IDF15
INTEL DEVELOPER FORUM

Intel® Architecture, Code Name Skylake Deep Dive: A New Architecture to Manage Power Performance and Energy Efficiency

Efraim Rotem

Senior Principal Engineer, Lead Client Power Architect, Intel Corporation

ARCS001

intel
experience
what's inside™

Intel® Architecture, Code Name Skylake: Energy Efficiency

- Maximize user experience within system constraints
- User experience:
  - Throughput performance
  - Responsiveness
  - Usage
- System constraints
  - Power, Thermal, Energy
  - Form Factor innovation

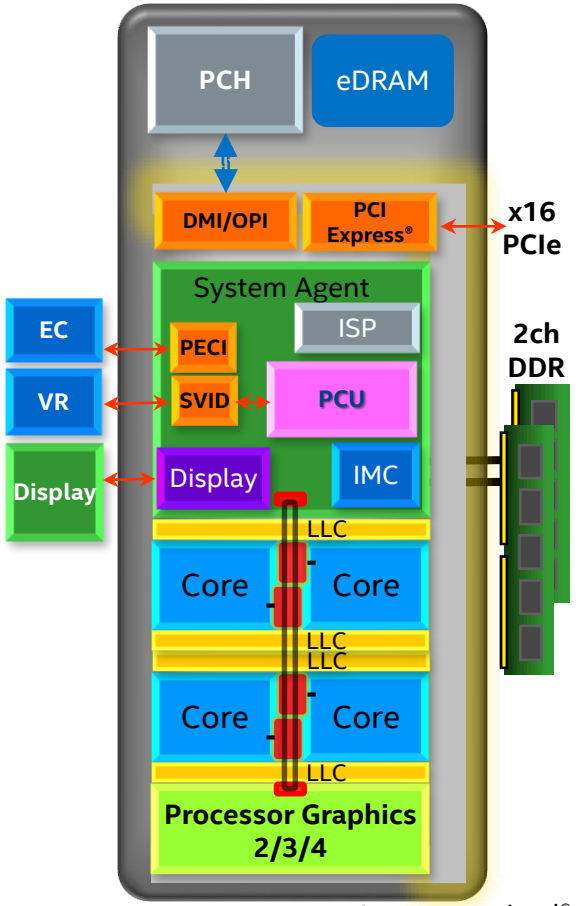
Agenda

- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

Agenda

- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

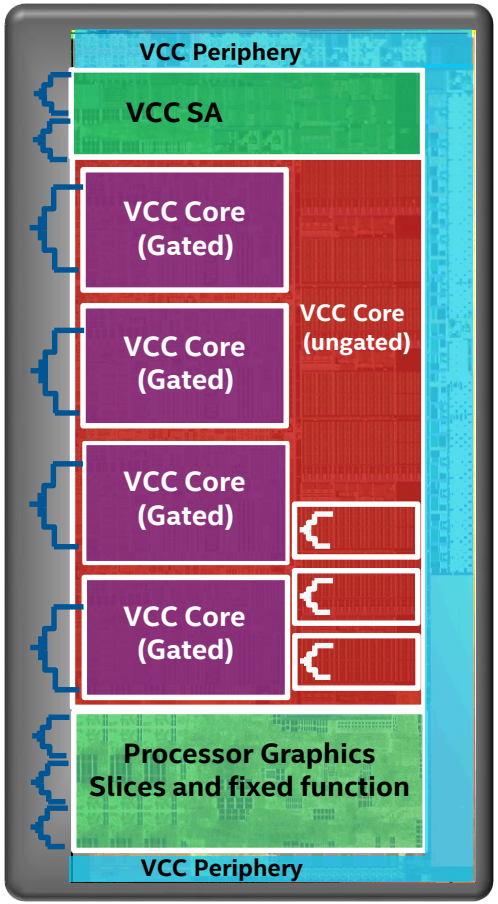
Skylake Overview – Power Management View



- Skylake is a SoC consisting of:
  - 2-4 CPU cores, Graphics, media, Ring interconnect , cache
  - Integrated System Agent (SA)
  - On package PCH and eDRAM
- Improved performance with aggressive power savings
- Package Control Unit (PCU) :
  - Power management logic and controller firmware
  - Continues tracking of internal statistics
  - Collects internal and external power telemetry: iMon, Psys
  - Interface to higher power management hierarchies: OS, BIOS, EC, graphics driver, DPTF, etc.

Note: Not to scale

Skylake Power Management ID Card



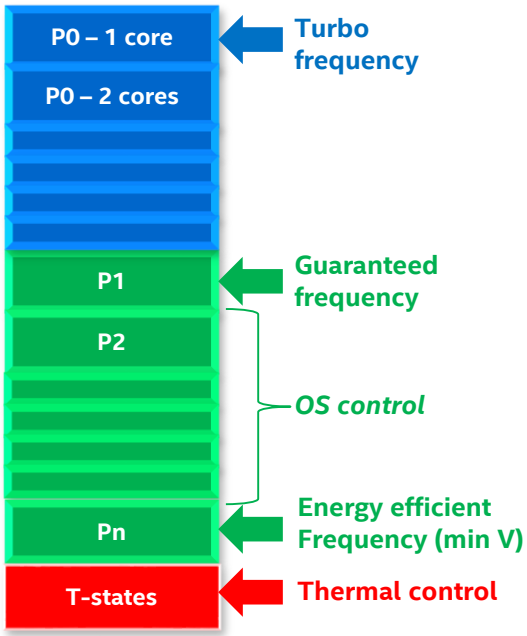
Note: Not to scale

- Up to four independent variable Power domains:
  - CPU cores & ring, PG slice, PG logic and SA
- Other fixed SoC and PCH voltage rails
- High granularity power gating
  - Partial and full core gating, Sub slice Graphics gating, System agent, cache, ring and package power off
- Shared frequency for all Intel® Architecture cores
- Independent frequencies for ring, PG slice & logic
- SA GV for improved performance and battery life

Agenda

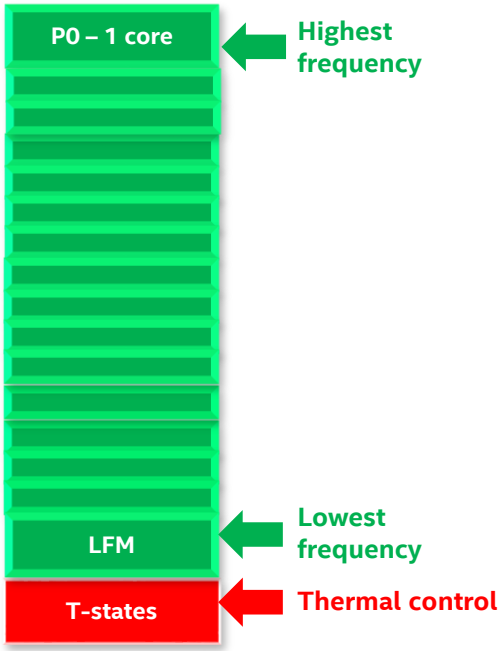
- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

Legacy Energy-performance Control (P-state)



