Brain-CA Technologies

Brain-CA:
A Cellular Automata-Inspired Architecture for Energy-Efficient CPU Design

ISCA 2025 DCEE Workshop
June 2025

Brain-CA Technologies

Who is Brain-CA Technologies?

Founded by three former engineers with 70+ years in system design with the HP Enterprise team,

We are addressing the AI Energy Crisis.

Brain-CA Technologies

Our Progress:

ISCA 2024:
Presented theoretical approach to cut energy consumption *by orders of magnitude*.

Since:
Success in software simulations.
FPGA implementation.
Smaller models.

Brain-CA Technologies

Our Motivation:

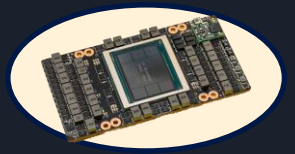
You don’t need equations to play catch.

The same goes for learning.

Math can describe it—
but that doesn’t mean
it’s the best way to do it.

Brain-CA Technologies

Comparing a mosquito brain to the H100



A mosquito brain is...

- ~40,000× smaller
- 1/400,000 the number of components
- ~700,000× more energy-efficient

Brain-CA Technologies

Re-think the lower levels of AI stack

- Application-Level Software (LLMs, Vision AI, etc.)
- Inference Logic (Prompting, APIs, etc.)
- Model Runtime (TensorRT, ONNX, etc.)
- Training Framework (PyTorch, TF, JAX)
- Hardware Interface Layer (CUDA, XLA, etc.)
- Compute Hardware (GPUs, TPUs — Von Neumann)

Brain-CA Technologies

Re-think the lower levels of AI stack

- Application-Level Software (LLMs, Vision AI, etc.)
- Inference Logic (Prompting, APIs, etc.)
- Model Runtime (TensorRT, ONNX, etc.)
- Training Framework (PyTorch, TF, JAX)
- Hardware Interface Layer (CUDA, XLA, etc.)
- Compute Hardware (GPUs, TPUs — Von Neumann)

331 Pages!

Von Neumann Architecture (JvN: 101 pages), Neural Networks (McCulloch & Pitts: 24 pages, Rosenblatt: 22 pages, Minsky: 136 pages), Backpropagation (Rumelhart, Hinton & Williams: 13 pages), Cuda & GPUs (Buck, Harris, Krizhevsky, Sutskever, Hinton: 35 pages)

Brain-CA
Technologies

1. We invented the world’s simplest device that learns and predicts.

Brain-CA
Technologies

1. We invented the world’s simplest device that learns and predicts.

2. We invented a scalable learning system using that elemental device.

Brain-CA Technologies

Re-think the Elemental Components

Element	Function
Bit	0 or 1
NAND Gate	Invert or Ignore
Transistor	Switch or Amplify
Estimator	Learn or Predict

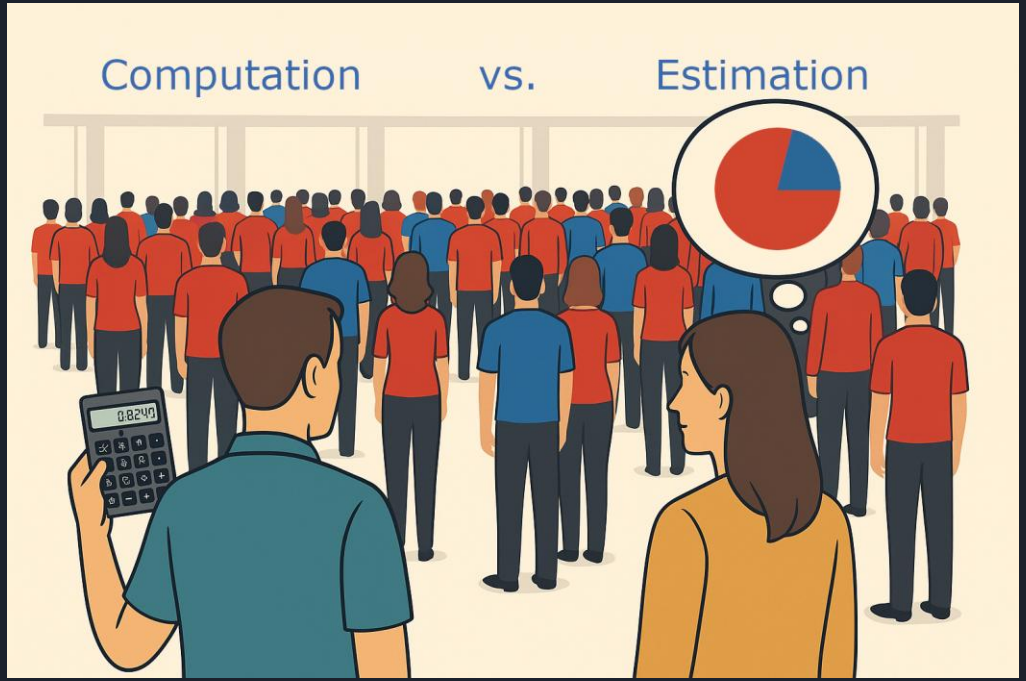
Brain-CA Technologies

Re-think AI using Information Theory

- Shannon: Entropy defines the minimal message
- Redundancy enables perfect reconstruction from noise
- Estimation = compressed memory with structure
- AI, reframed: a compression + prediction problem
- Partial observations → compact model → accurate inference
- Redundancy fills in what was never seen
- Brain-CA builds AI from these first principles

