UNSW UG Thesis seminar

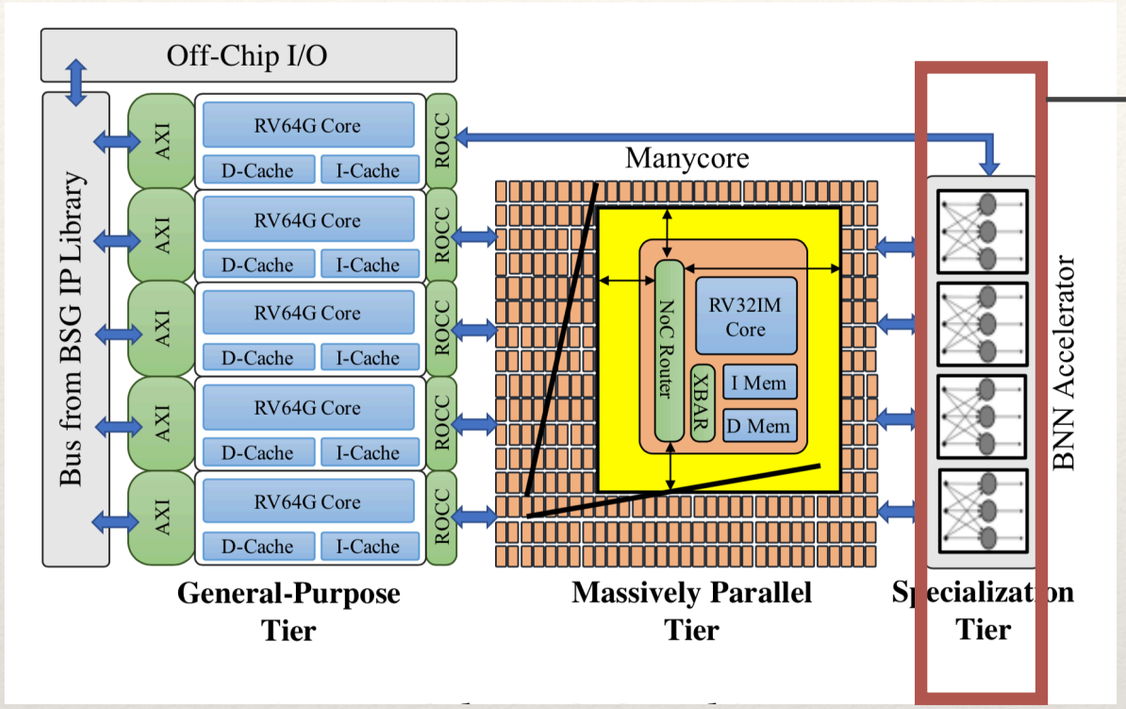
Deep Neural Network engine on low cost hardware accelerator

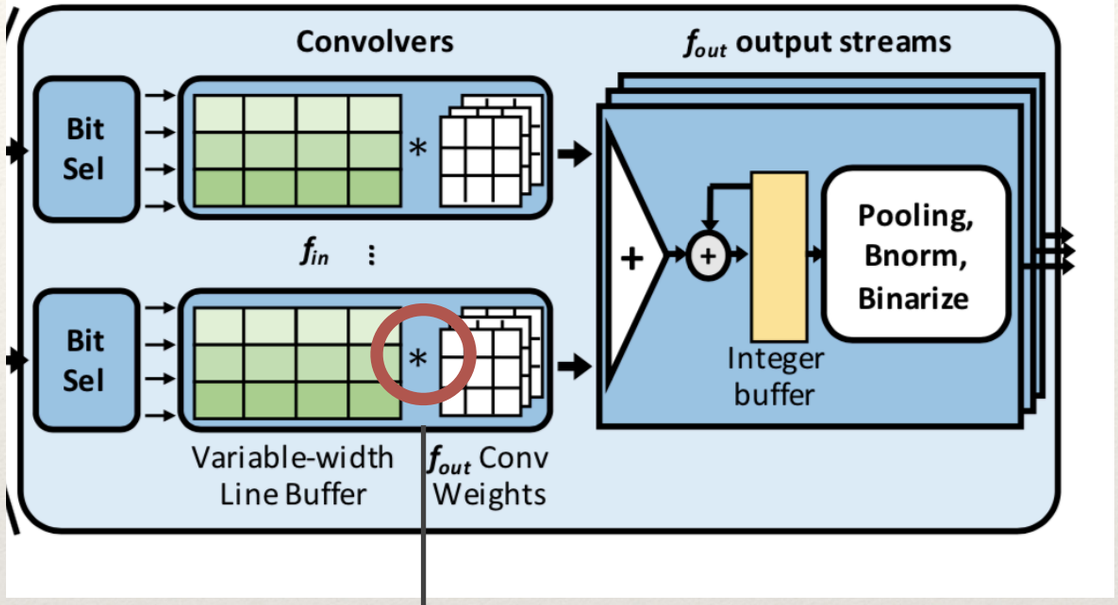
Yikai Wang
z5123540

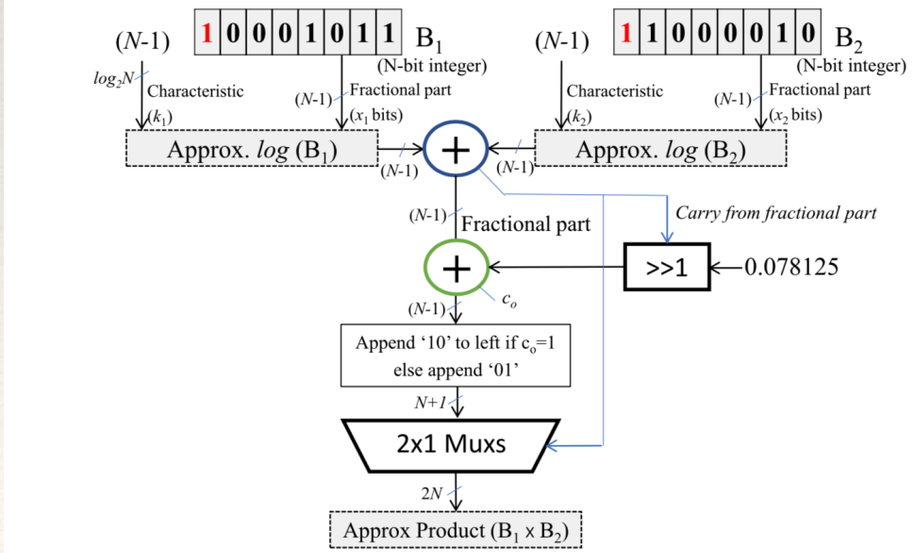
Table of content

- Re-cap
- Lenet-5 on RISC-V core software simulation
- MBM hardware waveform simulation
- Problems encountered
- System Architecture

Re-cap


Off-Chip I/O
Bus from BSG IP Library
AXI
RV64G Core
D-Cache
I-Cache
ROCC
Manycore
NoC Router
XBAR
RV32IM Core
I Mem
D Mem
BNN Accelerator
General-Purpose Tier
Massively Parallel Tier
Specialization Tier


Convolvers
f_{out} output streams
Bit Sel
f_{in}
Pooling, Bnorm, Binarize
Integer buffer
Variable-width Line Buffer
f_{out} Conv Weights


(N-1)
1 0 0 0 1 0 1 1
B1
(N-bit integer)
log2N
Characteristic
(k1)
Fractional part
(x1 bits)
Approx. log (B1)
1 1 0 0 0 0 1 0
B2
(k2)
(x2 bits)
Approx. log (B2)
Carry from fractional part
>>1
0.078125
co
Append '10' to left if co=1
else append '01'
N+1
2x1 Muxs
2N
Approx Product (B1 x B2)

Minimum biased error reduced approximate multiplier (MBM)

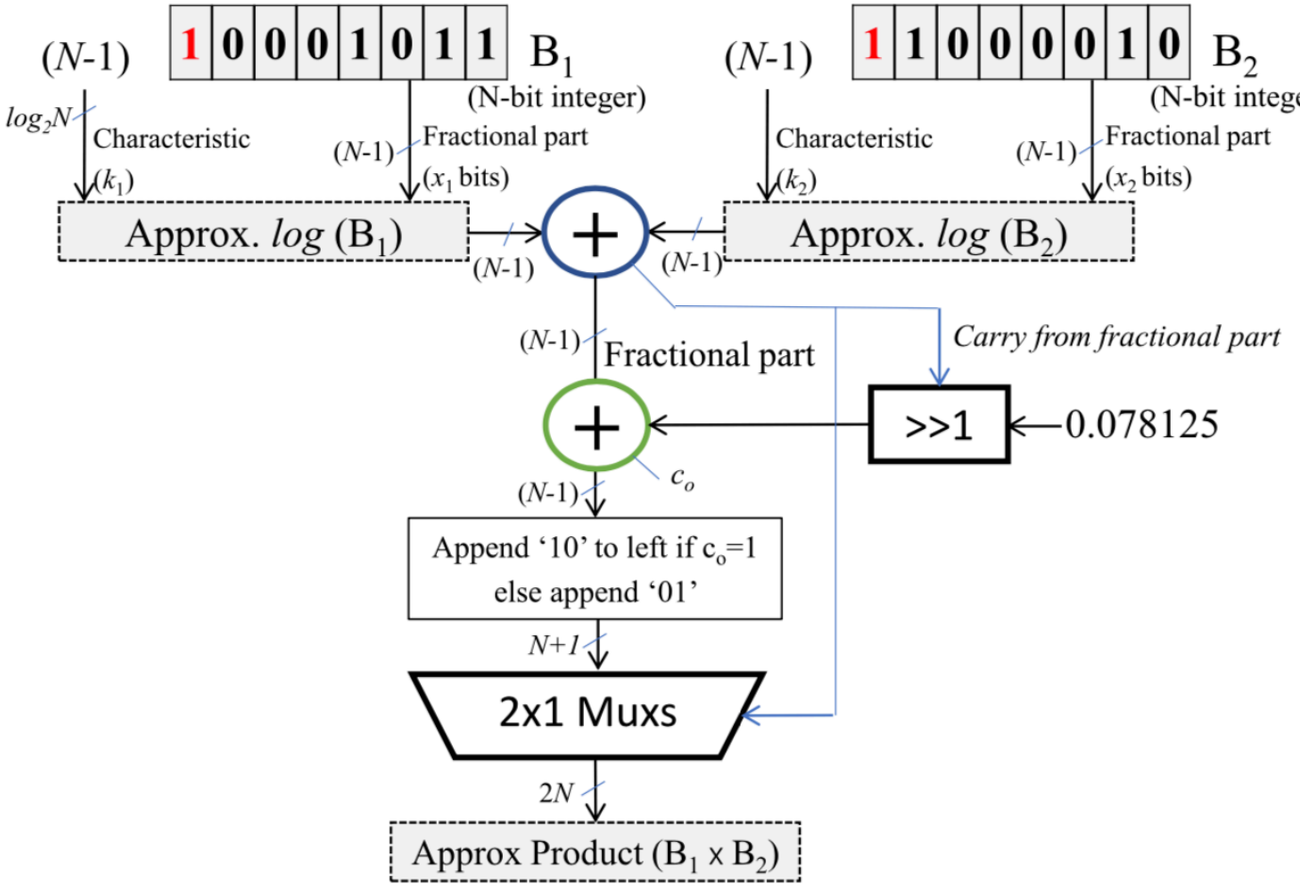
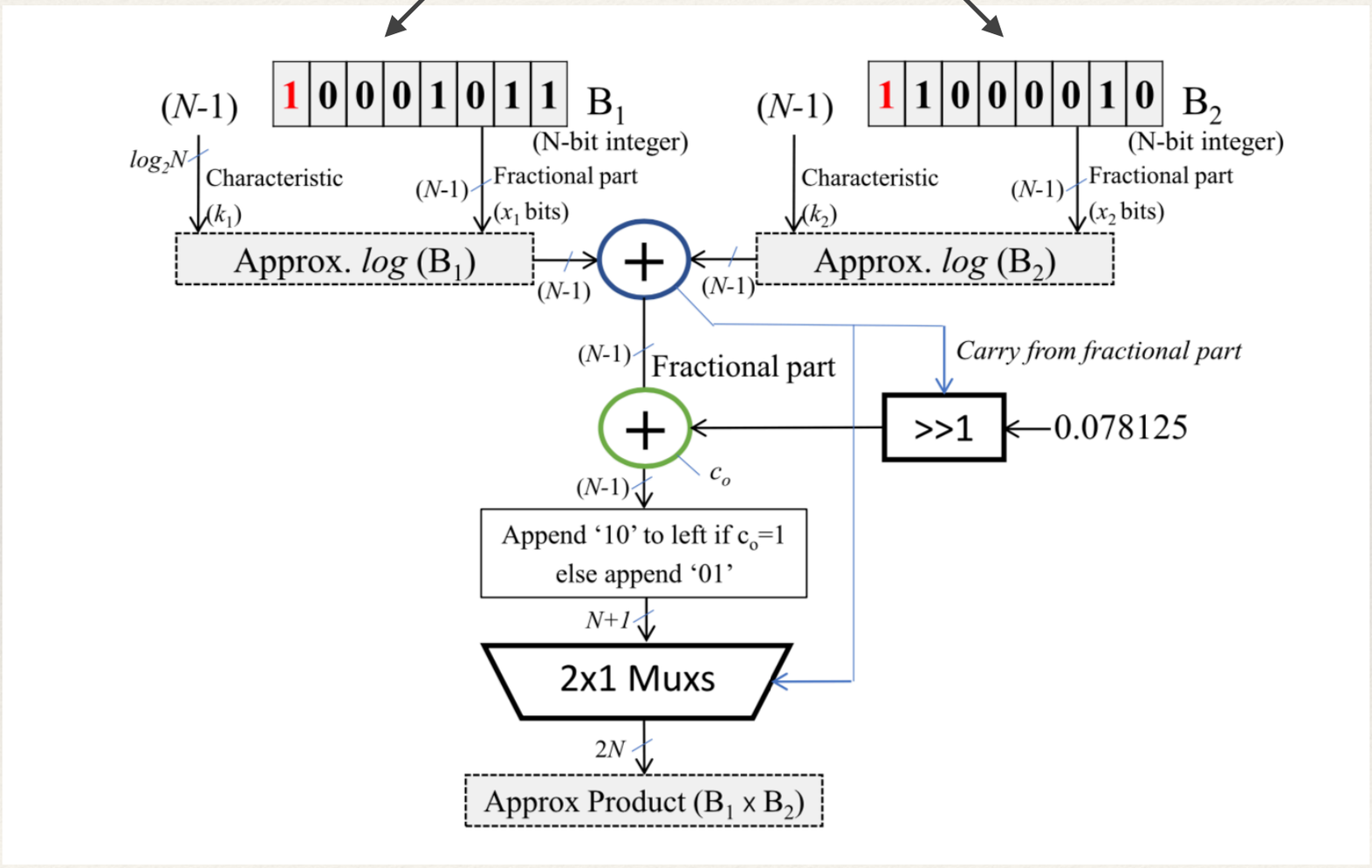


Fig. 6. Optimized MBM multiplier for FP arithmetic. The resource-hungry LODs and barrel shifters are eliminated.

H. Saadat (2018)

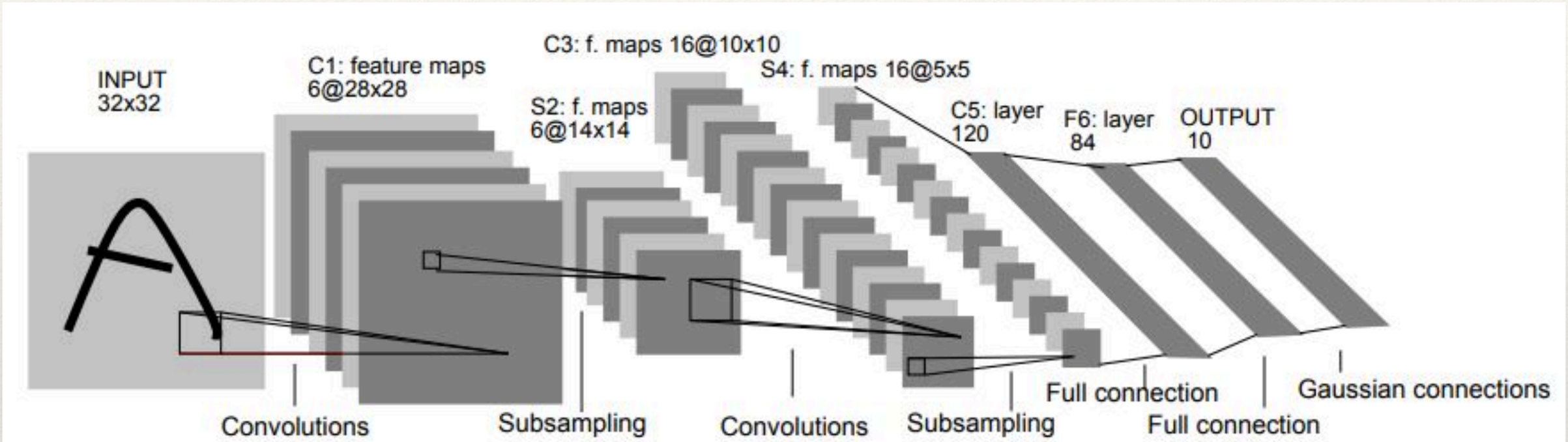
MBM precision scaling



H. Saadat, H. Bokhari and S. Parameswaran (2018)

Lenet-5 RISC-V simulation

Lenet-5



(LeCun et al., 1998)

Lenet-5

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	32x32	-	-	-
1	Convolution	6	28x28	5x5	1	tanh
2	Average Pooling	6	14x14	2x2	2	tanh
3	Convolution	16	10x10	5x5	1	tanh
4	Average Pooling	16	5x5	2x2	2	tanh
5	Convolution	120	1x1	5x5	1	tanh
6	FC	-	84	-	-	tanh
Output	FC	-	10	-	-	softmax

(Muhammad Rizwan 2018)

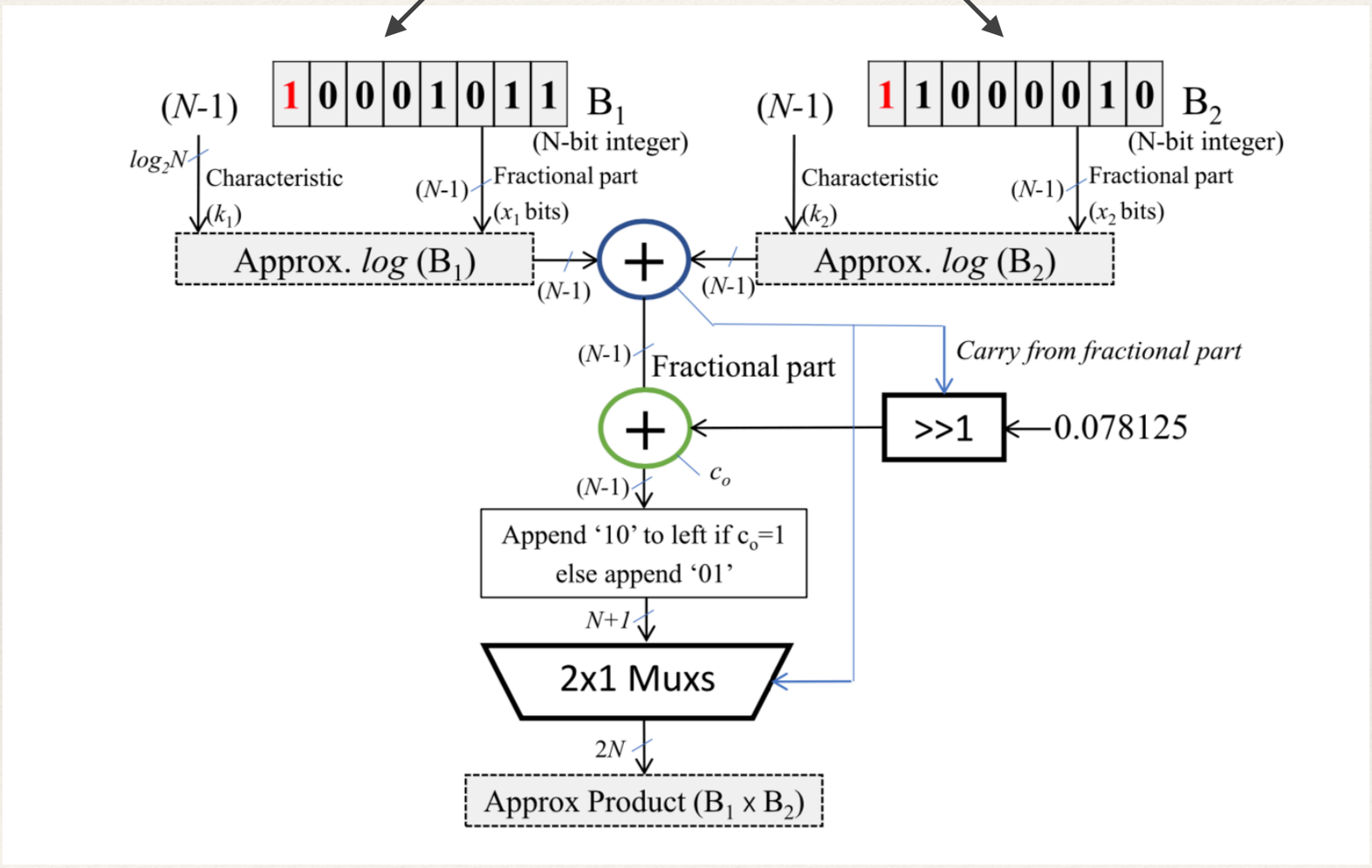
Simulation Result

Reference

Test case

MBM hardware waveform simulation

MBM



H. Saadat, H. Bokhari and S. Parameswaran (2018)

Simulation Result

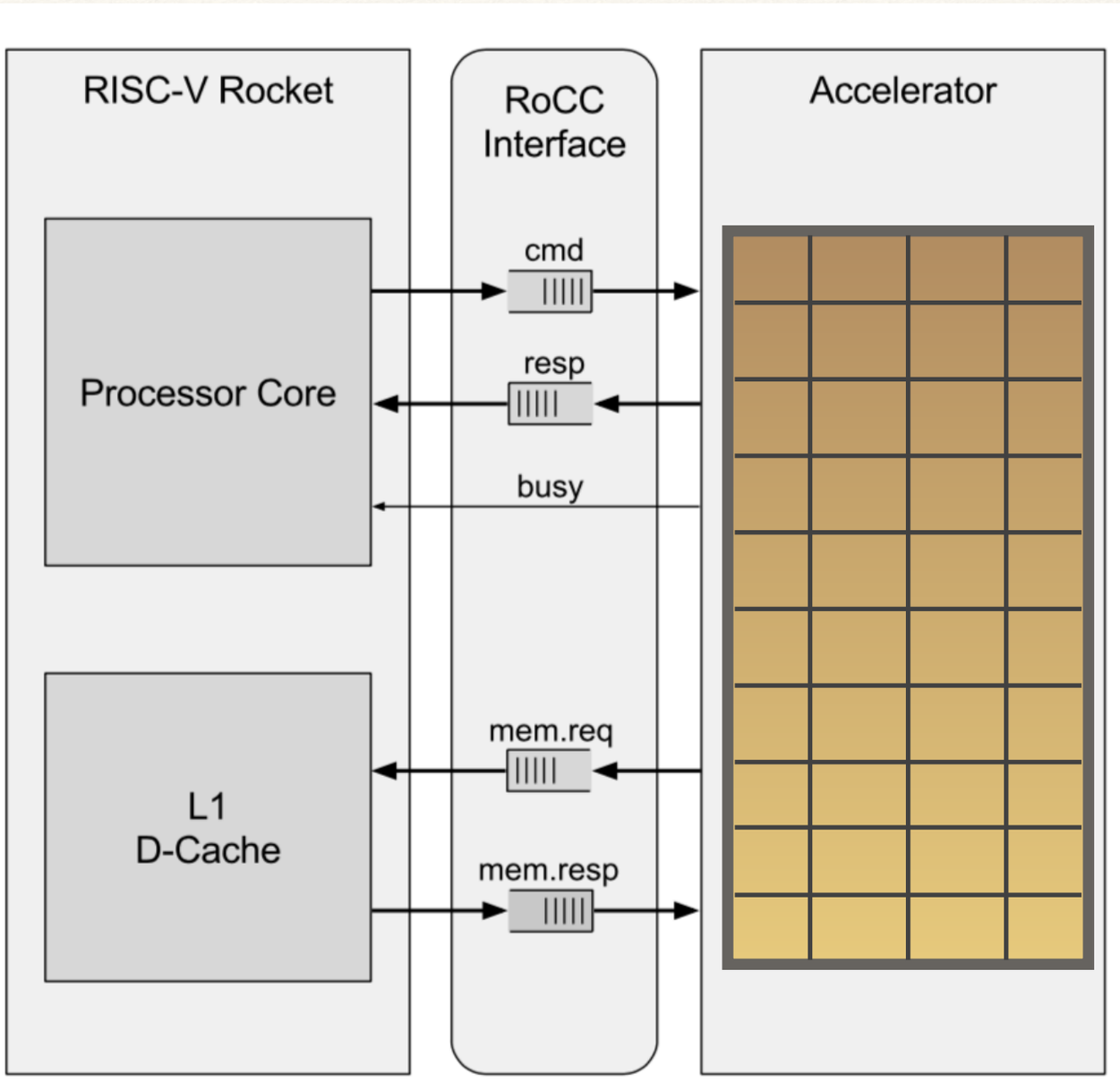
Reference

Test case

Problems and solution

RoCC

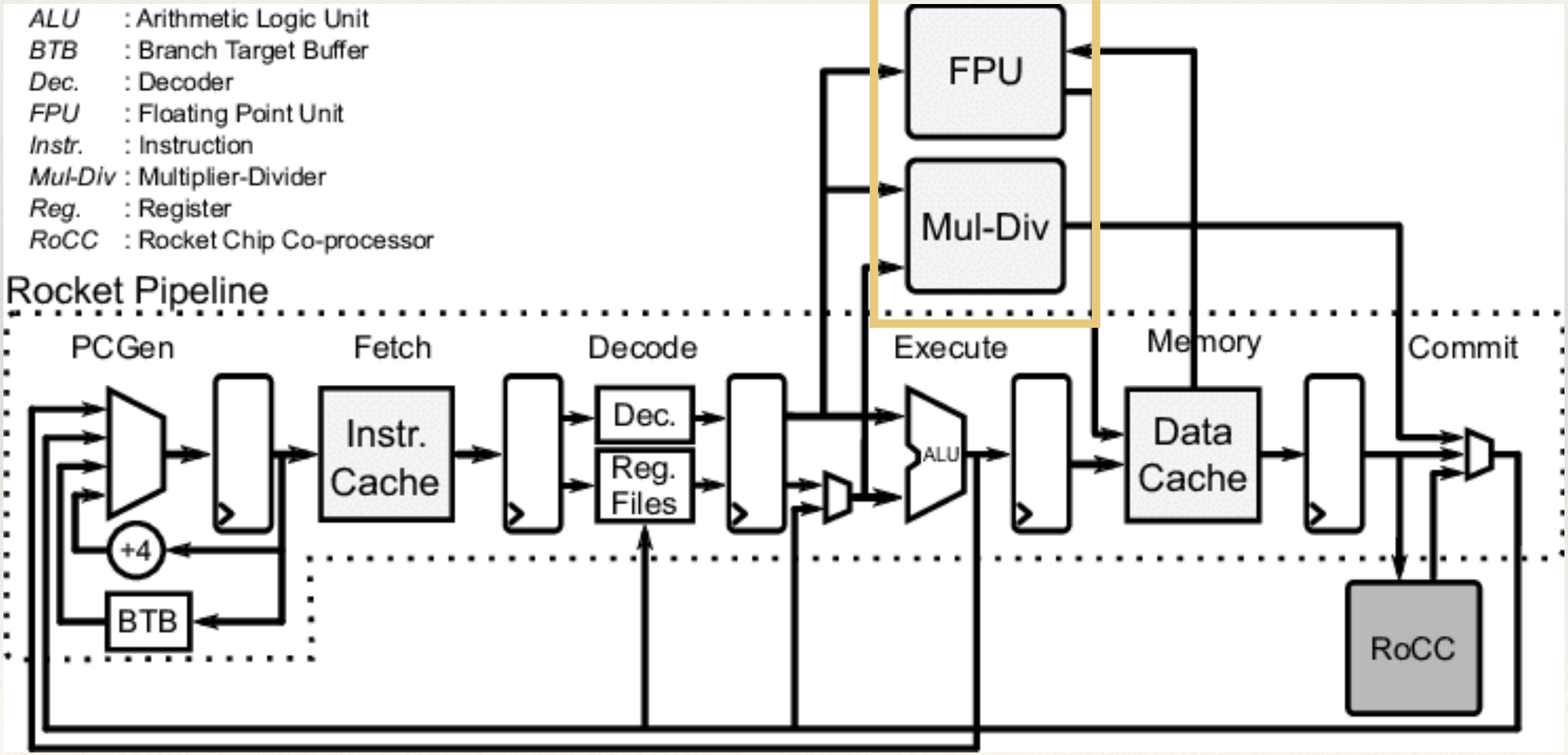
- Black-box API abandoned due to the complexity
- Replace FPU with accelerator in generated Verilog



Simplified View of RoCC

System View

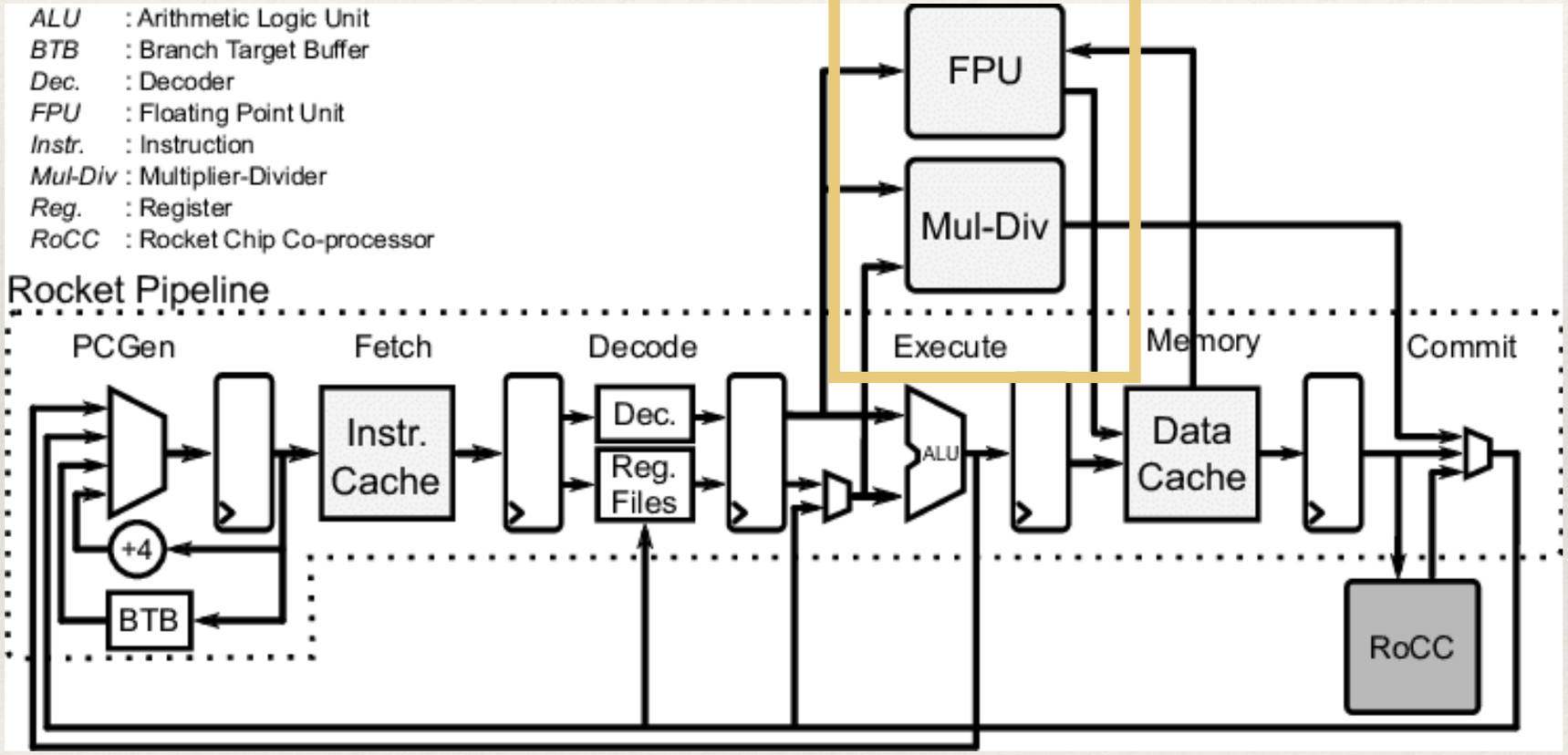
replaced by Single MBM



System Architecture

System View

replaced by



MBM Matrix

Power

Comparing MBM. & single precision accurate multiplier

	Power (uW/MBM)	Area (um^2)
MBM	21.6	288.4
Single precision	35.3	445.0
Reduction(%)	35.2	38.7

Thank you!

Reference

- Ajayi, T., Al-Hawaj, K., Amarnath, A., Dai, S., Davidson, S., Gao, P., Liu, G., Rao, A., Rovinski, A., Sun, N., Torng, C., Vega, L., Veluri, B., Xie, S., Zhao, C., Zhao, R., Batten, C., Dreslinski, R.G., Gupta, R.K., Taylor, M.B., & Zhang, Z. (2017). Experiences Using the RISC-V Ecosystem to Design an Accelerator-Centric SoC in TSMC 16 nm.
- James Martin (2017). RISC-V, Rocket, and RoCC
- H. Saadat, H. Bokhari and S. Parameswaran, "Minimally Biased Multipliers for Approximate Integer and Floating-Point Multiplication," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 11, pp. 2623-2635, Nov. 2018.
- Jouppi et al.(2017)]{2017arXiv170404760J} Jouppi, N.~P., Young, C., Patil, N., et al.\ 2017, arXiv e-prints, arXiv:1704.04760