



[12] 发明专利申请公开说明书

[21] 申请号 200410042286.6

[43] 公开日 2004 年 10 月 20 日

[11] 公开号 CN 1538694A

[22] 申请日 2000. 12. 28

[21] 申请号 200410042286.6

分案原申请号 00819186.7

[30] 优先权

[32] 1999. 12. 28 [33] US [31] 09/474039

[71] 申请人 ADC 宽带通路系统公司

地址 美国麻萨诸塞州

[72] 发明人 P·E·尼科利希 K·钦纳斯瓦米

P·H·多尔米策尔

W·A·H·恩格尔塞

W·G·马拉 H·恩艾

D·R·保利诺 K·B·皮尔斯

J·B·萨卡

[74] 专利代理机构 中国专利代理(香港)有限公司

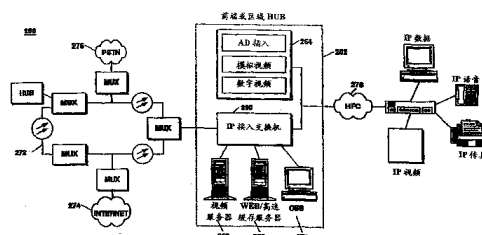
代理人 罗 朋

权利要求书 1 页 说明书 15 页 附图 19 页

[54] 发明名称 宽带网络中高可用性、直接、灵活
及可缩放数据分组交换的系统和方
法

[57] 摘要

一种系统和方法，用于在宽带数据处理系统和网络的节点之间直接和灵活地交换数据分组流的连接。系统作为单个 IP 交换机。



1. 一种机架，所述机架具有用于固定数据处理应用模块的多个相同槽位，每个槽位具有对底板网格的相同接口，所述底板网格含有网格线，其中每一个都提供用于每对所述接口之间数据通信的连续直接的信道，所述网格线具有以每秒 1.5Gbps 或更高的速率与数字数据分组传送相称的吞吐量。
2. 如权利要求 1 所述的机架，其特征在于还包括管理控制信号总线，所述相同接口与所述总线连接。
3. 如权利要求 1 所述的机架，其特征在于还包括电信总线。
4. 如权利要求 1 所述的机架，其特征在于在所述机架的底板侧为所述模块提供对外部数据端口的连接。
5. 如权利要求 1 所述的机架，其特征在于还包括所述模块的完全分布式 DC 电源。
6. 如权利要求 1 所述的机架，其特征在于所述相同槽位的数量大于四。

宽带网络中高可用性、直接、灵活及 可缩放数据分组交换的系统和方法

5

技术领域

本发明涉及通过数据处理系统的宽带网络进行的通信，所述宽带网络例如有采用 SONET 主干网络和混合光纤同轴电缆(“HFC”)接入网络的网络进行的通信，其中 HFC 把用户连接到主干的网络。

10 HFC 环境的分组通信标准为 CableLabs 的 DOCSIS(通过电缆业务接口规范的数据，文献号为 No. SP-RFI-I05-991105)。

背景技术

图 1 示出向 HFC 系统(其中的每个 HFC 系统与 200 - 2000 个通过住户(HHP: households passed)相连接)进行数据广域传送的典型配置。前端终接系统 2 经环 6 上的复用器分出与 SONET 环 4 连接。前端 2 通过光纤下行连接 7 和光纤上行连接 8 与光纤节点 9 连接。光纤节点 9 包含同轴连接器 10，与电缆调制解调器用户(即家庭 11)相连接。除前端终接系统的费用之外，这些复用器目前的费用约为

15 \$50000，并且对某个社区进行的业务扩充可能还需要新的复用器和服务器。

前端终接系统中某个组件的失效会使整个“下游”(从前端至终端用户)子网(图 7)无法与所连接的网络进行通信。

已经进行了对系统进行集成的尝试，以便降低费用及便于系统管理。目前的一种“集成”数据传送系统 20 在图 2 中示为功能块。

25 图 2 示出一种系统 20，它包含反向路径监视系统 22、IP 交换机 24、路由器 26、调制器和上变频器 28、供应系统 30、电话部分 32 以及多个 CMTS 34(电缆连接调制解调器终接系统)。这种系统通常采用多个厂家的多种系统来构成，具有不同的管理系统、大占地面积、高

功率要求以及高运营成本。

图 3 中示出目前的一种用于语音和数据传送的典型网络宽带电缆网。两个 OC-12 分组经 SONET(POS)链路 40、42 分别与两个主干路由器 44、46 连接,而这两个主干路由器 44、46 又与两个交换机 48、50 连网。交换机 48、50 与 CMTS 前端路由器 52、54 连网。CMTS 前端路由器 52、54 通过 CMTS 61 连接到多个光纤节点 56。交换机 48、50 还连接到多个电话中继网关 58,而电话中继网关 58 又连接到公共交换电话网(PSTN)60。当采用图 2 所示的“集成”系统时,这种网络通常也包含多个厂家的多种系统,具有不同的管理系统、大的占地面积、高的功率要求以及高的运营成本。仍然需要一种真正的集成解决方案来降低系统的尺寸、功率需求、成本,以及保证数据传送中较高的一致性和增强的可靠性。

发明内容

本发明的一个目的是提供一种集成系统和方法,用于高质量的语音、数据及视频业务的宽带传送。

本发明的另一个目的是提供一种系统和方法,用于电缆接入平台,具有高网络可靠性,同时具有可靠地支持重要的电话业务的能力以及提供分层语音、数据业务的能力。

本发明的另一个目的是提供一种系统和方法,用于安全及可缩放的网络交换机。

发明内容

本发明具有用于传送高质量语音、数据及视频业务的高度集成载波类宽带接入系统,本发明解决了为宽带网络环境中的数据处理提供集成网络解决方案的问题。

在本发明中提供了一种方法和系统,其中,利用由群集管理器所控制的完全网状公用机架(fully meshed common chassis)中广域网(WAN)接口卡、HFC 接口(电缆调制解调器终接)卡以及机架间链路

(ICL)卡的组合,在 IP(Internet 协议)或网络层(OSI 层 3)进行与 HFC 管道的交换。ICL 可以延伸数英里,而由群集管理器所控制的机架群集器则作为单个 IP 交换机。机架还可以连接其它宽带接入媒体,例如数字用户线(DSL)、固定无线本地多点分布业务(LMDS)、多信道
5 多点分布业务(MMDS)以及“光纤到路边”(FTTC)。

本发明提供了布置本地及广域配置的灵活性,使其适合于小型运营商要求已备路径(ready path)用于扩容。不同的媒体卡可以用于不同模式的宽带通信,诸如具有电话调制解调器回路的单路电缆。本发明还允许不同网络部件的多级冗余,使运营商能够通过具有特定
10 重要部件(stressed elements)的冗余来提供较高的可用性。本发明支持分层语音和数据系统。本发明提供一种小占地面积和较低功率要求的可靠的可缩放系统。

通过以下对附图所示本发明实施例的详细说明,可以更好地理解本发明和其上述优点及其它优点。

15

附图说明

图 1 示出一种先有技术网络,它具有 HFC 前端服务器,通过复用器分出与 SONET 环连接;

图 2 示出一种通过电缆传送系统的先有技术 HFC 数据;

20 图 3 示出一种先有技术数据传送网络;

图 4 是根据本发明原理的机架的方框图;

图 5 示出包含图 4 的机架的集成电缆基础结构;

图 6 示出一种网络,它在环型拓扑结构中采用多个诸如图 4 所示的机架;

25 图 7 示出第一分级网络,它包含按照本发明原理进行工作的主控机架和辅助机架;

图 8 示出一种星型网络,它包含多个诸如图 4 所示的机架;

图 9 示出第二分级网络,它包含多个诸如图 4 所示的机架;

- 图 10 示出一种网状网络，它包含多个诸如图 4 所示的机架；
图 11 是图 4 所示机架的应用卡和底板部分；
图 12 是底板互连的示意图，其中包括交换网格；
图 13 是根据本发明原理的 BAS 标题的方框图；
5 图 14 是千兆以太网机架间链路或出口应用模块的方框图；
图 15 是 ARP 表的方框图；
图 16 是机架中两个出口模块的方框图；
图 17 是路由服务器的方框图；
图 18 是机架控制器模块的方框图；
10 图 19 是网络管理体系结构的方框图；
图 20 是 CMTS 应用模块的方框图；
图 21 是底板网格接口的方框图；以及
图 22 是 MCC 标记的方框图。

15 具体实施方式

- 图 4 示出根据本发明原理进行工作的机架 200。机架 200 将多个网络接口和应用集成在单个交换机系统中。本发明的机架是完全网状 IP 交换机，具有高性能的分组转发、过滤以及 QoS/CoS(业务质量/业务类)能力，采用由机架控制器中的群集管理器所控制的低级嵌入式软件。分组转发、过滤以及 QoS/Cos 分布到机架中的应用模块(又
- 20 称作“卡”)。通过快速 IP 处理器在各模块中执行这些操作。机架控制器和群集管理器控制操作，并配置各模块。高层软件驻留在群集管理器中，包括路由服务器功能(RIPv1、RIPv2、OSPF 等)、网络管理(SNMP V1/V2)、安全、DHCP、LDAP 以及远程接入软件(VPN、PPTP、L2TP 以及 PPP)，并且能够容易地被修改或升级。
- 25

在本发明的该实施例中，机架 200 包含用于模块的十四(14)个槽位。其中十二个槽位安装应用模块 205，另外两个槽位安装机架控制器模块 210。各应用模块均包含板载 DC-DC 转换器 202，并且对于

机架是可热插拔的。机架控制器模块 210 用于冗余系统时钟/总线仲裁。机架还包含冗余电源模块。电源模块和 DC-DC 转换器构成机架中的完全分布式电源。可集成在机架中的应用的示例是 CMTS 模块 215、以太网模块 220、SONET 模块 225 以及电话应用 230。另一个应用可以是机架间链路(ICL)端口 235, 机架可以通过这个端口键接到另一个机架。ICL 端口是唯一的出口端口。可以采用任何出口模块, 例如 10/100 以太网、IG 以太网以及 SONET 上的分组(PoS: Packet-over-SONET)等, 来实现 ICL 端口。

图 5 示出包含图 4 所示机架 200 的集成电缆基础结构 260。机架 200 是用于语音和数据传送的区域集线器(HUB) 262 的一部分。集线器 262 包括视频控制器应用 264、视频服务器 266、Web/高速缓存服务器 268 以及操作支持系统(OSS)270, 它们连网到作为 IP 地址交换机的机架 200。机架 200 连接到集线器 262 外部的 SONET 环 272, 所述 SONET 环 272 包含 Internet 连接 274 和公共交换电话网(PSTN)连接 276。机架 200 还通过 HFC 链路 278 连接到电缆用户, 并提供基于 IP 的业务, 包括语音、视频数据以及传真业务。每个 HFC 应用模块能够处理多达 2000 个电缆调制解调器业务用户。用户数量的逻辑限制为 8192。机架可以支持多个 HFC 链路, 多个机架也可以连网在一起(如以下所述)以支持一百万以上的电缆调制解调器用户。

根据当前惯例, 有一条宽带信道(27-40Mbps)用于对用户(可以是台式计算机、传真机或电话机)进行传送(下载), 以及多条较窄信道(320 Kbps-10 Mbps)用于上装。这由 O/E 节点中具有双工的 HFC 卡来处理。本地 HFC 电缆系统或环路是具有对电缆调制解调器的引出的同轴电缆分布网络。

图 6 示出一种环型网络 300, 其中采用了多个图 4 所示的机架。多个机架 302、304、306 以及 308 经过全双工机架间链路 310、312、314 以及 316 连接为环。如果机架 304 或其中的 ICL310、312 出现故障, 则该环路配置允许机架 302 及其有关电缆网络通过机架 308 与

机架 306 及其有关电缆网络进行通信。

图 7 示出第一分层网络 328, 它包含按照本发明原理进行工作的主控机架 330 和辅助机架 332。主控机架连接到光纤通信线路(OC-n)334、336。图中给出了 OC-12 线路, 但同样也支持 OC-3 和 OC-48。
5 主控机架 330 还可以连接到 PSTN 338。主控机架 330 和辅助机架 332 通过全双工链路 340 进行键接, 其中, 全双工链路 340 可以是快速以太网(FE)、千兆以太网或 SONET 上的分组(PoS)类型连接。辅助机架 332 连接到多个光节点 342。在包含一个主控机架以及一个或多个辅助机架的网络中, 主控机架采用其群集管理器来管理群集。

