

清华信息科学与技术国家实验室（筹）

公共平台与技术部工作进展

孙茂松

2006. 2. 22

为有力支持基础科学研究和应用基础科学研究工作，国家实验室设立了公共平台与技术部，旨在在整个国家实验室范围内对重大仪器设备与科学实验环境进行统一规划、建设、使用与管理，进而实现面向全社会开放共享的公共平台架构。目前公共平台与技术部集中建设了两大平台：高性能计算平台和 EDA 平台（其余若干重要平台也正在规划、建设中）。

一、高性能计算平台

随着信息科学技术，特别是计算机硬件和软件的飞速发展，科学研究的方式发生了划时代的深刻变化。计算技术在各科学领域的广泛运用，正在改变许多传统科学的研究方式和面貌。计算科学被认为是传统的科学实验和科学理论之外的“科学的第三维”。目前，已经产生了大量冠以“计算”的科学领域，例如计算物理、计算化学、计算生物学、计算地球科学以及计算大气科学等。

我们的高性能计算平台主要由“清华探索 3 号”高性能机群计算机系统组成。该系统采用浮点运算速度高的 Itanium 2 的 64 位处理器，共 128 个结点，256 个 CPU，峰值速度达每秒 1.33 万亿次，效率为 84.6%，Linpack 持续性能达到 1.126T，内存为 514G，存储能力争

取近期内拓展到 200TB。“清华探索 3 号”为清华大学目前性能最高的计算机系统，可为提供大规模科学计算服务。系统配备了 LSF 作业管理系统，加强了对作业的管理、方便用户使用。此外，还建立了配套的计算可视化环境(同时成为 ChinaGrid 的监控中心)。在建设过程中，除学校本身的力量外，我们还得到了 EMC、Juniper、HP、Platform、SGI 等多家公司的全力配合与协助。

目前，“清华探索 3 号”已成为 China Grid 主节点，并面向校内外用户提供高水平的高性能计算服务。我们力求“以（优质）服务创造卓越（成果）”，并进一步以此为依托，建设前沿与交叉学科的重要资源与科学研究实验平台，以期形成特色(如规划中的基于 Web 的海量信息智能处理平台)。到 2006 年 1 月为止，机群计算机上在研项目 23 项（项目列表如上），有些项目已经完成。平台计算机目前已满负荷工作，单作业最大占用节点数达到 64 个。

高性能计算平台在研项目的分布如下：

国家实验室内部 7 项

院 系	项目负责人	项 目 名 称
电子系	何芸教授	高效视频编码中的关键技术
自动化系	李衍达院士	可变翻译基因的预测与分析
自动化系	李衍达院士	复杂系统意义下的生物信息学中若干问题研究
自动化系	张学工教授	人类基因组重组率估计及重组热点的定位
信研院 CPU 中心	张悠慧副研究员	可编程可配置的媒体处理器芯片设计
计算机系	陈文光副教授	大规模高性能计算机性能测试技术及方法研究
计算机系	舒继武教授	大规模三维三相油藏并行模拟

校内(国家实验室以外) 14 项

院 系	项目负责人	项 目 名 称
核研院	马远乐教授	核反应堆物理热工计算，稠油热采并行

核研院	杨永伟副研究员	加速器驱动的次临界堆中子学优化研究
航天航空学院	周力行教授	湍流燃烧的直接模拟
航天航空学院	崔桂香教授	建立高雷诺数的湍流数据库
航天航空学院	陈民副教授	合金熔体微观结构和热物理性质及其对凝固组织的影响
航天航空学院	符松教授	湍流大尺度相干结构的理论与模拟研究
生物系	孙之荣教授	转录调控研究
清华美院	杜大凯教授	三峡坝区艺术设计与文化建设总体规划设计
物理系	王崇愚院士	纳米大尺度、跨层次物理相关研究
物理系	段文辉博士	电子态研究
物理系	吴建教授	光电子材料与器件的微观结构设计
物理系	吴建教授	光电子材料及器件的结构设计与性能预测
物理系	罗豪甦、张孝文教授	高性能钙钛矿型铁电固溶体材料的制备科学和结构本质
材料系	朱静院士	Ni-Fe 合金表面化学吸附 H ₂ 的第一性原理研究

