Case Study

Brian Piercy
bjpcjp@gmail.com | 512-577-5269 |
bjpcjp.github.io | @brianpiercy

1. Assumptions
2. Philosophy
3. Ideas
4. Financials

Opening Assumptions

- $20B market cap
  - ~40 P/E ratio ⇒ $**0.5B** earnings
- Margin & revenue guesswork:
  - 50% gross
  - 33% EBITDA
  - 25% operating (pretax)
  - 20% net ⇒ $**2.5B** revenue
- My portfolio:
  - 10% of company ~$**250M**/yr
  - 50% GM target (ICs).
- Product “A”: In production.
  - 75 week dev cycle
  - New PDK, pkgs, substrate
  - *Some* outsourced IP.
- Product “A+1”: in development.
  - 54 week dev cycle. 70% complete.
  - Goal: 10% mfg cost reduction.
  - Fixed R&D headcount. Replaces “A”.
- Competitor announces “B”.
  - 30% performance advantage over “A”

Assumptions: Customers / Markets

- TAM:
  - $750M/year (75M units; blended ASP ~$10.0)
  - Mature market: (~10%/year ASP drop; ~5% unit volume growth.)
- Growth plan (as “performance leader”*):
  - Baseline: maintain ~1/3 share @ near-market ASP:
  - ~25M units/year, ~$225-250M (year 0)
- Customer profiles:
  - 5 Tier-1 customers (40/20/10/10/10% of shipments)
  - Distribution + emerging use cases (10% of shipments)

Assumptions: Timelines

- Product Development (A ⇒ A+1)
  - execution approval ==> tapeout 1.0: 39 weeks ⇒ 30 weeks*
  - tapeout v1.0 ⇒ tapeout v1.1 (m2 fix): 13 weeks ⇒ 10 weeks
  - tapeout v1.1 ⇒ alpha samples: 8 weeks ⇒ 7 weeks
  - alpha samples ⇒ JEDEC commercial qual (3 lots) 15 weeks ⇒ 7 weeks
  - total: 75 weeks ⇒ 54 weeks
- Tier-1 Customer Acceptance:
  - Samples to supplier approval (multiple applications): 6 months
  - Supplier approval ⇒ 100% ramp (or volume crossover): 9 months
  - Total market window: 5 years

* 70% complete ~ 9 weeks remaining to T/O 1.0

Product A+1: R&D Spend

HEADCOUNTS, NREs, TIMELINES

					0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
HW IP:	**FORM**	**CMPLXTY**	**REUSE?**	**#?**																																			
	HARD	HIGH	HIGH	1	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	RTL	VERYHIGH	HIGH	1	4	4	4	4	4	4	4	4	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	MEM	COMPILED	LOW	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	MEM	CUSTOM	LOW	1	4	4	4	4	4	4	4	4	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	I/OS	MID	HIGH	6	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	MXD SGNL	MID	MID	2	4	4	4	4	4	4	4	4	4	1	1	1	1	1	1	1	1	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	... etc...																																						
SW IP:	DRIVERS	MID	HIGH		2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	2	2	2				1	1	1									
	MODELING	HIGH	MID		6	6	6	6	6	6	6	6	1	1	1	2	1	1	1	1	1	2	2	2				1	1	1									
	DEVTOOLS	HIGH	HIGH		2	2	2	2	2	2	2	2	4	4	2	2	1	1	1	1	1	2	2	2				1	1	1									
SOC INTEGRATION:	ARCHITECTURE				3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2												
	STA, E2E, BIST, CLOCK TREE, ...				6	6	6	6	6	6	6	6	1	1	1	1	1	1	1	1	1	2	2	2				1	1	1									
	FLOORPLAN, DRC/LVS, POWER, ...				2	2	2	2	2	2	2	2	4	4	2	2	1	1	1	1	1	2	2	2				1	1	1									
	TAPEOUT (1.0, 1.1), FAB, PKG, FT												X	X	X	X	X	X	X	X	X			X	X	X	X												
PKG, TEST, BI, QA:	CZ, ERRATA LIST, ALPHA ES:				4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	4	4	3	3	3	3	4	4	4	4	2	2	2	2	2	2	2	
	JEDEC 3-LOT QUAL:																														X	X	X	X	X	X	X	X	

