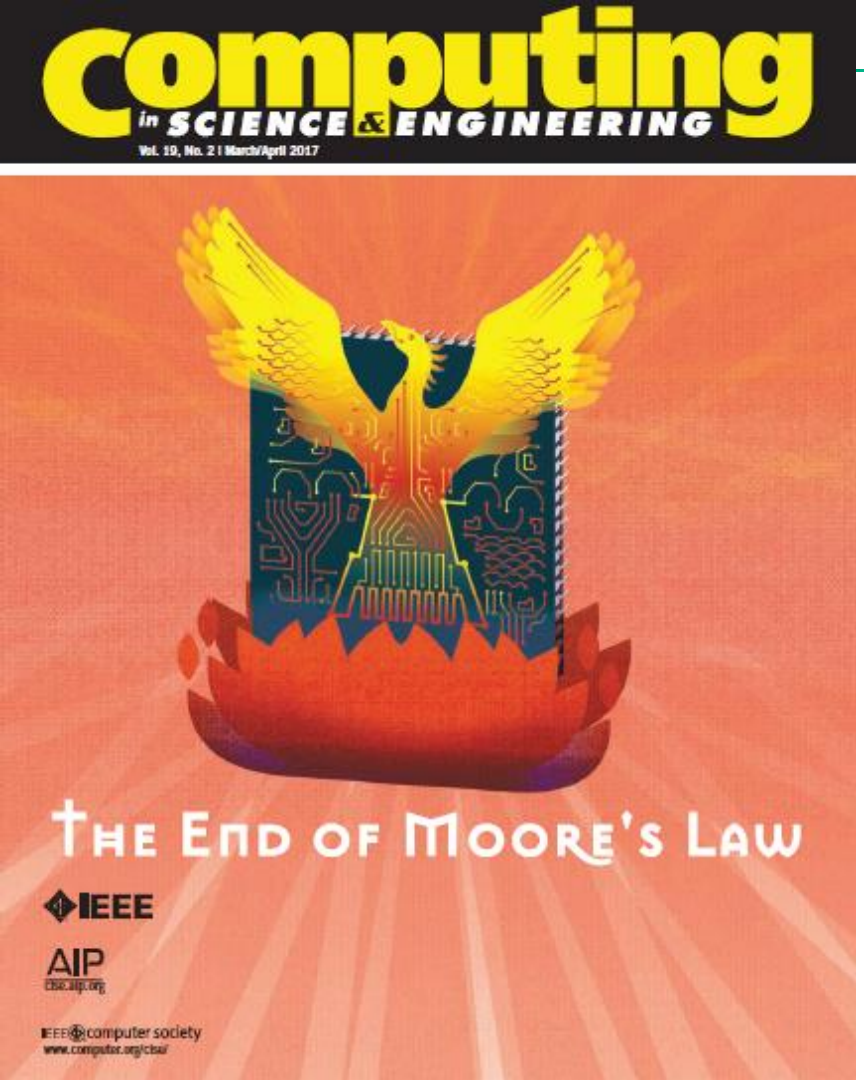
Hewlett Packard Enterprise

Computing for a new era

John Paul Strachan, Miao Hu, Cat Graves, Suhas Kumar, Naveen Muralimanohar, Dejan Milojicic, R. Stanley Williams

BESAC Meeting **July 13, 2017**



What's Next?

The end of Moore's law could be the best thing that has happened in computing since the beginning of Moore's law.

