

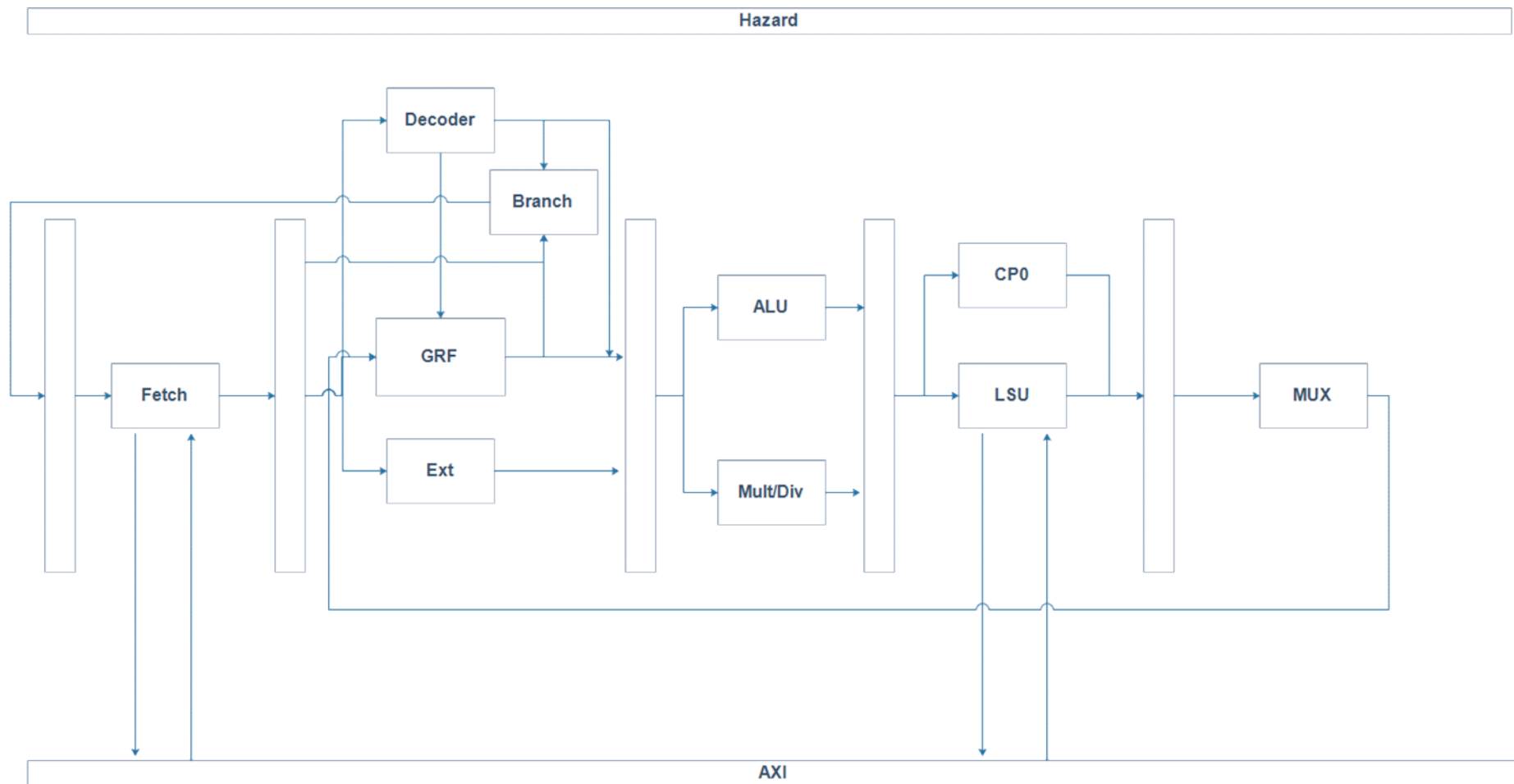
# MIPS32神经网络协处理器设计

北京航空航天大学1队

# 提纲

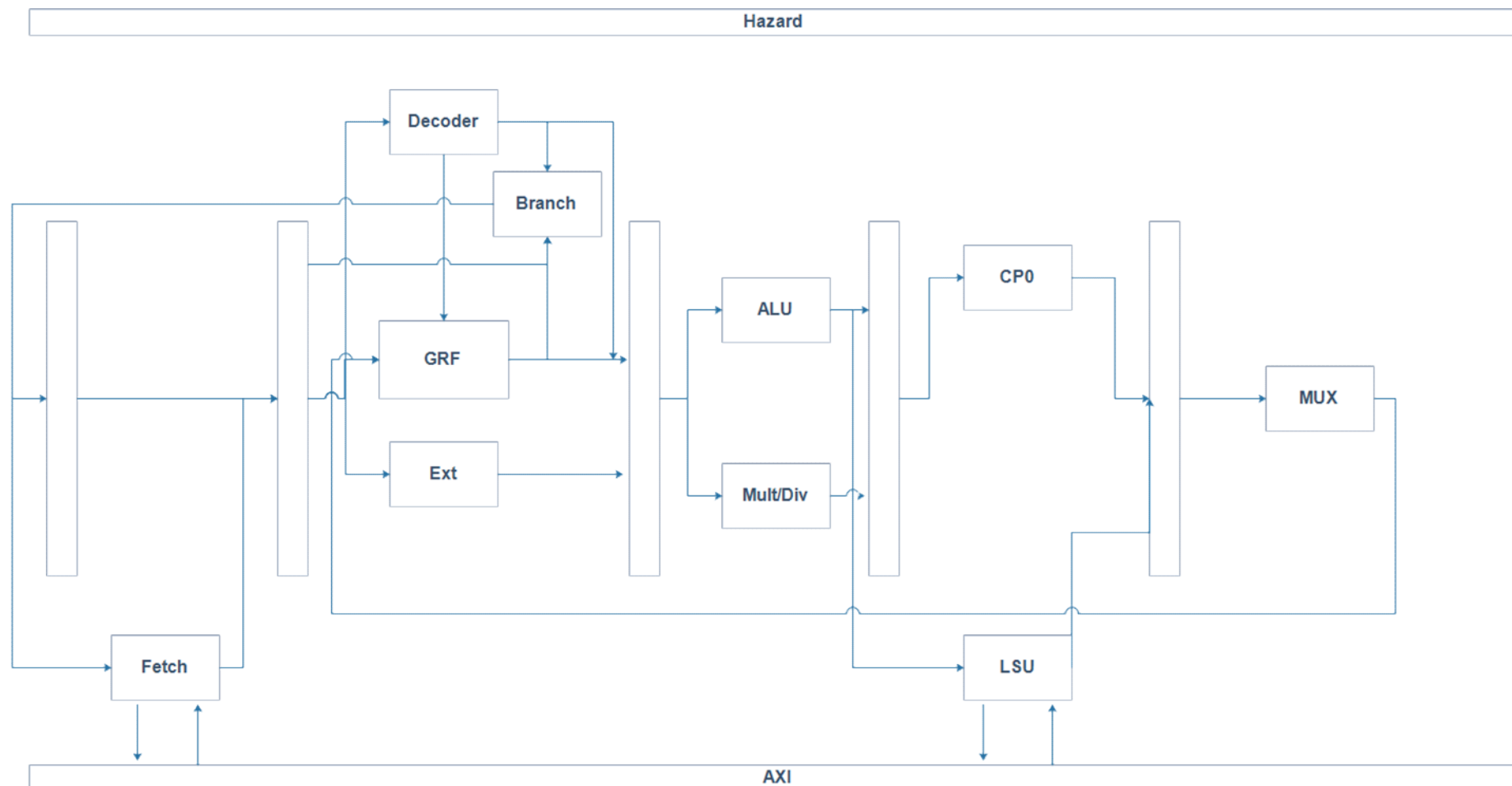
- CPU结构设计
- 协处理器结构设计
- 协处理器交互

# CPU结构 - 传统的5级流水



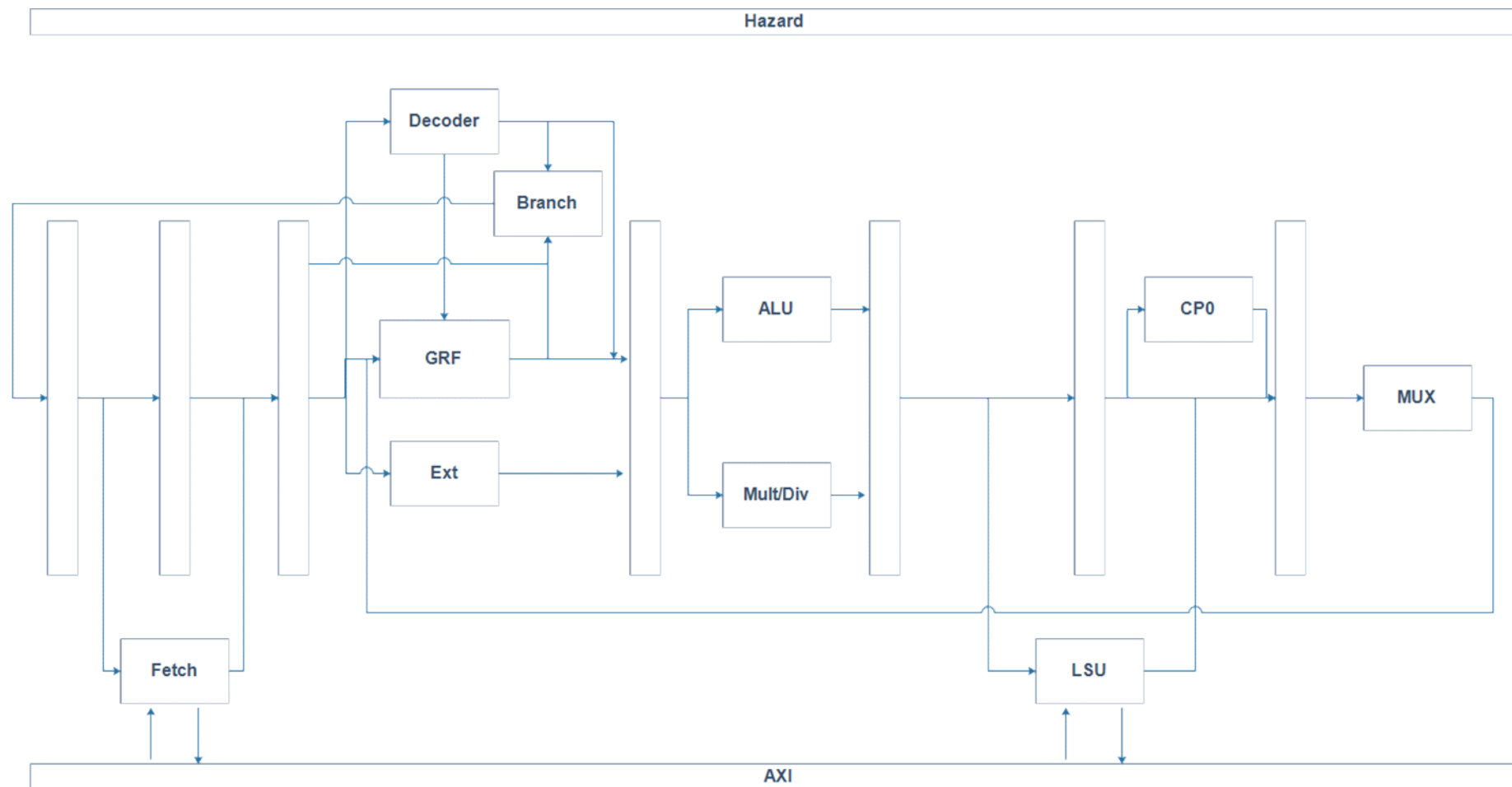
# CPU结构 - 初赛设计的5级流水

频率：42MHZ  
