Ascalon Flex Configuration

Date	Changes	Latest Status / Next Steps	Author
Apr 2, 2024	• Update page based on input from @Yasuo Ishii	• Use the current confluence page as a baseline for D8 Flex planning	James Ko
Oct 15, 2024	• Priority updated for LAC		James
Nov 12, 2024	• Updated format and requirements table to query New Feature with label “ascalon_flex”		Rae

Target release	Type // to add a target release date
Epic	Type /Jira to add Jira epics and issues
Document status	DRAFT
Document owner	@James Ko
Architect	@ architect
Engineering TPM	@ EngTPM
Product TPM	@ ProductTPM

🎯 Objective

Ascalon Flex provides customers with configuration options based on Ascalon D8 (and other variants). The target customer will have specific needs that are not met by the standard Ascalon IP offering. This product broadens the reachable market that Product IP portfolio can address.

This segment of the Ascalon portfolio will be known as “Variants”, and could be named as:

- D8 flexible
- D4 flexible
- D2 flexible

The intent of the document is to define the scope of features and requirements to describe Ascalon Flex as a product.

Reference Links

Overall Product Portfolio - Variants (not confirmed/committed)

🎨 Cluster Configuration Requirements for Ascalon

Configuration requirements:

1. Configuration requirements (to be aligned and be locked when product is committed)
   - MVP = A must for current generation
   - P1 = can be considered for future gen

Configuration	Value	Notes
CPU Per Cluster	• MVP: ~~1~~, 2, 4, 8 (default) • P1: 10, 14, 16, 1	

Private L1 Cache	• MVP: fixed
	  ◦ L1 I-Cache
	  ▪ 64k (default), 128k (samaritan)
	  ◦ L1 D-Cache
	  ▪ 256k only
	• P1: Flexible
		• Looking for feedback from architecture team if L1 config is useful
		• We may have more L1 cache options once team starts working on D4/D2 CPUs. Non-default / non-samaritan configs will need additional verification effort.

Shared L2 Cache (Cluster)	• MVP: 2M, 4M, 8M, 12M (default), 16M (samaritan) • MVP: Per core separation/partition • P1: 0, ~~512K~~, 32M, 1M,3M, 6M,	• Using RCID + Access Type for SC way partition • Using 1-way partition granularity
Physical Address	• 52 ~~56~~ PA ~~currently planned~~ fixed with Ascalon • P1: 40b, 48b possible?	• Mimic'ing Arm's PA space for easier transition - is there a restriction on RISC-V? • 56PA currently planned with Ascalon
Vector Width	• MVP: 2*256 (default) , 1*256	• Will need to limit Vector issue rate as well to meet Export restrictions

Cluster I/F	• MVP
	  ◦ CHI.E
	  ▪ 1 - 2 (default) ports
	  ▪ 256b (Default)
	  ◦ AXI5-Lite
	  ▪ 1 (default) - 2 ports
		• Confirmed CHI.E

	▪ 512b (default) • P1 ◦ 1-4 ports ◦ Flexible bit-width on interface ◦ ACP equivalent port	
Power Management	• MVP: Arm equivalent (except Async DVFS)	• Currently Ascalon only has T-states support (throttling issue rate)

Configuration Priority

2. List of config prioritization (to kickstart validation and be ready for future products):

Fixed values

Decoder Width	L1 I-Cache	L1 D-cache	CHI Count (w: 256b)	AXI Count (w: 512b)
D8	64K	256K	2	1
D4	32K	64K	2	1
D6	32K	64K	2	1
D2	32K	32K	2	1