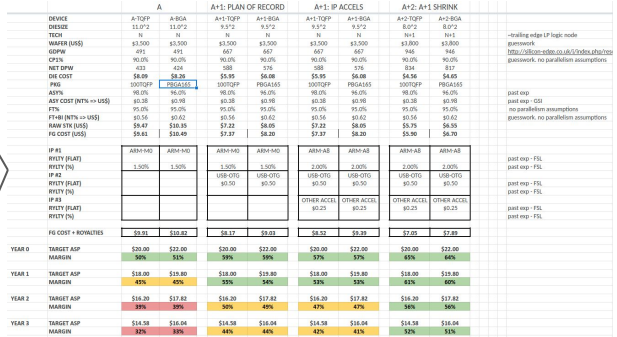
DEV TEAM FTE COST ASSUMPTIONS (FULLY LOADED):

	$FTE/WEEK	%WEIGHT
HIGH	$3,500	20%
MID	$2,200	30%
LOW	$1,400	50%

MAN-WEEKS SPENT	MAN-WEEKS REMAINING	MASKSET + TO 1.1 (3 LAYERS	ENGRNG WAFERS	PROBE HW	B/I HW	TOTAL
1032	955					
$2,125,920	$1,967,300	$900,000	$100,800	$15,000	$10,000	$5,119,020

IC Product Costs: “A”, “A+1”, “A+2”

		A+1: PLAN OF RECORD		A+1: IP ACCELS		
	DEVICE	A+1-TQFP	A+1-BGA	A+1-TQFP	A+1-BGA	
	DIESIZE	9.5^2	9.5^2	9.5^2	9.5^2	
	TECH	N	N	N	N	~trailing edge LP logic node
	WAFER (US$)	$3,500	$3,500	$3,500	$3,500	guesswork
	GDPW	667	667	667	667	http://silicon-edge.co.uk/j/index.php/res
	CP1%	90.0%	90.0%	90.0%	90.0%	guesswork. no parallelism assumptions
	NET DPW	588	576	588	576	
	DIE COST	**$5.95**	**$6.08**	**$5.95**	**$6.08**	
	PKG	100TQFP	PBGA165	100TQFP	PBGA165	
	ASY%	98.0%	96.0%	98.0%	96.0%	past exp
	ASY COST (NT% => US$)	$0.38	$0.98	$0.38	$0.98	past exp - GSI
	FT%	95.0%	95.0%	95.0%	95.0%	no parallelism assumptions
	FT+BI (NT% => US$)	$0.56	$0.62	$0.56	$0.62	guesswork. no parallelism assumptions
	RAW STK (US$)	**$7.22**	**$8.05**	**$7.22**	**$8.05**	
	FG COST (US$)	**$7.37**	**$8.20**	**$7.37**	**$8.20**	
	IP #1	ARM-M0	ARM-M0	ARM-A8	ARM-A8	
	RYLTY (FLAT)					past exp - FSL
	RYLTY (%)	1.50%	1.50%	2.00%	2.00%	past exp - FSL
	IP #2	USB-OTG	USB-OTG	USB-OTG	USB-OTG	
	RYLTY (FLAT)	$0.50	$0.50	$0.50	$0.50	past exp - FSL
	RYLTY (%)					past exp - FSL
	IP #3			OTHER ACCEL	OTHER ACCEL	
	RYLTY (FLAT)			$0.25	$0.25	past exp - FSL
	RYLTY (%)					past exp - FSL
	FG COST + ROYALTIES	$8.17	$9.03	$8.52	$9.39	
YEAR 0	TARGET ASP	$20.00	$22.00	$20.00	$22.00	
	MARGIN	**59%**	**59%**	**57%**	**57%**	
YEAR 1	TARGET ASP	$18.00	$19.80	$18.00	$19.80	
	MARGIN	**55%**	**54%**	**53%**	**53%**	
YEAR 2	TARGET ASP	$16.20	$17.82	$16.20	$17.82	
	MARGIN	**50%**	**49%**	**47%**	**47%**	
YEAR 3	TARGET ASP	$14.58	$16.04	$14.58	$16.04	
	MARGIN	**44%**	**44%**	**42%**	**41%**	



Scenarios:

* (TQFP & BGA opts for each die rev)
* “A”: existing device
* “A+1”: POR
* “A+1”: POR + selected 3rd party IP
* “A+2”: Traditional die shrink

A+1 Catchup: R&D Toolbox

Silicon Improvements

1. Process & Physical Design Tweaks
    a. Tighter process tolerances (ΔVt,ΔLeff?)
    b. Custom vs compiled memories
    c. HS vs LP std cells
    d. Clock tree & power grid optimization
    e. Upgraded I/O buffer strengths
2. Compute Architecture
    a. Optimized L1/L2/L3
    b. Improved branch prediction
    c. Simplified instruction decode

Compiler / Application SW Improvements

1. Compiler Tweaks
    a. Loop unrolling
    b. Matrix sub-blocking
    c. Pipelines: dynamic scheduling
2. Application-specific improvements
    a. (ML) weight sharing/pruning, reduced-precision math
    b. (HPC) optimized linear algebra (ex: BLAS, ...)
    c. (OS) boot time improvements

A+1 Catchup: Risk Management

- Assessment: “B”’s 30% performance lead comes from
  - 1) Advanced Si tech node
  - 2) Completely new architecture
- No Hail Marys..
  - 30% gain from one change = unlikely.
  - Use incremental approaches.

	simulated improvement	%success	weighted %success
Idea 1	1.0%	100%	1%
Idea 2	1.0%	100%	1%
Idea 3	2.5%	75%	2%
Idea 4	2.5%	75%	2%
Idea 5	2.5%	75%	2%
Idea 6	2.5%	50%	1%
Idea 7	5.0%	75%	4%
Idea 8	5.0%	50%	3%
Idea 9	5.0%	50%	3%
Idea 10	10.0%	75%	8%
Idea 11	10.0%	50%	5%
	47.0%		30%

Assessment: chances of closing 30% gap, given no change to POR, is low.

Roadmap Options

Hypothesis (assume equal % of being true)	Proposal	Tech Risk	Cost (Buy/Build)	"Moat"?	"Crowbar"?	Priority
Targeted performance #1.	**Own the devtools:** libraries & compilers	Medium	Low	Yes	Micro-architecture knowledge	1
	ASSPs A+2, A+3, ...	Medium	High	Yes	Use case knowledge (eg Google TPU)	3
Performance #1, but stagnant market.	**Moonshot idea #1** Core competence ==> adjacent market	High	Medium	Yes	Micro-architecture knowledge	2
Time-to-market >> Performance	**Finish A+1.** (minimal chg to POR, hold schedule)	Low	Low	No	Unknown	2*
	Vertical integration A+1,2 ==> accelerator cards	Medium	Medium	Maybe	BOM %share	3
	Very low cost A+2 Next-gen PDK, stripped-down IP BOM	Medium	High	Maybe	No	3
Performance is the driver.	**Finish A+1.** (close 30% gap; slip schedule)	Low	Low	No	Unknown	2*
	Buy "B"?	Unknown	High	Unknown	Unknown	4
Performance #1: customer lacks expertise / resources	**Own the interfaces/API** Memory controller IP; NVlink, ...	Medium	Medium	Medium	IP ==> Razor ICs ==> Blades	3
	Vertical integration A+1,2 ==> accelerator cards	Low	Medium	Maybe	BOM %share	3
	Moonshot idea #2 Chiplets	High	Medium	Yes	Unknown	4
Customers' customer is the real target.	**Own the devtools** IDE, debug, systems mgmt, profiler	Medium	Low	Maybe	Use case knowledge (eg Google TPU)	2
	Vertical integration A+1,2 ==> accelerator cards	Medium	Medium	Maybe	BOM %share	3

Estimated R&D Costs: Accelerator Card

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
HW DESIGN:	BOARD LAYOUT	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2																		
	SIMULATION	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1										
	VERIFICATION	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1										
SW DESIGN:	DRIVERS	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	SYS MONITORS, DEVTOOLS	2	2	2	2	2	2	2	2	1	1	1	2	1	1	1	1	1	2	2	2	2	2	2	2	2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
ONGOING / AE SUPPORT:															1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

