

## PCI 总线多用户数据缓冲区管理器的实现

乔庐峰<sup>\*\*\*</sup> 王志功<sup>\*</sup> 黄斌<sup>\*</sup> 陆园琳<sup>\*</sup>

<sup>\*</sup>(东南大学射频与光电集成电路研究所 南京 210096)

<sup>\*\*</sup>(南京通信工程学院 南京 210007)

**摘要:** 分析了一种 PCI 总线上支持多个用户的数据缓冲区管理器电路所采用的电路结构, 给出了关键点的仿真波形。从理论上分析了用户的缓冲区分配原则、端口总线带宽、用户最大服务等待时间和最小缓冲区数量等工作参数。给出了多用户缓冲区管理器中所需要最小存储区的计算方法。以分析为基础, 采用 XILINX 的 XCV600EPQ240 实现了 128 用户缓冲区管理器电路, 并在实际系统中进行了测试和验证。

**关键词:** VLSI, PCI总线, 先入先出存储器, 缓冲区管理, 现场可编程门阵列

中图分类号: TN47

文献标识码: A

文章编号: 1009-5896(2005)07-1162-05

## Implementation of PCI Bus Multi-user Data Buffer Manager

Qiao Lu-feng<sup>\*\*\*</sup> Wang Zhi-gong<sup>\*</sup> Huang Bin<sup>\*</sup> Lu Yuan-lin<sup>\*</sup>

<sup>\*</sup>(Institute of RF- & OE-ICs, Southeast University, Nanjing 210096, China)

<sup>\*\*</sup>(Institute of Communication Engineering, Nanjing 210007, China)

**Abstract** The circuit structure of a kind of PCI bus multi-user data Buffer Manager (BM) is analyzed in this paper, and typical simulating waveform is presented. The method to allocate the data buffers, port bandwidth, maximum user waiting time and minimum user buffer requirements are analyzed theoretically. The expression to calculate the minimum memory needed in the BM is given. Based on the analysis, a 128-user buffer manager is realized with XILINX XCV600EPQ240 and verified in application systems.

**Key words** VLSI, PCI bus, FIFO memory, Buffer manager, FPGA

### 1 引言

具有 PCI 接口的多用户 HDLC 通信控制器芯片在网络接入、工业控制、访问服务器等领域有着广泛的应用, 其基本电路结构和在计算机系统中的应用如图 1 所示。在接收方向上, 接收多用户 HDLC 处理器负责对多个 E1 链路上多个用户的低速数据进行与 HDLC 规范相关的处理工作<sup>[1]</sup>, 并将数据交给接收缓冲区管理器(Buffer Manager, BM)。接收 BM 将多个用户分散到达的数据进行汇集、排队, 形成多个用户数据队列, 并在数据累积到设定的门限时, 向直接存储器访问控制器(Direct Memory Access Controller, DMAC)发出操作请求, 由 DMAC 负责将数据读出并通过 PCI 接口电路以高速、突发的方式写入到主机内存中。在发送方向上, DMAC 以高速、突发的方式将不同用户的待发数据从内存中读出并写入到对应的发送缓冲区中, 发送多用户 HDLC 处理器根据链路上用户的数据速率, 将数据低速读出并进行与 HDLC 规范相关的发送帧处理后送到链路上。

在整个设计中, BM 的基本功能是对多个用户的数据进

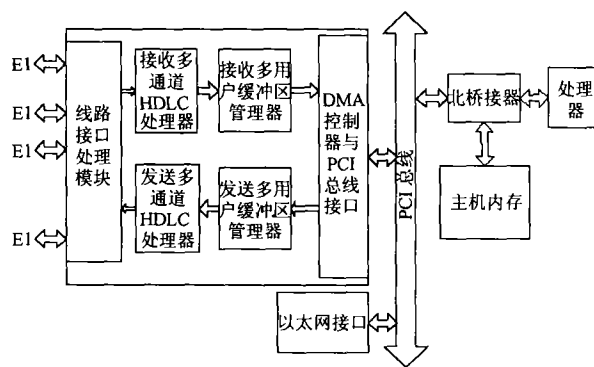


图 1 多用户 HDLC 数据帧处理器  
电路基本结构与系统应用示意图

行本地缓冲和规划, 在以高速突发为数据传输特征的 PCI 总线和具有低速连续为特征的物理链路之间进行操作缓冲。当用户数达到数十和上百时, 它应该具有端口操作时序简单、用户所占缓冲区深度可进行单独配置、端口和电路内部数据处理效率高等特点。

