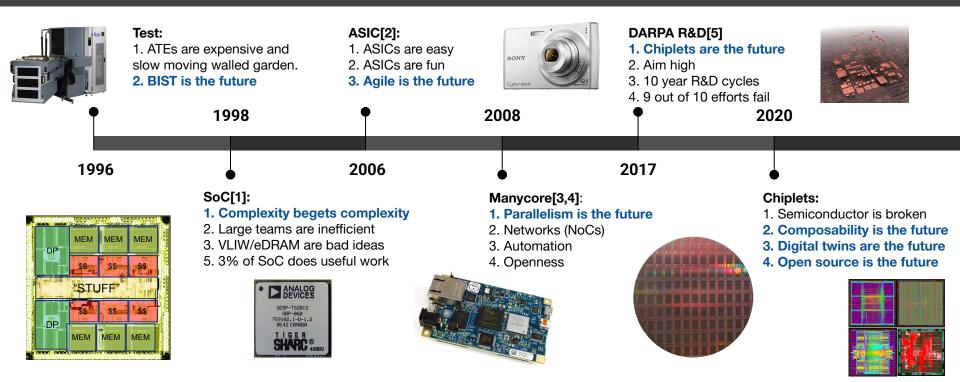
# zer | o |

# Building Composable Chiplets

Apr 16th -17th, 2025 Boston MA

Andreas Olofsson andreas@zeroasic.com

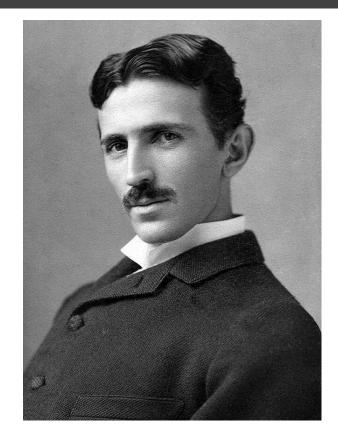
### Chip(let) lessons from the trenches



[1] Adelman, Olofsson et al (2004), "A 600 MHz DSP with 24 Mb embedded DRAM with an enhanced instruction set for wireless communication", ISSCC
 [2] Olofsson et al (2008). A variable width software programmable data pattern generator (U.S. Patent No. 8,006,114)
 [3] Olofsson et al (2011) A 1024-core 70 GFLOP/W floating point manycore microprocessor, High Performance Embedded Computing Conference
 [4] Olofsson et al (2014), "Kickstarting high-performance energy-efficient manycore architectures with Epiphany" 48th Asilomar Conference on Signals, Systems
 [5] Olofsson et al (2018), Enabling High-Performance Heterogeneous Integration via Interface Standards, IP Reuse, and Modular Design, IMAPS

"Den som är före sin tid får invänta framtiden på en obekväm plats." – Lennart Lubeck, CEO Swedish Space Corporation (1980's)

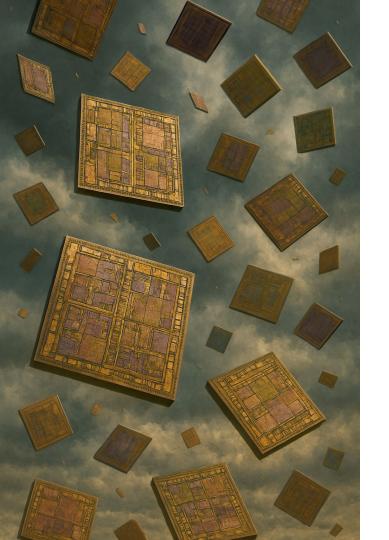
"Those who are ahead of their time often have to wait for it in uncomfortable quarters." – Stanislaw Lec (1909-1966), Polish aphorist, poet



### 2025 Chiplet Report Card

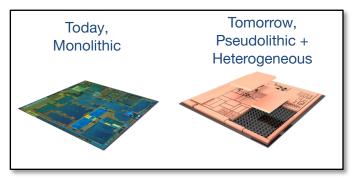
- No open access chiplets for sale
- No full-stack chiplet standard
- **No** solution to the PPM/KGD problem
- No solution to margin stacking problem
- **No** solution to the chip rework problem
- No 3rd party SOTA chiplet integrators
- No low volume advanced packaging
- **No** funding for chiplet ecosystem development
- No viable PCB like design ecosystem



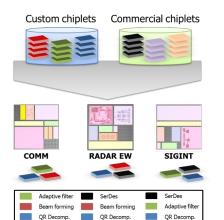


# How Did we Get Here?

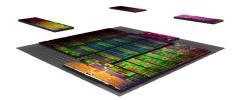
# (2016) DARPA CHIPS Program Launch



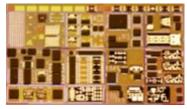
REF: DARPA @ Semicom Design West 2019



- A universal efficient interface standard
- SOTA manufacturing assembly
- A large and critical set of IP chiplets



Extend Moore's law Scale out and scale down while managing yield





Materials/processes, companies, geography, security

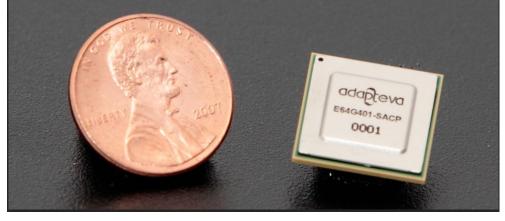


System Integration Democratize access to leading edge silicon for system integrators

