optimization point is for typical *integer* applications which have .6X the switching power of the worst case ➔ Amdahl's law

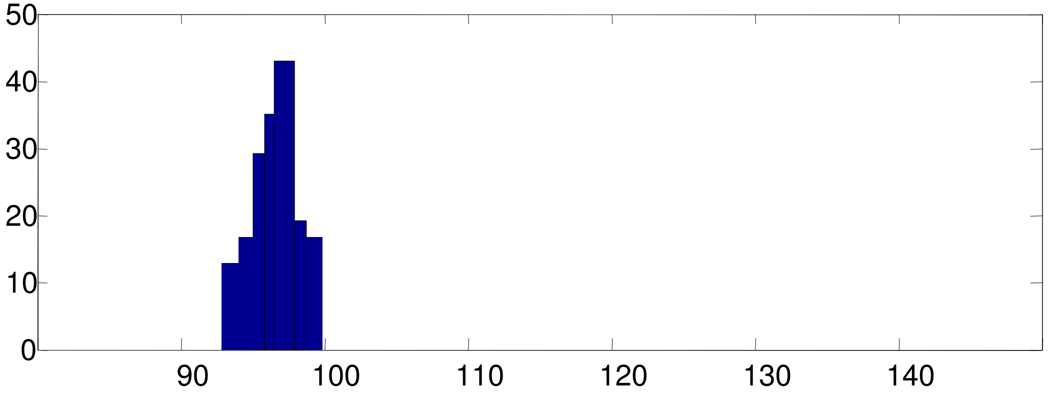
Manufacturing test is accomplished by observing the *self measured power*, and the *self-generated frequency* for typical code at the power limit

