Integrated Matrix Extension (IME)

Task Group Meeting

Guido Araujo
Jose Moreira

06/17/24

Agenda

- Update on TG schedule
- Revisiting CI under a datatype glass
- [Jose Moreira] Option C* with mixed datatypes

Agenda

- Update on TG schedule
- Revisiting CI under a datatype glass
- [Jose Moreira] Option C* with mixed datatypes

Update on TG schedule (Previous)

Task	Del.	Task Description	Meetings											
			1	2	3	4	5	6	7	8	9	10	11	12
1		Architectural features												
3		a. uArch: Overall analysis												
6		b. uArch: Memory access analysis												
2		c. ISA: Matrix data encoding												
4		d. ISA: Register usage and mapping												
6		e. ISA: Data type and geometry configuration												
5		f. ISA: Binary compatibility												
6		g. ISA: Computation operations definition												
7		h. ISA: Instruction encoding												
8		Workloads and bechmarking												
9		a. ML: T-Head profiling and ConvBench												
10		b. HPC: Polybench												
11		c. ML: POWER10 MMA transfers												
12		d. Workload analysis												
13		Quantitative analysis												
14		a. QEMU modelling												
15		b. Performace evaluation												
16		Definition of the final architecture												
17		a. RVM ISA v0												
18		b. RVM ISA v1												
19		RVM Spec writing												

Delivery date
Still discussing
Not touched
N Past meeting
Delayed

Update on TG schedule (Current)

Task	Del.	Task Description	Meetings											
			1	2	3	4	5	6	7	8	9	10	11	12
1		Architectural features												
3		a. uArch: Overall analysis												
6		b. uArch: Memory access analysis												
2		c. ISA: Matrix data encoding												
4		d. ISA: Register usage and mapping												
6		e. ISA: Data type and geometry configuration												
5		f. ISA: Binary compatibility												
6		g. ISA: Computation operations definition												
7		h. ISA: Instruction encoding												
8		Workloads and bechmarking												
9		a. ML: T-Head profiling and ConvBench												
10		b. HPC: Polybench												
11		c. ML: POWER10 MMA transfers												
12		d. Workload analysis												
13		Quantitative analysis												
14		a. QEMU modelling												
15		b. Performace evaluation												
16		Definition of the final architecture												
17		a. RVM ISA v0												
18		b. RVM ISA v1												
19		RVM Spec writing												

