

# 智能计算体系结构

课程编号：

课程中文名称：智能计算体系结构

课程英文名称：Intelligent Computing Architectures

开课学期：秋季

学分/学时：2/32

先修课程：计算机组成原理，数字系统设计，智能计算概论，机器学习导论等

建议后续课程：计算机体系结构，深度学习，数据挖掘，智能计算系统等

适用专业/开课对象：计算机专业、电子信息专业、高等工程学院/大三本科生

团队负责人：杨建磊

责任教授：

执笔人：杨建磊

核准院长：高小鹏

---

## 一、课程的性质、目的和任务

本课程是面向理工科三年级本科生开设的专业选修课程，为《计算机组成原理和实验》与《智能计算导论》等课程的后续延伸，尝试将智能计算硬件加速的最新研究前沿技术引入课堂。本课程涉及到的理论基础较为广泛，包括智能计算领域（人工智能、机器学习、数据挖掘、图像处理、计算机视觉等）与硬件加速领域（硬件描述语言、计算机组成原理、数字系统设计、集成电路设计等）的基本知识。同时，本课程亦需要较为扎实的工程技术实践能力（算法实现、系统仿真、硬件调试等），因此是一门软硬件协同兼顾且重于研究探索的实践课程。

本课程以智能计算领域几个典型应用为研究对象，讲述从软件算法到硬件结构的设计和映射原理及方法，使学生理解从系统结构的角度实现智能计算的工作过程，培养学生直接采用硬件系统结构分析、设计和实现智能计算算法的基本技能。本课程要求学生自主设计及开发几项典型的智能计算功能模块并最终搭建一项完整的功能原型机，使其掌握硬件加速智能计算的工程化方法，并具备软硬件异构协同设计的基本素养和学术视野，为今后更深入的理论学习和研究实践奠定良好基础。

本课程是学院规划和建设的“本科前沿课程”之一，实践性较强，课程教学目标高，知识内容前沿、知识面宽、知识运用综合，实验体系具有很高的难度与强度。本课程适合渴望学习前沿技术与敢于面对挑战的学生，选修本课程的学生应有充分心理准备。具体的教学目标可分解为以下几点：

- (1) 了解智能计算领域中包括人工智能、机器学习、模式识别、数据挖掘、图像及视觉处理等方向的几类基本模型和算法；了解智能计算硬件加速的科学问题、实验技术、工

具资源以及该领域最新的研究前沿；

- (2) 掌握数字系统设计的基本方法、原理和相关研究方法，注重工程思维与创新思维的培养，尤其是具备将智能计算方法按照功能层次和数据流图进行模块化实现及系统集成的思维能力，并采用数字系统设计方法进行硬件实现、调试及性能测试分析；
- (3) 具备利用 Verilog 语言和 Modelsim、Vivado 及 HLS 等工具设计和测试智能计算算法模块并进行系统集成、功能测试、性能分析和论证的能力；
- (4) 具有获取和利用标准、规范、手册、图册等有关技术资料的能力；
- (5) 掌握计算机和嵌入式数字系统工程中方案设计与性能测试及系统优化的实验方法，获得实验技能的基本训练；
- (6) 掌握智能计算系统模型建立、分析求解和设计方案论证的理论和方法，培养学生分析和解决智能计算加速方法等工程实践问题的创新设计能力。

#### 课程目标对毕业要求的支撑关系

毕业要求	课程目标对毕业要求的支撑关系
<b>毕业要求 1：工程知识</b> 能够将数学、自然科学、工程基础和专业知用于解决复杂计算机工程问题。	<b>课程目标：1、3</b>
<b>毕业要求 2：问题分析</b> 能够应用数学、自然科学基本原理，并通过文献研究，识别、表达、分析复杂计算机工程问题，以获得有效结论。	<b>课程目标：2、4、5、6</b>
<b>毕业要求 3：设计/开发解决方案</b> 能够设计针对复杂计算机工程问题的解决方案，设计满足特定需求的计算机系统，并能够在设计环节中体现创新意识，考虑法律、健康、安全、文化、社会以及环境等因素。	<b>课程目标：2、5、6</b>
<b>毕业要求 4：研究</b> 能够基于科学原理并采用科学方法对复杂计算机工程问题进行研究，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。	<b>课程目标：2、6</b>

