

基于RISC-V的开源GPU架构与设计探索

课程介绍

2025-10

课程基本信息

- 课程名称：基于RISC-V的开源GPU架构与设计探索
- 授课地点：深圳市学苑大道南山智园C3栋20层2005（清华大学深圳国际研究生院）
- 授课时间：十周（每周六，9:00-12:00 14:00-18:00）
- 作为全球芯片产业与技术大国战略的核心方向，RISC-V以其高效指令集、免授权费及灵活扩展能力，正重塑GPU在大模型计算中的技术格局。本课程聚焦RISC-V开源架构与GPU设计的**热门交叉前沿**，构建从指令集到众核计算架构的**探索性知识体系**，以**最短路径**实现RISC-V GPU体系架构入门，并探讨3D/3.5D Chiplet、开源EDA等技术 在开源GPU中的应用潜力
- 课程核心内容涵盖：RISC-V指令集架构（ISA）与GPU并行计算架构的融合设计，开源GPU核心模块（如流处理器、存储架构、TensorCore），核心编译器的原理与探索实现。通过C-Model仿真与FPGA实现，学生将掌握从架构建模、功能验证到部署实现的精简芯片设计流程，并探索开源RISC-V GPU的大模型（Transformer）应用，为进入AI芯片/GPU编程等前沿领域的研究生学习或职业发展奠定**核心竞争力与先发优势**

为培养下一代图灵奖获得者传递火种

- 以“培养下一代图灵奖获得者”为终极目标
 - 通过“**理论-实践-创新**”三位一体的培养模式，激发学生对核心架构的深度洞察与颠覆性技术创新的能力
- 为匹配“**图灵奖级**”人才培养需求，课程突破传统教学边界，采用“**问题驱动+学术共生**”模式
 - 设置“开源GPU架构探索”项目群，培养学生组队完成具备学术竞争力的原型设计——从分析现有开源GPU的性能瓶颈，到提出基于RISC-V扩展指令的新型并行计算单元，再到通过仿真与实测验证优化效果，全程模拟顶尖科研团队的攻关过程；
 - 与全球RISC-V基金会、开源GPU社区及顶尖高校实验室深度联动，学生可直接参与开源项目贡献，甚至主导子课题研究，积累国际学术影响力
 - 邀请GPU架构领域泰斗及头部AI芯片企业首席架构师开设专题研讨，围绕“开源RVGPU架构演进”、“开源生态对计算范式的重构”等命题展开思想碰撞

伯克利教育模式

- 注重“伯克利模式”的科研思维与产业素养的双重塑造
 - 培养学生直面技术挑战的韧性，通过“架构设计”引导思考技术创新的社会价值，通过“跨学科碰撞”（如与计算机体系结构、人工智能、集成电路的交叉研讨）**打破学科壁垒**
 - 将具备独立开展前沿架构研究的硬实力——既能深耕RISC-V GPU的微架构细节，又能跳出技术细节洞察计算本质；既掌握扎实的仿真与验证工具链，又能在开源社区中**推动技术共识**
- 致力于点燃学生对计算架构的激情，培养其成为兼具技术深度、创新勇气与学术视野的下一代计算架构引领者
- 我们相信，这里将走出定义未来十年处理器与AI计算范式的先锋，甚至孕育出**未来的图灵奖得主**

导师与助教



W. Chen



L. Ren



X. Peng



X. Shen

考评方式（无考试）

序号	考核科目	占比	考核依据
1	出勤	10%	每次出勤1分
2	调研报告 (书面)	10%	书面技术调研报告，重点考察 书面技术思维和调研能力
3	技术报告 (当堂讲述)	10%	课堂报告，重点考察技术逻辑 口头表达能力
4	设计报告	25%	探索项目的报告，重点考察工 程思维的完整度
5	设计代码	15%	探索项目的代码，重点考察电 路/架构/代码设计能力
6	小论文	30%	探索项目的小论文

探索项目

- 2110 4人 Open3DNoC (不少于4人)
- 2110 6人 OpenTensorCore (不少于6人)
- 2109 6人 OpenRVGPU (参考某一款编译生态好做或完善的RVGPU进行升级, 不少于5人, 可以参考商用IP思路)
- 2109 10人 OpenEDA+大模型 (不少于2人)
- 课程导师同意的其他探索项目, 需要与GPU有关

