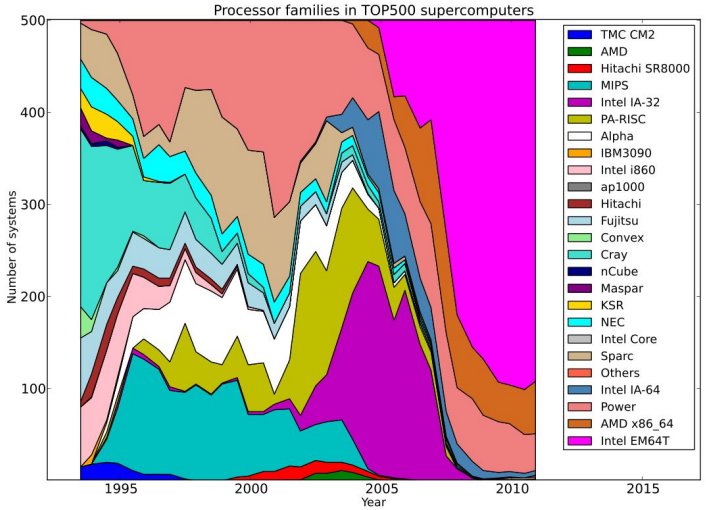
New and Future Processors

Arthur Vidineyev
CS462 - Parallel Programming
December 08, 2017

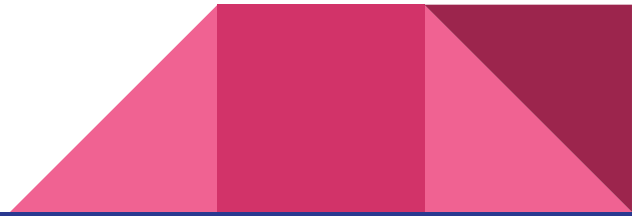
Current State

- General purpose CPUs - x86_64
- Graphics and simple parallelizable computation - GPU accelerators
- Specific or energy efficient workloads - FPGA/ASIC
- Mobile or embedded systems
  - ARM - Mali, Adreno, PowerVR GPUs
  - AVR
  - MIPS
  - PIC



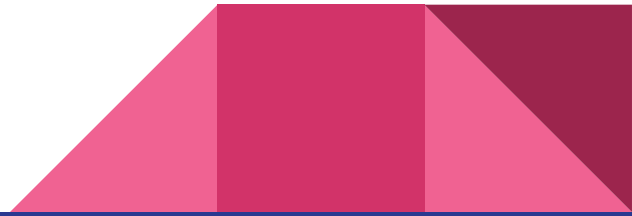
Possible Computing Paths

- Different CPU architectures
  - Take advantage of new research in compilers
  - Simplify hardware
  - Get rid of legacy instructions
- Greater parallelization facilitated by OS
  - Android interface and big.LITTLE architecture
  - Tasks can’t all be parallelized
- Offloading compute tasks
  - Accelerator cards
  - Purpose built chips



Emergent

- ARM Servers
- Xeon Phi
  - Many simpler x86 cores related to Atom
  - Provides AVX instructions
  - OpenMP, OpenCL
- Hardware optimized for AI/ML
  - Nvidia since Pascal, Radeon Vega offering fp16
  - Intel Nervana built for neural networks
- OpenPOWER and custom Xeon
  - Talos II workstation



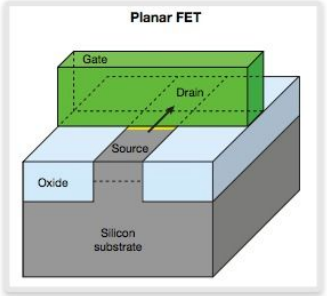
Future

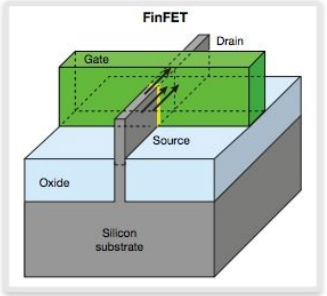
- RISC-V
  - Open source, scalable architecture
  - Similar in strategy to ARM, MIPS but open and no licensing fees
- Memristor based neuromorphic computing
  - Systems closer to brain structure
  - Remembers most recent magnitude, polarity, duration of voltage applied
  - Allows for more NV storage
- Quantum computing
  - Dwave uses quantum annealing via superconducting logic elements
  - Finding the global minimum of a given objective function
  - Currently not a good speedup over classical computing
- Transistor technologies

RISC-V

- Scalable, open-source ISA
  - Academic origin, used BSD license
  - Supports 32, 64, 128 bit words
  - Load/store architecture - PowerPC, SPARC, ARM, MIPS, GPUs
  - Vector instructions rather than fixed SIMD (i.e. x86)
  - Extensions picked based on use case
- Microcontrollers
  - SiFive, Western Digital, Nvidia
- Software support
  - GCC/Newlib toolchain
  - Supported by FreeBSD and upstream Linux 4.15

Transistor Tech





- Gate all around fets
  - Demonstrated by IBM in 5nm process
- Tunnel field-effect transistor
  - Switching allows tunneling through barrier instead of channel formation
- Carbon nanotube
  - Higher current and faster switching
- Novel materials
  - III-V semiconductors (e.g. indium gallium arsenide)
  - Germanium instead of silicon

New and Future Processors

Arthur Vidineyev
CS462 - Parallel Programming
December 08, 2017
avidiney@vols.utk.edu