CIM

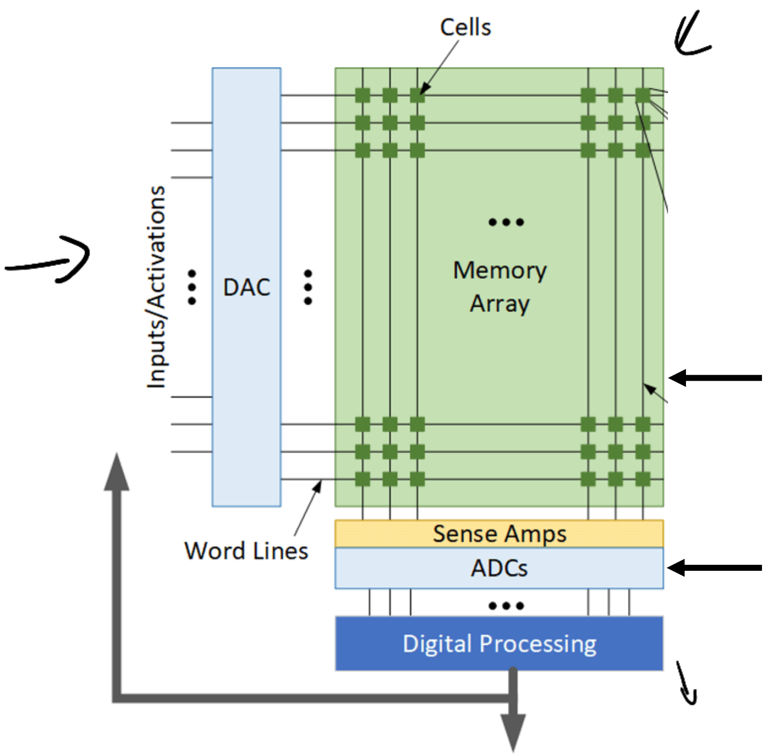
B11901027 王仁軒

Outline

- Analog vs Digital
- Analog CIM structure
  - Optimization techniques for analog designs
- Digital CIM structure
  - Optimization techniques for digital designs
- Overall comparison among 3 papers

Analog CIM Structure

- Typical analog CIM Structure
  - Require DAC & ADC

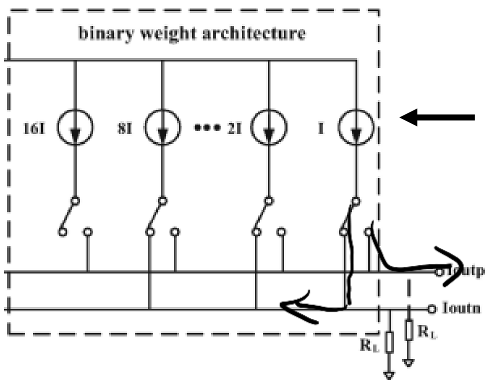


Multiply input with weight
(n-bit) (1-b)

Sum of all partial products

Optimization for Analog Designs

- Analog multiplication requires DAC
  - Current source may consume lots of power

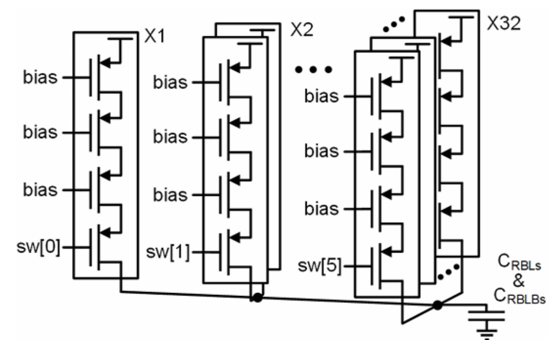


Differential output

Current source are always on

Single-ended output

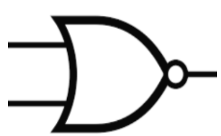
Can turn off current source



- Not using DAC
  - Digital multiplication

Input_B

Weight(QB)



