Macrochip：新型片上光网络

作者：陈峥，华北电力大学；李慧、顾华玺，西安电子科技大学ISN国家重点实验室

本文介绍了一种新的片上光网络——macrochip，并逐一介绍了其可行的三种网络实现方式，并就其良好的可扩展性和性能进行了说明。

随着半导体工艺技术的发展，集成电路设计者能够将越来越复杂的功能集成到单个硅芯片上。片上系统（SoC：System on Chip）正是在集成电路向集成系统转变下应运而生的。SoC 中可包含一个或多个处理器、存储器、模拟电路、数模混合电路以及片上可编程逻辑等 IP 核。

当芯片包含的 IP 核数目增至成千上万的时候，以总线结构为基础的 SoC 在性能、功耗、时延和可靠性等方面面临着巨大的挑战。为了解决 SoC 面临的相关问题，1999 年前后几个研究小组提出了一种全新的集成电路体系结构：片上网络（NoC：Network on Chip），其核心思想是将计算机网络技术移植到芯片设计中来，借用多处理机系统和常规通信网络中的一些基本思想，从体系结构上彻底解决总线结构带来的问题。

由于电连接存在电磁干扰、时间延迟、时钟歪斜、串扰、能耗高等问题，人们开始寻找一种新的连接方式来取代电连接。由于近几年光电集成器件的发展，使片上光网络（ONoC：Optical Network on Chip）成为可能。与 CMOS 兼容的光子技术的出现，使其成为在 22nm 条件下具有突出优势的互连方式。相比于传统的电连接方式，光互连有如下的明显优势：1、没有串扰；2、在无源光网络中可以提供更高的速度和带宽；3、能耗小；4、电磁干扰低。

本文将介绍美国 SUN 公司设计的 macrochip[1] 片上光网络。与以往的 ONoC 不同，macrochip 不是一个多核处理器芯片，而是将多个多核处理器芯片集成在一个封装里，芯片之间的通信采用光连接实现。这种设计可以使整个系统的性能与一个庞大的芯片相媲美。

macrochip 的结构

Macrochip 的网络结构嵌入在 SOI（Silicon-on-Insulator）衬底上，其基本组成单位是被称之为站点（site）的独立 CMOS 芯片。该设计旨在将多个大小为 225mm^2 左右的芯片集成在一起，使之具有与大小为 64×225mm^2 的单一芯片相同的性能。

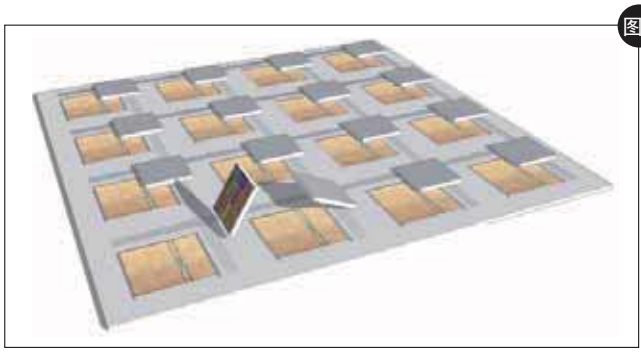
图 1 所示为一个 4×4 的 macrochip 网络。每个 site 均包含一个处理器芯片和一个内存芯片，所占面积为 225mm^2。内存芯片表面朝上地安装在 SOI 衬底的开槽上，一个大小 125mm^2 的处理器芯片面朝下地大部分覆盖在内存芯片上，一部分覆盖在 SOI 衬底上。site 内部使用电连接进行通信，而 site 之间利用硅光技术进行通信。

处理器芯片包含光发射机、波导和接收机，处理器芯片的波导和 SOI 衬底中路由使用的无源光波导通过 OpxC[2] 连接，其方法是将两个芯片面对面地放置，利用对齐的波导光栅实现信号耦合，从而实现波导间信号的传输。SOI 衬底路由层中使用的波导分为两种：本地波导和全局波导。薄 SOI 衬底处的波导用于本地通信，厚 SOI 衬底处的用于全局通信。波导布局时采用两层结构，上下两层波导正交，正交的结构可以有效减少物理交叉并避免信号串扰。

Macrochip 的五种典型网络结构

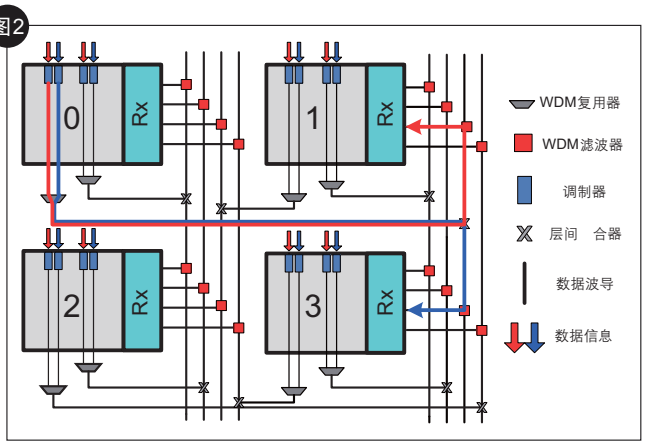
Macrochip 有三种新的网络实现方式：静态波长路由点对点网络（a static wavelength-routed point-to-point network）、两阶段仲裁网络（a-two-phase arbitrated network）和有限连接点对点网络（a limited-connectivity point-to-point network），还可以借鉴两种已有的网络：基于令牌环的交叉开关网络

