

# 典型 ATM 交换结构分析与交换机设计分类<sup>\*</sup>

Analysis of Typical ATM Switching Fabrics and Categories of Switch Design

陈永光

(解放军电子工程学院 合肥 230037)

**Abstract** This paper introduces ATM switching functions and several indexes for evaluating the performance of ATM switch. Some typical switching fabrics and their influence on switch performance are also analyzed. Furthermore, the paper discusses categories of switch design in detail and proposes a scheme to realize small type of ATM switch by using MB866 family chips.

**Keywords** ATM, Switching fabric, Shared medium, Self routing

目前的通信网正逐步向宽带化、综合化和智能化方向发展,为了传送高清晰度电视(HDTV)图像及高速数据,要求通信网高速宽带化。高速宽带通信与现有的电话及数据通信相比,应具有极宽的通信频带(几十Mbps以上)。另一方面,随着科技的发展,人们对通信的需求不断增加,导致未来通信业务的不确定性。因此,人们致力于研究新的通信方式以适应上述情况,其中,ATM可能是实现宽带ISDN的最佳方式,可以满足B-ISDN的要求,因而受到各发达国家重视。ATM的核心是交换技术,从商用ATM交换机产品的设计中可以发现,目前已经形成了若干种典型ATM交换结构,因此,对交换结构的分析和对交换机设计方法的归纳具有实际意义。

## 1 ATM 交换功能和交换机性能指标

ATM交换机需完成以下动作,即三个功能:(1)空分交换。信息从编号为N的入口通过交换通往编号为M的出口,这是一种路由选择功能。(2)时隙交换。时隙(信道) $i$ 中信息被交换到信道(信道) $j$ ,这里牵涉到信元头的转译。(3)排队。两个来自不同链路的信元可能同时到达ATM交换机,并去向同一链路,这时交换机中缓冲存储器就用以暂放那些暂时不能送出的ATM信元,即有些ATM信元在排队等待送出,这样就有可能对信元统计复用,这是ATM交换机的典型功能。

ATM交换机完成这些功能的质量如何,其性能的评价主要由三个指标来衡量:

·连接阻塞率:ATM是面向连接的,连接阻塞率就是在出入口之间没有足够的资源以保证现有连接和新建连接质量的概率。这是由交换机配置,诸如交换机内部连接数以及这些连接上的负载决定的。

·信元丢失/信元插入概率:在ATM交换机中,当有很多信元的去向是同一链路,而缓冲存储器容量又不足以同时存入所有要排队的信元时,就会发生信元丢失。信元丢失概率应有限度以保证语义之透明性:ATM交换机信元丢失概率典型范围是 $10^{-8} \sim 10^{-12}$ 。另一种可能是ATM交换机内部选择路由有差错,导致某些信元被错误地送往另一逻辑连接。信元插入概率典型值应小于信元丢失率的1000倍或更多些。

·交换时延:ATM交换机的交换时延范围典型值为 $100 \sim 1000 \mu s$ 之间,其抖动为几百 $\mu s$ 或更小一些。该指标通常以时延不超过某一值的概率来表示,如时延大于 $10 \mu s$ 的概率为 $< 10^{-10}$ 。这些性能指标受到ATM交换机排队方式的直接影响,根据缓冲存储器所处位置,排队方式有三种,即输入排队、输出排队和中央排队,有三种参数影响到排队系统的复杂性,即队列长短、存储器速度和存储器控制。

输入排队的有用负载最多只有58.6%<sup>[1]</sup>,即58.6%的来向信元可含有有用数据,其余皆为空的,因而性能较差。输出排队的有用负载为80%左右<sup>[2]</sup>。中央排队和它差不多,但可用较小规模缓冲存储器,例如当信元丢失率为 $10^{-8}$ 时,输出排队系统队列规模为40个信元,而中央排队系统中每条输出链路队列可短于10个信元,只是中央排队系统需要较快的速度和存取时

<sup>\*</sup>安徽省自然科学基金资助项目98312518。陈永光 博士(后),教授,研究方向为ATM交换技术、电子站仿真、电子对抗作战效能分析等。

间、较短的缓冲器和较复杂的控制逻辑。

## 2 典型 ATM 交换结构

### 2.1 段式交换结构

把几个小的交换单元进行多级连接,便形成实用型大规模交换机结构,这种方法在以往的电路交换中采用过。段式交换结构就是采用这种方式,将 $2 \times 2$ 个交换单元以 $\log_2 N$ 级连接,构成 $N \times N$ 的交换结构。