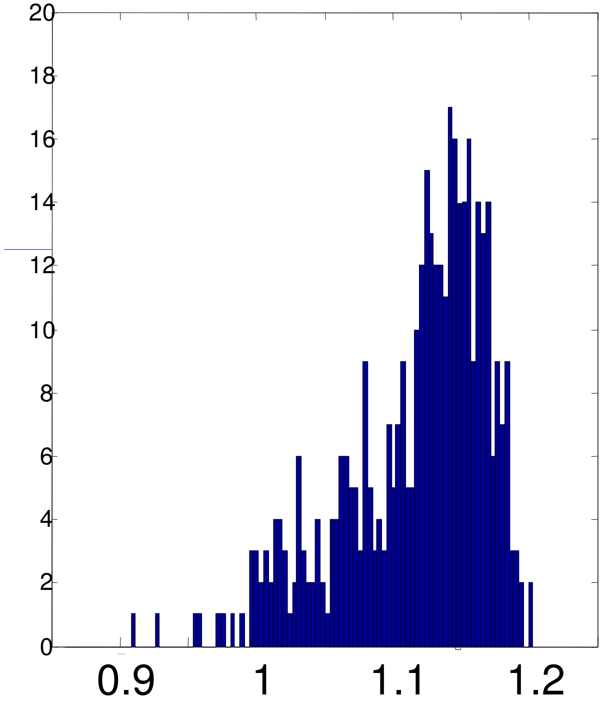
Per-Part Self-Optimization

Power Distribution at Fixed V/F



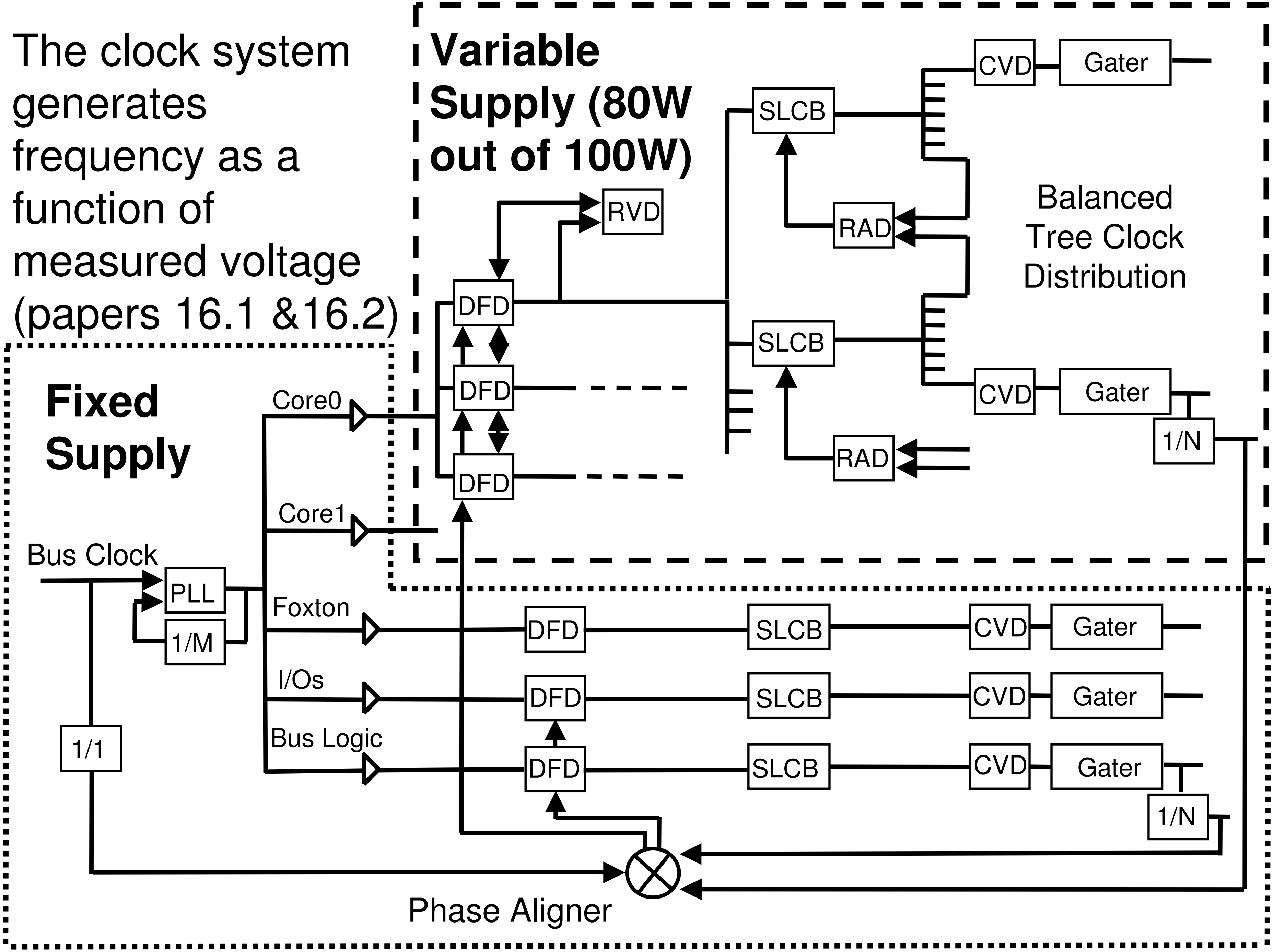
Power Range with Foxton



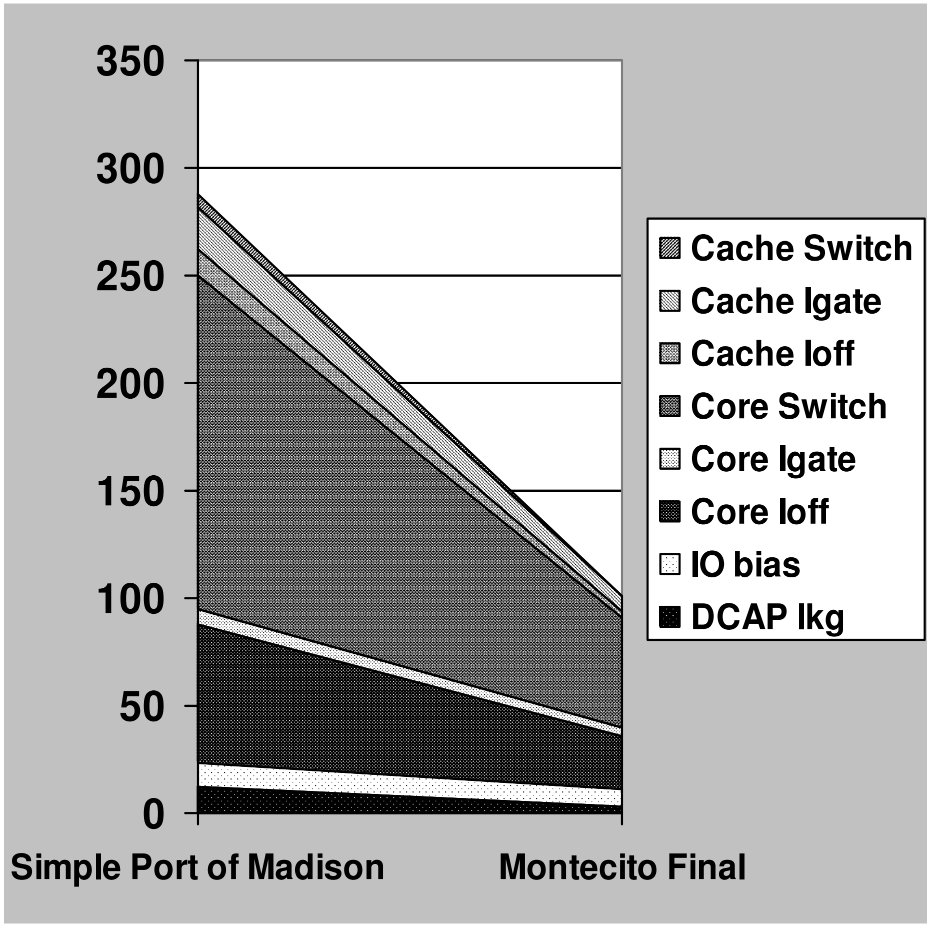


Resulting Vdd
Distribution

The clock system generates frequency as a function of measured voltage (papers 16.1 &16.2)



