

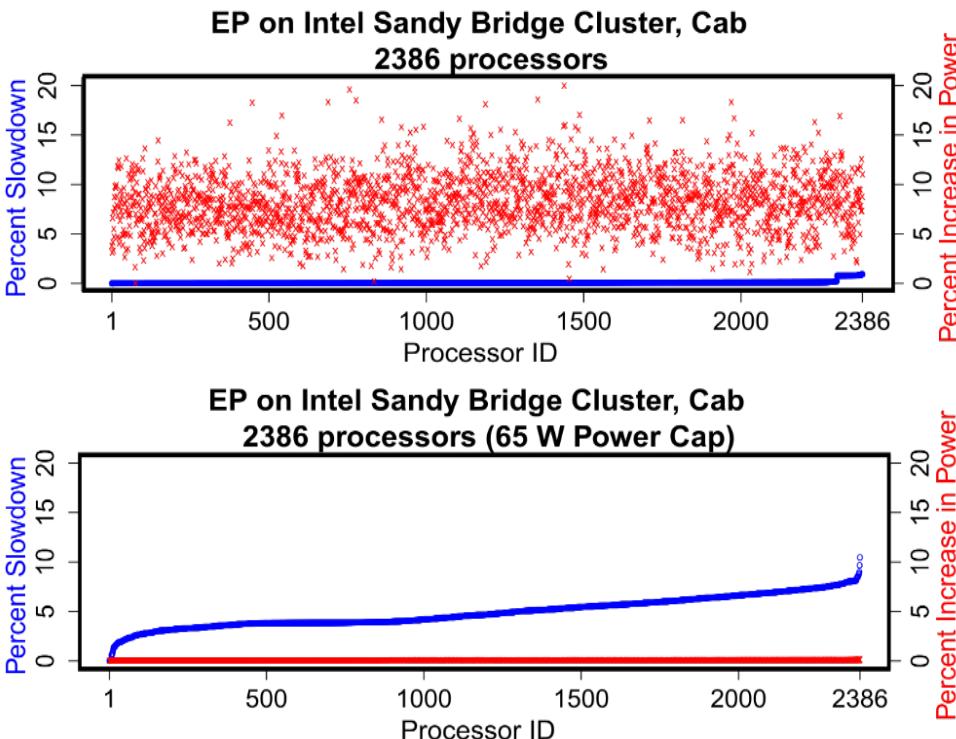
Řízení spotřeby procesoru, Historie procesorů Intel AVS – Architektury výpočetních systémů Týden 12, 2024/2025

Jirka Jaroš

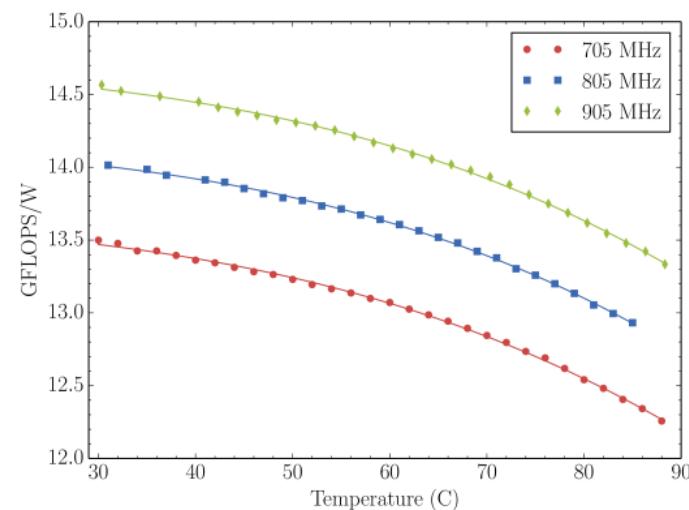
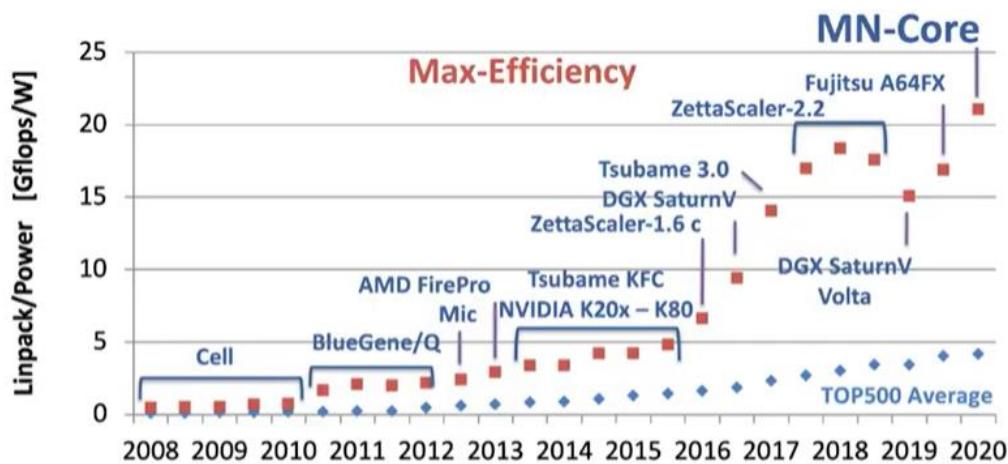
Vysoké učení technické v Brně, Fakulta informačních technologií
Božetěchova 1/2, 612 66 Brno - Královo Pole
jarosjir@fit.vutbr.cz

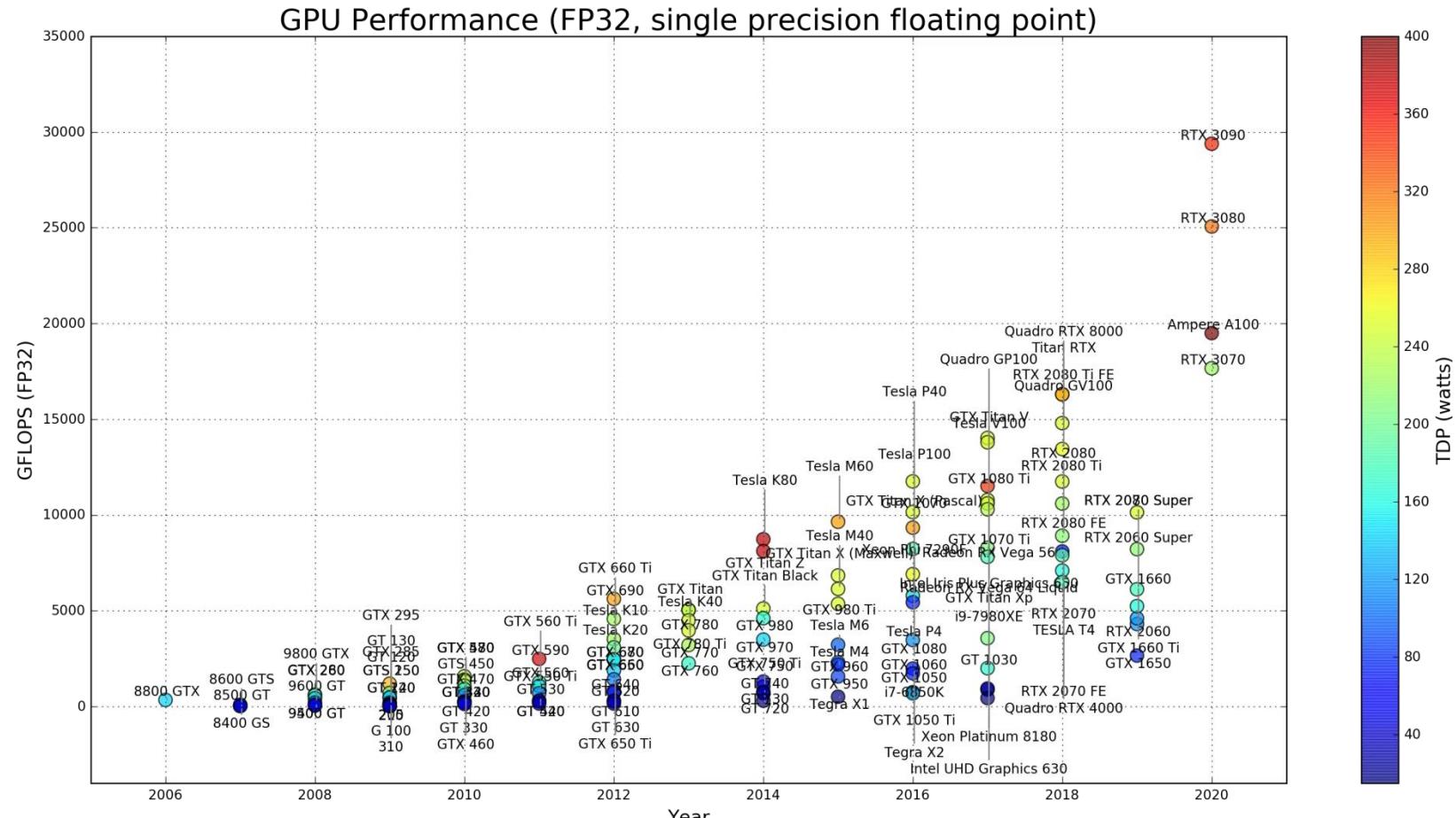


- **Limity chladicího systému**
 - Vzduchové, kapalinové, dusík, ...
 - S teplotou roste příkon
- **Rozvod napájení po čipu**
 - Širší dráty, vyšší vodivost, ...
- **Nestabilita výkonu při fixním příkonu**
 - Díky výrobním tolerancím má každý čip jinou spotřebu při dané frekvenci
- **Omezená kapacita baterií**
 - Pro mobilní zařízení je důležitá výdrž na baterii
 - Výkon bývá na druhém místě
- **Náklady na energii mohou snadno přesáhnout pořizovací cenu HW**
 - Intel Core i9-10900K \approx 125/250W
 - Spotřeba za rok: 1,1 MWh \approx 13.200,- Kč
 - Maximální spotřeba pro ExaScale stroj stanovena na 20 MW, reálně bude 50 MW



- Dynamické řízení příkonu DPM (Dynamic Power Management)
- Dynamické odstupňování napětí DVS (Dynamic Voltage Scaling)
- Spouštění instrukcí z mnoha vláken
 - V jednom cyklu SMT (simultaneous multithreading)
 - Na více jádrech procesoru typu CMP
- Základní metrika MIPS/W (GFlops/W)

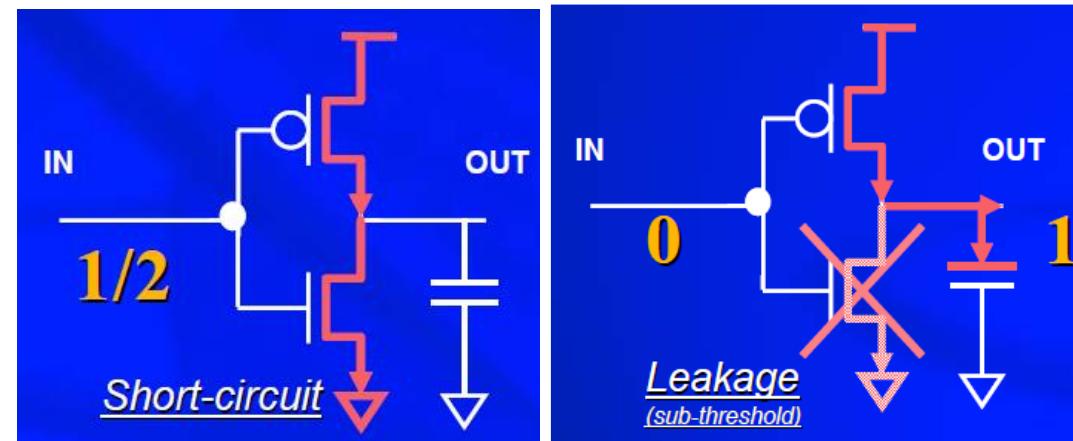
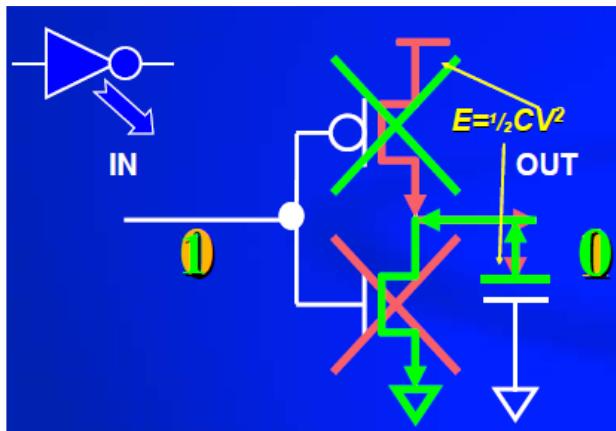




- Návrh na úrovni tranzistorů
 - Redukce energie pro přepnutí ON-OFF, klidový/zkratový proud
- Návrh na úrovni obvodu
 - Snaha o asynchronní řízení, různé frekvence různých bloků
- Návrh na úrovni bloků
 - Vypínání nevytížených bloků (jádra, cache, FX, FP, sběrnice)
- Návrh na úrovni systému
 - Alokaci architektury, mapování aplikací, plánování procesů...
- Na úrovni kompilátoru
 - Optimalizace strojového kódu pro optimální využití zdrojů
 - Poskytnutí informací pro CPU o náročích procesu

- Na tranzistorové úrovni lze formulovat celkový ztrátový výkon jako součet tří hlavních složek

- Přepínací ztráty
- Ztráty zkratovým proudem
- Ztráty klidovým proudem



Přepínací ztráty

$$P_{device} = \frac{1}{2} C \cdot V_{DD} V_{swing} a \cdot f$$

Ztráty klidovým proudem

$$I_{leakage} V_{DD}$$

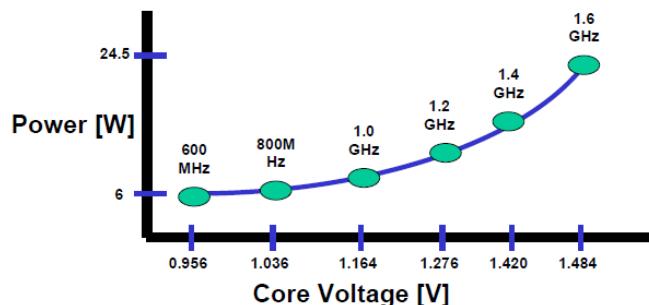
Ztráty zkratovým proudem

$$I_{SC} V_{DD}$$

- C je kapacita na výstupu tranzistoru
- V_{DD} je napájecí napětí
- f je hodinový kmitočet čipu
- a je faktor aktivity ($0 < a < 1$)
- V_{swing} je napěťový rozkmit na výstupní kapacitě
- $I_{leakage}$ je klidový proud
- I_{SC} je zkratový proud

Nejdůležitější metodou redukce příkonu je tedy

- Snižování napájecího napětí
- Snižování frekvence



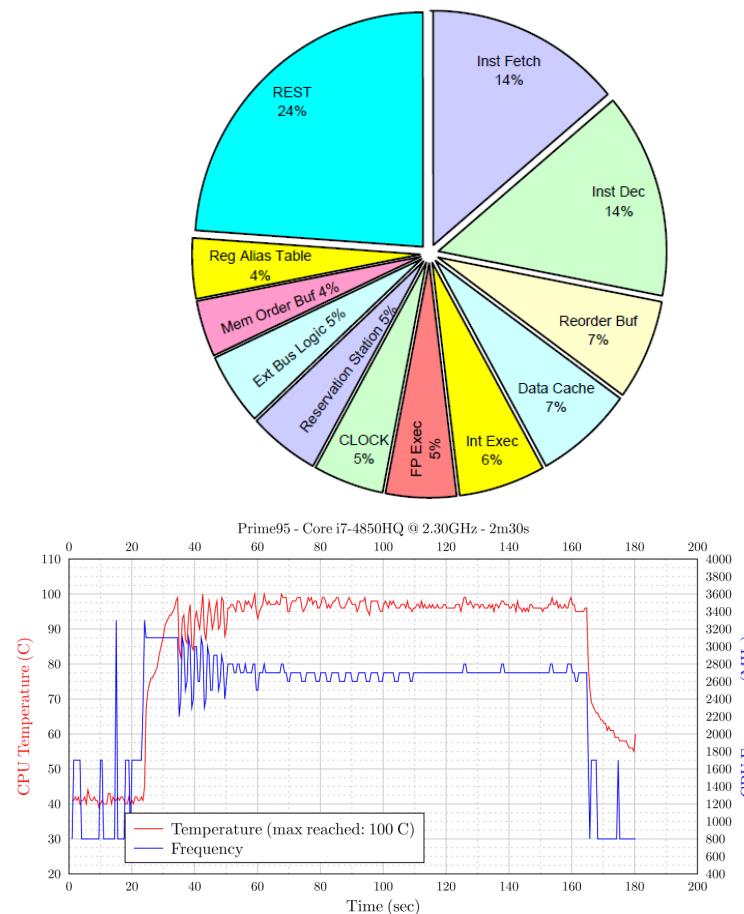
- **Paměti RAM, cache, registrové soubory, tabulky predikce skoků, atd...**

- Paměťové vodiče jsou dlouhé, představují určitou kapacitu a tím pádem spotřebu.
- Cache s vypínáním paralelních cest, tedy se snižováním stupně asociativity.
- Fronty instrukcí a tabulky přeložených adres

- **Filtrovací cache**

- Vychází se z předpokladu že většina přístupů do cache může být obslužena malou filtrovací cache.
- Omezení příkonu za cenu snížení výkonnosti způsobené zdržením přístupu do hlavní cache
- Využívají se např. cache s frontou požadavků s proměnnou délkou, řízenou dynamickou pracovní zátěží.
- Náklady na přídavné tranzistory jsou pod 2 %

- Hradlování umožní pozastavit některé stupně pipeline v případě, že procesor provádí instrukce z nesprávně předpovězené větve skoku.
 - Hradlování pipeline řídí stupeň spekulace superskalárních procesorů, u kterých se používá predikce skoků.
 - Rozhodnutí o pozastavení pipeline se provádí pomocí obvodu **odhadu konfidence**.
- **Zatlumení pipeline (throttling)**
 - Jestliže příkon překročí určitou mez a teplota procesoru se zvýší k nebezpečné hranici, dojde k zatlumení pipeline.
 - Provádí se buď snížením frekvence (pozastavením hodin).
 - Nebo vkládáním prázdných operací.