TIMELINE:																																				
	PCB FABRICATE	X	X	X	X	X	X																													
	ASSEMBLY + BRINGUP (10#)														X	X	X																			
	CZ																	X	X	X																
	VALIDATION (10#)																				X	X	X	X	X	X										

DEV TEAM FTE COST ASSUMPTIONS (FULLY LOADED):

	$FTE/WEEK	%WEIGHT
HIGH	$3,500	20%
MID	$2,200	30%
LOW	$1,400	50%

NRE (PCB, ?)	MAN-WEEKS	PCB + ICS, 10#	SW LICENSING	VALIDATION HARDWARE	PLACEHOLDER	TOTAL
0	199			ovens? rack mounting?		199
$5,000	$409,940	$15,000	$2,500	$30,000	$0	$462,440

Roadmap Strawman

		2020				2021				2022				2023			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
(ALL)	DEVTOOLS	DEF	DEV	REL "A+1"	DEV	REL "A"	DEF	DEV	DEV	REL "MOON"	DEF	DEV	REL "A+2"	TBD	TBD	TBD	TBD
"A"	IC	PROD						PHASE OUT	PHASE OUT	PHASE OUT	PHASE OUT	(continued production for accelerator card demand)					
	CARD	DEF	DEF	DEV	ES	ES	RISK RAMP	RISK RAMP	PROD	>>>	>>>	>>>	>>>	>>>	>>>	>>>	>>>
"A+1"	IC	DEV	DEV	ES	RISK RAMP	RISK RAMP	PROD	>>>	>>>	>>>	>>>	>>>	>>>	>>>	>>>	>>>	>>>
	CARD			DEF	DEF	DEV	ES	ES	RISK RAMP	RISK RAMP	PROD	>>>	>>>	>>>	>>>	>>>	>>>
MOONSHOT1	IC			DEF	DEF	DEF	DEV	DEV	DEV	ES	ES	ES	RISK RAMP	RISK RAMP	PROD	>>>	>>>
"A+2"	IC								DEF	DEF	DEV	DEV	ES	ES	RISK RAMP	RISK RAMP	PROD
	CARD										DEF	DEF	DEV	DEV	ES	ES	RISK RAMP
MOONSHOT2	2.5D / CHIPLET											DEF	DEF	DEF	DEV	DEV	DEV

Cash Flow Strawman

		2020				2021				2022				2023			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
R&D	SW (K$)	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500	-$500
	ICS (K$)	-$1,500	-$1,500	-$3,000	-$1,750	-$1,500	-$1,500	-$1,500	-$1,500	-$3,000	-$1,500	-$1,500	-$3,000	-$1,500	-$1,500	-$1,500	-$1,500
	BOARDS (K$)	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250	-$250
A	KUNITS	7000	6000	5000	4000	2000	500	50	0	0	0	0	0	0	0	0	0
	%SHARE LOSS TO	10%	15%	20%	25%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%
	ASP	$10	$10	$10	$10	$9	$9	$9	$9	$8	$8	$8	$8	$7	$7	$7	$7
A CARD	KUNITS	0	0	0	10	50	100	500	250	100	10	1	0	0	0	0	0
	ASP	$125	$125	$125	$125	$113	$113	$113	$113	$101	$101	$101	$101	$91	$91	$91	$91
A+1	KUNITS	0	0	50	250	500	4000	6000	6000	6000	5000	4000	3000	1500	750	100	50
	ASP	$10	$10	$10	$10	$9	$9	$9	$9	$8	$8	$8	$8	$7	$7	$7	$7
A+1 CARD	KUNITS	0	0	0	0	0	10	50	100	500	500	500	750	1000	500	250	50
	ASP	$125	$125	$125	$125	$113	$113	$113	$113	$101	$101	$101	$101	$91	$91	$91	$91
MOONSHOT	KUNITS	0	0	0	0	0	0	0	0	5	50	500	750	750	500	250	250
	ASP					$0	$0	$0	$0	$18	$18	$18	$18	$15	$15	$15	$15
MS CARD	KUNITS	0	0	0	0	0	0	0	0	5	25	100	250	500	1000	1500	2000
	ASP	$250	$250	$250	$250	$225	$225	$225	$225	$203	$203	$203	$203	$182	$182	$182	$182
A+2	KUNITS	0	0	0	0	0	0	0	0	0	0	0	5	50	500	750	750
	ASP	$10	$10	$10	$10	$9	$9	$9	$9	$8	$8	$8	$8	$7	$7	$7	$7
A+2 CARD	KUNITS	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	50
	ASP	$250	$250	$250	$250	$225	$225	$225	$225	$203	$203	$203	$203	$182	$182	$182	$182
TOTAL	KUNITS	7000	6000	5050	4260	2550	4610	6600	6350	6610	5585	5101	4755	3800	3251	2860	3150
	$REVENUE ($M)	$63	$51	$41	$34	$23	$52	$116	$93	$110	$98	$112	$164	$205	$245	$308	$388
	MFG MARGIN ($M)	$32	$26	$20	$17	$9	$22	$38	$34	$32	$28	$30	$35	$35	$39	$46	$57
	OP MARGIN ($M)	$29	$23	$17	$14	$7	$19	$36	$32	$28	$25	$28	$31	$32	$36	$43	$55

