

The BRASICA Chip Set Architecture

Foundation for a New Generation of High-Performance Ethernet and Fast Ethernet Switches



The BRASICA Chip Set Architecture

Foundation for a New Generation of High-Performance Ethernet and Fast Ethernet Switches

By John Hickey and Tim Jalland

3Com's state-of-the-art BRASICATM *chip, at* the heart of the LinkSwitch[™] 1000 and 3000 switches, delivers wire speed switching, optimized for workgroup environments at very low cost. This article is intended for the network manager who is interested in learning more about the advanced system architecture of the BRASICA switching ASIC chip set. The article describes the chip set and bus architecture—including the ATM-ready downlink and software-configurable switching and address learning modes-and presents results from independent performance testing. It also describes the comprehensive set of advanced switching features built into the switching ASIC, such as the sophisticated distributed buffering, Intelligent Flow Management, and multimedia congestion control capabilities; and value-added virtual LAN and RMON network management functionality.

3Com's Leadership in Switching ASIC Design Workgroup switches are special-purpose devices designed and tuned to address LAN performance problems resulting from bandwidth shortages and network bottlenecks in client-server environments. Switches solve congestion problems by switching network traffic to target ports. Providing dedicated bandwidth to each port adds bandwidth to the network infrastructure with minimal risk and no disruption to the cabling or the desktop. Switches also provide high-speed connections to servers, leveraging server bandwidth and equipment investment. The use of applicationspecific integrated circuit (ASIC) technology allows 3Com switches to simultaneously forward packets across all ports at much higher speeds, and at a lower cost per port than for traditional bridges or microprocessor-based

products. Prices start in the range of high-end repeating hubs, which brings switching within the budgets of LAN environments of any size.

ASIC technology has made a significant impact on the networking industry, especially in the area of high-speed LAN switches. With this technology, engineers can embed high levels of functionality, such as switching, bridging, and network translation, onto specialized microchips. This migration of functionality onto integrated circuits, although time-consuming and capital-intensive to implement, produces tremendous benefits in product cost, performance, and reliability. For the network manager, investing in ASIC-based switching technologies yields much higher performance with a much lower price point compared to processor-based solutions. The result is that network managers can have the performance and functionality of switching at the price of traditional repeating technology.

3Com embarked on ASIC development in early 1993, taking advantage of its expertise more than 150 person-years of switching experience—to migrate field-proven hardware and software designs onto silicon. 3Com's ASIC development program has produced several switching ASICs (Table 1), each chip designed for a particular switching environment. The BRASICA chip set in the LinkSwitch 1000 and 3000 switches is optimized for Ethernet and Fast Ethernet switching environments. More than 180 person-months and an \$8 million investment in research and development were devoted to BRASICA's chip set design.

Table 1. 3Com's Switching ASICs

ASIC	LAN Environment
BRASICA	10 Mbps Ethernet and 100 Mbps Fast Ethernet switching
ISE	10 Mbps Ethernet and 100 Mbps FDDI switching
TRSE	4 and 16 Mbps Token Ring switching
ZipChip™	10 Mbps Ethernet and 155 Mbps ATM switching

John Hickey is a principal ASIC design engineer for 3Com's premises distribution division. Prior to joining 3Com in 1993, John spent seven years working for Digital Equipment Corporation's networks and communications division, designing routing and terminal server products.

John holds a B.S and M.S. in engineering from Dublin City University.

Tim Jalland is the product line manager for 3Com's LinkSwitch workgroup switches. Tim has over 15 years of experience in data networking. He joined 3Com in 1992 in the BICC Data Networks acquisition, where he was an engineering manager in network management software development.

Prior to joining BICC, Tim was a WAN product engineering manager for Cray Communications in X.25 and packet switching software development. Tim holds a B.S. in physics from London University and an M.B.A. from the Open University.

3Com's LinkSwitch Ethernet and Fast Ethernet Workgroup Switches

The LinkSwitch 1000 10/100 Mbps switch is designed to connect 10 Mbps desktops or workgroup repeaters onto high-speed server or backbone links. It is available with 12 or 24 switched 10 Mbps Ethernet ports, and includes a built-in 100BASE-TX port.

The LinkSwitch 3000 100 Mbps switch is optimized for Fast Ethernet backbone and workgroup aggregation, connecting 10/100 switches, Fast Ethernet repeaters, and highspeed server links. It is available with five SC fiber or eight UTP Fast Ethernet ports.

