



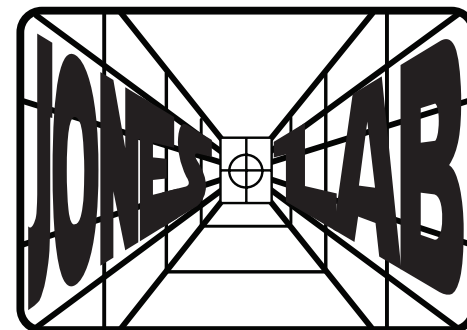
University of Pittsburgh

Collaborative AI at the Edge: IEEE Micro Magazine Jan/Feb 2023

Alex K. Jones

University of Pittsburgh
akjones@pitt.edu

February 9, 2023



Outline



- CORUSCANT and POD-RACING
 - Pretty cool Racetrack device
 - Let's make it do PIM
- Sustainable Computing
 - We're at Post Moore crossroads
 - Let's make machines that address sustainability metrics

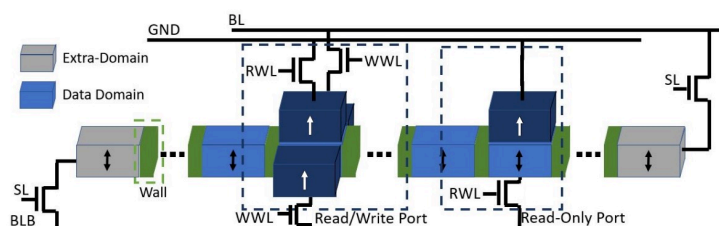
Outline



- CORUSCANT and POD-RACING
 - Pretty cool Racetrack device
 - Let's make it do PIM
- Sustainable Computing
 - We're at Post Moore crossroads
 - Let's make machines that address sustainability metrics

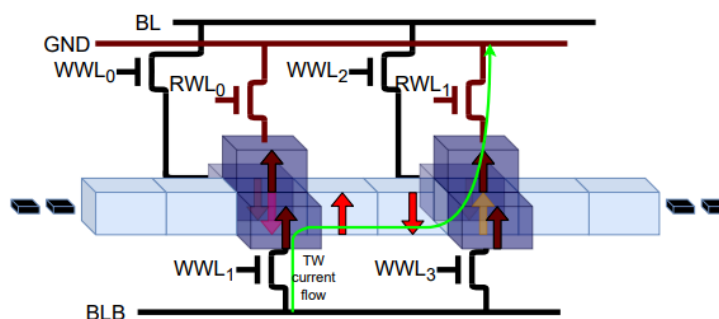


Racetrack Memory and Multi-domain Read

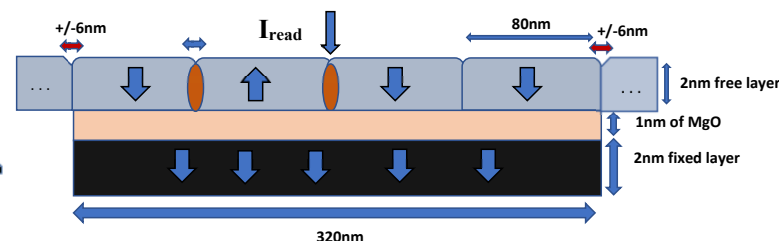


Access Speed	~1ns
Energy per write	0.1pJ
Area per Cell	~2F ²
Property	Non-Volatile Memory
Biggest Challenge	Shifting Latency

