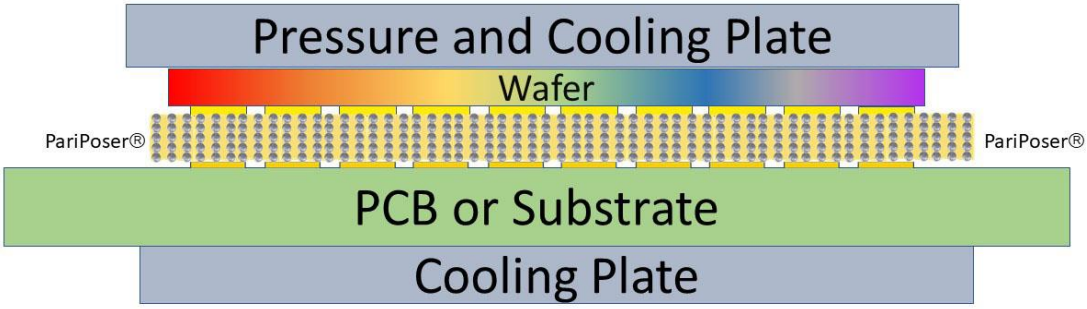
NEW: Wafer Connectors

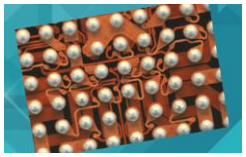
Rev: Sept 18, 2019



A wafer connector concept using the PariPoser® anisotropic elastomer

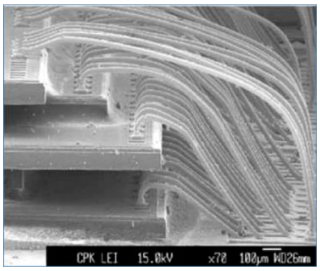
Why have a wafer connector ?

1) Wafers can be made bigger, and can be interconnected more efficiently.
2) The packaging density can improve (on the wafer, and also from the wafer).
3) Wafers can be tested before packaging.
4) Wafers can be connected in a grid array pattern.

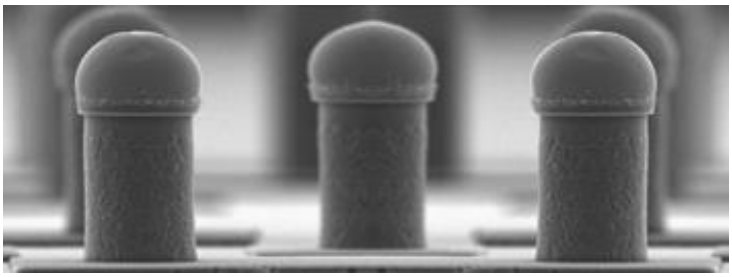


Grid pattern pads on an IC

5) Power and ground distributions to the wafer can be improved. (Wire bonding is messy and has a high inductance. Copper pillars require costly equipment.)



Wire bonding



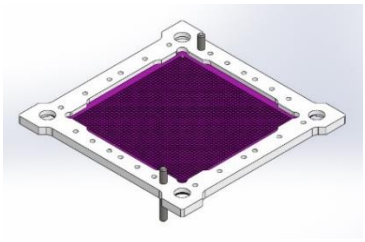
Copper pillars

500 Myles Standish Blvd. Taunton, MA 02780 (508) 676-6888 Fax (508) 676-8111 www.paricon-tech.com sales@paricon-tech.com

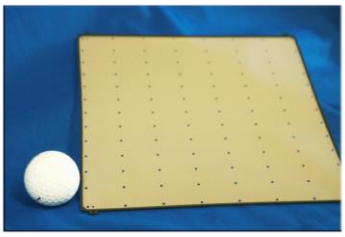
Who could use a wafer connector ?

A wafer connector can be used for small ICs or big ICs. For special function ICs or for multi-cores ICs. For single wafers or stacked memory packages.

The 219mm x 219mm (8.62” x 8.62”) connector shown below is used by Cerebras for their
very big ICs. (Ref: www.cerebras.net) It’s made with a continuous sheet of PariPoser® material mounted on a thin aluminum frame.



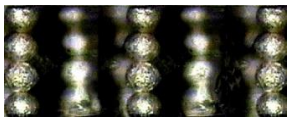
Small ICs



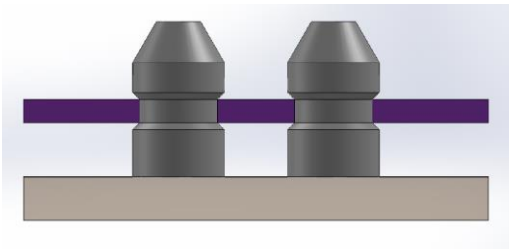
Big ICs (215mm x 215mm)

What can Paricon provide ?

PariPoser® — PariPoser® material is a dense and economical way to connect a wafer to a substrate



PariProbe® — PariProbes® can be added to the PariPoser® material to handle unusual target shapes.

