Computers as Information Channels

A Common Yardstick for the Post-Moore Hardware Zoo

Max Hawkins and Rich Vuduc (advisor) - Georgia Institute of Technology

The Problem: Heterogeneity

Computing innovations bring new, incomparable performance metrics.

- Datatype variety —> Bit-width scaling (regardless of data used)
- Emulation —> Number of slices
- Analog —> MAC/s
- Neuromorphic —> Spike/s

How do you compare a GPU, a noisy analog device, and a ternary computer?

Application-specific evaluation works but is narrow.

Our Proposal

Information theory grounds communication performance, and it can work for computation too.

- Communication and computation both map inputs to outputs.
- Communication is restricted to the identity function, while computation allows arbitrary mappings.
- Both allow for info-theoretic analyses:
  - Entropy, H(X), quantifies the average uncertainty of input/output states (e.g. bit strings) seen.
  - Mutual information, I(X;Y), measures the information transformed and retained across a channel.
  - Channel Capacity, C, upper bounds I. $C = \max_{p(x)} I(X;Y)$
  - **Peak Throughput (bit/s) = Capacity (bit/op) * Rate (op/s)**

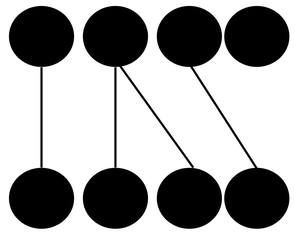
$$\mathrm{H}(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x) \qquad I(X;Y) = H(X) - H(X|Y)$$

Throughput Examples of FP64 vs Emulation

- Noiseless, digital HW with 64-bit output: $C \leq 64$ bit/op
  - *Bit-width scaling assumes $I = C$, which is usually not true.*
  - Throughput: 64 bit/op * 1.1 Tflop/s = 70.4 Tbit/s
- Ozaki emulation (OzIMMU with S=12): $C \lesssim 58$ bit/op
  - Throughput: 58 bit/op * 1.175 Tflop/s = 68.2 Tbit/s

Noisy Communication

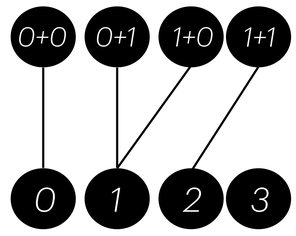
Input: X



f(X) = Y, where f is Identity

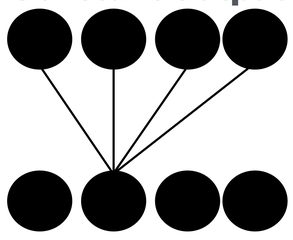
Computation: 1-bit Addition

Input: X



f(X) = Y, where f is Addition

Constant Output

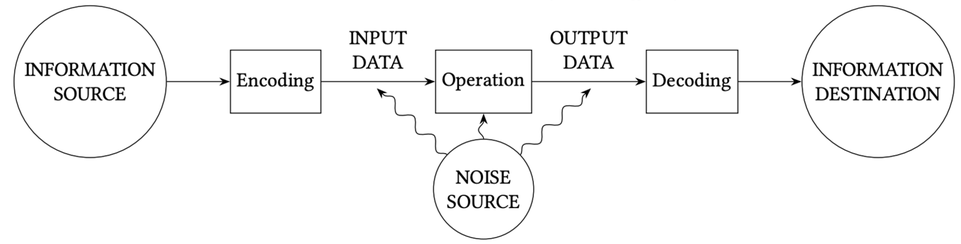


Noise and redundant inputs mapping to the same output (like multiple ways to sum to 1) both destroy information.

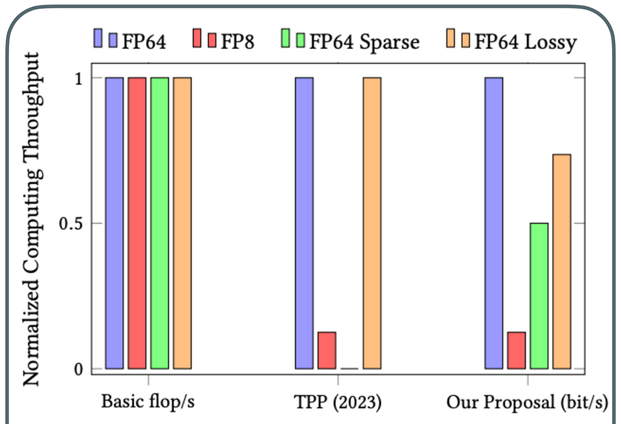
In the extreme, the output is always the same, I(X;Y)=0. There is no uncertainty about the output, and the computation can be compiled away.

A More General Measure for Work

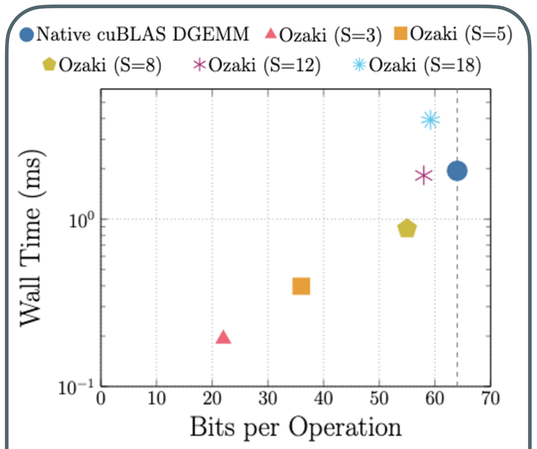
Modeling computers as information channels provides a common measure for comparing the performance of conventional, approximate, emulated, and novel computing systems.



Under this framing, communication is a specific case of computation.



Our proposal matches existing Total Processing Performance (TPP) U.S. export regulations and extends beyond it. TPP explicitly excludes sparse pipelines and doesn't penalize noise, both of which an info-theoretic approach handles naturally.



Channel capacity per operation of native DGEMM and ozIMMU emulation where S is Int8 slice count.

RTX 6000 Pro Blackwell with N=1024.

(github.com/enp1s0/ozIMMU)

Implications and Impact

This work aims to...

- Bound and fairly evaluate mixed-precision and emulation schemes
- Enable evaluation of sparsity, noise, and novel hardware
- Theoretically ground computing export controls

How will you measure computational work in a heterogeneous future?

Preprint: "Back to Bits: Extending Shannon's communication performance framework to computing", arXiv:2508.05621

Related work: Shannon (1948), von Neumann (1952), Elias (1958), Evans (1994), Ryabko & Fionov (2012), and Simon (2016)