

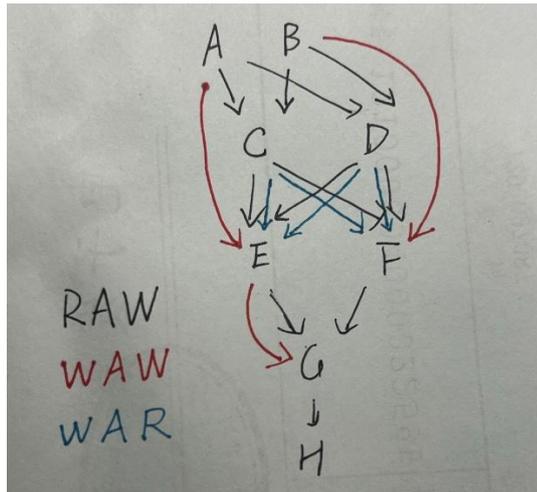
1、五级流水线，stall 与 forwarding 前馈优化

(1)															
Inst	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LD	IF	ID	EX	MEM	WB										
LD		IF	ID	EX	MEM	WB									
ADDD			IF	ID	ID	ID	EX	MEM	WB						
ADDD				IF	IF	IF	ID	ID	ID	EX	MEM	WB			
SD							IF	IF	IF	ID	ID	ID	EX	MEM	WB

(2)													
Inst	1	2	3	4	5	6	7	8	9	10	11	12	13
LD	IF	ID	EX	MEM	WB								
LD		IF	ID	EX	MEM	WB							
ADDD			IF	ID	ID	EX	MEM	WB					
ADDD				IF	IF	ID	ID	EX	MEM	WB			
SD						IF	ID	EX	MEM	WB			

2、数据依赖图与寄存器重命名

- 1) 假设 Mem[X1+0] 中存储值为 a, Mem[X1+8] 中存储值为 b, 则 $a+b+ab-ab(a+b)$
- 2)



3)

- A:** LD 0(X1), R1 **R1 = Mem[X1+0]**
- B:** LD 8(X1), R2 **R2 = Mem[X1+8]**
- C:** ADDD R1, R2, R3 **R3 = R1+R2**
- D:** MULTD R1, R2, R4 **R4 = R1*R2**
- E:** ADDD R3, R4, R5 **R5 = R3+R4**
- F:** MULTD R3, R4, R6 **R6 = R3*R4**
- G:** SUBD R1, R2, R7/R1 **R7/R1 = R5-R6**
- H:** SD 0(X2), R7/R1 **Mem[X2+0]=R7/R1**

根据寄存器重命名规则改写代码，数据依赖图 RAW 同 (2) 中，WAW 与 WAR 均被消除掉

3、 Winograd 卷积

$$Y = A^T [[GgG^T] \odot [B^TdB]] A$$

$$G = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \quad B^T = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$$

$$d \quad g = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} \quad g^T = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$$GgG^T = \begin{bmatrix} 3 & 3 & 1 & 1 \\ 3 & \frac{19}{4} & \frac{5}{4} & 3 \\ 1 & \frac{5}{4} & \frac{3}{4} & 1 \\ 1 & 3 & 1 & 3 \end{bmatrix} \quad B^TdB = \begin{bmatrix} 0 & -16 & 0 & 0 \\ -4 & 34 & 2 & -4 \\ 0 & 8 & 0 & 0 \\ 0 & -16 & 0 & 0 \end{bmatrix}$$

$$Y = A^T [[GgG^T] \odot [B^TdB]] A = \begin{bmatrix} 114 & 133 \\ 190 & 209 \end{bmatrix}$$

Winograd 中乘法在 \odot 操作中为 16 次；加法在 A^TMA 中（因 GgG^T 与 B^TdB 可以通过预计算方式完成，因而其中 64 次加法忽略）为 36 次

普通卷积中乘法 36 次，加法 32 次

注：winograd 中因 A 矩阵中只有 6 个非 0 元素，因而计算在 A^TMA 计算中只有 24 次加法，这里 24 和 36 我们都判对