

基于超长指令字的顶点染色处理器设计

徐起超, 杜慧敏, 刘青楠

(西安邮电大学 电子工程学院, 陕西 西安 710121)

摘 要: 为了提高嵌入式图形处理器 GPU(Graphic Process Unit)中顶点染色处理器, 设计了一款超长指令字格式的可编程顶点染色处理器, 采用六级流水线实现, 每条指令在同一个周期最多执行 7 种操作, 软硬件协同设计, 降低了功耗. 采用基于 FPGA 的验证方式, 可编程顶点染色处理器在 Xilinx Virtex-7 FPGAs V2000T 上最大工作频率达到 50 MHz, 顶点的处理速度达到 0.16 M/s, 处理一个顶点平均 44 个周期, 在 Synopsys 公司 Design Compiler 工具 130 μm 工艺综合下, 主频 150 MHz, 功耗约为 177.7428 mW.

关键词: 图形处理器; 顶点染色; 超长指令字(VLIW); 可编程

中图分类号: TP399

文献标识码: A

文章编号: 1000-7180(2018)10-0013-06

Vertex Shader Processor Design Based on Very Long Instruction Word

XU Qi-chao, DU Hui-min, LIU Qing-nan

(School of Electronics Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: In order to improve the vertex shader processor in the Graphic Process Unit (GPU), a programmable vertex shader processor with a very long instruction word format is designed and implemented in a six-stage pipeline. Each instruction has a maximum of one cycle Implementation of seven kinds of operations, hardware and software co-design, reducing power consumption. Using FPGA-based verification, the programmable vertex shader processor achieves a maximum operating frequency of 50 MHz on the Xilinx Virtex-7 FPGAs V2000T, a vertex processing speed of 0.16 M/s, and an average of 44 cycles for a vertex. At Synopsys Design Compiler tools 130 μm process synthesis, clocked at 150 MHz, power consumption is about 177.7428 mW.

Key words: GPU; vertex shader; very long instruction word; programmable

1 引言

随着智能手机、平板电脑、游戏机快速发展, 高性能的嵌入式系统中都包含有图形处理功能, 一般采用高性能的嵌入式 CPU(Central Processing Unit)和专用嵌入式图形处理单元 GPU(Graphic Process Unit)实现. 但是, 在嵌入式系统中, 由于受到面积、功耗的限制, CPU 和 GPU 如何配合高效地完成图形处理功能是业界关注的热点问题之一^[1].

图形流水线一般包括顶点染色、图元装配、光栅

化、像素染色和现实等阶段. 在现代 GPU 中, 顶点染色和像素染色是用户可编程处理器实现的. 顶点染色处理器主要完成浮点的光照计算、法向量计算以及旋转、缩放、平移等几何变换等功能, 其处理速度对 GPU 性能产生很大的影响, 因此, 学术界在可编程顶点染色处理器的体系结构方面进行了大量的研究^[2-8].

本文通过分析顶点染色处理功能, 提取其完成的功能和操作, 提出用于顶点染色的 VLIW 指令集, 同时结合高效能的 ARM 处理器, 合理划分了软

硬件实现,降低了 VLIW 处理器的复杂度. 本文设计采用六级流水线,实现了顶点染色处理器功能,并在开发板上进行了验证.

2 顶点染色处理器的功能及算法分析

2.1 顶点染色处理器的功能

顶点染色处理器在图形处理器中占据重要地位,主要实现顶点的光照计算、几何变换以及法向量归一化计算. 在计算机图形学中^[9],光照计算就是利用一些特定的算法来计算出各个顶点经过光线照射后产生的颜色值,进而来模拟物体表面实现光线照射后,使在人眼中呈现的图像更加真实. 几何变换实现空间坐标的变换,法向量归一化计算是在 Gouraud 模型与 Phone 模型^[9-10]中,使法向量单位化.

