Verification Futures

July 2025

SRAM: Dead or alive?

© Chipletti 2025

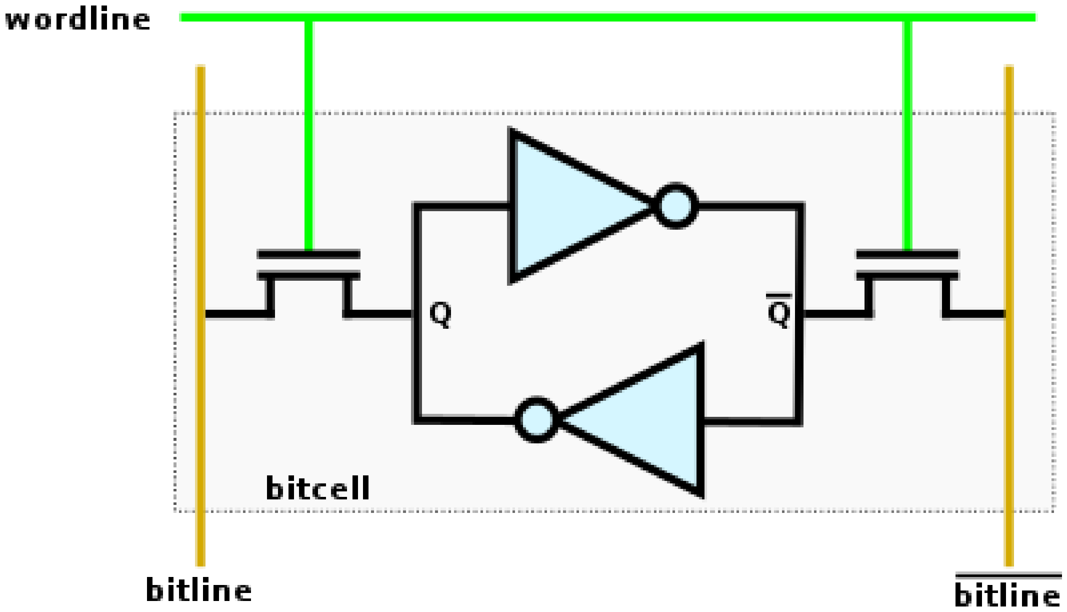
What is an SRAM?

Static

Volatile

Six transistors per bit cell

A 6T SRAM bit cell layout.



Source: https://en.wikichip.org/wiki/static_random-access_memory - google_vignette

© Chipletti 2025

SRAM: History and timeline

1963

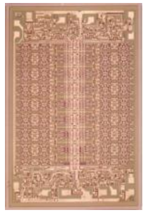
Robert H. Norman at Fairchild Semiconductor Bipolar SRAM

1964

John Schmidt at Fairchild Semiconductor MOS-SRAM

First commercial SRAM chip

1969



Intel released The 3101 SRAM chip

64bit capacity!

1971

Intel 4004: the world's first commercially available microprocessor

1980s

SRAM integrated into processor dies – transition to multi-level cache hierarchies

1990s

Embedded SRAMs become common in ASICs, FPGAs and SoCs

2000s

With the rise of mobile devices, low power SRAMs become essential

2010s-Present

FinFET based SRAM cells. SRAM faces scaling challenges.

Why is it so ubiquitous?

1. Best read/write latencies
2. Fully deterministic
3. Ease of integration with logic
    1. SRAM memory compiler for every process node
    2. Generate any shape and the size you want for your design

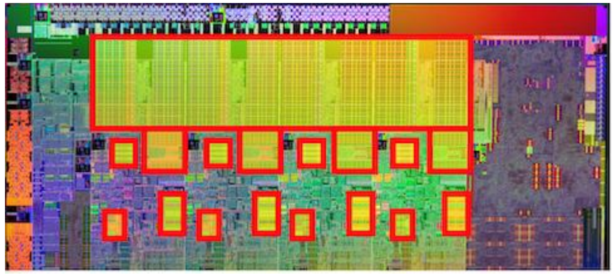
Memory	Read latency	Density	
SRAM	0.5-2ns	~60-120F2	Available on every process node as part of logic foundries SRAM compilers.
DRAM	10-15ns	~6-8F2	Special foundries that manufacture DRAM chips - currently cannot be integrated together with logic which adds additional off-chip latencies in the order of 50-100ns.
MRAM	10-35ns	~20-40F2	Very limited foundries and process nodes.
ReRAM	10-100ns	~4F2	Limited availability: 22nm FDSOI, 130nm etc