### (2016) CHIPS Proposers Day

#### **Cost Constrained CHIP Design**

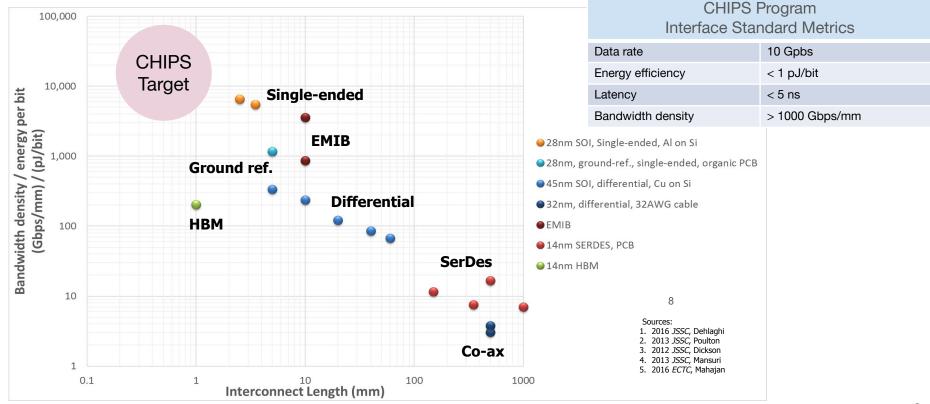
by Andreas Olofsson, Adapteva (9/17/16)



#### **Our I/O Challenges** Needed massive IO to support 4 TFLOPS! Initial plan was 128 x 10Gbit SERDES lanes ...but cost made monolithic integration impossible fallback was 1024 1.8V CMOS pins running at @ 150MHz... **One Possible Chip to Chip Interface** Make sure 50um bumps are available to all customers Drive parallel interfaces (clk, frame, wait, data[N-1:0]) Reference RTL: github.com/parallella/oh CMOS signaling using thin oxide transistors (0.8V) Energy Target: 0.2pJ / bit

Density Target: 2Tbit / mm<sup>2</sup>

#### (2017) CHIPS Standards War



[REF] Olofsson et al (2018), Enabling High-Performance Heterogeneous Integration via Interface Standards, IP Reuse, and Modular Design, IMAPS

# (2018) CHIPS Open source AIB FTW!

- AIB (Advanced Interface Bus) is a PHY-level interface standard for high bandwidth, low power die-to-die communication
  - Clock-forwarded parallel data transfer like DDR DRAM
  - High density with 2.5D interposer (e.g., CoWoS, EMIB) for multi-chip packaging
  - PHY level only (OSI Layer 1)
  - Protocols like AXI-4 can be built on top of AIB

#### • AIB Performance:

- 1 Tbps/mm shoreline
- ~1pJ/bit
- <5ns latency</li>

#### **Open Source!**

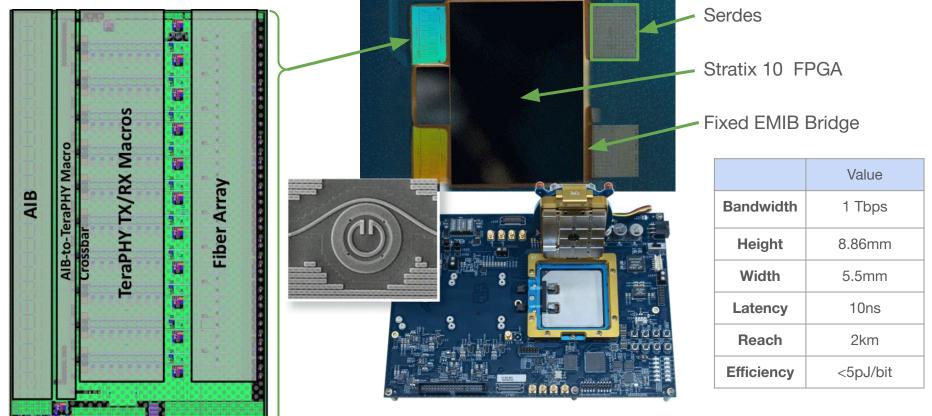
- Standard and reference implementation
- https://github.com/chipsalliance/aib-phy-hardware



#### **AIB Adopers**

- Boeing
- Intrinsix
- Synopsys
- Lockheed Martin
- Sandia
- Jariet
- NCSU
- U. of Michigan
- Ayar Labs

### (2019) CHIPS HI Win #1, Photonic Interconnect



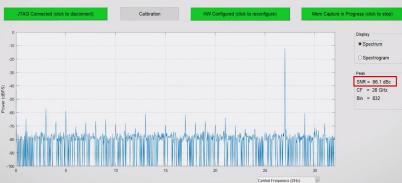
[REF] Wade (2019), "A Chiplet Technology for Low-Power, High-Bandwidth in-Package Optical I/O", Hot Chips 10

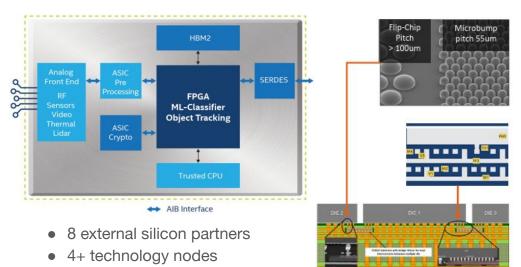
### (2019) CHIPS HI Win #2, Mixed Signal FPGAs



#### intel.

64.0G 1024 Channels Analysis Filter Bank

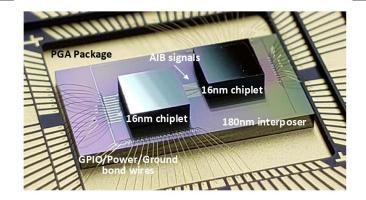




- 3 FPGA families
- 3 data converter chiplets
- 2 ASIC compete chiplets
- 9 serdes/optical IO chiplets

[REF] Shumarayev (2022), "Heterogenous Integration Enables FPGA Based Hardware Acceleration for RF Applications", Hot Chips

# (2019) CHIPS HI Win #3, Collaborative Innovation



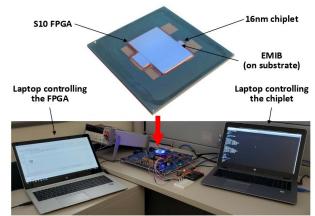
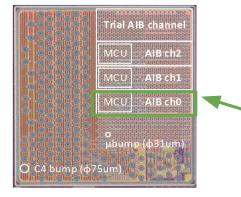


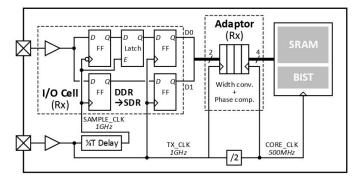
Fig. 5. A 16nm chiplet is integrated with an Intel Stratix 10 FPGA via EMIB on the package substrate.