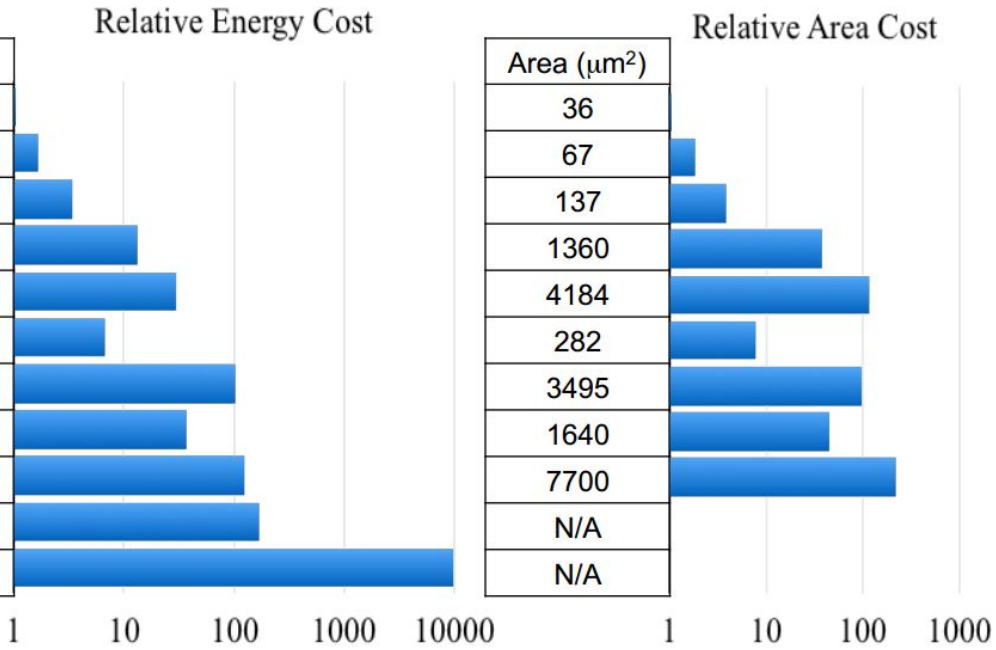
Backup Material

Sometimes, Performance can Outstrip a Customer's Expertise

- 3rd-gen QDR SRAMs (GSI Technology, t_clk ~ 650-800MHz)
- Beta customers: Sophisticated, Tier-1, ASIC-based platforms
- Follow-on customers: Interested, but unable to take advantage
  - Using FPGA EDA tools
    - "plug & play" memory solutions (drop-down menus, preselected memory IC callouts)
  - Lack of controller design expertise
  - Lack of PCB layout & Thermal tolerance experience.