Feature	Frequency Gain @ 100W
Manage to application power vs. max power	12%
Adapt Vcc to optimal value for each part	5%
Manage junction temperature to the minimum possible	3%
Adapt frequency to Vcc to operate at frequency(V_{avg}) vs. frequency(V_{min})	7%
Optimize circuits for low switching and low leakage	11%
Cache power optimization (low V, asynch. etc.)	5%



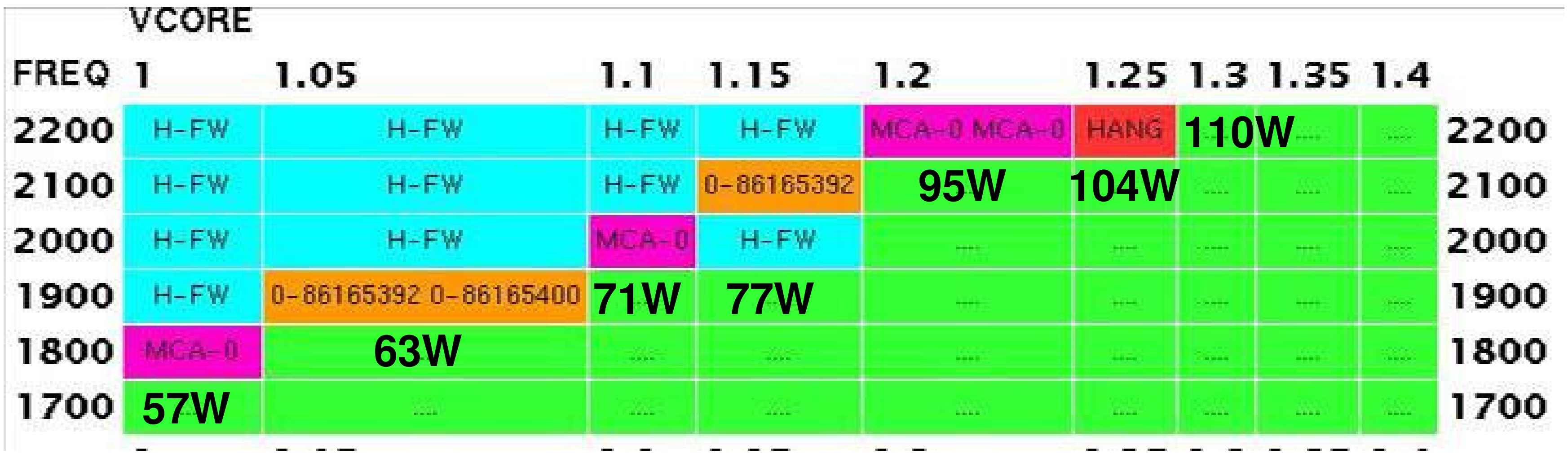
For Power, since P ~ F^3:

$$P = P_{orig}/(1+.12+.05+.03+.07+.11+.05)^3 = P_{orig}/2.92$$

This improvement is for a *typical* application.

High power floating point code may see up to a 10% reduction in frequency

Shmoo



FREQ	VCORE 1	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	
2200	H-FW	H-FW	H-FW	H-FW	MCA-0 MCA-0	HANG	110W	.....	.....	2200
2100	H-FW	H-FW	H-FW	0-86165392	95W	104W	.....	.....	.....	2100
2000	H-FW	H-FW	MCA-0	H-FW	.....	.....	.....	.....	.....	2000
1900	H-FW	0-86165392 0-86165400	71W	77W	.....	.....	.....	.....	.....	1900
1800	MCA-0	63W	.....	.....	.....	.....	.....	.....	.....	1800
1700	57W	.....	.....	.....	.....	.....	.....	.....	.....	1700

Power Consumption is for all 4 voltage domains at ~40C Tj

Summary

- Design builds on Itanium's performance strengths:
  - Improves leadership cache hierarchy
  - Improves instruction throughput further with dual-core, multi-threading and new execution units
- Bolsters reliability
  - Parity in register files, improved ECC
  - Seamless adaptation to power or thermal overages
- Achieves leadership performance / Watt
  - Breakthrough power measurement and management technology provides flexibility and efficiency