	This Work
Technology	16nm FinFET
Voltage swing	0.9V
Bump pitch	55um
Chiplet carrier	Silicon interposer 3-layer / EMIB 4-layer
Reach	2mm
I/O size	203.2um <sup>2</sup> /b
Data rate per pin	2Gb/s
Energy efficiency	0.83pJ/b
Shoreline BW density	256Gb/s/mm
Area BW density	614.4Gb/s/mm <sup>2</sup>
Latency	4ns



Each AIB channel contains 96 signal and 42 power/ground µbumps, occupying 312.5µm × 1246.5µm

[REF] Liu, et al (2021), "A 256Gb/s/mm-shoreline AIB-Compatible 16nm FinFET CMOS Chiplet for 2.5D Integration with Stratix 10 FPGA on EMIB and Tiling on Silicon Interposer", IEE CICC



### (2019) CHIP $\rightarrow$ SHIP $\rightarrow$ STEAMPIPE Transition

#### NEWS | Oct. 31, 2019

NSWC Crane leverages OTA to ensure that the U.S. Government has access to secure state-ofthe-art design, assembly, packaging and test for state-ofthe-art microelectronics

By NSWC Crane Corporate Communications

Andreas Olofsson, DARPA PM for the Common Heterogeneous Integration and IP Reuse Strategies (CHIPS) program said, "The future of computing hardware is specialized, heterogeneous and parallel."

CHIPS is a precursor for SHIP, and with the below stated goals it is serving as a transition partner to SHIP:

- · Establish and demonstrate common interface standards
- Enable the assembly of systems from modular IP blocks built with these established standards
- Demonstrate reusability of the modular IP blocks via rapid design iteration

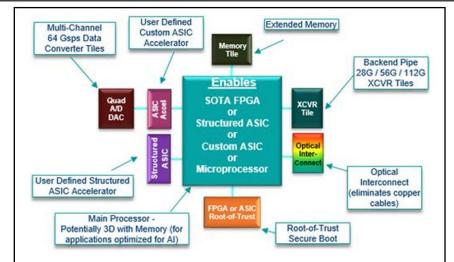
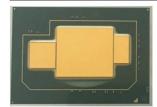
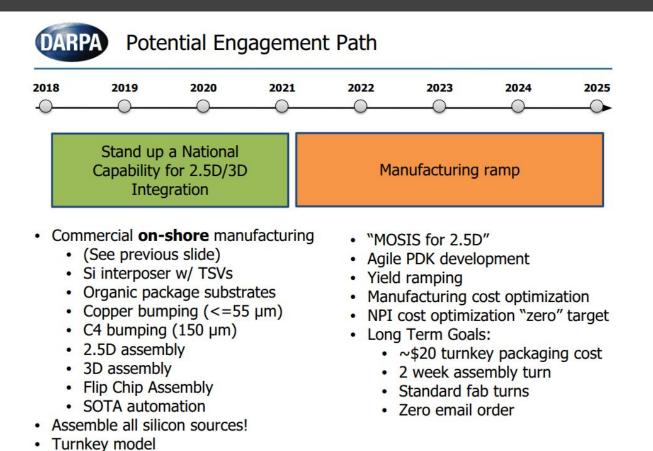


Figure 1. Notional Heterogeneous Integration Example



[REF] Shenoy et al (2023), "DoD Microelectronics: Heterogeneous Integration with Compound Semiconductors and Photonics", MANTECH

### (2019) DARPA CHIPS 2.0 Workshop



It has been 6 years... how long until we have this in place?!

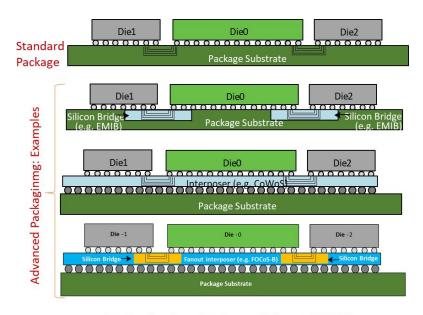
### (2019) My DARPA 2025 Predictions



Conclusion: My 2025 Predictions...

- We will have no human in the loop general purpose silicon compiler (RTL/schematic  $\rightarrow$ GDSII)
- We will experience FOSS "GCC/LLVM" for ASIC and FPGA design
- · Domain specific compilers sitting on top of the silicon compiler will proliferate
- PCBs will be designed using programming languages, not schematic entry tools
- You will be able to download production quality analog & digital FOSS IP
- Building heterogeneous System-In-Package will be as easy as easier than PCB design
- ML ASICs will be ubiquitous
- All major system companies will design their own silicon
- Consumers will order custom "N=1" silicon

### (2022) UCIe Standard

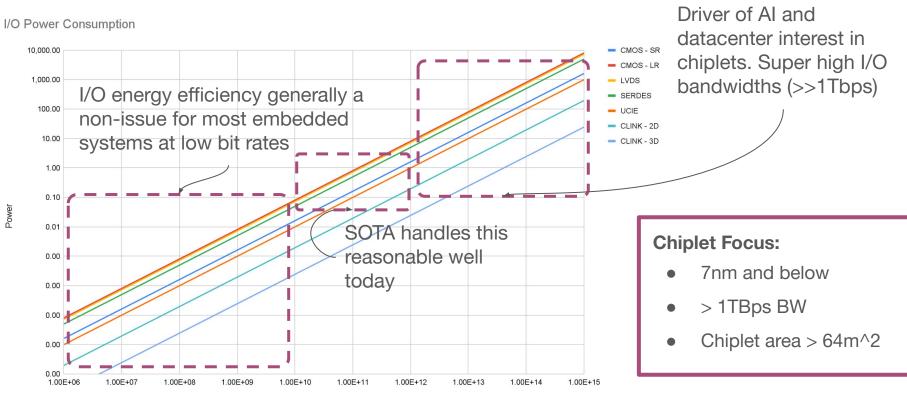


(b. Packaging Options: 2D and 2.5D)

Characteristics / KPIs	Standard Package	Advanced Package
Characteristics		
Data Rate (GT/s)	4, 8, 12, 16, 24,	32
Width (each cluster)	16	64
Bump Pitch (um)	100 - 130	25 - 55
Channel Reach (mm)	<= 25	<=2
Target for Key Metrics		
B/W Shoreline (GB/s/mm)	28 – 224	165 – 1317
B/W Density (GB/s/mm <sup>2</sup> )	22-125	188-1350
Power Efficiency target (pJ/b)	0.5	0.25
Low-power entry/exit	0.5ns <=16G, 0.5-1ns >=24G	
Latency (Tx + Rx)	< 2ns	
Reliability (FIT)	0 < FIT (Failure II	n Time) << 1

TLDR: Big, fragmented, complex, expensive, not composable...but will likely find sockets in datacenter. What about everyone else?