- 旁听学生在助教处登记后可自愿参与探索项目

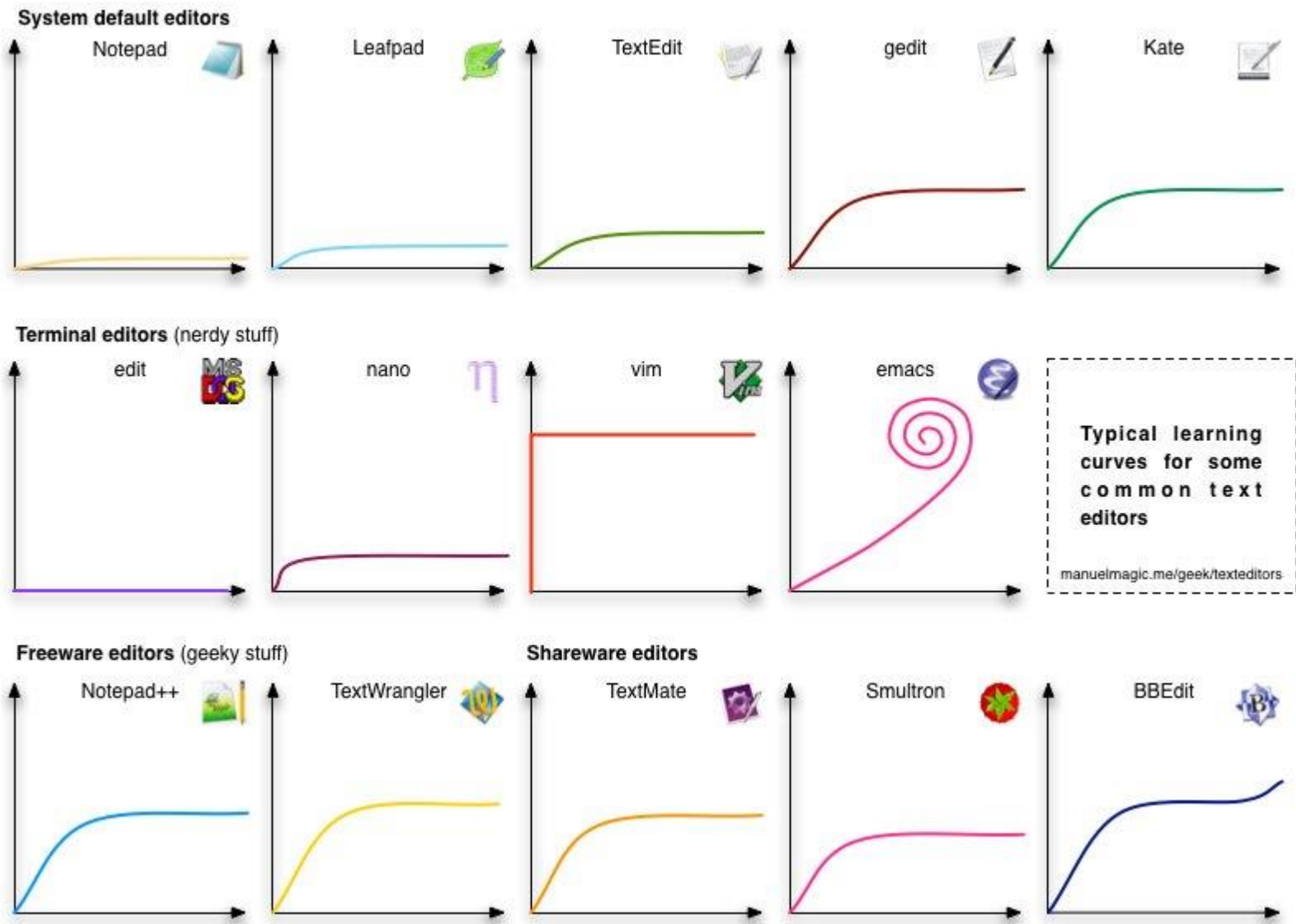
Week	Date	Daily Topic of Teaching Content	Guest/Video Sharing Content	Technical Report Content
课周	日期	授课内容当天主题	嘉宾分享内容	技术报告内容
		9:00-9:50	10:00-10:50	11:00-11:50
Week 1	10-18	Course Overview and GPU Introduction	Berkeley Approach to Graduate Education in EECS (David Patterson Video)	Data Format and Converter Design in GPU
		课程概述与GPU简介 (陈)		GPU中的数据格式与转换
Week 2	10-25	SoC Design Flow and OpenEDA	Real Chip Design Case	Technical Report
		SoC设计流程与OpenEDA (陈)		
Week 3	11-01	RISC-V Instruction Set and Architecture	A RISC-V Prodcut Chip	Technical Report
		RISC-V指令集与架构 (任)		
Week 4	11-08	GPU Architecture and Open-Source RISC-V GPU	Tsinghua University: Introduction to Ventus GPGPU Architecture	Technical Report
		GPU架构与开源RISC-V GPU (陈)		
Week 5	11-15	2.5D/3D/3.5D Architecture and Memory Compute Integration	DeToolIC: EDA Design Flow for Advanced Packaging/ Integration	Technical Report
		2.5D/3D/3.5D架构与存算一体		
Week 6	11-22	Tensor Core (TensorCore) and Transformer	Introduction to TPU Architecture	Technical Report
		张量核心 (TensorCore) 与Transformer		
Week 7	11-29	Memory Architecture and On-Chip Network (NoC)	GigaDevice: Advanced Memory (HBM or Hybrid Bonding DRAM)	Technical Report
		存储架构与片上网络 (NoC)		
Week 8	12-06	GPU Programming with CUDA/OpenCL	A Compilation Platform	Technical Report
		GPU编程与CUDA/OpenCL		
Week 9	12-13	GPU Compiler and Triton	High-Level Compiler Platform	Technical Report
		GPU编译器与Triton		
Week 10	12-20	GPU Deployment Acceleration and DeepSeek Optimization Principles	Cross-Platform Acceleration Framework	Technical Report
		GPU部署加速与deepseek优化原理		