10 其它配置也是可行的, 包括: 如图 8 所示的“星型”配置 350, 其中包含主控机架 352 以及辅助机架 354、356、358、360、362, 或如图 9 所示的分层布置 380, 其中包含主控机架 382, 第二层的辅助机架 384、386 以及第三层的辅助机架 388、390、392、394。图 10 中示出网格配置 400, 包含机架 402、404、406、408 以及 410, 其中, 每个机架均包含键接其它机架的机架间链路。主控和辅助机架
15 由系统配置所确定。每个机架基本上与其它机架相同。

上述配置通过采用长距离光纤链路可以延伸数十英里, 或者可以布置在相同主前端站点的单个机柜中。

图 11 示出与图 4 所示机架 200 的底板 420 连接的应用模块。每个应用模块 422 通过网格通信芯片(MCC)424 与底板 420 连接, 将在下面对 MCC 424 进行详细说明。各 MCC 424 包含十二(12)个串行链路接口 426, 其中的十一个通向底板 420。各应用模块中的十一个通向底板的串行链路用于将应用模块连接到该机架中的其它每个应用模块。一条链路用于模块其自身的连接, 即环回。应用模块 422 通
20 过机架管理总线 432 与机架控制器 428、430 连接。辅助机架控制器 430 可选地用于冗余, 以便使系统更为可靠。在一个最佳实施例中提供了第二机架管理总线(未示出), 同样用于冗余/可靠性。

底板是完全网状的, 这就是说, 通过串行链路, 各应用模块均

具有与机架中其它应用模块的直接点对点链接。网状底板中的各网格线均以 1.5 Gbps 或更高的速率为数据通信提供连续的直接信道。作为示例, 图 11 中仅示出连接 200 的一部分。图 12 示出底板网格。

5 具有 MCC 串行链路的 12 条信道编号为 0 至 11。该编号称作信道 ID 或 CID。下面将对信道进行详细说明。底板中的槽位同样编号为 0 至 11(槽位 ID 或 SID)。但是, 机架系统并不要求信道 0 连接底板的槽位 0。串行链路可以连接到任意的槽位。槽位 ID 根据系统拓扑结构动态地进行配置。这允许底板接线自由, 减少路由复杂度。

10 再来看图 11, 各应用模块还连接到机架管理总线 432, 其中, 机架管理总线 432 为模块提供与机架控制器的连接。

15 为了通过机架间链路(ICL)进行机架间的分组交换, 以及为了通过 MCC 链路进行机架内的分组交换, 该机架包含一个机架间交换层。机架间交换层位于 IP 层 3(L3, 网络层)的下面。机架中的分组处理分为两种类型的业务: 单播和广播/多播。通过机架间交换层进行的交换是利用图 13 所示的机架间标题 500(又称作机架间标记)来实现的。

Bas_Type 字段 502 具有 x 位, 并表示机架间标题类型。该字段可以用来表示光纤接口地址(FIA)具有逻辑端口, 表示分组是广播还是单播分组, 并表示 BAS 标题中的相关字段。Bas_Type 字段 502 用于对 BAS 标题进行解码, 并协助分组转发。

“Keep (保持)” 字段 504 具有 x 位, 并确定是否由于拥塞而失掉分组。

优先级字段具有 x 位(图 13 中未示出), 并确定分组优先级以及由于拥塞而对分组进行排队的位置。

25 分段字段 506 具有 x 位, 并表示分组分段以及该分组是否由两帧组成。

next_Hop 字段 508 具有多位, 并被用于机架内分组传送的下一跳信息。该字段是无效的, 除非 Bas_Type 字段另有表示。

Encap_Type 字段 510 是一位字段，表示分组是否还需要层 2(数据层，在机架间交换层的下面)处理，或者指示该分组是否可以不进行其它处理而被转发。

Mcast_Type 字段 512 是一位字段，表示分组是广播的还是多播。

- 5 Dest_FIA_Type 字段 514 是一位字段，表示目的地 FIA 是以短形式(<机架/槽位/端口>)还是以长形式(<机架/槽位/端口/逻辑端口>)来提供的。

Src_FIA_Type 字段 516 是一位字段，表示源 FIA 字段是以短形式还是以长形式来提供的。

- 10 Data_Type 字段 518 具有多位，并被用来表示净负荷中所携带的业务类型。

TTL 字段(未示出)具有多位，并且是一种故障安全机制，以便防止分组不确定地停留在系统中。每次在 ICL 端口接收到 TTL 时，TTL 递减。如果该字段在递减之后为零，则丢弃所述分组。

- 15 “Forwarding Info (转发信息)” 字段 520 具有多位，并包含转发表修正、forward_table_entry 密钥/标识以及下一跳信息。

“Destination FIA (目的地 FIA)” 字段 522 具有多位，并表示分组的最终目的地。该字段包含机架/槽位/端口以及逻辑端口信息。零机架值具有特定含义，用“主代理”(下面将进行说明)来表示机架。

- 20 零端口值也具有特定含义，它表示所述分组的接收方是应用模块。逻辑端口可以用来表示模块中接收该分组的栈/实体。因此，所有边缘端口和 ICL 都是基于 1 的，即端口号大于零。

“Src FIA” 字段 524 具有多位，并表示分组的源。该字段由路由服务器用来标识输入分组的源。

- 25 根据系统体系结构，插入机架中的所有应用卡共享公共特性。这些应用卡也称作数据处理器、数据应用模块以及数据通信模块。图 14 是机架中以太网应用模块 550 的方框图。该应用模块连接到网格底板 552 以及管理总线 A 554、B 556。在以太网应用模块 550 中，

快速 IP 处理器(FIPP)558 连接到 PCI 总线 560 以及 F 总线 562。FIPP 558 具有高级 RISC 机(ARM)处理器核 564 和六个微引擎 566。FIPP 558 与储存分组的 DRAM 568 和储存路由表的 SRAM 570 连接。包含 8 个端口的以太网装置 572 连接到 F 总线 562。MCC 574 提供与网格底板 552 的连接。以太网装置 572 提供与机架外部的连接。两个 MAC 装置连接在 PCI 总线 560 和管理总线 A 554 之间。MAC A 576 连接在 PCI 总线 560 和管理总线之间。MAC B 578 连接在 PCI 总线 560 和管理总线 B 556 之间。

当数据分组进入以太网装置 572 时，它经过 FIPP 558 并储存在 DRAM 568 中。微引擎 566 并行检测数据分组。微引擎 566 在分组中查找 IP 地址，然后在储存于 SRAM 570 的转发表中查找目的地地址。转发表提供分组离开的机架、槽位以及端口。当分组通过以太网装置进入时，该分组具有以太网标题。如果分组离开其进入的端口，则不会加入机架间标题。FIPP 确定发送端口，并经该端口发出分组。如果分组由不同的槽位发出，则消除以太网标题，并将 BAS 标题加入到该数据分组。BAS 标题中包含的最少限度的信息为目的地址数据、机架、槽位以及端口信息。分组还包括 IP 标题和净荷。FIPP 具有发送队列，发送前对分组进行排队。分组经 F 总线 562 发送到 MCC 574，后者以 64 字节块发出数据。

当分组通过以太网模块 550 的 MCC 574 进入时，应用模块 550 将该分组或该分组的指针加入 DRAM 570 的队列中。如果分组离开一个串行端口，则 FIPP 558 在 SRAM 570 的 ARP 表中查找目的地。ARP 表如图 15 所示。FIPP 查找机架、槽位以及端口。机架地址不会改变，除非目的地地址为 ICL。在相同机架的情况下，应用模块查找 IP 分组的 MCC 地址，并将其发出。如果分组将发送到其它机架中，则应用模块询问是否有任何端口为 ICL。如果是，则该应用模块将分组经该端口发出。

图 20 示出 CMTS 应用卡，它又称作 HFC DOCSIS 接口卡。由

于当前可利用的 Broadcom(宽带通信)技术的限制, 所以目前采用基于 PCI 的 HFC 技术(HFC-on-PCI technology), 但将来可采用其它类型的接口技术。

图 16 示出机架中的两个应用模块 600、602。应用模块 600、602。通过管理总线 604、606 以及网格底板 608 相连接。第二应用模块 602 包含连接到 PCI 总线 612 的路由服务器 610。机架的路由服务器可以布置在任何应用模块中, 在本发明的该实施例中, 路由服务器为 Pentium®处理器。机架的网络中的各机架包含路由服务器。但是在网络中, 一次只能有一个路由服务器是活动的。路由服务器通过管理总线以及通过作为路由器的 ICL 来进行广播。

在多个机架的网络中, 指定一个路由服务器为主控路由服务器。其它路由服务器为辅助的。网络中的路由服务器每 30 秒发出路由信息。主控路由服务器将其机架、槽位以及端口向其它所有路由服务器广播。

图 17 从概念上示出由“控制”或主控路由服务器进行的路由服务器的更新。

所有转发表都知道其机架中路由服务器的位置。主控路由服务器包含等效于以太网端口的逻辑端口。

当分组被发送到连网在一起的机架的其中之一中的路由服务器时, 主控路由器接收分组信息, 就像主控路由器正在接收分组本身。主控路由器建立路由表和转发表, 并将它们向所有应用模块及网络中所有其它机架广播。转发表是从路由表的信息中建立的。转发表包含诸如分组机架、槽位、端口、以及 QoS 数据、CMTS 信息、多播域信息和下一跳 ICL 信息的信息。

图 18 示出连接到机架中管理总线的机架控制器。为了冗余, 各机架具有两个机架控制器。机架控制器没有连接到网格底板。机架控制器承担网络管理任务和供应任务。机架控制器使整个机架群集器能够看起来像一个被管理部件。机架控制包含处理器、存储器以

及 CI(Craft Interface)。在本发明的该实施例中，处理器为 Pentium® 处理器。CI 是采用基于 10/100 以太网的网络管理接口。

逻辑上，图 8-10 所示的群集共同起作单个路由器的作用，即 IP 层 3 交换机，具有多个外部端口。物理上，这些外部端口分布在多个机架上，可以分布在远程位置。ICL “虚拟化”外部点的物理分布。

逻辑上，将群集器作为单个实体进行管理(图 19)。从 IP 的角度来看，每个机架中的每个卡具有管理子代理。各卡本地的子代理代表用于管理目的的卡。群集管理器的管理服务器(主控代理)与子代理进行通信，同样也“虚拟化”子代理的物理分布。

主控和辅助机架两者均能支持各种传送媒体，最初集中于 HFC 接口来支持电缆调制解调器的连接，以提供 Internet 接入和 IP 电话。这些机架能够支持某些电缆调制解调器基础结构的基于电信的返回路径、以及常规(PSTN)的基于调制解调器的 Internet 接入。

群集管理器 100 运行诸如 Windows NT 的通用操作系统。它是一组更为具体的实体，包括：(1)主控代理(MA)或网络管理器；(2)应用模块；(3)资源管理器(RM)，用于提供卡资源的统计；(4)用户管理器(UM)，它维护由群集所服务的 IP 终端用户的动态观察；(5)转发表管理器(FT)，用于创建/分布转发表；(6)过滤管理器(FM)，用于类属 IP 过滤，(7)带宽(业务质量)/业务类管理器；以及(8)负荷均分管理器(LS)。

群集管理器中的系统管理服务器还将群集机架表示为单个虚拟实体。群集管理器还可作为拨入调制解调器的 RAS 服务器。群集管理器中的 RADIUS 客户机与网络管理员提供的 RADIUS 服务器进行通信。

象 Web 服务器和邮件服务器的通用功能可以在群集管理器中运行。FTP 和 Telnet 允许管理员远程访问群集管理器。

本发明允许资源共享。例如，进入辅助机架的 HFC 接口的 IP 业务有可能将语音作为其净荷。该 IP 分组可被交换到不同位置的不

同机架中的 VOIP/PSTN 卡上。具有语音净荷的 IP 分组由分布式 IP 交换结构路由选择到适当的 VOIP/PSTN 卡中。始发机架用最高“业务类”优先级来标记分组，允许以最小等待时间进行路由/交换决定。群集管理器保持群集系统中所有资源(包括可用的 DS0 的位置)的集中资源分布图。

机架控制器通过管理总线与机架中的应用模块进行通信。机架控制器和应用模块均具有用于通信的代理。机架控制器具有主控代理，应用模块具有子代理(图 19)。子代理将其机架、应用模块以及槽位信息传送给所述机架中的主控代理。

图 19 示出网络中的多个机架。在多个机架的环境中，上面提到的第一机架被指定为主控机架。主控机架具有附加的主控代理，称作“主-主”。网络中的所有机架都与主-主代理进行通信。每个机架中的各应用模块都将其 IP 地址(“10.机架.槽位.端口”)传送给其机架的主控代理以及主-主，如图 19 所示。IP 地址中的数字“10”表示专用网地址。

