UT ECE | THE UNIVERSITY OF TEXAS AT AUSTIN
DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

Laboratory for Computer Architecture

The University of Texas at Austin

Predictive Power Management for Multi-Core Processors

Workshop on Energy-Efficient Design (WEED), June 2010

W. Lloyd Bircher and Lizy John

Presented by Jeff Stuecheli

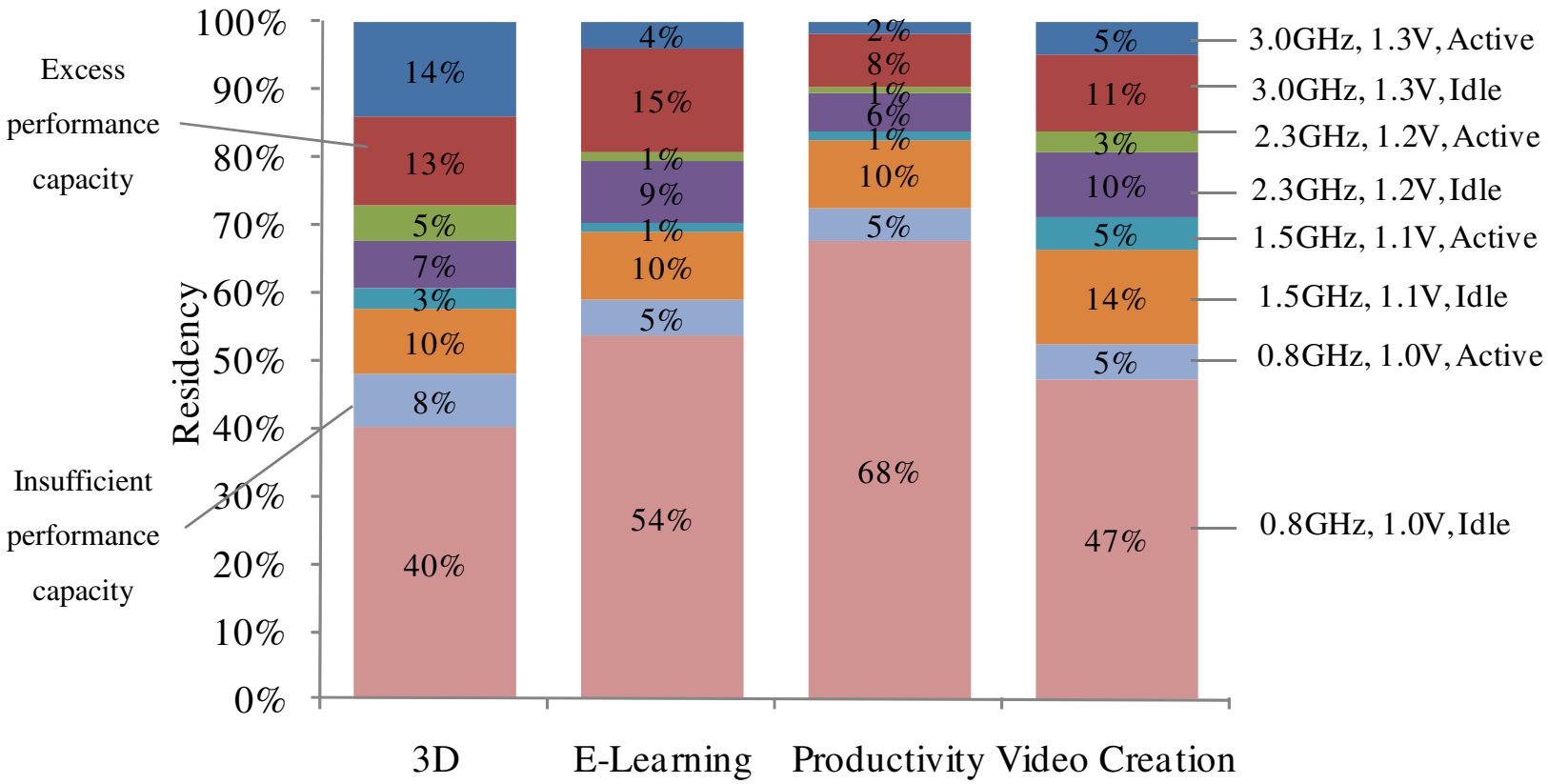
Motivation

- Dynamic power management improves energy efficiency by adapting performance capacity to match performance demand
- Adapting performance capacity imposes power and performance penalties
- Higher efficiency can be achieved by avoiding adaptations for short-duration phases and “pre-fetching” adaptations for impending phase changes

