Power and Performance Challenges for High Performance Computing Architecture

Tawfik Arabi
Sr. Principal Engineer
Manufacturing and Validation Group, Intel Corporation

Mesa, Arizona ● March 3 - 6, 2019

Legal Notices and Disclaimers

- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.
- Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.
- No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.
- All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.
- Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.
- The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.
- Intel, the Intel logo and others are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.
- *Other names and brands may be claimed as the property of others.

Acknowledgements

- Mo Bashir
- Barnes Cooper

Agenda

- Power and Performance Requirement Historical and Future Trends
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

Agenda

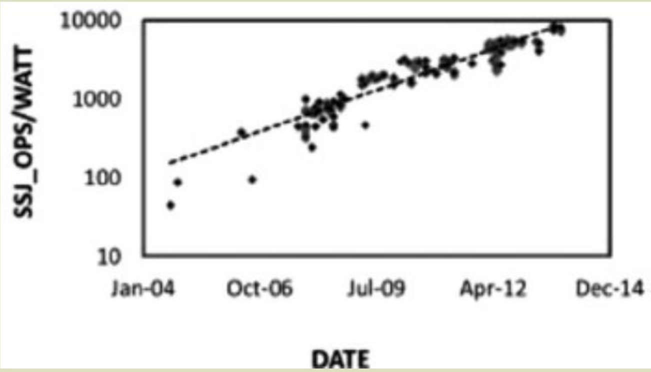
- Power and Performance Requirement Trends for the Next 5 Years
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

TestConX™

Power and Performance Challenges for High Performance Computing Architecture 5

2000-2019 20 YEARS ANNIVERSARY

Historical Trends for Data Center



"Dual-socket server energy efficiency, as measured by SPECPower, Intel-Xeon based systems versus their "hardware available" date. Note the logarithmic scale, indicating an exponential trend"- Energy Efficient Servers, Blue Print for Data Center Optimization by Corey Gough, Ian Steiner, and Winston Saunders

Facebook estimated that the energy used to sustain an average account for a month is about equal to the energy used to make a cup of coffee www.facebook.com/green/app_439663542812831

eBay's published data center energy use shows that the amount of carbon produced per transaction is about 50 times lower than the carbon produced in a short drive to the store to complete the same purchase http://tech.ebay.com/dashboard

"online purchasing of music uses 40%–80% less energy than any of multiple methods for delivering music by CD, even though that calculation used an upper bound estimate for the electricity intensity of Internet data transfers", Christopher Weber, Jonathan G. Koomey, and Scott Matthews, "The Energy and Climate Change Impacts of Different Music Delivery Methods," *Journal of Industrial Ecology* 14, no. 5(October 2010): 754–769 http://dx.doi.org/10.1111/j.1530-9290.2010.00269.x

All above quotes from *Energy Efficient Servers, Blue Print for Data Center Optimization by Corey Gough, Ian Steiner, and Winston Saunders*