Transverse read – Count the Number of 1's



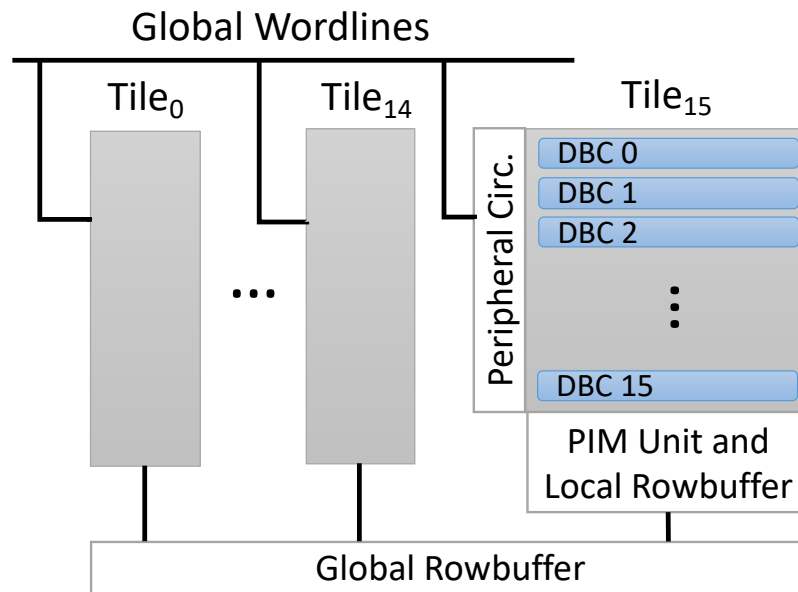
Transverse Read, TNANO, 2022



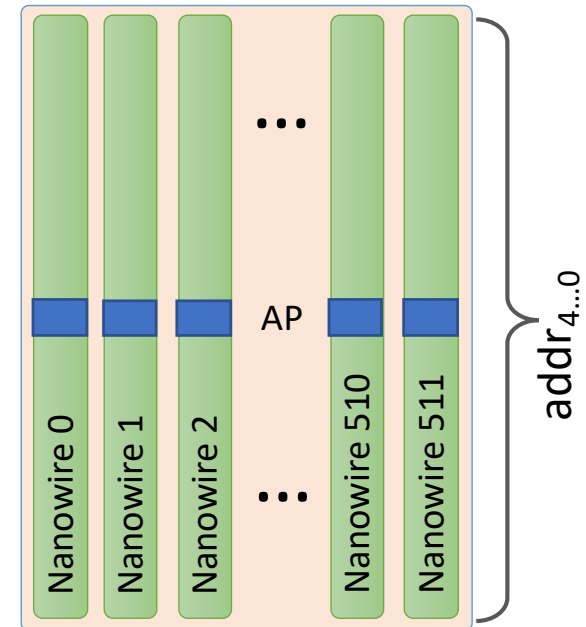
Multi-domain MTJ: arXiv: 2205.12494



DWM Architecture



Subarray

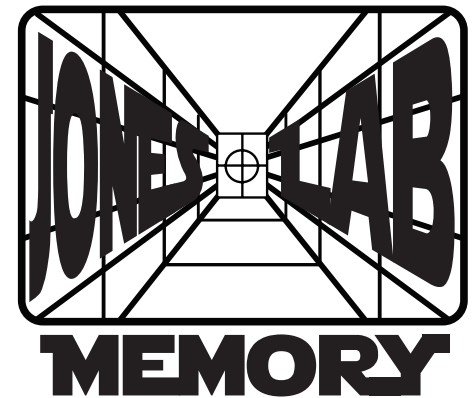
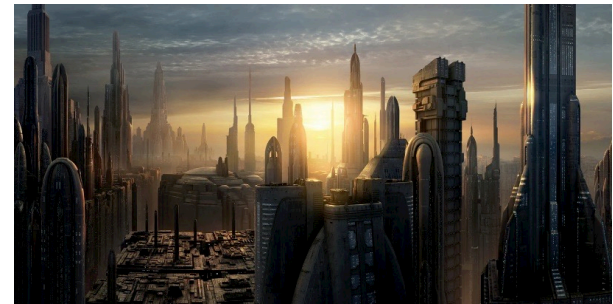
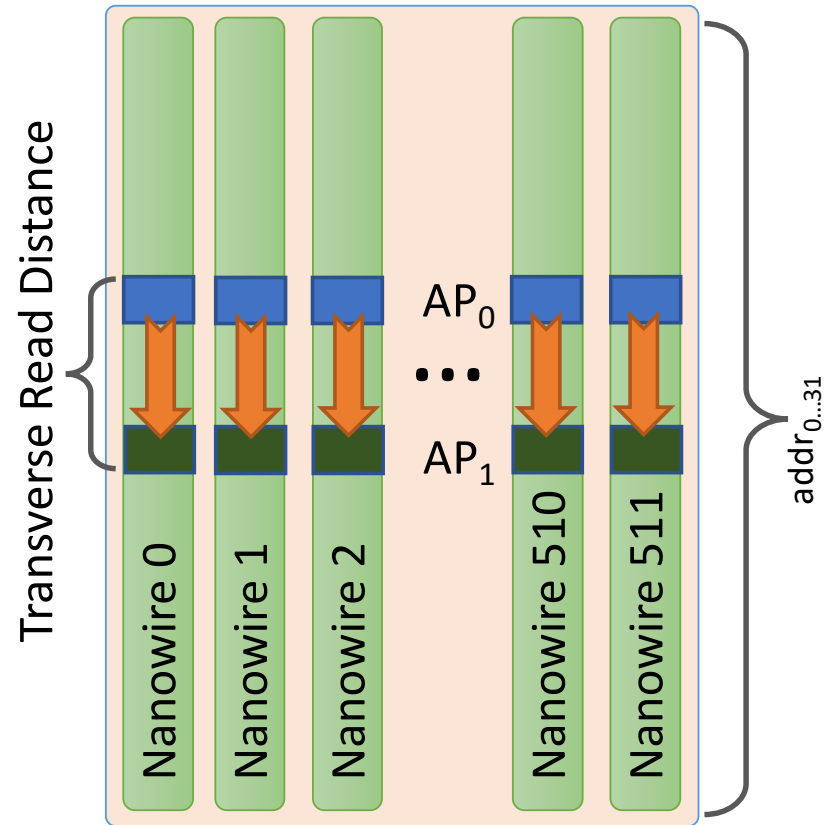


Data Block Cluster (DBC)