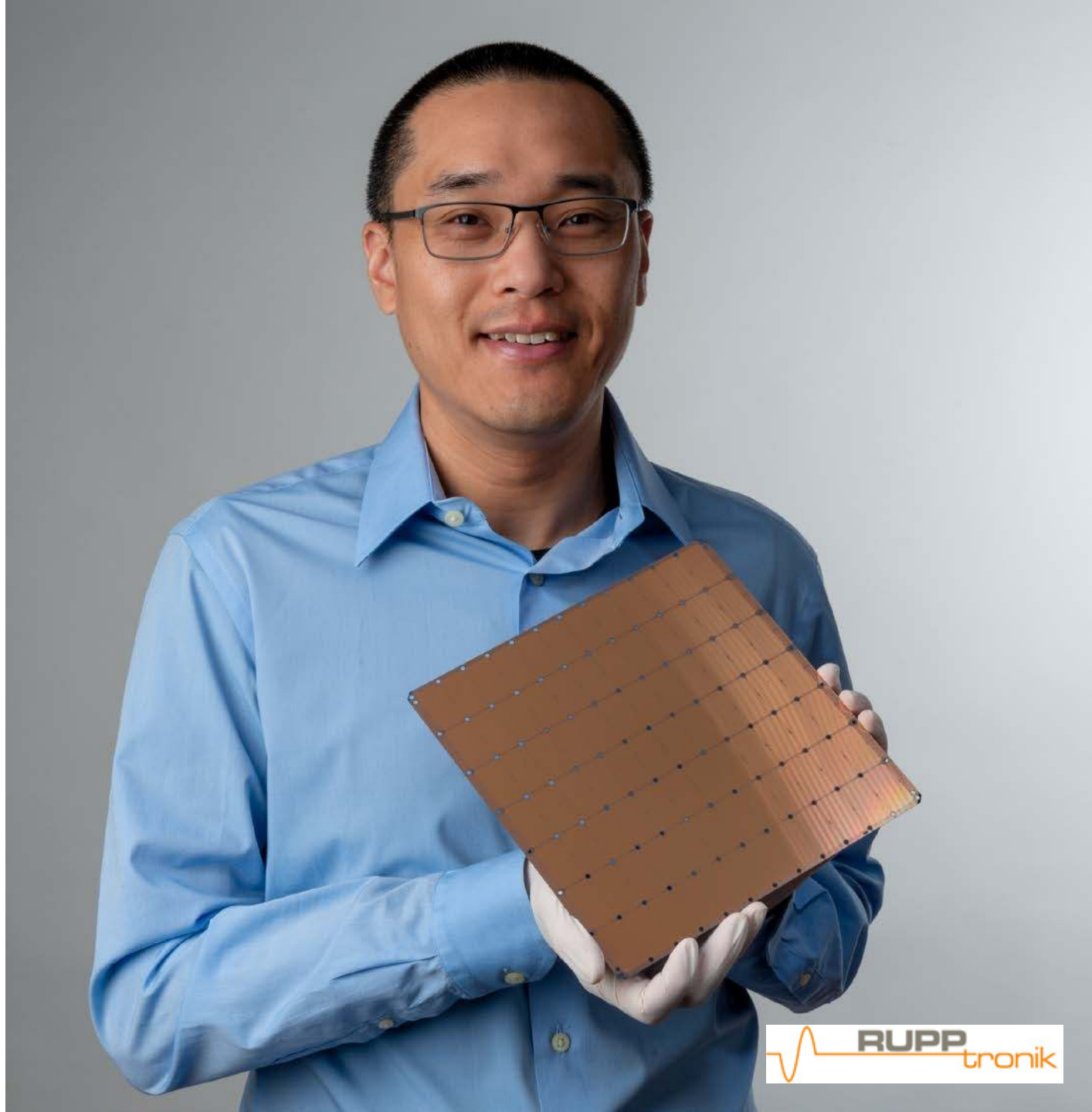


For more Details please contact:

RUPPtronik
Bernd Rupp
+49 (0)8062 80 96 960
Bernd.Rupp@RUPPtronik.de

Largest Chip Ever Built

- 46,225 mm^2 silicon
- 1.2 trillion transistors
- 400,000 AI optimized cores
- 18 Gigabytes of On-chip Memory
- 9 PByte/s memory bandwidth
- 100 Pbit/s fabric bandwidth
- TSMC 16nm process



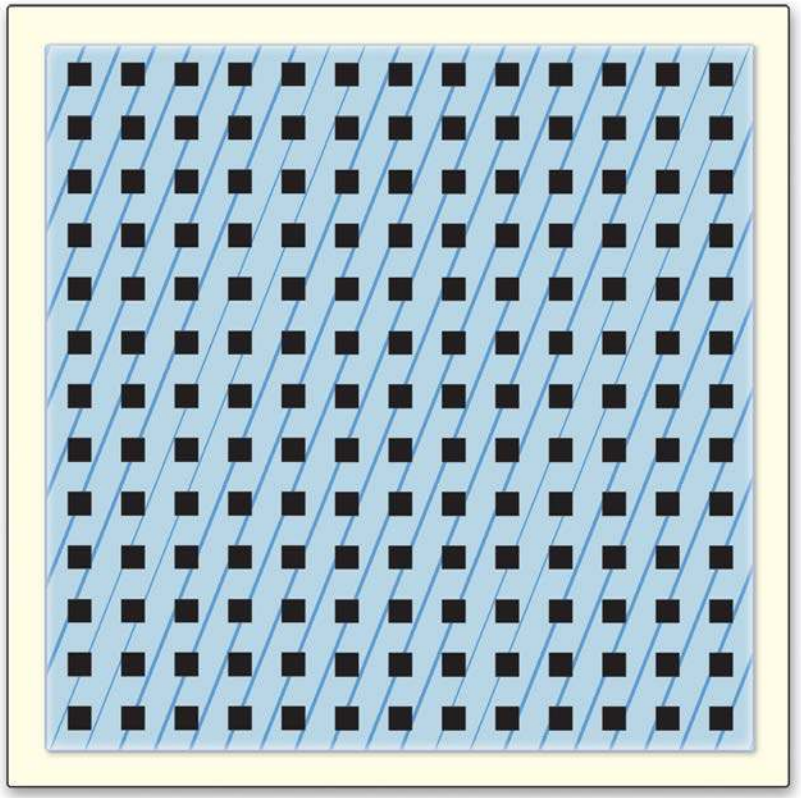
cerebras

A Memory Architecture that *is* Optimized for DL

In neural networks, weights and activations are local, with low data reuse

The right answer is distributed, high performance, on-chip memory

- All memory is fully distributed along with compute datapaths
- Datapath has full performance from memory