#### (2025) Datacenter consuming all the chiplet oxygen



# zer | o |

# Zero ASIC Composable Chiplet Journey

(2020 - present)

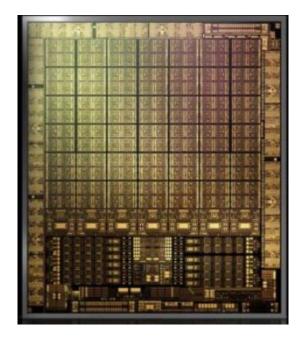
#### (2020) Heilmeier Questions

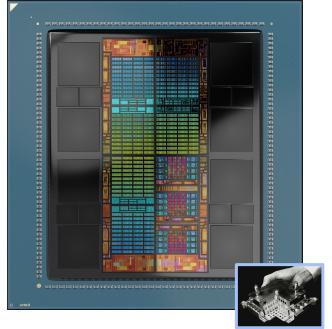
- 1. What are we proposing? "LEGO for chiplets"
- 2. How is it done today? Tower of Babel of bespoke chiplets
- 3. What is new in our approach? A system of composable chiplets
- 4. Why does it matter? Potentially cuts design time and cost by a factor of 100
- 5. What are the risks? Disrupting 50 years of Moore's law inertia
- 6. How much will it cost? \$100M \$1B
- 7. How long will it take? 5 years
- 8. What are key milestones? First viable composable chiplet based system

### Three Er(r)a(r)s of Chip Design



**Discrete Era** Tyranny of Wires (1940 – present)



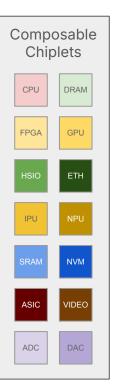


Monolithic Era \$1-10B Per Generation (1960 – present)

Chiplet Era Private Bespoke Islands (1980 - present)

20

#### Could we build "Amino acids for silicon systems"



CPU	CPU	CPU	CPU
DRAM	DRAM	DRAM	DRAM
DRAM	DRAM	DRAM	DRAM
DRAM	DRAM	DRAM	DRAM

(CPU)

ASIC	ASIC	ASIC	ASIC
ASIC	ASIC	ASIC	ASIC
ASIC	ASIC	ASIC	ASIC
ASIC	ASIC	ASIC	ASIC

Application Specific Integrated Circuit (ASIC)

CPU	DRAM	CPU	DRAM
DRAM	CPU	DRAM	CPU
CPU	DRAM	CPU	DRAM
DRAM	CPU	DRAM	CPU

Processing-In-Memory (PIM)

DRAM	DRAM	DRAM	DRAM
PE	PE	PE	PE
PE	PE	PE	PE
PE	PE	PE	PE

General Purpose GPU (GPGPU)

LUT	SRAM	LUT	SRAM
SRAM	LUT	SRAM	LUT
LUT	SRAM	LUT	SRAM
SRAM	LUT	SRAM	LUT

Field Programmable Gate Array (FPGA)

	SRAM		SRAM
SRAM		SRAM	
	SRAM		SRAM
SRAM	AI	SRAM	AI

Application Specific Processor (ASIP)

CPU	SRAM	CPU	SRAM	
SRAM	CPU	SRAM	CPU	
CPU	SRAM	CPU	SRAM	
SRAM	CPU	SRAM	CPU	
Manycore				

CPU

PE SRAM PE SRAM SRAM PE SRAM PE PE SRAM PE SRAM SRAM PF SRAM PE

Coarse Grained Reconfigurable Array (CGRA)

CPU (S)	CPU (M)	CPU (XL)	DRAM
SRAM	NVM	GPU	IPU
	PE	FPGA	VIDEO
ASIC	CRYPT	DPU	HSIO

Heterogeneous System-On-Chip (SoC)