2.2 算法分析

2.2.1 几何变换

几何变换主要实现模型视图变换,为了便于在屏幕上显示图形对象,将所有在模型空间中形成的对象放在同一参考系视觉空间中,要进行坐标系转换^[9]. 这种空间坐标系的转换是通过顶点坐标与旋转矩阵、平移矩阵或者缩放矩阵相乘来实现的,涉及的运算矩阵类型包括绕 x 轴旋转矩阵:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

绕 y 轴旋转矩阵

$$\begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

绕 z 轴旋转矩阵

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

平移矩阵

$$\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

缩放矩阵

$$\begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

2.2.2 法向量归一化

法向量归一化计算^[9-11]就是把一个长度不为 1 的向量单位化,使该法向量的单位长度变为 1,即

$$\vec{N}_v = \frac{\vec{N}}{|\vec{N}|} \quad (1)$$

式中, $\vec{N}$ 表示归一化前的法向量; $|\vec{N}|$ 表示法向量的长度; $\vec{N}_v$ 表示归一化后的单位法向量.

2.2.3 光照处理

在光照处理过程中,Gouraud 模型和 Phone 模型^[9-10]是两种经典的光照模型,两种模型计算顶点的光照强度的方式都是计算光照方程^[11],即

$$I = I_a + I_d + I_s =$$

$$k_a I_a + f(d) I_p [k_d (\vec{N} \cdot \vec{L}) + k_s (\vec{V} \cdot \vec{R})^n] \quad (2)$$

式中, I 表示计算顶点的最终光照强度; I_a 表示环境光强度; I_d 表示漫反射光强度; I_s 表示镜面反射光强度; I_p 表示聚光灯光照强度; k_a 、 k_d 和 k_s 分别表示环境光系数、漫反射光系数和镜面反射光系数; $\vec{N}$ 表示顶点单位法向量; $\vec{L}$ 表示反射光单位法向量; $\vec{V}$ 表示进入人眼光的单位法向量; $\vec{R}$ 表示镜面反射光的单位法向量; n 表示镜面反射光指数; $f(d)$ 表示衰减因子,是一个跟光源与顶点之间距离 d 相关的函数^[11],即

$$f(d) = \frac{1}{k_c + k_1 d + k_q d^2} \quad (3)$$

式中, d 表示顶点与光源之间的距离; k_c 表示常量衰减因子; k_1 表示线性衰减因子; k_q 表示二次衰减因子.

在实际的关照计算中,可能会存在有多盏聚光灯对顶点进行照射,因此式(2)需要修改为

$$I = k_a I_a + f(d) \sum_{i=0}^m I_i [k_d (\vec{N} \cdot \vec{L}_i) + k_s (\vec{V} \cdot \vec{R}_i)^n] \quad (4)$$

将所有聚光灯的光照结果进行累加,计算顶点的最终光照结果.

Gouraud 光照模型的核心思想是先对顶点的法

向量和与顶点相邻的多边形的法向量计算一个平均值,再把平均值代替顶点的法向量,最后将其代入光照式(4)计算顶点的最终光照强度.而 Phone 模型不需要计算平均法向量,而是直接将顶点的法向量代入光照式(4),最终计算顶点光照强度.在上述两种光照模型中,都需要做除法运算、超越函数计算,计算量都比较大.在实际应用中,为了支持对选择的光照模型进行更快的处理,定义了顶点染色处理器指令集.

3 指令集设计与指令格式设计

3.1 指令集设计

根据顶点染色处理器的功能以及对顶点处理算法的分析,并针对 OpenGL ES Shading Language Space^[12]中所规范的定义,本文定义了适用于嵌入式可编程顶点染色处理器的一组 128 位指令长度的 VLIW 指令集,总共分为五大类指令.

- (1)用于计完成超越函数计算的查找表指令;
- (2)用于完成程序跳转的分支控制指令;
- (3)完成数据运算类指令;
- (4)数据控制类指令;
- (5)地址加载类指令.

在以上设计的指令中,除了超越函数的逼近运算需要两条指令在三个周期内完成,其他指令都是在一个周期内完成运算.且上述表格中的指令用于矢量数据,针对长度为 4 的矢量和较小的矢量的操作进行优化.

在本设计中,考虑到除法运算实现的复杂度高,硬件成本高,用软件实现.