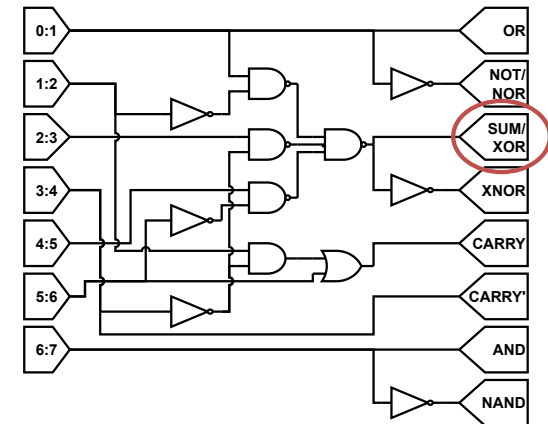
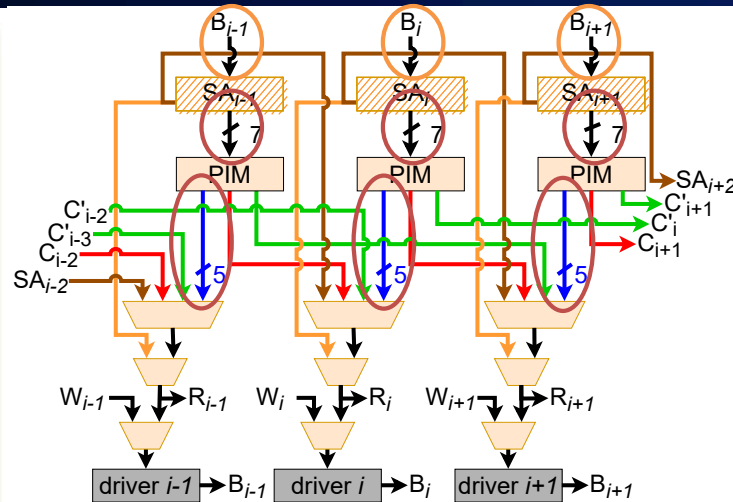
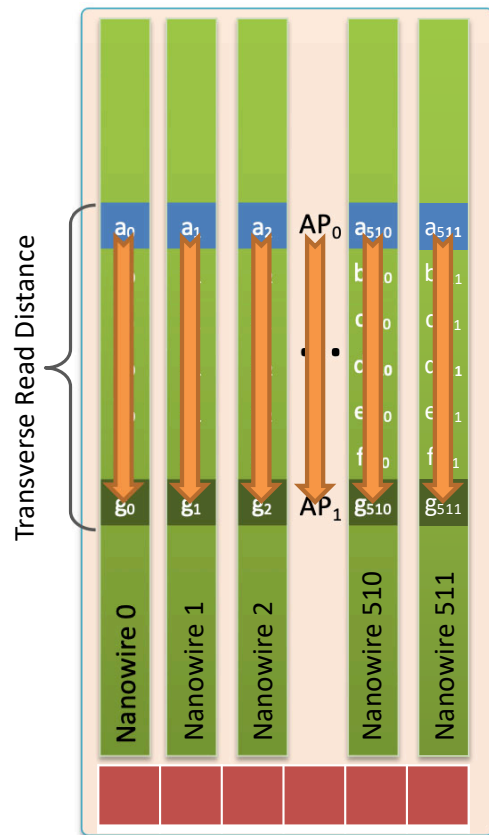
CORUSCANT



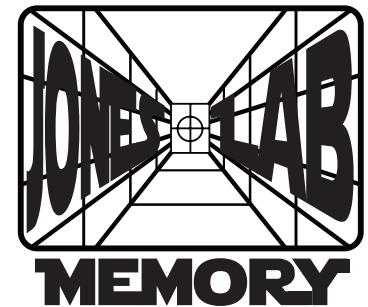
- **CORUSCANT** –
Computing
Optimized
Racetracks Using
Specialized
Clusters
Accessing
Nanowires
Transversely



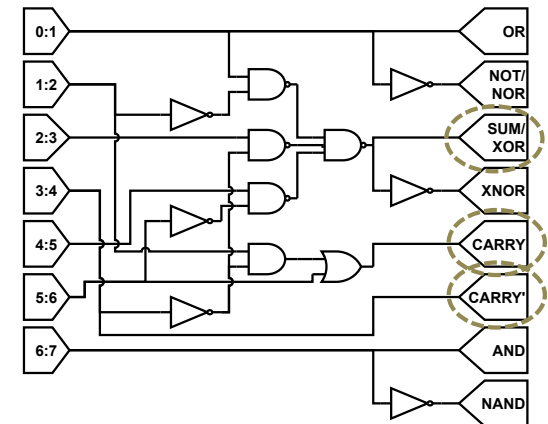
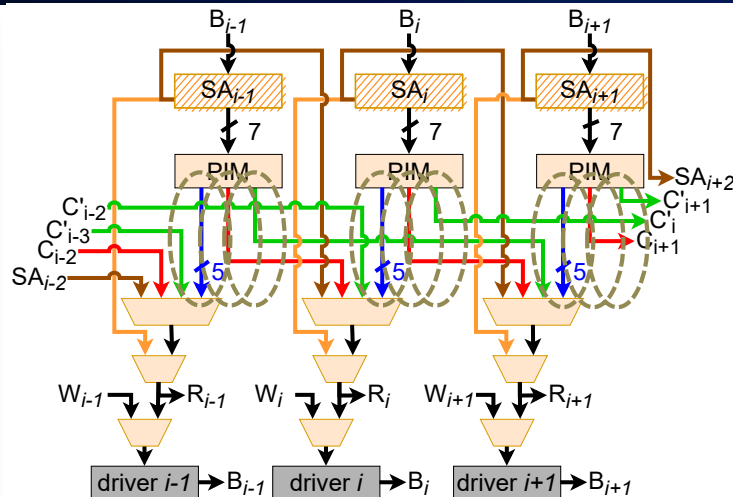
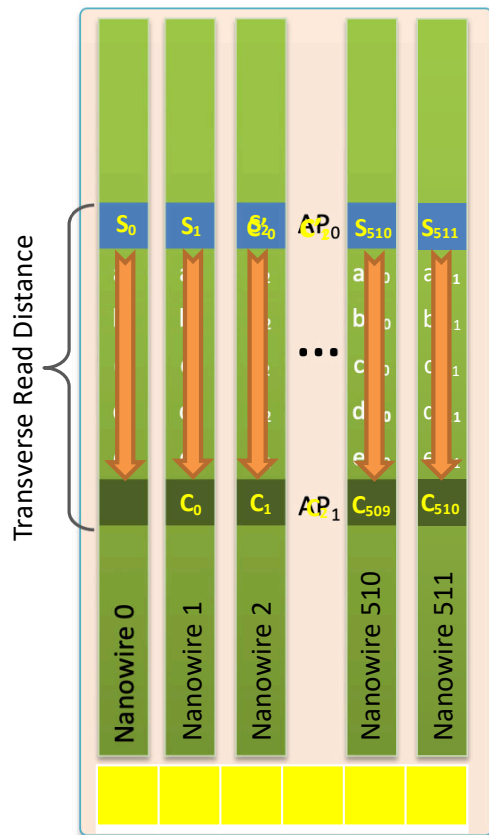
CORUSCANT: Bulk Bitwise