主-主代理通过管理总线来分布其机架 ID。在机架被指定为主控机架之后，该主控机架的机架 ID 变到一。当第二机架出现在网络中时，链路检测协议(LDP)被用来确定其存在。LDP 消息经当前机架的每条链路发出。如果发送机架接收到返回信号，则机架控制器将其标识为 ICL。LDP 使每个机架都能标识其 ICL 链路。主控机架通过其 ICL 链路广播它具有主-主代理。

图 21 是网格通信芯片(MCC)的方框图。MCC ASIC 经表示为全双工串行链路 215 的高速差分对(high speed differential pairs)来提供对机架中所有其它卡的连接。每个差分对以大于每秒 1Gbps 的数据吞吐量进行工作。F 总线接口 805 将 MCC 300 连接到 FIFO 总线(F 总线)。十二个发送 FIFO 810 和 12 个接收 FIFO 815 连接到 F 总线接口 805。每个发送 FIFO 具有并行-串行数据压缩器(总共 12 个数据压缩器，820)，每个接收 FIFO 具有数据扩展器(总共 12 个数据扩展器，

825)。十二个串行器 830 为数据压缩器 820 和数据扩展器 825 提供服务, 每个串行器用于一个压缩器和一个扩展器。MCC 中的信道被定义为串行链路, 连同其编码/解码逻辑、发送队列及接收队列。从信道引出的串行线路连接到底板网格。所有信道可以同时发送数据。

5 本发明的该实现采用网格通信芯片与采用串行链路技术的完全网格中多达十三条 F 总线互连。每个 MCC 具有两个 F 总线接口和十二个串行链路接口。MCC 以从 64 字节至整个分组的可编程大小增量, 在 F 总线上发送和接收分组。它包含十二个虚拟发送处理器 (VTP), VTP 从 F 总线得到分组并经串行链路将其发出, 允许同时发出十二个分组。VTP 读取分组前面的 MCC 标记, 并动态地将其连接到标题所表示的目的地槽位。

10 卡/槽专用处理器、卡/槽专用 MAC/PHY 对(以太网、SONET、HFC 等)及 MCC 在双向 F 总线(或多条单向 F 总线)上进行通信。分组发送路径是从 PHY/MAC 到处理器, 然后从处理器到 MCC 并离开网格。处理器在 FIPP 中进行层 3 和层 4 查找, 以确定分组的目的地和业务质量(QoS), 根据需要修改标题, 并在发送前预先考虑对分组的 MCC 标记。

15 分组接收路径是从网格到 MCC 再到处理器, 然后从处理器到 MAC/PHY 并离开信道。在把分组发送到 MAC 之前, 处理器移去 MCC 标记。

20 图 22 示出分组标记, 又称作 MCC 标记。MCC 标记是 32 位标记, 用来通过底板网格对分组进行路由选择。标记由槽处理器在向 MCC 发送分组之前加入到该分组前面。标记包含四个字段: 目的地掩码字段、优先级字段、保持字段以及保留字段。目的地掩码字段是保持分组被指定到的当前机架中槽位的掩码用的字段, 它可以是或不是系统中的最终目的地。对于发送分组, MCC 利用目的地掩码来确定分组被指定的发送队列。对于接收分组, MCC 利用优先级字段和保持字段来确定在过度被提交的槽位中要丢弃的分组。在本发

明的该实施例中没有使用保留字段。

MCC 具有两种独立的发送模式选择器，即槽位-信道映射和虚拟发送模式。在槽位-信道映射模式中，MCC 将 SID 透明地映射到 CID，并且软件不需要跟踪映射。在虚拟发送模式下，MCC 半透明地处理多播分组。MCC 取单个 F 总线流，并将其送往多个信道。MCC 中的发送端口对虚拟发送处理器(VTP)而不是对槽位进行寻址。F 总线接口把分组送往被选虚拟发送处理器。VTP 保存来自 MCC 标记的目的地掩码字段，并将分组数据(包括 MCC 标记)转发给目的地掩码中所表示的发送队列集。分组的所有后续 64 字节“块”由槽处理器采用相同端口 ID 来发送，并被送往相同的 VTP。VTP 把分组的块转发给从 MCC 标记所保存的目的地掩码字段所表示的发送队列集。当块到达并设置了 EOP 位时，VTP 清除其目的地掩码。如果指向该端口的下一个块不是新分组的开始(即设置了 SOP 位)，则 VTP 不会将该块转发给任何队列。

MCC 维护一组“信道忙”位，利用它们来防止多个 VTP 将分组同时发送给相同的 CID。这种防止冲突机制不是为了协助槽位处理器管理忙信道，而是为了在槽位处理器意外地将两个分组同时发送给同一个槽位的情况下防止分组的完全破坏。当 VTP 得到新的分组时，将目的地 CID 掩码与信道忙位进行比较。如果任何一个信道忙，则从目的地掩码中将其去除，并对该 CID 记录错误。然后，VTP 对其余所有目的地信道设置忙位，并发送分组。当 VTP 看到 F 总线上分组的 EOP 时，它清除其目的地 CID 的信道忙位。

F 总线接口在 MCC 和应用模块其余部分之间执行 I/O 功能。应用模块将图 22 所示的 32 位分组标记(MCC 标记)加到每个要通过网格被路由选择的分组。

F 总线上接收或发送的数据达 64 位宽。在数据传送中，F 总线接口把 4 个状态位加到发送数据，以形成 68 位数据段。F 总线接口把 68 位数据段分出到从分组标记所确定的适当发送 FIFO 中。来自

发送 FIFO 的数据被传送到有关数据压缩器中，其中 68 位数据段减少至 10 位数据段。然后再将数据传送给有关串行器，其中进一步将数据减少为串行流。串行流经串行链路发送给底板。

5 来自底板的数据通过串行链路到达有关信道。该信道的串行器把该数据扩展为 10 位数据段，并且有关数据扩展器将该数据扩展为 68 位数据段，68 位数据段被传送给有关 FIFO，然后再从 FIFO 传送到 F 总线接口。

快速 IP 处理器(FIPP)配备了 32/64 Mb 高速同步 SDRAM、8 Mb 高速同步 SRAM 以及引导快闪(boot flash)。FIPP 具有 32 位 PCI 总线
10 和 64 位 FIFO 总线(F 总线)。FIPP 向所有 F 总线连接装置传送分组数据以及从所有 F 总线连接装置传送分组数据。它以单播和多播两种模式提供 IP 转发。路由表是通过管理总线从机架路由服务器中接收的。FIPP 还提供高层功能，诸如过滤和 CoS/QoS。

15 各线路卡具有产生各卡所需的所有时钟的时钟子系统。这将会锁定到系统时钟和管理总线仲裁卡所提供的基准时钟。

各卡具有热插拔、加电复位电路以及健全定时器(Sanity Timer)功能。所有卡都具有板载 DC-DC 转换器，以便从底板中的-48V 导轨转换为应用所需的电压。某些卡(诸如 CMTS 卡)同样具有两个单独、绝缘的电源，以便使卡的模拟部分的性能达到最佳。

20 以上按照编排时序说明了本发明的最佳实施例。应该清楚，在不脱离本发明的主旨的情况下，可对等效的组件和功能进行替换。在保持本发明权益的情况下，硬件和软件实现的各种组合均是可行的。由于本发明的高度灵活性和可缩放性，所以重要的是本文所述模块的合作，而不是模块和端口的数量。

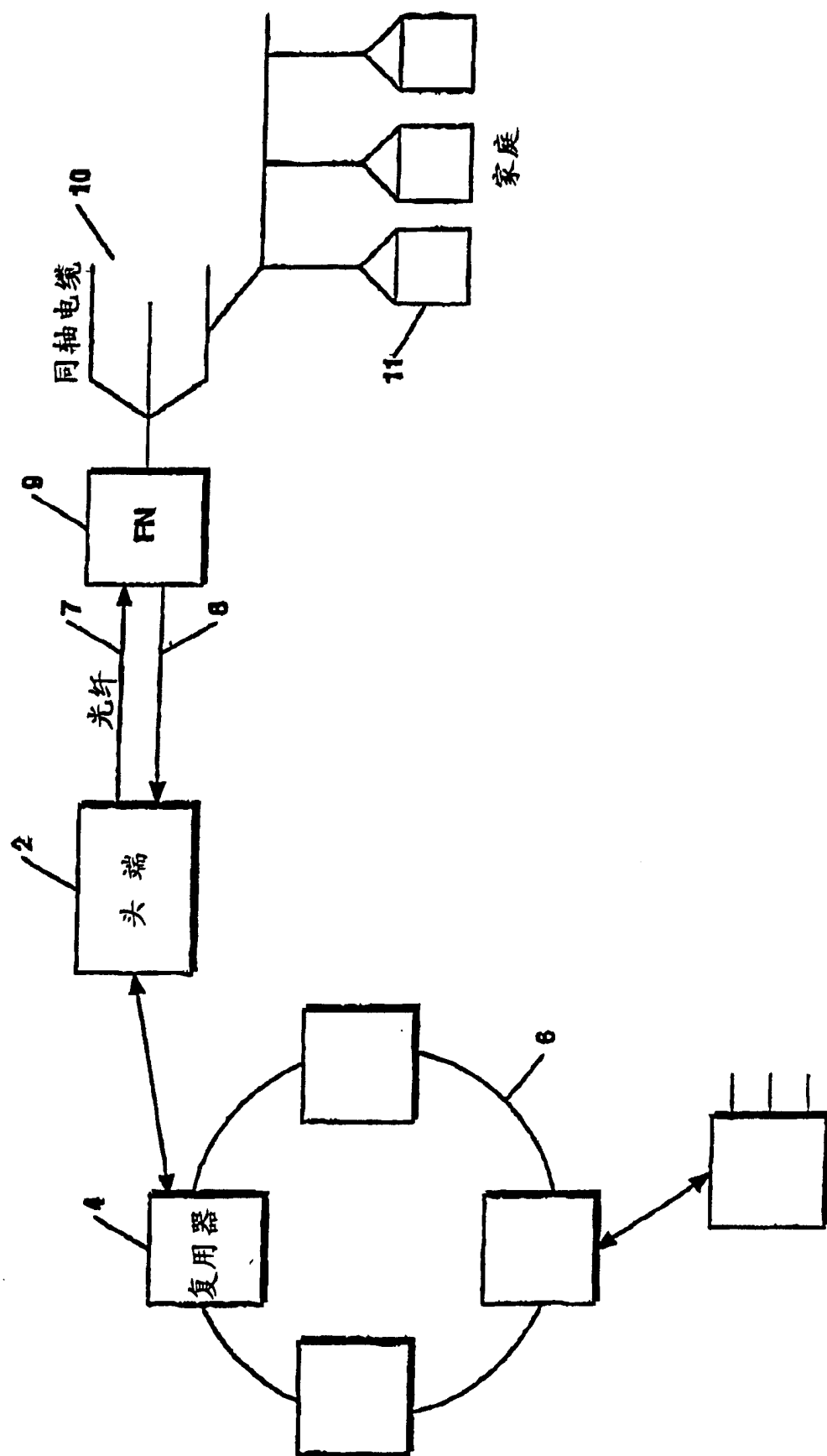


图 1 (现有技术)

