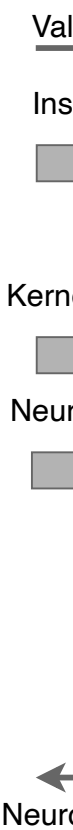
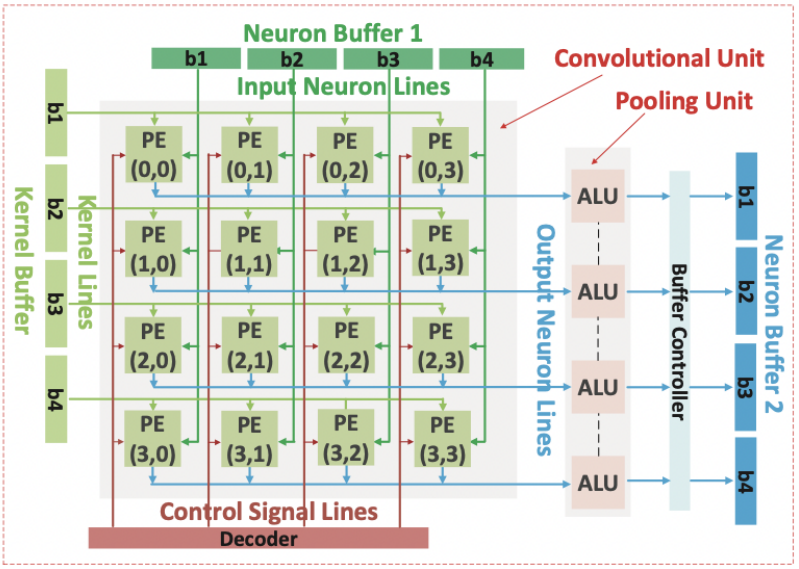
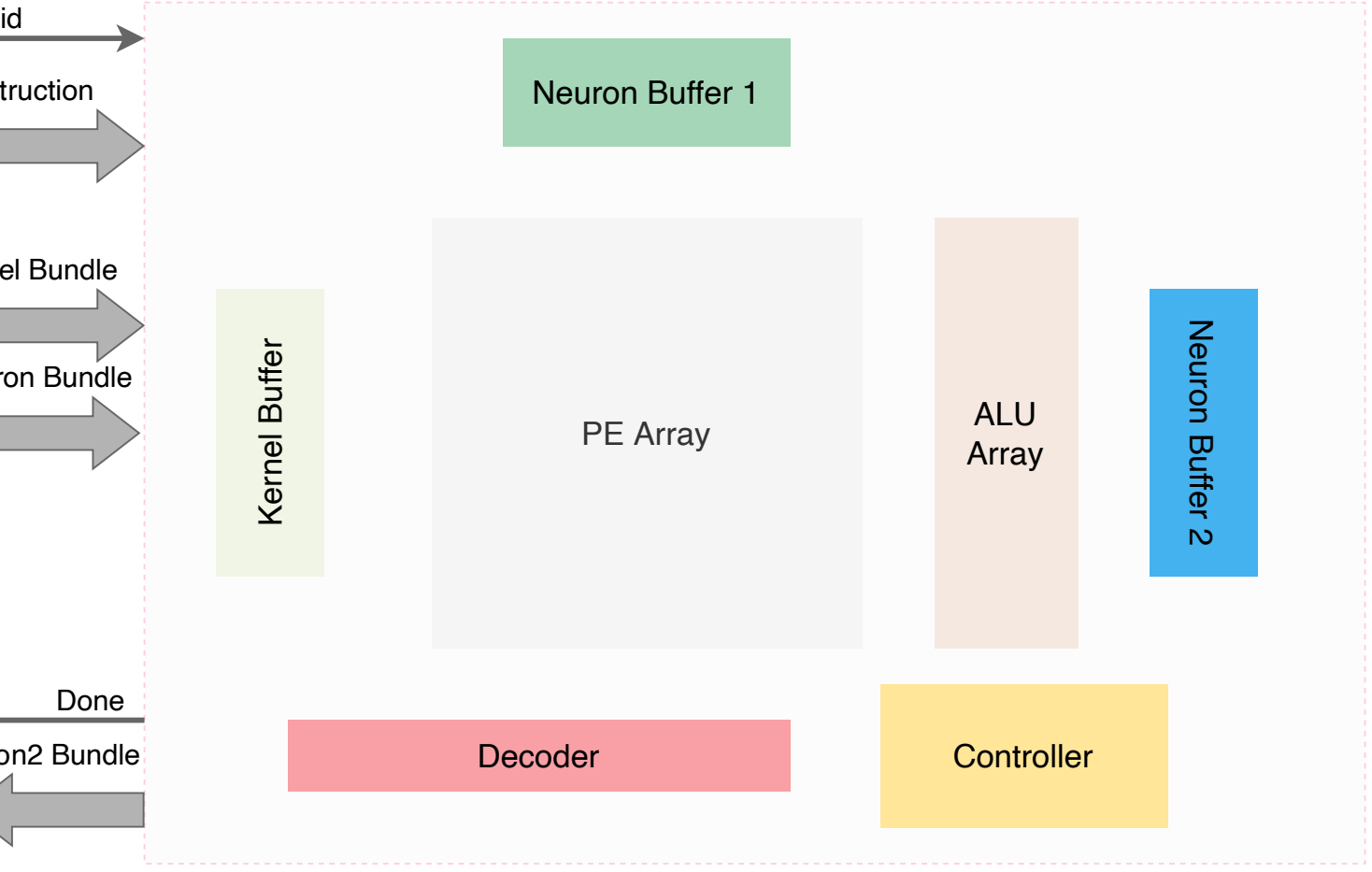
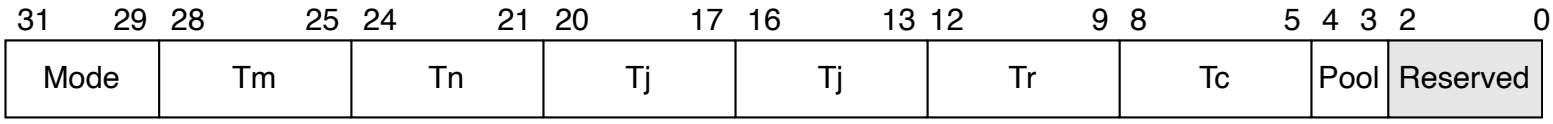

Val
Ins
Kern
Neur
Neuro


Neuron Buffer 1
b1
b2
b3
b4
Input Neuron Lines
Convolutional Unit
Pooling Unit
Kernel Buffer
Kernel Lines
PE (0,0)
PE (0,1)
PE (0,2)
PE (0,3)
PE (1,0)
PE (1,1)
PE (1,2)
PE (1,3)
PE (2,0)
PE (2,1)
PE (2,2)
PE (2,3)
PE (3,0)
PE (3,1)
PE (3,2)
PE (3,3)
ALU
ALU
ALU
ALU
Output Neuron Lines
Buffer Controller
Neuron Buffer 2
Control Signal Lines
Decoder

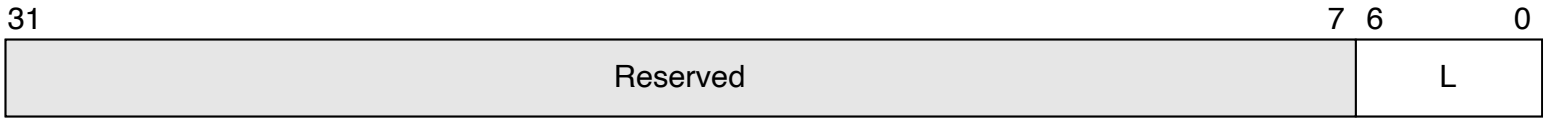

id
truction
el Bundle
ron Bundle
Done
on2 Bundle
Neuron Buffer 1
Kernel Buffer
PE Array
ALU Array
Neuron Buffer 2
Decoder
Controller