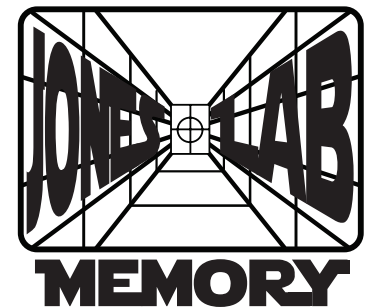
$$A \oplus B \oplus C \oplus D \oplus E \oplus F \oplus G$$



CORUSCANT: Addition



A plus B plus C plus D plus E



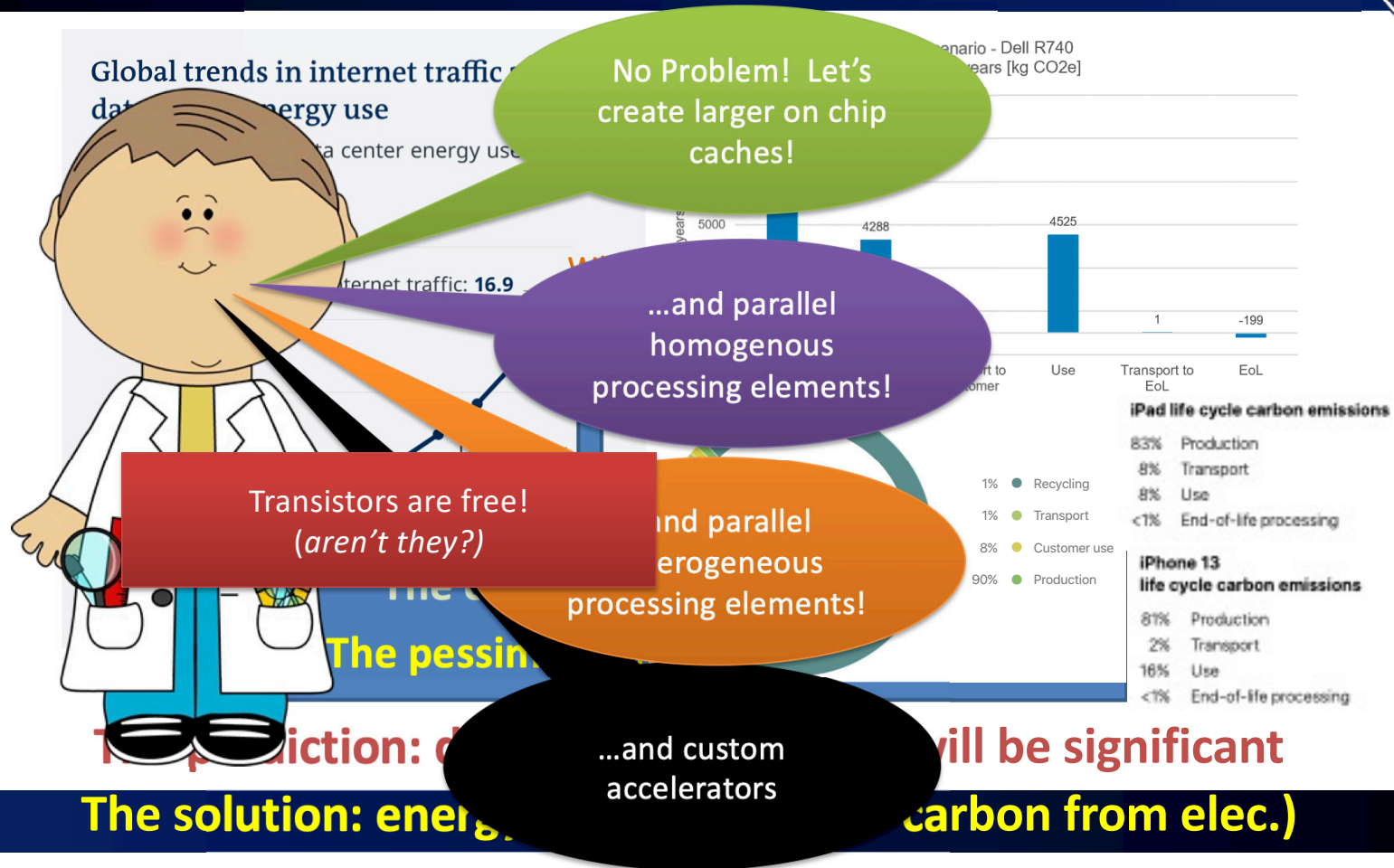
Outline



- CORUSCANT and POD-RACING
 - Pretty cool Racetrack device
 - Let's make it do PIM
- Sustainable Computing
 - We're at Post Moore crossroads
 - Let's make machines that address sustainability metrics