In this Presentation…

- Characterization of multi-core CPU activity in recent desktop productivity workloads
- Table-based predictor for CPU active-idle and idle-active transitions
- Application of predictor to improve selection of DVFS states for power efficiency and frequency boosting

Windows Vista Reactive Power Management



SYSMark2007 P-State and C-State Residency

Power and Performance Impact

- Reactive power management performs well for steady-state activity patterns
- Bursty activity patterns cause reactive schemes to over or under allocate performance

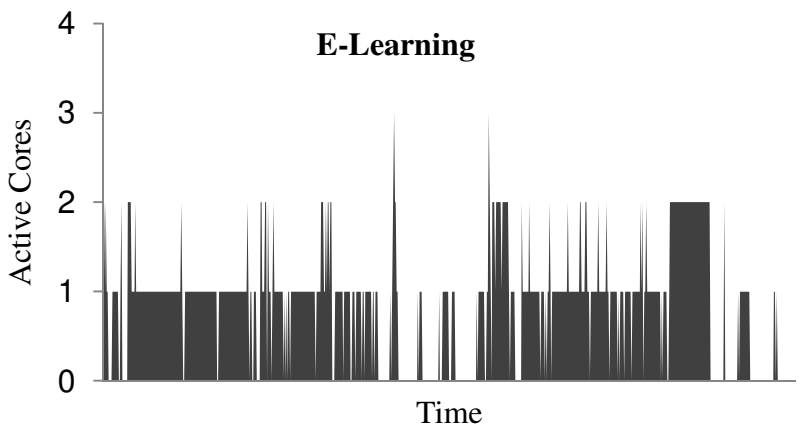
Reactive Power Management versus Oracle*

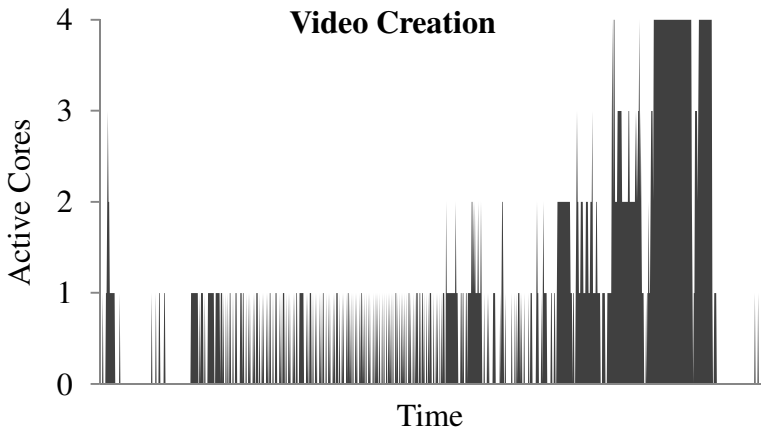
	3D	E-Learning	Productivity	Video Creation
Performance	-8.8%	-6.2%	-9.5%	-5.9%
Power	+24%	+35%	+21%	+28%

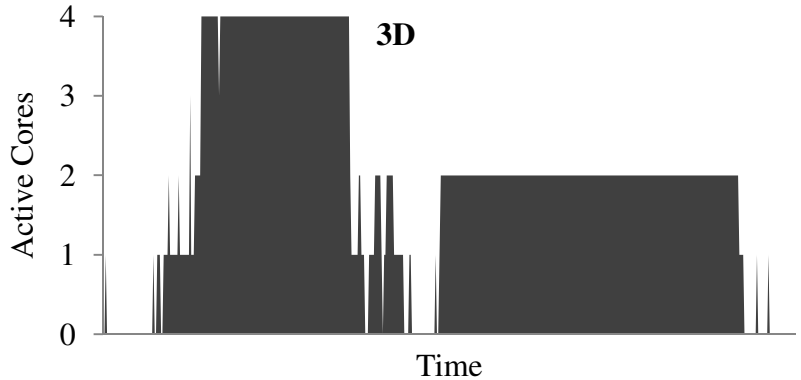
*Measured performance loss for Windows Vista balanced mode versus fixed frequency at Fmax

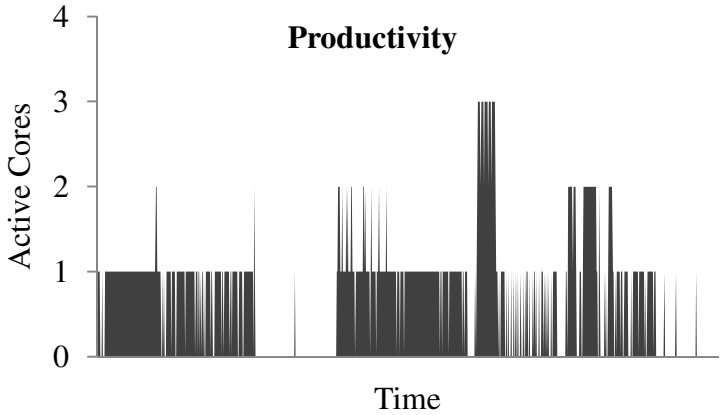
*Core power increase due to idling at > minimum frequency