Brain-CA Technologies

Estimation replaces Computation for Energy Savings



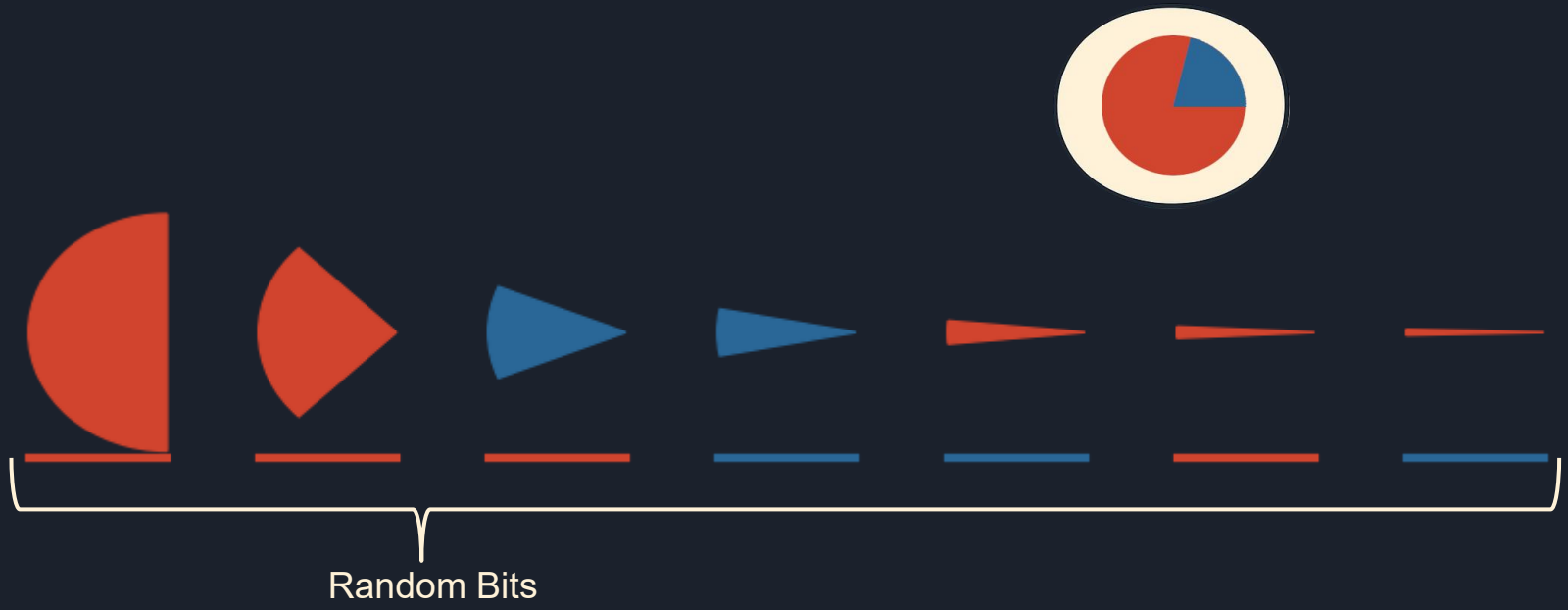
Brain-CA Technologies

Re-think the Pie Chart

1/2 1/4 1/8 1/16 1/32 1/64 1/128