	10	10	
10	FPGA	FPGA	10
10	FPGA	FPGA	10
	10	10	
	FP	Cost GA	
	10	10	
10	PE	PE	10
10	HSIO	SRAM	10
_	10	10	

	D.	36	
	10	IO	
10	HSIO	CPU	10
10		FPGA	10
-	10	10	



	10	10	10	10	10	
10	USB	DDR	PCIE	MIPI	ЕТН	10
10	SRAM	GPU	IPU	VIDEO	CPU	10
	10	10	10	10	10	

Low Cost SoC

	10	10		
0	CPU	PE	ю	ю
0	SRAM	FPGA	10	ю
	10	10		
Н	leterog	geneo	us	Hig

IO CPU

IO SRAM

10 10

PE

10 10

10 10

IO ADC

Heterogeneous

DSP

10 10

AI

Microcontroller

NVM 10

CPU IO

IO HSIO

FPGA

CPU IO

IPU IO



SRAM FPGA IO

10 10

10 10

HSIO PE

> FPGA 10 10 IO PCIE DDR IO ETH 10 10

> > Microprocessor

10 10

PE

IO ADC FPGA IO

10 10

Mixed Signal

Microcontroller



DSP



AI

ASIC

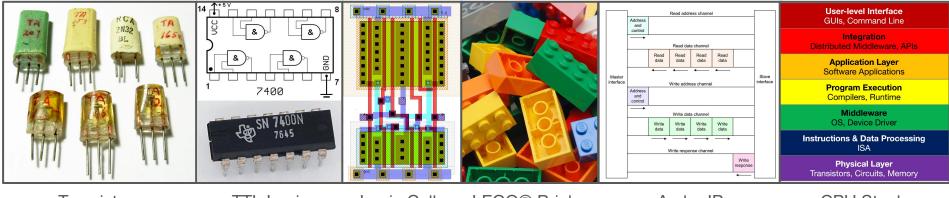


High Performance SoC



21

#### **Composable Hardware Inspiration**



Transistors

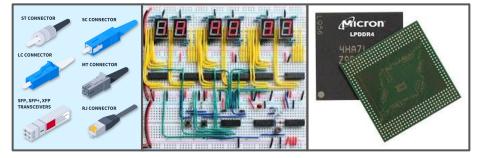
TTL Logic

Logic Cells

**LEGO® Bricks** 

Amba IP

**CPU Stack** 



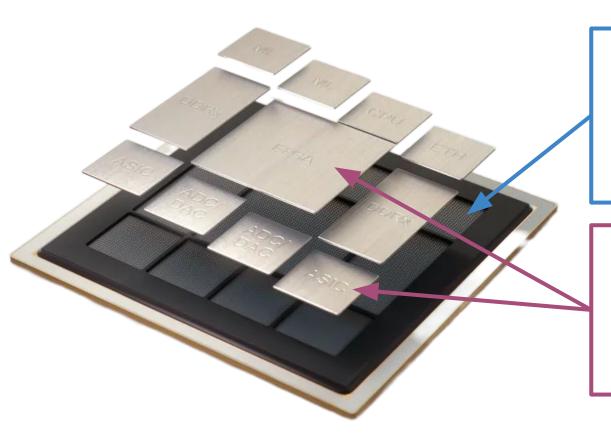
Composable hardware systems can be **effectively** constructed by connecting together independently developed modular and reusable components.

Ethernet Breadboard JEDEC DRAM

### Key Composable Chiplet Optimization Questions

Question	Range	Conclusion
1. Mechanical Structure	2D, 2.5D, 3D	3D
2. Substrate Technology	Organic, glass, Si (active/passive),	Active Silicon
3. Chiplet Types	FPGA, CPU, ML, SRAM, DRAM,	Many
4. Chiplet sizes	1 mm <sup>2</sup> to 858 mm <sup>2</sup>	Discrete grid
5. Interconnect Pitch	1 um to 150 um	45 um →8um →4um →
6. Standard	UCIe, BOW, AIB, HBM,	CLINK + EBRICK + UMI

#### Zero ASIC's Composable Chiplet Approach



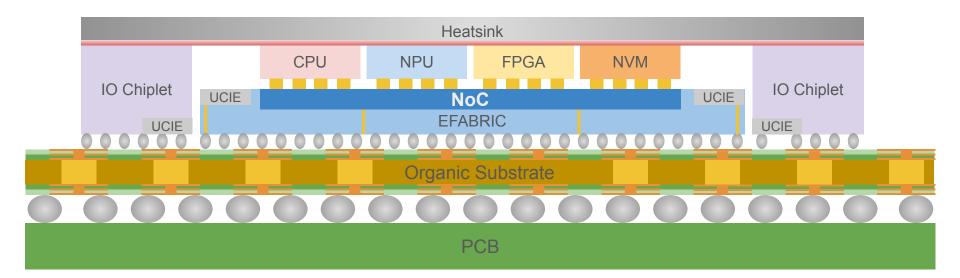
#### **EFABRIC**

- Active silicon substrate
- Fixed mechanical grid connections
- Built in NoC, clocking, management
- Shared memory architecture
- 3D chiplet links
- Scale out 2D I/O

#### **EBRICKS**

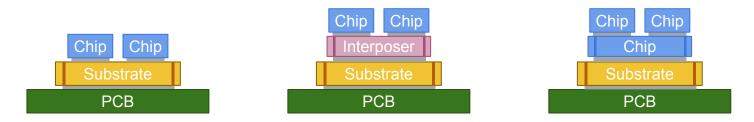
- Discretized chiplet sizes
- 3D chiplet point-to-point links
- CPU, FPGA, NPU, etc, ...
- 100% interchangeable/swappable
- Rotationally symmetric footprints
- Rigid specification (aka like ethernet)

#### EFABRIC Cross Section (v2)



- Optimized for manufacturability, performance, and supply chain security
- 45um 3D bumps, 110um I/O bumps, 100um chiplet spacing

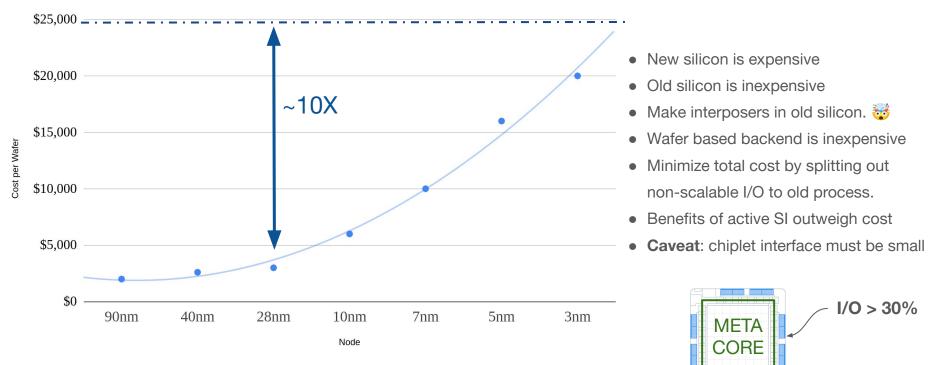
### Q{1,2}: Mechanical Topology



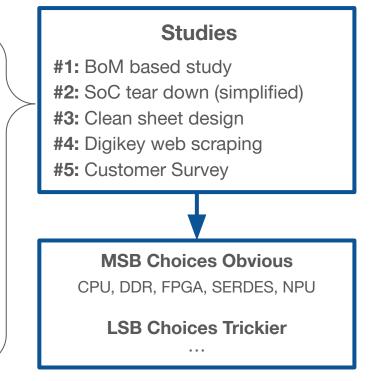
	2D	2.5D	3D
Wire Length	1000um - 5000 um	1000um - 5000 um	< 100 um
Wire Density	50 wires/mm/layer	500 wires/mm/layer	500 - 10,000 wires / mm <sup>2</sup>
Cost	Low	Medium	Medium
Mfg Risk	Low	High	Medium