图1



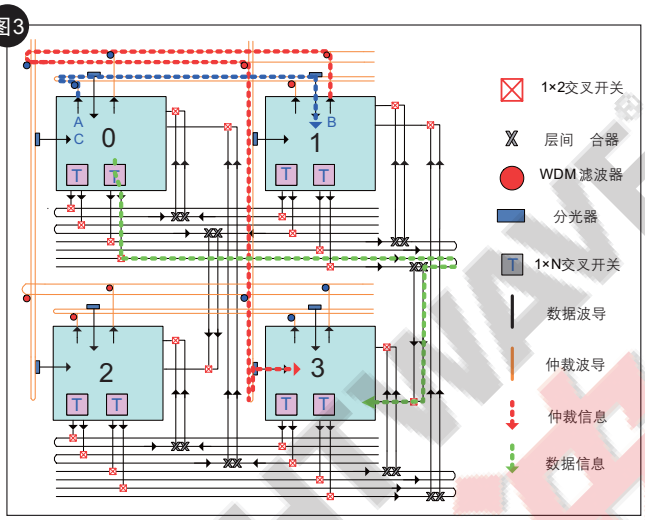
4×4 的 macrochip 网络。

图2



静态波长路由点对点网络。

图3



两阶段仲裁网络。

和电路交换网络。

1、静态波长路由点对点网络。图 2 所示的是一个 2×2 的点对点网络。该网络中的每个 site 与其他任何一个 site 之间都由专用数据通道连接。图中，site 0 分别发送一个数据分组到 site 1 和 site3，虽然分组在传输时使用同一条波导，但是所用波长不同，蓝色代表到达 site3 所使用的波长，红色则代表到达 site1 使用的波长。

网络由水平波导和竖直波导组成，前者在 SOI 路由衬底的底层，后者在顶层，两者用层间的耦合器进行通信。竖直波导中对每列的各个 site 都会分配一个波长。A site 与任何一个 S site 之间通信时，首先找到通往 S site 所在列的波导，再根据 S site 的特定波长，在列中找到 S site 对应的波长信号。

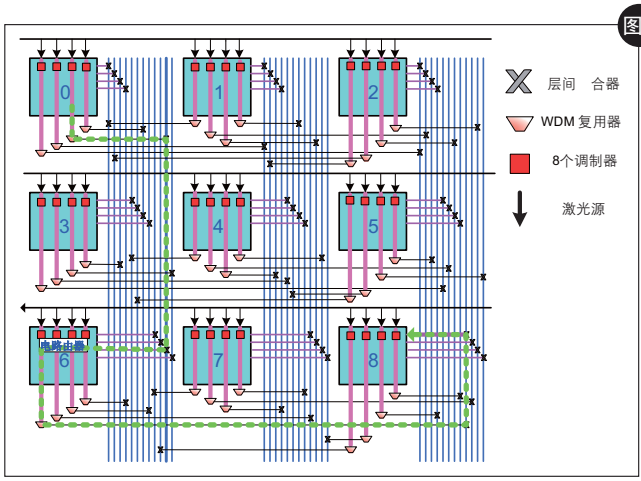
虽然在电连接网络中，增加一个节点就意味着增加二次方的连线，但是由于 WDM 的使用，光互连网络中每条波导可以复用多个波长，从而有效降低波导所占面积和路由复杂度。由于该设计是一个全连接网络，因而，静态路由点对点网络结构没有交换和仲裁的开销，但是其带宽比较低，而且 site 到 site 的数据通道较窄。

2、两阶段仲裁网络。为解决全连接带来的低带宽等问题，进而提出了两阶段仲裁网络。该网络的 site 之间通过共享数据通道来提高带宽（图 3）。每行的各个 site 与同一个 site 通信时，共享一条光数据通道。该网络的波导包括图 3 中红色代表的仲裁波导和黑色代表的数据波导。与点对点网络一样，水平波导和竖直波导之间也是利用层间耦合器进行信号的传输。每个 site 对每列都会分配一个图 3 中的“T”形的树状 1×N 的宽带交叉开关，用以选择对应于该列的目的 site。

当任意 site A 给任意一个 site B 发送数据时，先要在 site 所在的仲裁域解决竞争问题，然后通过控制操作为数据传输设置适当的交叉开关。仲裁网络由在每行负责解决竞争问题的请求波导和在每列负责设置合适的目的交叉开关的通知波导组成。每个 site 发送的信息在请求波导和通知波导中传输时，首先进行波长预分配。例如在图 3 中，site 0 与 site 3 通信时，过程如下：首先 site 0 发送一个仲裁请求，即图中蓝线所示，然后仲裁网络为这个请求分配一个仲裁槽，site 1 沿着列方向的红线发送交叉开关请求，site 3 设置输入交叉开关，site0 也要设置自己的“T”型宽带交叉开关，选择正确的输出端口。最后 site 0 在数据波导中发送信息到 site 3。所谓的两阶段就是：1. 仲裁阶段；2. 设置交换通路并传输数据。