本文借鉴文献[2-4]中对BM基本工作机制的说明,根据BM的功能设计要求,从芯片设计角度给出了多用户条件下BM统一的内部电路结构,分析了电路的工作机制,并在实验基础上给出了内部关键节点的仿真工作波形。从电路的实际工作特点出发,对缓冲区在多个用户之间的分配原则、端口工作带宽进行了分析,给出了数据缓冲区最小需求量的理论分析公式,以期在满足设计要求的情况下尽量减小BM中存储器的容量,从而减小BM的芯片面积。整个设计在我们采用FPGA建立的实验环境中进行了验证。

## 2 缓冲区管理器的电路结构与工作流程

借鉴传统的 FIFO 所采用的电路结构<sup>[5]</sup>和文献[2,3]所采用的链表管理方式,我们将这类电路统一为图 2 所示的结构。

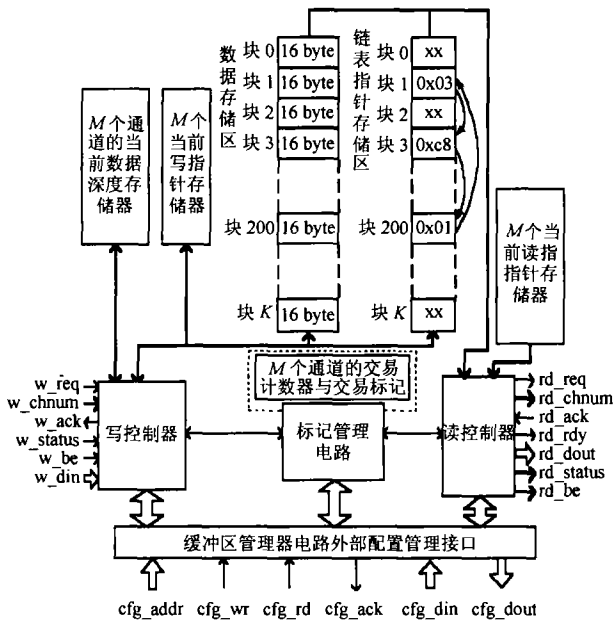


图2 缓冲区管理器的内部电路结构

$M$ 个用户的数据缓冲区被划分为 $N$ 个16byte的数据块。每一个数据块在链表指针存储区中都对应着一个指针,该指针指向该块所链接的下一个数据块。分配给每个用户的数据缓冲区通过指针构成循环结构。在图2中,某个用户所占用的数据缓冲区的循环写入关系为:块1→块3→块200→块1,通过外部配置管理接口修改指针存储区的链接关系就可以方便地改变某个用户占用的缓冲区总量。

写控制器负责数据的写入操作。收到用户发出数据写入请求后,写控制器首先会根据其用户号从当前写指针存储器中读出该用户的当前写指针,同时从链表指针存储区中读出指向下一个数据块的链接指针。当写入的数据将当前数据缓冲区填满后,继续进入的数据将被写入到链接指针所指向的空间内,直到操作完成。数据写入过程中,写控制器不断监视当前数据缓冲区深度和写入数据的状态,如果发现数据缓

冲区深度达到了为该用户所设定的深度门限或数据状态满足触发条件,写控制器就会向标记管理电路发出通知,请求DMAC读出该用户的数据。

读控制器负责数据的读出。它从标记管理电路得到操作要求和请求用户号后,首先会向DMAC发出操作请求,得到确认后,它根据当前读出指针存储器和链表指针存储器所指向的数据缓冲区,以突发方式连续的将数据读出到外部数据总线上。当本次读操作完成后,读控制器向标记管理电路发送操作完成确认信号。

标记管理电路的任务是对读、写控制器的操作进行协调。它管理着 $M$ 个用户的操作申请标记和每个用户发出申请的次数。它按照平等或有优先级两种规则对所有用户的请求标记进行查询,如果发现某一个用户的通道标记为1,那么它就会向读控制器送出操作请求。由于交易标记和申请次数与读、写控制器的操作都有关系,可能引起操作冲突,此时需要标记管理电路加以协调。