Articulating the Sustainability Problem



Indifference Point Analysis

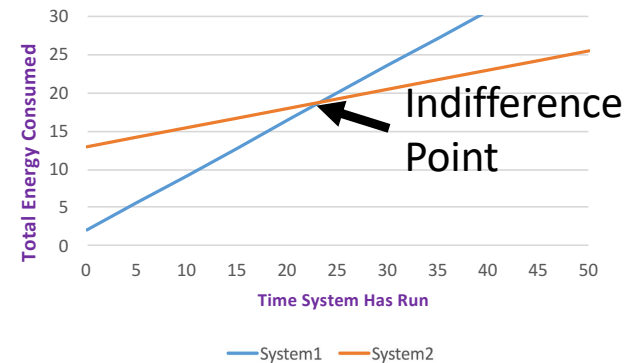
Energy & GWP
Evaluation



- Make Comparisons
 - Absolute evaluation is still hard
- Borrow from Economics

In the context of computing systems

- Time to amortize initial manufacturing investment

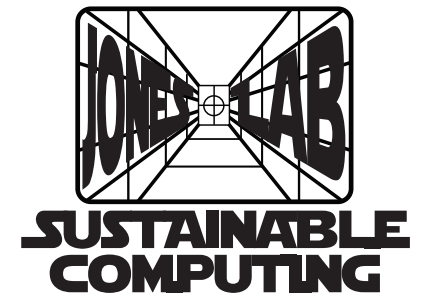


- **Indifference Analysis:**

Time (t_I) when there is no difference in cost between two alternatives

- **Break Even Analysis:**

Time (t_B) when new system will reach the same energy consumption of the system it will replace