Data Center Performance Trends

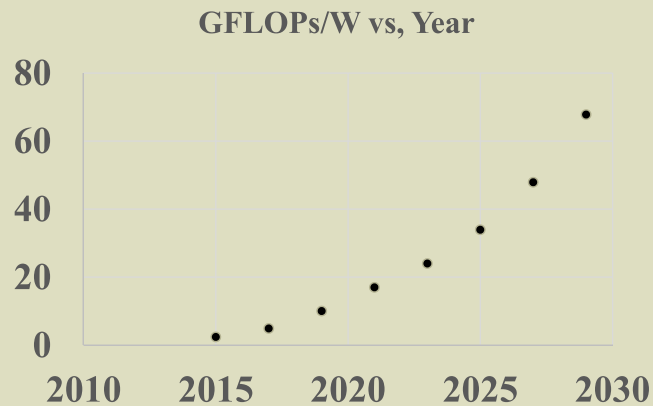
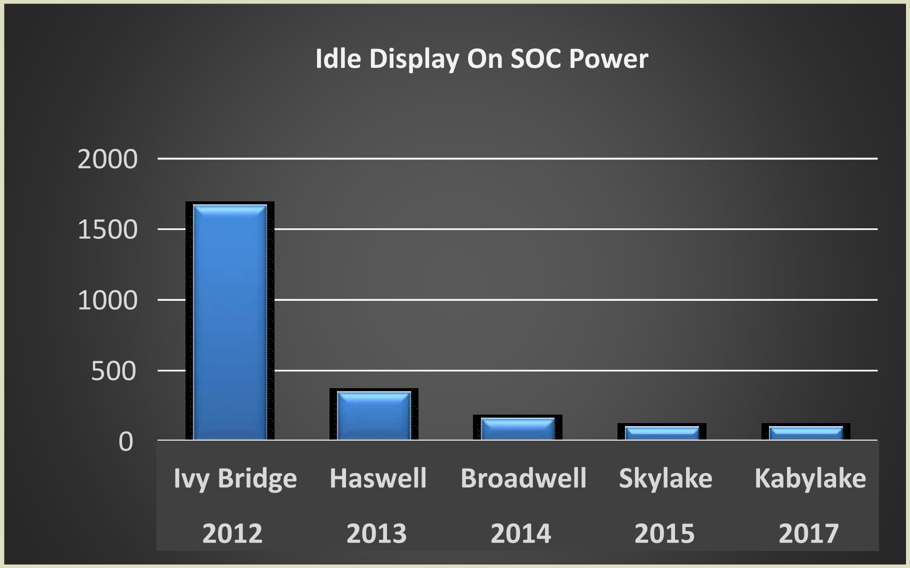
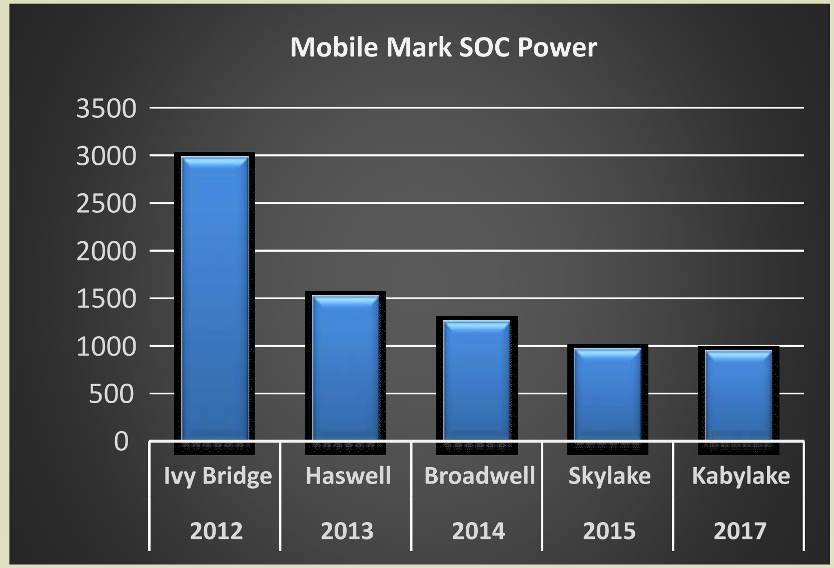


Figure 1: 2015 ITRS System Integration Data Center Efficiency Roadmap

- Demand for More Performance Per Watt Continues to Increase Across Market Segments, From Data Center to Mobile Devices –

Notebook Battery Life Trends





Battery Life Increasingly a Factor in Purchasing Decisions with Devices Operating Multiple Weeks In Standby Modes

Benchmark results were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown". Implementation of these updates may make these results inapplicable to your device or system.

