**数据复用作业题目：共2题**

**第一题：**

在图1所示的卷积计算中，卷积核大小为KxK=2x2, 输入特征图大小为HxL=4x4，步进为1。现基于图2所示的架构实现此计算实例的权重复用，假设在该架构中所有硬件及操作均只消耗一个时间单位，并且在时刻0时输入行缓存1、输入行缓存2、蓝色部分卷积核加载好值。求（1）时刻0时输入行缓存1、输入行缓存2、卷积核缓存内每个寄存器所存的数值？（2）求时刻1时输入行缓存1最右端寄存器里的值、输入行缓存2最左端寄存器里的值？（3）求时刻3时寄存器a, b, c, d及f, g, h, i里的值？（4）寄存器k得到与图1卷积示例对应的第一个输出结果需要多少个时间单位？（5）计算完全部的输入特征图需要多少个时间单位？

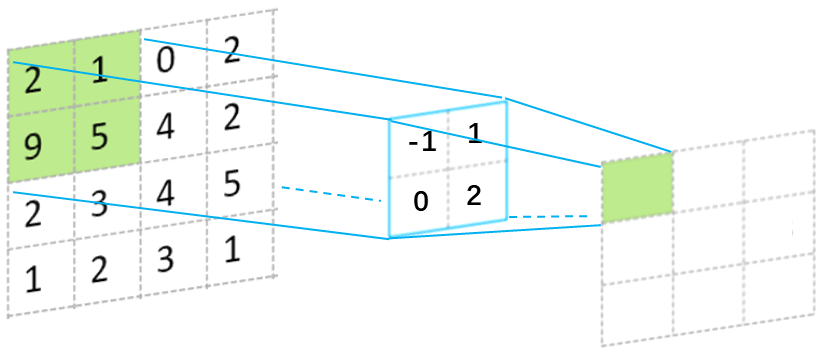


图1 卷积计算实例

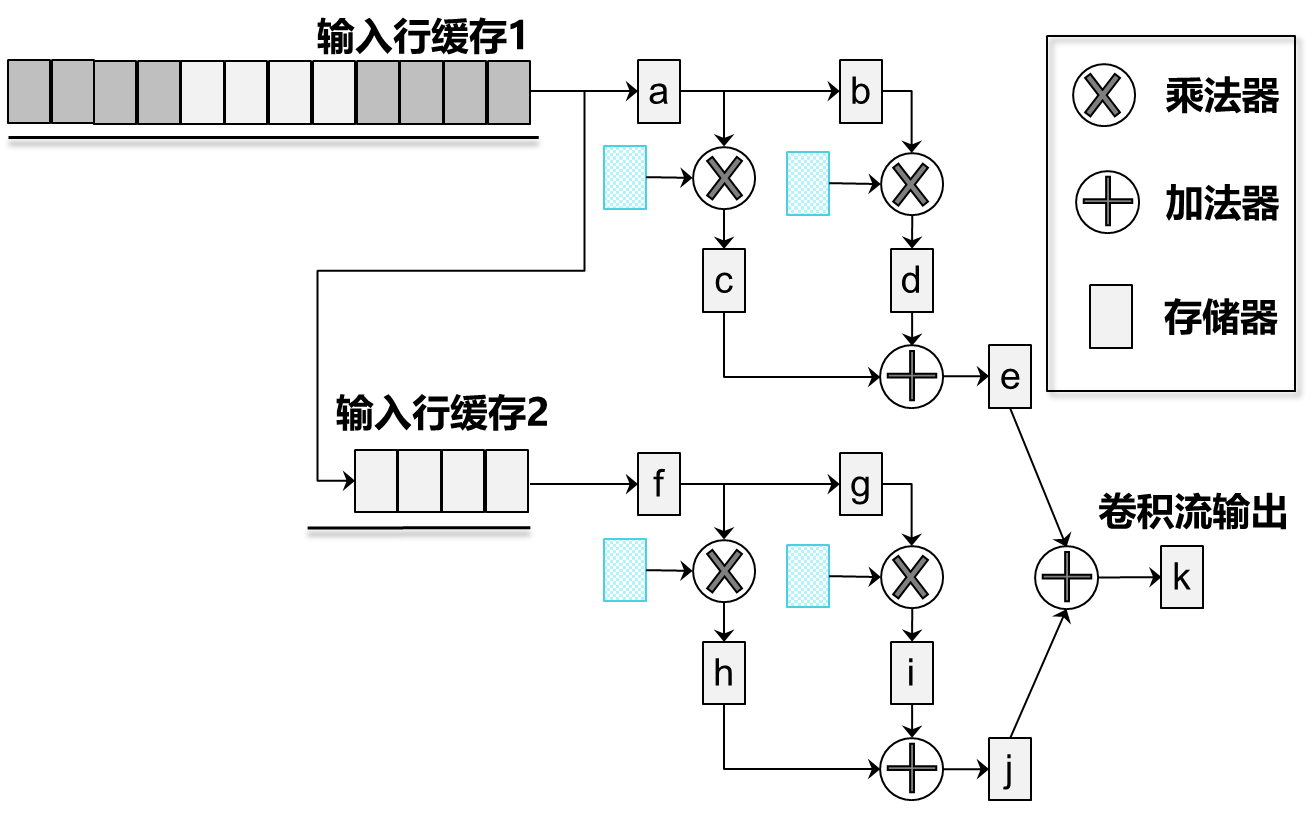
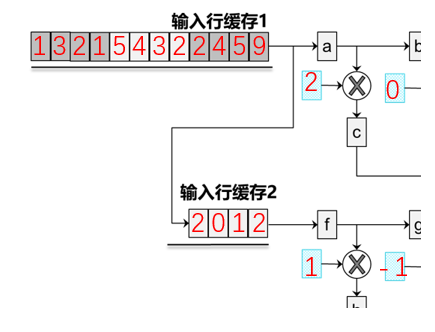