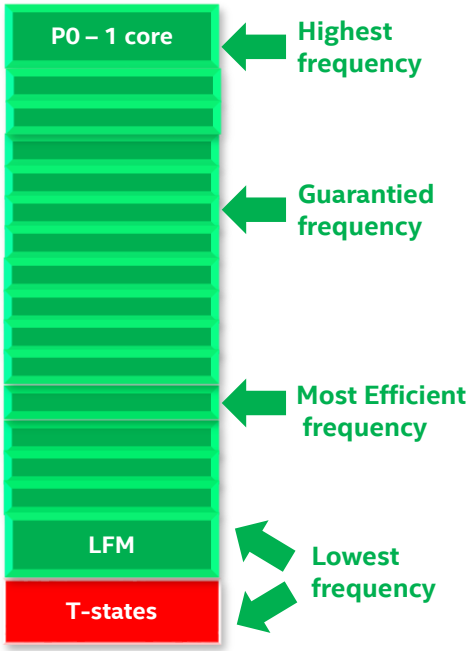
- DVFS – Intel SpeedStep® Technology
  - $P \sim V^2 \cdot f \cdot C_{dynn} + leakage(V) \sim f^3$
  - Performance comes at a cost of energy
- Operating System performs P-state control
  - P1-Pn frequency table enumerated via ACPI tables
  - Explicit P-state selection
- Typically demand based algorithm
  - Policies (AC/DC/Balanced, etc.)
  - Non regular workloads are hard to manage
  - Lower than Pn is used for critical conditions only

Intel® Speed Shift Technology - Hardware P-state



- Why change:
  - Highly dynamic power – Multi core, AVX, accelerators
  - Small form factors → large turbo range
  - Smarter power management enables better choices
    - Finer grain and micro architectural observability
- How:
  - Expose entire frequency range
  - A new deal - OS and hardware share power/perf. control
    - OS direct control when and where desired
    - Autonomous control by PCU elsewhere

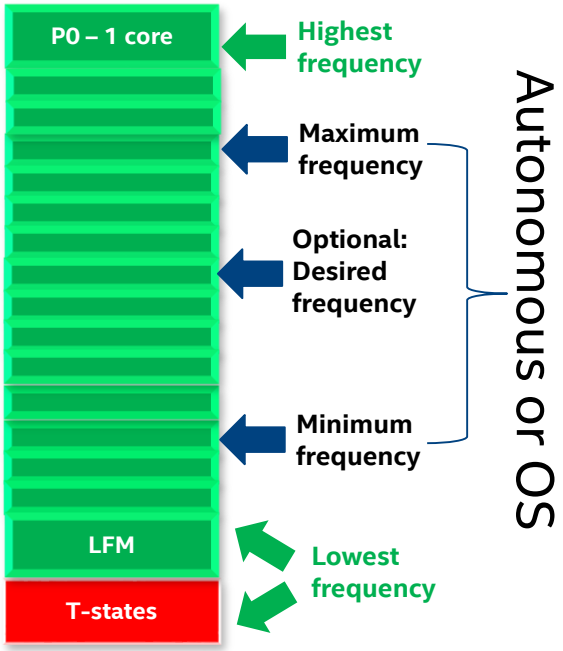
Intel® Speed Shift Technology - Enumeration



- CPU ID and MSR IA32_HWP_STATUS
- Highest frequency[1] – up to P0-1 core
  - Controlled by OEM: Turbo ratio Limit MSR and OC
- Guarantied frequency – resembles legacy P1
  - Controlled by configurable TDP, etc.
- Most efficient frequency (Pe) – calculated at run time
  - A function of system and workload characteristics
- Lowest frequency - in Skylake is set for 100MHz

[1]*Note: actual semantics is performance. Frequency used in Skylake is for convenience only.*

Intel® Speed Shift Technology - Control

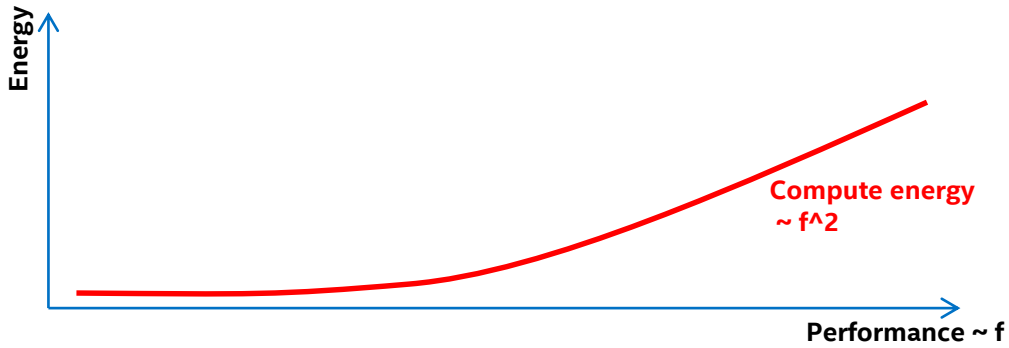


- OS control via IA32_HWP_REQUEST MSR
- Minimum – QoS request, Maximum – Upper limit
- Skylake implements fully Autonomous P-state
  - Demand Based algorithm with responsiveness detection
- Desired – “Soft” request can be overwritten
  - Desired == 0, Full range Autonomous
  - Desired != 0, Autonomous disable
- EPP – Energy Performance Preference
  - OS directive on energy efficiency preference
- Intel® Speed Shift technology OS enabling
  - Work in progress

Agenda

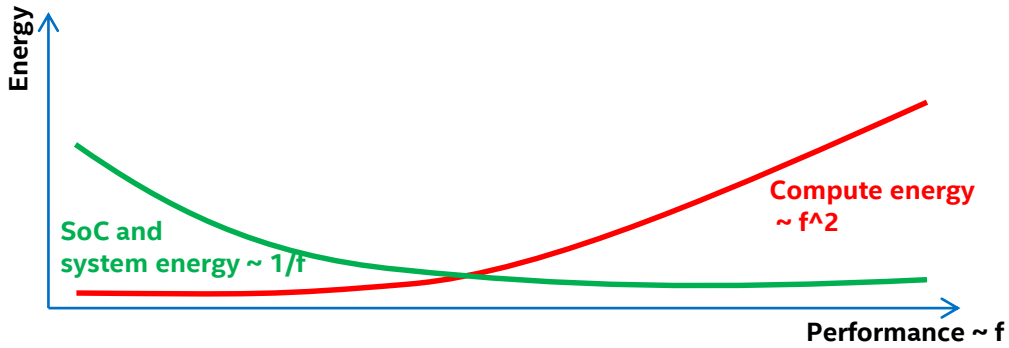
- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

System Active Energy Efficiency



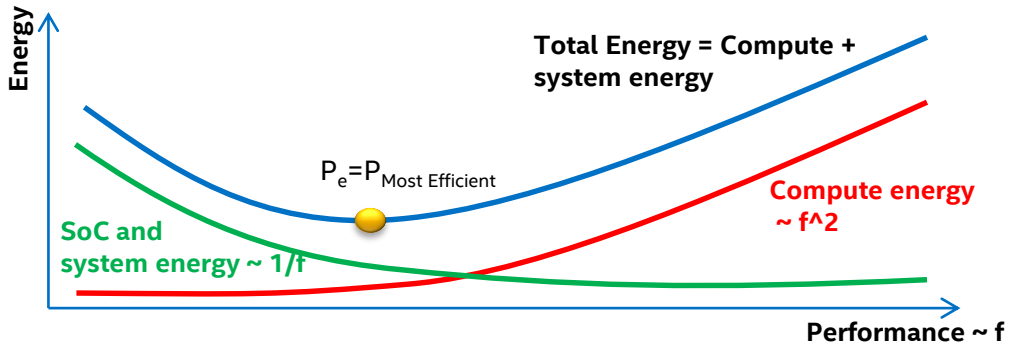
- Compute power $P \sim f^3$,Runtime $T \sim 1/f$ x→ Energy $E \sim f^2$
- Running fast and closing system power → Energy~1/f
- Total energy therefore have a global minimum → P_e
  - Function of System to compute power and workload characteristics
  - System configuration is set by BIOS