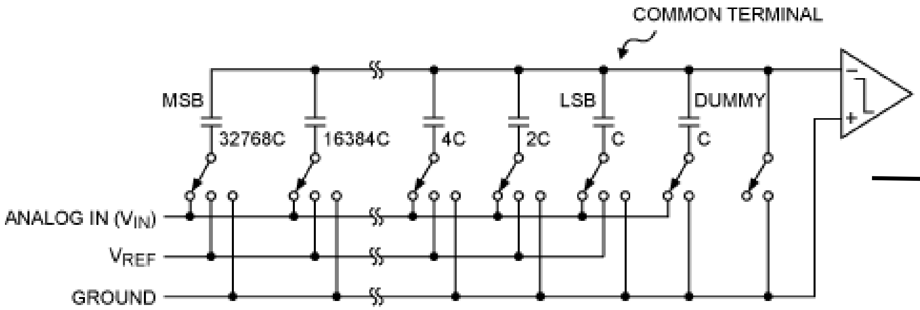
1-bit partial sum

Require shift and sum to do n-bit x n-bit from 1-bit x 1-bit

4b x 4b → 16

Optimization for Analog Designs

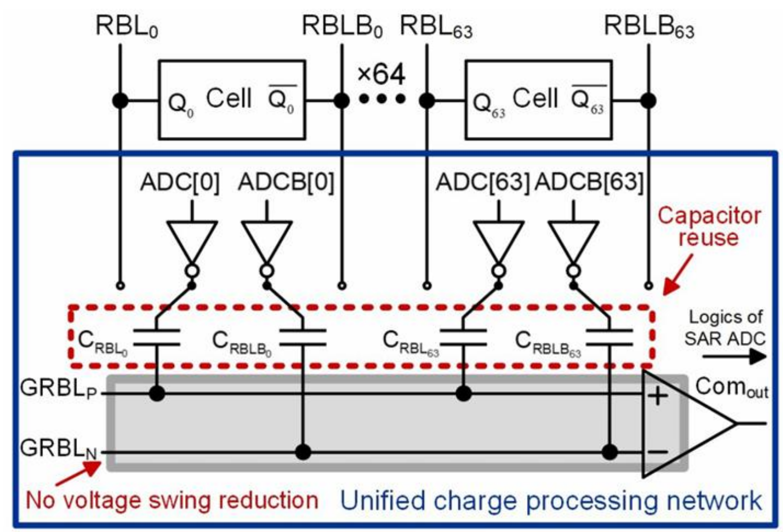
- ADC is required



Capacitance $\propto 2^{\text{Bit precision}}$

→ Large area for caps

- Reuse of Capacitors



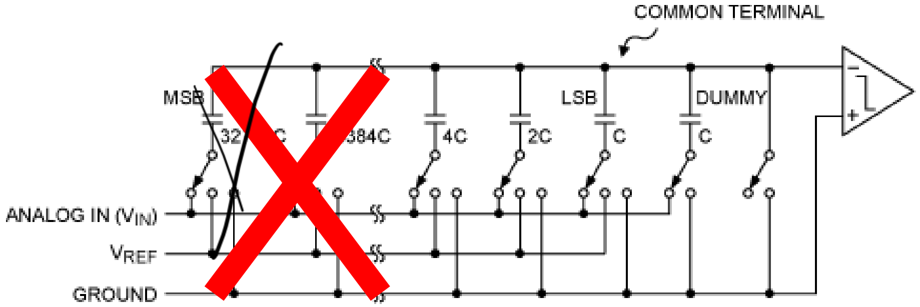
→ Reuse of caps for ADC allows us to save area

Optimization for Analog Designs

- Skipping iterations

86

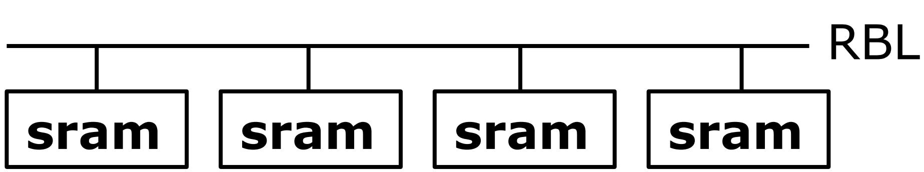
127



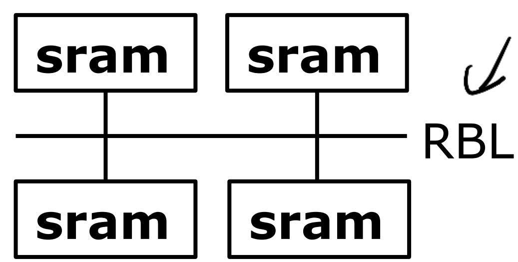
Input has 32 1's
Partial sum < 32
Skip first few iterations
saves time

- Interleaving (memory cell)