课程技巧

- 不要求立刻理解教学内容
- 学会组队
- 善用搜索引擎
- 善用大模型
- 时间管理
- 2/8法则
- 七个习惯

不要求立刻理解教学内容

- 每种技术有不同的学习曲线
- GPU技术具备挑战性，体系庞大
- 大模型带来GPU技术飞跃，新技术层出不穷



学会组队

贝尔宾团队角色模型

- 能力互补
- 性格互补
- 有人带头
- 有人当军师
- 有人当组织委员
- 一起添砖加瓦



善用搜索引擎



- 双引号 "" 。在搜索框中输入文字时在文字的前面跟后面加上双引号，这样可以精准匹配，只能一样的文字内容及一样的文字顺序才能被搜索到。
- 书名号 《》 。在搜索框中输入文字时在文字的前面跟后面加上书名号，这样书名号内的文字不会被拆分，只搜索书名号内的字及包括书名号的文字。
- 加号+。在搜索时如果想搜索到包含有两个或三个以上的间隔文字，则用加号连接，比如搜索双11，还要包括女装，则输入双11 + 女装，注意加号前面要有个空格。
- 减号-。跟上面相反，如果在搜索中想要排除某些字眼，则用减号。
- 要搜索带有某些字母的网址，则用 “inurl:要搜的字母” 的格式。
- 要搜索带有不连续的两段字母的网址，则用 “allinurl:要搜的第一段字母 要搜的第二段字母” 。
- filetype用来搜索某种格式文件。
- 用通配符 "*" 和 "?"，一个? 代表一个字符，一个*代表多个字符。比如搜索li和o，之间间隔两个字符，那就输入li??o

善用大模型

- 能力和角色
 - 假设你是一个集成电路Verilog专家
- 洞察
 - 参考互联网检索结果，针对开源GPU
- 个性
 - 以专业的代码风格
- 实验
- 注意大模型的幻觉问题
 - 对结论和引用务必进行人工检查/确认