Delivery date
Still discussing
Not touched
N Past meeting
Delayed

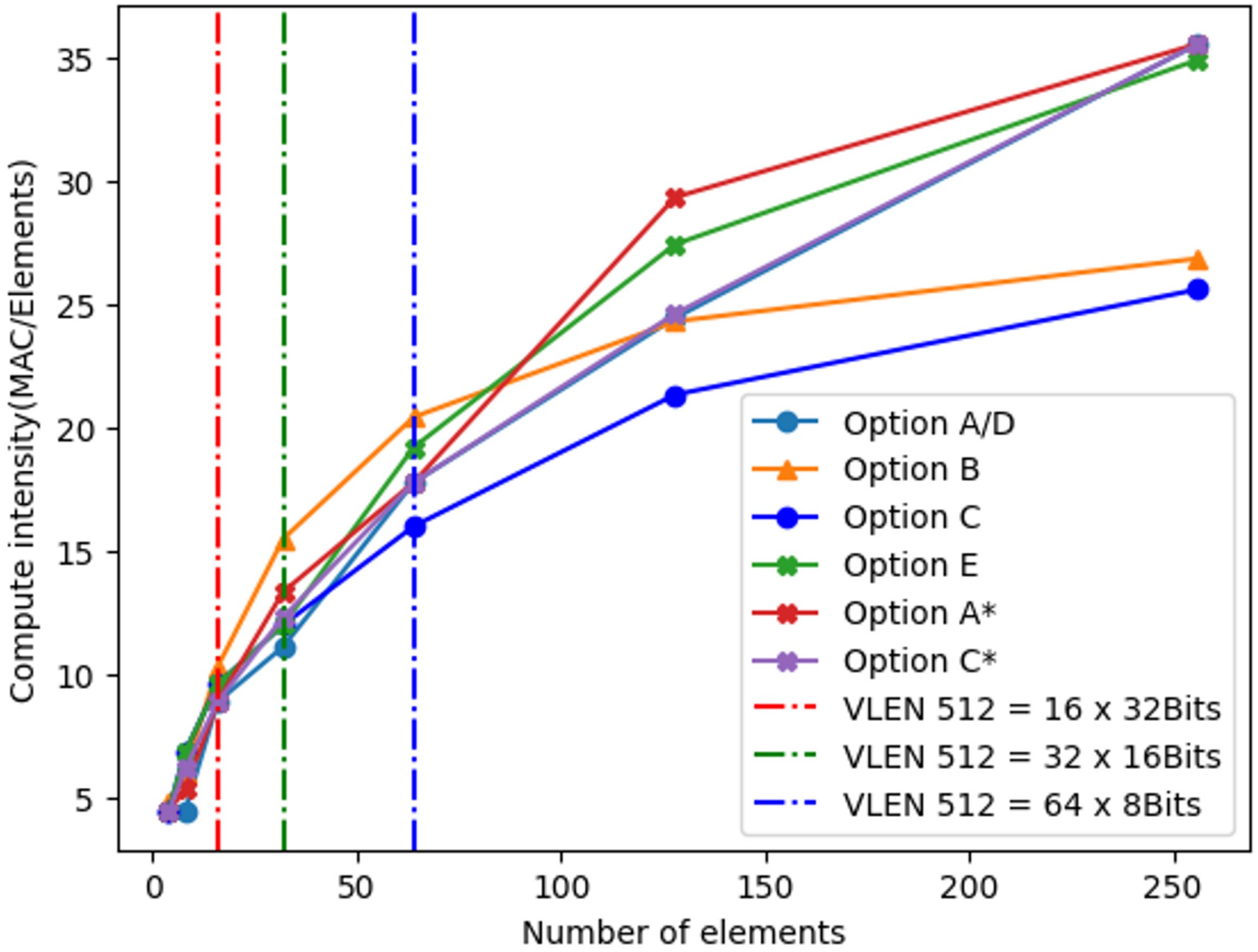
Working groups

Group	Coordinator	Members
Option A and A*	Marc Casas	Huayue Liang, Erich Focht
Option B		
Option C and C*	Jose Moreira	
Option D	Abel Bernabeu	
Option E	Jim (CN.Ke)	Yi-Xuan.Huang
Workloads and benchmarking	Guido Araujo	