Core Activity Patterns





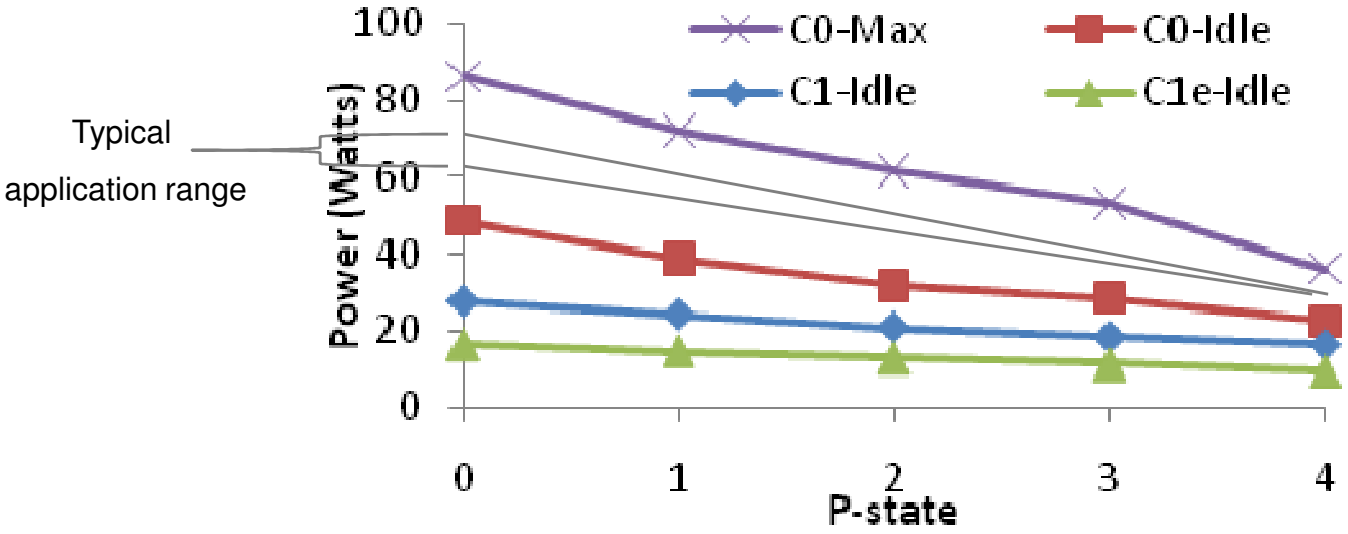




Concurrent Core Activity by SYSmark® 2007 Benchmark

Core Power Variation

- Variation in core power consumption is dominated by the difference between active execution and explicit idle phases



C0-Max	All Cores Active IPC ~ 3
C0-Idle	All Cores Active IPC ~ 0
C1-Idle	At Least One Active Core, Core ~ 0 MHz
C1e-Idle	All Idle, Core ~ 0 MHz, MemCntrl ~ 0 MHz