System Active Energy Efficiency



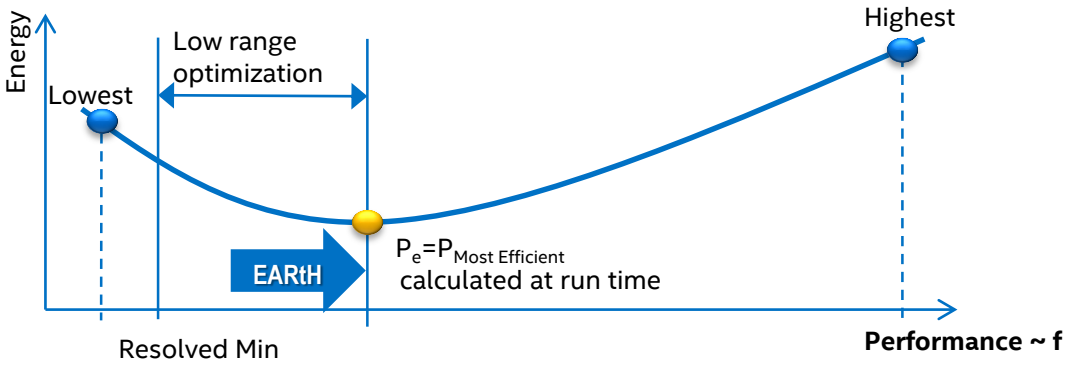
- Compute power $P \sim f^3$,Runtime $T \sim 1/f$ x→ Energy $E \sim f^2$
- Running fast and closing system power → Energy~1/f
- Total energy therefore have a global minimum → P_e
  - Function of System to compute power and workload characteristics
  - System configuration is set by BIOS

System Active Energy Efficiency



- Compute power $P \sim f^3$,Runtime $T \sim 1/f$ x→ Energy $E \sim f^2$
- Running fast and closing system power → Energy~1/f
- Total energy therefore have a global minimum → P_e
  - Function of System to compute power and workload characteristics
  - System configuration is set by BIOS

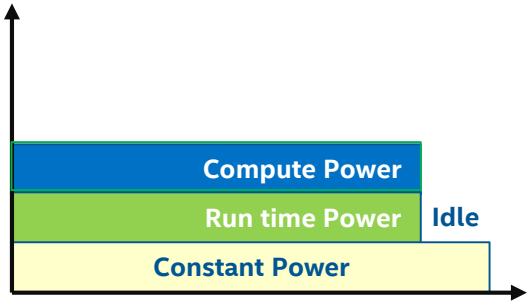
Autonomous Algorithms – Low Range



- At low compute demand, frequency is lowered to conserve energy
- No benefit is running lower than P_e and loose energy
  - Energy Aware Race to Halt (EARtH) unless critical power saving is needed
- Autonomous EARtH[1] algorithm overrides low P-request – run at P_e
  - P_e is calculated every mSec based on workload and system characteristics
  - EARtH only if possible to enter package sleep state (**Consumer Producer**)

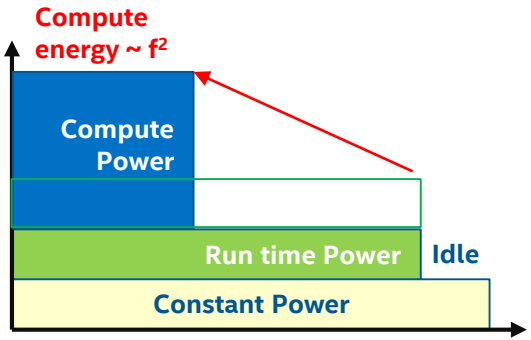
[1] *E. Rotem, R. Ginosar, U. C. Weiser and A. Mendelson, "Energy Aware Race to Halt: A Down to EARtH Approach for Platform Energy Management," IEEE Computer Architecture Letters, vol. 99*

Autonomous Algorithms – Consumer Producer



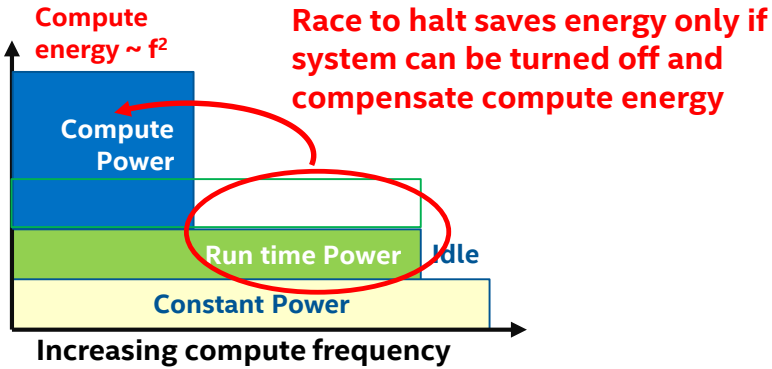
- Increased compute voltage and frequency comes at an energy cost
  - If another component prohibits idle state – there is no reason to run faster

Autonomous Algorithms – Consumer Producer



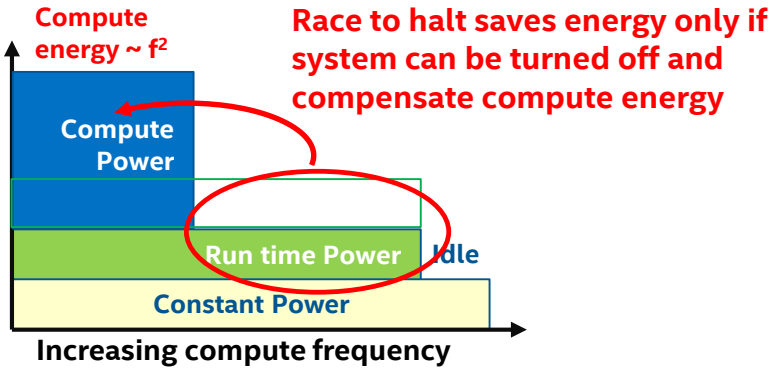
- Increased compute voltage and frequency comes at an energy cost
  - If another component prohibits idle state – there is no reason to run faster

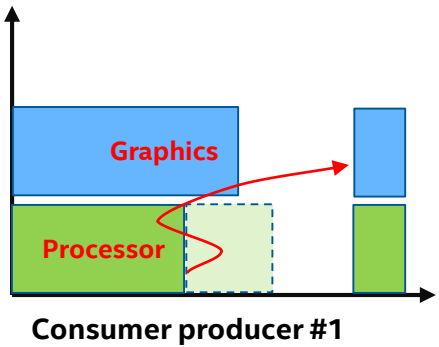
Autonomous Algorithms – Consumer Producer

