RISC-V Chip for Mobile DNA Sequencing

Sebastian Magierowski

Electrical Engineering & Computer Science Department

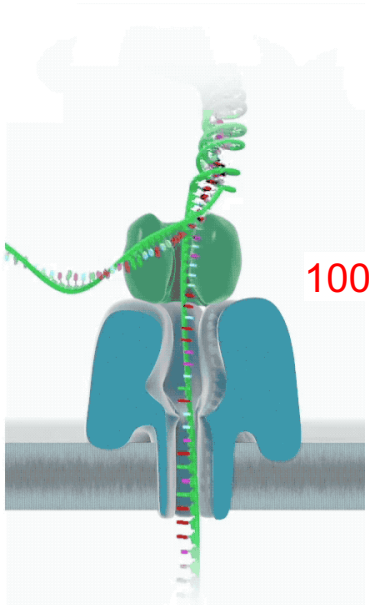
York University

Nutshell

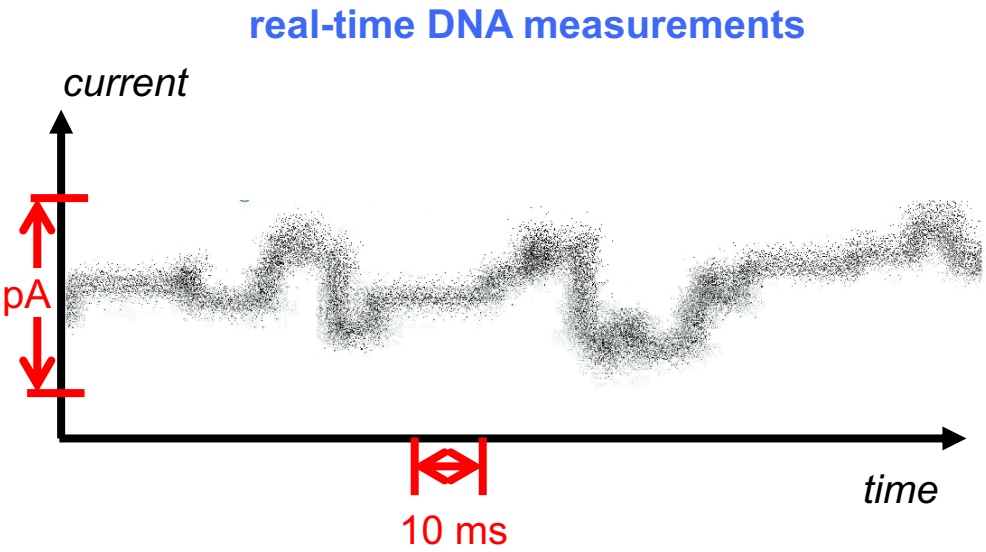
- We’re interested in ASIC for embedded molecular measurements
- We’re planning a RISC-V based SoC
  - Rocket + HW accelerator for now
- Lots of Open Source stuff out there to help
  - documentation might be good
  - but we’re still having a hard time working through it

Background

- There's a "simple" way to measure some molecules

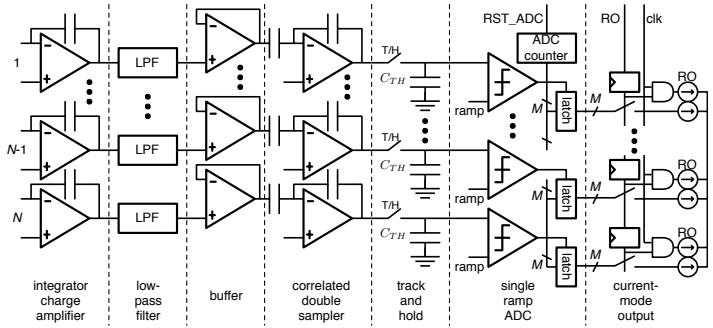


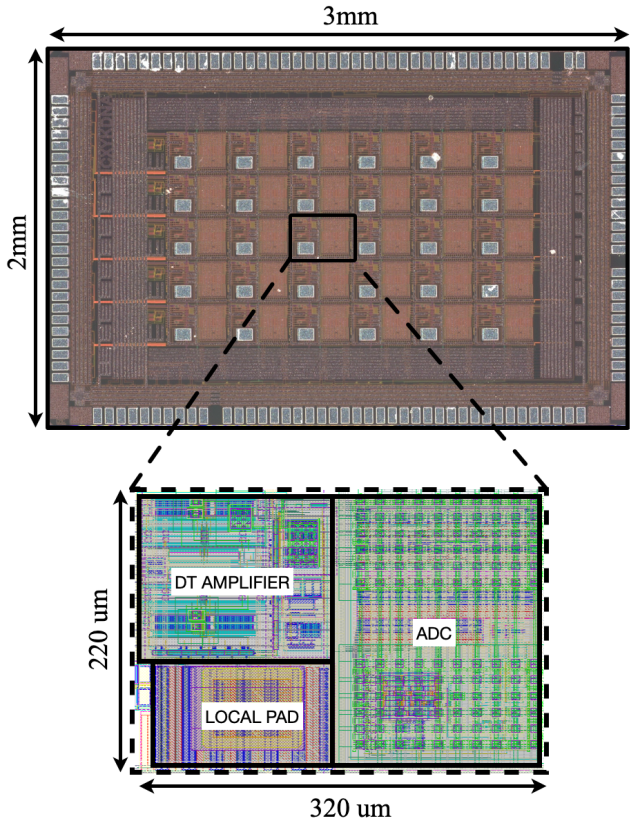
[©Oxford Nanopore Tech]



