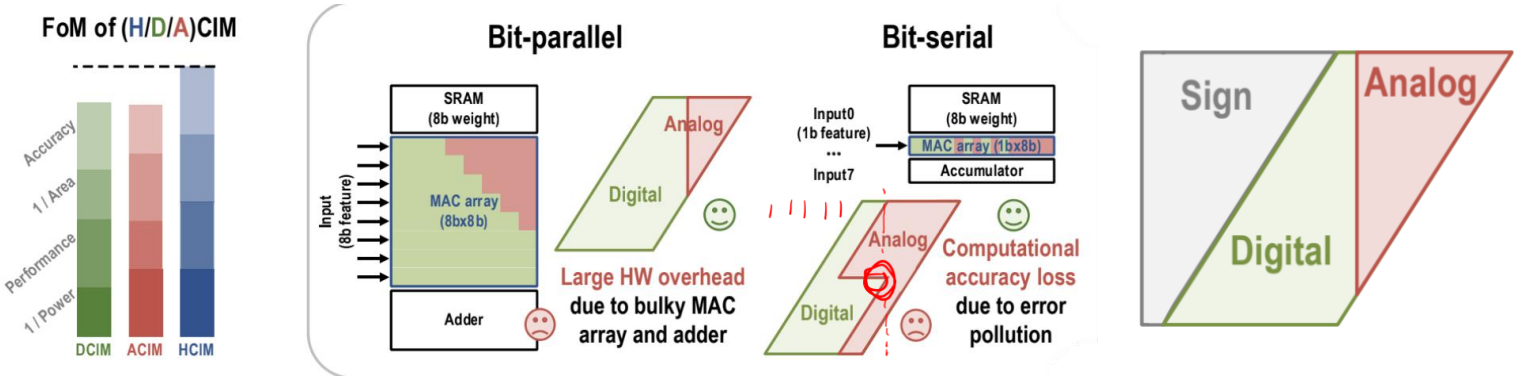
A 28nm 64kb Bit-Rotated Hybrid-CIM Macro with an Embedded Sign-Bit-Processing Array and a Multi-Bit-Fusion Dual-Granularity Cooperative Quantizer

B11901027 王仁軒

Motivation & Challenge

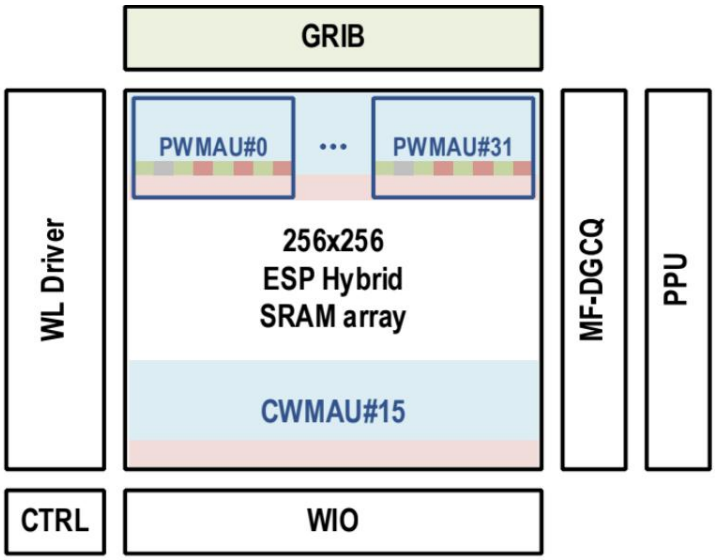
- Hybrid CIM
  - Balance between DCIM and ACIM
  - Challenge1 : definition problem of boundary between digital and analog part