图1是 $8 \times 8$ 三级段式交换结构;各交换单元依次自动读出相应的信元选路比特,如图1(a)所示。当选路比特为“0”时,在交换单元上方输出;为“1”时则在下方输出;经多级连接后成图1(b),当有“010”选路信息段信元从输入端口“4”输入时,该信元沿图中粗线所示路径到达输出端口“2”。

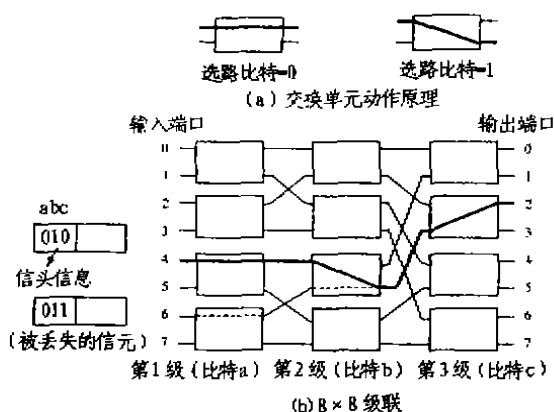


图1 段式交换结构

这种交换结构具有信元选路功能,在输入和输出端口之间有多条路由,多条ATM信元流可以同时在不同路由上通过,LSI硬设备容易实现。但是会出现交换结构内部的链路竞争,内部阻塞不可避免。例如图1(b)中,当有“011”选路信息段的信元从输入端口“6”到输出端口“3”时,即使输出端口“3”空闲,也会因第2~3级间的链路被占而无法通过,造成阻塞。为了减少内部阻塞,可以再增加链路,但需要合适的选择路径的算法。比较适用的是采取在各交换单元的输入链路上专门设计缓冲器(或暂存器)的方法,被阻塞的信元在此暂时等待,以免丢失。另外还有减少内部阻塞、提高吞吐量的办法,如在输入端口之前一级增加分散业务量功能,或者追加按输出端口地址的分类存储功能。当然,即使在交换机内部不产生信元阻塞,从输入、输出端口信元竞争的角度出发,也需要有信元存储功能。

### 2.2 完全连接型交换结构

ATM交换结构的另一个例子如图2所示,这是完全连接型交换结构。这种交换结构在输入输出矩阵的输出侧设信元缓冲器;当信元在输出端口发生竞争时,可在缓冲器内暂存等待;因为信元在输出端口竞争时进行存储调整,故又称输出缓冲型交换结构,在 $N \times N$ 交换矩阵中,相对于一个输出端可转送来自 $N$ 个输入端口的信元,但是,一旦缓冲器中信元装满,则从缓冲器溢出的信元会丢失。

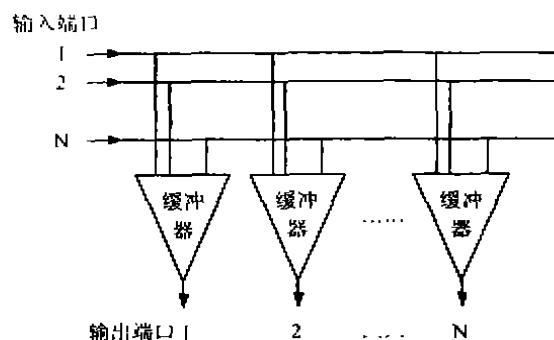


图2 完全连接型交换结构

## 3 ATM 交换机设计分类

交换机设计分两大类:时分和空分。时分容量小、速率低;空分容量大、速率高、可并行处理。

时分结构是指所有的输入输出端口共享一个高速的信元流通路,这个共享高速通路可以是共享介质(总线、环),也可以是共享存储器。该结构的交换容量受共享通路吞吐量的限制,如总线的速度限制、存储器容量和存储速度的限制等,但该结构易于实现点到多点的操作;时分结构沿袭了计算机结构的设计思路。

空分结构是指在输入和输出端口之间有多条通路,多条ATM信元流可以同时在不同通路上通过。这种结构的交换容量由每条通路的带宽和同时传送信元的通路平均数决定,从理论上讲是无限的,但实际上受芯片等一系列实际问题的限制,空分结构多沿袭于电路交换的设计思想。

### 3.1 共享介质交换机

这类交换机通过共享一个高速介质(如总线或附加存储器系统的总线)而运作<sup>[1]</sup>,共享介质也可以是环(如令牌环或FDDI)或带介质存取控制的串行总线(如以太网)。这类交换机的带宽受限于总线经过的物理距离、负荷类型和分布以及总线位数等电气驱动类因素,现在的共享介质系统每个总线字节宽度支持的典型交换带宽为200Mb/s;共享介质交换机的构造如