$\frac{256\,C}{256C}$



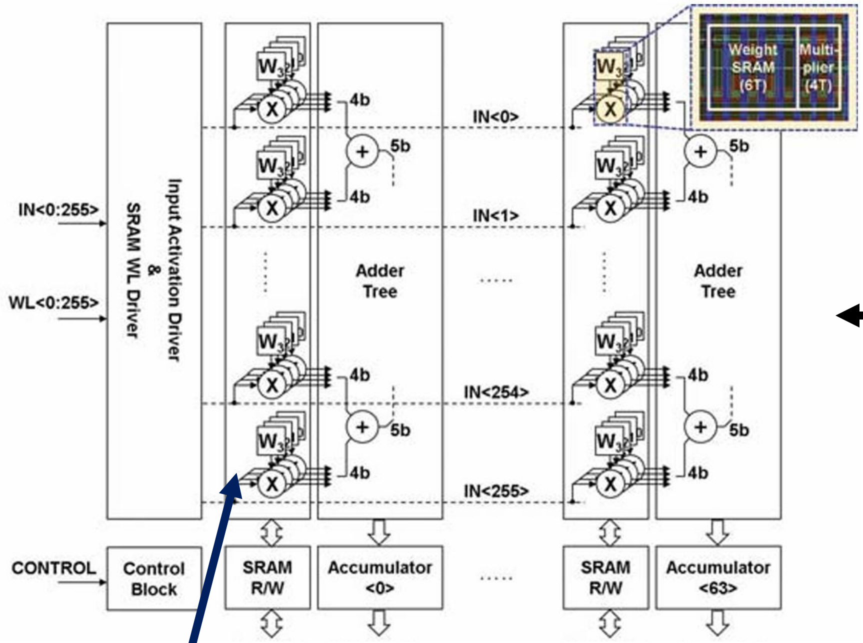
RBL length = L
Parasitic caps = C



RBL length = L/2
Parasitic caps = C/2

Digital CIM Structure

- Typical digital CIM design
  - Serial input



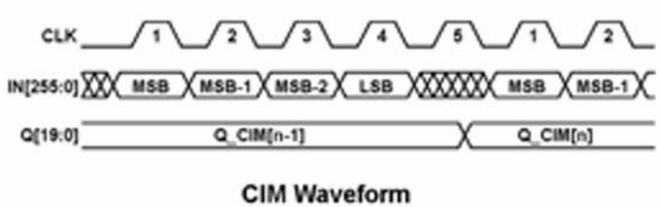
Adder tree to sum all partial sums

Digital multiplication (1b input x 4b weight)

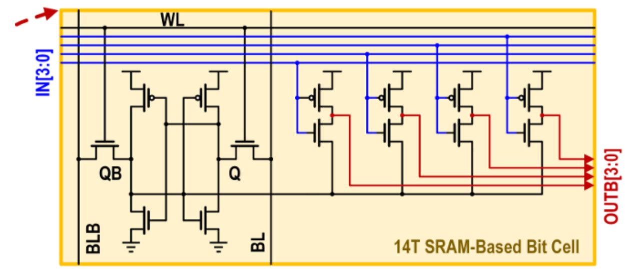
Optimization for Digital Designs

- Increase throughput
  - Increase input BW

46 x 46
16x 46

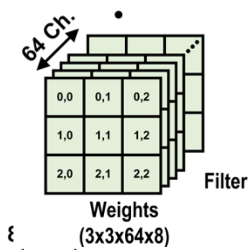


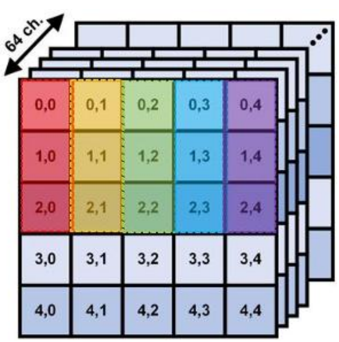
Serial input
Require bit-precision+1 cycles

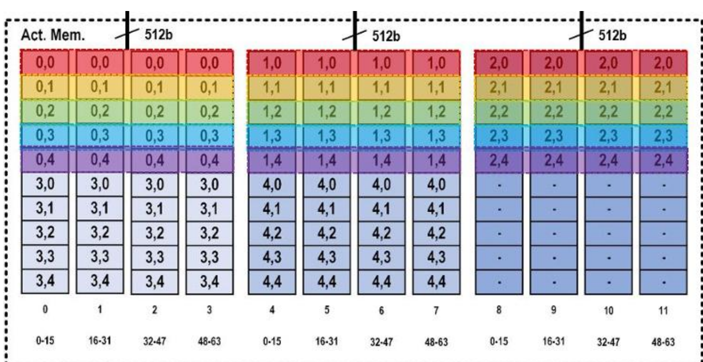


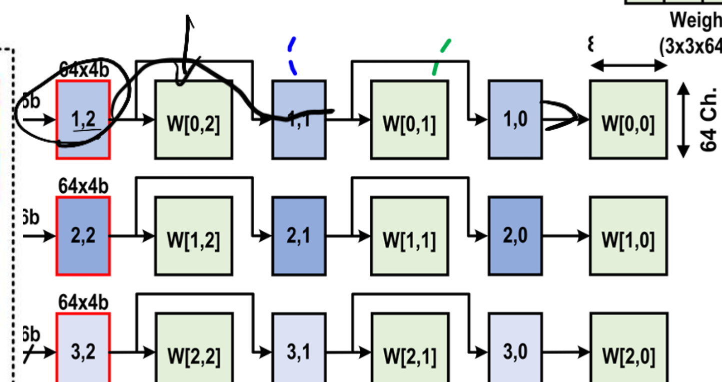
Parallel input
Require bit-precision/4+1 cycles

