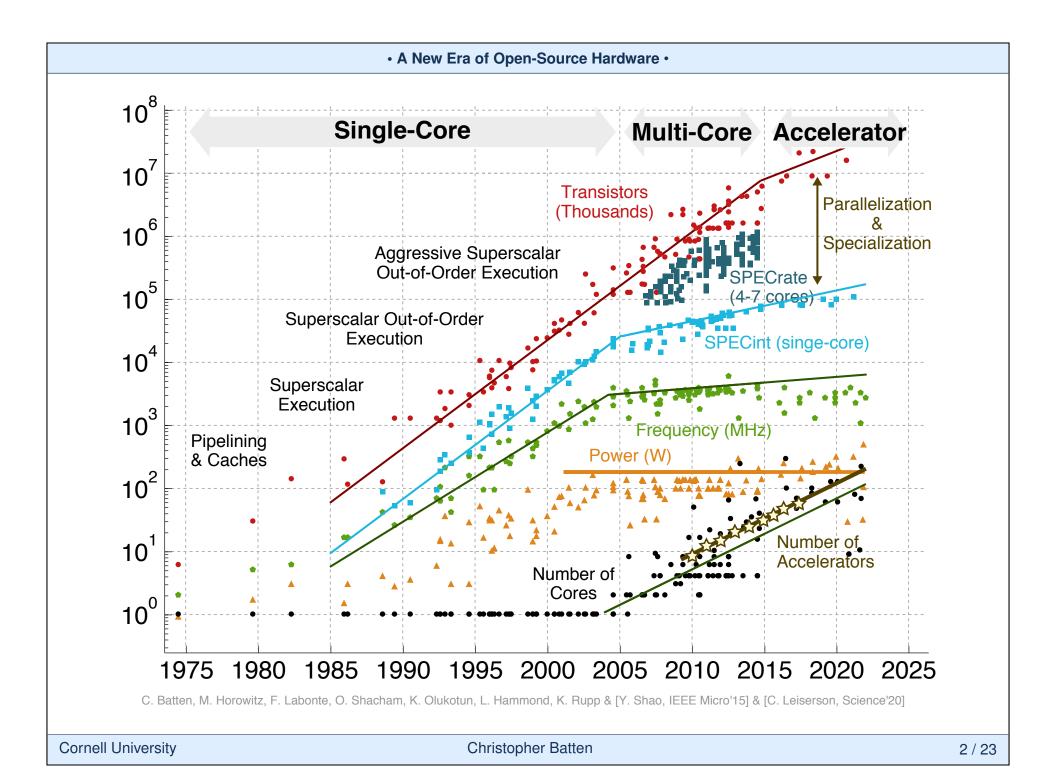
Christopher Batten

Computer Systems Laboratory Electrical and Computer Engineering Cornell University

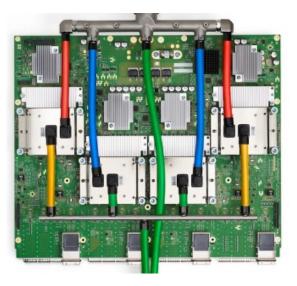


Accelerators for Machine Learning in the Cloud



NVIDIA DGX Hopper

- Graphics processor specialized just for accelerating machine learning
- Available as part of a complete system with both the software and hardware designed by NVIDIA



Google TPU v4

- Custom chip specifically designed to accelerate Google's TensorFlow C++ library
- Tightly integrated into Google's data centers



Microsoft Catapult

- Custom FPGA board for accelerating Bing search and machine learning
- Accelerators developed with/by app developers
- Tightly integrated into Microsoft data center's and cloud computing platforms

Accelerators for Machine Learning at the Edge



Amazon Echo

- Developing AI chips so Echo line can do more on-board processing
- Reduces need for round-trip to cloud
- Co-design the algorithms and the underlying hardware

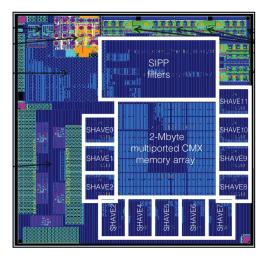


Facebook Oculus

- Starting to design custom chips for Oculus VR headsets
- Significant performance demands under strict power requirements