图3所示。

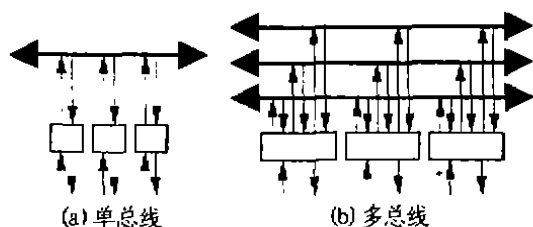


图3 共享介质交换机

共享存储器交换机属于共享介质交换机的范畴，它在总线上只设一个存储器子系统，而并非每个口都有一个存储器系统。这样做是为了最有效地利用存储器，因为按此配置，存储器可以在所有输入和输出上共享。如果存储器的输入、输出数据线连到同一总线上，总线作为交换机的有效带宽则会减半。

共享介质交换机被认为是性能-价格比最大的交换机类型，在用现有技术实现适度带宽时尤其是这样。但由于其带宽被限制在10Gb/s左右，故使用范围也有限，大多数用共享介质构造的交换机具有800M~3.2G的带宽。

多播和广播相对而言容易适应这类交换机，因为所有输出都可以获取放在共享介质上的数据。许多路由器可被看成是共享存储器交换机，许多多端口桥也属于共享介质交换机。

### 3.2 多级自主路由阵列

这类交换机已研究了多年，Batcher-Banyan 是其中研究最多的类型。研究动力来自于对1000~10000线的中央办公室级交换机的需求。由于芯片脚、PC板尺寸以及连接系统的限制，交换机的规模被限制在128/128之下<sup>[4]</sup>，商业上对其兴趣已稍减弱，因为市场踏上了用于LAN方面的交换。

自主路由阵列交换机在处理拥塞时比其他类型交换机的困难大，部分原因是交换机的结构中存在大量管线和缓冲。竞争只有当信元深入到了交换机构造内部某个分选阵列的输出时才能被检测出来，此时，该信元必须返回到阵列的输入或保留在存储器里。另外，如果没有专门的硬件来复制信元，阵列交换机也难以实现多播和广播。

当阵列交换机可以达到所期望的大端口数时，能够提供的潜在带宽为2~400Gb/s；大多数样机有16~64个端口。

至此可以说：应该通过小规模交换机用不断增速的链路来满足商业数据的要求，而不是用单个的有成

千上万端口的交换机来实现。自主路由阵列交换机的构造如图4所示。

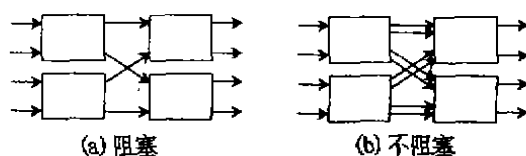


图4 自主路由阵列

### 3.3 纵横式交换机

用几个VLSI芯片即可构成大型全交叉(纵横式)交换机。如一种128/128口的实验型芯片有447个脚，六片这种芯片即可组成支持128个Sonet 155Mb/s链路的纵横式交换机，带宽达20Gb/s<sup>[5]</sup>。

制造纵横式芯片的经验表明：纵横式交换机的每个阵列连同相应的交换状态存储单元已变得如此之小，使得焊点尺寸和可封装脚数成了限制大型交换芯片问世的因素。现在，人们不再单纯追求交叉点的数量了，因为封装技术尚不允许脚数以同样的速度增加。

很难实现纵横式交换机的快速控制机制，许多采用了纵横式交换芯片的交换机以相对低的速率交换。为了提高连接速度，可并行控制纵横式交换机的每一输出口；用并行重复配对和控制的方法能使连接速度达到3080万连接/秒。应用多个芯片的交换机实际上形成了有并行数据通路的三维配置结构(图5(b))。例如用4块芯片可以构成一个16×16口、32位宽的具有800Mb/s端口的纵横式交换机。用纵横式交换芯片可以很容易地构成带宽高达80Gb/s的交换机<sup>[6]</sup>。

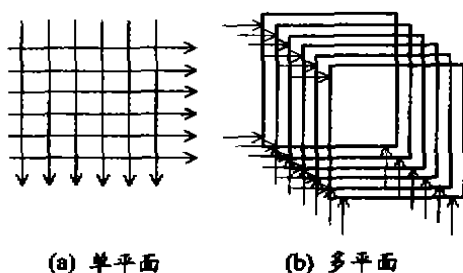


图5 纵横式交换机

## 4 一种小型ATM交换机设计方案

图6是用富士通公司MB866系列芯片设计的  
(下转第79页)