周期性能：1.229



# CPU结构 - 决赛设计的7级流水

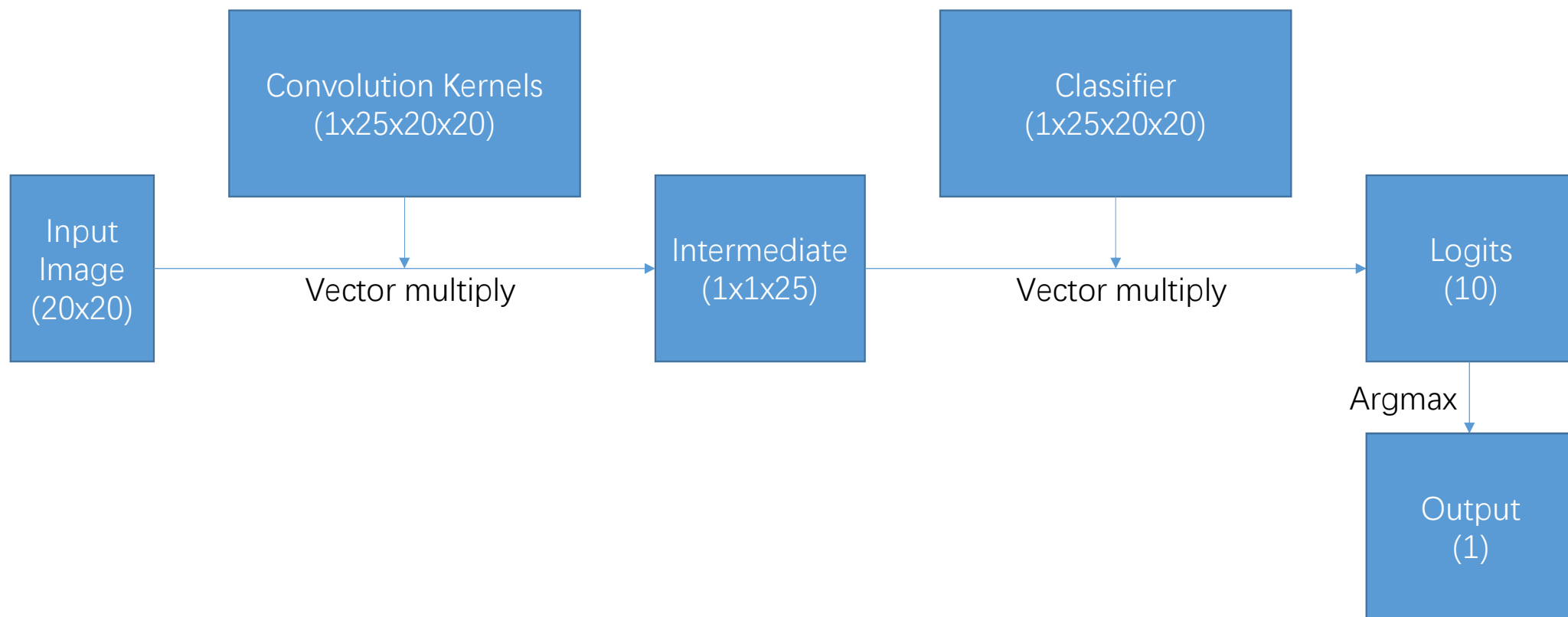
频率：75MHZ  
周期性能：1.007



# 协处理器结构 – 主要特点

- 协处理器在内部RAM进行计算
  - 避免与主CPU争用访存周期，提高双方性能
  - 内部RAM单独编址，实现针对多分支等复杂网络结构的灵活定制
- 协处理器使用定点数（16bit）进行运算
  - 对输入图片进行归一化，使得神经元激活的动态范围可以接受
  - 在FPGA上，使用很少的资源快速实现乘法器
  - Sigmoid函数定点化：线性插值
- 向量化计算
  - 针对NN规整的结构，堆积运算单元，增加吞吐

# 协处理器结构



# 协处理器-CPU交互

- 使用访存指令与Coprocessor进行交互
  - CPU将输入写入Coprocessor对应地址段
  - CPU向Coprocessor发送开始命令
  - 以轮询或中断方式得知计算是否完成
  - 取结果，进行后续处理



# 协处理器性能

- On mnist dataset:
  - Top-1 accuracy: 90%
- Input -> fc -> sigmoid -> fc -> argmax
  - Coprocessor performance: ~450 cycles / inference (without data loading cycles)
  - CPU performance: ~100 000 cycles / inference (without data loading cycles)