



附件二：

## 2022 年 CCF-飞腾基金申报课题

### 目 录

分类一：开放课题 .....	2
分类二：半开放课题 .....	2
课题一：分支预测及数据预取前沿技术研究 .....	2
课题二：支持 Chiplet 设计的 CPU 芯片级性能模拟器 .....	3
课题三：基于机器学习的 MMMC STA 加速方法 .....	4
课题四：机器学习指导的 PGO 编译优化技术 .....	5
课题五：后量子密码算法的硬件设计与优化 .....	6
课题六：面向处理器安全防护的频率/电压/可见光传感器设计技术研究 .....	7
课题七：大尺寸复杂封装基板的可靠性研究 .....	8
课题八：飞腾 NPU 国内开源推理框架适配与优化 .....	9

## 分类一：开放课题

在通用及专用微处理器设计、优化、制造、应用等方向自由选题。

## 分类二：半开放课题

### 课题一：分支预测及数据预取前沿技术研究

#### 【研究背景】

现代高性能微处理器中，分支预测和数据预取算法的准确度对处理器的性能有至关重要的影响。近年来，关于这两种算法的最新研究成果也层出不穷。如何在现有研究的基础之上更上一层楼，进一步提升分支预测和数据预取的准确率，对处理器性能的持续提升非常重要。

#### 【研究内容】

在目前学术界最先进分支预测算法和数据预取算法的基础上更进一步，提高分支预测和数据预取的准确率：

1. 对学术界最新分支预测算法进行改进优化，针对 SPEC CPU2017 INT Benchmark 进行模拟测试，综合性能较学术界最新算法有明显提高，实验评估需要基于一条实际的乱序执行流水线；

2. 对学术界最新数据预取算法进行改进优化，针对 SPEC CPU2017 INT Benchmark 进行模拟测试，综合性能较学术界最新成果有明显提高，实验评估需要基于一条实际的乱序执行流水线；

3. 要求改进后的算法具有较高的硬件可实现性，包括较好的时序特性，消耗的硬件存储资源适中；或者将准确率高但硬件实现性低的现有算法进行改进，在保持准确率的前提下大幅增加硬件可实现性。

## 课题二：支持 Chiplet 设计的 CPU 芯片级性能模拟器

### 【研究背景】

目前学术界和工业界已有一些面向处理器核心的模拟器，例如 Gem5、Gem5 Graphic 和 MPGPUSim 等，这些模拟器对于指导单核、多核甚至异构结构研究有很大帮助。在 SoC 层面，则需要一个能够对片上网络拓扑、一致性协议、本地和跨路访存带宽及延迟等体系结构特征和参数进行精确仿真的模拟器，从而评估芯片的整体性能。另外，Chiplet 技术及设计方法学为体系结构设计带来更多的选择，模拟器应能够对多种 Chiplet 划分方式进行性能仿真，从而确定最佳的全芯片架构方案。

### 【研究内容】

研究支持 Chiplet 设计的 CPU 芯片级性能模拟器。具体如下：

1. 与处理器核心的接口设计。可以接受真实处理器核心或者处理器核心模拟器产生的 Trace。
2. 片上网络拓扑的建模方法和模型，能够支持从事务级到时钟精确级等不同抽象层级。
3. Chiplet 接口的建模方法和模型，能够支持不同 Chiplet 实现方式。
4. 完成性能模拟器开发，基于典型应用程序进行模拟，给出性能评估结果。

## 课题三：基于机器学习的 MMMC STA 加速方法

### 【研究背景】

静态时序分析 STA 要考虑芯片工作模式和工艺角,Multi-Mode Multi-Corner STA 是指在多个模式和多个工艺角下进行静态时序分析的方法。先进工艺的 Corner 数目进一步增加,给设计带来很大的挑战。静态时序分析组合越多,计算资源和时间的耗费会越来越大。对于设计规模较大的 CPU 芯片,难以在每次设计迭代时把所有的模式和所有的 Corner 都完全覆盖,通常选取部分 Corner 在时间和资源开销与良率间进行平衡。近年来基于深度神经网络的机器学习方法在模式识别、行为预测等方面取得了很大的成功,开始应用于 MMMC STA 中。通过机器学习的方法,减少 Corner 迭代数目,提高 Corner 覆盖率,同时保证时序分析的准确性和 Signoff 的完备性。