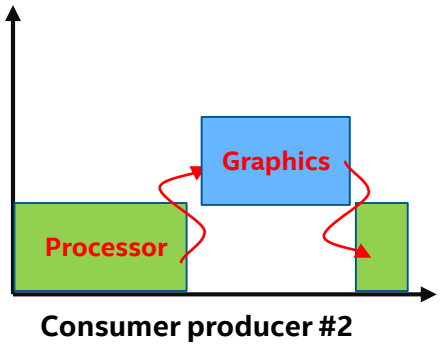


- Increased compute voltage and frequency comes at an energy cost
  - If another component prohibits idle state – there is no reason to run faster

Autonomous Algorithms – Consumer Producer

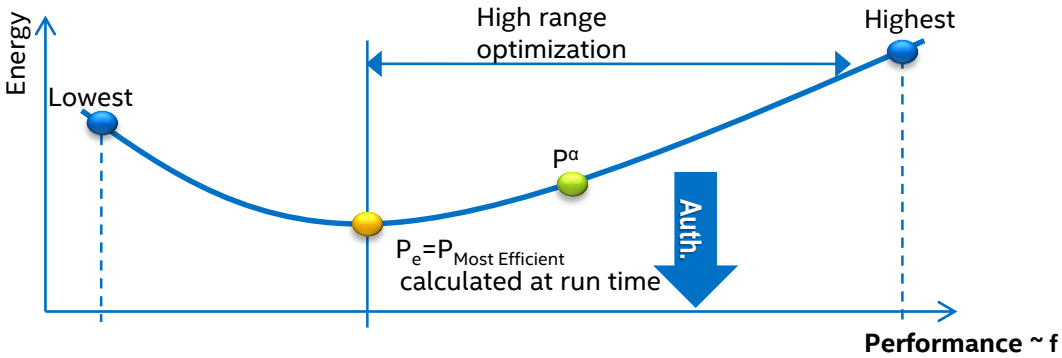






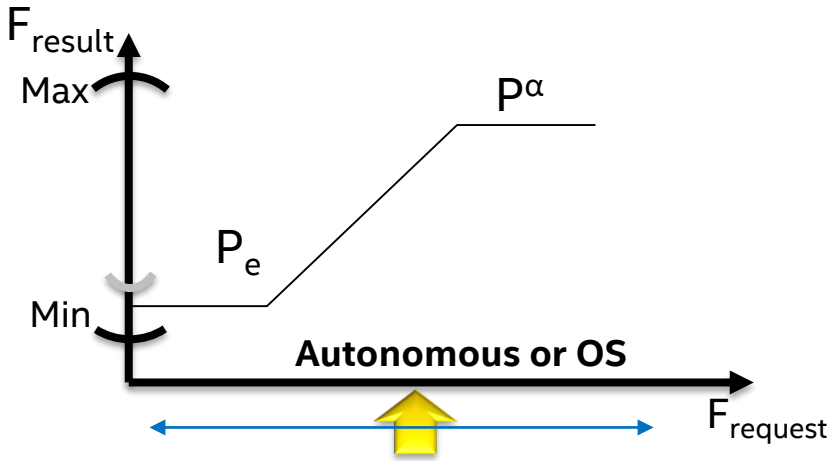
- Increased compute voltage and frequency comes at an energy cost
  - If another component prohibits idle state – there is no reason to run faster
- An example - consumer producer
  - Example #1 – if the processor runs in the “shadow” of the graphics , system will not idle
  - Example #2 – Demand based algorithms will make wrong P-state request
- Autonomous algorithms detect these two profiles

Autonomous Algorithms – High Range



- Performance comes at an increased energy cost
  - Preference (α) allows limiting the energy cost by limiting frequency (P^{α})
    - Semantics: frequency that meets ΔPower/Δperformance ≤α
    - Controlled by an OS and user preference (e.g. max performance, balanced)
    - Function of workload characteristics (power and scalability)

Frequency Resolving



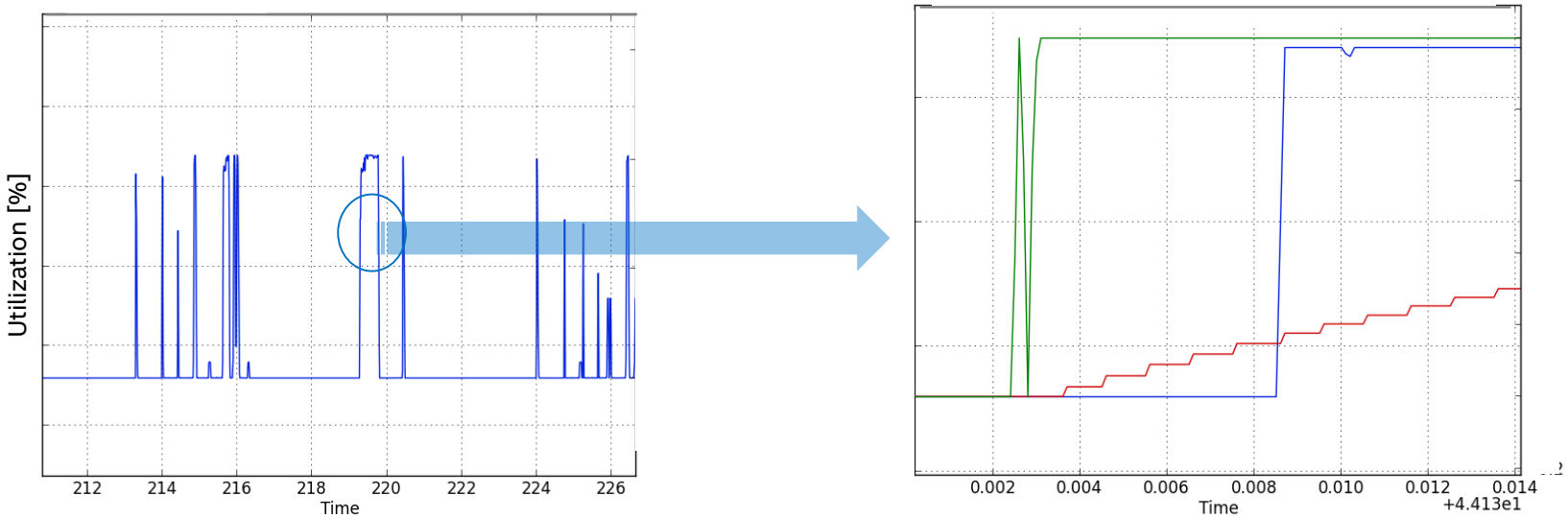
- Either desired request or autonomous set frequency demand
- Shaped by energy efficiency algorithms P_e and P^{α}
- OS Min and Max act as “brackets”
- Physical power or thermal condition can override performance control down