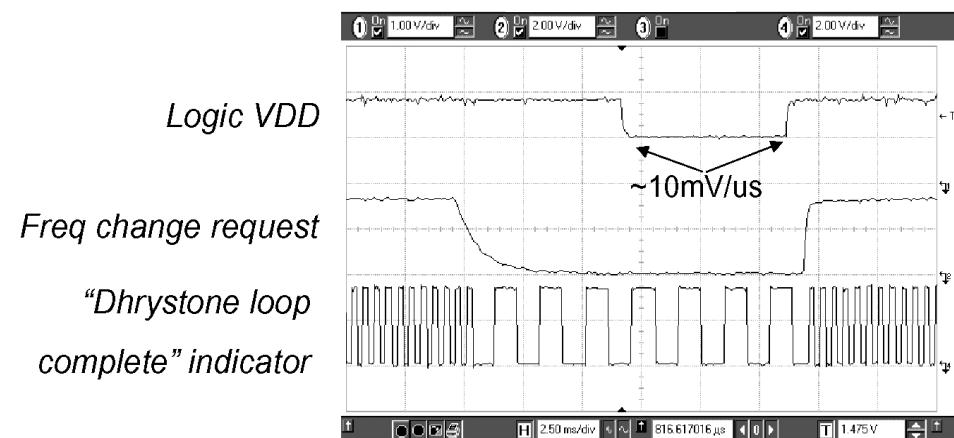
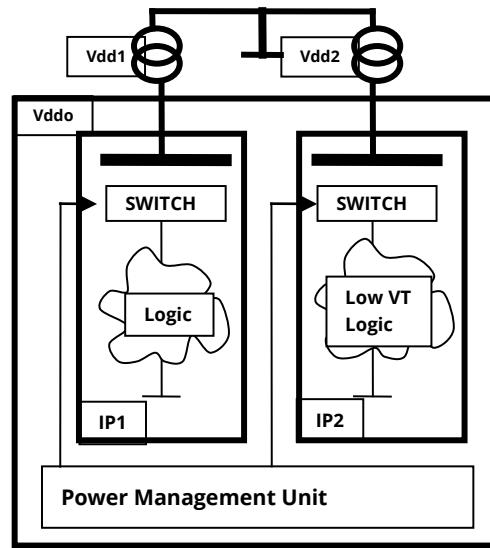


- **Na úrovni bloků**

- Odpojování nepotřebných částí (power gating)
- Snižování frekvence nepotřených jednotek (clock gating, frequency stepping)
- Pokročilé techniky (detekce cyklů, fúzování mikroinstrukcí)

- **Na úrovni systému**

- Zpracování instrukcí z více vláken



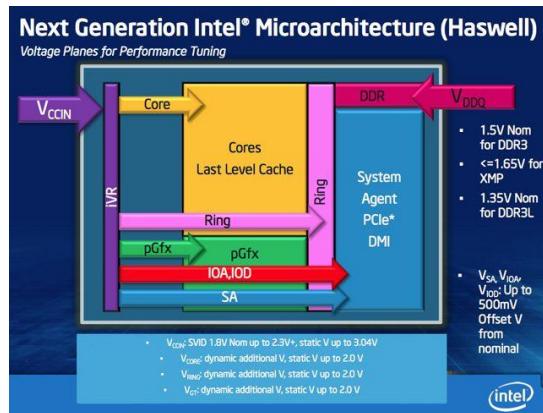
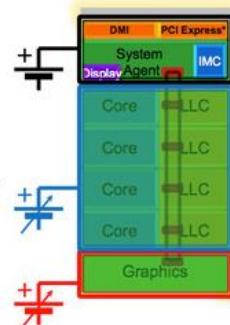
- Jedná se především o optimalizace mezikódu
 - Snaha o zabránění vykonávání zbytečných instrukcí
 - Rozbalování smyček
 - Propagace konstant
 - Odstranění mrtvého kódu
 - Automatická detekce paralelizmu
 - Pokud nelze některé bloky vypnout jednotlivě (např. pouze celá jádra) je vhodné využít všechny jednotky
 - Snižování počtu přístupů do paměti
 - Velice důležitá alokace registrů a plánování instrukcí

PŘÍKLADY TECHNIK POWER MANAGEMENTU

I Sandy Bridge a Haswell – System Agent

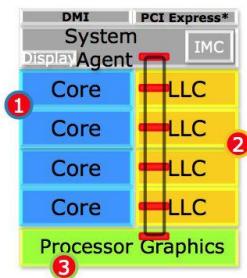
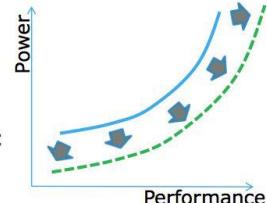
Lean and Mean System Agent

- Contains PCI Express*, DMI, Memory Controller, Display Engine...
- Contains **Power Control Unit**
 - Programmable uController, handles all power management and reset functions in the chip
- Smart integration** with the ring
 - Provides cores/Graphics /Media with high BW, low latency to DRAM/IO for best performance
 - Handles IO-to-cache coherency
- Separate **voltage and frequency** from ring/cores, **Display integration** for better battery life
- Extensive **power and thermal management** for PCI Express* and DDR



Maximizing Power-Limited Performance

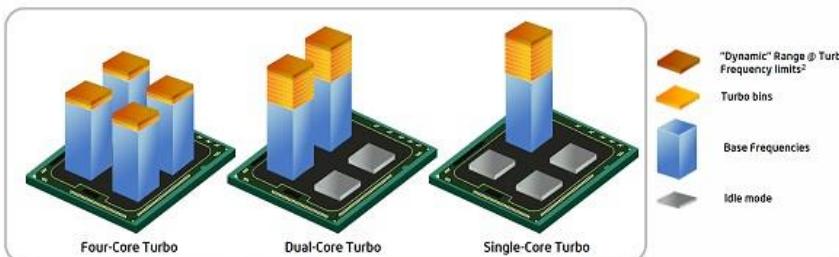
- Extended operating range
 - Power efficient features: better than voltage / frequency scaling
 - Continued focus on gating unused logic and low-power modes
 - Optimized manufacturing and circuits
- Independent frequency domains
 - Cores separated from LLC+Ring for fine-grained control
 - Power Control Unit dynamically allocates budget when power-limited
 - Prioritization based on run-time characteristics selects domain with the highest performance return



- Do čipu jdou pouze 2 různá napětí.
- Procesor již sám vytváří všechna další
- V_{DDQ} napájí paměťový řadič
- V_{CCIN} napájí zbytek procesoru

- Technologie umožňující vyšší výkon pro sekvenční aplikace (jádro Nehalem+)
 - Aktivuje se pokud některé jádro pracuje na plný výkon, ale nejsou vytížena všechna jádra
 - Zvyšování frekvence je závislé na
 - Počtu aktivních jader
 - Odhadovaném proudovém zatížení
 - Odhadované spotřebě příkonu
 - Teplotě procesoru

Intel® Turbo Boost Technology 2.0



Efficient.

Adapts by varying turbo frequency to conserve energy depending upon the type of instructions

Dynamic.

Boosts power level to achieve performance gains for high intensity "dynamic" workloads

Intelligent.

Power averaging algorithm manages power and thermal headroom to optimize performance

Intel® Turbo Boost Technology 2.0 delivers intelligent and energy efficient performance on demand

INTEL® TURBO BOOST MAX TECHNOLOGY 3.0



- In-Die Variation naturally produces parts with some cores that are faster than others (higher performance/ lower voltage)
- Intel® Turbo Boost Max Technology 3.0
 - Identifies the best performing core to provide increased single threaded performance
 - Requires OS awareness or Intel's core affinization driver to get the performance benefits.
- Processor continues to operate within specifications/warranty, this is not "overclocking"

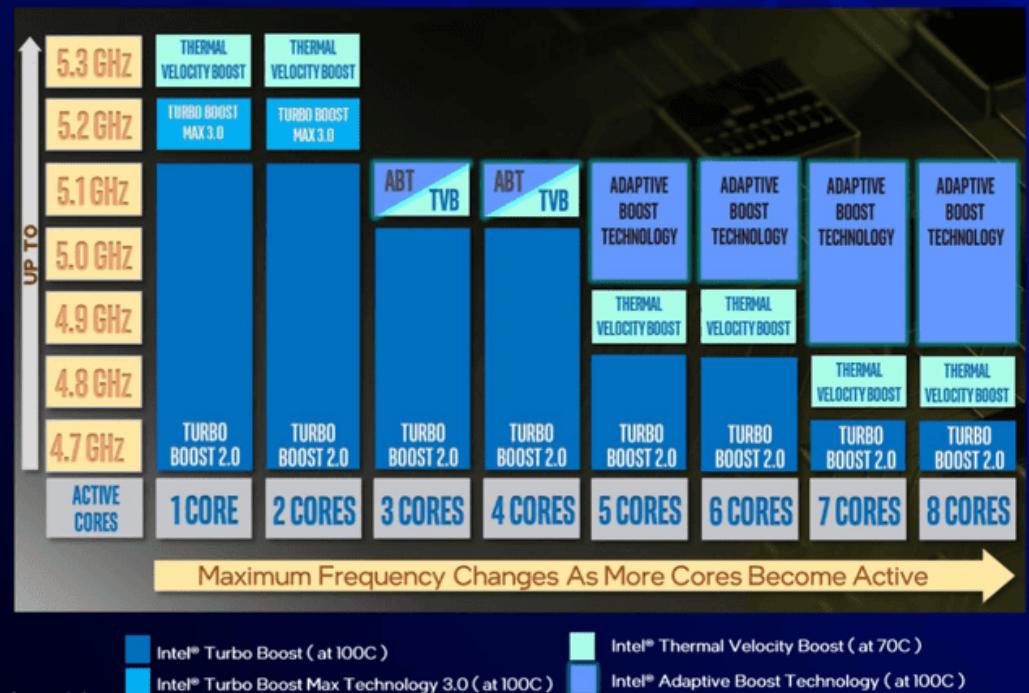
Intel® Turbo Boost Max Technology 3.0 improves single thread performance more than 15% vs. previous gen¹

Intel® Adaptive Boost Technology Unleashing Multi-Core Turbo Performance

Intel Adaptive Boost Technology improves the 11th Gen Intel® Core™ i9 K and KF desktop processors performance by opportunistically allowing higher multi-core turbo frequencies.

In systems equipped with enhanced power delivery and cooling solutions, Intel Adaptive Boost Technology allows additional multi-core turbo frequency while still within the spec's current and temperature limits.

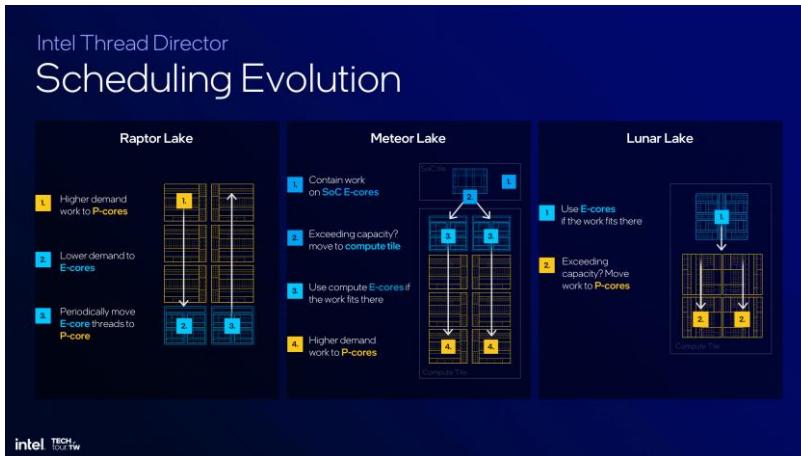
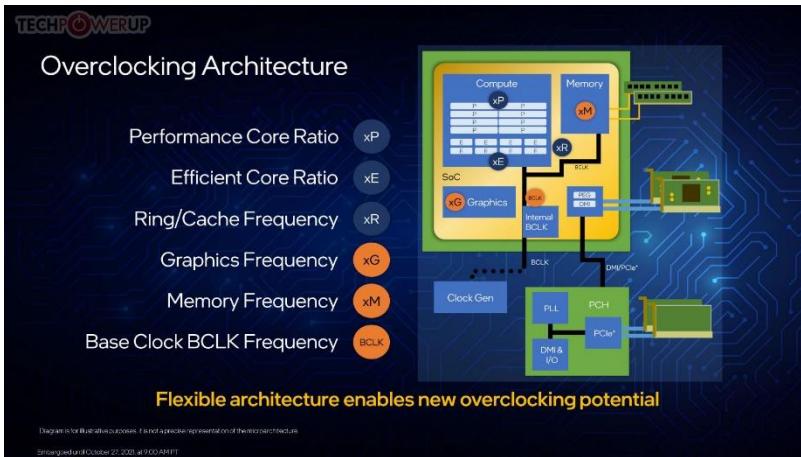
Like past Intel turbo technologies, Intel® Adaptive Boost Technology will be within spec operation and is not considered overclocking.



Intel® Adaptive Boost Technology Disclaimer: When enabled, Intel® Adaptive Boost Technology (Intel® ABT) is a feature that opportunistically allows additional multi-core Intel® Turbo Boost Technology frequencies, while operating within system power and temperature specifications, when current, power and thermal specification headroom exists. The frequency gain and duration is dependent on the workload, capabilities of the processor and the processor cooling solution.

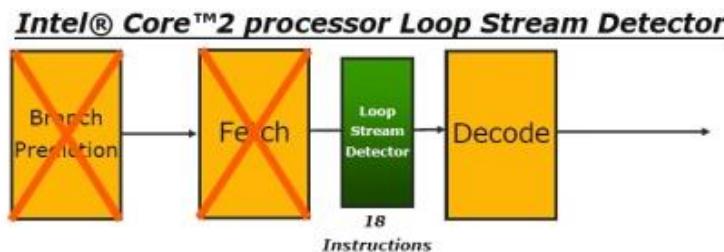
Performance varies by use, configuration and other factors. Learn more at www.intel.com/PerformanceIndex.

Overclocking Architecture and Thread Scheduling



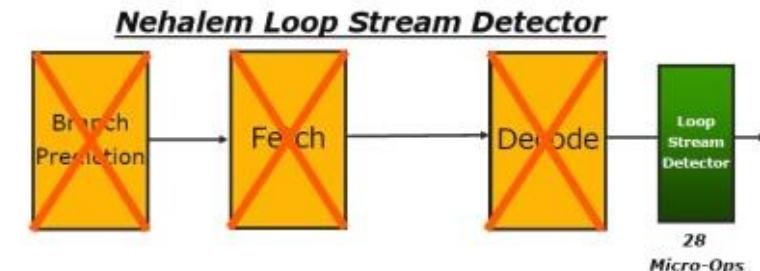
Loop Stream Detector Reminder

- Loops are very common in most software
- Take advantage of knowledge of loops in HW
 - Decoding the same instructions over and over
 - Making the same branch predictions over and over
- Loop Stream Detector identifies software loops
 - Stream from Loop Stream Detector instead of normal path
 - Disable unneeded blocks of logic for **power savings**
 - **Higher performance** by removing instruction fetch limitations



Nehalem Loop Stream Detector

- Same concept as in prior implementations
- **Higher performance:** Expand the size of the loops detected
- **Improved power efficiency:** Disable even more logic



- U starších čipů (Core 2, Nehalem) jsou data součástí mikroinstrukce a cestují po OOO enginu společně s instrukcí (RoB).
- Fyzický registrový soubor drží data na jednom místě
- V instrukci je tak pouze pointer na fyzický registr.