## 3 BM的设计实现与关键工作波形分析

上面的分析为基础,我们采用现场可编程门阵列设计了 $M=128$ 时的缓冲区管理器电路,下面以接收方向上BM的工作仿真波形为例对电路的核心功能加以分析。

### 3.1 BM配置功能的后仿真分析

BM正确工作的第一步是为每一个用户建立多块链接的数据缓冲区。图3是对用户0进行配置的后仿真波形,它包括配置端口工作时序和电路内部节点的工作时序(操作①-④)。在这里,操作①表示在配置空间地址0中写入用户0的当前写指针“0”;②是将其当前读指针值设为“0”(配置地址为0x080),它与写指针相同,说明该用户缓冲区内没有数据;③是配置用户0数据缓冲区链接关系的操作,配置地址0x200-0x203中是各数据块所链接的下一个数据缓冲区。可以看出,0x203中的内容为0x00,又指向了第一块缓冲区,从而形成了循环链接的数据缓冲空间。在④中,我们将用户0的数据读出门限设定为2,写入到地址0x180中。

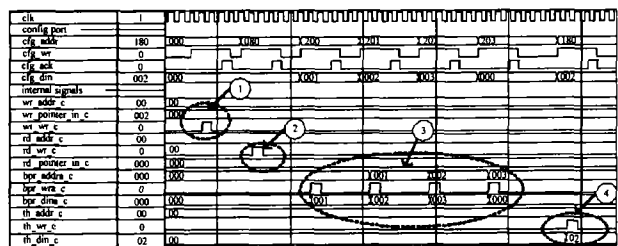


图3 系统配置操作过程

### 3.2 用户数据写入与内部节点的后仿真分析

在配置完成后,将8个32位数据分别写入到用户0中,如

图4所示。①是BM写入操作的端口波形，其操作时序规则非常简单。②是写入一个数据后，该用户的当前写指针更新为1。操作③是每4次写操作后，通过链接指针得到下一块数据缓冲区首地址的操作过程。4次写操作后，用户0的数据深度升为1，表示有一个数据块被填充满，又写入4次后，深度值应为2，由于此时达到了申请门限，所以深度值记录为0，同时通过wr\_roamer向中央控制器发出一次读操作请求(见④)，由标记管理电路将该用户的申请标记反转为1，并将申请次数置为1。

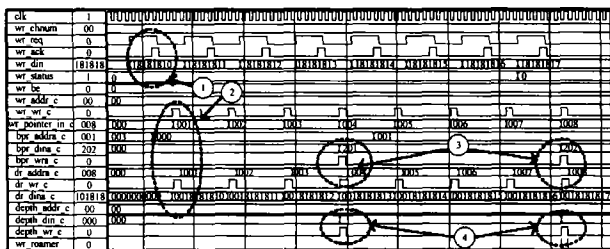


图4 用户0数据写入操作后仿真波形

### 3.3 缓冲区用户操作申请与数据连续读出的后仿真分析

图5是缓冲区数据读出处理流程。标记管理电路轮流查询每一个用户的申请标记，当发现用户0的申请标记为1时，就会将该用户的用户号和申请门限提交给读控制器(图5中操作①)，由读控制器通过总线向外部电路发出操作请求(rd\_req)。当外部电路给出确认信号后(rd\_ack)，读控制器就可以按照先入先出的原则将数据连续从内部缓冲区中读出(图5中操作④)。在连续读取数据的过程中，我们可以看到当前数据读地址和读指针不断发生变化(图5中操作②和③)。

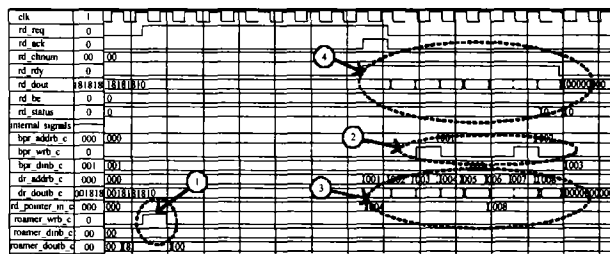


图5 用户0数据连续读出操作后仿真波形

上面的分析都是以接收方向为例加以分析的，在发送方向上，DMAC将数据连续突发地写入到BM中，用户将数据以低速读出，内部电路的工作机制完全相同。

## 4 缓冲区管理器的工作参数与性能分析

### 4.1 缓冲区分配的要求

由于BM所管理的缓冲区总数量是固定的，这就存在如何在多个用户之间进行缓冲区分配的问题。这里使用的一个基本参数是某个用户所占缓冲区块数( $B_{number}$ )与数据到达速度( $V_{channel}$ )的比值，即： $R = B_{number} / V_{channel}$ 。如果某一个用