All LinkSwitch 1000 and 3000 switches offer a high-speed backbone slot for a Fast Ethernet

fiber (100BASE-FX) or copper twisted-pair (100BASE-TX) plug-in module. An ATM plug-in module for connecting into ATM infrastructures will be available with a software upgrade in the second half of 1996.

The switches come complete with built-in PACE[™] multimedia capability and comprehensive RMON network management. They are also members of 3Com's integrated SuperStack[™] system, which includes hubs, routers, remote access servers, and other devices. SuperStack devices have a common footprint and features such as optional redundant power and can be managed with 3Com's Transcend[®] integrated network management software.

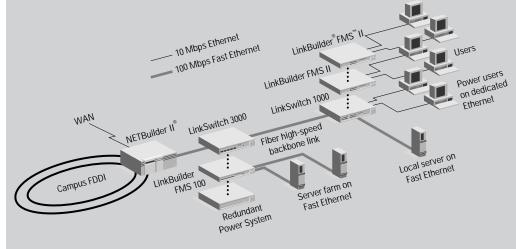


Figure 1. LinkSwitch 1000 and 3000 Workgroup Switch Configurations

BRASICA Chip Set Architecture

ASICs are used as scalable building blocks for designing integrated switching capabilities into the switch. BRASICA's built-in capabilities offload processing tasks, resulting in wirespeed performance with low latency and zero packet loss at a very low cost per port. The BRASICA chip set architecture features, listed below, are described in detail in this article:

- ATM-ready downlink capability
- · Software-configurable switching modes
- · Address learning and forwarding modes
- Distributed buffering and Intelligent Flow Management active congestion control
- PACE-based multimedia networking capability

- Virtual LAN (VLAN) capability
- Comprehensive RMON management

The BRASICA chip set (Figure 2) consists of three ASICs connected by a highspeed data D-bus, capable of 800 Mbps throughput, and a 125-Mbps control and status bus called STATBus:

 BRASICA/CNTRL (control) chip. With its associated switching database, this chip is the heart of the chip set. It controls the highspeed switching of data and information packets within the switch. It also controls PACE-based multimedia class of service management, VLAN, and RMON support.

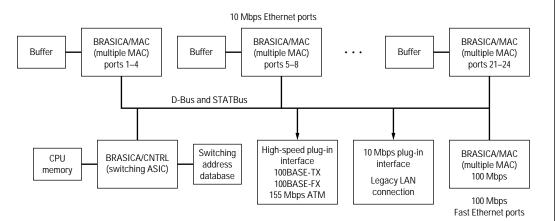


Figure 2. The BRASICA Chip Set and Bus Architecture

• BRASICA/MAC (media access control) chips. Control the wire-speed transfer of data between the LAN connections and the switch memory, and control Intelligent Flow Management within the switch. They also implement the PACE Interactive Access protocol that manages latency-sen-

sitive multimedia traffic. Each BRASICA/MAC chip supports four 10 Mbps ports or one 100 Mbps independent port. Each port contains an 802.3 MAC controller that has the flexibility to be configured in either 10 or 100 Mbps Ethernet

mode. The controller also manages port buffering and collects per-port RMON statistics for device monitoring and troubleshooting.

• *The BRASICA/PHY (physical) chip.* This chip (not shown) implements the physical network interface to the 10 Mbps Ethernet ports, including encoding/decoding, clock recovery, and link detection and regeneration. The BRASICA/PHY chip is the companion interface to the BRASICA/MAC chip; each chip supports four independent 10 Mbps Ethernet ports.

High-Speed and ATM-Ready Downlink Capability High-speed capability is integrated into the BRASICA chip set architecture. A high-speed plug-in module slot off the D-bus provides switched Fast Ethernet downlink capability for either fiber (100BASE-FX) or copper twistedpair (100BASE-TX) wiring. This high-speed downlink port is also ATM-ready, providing a low-cost upgrade for scaling performance. The ATM plug-in module, available in the second half of 1996, supports a single 155 Mbps

The high-speed downlink port is ATM-ready, providing a lowcost upgrade for scaling performance. OC-3c connection for connecting to any standardsbased ATM infrastructure. Onboard LAN Emulation software ensures full integration with existing legacy LAN environments and access to ATM-based server resources.

BRASICA's Built-In Flexible Switching Modes The BRASICA chip set can operate in one of three software-configurable switching modes to accommodate different workgroup switching environments (see Table 2 on page 4 for performance information). The switching mode selected determines the BRASICA chip set packet forwarding latency (defined as first bit in to first bit out) and error-handling characteristics. In all cases, switching performance is limited by the packet transmission rate of the network itself, not the packet forwarding rate of the BRASICA chip set.