超越函数(如均方根,倒数,对数,三角函数和指数等)用二次函数逼近方法实现.超越函数的逼近计算,由软硬件协同共同完成,硬件实现二次函数的系数查找与乘加计算,软件通过编译器编译指令来实现超越函数.

3.2 硬件指令格式设计

可编程顶点染色处理器采用 128 位固定长度的指令系统,即每条指令字总共 16 个字节.其中高 64 位设置了跳转指令的程序地址和各功能单元执行指令的操作码,低 64 位设置了寄存器文件的读数据地址以及工作寄存器与常量寄存器的写数据地址.

在每条超长指令字中,有多个并行操作,每个操作都是对每个向量数据的分量进行操作.在顶点染

色处理器处理顶点数据过程中,如果读取指令标识无效,进入流水线的指令为空指令.并且顶点染色处理器的指令寻址方式总共五种,包括寄存器直接寻址、寄存器间接寻址、相对寻址、寄存器偏移量寻址和自动递增寻址.

4 顶点染色处理器的 VLIW 设计

4.1 流水线设计

流水线是当今设计处理器的一种重要技术,能够将多条指令重叠执行^[11].一条指令的执行需要多个操作,流水线技术能够是这些指令操作并行,能够加快处理器的速度.

依据所设计的指令和指令格式,本项目组设计的顶点染色处理器采用六级宏流水线,结构如图 1 所示.

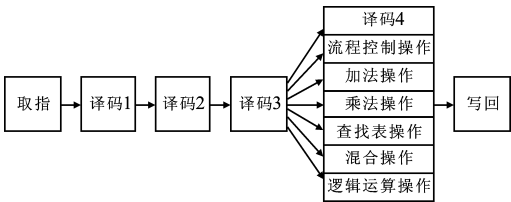


图 1 六级流水线结构

在整个流水过程中,在取指令之后,该处理器是对每一阶段所需操作的指令部分进行译码,这样可以减少处理器的功耗.除此之外,流水结构还包括计算常量寄存器地址、读寄存器文件、选择执行单元所需要的数据、执行计算、写回.

第一级流水线:访问指令存储器,读取将要执行的指令数据;

第二级流水线:对指令寄存器中的指令进行译码,并计算访问常量寄存器的地址;

第三级流水线:访问寄存器组获取操作数;

第四级流水线:根据指令结构选择执行阶段所需要的操作数;

第五级流水线:根据指令结构,译码出各执行单元的操作码,执行相应运算;流程控制操作根据跳转指令,产生新的 PC 值,加载访问常量寄存器的地址;

第五阶段:写回阶段将执行单元中的计算结果按照规则写入寄存器文件中.

处理器的指令时序如下图 2 所示:

对超长指令字格式的顶点染色处理器的设计,

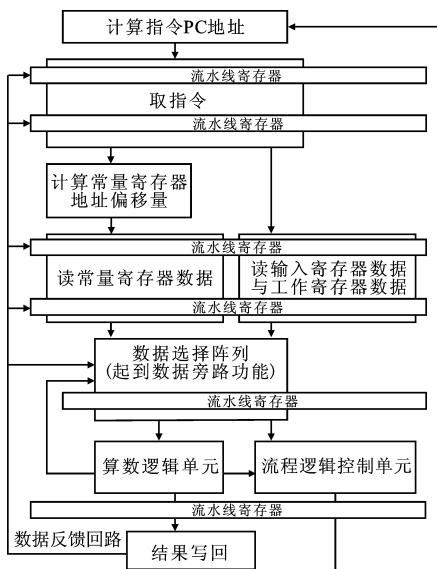


图2 指令时序结构图

为了处理流水线冒险,进行了合理的软硬件划分. 为了防止结构冒险,设计的顶点染色处理器采用哈佛结构^[13-14],将指令存储与数据存储分开. 由于超长指令字格式的处理器高度依赖编译器,数据冒险的判断放在编译器中完成,把判断的结果以指令的形式通过硬件电路输入数据选择器为各算术逻辑单元选择出所需要的数据. 控制冒险的判断在控制逻辑电路中执行,而对控制冒险的处理在编译器中完成,如果判断跳转条件真实,则编译器会在指令中给出程序跳转的地址.