Command and Parameters



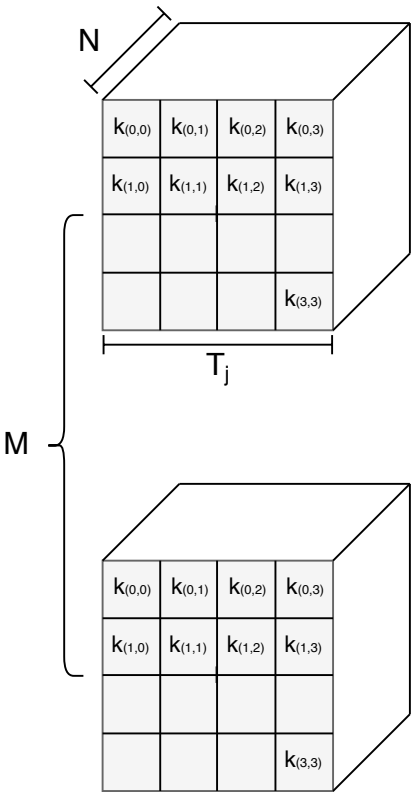
Mode: Nop
Load 1: Load kernel and neuron data from off-chip memory to on-chip outer buffer.
Load 2: Load kernel and neuron data from outer buffer to local store in PE.
Clear: Clear all settings and data.