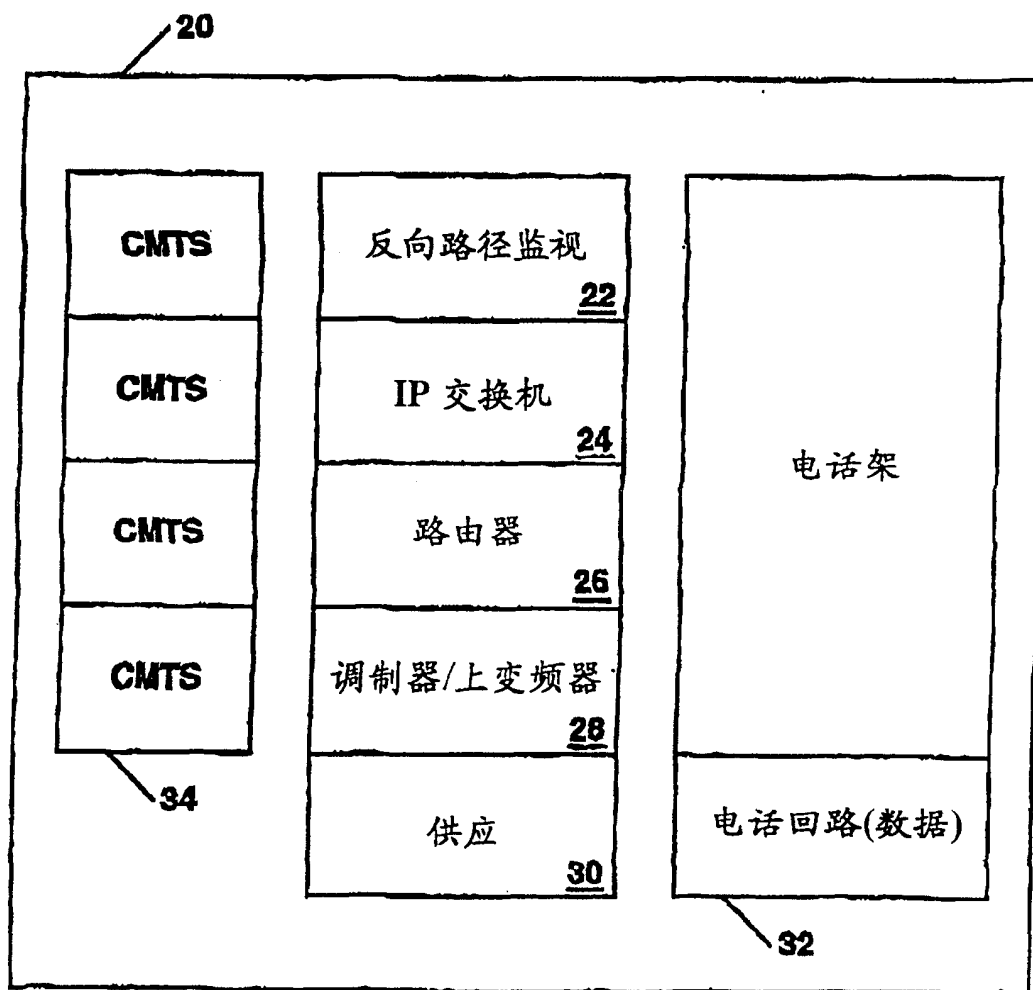


图 2 (现有技术)

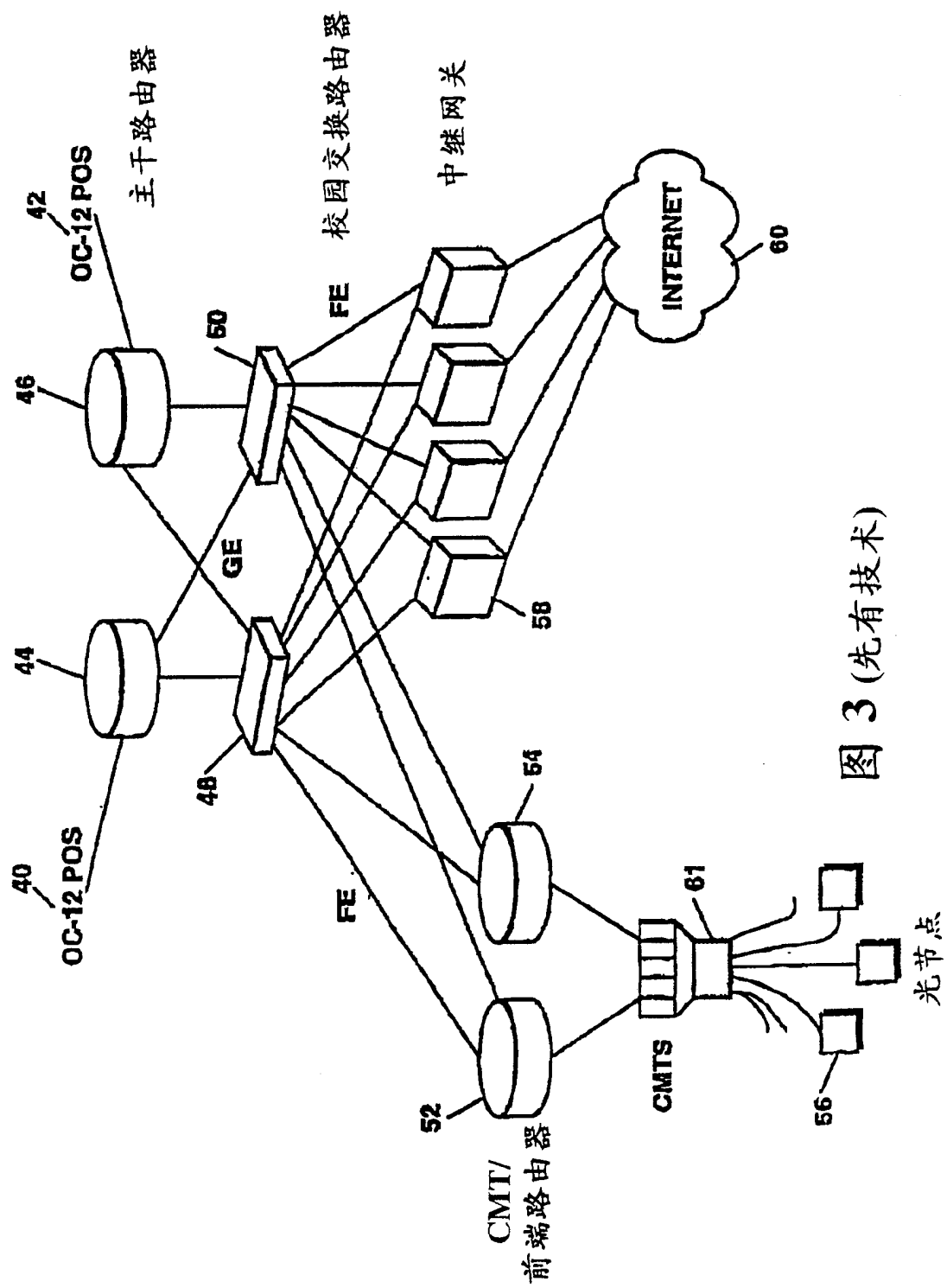


图 3 (现有技术)

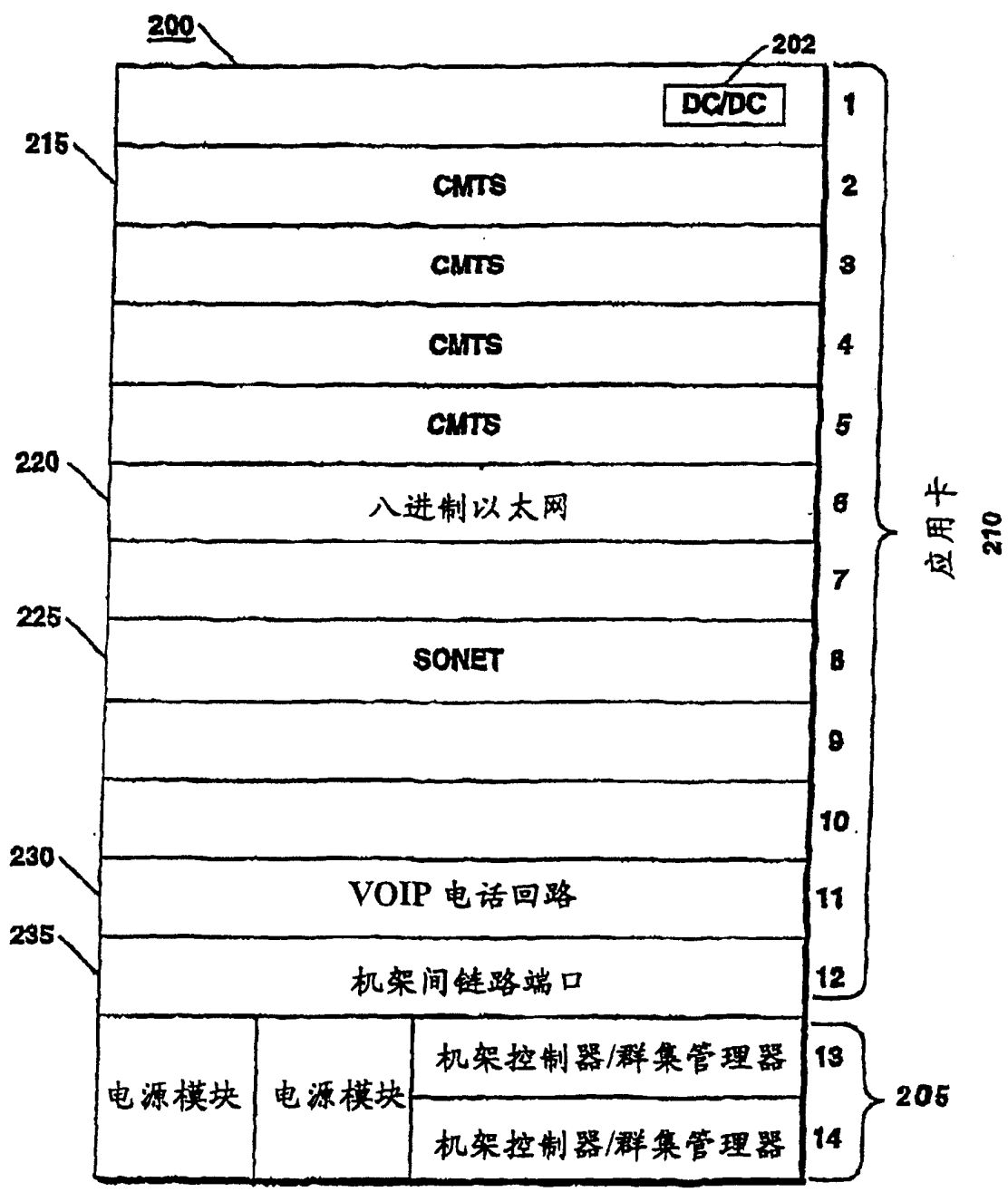
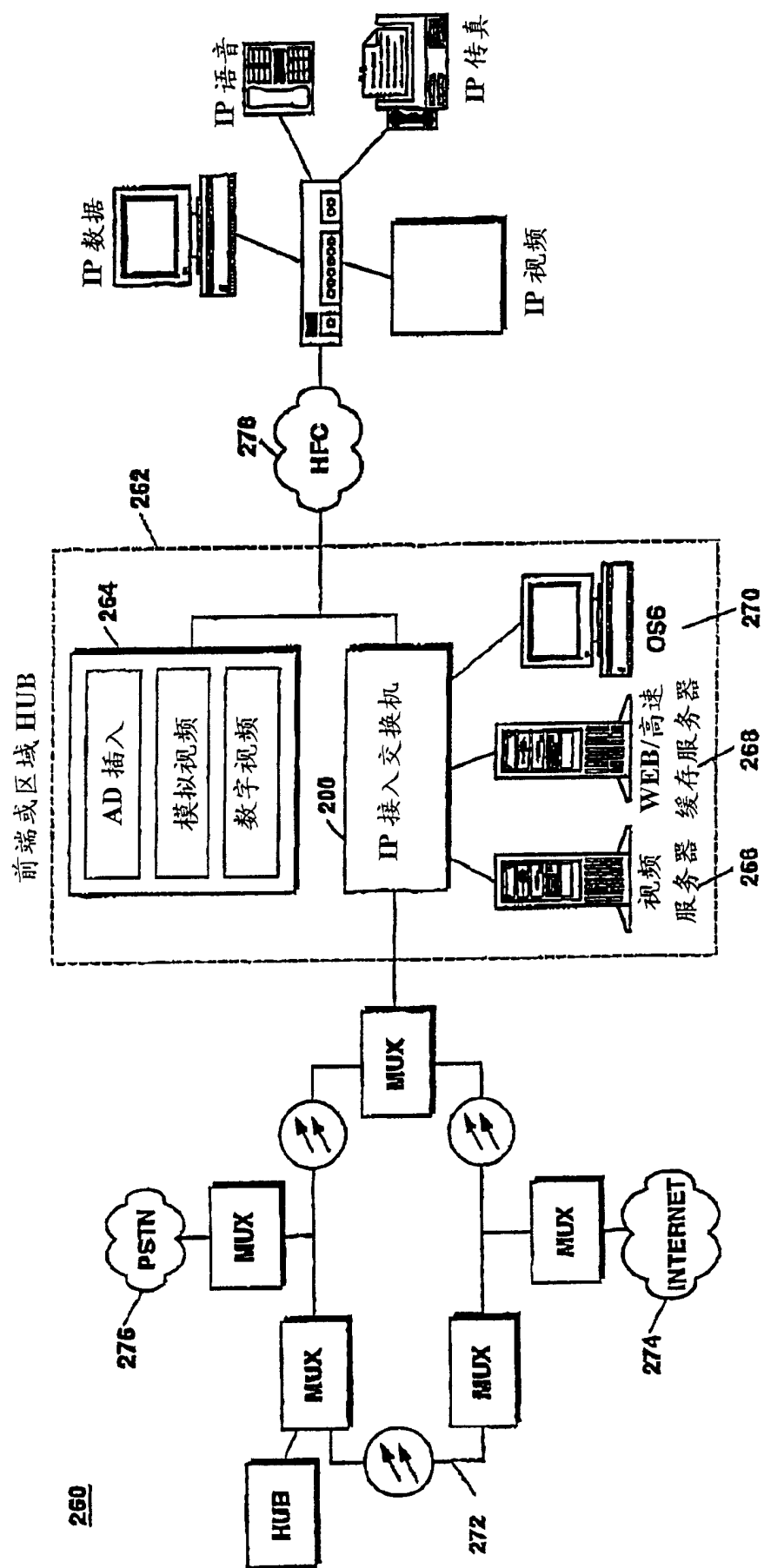