- High BW input memory
  - In order to support MAC pipeline



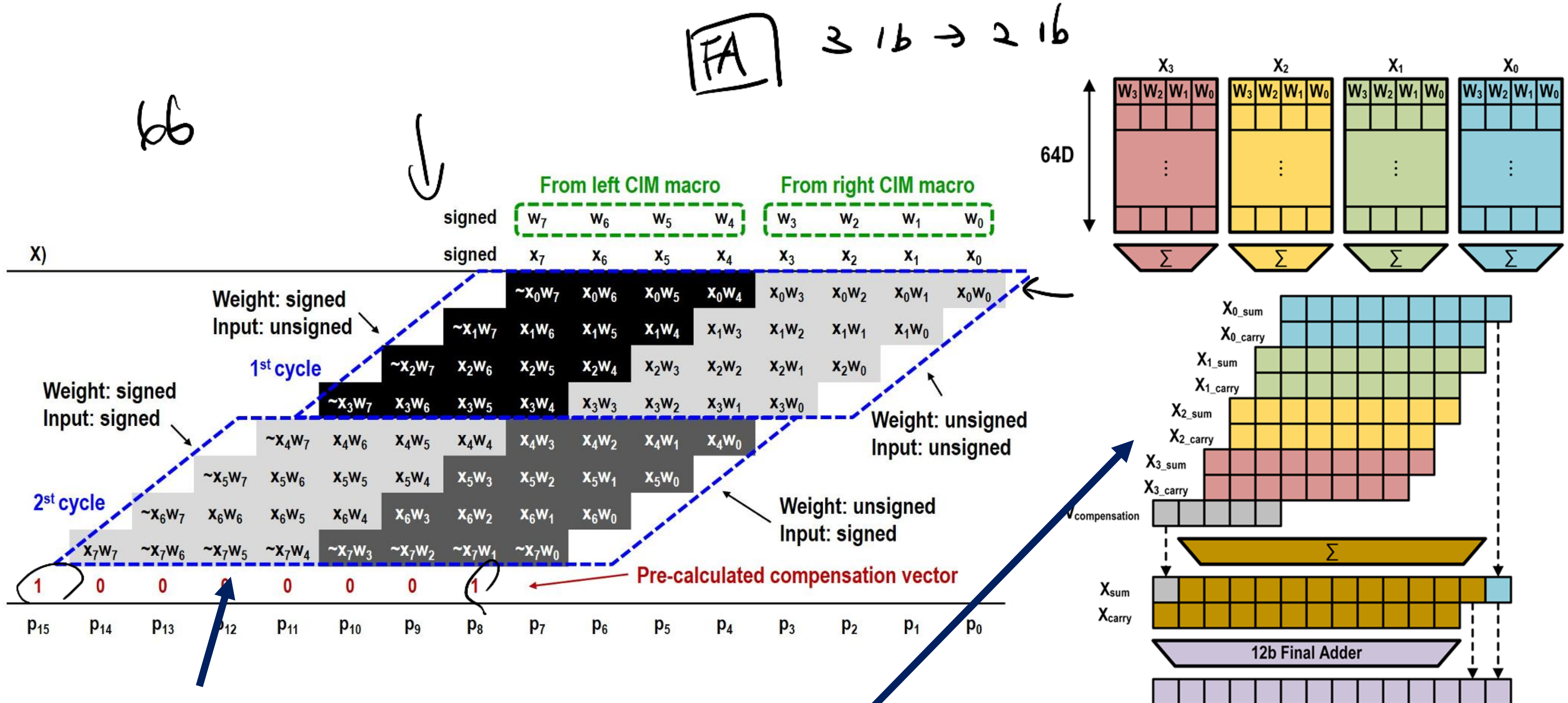






Optimization for Digital Designs

- Bit-flexibility & Adder tree optimization



Each color is produced by a CIM block

Do not propagate carry
Faster than ripple-carry tree

Overall Comparison

	2021/Analog		2023/Analog		2022/Digital
Throughput TOPS (1b x 1b)	Cycle time ↑ 1.30 7b x 1b	← x7	1.31 1b x 1b	Parallel input ↔ Freq increase Pipelining	4x4 98.3 MAC+ Adder
Energy Efficiency TOPS/W (1b x 1b)	293		291	ADC? ↔	625
Area Efficiency TOPS/mm2 (1b x 1b)	23.01	DAC ↔	27.7	Adder tree ↔	10.49