3、有限连接点对点网络。图 4 所示为一个 3×3 的有限连接点对点网络，其拓扑结构和点对点网络的很相似。每个 site 只和与其同行或是同列的 site 有直接的光连接，

图4



有限连接点对点网络。

和其他行或列的 site 没有直接连接。与之前提到的点对点网络的主要区别在于 site 内部使用电路由，故为了解决处于不同行与列的 site 之间的通信问题，需要给每个 site 增加两个电路由器，一个负责把数据分组从同列转发到同行方向上，另一个负责把同行的数据分组转发到同列方向。不同行但同列的 site 间通信时，首先要把分组数据发送到与目的 site 同行或同列的 site，然后在中间 site 进行转发，先将光信号转化为电信号，后将其发送到正确的输出端口，再转化为光信号，图 4 所示的是 site 0 与 site 8 通信的过程。

4、基于令牌环的交叉开关网络。Corona 是一个基于令牌环仲裁的光交叉开关网络，其环形拓扑不会引起波导交叉。Corona 结构中的各 site 或是 cluster 都有一个专用波导束，由向其发送数据的所有 site 或是 cluster 所共享，故进入共享波导束的信号需要通过使用令牌环进行仲裁。如果一个 site 要发送数据到目的 site，需将其在目的 site 令牌通道的接收机调节到合适的波长来吸取令牌。一旦数据传输结束，发送数据的 site 向令牌总线再次注入一个光脉冲，从而释放令牌。

5、电路交换网络。电路交换网络中传输信号时，首先需要建立传输路径，即设置源 site 到达目的 site 沿途的光交叉开关，路径建立完毕就可传输数据，传输结束再拆链。这种网络一般是在电 torus 网络附加光 torus 网络，其中，电网络通常为低带宽的分组交换网络，负责建立从源到目的的光传输路径，而光网络用于实现无阻塞的数据传输。电网络中的每个交换节点都控制着一个与其自身连接的 4×4 的光交叉开关。建立光路径时，源节点发送一个建链分组，该分组经由源 site 到达目的 site，在每个中间交换点，路由器设置对应的光交叉开关，并为数据分组预留到达目的 site 的路由。通信结束后，拆除先前建立的路径。电网络中 site 之间的互连需要很多长的电连线，这会增加设计难度，因而，macrochip 在设计中用一个低带宽的光网络代替相应的电网络，以完成路径建立的功能。

总结

随着集成电路技术和光器件的飞速发展，光互连技术以其高带宽、低能耗、低电磁干扰的突出优势，必然会取代传统电连接成为 NoC 的主流互连方式。Macrochip 的点对点网络比其他网络有着十倍以上的功耗有效性，而且网络复杂性最低，这是 macrochip 未来的主要方向。LWC

注：本文得到基金项目的资助：国家自然科学基金（No.60803038、No.60725415），国家重点实验室专项基金（No.ISN090306）

参考文献：

1. Pranay Koka, Michael O. McCracken, Herb Schwetman,et al. Silicon-photonic network architectures for scalable, power-efficient multi-chip systems. Computer Architecture News, v 38, n 3, p 117-28, June 2010.
2. Zheng Xuezhe, P. Koka, H. Schwetman, et al. Silicon photonic WDM point-to-point network for multi-chip processor interconnects. Group IV Photonics, 2008 5th IEEE International Conference on. 2008.

上接第17页

尽管如此，智能电网技术的不断发展将会在未来打破这个平衡。Janson 说："仅仅坐在这里读电表显然需要的码率较低。"随着电网机构需要汇聚数以千计的自动电表，加上整个网络的电力状态单元和监控单元，以及有合适的冗余备份这些链路，他说："很快地，我们将会看到和传统电信类似的情形。"

电网所需的"合适的冗余"也和传统电信网有所不同。"在电信环境下只需要双重保护，"Johnson 解释到，"但在电网环境中，必须设计三重甚至四重保护以应对同时失效。"

电网市场还有特殊的标准和认证流程。在美国，这些标准或者来自联邦能源监管委员会（FERC），或者来自北美电力可靠性公司（NERC），以及州或本地公共设施监管委员会。国际上，国际电工委员会（IEC）和 IEEE 扮演着重要角色。

然而对于大部分公司来说，给电信运营商开发的设备都能很好地为电网应用服务——只要它有足够的环境容忍度，可根据相关认证做出一些调整。**Alcate-Lucent** 和 **Ciena** 都沿用了这个模式。当然，FTTH 系统商可以给那些想将宽带业务当作不动产摊销的电网提供更广泛的平台。

JDSU 测试设备市场战略总监 Jon Bechman 总结，即使没有 FTTH 的参与，"光通信和光纤在电网的通信架构和基础设施的整体现代化进程中也将扮演重要角色"，很多公司都等待着帮助。LWC