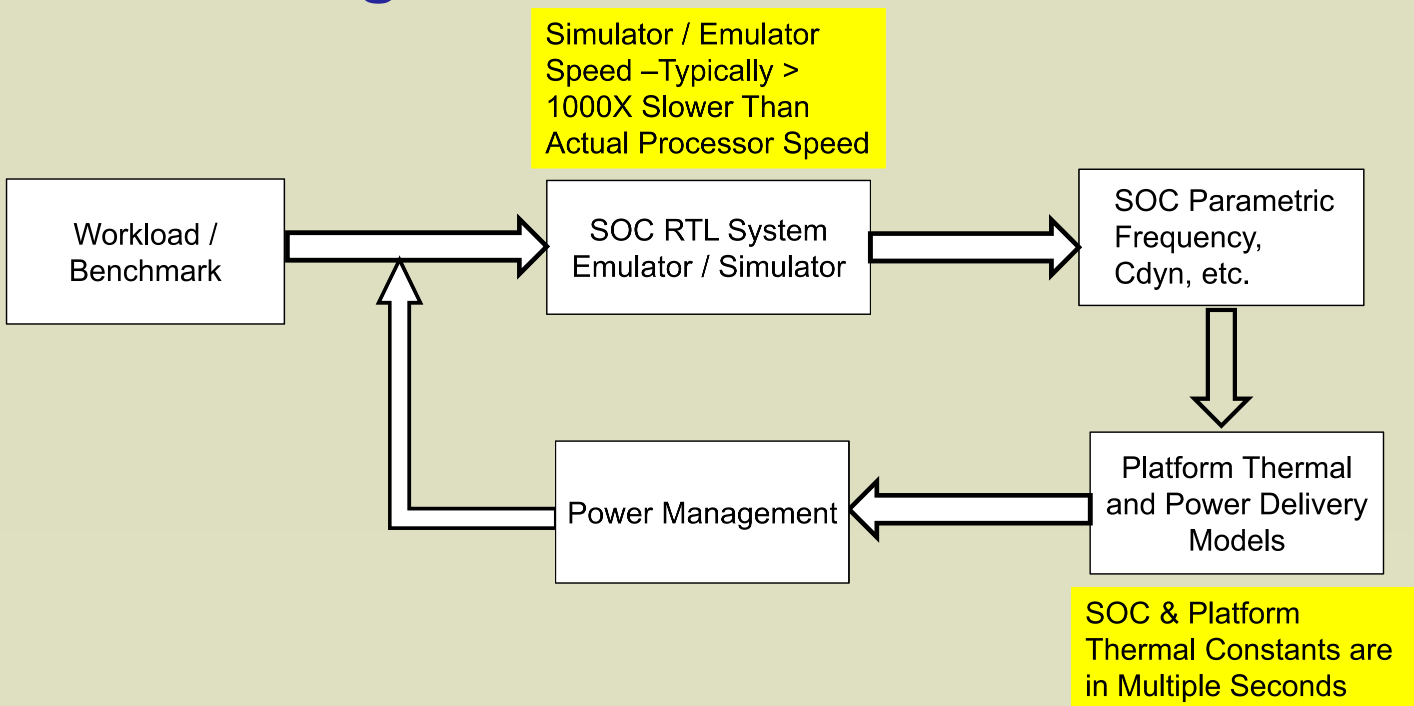
- Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.
- Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
- **For more information go to** http://www.intel.com/benchmarks

TestConX™

Agenda

- Power and Performance Requirement Trends for the Next 5 Years
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

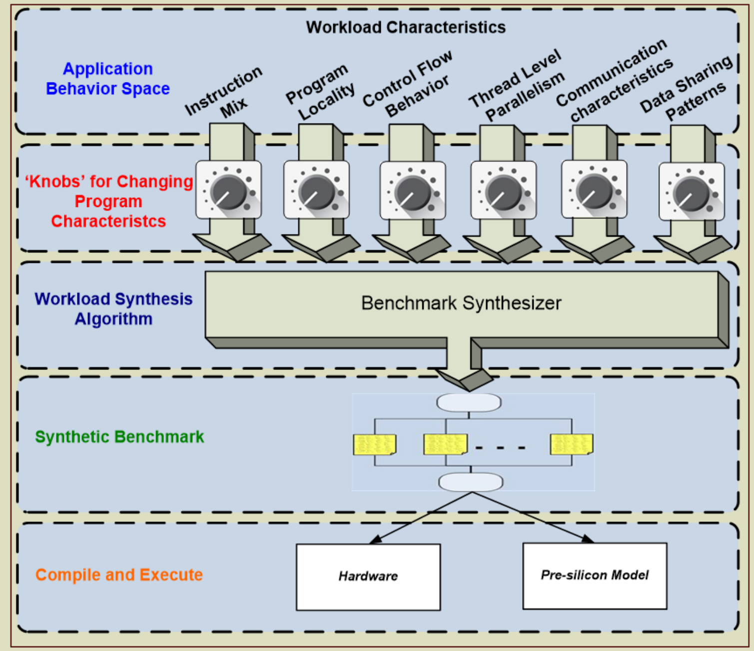
Modeling and Simulations in Pre-Silicon



Brute Force Simulation or Emulation of a Real Workload in a Closed Loop System is not Feasible