图 4



五
四

图 6

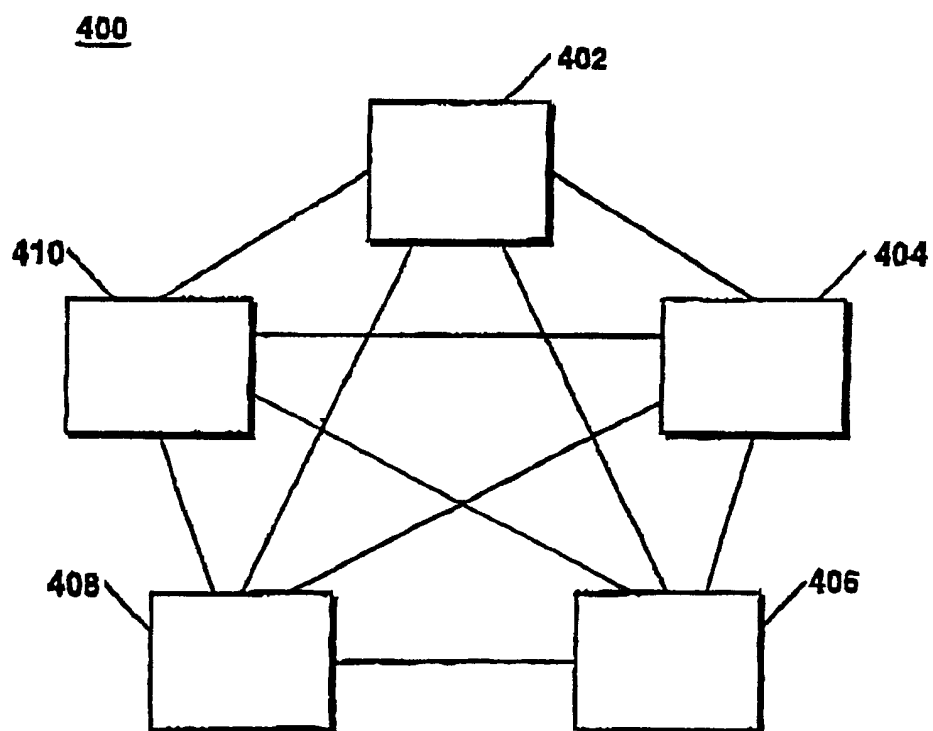
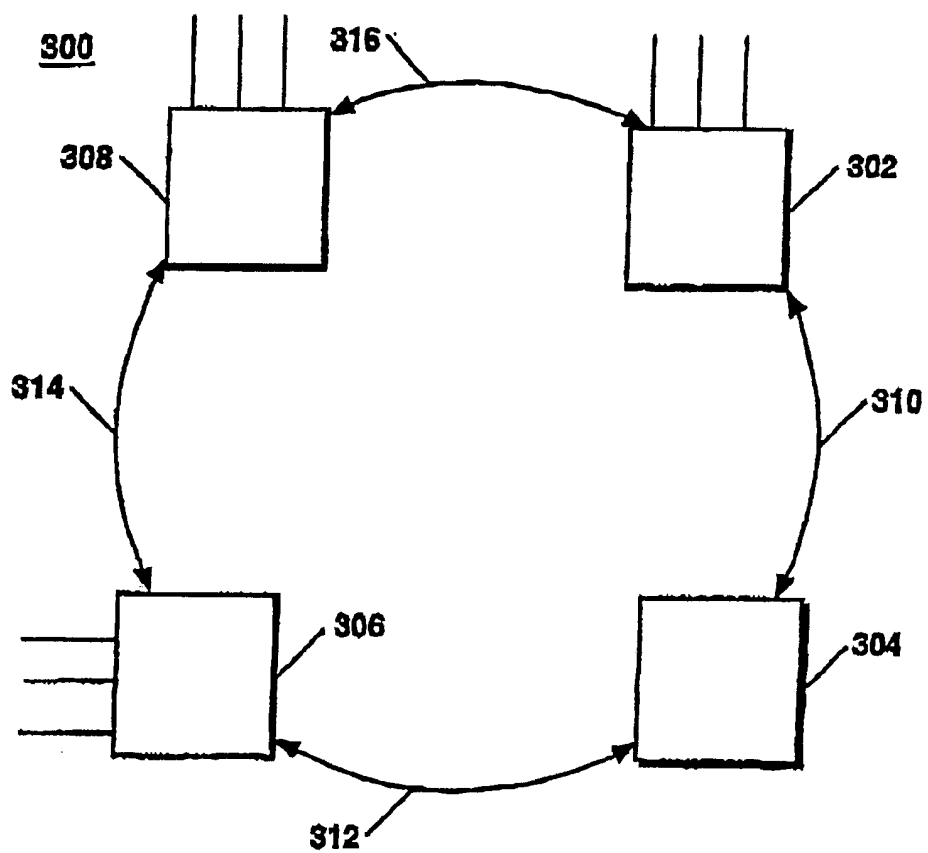


图 10

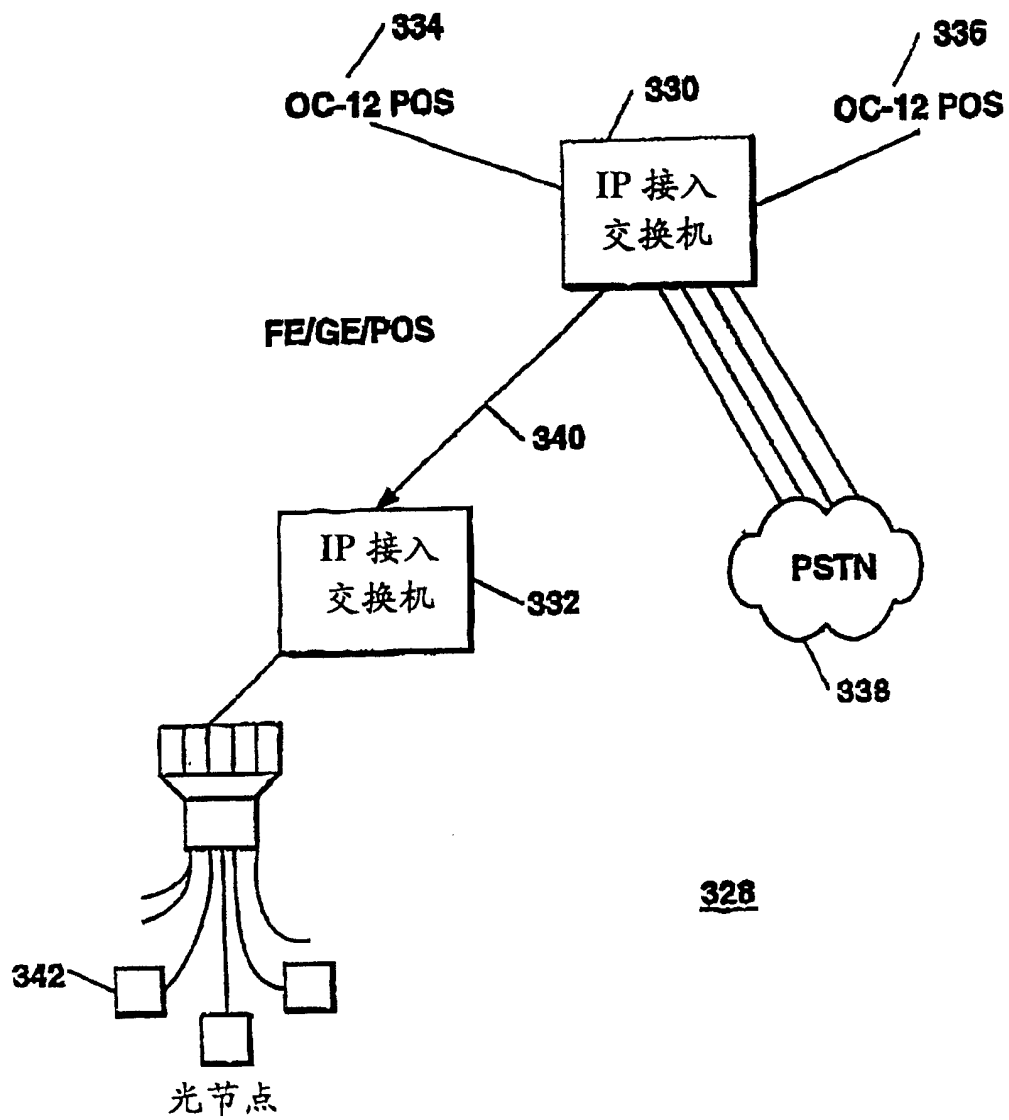


图 7

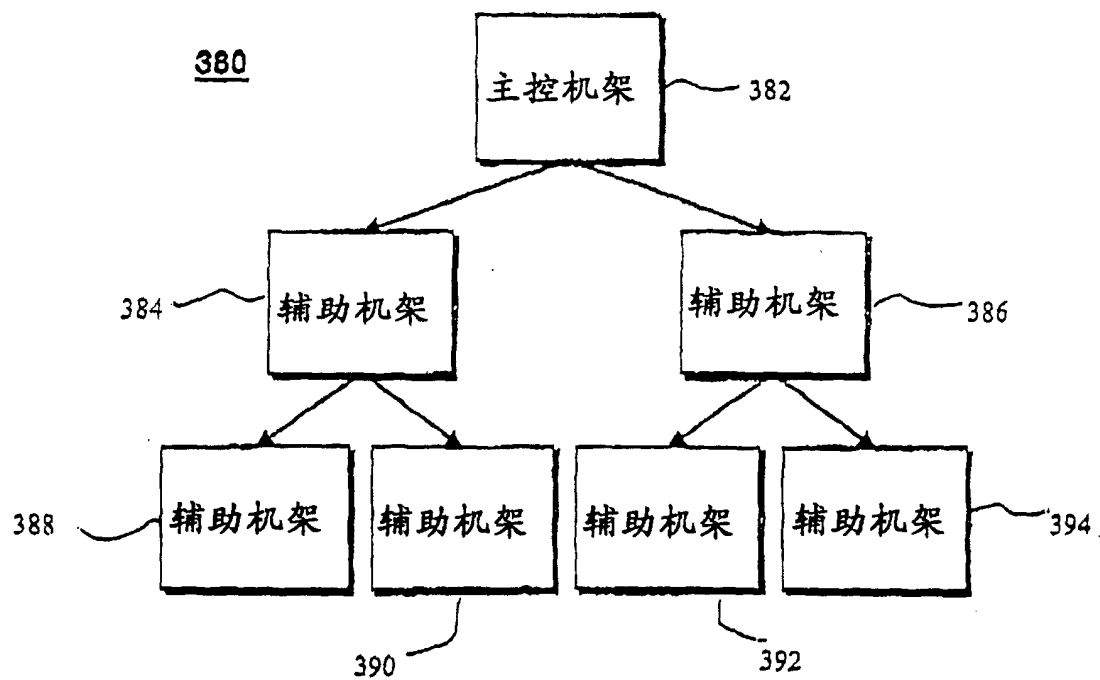
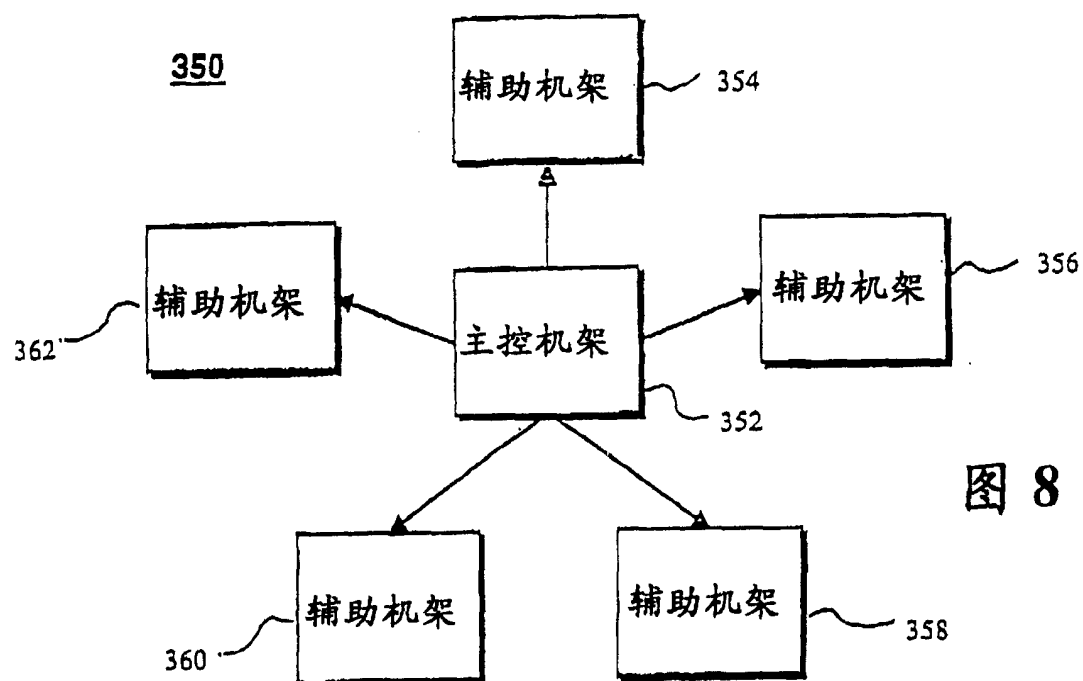