Movidius Myriad 2



Top-five software companies are all building custom accelerators

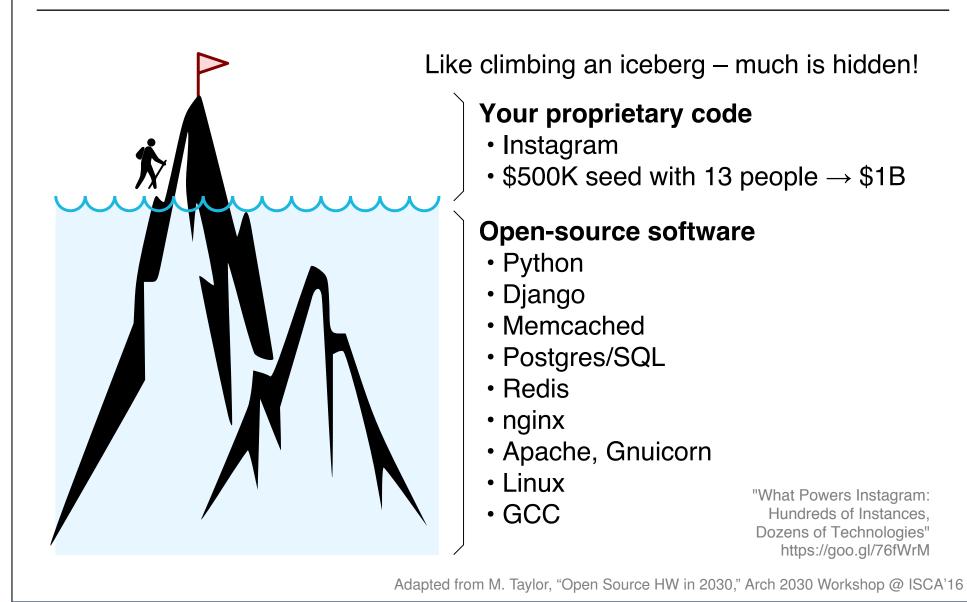
Facebook: w/ Intel, in-house AI chips
 Amazon: Echo, Oculus, networking chips
 Microsoft: Hiring for AI chips
 Google: TPU, Pixel, convergence
 Apple: SoCs for phones and laptops

Chip startup ecosystem for machine learning accelerators is thriving!

How can we accelerate innovation in accelerator-centric hardware design?

- Graphcore
- Nervana
- Cerebras
- Wave Computing
- Horizon Robotics
- Cambricon
- DeePhi
- Esperanto
- SambaNova
- **Eyeriss**
- Tenstorrent
- Mythic
- ThinkForce
- Groq
- Lightmatter

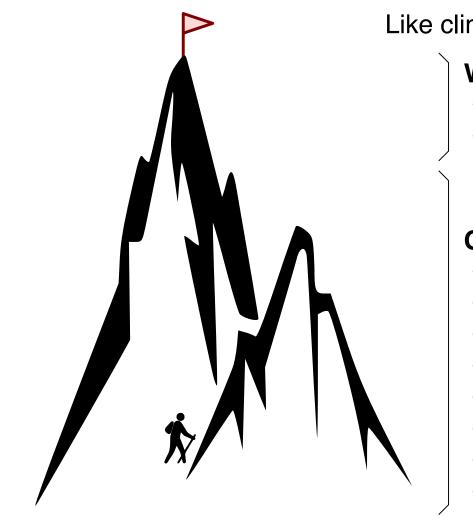
Software Innovation Today



Cornell University

Christopher Batten

Hardware Innovation Today



Like climbing a mountain – nothing is hidden!

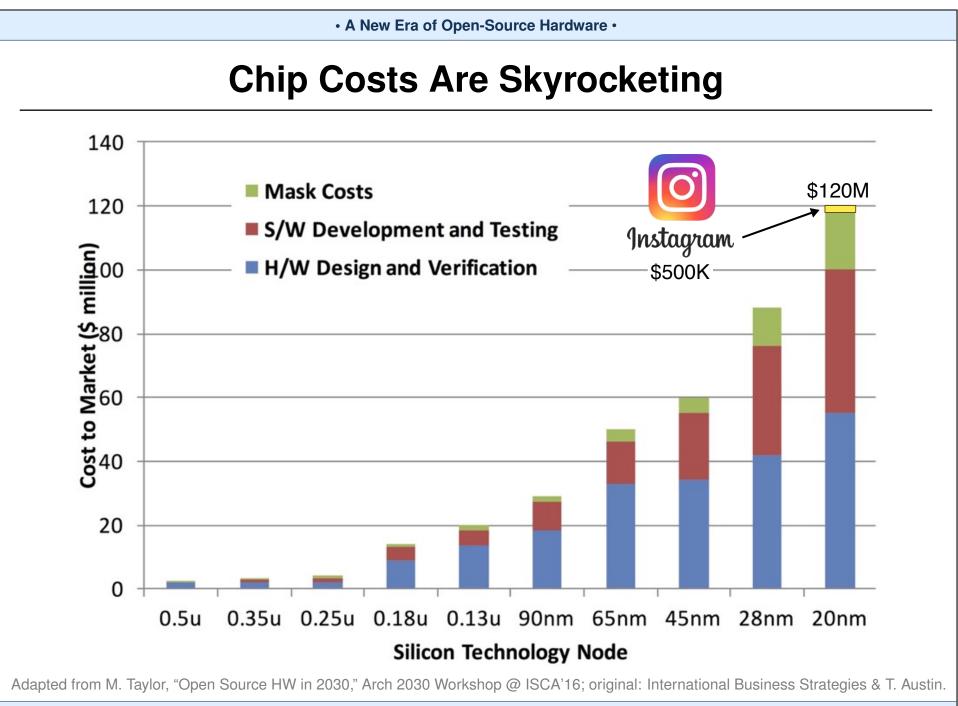
What you have to build

- New machine learning accelerator
- Other unrelated components, anything you cannot afford to buy or for which COTS IP does not do

