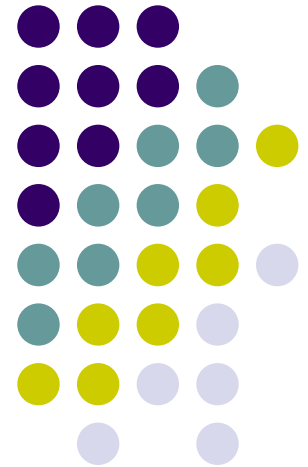
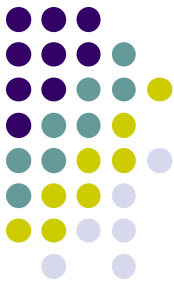


Vícejádrové procesory

Intel Core



Intel Core



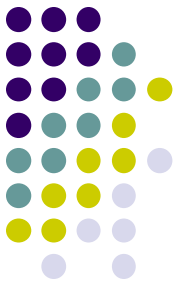
- Architektura **Intel Core** představuje razantní změnu dříve nastoleného kurzu - Intel chtěl mít až *10 GHz* procesory založené na architektuře **NetBurst**, která se komerčně prodávala pod značkou **Pentium 4 a Pentium D**
- Rozhodnutí Intelu jít cestou **vysokých frekvencí** se ale ukázalo jako nešťastné
- Problém velké **energetické náročnosti** byl stále více zřetelný - 3 GHz Pentium 4 a pozdější modely Prescott spotřebou překračovaly **100W**
- Bylo stále více zřejmé, že architektura nazvaná **NetBurst** je dlouhodobě neudržitelná
- Celá technologie NetBurst byla vlastně navržena s jediným cílem - **dosahovat vyšších frekvencí než konkurence**, protože zákazníci laici dají spíše na hodnotu taktovací frekvence než na skutečný výpočetní výkon procesoru – šlo o marketingový trik
- AMD ale začal své procesory označovat ekvivalentní frekvencí (**Performance rating**)

Intel Core

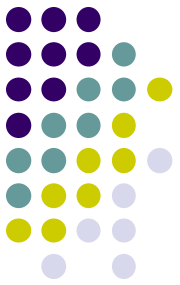


- **Intel Core** nemá se svými předchůdci Pentium 4 a Pentium D společného skoro nic (dokonce ani hyperthreading, ten se znovu objeví až u další generace procesorů)
- Nové jádro je založeno na procesorech **Pentium M**, které jsou zase odvozeny od původních **Pentii 3** a ta z procesoru **Pentium Pro**
- Hlavní snahou při vývoji nového jádra bylo **snížení el. příkonu** a Intel opouští od honby za dosažením co nejvyšších frekvencí (i při nižší frekvenci lze pracovat výkonně a při tom méně „topit“)
- Proti architektuře **Netburst** byla zkrácena instrukční **pipeline na 14 stupňů**
- Procesory architektury Intel Core proto nedosahují pracovních frekvencí předchozích procesorů (PentiumD, Pentium4), jejich reálný výpočetní výkon je ale vyšší
- Snížením frekvence a vylepšeným řízením spotřeby bylo omezeno produkované teplo

Intel Core



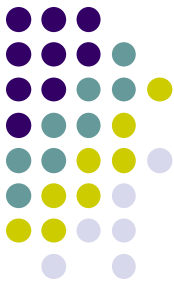
- **Intel Core** používá **RISC** jádro s vlastními microOPs (to není nic převratně nového)
- Komplexní **CISC** instrukce x86 jsou po načtení z instrukční cache přeloženy do vnitřních mikrooperací
- **MicroOPs** jsou vnitřní malé instrukce, které vstupují do pipeline (přitom několik instrukcí vstupuje paralelně do **4 pipeline** a navíc v promíchaném pořadí **out-of-order**)
- V jednom taktu tak lze dokončit **4 instrukce současně** (všechny ostatní procesory dosud umí pouze 3)
- Některé CISC instrukce jsou přeloženy na jedinou microOP jiné na více mikroOPs
- Architektura **Intel Core** nově umožňuje dokonce dvojici x86 instrukcí přeložit společně na efektivnější microOP – tomu se říká **MacroFusion** (dvě po sobě jdoucí navazující instrukce vytvoří kombinovaný mikrokód – je to podobné jako když do cizího jazyka přeložíte význam celé věty a ne jen postupně jednotlivá slovíčka)



Intel Core

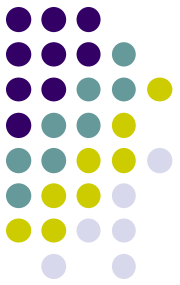
- **Out-of-Order** - Umožňuje zpracování instrukcí v jiném pořadí, než jak je v programu stanoveno, aby se co nejvíce využilo dostupného výpočetního výkonu
- Toto umělo poprvé už Pentium PRO
- Out-of-Order zpracování probíhá pomocí mikrooperací, což jsou jednoduché RISC operace, na které se převádějí instrukce předstírané instrukční sady x86, SSE atd.
- Překladem vzniklé mikrooperace putují do **Reorder Buffer**, což je paměť, v níž se microOPs přeřadí tak, aby je bylo možné vykonávat co možná nejvíc paralelně.
- Velikost tohoto bufferu se udává jako **in-flight microOPs** (někdy také in-flight (micro)instructions nebo instruction window)
- Jde vlastně o počet MicroOPs, které jsou dohromady k dispozici k přeskupení pořadí
- Tento počet je jedním z největších faktorů výpočetního výkonu
- při velkém počtu in-flight MicroOPs lze obcházet závislosti, které znemožňují paralelismus

Intel Core



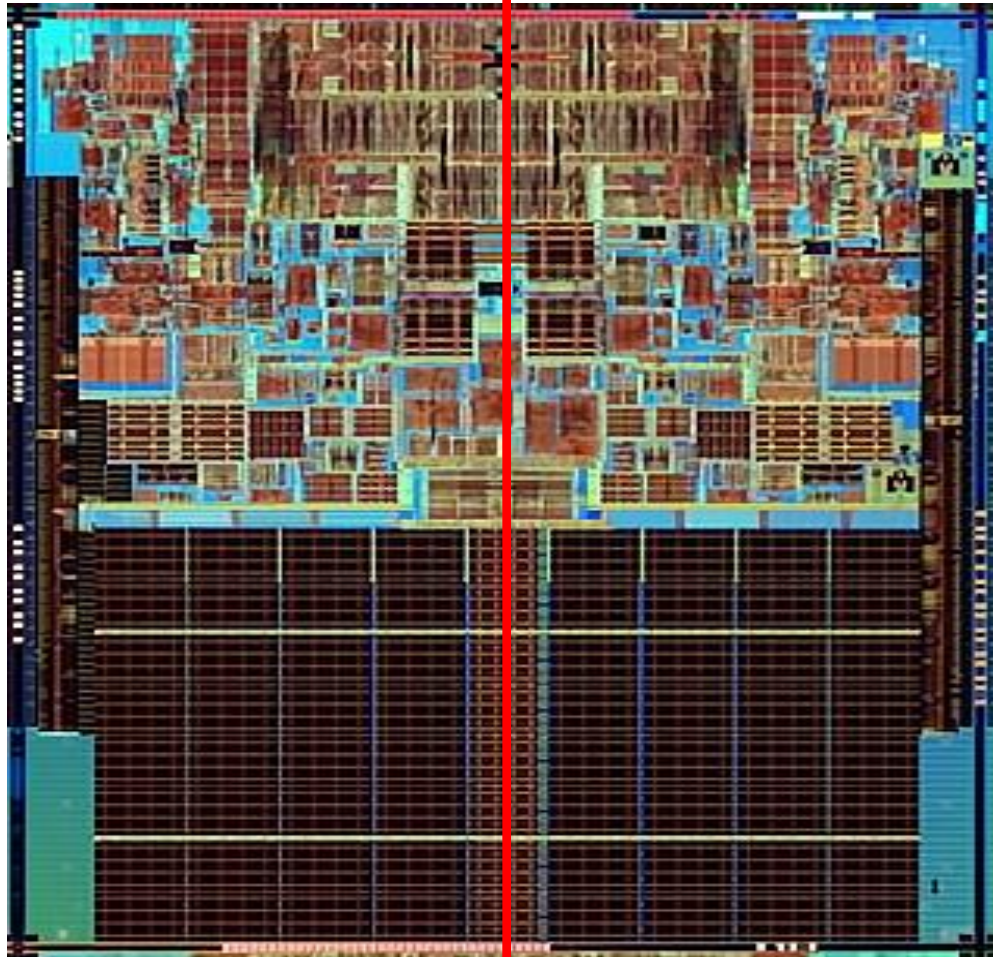
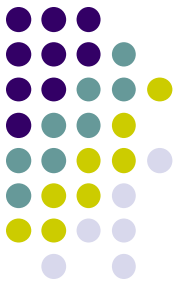
- Limitace paralelismu plyne z toho, že v nějaké právě prováděné části programu se řeší problém, ve kterém se často nachází značné datové závislosti, takže povely nelze vykonávat najednou.
- Pokud procesor není schopen přeřadit pořadí hodně instrukcí, je tímto silně degradován jeho výkon.
- Pokud má ale velké zdroje (zásobárnu MicroOPs), je schopen pracovat najednou třeba na několika různých podproblémech, které jsou v programu napsané postupně za sebou. Díky tomu pak optimálně využívá svůj potenciál
- Architektura Core umožňuje v jednom okamžiku přeřazování až **96** připravených microOPs, které budou vykonávány po čtyřech naráz
- Po přeřazení putují microOPs do **Reservation Station**, kde čekají na dodávku dat k nim přiřazených. Počet takto čekajících instrukcí činí až **32**
- Poslední fází zpracování microOPs je jejich opětovné přeřazení do původního pořadí (Retirement).
- Procesor posune programový čítač, čímž dá navenek najevo, že všechny instrukce do daného místa v programu byly dokončeny

Intel Core



- Procesory jsou vyráběny stejnou technologií jako předchozí Pentium4 a Pentium D – rozměr tranzistoru 65 nm
- Kódové označení jader první generace (rok 2006) založených na architektuře Intel Core jsou **Conroe** pro stolní počítače, **Merom** pro mobilní procesory a **Woodcrest** pro serverové Xeony
- Všechny mají stejnou architekturu. Liší se jen rozměrem pouzdra, rychlostí sběrnice a spotřebou energie
- **Conroe** je kódové označení pro rozsáhlou skupinu procesorů.
- Byly prodávány jako
 - Core 2 Duo, Intel Core Xeon (výkonnější varianty, L2 Cache 2 až 4 MB)
 - Pentium Dual-Core (1 MB L2 Cache)
 - Intel Core Celeron (512 KB L2 Cache)
 - Intel Core Solo – jednojádrový procesor (kódové označení Conroe-L)

Intel Core – jádro Conroe



1.jádro

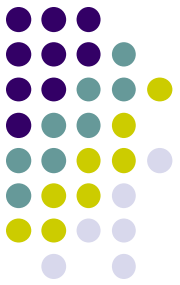
2.jádro

Intel Core



- V roce 2007 se objevují nová jádra
- Tato řada procesorů je již vyráběna modernější technologií 45 nm
- **Kentsfield** - čtyřjádrové procesory, které byly fyzicky spojení dvou procesorů Conroe, L2 cache 8 MB
- **Clovertown** – čtyřjádrové procesory vzniklé složením dvou čipů Woodcrest
- Tyto 4-jádrové procesory se prodávají jako **Intel Core 2 Quad**

Procesory x86



- Architektura Intel Core je 8. generací procesorů rodiny x86

- 1. 8086
- 2. 80286
- 3. 80386
- 4. 80486
- 5. Pentium
- 6. Pentium PRO
 - Pentium MMX
 - Pentium II
 - Pentium III
 - Pentium M
- 7. Pentium IV
- 8. Intel Core

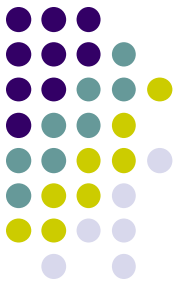


Procesory Intel Core

- Všechny další mikroprocesory až do současnosti se označují jako **IntelCore**

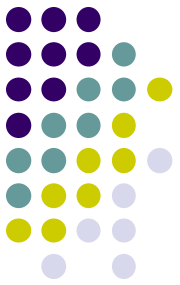
Procesory **Intel Core** pak mají několik generací

- Intel Core (2006)
- Penryn (2007)
- Nehalem (2007)
- Westmere (2008)
- SandyBridge (2010)
- IvyBridge (2012)
- Haswell (2013)
- Broadwell (2014)
- Skylake (2015)
- Kaby-Lake (2017)
- Coffee-Lake (2018)
- Cannon-Lake (2019)
- Ice-Lake (2020)
- Rocket-Lake (2020)
- Tiger-Lake (2020)
- Alder-Lake (2021)
- Raptor-Lake (2022)



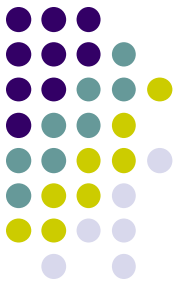
Architektura Nehalem

- Architektura **Nehalem** – v roce 2007 podstatné zdokonalení původní architektury Intel Core
- 4-jádrový procesor obsahuje **731 milionů** tranzistorů (včetně cache)
- **Turbo Mode** pro zvýšení výkonu **singlethreaded** aplikací (přetaktování za běhu)
- nové instrukce **SSE 4.2**
- procesor je silně zaměřen na **multithreading** - každé jádro Nehalemu je schopné zpracovávat dvě vlákna (čtyřjádro tedy zvládne osm threadů)
- Intel tvrdí: „*Nehalem přináší stejné jádro pro všechny segmenty trhu, vysoký výkon a škálovatelnost při zachování spotřeby minulé generace procesorů*“
- **Škálovatelnost** – možnost výroby různě výkonných konfigurací procesoru se stejným jádrem Nehalem (různá Cache, frekvence, příkon...)
- **FSB** sběrnici nahrazuje **QPI** (QuickPath Interconnect), která je full-duplexní
- Výkon **QuickPath** sběrnice je 4,8 až 6,4 Gtakt/s na linku (možná konfigurace je 5, 10 a 20 bit), s celkovou propustností až 25,6 GB/s.