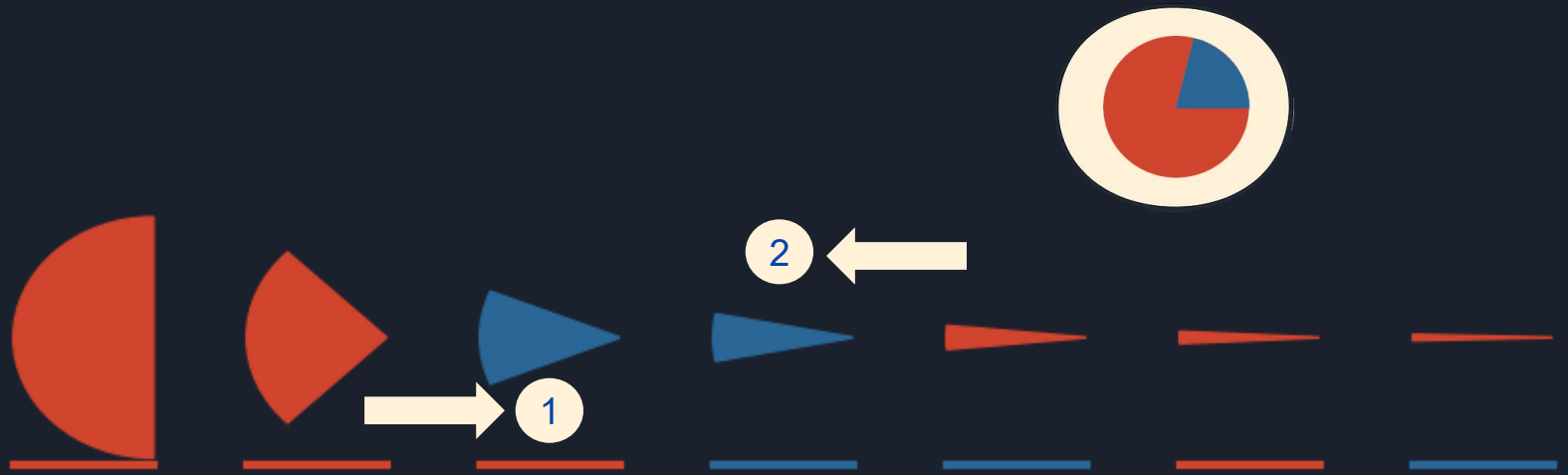
Brain-CA Technologies

Re-think Memory: add Randomness



Brain-CA Technologies

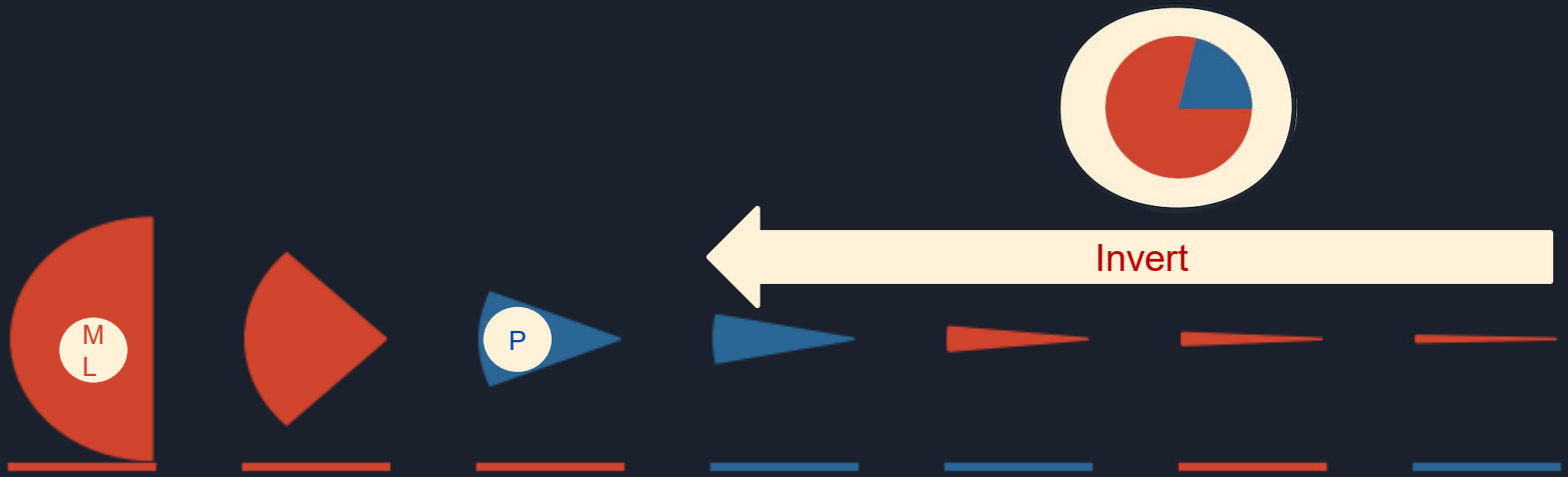
Prepare (off-cycle) to be ready for action