F^2 represents feature size relative to process node

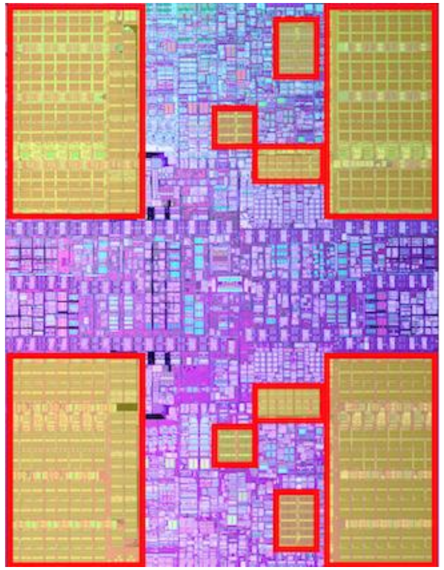
© Chipletti 2025

Example designs

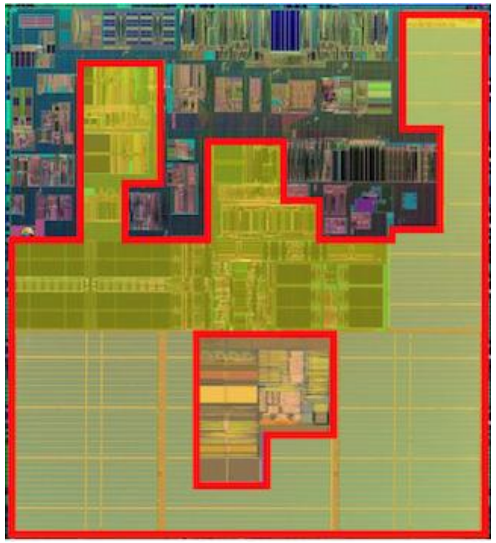
Die shots of microprocessors using a large quantity of SRAMs. Some of the smaller macros not shown.



Sandy Bridge
[32nm]



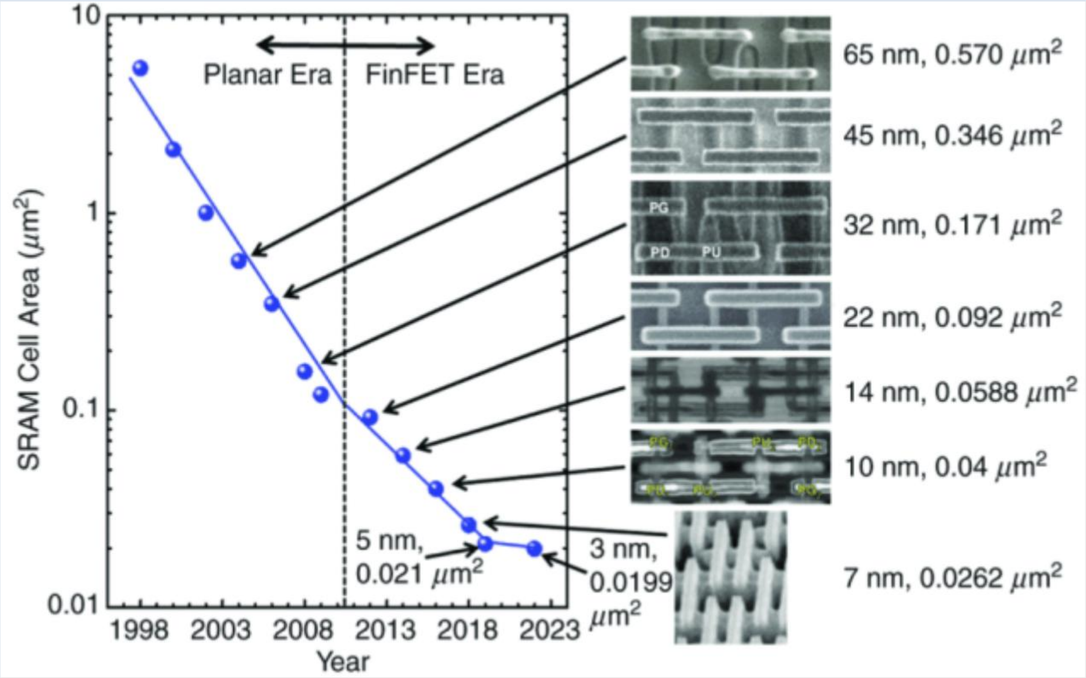
Power6
[65nm]