Closed source

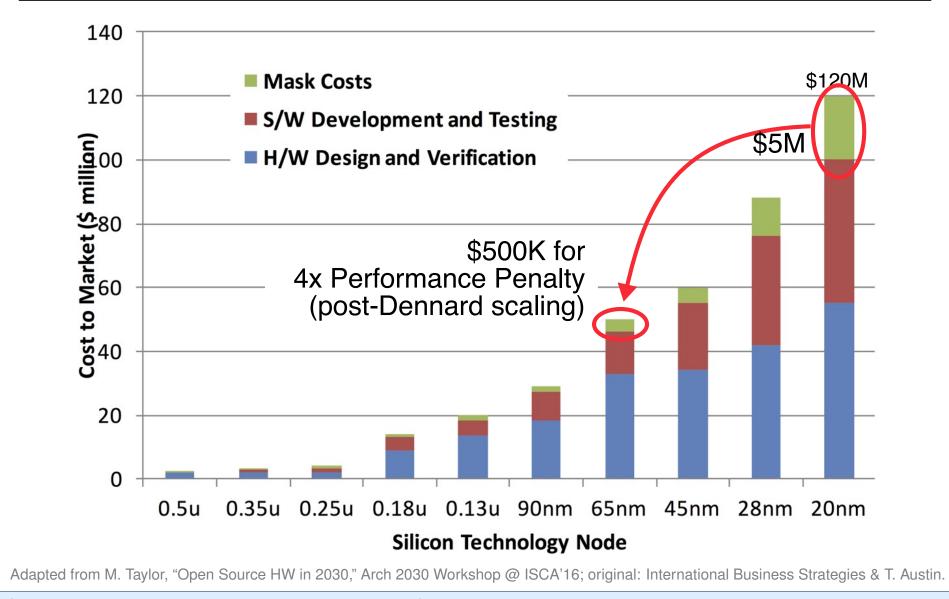
- ARM A57, A7, M4, M0
- ARM on-chip interconnect
- Standard cells, I/O pads, DDR Phy
- SRAM memory compilers
- VCS, Modelsim
- DC, ICC, Formality, Primetime
- Stratus, Innovus, Voltus
- Calibre DRC/RCX/LVS, SPICE