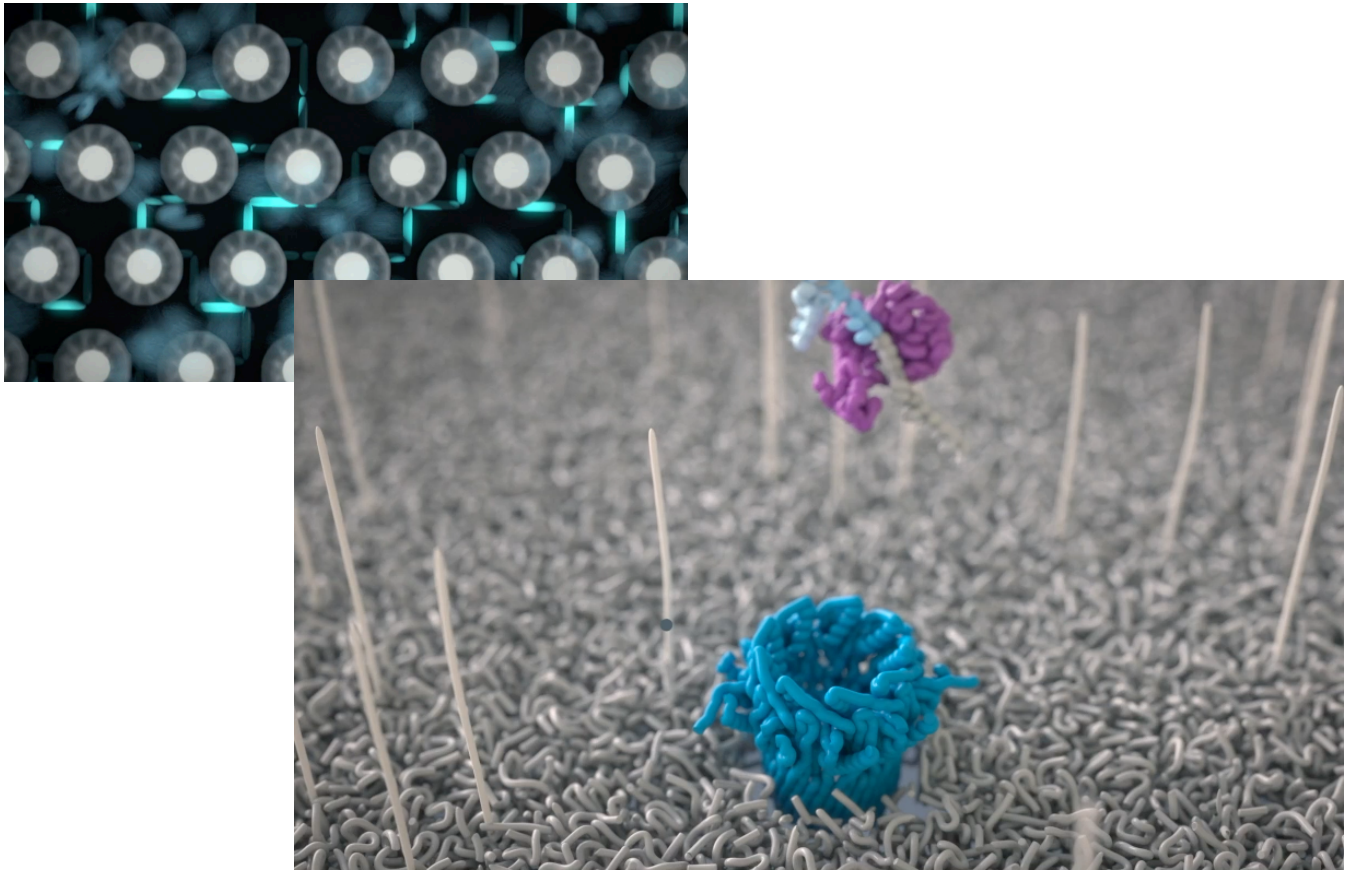
Sensing

- Getting and processing this current is difficult!