校外 2 项

院 系	项 目 名 称
胜利油田	胜利油田地质院油藏模拟
广电研究院	AVS 编码测试

这里给两个典型案例：

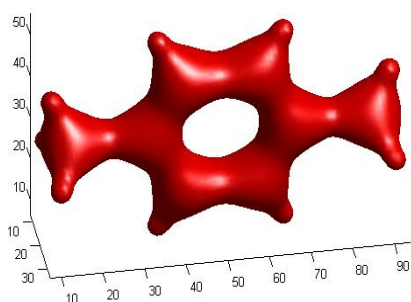
A. 纳米大尺度、跨层次物理中并行 DVDAC 方法的研究（物理系王崇愚院士）

物质材料的构成性质及应用涉及复杂大尺度跨层次问题。第一原理离散变分线性标度方法 (DV-DAC) 是将 D. E. Ellis 教授发展的第一原理离散变分 (DV) 方法与 W. Yang 教授提出的线性标度理论 - Divide and Conquer (DAC) 相结合，适合研究大尺度体系的一种新的快速电子自洽场方法。我们将其应用于大尺度体系研究，重点是碳纳米管的场发射特性研究。

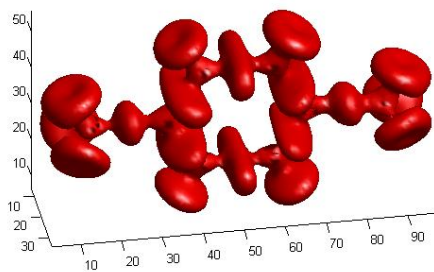
物理系开发的串行 DVDAC 程序，在原有环境下计算 1500 个原子的体系需要近两周的时间，因此迫切需要高性能计算机帮助解决大规模

计算问题。目前国际上最大体系原子数为 8000，我们的目标是扩展为可以计算 $10^4 \sim 10^5$ 原子数的体系。国家实验室平台部与物理系联合设计、开发了并行 DVDAC 软件，目前已进入大规模数据测试阶段。

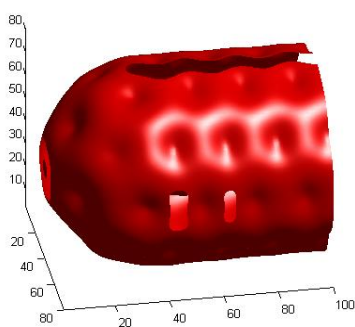
下图是我们初步利用并行 DVDAC 方法获得的苯分子和碳纳米管分子的电荷密度及电荷差分图。从图中我们可以清楚地看到电荷分布形态，可为大尺度体系研究提供必要的依据。



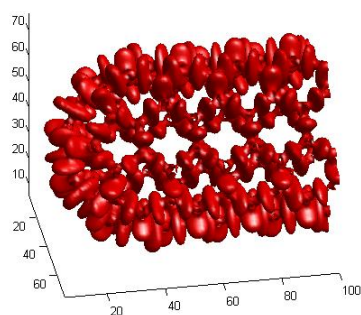
苯电荷密度图



苯电荷差分度图



碳纳米管电荷密度图



碳纳米管电荷差分度图

B. 可变翻译基因的预测与分析(自动化系生物信息实验室)

可变翻译是增加基因多样性的一种重要机制。核糖体通过识别 mRNA 序列上不同的 AUG 位点，可以产生出不同的蛋白质产物。目前已经报道了一些可变翻译的基因，这些数据都是利用实验手段得到的。我们希望通过对整个人类蛋白质组进行分析，用比对的方法从中

预测出可能的可变翻译的基因。目前，我们通过计算的方法已经预测出大量的可变翻译事件，大大提高了可变翻译基因的数目。

由于人类蛋白数据非常多，进行蛋白质数据之间的比对需要大量的计算。在本地单机上计算需时 1~2 个月左右，使用机群计算机一个星期内便可完成所有的计算和分析。目前已完成全部计算工作，正在论文撰写中，准备投给生物信息学领域最有影响的杂志 *Bioinformatics* (影响因子 5.74)。