## A{1,2}: Debunking Myth of Expensive Si

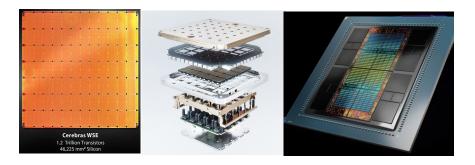
Cost per Wafer vs. Node



	Snapdragon 8	NXP MX8+	NVDA Orin	AMD Zynq
CPU				
DDRx				
NPU				
GPU				
DSP				
FPGA				$\checkmark$
Serdes				
Other				



### Q4: What Is the optimal chiplet size?



I/O communication costs are prohibitive, so maximum die are optimal for large problems

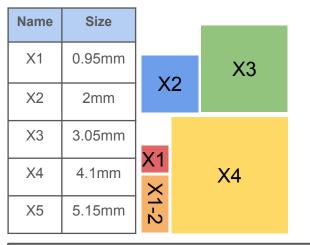
#### **Conclusion:**

Composability favors small dies, HPC favors large dies. No optimal chiplet size so we need to support multiple sizes.

Lib size (n)	5	10	5	10
Fabric Area (A)	100 mm <sup>2</sup>	100mm <sup>2</sup>	858mm <sup>2</sup>	858mm <sup>2</sup>
Chiplet (C)		Composabil	ity (n^(A/C))	
1	7.89E+69	1.00E+100	#NUM!	#NUM!
4	2.98E+17	1.00E+25	3.80E+149	1.00E+214
9	4.88E+07	1.00E+11	2.52E+66	1.00E+95
16	1.56E+04	1.00E+06	1.11E+37	1.00E+53
25	625	10,000	5.82E+23	1.00E+34
36	25	100	1.19E+16	1.00E+23
49	25	100	7.63E+11	1.00E+17
64	5	10	1.22E+09	1.00E+13
81	5	10	9.77E+06	1.00E+10
100	5	10	3.91E+05	1.00E+08

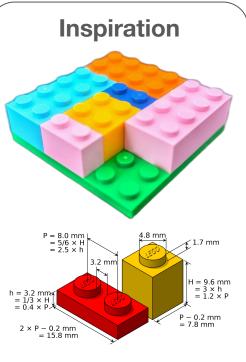
Composability ("solution diversity") achieved via small dies and large substrates.

### A4: Standardized Discretized Chiplet Grids

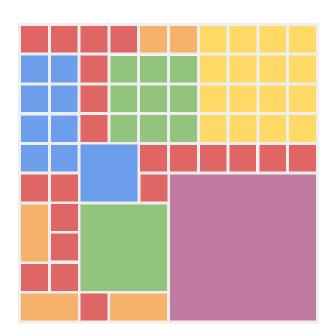


#### **Key Considerations:**

- Cost/density of 18A/N3 silicon
- 100Mtr/mm^2
- 100um safe chiplet spacing
- Minimum handling size
- Composability
- "Forever standard"



Lego Brick Standard Unchanged Since 1958!



Fixed Forever Chip Grid!

## Q&A5: 3D Chiplet Interconnect Pitch

	BGA package		DRAM dia Microbump Dia underfit DRAM dia Tay Dia underfit Dia underfit Dia underfit Dia underfit	(2 µm) Misalignment (<1 µm) Cu pillar (5 µm) (1.5 µm) Si JiF (2 µm) PECVD oxide (10 µm) Si JiF	
Pitch	150um	110um	45um	10um	5um
Pins/mm^2	44	82	493	10,000	40,000
Interface	Solder Ball	Cu+SnAg	Cu+SnAg	Cu	Cu
Assembly	Reflow	Reflow	Reflow	TCB	Hybrid
Cost	Low	Low	Medium	High	High
Tech Risk	Low	Low	Medium	High	High
REF	OSAT	OSAT	НВМ	UCLA	AMD

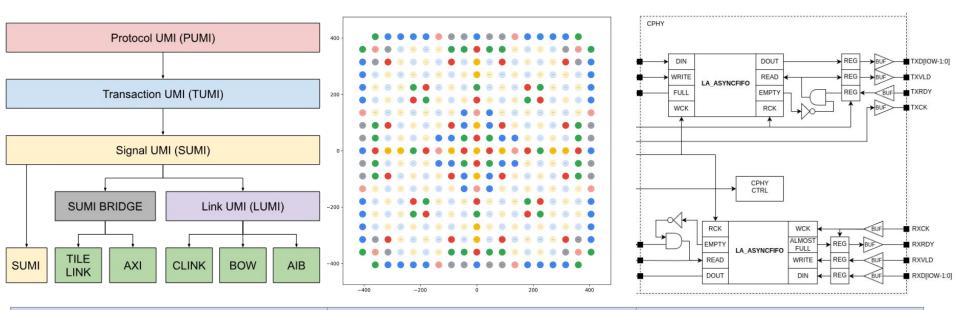
No right answer, but many wrong answers...