Some Fun Details

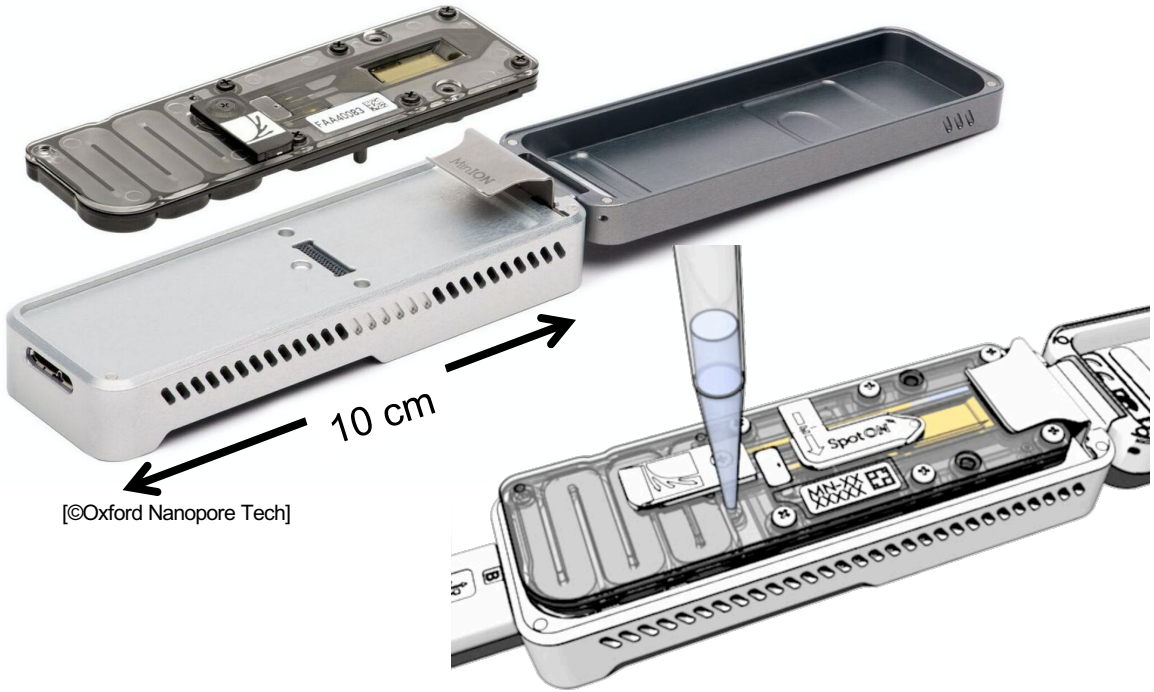


[©Oxford Nanopore Tech]

Small DNA Meters

- These sensors can be packaged in a small footprint
  - ~500 channels per cm^2

90 grams

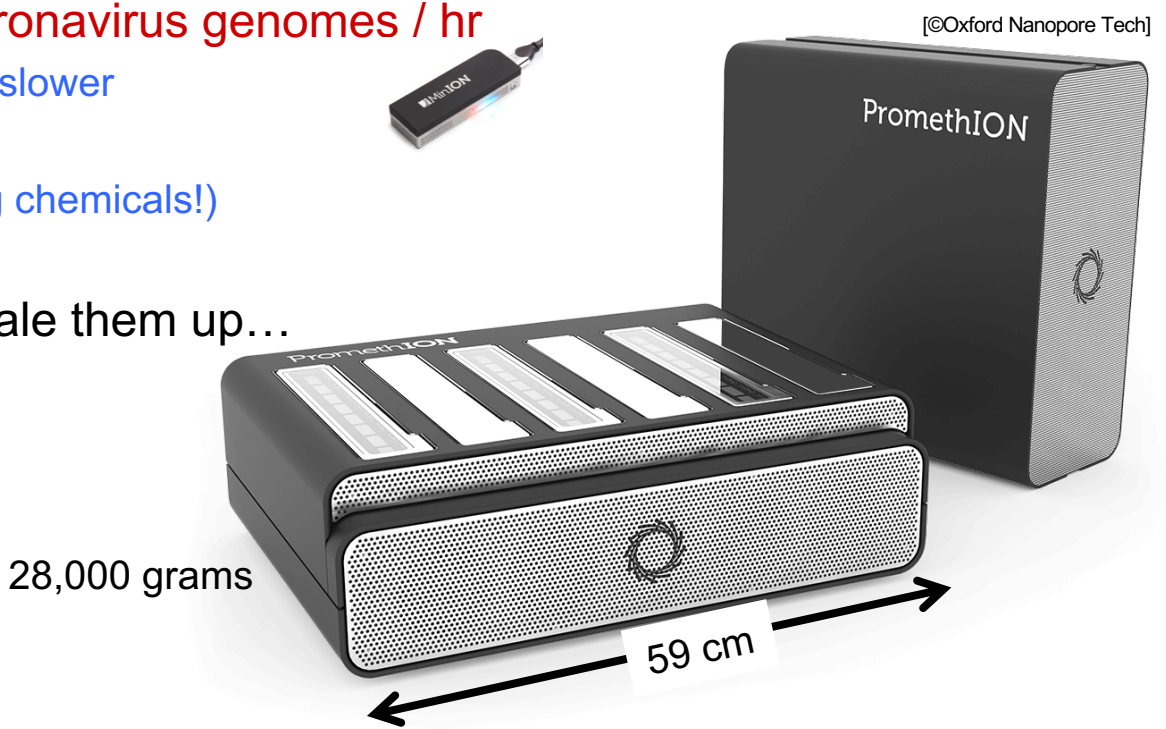


[©Oxford Nanopore Tech]

Fast DNA Meters

- And they can measure a lot (per cm^2 of sensor area)…
  - ideally: ~0.25 human genomes / hour
    - ~25,000 coronavirus genomes / hr
  - realistically: ~2X slower
  - 1 Watt
  - $1k (not counting chemicals!)