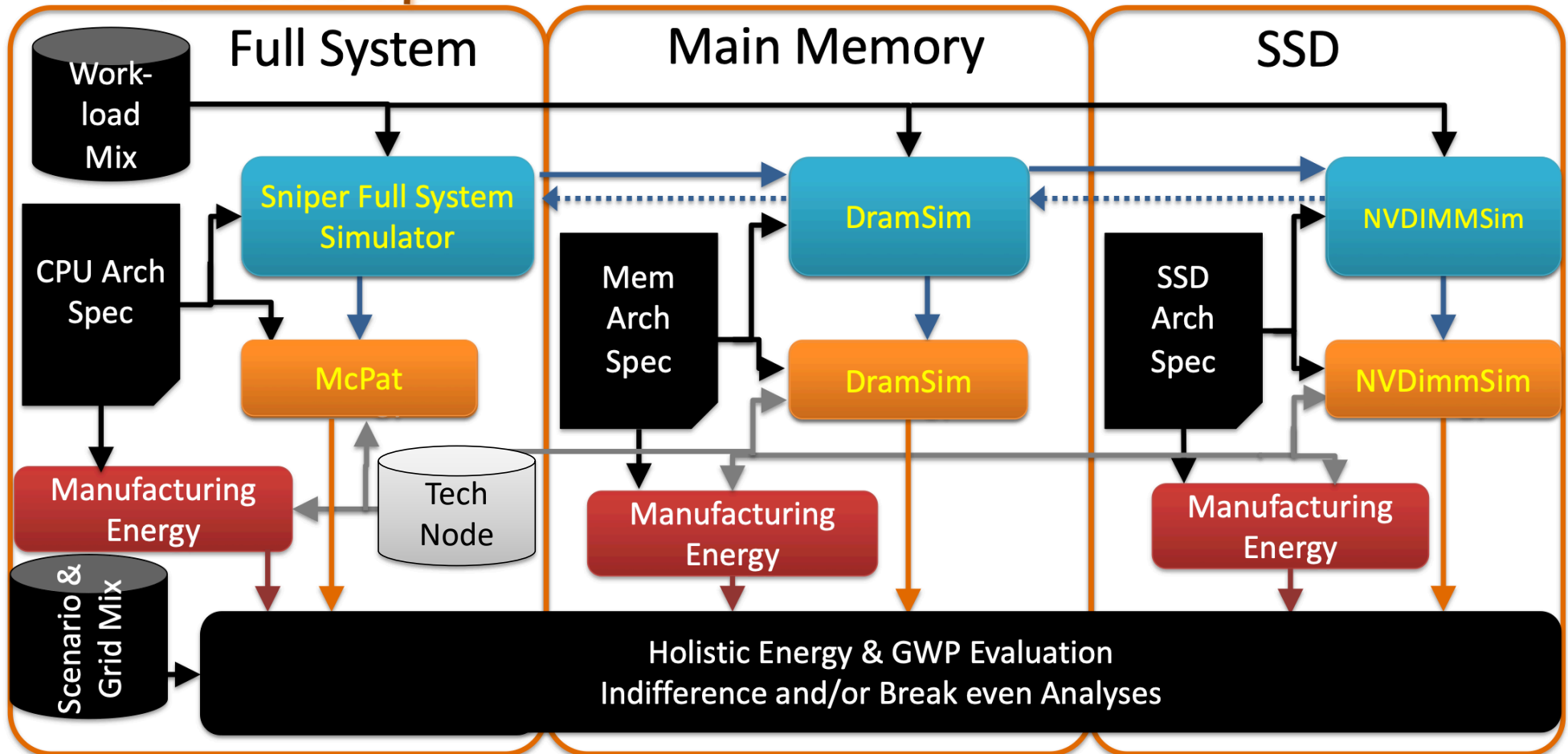
$$t_I = \frac{M_1 - M_0}{P_0 - P_1}$$

$$t_B = \frac{M_1}{P_0 - P_1}$$

$$t_I = t_B - \frac{M_0}{P_0 - P_1}$$

GreenChip

<https://github.com/Pitt-Jones-Lab/Greenchip>

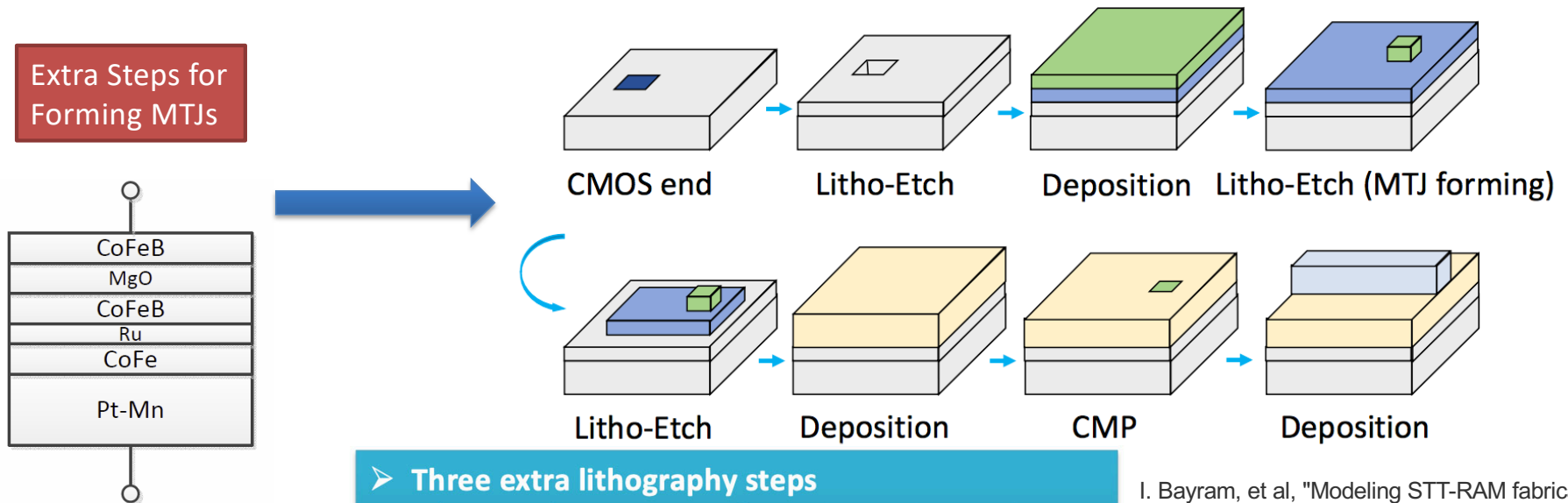


D. Kline, et al., "GreenChip: A tool for evaluating holistic sustainability of modern computing systems," *Sustainable Computing: Informatics and Systems*, Vol. 22, 2019, pp. 322-332, <https://doi.org/10.1016/j.suscom.2017.10.001>