户的R值与其它用户的R值相比过小，说明为其分配的缓冲区深度需要加大。

### 4.2 输入接口的数据吞吐量

在用户侧端口的数据操作上，我们改进了目前常用的以块为单位进行读写的方式，用户可以按照以块为单位和以32bit双字为单位两种方式进行操作。这增加了BM用户端口的设计复杂程度，但给用户电路的设计带来的方便，可以减小用户电路的本地存储空间。

根据图4的仿真结果，一次数据操作申请如果只进行一次写操作，需要占用6个时钟周期，总线带宽利用率只有16.7%，在33MHz，32bit宽度的PCI总线条件下，输入数据带宽约为167Mbit/s。如果一次数据操作处理包括4个数据周期，那么总线利用率就可以达到44%，输入数据带宽约为444Mbit/s，但此时用户电路需要更大的本地数据缓冲能力。

### 4.3 最大用户等待时间与最小缓冲区量的分析

BM中，可能会由于PCI总线上其它设备和本设备中其它用户的存在而造成服务延迟过大，出现缓冲区“上溢”或“下溢”错误。为了减少这类错误的发生，需要分析从一个用户发出操作请求到数据传输完成所需要的时间，它由下面6个部分组成，其中参数的单位为时钟周期数：

(1) 用户操作请求记录时间  $T_{rec}$ ：标记管理电路收到一个用户的操作请求时，如果其处于标记检索状态，那么可以在一个时钟周期内完成该用户的标记位置；如果此时读控制器需要对某个标记和交易计数器进行修改，则存在操作冲突，此时需要更多的时间。

(2) 用户标记检索时间  $T_{find}$ ：它是标记管理电路对所有用户标记进行检索并发现用户请求所需要的时间。从硬件实现角度来说，它是一个条件判断电路，这类电路的特点是随着用户数的增加，对所有用户标记查询一次需要的时间会同时增大，在33MHz PCI总线时钟的条件下，我们取每进行32个用户标记的判断需要一个时钟周期，通过延长判断时间来降低电路设计的复杂程度。

(3) 标记管理电路根据检索结果，读出该用户的状态信息，并向读控制器发出操作请求、请求用户号和操作块数所需要的时间  $T_{load}$ 。

(4) 读控制器根据标记管理电路的操作申请和用户号，读出其当前读指针和链接指针并通过读出端口向DMAC发出操作请求需要的时间  $T_{req}$ 。

(5) DMAC根据BM的操作请求和该用户的状态信息，在BM与主机内存之间传输X块数据所需要的时间为： $T_{op}$ 。如果每次读出的块数为X，平均收发一个数据块DMAC需要的平均额外时钟周期开销为  $T_{OH}$ ，那么有平均收发X块数据时： $T_{op} = 8X + T_{OH}X$ 。

(6) 操作完成后, 读控制器通知标记管理电路进行该用户信息的更新:  $T_{ref}$ 。

以上分析的各项, 除PCI总线操作的额外开销, 其它与BM相关的各项, 在收发方向上完全相同。由此可以得到整个芯片为一个用户平均收发 $X$ 块数据时所需要的时钟周期数为

$$T_X = 2 \cdot (T_{rec} + T_{find} + T_{load} + T_{req} + T_{ref}) + (8 + T_{OH}) \cdot X \quad (1)$$

如果取  $T_{moh} = T_{rec} + T_{find} + T_{load} + T_{req} + T_{ref}$ , 则有

$$T_X = 2 \cdot T_{moh} + (8 + T_{OH}) \cdot X \quad (2)$$

一个用户的数据在接收或发送方向上达到缓冲区深度门限时会发出数据传输请求, 如果此时其它所有用户和该用户的另外一个传输方向都已经发出了操作请求, 那么此操作需要等待的时间最长, 称为最长用户等待时间, 用  $W_{max}$  表示。如果每个用户一次操作的数据量均为 $X$ 块数据, 那么  $W_{max}$  表示为

$$\begin{aligned} W_{max} &\approx \left( \frac{1}{2} \cdot T_X + \sum_{i=1}^{M-1} T_X \right) \cdot T_{PCI} \\ &= \left( M - \frac{1}{2} \right) \cdot [2 \cdot T_{moh} + (8 + T_{OH}) \cdot X] \cdot T_{PCI} \\ &= (2M - 1) \cdot T_{moh} \cdot T_{PCI} + \left( M - \frac{1}{2} \right) \cdot (8 + T_{OH}) \cdot T_{PCI} \cdot X \quad (3) \end{aligned}$$