4.2 顶层结构设计

根据顶点染色处理器的功能和设计的指令集,设计出顶点染色处理器的硬件架构,如图3所示.

由于浮点除法运算占用硬件资源太多,综合考虑该顶点染色处理器的效率,在该设计中不包含除法运算单元和超越函数,而是把除法运算和超越函数的实现放在高效能的 ARM 处理器上完成.

本文提出的基于超长指令字的顶点染色处理器由两个加法器、两个乘法器、一个查找表单元(LUT)、一个混合单元、一个数据输入选择阵列、一个预处理单元、一个输出选择阵列、一个工作寄存器组、一个常量寄存器、一个输入寄存器组、一个输出寄存器组、一个顶点加载单元以及一个指令存储单元.

(1)指令存储器:指令存储器存储将要执行每个顶点计算的指令,最多可以存储 512 条指令.

(2)加法操作单元:顶点染色处理器包含两个完

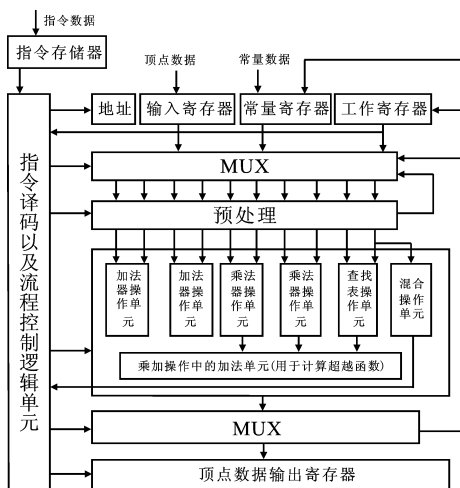


图3 顶点染色处理器结构图

整的 32 位浮点加法器. 加法器单元不仅可以执行标准的浮点加法操作,可以执行其他操作,例如计算最小值、最大值和取整.

(3)乘法操作单元:顶点染色处理器包含两个完整的 32 位浮点乘法器. 乘法器单元不仅执行标准的浮点乘法运算,还可以作为在查找表单元查找出系数后计算超越函数逼近计算过程中特殊乘法单元的一部分.

(4)查找表单元:LUT 单元执行查表作为计算倒数,对数和其他超越函数的基础. LUT 单元依据输入的数据,查找出所选超越函数近似方程的系数,最后的结果由乘加单元计算出.

(5)混合操作单元:除了一些其他操作之外,混合操作单元还处理定点数据与浮点数之间的转换.

(6)流程控制逻辑单元:流程控制逻辑单元处理程序流程,并控制一些特殊的操作. 除了正常的程序流程操作(例如顺序流程,有条件或无条件的分支以及函数调用和返回)外,流程控制单元逻辑还处理:硬件循环、输入寄存器组切换以及常量寄存器写入.

硬件循环功能包含一个专用的循环计数器和一个特殊的地址寄存器. 可以重复一个预定义次数的代码序列,有选择地更新地址寄存器,而无需使用正常的算术单元来更新循环计数器. 在流程控制逻辑单元中,硬件环路不能嵌套,而且还要维护一个 4 元素的调用堆栈.

(7)工作寄存器组:工作寄存器是一组 16 个向量寄存器,相当于临时寄存器,可以存储指令处理的中间值. 每个寄存器包含四个 32 位浮点值.

(8)输入寄存器组:输入寄存器存储将要处理的顶点数据(x, y, z, w),这些顶点数据被顶点染色

处理器读取并执行计算. 该寄存器文件有 16 个输入寄存器,每个寄存器存储的数据都是四个 32 位浮点数值 的向量.

(9)输出寄存器:输出寄存器是 16 个矢量寄存器,顶点染色处理器可以存储计算结果. 每个寄存器都是四个 32 位浮点数值 的向量.

(10)常量寄存器:常量值寄存器是一组存储常量数据的向量寄存器,存放顶点染色过程中需要的顶点各种属性.

(11)输入多路复用器:输入多路复用器把每个功能单元能够使用寄存器值或先前计算的结果作为输入.