毕业要求	课程目标对毕业要求的支撑关系
<p><b>毕业要求 5：使用现代工具</b></p> <p>能够在计算机工程实践中开发、选择与使用合理有效的技术、资源、现代工程工具和信息技术工具，并了解其局限性。</p>	<p>课程目标：3、4</p>
<p><b>毕业要求 6：工程与社会</b></p> <p>具有追求创新的态度和意识，掌握基本的创新方法，以及综合运用理论和技术手段设计复杂计算机系统与过程的能力；设计过程中能够综合考虑社会、经济、文化、环境、法律、安全、健康、伦理等制约因素。</p>	<p>课程目标：1</p>
<p><b>毕业要求 7：环境和可持续发展</b></p> <p>了解与本专业相关的职业和行业的生产、设计、研究与开发、环境保护和可持续发展等方面的方针、政策和法律、法规；能够正确认识专业工程实践对环境和社会可持续发展的影响，合理评价专业工程实践和复杂工程问题解决方案对社会、健康、安全、法律及文化的影响。</p>	<p>课程目标：4</p>
<p><b>毕业要求 8：职业规范</b></p> <p>具有坚定正确的政治方向，良好的思想品德、社会公德和职业道德；具有人文社会科学素养、社会责任感；具有良好的身体素质和心理素质，能履行建设祖国和保卫祖国的神圣义务。</p>	<p>课程目标：4、5、6</p>
<p><b>毕业要求 9：个人和团队</b></p> <p>具有在多学科团队中发挥重要作用的能力。</p>	<p>课程目标：5</p>
<p><b>毕业要求 10：沟通</b></p> <p>能够就复杂计算机工程问题与业界同行及社会公众进行有效沟通与交流，包括撰写报告和设计文稿、陈述发言、清晰表达个人见解等，并具备一定的国际视野，能够在跨文化背景下进行沟通和交流。</p>	<p>课程目标：4、5</p>
<p><b>毕业要求 11：项目管理</b></p> <p>具有一定的组织与工程管理能力、表达与人际交往能力以及在多学科背景下的团队中发挥作用的能力。</p>	<p>课程目标：4</p>

毕业要求	课程目标对毕业要求的支撑关系
<b>毕业要求 12：终身学习</b> 具有自主学习和终身学习的意识，有不断学习和适应发展的能力。	<b>课程目标：4、5、6</b>

## 二、课程内容、基本要求及学时分配

本课程介绍智能计算系统结构知识和工程实践方法，主要包括算法实现、硬件加速、系统架构、软硬件实现等问题。

序号	教学内容	基本要求及重点和难点	学时	教学方式	对应的教学目标	支持毕业要求指标点
1	<b>智能计算加速前沿</b> 深度学习加速器； 大数据处理加速； 计算机视觉加速	<b>基本要求：</b> 了解智能计算加速器未来发展潜力和应用前景。 <b>重点：</b> 进行充分的文献调研和阅读，意识到当前人工智能的大趋势。 <b>难点：</b> 理解传统智能计算系统结构所遇到的计算能耗效率及性能瓶颈问题。	2	课堂讲授+课后调研+答疑	1, 4	1, 6
2	<b>智能计算模型与方法</b> 聚类；分类器；回归分析；大图计算；卷积神经网络；玻尔兹曼机；贝叶斯推理；特征提取；图像分割、变换等	<b>基本要求：</b> 掌握机器学习和数据挖掘的基本数学模型和理论分析方法。 <b>重点：</b> 理解智能计算的基本原理，实现基本的分类器和神经网络算法。 <b>难点：</b> 编程实现及调试，帮助理解算法基本原理。	4	课堂讲授+课后作业+答疑	1, 4, 6	2, 3, 4