# Q6: 3D Chiplet Standard

	AIB	BOW	UCIE
Adoption Rate	Abandoned?	Unclear	High
Electrical standard	Yes	Yes	Yes
Footprint standard	No	No	No
Protocol Standard	No	No	No
3D Standard	No	No	Yes
Symmetrical	No	No	No
Suitable as AXI replacement	No	No	No

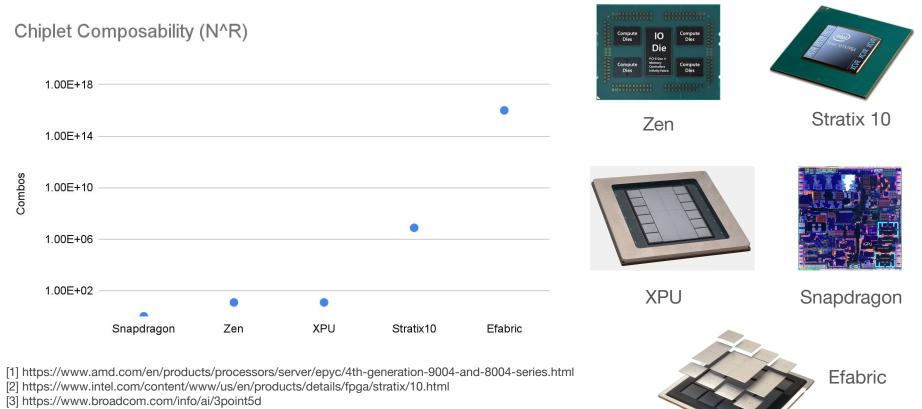
Existing chiplet standards don't support composability.

## A6: A full stack 3D chiplet standard



UMI Protocol	EBRICK Footprint	CLINK Electrical
Memory mapped packets	64b datapath	Source synchronous
Latency Insensitive	Rotational Symmetry	8b - 1024b
github.com/zeroasiccorp/umi	Analog, multi-power, passthrough	0.04mm2 in ASAP7 (512b)

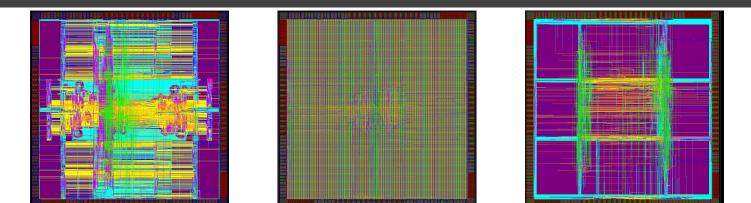
### **EFABRIC**: Composability Comparison



[4] https://chipsandcheese.com/p/inside-the-snapdragon-855s-igpu

[5] Zero ASIC, N = 10 (library size), R = 16 (number of sockets)

## **EBRICK**: Composable Chiplet Prototypes



	GOTLAND	MAUI	KODIAK
PROCESS	12nm	12nm	12nm
STANDARD	EBRICK_2x2	EBRICK_2x2	EBRICK_2x2
TYPE	CPU	FPGA	MEMORY
SIZE	2 x 2 mm	2 x 2 mm	2 x 2 mm
METRIC	Quad Core RV64GC CPU	(now 2K LUTs/mm^2)*	3MB
DESIGNERS	2	2	2
WALL TIME	< 4 weeks	< 8 weeks	< 8 weeks
RUN TIME	< 24hrs	< 24hrs	< 24hrs

# Switchboard: Chiplet Design Abstraction

- Heterogeneous simulation framework
- Latency insensitive protocol (ready/valid)
- Fast shared memory queues
- Supports RTL, FPGAs (HIL), SW models)
- UMI implementation
- Python bindings
- <u>10x faster than commercial emulators</u>
- <u>1000X</u> build time improvement over Verilator
- Deployed in AWS
  - 0.2us host latency
  - 4us host-fpga latency
- Source: github.com/zeroasiccorp/switchboard
- Demo: <u>zeroasic.com/emulation</u>

S.Herbst, et al, Switchboard: An Open-Source Framework for Modular

Simulation of Large Hardware Systems, arXiv preprint

arXiv:2407.20537, Jul 2024

Build a simulator for each, exposing latency-insensitive channels

Block A

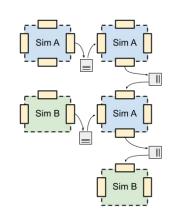
Block B

Sim A

Sim B

Connect simulator instances through shared-memory queues





#### TIMING BREAKDOWN FOR THE MILLION-CORE SIMULATION

Name	Time	Percentage
Launch 250 ECS tasks	2m 30s	23%
Wait for ECS tasks to boot	1m 20s	12%
Run simulation	7m 4s	65%
Total	10m 54s	100%

### **Emulator:** Chiplet Digital Twin Demo

**CPU** info

\$ cat /proc/cpuinfo

#### **Removing Barriers:**

- No EDA licensing
- No IP licensing
- No code
- No layout
- No mask costs
- No fabrication
- No installation

#### **New Capabilities:**

- White box validation
- Real time emulation

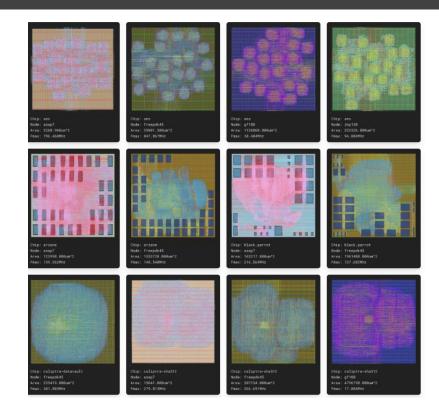
#### zerlo Emulation Demos Step 1: Select a demo ZA2011 zerlo Linux FPGA SDK Manycore efabric Custom Chiplet-Based SoC Fonturos Step 2: Select Components 4 RISC-V CPU cores Drag the iobricks and ebricks you want onto the eFabric canvas. 256 KB L1 Cache 1 MB L2 Cache This demo supports a maximum of four jobricks and four ebricks and 1 DDR PHY must have at least one cpu, eth, and memif chiplet. 1 Ethernet PHY Suggest Lavout CLINK LICTE Step 3: Inspect Datasheet Review the features of the chip you just designed. (c) 2025 by Zero ASIC Corporation Step 4: Emulate ebricks Press the "Emulate" button to launch an EPGA based emulation of the new chip. iobricks Step 5: Test Status: Please login to run emulations. Interact with the Terminal window to verify that the machine configured in Step 1 performed as expected. The terminal runs a Yocto-generated Emulate Clear version of Linux with a minimal set of packages installed. Here are some examples to get you started.

Output

#### https://emulation.zeroasic.com/emulation

Logs will appear when instance is running.