Agenda

- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

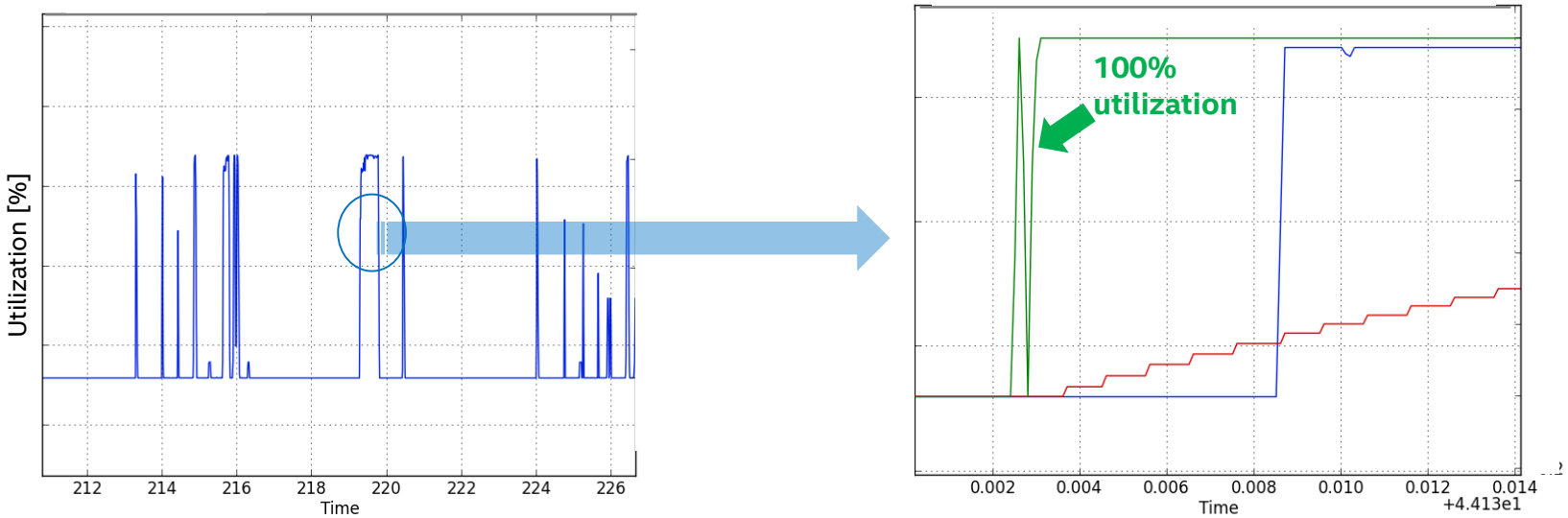
Responsiveness

- Fast burst response while performing interactive work
  - Filter out short interrupts and repeated work such as Video playback
  - Filter cyclic workloads e.g. video playback



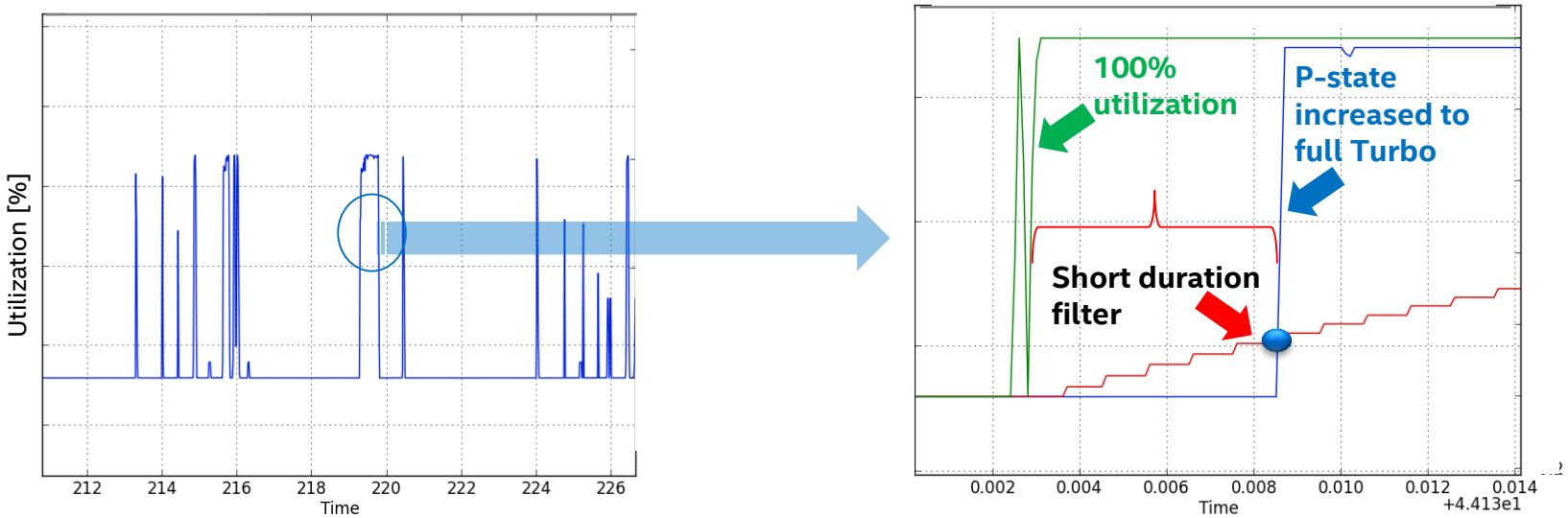
Responsiveness

- Fast burst response while performing interactive work
  - Filter out short interrupts and repeated work such as Video playback
  - Filter cyclic workloads e.g. video playback

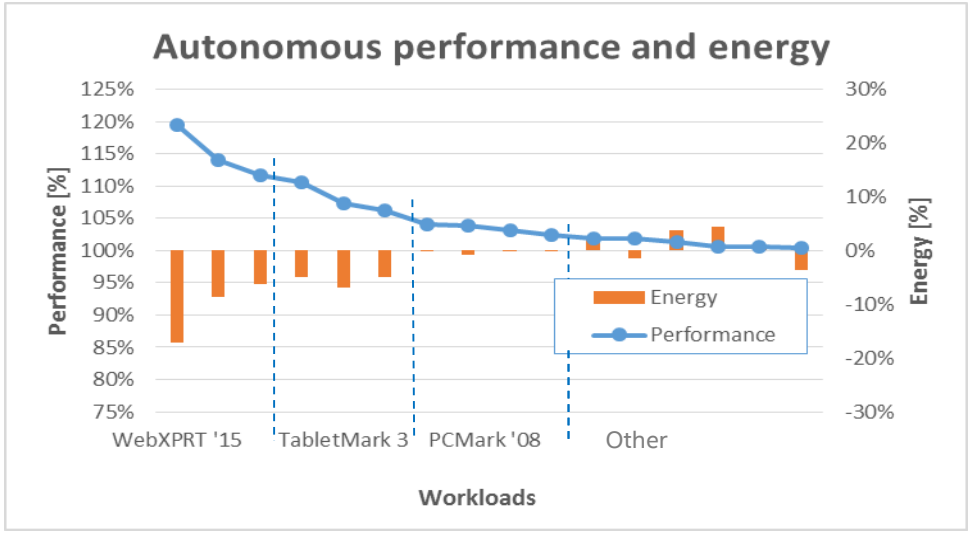


Responsiveness

- Fast burst response while performing interactive work
  - Filter out short interrupts and repeated work such as Video playback
  - Filter cyclic workloads e.g. video playback



