人工智能的硬件基石：从物理器件到计算架构

Lab2 bonus: AI卷积加速器设计

lab2 bonus 占比Lab2的50%，要求设计简单AI卷积加速核，尽可能完整的实现加速核如控制逻辑、所需缓存等等，

此任务为开放式，但我们给出以下思路与建议供参考

1. AI卷积加速器的几种可能实现方案

1. Input-Stationary/Output-Stationary/Weight-Stationary传统卷积实现方案，输入/权重/输出缓存的控制方案应合理考虑（缓存可以用大规模reg来简单表示），此方案较为简单，最好能支持参数化的输入/权重/输出矩阵尺寸以及stride/padding等参数（即可以用parameter来表示与调整）

2. Systolic Array脉动阵列实现方案：需要实现小的计算单元PE，然后调用PE实现大规模卷积运算阵列，同时需要实现输入/权重矩阵转换为阵列输入的逻辑与可能的控制逻辑

3. Winograd卷积实现方案：可以调用Lab1中2D Winograd核（如果写了的话），来实现更大规模卷积核尺寸（最好能支持多个大的尺寸）的通用卷积计算。提醒：因为Winograd面向的是小kernel size，所以这种方式会有些复杂

其他新颖的方式如存内计算实现方案等等都是可以的，但需要指出你所使用的方案相比于传统方案的优势在哪里；额外设计的模块只要说明用处与操作方式可加分

2. Lab2 bonus的评分依据

1. **因为本实验为开放性的，因此实验报告很重要，说明采取的方案，设计的加速器架构、流水和操作方式等等重要设计点，助教根据实验报告中你对于AI卷积加速器的设计与理解评分**

2. **代码与代码的测试程序（testbench）由你自己完成，同样在实验报告中应包含testbench完成了什么任务（任务由你自己设定）；确保实验报告中描述的tb操作（波形图与解释）用你提交的代码可以复现，由助教验证；只有代码但没有testbench将会导致无法评定代码功能**
3. **如果有的部分未经代码实现也可以写下设计思路与理解**
4. **实验报告最好精简，体现核心设计点即可；过分冗余的报告可能会导致核心设计点被遮盖**

截止日期：2025.6.15 23:59