序号	教学内容	基本要求及重点和难点	学时	教学方式	对应的教学目标	支持毕业要求指标点
3	<p><b>专用计算架构平台</b></p> <p>基于 GPU/FPGA/ASIC 加速; 开发语言、环境与工具; 逻辑仿真; 逻辑综合; 高层次综合; 物理设计; 设计实例</p>	<p><b>基本要求:</b> 了解目前主流基于 GPU/FPGA/ASIC 平台的加速方法, 掌握 FPGA 开发环境和调试技能。</p> <p><b>重点:</b> 理解如何突破传统计算架构的瓶颈并直接以硬件实现算法的优势, 实现在可编程逻辑硬件平台上实现智能算法。</p> <p><b>难点:</b> 实践 FPGA 开发环境并掌握基本流程。</p>	4	课堂讲授+课后实践+答疑	3, 4, 5	1, 3, 5
4	<p><b>算法到硬件映射方法</b></p> <p>系统建模; 功能划分; 数据流图; 模块设计; 系统集成; 浮点精度分析; 流水线; 并行化; 性能分析; 能耗效率分析</p>	<p><b>基本要求:</b> 掌握算法到硬件的映射实现基本原理和方法流程, 并掌握将系统结构优化方法应用到设计思想中。</p> <p><b>重点:</b> 理解映射方法的原理和依据以及优化技巧。</p> <p><b>难点:</b> 以数字系统的硬件系统结构思维来理解如何实现软件算法, 跨越了操作系统和编译原理等中间层, 挑战较大。</p>	6	课堂讲授+课后作业+答疑	2, 5, 6	2, 3, 4
5	<p><b>神经网络加速器设计</b></p> <p>模型训练; 数据集; 系统 I/O; 加速器结构; 时序控制; 数据流图;</p>	<p><b>基本要求:</b> 了解主流深度学习架构, 包括寒武纪、Eeyriss、TPU 等; 手动实现单层神经网络算法或基于 Caffe 框架实</p>	10	课堂讲授+课后实践+答疑	2, 3, 4, 5, 6	1, 2, 3, 4, 5

序号	教学内容	基本要求及重点和难点	学时	教学方式	对应的教学目标	支持毕业要求指标点
	模型压缩; 存储访问瓶颈; 数据复用; 卷积单元设计; FPGA 平台验证	<p>现深度神经网络的训练和测试, 完成神经网络测试过程的结构设计, 用 Verilog 实现模块单元设计, 并实现整个测试过程的系统结构。</p> <p><b>重点:</b> 理解典型神经网络架构设计原理与技巧; 掌握神经网络顶层结构设计思想, 模块实现并进行单元测试之后再行系统集成。</p> <p><b>难点:</b> 仿真通过后采用 FPGA 平台进行实测, 时序控制和数据通路等调试技巧尤其需要注意。</p>				
6	<p><b>专用集成电路实现</b></p> <p>行为描述; 逻辑综合; 系统仿真; 物理设计; 时序优化; 性能分析</p>	<p><b>基本要求:</b> 了解采用 ASIC 实现神经网络加速器的基本思路、方法以及最新研究成果。</p> <p><b>重点:</b> 加速器设计经过 FPGA 平台验证后, 基于工艺库进行物理设计实现 ASIC。</p> <p><b>难点:</b> 针对 ASIC 的物理设计流程比 FPGA 复杂许多。</p>	4	课堂讲授+课后调研+答疑	2, 4, 5, 6	1, 3, 5, 6
7	<p><b>加速器应用前景</b></p> <p>自动驾驶; 无人飞行器; 视听觉信息处理; 机器人; 大数据处理</p>	<p><b>基本要求:</b> 了解智能计算加速器在未来人工智能应用中的重要性。</p> <p><b>重点:</b> 如何开阔视野, 以智能硬件的角度来理解未来高</p>	2	课堂讲授+课后调研+答疑	1, 4	1, 6

序号	教学内容	基本要求及重点和难点	学时	教学方式	对应的教学目标	支持毕业要求指标点
		效能计算系统的潜力。 <b>难点：</b> 人工智能新概念、智能硬件新模式的理解。				

### 三、课内外教学环节及基本要求

在教学过程中体现“学生主体、教师主导”的教学思想，提倡启发式、讨论式教学，突出对学生逻辑思维、工程创新及实践能力的培养。在讲授过程中做到由浅入深、由表及里、循序渐进，同时注重举例和类比，并加入该领域最新研究进展，活跃课堂，使课堂讲授生动有趣。

本课程课堂授课 32 学时，周学时为 2 学时，教学方式包括课内教学和课外教学两部分，授课语言为中文。课内教学采用多媒体讲义，包括课堂讲授、课堂讨论、课堂展示等。课外教学包括课下自学、课外作业、系统开发、实验报告等环节。

部分教学内容采用项目式教学模式，注重学生实践动手能力，为学生提供 FPGA 等软硬件实验开发平台，提高学生利用现代工具解决工程实践问题的能力，并将前沿研究带入课堂，给学生提供接触最新研究成果的机会，以及亲自参与最新科技进展的实践活动中来，培养其浓厚的学术兴趣和优良的科研素养。