Autonomous Algorithm Benefits vs. Legacy



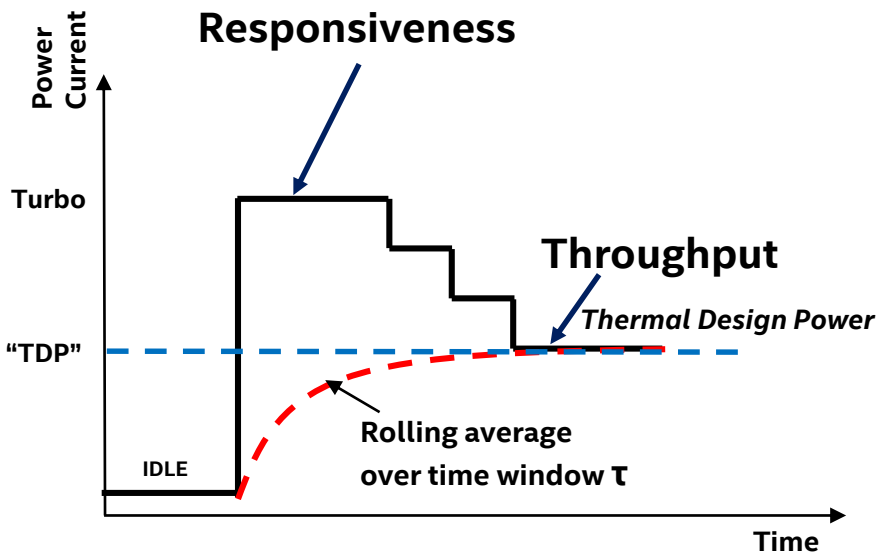
Note: Example only, tested on limited number of workloads and products.

- Energy and performance benefits responsiveness and consumer producer
- Improved performance → shorter run time → better energy
  - Also improves platform energy
- Lower energy → cooler system

Agenda

- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

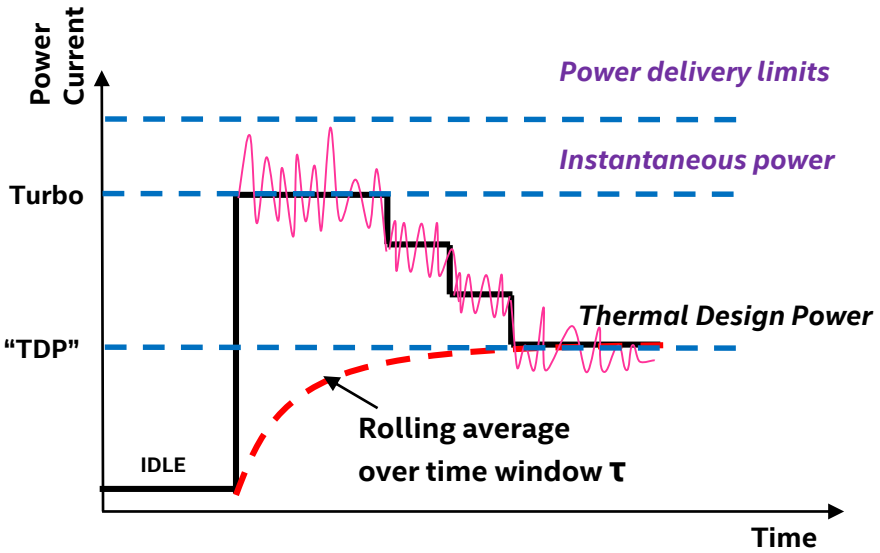
Intel® Turbo Boost Technology 2.0



- Maximize user experience within system constraints[1]
- User experience:
  - Throughput performance
  - Responsiveness

[1]*E. Rotem, A. Naveh, A. Ananthakrishnan, E. Weissmann, and D. Rajwan, "Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge," IEEE Micro, vol. 32, no. 2, pp. 20-27, March-April 2012*

Intel® Turbo Boost Technology 2.0

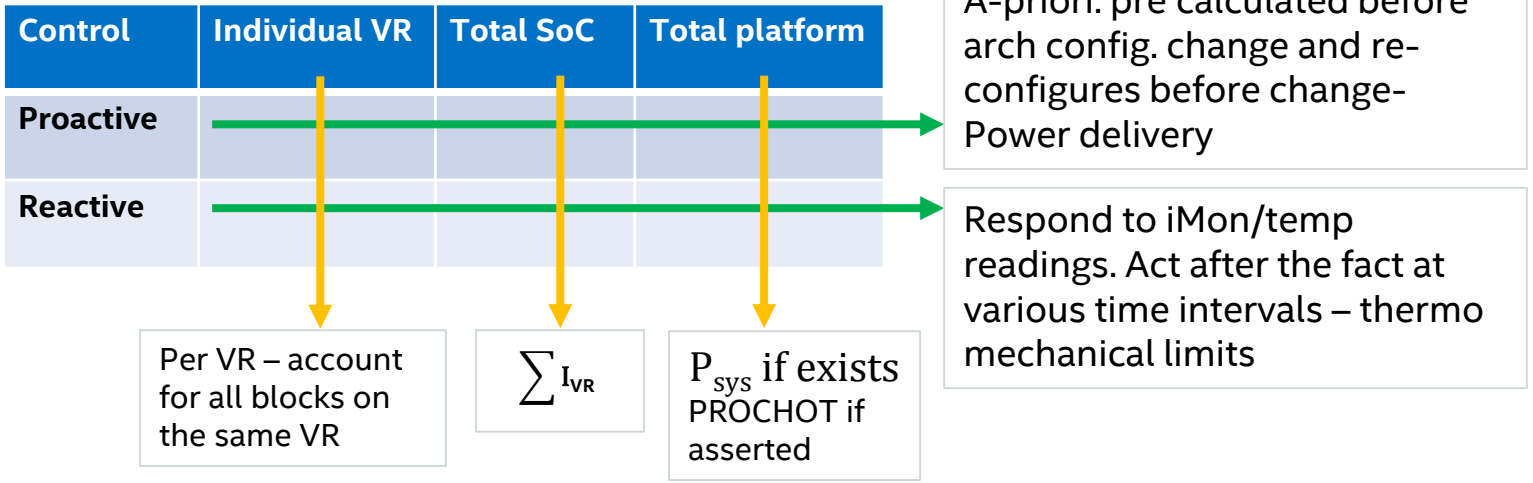


- Maximize user experience within system constraints[1]
- User experience:
  - Throughput performance
  - Responsiveness
- System constraints
  - Power, Thermal, Energy
  - Power delivery – wall outlet to die
  - Form Factor

[1] *E. Rotem, A. Naveh, A. Ananthakrishnan, E. Weissmann, and D. Rajwan, "Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge," IEEE Micro, vol. 32, no. 2, pp. 20-27, March-April 2012*

Power Control

Taxonomy of controls:

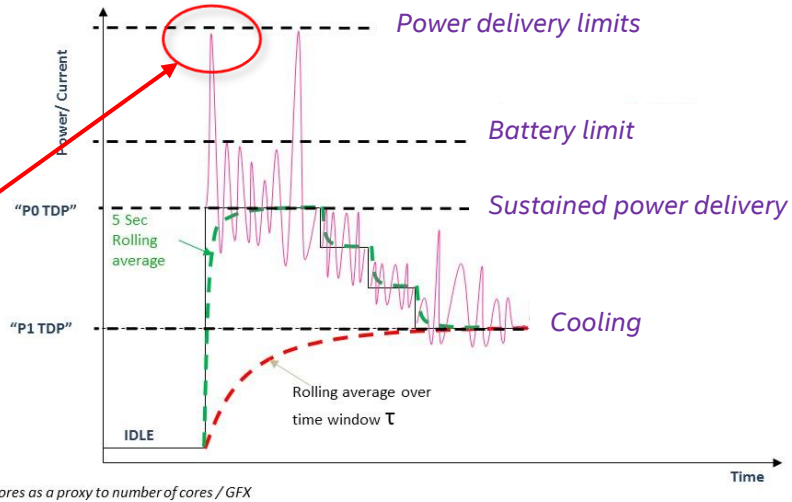


Power Control Capabilities

Control	Individual VR	Total SoC	Total platform
Proactive	$\sum_{domains} I_{Virus}$	$\sum_{All\_VRs} P_{virus}$ (PL4)	N/A
Reactive	VR sustained: single current controller (PID)	PL1, PL2 PL3	P_sys PPL1, PPL2 PPL3

Note PL = Power Limit

Proactive control Features
Configurable VR Topology (described to the PCU)
Dynamic power range management
Safe workload calculation before C/P-state change
Total SoC (for battery & Brick) max Icc control
VR Max Icc budgeting (Intel® Architecture vs. processor graphics



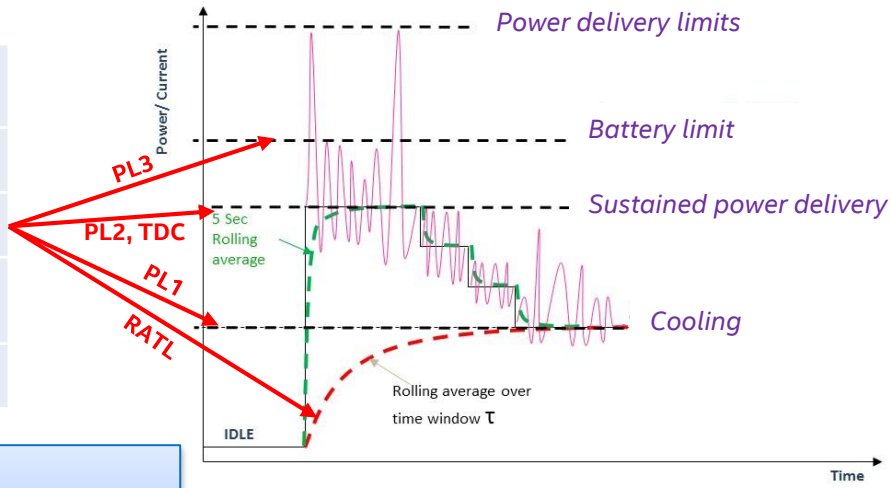
**4 cores as a proxy to number of cores / GFX*

Max instantaneous current limit cannot be violated ever. Therefore the SoC will not run at a configuration that may, even rarely, violate that limit.

Power Control Capabilities

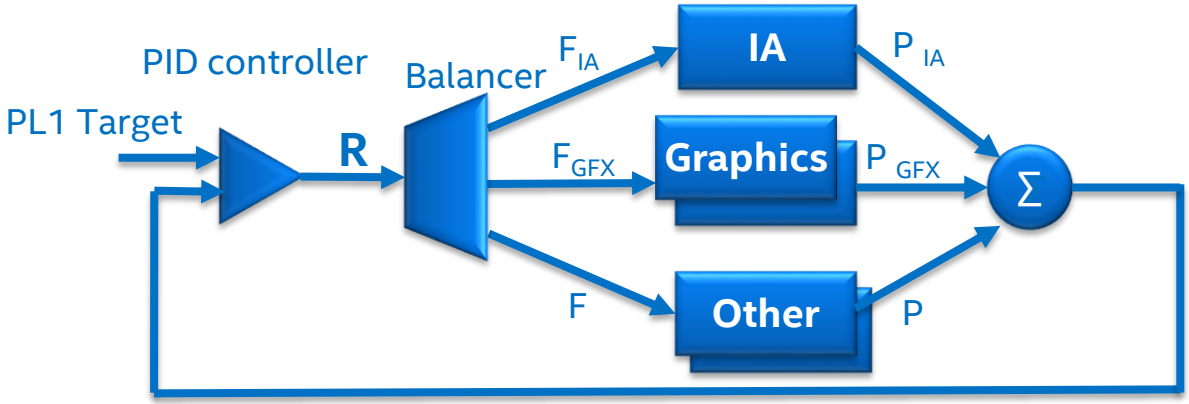
Control	Individual VR	Total SoC	Total platform
Proactive	$\sum_{domains} I_{Virus}$	$\sum_{All\_VRs} P_{virus}$ (PL4)	N/A
Reactive	VR sustained: single current controller (PID)	PL1, PL2 PL3	P_sys PPL1, PPL2 PPL3

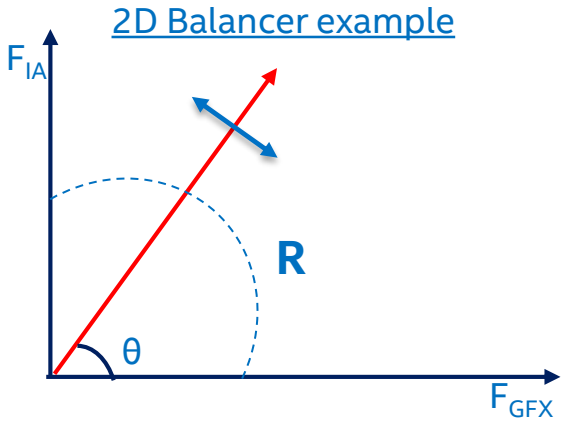
Reactive control Features
Battery protection - PL3
Voltage Regulator sustained current - PL2
Dynamic and average temperature response RATL – Rolling Average Temperature Limit
SoC and enclosure temperature - PL1



Attention given in SkyLake to the questions:
1. How to share the power budget between consumers
2. Optimize low power form factors

Balancer – Workload Aware Control



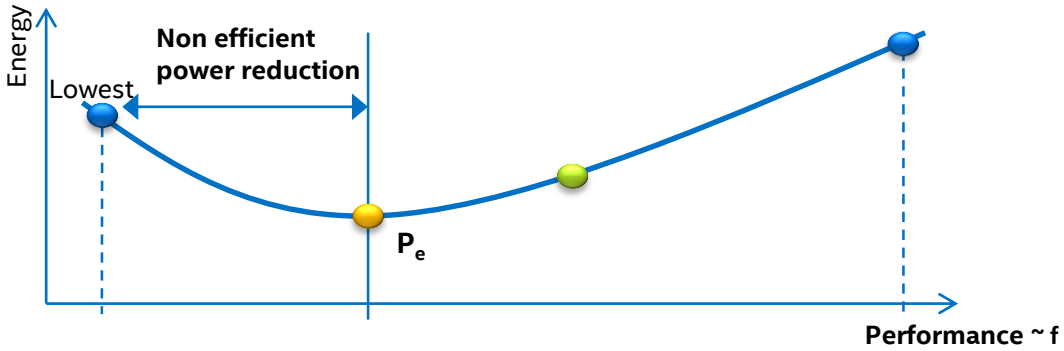


- Keeping balanced state under constraints:
  - Workload characteristics set compute and data transfer demand
  - Controlling the power while keeping the right balance