二、EDA 平台

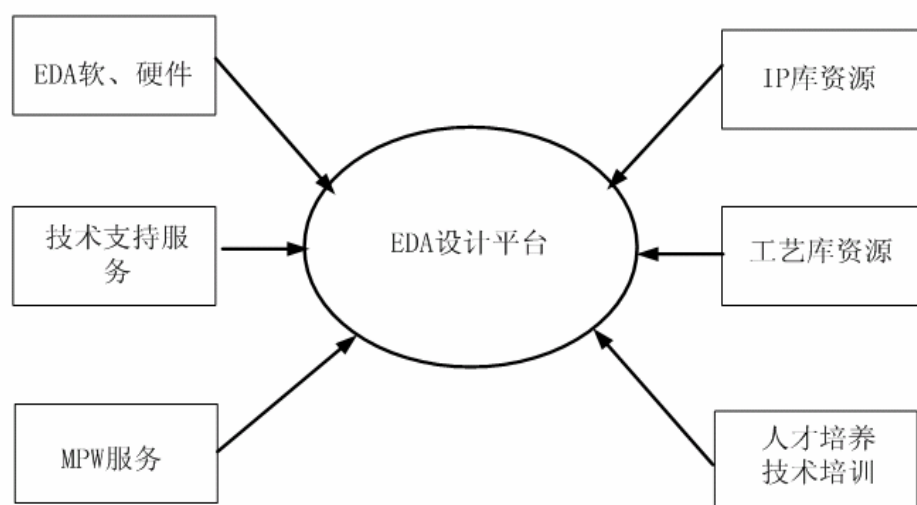
随着集成电路 (IC) 设计技术的不断发展，先进的 EDA 设计环境已成为集成电路设计、人才培养、科学技术研究必不可少的工具，这个环境中不仅包含计算机硬件、EDA 软件、集成电路设计资源 (IP-intellectual property 模块)、工艺资源，随着网络的发展，还包含更便捷的使用方式和完善的服务。建立一个为学科建设服务、为教学科研服务、软硬件环境一流且全面开放的 EDA 平台，是适应学校建设世界一流大学的需要，也是发展我国微电子技术适应信息时代的需要。

先进的集成电路设计环境对于集成电路设计、人才培养、科学技术研究是必须的，但是建设这种设计环境所需的经费非常昂贵。要让现有资源得到充分利用，避免重复建设，正是目前国家实验室集中优势、整合各方资源的目的所在。

EDA 平台的建设目标是：集中力量，建设先进开放的 IC 设计环境，包括设计工具、设计库与其他资源的平台，使该平台具有以下三种基

本职能：①高水平集成电路设计人才培养研究环境，②相关领域重大科研项目集成电路设计条件支撑，③专业技术人员培训实验条件。

作为清华大学 IC 设计领域教育与科研的网络化公共平台，EDA 平台提供 EDA 工具应用、设计资源、技术支持、人才培养等服务。该平台的基本框架如下：



以丰富的 EDA 软件、高性能服务器和高可靠性的数据存储中心为基础的先进的 EDA 平台，将大大改善我校集成电路设计的科研和教学环境，提高了 IC 设计的整体研究实力和水平，更好地为集成电路设计领域本科生、研究生课程的教学实践环节服务。我们建立了全面广泛的技术服务平台，提供专业的服务和高水准的技术支持，推动优化的设计流程、共享单元库、IP 库和相关技术资源库的建立，开展多项目芯片（MPW）服务。在此基础上，我们充分发挥国家集成电路人才培养（清华）基地的优势，加大人才培养力度，培养高水平实用化集成电路人才。同时进行一系列专业技术培训，通过与 EDA 公司合作、

开展国际交流与合作等方式，针对校内外不同群体的需要，举办不同层次的培训，内容从 EDA 工具的使用、EDA 新技术介绍到设计流程培训等等。

EDA 平台目前已拥有世界上主要 EDA 厂商(synopsys 公司、mentor 公司、cadence 公司)丰富的设计软件，涵盖了电子设计的整个流程，提供从系统级到版图级的全套设计工具和验证系统，包括系统级设计、功能验证、IC 综合及布局布线、模拟、混合信号及射频 IC 设计、全定制 IC 设计、物理验证和硬件仿真建模等，为集成电路设计提供有力的设计和分析手段。平台即将完成硬件升级更新，主要新增 SUN 高端服务器和大容量的存储设备，同时建设了一个多媒体培训教室，该教室有 20 台 IBM PC 机，可完成 EDA 和其它类型的教学培训和实验上机。

目前，EDA 平台已形成了这样一种基本格局：服务器、数据中心、EDA 工具相结合，可以提供从系统到电路各个层次的集成系统和专用集成电路仿真验证能力，系统算法研究和验证、虚拟原型设计直至实际工艺库支持的专用集成电路芯片物理设计与制造的完整环境，可提供相应工艺库以及 MPW 服务。服务器连接校园网，提供 24 小时开放、百人以上的上机环境，用户可以在任何时间、连接 Internet 的任何地点使用服务器上的 EDA 资源。