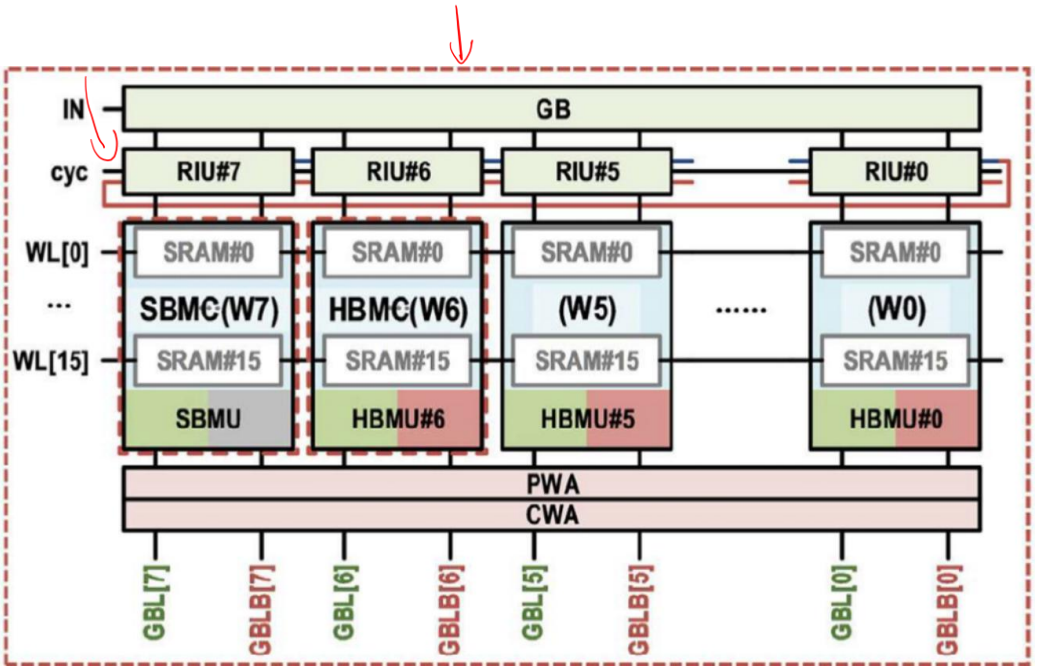


  - Challenge2 : more expensive overhead of sign-bit processing on digital part
  - Challenge3 : energy wasted on low-accuracy-contributed quantization on analog part