1. Find first mis-match Memory / Random from left
2. Find first matching Memory #1 / Memory from right

Brain-CA Technologies

Designed to learn or predict instantly



Learn: Red → Invert last four Blue → Ignore

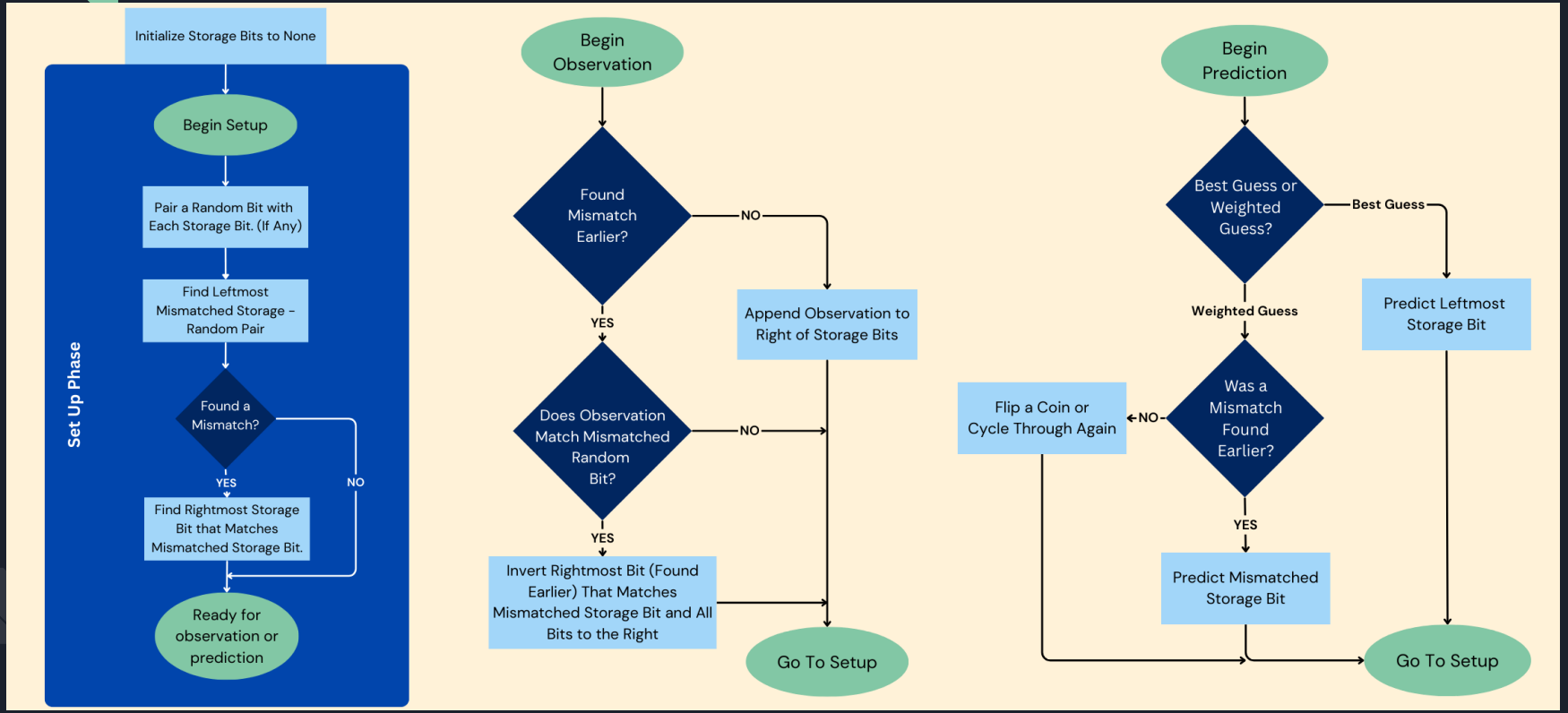
Predict: M L Most Likely → Red P Probabilistic → Blue