其中的  $T_{PCI}$  是PCI总线的时钟周期。文献[6]中分析了采用目前常用的链表结构时, DMAC平均收发一个字节所需要的时钟开销数, 由此可以方便地得到  $T_{OH}$ 。需要说明的是  $T_{OH}$  只有在进行最长等待时间分析时, 取数据包长度等于 $X$ 的条件下, 才是一个常量, 否则它与 $X$ 是相关的。对于一个链路速率为  $V_i$  的用户, 如果不发生缓冲区错误, 那么要求  $W_{max}$  时间内可用于继续填充数据的缓冲区量  $B_L$  应满足下面条件:

$$B_L \geq \left[ \frac{V_i \cdot W_{max}}{128} \right] \quad (4)$$

$B_L$  通常与一次突发操作的数据量 $X$ 相同, 此时有

$$\begin{aligned} B_L \geq \left[ \frac{V_i \cdot W_{max}}{128} \right] &\Rightarrow X \geq \left[ \frac{V_i \cdot W_{max}}{128} \right] \\ &\Rightarrow X \geq \left[ \frac{(2M - 1) \cdot V_i \cdot T_{moh} \cdot T_{PCI}}{128 - \left( M - \frac{1}{2} \right) \cdot (8 + T_{OH}) \cdot T_{PCI} \cdot V_i} \right] \quad (5) \end{aligned}$$

由此我们可以得到一个用户所需要分配的最小缓冲区数量  $B_{min}$  为

$$\begin{aligned} B_{min} = B_L + X &\Rightarrow B_{min} = 2 \cdot X \Rightarrow B_{min} \\ &= 2 \cdot \left[ \frac{(2M - 1) \cdot V_i \cdot T_{moh} \cdot T_{PCI}}{128 - \left( M - \frac{1}{2} \right) \cdot (8 + T_{OH}) \cdot T_{PCI} \cdot V_i} \right] \quad (6) \end{aligned}$$

根据式(5)和式(6), 可以方便地根据电路的工作特点得到电路中每个用户所需要的最少缓冲区块数, 从而指导不同用户数量和工作条件下BM中数据缓冲区数量的选定。图6是根据式(5)所得到  $M$  分别取64, 128和256时,  $X$  与  $T_{OH}$  的关系。根据图3-图5中BM的工作波形, 可以得到式(5)中各部分开销所占用的时钟周期数分别为:  $T_{rec} = 2, T_{find} = 4, T_{load} = 2, T_{req} = 2, T_{ref} = 2$ , 另外取  $T_{PCI} = 30ns, V_i = 64kbit/s$ 。

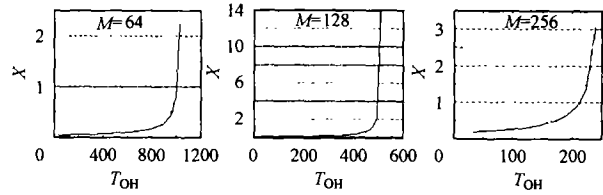


图6  $M$ 取不同值时  $T_{OH}$  与  $X$  的关系

根据式(5)和式(6), 可以看出  $X$  或  $B_{min}$  在  $M$  较大时与用户数量、用户数据速率、BM数据处理时间开销以及计算机系统所支持PCI规范的  $T_{PCI}$  都呈线性关系。减小  $T_{moh}$  可以直接减小BM中的缓冲区需要量, 因此要求BM设计时应注意提高电路的处理效率, 将  $T_{moh}$  控制在一个尽量小的范围内。从图6可以看出, 在不同用户数量的条件下,  $T_{OH}$  较小时,  $X$  值增加的十分缓慢, 当  $T_{OH}$  大于某个门限后, 随着  $T_{OH}$  的增大,  $X$  迅速增加, 并且对  $T_{OH}$  的变化十分敏感, 同时这一门限随着用户数量的增加以近似线性的关系减少。如果系统设计过程中, 能够将  $T_{OH}$  控制在此门限之下, 那么对减小BM中缓冲区的需求量, 提高整个系统性能都有直接影响。  $T_{OH}$  与DMAC所采用的电路结构、数据结构和计算机软硬件系统结构都有直接关系, 文献[6]对其计算方法进行了详细的分析。在得到  $X$  后, 就可以得到  $B_{min}$  的值, 从而确定BM所需要的本地缓冲区的最小数量。