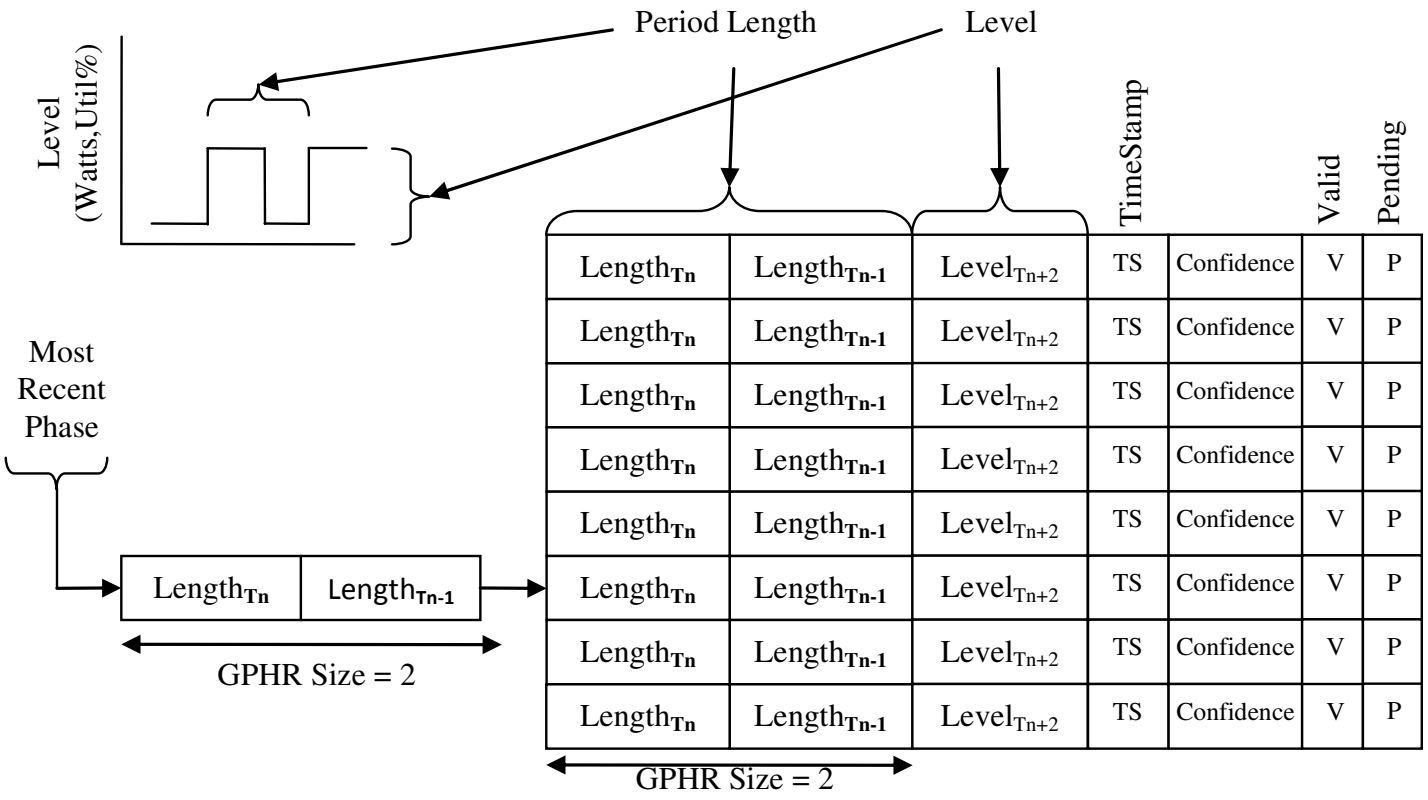
Predictable Workload Characteristics

- Many active-idle, idle-active events are regular and predictable
- Idle-Active caused by interrupts
  - I/O device interrupts, End-of-DMA events, Inter-Processor-INTs
- Active to idle caused by lack of work in ready queue
  - OS or application controlled affinity
  - Workload intensity, audio/video processing
  - One page fault, OS may idle a core while data is brought into memory