### 【研究内容】

结合部分 Mode 和 Corner 下时序分析结果样本,使用机器学习的方法,组建深度神经网络,对样本数据进行学习训练,完成训练后的神经网络模型能够对其它 Mode 和 Corner 下的时序路径预测时序分析结果。课题要求完成算法的设计与验证,包括:

1. 输入数据集的选择和构造;
2. 深度神经网络的搭建;
3. 对单条时序路径的学习和预测;
4. 对多条时序路径的学习和预测;
5. 模型能支持对 30 万条时序路径的模块进行参数提取、训练和预测。

## 课题四：机器学习指导的 PGO 编译优化技术

### 【研究背景】

PGO (Profile Guided Optimization) 是提高编译器优化能力的有效技术, 但从收集 Profile 到使用 Profile 数据进行编译优化的全过程相对复杂、繁琐, 且需要根据软件代码更新而不断重复, 导致 PGO 技术的应用场景受限。随着机器学习技术的发展, 编译领域拥有了更多可以捕捉程序语义的方法与工具, 比如自然语言理解、深度学习等。在不需要 Profile 的情况下, 以机器学习的预测数据为指导, 为编译器提供近似等价的 Profile 预测信息, 指导编译器产生更好的优化代码, 可以有效的将 PGO 优化技术推向实用。

### 【研究内容】

基于 LLVM 编译框架, 研究以下内容:

1. 面向 PGO 优化的应用程序语义特征的抽取技术。研究应用程序源码语义的特征抽取, 结合源码语义、指令语义、控制流与数据流语义等多种信息, 构建一个或多个面向 PGO 所需信息的预测模型, 包括但不限于: 分支跳转概率、循环次数概率、程序可能热点区域的识别等;
2. 研究并收集相关训练代码集合, 结合真实 Profile 的结果, 形成训练模型的方法;
3. 以 LLVM PGO 框架为基础, 使用训练好的模型, 在 PGO 编译优化过程中, 用预测数据替换真实 Profile 的数据, 研究并评估预测准确性与优化效果。

## 课题五：后量子密码算法的硬件设计与优化

### 【研究背景】

密码技术是保障网络空间安全秩序的核心基础，是维护国家安全的重要战略资源。当前适用的公钥密码体系建立在以大整数分解、离散对数等为代表的经典数学困难问题之上。随着量子计算技术的发展，经典计算困难问题在量子计算模型下可以被快速有效求解，经典密码算法和协议将面临严峻的安全挑战。后量子密码（Post-quantum Cryptography）是有效解决量子攻击威胁的一种密码学解决方案。目前美国国家标准与技术研究所 NIST 正面向全球开展的后量子密码标准化工作，中国密码学会也在逐步启动我国的后量子密码算法标准化工作。

### 【研究内容】

面向具有抗量子攻击的后量子密码算法体系，针对后量子密码算法数学困难问题多样、计算密集、存储开销大及硬件安全实现研究不足的技术难题，开展基于先进体系架构下的后量子密码指令集扩展、算法引擎设计与优化、SDK 开发等，推动后量子密码硬件在国产处理器上的实现与应用。

1. 基于格理论、编码、哈希、多变量及超奇异同源等成熟后量子密码算法中的多种算法进行研究和设计优化；

2. 对后量子密码算法的指令集扩展方案进行研究。

### 【课题提供】

1. 密码算法在处理器上的常规实现方法；

2. 在非指令扩展层面，提供 FT 开发板的开发环境。

## 课题六：面向处理器安全防护的频率/电压/可见光传感器设计技术研究

### 【研究背景】

安全处理器是目前各种电子设备的基础组件，但安全处理器所在的外部运行环境不能保障处理器自身的物理安全，将面临诸如功耗攻击、探针攻击等物理攻击的风险。当处理器工作电压、频率发生改变或封装盖板被打开等可导致故障的工作参数改变时，处理器处于易受攻击状态，导致密钥和敏感信息泄露。为应对物理攻击，业内常采用集成频率、电压以及可见光传感器的方式，来检测可能面临的一些物理攻击。