Brain-CA Technologies

Estimator converges on actual ratio

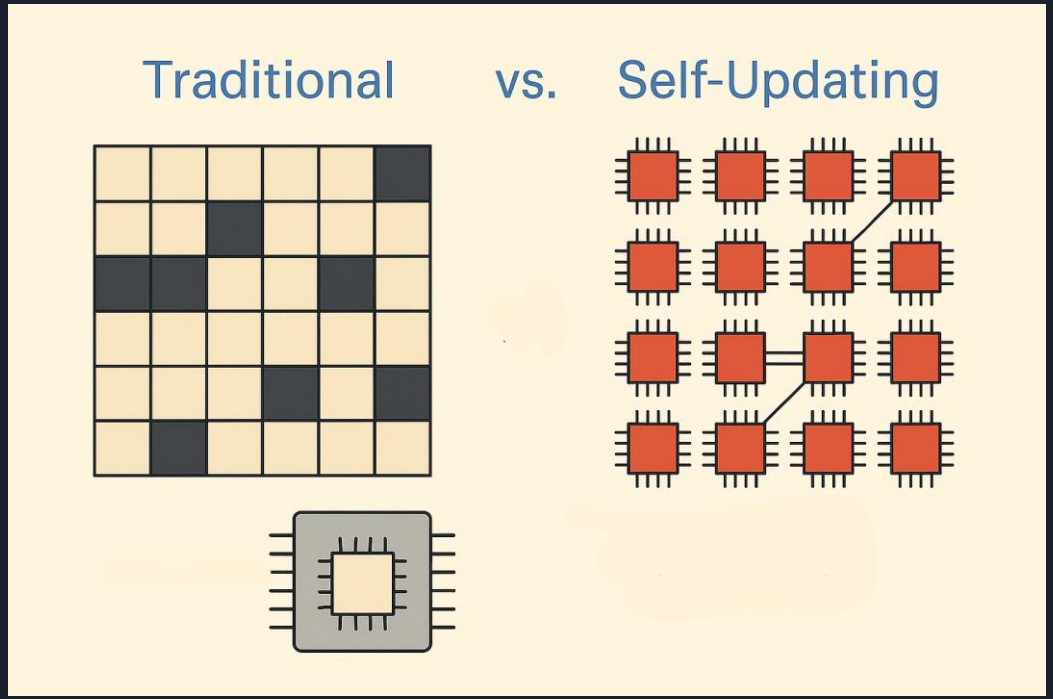
Brain-CA Technologies

Simple Architecture: Estimator Logic



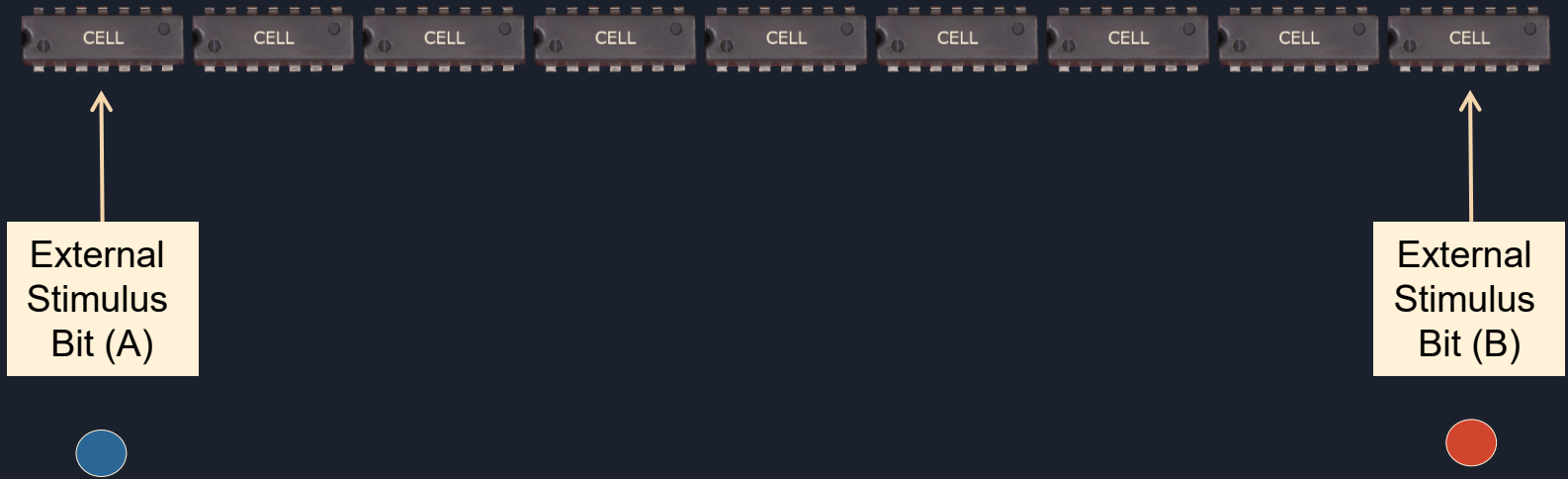
Brain-CA Technologies

Re-think Cellular Automata



Brain-CA Technologies

One-dimensional view



Brain-CA Technologies

Learn through observation

How does “A” behave?

How does “B” behave?

External Stimulus Bit (A)

External Stimulus Bit (B)

Memory Logic:

- Remember the characteristics of external stimuli

Brain-CA Technologies

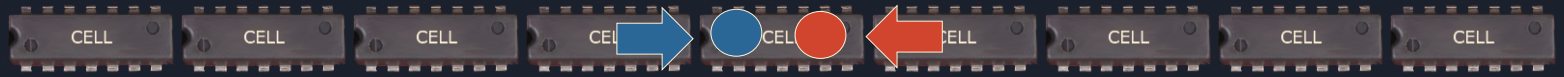
Discovering Relationships

Communication Logic:

- Pass any external stimulus to neighbor(s)
- If something is passed from the left, pass it to the right (and vice versa)

Brain-CA Technologies

Collision Cells can track relationships

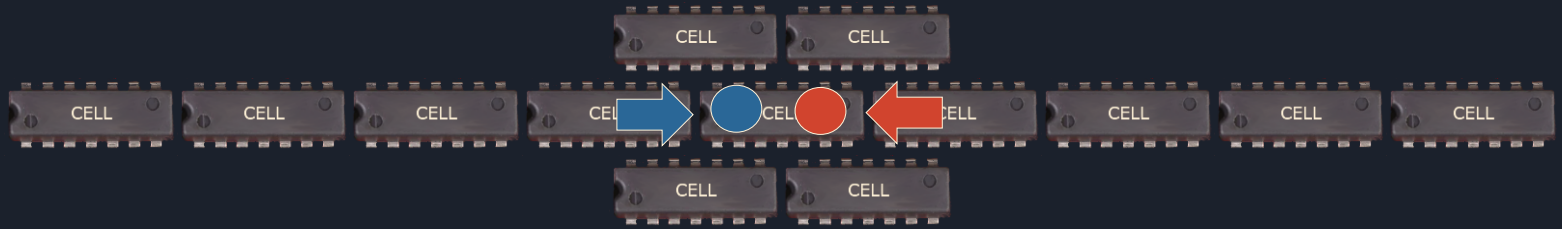


Communication to Memory Logic:

- When signals collide, bifurcate and remember relationship

Brain-CA Technologies

Collision Cells can track relationships

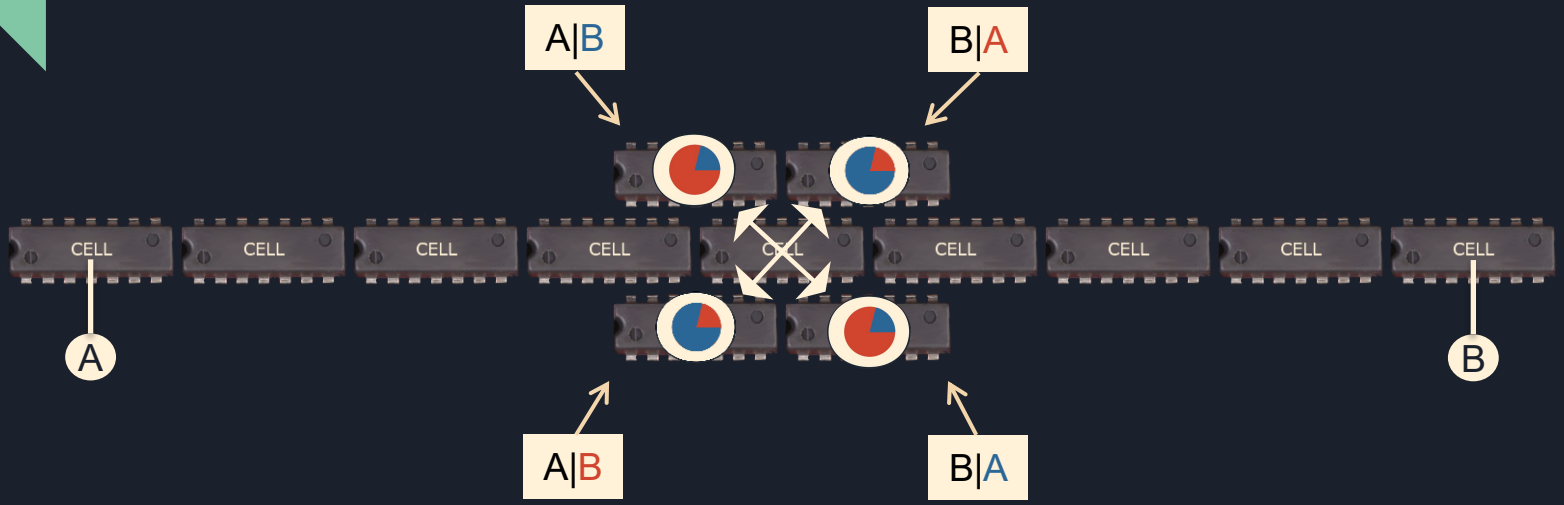


Communication to Memory Logic:

- When signals collide, bifurcate and remember relationship

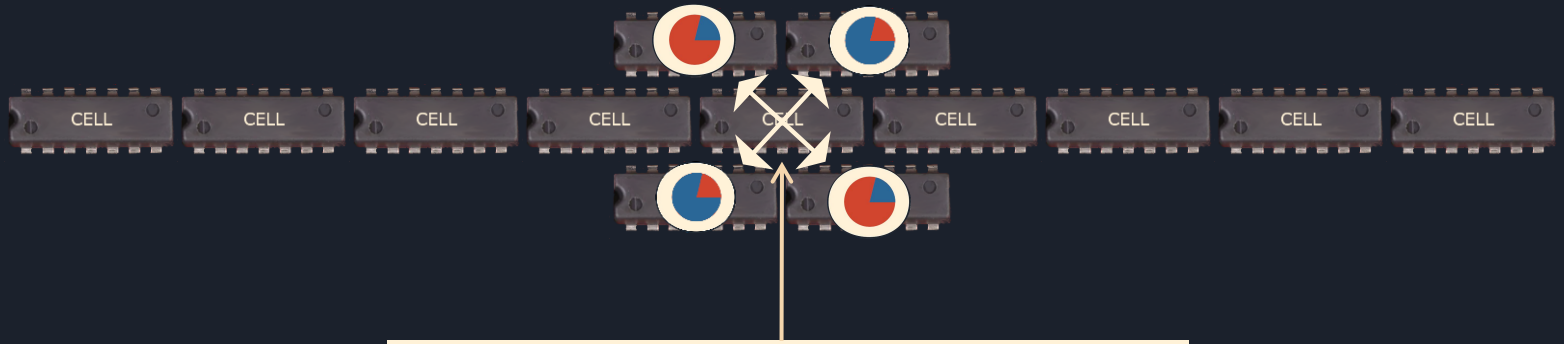
Brain-CA Technologies

Use Neighbors to track conditional relationships



Brain-CA Technologies

Collision Cells coordinate the memory

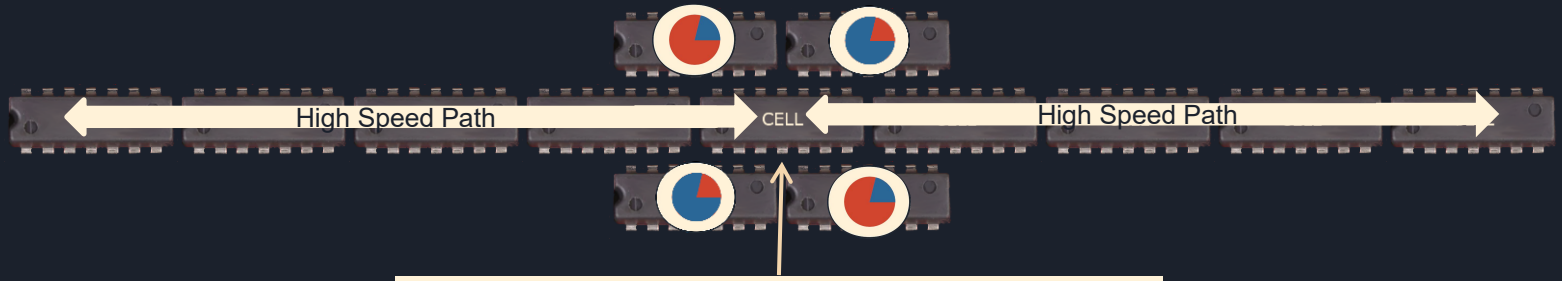


This cell can now predict A|B and B|A.