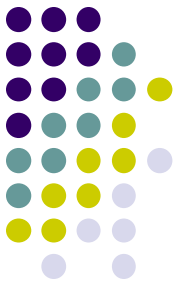
Intel Core Nehalem

- **16-stupňový pipelining** – počet stupňů vzrostl jen kvůli větším rozměrům chipu, aby signál stíhal doputovat během jednoho taktu z jedné sekce procesoru do další
- Nehalem podědil **fúzování makroinstrukcí** po IntelCore a dále jej rozšiřuje
- Přibylo fúzování i v **64bit režimu** (Core 2 umí pouze 32bit)
- **LSD (Loop Stream Detektor)** - umožňuje odhalit cyklus v probíhajícím kódu a zabránit tak opakovanému dekódování instrukcí a predikci větvení.
- LSD zvyšuje výkon v průběhu cyklu a na jeho základě také případně vypíná nepotřebnou logiku (jestliže se cyklus např. 1000x opakuje, dopředu víme, které jednotky procesoru budou kdy potřeba zapnout a které lze kdy vypnout)



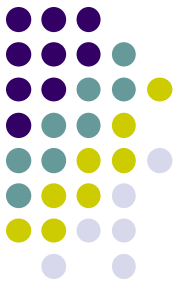
Nehalem – inkluzivní L3 cache

- Intel Core 2 obsahoval standardně 2 jádra s vlastní L1 cache, která mají L2 sdílenou mezi obě jádra (velikosti obvykle 6 MB)
- Nehalem má cache:
- **32 kB L1 instrukční** a **32 kB L1 datové** cache (zpoždění 1T)
- **L2 Cache** má kapacitu sniženou na pouhých 256 kB na jádro (takže je 24x menší než u Intel Core – zpoždění 4T)
- Nově je v Nehalemu velká **L3 cache** (8MB u čtyřjádrové varianty) sdílená pro všechna jádra - zpoždění 10T
- **inkluzivní cache** - obsah L1 je obsažen v L2 cache a L2 je obsažen v L3



Inkluzivní x Exkluzivní cache

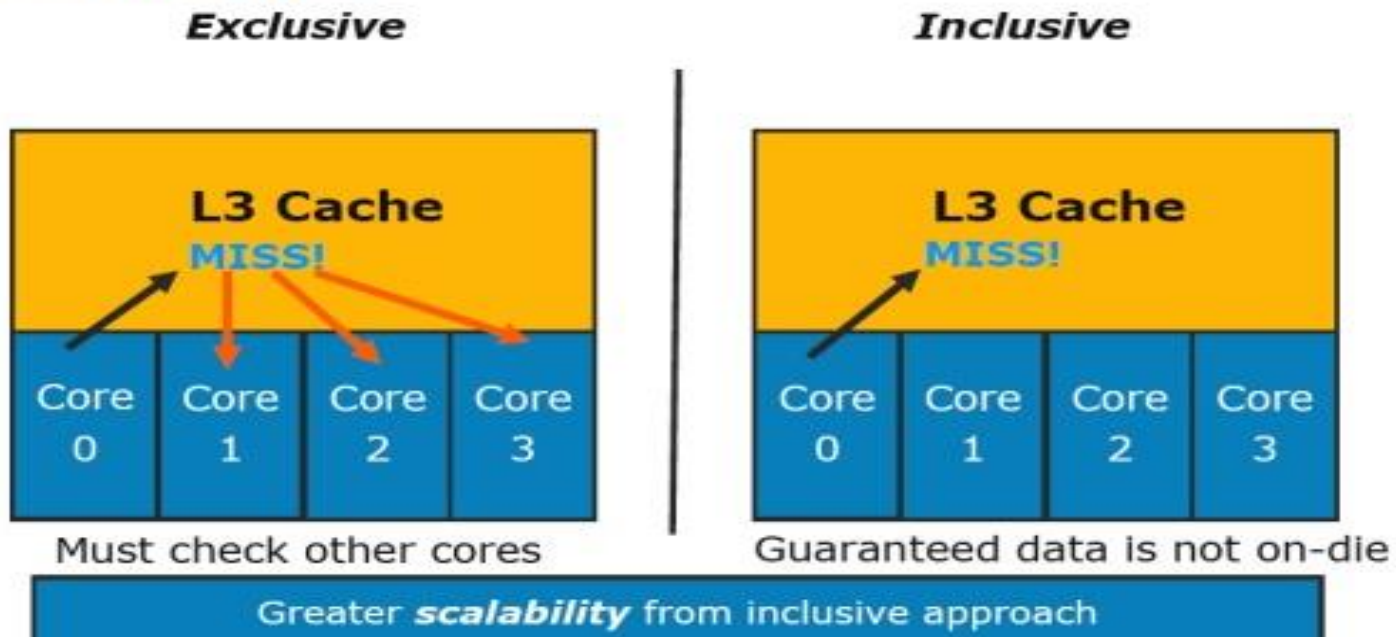
- Pro srovnání **AMD** preferuje přístup **exkluzivní cache**, takže data nejsou v cache uložena duplicitně a celkem tak lze cachovat více dat
- Hlavní nevýhoda exkluzivní cache spočívá v tom, že cache se chovají jako „nepropojené“. Pokud nejsou data nalezena ve velké sdílené cache, pak to ještě neznamená, že v některé z lokálních cache jiných sousedních jader se hledaná data nenachází. Takže je třeba všechny je projít, což stojí čas navíc
- K situaci, že jedno z jader chce pracovat s daty cachovanými v sousedním jádře by ale u správně napsaného vícevláknového programu nemělo docházet (každé vlákno a tím pádem i každé jádro by mělo pracovat s úplně jinými daty), takže prohledání cache sousedních jader lze dělat opožděně spekulativně, až po vydání požadavku na čtení dat z hlavní operační paměti při jejich nenalezení ve sdílené cache



Inkluzivní a exkluzivní cache

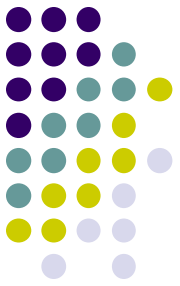
- Příklad:
- potřebná data požadovaná v jádru 0 nejsou v L1 ani L2 cache
- Data nejsou nalezena ani v L3 Cache (po cca 10 taktech) = „cache miss“
- Zatímco v případě exkluzivní cache se bude zjišťovat stav v cache u jader 1 až 3, v případě inkluzivní cache lze rovnou generovat požadavek do RAM

Inclusive vs. Exclusive Caches – Cache Miss



Nehalem

Memory controller



- O přístup k paměti RAM se u Nehalemu stará **integrovaný paměťový řadič** (Memory controller)
- Dříve byl umístěn v **severním můstku** (což je část chipsetu na základní desce)
- Mikroprocesor komunikuje s paměťovými moduly **přímo** a ne prostřednictvím dalších chipů na základní desce
- Integrace řadiče paměti do procesoru má za následek značné zrychlení přístupu k RAM paměti

Power management

Self monitoring

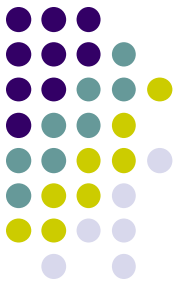


- Nehalem je první procesor Intelu, který zavádí **nezávislé napájení některých částí**. Celý čip je nyní rozdělen do třech oblastí: výpočetní jádro, paměťový řadič, zbytek. Každá z těchto částí má přitom svoje vlastní napětí
- Novinkou u Nehalemu je **Power Gating**
- Část procesoru, která aktuálně nic nedělá, by také neměla spotřebovávat elektrickou energii
- Bohužel i nepracující zavřené tranzistory spotřebovávají energii, protože přes ně uniká proud **leakage current** – tranzistory u moderních technologií jsou tak malé, že přes ně elektrony mohou „tunelovat“
- **Power Gate** umožňuje efektivně ze systému zcela odpojit neaktivní jádro a dále tak snížit spotřebu (přes tranzistory v odpojené části procesoru žádný proud neuniká)
- Nehalem obsahuje řadu senzorů pro zjištění teploty jádra, napětí a proudu a také spotřeby
- **PCU** (Power Control Unit) - řídicí jednotka spotřeby (samostatný mikrořadič, který obsahuje cca 1 milion tranzistorů)
- **PCU** je schopen velice sofistikovaně řídit dle potřeby napájecí napětí a frekvenci jednotlivých jader

Turbo

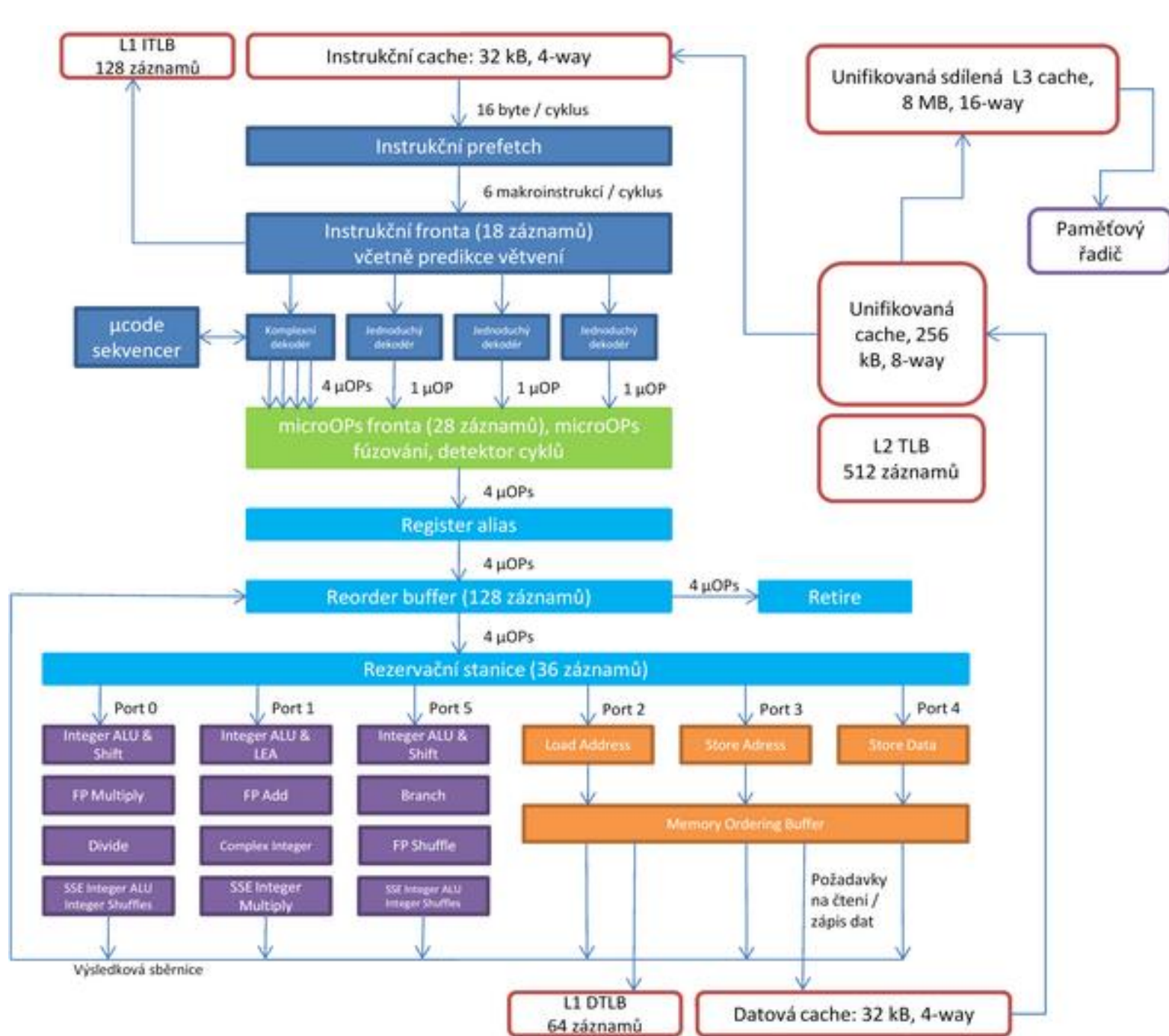


- **Turbo mode** nebo **Turbo boost**
- Název pro technologii, kterou mají všechny moderní vícejádrové procesory
- Pokud není zpracovávána vícevláknová aplikace, některá jádra vícejádrového procesoru jsou **nezatížená**
- Nezatížená jádra může PCU zcela **vypnout**
- Elektrický příkon procesoru s vypnutými jádry zdaleka nedosahuje hodnoty TDP
- Zapnutá jádra je tak možné přepnout do „**Turbo režimu**“ – zvýšit frekvenci jejich běhu nad **nominální frekvenci** procesoru a přesto se procesor **nebude přehřívat**
- U některých moderních vícejádrových procesorů pak často bývá úvaděna **nominální frekvence** procesoru při běhu všech jader a dále také **maximální možná frekvence** v **Turbo** režimu, na které může běžet jedno jediné jádro, když ostatní budou vypnutá