Memory uniformly distributed across cores

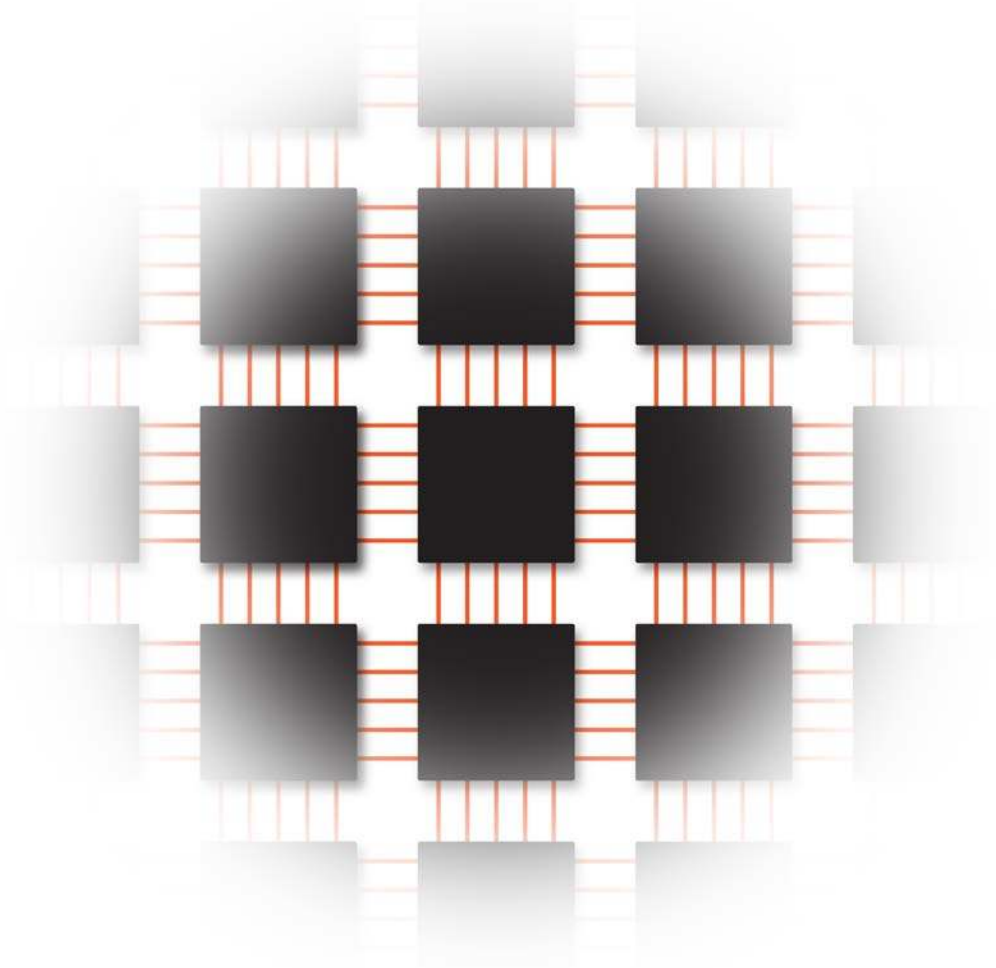
cerebras

RUPP tronik

High-Bandwidth Low-Latency Interconnect

Low latency intra/inter-layer local connectivity with high bandwidth

- Fast and fully configurable fabric
- Small single-word messages
- All HW-based communication avoids SW overhead
- **2D mesh topology** effective for local communication
  - High bandwidth and low latency for local communication
  - Higher utilization and efficiency than global topologies



cerebras

RUPP tronik

The Challenges Of Wafer Scale

Building a 46,225 mm^2, 1.2 Trillion Transistor Chip

Challenges include:

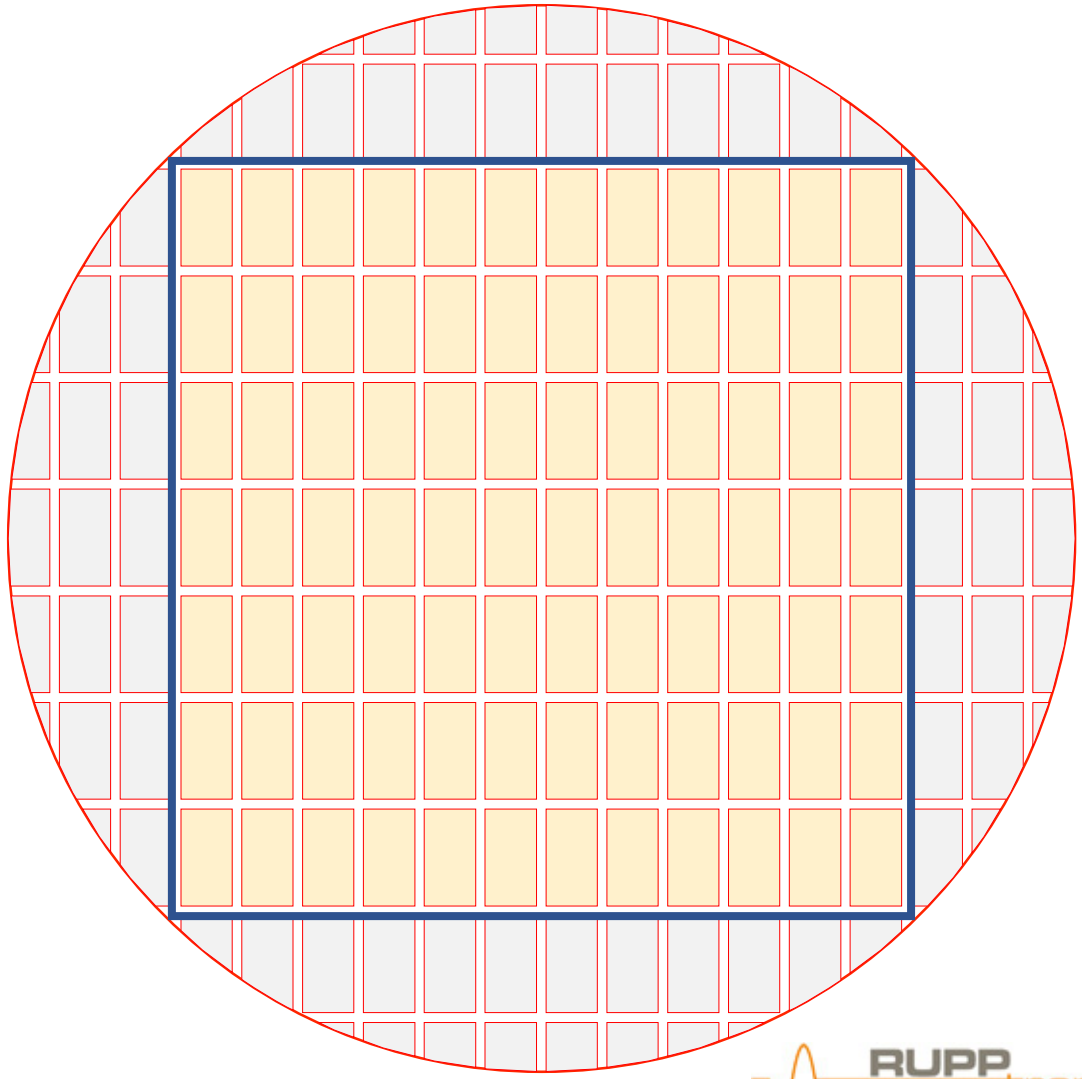
- Cross-die connectivity
- Yield
- Thermal expansion
- Package assembly
- Power and cooling

cerebras

RUPP tronik

Challenge 1: Cross Die Connectivity

- Standard fabrication process requires die to be independent
- Scribe line separates each die
- Scribe line used as mechanical barrier for die cutting and for test structures

