ECE 571 – Advanced Microprocessor-Based Design Lecture 22

Vince Weaver

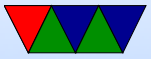
http://web.eece.maine.edu/~vweaver

vincent.weaver@maine.edu

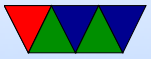
21 November 2019

Announcements

- Consensus is we will have class on Tuesday
- HW #11 will be a GPU reading
- Grad seminar on Friday
- Project status due Tuesday
  Topic, plus related work

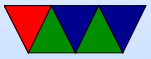


Reading of the Articles



Intel Ice Lake / Sunny Cove

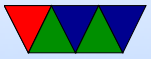
Generation	Name	tech		Year
-	Conroe/Merom	65nm	Tock	2006
-	Penryn	45nm	Tick	2007
1	Nehalem	45nm	Tock	2008
1	Westmere	32nm	Tick	2010
2	Sandy Bridge	32nm	Tock	2011
3	Ivy Bridge	22nm	Tick	2012
4	Haswell	22nm	Tock	2013
5	Broadwell	14nm	Tick	2014
6	Skylake	14nm	Tock	2015
7	Kaby Lake	14nm	Tock	2016
8/9	CoffeeLake	14nm	Tock	2017
9	CannonLake (mobile,rare)	10nm	Tick	2018
10	IceLake/SunnyCove	10nm+	?	2019



Sunny Cove Article

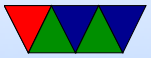
https://www.anandtech.com/show/14514/examining-intels-ice-lake-microarchitecture-and-sunny-cove

Examining Intel's Ice Lake Processors: Taking a Bite of the Sunny Cove Microarchitecture by Ian Cutress.

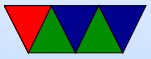


Page 1 – Ice Lake

- First “widely-available” 10nm core
- 18% performance increase over Skylake
- Gen11 Graphics, 1TFLOP
- 10th generation core
- Up to four cores
- Ice-Lake U and Ice-Lake Y (mobile)
- Spectre and Meltdown fixes
- AVX-512
- Thunderbolt 3 (PCIe and displayport)

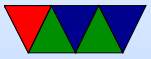


- Turbo up to 4.1GHz
- LPDDR4X-3733 (32 GB) or DDR4-3200 (64 GB)
- 16 PCI 3.0 lanes (32, but 16 needed for Thunderbolt)
  Note this is much less (and not 4.0) like on Zen 2
- Project Athena – sort of a common minimal spec for laptop designers
- This article was fro July. How does it stand?
  You can buy laptops with this, but desktops still 9th generation (Coffee Lake)



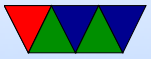
Page 2 – Ice Lake

- For laptop, combine CPU and chipset in one package, 10nm+ and 14nm
- Up to 4cores/8threads, Up to 4.1GHz, can use LPDDR4 RAM
- Chipset
  - Up to 64GB of DDR4-3200 or LPDDR4/X with 32GB 3733 MT/s
  - "New Converged Chasis Fabric" – power control
  - Integrated WiFi 6

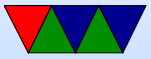


- Voltage Regulators integrated (reducing parts)
- Audio DSP
- eMMC
- 6 USB 3.2 ports or 10 USB 2.0 or some combo
- SATA, eMMC2, UFS? – universal flash storage (replacing eMMC since 2016? Speed rivaling SATA)?

- 10nm process
  - Cannonlake did not go well
  - Have added metal layer
  - New transistor lib, MIM caps
  - Thin magnetic inductor array

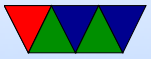


- Reduced package Z height

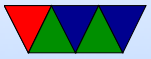


Page 2 – Ice Lake

- 10th gen core
- milliwatt to Megawatt strategy
- Sandybridge hadn't really been updated in 4 years
- Prefetchers and Branch predictors Intel keeps secret
  Brpred bigger and more accurate
- L1 icache still 32k/8-way, still 5 decoders (1 complex, 4 simple)
- uop cache from 1.5k to 2.25k (but buggy? recent news)
- uop queue now 352 uops in size

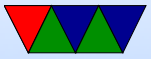


- 128 in-flight loads, 72 in-flight stores
- Back End – four reservation stations
  - 1 for 4 ALU ports (4op/cycle)
  - 1 for 2 store ports (2op/cycle)
  - 1 for 2 AGU ports (load/store) (2op/cycle)
- Up to 10 uops/cycle dispatched
- Increased bandwidth to L1 cache
- Improved L1 Data cache
  - Up from 32k, 8 way to 48k 12-way
    still works for VIPT
- L2 cache up from 256k to 512k, 8-way non-inclusive L1,