Predictor Characteristics

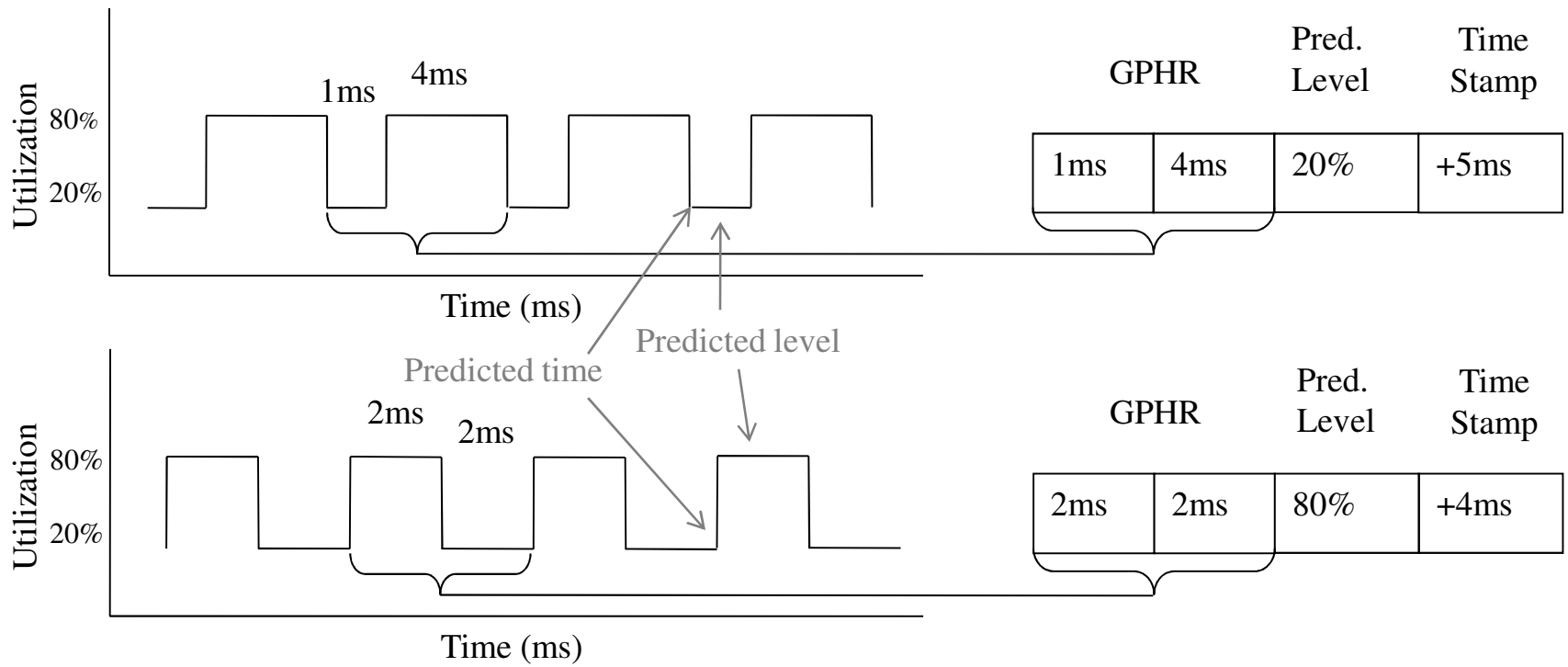
- Detects active-idle and idle-active patterns of particular length
- Detection uses quantized activity levels of 0% or 100%
- Predicts activity level or power in phase immediately following detected pattern. Predicted value is high-resolution (not quantized to two levels).
- When a repeated pattern is detected in the GPHR (Global Pattern History Register), it is predicted to be the next pattern.

Predictor Description



Power Phase Predictor

Predictor Description



Example of Program Phase Mapping to Predictor

Predicting Activity Levels for DVFS Selection

- Predictor compared to Windows Vista Reactive P-State selection algorithm on basis of ideal frequency selection
- Windows reactive scheme is comparable to last value predictor. Next value is same as last observed.
- Vista target core activity range is 30% to 50%.
  - If < than 30% activity then $Frequency_{Next} = Frequency_{Current} * Activity_{Current} / 30\%$
  - If > than 50% activity then $Frequency_{Next} = Frequency_{Current} * Activity_{Current} / 50\%$

Predicting Activity Levels for DVFS Selection

- A prediction is considered correct (a hit) if the selected p-state caused the system to operate within the ideal range
- The reactive p-state selections and core activity levels were traced on 1ms intervals
- P-state selections were performed similarly for our proposed predictor. Based on predicted core utilization level, p-state selections are made to in order to meet the 30%-50% target.