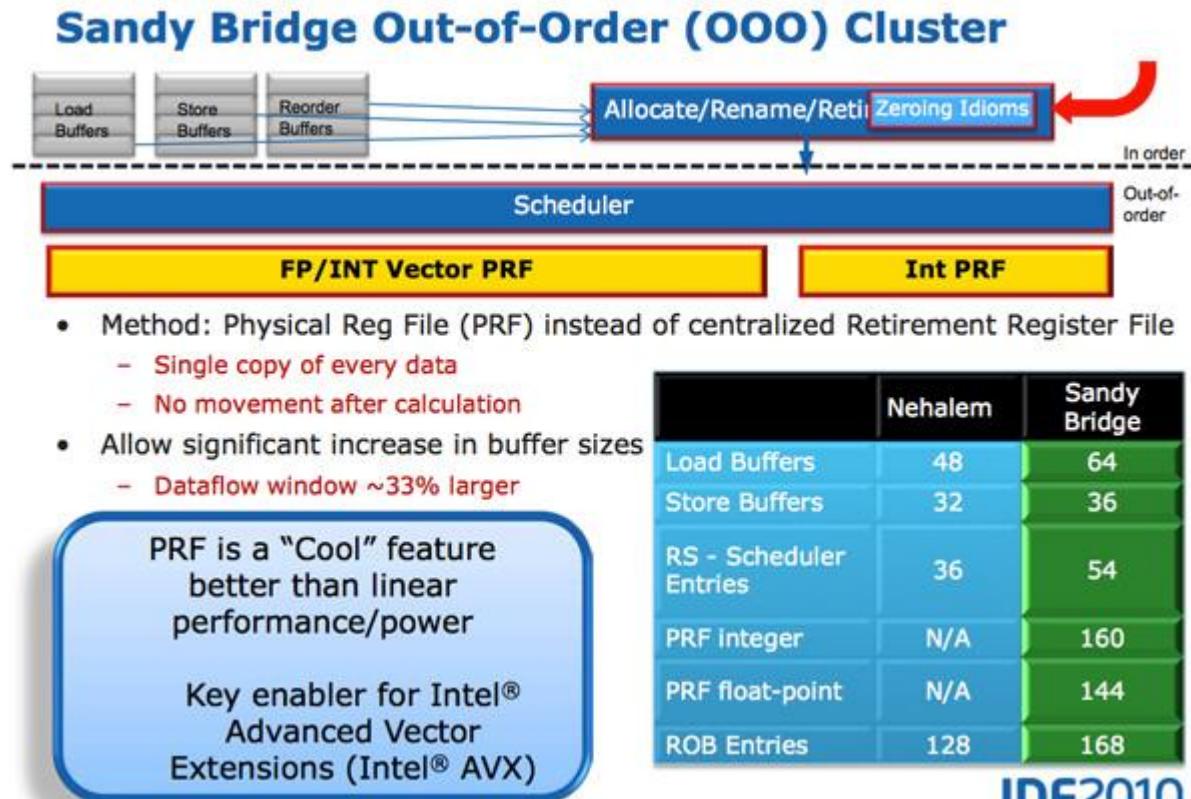
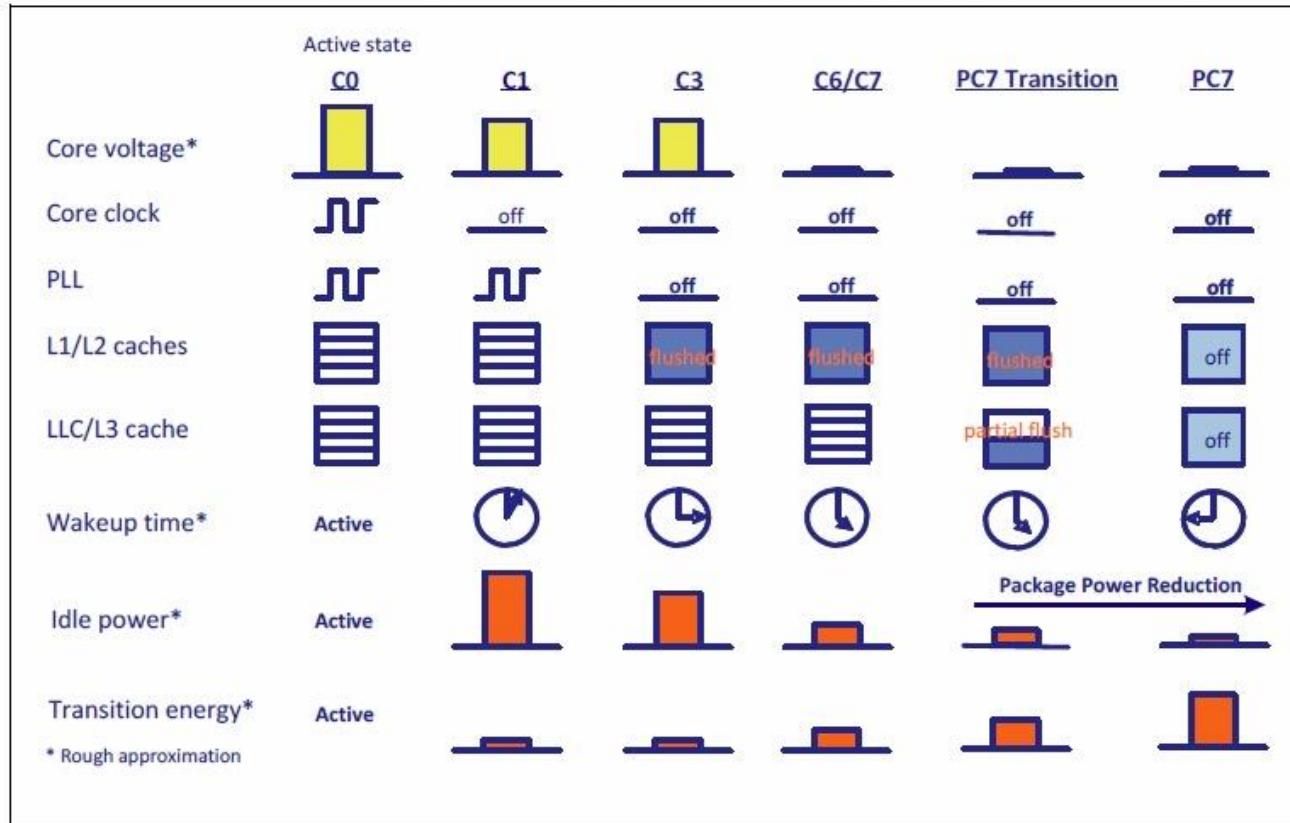
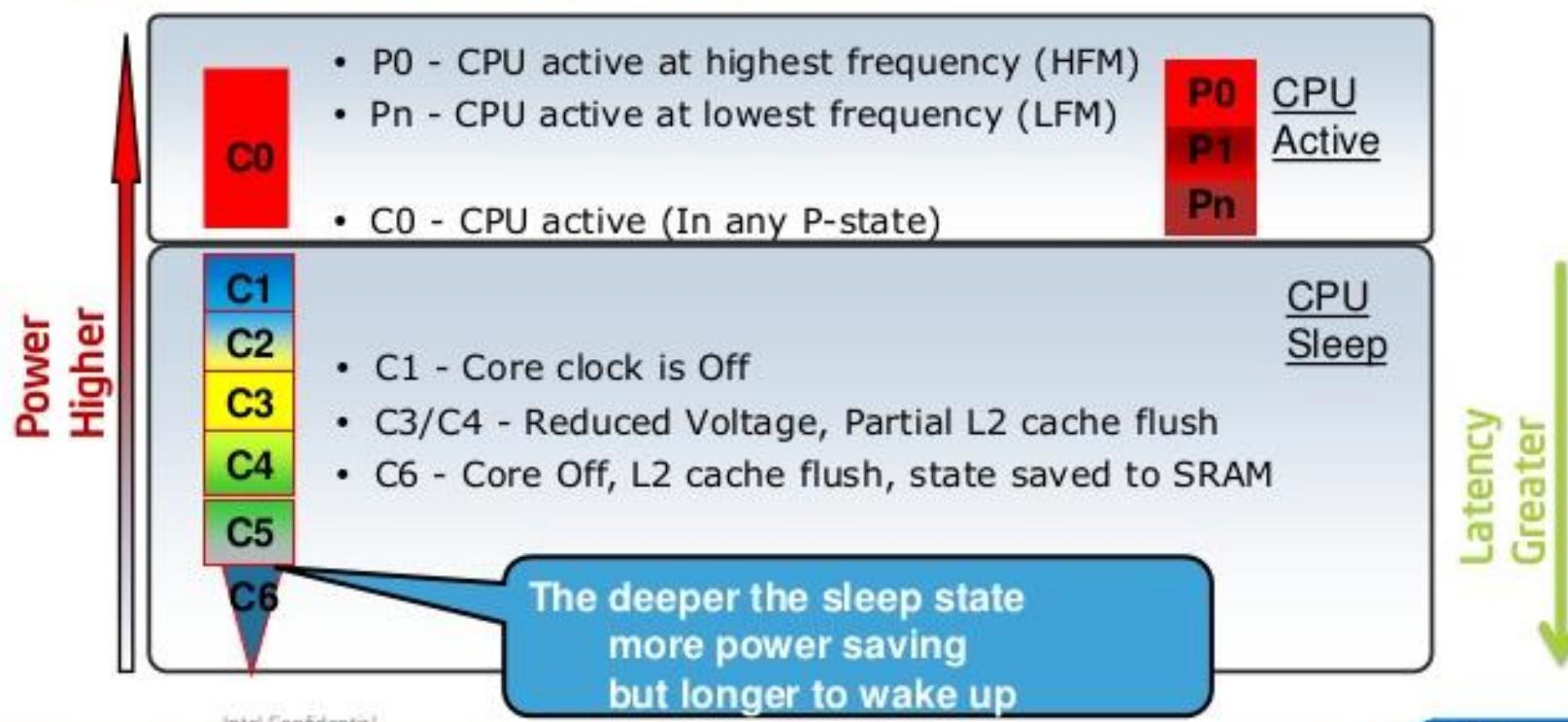


Figure 4: Flexible C-states to select Idle Power Level vs. Responsiveness



CPU C-States / P-States

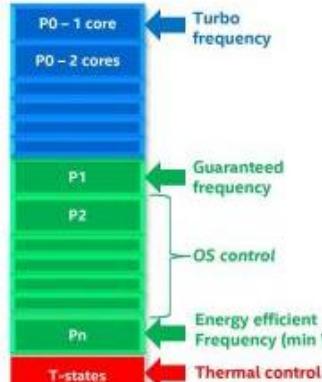


Intel Confidential

Intel Speed Shift (Skylake)

P-state

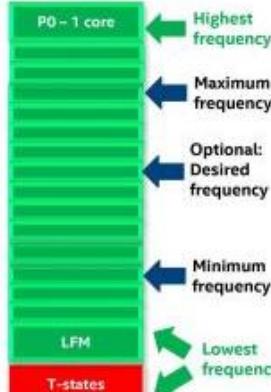
(All CPUs, plus Skylake)



Fixed, Limited
Discrete P-states

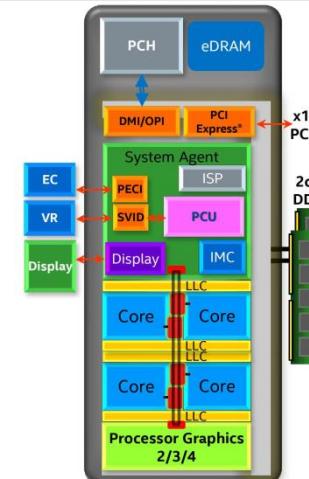
Speed Shift

(Skylake Only)

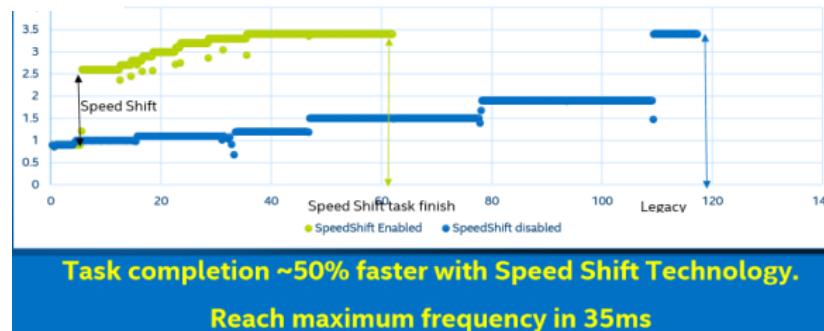


Full Multiplier Range

- P-state defined in BIOS, Managed by OS
- Speed Shift gives full or partial frequency control back to CPU
 - Can give full range or narrow window
- Speed Shift requires OS support
 - Windows 10 via new update
 - Others perhaps later
- Speed Shift means quicker response to performance burst requests
 - Javascript, Office Tools, Web Browsing
- Performance increase in regular tasks
- Slight overall power reduction
- Requires any Intel 6th Gen Skylake CPU
- Collaboration between Intel and Microsoft specifically for W10 + Skylake



Speed Shift Frequency transitions
Frequency (GHz) vs Time (ms)
Representative task with and without Speed Shift Technology



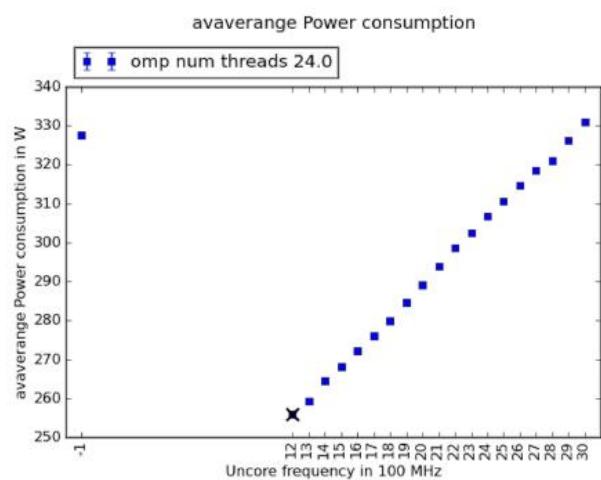
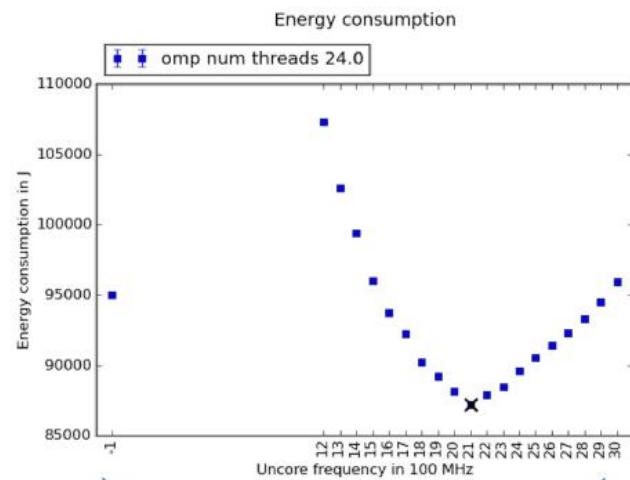
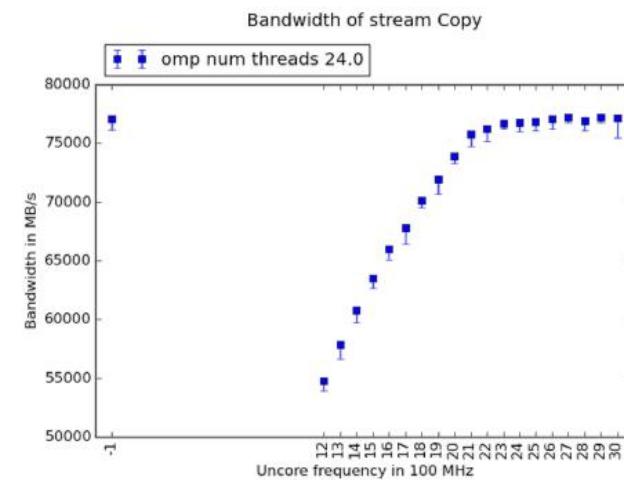
Task completion ~50% faster with Speed Shift Technology.

Reach maximum frequency in 35ms

- **Taktování pod Linuxem**
 - Zjištění aktuálního nastavení
 - `cpupower frequency-info`
 - Nastavení řídicího governoru
 - `sudo cpupower frequency-set --governor performance`
 - Nastavení rozsahu frekvence
 - `sudo cpupower frequency-set --min 1600MHz --max 2000MHz`
- **Měření spotřeby – nástroj *Intel Performance Counter Monitor***
 - `sudo modprobe msr`
 - `sudo ./pcm-power.x -- ./my_app my_arguments`
- Alternativně s pomocí knihovny PAPI RAPL

- Vliv CPU uncore frekvence (Cache, propojení jader, řadič paměti) na propustnost stream benchmarku a spotřebu energie
 - Optimální frekvence má minimální dopad na výkon, ale uspoří energii

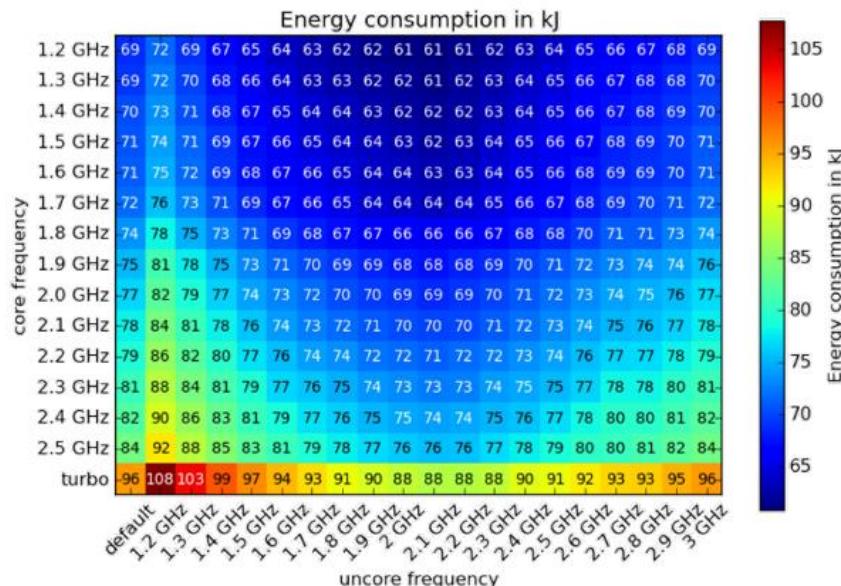
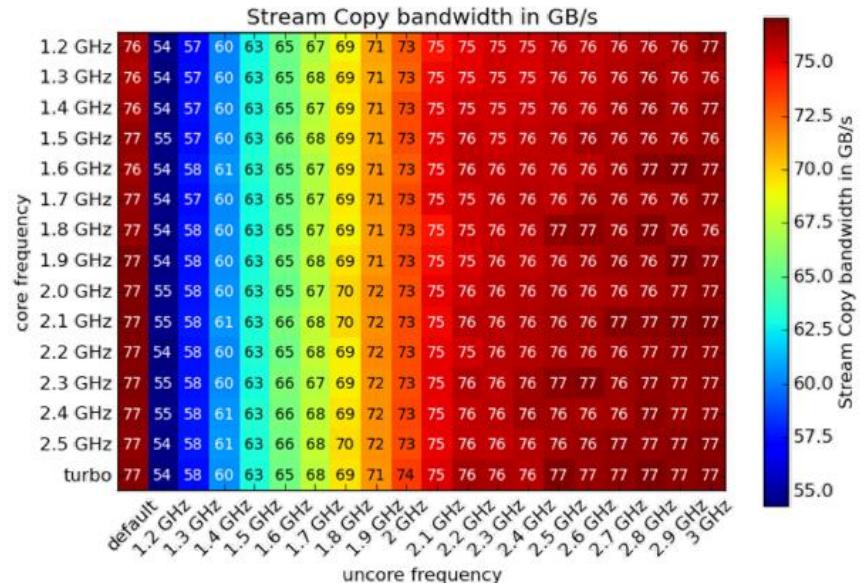
Evaluation using STREAM Copy benchmark



| Praktické ukázky – Benchmark stream (memcpy) | FIT

- Vliv změny core frekvence (jádra) a uncore frekvence (paměť) pro memory bound problémy (nízká aritmetická intenzita)

Evaluation using STREAM Copy benchmark

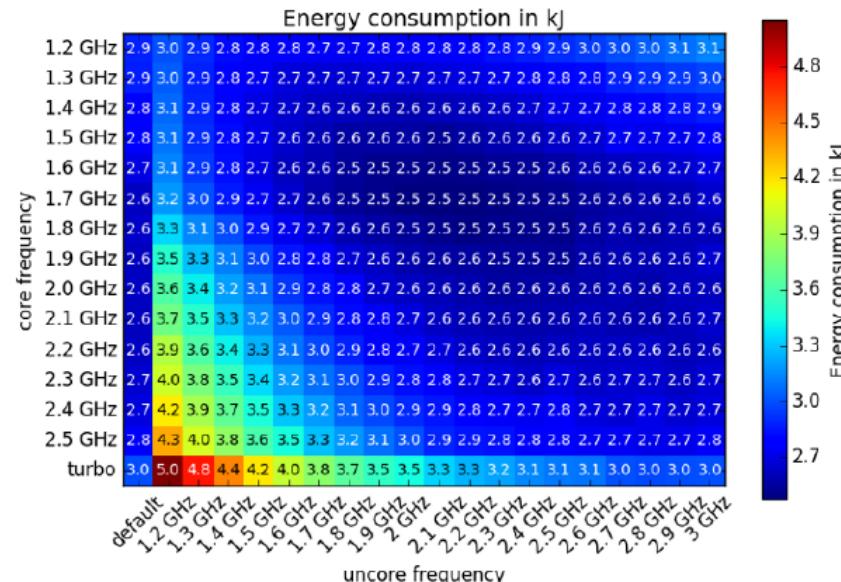
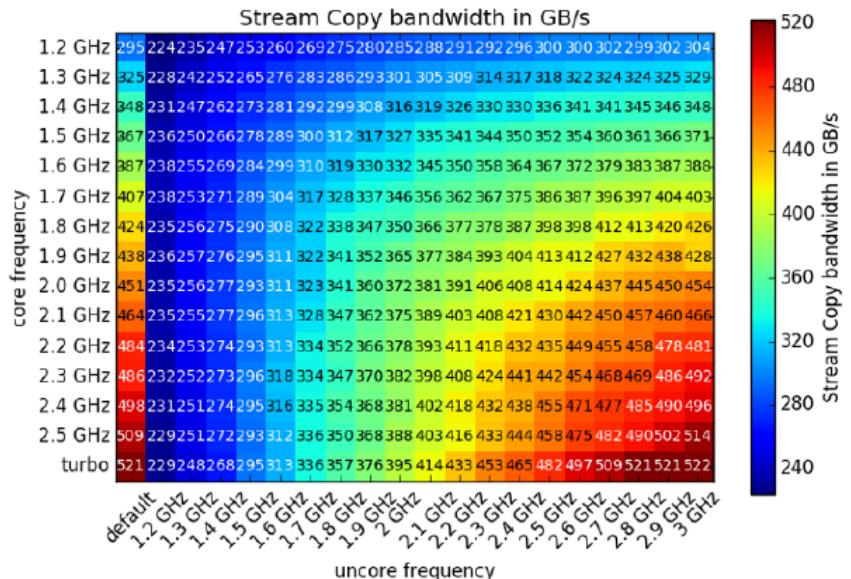