Work Load Proxies

- Intel and UT-Austin Teaming Up to Generate Workload Proxies > 1000X Smaller yet Retain the PnP Behavior of the Original Workloads
- This makes The Open Loop Simulation / Emulation of Benchmarks Possible
- Closed Loop Simulation / Emulation Remains a big challenge given order of magnitude Difference Between Thermal and Electrical Time Constants



Intel Sponsored Research at UT-Austin (Professor Lizy John)

Examples of Workload Proxies

Example: Geek Bench	Milli-Seconds
Frequency	2.35
Overall SC	3500
AES	2420
LZMA	3426
JPEG	3552
Canny	3532
Lua	4115
Dijkstra	4200
SQLite	3536
HTML5 Parse	3735
HTML5 DOM	5209
Histogram Equalization	3287
PDF Rendering	3494
LLVM	4670
Camera	4074
SGEMM	2270
SFFT	2798
N-Body Physics	3216
Ray Tracing	3517
Rigid Body Physics	3605
HDR	4041
Gaussian Blur	3698
Speech Recognition	3507
Face Detection	3787
Memory Copy	3499
Memory Latency	3790
Memory Bandwidth	2449

- Some Workloads Are Simple and Parallelizable Such As Geek Bench
- Other Workloads are More Complex and Have Serial Dependencies

- Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.
- Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
- **For more information go to** http://www.intel.com/benchmarks

TestConX™ Power and Performance Challenges for High Performance Computing Architecture 12

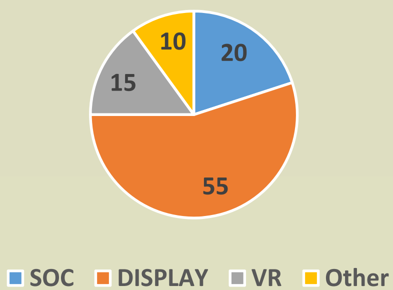
Agenda

- Power and Performance Requirement Trends for the Next 5 Years
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

Power Delivery Efficiency

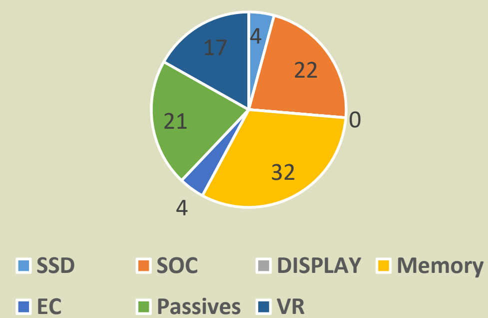
- Notebooks Need Multiple Weeks of Connected Standby Power on a Single Charge
- SOC and Other Ingredients Power Consumption Has Dramatically Reduced in Connected Standby Mode
- Display Power Also Reduced Significantly in Active Mode
- Voltage Regulator (VR) Efficiency Has not Significantly Improved
- Need High Efficiency Conversion from 60V to Sub 1Volt High Power in Data Center, 20/12V to Sub 1Volt in Client, and High Efficiency at Low Amperage for Standby Power

% Power Consumption of a Typical Notebook for Active Work loads



Numbers Can Vary significantly Based on Configuration, Display Resolution, Screen Size, and Display Technology

Typica; % Power Consumption of Notebooks in Moden Standby



Numbers Can Vary significantly Based on Configuration, SOC, Memory Size and Other Ingredients