图 9

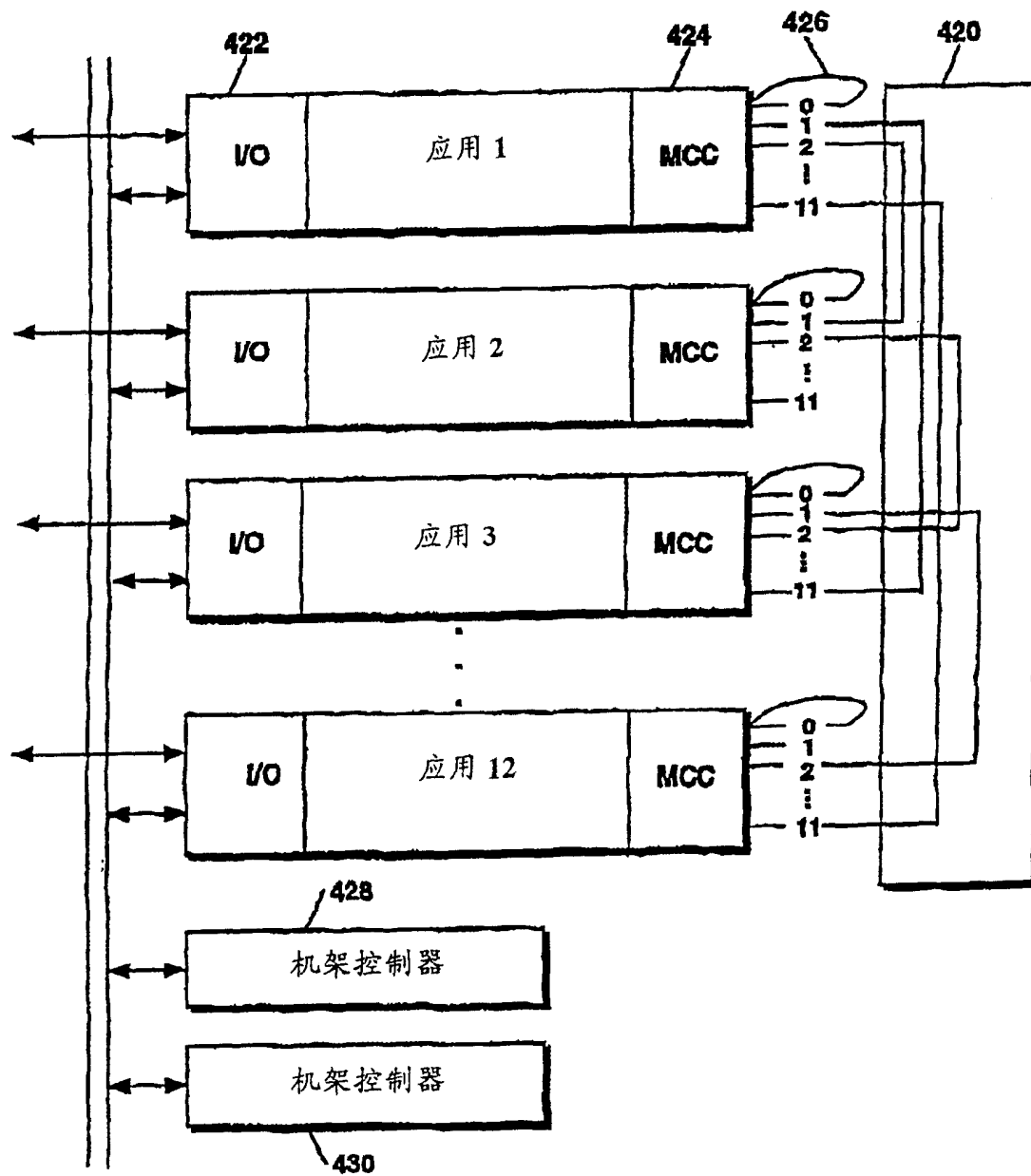


图 11

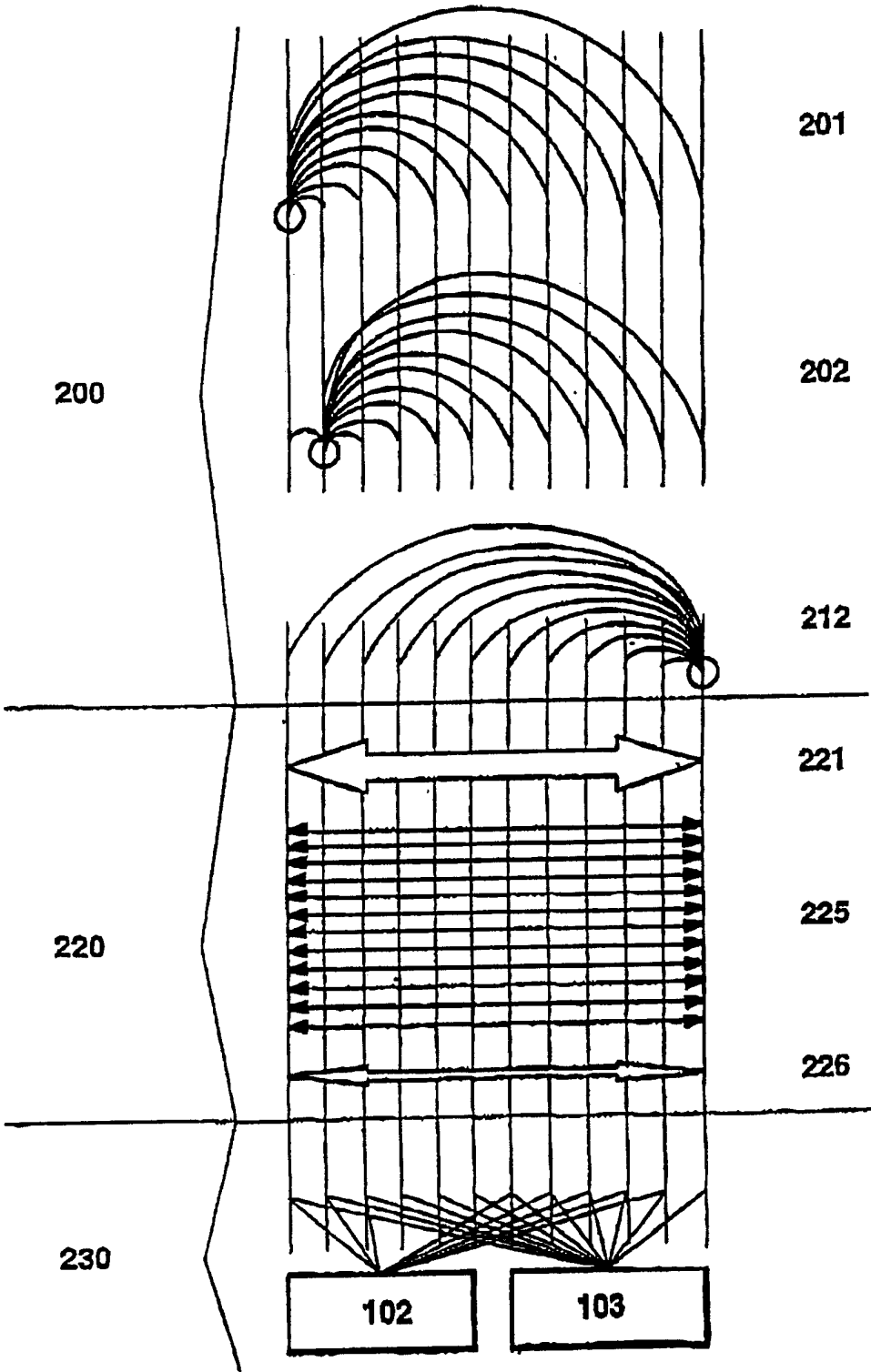


图 12

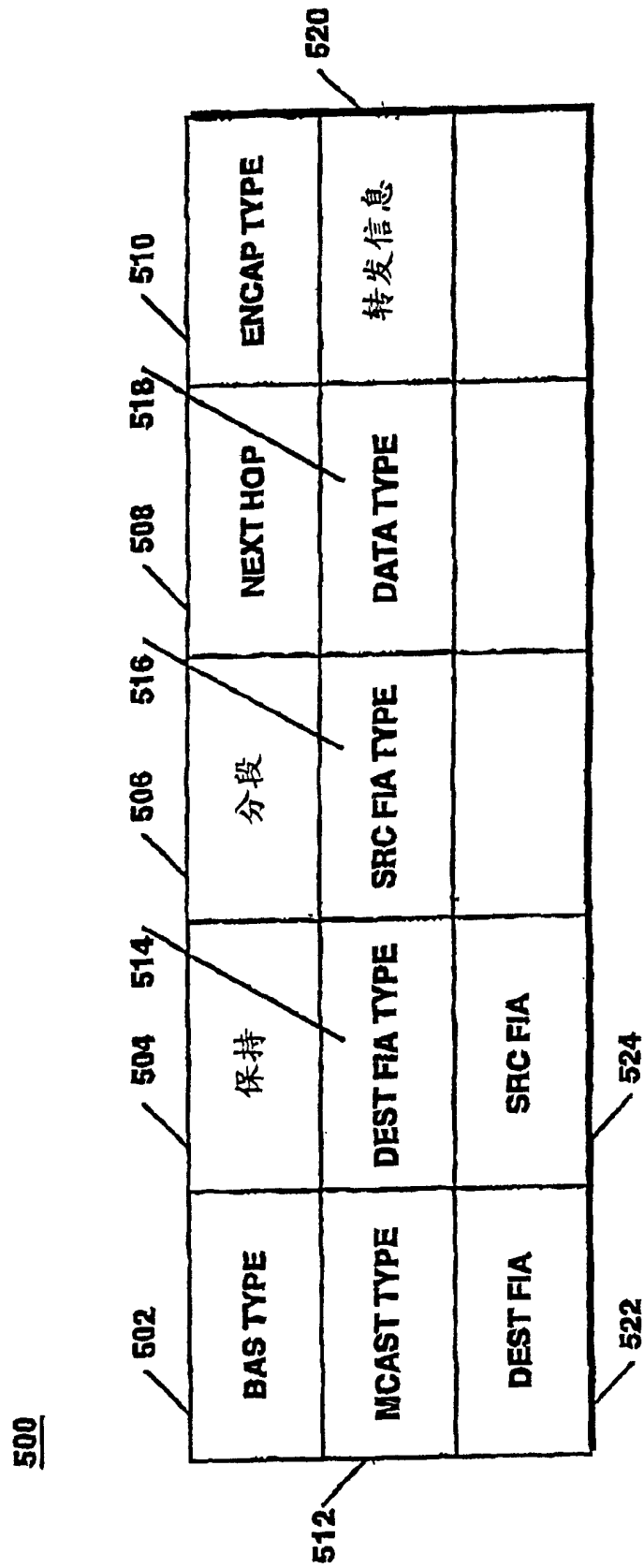


图 13

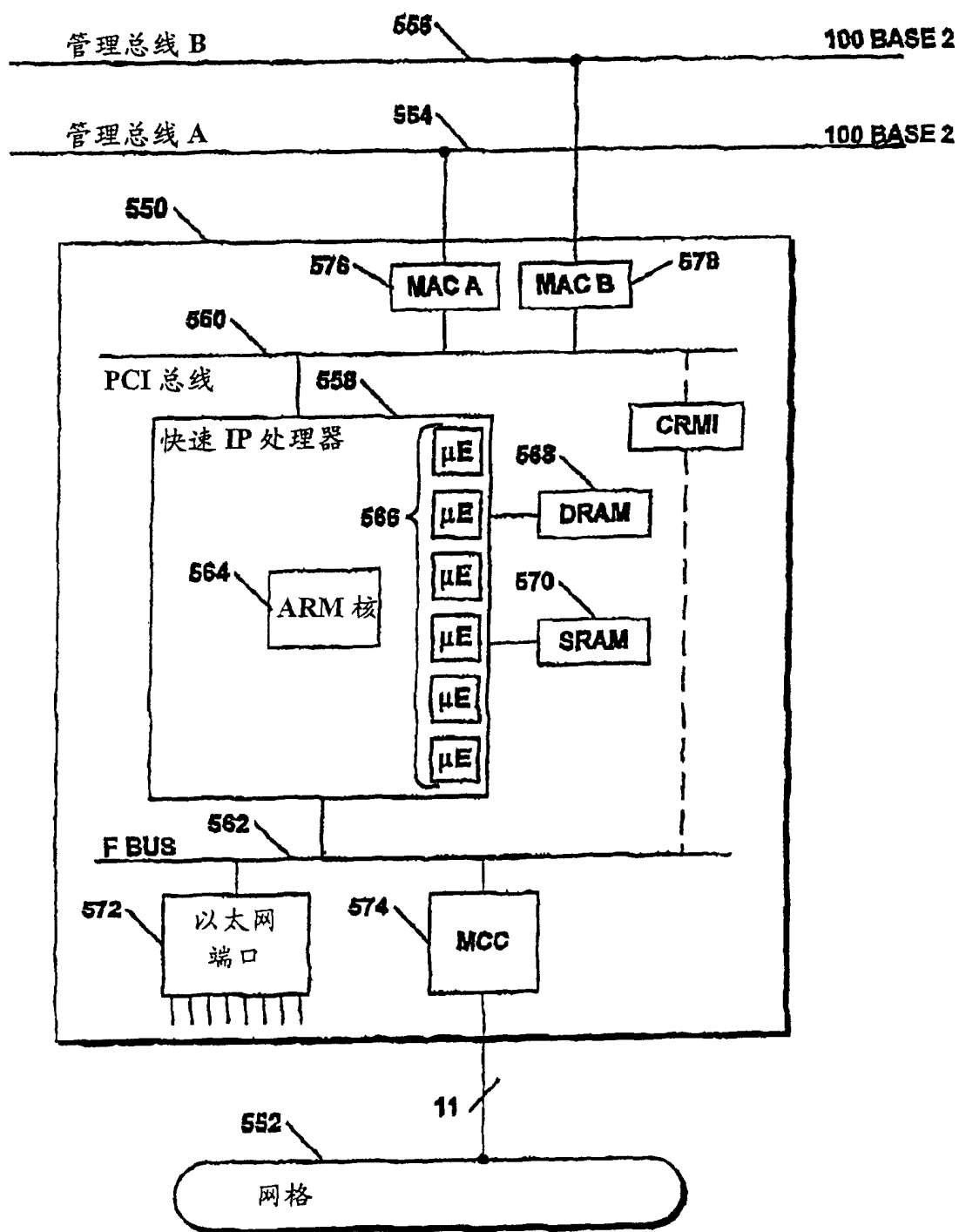