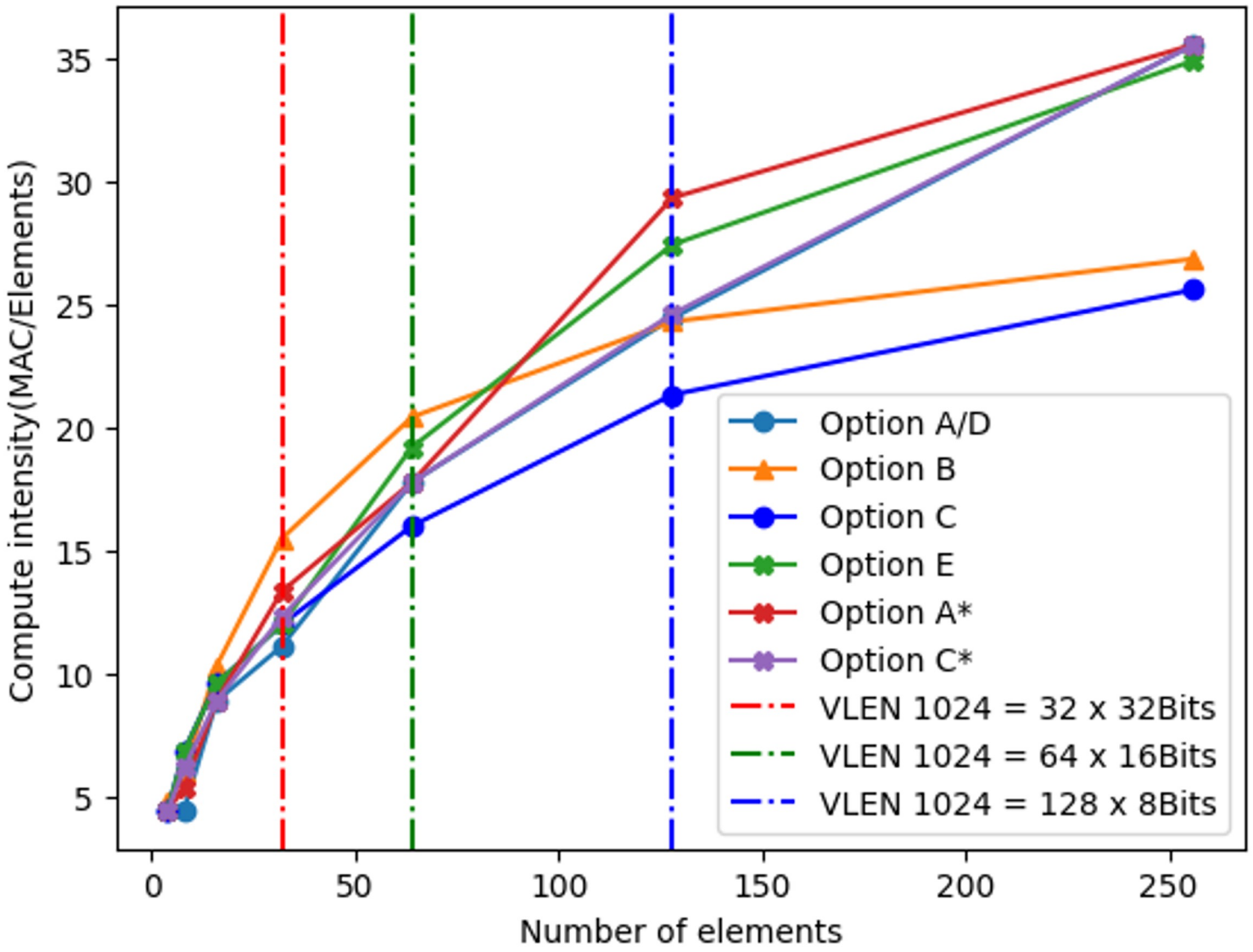
Agenda

- Update on TG schedule
- Revisiting CI under a datatype glass
- [Jose Moreira] Option C* with mixed datatypes