Turbo mode

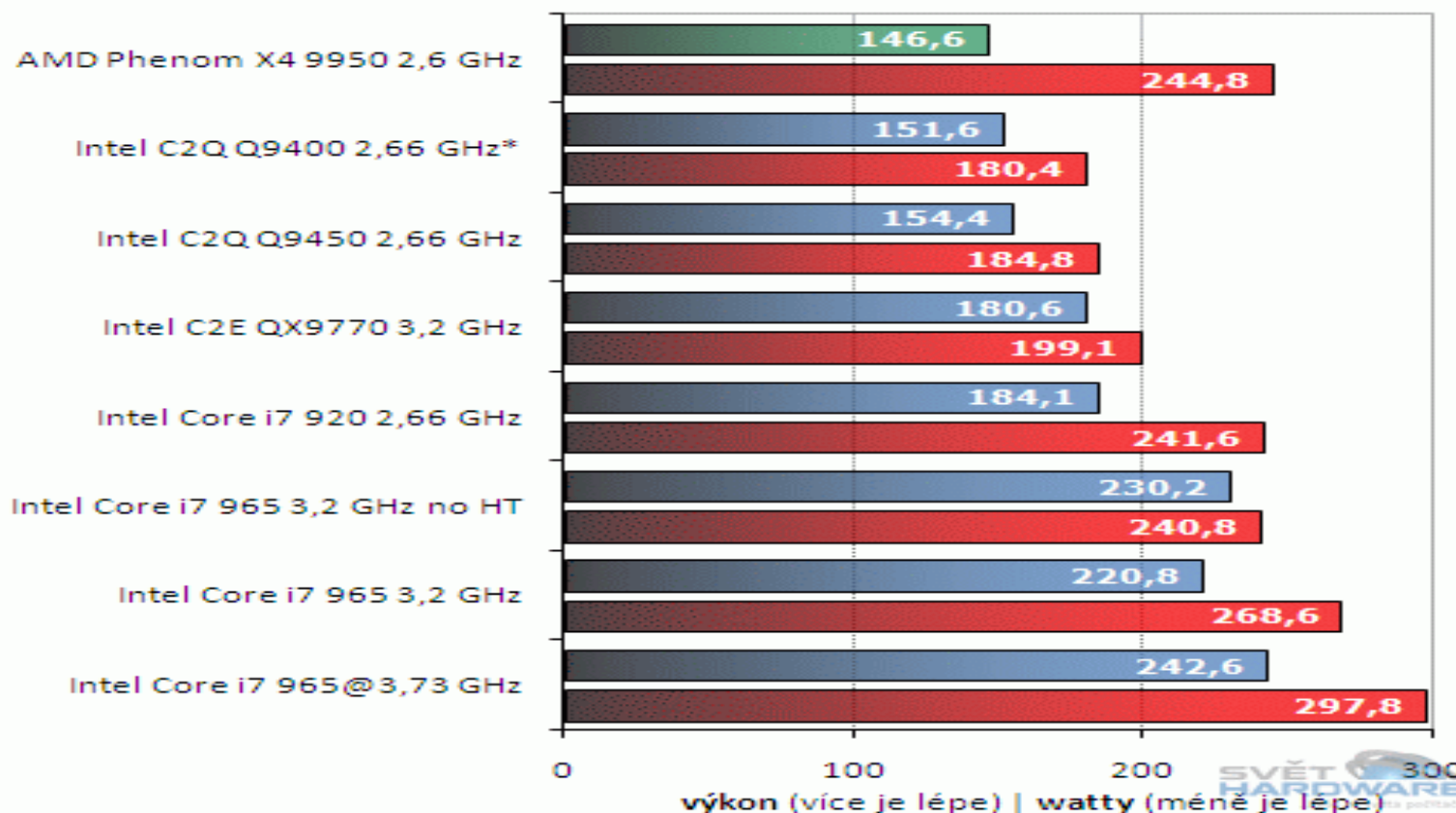
- Dvoujádrový procesor s TDP 40 W bude mít při vypnutí jednoho jádra příkon okolo 20 W
- Bohužel to ale neznamená, že frekvenci toho jednoho zapnutého jádra pak lze zvýšit na dvojnásobek
- Při dvojnásobném zvýšení frekvence vzroste totiž příkon mnohem více než na dvojnásobek
- Frekvenci nelze zvednout ani právě tak, aby příkon vzrostl přesně na 40W, přestože procesor má TDP 40 W
- Problém by byl v tom, že těchto 40W tepla by vznikalo na velmi malé ploše jednoho jádra (poloviční oproti ploše celého procesoru)
- TurboBoost tedy umožňuje krátkodobé zvýšení frekvence jednoho jádra maximálně o pár desítek procent



Srovnání výpočetního výkonu



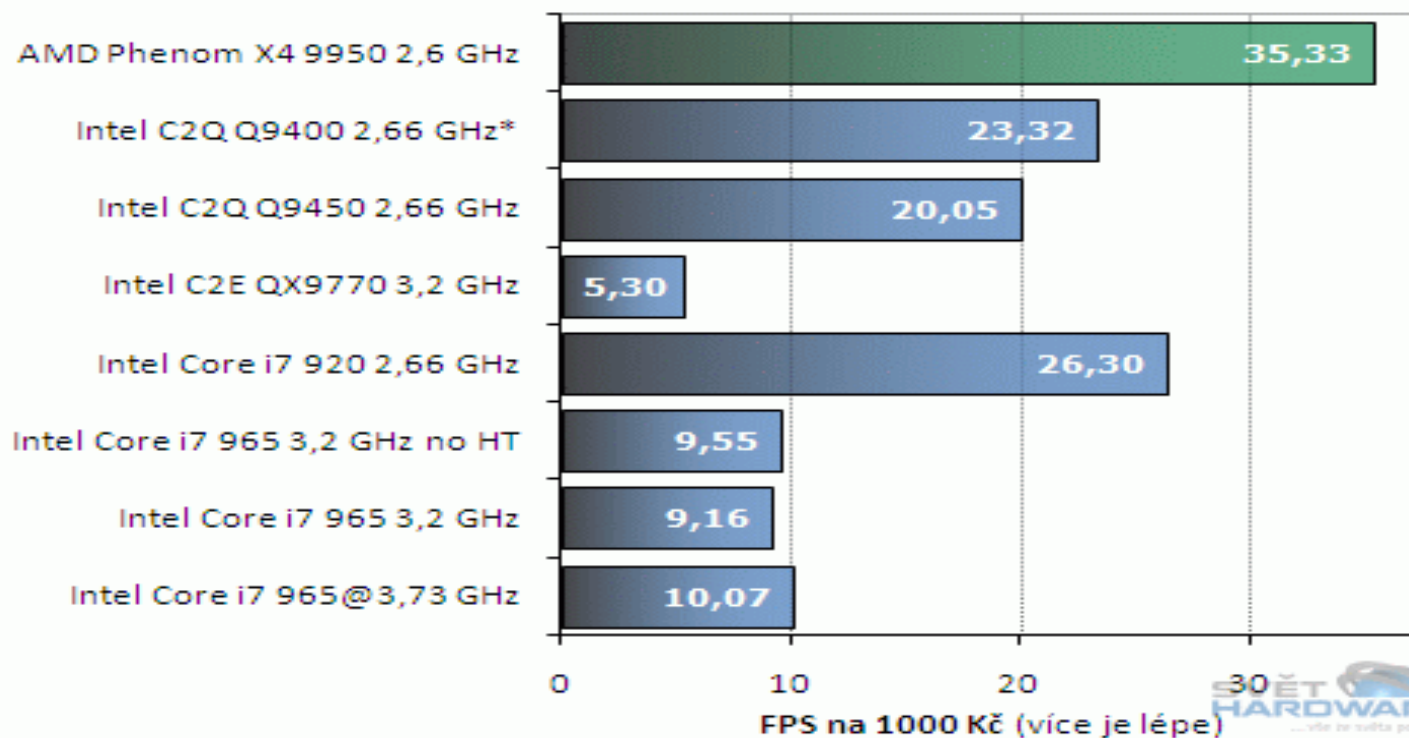
Výkon a spotřeba - x264 benchmark encoding videa