## 5 BM的实现与系统功能验证

在以上分析的基础上, 我们设计了支持4个E1链路, 最多128个用户的BM电路, 用于图1所示的应用中。当PCI总线上最多有4个相同设备时  $T_{OH}$  值约为400, 考虑到一定的裕量, 最终每个用户取  $X=2, B_{min}=4$ 。此时, BM在收发方向上各需要的缓冲区总量为512个数据块。这与文献[2,3]所采用的缓冲区量是吻合的。

为了对整个设计进行验证, 我们采用Xilinx公司的FPGA xc2s200PQ208实现的具有PCI总线接口的DMAC和使用xcv600eHQ240实现128用户HDLC处理器和缓冲区管理器。同时, 采用Conexant的CN8474和PMC-Sierra的PM7366设计的对比测试板。实验板卡和对比测试板卡被插入到两台运行嵌入式实时操作系统(Vxworks)的目标PC机中, 两个板卡之间采用4条E1链路进行连接, 通过编写各自的高层应用程序,

两台目标机中128个用户不同长度的数据包可以按照所需要的配置进行互通,对比分析表明BM的设计达到了系统性能要求。

本设计所采用的缓冲区深度与文献[2,3]所采用的缓冲区深度是一致的,说明了分析的正确性。同时,通过实验和仿真可以看出,在用户数量增加时,只需要按比例增加BM中的缓冲区块数,而整个系统的优化重点应该放到提高PCI桥接器的数据传送效率上。

## 6 结束语

我们所设计的BM是“基于PCI总线的128用户HDLC数据帧处理器”电路的一个重要组成部分,整个电路已经在实际应用系统中完成了测试。在保持现有电路结构的情况下,改变缓冲区管理器中的存储器深度,可以方便地扩充用户数量。由于该电路具有良好的外部操作端口并且可以进行灵活的配置,所以具有广泛的应用价值。

## 参考文献

- [1] 陆园琳, 乔庐峰, 王志功. 多通道高速HDLC处理器的设计与实现[J]. 电子学报, 2003, 31(11): 1630 - 1633.
  - [2] Dallas Semiconductor, Inc., DS3131 BoSS 40-Port, Unchanneled Bit-Synchronous HDLC Controller [EB/OL]; <http://pdfserv.maxim-ic.com/ds/en/DS3131.2002-11-20/2003-2-5>.
  - [3] PMC-Sierra, Inc. FREEDM-32 Data Sheet [EB/OL]; <http://www.pmc-sierra.com/products/details/pm7364/>. 2003-6-6/2003-7-30.
  - [4] Varada S, Oduol V, Shelat A. Data flow and buffer management in multi-channel data link controller[A]. In: 24th Conference on Local Computer Networks[C]. Lowell, Massachusetts, 1999: 132 - 141.
  - [5] Integrated Device Technology, Inc. IDT 72230 Data Sheet [EB/OL]; <http://www.idt.com/products/pages/FIFOs-72230.html>. 2002-9-12/2003-11-20.
  - [6] PMC-Sierra, Inc. FREEDM PCI Bus Utilization and Latency Analysis [EB/OL]; <http://www.pmc-sierra.com/products/details/pm7364/>. 2003-6-6/2003-7-30.
- 乔庐峰: 男, 1971年生, 博士生, 从事超大规模集成电路的设计, 特别是计算机网络的接入芯片设计。
- 王志功: 男, 1954年生, 博士, 教授, 博士生导师, 从事用于光纤通信、卫星通信和雷达领域的超高速、微波毫米波集成电路、光电集成电路的设计; 现为国家863计划专家组专家, 东南大学射频与光电集成电路研究所所长。
- 黄斌: 男, 1973年生, 硕士, 主要从事超大规模集成电路设计工作。
- 陆园琳: 女, 1977年生, 硕士, 主要从事超大规模集成电路设计工作。