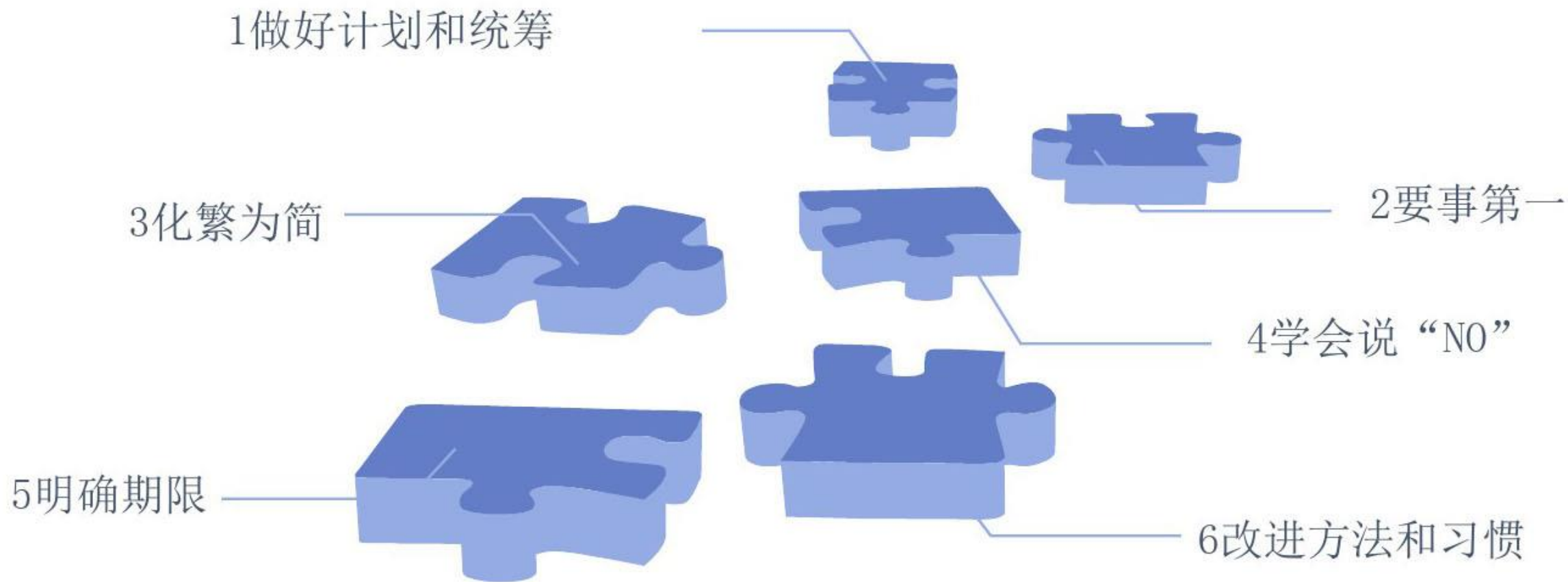


时间管理四象限法

- 四象限法则是时间管理理论的一个重要观念
- 对任务进行重要/紧急两个维度的划分
- 首先从重要性入手
- 应有重点地把主要的精力和时间集中地放在处理那些重要的工作上，这样可以做到未雨绸缪，防患于未然