Heatmap of the energy consumption of a stream benchmark for different core and uncore frequencies.
The data array does not fit in the processor's L3 processor cache

- Efektivita práce s L3 cache při různých core a uncore frekvencích**

- Data se vlezou do cache

Evaluation using STREAM Copy benchmark



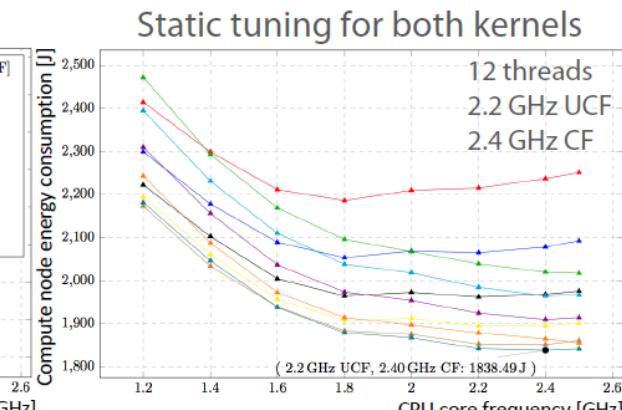
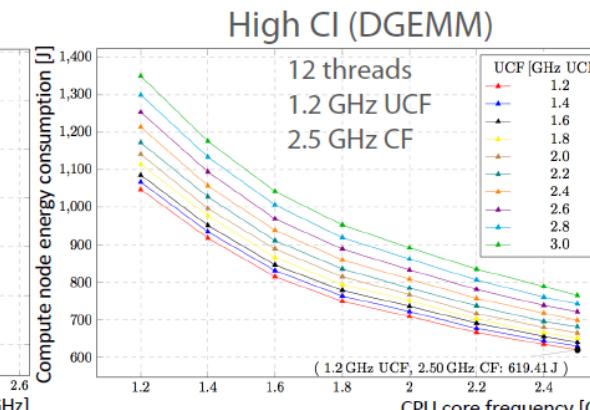
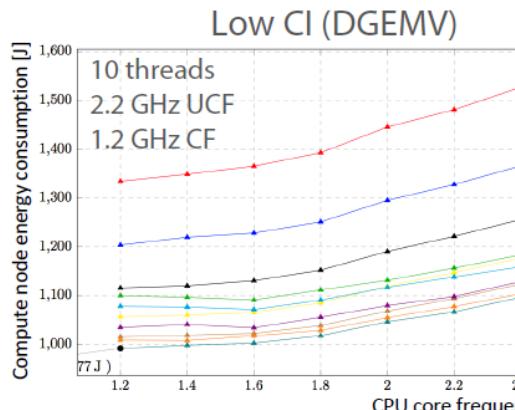
Heatmap of the energy consumption of a stream benchmark for different core and uncore frequencies.
The data array does fit in the processor's L3 processor cache

I Hledání optimálních parametrů procesoru

Behavior of the simple application with two kernels

- Low computational intensity – DGEMV
- High computational intensity – DGEMM
- Tuning of three parameters
 - Core frequency
 - Uncore frequency
 - Number of OpenMP threads
- Visualized by RADAR

Two kernels with 1:1 workload ratio	Energy consumption	Energy savings
Default settings	2017J	- -
Static optimal	1833J	179J 9%
Dynamic optimal	1612J	221J 12%
Total savings	-	400J 20%



Note: runtime of both kernels was equal for default settings

HISTORIE A PŘÍKLADY SUPERSKALARNÍCH ARCHITEKTUR

- **P5 (1993):**

- první superskalární IA-32 mikroarchitektura – 1993:
 - In Order, dvojitá integer pipeline (**U** a **V**) 5 stupňů.
 - Dokončují až 2 instrukce/takt. Kompilátor plánoval dvojice staticky.

- **P6 (1995):**

- **OOO**, zavedeno **super-řetězení** (14 stupňů).
- **Procesory**: Pentium Pro, Pentium II, III; MMX a SSE.
- **Modernizovaná P6**: Pentium M, Core Solo, Core Duo.

- **NetBurst (2000):**

- Trace cache, 31 stupňů, SSE2, SSE3, hyper-threading HT, EM64T.
- **Procesory**: Pentium 4, Pentium D, Xeon

- **Core (2006):**

- příkon ↓, 14 stupňů pipeline, 65 nm, multi-core, SSE3, Intel 64
- Procesory: Pentium dual core, Celeron, Xeon, Core 2

- **Nehalem (2008): řady: i3, i5, i7**

- 45 nm, HT, L3C, Quick Path, integrované MemCtrl, bufer μops.
- 32 nm Nehalem = Westmere: IGP (Integrated GPU).

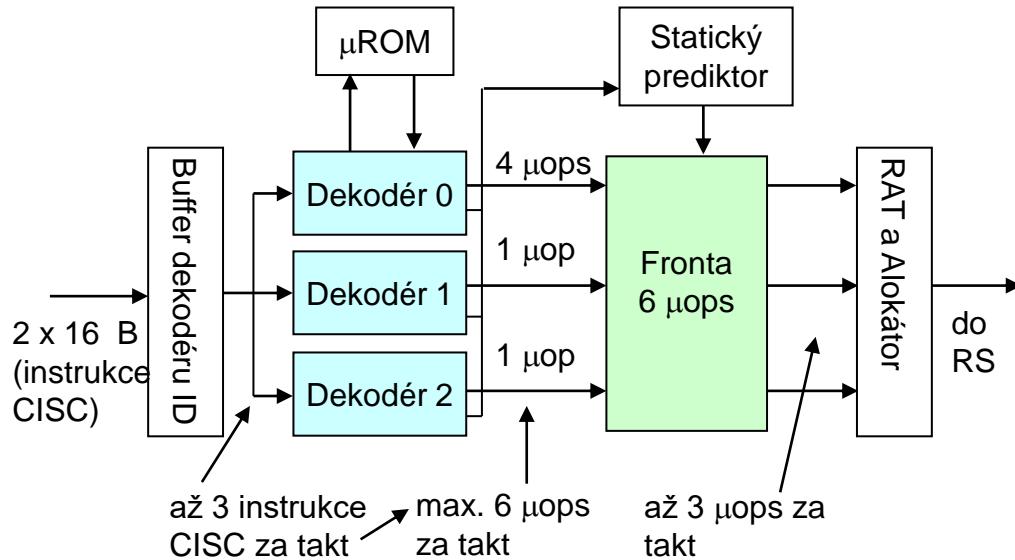
- **Sandy Bridge 2010:**

- 32 nm, AVX 256 bitů, μop-cache, HT.
- 22 nm Sandy Bridge = Ivy Bridge: 3D-tranzistor.

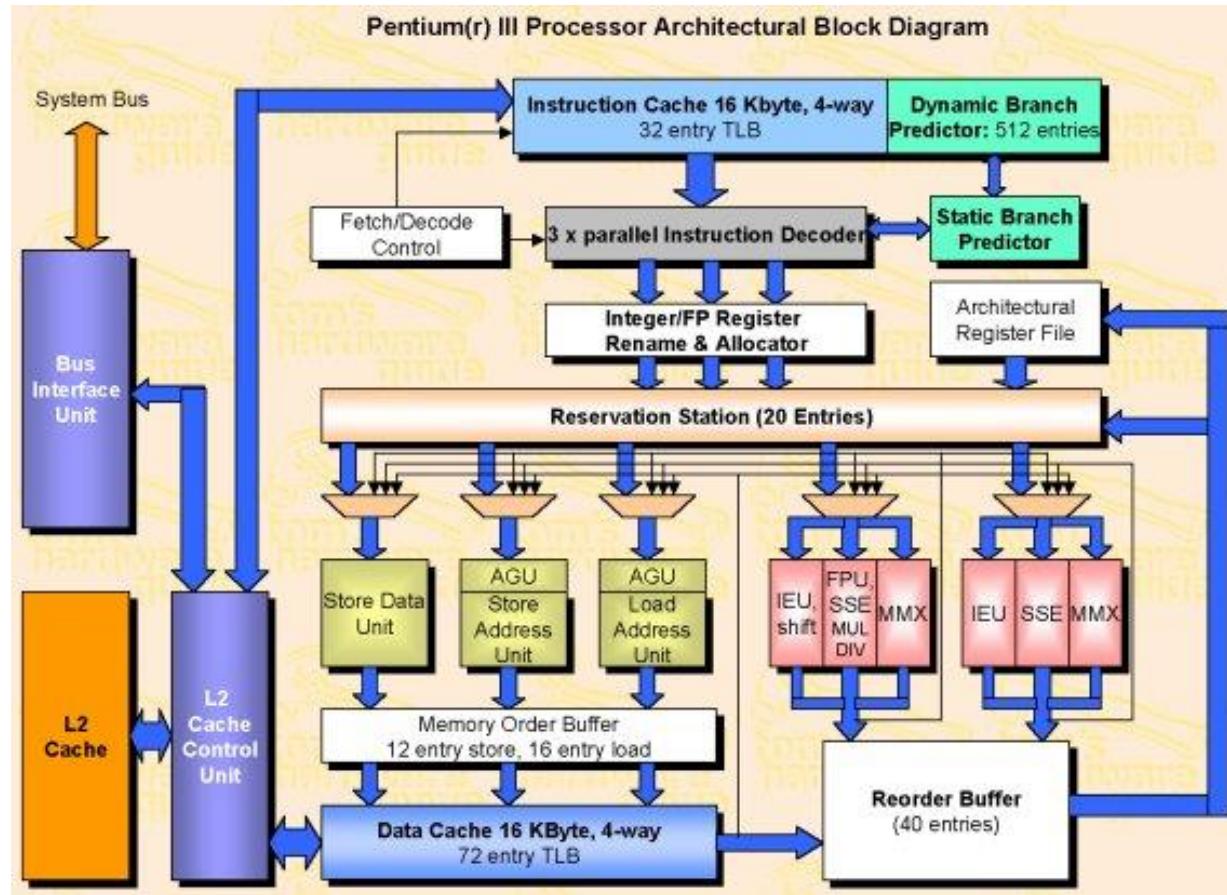
- Haswell 2013:
 - 22 nm, 4 ALU, 3 AGU, 2 jednotky predikce skoků, AVX2, FIVR (Fully Integrated Voltage Regulator)
 - 35–40 MB LLC. Server procesory až 20 jader, možnost rozdělit jádra do 2 uzlů NUMA (COD, cluster on die)
 - 4 verze integrované GPU (až 40 EU), TDP 35–140 W.
 - 14 nm Sandy Bridge = Broadwell
- Skylake 2015:
 - 14 nm, 4 typy Y, U, H a S (TDP 4–95W) integrovaná L4 eDRAM cache (64/128 MB), podpora DDR3/4
 - Změna struktury cache – výrazný nárůst L2 na úkor L3
- Sunny Cove 2015:
 - Zvětšena trace cache, 384 položek ROB, Rozšířeny L/S jednotky a L/S bufery
- Alder Lake 2021:
 - Heterogenní procesor – silná (Golden Cove) a slabá jádra (Gracemont)
 - Výrazný nárůst zdrojů na silných jádrech (512 položek ROB, 12 exekučních portů, 6-wide dekodér)
 - AVX-512 vypnuto (slabá jádra totiž nepodporují)

P5 (Pentium)	superskalární, „in-order“	5
P6 (Pentium Pro)		14
P6 (Pentium III)		10
NetBurst Pentium 4 (180 a 130 nm)		20
NetBurst Pentium 4 (90 a 45 nm)		31
Core	superskalární	14
Nehalem	„out-of-order“	16
Sandy Bridge		14–19
Ivy Bridge		14–19
Haswell		14–19
Bonnell (Atom)		16
Quark	skalární	5

- Řetězené zpracování CISC-ových instrukcí x86 se řeší transformací (dekódováním) na RISC-ové mikrooperace délky 72 bitů.
- Délka instrukcí x86: 1–15 B, **dekodér délky instrukcí** posílá až 3 instrukce x86 na 3 dekodéry:
 - D0 zpracovává 1. instrukci, která generuje až 4 µop/takt.
 - D1 a D2 zpracovávají jednodušší 2. a 3. instrukci, které nejsou delší než 8 B a generují jen 1 µop.
 - 2. a 3. instrukce musí čekat na D0, pokud to nesplňují.
 - Pro dekódování instrukcí, které generují více než 4 µop je použita paměť mikrokódu a generování trvá 2 nebo více taktů.



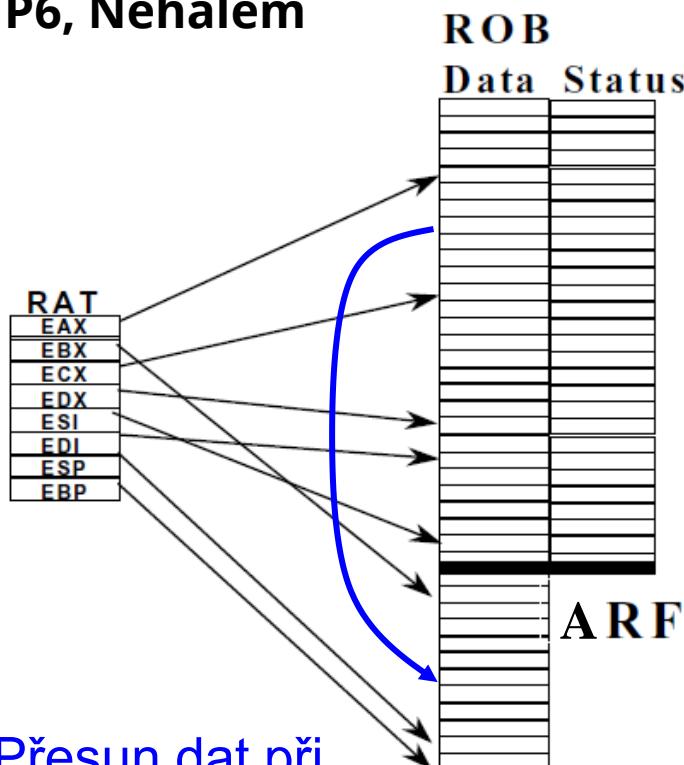
- Prediktor podmínky je 2 úrovňový adaptivní (autoři Yeh a Patt) s $k = 4$ -bitovým lokálním BHSR
 - perfektně predikuje libovolné periodické sekvence až $k + 1 = 5$ bitů
 - na 1 skok je třeba 36 bitů (= 16 dvoubitových prediktorů + 4 bity BHSR).
- BTB je organizován jako skupinově asociativní cache (128 skupin, 4 cesty, tj. 512 položek)
- Položka obsahuje adresu skokové instrukce b , cílovou adresu skoku t a 4 bitový lokální BHSR. Jeden index do PHT je část adresy skoku b , jako druhý index se použije obsah BHSR.
- **Pokuta za špatnou predikci je 10–20 taktů.**
- Není-li skok v BTB, použije se statická predikce (skok v kódu dopředu -, skok dozadu +)



- r. 2000, IA-32 procesor (adresa 32 bitů, instrukce x86)
- SSE2 (Streaming SIMD Extension 2)
- Trace Cache (kapacita 12k μ ops, cca 64 bitů / μ op
 - TC může rozeslat do RS 3 μ ops/takt, vydat do FJ se může až 6 μ ops/takt a propustit opět 3 μ ops/takt.
- Přejmenování mapuje 8 standardních registrů x86 na 128 vnitřních fyzických registrů PRF, 2 tabulky RAT (front-end a propouštěcí) → není nutno kopírovat registry při propouštění.
- ROB: až 126 μ ops v pořadí bez hodnot dst operandů
- **Co bylo špatné:** chyběla L3 cache na čipu, malá L1 D-cache (8 KiB), výkonnost \approx Pentium III, velký příkon