- And you can scale them up…
  - ~550X faster
  - 2 kWatt
  - $285k



And They Produce Lots of Data

- Generate ~ 10 GB / hour
  - ~ 3 HD NETFLIX streams

- This is handled by an external computer
  - only analog and USB comms in the box

[©Oxford Nanopore Tech]

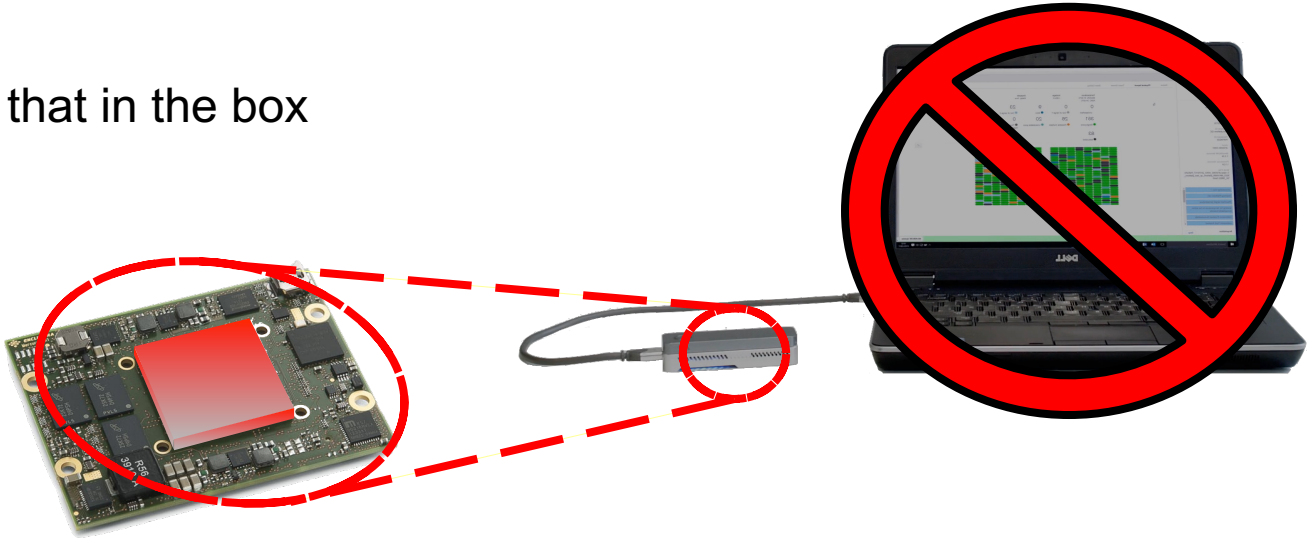
- The computer isn't there just to store the measurements
  - it needs to compute

Bioinformatics in the Box

- A bunch of signal processing needs to be done on the measurements

[©Oxford Nanopore Tech]

- We want that in the box

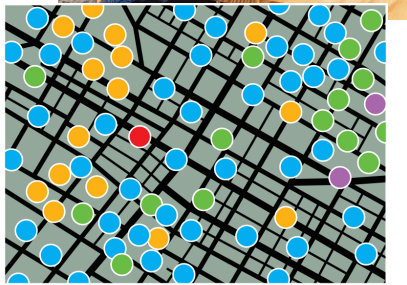


Broadened Applications

- Field genomics

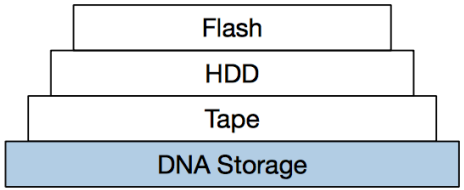
[©The Economist]

- Pervasive sensing



[©Wired]

- New IT tech

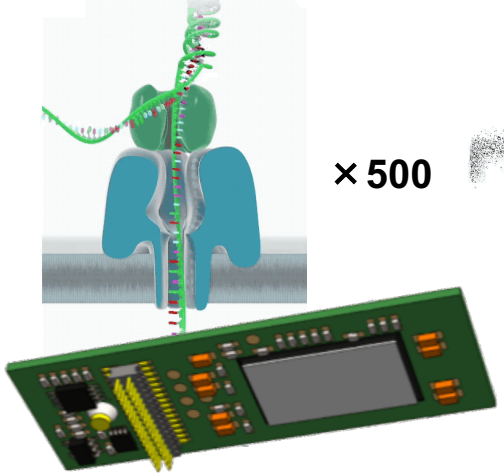


	Access Time	*Durability*
Flash	ms	~5 yrs
HDD	10s ms	~5 yrs
Tape	minutes	~15-30 yrs
DNA Storage	10s hrs	centuries

[©Microsoft]

Basecalling in the Box

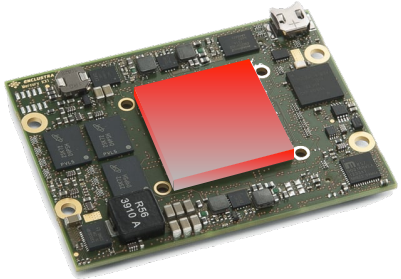
- We want basecalling in the box
  - current-to-text conversion



[©Oxford Nanopore Tech]