图 14

图 15

IP 地址	MAC 地址

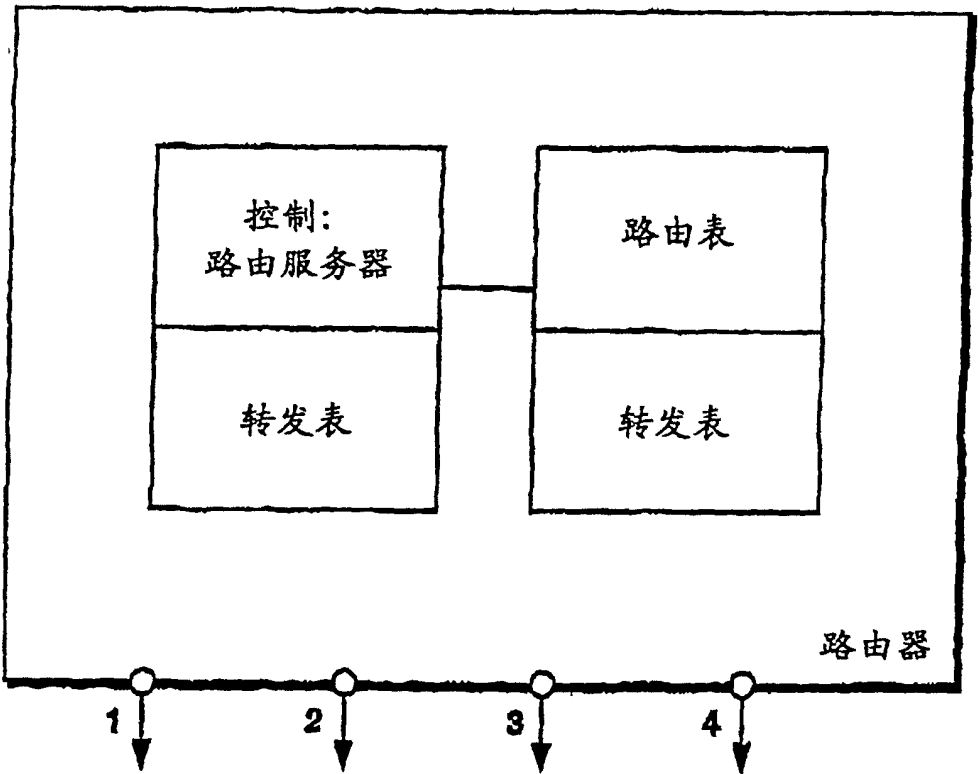


图 17

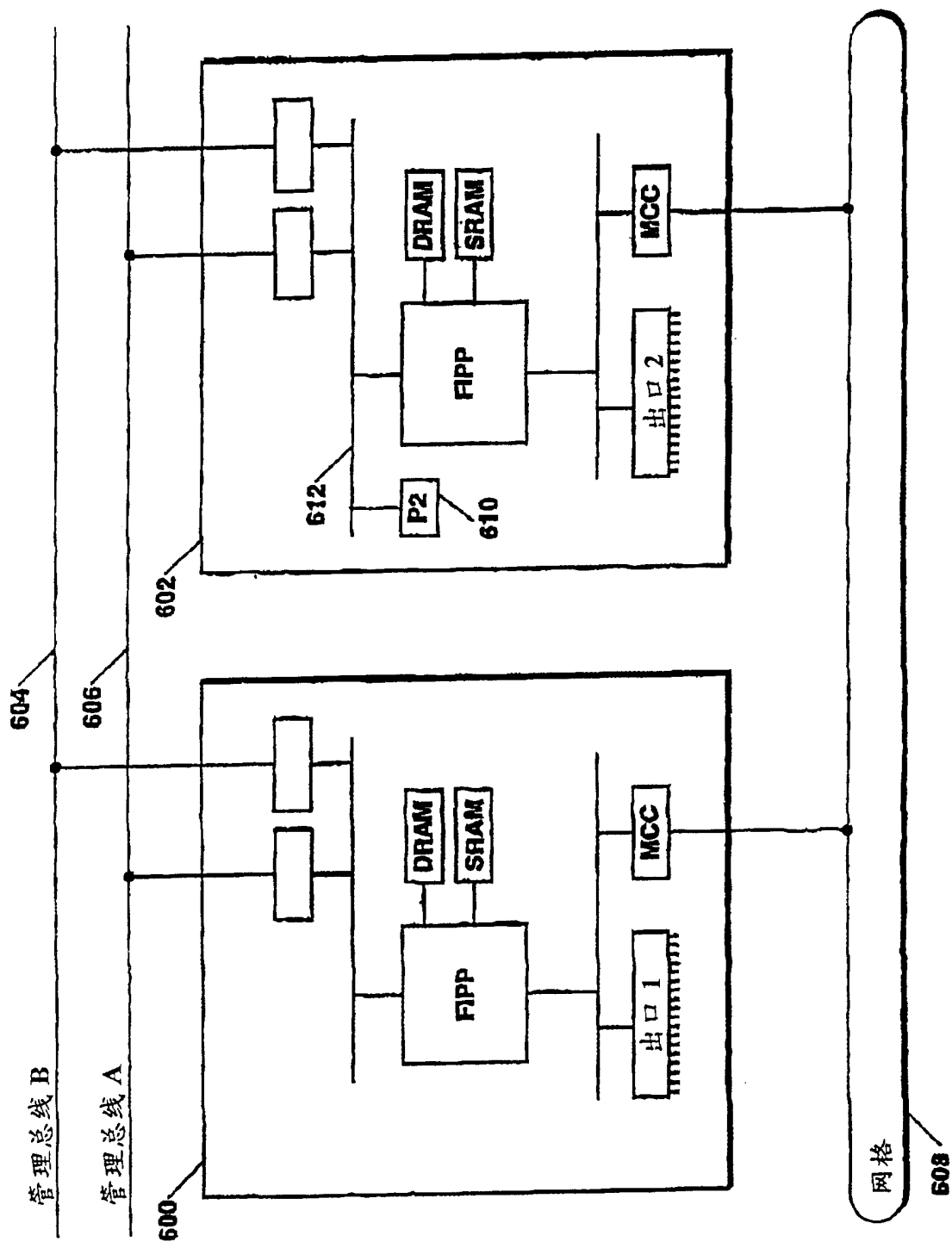
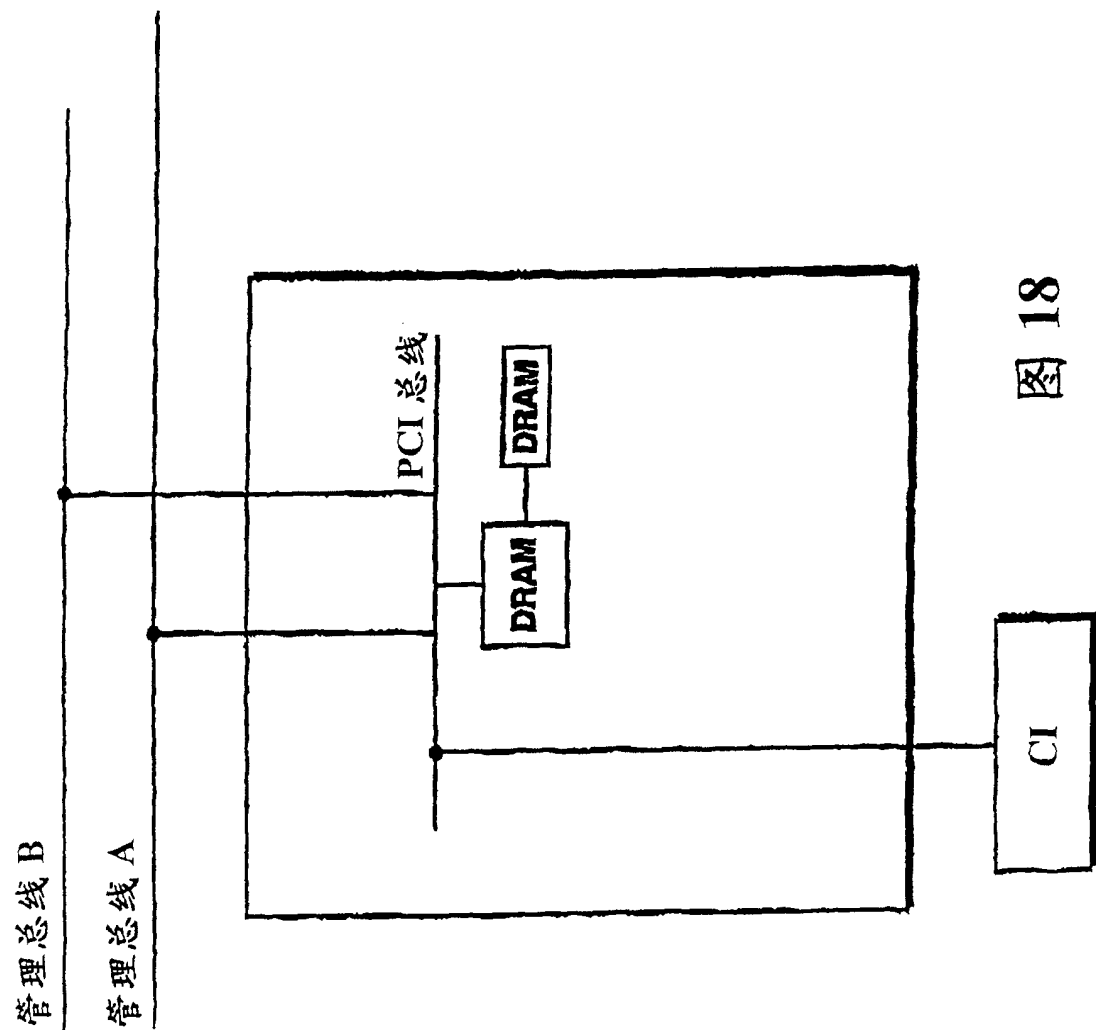