### 四、考核方式及成绩评定

本课程以平时成绩与项目实践相结合的方式考核，总分为 100 分。

**平时成绩占 40%：**包含课程作业与交流讨论等。其中包含 3 次作业与 1 次文献调研：

作业 1: Zynq FPGA 开发平台与实验环境，占 10 分；

作业 2: 深度学习框架，Caffe/TensorFlow/PyTorch，占 10 分；

作业 3: Verilog 数字模块设计（稀疏矩阵向量乘法），完成仿真、综合、验证，占 10 分；

文献调研：阅读深度学习处理器领域相关代表性论文，展示 PPT 并进行交流讨论，占 10 分。

**项目实践占 60%：**包含算法、系统实现、测试及报告、展示等。其中包括 3 次课程实践：

项目 1: 基于 Zynq FPGA 平台的 MAC 阵列设计，要求经过 FPGA 上板测试验证，占 15 分；

项目 2: 基于 Zynq FPGA 平台的 LeNet 网络架构设计，要求上板测试验证，占 20 分；

项目 3: 基于 Zynq FPGA 平台的 VGGNet 网络架构设计，要求上板测试验证，占 25 分。

### 五、教材和参考资料

## 参考资料:

- [1] 智能科学 (第二版), 史忠植著, 清华大学出版社, 2013 年。
- [2] Verilog 数字系统设计教程, 夏宇闻著, 北京航空航天大学出版社。
- [3] Verilog HDL 高级数字设计 (第二版), Michael D. Ciletti 著, 李文军、林水生、波等译, 电子工业出版社, 2014 年。
- [4] Verilog HDL 数字设计与综合, 帕尔尼卡著, 电子工业出版社, 2004 年。
- [5] Vivado Design Suite User Guide, Xilinx 公司, 2016 年。
- [6] 昇腾 AI 处理器: 架构与编程, 梁晓峤著, 清华大学出版社, 2019 年。
- [7] 当计算机体系结构遇到深度学习, 杨海龙、王锐译, 机械工业出版社, 2019 年。
- [8] Computer Architecture: A Quantitative Approach, John L. Hennessy 著, Morgan Kaufmann Publishers, 2011 年。
- [9] Computer Systems: A Programmer's Perspective, Randal E. Bryant 和 David R. O'Hallaron 著, Pearson, 2015 年。
- [10] 深入理解计算机系统, Randal E. Bryant 和 David O'Hallaron 著, 机械工业出版社, 2011 年。
- [11] Computer Organization and Design: The Hardware/Software Interface, David Patterson 著, Morgan Kaufmann Publishers, 1997 年。
- [12] 机器学习, 周志华著, 清华大学出版社, 2016 年。
- [13] 深度学习: 21 天实战 Caffe, 赵永科著, 电子工业出版社, 2016 年。
- [14] Learning Deep Architectures for AI, Yoshua Bengio 著, Now Publishers Inc., 2009 年。
- [15] 统计学习方法, 李航著, 清华大学出版社, 2012 年。
- [16] Hardware Acceleration of EDA Algorithms: Custom ICs, FPGAs and GPUs, Sunil P Khatri 和 Kanupriya Gulati 著, Springer, 2010 年。
- [17] Machine Learning in Silicon, UIUC ECE 598 NS 课程, Naresh R Shanbhag, 2016 年。
- [18] Hardware acceleration for machine learning and big data analytics, 哥伦比亚大学 E6895 Advanced Big Data Analytics 课程, Ching-Yung Lin, 2015 年。
- [19] High Performance Computing with a Focus on Hardware Acceleration Technologies, University of Western Ontario 大学 CS4435b-CS9624b 课程, Marc Moreno Maza, 2010 年。
- [20] Advanced System Design using HDLs, University of North Carolina at Charlotte 大学 ECGR 4146/ECGR5146 课程, Arindam Mukherjee, 2011 年。

[21] High-Performance Hardware for Machine Learning, NIPS Tutorial, William Dally, Stanford University & NVIDIA Corporation, 2015 年。

## 六、其它

本课程提供 Xilinx FPGA/Zynq 开发硬件平台, 包括支持 HLS 高层次综合与嵌入式双核 ARM 处理器的 SDK 开发环境, 以及深度学习框架及 GPU 服务器等环境支持。