× 500

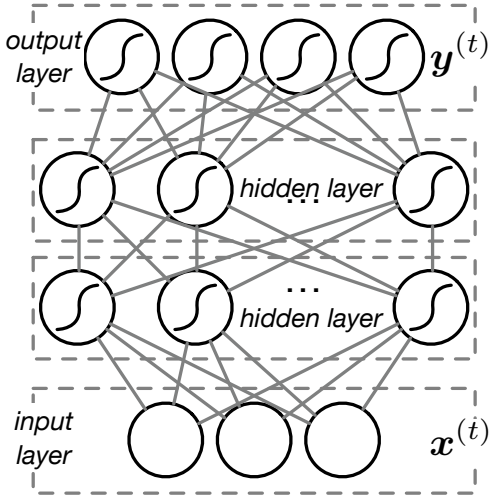
basecalling



:GTTGATTTGCTGGGGGCAGGGCCG

Real-Time Basecalling

- But real-time basecalling is very compute intensive
  - HMM Basecallers
    - ~ 70 GOPS
      - 60-70% accuracy
  - DNN Basecallers
    - ~ 250 GOPS
      - 80-90% accuracy
  - e.g. heard real-time basecalling work ok on a $25,000 machine
    - 24-core; 500-GB RAM

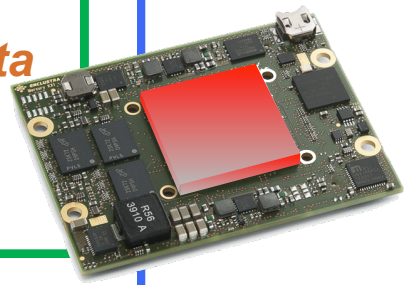


- And power hungry
  - probably need ~2 kW to keep up with full-throttle MinIC
    - all 500 channels continuously working

[©Oxford Nanopore Tech]

Basecalling Algorithm

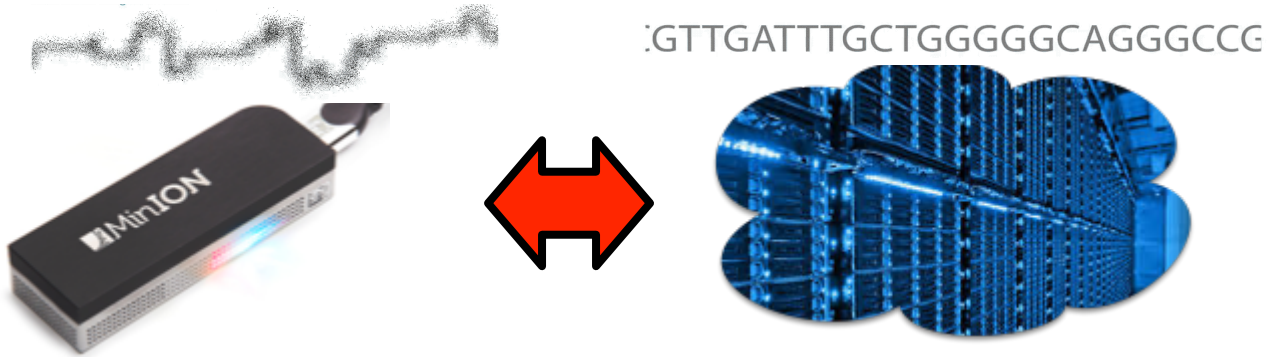
```
for: L = 0 to # of channels ~ (1 – 10^3)
 for: i = 0 to # of events/DNA strand ~(10^3 – 10^6)
  for: j = 0 to # of states/model ~(10^2 – 10^4)
   for: k = 0 to # transitions ~(4 – 10^2)
    load T(k,j,L)
    calc E(j,event(i))   <- measured data
    calc P(k,j) = T × E
   end
   calc P(j) = max{P(k,j)}
  end
  calc max{P(j)}
 end
end
```



30B iterations / sec. → 70 GOPS

Maybe just 5G or 6G It?

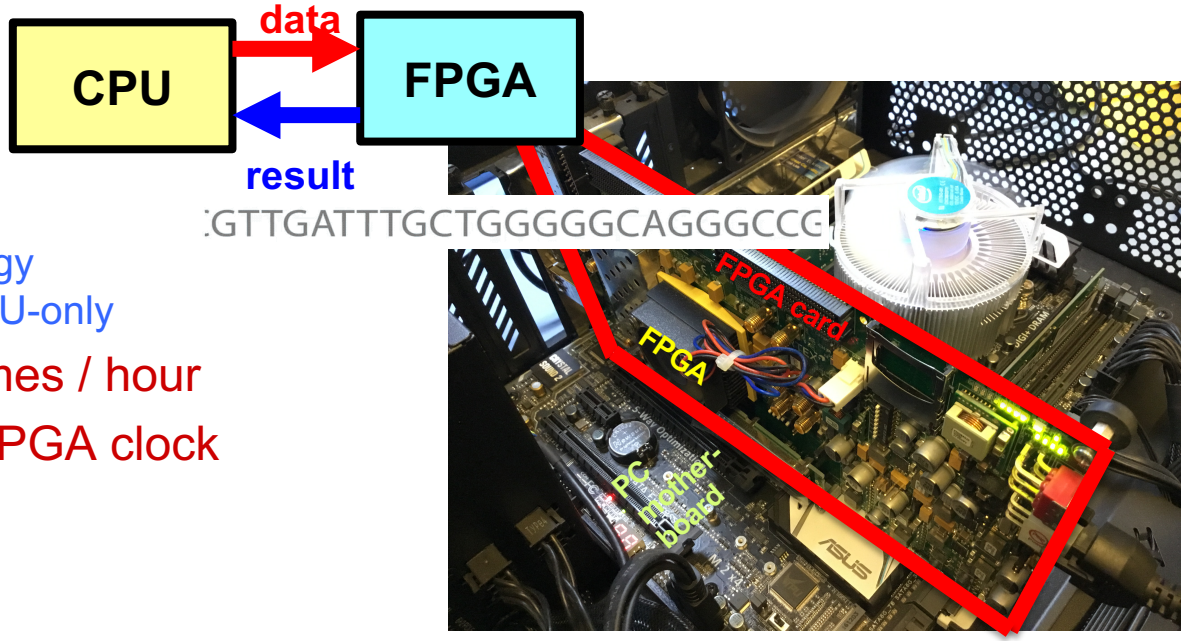
- Basecall in the cloud?