Future exponential increases in computer performance and efficiency will require multiple advances:

Systems and Architectures

- Memory-centric computing – no von Neumann bottleneck

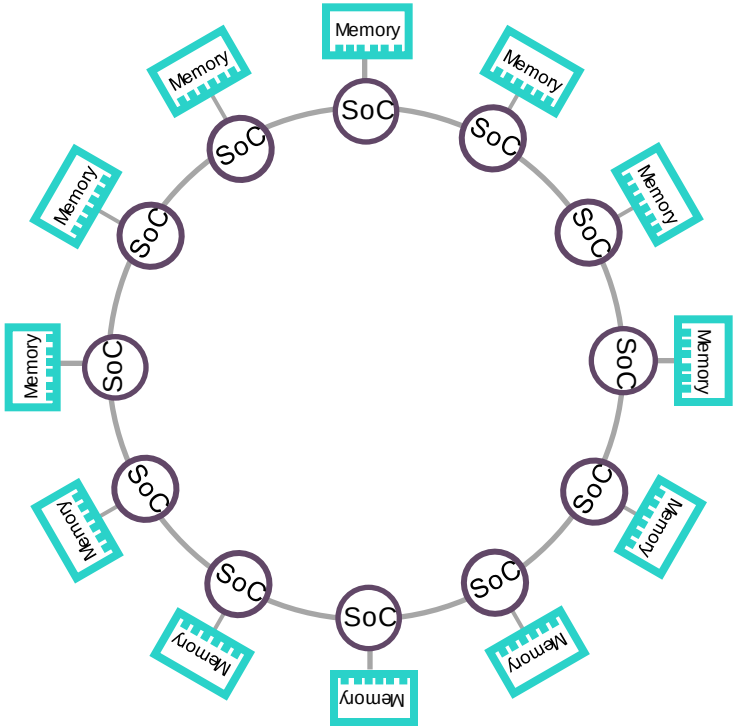
- Gen-Z – high performance open fabric to democratize computing
  (http://genzconsortium.org/ - ~40 companies so far)

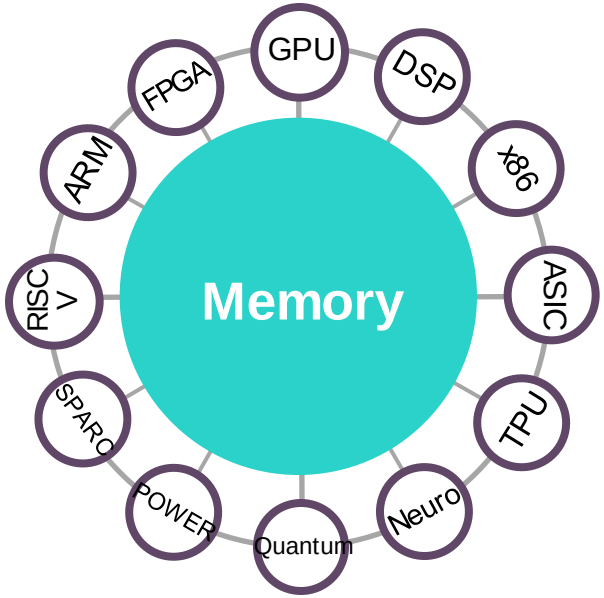
- Dot-product engine: memristor-based vector-matrix multiplication accelerator for neural nets and signal processing

- Chaos as a computing resource for constrained optimization problem solving (Hopfield network)

Devices

Hewlett Packard Enterprise





From processor-centric computing…

the traditional von Neumann architecture

…to Memory-Driven Computing

with Gen-Z open (nonproprietary) fabric, computing is plug and play

Memory-Driven Computing (MDC) is a reality: The Machine

Fast, persistent memory	Fast memory fabric	Task-specific processing	New and Adapted software
Combining memory and storage in a stable environment to increase processing speed and improve energy efficiency	Using photonics where necessary to eliminate distance and create otherwise impossible topologies	Optimizing processing from general to specific tasks	Radically simplifying programming and enabling new applications that we can't even begin to build today

Gen-Z fabric: ultra-high bandwidth open bus to democratize computing

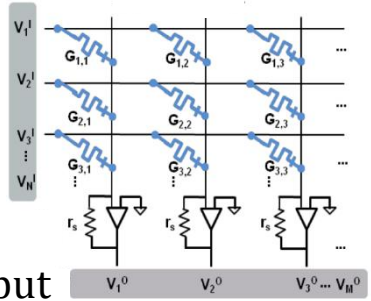
Up to 8000x speed-up for Monte Carlo simulations