Adapted from M. Taylor, "Open Source HW in 2030," Arch 2030 Workshop @ ISCA'16



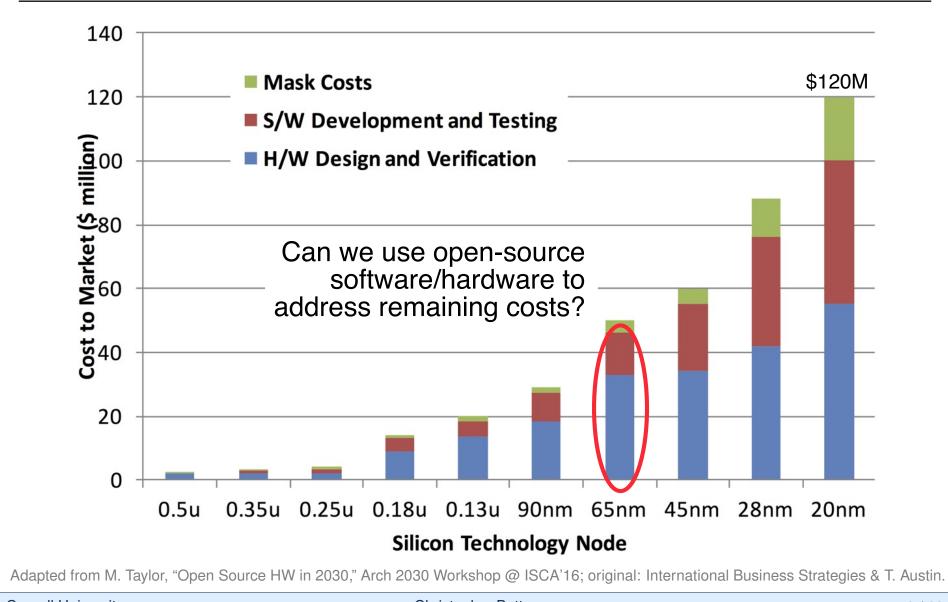
Cornell University

Minimum Viable Product/Prototype



Cornell University

Minimum Viable Product/Prototype



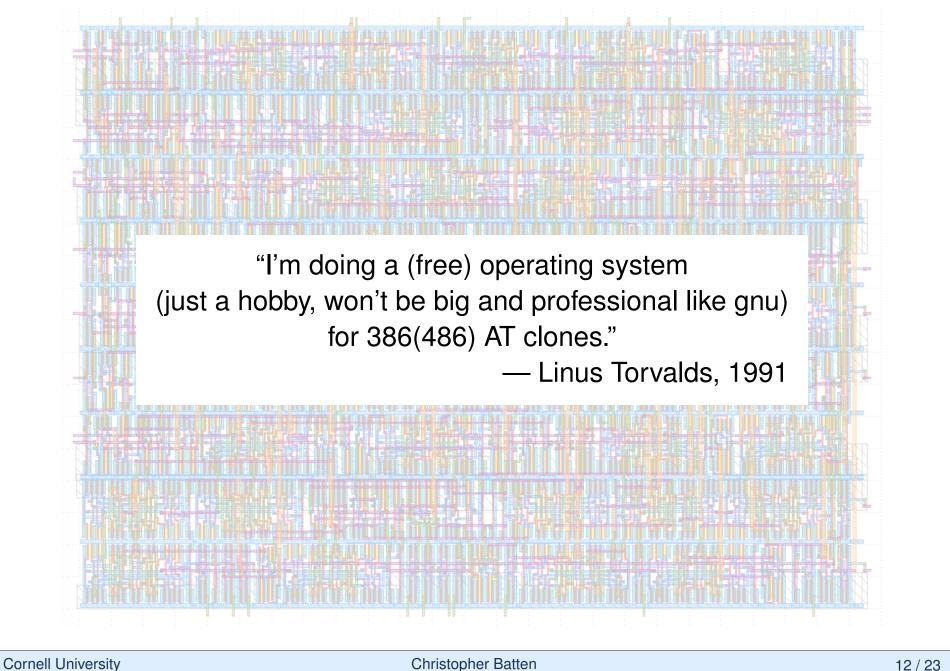
Cornell University

How can HW design be more like SW design?

Open-Source	Software	Hardware	
high-level languages	Python, Ruby, R, Javascript, Julia	Chisel, PyMTL, PyRTL, MyHDL, JHDL, Cλash, Calyx	,
libraries	C++ STL, Python std libs	BaseJump	
systems	Linux, Apache, MySQL, memcached	Rocket, Pulp/Ariane, OpenPiton Boom, FabScalar, MIAOW, Nyuz	
standards	POSIX	RISC-V ISA, RoCC, TileLink	
tools	GCC, LLVM, CPython, MRI, PyPy, V8	Icarus Verilog, Verilator, qflow, Yosys, TimberWolf, qrouter, magic, klayout, ngspice	
methodologies	agile software design	agile hardware design	
cloud	laaS, elastic computing	laaS, elastic CAD	
ell University	Christopher Bat	ten	10

```
# Ubuntu Server 16.04 LTS (ami-43a15f3e)
% sudo apt-get update
                                                           c. 2018
% sudo apt-get -y install build-essential qflow
% mkdir qflow && cd qflow
% wget http://opencircuitdesign.com/qflow/example/map9v3.v
% qflow synthesize place route map9v3 # yosys, graywolf, grouter
% wget http://opencircuitdesign.com/qflow/example/osu035_stdcells.gds2
% magic
                       # design def/lef -> magic format
>>> lef read /usr/share/qflow/tech/osu035/osu035_stdcells.lef
>>> def read map9v3.def
>>> writeall force map9v3
                       # stdcell qds -> magic format
% magic
>>> gds read osu035_stdcells.gds2
>>> writeall force
% magic map9v3
>>> gds write map9v3  # design + stdcells magic format -> gds
% sudo apt-get -y install libqt4-dev-bin libqt4-dev libz-dev
% wget http://www.klayout.org/downloads/source/klayout-0.24.9.tar.gz
% tar -xzvf klayout-0.24.9.tar.gz && cd klayout-0.24.9
% ./build.sh -noruby -nopython
% wget http://www.csl.cornell.edu/~cbatten/scmos.lyp
% ./bin.linux-64-gcc-release/klayout -l scmos.lyp ../map9v3.gds
```