Predicting Activity Levels for DVFS Selection

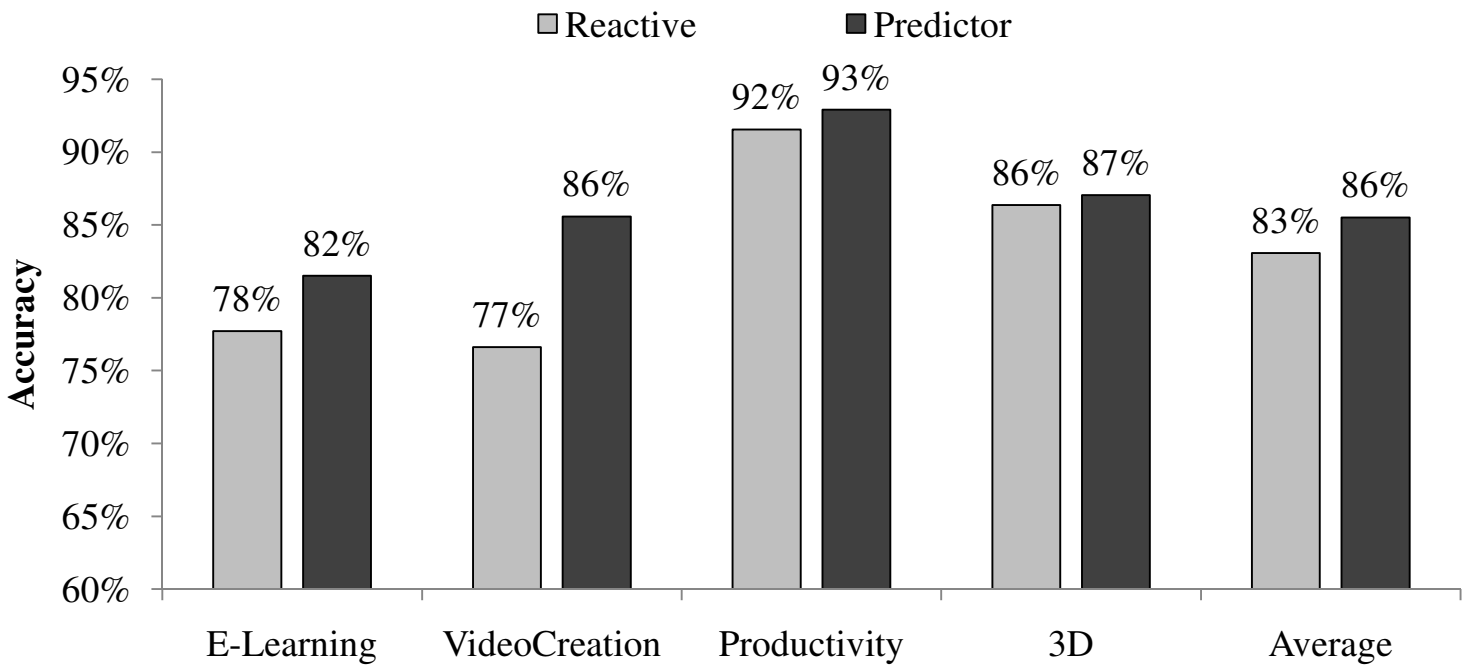
Prediction Outcomes – Core Activity

Predictor, 48-Entry PHT				
Workload	**Weighted Match Ratio**	**Weighted Hit Ratio**	**Prediction Coverage**	**Effective Hit Ratio**
E-Learning	92.8%	98.8%	57.0%	82.6%
Productivity	95.5%	95.8%	33.5%	73.8%
Video Creation	95.1%	97.7%	43.0%	76.4%
3D	98.3%	99.1%	37.9%	72.7%
Reactive				
Workload	**Weighted Match Ratio**	**Weighted Hit Ratio**	**Prediction Coverage**	**Effective Hit Ratio**
E-Learning	100%	66.4%	100%	66.4%
Productivity	100%	65.2%	100%	65.2%
Video Creation	100%	63.5%	100%	63.5%
3D	100%	59.7%	100%	59.7%

Predicting Power Core Power Consumption

- Alternative application of predictor: Predict core power consumption instead of core activity level
- Predicted power level is compared to modeled per-core power level
- Power prediction is applied to frequency boosting to improve performance during low power phases

Predicting Power Core Power Consumption



Core Power Accuracy Predictive versus Reactive

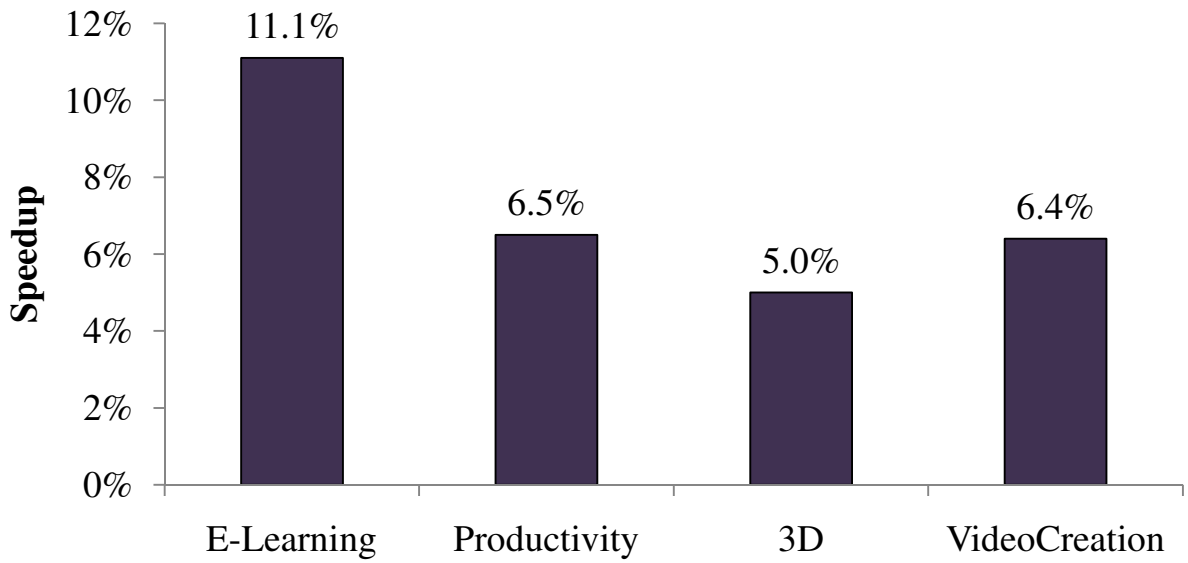
Predicting Power Core Power Consumption

- Predicted power level consistently more accurate than last value estimate
- Improvement range from 1% to 9% with an average of 3% across SYSMark2007
- Power prediction is easier for Productivity and 3D workloads due to high percentage of steady state execution
  - Productivity has large phases with all cores idle
  - 3D has large portion of time with all cores active
  - For these workloads, the benefit of prediction is much less