(12)输出多路复用器:对于输出寄存器,每个分量 x,y,z,w 都有一个输出值选择器,选择 ALU 单元各功能模块计算出的结果

(13)顶点加载单元:加载顶点染色处理器将要处理的顶点数据.

5 验证与结果分析

5.1 顶点染色处理器验证

所实现的顶点染色处理器通过仿真后,采用 PPGA 开发板形式进行板级验证,如图 4 所示. 搭建的硬件平台主要由 Xilinx Zynq-7000 ZC706 开发板、FMC 板卡和 Xilinx Virtex-7 FPGAs V2000T 开发板三大部分组成.

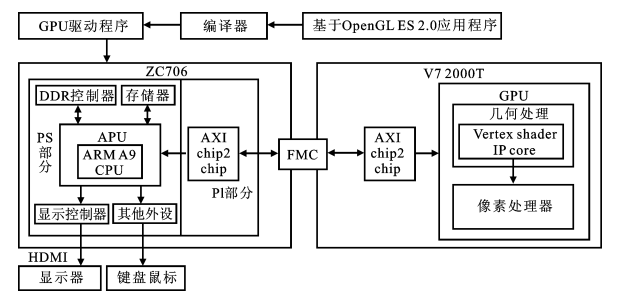


图 4 硬件系统平台的示意图

在 Virtex-7 FPGAs V2000T 开发板上,把设计的顶点染色处理器封装成 IP 核,并嵌入在 GPU 中,以此用于实现图形处理;ZC706 开发板视为整个系统的上位机,控制着嵌入式图形处理器;FMC 板卡用于连接 ZC706 开发板和 V2000T 开发板,在两个开发板之间,完成数据交互. 在搭建的硬件系统开发平台中,在 ZC706 开发板的 ARM 处理器上移植 Linux 操作系统,挂载 GPU 设备驱动程序,并且开

发 OpenGL ES 2.0 应用程序进行测试,从而验证顶点染色处理器的性能.

5.2 性能分析

综合测试结果如图 5 所示,给出了光照球、光照环、Box、Teapot、Cow 和 Rabbit 六种典型的图形应用程序在硬件平台上的运行结果,并统计了图形应用程序的顶点处理速度. 其中 Box 和 Teapot 图形应用程序每帧分别需要处理大约 4 404 和 6 731 个顶点数据,Cow 和 Rabbit 图形应用程序每帧分别需要处理大约 21 556 和 208 381 个顶点数据. 这些测试程序都需要坐标变换和光照计算,包含了各种顶点数据和各种运算,处理多达十几万个顶点数据. 在充分测试了顶点染色处理器的各种功能和稳定性情况下,本项目组设计的 GPU 在 FPGA 开发板上运行 50 MHz,顶点染色处理器的处理顶点速度达到 0. 16 M/s,处理一个顶点大约需要 44 个周期,每帧顶点染色花费的时间约 1. 3 s (视 OpenGL ES 2. 0 应用程序复杂度而定).

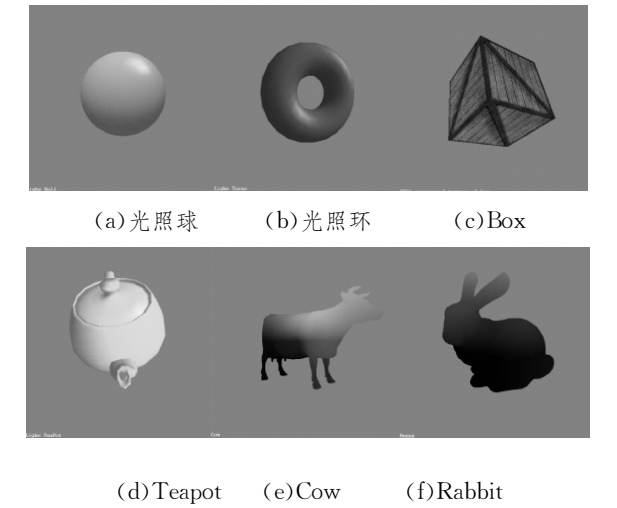


图 5 应用程序运行结果图

采用 Synopsys 公司 Design Compiler 综合工具在 0. 13 μm CMOS 逻辑工艺下进行综合,主频达到 150 MHz,功耗约 177. 7428 mW,比文献[15]所设计的染色处理器主频高 50%,功耗低约 8. 85% .