### SiliconCompiler: Automated chiplet compilation



\$ pip install siliconcompiler
\$ sc heartbeat.v -remote

import siliconcompiler
<pre>chip = siliconcompiler.Chip('heartbeat')</pre>
chip.load_target('skywater130_demo')
chip.input('heartbeat.v')
chip.clock('clk', period=10)
chip.set('option','remote', True)
chip.run()
chip.summary()
chip.show()

#### **PDKs:** GF12LP, GF22FDX, SKY130, GF180 **Tools:** Yosys, Openroad, VPR, Verilator, Icarus, Xyce, GHDL, Slang, Klayout, Cadence, Synopsys, Siemens, and many more

https://github.com/siliconcompiler

A. Olofsson, et al. "Invited: A Distributed Approach to Silicon Compilation", DAC 2022,

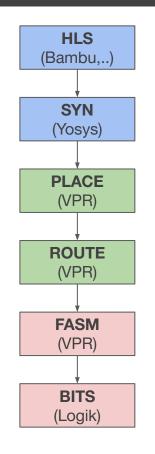
### Platypus: Because we need a "RISC-V for FPGAs"

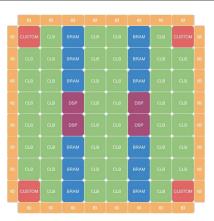
#### Zero ASIC · Mar 18, 2025

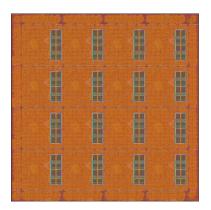
#### Zero ASIC launches world's first open standard eFPGA product

Cambridge, MA – March 18, 2025 – Zero ASIC, a U.S. semiconductor startup on a mission to democratize silicon, today announced Platypus<sup>TM</sup>, the world's first open standard eFPGA product. Platypus addresses a long standing critical issue of FPGA obsolescence and vendor lock that has put critical infrastructure at risk.

- 100% open standardized FPGA architecture
- 100% open source FPGA bitstream format
- 100% open source FPGA development tools
- 2K LUTs in 1mm<sup>2</sup>
- Support for BRAM and DSPs
- GF12LP process node (other nodes in development)
- OpenRoad based PNR implementation
- <u>Will become a standardized chiplet!!</u>
- <u>https://github.com/siliconcompiler/logik</u>
- <u>https://github.com/siliconcompiler/logiklib/releases</u>



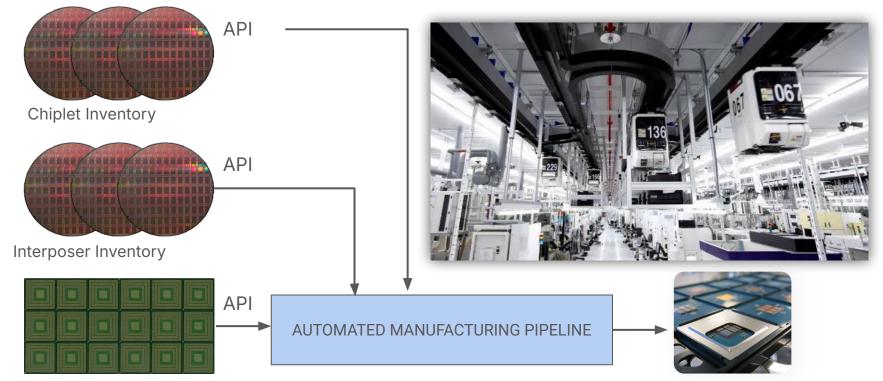




# zer | o |

# Predicting The Future of Chiplets

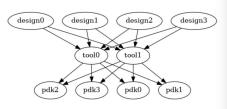
#### Lights Out Chiplet Assembly



Substrate Inventory

Standardized automation is the only way to fix the broken economics of low-volume high-mix manufacturing

#### New Era of Mechanical Configurability



**100%** Automated Silicon Compilers





**100%** Automated System-In-Package



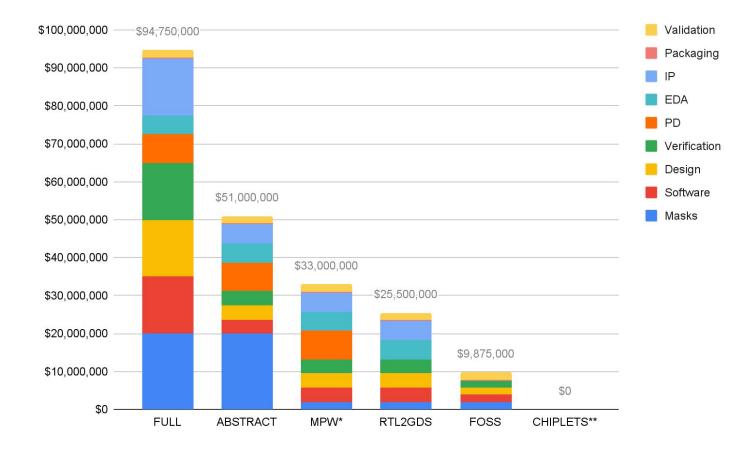
Modular Device Library

**100%** Automated Robotic Reconfiguration

New Silicon	90 days
New Device	1 day
New Supercomputer	1 day

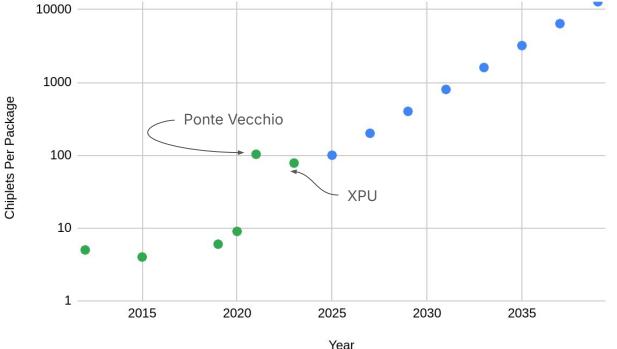


#### One Day We Will Spin SiPs at a cost of \$1K in 24Hrs



43

#### Olofsson's Chiplet Roadmap



"The number of chiplets in a package will double every two years."

44

#### Conclusion

#### "The most reliable way to predict the future is to create it."

– Abe Lincoln