	2 RISC-V International × +			
$\leftarrow \rightarrow$	C riscv.org		🗅 🖈 🕼 🛊 🗊 🗖 🥞	
	🌶 in 🖸 🔊 🗖 🜌	Languages \vee Tech Meetings Comm	nunity Meetings Working Groups Portal Join	
	RISC-V® About RISC-V	Membership RISC-V Technical / ~ ~ Exchange ~	News & Community Q Events ~ ~	
	What's New!		Check out our local language pages!	
	RISC-V°	RISC-V: The Open era	日本語のページをご覧くださ い!	
	RISC-V Announces First New Specifications of 2022, Adding to 16	of computing	访问我们的中文页面!	
	Ratified in 2021 RISC-V International RISC-V Community News Announcements, What's New	RISC-V: The Op	Get in touch!	
	Efficient Trace, Supervisor Binary Interface, Unified Extensible Firmware Interface, and Zmmul Multiply-Only Extension Accelerate	Computing RISC-V	Press: press@riscv.org	
Embedded- and Large-System Design. Six Additional Specifications Already In the Pipeline As Development Extends Into			Analysts: analysts@riscv.org	
	Vertical		General: info@riscv.org	

Industry Interest in RISC-V is Growing



Cornell University

Christopher Batten

RISC-V Hardware and Software Ecosystem

Open-source software: Gcc, binutils, glibc, Linux, BSD, LLVM, QEMU, FreeRTOS, ZephyrOS, LiteOS, SylixOS, ...

Commercial software:

Lauterbach, Segger, IAR, Micrium, ExpressLogic, Ashling, AntMicro, Imperas, UltraSoC ...

Software



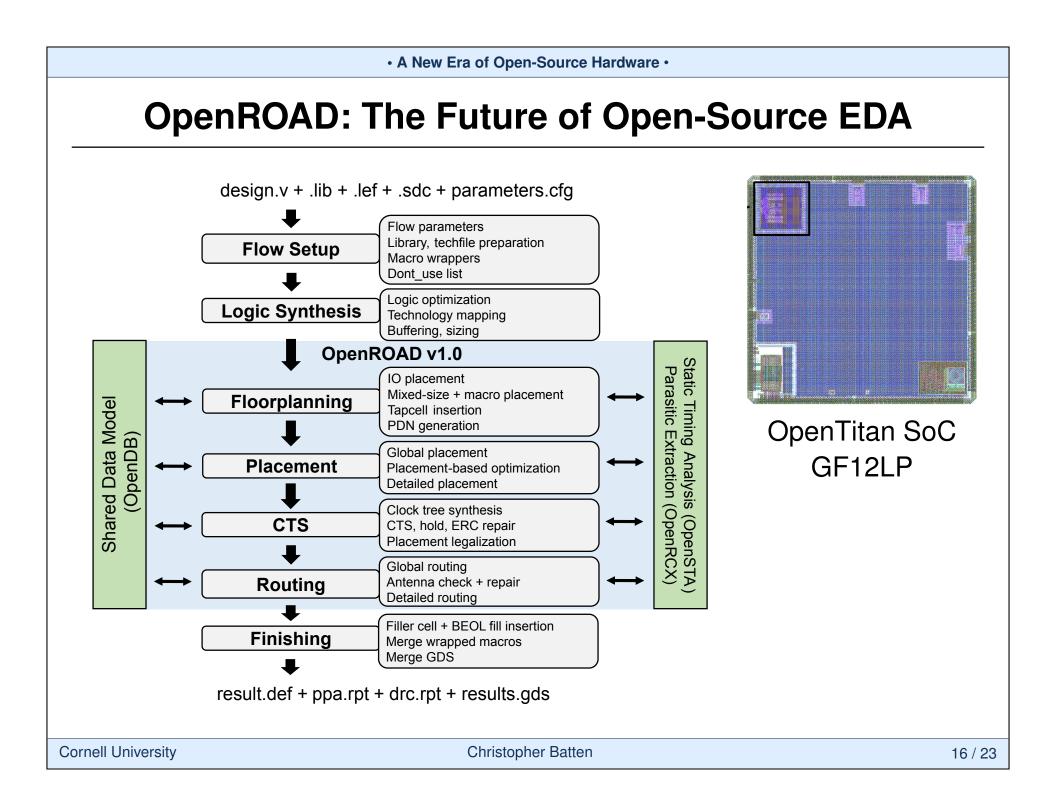
ISA specification Golden Model Compliance