TestConX™

Power and Performance Challenges for High Performance Computing Architecture 14

Thermal and Power Delivery Challenges

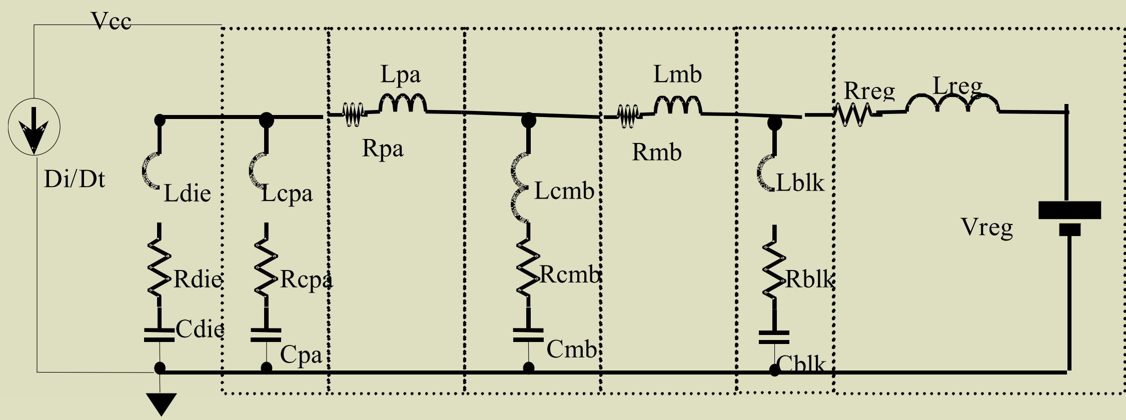
$$V_j = V_{id} - Z_e I$$

Power Delivery Cost

TestConX™

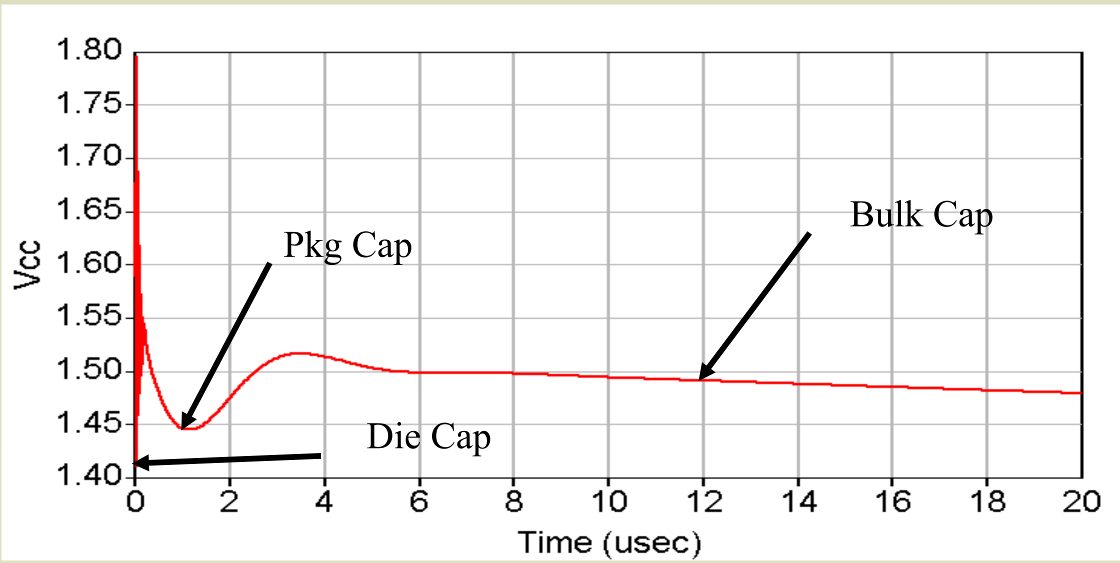
2000-2019 20 YEARS ANNIVERSARY

Power Delivery Impedance



Circuit Representation of a Typical power delivery network

Power Delivery – Noise Containment



Typical System Response to Sudden Current Demand by the Processor
Showing Time Constants for die, package and board time constants

Droops Are Generally Different in Test and Actual Application Environments

Agenda

- Power and Performance Requirement Trends for the Next 5 Years
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