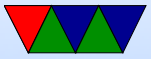
write-back

- Improved TLB
- VNNI? AVX-512 Vector Neural Network Instructions VPDPBUSDS, VPDPWSSD
- 18% IPC. Crazy graph, just shows x speedup for SPEC2016, 2017, cinnebench, but no labels. ALso one benchmark worse. Huge disclaimer.
- DLboost vs whiskey lake "deep learning boost", Whiskey Lake after Coffee Lake



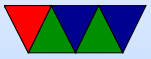
Page 3 – Gen 11 Graphics

- Xe designs? Intel back to discrete graphics
- 64 execution units, 1100MHz, 1TFLOP
- Low-end DDR4 (vs LPDDR4) designs might have worse graphics perf due to lack of bandwidth
- Variable rate shading, change shading detail depending on importance



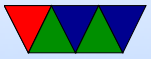
Page 4 – DL Boost

- AI, Machine learning
- Lower resolution calc, FP32 to FP16 to INT8, to INT2???
- IFMA
- AVX-512



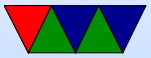
Page 5 – Two Power Targets

- H, Y, and U power budgets
- 5-28W
- binning for power?
- Packages 1526 balls, 0.65mm pitch, 1377 balls, 0.43mm pitch
- ANother try at FIVR (fully integrated voltage regulator), holes in motherboard for caps to stick down, inductors



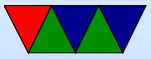
Page 6 – Dynamic Power Tuning

- Intel has TDP, then PL1 and PL2. Can maintaing PL2 (turbo) before dropping to PL1 (sustained)
- New way of doing things.. machine learning, predicts future power draw?



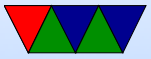
Page 7 – Thunderbolt on CPU

- TB3
- Cannonlake had it, didn't work?
- USB4 support?



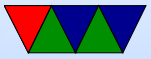
Page 8 – Wifi-6

- CVNI proprietary connector
- external 28nm chip for analog, better performance



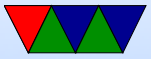
Page 9 – Benchmarks

- Intel complain, probably because not doing great
- 18% IPC improvement good, but over 4 years?



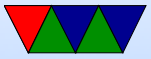
Page 10 – Comet Lake confusion

- 10th gen chip but 14nm?



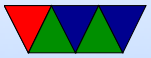
ARM Article

https://www.anandtech.com/show/14384/arm-announces-cortexa77-cpu-ip ARM's new Cortex-A77 CPU Micro-Architecture: Evolving Performance by Andrei Frumusanu.



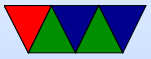
Note – A77 an update over A76

- A76 clean sheet design
- “Laptop-class, Mobile Efficiency”
- Super-scalar OoO 4-wide 8 execution units, 13-stage pipeline
- Branch-predictor hybrid indirect
- 16-deep Issue queue
- Cache: L1: 64k 4-way 4-cycle
  L2 9-cycle

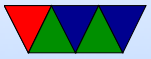


Page 1

- From May
- Cortex-A76 last year (maybe should have read up on that one)
- Austin Core – Deimos
  - A73+A75 designed in France
  - A53+A55 designed in Cambridge England
  - A76+A77 designed in Austin, TX
- Confusing plots

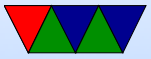


- 7nm TSMC (same as Zen2?), 3GHz
- Competition. Samsung M4? (recently shut down?) Apple A11 and A12
- Best PPA in industry? PPA = Power, Performance, Area
- ARMv8.2
- Can pair with Cortex-A55 biglittle DynamIQ shared unit (DSU)
- 64k L1 I and D, 256k or 512k L2 (1MB vesion for non-mobile)

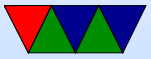


Page 2 – Microarchitecture

- higher fetch bandwidth
- macro-op cache (L0 instruction cache) fuse instructions? Located after decode, feed directly into core, mispredict latency down to 10 from 11 (other Zen1, skylake 16 cycles)
- Improved branch predictor – bigger run ahead (32B vs 64B) up to 16 instructions per cycle, allow to catch up on "branch bubbles"
- Bigger BTB, got rid of split nano/micro from A76

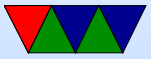


- Decoder still 4 wide, but 6 can launch if hit in MOP cache



Page 2 – Microarchitecture

- Added ALU, single cycle and some double cycle instructions
- Surprising perf increase from added ALU
- Faster multiply, 2-3 cycles instead of 4
- additional branch port
- Not much about FPU, good enough?
- Improved data prefetcher



Page 2 – Microarchitecture

- SPEC benchmarks, SPEC2006 int 23% IPC, fp 35%. Compare to Sunny COve
- Energy efficiency same as Cortex-A76