Přejmenování registrů v P6 a NetBurst

P6, Nehalem



Přesun dat při propuštění instrukce

Sandy Bridge

NetBurst

přepis při špatné predikci skoku

Frontend RAT

EAX
EBX
ECX
EDX
ESI
EDI
ESP
EBP

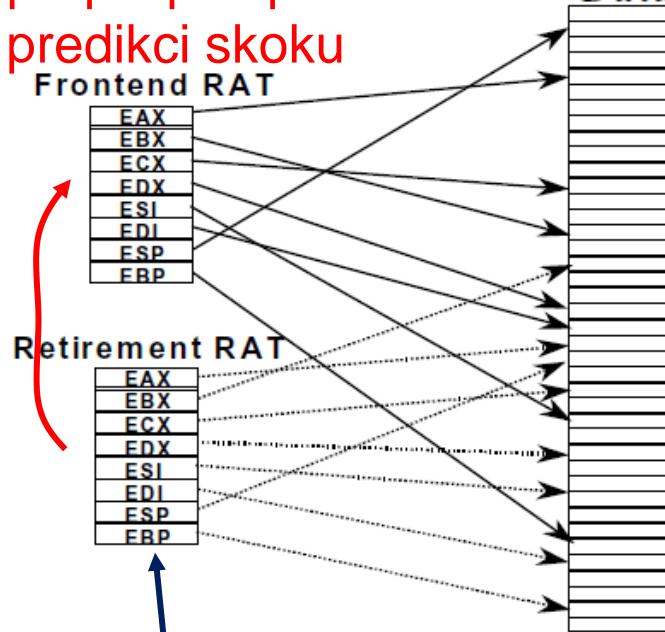
Retirement RAT

EAX
EBX
ECX
EDX
ESI
EDI
ESP
EBP

PRF

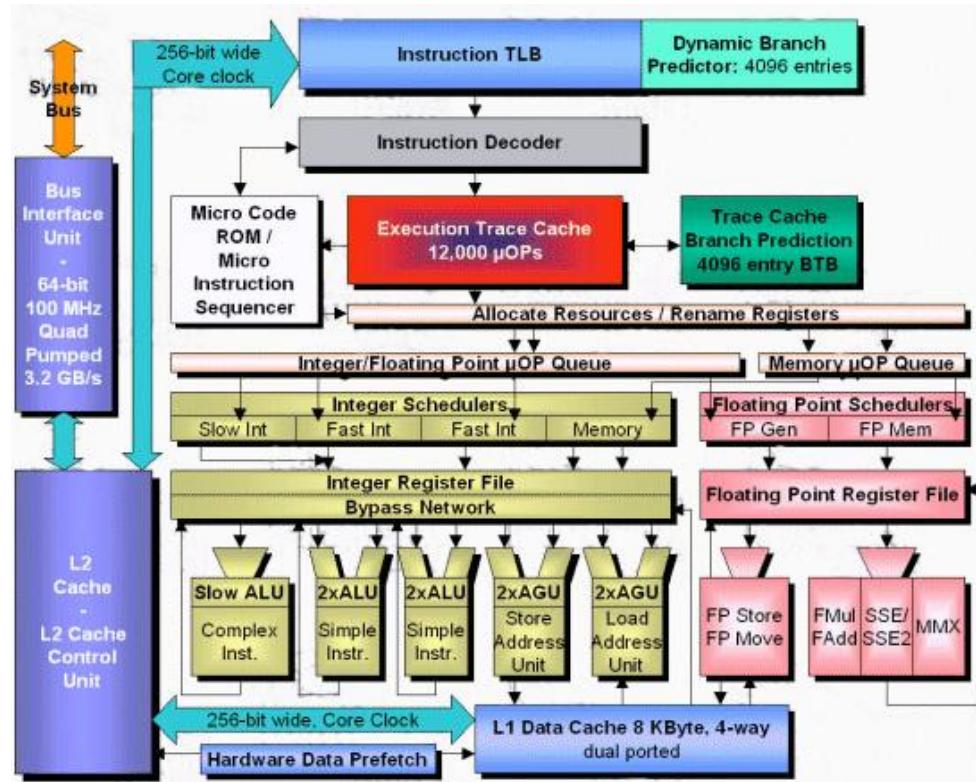
ROB

Status



Při propuštění instrukce se sem vloží mapování jejího dst registru

Pentium 4 – architektura NetBurst

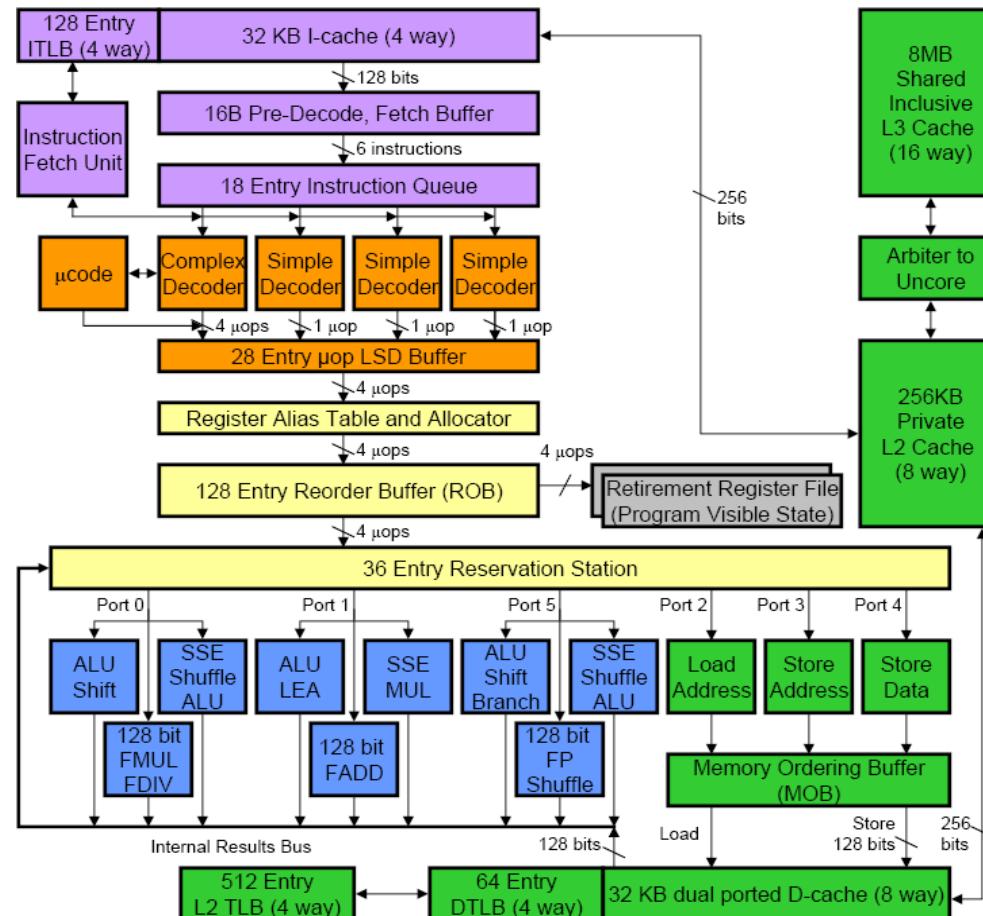


1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
TC Nxt IP	TC Fetch	Drive Alloc	Rename	Que	Sch	Sch	Disp	Disp	RF	RF	Ex	Flgs	Br Ck	Drive					

- Použita u vícejádrových procesorů se sdílenou pamětí cache L2. Snížení příkonu a zvýšení výkonnosti činí kolem 40 %. **Vychází z P6.**
- **Široká instrukční linka.**
 - Dekóduje a propouští až **4 instrukce za takt**, rozesílat a provádět může až **5 µops**.
 - Umí sdružovat instrukce x86 (*macrofusion*) a také sdružovat µops vzniklé z jedné x86 (*microfusion*), čímž lze dosáhnout až 6 µops za takt.
 - Ukazovatel zásobníku je modifikován speciálním HW. To dovoluje **načítání dat ze zásobníku již na začátku linky** (25 % všech načítání je ze zásobníku).
 - Tyto inovace zachovány i v novějších mikroarchitekturách
- **Pokročilá práce s multimédii.** Instrukce **MMX, SSE, SSE2, SSE3** se **128 bity** provedené **za 1 takt** znamenají výkonnost až 24 GFLOP/s (1 jádro na 3 GHz, SP).
- **Inteligentní napájení.** Dynamické odpojování subsystémů dle potřeb nebo přepojování do úsporného režimu neovlivňuje responzivitu.
- **Pokročilá chytrá cache.** Sdílená sjednocená cache úrovně 2 může být celá k dispozici jen 1 jádru, když druhé není aktivní. Špičková přenosová rychlosť je 96 GB/sec @ 3 GHz.
- **Chytrý přístup do paměti.** Je zavedena podpora pro **RPW** i pro případ dosud neznámé adresy zápisu (dynamické rozlišování adres, *memory disambiguation*)
- **Přednačítání dat** do L1/L2 D-cache pomocí tabulky historie čtení

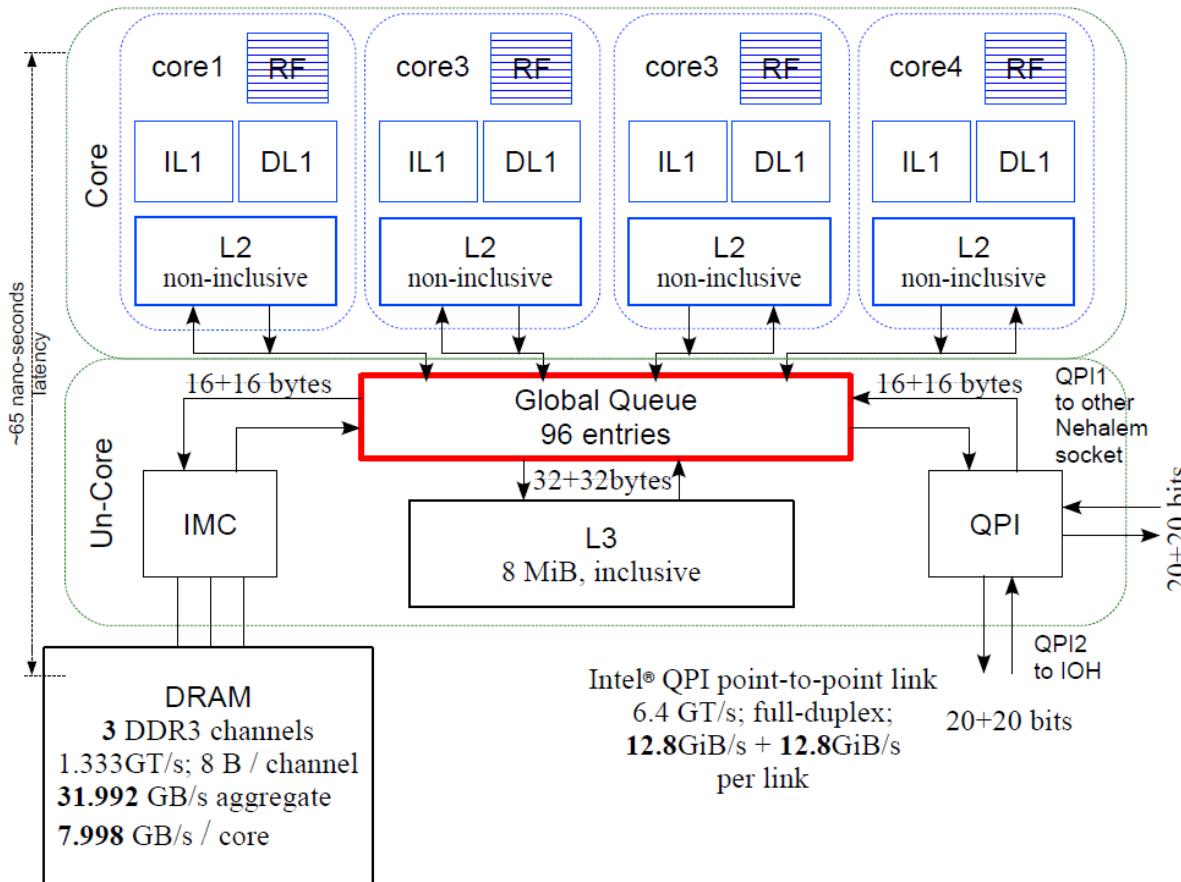
- Druhá generace architektury Core, orientace na výkonnost: 2, 4, 6 nebo 8 jader (45 nm, 4 jádra: 731 M tranzistorů)
- 16 stupňová linka, 6 FJ (3 paměťové, 3 výpočetní) a **HyperThreading (HT)**
- Větší cache a vyšší propustnost pamětí 32 KiB L1 I-cache, 32 KiB L1 D-cache, L2C: 256 KB.
- **Register Alias Table (RAT)** může přejmenovat až 4 μop za takt a každé přidělit dst. registr v ROB více rozpracovaných mikrooperací.
- **Dvouúrovňový prediktor skoků i dvouúrovňový TLB**.
- **Loop Stream Detector LSD**: ve frontě 18 před-dekódovaných instrukcí detekuje každé tělo smyčky, uloží je dekódované do LSD buferu (až 28 μop) a pak opakovaně používá (bez načítání a dekódování) až do špatné předpovědi skoku (malá náhrada Trace cache).
- **Turbo režim**: kmitočet hodin se zvyšuje, pokud není překročena teplotní mez.

Mikroarchitektura Nehalem

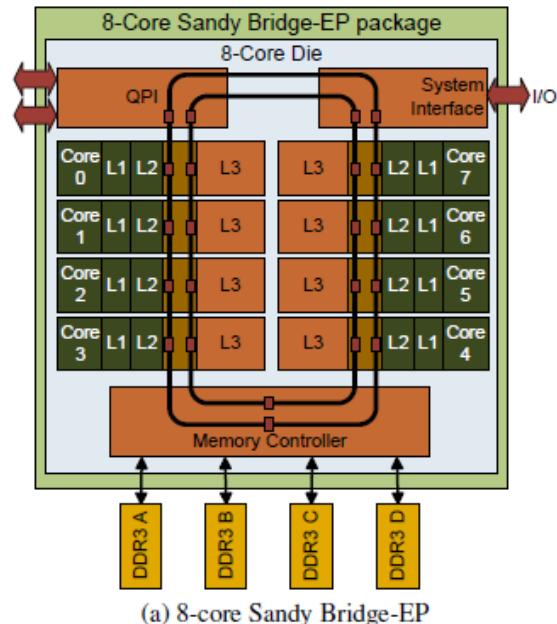


- **Inovace:** nové propojení soketů: front-side bus FSB nahrazen linkami (QPI, Quick Path Interconnect).
- **Inovace:** integrovaný řadič paměti, podporuje 3 paměťové kanály DDR3 SDRAM nebo 4 FB-DIMM, severní most eliminován.
 - Firma AMD zavedla linky HyperTransport (HT) a integrované řadiče paměti již v roce 2003, o 5 let dříve.
- **Sdílená L3C:** 4–8 MB je inkluzivní, obsahuje data z L1 i L2 a info, kde jsou lokálně (menší komunikace).
- **Čip procesoru grafiky** v témže pouzdro jako CPU.
- **Řízení příkonu:** vestavěný mikrořadič a senzory teploty, proudu a příkonu, odepínání jader, možnost redukce příkonu pamětí a QPI.
- Global Queue (GQ) uchovává, spravuje a plánuje tok dat v „uncore“. Má 3 fronty požadavků:
 - WQ, žádosti zápisu z lokálních jader, 16 položek
 - LQ, žádosti čtení z lokálních jader, 32 položek
 - QQ, fronta QPI, žádosti jdoucí mimo čip, 12 položek
- Obsahuje **křížový přepínač** pro výměnu dat mezi propojenými částmi (L2, L3, IMC, QPI).
- **Funkce:**
 - lokální žádost jádra o čtení: GQ sonduje další jádra. Z více vlastníků kopí jedno jádro dodá data.
 - Když nikdo nemá kopii a L3 ano, dodá data inkluzivní L3
 - Výpadek v L3: data dodá lokální IMC (Integrated Memory Controller) za 65 ns, popřípadě vzdálený IMC přes QPI za 105 ns

Celkový diagram procesoru Intel Nehalem



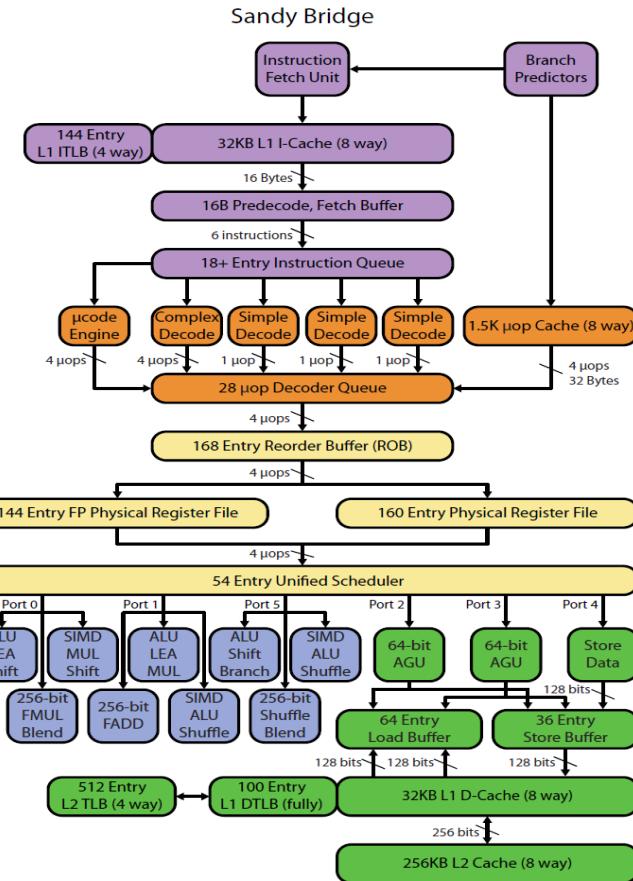
- 4, 6 a 8 jader na 3,0–3,8 GHz s podporou HyperThreadingu (HT) a s technologií Turbo Boost
- Jádra, grafika, L3 cache a systémový agent jsou propojeny **kružnicovým propojením** s propustností 256 bitů/takt.
- Podpora **Advanced Vector Extension (AVX): 256 bitů, 32 GFLOPS/jádro (8 FP/takt),**
- Každé jádro: 32 KiB L1 D-cache + 32 KiB L1 I-cache (3 taktů), 256 KiB L2 cache (8 taktů).
- 8 MiB **sdílená L3 cache** (25 taktů). Je též sdílena s integrovaným grafickým jádrem. Blok cache 64 byte.
- **Integrované jádro grafiky na 1–1,4 GHz, 16 ex. jednotek.**
- **Integrovaný řadič paměti** s max. propustností 25,6 GB/s, podporuje DDR3-1600 dual channel RAM.
- CPU ↔ L3 cache: průměrně jen 1,5 skoků (do lokálního bloku cache není třeba jít po kružnici). Latence sdílené L3 cache je 26 až 31 taktů.



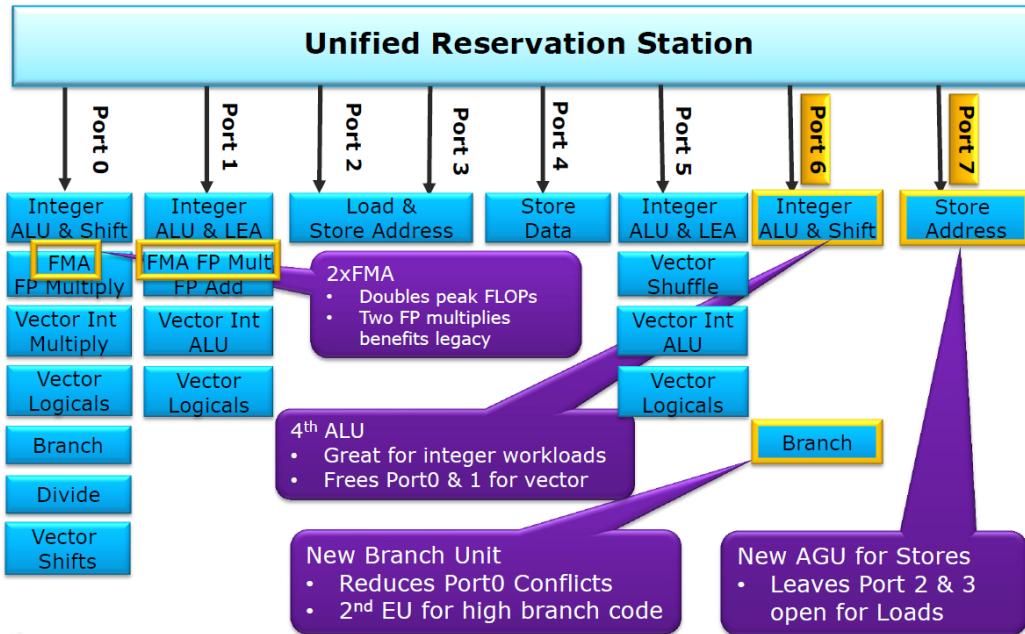
Dvě propojovací kružnice, jedna pro směr nahoru, druhá dolů. Pro každý přenos je vybrán směr kratší cesty do cíle, nejvíce 4 skoky.

μ op-cache (dekódovaná I-cache) v Sandy Bridge

- Je částí L1 I-cache, zachovává výhody Trace cache, eliminuje složité dekódování při mnohem nižším příkonu.
- μ op-cache má kapacitu 1536 μ ops, 10 % velikosti Trace cache Pentia 4.
- Mapování instrukcí do μ op-cache probíhá po blocích 32 B instrukcí, 1 blok může zabrat až 18 μ ops. Každý blok μ op-cache uchovává „metadata“ včetně počtu platných μ ops v bloku a délku odpovídajících x86 instrukcí.
- Jestliže okénko 32 B instrukcí má více než 18 μ ops, musí jít přes tradiční front-end.
- **Mikrokódované** instrukce nejsou v μ op-cache – jsou reprezentovány ptr do ROM mikrokódu a případně několika prvními μ ops.