Impact of Power Supply Noise on Power and Performance

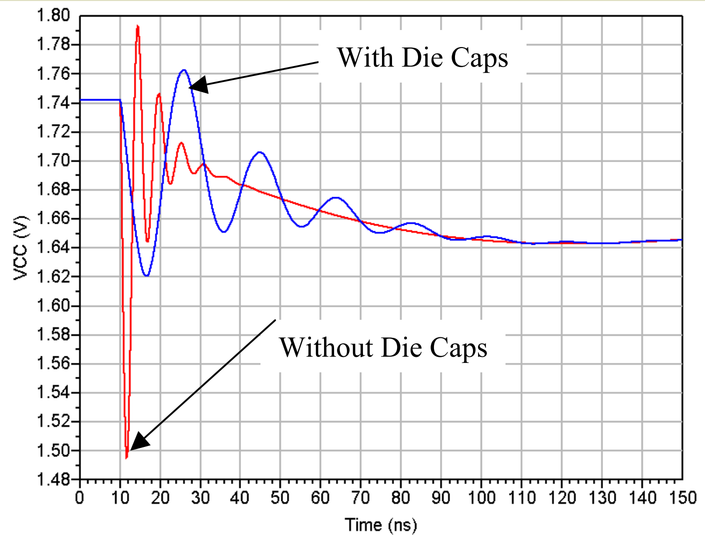
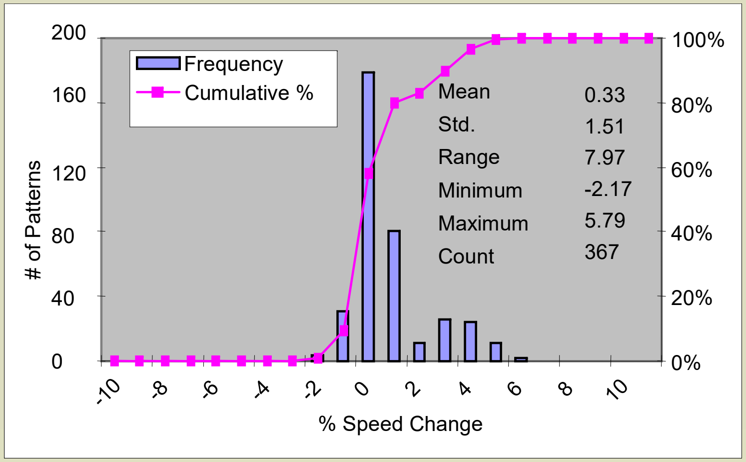
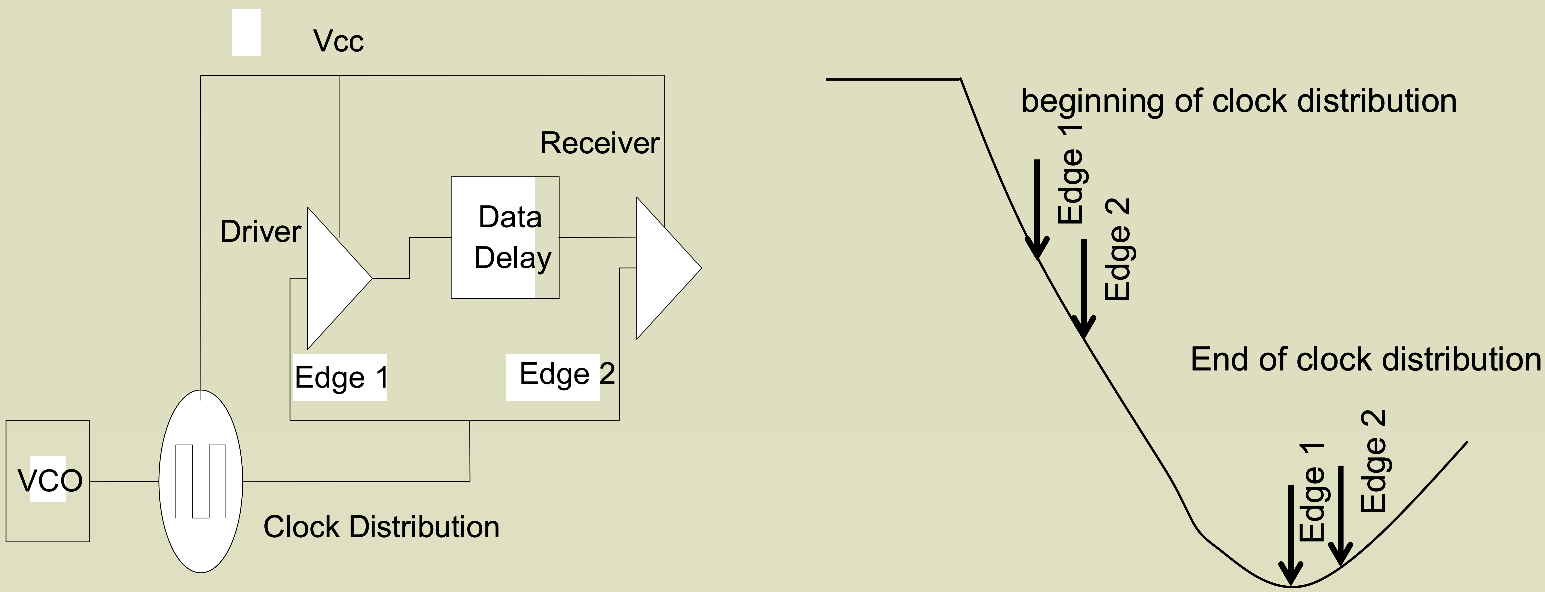