Agenda

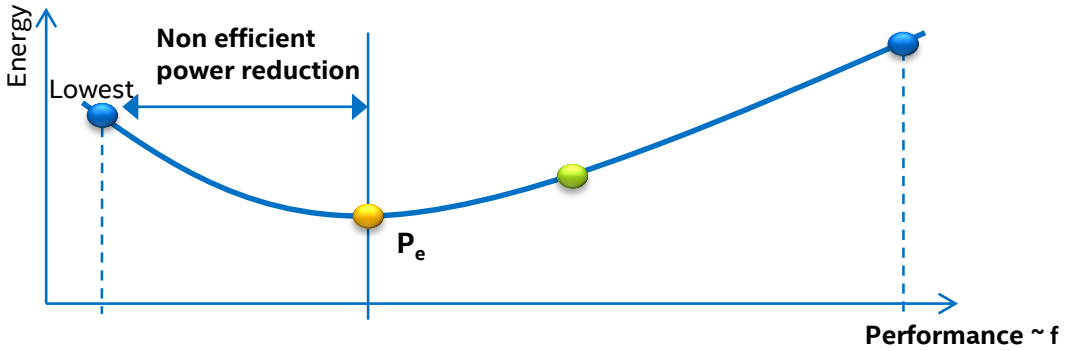
- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

SkyLake SoC Duty Cycling



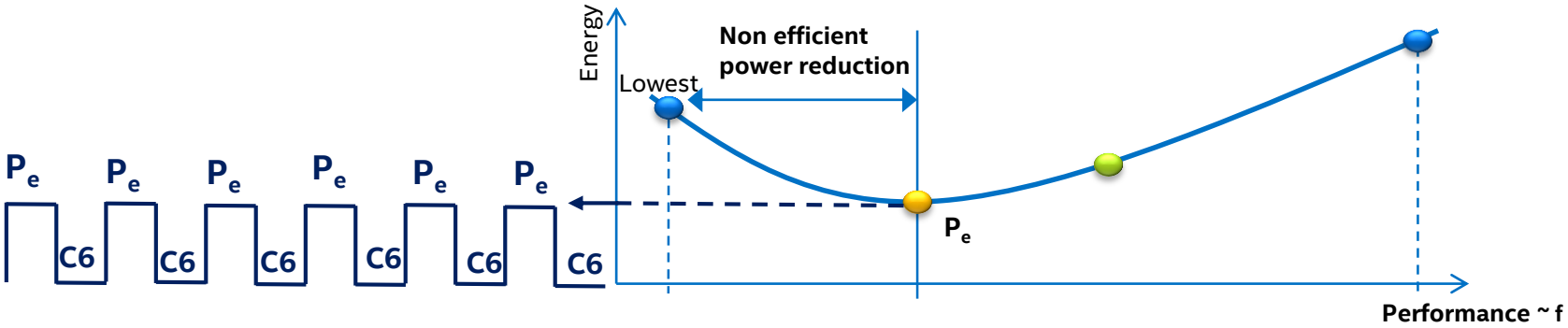
- At small form factors – it may be needed to significantly reduce power
- Going below P_e reduces power inefficiently – fixed power components

SkyLake SoC Duty Cycling



- At small form factors – it may be needed to significantly reduce power
- Going below P_e reduces power inefficiently – fixed power components
- Running at Pe is the energy most efficient point
- Turning package on and off reduces power proportionally to run time
  - Providing power savings at ~no energy cost
- Intel® Architecture forced sync. Idle – Graphics duty cycle on frame boundaries

SkyLake SoC Duty Cycling



- At small form factors – it may be needed to significantly reduce power
- Going below P_e reduces power inefficiently – fixed power components
- Running at Pe is the energy most efficient point
- Turning package on and off reduces power proportionally to run time
  - Providing power savings at ~no energy cost
- Intel® Architecture forced sync. Idle – Graphics duty cycle on frame boundaries

Agenda

- Overview – Power Management View
- Intel® Speed Shift Technology
  - Autonomous Algorithms
  - User Interaction and Accelerating Responsiveness
- Managing Physical Constraints
  - SoC Duty Cycling
- Summary and Conclusions

Summary and Next Steps

- Intel® Architecture code name SkyLake is built for best user experience in a form factor
- Intel® Speed Shift Technology: performance, responsiveness and energy
  - Operating system vendors enabling
  - New dimensions of user visible value
- Rich set of controls allow OEM innovation
  - Allow design choices

Other Technical Sessions

Session ID	Title	Day	Time	Room
✓ ARCS001	Intel® Architecture, Code Name Skylake Deep Dive: A New Architecture to Manage Power Performance and Energy Efficiency	Tues	1:15	2006
ARCS002	Software Optimizations Become Simple with Top-Down Analysis Methodology on Intel® Microarchitecture, Code Name Skylake	Tues	2:30	2006
ARCS003	Intel® Architecture Code Name Skylake Deep Dive: Hardware-Based Security for Windows® 10	Tues	4:00	2006
Special	Zoom-in on Your Code with Intel® Processor Trace and Supporting Tools	Tues	5:30	Showcase Networking Plaza
SFTS002	Bringing Energy Efficiency Improvements Through Windows 10 and Intel Architecture Based Platforms	Tues	2:30	2009

✓ = DONE

A PDF of this presentation and the others is available from our Technical Session Catalog: www.intel.com/idfsessionsSF. This URL is also printed on the top of Session Agenda Pages in the Pocket Guide.

Legal Notices and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, 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.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, SpeedStep, and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the second quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as "anticipates," "expects," "intends," "plans," "believes," "seeks," "estimates," "may," "will," "should" and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel's actual results, and variances from Intel's current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be important factors that could cause actual results to differ materially from the company's expectations. Demand for Intel's products is highly variable and could differ from expectations due to factors including changes in business and economic conditions; consumer confidence or income levels; the introduction, availability and market acceptance of Intel's products, products used together with Intel products and competitors' products; competitive and pricing pressures, including actions taken by competitors; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel's gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; and product manufacturing quality/yields. Variations in gross margin may also be caused by the timing of Intel product introductions and related expenses, including marketing expenses, and Intel's ability to respond quickly to technological developments and to introduce new products or incorporate new features into existing products, which may result in restructuring and asset impairment charges. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Results may also be affected by the formal or informal imposition by countries of new or revised export and/or import and doing-business regulations, which could be changed without prior notice. Intel operates in highly competitive industries and its operations have high costs that are either fixed or difficult to reduce in the short term. The amount, timing and execution of Intel's stock repurchase program could be affected by changes in Intel's priorities for the use of cash, such as operational spending, capital spending, acquisitions, and as a result of changes to Intel's cash flows or changes in tax laws. Product defects or errata (deviations from published specifications) may adversely impact our expenses, revenues and reputation. Intel's results could be affected by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel's ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. Intel's results may be affected by the timing of closing of acquisitions, divestitures and other significant transactions. A detailed discussion of these and other factors that could affect Intel's results is included in Intel's SEC filings, including the company's most recent reports on Form 10-Q, Form 10-K and earnings release.

Rev. 4/14/15