McKinley
[180nm]

Source: [https://en.wikichip.org/wiki/static_random-access_memory - google_vignette](https://en.wikichip.org/wiki/static_random-access_memory#google_vignette)

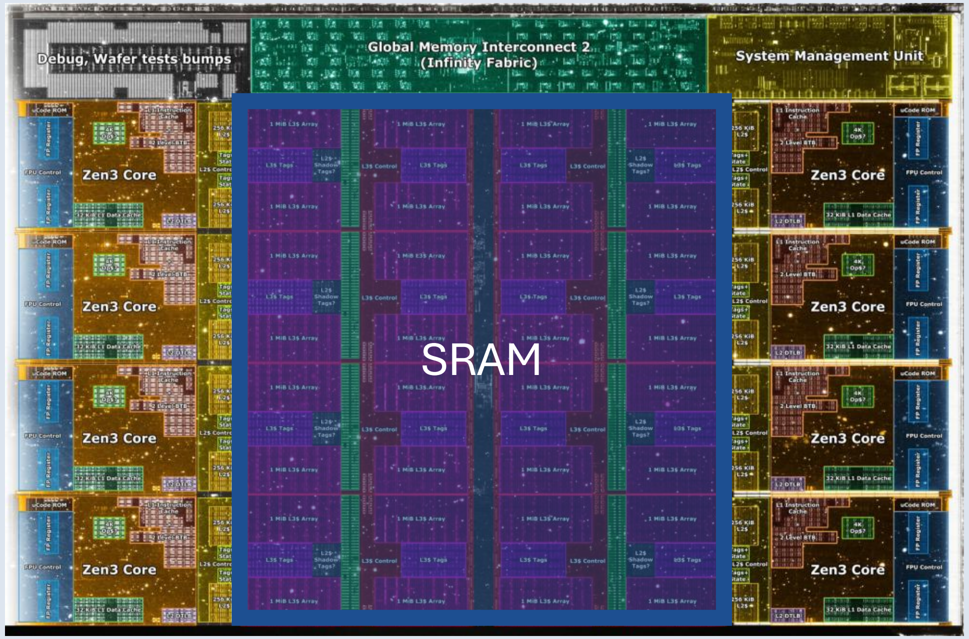
However!



Source: [H. Chang et al.]

SRAM scaling has *significantly slowed* down below 5nm in density and performance