图 16



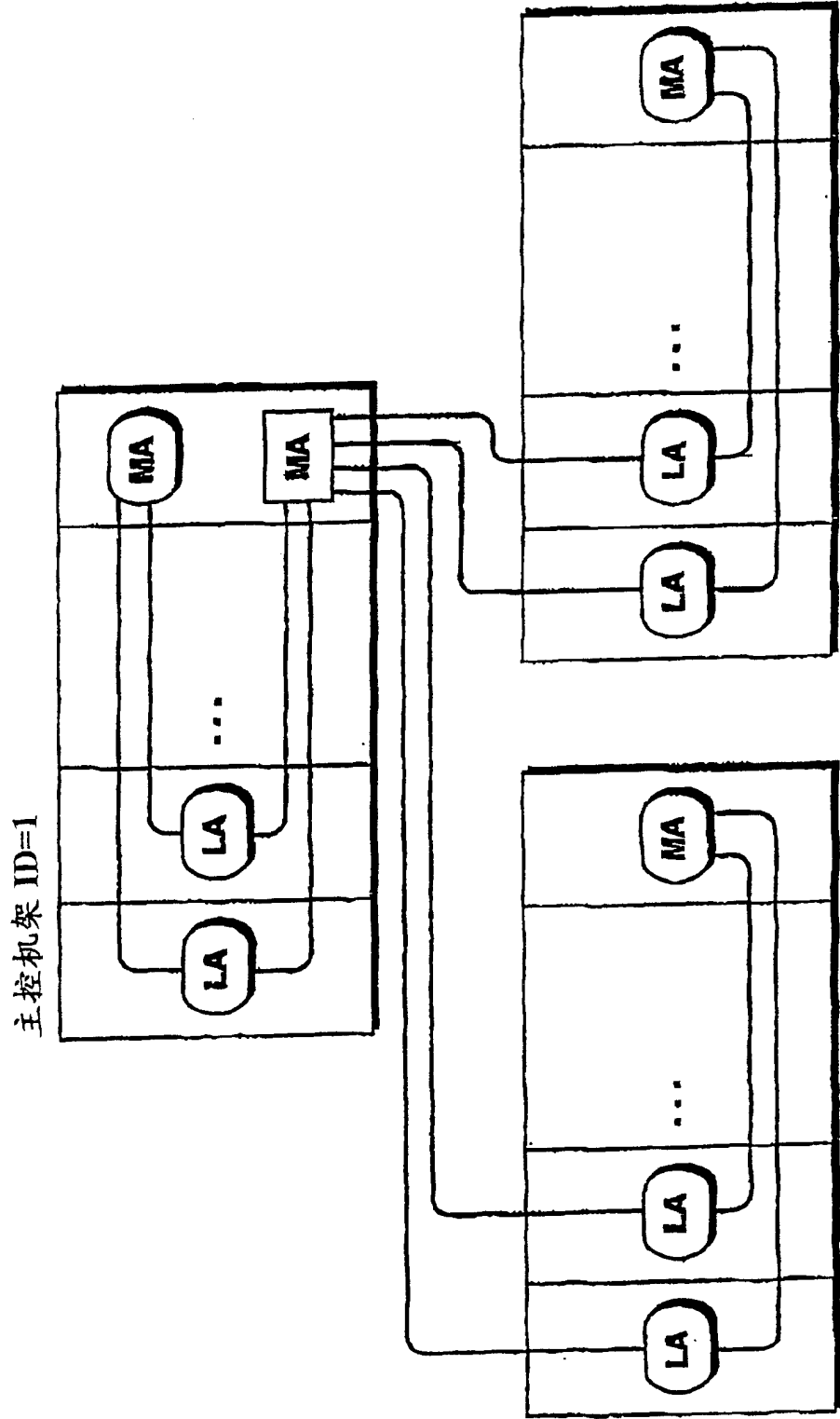


图 19

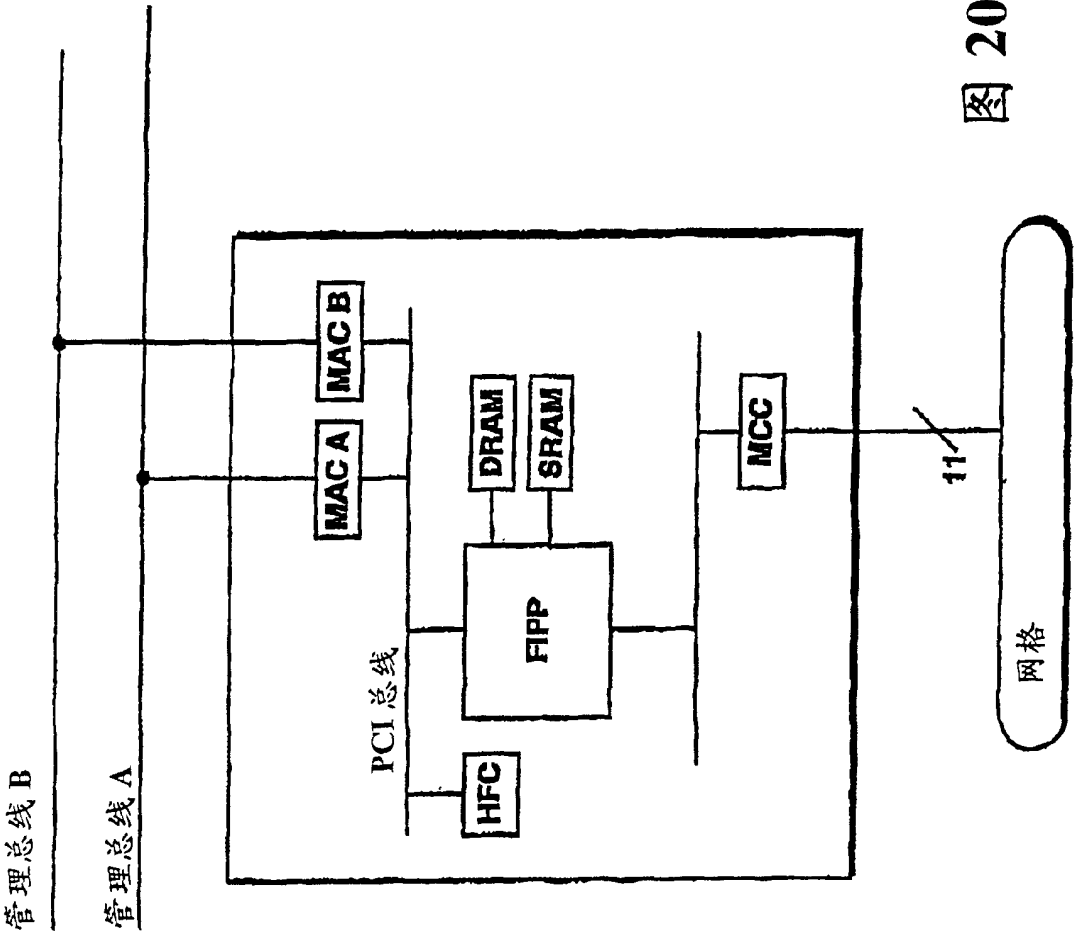


图 20

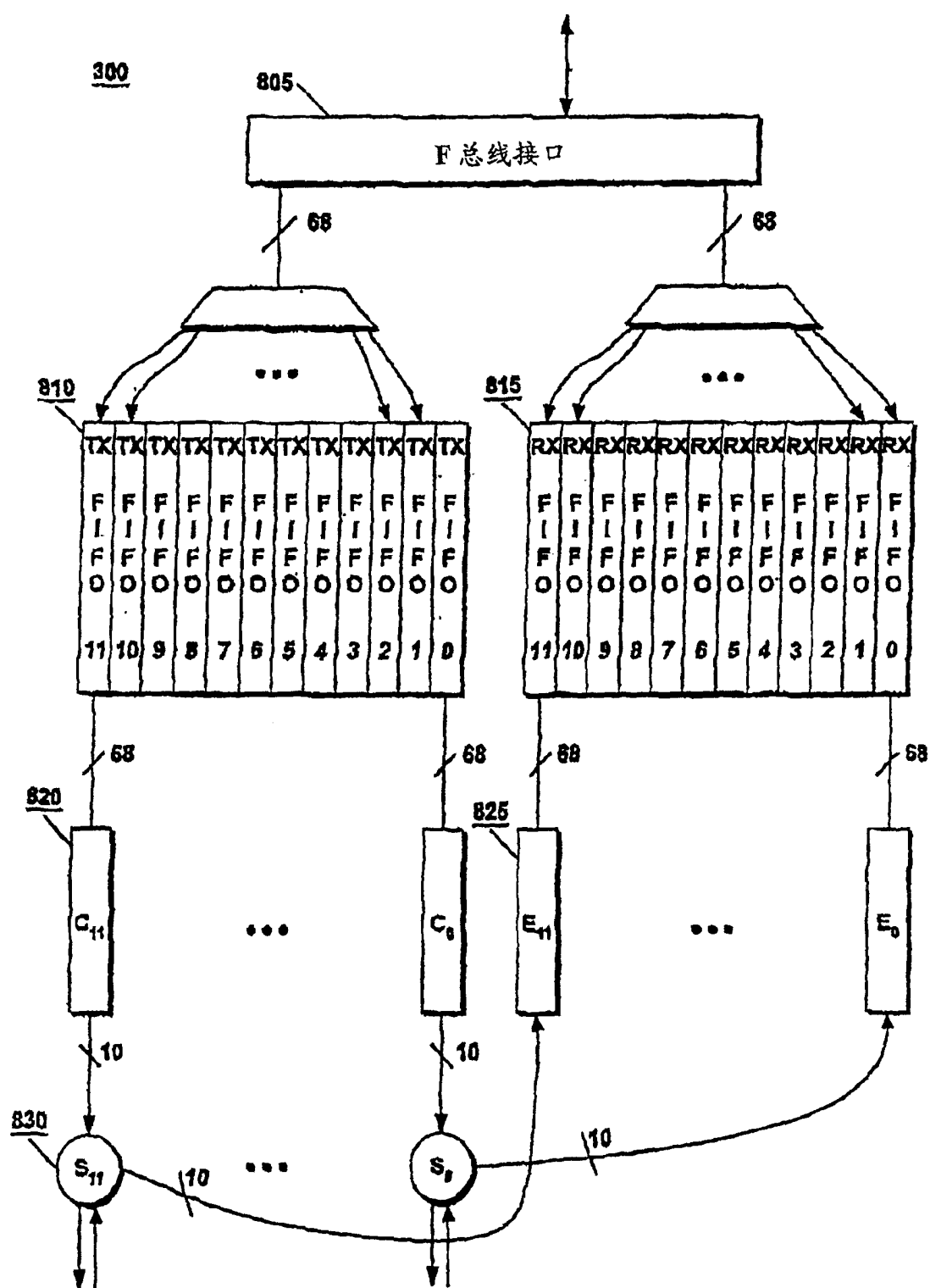


图 21

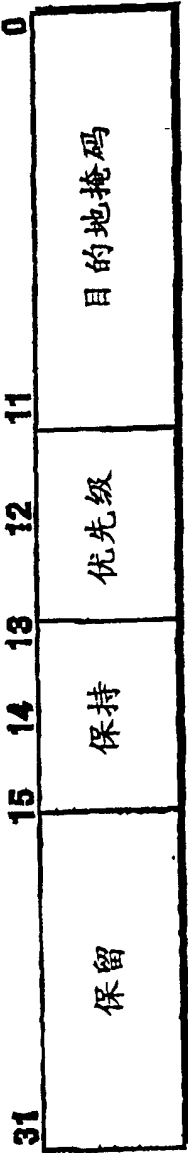


图 22