x264 benchmark - výkon na 1000 Kč

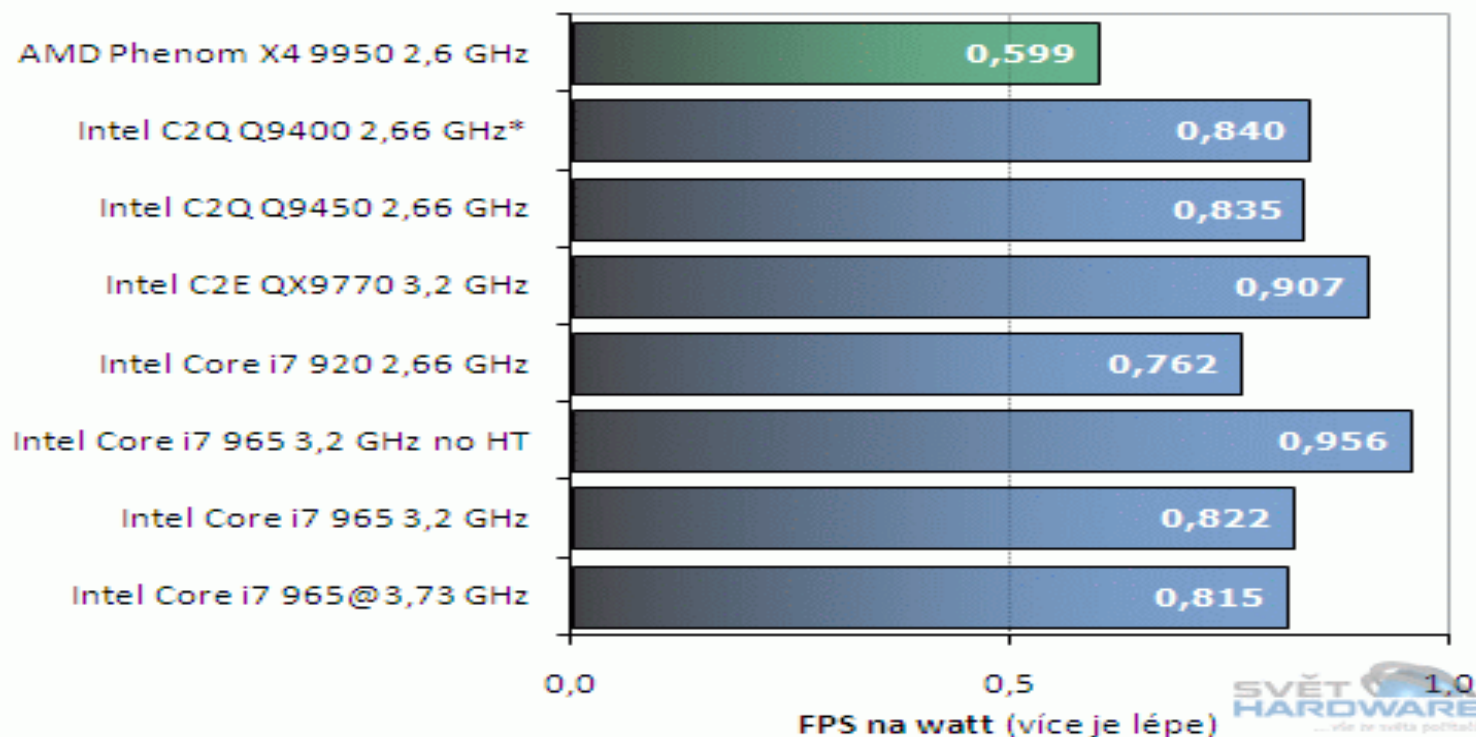
encoding videa





x264 benchmark - výkon na watt

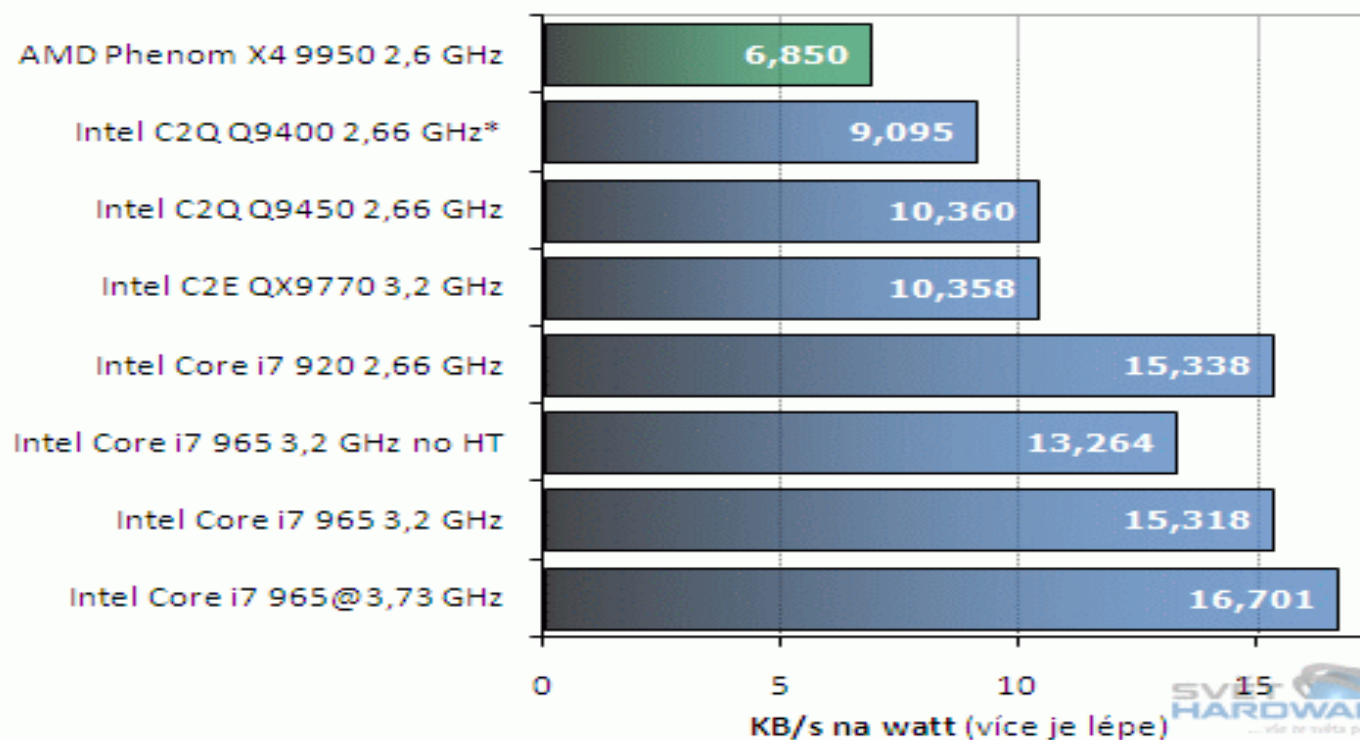
encoding videa





WinRAR 3.80 - výkon na watt

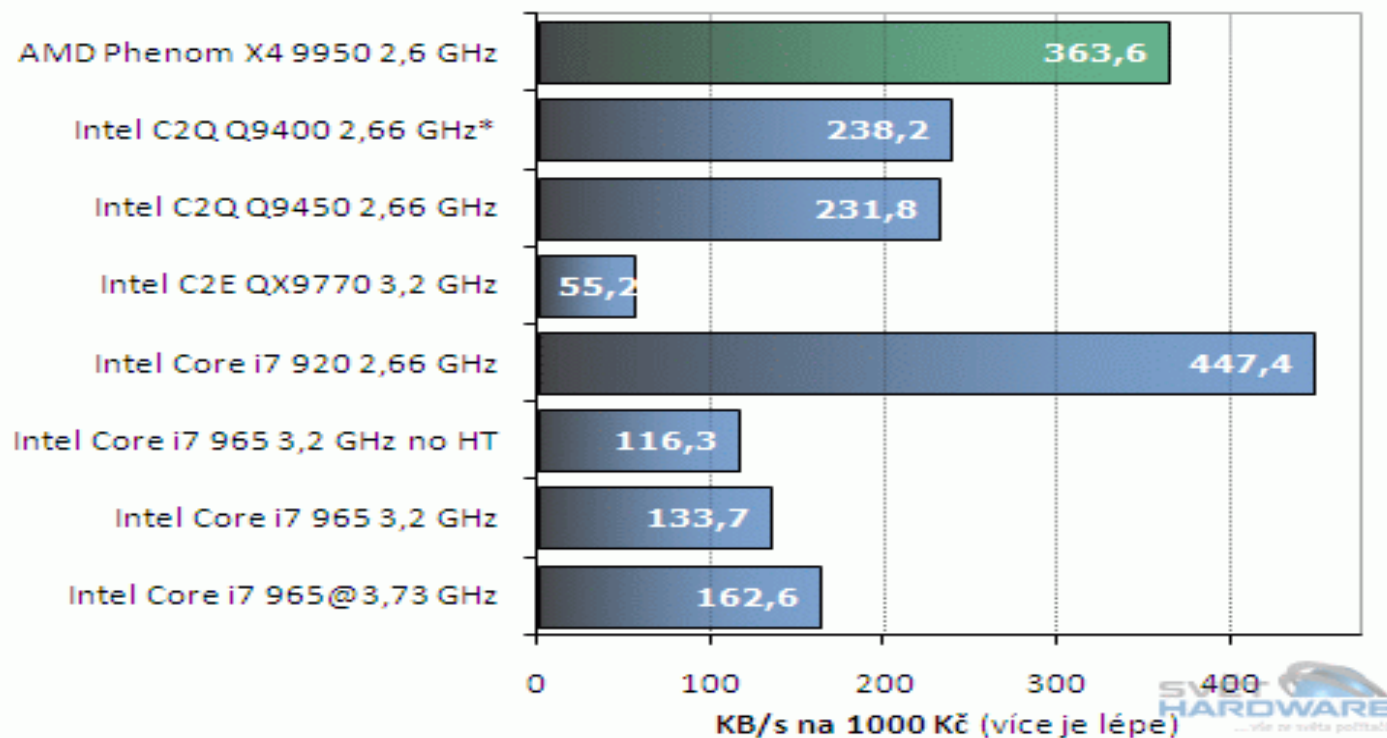
benchmark pakování

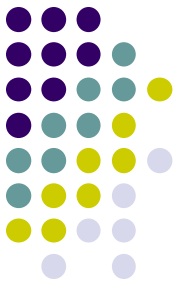




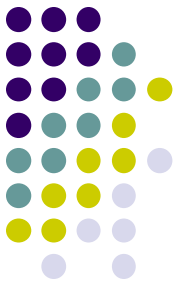
WinRAR 3.80 - výkon na 1000 Kč

benchmark pakování



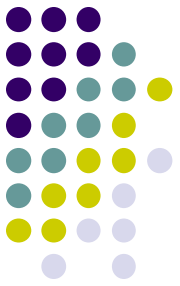


i3, i5, i7



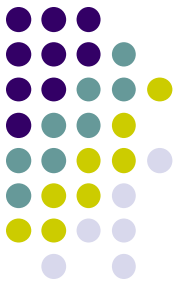
- V současné době firma Intel vyrábí vícejádrové mikroprocesory s označením **i9, i7, i5 a i3**
- Označení **i5 a i7** se poprvé objevilo právě u mikroprocesorů generace Nehalem
- Označení **i3** se pak objevilo u procesorů následující generace Westmere
- Procesory i7 obvykle představují varianty mikroprocesoru s vysokým výpočetním výkonem, není zde kladen důraz na spotřebu elektrické energie a cenu. Takové procesory jsou vhodné například do serverů
- Procesory i3 naopak představují varianty procesoru optimalizované pro nejnižší spotřebu elektrické energie a také nízkou cenu. Takové mikroprocesory jsou vhodné například do notebooků
- Procesory i5 jsou pak jakýmsi kompromisem mezi všemi protichůdnými požadavky
- Hranice mezi procesory i3, i5 a i7 není nijak striktní
- Procesory i7 jsou samozřejmě použitelné i v notebookech

i3, i5, i7



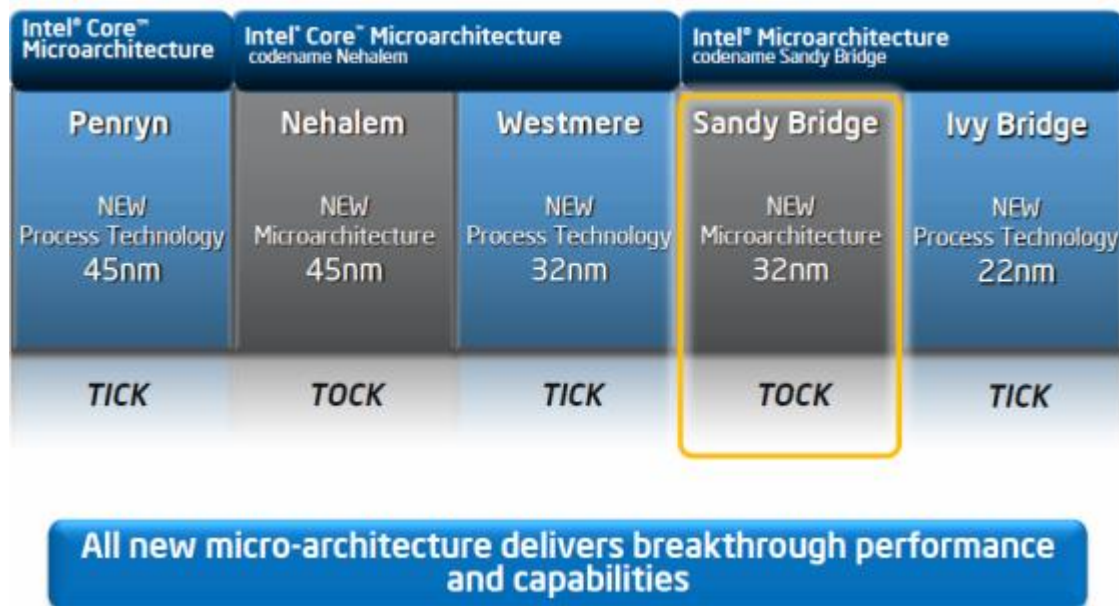
- Samotné označení i5 nestačí a je nepřesné – vždy je třeba dodat o jakou generaci čipu s tímto označením se jedná a jaké jádro je zde použito
- Po architektuře **Nehalem** následuje několik dalších generací a ve všech byly vyráběny procesory **i7, i5 a i3**
- **i5 Westmere** je úplně jiný procesor než například **i5 Sandy Bridge**

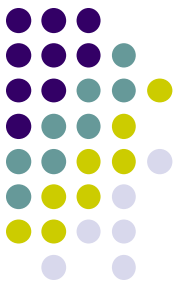




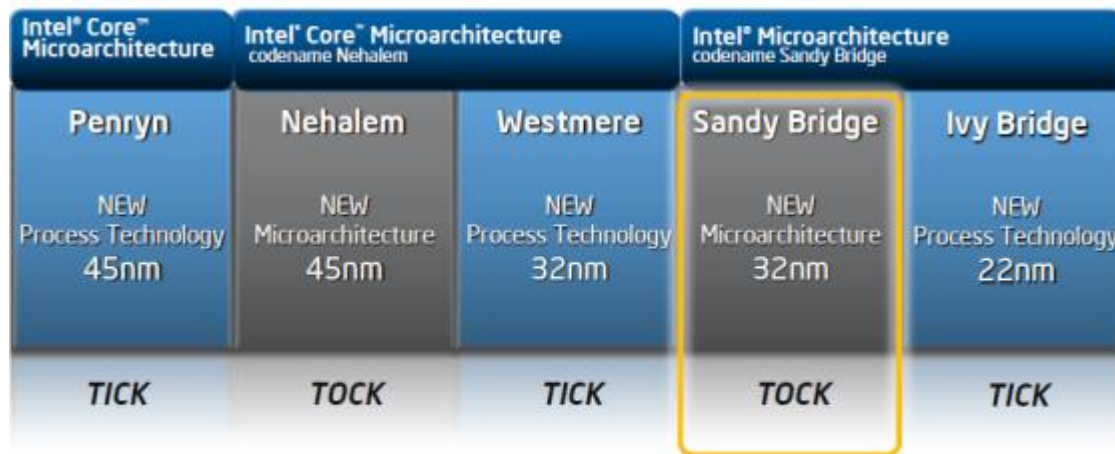
Tick - Tock

- Ve vývoji procesorů Intel uplatňuje model, který sám nazval „**tick-tock**“
- **střídání miniaturizace** výrobního procesu a **uvádění nových architektur**
- Jeden rok tak Intel **zmenšuje velikost tranzistoru** - fáze „**tick**“
- další rok pak **uvede novou architekturu**, která využívá výhod miniaturizovaného výrobního procesu - fáze „**tock**“

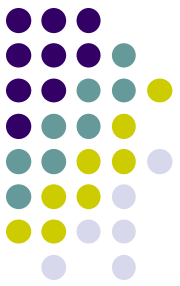




- Po architektuře **Nehalem** přichází (rok 2009) její vylepšená verze **Westmere** – nejde o změnu koncepce, ale především o zlepšení výrobní technologie (fáze Tick)
- Nový milník nastává vždy ve fázi **Tock**
- Tím je příchod nové architektury na současném výrobním procesu.
- První procesory architektury **Bridge** s názvem **Sandy Bridge** jsou tedy vyrobené na **32nm** technologii.
- Za rok se pak překlápíme do fáze **Tick**, kdy se čipy zmenší pomocí 22nm výroby, nazývat se budou **Ivy Bridge**.



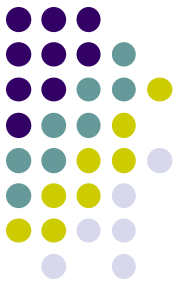
All new micro-architecture delivers breakthrough performance and capabilities



Intel's Tick Tock Development Model

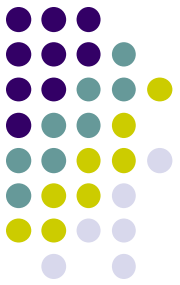


Westmere i3

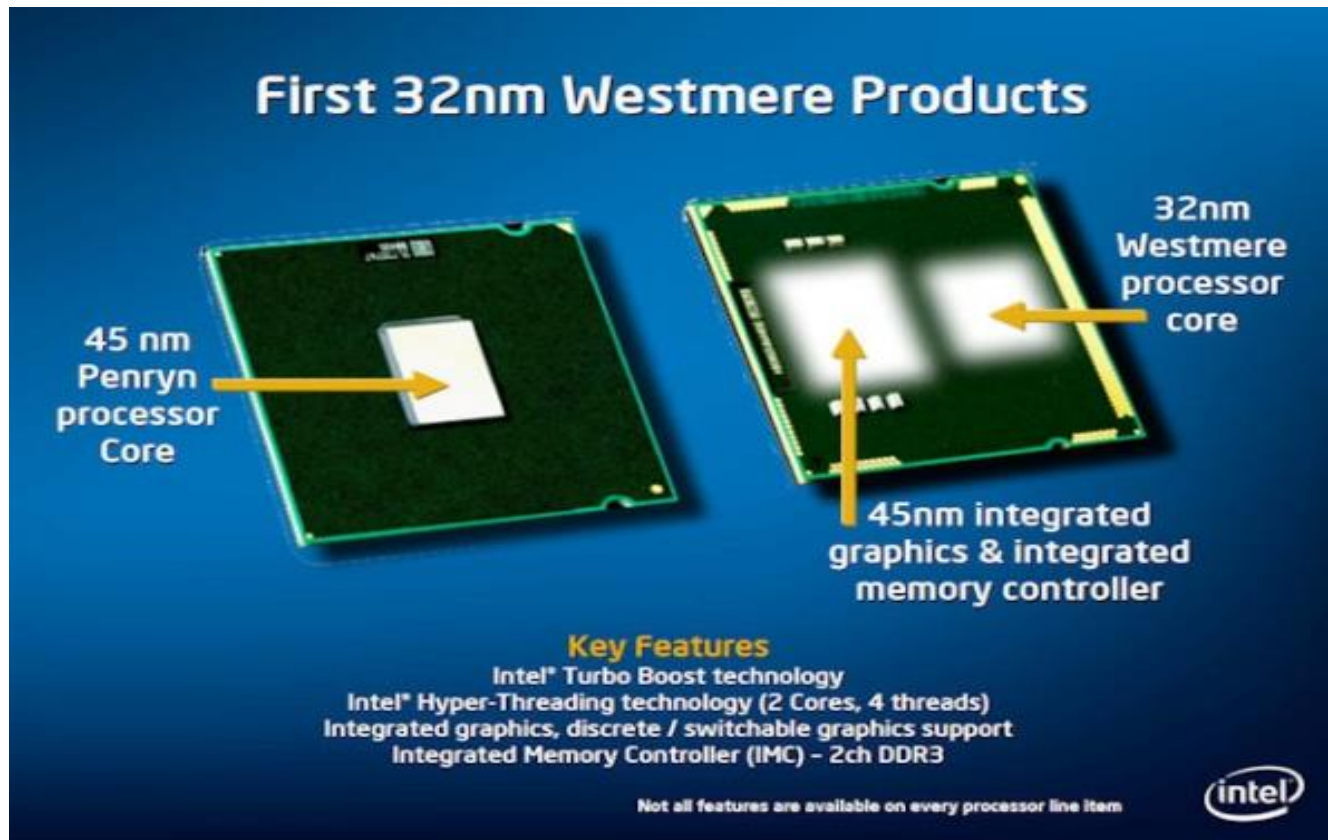


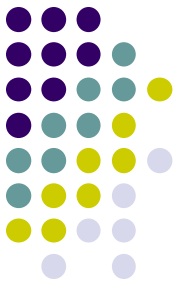
- Na architektuře **Westmere** se objevuje kromě označení **i5** a **i7** nová řada **i3**
- Výrobci notebooků osazují mobilní procesory **Westmere Core i3** zpravidla do svých levnějších modelů.
- Řada **Westmere Core i3** je považována za low-endovou kategorii
- U procesorů i3 je samozřejmě mnoho podobností s výkonnějšími řadami procesorů i5 a i7
- Všechny procesory i3 jsou dvoujádrové a podporují hyperthreading
- funkce **TurboBoost** zde narozdíl od i5 a i7 schází