Extending NVSim with Sustainability Metrics



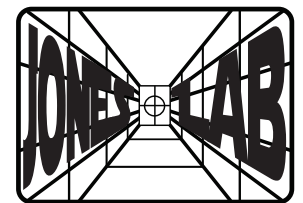
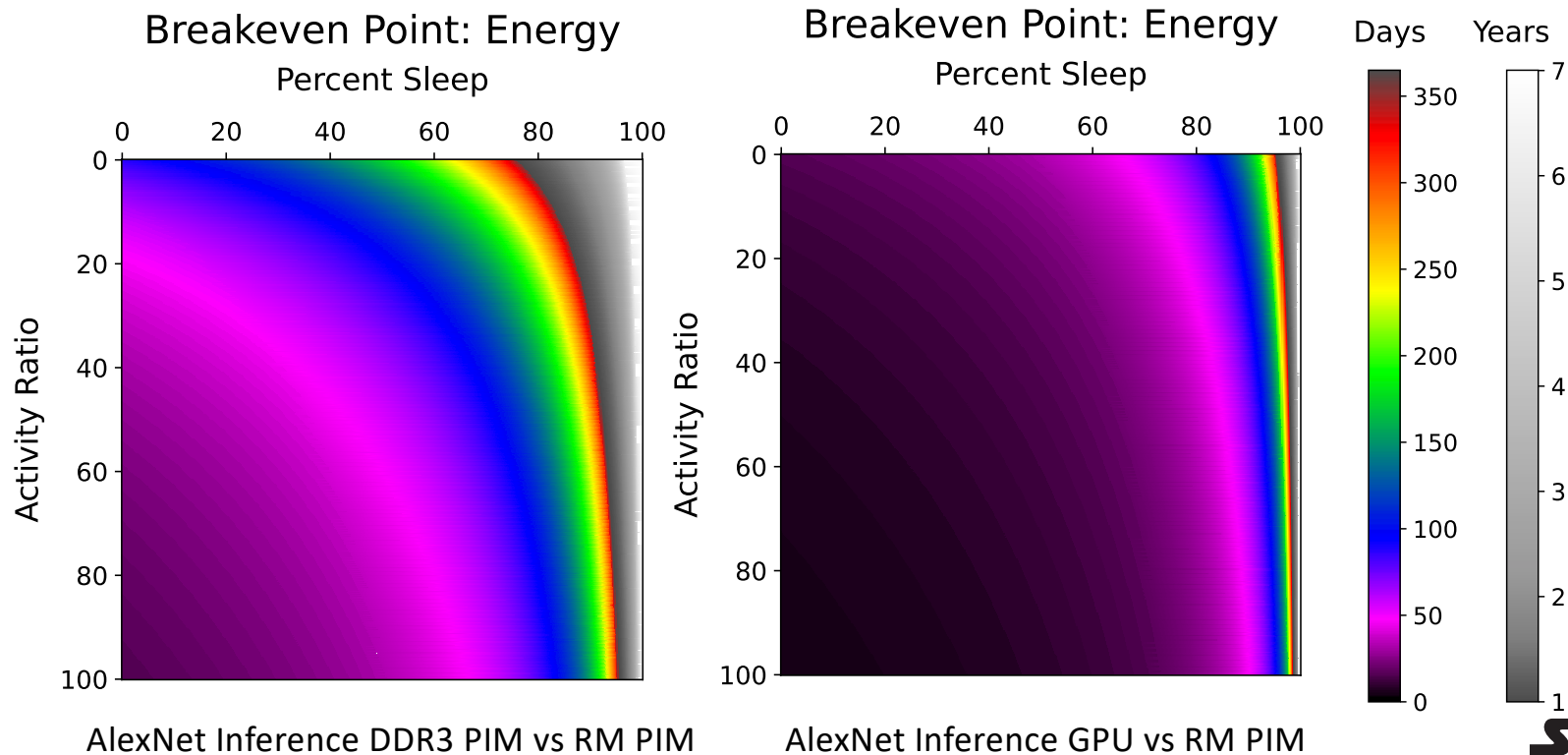
- NVSim: area/perf/use-phase energy estimates for SRAM and NVMs
- Combine with parameterized model to determine cost and environmental impact data for arbitrary memory design



- Three extra lithography steps
 - One for MTJs, two for dedicated metal layers
- Multiple materials required to form MTJ (CVD)
 - Multiple etching chemicals and sequences

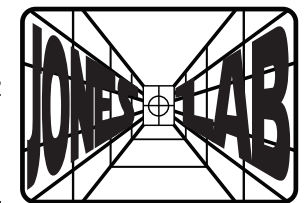
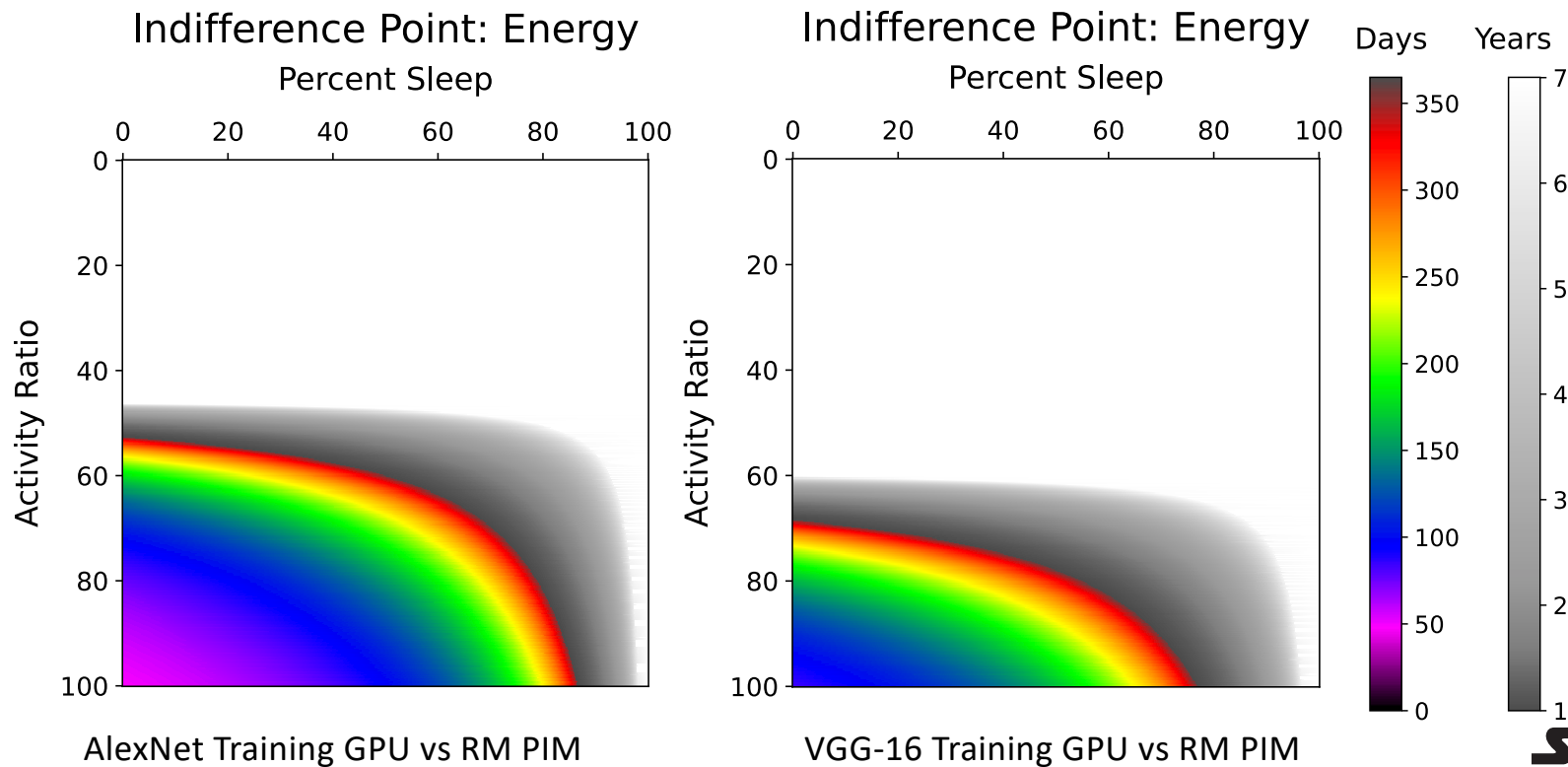
I. Bayram, et al, "Modeling STT-RAM fabrication cost and impacts in NVSim," /GSC, 2016