  - comms isn't cost/energy free
    - how will you know what to send and what not to send?
  - neither is the cloud
    - $5/hr on AWS for real-time facilities
      - ~$40/genome
    - how many measurements before its better to invest in an embedded computer?

FPGA Basecaller

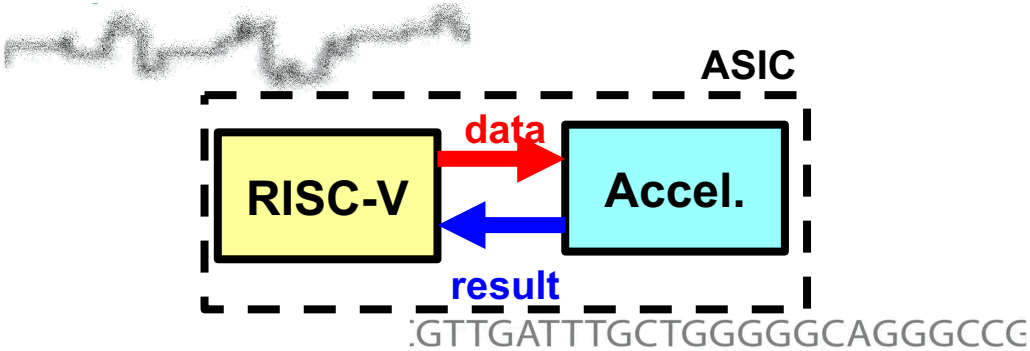
- We think a basecaller ASIC has a lot of potential
  - at least FPGA accelerators have looked pretty good



  - 550X more energy efficient than CPU-only
    - 0.45 genomes / hour
    - 250 MHz FPGA clock
    - 6 W

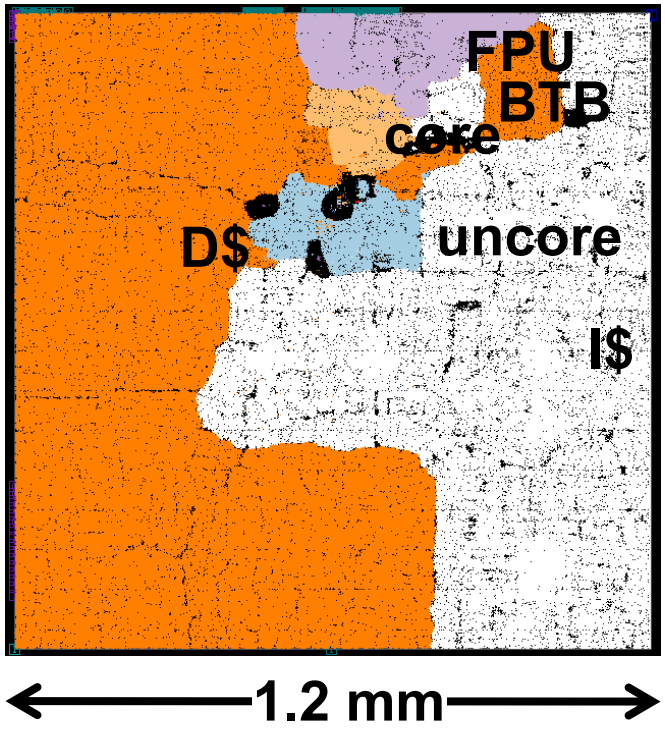
RISC-V Basecaller

- We want the same in an ASIC
  - HW accelerated RISC-V



RISC-V So Far

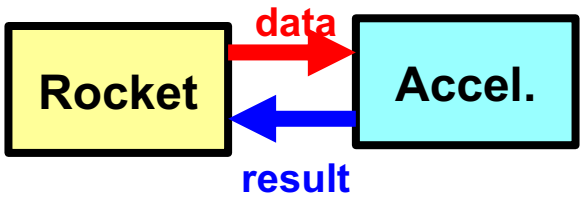
- RV64G ISA
  - 64-bit
  - integer multiplication and division
  - single & double precision FP
- Open source Rocket
  - from Chipyard
    - Berkeley Arch. Research
  - from Bespoke Silicon Group
  - I$ 4 KiB; D$ 4 KiB
    - 64 set, 4-way
- Technology
  - FD-SOI GF 22-nm
  - SSS (0.72 V) → 500 MHz
  - FFF (0.88 V) → 950 MHz