Westmere i5



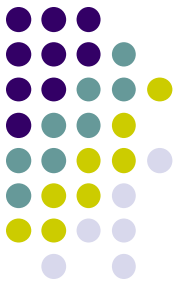
- Procesory i3 a i5 Westmere **mají integrovaný grafický adaptér**
- Procesory i7 Westmere určené do desktopů většinou integrované grafické jádro nemají





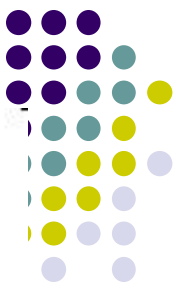
Sandy Bridge

- Další důležitá „**tock**“ fáze ve vývoji procesorů
- Všechny procesory mají **integrované grafické jádro GPU**
- **GPU** – Graphics processing unit
- Dále je nově integrovaný hardwarový **Media engine** – integrovaný čip pro kódování videa a čipu pro dekódování videa
- Předchozí generace procesorů měla bez **Media engine** problém přehrávat plynule **videa ve vysokém rozlišení** (1920x1080px), nyní to jde zcela hladce a procesor je pouze minimálně zatížen
- Instrukční sada **AVX** (Advanced Vector Extension) – nástupce SSE
- **Turbo Boost 2.0** - možní jednomu „studenému“ jádru zvýšit na několik sekund svůj výkon až nad úroveň maximální povolené TDP – teplo se „rozloží v čase“
- Změna L3 cache - přístup k ní je skrze **ring-bus**
- Přímo do procesorů je integrováno rozhraní **PCI-Express**
- **System Agent** - ekvivalent **Severního můstku** umístěný přímo v CPU
 - přístup k PCI Express 2.0 x16 lince, kterou je možné rozdělit na dvě x8.
 - 2kanálový řadič DDR3 (až 1333 MHz) operační paměti s podstatně sníženou latencí a dvojnásobným počtem operací za 1 cyklus
 - DMI rozhraní, Display Engine (ovládání displeje signálem přímo z procesoru)
 - Obsahuje jednotku řízení spotřeby PCU (Power Control Unit)

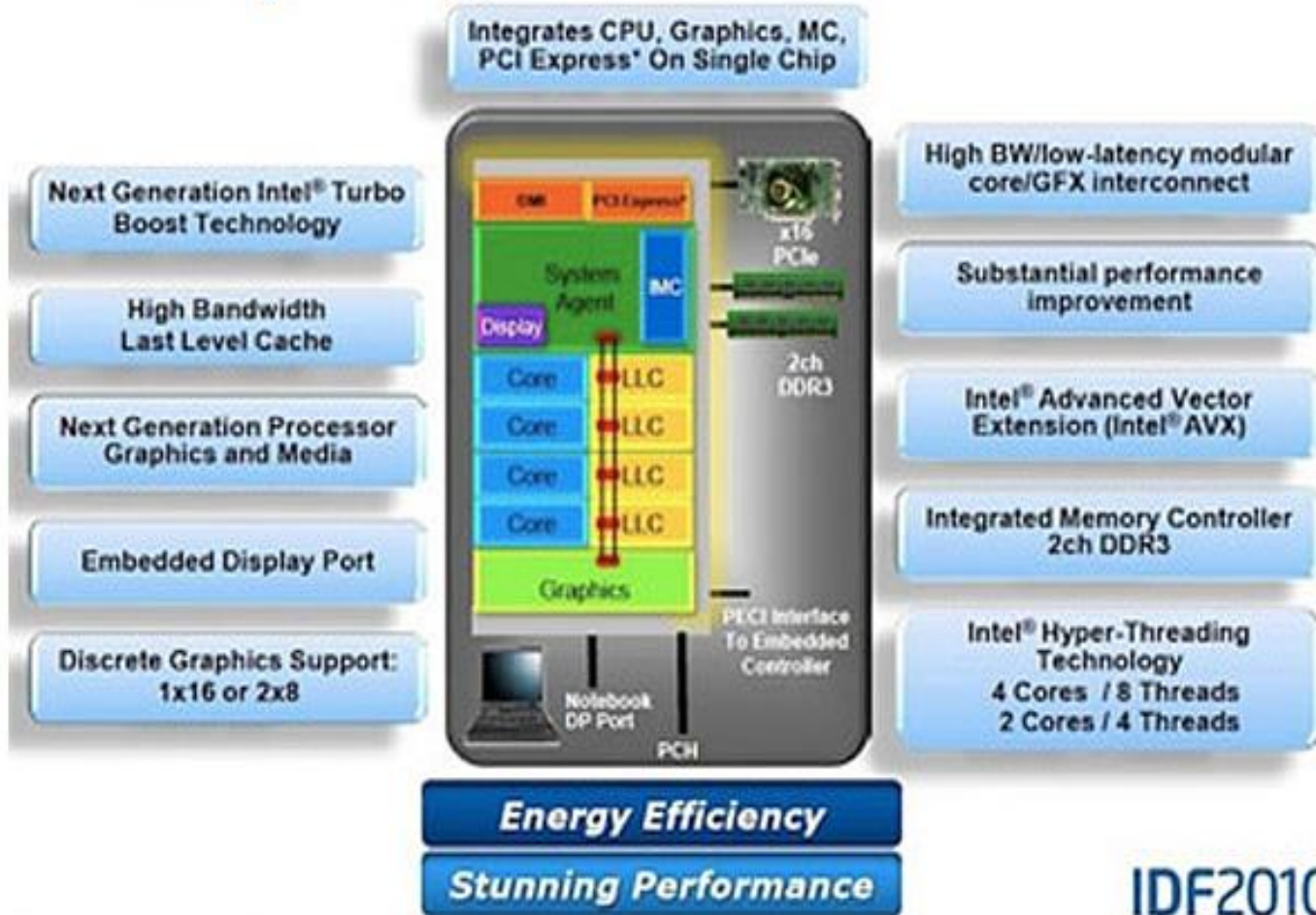


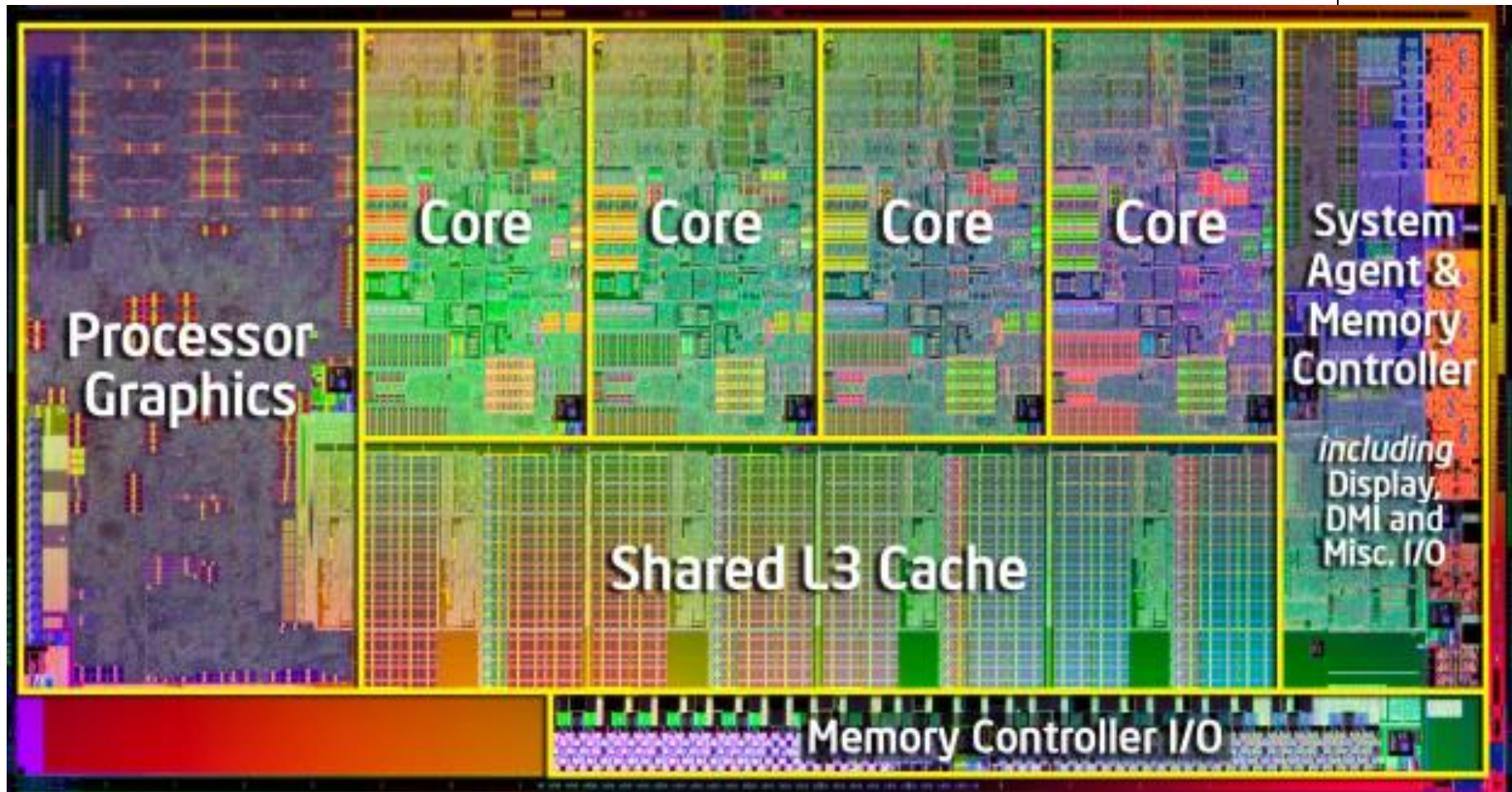
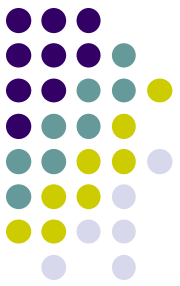
L3 Cache ringbus

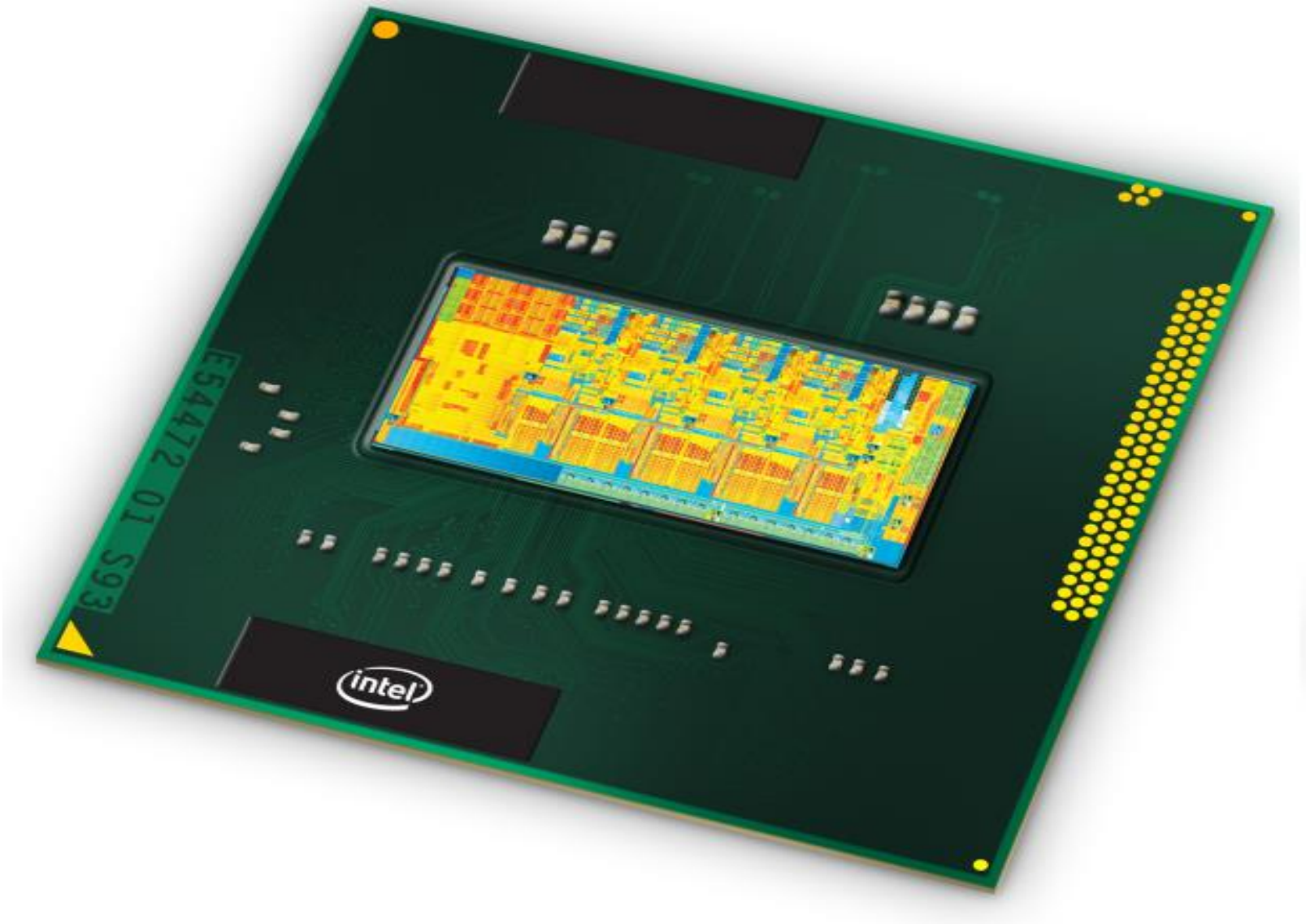
- Intel u architektury Nehalem započal se sdílenou **L3 cache**, do které měla přístup všechna jádra
- S novou architekturou nastala potřeba napojit na **L3 cache** další komponenty než jen jádra CPU (grafika, media engine, System Agent...)
- **RingBus** představuje společnou sběrnici pro každé z jader, L3 Cache, grafické jádro a System Agent
- Uspořádání jádra, řadiče a grafického čipu vytváří kruh kolem L3 cache
- Přesné fungování RingBus nebylo nikdy uspokojivě vysvětleno – otázkou je, jestli nejde spíše o bezkolizní přístupovou metodu ke sdílené Cache