CIM Macro

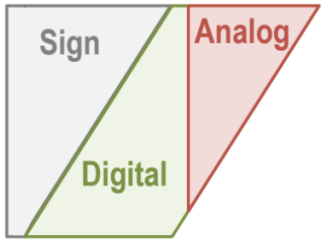
- Hybrid CIM macro

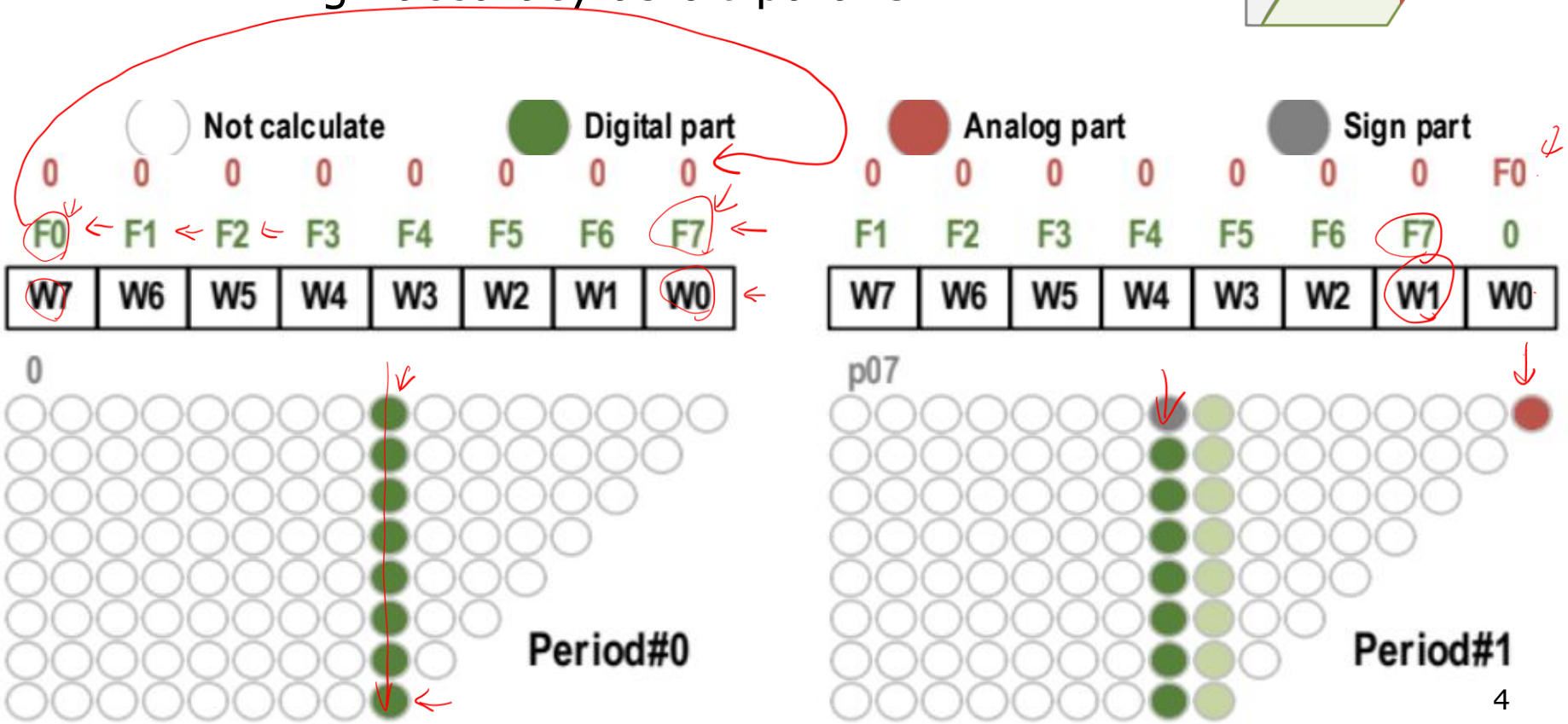




Proposed Scheme

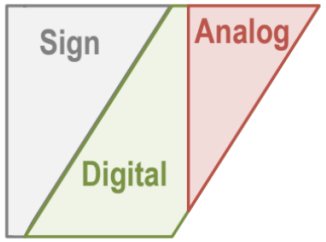
- Bit-rotated feature-in scheme
  - Low hardware overhead as bit-serial
  - High accuracy as bit-parallel