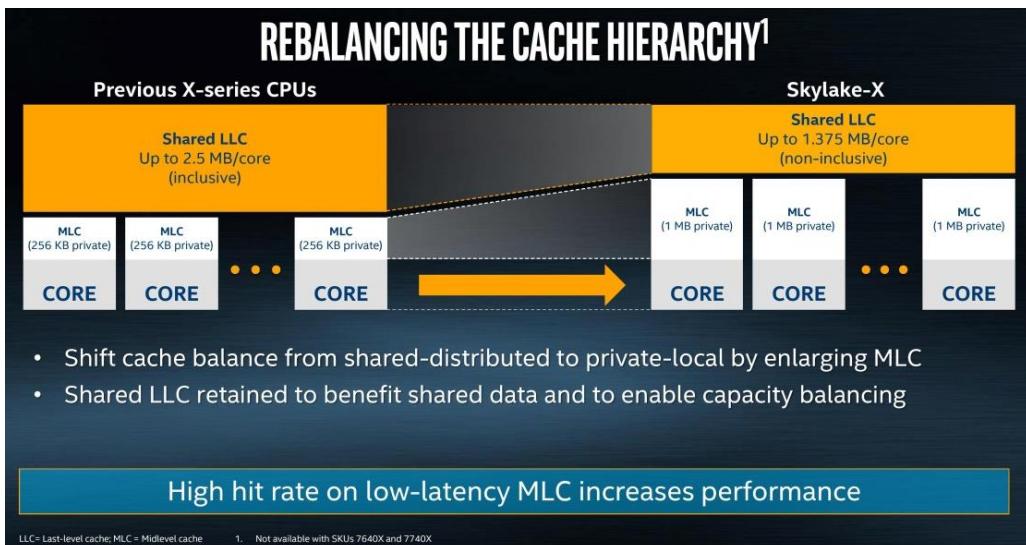


- Haswell je zaměřen na **nižší příkon pro mobilní zařízení** (hybridní laptop-tablety) ale i pro superpočítáče. Dřívejší TDP (Thermal Design Power) 35 až 45 W pro mobilní procesory je redukován na ULT: 13,5 W a 15 W TDP, ULTX: 10 W TDP.
- Superpočítáč v Ostravě obsahuje 24 192 jader Haswell-EP!
- Podpora pro AVX2 a MAD operace.**
- Haswell má výkonnější grafiku GT3e, ze 16 jednotek na 1150 MHz (GT1 u Sandy Bridge) narostla na 40 ex. jednotek a 1300 MHz.
- eDRAM (embedded DRAM) 128 MiB je na vlastním čipu, ale ve stejném pouzdru jako procesor. Pracuje jako sdílená L4 cache jak pro grafiku, tak pro jádra procesoru. Vylepšuje paměťovou propustnost.



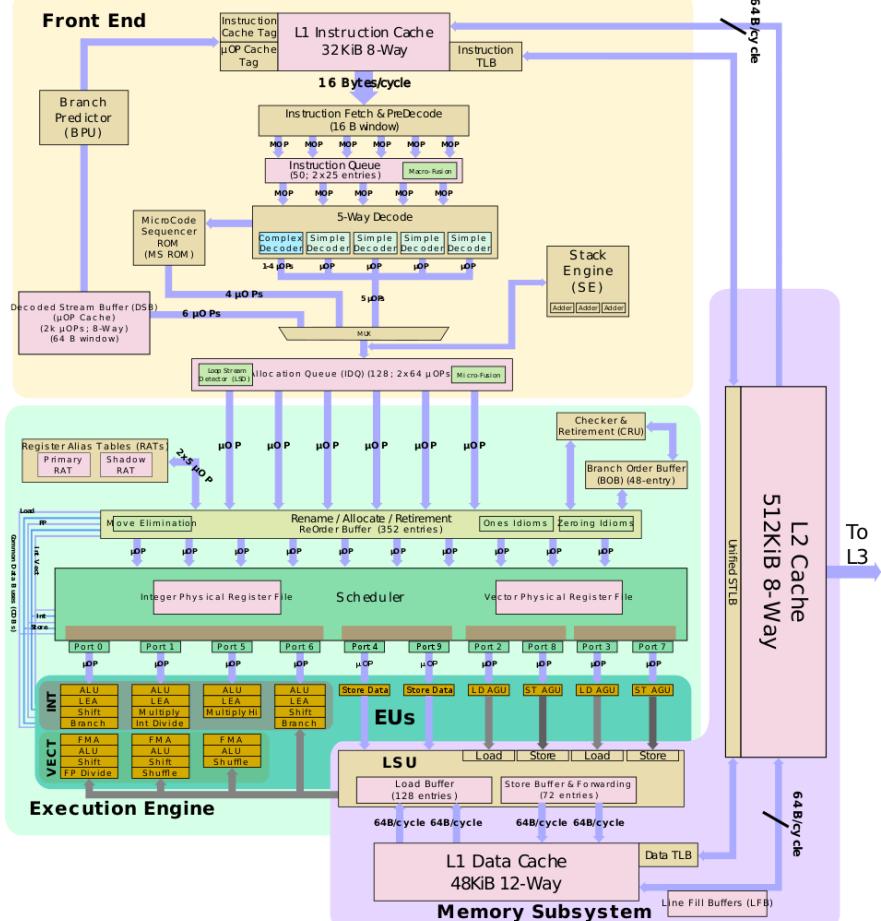
- Broadwell = Haswell předělaný na 14 nm (2014).

- Podpora pro AVX-512**
- Nová organizace cache**
 - Zvětšení L2 cache na úkor L3 cache
- Změna propojovací sítě**
 - Z hierarchických kruhů na 2D mřížku



Comparison: Skylake-S and Skylake-SP Caches		
Skylake-S	Features	Skylake-SP
32 KB 8-way 4-cycle 4KB 64-entry 4-way TLB	L1-D	32 KB 8-way 4-cycle 4KB 64-entry 4-way TLB
32 KB 8-way 4KB 128-entry 8-way TLB	L1-I	32 KB 8-way 4KB 128-entry 8-way TLB
256 KB 4-way 11-cycle 4KB 1536-entry 12-way TLB Inclusive	L2	1 MB 16-way 11-13 cycle 4KB 1536-entry 12-way TLB Inclusive
< 2 MB/core Up to 16-way 44-cycle Inclusive	L3	1.375 MB/core 11-way 77-cycle Non-inclusive

Architektura SunnyCove (2019)



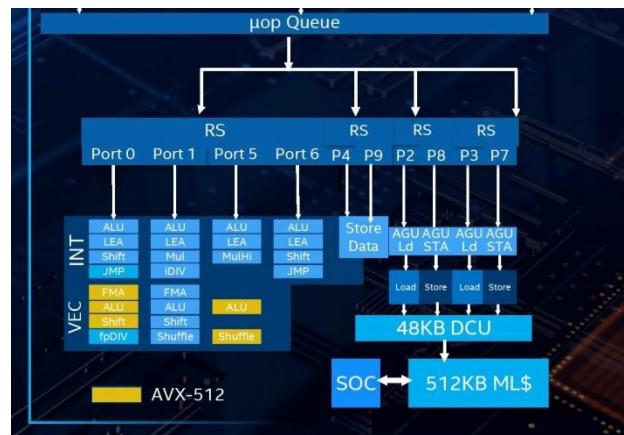
Změna velikostí cache

- L1 z 32 KB → 48 KB
- L2 z 256 KB → 512 KB
- Zvětšena TraceCache (2,25k)

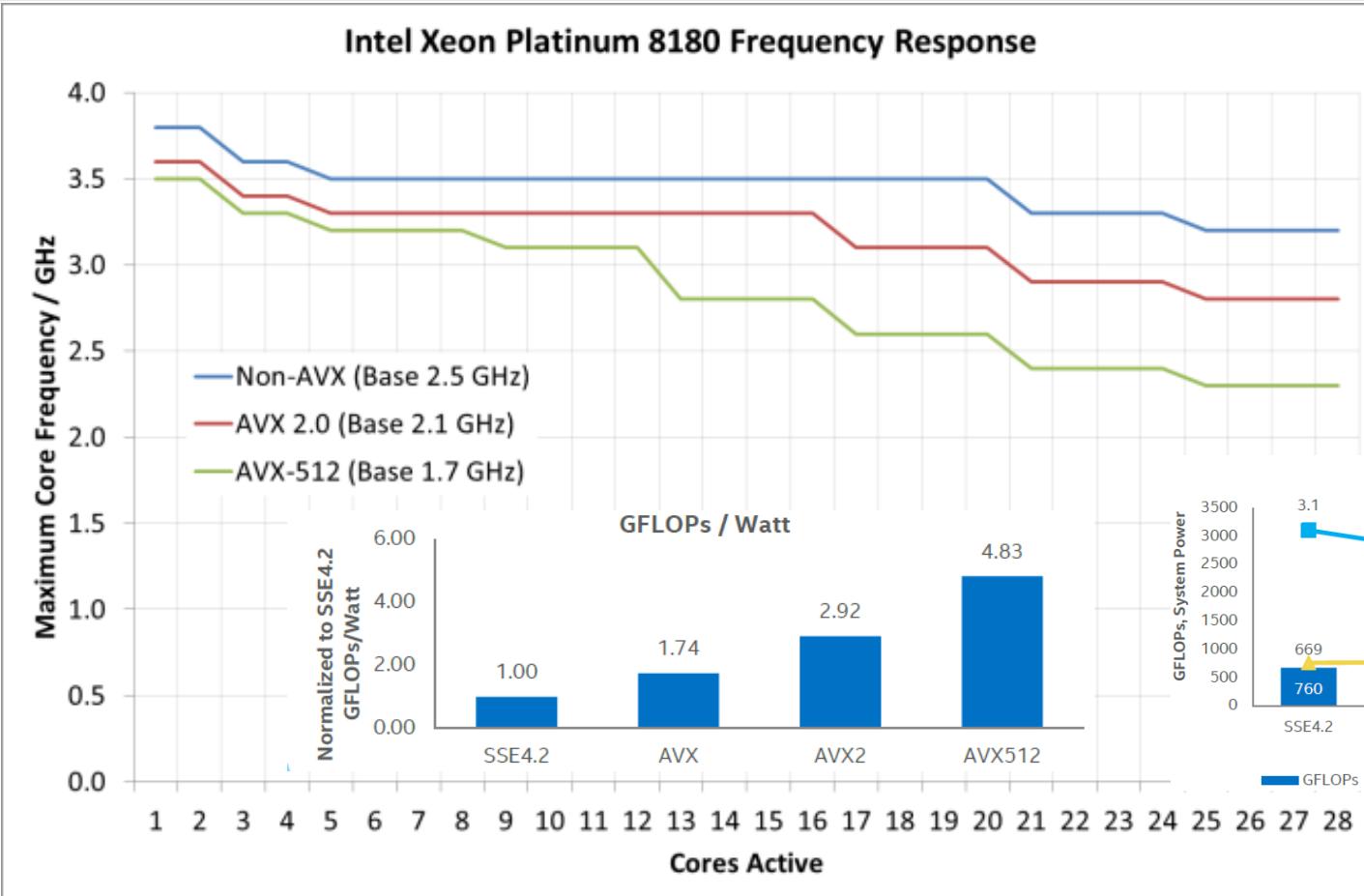


Back-end

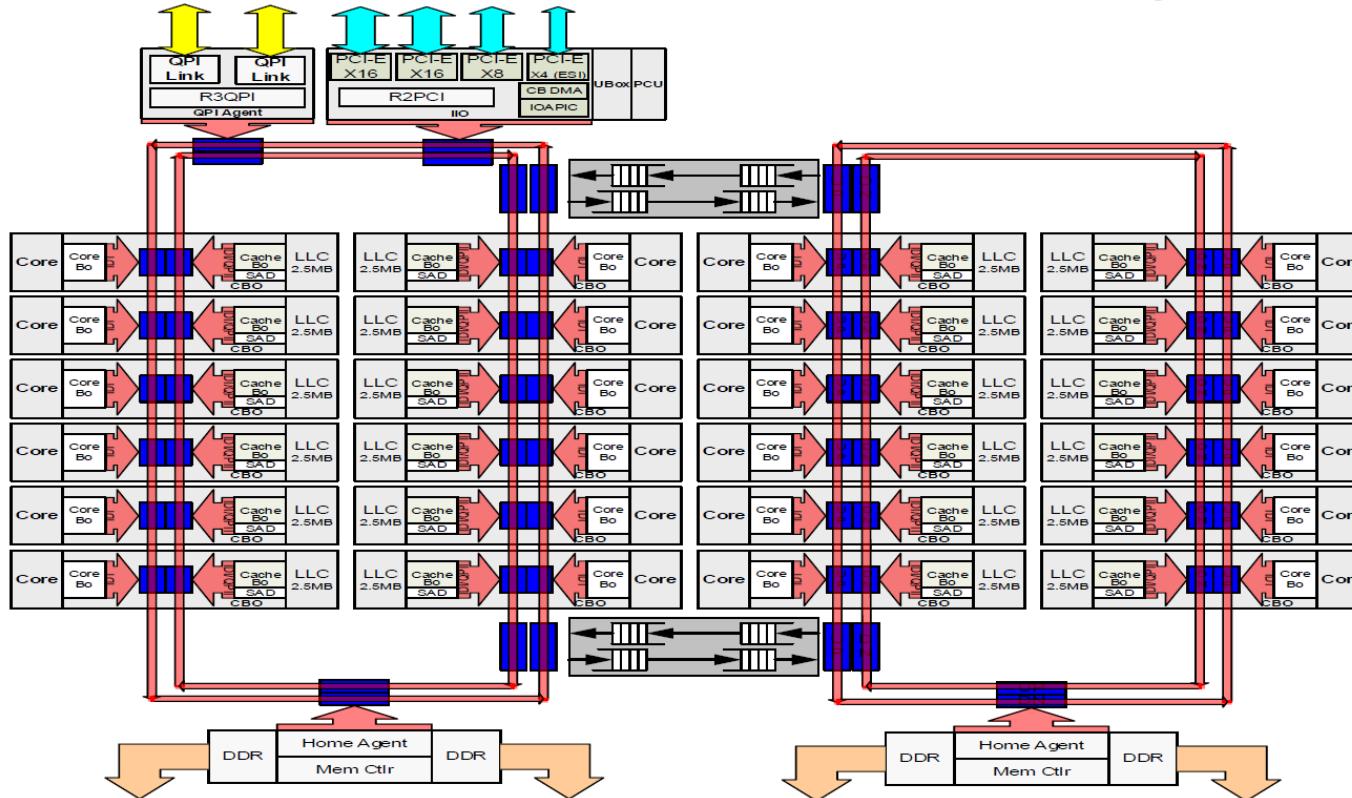
- Zvýšeno IPC o 18 %
- 8 → 10 výpočetních linek
- ROB zvětšen z 224 na 352 záznamů
- Nová AGU jednotka (4)
- Výrazně zvětšeny load/store buffery



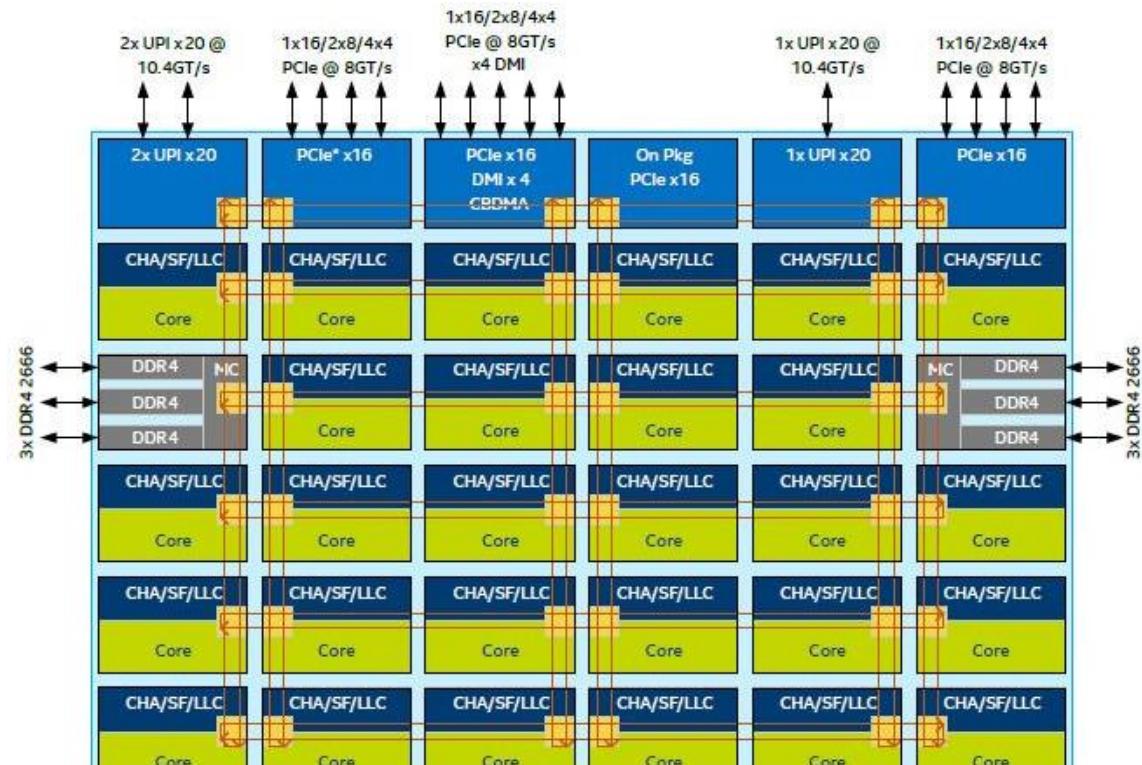
Frequency Scaling with Instruction Types



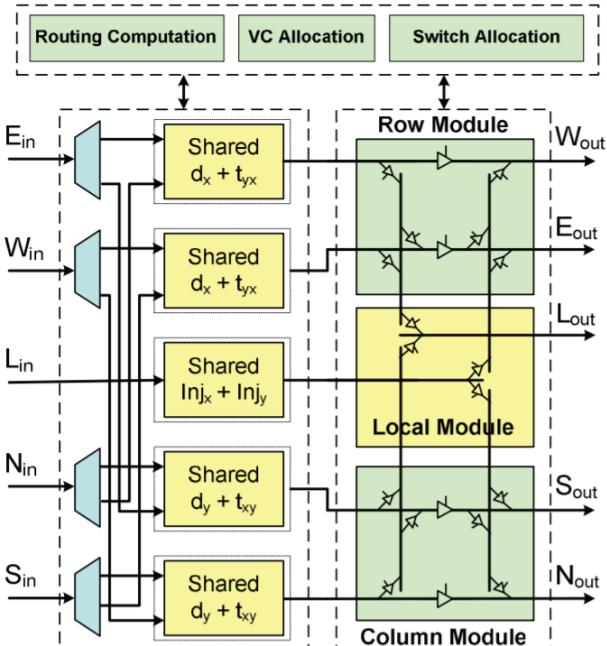
Intel® Xeon® Processor E5 v4 Product Family HCC