- 12 Sistla A P, et al. Modeling and Querying Moving Objects. In Proc. IEEE Intl. Conf. on Data Engineering, Birmingham, U.K., 1997. 422~432
- 13 Kampke T. Storing and Retrieving Changes in a Sequence of Polygons. Int. Journal of Geographical Information Systems, 1994, 8(6):493~513
- 14 Price S. Modeling the Temporal Element in Land Information Systems. Int. Journal of Geographical Information Systems, 1989, 3(3):233~243
- 15 Raafat H, et al. Relational Spatial Topologies for Historical Geographical Information. Int. Journal of Geographical Information Systems, 1994, 8(2):163~173
- 16 Vazirgiannis M, et al. Spatio-Temporal Composition and Indexing for Large Multimedia Applications, 1997
- 17 Nabil M, et al. Modeling Moving Objects in Multimedia Database. Proc. of the 5th on Database Systems for Advanced Applications, Melbourne, Australia, 1997
- 18 Albers G, Roos T. Voronoi Diagrams of Moving Points in Higher Dimensional Spaces, 3rd Scandinavian Workshop on Algorithm Theory, 1992. 399~409
- 19 Fu J J, Lee R C T. Voronoi Diagrams of Moving Points in the Plane. Int. Journal of Computational Geometry and Applications, 1991, 1(1):23~32
- 20 Guibas L, et al. Voronoi Diagrams of Moving Points in the Plane, 17th Workshop on Graph-Theoretic Concepts in Computer Science, LNCS 570, 1991. 113~125
- 21 Guting R H, Schneider M. Realn-Based Spatial Data Types: The ROSE Algebra. VLDB Journal, 1995, 4:100~143
- 22 Svensson P, Huang Z. Geo-SAL: A Query Language for Spatial Data Analysis. In: Proc. 2nd Intl. Symp. on Large Spatial Databases, Zurich, Switzerland, 1991. 119~140
- 23 Berry J K. Fundamental Operations in Computer-Assisted Map Analysis. Int. Journal of Geographical Information Systems, 1987, 1(2):119~136
- 24 Tomlin C D. Geographic Information Systems and Cartographic Modeling, Prentice Hall, 1990
- 25 Raper J R, Maguire D J. Design Models and Functionality in GIS, Computers & Geosciences, 1992, 18(4):387~394
- 26 Smith T R, et al. Requirements and Principles for the Implementation and Construction of Large-Scale Geographic Information Systems. Int. Journal of Geographical Information Systems, 1987, 1(1):13~31
- 27 Karlsson J S, et al. AMOS User's Guide, Linköping University, Sweden, 1996
- 28 Qin Xiaolin. AMOROSE User's Guide, Royal Institute of Technology, Sweden, 1998

(上接第74页)

ATM 交换机方框图,其中 MB86683 NTC(Network Termination Controller)是一个用于宽带 ATM 网络的高度集成的网络终端控制器,它执行与物理介质相关的传输集中功能,MB86689A ATC(Address Translation Controller)则以 155Mb/s 的速率实时翻译 ATM 信元头信息、替换 VCI/VPI 域及添加 24 位路由标记,而 MB86680 SRE (Self Routing Switch Element)乃是自主路由交换矩阵。关于这些器件的详情参见有关资料<sup>[7,8]</sup>。

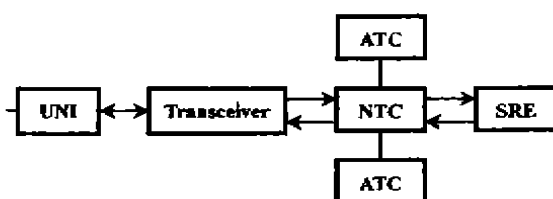


图6 基于 MB866 系列芯片的 ATM 交换机设计方案

#### 参考文献

- 1 陈永光,卢锡城. ATM 交换机缓冲策略研究. 计算机科学, 1998, 25(3):66~69
- 2 Re E D, et al. Performance Evaluation of Input and Output Queuing Techniques in ATM Switching Systems. IEEE Trans. on Communications, 1993, 41(10):1565~1575
- 3 Kuwahara H, et al. A Shared Buffer Memory Switch for an ATM Exchange. In: Proc. ICC' 89, 1989. 118~122
- 4 Banwell T, et al. Physical Design Issues for Very Large ATM Switching Systems. IEEE J. Selected Areas in Communications, 1991, 9(8):1227~1238
- 5 Simcoe R J. Perspectives on ATM Switch Architecture and the Influence of Traffic Pattern Assumptions on Switch Design. ACM SIGCOMM' 94, 1994. 93~105
- 6 Furhmann S W. Performance of a Packet Switch with Crossbar Architecture. IEEE Trans. on Communications, 1993, 41(3):486~491
- 7 FUJITSU, MB86683 Network Termination Controller Data Sheet, October 1993
- 8 FUJITSU, MB86689A Address Translation Controller Data Sheet, June 1995