Slight problem: the cellular propagation delay.

Brain-CA Technologies

Dual Speed Architecture

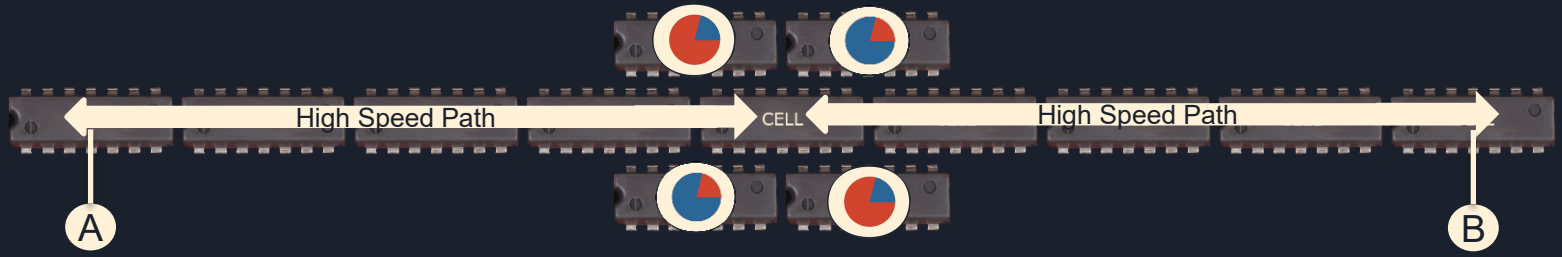


Connection Logic:

- Build high speed paths between correlated cells

Brain-CA Technologies

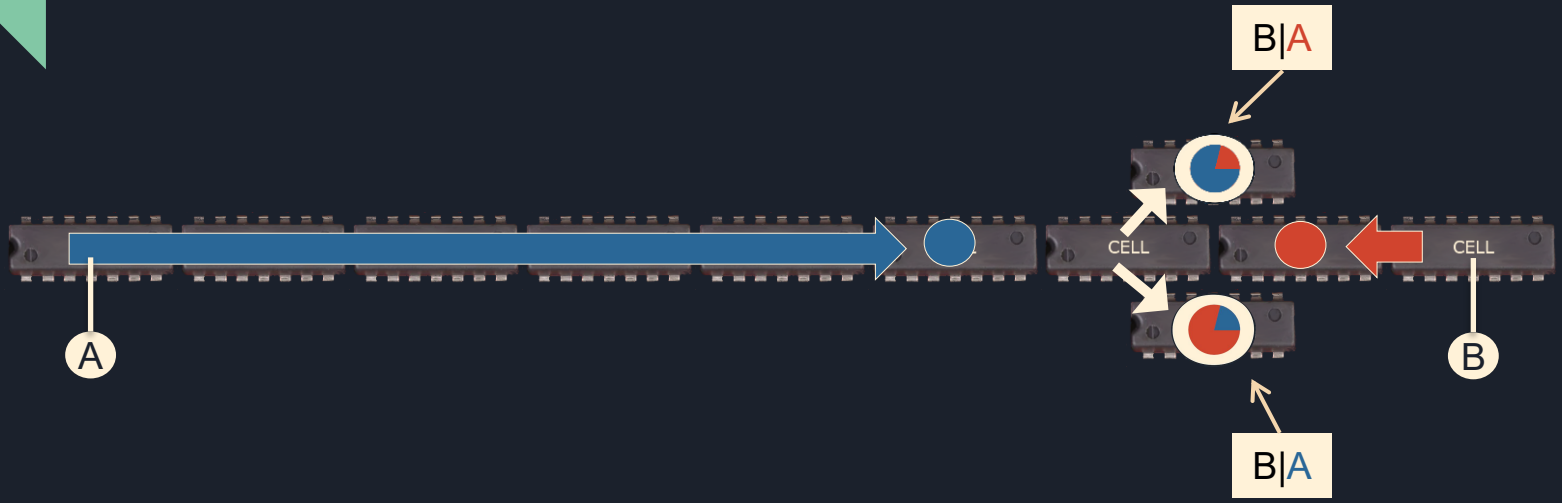
Inference with concurrent signals



1. If “A” is stimulated, the signal is immediately sent to the collision cell.
2. “B” can be predicted based on the value of “A”.
3. The prediction is rapidly sent to “B”.