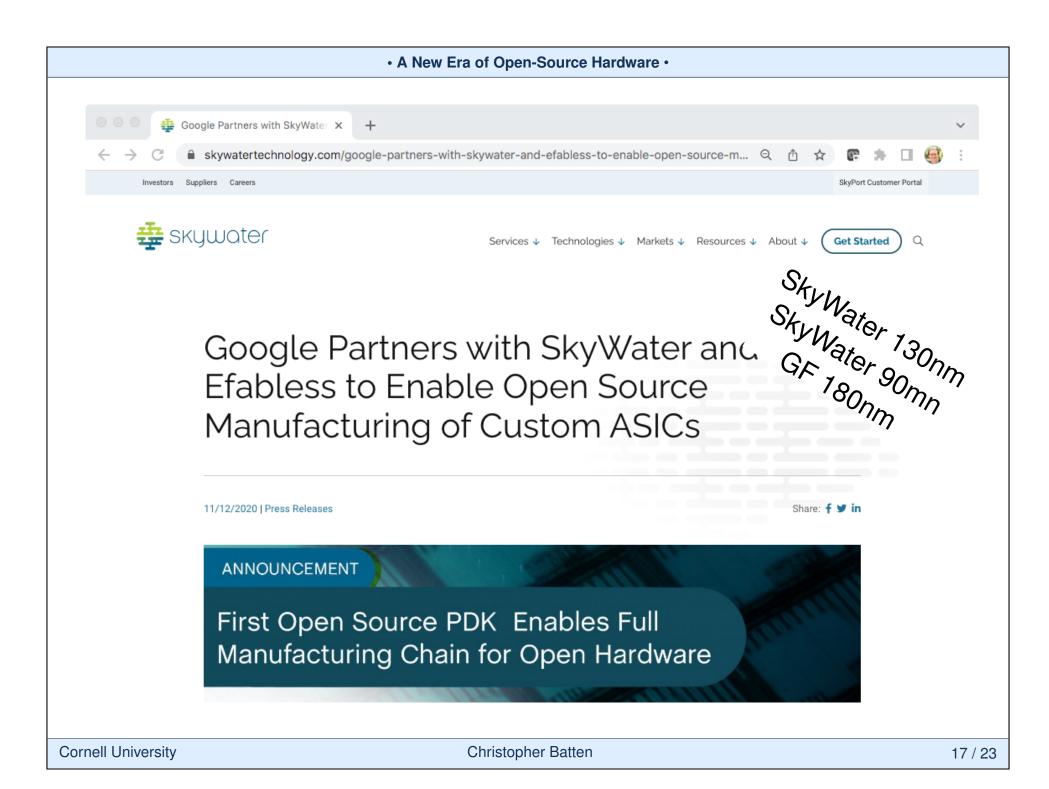
Hardware

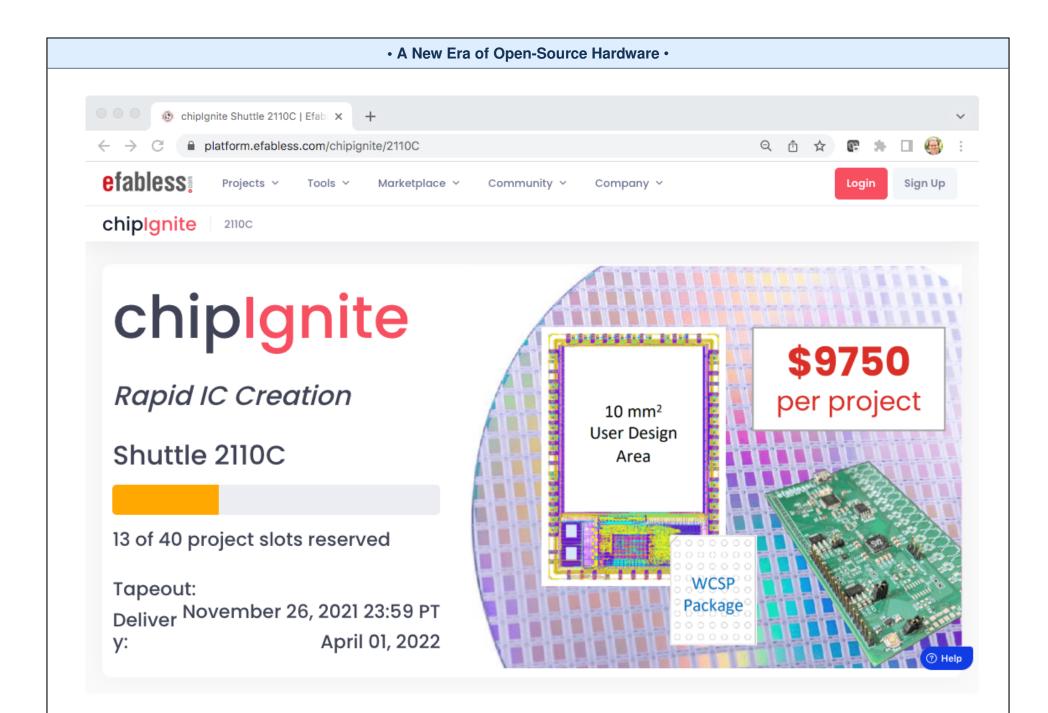
Open-source cores: Rocket, BOOM, RI5CY, Ariane, PicoRV32, Piccolo, SCR1, Shakti, Swerv, Hummingbird, ...