Priority config

Decoder Width	Core Count	L2 Size	Priority	Build configs	Schedule from RV Eng	Notes
D8 ASC X8c12m	8	12M	P0.0 PoR (LA/EA releases)	v8_n2_32B_c8_12mb v8_er_n2_32B_c8_12mb	Beta: April 2025 LA: July 2025 EA: Sept 2025	Ascalon default config (Aegis)
D8 ASC X4c4m	4	4M	P0.1	v8_n2_32B_c4_4mb v8_er_n2_32B_c4_4mb	Alpha: 10/30/25 (For Aria) Beta: LA: EA:	Alexandria config. standalone high-performance DTV solution, etc... might be difficult to differentiate

						from Aegis config
D4 ASC S4c4m	4	4M	P0.2	v4_n2_32B_c4_4mb v4_er_n2_32B_c4_4mb	Beta: LA: EA:	Targeting A78 replacement with 512K L2/core. Treating as mainstream product for us MBLY, MSFT, Axelera, AMD, Positron)
D8 ASC X2c4m	2	4M	P1	v8_n2_32B_c2_4mb v8_er_n2_32B_c2_4mb	Beta: LA: EA:	highest single thread CPU performance (compare to Apple A15) P1 only if this config DV is large effort. If the 2 core config falls out of the 4 core config efforts easily, promote priority to P0.
D6 ASC H4c8m	4	8M	P1	v6_n2_32B_c4_8mb v6_er_n2_32B_c4_8mb		
D6 ASC H2c8m	2	8M	P2	v6_n2_32B_c2_8mb v6_er_n2_32B_c2_8mb		
D4 ASC S4c2m	4	2M	P3 → P1	v4_n2_32B_c4_2mb v4_er_n2_32B_c4_2mb		Unless we have some trick = uarch change, L2 cache becomes 4-way

D2 ASC U4c2m	4	2M	P3	N/A		

Market Availability

(not a product MVP, captured as reference what's available in the market)

	Client	Mobile	Automotive	Infrastructure
Application device	• Laptop, tablet, etc.	• Mobile phones	• Likely IVI (or ADAS / IVI Combo)	• Cloud Computing, etc.
Priority for TT1	• Priority (Mid) • Not as high as a couple of blocking factors at the initial generation (OS, LITTLE cores)	• Priority (Mid) • Same as Client •	• Priority (High)	• Priority (High) •
Notes	• bL configuration needed to be power and area efficient • Given that we don't have "LITTLE" cores, we may need a dual cluster solution and target prime cluster	• Similar to Cli configuration •	• Not as stringent requirement on PPA compared to Client/Mobile/Infra yet •	• Usually have smaller number of cores (and LLC) per cluster
Decoder Width	• D8 • D4/~~D2~~: P1	• D8 • D4/D2: we are at the A78 (2019/2020) generation PPA (market for it?)	• D8, D4	• MVP: D8, D4 (?) • P1: D2
CPU/cluster (TT1)	• MVP: 2, 4, 8 • P1: 10	• MVP: 2, 4 • P1: 8	• MVP: 4, 8, 10 (?)	• MVP: 1, 2, 4, 8

L2 size	• MVP: 0, 512K, 1M, 2M, 4M, 8M, 16M, 32M	• MVP: 0, 512K, 1M, 2M, 4M, 8M, 16M, 32M	• MVP: 0, 512K, 1M, 2M, 4M, 8M, 16M • P1: 32M	• MVP: 512k, 1M, 2M, 4M, 8M, 16M •
PA	• MVP: 40b	• MVP: 40b	• MVP: 40b	• MVP: 48b, 52b
Vector Width	• 2*256, 2*128	• 2*256, 2*128	• 2*256, 2*128	• 2*256, 2*128
Cluster I/F	• AMBA5 CHI.E	• AMBA5 CHI.E •	• AMBA5 CHI.E •	• AMBA CHI.E •

🗒️ Lead Partner Requirements (label = ascalon_flex)

We couldn't find anything matching your search

Try again with a different term.

D8	Samaritan	IOMMU and AIA	Need as a deliverable	
		Trace and debug as a product	Need as a deliverable	
		L2 Cache size	16M	
		L1 Cache size update	128K	

❓ Open Questions

Question	Answer	Date Answered

⚠️ Out of Scope