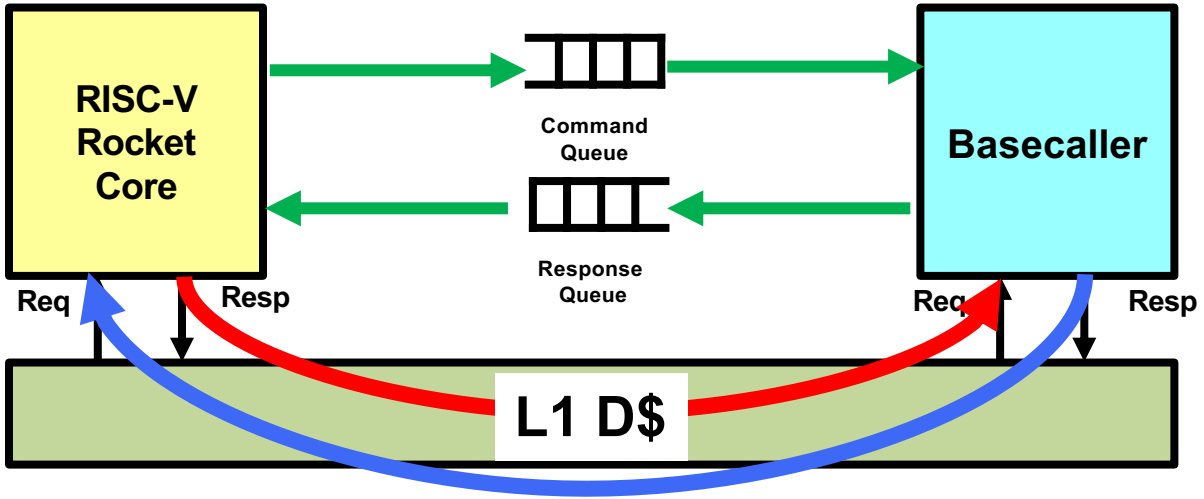


Rocket

Accelerating RISC-V

- Lots of open-source help to facilitate HW acceleration



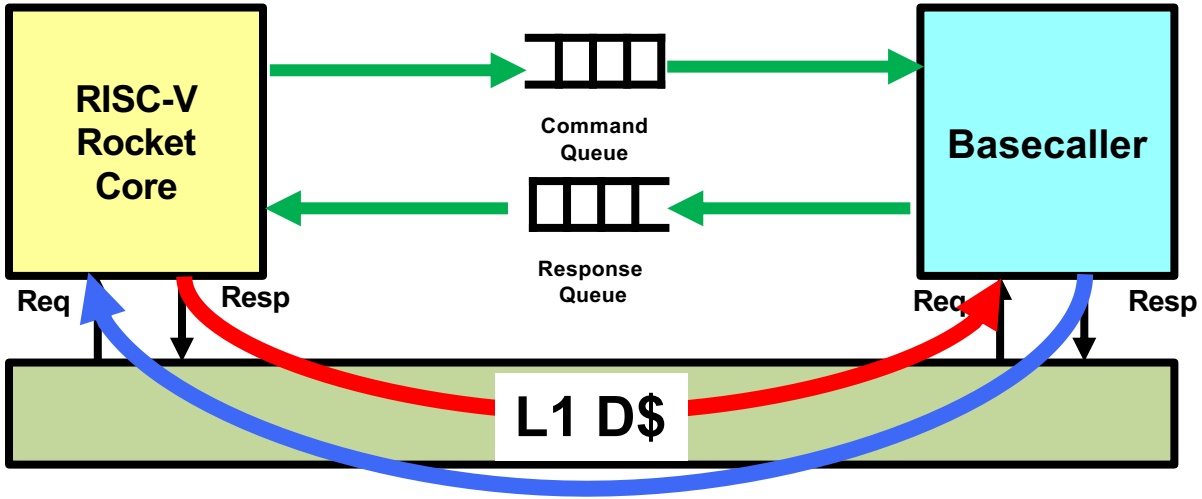


Open-Source Help

- RISC-V
  - but also special instructions for core/rocket commands (RoCC)
    - and means for accelerators to process those instructions