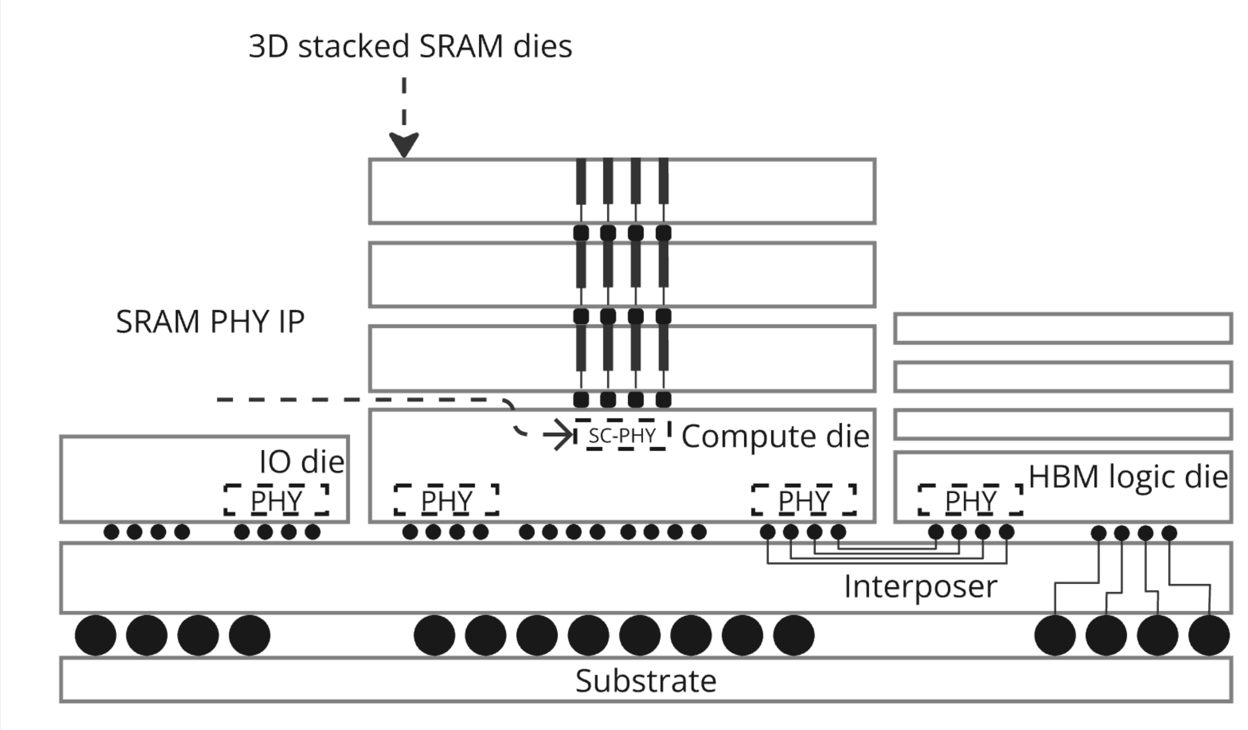
SRAM uses large chip area which could be used for logic



Source: [AMD Ryzen 5000 Zen 3 Desktop CPU Gets First High-Res Infrared Die Shot, Vermeer Fully Detailed]

Solution?

1. When large arrays of SRAMs are needed, put them on a separate die
2. Vertically stack them on top of compute
   1. Why? Connections via silicon interposers require SerDes – latencies in the range of 50-100ns



Why now?

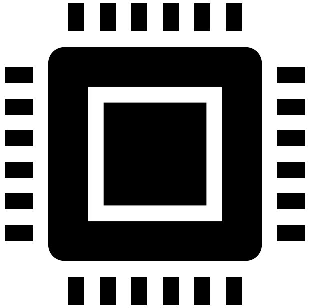
1. Designs are hitting reticle limits
2. Advanced 3D integration with very high yields
    1. Bumpless sub 10um pitch IOs in mass production for several years
3. Wafer costs disparity
4. Reliability/scaling issues

Let's dive in a bit more.

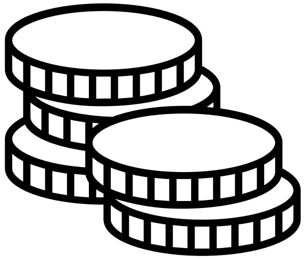
© Chipletti 2025

Break-even

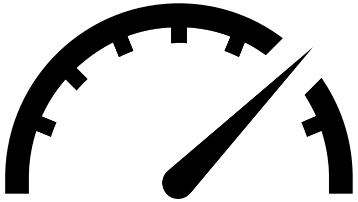
3 break-even factors to consider for switching to 3D stacked SRAMs.