Predicting Power for Frequency Boosting

- Estimate performance benefit from applying power prediction to frequency boosting
  - Trace core utilization and modeled power levels
  - Simulate performance benefit by accelerating execution of active phases when last and next power level is predicted to be below limit
  - Performance increase is 75% of frequency increase due to sub-optimal frequency scaling of application (measured).
  - A single boosted frequency of +20% is provided. Power consumption at this frequency was correlated to actual silicon

Predicting Power for Frequency Boosting



SYSmark 2007 Speedup

Predicting Power for Frequency Boosting

- Given that the prediction accuracy for the four SYSMark subtests are similar, most of the difference is due to baseline power consumption
- The mostly single-threaded E-Learning has the highest performance uplift at 11.1%. In the best case, a truly single-threaded application would obtain +15% (20% Freq x 75% sensitivity).
- The lowest performance was obtained in 3D which has nearly 50% of the time spent with all cores active.

Summary

- Identified opportunity to increase performance and efficiency by predicting transitions from active-idle and idle-active in the SYSMark2007 productivity suite
- Presented a table-based predictor for CPU active-idle and idle-active transitions and power levels
- Applied prediction of power levels to frequency boosting

Backup

Quad-Core Performance Counter Power Model

- Power delivering for existing multi-core processors provided by a single shared rail.
- It is not possible to directly measure core-level power consumption
- Solution is to develop core-level power model based on performance counters
  - This model improves on existing models through the addition of temperature and voltage awareness

Quad-Core Performance Counter Power Model

- Core-level performance events and aggregate core power measured on a 3GHz Phenom2 processor
- Performance events include fetched ops, floating point ops executed, and level 1 data cache accesses
- Voltage and temperature awareness provided by modeling relationship between DVFS state and temperature to power
- Idle power management states also traced to account for coarse-grain clock gating

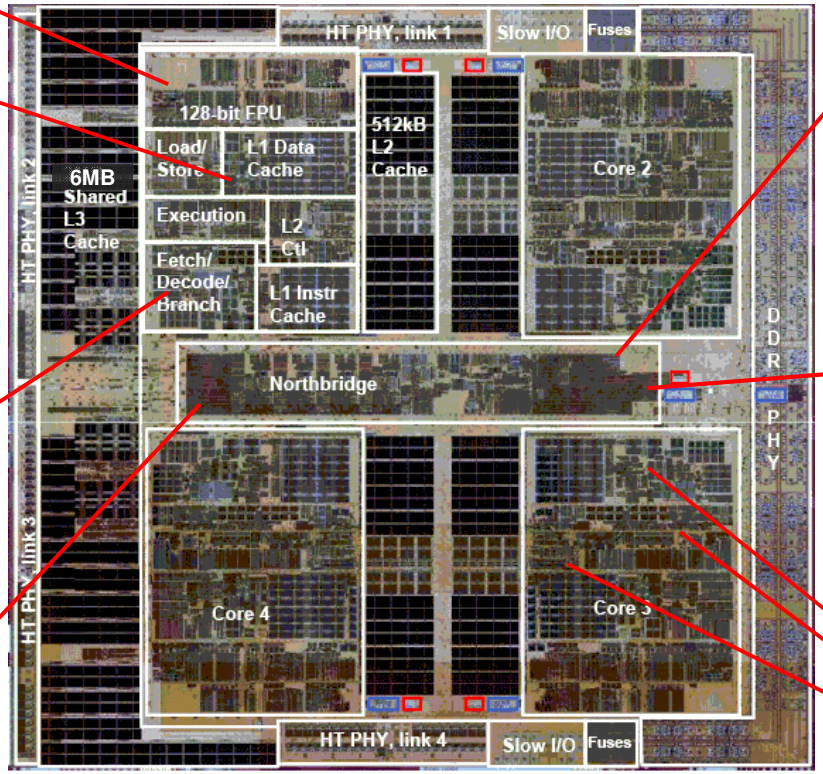
Quad-Core Performance Counter Power Model

Floating Point Ops Retired: higher power than integer

L1 DC Accesses : Proxy of memory subsystem

Fetched Ops : largest C0 impact on power. Includes speculation

Halt % : Architected HLT instruction can reduce dynamic power to ~0



AMD 45nM "Ridgeback" Processor

Frequency: Each core operates at independent frequency

Voltage: Drastic impact on leakage and dynamic power.