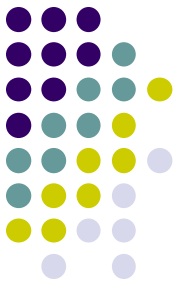


Sandy Bridge: Overview









Haswell

- Výsledkem další fáze **tock** je architektura **Intel Haswell**
- Používá se stejná výrobní technologie jako u předchozí architektury Ivy Bridge (22 nm)
- Intel vyvíjel architekturu Haswell jako velmi flexibilní, aby bylo možné snadno škálovat různé vlastnosti čipu.
- Haswell nabízí různě výkonné varianty pro tablety (7 W), mininotebooky (15 W), notebooky, desktopy (až 84 W)
- Pro tablety a ultrabooky jsou modely označené písmenem **U**.
- Tyto modely jsou vlastně **SoC** (Systém on Chip) – viz prezentace o mobilních technologiích. Integrovaná je zde výkonná grafika Intel Iris a součástí procesoru je integrovaná operační paměť **eDRAM** s kapacitou 128 MB sdílená všemi jádry a grafikou.
- Intel musí reagovat na popularitu čipů **ARM** a jejich úspěch v mobilních zařízeních a kromě snížení spotřeby procesoru se zaměřil i na celkovou spotřebu platformy, která zahrnuje i chipset na základní desce
- Nejúspěšnější řada procesorů Haswell, která má konkurovat procesorům ARM je označená písmenem **Y**

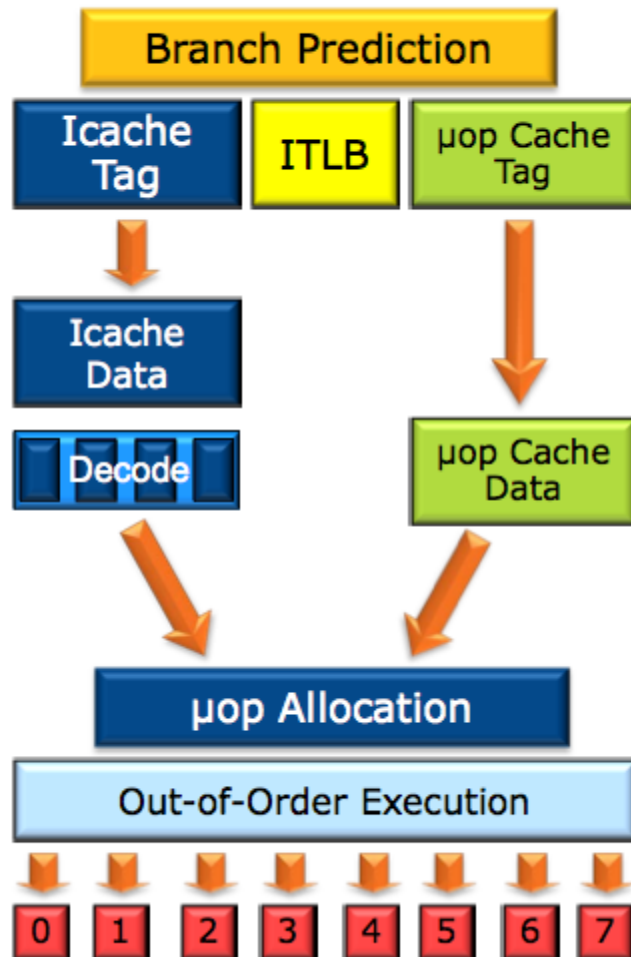
Haswell



- Pro standardní notebooky jsou určeny čtyřjádrové procesory označené písmenem **M** nebo **H** (s výkonnější grafikou)
- Úsporné varianty pro desktopy (35 – 65 W) jsou označené písmenem **S** a **T**
- Pro nejvýkonnější desktopy je určena řada označená písmenem **K** (spotřeba až 85 W)
- Procesory s označením Y jsou jen velmi málo podobné procesorům s označením K. Pouze protože jde o stejnou generaci procesorů, které jsou vyráběny ve stejném období, jsou označeny stejně jako Haswell – v minulosti by spíše bylo běžné označit tak odlišné procesory úplně jiným názvem
- Do toho se samozřejmě dále používá obchodní rozdělení těchto procesorů na i3, i5 a i7



Haswell Core at a Glance



Next generation branch prediction

- Improves performance *and* saves wasted work

Improved front-end

- Initiate TLB and cache misses speculatively
- Handle cache misses in parallel to hide latency
- Leverages improved branch prediction

Deeper buffers

- Extract more instruction parallelism
- More resources when running a single thread

More execution units, shorter latencies

- Power down when not in use

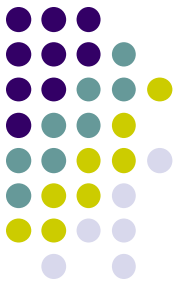
More load/store bandwidth

- Better prefetching, better cache line split latency and throughput, double L2 bandwidth
- New modes save power without losing performance

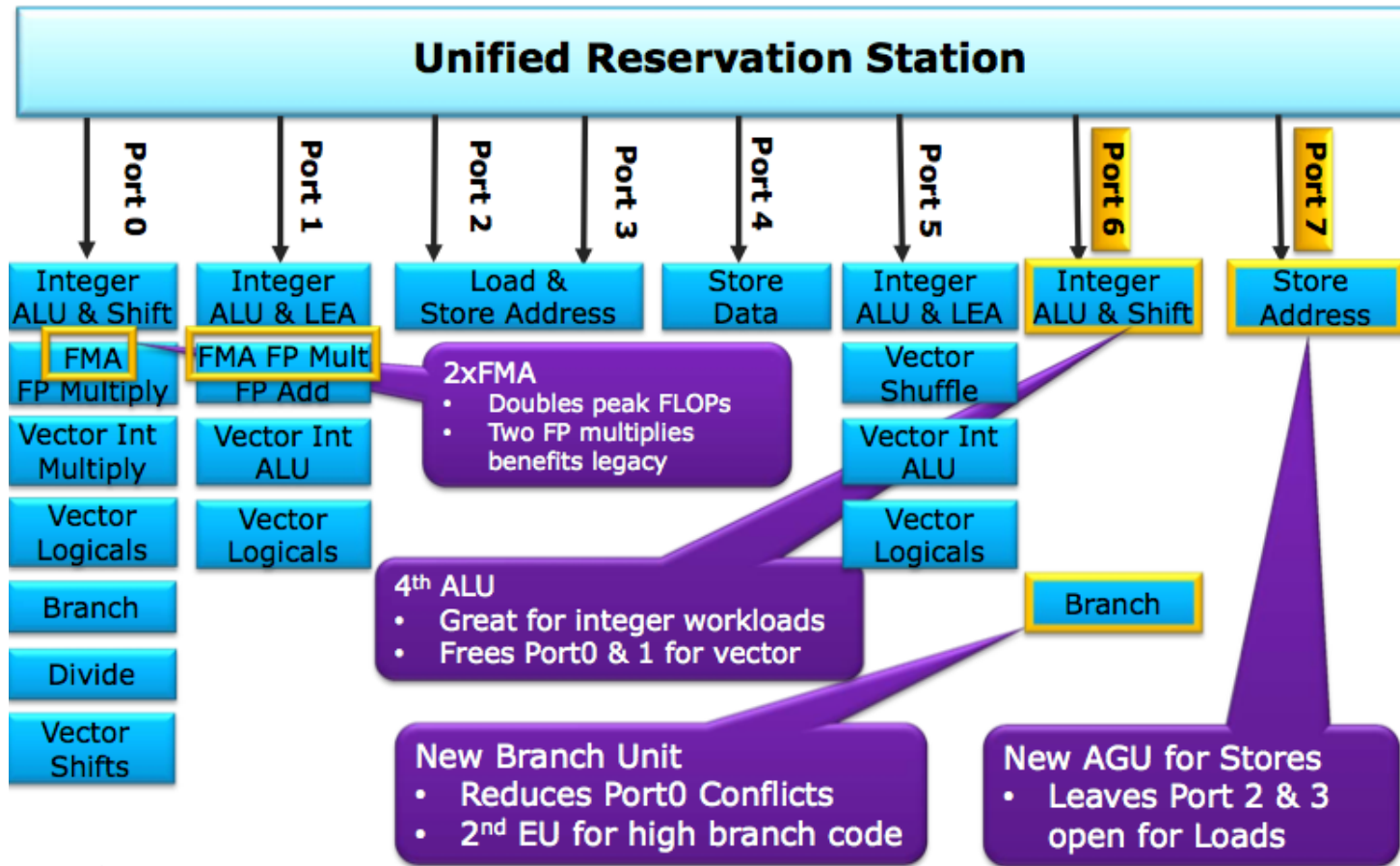
No pipeline growth

- Same branch misprediction latency
- Same L1/L2 cache latency

V jednom taktu lze dokončit až 8 instrukcí (mikrooperací) naráz



Haswell Execution Unit Overview



PCH = Platform Controller hub

Dříve byl mikroprocesor propojen přes **FSB se severním můstkem** na základní desce
Na architektuře Sandy Bridge došlo k integraci severního můstku do mikroprocesoru
(System Agent)

Procesory Haswell jsou připojeny na základní desce k chipu PCH a komunikují s ním přes nové rozhraní **PCH link**

Active Power

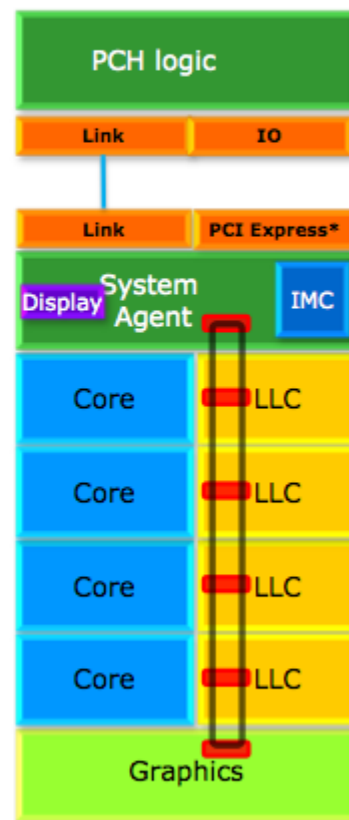
- Turbo Improvements
- Finer grain voltage/frequency islands
 - Decoupled Frequency: Power/Performance where it is needed
- Power optimized CPU to PCH link

Idle Power

- Optimized power delivery to allow fine grain power gating
- New C-States and improved entry/exit latencies
- Manufacturing process optimizations

Platform Power

- New platform power allocations (active and idle)
 - E.g., embedded controller idle power < 5 mW and high-efficiency VRs
- Improved platform power architecture
 - Panel Self Refresh (PSR)
 - Additional interfaces (I2C, SDIO, I2S, UART)
 - New Link Power Management states (USB, SATA*, PCI Express*)





FLOPS = floating point operations

SP = single precision (reálné číslo zakódované pomocí 32 bitů)

DP = double precision (reálné číslo zakódované pomocí 64 bitů)

New Compute Instructions

Intel® Advanced Vector Extensions 2

- FMA: Fused Multiply-Add
 - Throughput: 2xFMA per core
 - Latency: 5 cycles
- Extend AVX to Integer Vector
- Enhanced Data Rearrangement
- New Vector Operations
- Gather

Additional New Instructions

To Accelerate Key Algorithms

- Parallel Extract/Deposit
- Bit Manipulation Operations
- New Rotates/Multiplies/Shift

	Instruction Set	SP FLOPs per cycle per core	DP FLOPs per cycle per core
Nehalem	SSE (128-bits)	8	4
Sandy Bridge	AVX (256-bits)	16	8
Haswell	AVX2 (256-bits)	32	16

Haswell



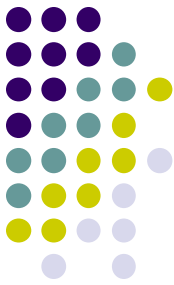
- Sdílená L3 cache je propojená se všemi procesorovými jádry i grafickým čipem a systémovým agentem
- Každé jádro má osmicestnou L1 a L2 cache, ze kterých se data čtou po blocích velkých 64 B
- Nová instrukční sada AVX2 – SIMD s 256-bitově zakódovanými vektory. Orientováno na provádění instrukcí Multiply-Add (součet součinů). Nově podporuje výpočty používané v kryptografii a hashování.
- Intel optimalizoval jednotlivé fyzické obvody a jejich výrobu - cílem bylo, aby nepoužitá logika čipu v co nejkratším čase snížila svou spotřebu na naprosté minimum, ale zároveň aby byla co nejdříve k dispozici v aktivním stavu
- Spotřeba v nejúspornějším stavu spánku se snížila až dvacetinásobně oproti Sandy Bridge
- Díky tomu mají notebooky a další zařízení s čipy Haswell velmi dlouhou výdrž ve spánku (až dva týdny)
- Stavby se za běhu přepínají až o 25 % rychleji