Brain-CA Technologies

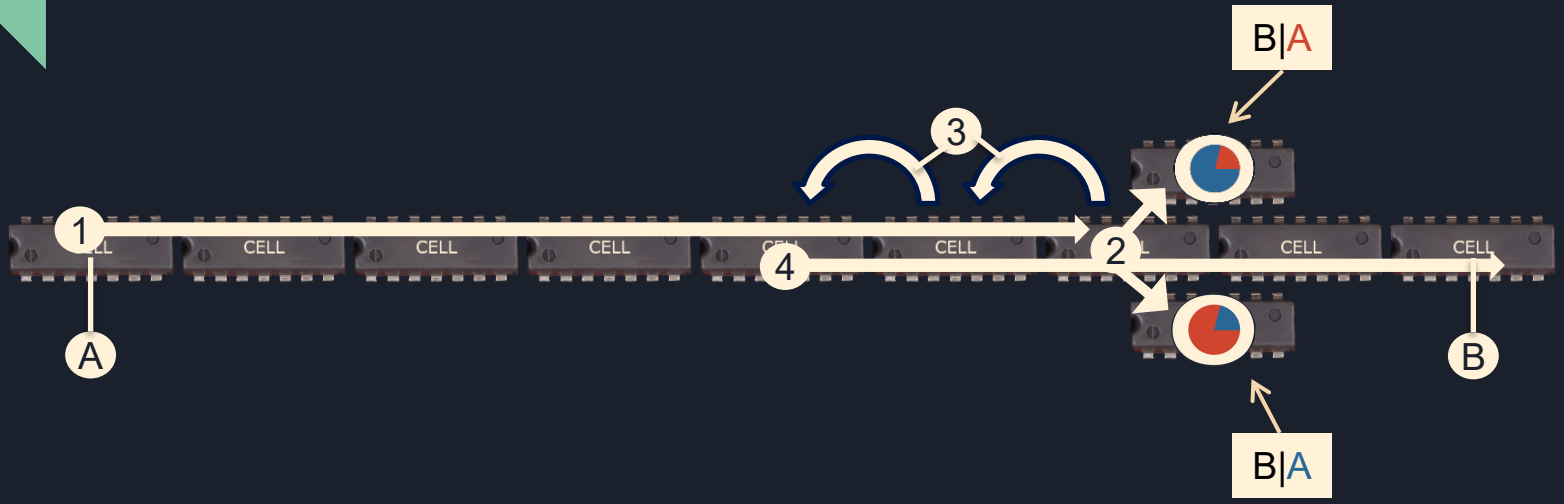
Learning with temporal offsets



“A” was stimulated four clock ticks earlier than “B”.

Brain-CA Technologies

Inference with temporal offsets



1. If “A” is stimulated, the signal is immediately sent to the collision cell.
2. “B” can be predicted based on the value of “A”.
3. The prediction can is slowly routed to the mid-point.
4. The prediction is rapidly sent to “B”. Right place, right time.

Brain-CA Technologies

All cells run the same logic

- Waves encode time and distance using quantized values
- Relationships are discovered at collision points of waves
- Fast Paths are built between related stimulus points
- Predictions are easy

Brain-CA Technologies

Each cell participates in three Subsystems

Communication: Broadcast signals

Memory: Remember relationships

Connection: Build Inference fast paths

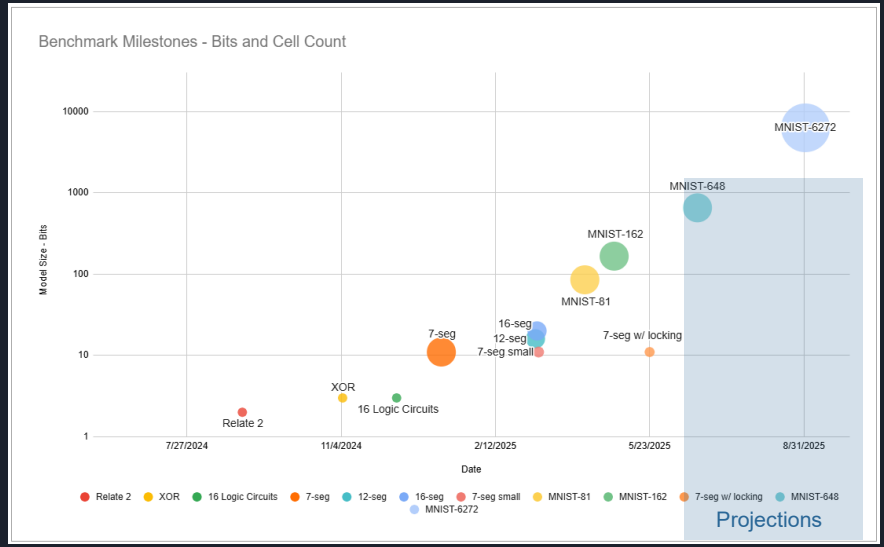
Brain-CA Technologies

Progress to date

Successful simulations on small models

FPGA implementation underway

MNIST expected to be first public results

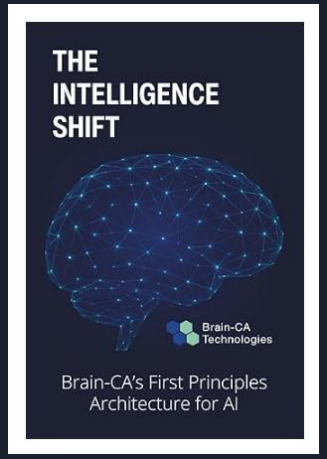


Brain-CA Technologies

Learn More

5:20 – 6:00 Today – Poster Presentation for MLArchSys
1:45 – 2:10 Tomorrow – Presentation for CogArch

Book, Papers, Videos at brain-ca.com/isca



amazon

brain-ca.com/isca