6 结束语

本文设计了一种基于 VLIW 架构的嵌入式可编程顶点染色处理器,定义了指令集,设计了超长指令字格式与顶点染色处理器的整体架构. 在与高性能的 ARM 处理器软硬协同工作基础上,本文设计的顶点染色处理器更适合进行向量操作. 可编程的

设计提高了顶点染色处理器的灵活性. 能够实现 OpenGL ES 2.0 应用程序中的各种染色程序.

参考文献:

- [1] 韩俊刚, 蒋林, 杜慧敏, 等. 一种图形加速器和着色器的体系结构[J]. 计算机辅助设计与图形学学报, 2010, 22(3): 363-372.
- [2] 申智波, 陈杰, 刘建, 等. 适用于嵌入式应用的多线程顶点处理器设计[J]. 微电子学与计算机, 2012, 29(7): 76-80.
- [3] Kim K S, Lee Y S, Choi S B. Multi-port register file design and implementation for the simd programmable shader[C]// IT Convergence and Security 2012. Springer Netherlands, 2013: 267-272.
- [4] CH Yu, Chung K, Kim D, et al. An energy-efficient mobile vertex processor with multithread expanded vliw architecture and vertex caches[J]. IEEE Journal of Solid State Circuits, 2007(42): 2257-2269.
- [5] Chang Y, Wei J, Guo W, et al. Full length article: a high performance, area efficient tta-like vertex shader architecture with optimized floating point arithmetic unit for embedded graphics applications[J]. Microprocessors & Microsystems, 2013, 37(6/7): 725-738.
- [6] CH Yu, Chung K, Kim D, et al. A 186Mvertices/s 161mW floating-point vertex processor for mobile graphics systems[C]// 2007 IEEE Custom Integrated Circuits Conference (CICC 07). San Jose, 2007: 579-582.
- [7] 李涛, 肖灵芝. 面向图形和图像处理的轻核阵列机结构[J]. 西安邮电学院学报, 2012, 17(3): 41-47.
- [8] Lee Y, Avizienis R, Bishara A, et al. Exploring the tradeoffs between programmability and efficiency in data-parallel accelerators[J]. Acm Sigarch Computer Architecture News, 2011, 39(3): 129-140.
- [9] Edward Angel. 交互式计算机图形学-基于 OpenGL 的自顶向下方法[M]. 四版. 北京: 清华大学出版社, 2007.
- [10] Dan Ginsburg, Budirijanto Purnomo. OpenGL ES 3.0 编程指南[M]. 北京: 机械工业出版社, 2015.
- [11] 杜慧敏, 李萌. 面向顶点染色算法的专用指令集优化处理器[J]. 西安邮电大学学报, 2014, 19(2): 60-66.
- [12] Khronos Group. GLSL ES Specification 1.0. 17. Unpublished. [EB/OL]. [2012-10-15]. (2018-01-04). <http://www.khronos.org/>.
- [13] Joho L Hennessy, David A Patterson, 亨尼西, 等. 计算机体系结构: 量化研究方法[M]. 北京: 机械工业出版社, 2012.
- [14] 胡伟武. 计算机体系结构[M]. 北京: 清华大学出版社, 2011.
- [15] Woo J H, Sohn J H, Kim H, et al. A 195 mW, 9.1 MVertices/s fully programmable 3 D graphics processor for low power mobile devices[J]. IEEE Journal of Solid-State Circuits, 2007, 43(11): 2370-2380.

作者简介:

徐起超 男, (1990-), 硕士. 研究方向为集成电路设计. E-mail: xuqichao@126.com.

杜慧敏 女, (1966-), 博士, 硕士生导师, 教授. 研究方向为集成电路设计、计算机体系结构.

刘青楠 男, (1992-), 硕士研究生. 研究方向为电路与系统.