Haswell

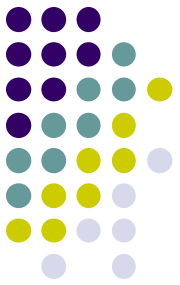


- Nový grafický chip v procesoru je zaměřen na optimalizaci akcelerace přehrávání videa a jeho převodu
- Nový je hardwarový dekodér pro MJPEG (určeno především pro videohovory), hardwarový enkodér pro MPEG2, akcelerace videa v rozlišení 4K....
- Novinkou je dále hardwarová stabilizace obrazu, převod snímkovací frekvence, vylepšení barev a rozšíření gamutu, deinterlacing

Broadwell

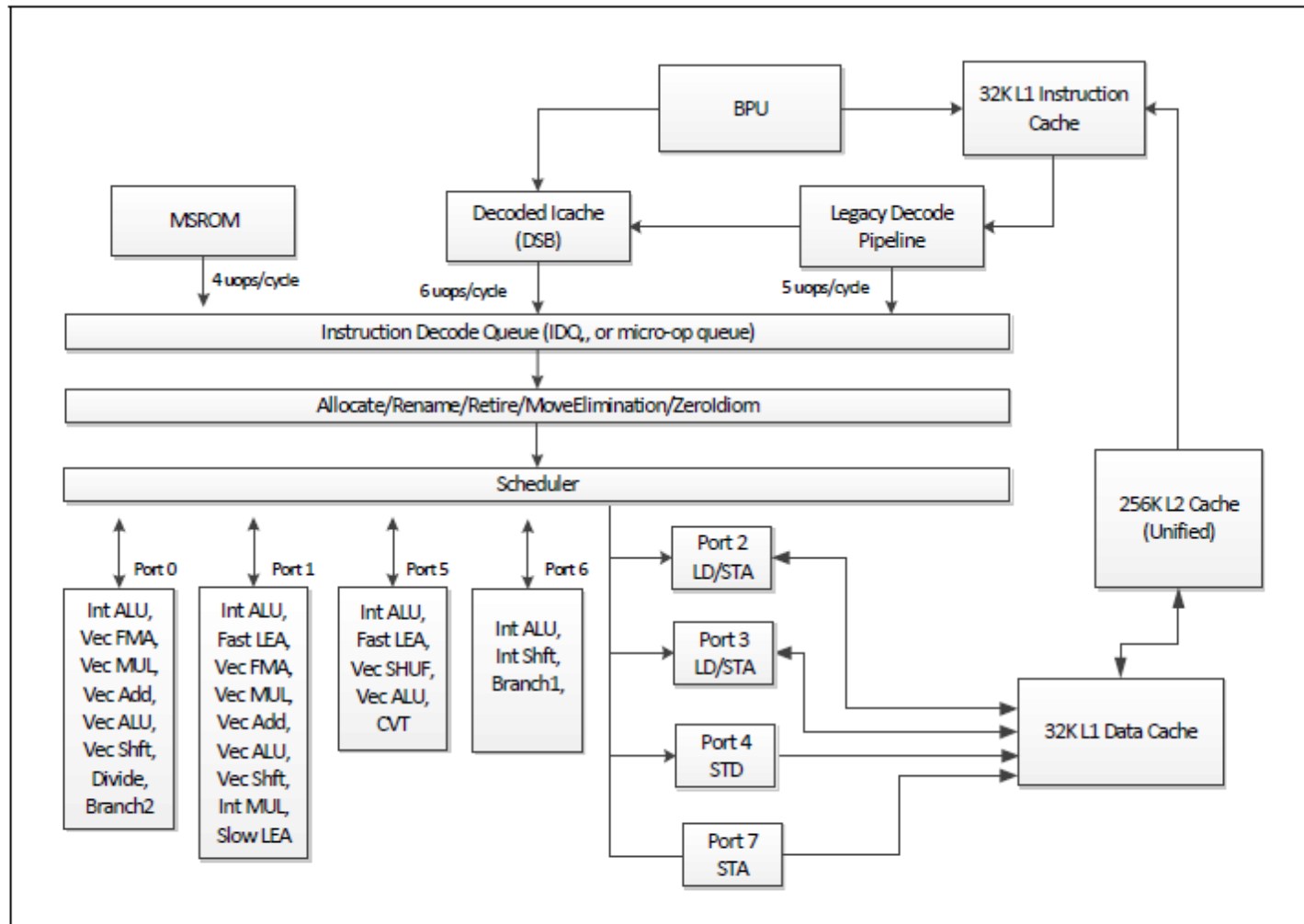
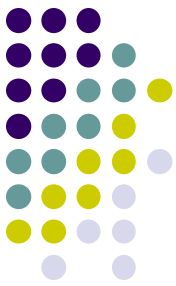


- **Broadwell** je pátá generace procesorů Intel Core – všechny procesory mají ve svém číselném čtyřciferném označení na začátku pětku (např. i7-5950)
- Jedná se o fázi **Tick**, tedy zefektivnění předchozí architektury Haswell
- Haswell architektura využívala **22nm** tranzistory
- Broadwell přechází na výrobní technologii **14nm**
- Nová technologie je až o 30 procent úspornější při stejném výpočetním výkonu
- Naprostá většina této generace procesorů (25 variant) je určena pro notebooky
- Pouze dvě varianty tohoto procesoru (5775C a 5675C) jsou určeny do desktopů. Tyto čipy mají 128 MB **L4 Cache**
- **L4 Cache** byla vyrobena technologií eDRAM (embedded DRAM) - není to rychlá SRAM, ale DRAM integrovaná v procesoru
- **L4 Cache** je využívána zejména GPU jednotkou (grafická data mohou být uložena přímo v procesoru a jsou tak dostupná s nižším zpožděním než z paměťových modulů)
- Někdy také bývá označována jako **pseudo-L4 cache**
- Tato řada procesorů byla uvedena na trh se značným zpožděním oproti původnímu plánu a byla zastíněna novější generací Skylake, která přichází o pouhé tři měsíce později

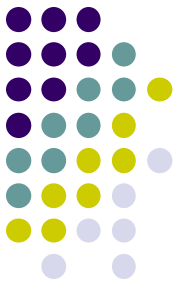


Skylake

- Podzim 2015 - Šestá generace procesorů Intel Core
- všechny procesory mají ve svém číselném čtyřciferném označení na začátku šestku (např. i7-**6**700K)
- Plnohodnotná fáze **Tock** (alespoň Intel si to myslí)
- Výrobní proces **14 nm**
- procesory dostupné v pěti variantách: SKL-S, SKL-X, SKL-H, SKL-U a SKL-Y
- Procesory s koncovkami S a X jsou přetaktovatelné díky odemknutému násobiči
- Varianty H, U a Y byly navrženy především pro přenosná zařízení
- U nejvýkonnějších variant se objevuje **nové označení i9**
- Procesory **i9** mají 10, 12, 14, 16 nebo 18 jader
- Procesory **i7** mají 4, 6 nebo 8 jader
- Procesory **i5** mají 4 jádra
- Procesory **i3** mají 2 jádra
- Procesory označené jako **Pentium** a **Celeron** mají 2 jádra bez hyperthreadingu

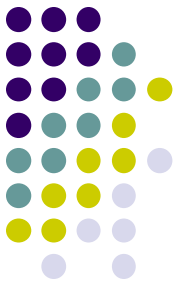


Skylake



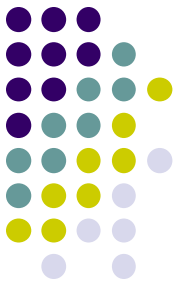
- Na architektuře Skylake se objevil podstatný bezpečnostní problém
- **Spectre vulnerability** – tato řada procesorů je zranitelná sofistikovaným typem útoků, který zneužívá chování procesoru při spekulativním řešení podmíněných skoků
- Procesor při spekulativním provádění podmíněných skoků provádí dopředu i ty části programu, které se vlastně provádět nemají a existuje zde možnost, aby se program dostal k datům, pro která nemá proces v chráněném režimu dostatečná přístupová práva
- Zranitelné jsou i další generace procesorů, ale u Skylake se problém projevuje nejvíce
- Intel problém začal důsledně řešit až v roce 2018 přepracováním predikce podmíněných skoků

Kaby-Lake



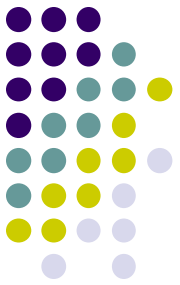
- Podzim 2016 - Sedmá generace Intel Core mikroarchitektury
- Stejně jako předchozí generace Skylake a Broadwell využívá 14nanometrový výrobní proces
- Architekturu Skylake končí několikaletý model **tick-tock**, kde se v jednotlivých generacích střídá nový výrobní proces a nová architektura
- Nový model: **výrobní proces-architektura-optimalizace**
- Po vytvoření nové architektury je tato nová architektura **optimalizována** - Kaby Lake je právě touto optimalizací
- Kaby Lake je první generací procesorů od Intelu, která postrádá oficiální podporu Microsoftu pro verze Windows starší než je Windows 10
- Oproti architektuře Skylake došlo k zvýšení taktovacích frekvencí až o 300 MHz a nejvýraznější změny nastaly u grafického jádra, kde je nyní plná hardwarová podpora HEVC/VP9 dekódování, včetně 4K videí při snímkové frekvenci 60 fps
- Narozdíl od Skylake podporují hypethreading i varianty procesorů označené jako Pentium

Coffee-Lake



- 2018 – Osmá generace procesorů Intel Core
- Druhá optimalizace architektury Skylake
- Očekávaná fáze **tick**, ve které by došlo ke zdokonalení výrobního procesu nebo změně architektury stále nenastává....
- Nyní to vypadá takto: ***výrobní proces-nová architektura-optimalizace-další optimalizace...***
- Některé procesory Coffee Lake mají oproti předchozím generacím procesorů více jader - i3 mají 4 jádra, Core i5 jsou 6jádrové

Cannon-Lake



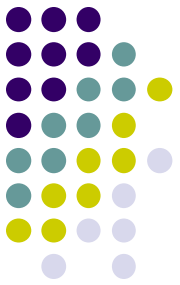
- 2018/2019 - Došlo k zdokonalení výrobního procesu a zmenšení tranzistorů na **10 nm** - odpovídá dřívější fázi Tick
- Takže nyní sekvence posledních fází je takováto: ***výrobní proces-nová architektura-optimalizace-další optimalizace-výrobní proces***
- **AVX512** – dále rozšířená SIMD instrukční sada o vektorové výpočty
- Pracuje s 32 registry o šířce **512 bitů**
- Do jednoho registru lze jako pakovaný bajt uložit 64 čísel, se kterými se vykoná naráz stejná operace nebo s nimi pracuje jako s vektorem či řádkem matice
- Do roku 2020 chtěl Intel vyrábět čipy **5nm** technologií, což se nepodařilo
- Zdá se, že vývoj nových architektur je stále složitější a zpomalilo se i zdokonalování výrobního procesu. Složitost mikroprocesorů stále roste, ale čas potřebný pro zdvojnásobení hustoty integrace se stále prodlužuje z původních 18 měsíců
- Po honbě za co nejvyšší taktovací frekvencí, nastala honba za co největší paralelizací (co nejvyšší IPC, co nejvíce jader...), která také pomalu dosahuje hranice možností dalšího rozvoje
- Dalším směrem vývoje bude zřejmě snaha o dosažení, co nejnižší ceny (nové výrobní technologie zaměřené na extrémně levnou výrobu procesorů)

Ice-Lake



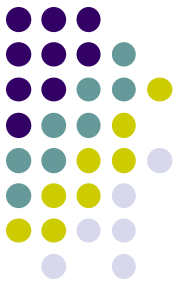
- 10. generace procesorů Intel-Core
- Výrobní proces 10 nm
- 80 kB L1 Cache (48 kB data + 32 kB strojový kód)
- Procesory vyráběné v letech 2020-2021
- Ve stejném období si vyrábějí i chipy označené jako Comet-Lake se starší technologií 14 nm
- Architektura 10 nm procesorů se také označuje jako SunnyCove
- Architektura 14 nm procesorů se označuje jako CypressCove
- Procesory CypressCove jsou novější, mají v průměru o 19 procent vyšší IPC než SunnyCove a rychlejší cache, ale jinak jde o přenos SunnyCove zpět na technologii 14 nm

Rocket-Lake



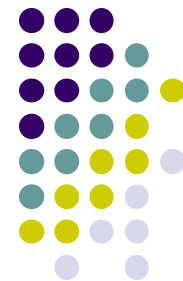
- 11.generace procesorů Intel-Core (březen 2021)
- Výrobní proces 14 nm, architektura CypressCove
- Taktovací frekvence až 5,3 GHz (Turbo Boost)
- Základní taktovací frekvence do 3,9 GHz

Tiger-Lake



- Také 11. generace procesorů Intel-Core
- Tyto procesory jsou na rozdíl od Rocket-Lake vyráběny s tranzistory 10 nm
- Na mikroprocesorech 10. a 11. generace Intel-Core se objevují nové metody umělé inteligence DL-Boost
- DL-Boost = Deep learning boost
- Součástí čipů Tiger Lake jsou i speciální jádra nazvaná GNA, která jsou specializována na úlohy využívající neuronové sítě a strojové učení. Využít ho mohou v současné době například úlohy pro odstranění šumu z obrazu a zvuku nebo algoritmy pro umělé zvětšování rozlišení obrázků (upscaling)
- GNA = Gaussian and Neural Accelerator