Memristor technology, ongoing transfer from lab to commercial product

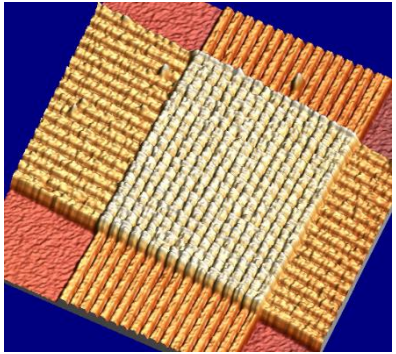
Developing new accelerators from novel device behavior

Dot Product Engine: memristor arrays accelerate vector-matrix multiplication

Input Voltage vector



Output current $I_j^O = \sum_i G_{ij} \cdot V_i^I$



- **Parallel multiply & add through Kirchoff's and Ohm's laws**

  1961, K. Steinbuch "*Die Lernmatrix*" – suggests using "ferromagnetic toroids"

- **Memristors as highly scalable, tunable analog resistors**

  High ON/OFF ratio (~10^5), supporting multiple levels

- **Well suited for streaming workloads like neural nets**

- **Many ways to scale up**

  Memristor levels, array size, wire pitch, 3D layer, DAC/ADC speed & width etc.

- **Performance (execution time) improvements >1000x and**

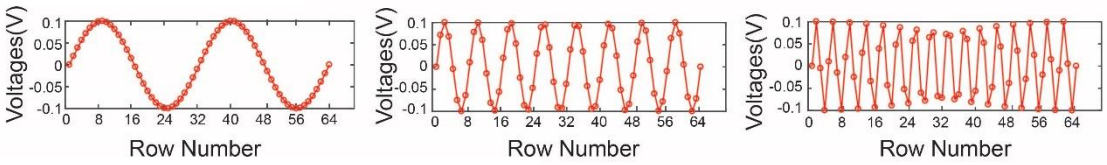
  energy efficiency >100x over GPUs for <u>particular applications</u>

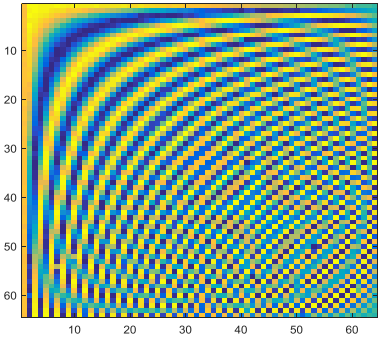
Hewlett Packard Enterprise

Experimental Demo - Signal processing on DPE system

(UMass collaboration) – instantaneous cosine transform

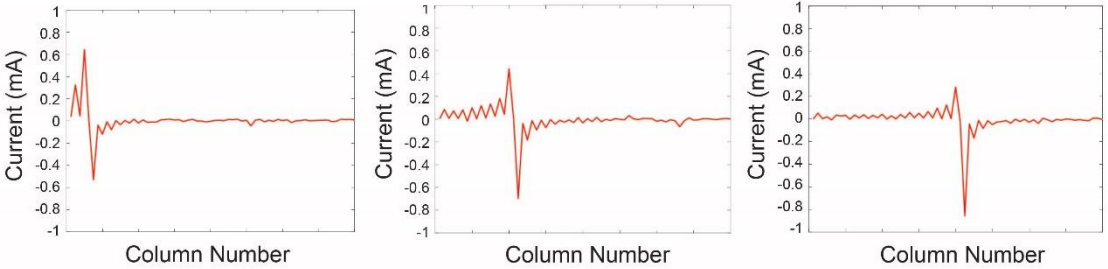
Time-domain inputs (applied to rows)