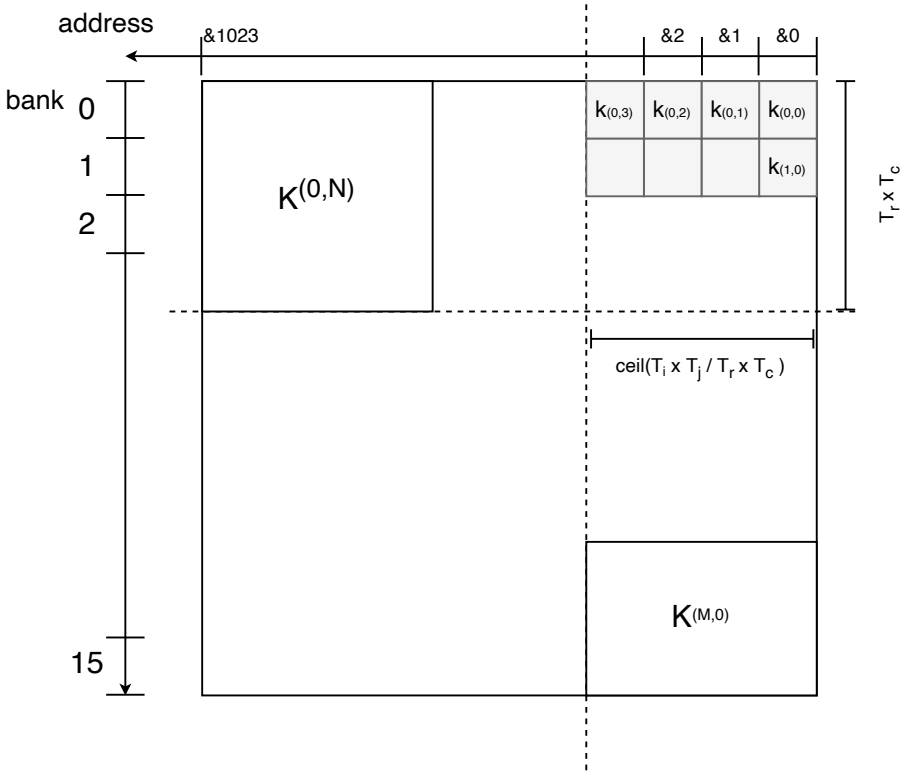
Size



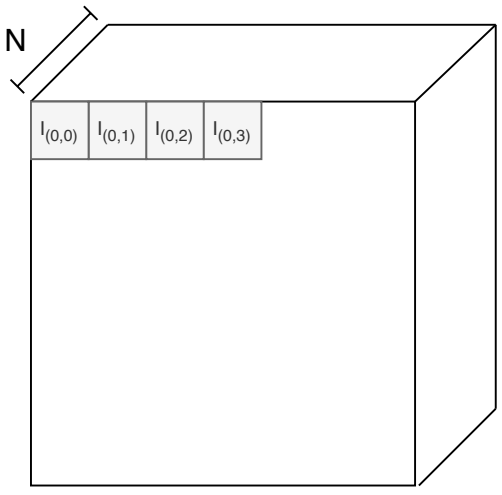
Neuron Size

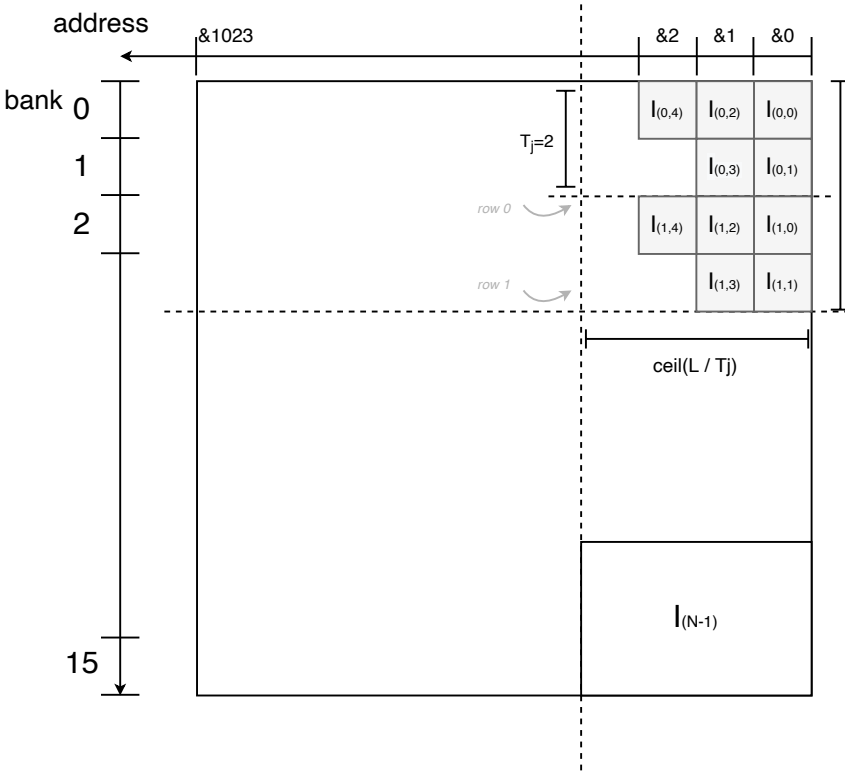
Kernel Buffer

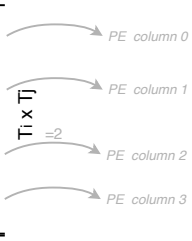




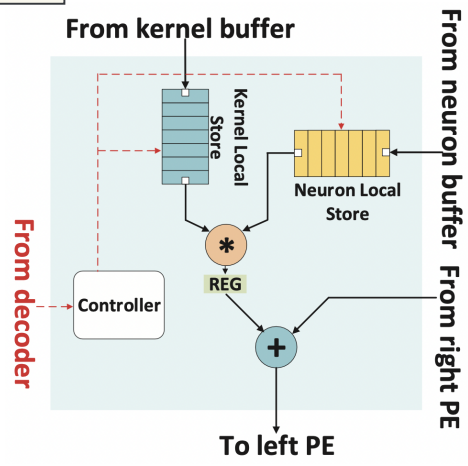
Neuron Buffer 1

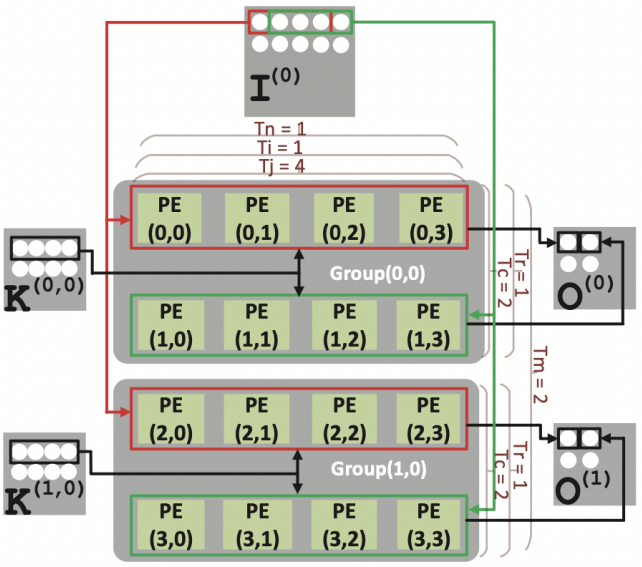





PE column 0
PE column 1
Ti x Tj
=2
PE column 2
PE column 3

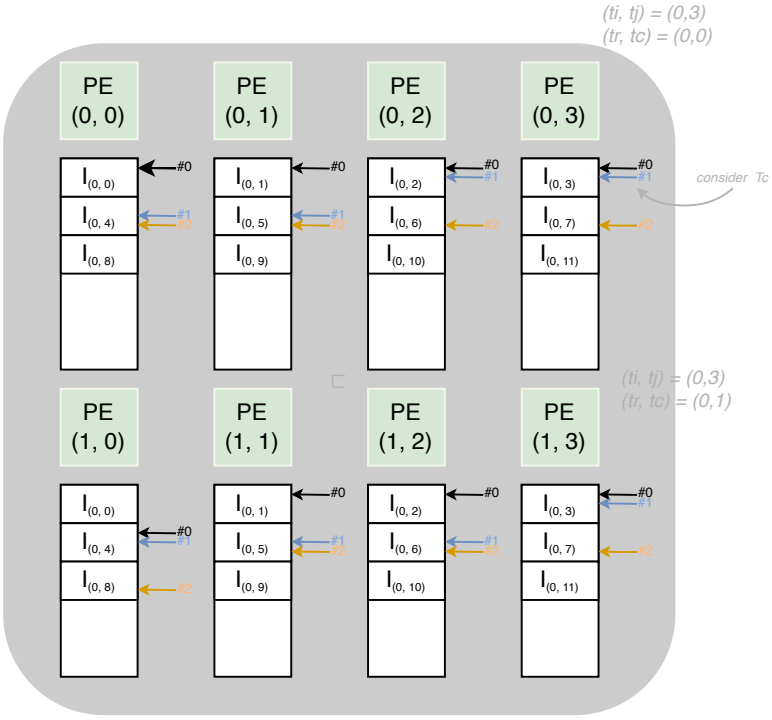
PE

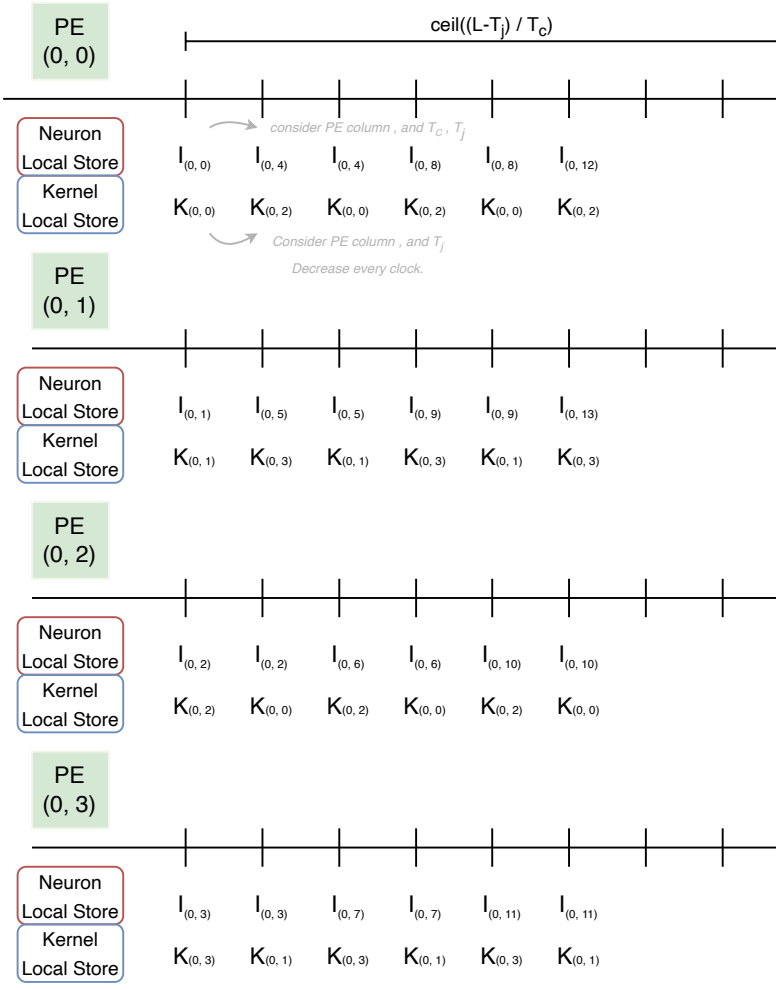




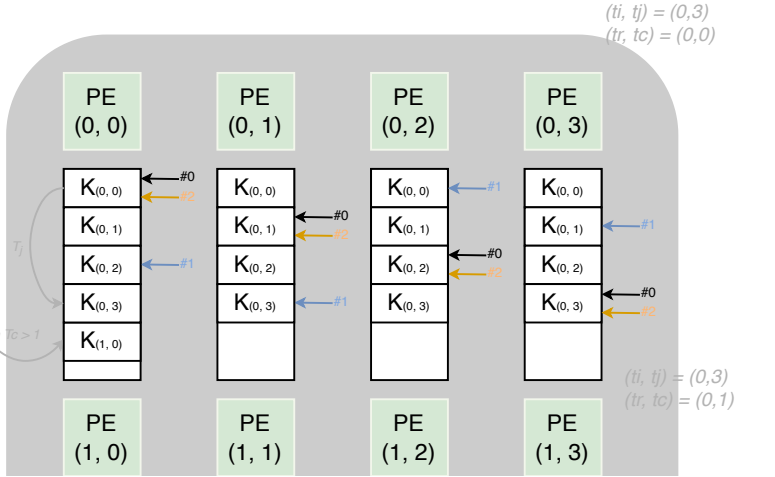
Example:
(Tm, Tn, Ti, Tj, Tr, Tc) = (2, 1, 1, 4, 1, 2)

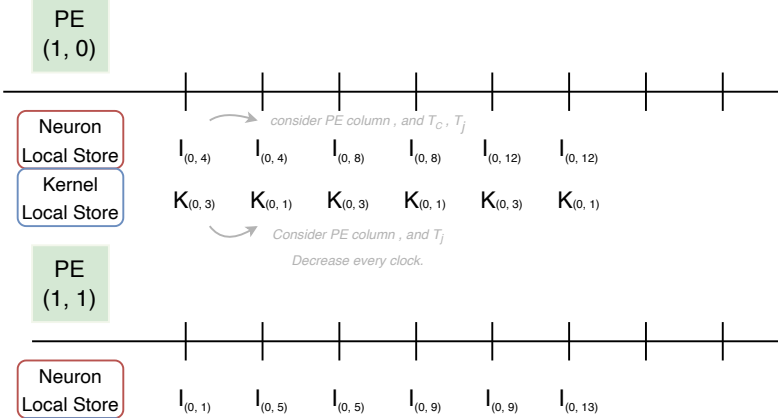
Neuron Local Store

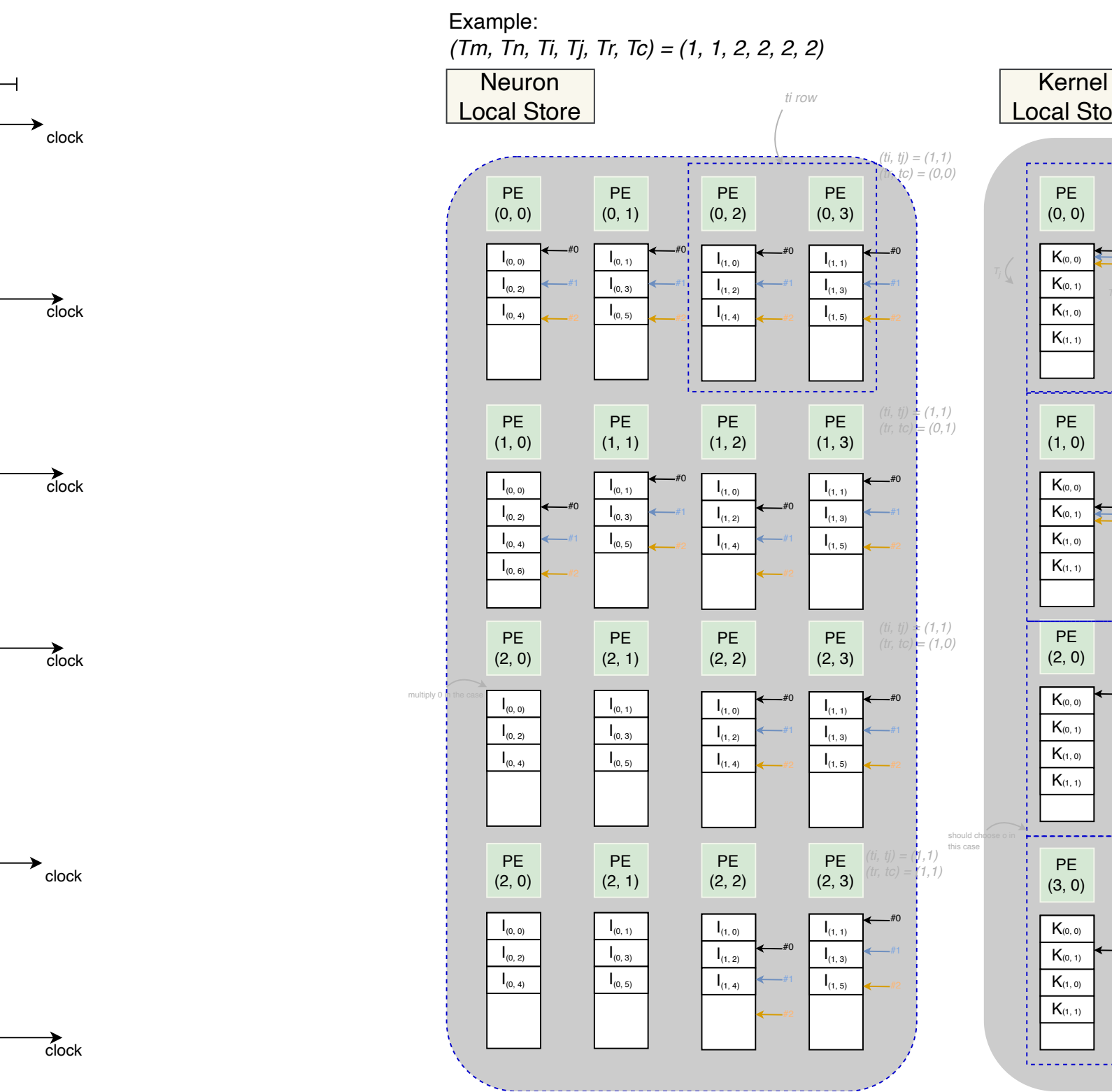




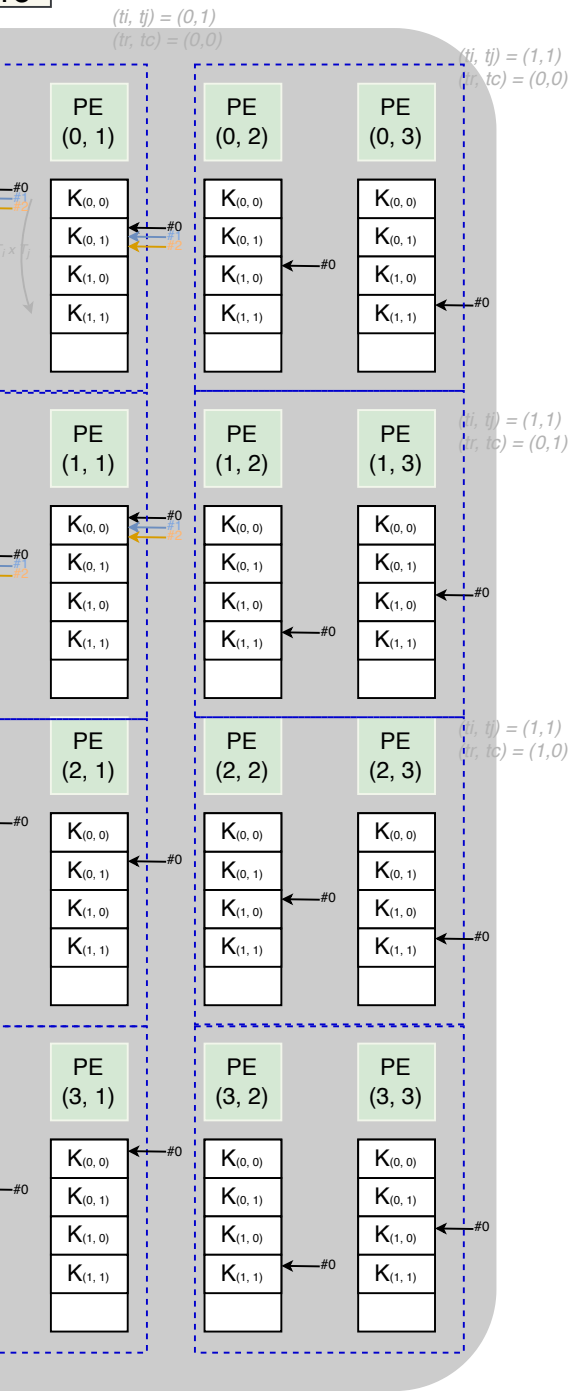
Kernel Local Store

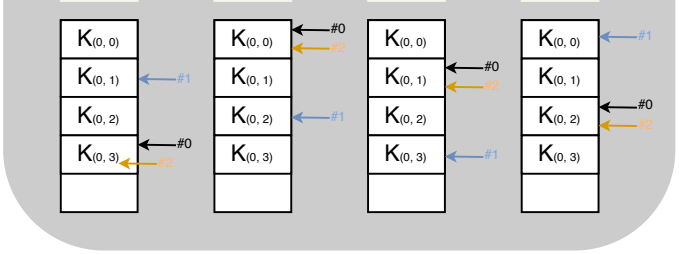


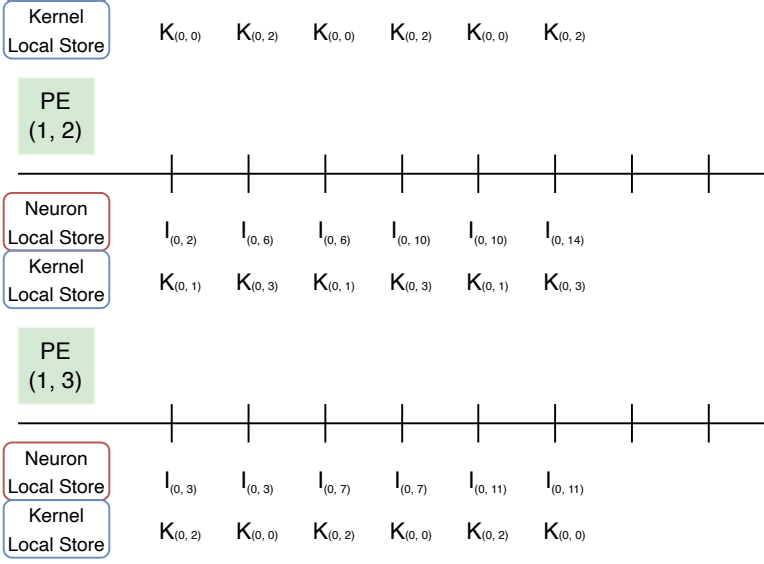



Example:
(Tm, Tn, Ti, Tj, Tr, Tc) = (1, 1, 2, 2, 2, 2)
Neuron Local Store
Kernel Local Sto
ti row
(ti, tj) = (1,1)
(tr, tc) = (0,0)
(ti, tj) = (1,1)
(tr, tc) = (0,1)
(ti, tj) = (1,1)
(tr, tc) = (1,0)
(ti, tj) = (1,1)
(tr, tc) = (1,1)
PE (0, 0)
PE (0, 1)
PE (0, 2)
PE (0, 3)
PE (1, 0)
PE (1, 1)
PE (1, 2)
PE (1, 3)
PE (2, 0)
PE (2, 1)
PE (2, 2)
PE (2, 3)
PE (3, 0)
#0
#1
#2
clock
clock
clock
clock
clock
clock

re


(ti, tj) = (0,1)
(tr, tc) = (0,0)
(ti, tj) = (1,1)
(tr, tc) = (0,0)
(ti, tj) = (1,1)
(tr, tc) = (0,1)
(ti, tj) = (1,1)
(tr, tc) = (1,0)
PE (0, 1)
PE (0, 2)
PE (0, 3)
PE (1, 1)
PE (1, 2)
PE (1, 3)
PE (2, 1)
PE (2, 2)
PE (2, 3)
PE (3, 1)
PE (3, 2)
PE (3, 3)
K(0, 0)
K(0, 1)
K(1, 0)
K(1, 1)
#0
#1
#2

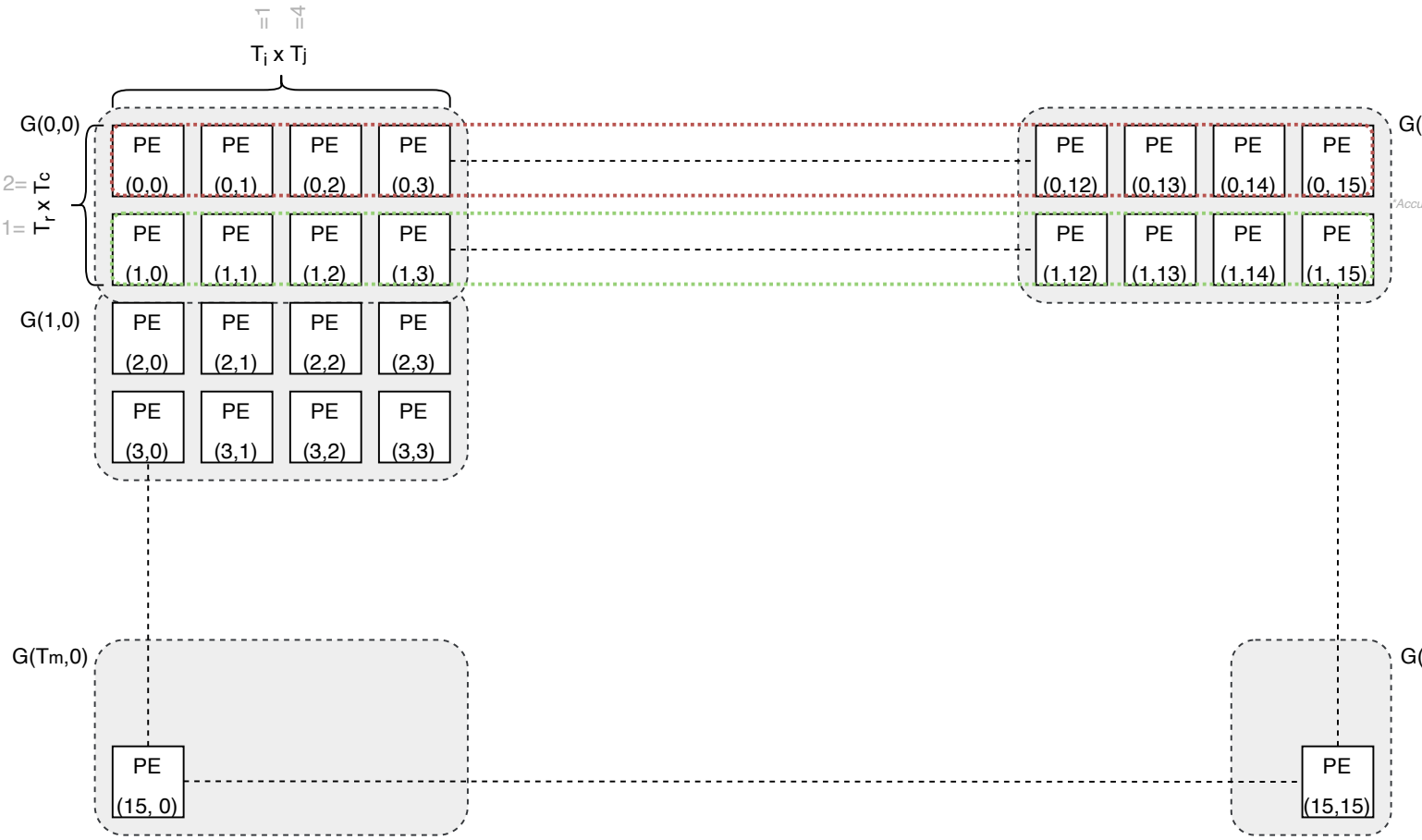


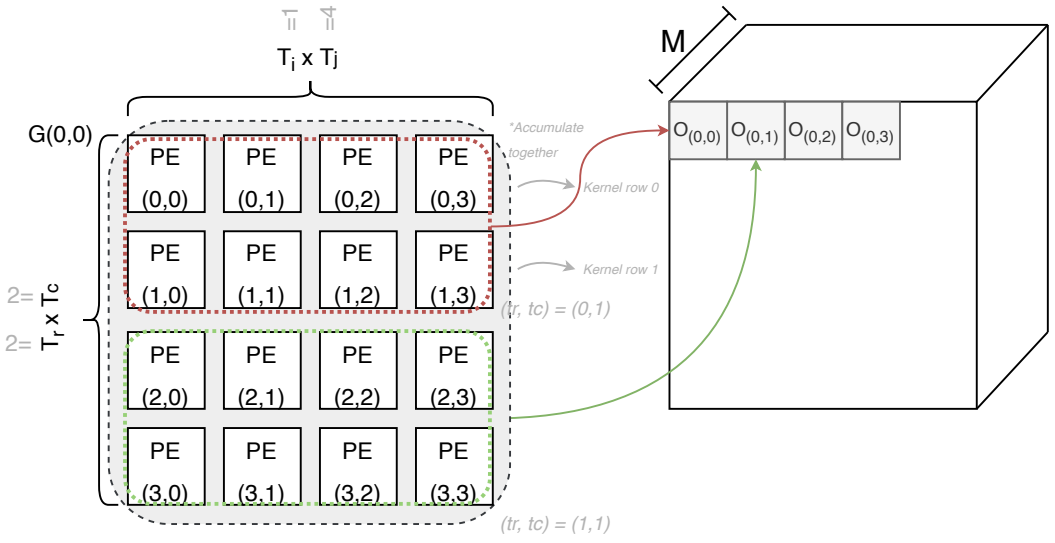


Algorithm Kernel Control

Ini

clock
clock

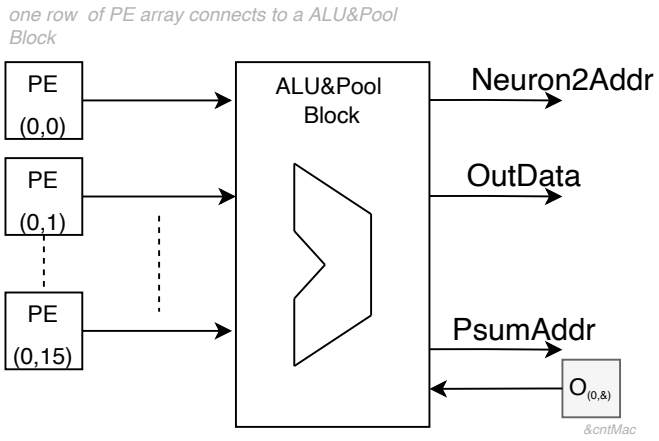

=1
=4
Ti x Tj
G(0,0)
2=
Tr x Tc
1=
PE (0,0)
PE (0,1)
PE (0,2)
PE (0,3)
PE (0,12)
PE (0,13)
PE (0,14)
PE (0, 15)
PE (1,0)
PE (1,1)
PE (1,2)
PE (1,3)
PE (1,12)
PE (1,13)
PE (1,14)
PE (1, 15)
G(1,0)
PE (2,0)
PE (2,1)
PE (2,2)
PE (2,3)
PE (3,0)
PE (3,1)
PE (3,2)
PE (3,3)
G(Tm,0)
PE (15, 0)
PE (15,15)

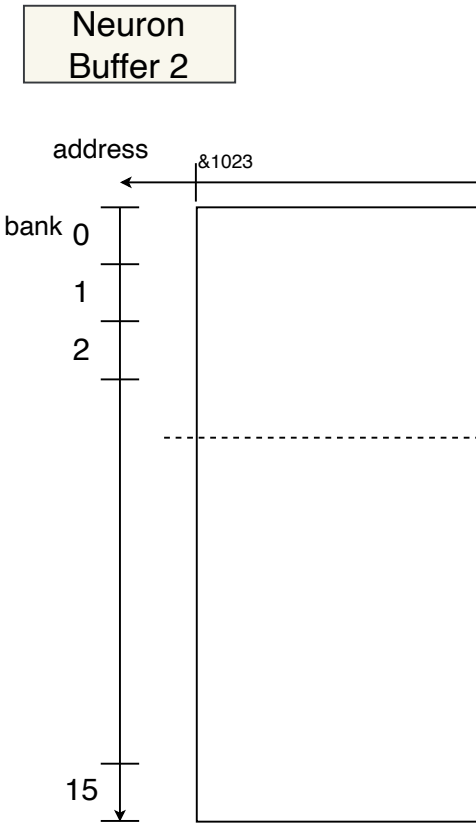

=1
=4
Ti x Tj
M
G(0,0)
2=
2=
Tr x Tc
PE (0,0)
PE (0,1)
PE (0,2)
PE (0,3)
PE (1,0)
PE (1,1)
PE (1,2)
PE (1,3)
PE (2,0)
PE (2,1)
PE (2,2)
PE (2,3)
PE (3,0)
PE (3,1)
PE (3,2)
PE (3,3)
O(0,0)
O(0,1)
O(0,2)
O(0,3)
*Accumulate together
Kernel row 0
Kernel row 1
(tr, tc) = (0,1)
(tr, tc) = (1,1)

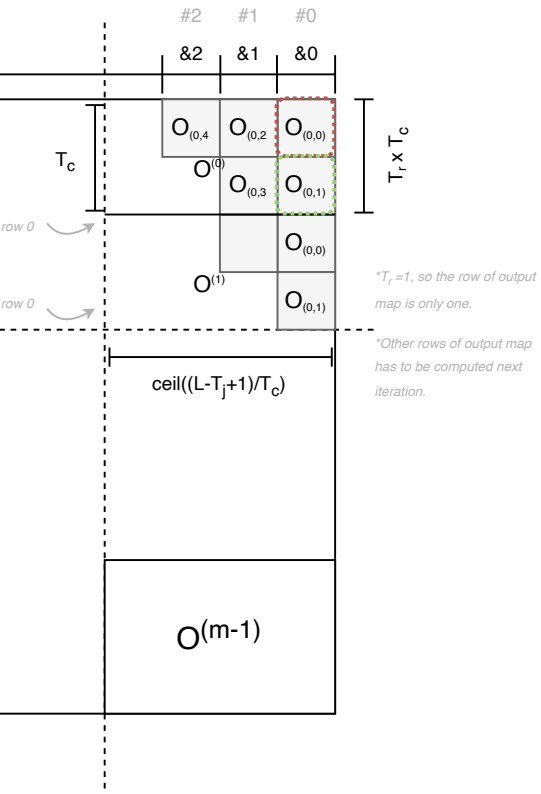
0,Tn)

mulate

(Tm,Tn)

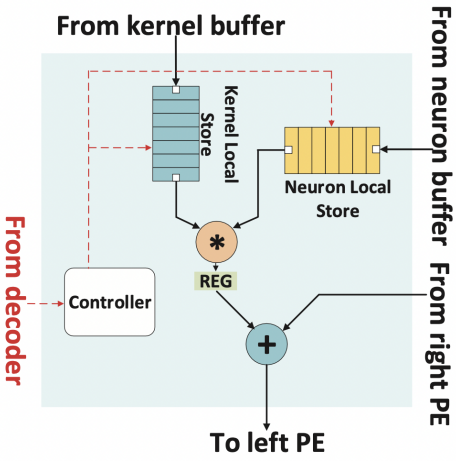


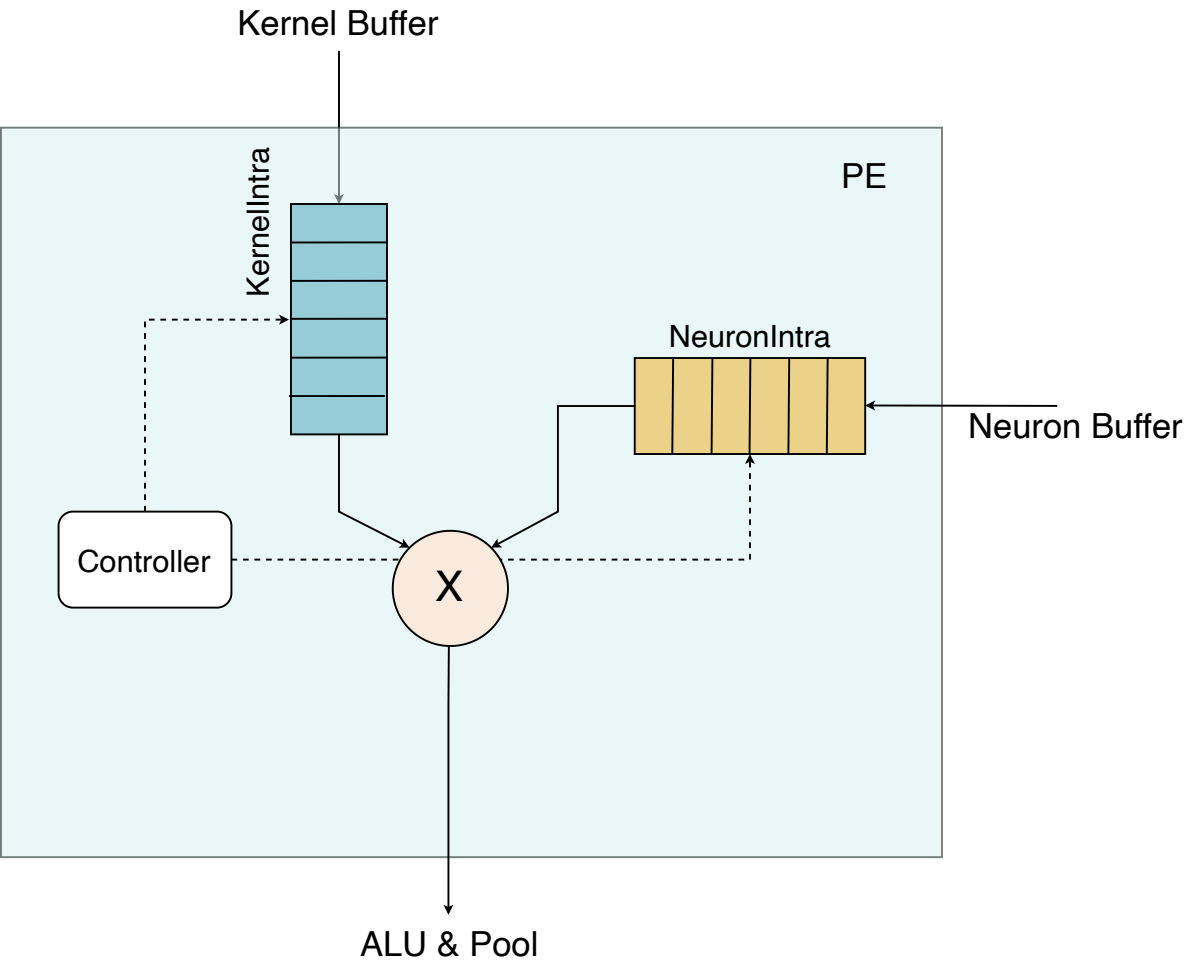


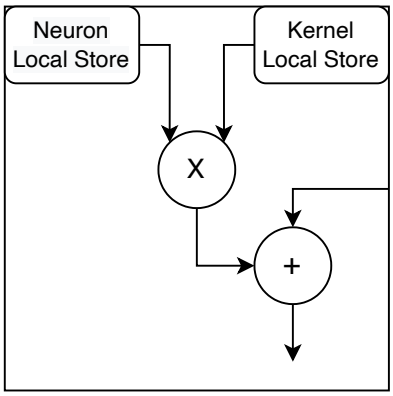


*T_r =1, so the row of output map is only one.

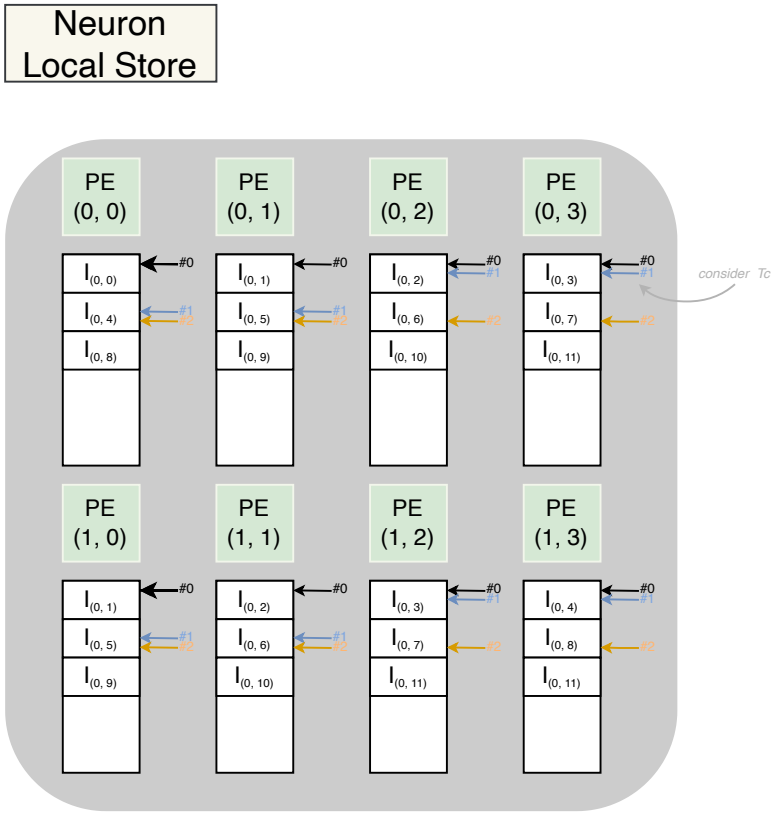
*Other rows of output map has to be computed next iteration.

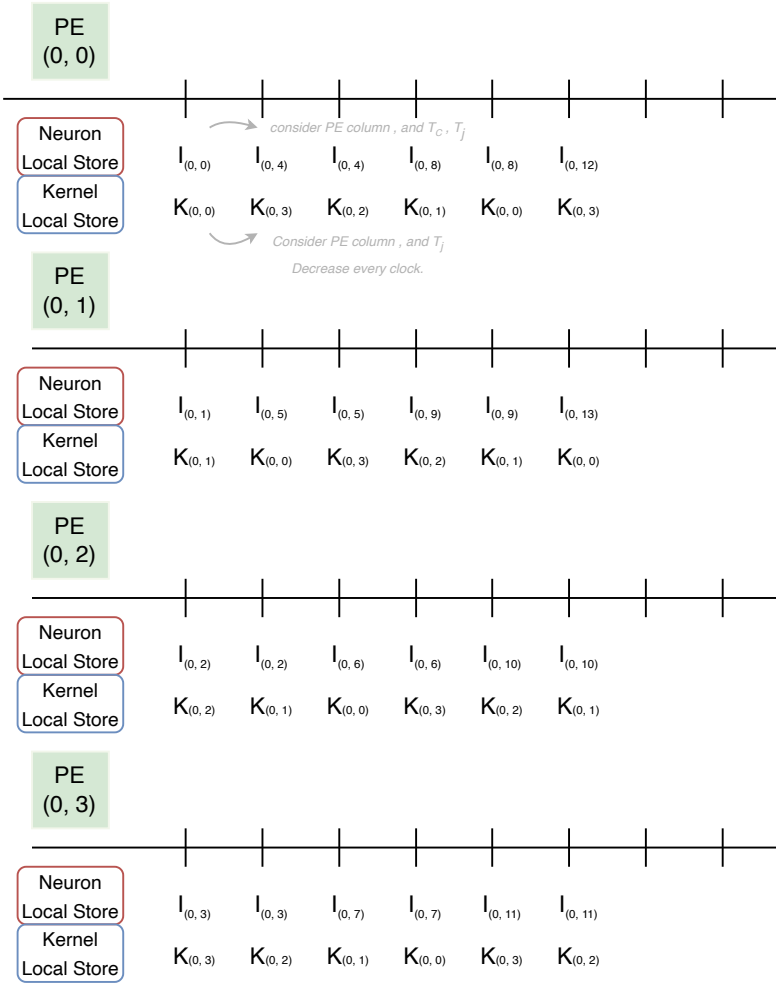

From kernel buffer
Kernel Local Store
Neuron Local Store
From neuron buffer
From right PE
From decoder
Controller
REG
To left PE

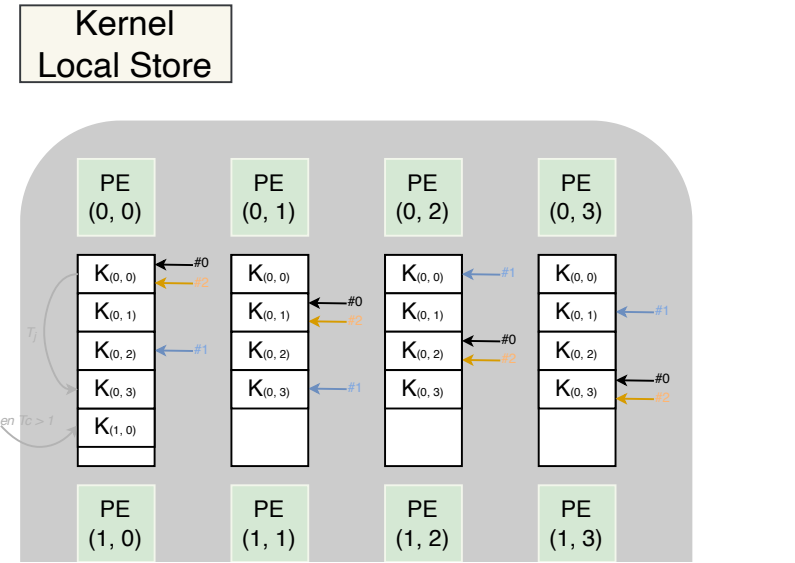

Kernel Buffer
PE
KernelIntra
NeuronIntra
Neuron Buffer
Controller
X
ALU & Pool

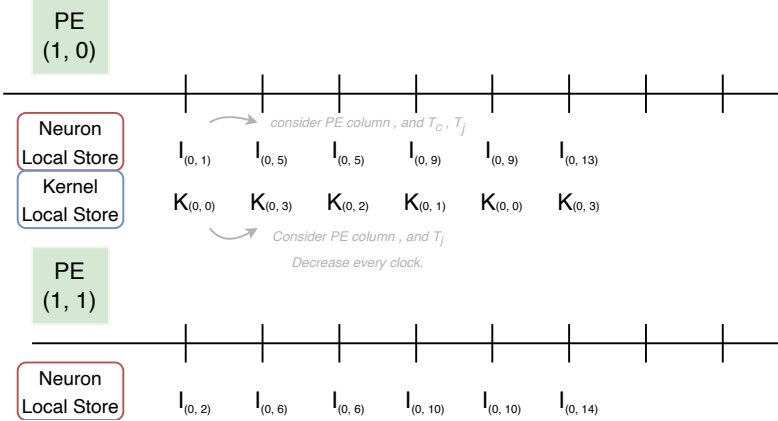


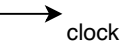
Example:
(Tm, Tn, Ti, Tj, Tr, Tc) = (2, 1, 1, 4, 1, 2)







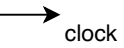



clock

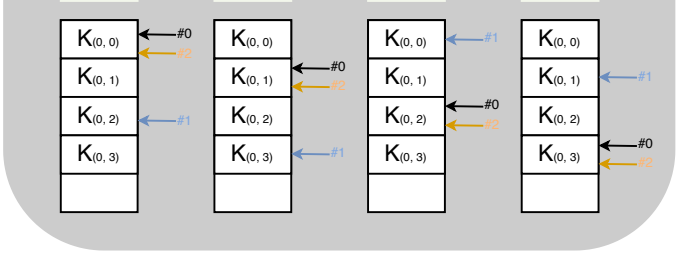
clock

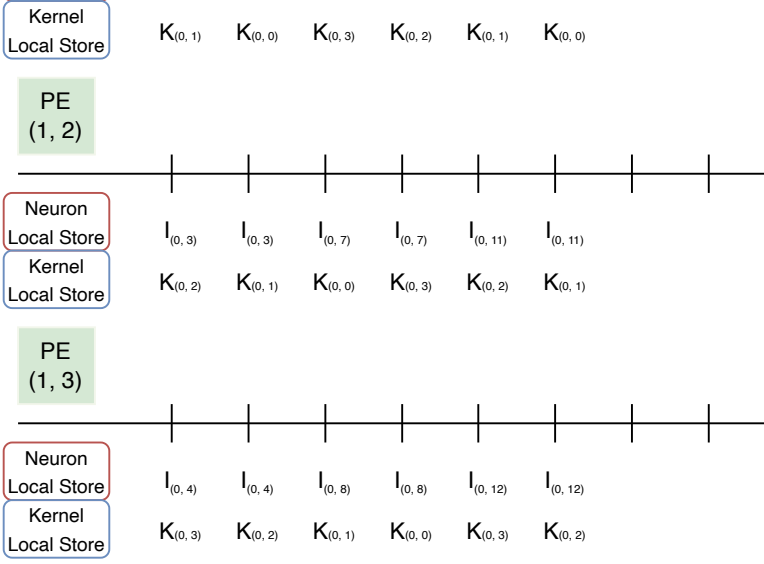
clock

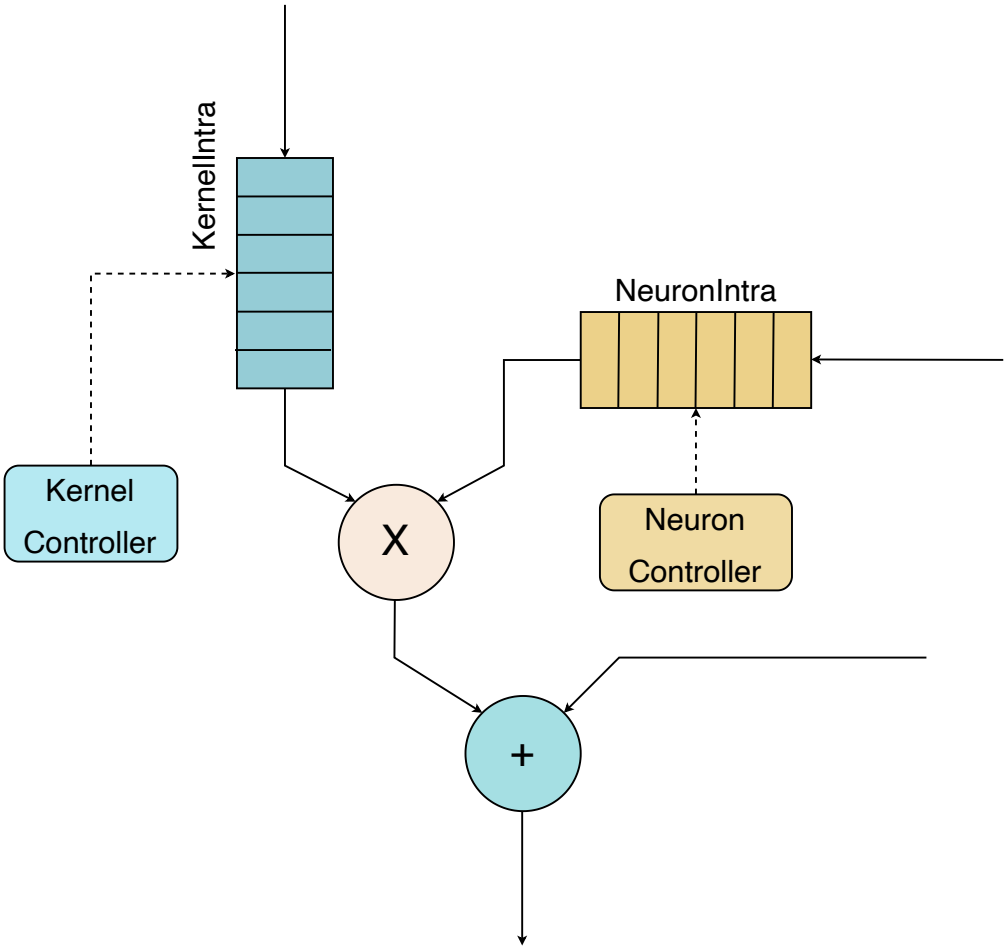
clock

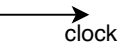

clock

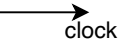
clock


K(0, 0)
K(0, 1)
K(0, 2)
K(0, 3)
#0
#1
#2


Kernel Local Store
K(0, 1) K(0, 0) K(0, 3) K(0, 2) K(0, 1) K(0, 0)
PE (1, 2)
Neuron Local Store
I(0, 3) I(0, 3) I(0, 7) I(0, 7) I(0, 11) I(0, 11)
Kernel Local Store
K(0, 2) K(0, 1) K(0, 0) K(0, 3) K(0, 2) K(0, 1)
PE (1, 3)
Neuron Local Store
I(0, 4) I(0, 4) I(0, 8) I(0, 8) I(0, 12) I(0, 12)
Kernel Local Store
K(0, 3) K(0, 2) K(0, 1) K(0, 0) K(0, 3) K(0, 2)


KernelIntra
NeuronIntra
Kernel Controller
X
Neuron Controller
+


clock


clock