Alder-Lake



- **12. generace** mikroprocesorů Intel-Core
- V roce 2021 je po 7 letech považována za skutečně novou zlomovou architekturu (což dříve přicházelo ve fázích Tock)
- Nová architektura jádra GoldeCove
- Technologie **10 nm**
- L1 Cache 96 KB (64 KB instrukce + 32 KB data)
- U procesorů se udávají tři frekvence(base, turbo, vyšší turbo)
- Nová **hybridní architektura** - vybaveny dvěma typy jader - **jádry malými a jádry velkými**
- V procesoru jsou **výkonná jádra (P-cores)** a **efektivní jádra (E-cores)**.
- V praxi to funguje např. tak, že vlákna aktivního okna, se kterým pracujete a probíhající náročné výpočty obsluhují P-cores, a vlákna procesů a oken na pozadí běží na E-cores

- Výkonná jádra mají průměrné IPC=3,4
- Efektivní jádra mají průměrné IPC=2,2

Alder-Lake



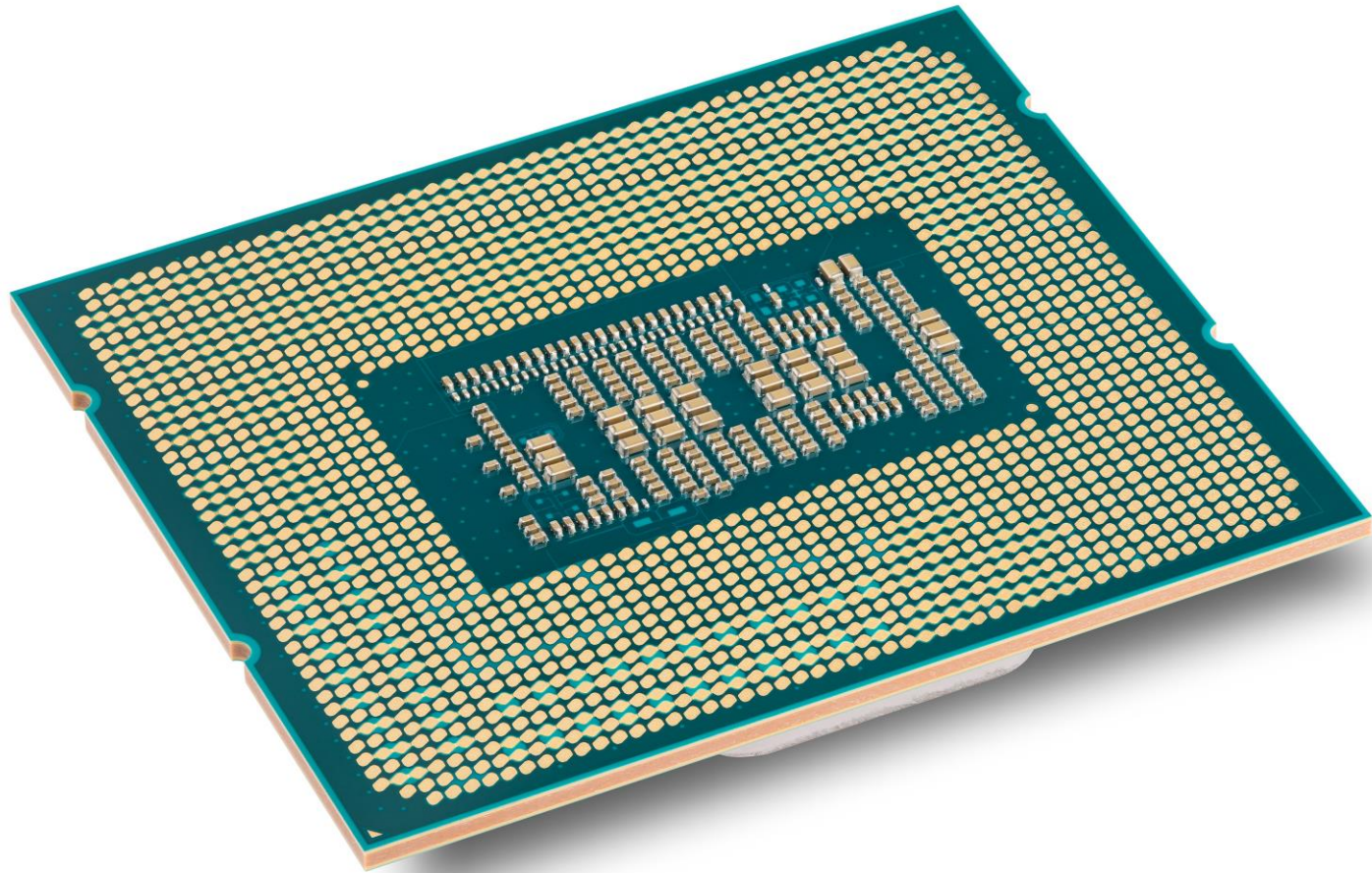
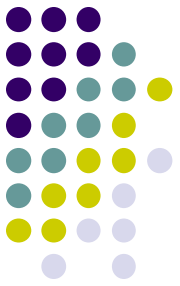
- Hardwarový mikrokontrolér **Intel Thread Director** - monitoruje vše, co se v procesoru děje. Má přehled o právě prováděných instrukcích, budoucích načtených instrukcích, taktech, teplotách, spotřebě a dalších stavech procesoru, které v reálném čase každou nanosekundu
- Thread Director reálně nic neřídí, přidělování vláken jednotlivým jádrům zůstává na softwarovém plánovači operačního systému. Thread Director pouze poskytuje zpětnou vazbu operačnímu systému a radí mu, jak podle něj dělat věci lépe.
- Aby mohl operační systém Thread Directoru vůbec naslouchat, bylo potřeba správce úloh a vláken ve Windows pro tyto účely upravit. Toto vylepšení se nachází pouze v novém **Windows 11**.
- Procesory Intel Alder Lake fungují samozřejmě i pod Windows 10, ale méně efektivně

Alder-Lake



- **Duální řadič paměti** – umí pracovat s pamětí DDR4 i DDR5
- **Intel Dynamic Memory Boost Technology** - funguje podobně jako turbo u jader procesoru.
- V momentě, kdy je potřeba velmi vysoký výkon, lze paměť krátkodobě přetaktovat.
- Pokud mikroprocesor není zatížený, lze přepnout do úsporného režimu i paměťové moduly
- Technologie funguje s novými paměťovými moduly DDR a DDR5, které mají Intel XMP certifikát
- **XMP** = Extreme Memory Profiles. Pro paměť lze nastavit různé „výkonnostní profily“ (napětí, časování) a přepínat je.
- XMP 3.0 nabízí celkem tři XMP profily nastavené od výrobce a navíc dva další pro samotného uživatele, které si může nastavit podle potřeby.

Alder-Lake

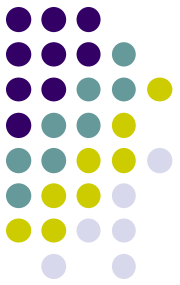


Raptor-Lake



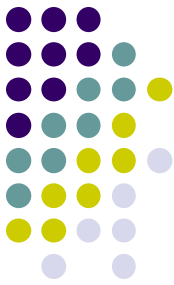
- 13. generace procesorů Intel-Core
- Uvedení na trh v říjnu 2022
- Výrobní proces 10 nm
- Procesor obsahuje
 - **RaptorCove** high-performance jádra (P-core) s taktovací turbo frekvencí až 6 GHz
 - **Gracemont** úsporná jádra (E-core)
 - Až 36 MB L3 Cache

IA-32

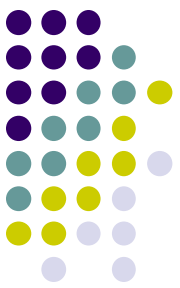


- **Intel architecture – 32 bit**
- Souhrné označení pro 32-bitové ale i 64-bitové procesory řady **80x86** – počínaje procesorem **80386**
- **32-bitové aplikace** jsou spustitelné na libovolném z těchto vzájemně **kompatibilních** procesorů
- Všechny procesory nabízejí 32-bitové registry EAX, EBX, ECX, EDX, EDI, ESI...
- Pro adresaci paměti se používají segmentové registry CS, DS, ES, SS, FS, GS
- K dispozici jsou 32-bitové **CISC instrukce**, které vycházejí z původního procesoru 8086
- Adresy v programu jsou relativní a vztahují se k počátku segmentu, který může být při spuštění programu operačním systémem alokován kdekoliv v paměti
- Segmenty mohou být velké až 4 GB a při zapnutém stránkování se dělí na 4 kB stránky
- Strojový kód (způsob jakým jsou instrukce zakódované pomocí bajtů) je u všech mikroprocesorů vycházejících z 80386 kompatibilní
- Vnitřní architektura jednotlivých procesorů je často zcela odlišná, přesto se navenek ovládají a programují v podstatě stejně – mluvíme tedy o **jednotné architektuře IA-32**

x86-64

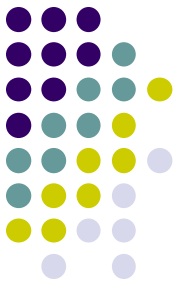


- **64 bitová** verze původní instrukční sady **IA-32**
- K dispozici jsou 64 bitové **datové registry** RAX, RBX, RCX...
- Umožňuje adresovat až **2⁶⁴ B** paměti
- Poprvé použito firmou **AMD** u mikroprocesorů **K8**
- Firma Intel použila 64-bitový režim později, u Pentia4 a novou koncepci nazvala **Intel64**
- Mikroprocesor umožňuje **reálný, chráněný a V86** mód - **Legacy** (zdeděné) módy
- Nový režim se nazývá **long-mode**
- Mikroprocesor po přepnutí do long-modu přejde do **64-bitové** verze **chráněného režimu**
- Přepnout mikroprocesor do long mode může pouze **64-bitový operační systém**, který podporuje tento nový 64-bitový chráněný režim
- Program napsaný pro **long mode** se nazývá také jako **64-bitová aplikace** – takový program nepůjde spustit v klasickém chráněném režimu IA-32
- Procesor je možné provozovat s **32bitovým** jádrem operačního systému (kterým může být i systém určený pro i386) ve starých **legacy módech**
- Větší výkon bude dosažen s **64bitovým jádrem** operačního systému v Long módech - jádro potom běží v **64bitovém módu** a aplikace v 64bitovém nebo v kompatibilním
- Mnoho aplikací je dnes k dispozici ve dvou verzích
 - 32 bitové verze používá instrukční sadu IA-32
 - 64 bitová verze vyžaduje 64-bitový operační systém a využívá 64-bitové registry a novou instrukční sadu x86-64



Kontrolní otázky

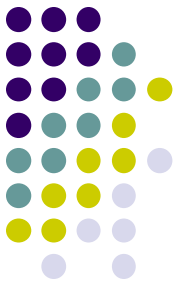
- Proč se architektura Netburst nehodí pro vícejádrové mikroprocesory ? – **má zbytečně dlouhý pipeling**
- Kolik instrukčních front používá architektura Netburst ? – **3 + 1 (FPU)**
- Vysvětlete rozdíl mezi inkluzivní a exkluzivní víceúrovňovou cache – **inkluzivní – data z L3 cache se kopírují do cache jednotlivých jader, exkluzivní – data z L3 cache se přesunují do cache daného jádra a dále nejsou dostupná pro ostatní jádra**
- Co je to Powergating ? – **odpojuje neaktivní jádro a tím šetří elektriku, architektura Nehalem**
- K čemu slouží jednotka PCU ? – **řídící jednotka spotřeby, ovládá Turbo boost, Powergating**
- Jak funguje Turbo boost ? – **pokud počítač nepoužívá všechny jádra, dokáže zvýšit frekvenci jiného jádra**
- Proč Turbo boost neumožní zvýšení frekvence používaného jádra na dvojnásobek, když druhé jádro je nečinné ? – **dvojnásobek výkonu spotřebuje 4x více energie, tudíž by se procesor přehřál**
- Proč Turbo boost neumožní zvýšení frekvence používaného jádra na hodnotu, při které příkon dosáhne TDP, když druhé jádro je nečinné ? – **teplo by vznikalo na velmi malé ploše**
- Seřadte chronologicky architektury mikroprocesorů Intel Core, Netburst, IvyBridge, Westmere, Nehalem, SandyBridge, Skylake, Broadwell – **Netburst, Intel Core, Nehalem, Westmere, SandyBridge, IvyBridge, Broadwell, Skylake**
- Vysvětlete fáze Tick - Tock – **Tick – zlepšení výrobní technologie, Tock – optimalizace výrobní technologie a nové funkce**
- Ve které architektuře mikroprocesoru se objevuje integrovaný Media Engine a k čemu slouží ? – **SandyBridge – kódování videí ve fullHD**
- Co je to IA-32 ? – **mód pro aplikace, které fungují od 80386**
- Co je to long-mode ? – **procesor může používat až 64 bitové registry**



Kontrolní otázky

- U kterých procesorů se poprvé objevila L1 cache, L2 cache, L3 cache, L4 cache? **L1 – 80486, L2 – Pentium Pro, L3 – Nehalem, L4 – Broadwell**
- Jaký rozměr tranzistoru mají přibližně současné mikroprocesory? – **pod 10 nm**
- Co je to leakage current? – **napětí protéká i skrz zavřené tranzistory**
- Vysvětlete pojem in-flight MicroOPs – **Jedná se o buffer, kam se řadí instrukce před zpracováním procesorem – tedy instrukce, které čekají na srovnání k optimálnímu zpracování CPU**
- Co je to AVX? – **nástupce MMX/ SSE – registry, se kterými jde dělat hromadné operace, lze je chápat jako vektory, advanced vector extension**
- V čem spočívá rozdíl mezi 32-bitovou a 64-bitovou aplikací? – **64 bitová aplikace může adresovat více paměti**
- Kolikrát více tranzistorů se vejde na stejnou plochu, pokud se zdokonalením výrobního procesu zmenší rozměr tranzistoru o 50 procent? **4x**
- Co je to GPU? – **grafická karta (grafický čip)**
- Co je to PCH? – **platform controller hub – čip pro komunikaci se základní deskou**

Použitá literatura



- <http://en.wikipedia.org/wiki/X86-64>