

## [HW4\_prob1] Resnet-20 Post-training Quantization (10pts)

- Apply quantization for all the conv's weights by applying the functions in HW3\_prob1's Resnet-20 by using following for loop.

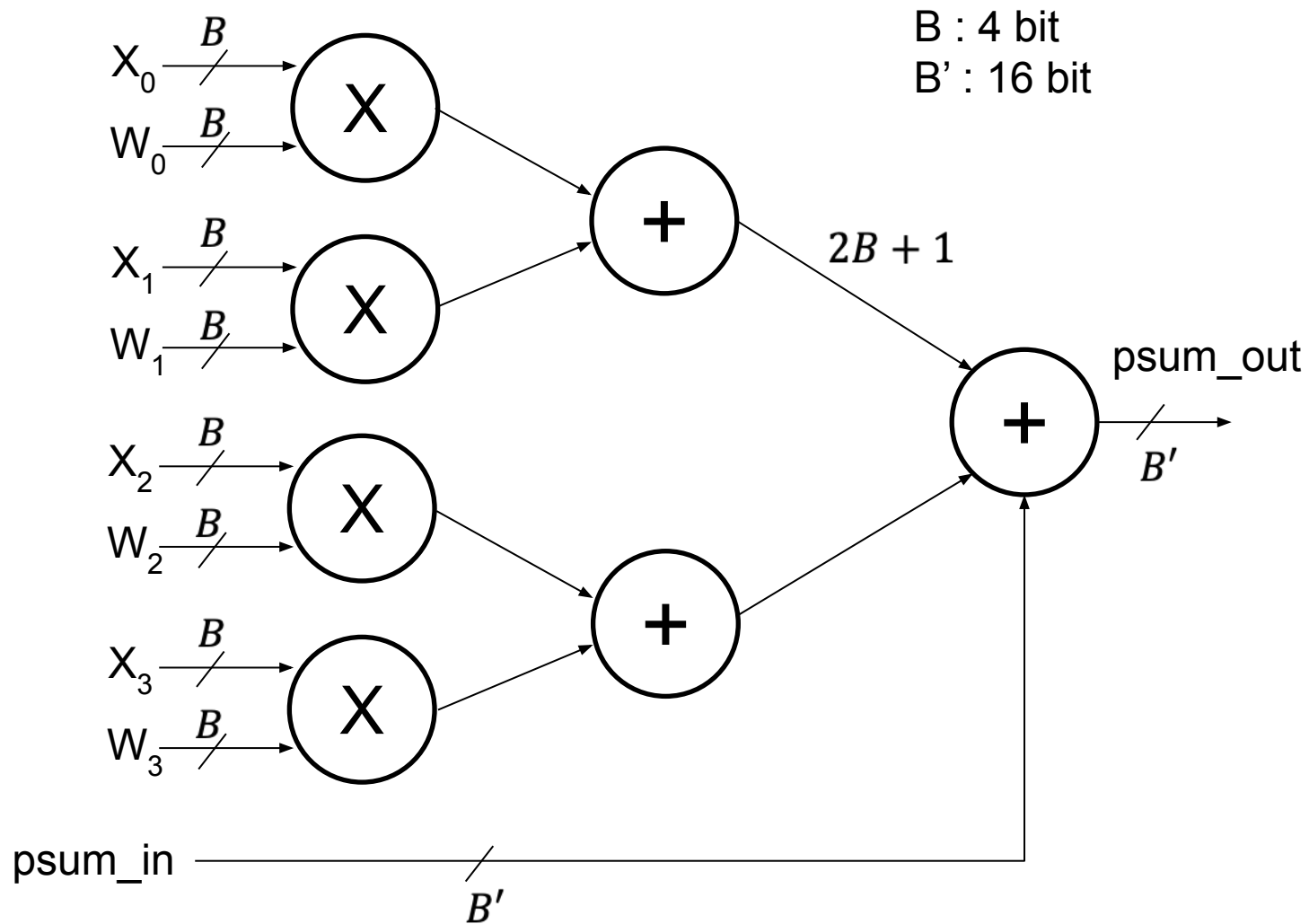
```
for layer in model.modules():  
    if isinstance(layer, torch.nn.Conv2d):
```

- Try 4 bit and 8 bit quantizations
- Report your observation
- Download the notebook as pdf and submit it

## [HW4\_prob2] MAC Design (A)(10pts)

- Weight: 4-bit signed
- Activation: 4-bit **unsigned**
- psum / output: 16-bit
  
- Please fill out the empty part in hw/w3/prob3/verilog to support above number representation
- Need to create another “dec2bin” function in testbench for unsigned number (w\_bin and x\_bin)
- Need to modify “mac\_predicted” function accordingly for sign \* unsign multiplication
- Now, a\_data.txt includes only positive numbers.
  
- Include your \*.v files and capture the final waveform to show output

## [HW4\_prob3] MAC Design (B)(10pts)



- Now extend your previous solution to support the left schematic
- Again, your `psum_out` bit precision is 16 bit
- Use the first four numbers in `a_data.txt` as `x0`, `x1`, `x2`, `x3`. Then, the next four numbers are `x0`, `x1`, `x2`, `x3` for the following cycle.
- You need to modify both `_tb.v` and other `*.v` files accordingly.
- Include your `*.v` files and capture the final waveform to show output
- all the multiplication and accumulation should happen in 1 cycle.