PHY area



Costs



Performance

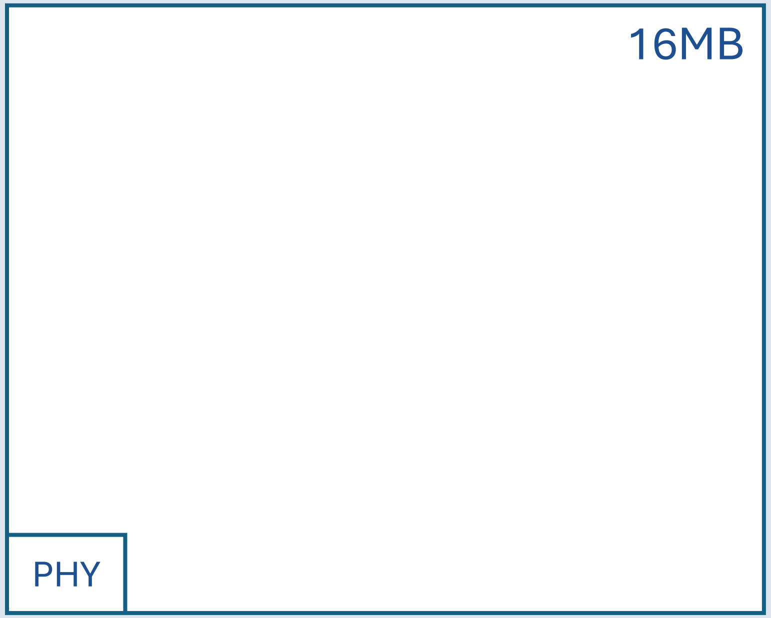
Break-even: PHY area

PHY area > SRAM macro

PHY area depends on the **number of IOs**, **process node**, the **type of bonding** and the **pitch** between the IO cells.

1MB

Hybrid bond, 9um pitch, 300 IO cells at 7nm



© Chipletti 2025

Break-even: Costs

$8.1	$7.5	**$5.3**	$7.7	$7.7	$11.3
28nm	16nm	**7nm**	5nm	3nm	2nm

Volume production costs for a 128MB SRAM chiplet

Doesn’t factor the exorbitant NRE costs!

Break-even for costs is very subjective.

© Chipletti 2025

Break-even: Performance

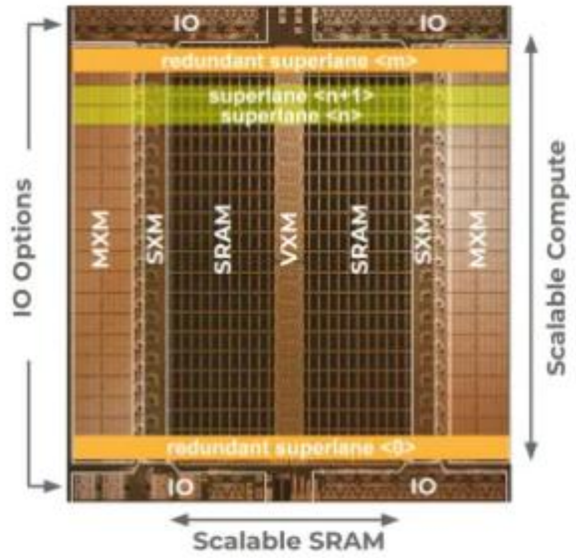
Need to measure system level performance. Factors to consider include **Fmax** improvements due to smaller compute die, **lower latencies** within the compute die, **higher latencies** to off chip SRAMs and **higher SRAM capacity**.