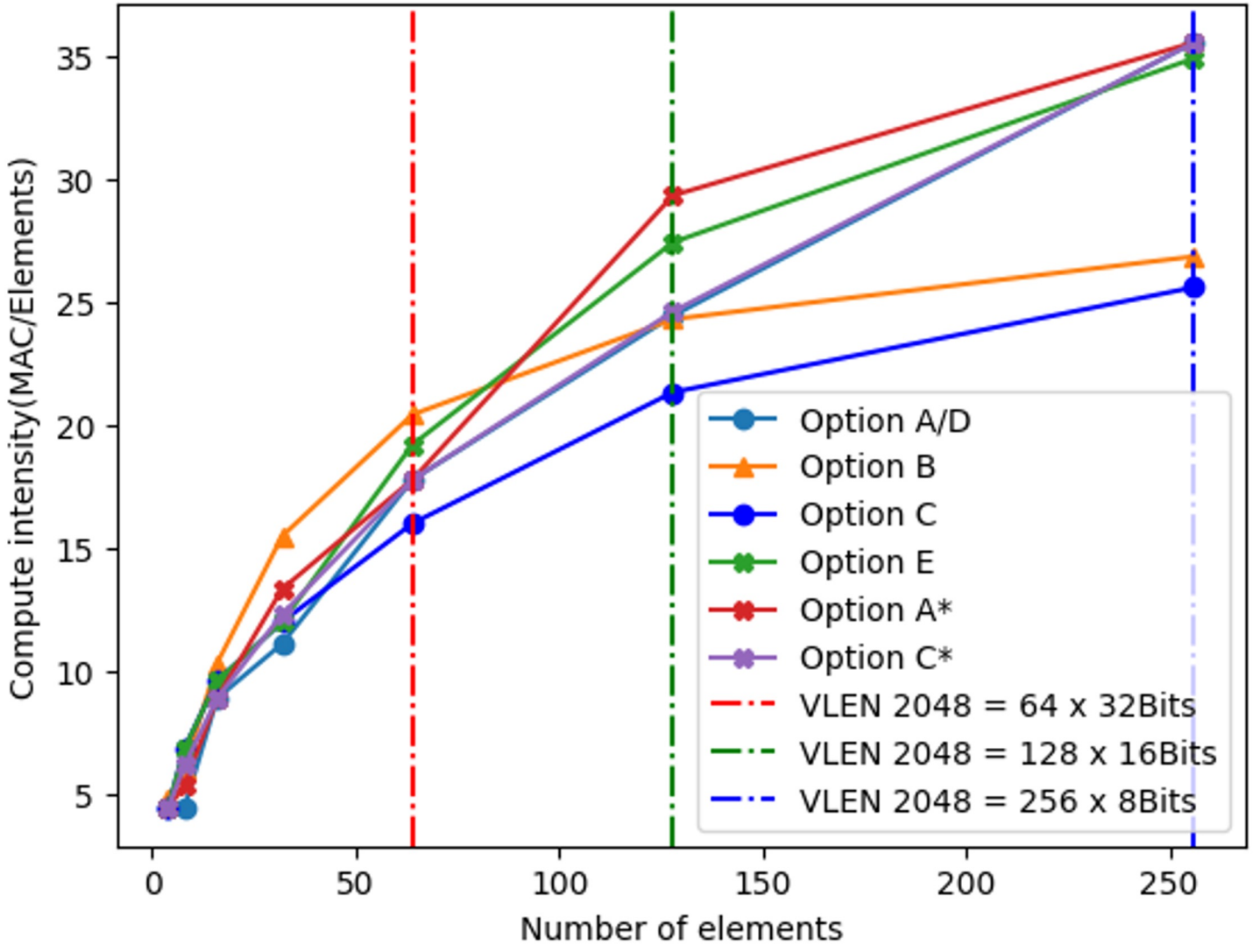
VLEN = 512


Compute intensity(MAC/Elements)
Number of elements
0
50
100
150
200
250
5
10
15
20
25
30
35
Option A/D
Option B
Option C
Option E
Option A*
Option C*
VLEN 512 = 16 x 32Bits
VLEN 512 = 32 x 16Bits
VLEN 512 = 64 x 8Bits

VLEN = 1024


Compute intensity(MAC/Elements)
Number of elements
0
50
100
150
200
250
5
10
15
20
25
30
35
Option A/D
Option B
Option C
Option E
Option A*
Option C*
VLEN 1024 = 32 x 32Bits
VLEN 1024 = 64 x 16Bits
VLEN 1024 = 128 x 8Bits

VLEN = 2048


Compute intensity(MAC/Elements)
Number of elements
0
50
100
150
200
250
5
10
15
20
25
30
35
Option A/D
Option B
Option C
Option E
Option A*
Option C*
VLEN 2048 = 64 x 32Bits
VLEN 2048 = 128 x 16Bits
VLEN 2048 = 256 x 8Bits

Agenda

- Update on TG schedule
- Revisiting CI under a datatype glass
- [Jose Moreira] Option C* with mixed datatypes