时间管理的六步法



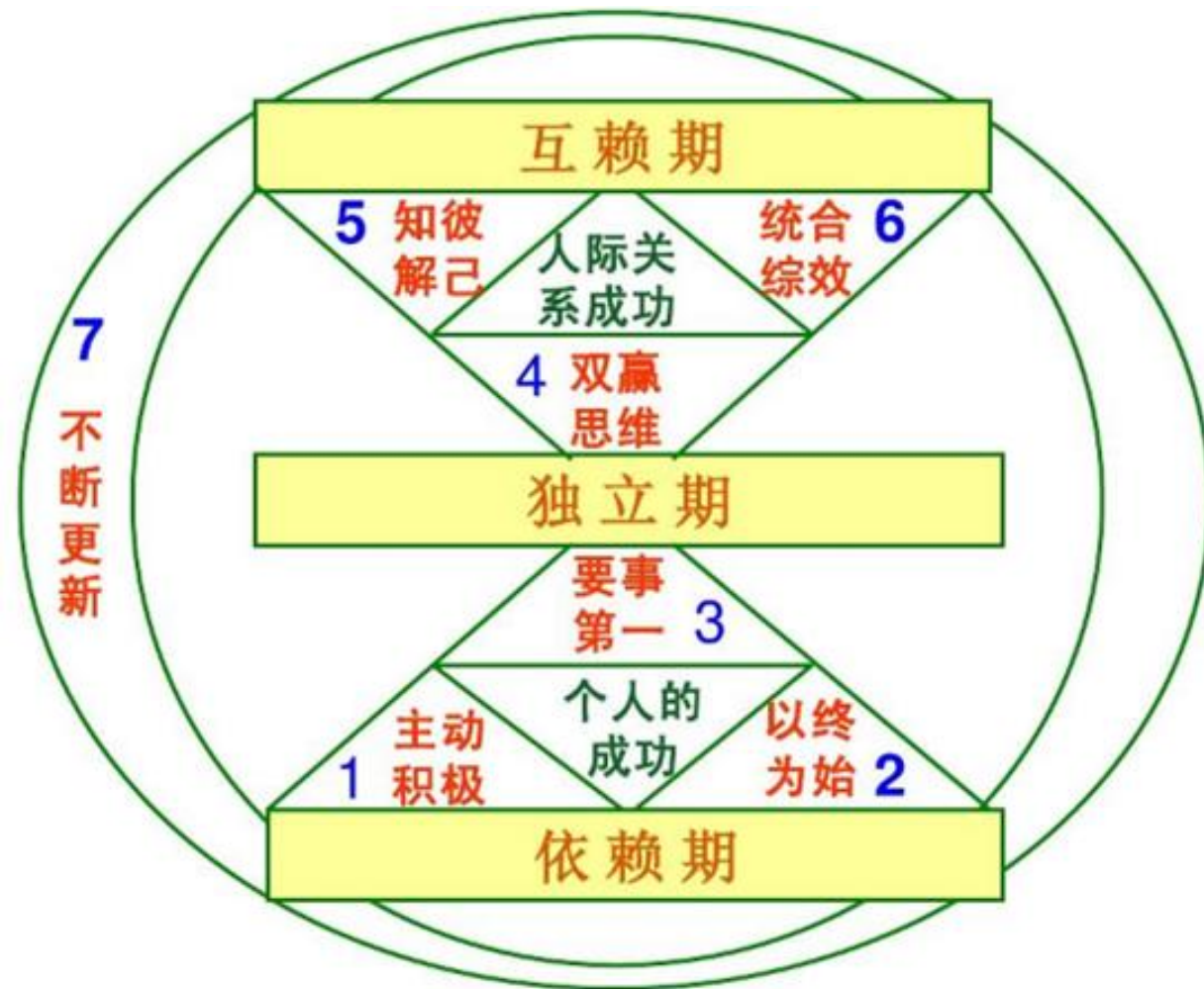
2/8法则

- 二八定律是19世纪末20世纪初意大利经济学家帕累托发现的。他认为，在任何一组东西中，最重要的只占其中一小部分，约20%，其余80%尽管是多数，却是次要的，因此又称二八定律
- 学会避免将时间和精力花费在琐事上，要学会抓主要矛盾。一个人的时间和精力都是非常有限的，要想真正“做好每一件事情”几乎是不可能的，要学会合理地分配时间和精力。要想面面俱到还不如重点突破。把80%的资源花在能出关键效益的20%的方面，这20%的方面又能带动其余80%的发展



七个习惯

- 习惯一：积极主动；Be proactive.
- 习惯二：以终为始；Begin with the end in mind.
- 习惯三：要事第一；Put first thing first.
- 习惯四：双赢思维，Win-win.
- 习惯五：知彼解己；Seek first to understand, then to be understood.
- 习惯六：统合综效；Synergize.
- 习惯七：不断更新；Sharpen the saw





清华大学深圳国际研究生院
Tsinghua Shenzhen International Graduate School



为培养下一代图灵奖获得者传递火种