$V = \max(V_1, V_2, V_3, V_4)$

Temperature: Sensed throughout die. Major impact on leakage power.

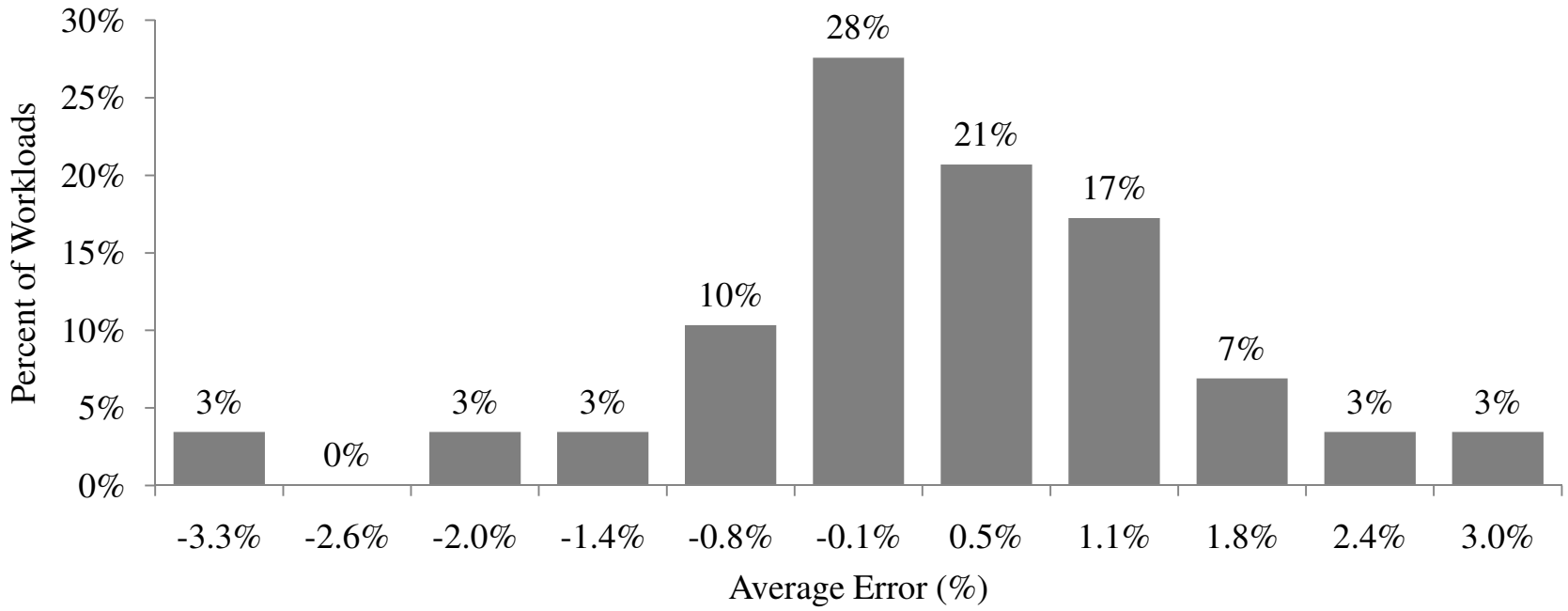
Quad-Core Performance Counter Power Model

Quad-Core Power Model

Power Models	Equation
Total Power	$\left(\sum_{N=0}^{3} WorkloadDependent_N + Ungateable_N + Gateable_N\right) + Static_{Volt,Temp}$
Workload Dependent Power	$((FetchOps_N/Sec)\cdot Coeff_F+(FloatPointOps_N/Sec)\cdot Coeff_{FP}+(DCAccess_N/Sec)\cdot Coeff_{DC)}\cdot Voltage^2$
Gateable Power	$(\%Halted_N)\cdot Coeff_{Gateable}\cdot Voltage^2\cdot Frequency_N$
Ungateable Power	$(\%NonHalted_N)\cdot Coeff_{Ungateable}\cdot Voltage^2\cdot Frequency_N$
Static Power	$(Temp^2\cdot Coeff_T{}^2+Temp^1\cdot Coeff_T{}^1+\cdot Coeff_T{}^0)_{Voltage_x}$

Quad-Core Performance Counter Power Model

Average error 0.89%, Worst-case error 3.3%



Model Error Analysis – SPEC CPU 2006 and SYSMark 2007