Performance Gap Reduction Ideas: Reduced-Precision Math Ops are your Friends

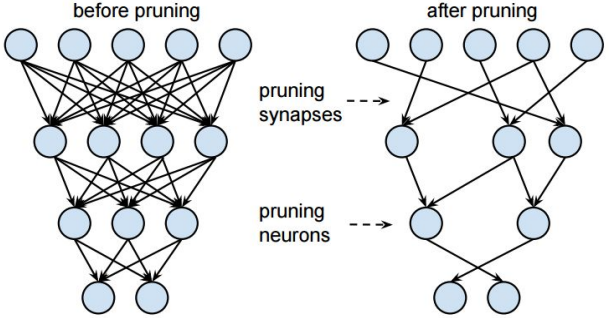
Operation:	Energy (pJ)	Area (μm²)
8b Add	0.03	36
16b Add	0.05	67
32b Add	0.1	137
16b FP Add	0.4	1360
32b FP Add	0.9	4184
8b Mult	0.2	282
32b Mult	3.1	3495
16b FP Mult	1.1	1640
32b FP Mult	3.7	7700
32b SRAM Read (8KB)	5	N/A
32b DRAM Read	640	N/A

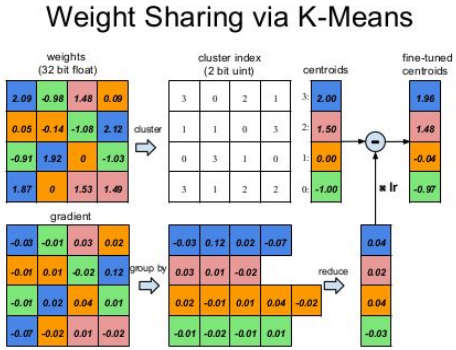


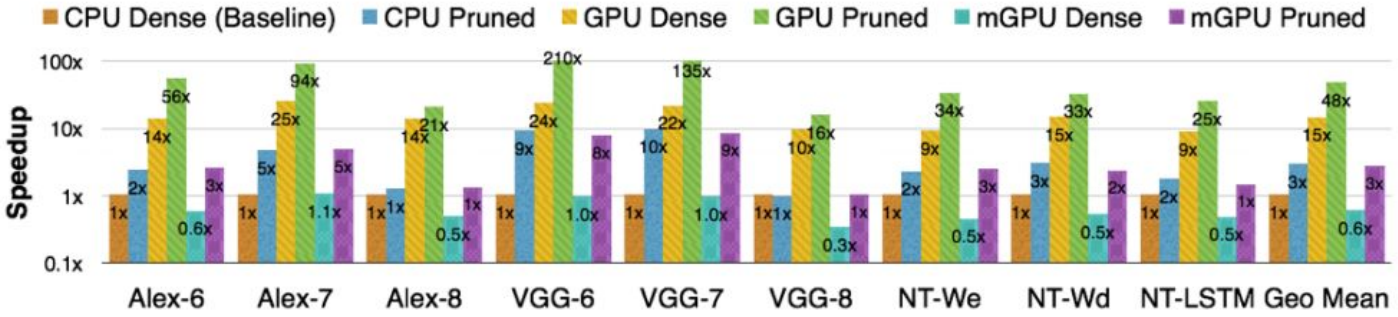
* "Computing's Energy Problem (and what we can do about it)", ISSCC 2014 (Horowitz)
* Area #s from Design Compiler @ TSMC 45nm

Performance Gap Reduction Ideas: Memory Pruning / Sharing

- 1GB model (250M weights) ⇒ 20-30MB (mobile memory footprint capable)
- Less memory B/W needed (esp. FC layers)