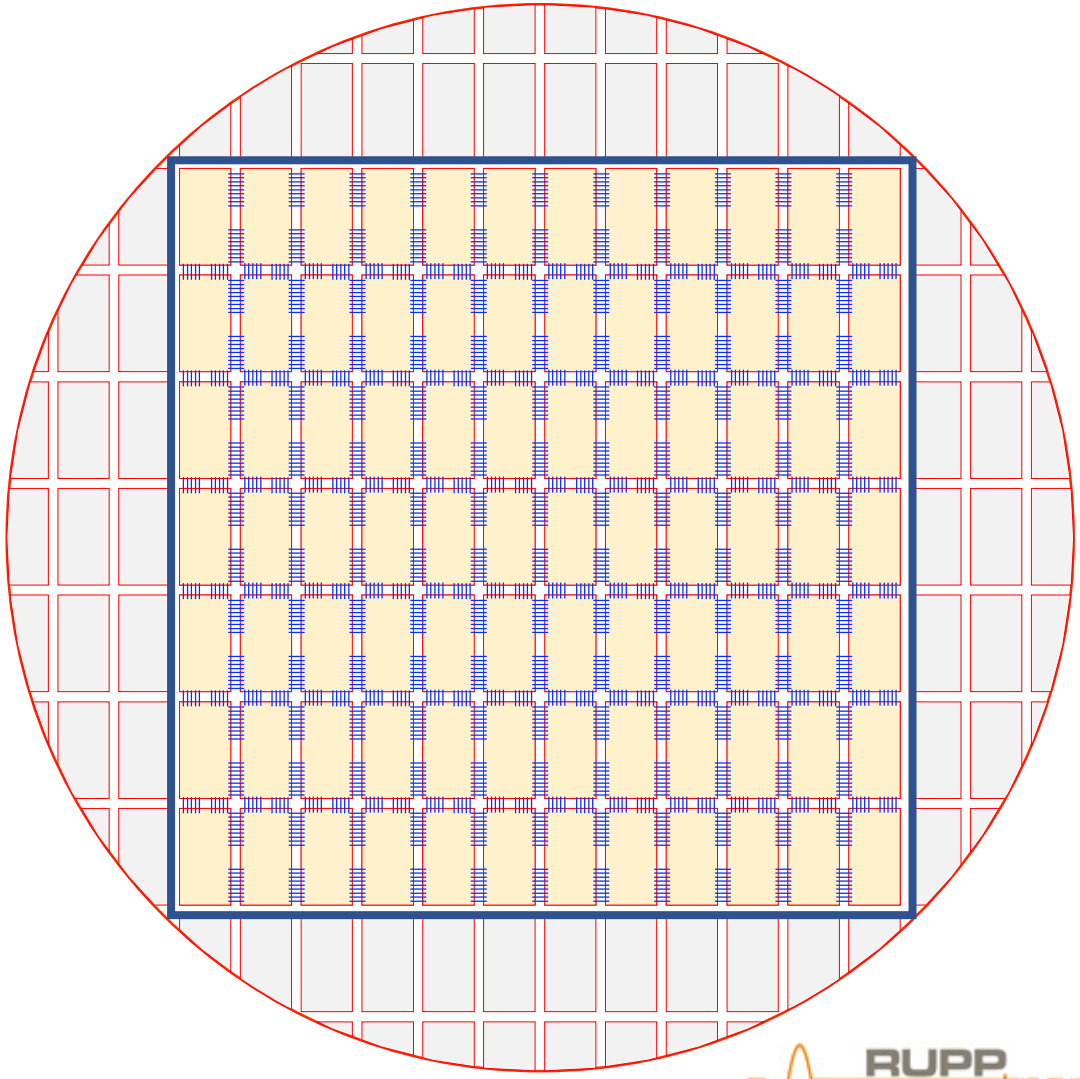


cerebras

RUPP tronik

Cross-Die Wires

- Add wires across scribe line in partnership with TSMC
- Extend 2D mesh across die
- Same connectivity between cores and across scribe lines create a homogenous array
- Short wires enable ultra high bandwidth with low latency



