Luis Ceze

computer architecture, systems, PL
Planet-scale, realtime?
Estimate capture capability of the world in 2030: 12B streams, 6 x 10^{17} pixels = 600 quadrillion pixels.

Planet-scale DNA sequencing and analysis

What about processing?
Sequencing and alignment of human genome: ~12 hours on a beefy server.
~100M servers worldwide => ~35 days to crunch genome for 7B people.
(EC2 instance suitable for genomics: $2.66/hr. $200B+ just in compute power)

What about clinical sequencing? And non-human genomic analysis? Maybe computational cost of medicine in general?
heterogeneity is natural

spatial fabrics programmed in space

circuits/devices

compute + storage deeply integrated

noise and nondeterminism is unavoidable

architecture

machine learning

model

data
don’t know the code

program synthesis

spec, sketch, examples

code

HW/SW interface too complex

Based on a slide from Vijay Ganesh
Blur distinction between HW and SW

System design specification language

- Machine Learning: don't know the code
- Model from data
- Spec, sketch, examples
- Program synthesis: HW/SW interface too complex
- Code
- Glue: Compose, reason about uncertainty & quality

Architecture
- Heterogeneity is natural
- Spatial fabrics programmed in space
- Compute + Storage deeply integrated
- Noise and nondeterminism is unavoidable

Circuits/Devices
- Right primitives?
- Behavior model?
Specialization

compute

storage

communication

accelerators, reconfigurable logic

tune to data type

tune coding to data type

source-channel joint coding

topology to match app

quality vs. resource utilization trade-offs

@Approx
int a = ...;

@Precise
int p = ...;

passert a != b, P, C;
A storage gap?

[YouTube Upload Rate graph]

[Disk cost-per-byte graph]

[Dense 3D molecular storage]

[Credit: David Rosenthal (CMU) and Preeti Gupta (UCSC), 2014]