Zapojení do mřížky 6x6 (28 jader)

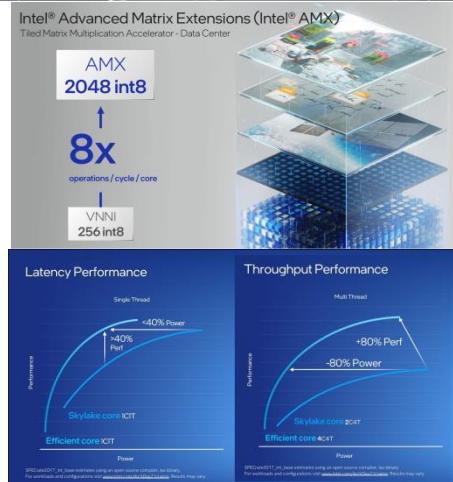
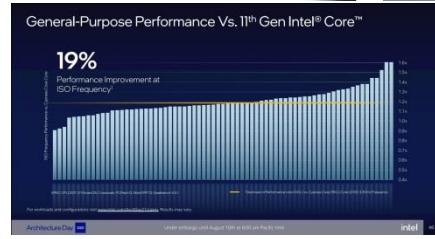
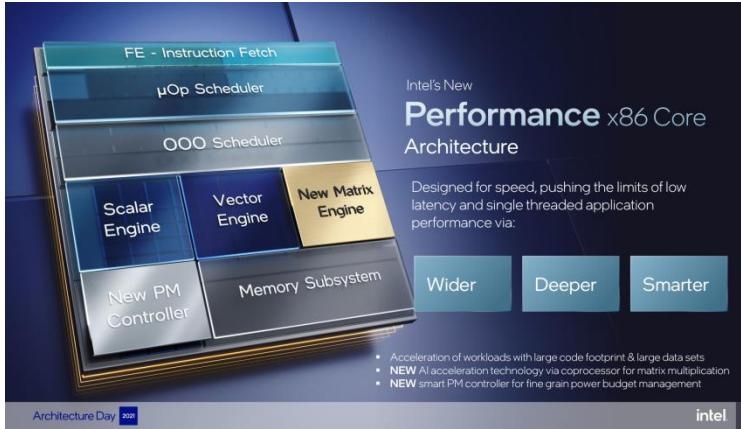
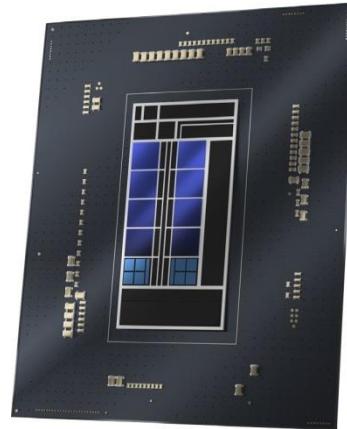


CHA – Caching and Home Agent ; SF – Snoop Filter; LLC – Last Level Cache;
Core – Skylake-SP Core; UPI – Intel® UltraPath Interconnect



(c) 5x5 MoDe-X-Single Router Design (Single Injection)

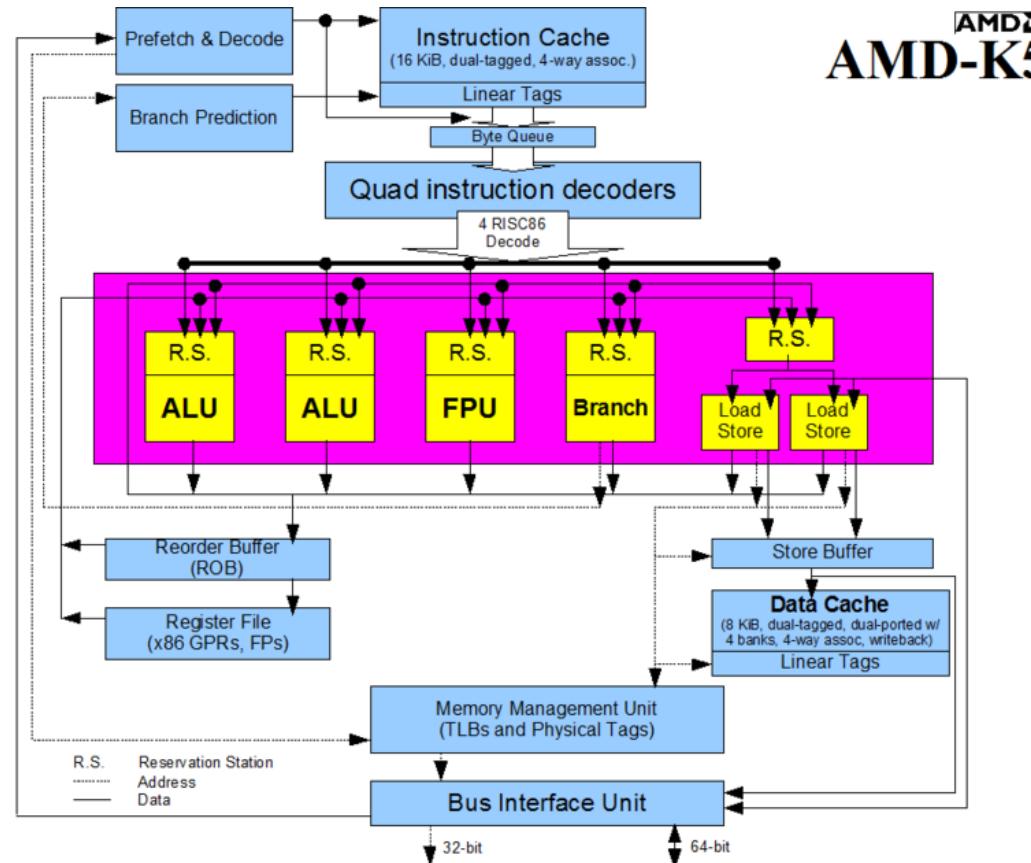
Intel Alder lake



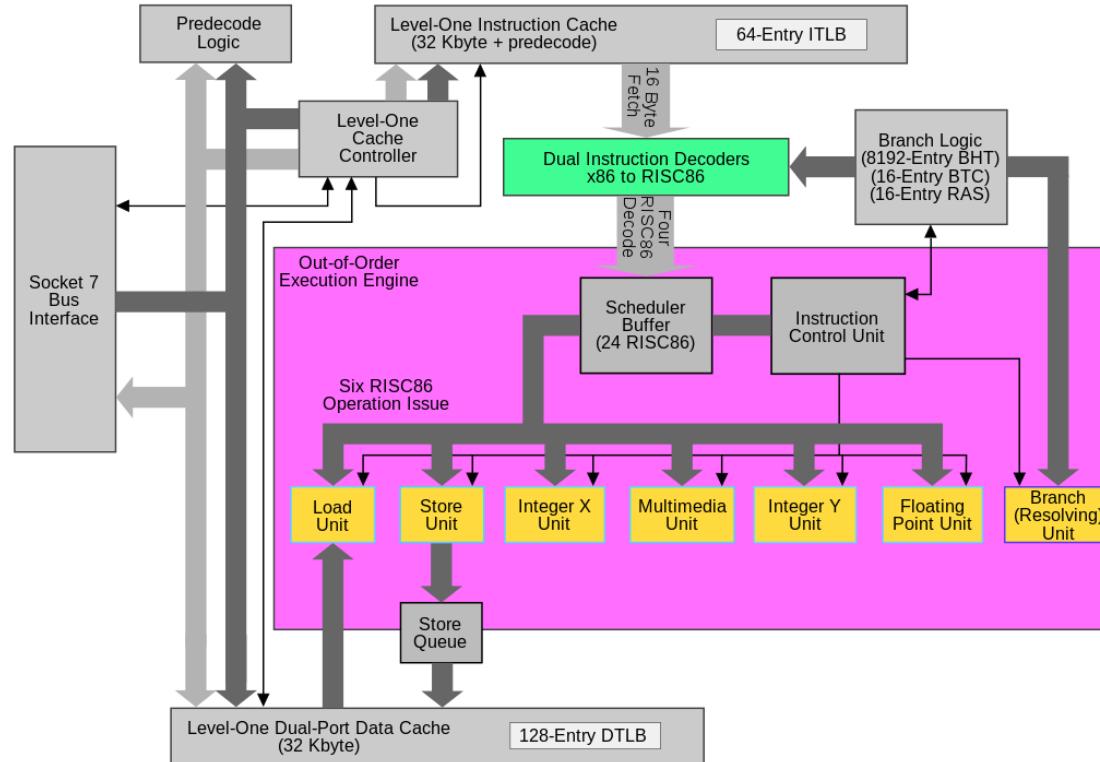
PROCESORY AMD

- **AMD K5 (1996)**

- První vlastní OOO procesor AMD
- 6 výpočetních jednotek, 4 vydání instrukce za takt, 5 stupňů linky.
- Spekulativní provádění podél 3 predikovaných větví
- Penalta 3 takty při špatné predikci
- Přejmenování registrů
- 16 KB L1, přístup do L1 v 1 taktu!
- Podpora MESI cache coherent protokolu

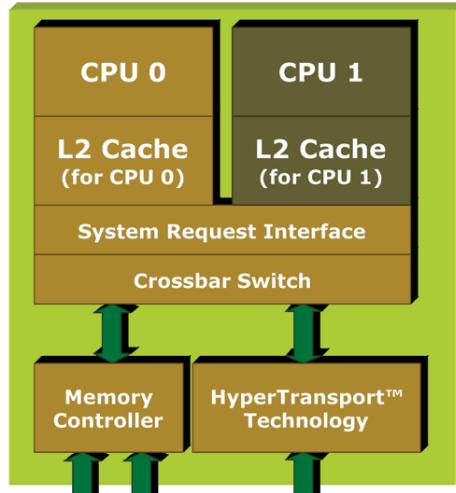
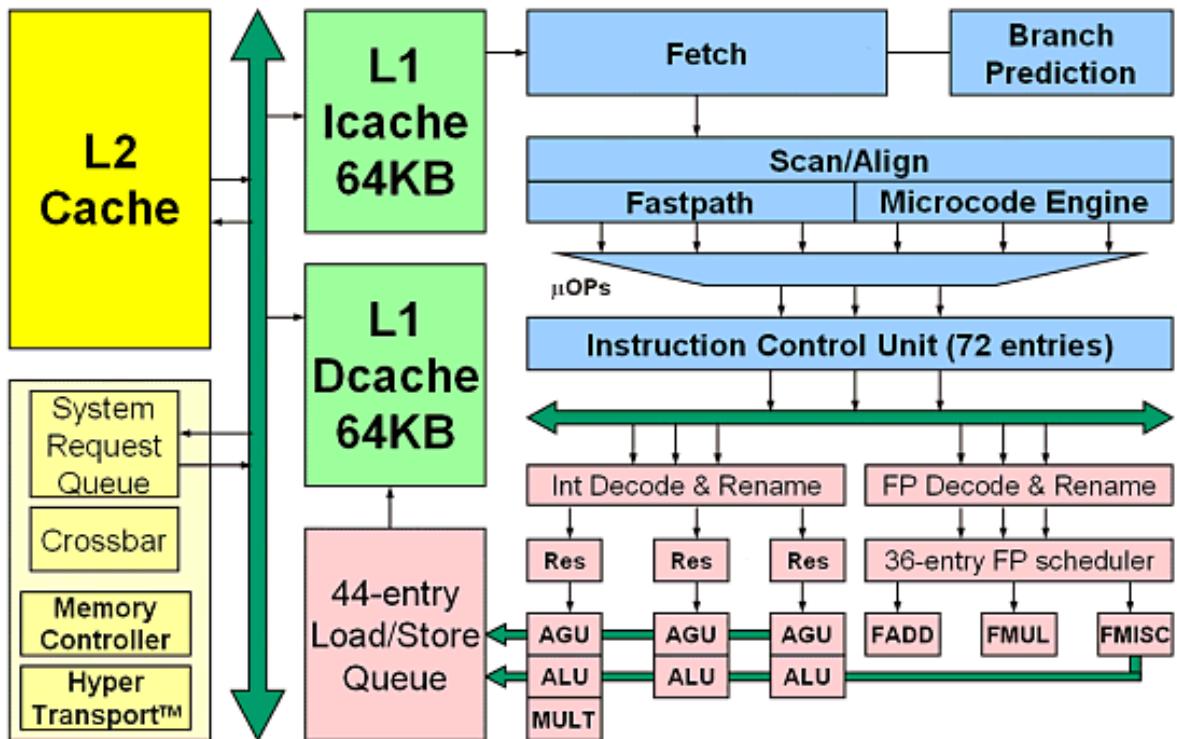


- Uvedena na trh v roce 1997
- Přináší instrukce MMX, později 3DNow

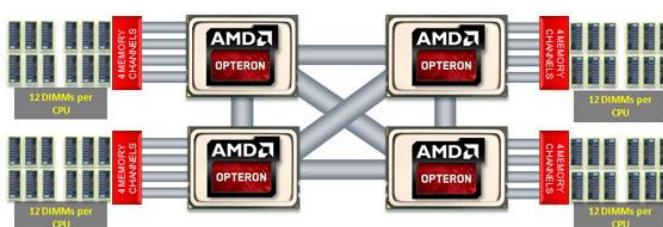


První CPU s instrukcemi x86 na 64 bitech, kompatibilní s Windows (2003), 32 bitové i 64 bitové aplikace, SW investice neznehodnoceny. Reakce Intelu: Extended Memory 64-bit Technology) EM64T a pak Intel® 64.

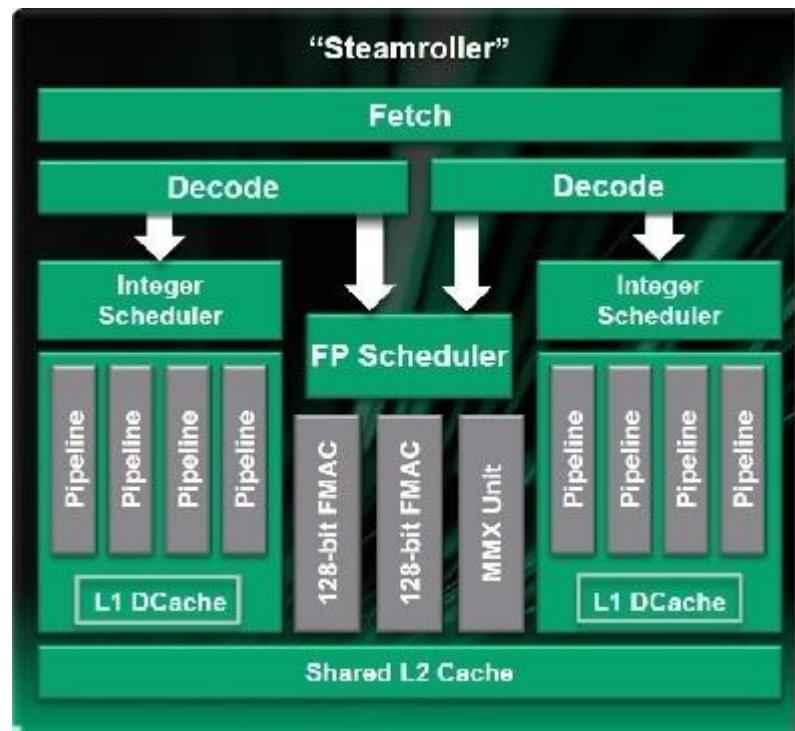
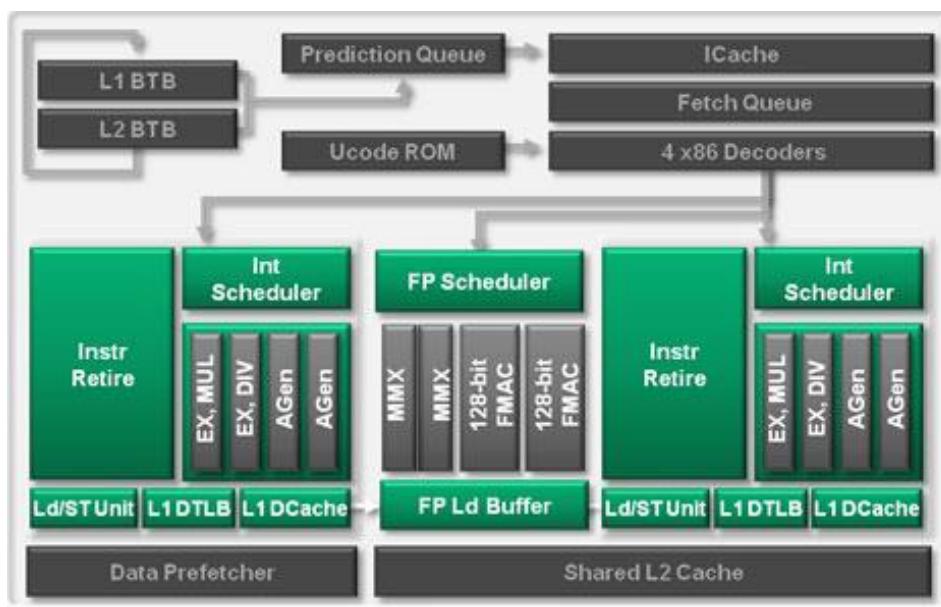
- **2, 4, 6, 8, 12, nebo 16 jader**, linka 12 stupňů, technologie SOI (Silicon on Insulator)
- **Linky Hyper Transport – 2003** (point-to-point) pro propojení s dalšími CPU (nahradily sběrnici) nebo I/O. Šířka 16 bitů, při 800 MHz to znamená 3,2 GB/s. Umožněna stavba multiprocesorů bez dalších součástek.
- **Řadič paměti DDR na čipu – 2003**, 128 bitové rozhraní na 333 MHz paměť.
- **Mikroarchitektura K10: Phenom II, 2,3,4 nebo 6 jader, 2008–12.**