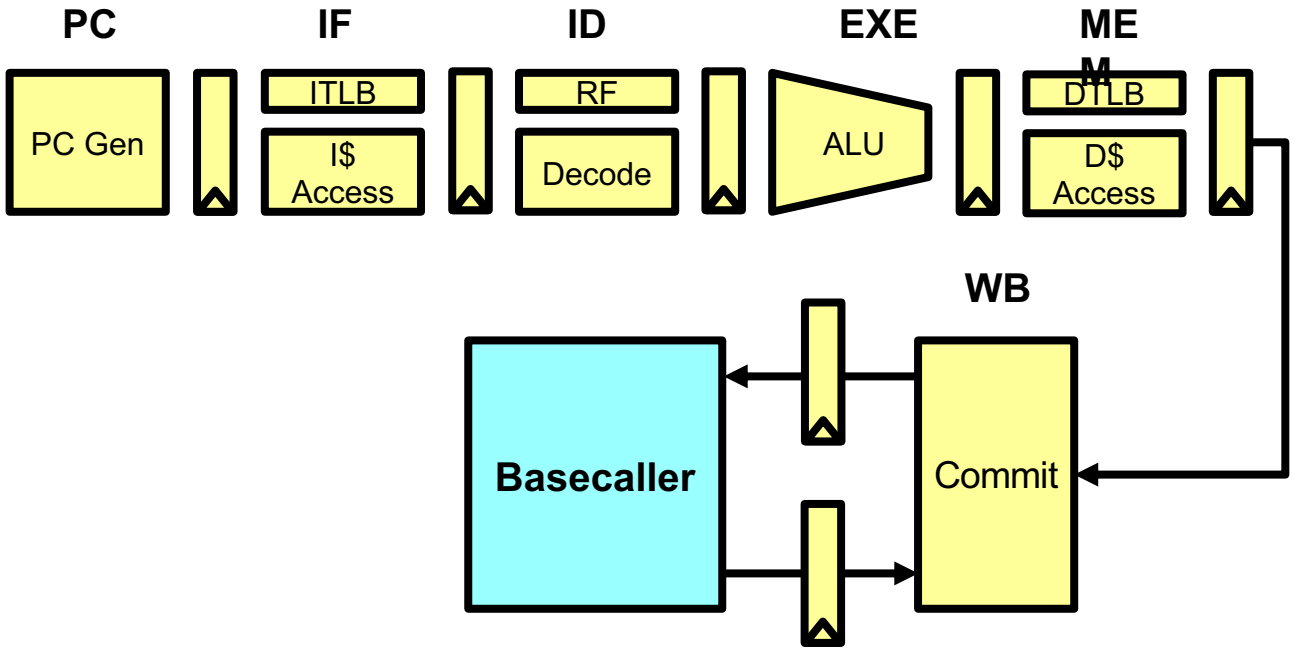
funct7	rs2	rs1	xd	xs1	xs2	rd	opcode
7	5	5	1	1	1	5	7

- Cache
  - and means for the accelerator to talk to the cache



Rocket Acceleration in a Bit More Detail

- RISC-V Rocket acceleration scheme

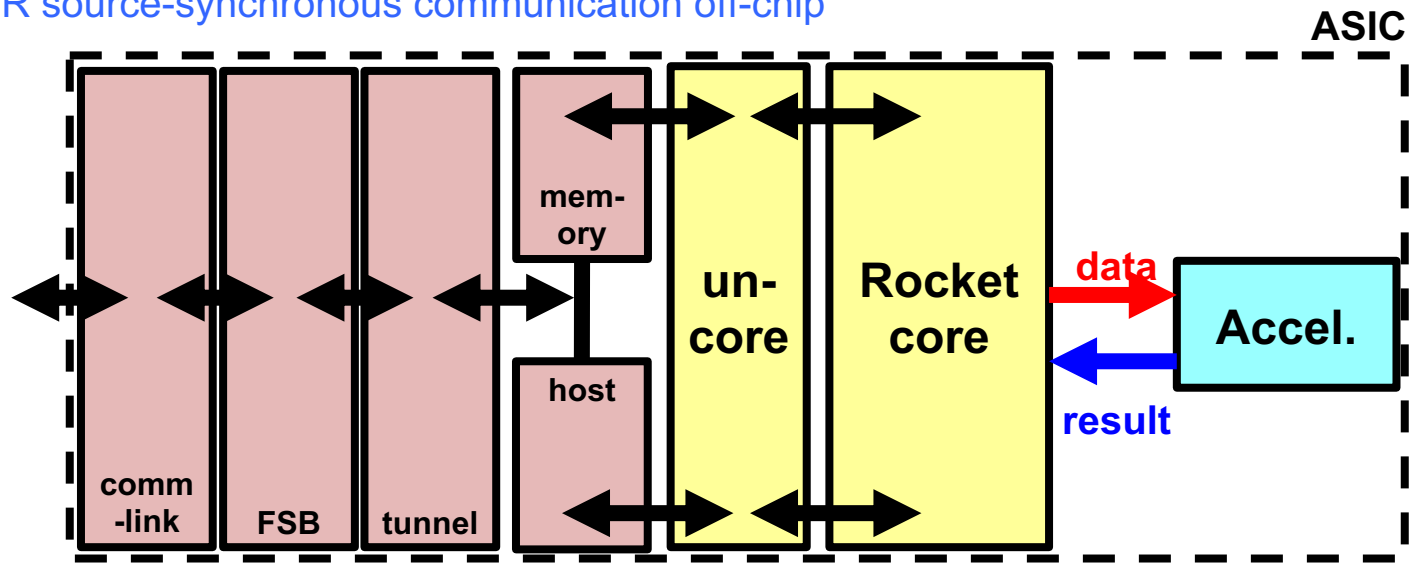


Accelerated Performance

- First (very rough) impressions look good
  - at 500 MHz clock
  - 50 mW
  - 0.5 genomes / hour
    - ~ 100X more energy efficient than FPGA-accelerated
  - room for improvement?
    - running at CPI ~ 3

Rocket: Talking to the Outside World

- Open-source here too
  - System Verilog communications blocks from BSG
  - DDR source-synchronous communication off-chip



FPGA Gateway and Bridge

- Board designs available from Bespoke Silicon Group
  - 400 Mbps data rates