

CCF-华为胡杨林基金系统软件专项 – 2022年编译器与编程语言方向指南介绍

华为技术有限公司 / 编译器与编程语言Lab

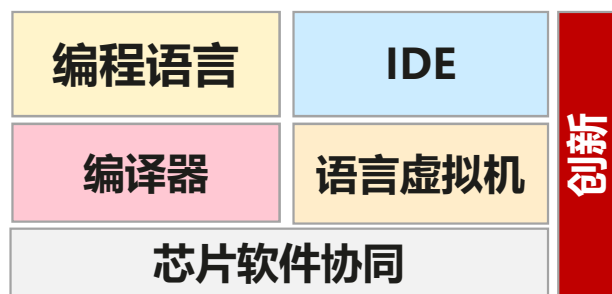
曾建江



系统软件专业委员会
Technical Committee of Systems Software



研究先进编译器与编程语言，探索面向先进体系结构的编程支持



> 
性能最优

编译器

极致发挥芯片能力

> 
开发者新生态

编程语言

语言抽象与应用及平台的垂直整合优化

> 
轻量/高效

语言虚拟机

轻量化高效语言虚拟机

> 
体验最优

IDE

开发者极致体验

> 
协同最优

芯片软件协同

先进体系结构支持

研究方向概览

编译器	编译核心框架	统一IR	静态编译优化	编程编译模型		
编程语言	编程范式	语言扩展	领域抽象	安全可信	实现和优化	工具和框架
语言虚拟机	程序分析	安全增强	垃圾回收	运行时优化		
IDE	低代码及元模型底座	智慧化软件工程	IDE基础设施中的协议及协议增强	IDE基础设施中的基座先进性		

编程语言研究方向（1）

方向	介绍
语言扩展	<p>【场景说明】 元编程技术可以构建EDSL（Embedded DSL），提升目标领域的软件生产力，但面向开发者友好的诊断调试是其推广的主要技术瓶颈。</p> <p>【子课题】 基于元编程构建EDSL，支持源码级诊断调试技术：提供屏蔽宿主语言语义的诊断调试信息的映射系统</p>
编程范式	<p>【场景说明】 跟踪软件发展趋势，探索和构建声明式、数据流等编程范式，提升开发效率并且能够隐式并行化。</p> <p>【子课题】</p> <p>课题一：探索声明式、数据流范式</p> <p>课题二：声明式、数据流范式自动并行化，达到高性能目标</p>
编程范式	<p>【场景说明】 终端开发者使用单一语言开发分布式应用，不感知多端分发部署，可获得“单机”开发体验，降低开发门槛</p> <p>【子课题】</p> <p>课题一：分布式语言特性的研究和探索，包括分布式语言的定义、代码的部署调度、测试等</p> <p>课题二：分布式语言支持多后端多场景，包括：Native, 云，边，轻量化IoT</p>

编程语言研究方向（2）

方向	介绍
领域抽象	<p>【场景说明】 面向AI+HPC的编程模型的支持和探索（如数组编程、自动微分等）。</p> <p>【子课题】</p> <p>课题一：面向华为自研编程语言的数组编程模型设计和探索，对标NumPy</p> <p>课题二：Whole-Program Automatic Differentiation</p> <p>课题三：华为自研编程语言自动微分的研究，如高阶微分、偏微分、增量计算模型、自动微分内存优化等</p>
安全可信	<p>【场景说明】 类型推导是编程语言的一个十分重要的特性，它让编程人员从繁重的类型标注中解放，极大地提高生产力。</p> <p>【子课题】 支持 intersection types 与 union types，且支持全局类型推导</p>
安全可信	<p>【场景说明】 借助软件分析技术，系统构建面向华为自研编程语言代码的程序分析能力。</p> <p>【子课题】</p> <p>课题一：基于华为自研编程语言的程序分析技术的设计与实现，如data race分析、逃逸分析等</p> <p>课题二：符号执行技术</p> <p>课题三：基于华为自研编程语言的测试技术，如 fuzzing 测试技术、基于语义规范的自动化测试用例生成技术、评价机制及指标设计、Benchmark分析工具/框架等</p> <p>课题四：基于华为自研编程语言 spec 的代码自动修复/推荐</p>

编译器研究方向（1）

方向	介绍
编译核心组件	<p>【场景说明】 构建编译器宏架构，打造可扩展，可重构，可定制，弹性化的新编译内核，帮助编译器进行优化。包括编译前端、中端和后端及相应的工具链和libc。</p> <p>【子课题】</p> <p>课题一：多级内存感知，建立多级内存访问代价模型，依据访存时延优化</p> <p>课题二：多模态(源码、二进制)、全生命周期(链接、运行时) 的持续优化</p> <p>课题三：超投机指令调度，深度协同指令调度</p> <p>课题四：静态反馈优化编译</p> <p>课题五：库函数精度优化</p>
统一IR	<p>【场景说明】 用于构建可复用和可扩展的编译器的框架，旨在统一的基础架构中支持多种不同需求的混合中间表示，解决软件碎片化、改善异构硬件的编译、降低构建领域特定编译器的成本，以及帮助将现有的编译器连接到一起。</p> <p>【子课题】</p> <p>课题一：融合现代AI低精度硬件架构以及异构统一编程编译技术，研究计算物理领域特定(数值风洞等)的隐式混合精度数值算法设计</p> <p>课题二：DSL语言实现软件框架，解决兼容性问题</p>