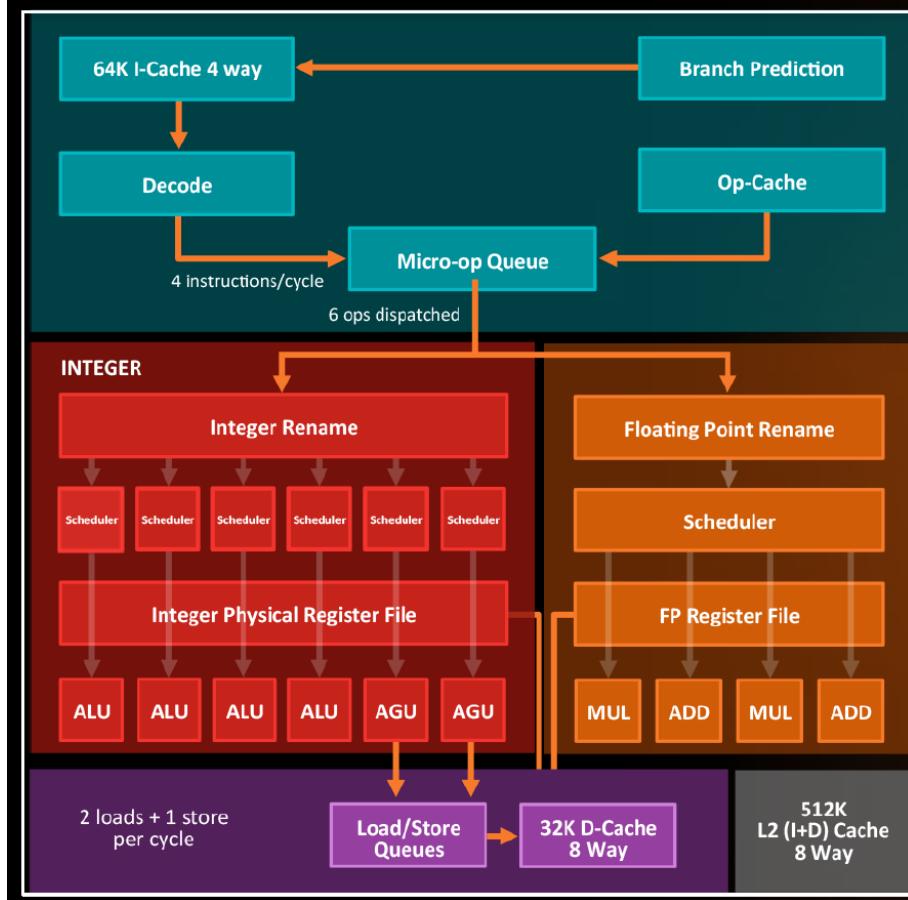


AMD Athlon™ 64 X2
Dual-Core Processor Design



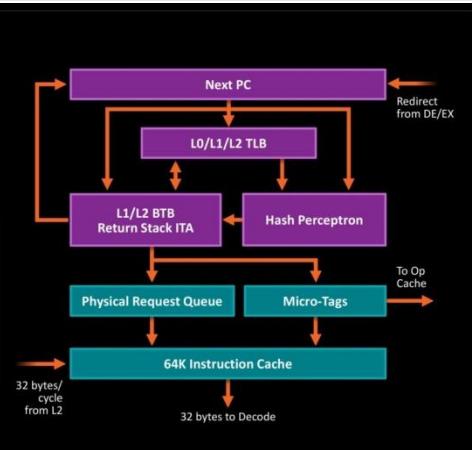
- Mikroarchitektura Bulldozer: 1 až 4 moduly se 2 jádry (tzv. CMT, Clustered MultiThreading, 1 až 2 vlákna/modul).
 - 10 až 100 W, 3,6–4 GHz, 2012.
 - Každý 2-jádrový modul sdílí L1-I cache, stupně načítání a dekódování, L2 cache, FPU.
 - Později přidány dedikované dekodéry (Steamroller)





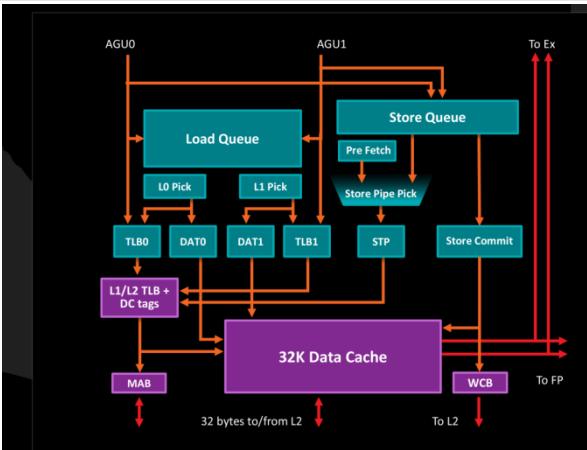
ZEN MICROARCHITECTURE

- ▲ Fetch Four x86 instructions
- ▲ Op Cache instructions
- ▲ 4 Integer units
 - Large rename space – 168 Registers
 - 192 instructions in flight/8 wide retire
- ▲ 2 Load/Store units
 - 72 Out-of-Order Loads supported
- ▲ 2 Floating Point units x 128 FMACs
 - built as 4 pipes, 2 Fadd, 2 Fmul
- ▲ I-Cache 64K, 4-way
- ▲ D-Cache 32K, 8-way
- ▲ L2 Cache 512K, 8-way
- ▲ Large shared L3 cache
- ▲ 2 threads per core



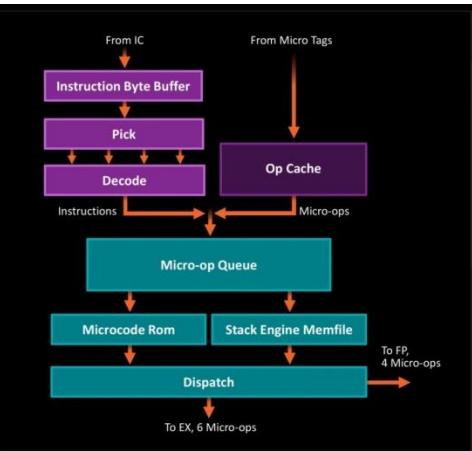
FETCH

- ▲ Decoupled Branch Prediction
- ▲ TLB in the BP pipe
 - 8 entry L0 TLB, all page sizes
 - 64 entry L1 TLB, all page sizes
 - 512 entry L2 TLB, no 1G pages
- ▲ 2 branches per BTB entry
- ▲ Large L1 / L2 BTB
- ▲ 32 entry return stack
- ▲ Indirect Target Array (ITA)
- ▲ 64K, 4-way Instruction cache
- ▲ Micro-tags for IC & Op cache
- ▲ 32 byte fetch



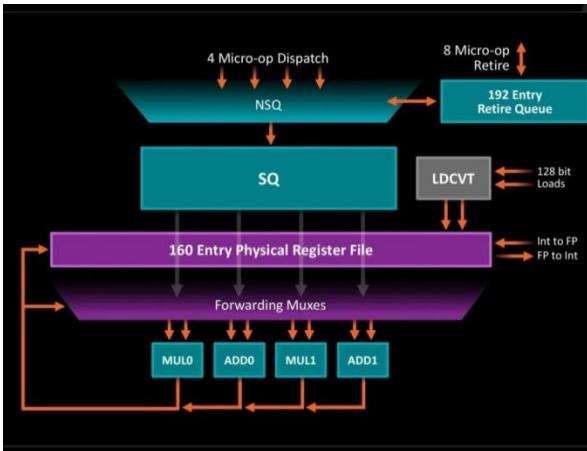
LOAD/STORE AND L2

- ▲ 72 Out of Order Loads
- ▲ 44 entry Store Queue
- ▲ Split TLB/Data Pipe, store pipe
- ▲ 64 entry L1 TLB, all page sizes
- ▲ 1.5K entry L2 TLB, no 1G pages
- ▲ 32K, 8 way Data Cache
 - Supports two 128-bit accesses
- ▲ Optimized L1 and L2 Prefetchers
- ▲ 512K, private (2 threads), inclusive L2



DECODE

- ▲ Inline Instruction-length Decoder
- ▲ Decode 4 x86 instructions
- ▲ Op cache
- ▲ Micro-op Queue
- ▲ Stack Engine
- ▲ Branch Fusion
- ▲ Memory File for Store to Load Forwarding

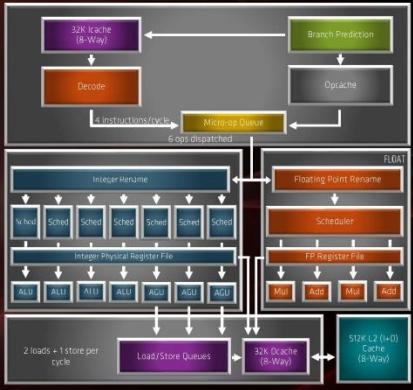


FLOATING POINT

- ▲ 2 Level Scheduling Queue
- ▲ 160 entry Physical Register File
- ▲ 8 Wide Retire
- ▲ 1 pipe for 1x128b store
- ▲ Accelerated Recovery on Flushes
- ▲ SSE, AVX1, AVX2, AES, SHA, and legacy mmx/x87 compliant
- ▲ 2 AES units

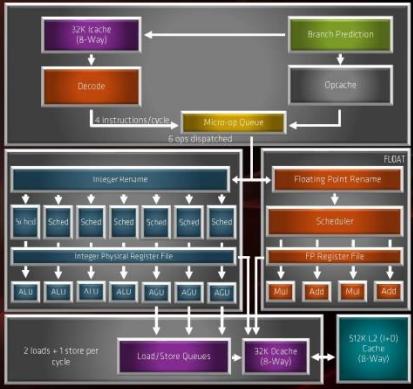
"ZEN 2" MICROARCHITECTURE OVERVIEW

- 2 threads per core (SMT) carried forward
- New TAGE branch predictor
- Larger Micro-Op Cache, now 4K instructions
- Larger L3 cache, now 2X "Zen" and "Zen+"
- 4 integer units
 - Large rename space ~ 180 registers
 - Increased AGUs from 2 to 3
- 3 AGENs per cycle
- 2 loads and 1 store per cycle
- 2 floating point units x 256 Fmacs
 - built as 4 pipes, 2 Fadd, 2 Fmul
 - Now supports single-op AVX256



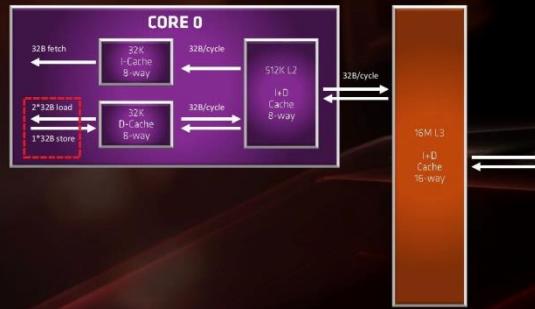
"ZEN 2" MICROARCHITECTURE OVERVIEW

- I-cache 32k, 8-way
- D-cache 32k, 8-way
- L2 cache 512k, 8-way
- TLBs
 - L1 64 entries I & D, all page sizes
 - L2 512 I, 2K D, everything but 1G
- Faster Virtualization Based Security
 - With Guest Mode Execute Trap
- Hardware-enhanced Security mitigations



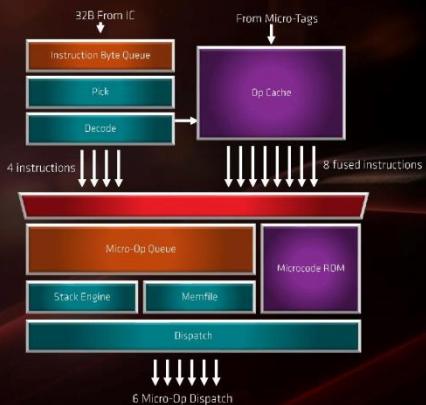
"ZEN 2" CACHE HIERARCHY

- Doubled L1 load/store bandwidth over Zen
- Improved L1 and L2 prefetch throttling
- Fast private 512K L2 cache
- Fast shared L3 cache
- High bandwidth enables prefetch improvements
- L3 is filled from L2 victims
- Fast cache-to-cache transfers
- Large Queues for Handling L1 and L2 misses



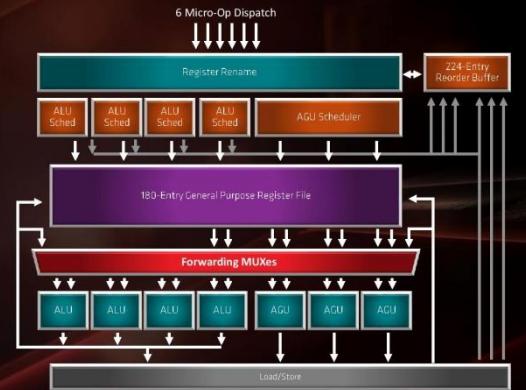
DECODE

- Op cache improvements
- Doubled capacity to 4K fused instructions
- Better instruction fusion
- Increased effective throughput



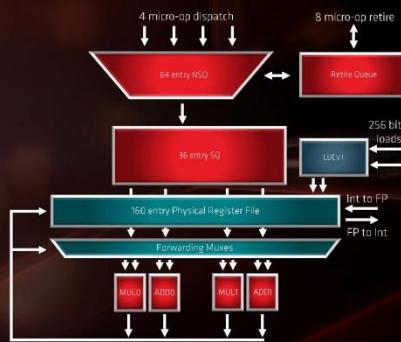
INTEGER EXECUTION

- 92 entry integer scheduler, up from 84
- 4, 16-entry ALU queues
- 1, 28-entry AGU queue
- 180 entry physical register file (up from 168)
- 7 issue per cycle, up from 6
- 4 ALUs, 3 AGUs
- 224 entry ROB, up from 192
- Improved SMT fairness for ALU and AGU schedulers
- Watermarked ALU tokens to manage spinlocks



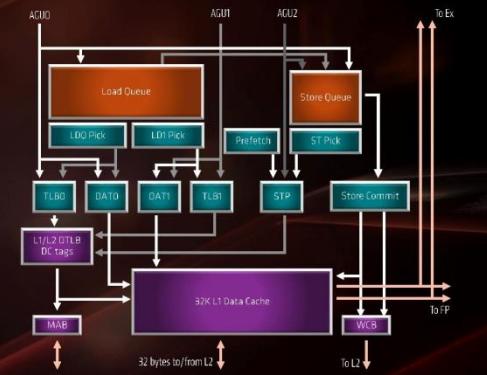
FLOATING POINT UNIT

- Doubled Floating Point & Load Store bandwidth from 128b to 256b
- Improved performance for instructions using 256b ymm registers which are generated by AVX intrinsics or /arch:[AVX|AVX2] compiler flags
 - Faster inline memcpy & memset
 - Faster physics simulation
 - Faster audio effects processing (Microsoft™ XAudio2_9)
- Improved mul latency from 4 to 3 cycles



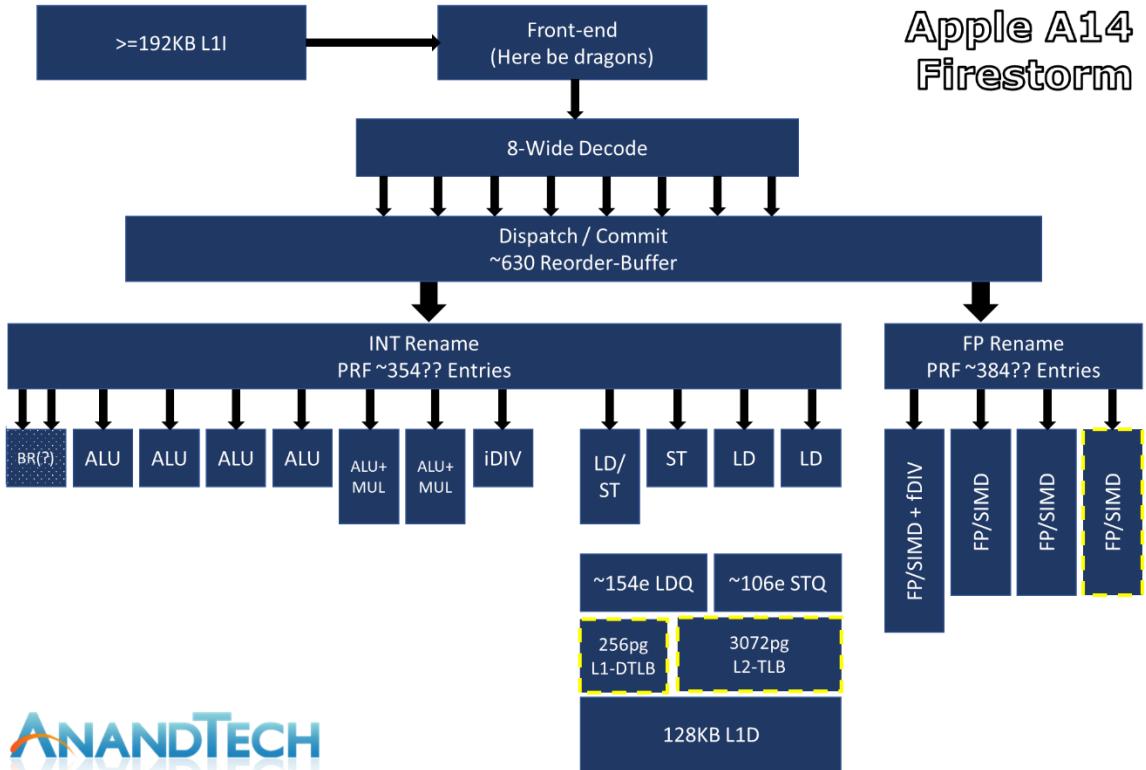
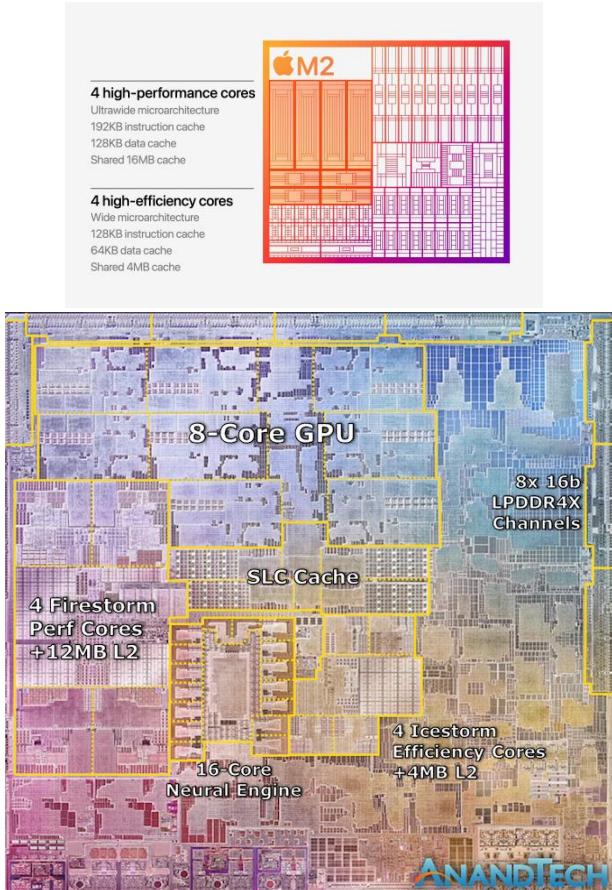
LOAD/STORE

- 48 entry store queue, was 44
- 2K entry L2 DTLB, 1G as 2M, was 1.5K no 1G
- Improved L2 DTLB latency
- 32KB, 8-way L1 data cache
 - Two 256-bit reads
 - One 256-bit write
 - 64B load, 32B store alignment boundaries
- Increased Load/Store bandwidth to 32B/clk (up from 16B/clk)
- Faster string copy and float-point point performance
- Improved write-combining buffer performance
 - While using multiple streams, the hardware avoids closing buffers before they are completely full
- Improved prefetch throttling



PROCESORY APPLE/ARM

Apple M1 – ARM (RISC) procesor



Pokračování příště