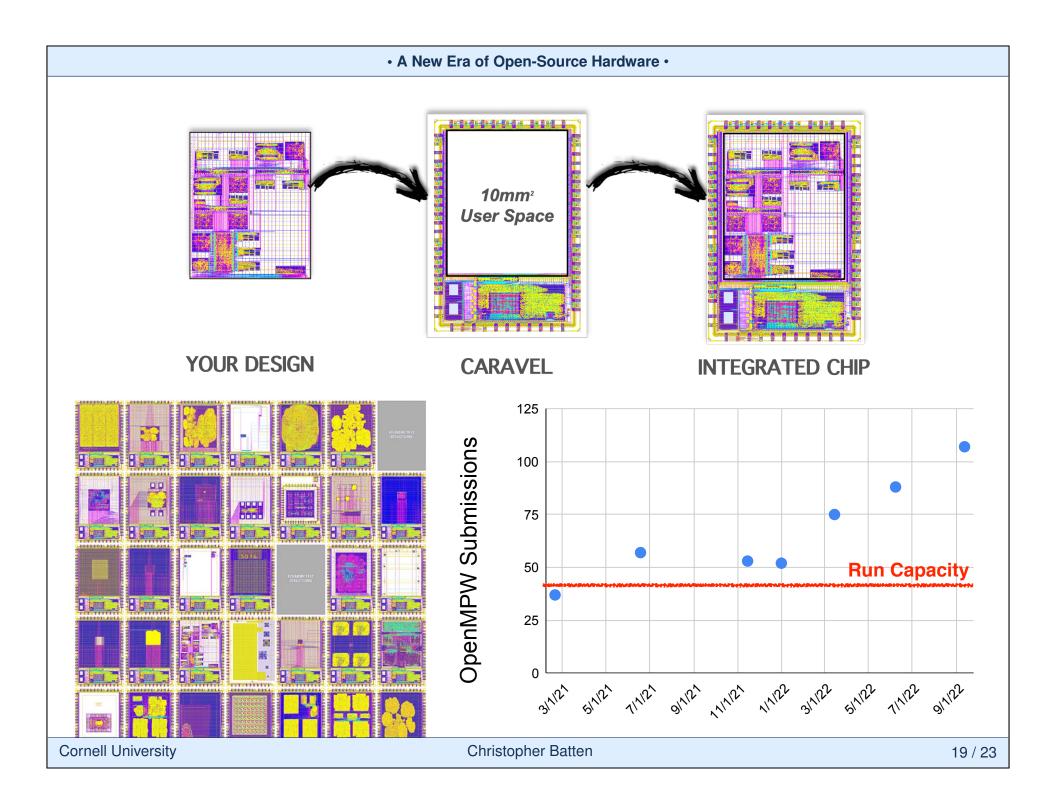
Commercial core providers: Andes, Bluespec, Cloudbear, Codasip, Cortus, C-Sky, InCore, Nuclei, SiFive, Syntacore, ...

Inhouse cores: Nvidia, +others









A New Era of Open-Source Hardware						
	🖄 Call for new engineering project	t team proposals — Exchange				
Associate Dean for Undergraduate Call for new engineering project team p To: ENGRFACULTY-L, ENGRFACULT Reply-To: Associate Dean for Undergr	proposals YINCSANDBEE-L, ENGRACADEMICS-L, Co		22, 2022 at 12:08 PM Details			

Colleagues,

Through a generous donor gift creating the **Shen Fund for Social Impact** we have the opportunity to fund multiple new engineering project teams. This program is designed to bring together new student teams under a faculty member's mentorship to address significant social challenges through novel and/or advanced engineering solutions. Falling under the Project Team Umbrella, the program will fund up to three new teams per year, with each supported for a three-year period at \$30K/yr. The teams will also be provided space and support to design and implement these projects.

Proposals may be submitted by either faculty looking to guide a group of students, or by students who will engage with a faculty member to form the teams.

Attached to this e-mail are three documents:

- Shen Fund FAQ Sp22.pdf: More fully describes the nature of the projects and the goals of the program (also copied to the e-mail below).
- Shen Fund Proposal Template Sp22.docx: Short project proposal form.
- Shen Funded Projects Summary_Sp22.pdf: A summary document of a currently funded teams.

The ideal project will likely develop through discussions with Lauren Stulgis (as director of the project teams) and me. Feel free to reach out to us with rough ideas and concepts and we can help to try to develop a viable proposal.

Proposals will be considered as they arrive, with discussions to strengthen each within the program constraints. The initial application is a simple document identifying the primary goals, technical challenges and plans, timeline and budget, and currently engaged personnel.

Proposals must be uploaded directly to Box by email to: Proposa.zeuyhp9wqg5p8teo@u.box.com. The first round of decisions will be made based on submissions received by 11:59pm on Sunday, March 13, 2022.

Again, please feel free to contact me or Lauren Stulgis with any questions or to discuss potential projects.