编译器研究方向（2）

方向	介绍
静态编译优化	<p>【场景说明】 在编译可执行文件的时候，将可执行文件需要调用的对应静态库(.a或.lib)中的部分提取出来，链接到可执行文件中，使可执行文件在运行的时候不依赖于动态链接库。</p> <p>【子课题】</p> <p>课题一：编译选项优化，根据不同模块，选用不同的编译选项，达成性能最优</p> <p>课题二：编译器函数冷热分区优化</p> <p>课题三：伪共享检测工具和优化，解决全局共享变量频繁的被多个线程在硬核间访问和修改的false-sharing问题</p> <p>课题四：全局函数优化，实现优势场景性能倍增</p>
编程编译模型	<p>【场景说明】 突破跨异构算力融合优化，远近多级内存感知计算等关键技术，跨算力场景SPEC Accel ratio平均提高30%。</p> <p>【子课题】</p> <p>课题一：跨异构算力统一编程以及隐式数据流编程</p> <p>课题二：算子自适应编程和编译，支持C/C++和python的混合编程</p> <p>课题三：夯实平台满足度，支持OpenCANN算子标准和基础设施能力构建</p> <p>课题四：构建动态shape编译和泛化能力，训练+推理，在可销售场景满足度对齐静态shape</p>

IDE研究方向（1）

方向	介绍
低代码开发及元模型底座	<p>【场景说明】 MDA（Model Driven Architecture）是一种软件设计开发方法，广泛应用于航空航天等安全关键领域，国内在MDA基础工具方面还是空白基于Eclipse EMF进行模型驱动研究与开发，存在禁用风险、兼容性差、独立性差等问题。</p> <p>【子课题】</p> <p>课题一：在不依赖Eclipse等第三方技术的前提下，自主研发国产MOF元模型基础设施及相关工具。</p> <p>课题二：领域特定图形建模语言GML生成框架。</p> <p>课题三：领域特定文本建模语言TML生成框架打下基础</p>
智慧化软件工程	<p>【场景说明】 软件泛在化的趋势使得软件逐渐成为传统生产生活等基础设施的核心，软件的可靠性和是否易受到攻击成为社会和个人安全的重要组成部分。因此，如何高效利用已有软件数据中的缺陷知识并用于智能化缺陷检测成为软件工程和信息安全领域的重要研究问题。</p> <p>【子课题】</p> <p>课题一：源代码缺陷库构建。</p> <p>课题二：基于图神经网络的代码缺陷检测模型。</p> <p>课题三：基于神经网络的源代码缺陷检测原型系统。</p>

IDE研究方向（2）

方向	介绍
IDE基础设施中的协议及协议增强	<p>【场景说明】当前的软件调试工具大都只是将程序状态信息以一种界面友好的方式提供给程序开发者或维护者，而缺乏对于每个节点程序错误状态的原理性解释从而辅助调试过程。当前的软件修复技术大都只是生成程序补丁，而缺乏对于补丁语义的解释（解释补丁以辅助判断补丁正确性）。</p> <p>【子课题】</p> <p>课题一：基于程序状态解释的交互式程序调试方法。</p> <p>课题二：可解释的交互式补丁推荐技术。</p>
IDE基础设施中的基座先进性	<p>【场景说明】软件项目开始之初，软件体系结构总是被精心设计，但随着项目的开展，架构侵蚀不可避免。如果IDE能够对体系结构关键代码进行明显标识，将极大的改善上述情况。</p> <p>【子课题】</p> <p>课题一：面向差异化样本密度的设计模式代码挖掘技术。</p> <p>课题二：基于小样本的体系结构策略代码挖掘技术。</p> <p>课题三：针对案例项目的应用验证。</p>

语言虚拟机研究方向（1）

方向	介绍
程序分析	<p>【场景说明】</p> <p>现有静态分析手段在处理千万级大规模代码时，存在开销高、精度低、扩展性差等问题，为了更加全面的解决程序安全问题，需构建一套完善可落地的静态分析能力</p> <p>【子课题】</p> <p>课题一：全局指针分析：千万行代码级的高精度模块内全局指针分析</p> <p>课题二：跨模块程序分析：支撑高精度指针分析的高效跨模块程序分析框架</p>
安全增强	<p>【场景说明】</p> <p>内存泄漏、野指针、重复释放等内存安全问题严重威胁了C程序的安全性，给开发调试和上线维护都带来巨大成本。</p> <p>【子课题】</p> <p>课题一：低内存开销的指针标记与追踪</p> <p>课题二：ASAN机制下运行时性能开销的安全优化</p> <p>课题三：高安全低代码改造的自动内存管理</p>

语言虚拟机研究方向（2）

方向	介绍
垃圾回收	<p>【场景说明】</p> <p>GC停顿时间长和因跨NUMA节点访问内存造成的总线消息挤压问题，理论上，线程本地回收有较好的效果。</p> <p>【子课题】</p> <p>课题一：适合java语言特性的逃逸分析方案的选定和实现</p> <p>课题二：分配和回收方案的选定和实现</p> <p>课题三：优化和开销的对比及适用场景确定</p>
运行时优化	<p>【场景说明】</p> <p>虚拟机自身的开销逐步成为业务应用的重要考量因素，构建低干扰、低开销的内存管理技术，解决痛点场景的轻量级并发问题。</p> <p>【子课题】</p> <p>课题一：轻量、高性能、低功耗的自动内存管理技术</p> <p>课题二：轻量级并发的Runtime技术研究</p>

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and
organization for a fully connected,
intelligent world.

**Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