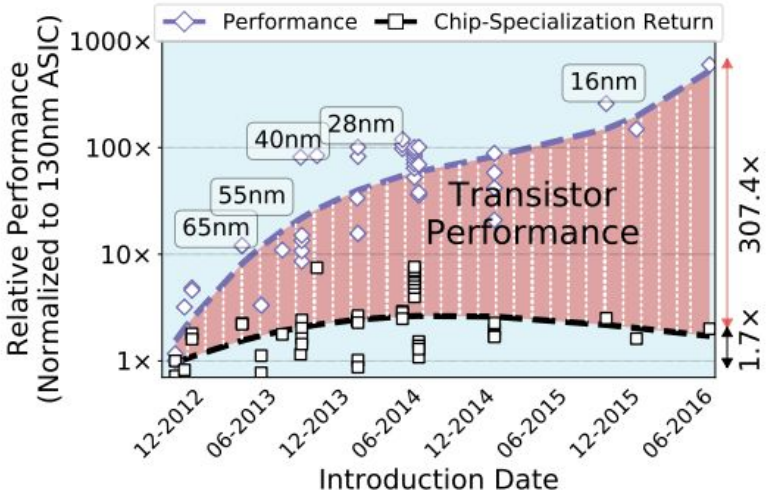






Principle: Transistor Gains >> Architecture Gains

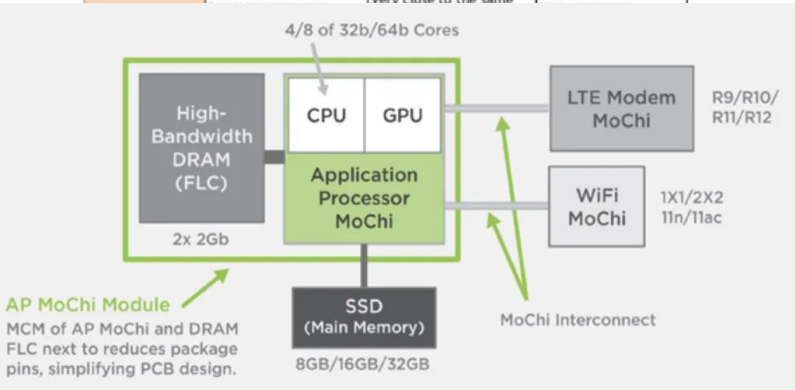
- Hardware accelerators have been a standard tool for increasing silicon value.
- #transistors budget is stagnating ⇒ limits ability to optimize accelerator returns (ie the “accelerator wall”.



http://parallel.princeton.edu/papers/wall-hpca19.pdf

Integration Option: Chiplets

	Monolithic SoC	Chiplet on Substrate	Discrete IC on PCB
Design Cost	Highest design cost 7nm >$200 million	Less costly than monolithic SoC	Least expensive
Design time	Longest design cycle 18+ months	Shorter design time 12 months faster to make derivative design	Shortest design time 6 months
Design risk	Highest risk, including missing future features, having to do a redesign	Lower risk, can change out chiplets to add or subtract features, redesigns are simpler	Lowest risk
Performance	Highest performance, but some functions might not scale so would use larger design rule for that portion of the chip, leads to inefficient use of resources	Good performance, can use the most appropriate process for the function needed in the chiplet	Lowest system performance, problems with timing, latency, more complex PCB
		Very close to the same	

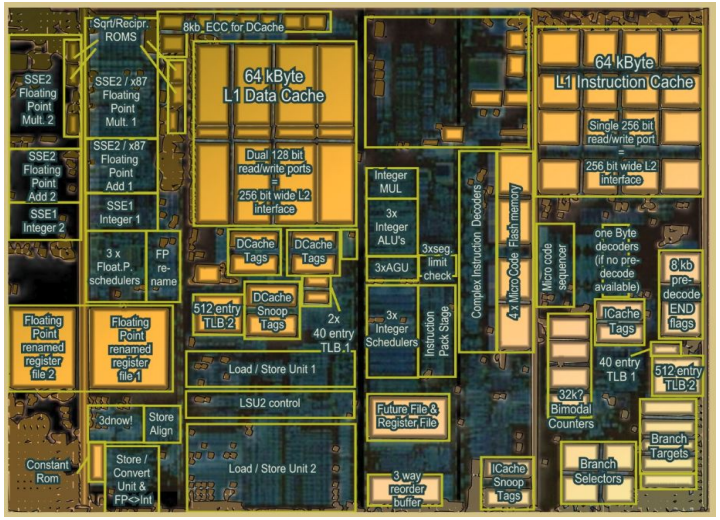


https://semiengineering.com/getting-serious-about-chiplets/

- Yield: need KGD agreement with component suppliers. (allowing escapes at wafer-level probe ⇒ vastly more expensive at chiplet level)
- Testability: 1) need to bring scan chains to outside world. 2) need well-defined inter-die signal stds

Product “A” Description (Example)

- Special-purpose compute engine
- Compiled SRAM: 40% area
- L1/L2 cache complex
- NVM: n/a
- Single power domain
- Single clock domain
- Accelerators: Hardened RTL: 10 blocks, ~20% area
- Std foundry CMOS, C040LP, 6LM?
- Timing, functional options ⇒ metal fuses
- TQFP & PBGA options



Way-too-ambitious example