Prof. Alan Zehnder Associate Dean for Undergraduate Programs 177 Rhodes Hall Phone: (607) 255-9181 email: <u>eng_ugdean@cornell.edu</u>

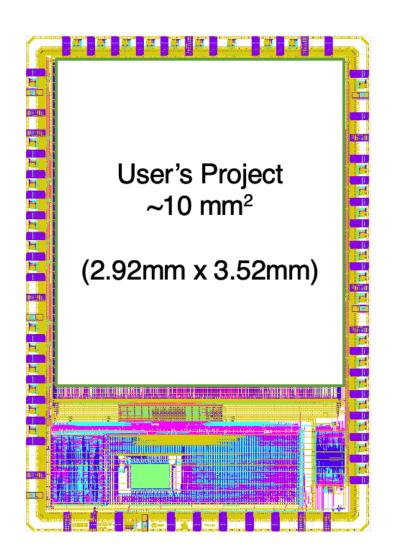
C2S2: Cornell Custom Silicon Systems Project Team

Three-year student-led project team to tapeout a custom chip in SkyWater 130nm to implement a proof-of-concept system for a campus partner

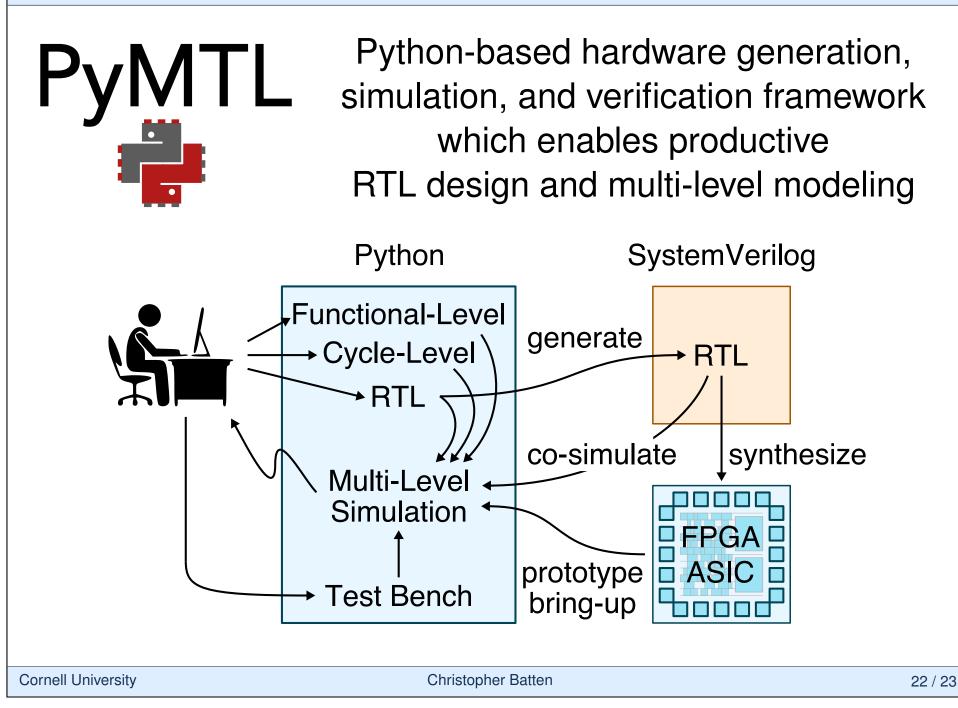
- Open RISC-V ISA
- Open-Source VexRISCV microcontoller
- Open-Source OpenROAD chip flow
- Open PDK for SkyWater 130nm
- OpenMPW + ChipIgnite w/ efabless

100+ applications \rightarrow 25 team members

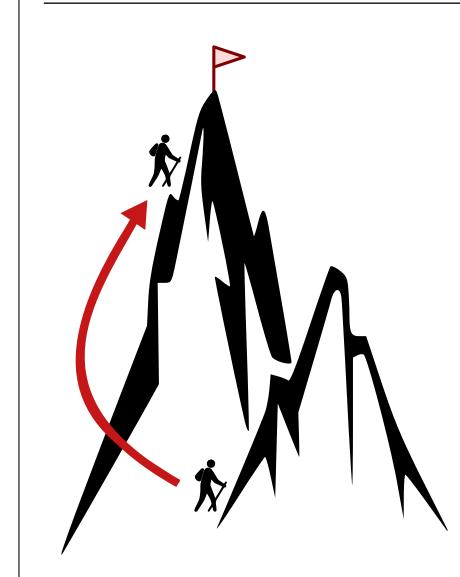
- Digital & Verification Subteam
- Analog Subteam
- Software Subteam
- System Architecture Subteam







A Call to Action



- Open-source hardware needs developers who
 - ▷ ... are idealistic
 - ▷ ... have lots of free time
 - ▷ ... will work for free
- Who might that be?

Students!

Academics have a practical and ethical motivation for using, developing, and promoting open-source electronic design automation tools and open-source hardware designs