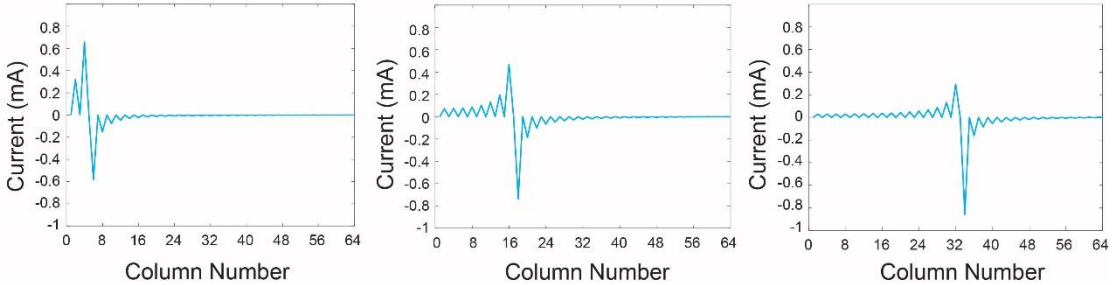


Freq-domain (DCT) outputs (read from columns)

Real-time experimental data from DPE

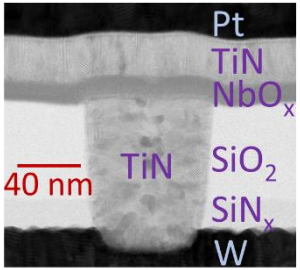


Software cosine transform



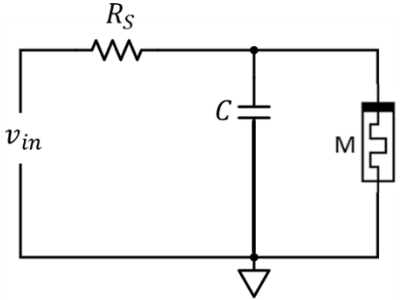
C. Li, et. al, *manuscipt under review*

NbO_2 memristor: computing with nonlinear dynamics and chaos

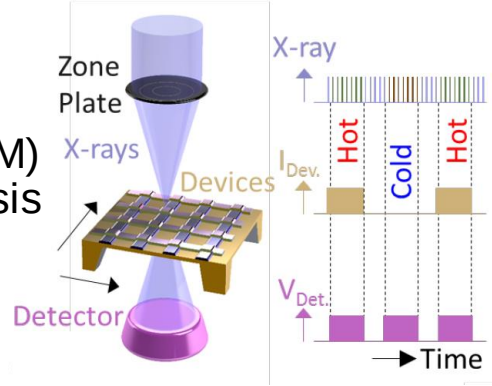


=

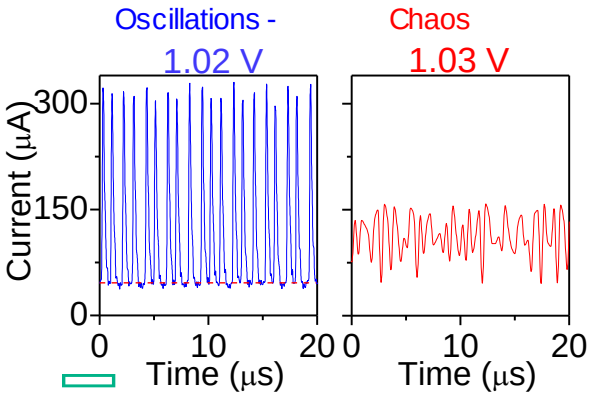
nanoscale oscillator



in situ and *in operando* scanning transmission x-ray microscopy (STXM) at ALS critical for analysis of memristor operation



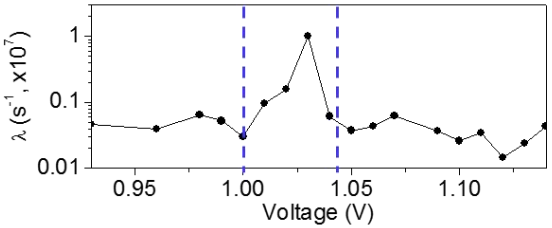
Dynamical Behavior:



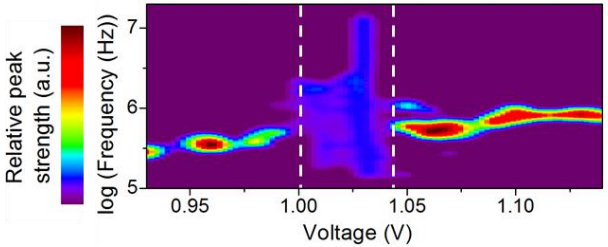
Hewlett Packard Enterprise

Lyapunov exponent

$$\lambda = \lim_{t\to\infty} \lim_{\delta \mathbf{Z}_0 \to 0} \frac{1}{t} \ln \frac{|\delta \mathbf{Z}(t)|}{|\delta \mathbf{Z}_0|}$$



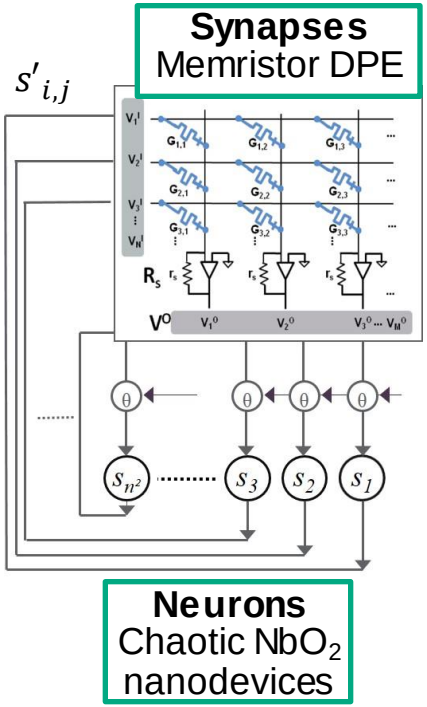
Frequency analysis



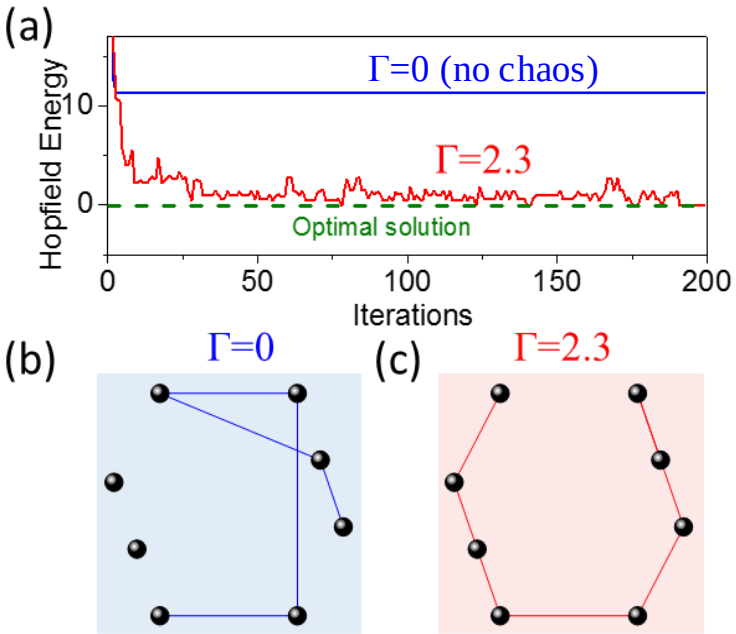
Chaos enabled Hopfield network for optimization problems

Traveling Salesman problem:

Find shortest route visiting all cities



Example solutions w/ and w/o chaos



S Kumar, et al. Nature, in press

Acknowledgments

HPE Labs

Miao Hu
Suhas Kumar
Catherine Graves
Emmanuelle Merced
Dejan Milojicic
Ali Shafiee
Naveen Muralimanohar
Stan Williams

UMASS AMHERST

Jianhua Yang
Qiangfei Xia
Can Li
Yunning Li

Izzat El Hajj (UIUC)
Pedro Bruel (IME)

Berkeley Lab

David Kilcoyne
David Vine

SLAC

Johanna Weker
Hendrik Ohldag

Funding assistance

Synchrotron at LBNL

To study the very small, we had to use the very large

The Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory