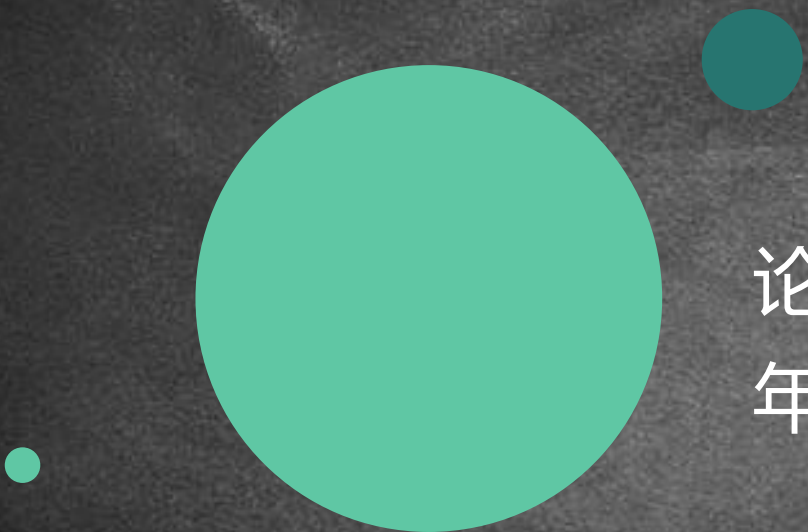


# 学习汇报

焦强



# 论文 年前学习计划





论文

2点贡献:

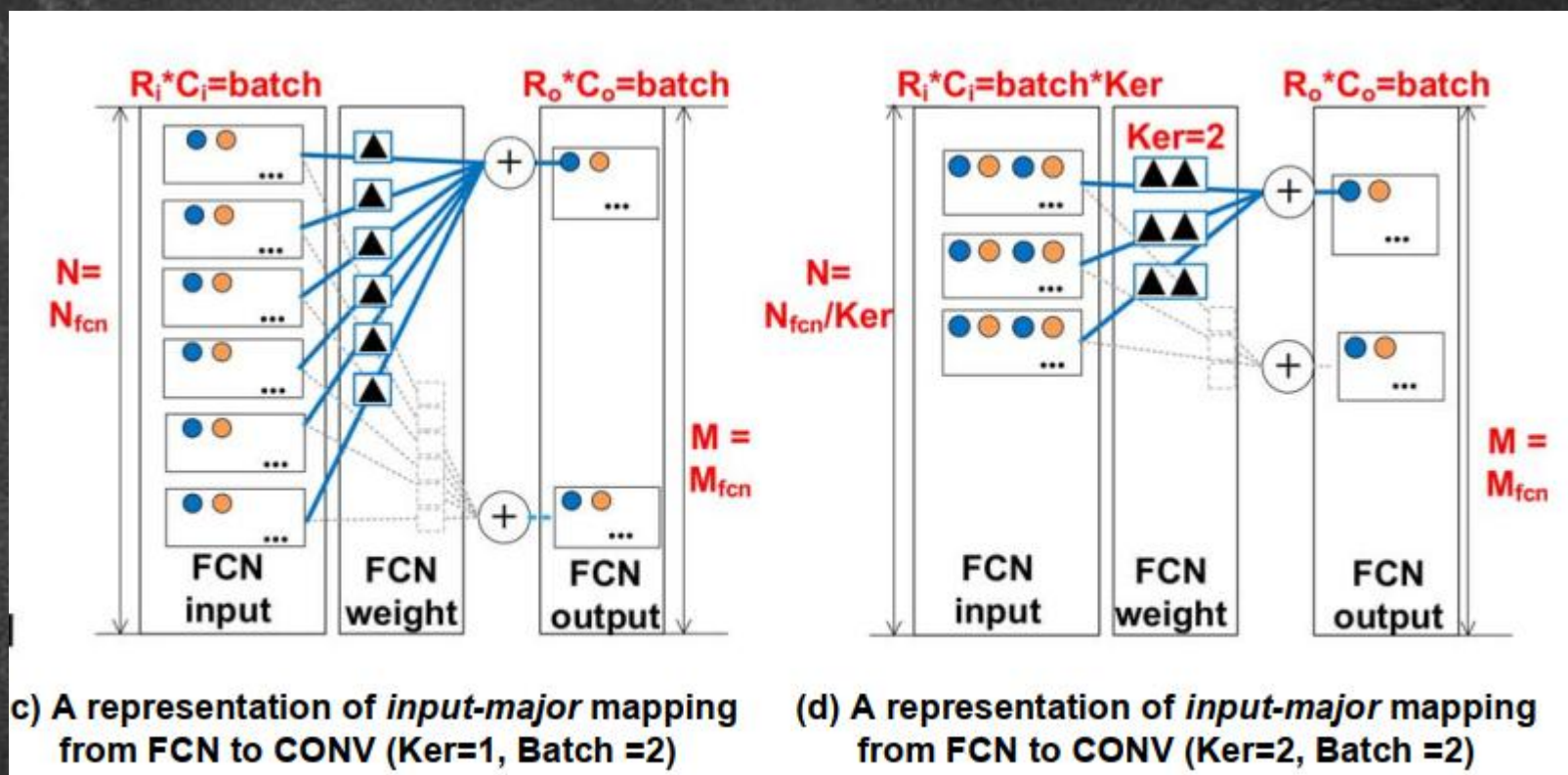
1, UNIFORMED CNN REPRESENTATION.

统一的CNN表示。

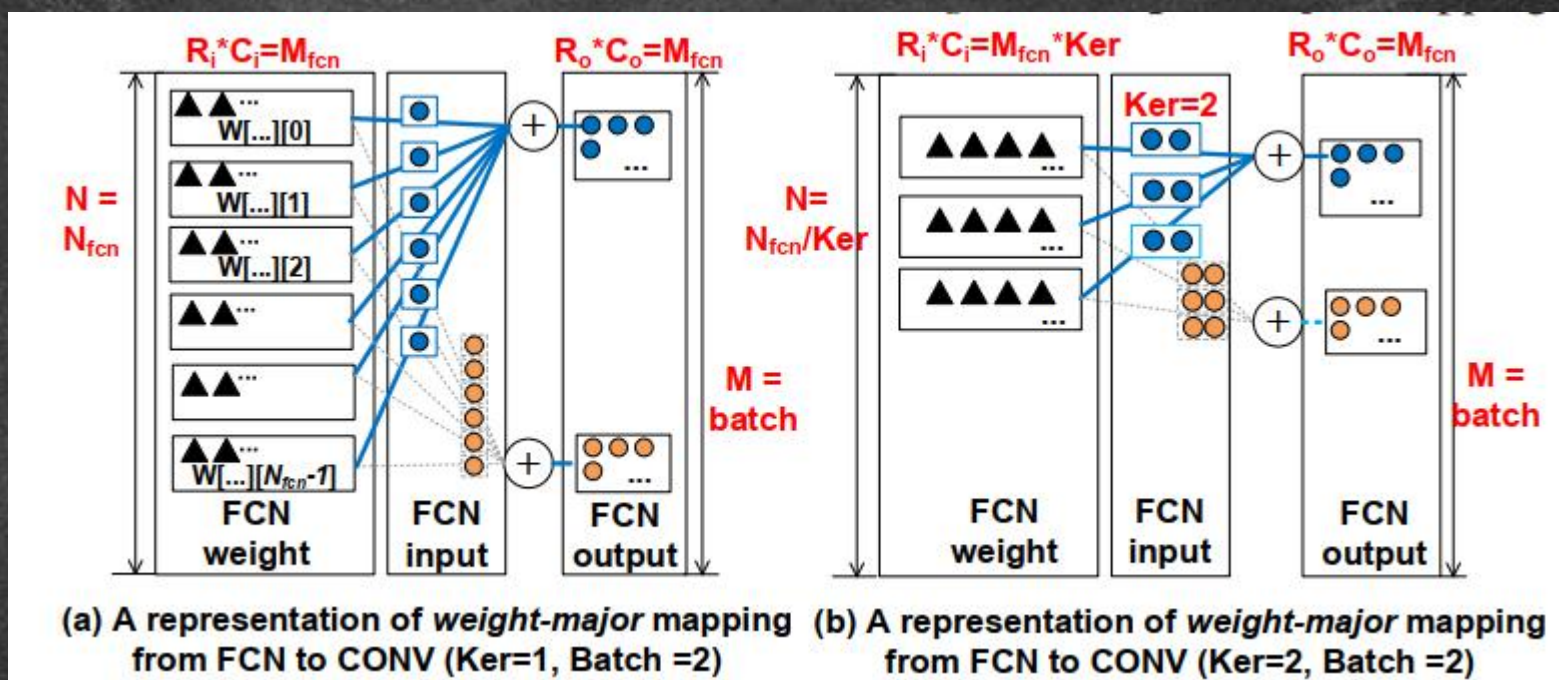
2, 通过动态随机存取存储器布局重组优化带宽。



# UNIFORMED CNN REPRESENTATION: 输入映射

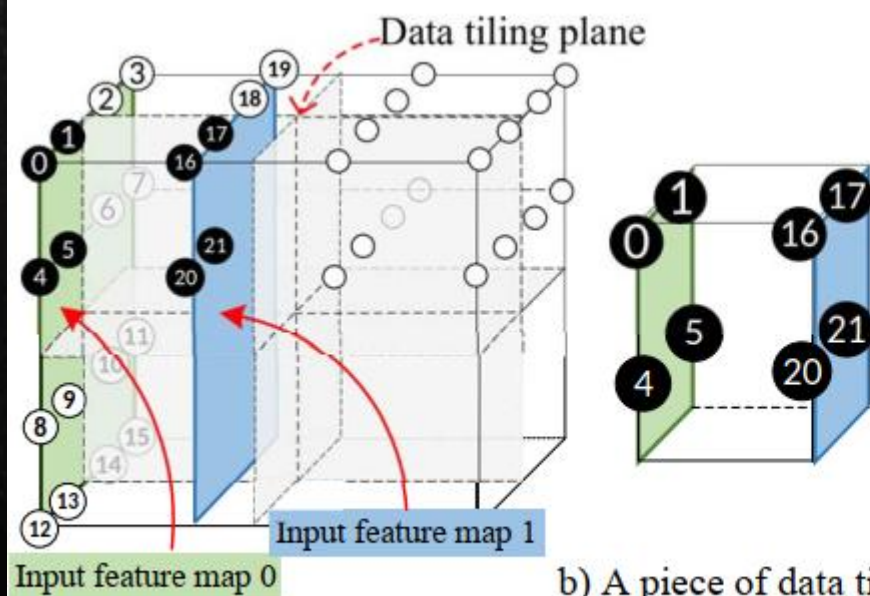
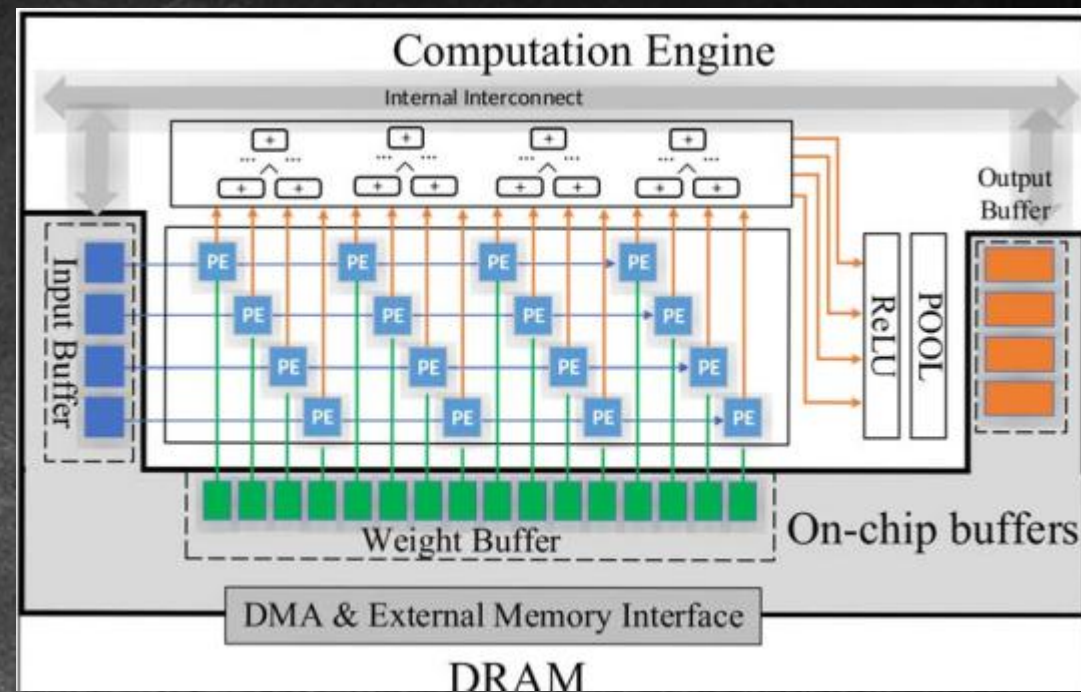


# UNIFORMED CNN REPRESENTATION: 权重映射



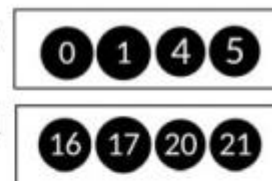


	Uniformed	Conv	FCN-Input	FCN-Weight
Input FM #	N	$N_{conv}$	$N_{fcn}/ker$	$N_{fcn}/ker$
Input FM size	$R_i \cdot C_i$	$R_{conv}^{in} \cdot C_{conv}^{in}$	batch $\cdot$ ker	$M_{fcn} \cdot ker$
Output FM #	M	$M_{conv}$	$M_{fcn}$	batch
Output FM size	$R_o \cdot C_o$	$R_{conv}^{out} \cdot C_{conv}^{out}$	batch	$M_{fcn}$
Kernel size	$K_1 \cdot K_2$	$K_1 \cdot K_2$	ker	ker
Stride	$S_1 \cdot S_2$	$S_1 \cdot S_2$	ker	ker



Data tile of input feature map 0  
is buffered in BRAM Bank 0

Data tile of input feature map 1  
is buffered in BRAM Bank 1



c) Physical data layout in on-chip buffer per BRAM bank

Data	0	1	...	4	5	...	...	16	17	...	20	21
DRAM Addr.	x0	x4	...	x10	x14	...	...	x40	x44	...	x50	x54

d) Row-major data layout in DRAM space

Data	0	16	1	17	4	20	5	21	...	...	...	...
DRAM Addr.	x0	x4	x8	xc	x10	x14	x18	x1c	...	...	...	...

e) Proposed data layout in DRAM space

a) A logical 3D data layout

b) A piece of data tile  
(input feature maps)





年前计划



年前计划：论文，视频。UG902实验