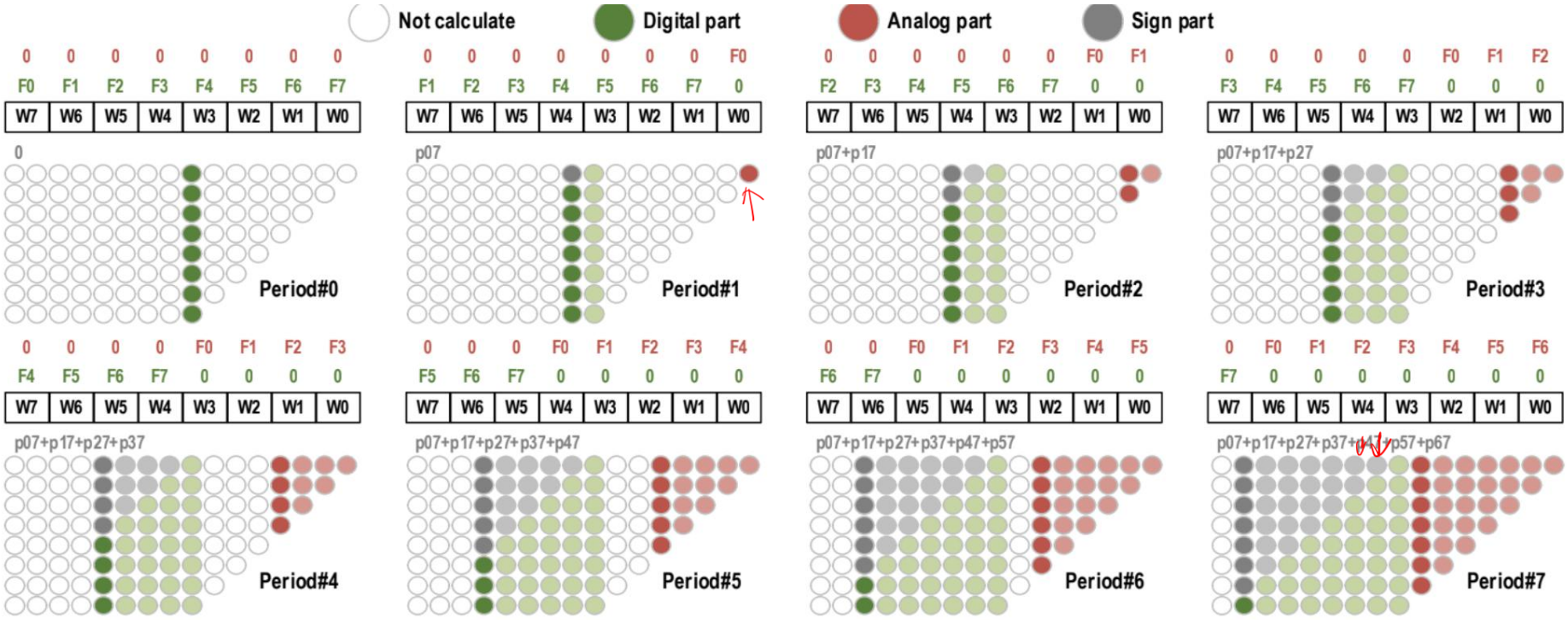




Proposed Scheme

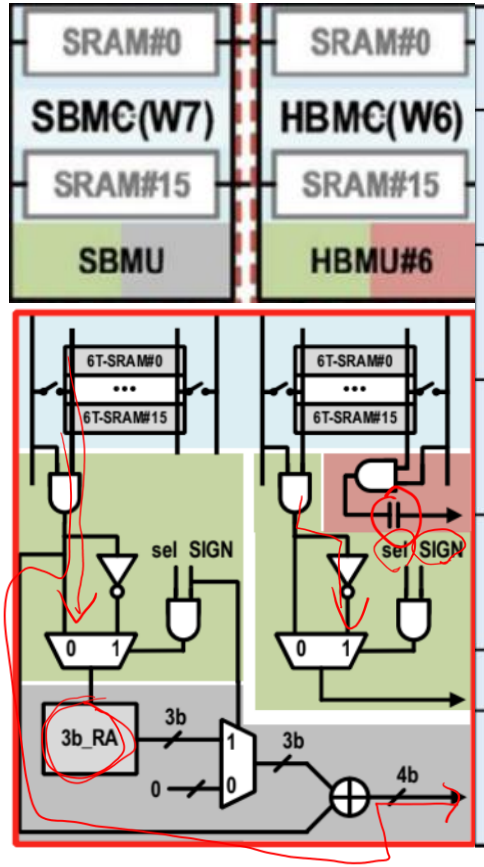
- Bit-rotated feature-in scheme
  - Low hardware overhead as bit-serial
  - High accuracy as bit-parallel



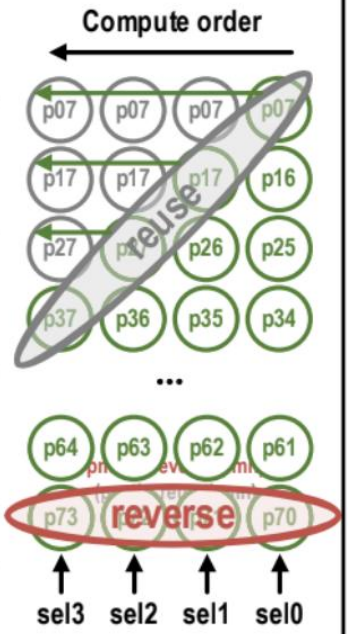


Proposed Scheme

- Embedded sign-bit-processing hybrid array



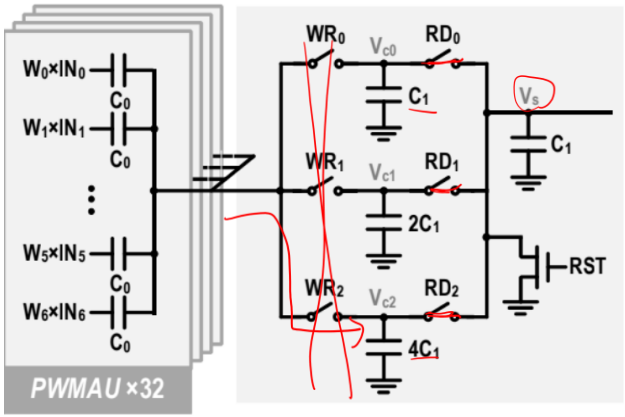
	Signed mode 1/selective reversion; 2/sign-bit extension
Period 0	p07+p16+p25+ p34+p43+p52+ p61+p70*+(0)
Period 1	p17+p26+p35+ p44+p53+p62+ p71*+(p07)
Period 2	p27+p36+p45+ p54+p63+p72* +(p07+p17)
Period 3	p37+p46+p55+ p64+p73*+(p07 +p17+p27)
...	...
Period 7	p77*+(p07+p17 +p27+p37+p47 +p57+p67)