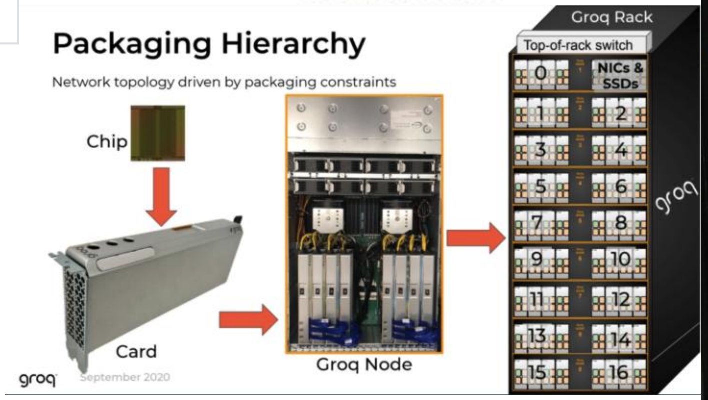
© Chipletti 2025

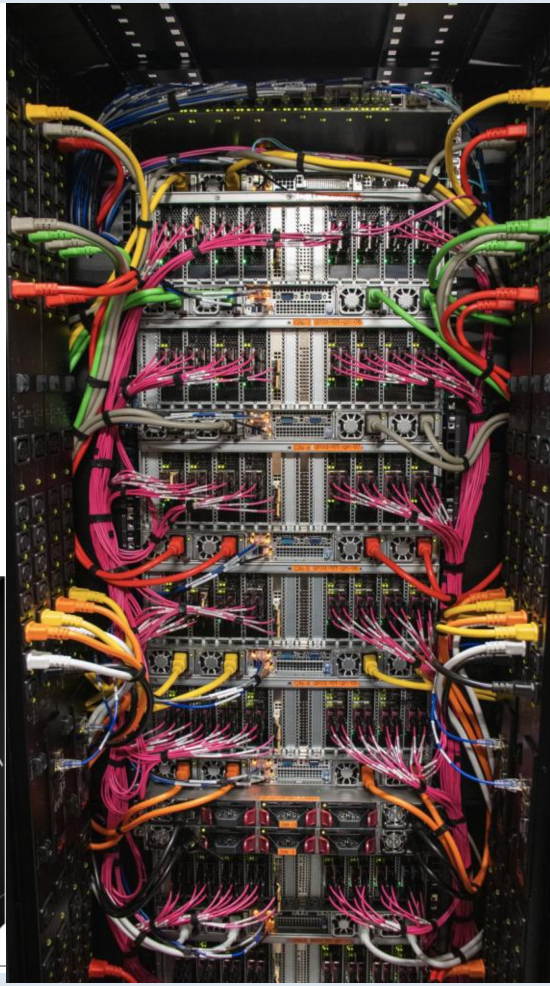
Example: AI inference

AI Model size	Memory		#Chips	Optical cables	
	Size	Stack		#Number	Miles
7B	16GB	2D	70	315	1.9
		3D	16	70	0.4
10T	20TB	2D	95,325	428962	2659
		3D	16,850	75825	470

A memory stacked approach can save 4-10x number of chip modules or optical cables for the same AI model.

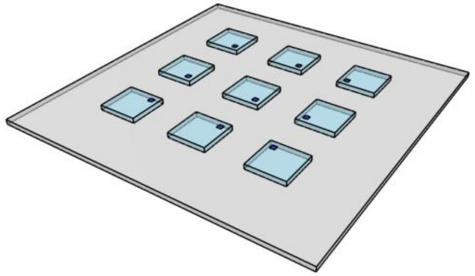


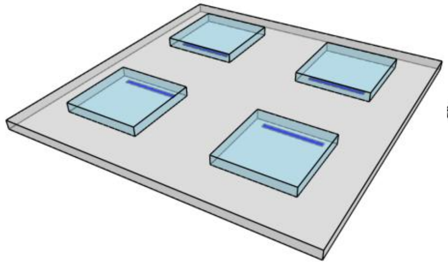


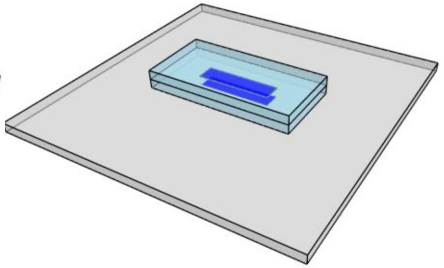


SRAM chiplets: configurations

Products	Small	Medium	Large
Memory	16MB	64MB	128MB
Chiplet Area	4.5mm^2	16.8mm^2	33.7mm^2
I/Os	256	512	1024
Macro area	0.02mm^2	0.05mm^2	0.05mm^2
Bandwidth	128Gb/s	256Gb/s	512Gb/s
Process	7nm		
Bond pitch	9um		







© Chipletti 2025

Thank you

kauser.johar@chipletti.com

© Chipletti 2025