Inference Breakeven Comparison



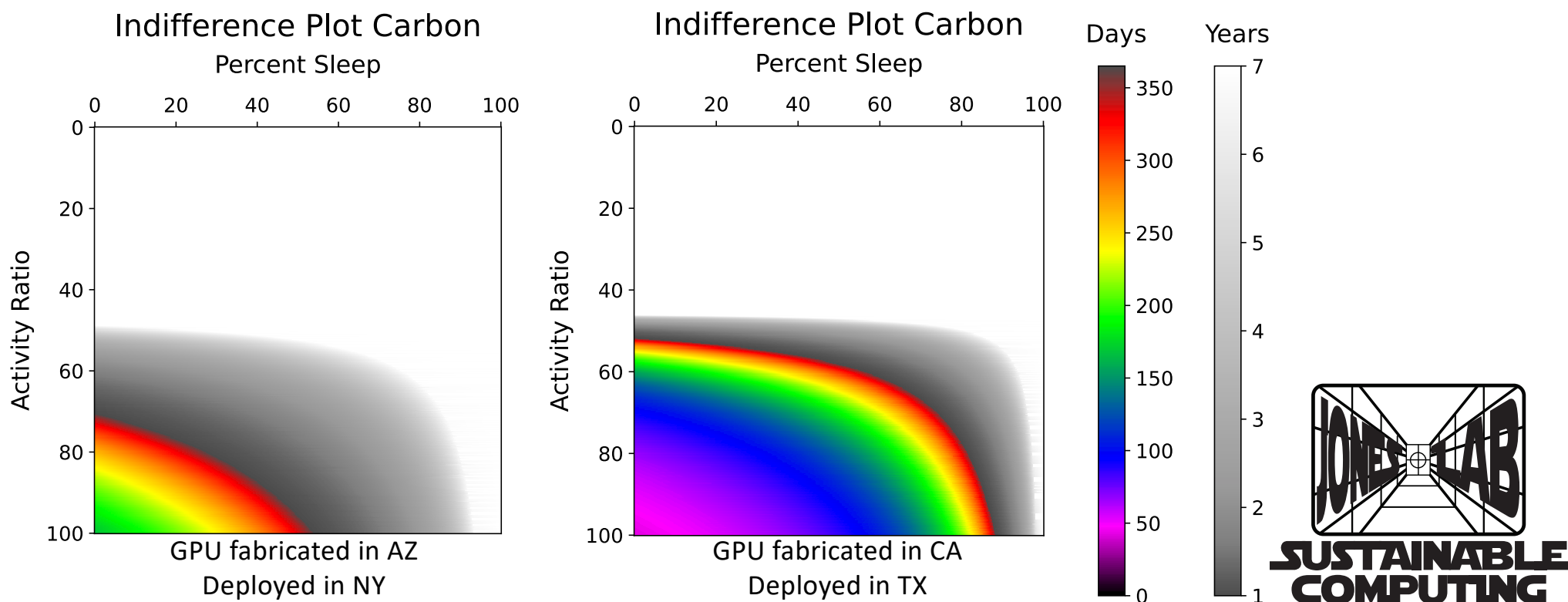
**SUSTAINABLE
COMPUTING**

Training Indifference Comparison



**SUSTAINABLE
COMPUTING**

Grid Mix Impact



Reading List



- CORUSCANT – MICRO, 2022
- POD-RACING – IEEE Micro Mag, Nov./Dec 2022
- Sustainable AI Processing at the Edge – IEEE Micro Mag, Jan.-Feb. 2023
- Dark Silicon Considered Harmful – IGSC 2018
- Considering Fabrication in Sustainable Computing – ICCAD 2013
- GreenChip -- SUSCOM, 2019



akjones@pitt.edu
University of Pittsburgh

alejones@nsf.gov
National Science Foundation