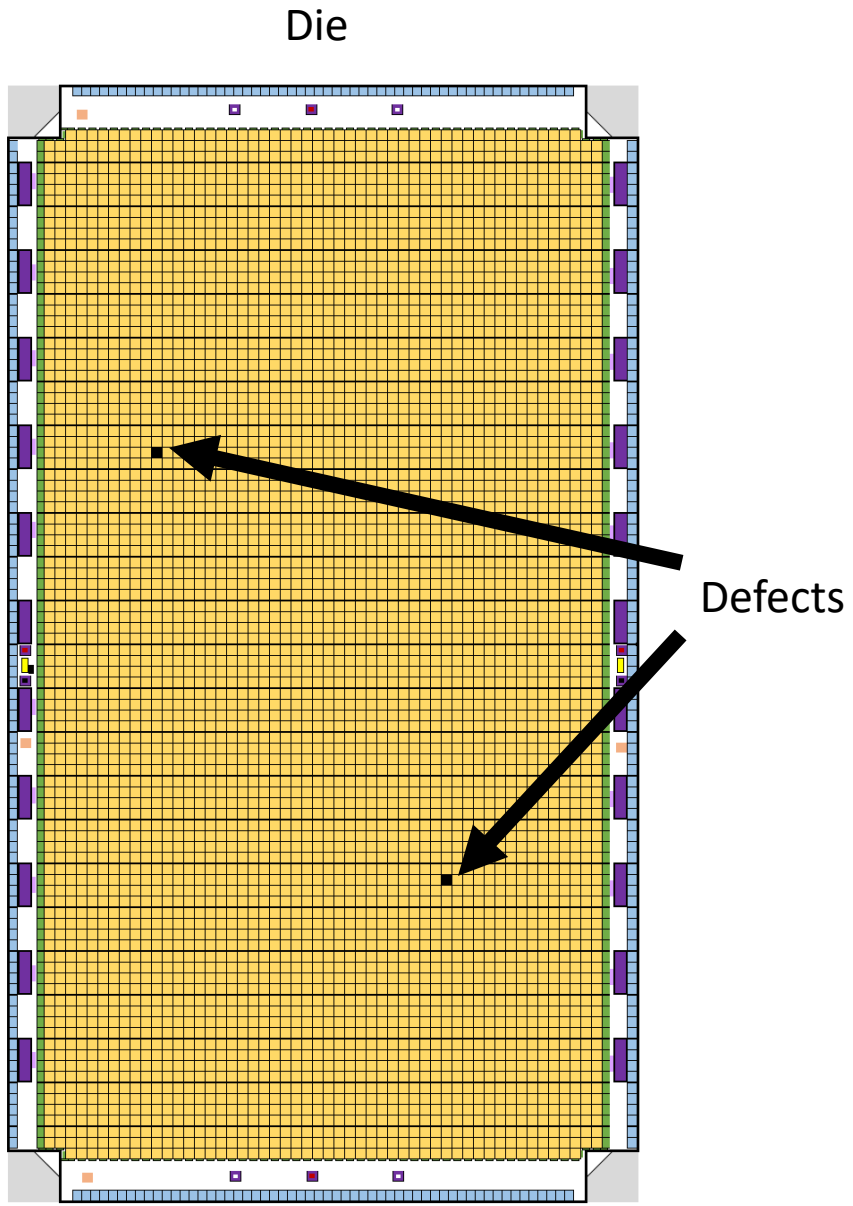
cerebras

RUPPtronik

Challenge 2: Yield

Impossible to yield full wafer with zero defects

- Silicon and process defects are inevitable even in mature process

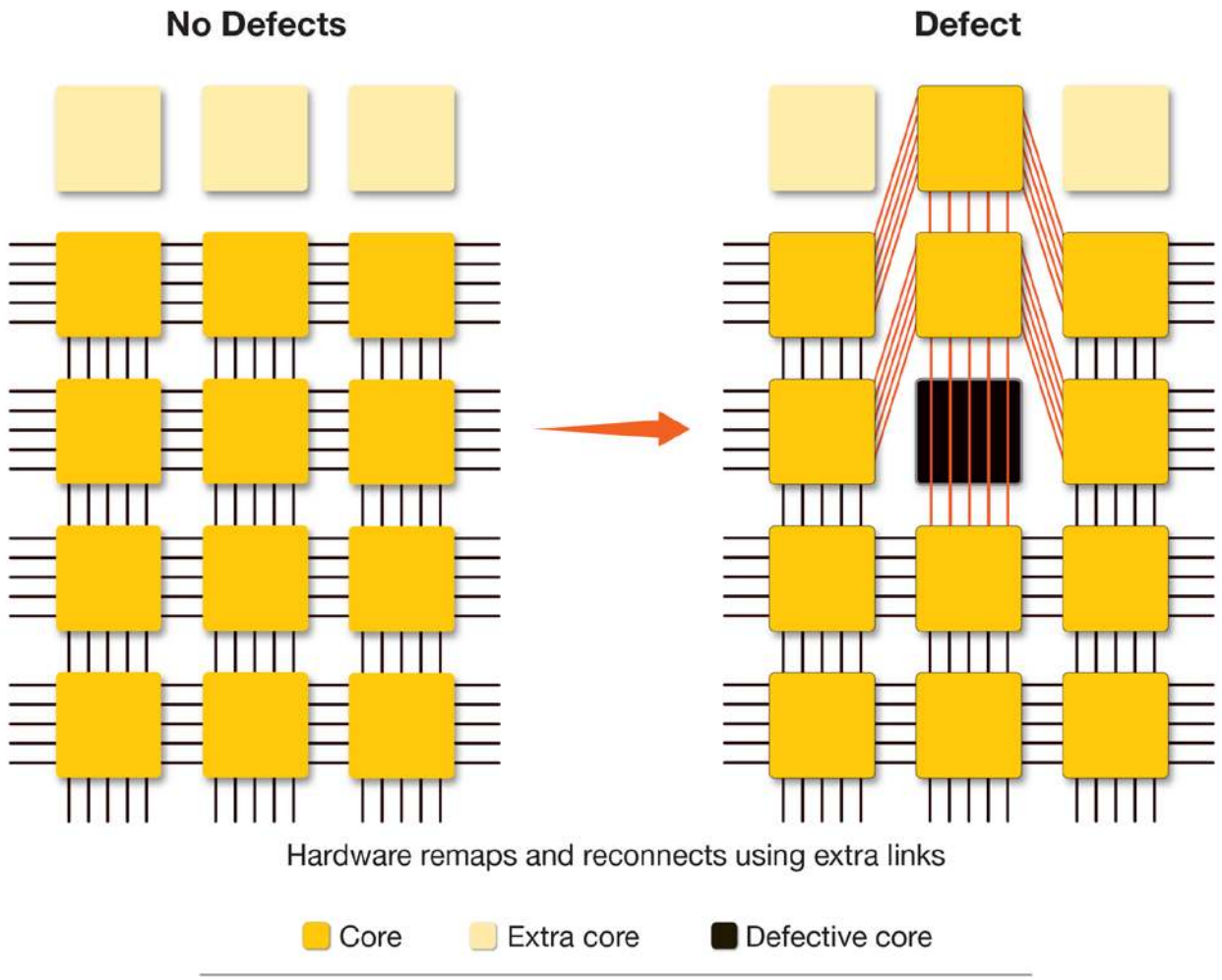


cerebras

RUPP tronik

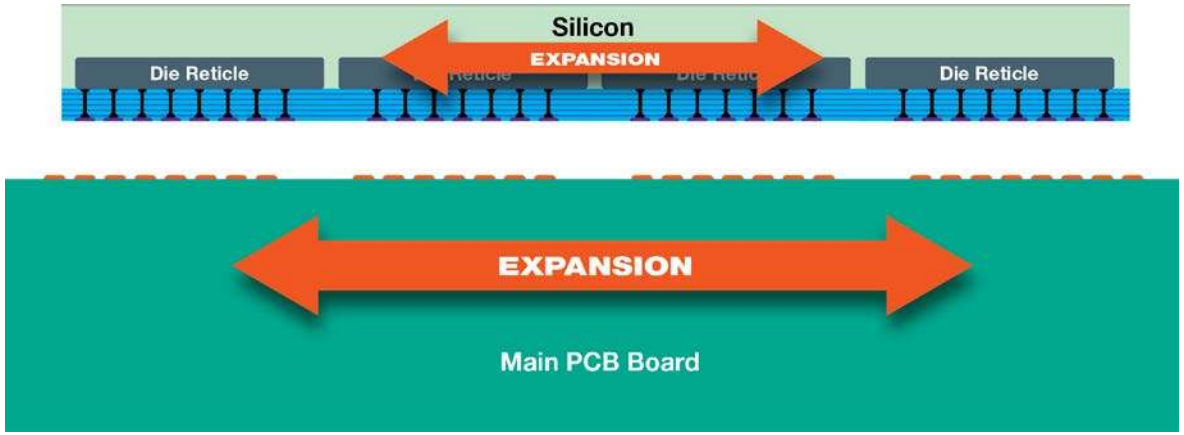
Redundancy is Your Friend

- Uniform small core architecture enables redundancy to address yield at very low cost
- Design includes redundant cores and redundant fabric links
- Redundant cores replace defective cores
- Extra links reconnect fabric to restore logical 2D mesh