图2 权重复用的流处理实现架构。

解：（1）由9x0+5x2+2x(-1)+1x1的对应关系可知，结果如图解1所示，



图解1 行缓存及权重缓存在时刻0的值

（2）由于行缓存1的值会不断地从其右端推入行缓存2左端，

故时刻1时输入行缓存1最右端寄存器里的值：5；

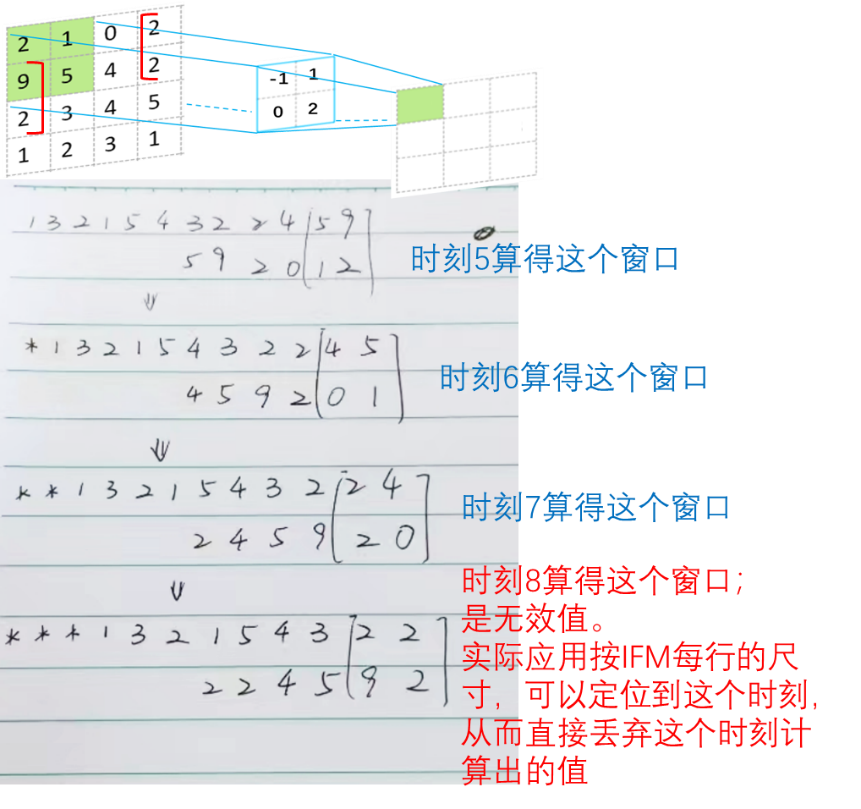
输入行缓存2最左端寄存器里的值：9

（3）时刻3时已经计算出IFM与权重的各个点乘积，故寄存器的值为a=4, b=5, c=10, d=0及f=0, g=1, h=1, i=-2。

（4）寄存器k得到与图1卷积示例对应的第一个输出结果需要多少个时间单位？5个

（5）由于时刻5得到第一个卷积窗口的值，此后每一个时刻得到一个对应卷积窗口的值，总共的卷积窗口数为9，故计算完所有的输入特征图需要5+9-1=13个时间单位

更正：此架构在窗口换行时会输出一个无效值，总共2个换行时刻，故正确结果是在上面13的基础上再加2，第15个时刻完成该输入特征图卷积。此处原理如下图所示：



**第二题：**

在上题的基础上，假设输入特征图及卷积核大小不变，特征图通道数等于卷积核通道数N=32, 卷积核个数M为16个，计算核心每次只能加载Tm=4个卷积核及Tn=4层通道进行运算。定义读取完全部输入特征图（HxLxN），或权重（KxKxN），或输出特征图（3x3xM）为一次完整读取，（1）求该HxLxN=4x4x32的输入特征图在计算核心内被完整读取的次数？（2）若计算核心内输入特征图缓存存储能力小于4x4x32个值，求在此最坏情况下该输入特征图从外部DRAM被完整读取的次数？（3）卷积核参数从外部DRAM被完整读取的次数？（4）输出特征图在计算核心内被完整读取的次数？（5）若对输出特征图的单次更新需要1次读和1次写共两次访问，计算核心内输出缓存能力小于3x3xTm个值，求在此最坏情况下的输出特征图从外部DRAM被完整访问的次数？

解：（1）由于采用了权重复用方式，输入特征图复用系数为，故其在计算核心内被完整读取的4次。

（2）影响输入特征图大小的相关维度为H、L、N, 而M维度在最外层遍历，为了让输入完整地与每一个最外层的Tm分块计算而无需从外部读取，那片上输入缓存得至少得存所有的输入特征图。题中输入缓存容量小于输入特征图，故需要每个Tm分块都对应的完整地从外部DRAM读取一次，共需读取4次。

（3）由于是纯权重复用方式，卷积核只需从外部读取1次。

（4）输出特征图在N维度上总共有个分块，这些分块需要逐个累加得到部分和，每次累加都需要读取当前部分和，故片上读取次数等于累加次数等于次。

（5）影响输出特征图大小的相关维度为M, R, C, 而权重复用遍历顺序为R, C, N, M, 由于M维度在最外层遍历且M维度与输出特征图大小相关，为了使得输出部分和不用往外部DRAM存取，故输出缓存只需要存一个最外层的Tm分块及内层大小相关维度所有尺寸即可，即TmxRxC。如果片上输出缓存小于此，即容量小于存TmxRxC=4x3x3个值，那每次更新部分和都需要从外部读取，总的外部DRAM访问（包括读、写）次数为次。