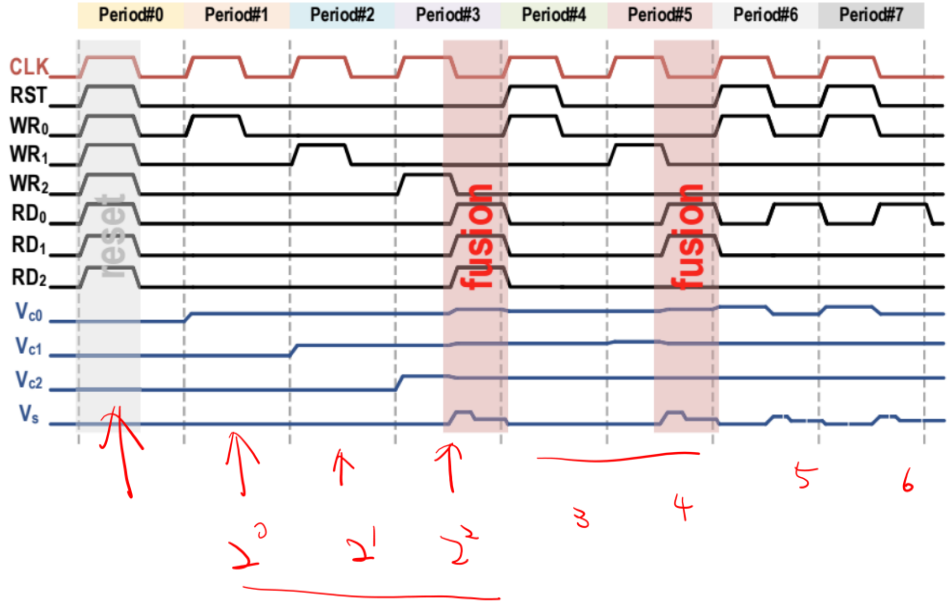



```
  1 0 1 1
  1 1 0 0
-------------
+ 0 0 0 0 0 0 0
+ 0 0 0 0 0 0
+ 1 1 0 1 1
- 1 0 1 1
-------------
```

Proposed Scheme

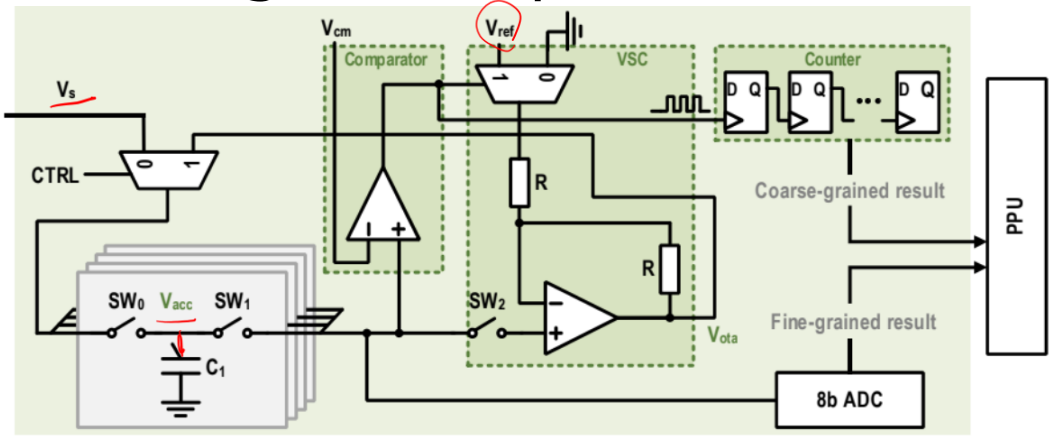
- Multi-bit-fusion
  - Period1 : $V_{c0} = \frac{C}{C+C_1}$, Period2 : $V_{c0} = \frac{C}{C+2C_1}$, Period3 : $V_{c0} = \frac{C}{C+4C_1}$
  - $V_S = \frac{V_{c0}\cdot C_1 + V_{c1}\cdot 2C_1 + V_{c2}\cdot 4C_1}{C_1+C_1+2C_1+4C_1} = \frac{(V_{c0}+2V_{c1}+4V_{c2})C_1}{8C_1}$

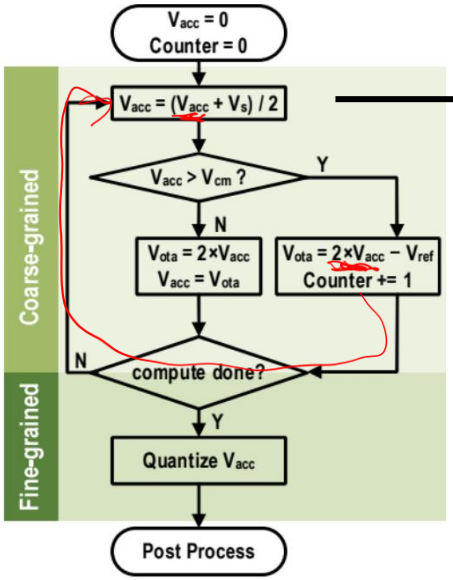




Proposed Scheme

- Coarse/fine-grained quantizer





$$V_{acc} \mathrel{+}= \frac{V_s}{2}$$

Total : N iterations

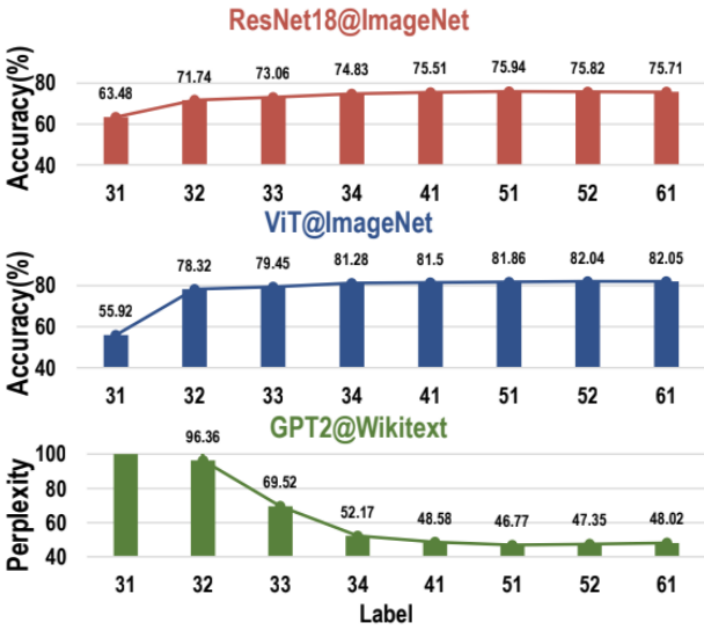
Total V : N · (Vs/2)

Counter : N · (Vs/2) / Vcm

Results

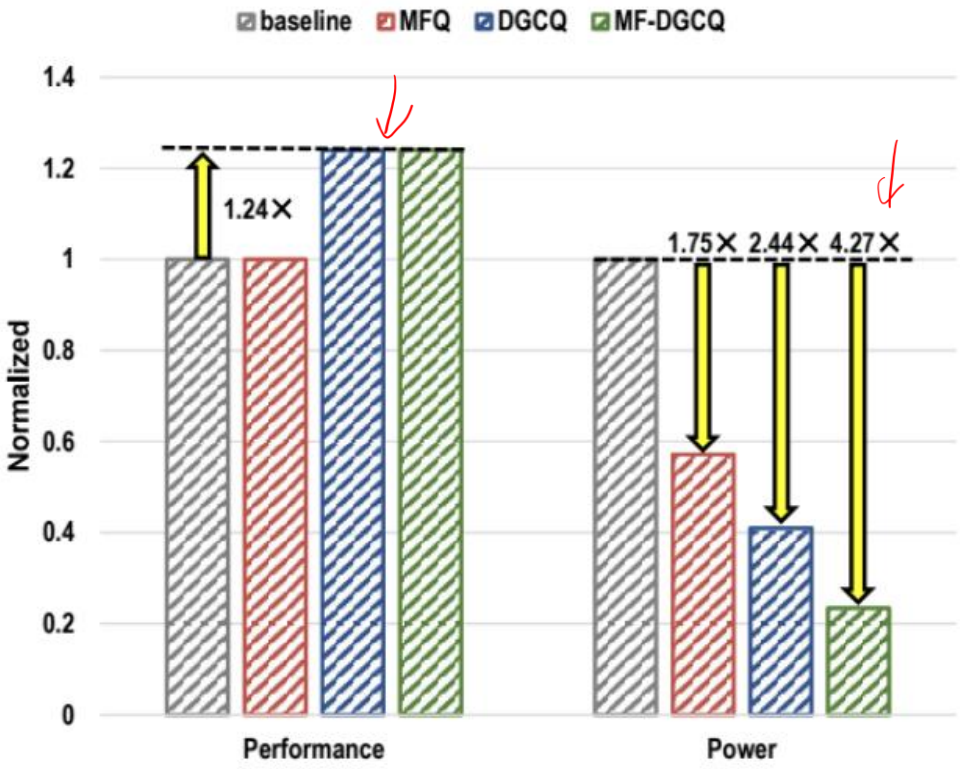
- Low accuracy loss but great energy reduction
  - 4-times readout combination incurs almost no accuracy loss and outperforms 5/6-times readout combinations regarding energy consumption

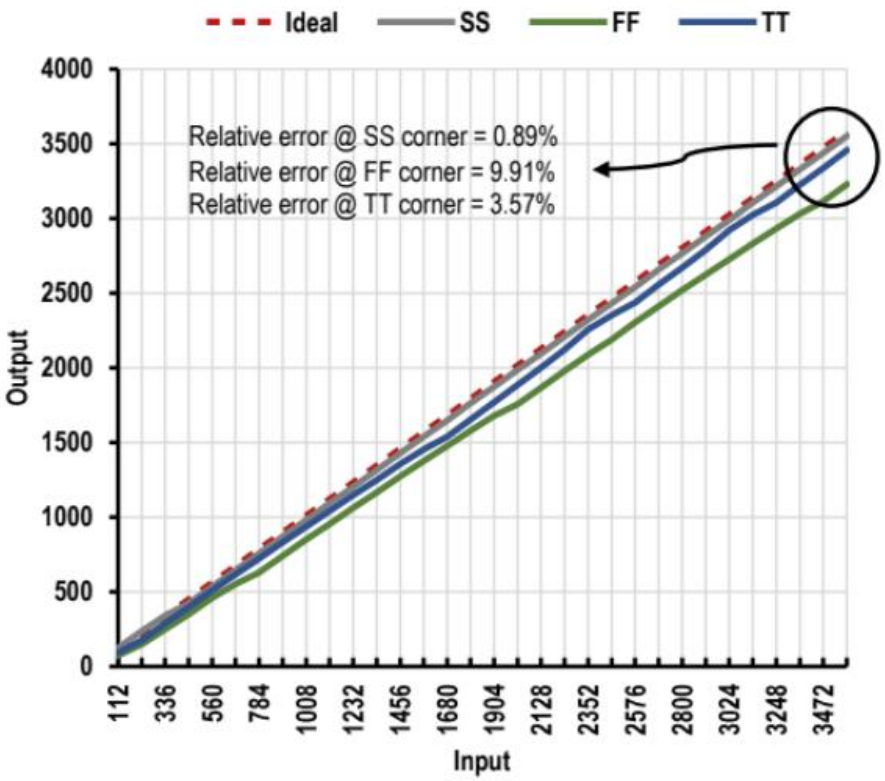
Readout times	Label	Readout combinations
3	31	(O6+O5), (O4+O3), (O2+O1+O0)
	32	(O6), (O5+O4+O3), (O2+O1+O0)
	33	(O6), (O5+O4), (O3+O2+O1+O0)
	34	(O6), (O5), (O4+O3+O2+O1+O0)
4	41	(O6), (O5), (O4+O3), (O2+O1+O0)
5	51	(O6), (O5), (O4), (O3+O2), (O1+O0)
	52	(O6), (O5), (O4), (O3), (O2+O1+O0)
6	61	(O6), (O5), (O4), (O3), (O2), (O1+O0)



Results

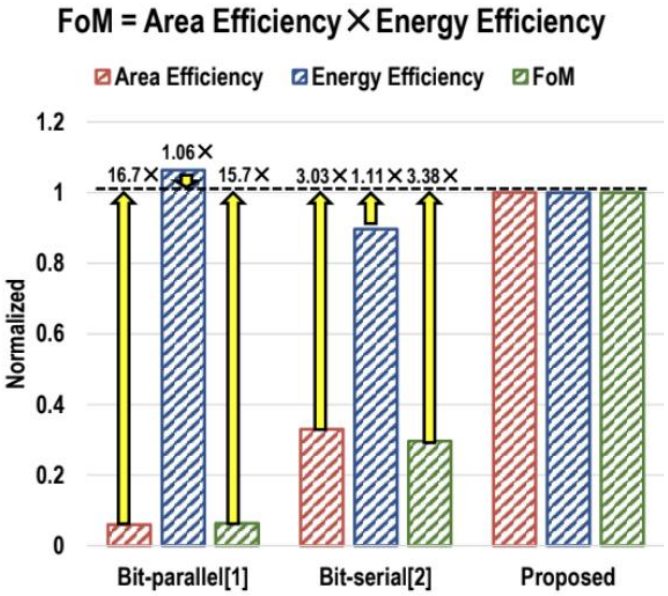
- MF-DGCQ achieves 24% performance enhancement and 4.27× power reduction with relative error of 3.57% in TT corner





Results

- This design outperforms typical bit-parallel / serial work on the FoM (area efficiency × energy efficiency) by 15.7× and 3.38×.