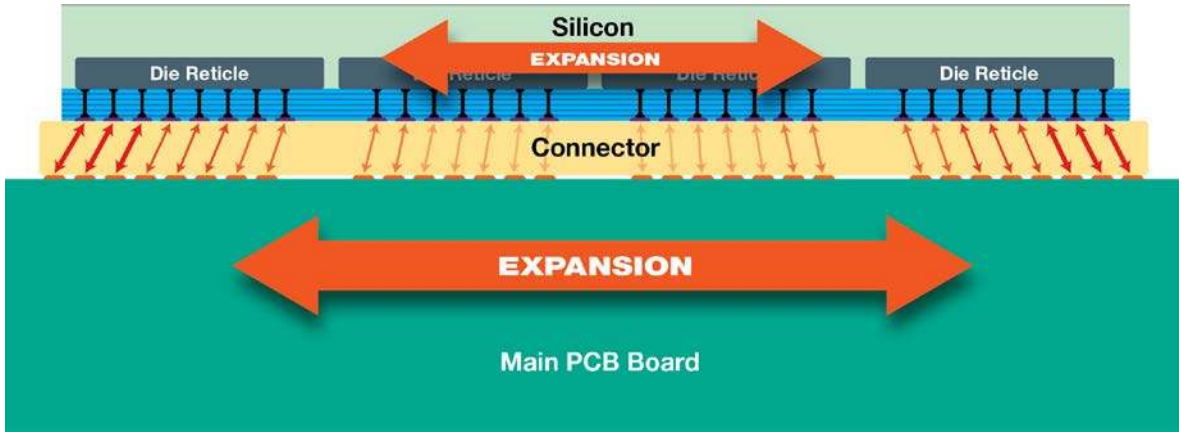
Challenge 3: Thermal Expansion in the Package

- Silicon and PCB expand at different rates under temp
- Size of wafer would result in too much mechanical stress using traditional package technology



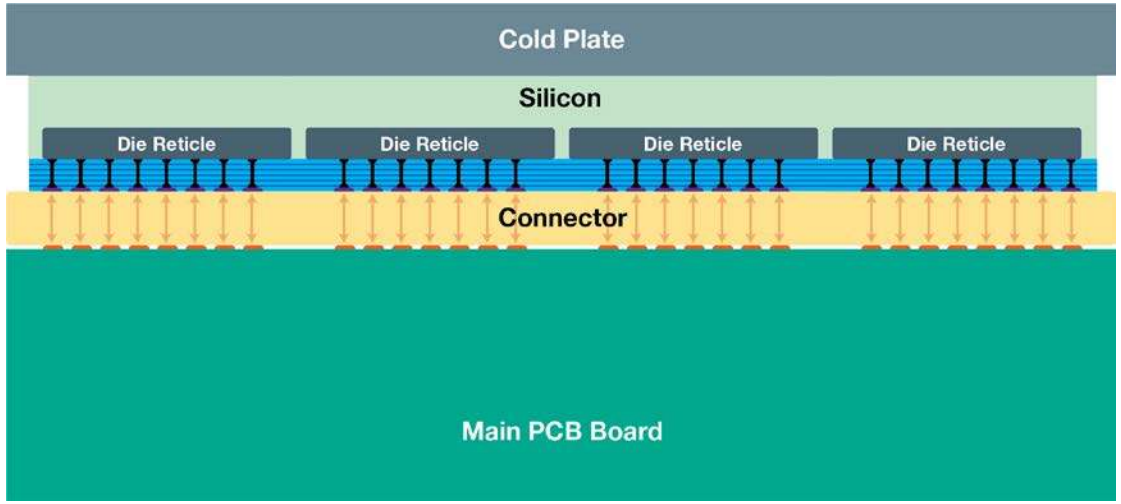
Connecting Wafer to PCB

- Developed custom connector to connect wafer to PCB
- Connector absorbs the variation while maintaining connectivity



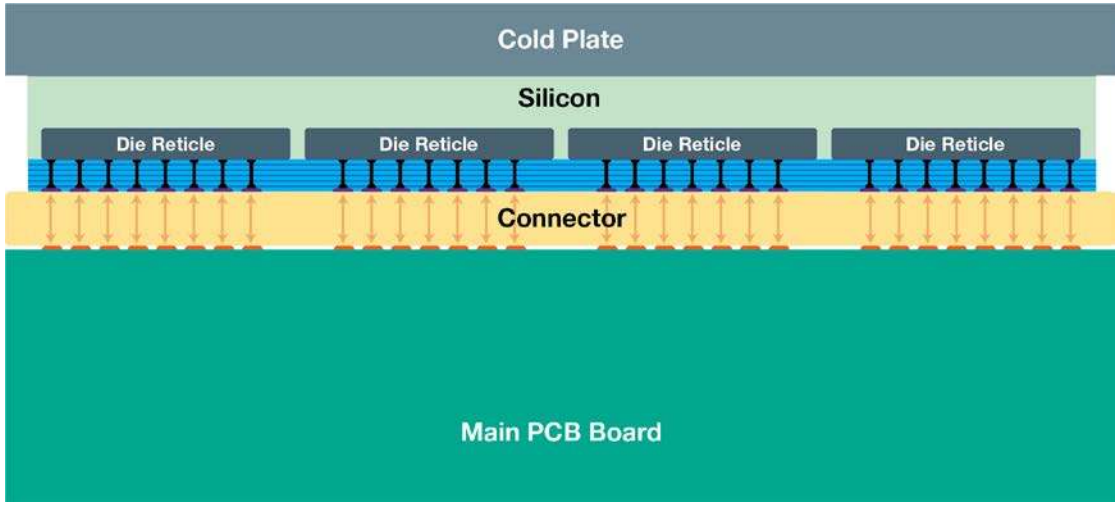
Challenge 4: Package Assembly

- No traditional package exists
- Package includes:
  - PCB
  - Connector
  - Wafer
  - Cold plate
- All components require precise alignment