Figure 10: Simulated first droop with and without die caps

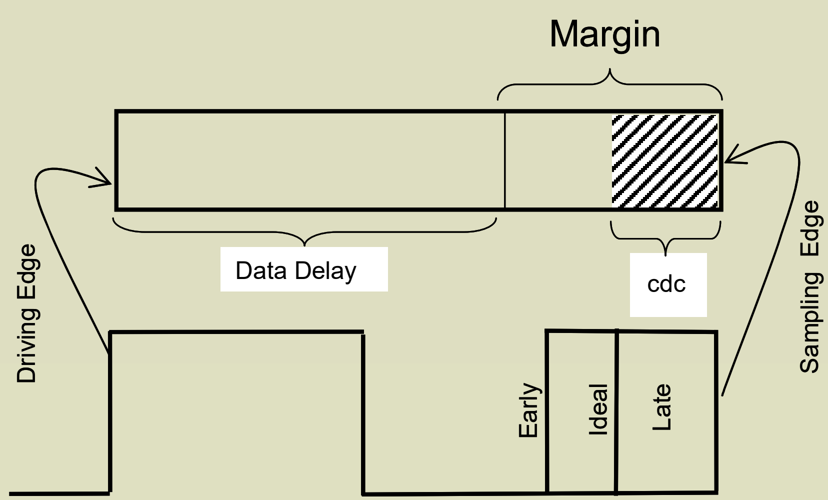


Speed change for individual due to on die cap removal

Architectural Mitigations to Power Supply Noise



Architectural Mitigations to Power Supply Noise – Clock / Data Compensation



Margin = Ideal cycle time – data delay + cdc

Agenda

- Power and Performance Requirement Trends for the Next 5 Years
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

Thermal Challenges / Temperature Monitoring and Control Accuracy

$$T_j = T_a + \theta_{ja} P$$

Power Removal
Cost

- Goal of Thermal Solution is to Guarantee Junction Temperature Does not Exceed Certain Limits to Maintain Reliability Goals
- Die Temperature is Dynamically Read and Managed by the Processor to Keep it Below Tj Max at All Times
- Temperature Inaccuracy Readings Means Guard banding Temperature Reading Leading to Wasted Power and Lower Performance

Thermal Challenges / Temperature Monitoring and Control Accuracy

$$T_j = T_a + \theta_{ja} P$$

Power Removal Cost

- A Key to More Accurate Temperature Reading is Validation and Calibration in Post-Silicon - Industry Should Aim for Temperature accuracy of +/- 1°C, Especially in Tj Constrained Systems
- IREM imaging is the industry standard for power imaging of CMOS chips (both for fault detection and thermal density). However, there is no direct method for a similar imaging of a 3D stacked die
- New Methods for Temperature and Power Maps Reading on Multi-Stacked Die with High Accuracy Will be Required

Agenda

- Power and Performance Requirement Trends for the Next 5 Years
- Modeling and Simulation Challenges
- Power Delivery Challenges
  - Impact of Power Delivery Quality on Performance & Mitigation Methods
- Thermal Challenges
- Summary and Conclusions

Summary

- ❑ Running OS-based workloads (for power, performance, and power management) in pre-silicon environment, simulation and emulation
  - Emulation speed is less than 1MHZ and RTL simulators are slower – Need 1000x speedup to run OS
  - Need proxies that very accurately represent actual benchmarks
- ❑ Power Delivery, Droops Measurements and Management
  - Need to ensure test environment produces same droops as intended system – Lowering Guard band Without Impacting Quality
  - Need Droop Containment Methods as Well as Architectural Droop Mitigation Methods

Summary

- ❑ Power Efficiency
  - Many components (SSD, EC, WIFI, SOC, etc.) expected to consume few milli-Watts and in some cases sub-mW
  - Measuring sub milliWatts Power in a Modern-Standby Sessions, in Validation and HVM Testing

- ❑ Temperature and Voltage Control
  - Temperature accuracy is very key to performance, especially in burst mode; need accurate monitoring and control in a temperature range From -40C to +125C