### 【研究内容】

面向安全处理器面临的多种物理攻击风险，研究频率 sensor、电压 sensor 及可见光学 sensor 等检测手段，主要包括以下几个方面的内容：

1. 研究灵敏可靠的频率、电压和可见光三种 sensor 的设计方案及原型；

2. 频率 sensor 设计目标要求：

(1) 检测频率是否处于 $\pm 20\%$ 正常范围，超出则给予报警；

(2) 检测系统是否受到时钟毛刺攻击。

3. 电压 sensor 设计目标要求：

(1) 检测外部电压是否处于目标电压 $\pm 20\%$ 正常范围，超出则给予报警；

(2) 检测电源系统是否受到电压毛刺攻击。

4. 可见光 sensor 设计目标要求：

(1) 当攻击者打开芯片封装暴露于可见光下时，在通电状况下，可见光 Sensor 会及时监测到攻击并向系统报警；

(2) 检测波段包括可影响光电效应的可见光、近紫外光、近红外光波段。

## 课题七：大尺寸复杂封装基板的可靠性研究

### 【研究背景】

随着芯片功能不断增强、性能不断提升，芯片封装尺寸不断增大，功率密度也越来越高，封装面临的可靠性问题愈发严峻。在封装规划与设计过程中，只有通过科学的仿真方法，充分考虑应力、翘曲、金属疲劳等综合特性，对封装方案进行设计评估，将封装环节风险降低到可控范围。

### 【研究内容】

1. 芯片翘曲应力仿真分析，包括制造过程中的翘曲应力、贴装时的翘曲应力、使用时的翘曲应力。提出优化方案，使最终应力和翘曲指标满足行业标准要求。

2. 探索复杂芯片结构的建模方法，包括微小结构的等效建模、基板线路的等效建模。实现芯片多尺度建模，提升仿真精度和效率。

3. 对芯片样品开展变温情况下的翘曲测试，分析测试结果与仿真的拟合情况，通过合理修正仿真参数，使仿真与测试结果吻合。

4. 研究芯片焊点、Tim 疲劳寿命，提出优化设计的解决方案。

### 【课题提供】

1. 封装设计模型；
2. 封装参数（结构，设计，材料参数等）；
3. 芯片样品。

## 课题八：飞腾 NPU 国内开源推理框架适配与优化

### 【研究背景】

AI 应用开发通常基于一定的推理框架进行。目前飞腾 NPU 软件栈为 AI 应用提供了两种开发方式：1) 基于 TVM Runtime 接口；2) 直接基于 FT-DNN API。在 AI 应用开发过程中存在一定的工程化难点，特别是应用移植较难，对于基于某种推理框架开发好的 AI 应用，需要重新开发。

开源推理框架在业界有较好的应用生态，在飞腾 NPU 上适配这些推理框架，能够充分利用这些框架的生态，解决飞腾 NPU 在 AI 应用工程支持上的问题，向更多用户提供服务，推广飞腾 NPU 的应用。

### 【研究内容】

进行飞腾 NPU 开源推理框架适配工作，主要包括神经网络编译技术、开源推理框架 AI 芯片适配技术、飞腾 NPU 算子与开源推理框架算子适配方法以及飞腾 NPU 开源推理框架适配实现。基于相关 API 在飞腾 NPU 上完成 3 种以上推理框架的适配和优化工作，能够在飞腾 NPU 上运行适配框架所支持的典型 AI 应用，包括但不限于：

1. 图像分类；
2. 人脸检测、识别；
3. 目标检测、跟踪；
4. 图像语义分割、实体分割；
5. 人体姿态估计；
6. 场景文字识别；
7. 图像生成（超分辨率、风格迁移等）；
8. 图像去模糊、去雨、去雾；
9. 语音识别；
10. 机器翻译。

可考虑（但不限于）如下推理框架：

1. PaddleLite（必选）
2. TensorFlow Lite（可选）
3. OpenCV DNN（可选）

4. TEngine (可选)
5. TNN (可选)
6. NCNN (可选)
7. MNN (可选)

**【课题提供】**

1. 飞腾 NPU 底层软件栈框架文档;
2. FT-DNN 库及 API 文档;
3. 飞腾 D2000+X100 开发板 (配套 OS、NPU 驱动)。