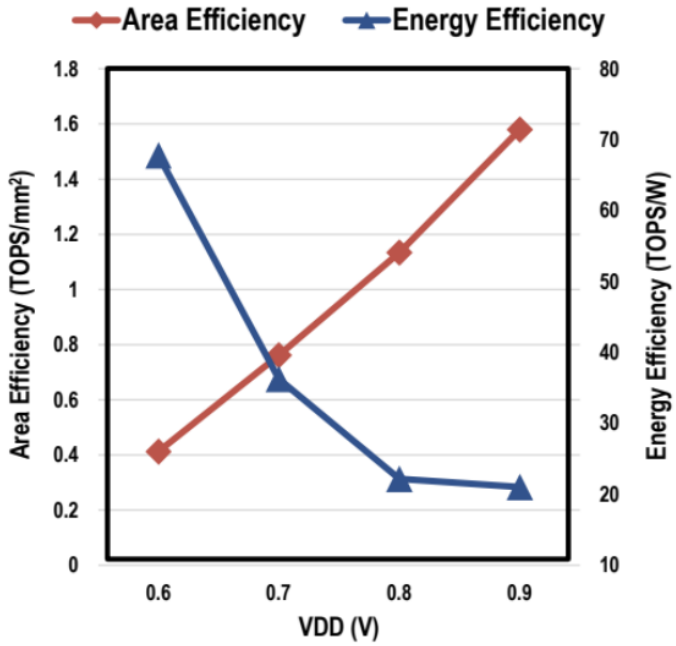


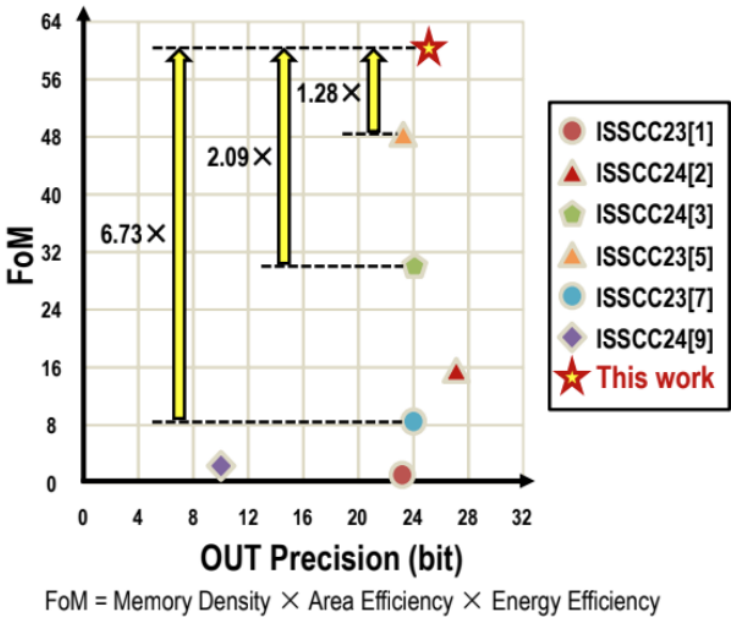
Model	ResNet-18	ViT	GPT-2
Dataset	ImageNet	ImageNet	Wikitext-102
Data precision	INT8	INT8	INT8
Task	Classification	Classification	NLP
Metric	Accuracy	Accuracy	Perplexity
Baseline	69.54%	80.24%	48.39
Proposed	68.48%	78.49%	48.58

- Accuracy loss -1.06% for ResNet-18@ImageNet, -1.75% for ViT@ImageNet, 0.19 for GPT-2@Wikitext-102.

Results

- 21.04TOPS/W@0.9V, 1.57TOPS/mm2@0.9V.
- Access time is 12ns@0.9V, 8bIN–8bW–21bOUT.





FoM = Memory Density × Area Efficiency × Energy Efficiency