Custom Packaging Tools

- Developed custom machines and process
- Tools to ensure precision alignment
- Tools for special handling

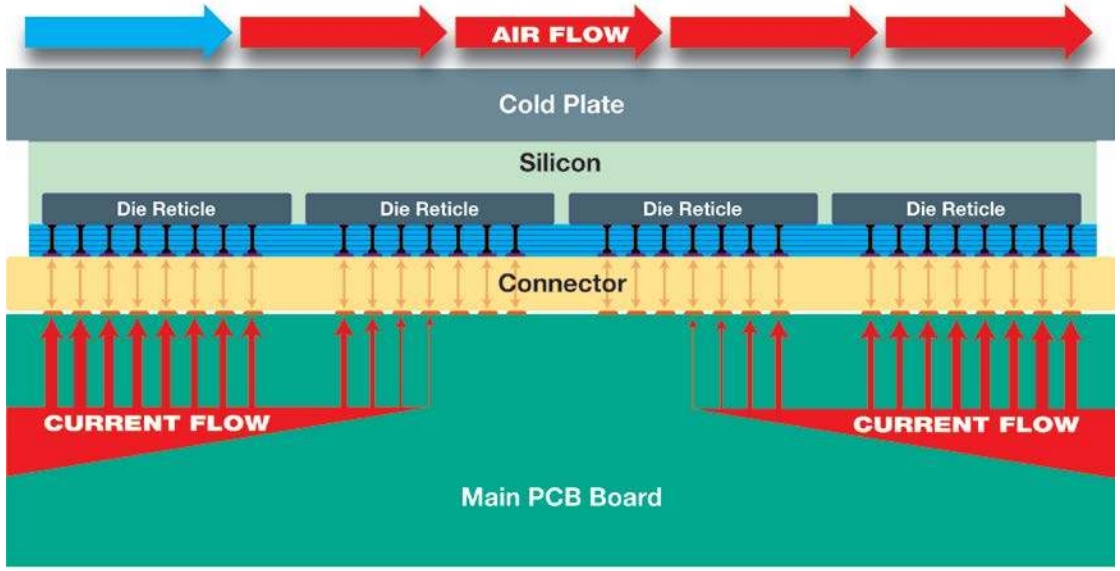


cerebras

Challenge 5: Power and Cooling

Concentrated high density exceeds traditional power & cooling capabilities

- Power delivery
  - Current density too high for power plane distribution in PCB
- Heat removal
  - Heat density too high for direct air cooling

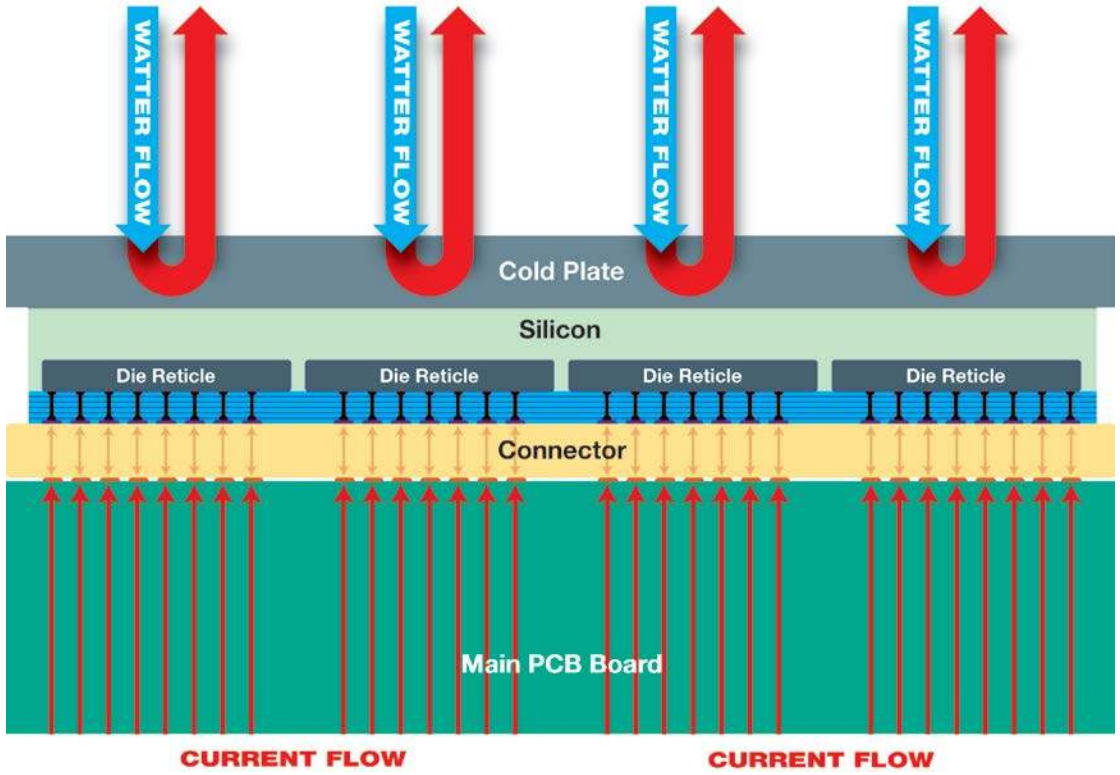


cerebras

RUPPtronik

Using the 3rd Dimension

- Power delivery
  - Current flow distributed in 3rd dimension perpendicular to wafer
- Heat removal
  - Water carries heat from wafer through cold plate



It’s working,
running customer workloads.

Stay tuned…

RUPPtronik