Fast-forward has the lowest latency; the packet is forwarded after destination and source addresses are received. Store-andforward has the best error protection; the entire packet is received and checked before it is for-

Abbreviations and Acronyms

ASIC Application-specific integrated circuit

ATM Asynchronous Transfer Mode

COS Class of service

IFM Intelligent Flow Management

ISE Intelligent switching engine

MAC Media access control

PACE Priority Access Control Enabled

RMON Remote Monitoring

TRSE Token Ring Switching Engine VLAN Virtual local area network

3

Table 2. BRASICA's Flexible Switching Modes			
Switching Mode	Description	Latency	
Cut-through fast-forward	Highest performance, lowest latency switching mode.	40 microseconds	
Cut-through fragment-free	Combines the performance benefits of cut-through switching with the error protection of store-and-forward. Cuts through after 64 bytes received.	64 microseconds	
Store-and-forward	Stores and checks packet integrity before forwarding.	8 microseconds (last bit in, first bit out; add packet size for equivalent cut- through latency)	

warded, allowing all runts and error packets to be filtered. Error protection comes at the price of larger latency, however, since the whole packet must be buffered before it can start to be transmitted. Fragment-free is a compromise between the performance of fast-forward and the error protection of store-and-forward; 64 bytes are received before forwarding to filter out runts and still maintain low latency.

The switching mode performance versus error packet filtering trade-off depends on user requirements. Applications such as Novell NetWare[®] 3.11 (which has a nonwindowed protocol), Microsoft Excel[™], and Lotus Notes[®] benefit from fast-forward mode due to the reduced transfer time. Networks with a large number of errors (error packets > 1% of total packets) benefit from store-and-forward mode. If the majority of error packets are runts, fragment-free mode is the best choice.

BRASICA Chip Set Performance Results The BRASICA chip set architecture optimizes Ethernet and Fast Ethernet switching, enabling the LinkSwitch 1000 and 3000 switches to forward packets at full wire speed with no packet loss. Actual performance has been rigorously tested in independent testing facilities.

Independent performance testing by the Tolly Group measured packet throughput between Novell NetWare clients and servers using Novell's Perform3[™] performance benchmarking utility. The two tests were designed to emulate a typical real-life client-server environment (Figure 3).

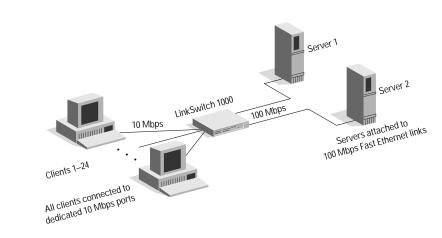


Figure 3. Switching Performance Testing Configuration

The LinkSwitch 1000 is connected to PCI bus, 16 MB RAM Dell Dimension XPS 90 PC servers running Novell NetWare 4.1. Clients are a mix of Pentium[®]- and 486-based PCs, with either EISA or PCI buses and 4 or 8 MB RAM. All computers have 3Com Fast EtherLink[®] Parallel Tasking[®] EISA or PCI 10/100BASE-T network adapters.

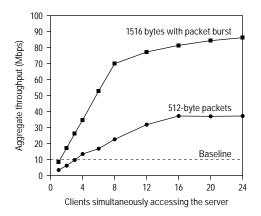


Figure 4. Full Load Performance with Single Server Access

The LinkSwitch 1000 is configured to switch 512-byte packets and 1516-byte packets with Novell's packet burst feature enabled. Up to 24 clients simultaneously accessed a high-speed server through the switch.

Both performance test results show dramatic improvement over traditional 10 Mbps Ethernet networks (Figures 4 and 5). In the first test, the LinkSwitch achieved 86 Mbps throughput with 24 clients accessing a single high-speed server through the switch. In the second test, the LinkSwitch achieved 150 Mbps throughput with 24 clients accessing two high-speed servers through the switch.

BRASICA's Flexibility for Address Learning and Forwarding Features

The switching database on the BRASICA/ CNTRL chip supports two modes of address learning and forwarding of unknown addresses: 802.1d mode or express switching mode.

802.1d, the standard for transparent bridging used in the LinkSwitch 3000, is appropriate for backbone and workgroup aggregation applications. In these applications, unknown packets are flooded to all backbone and server ports.

Express switching, used in the LinkSwitch 1000, is optimized for workgroup and clientserver environments. In this mode, traffic switches primarily between 10 Mbps clients and high-speed backbone/server links. Unknown packets are filtered and flooded only to the downlink port (any port can be designated the downlink). This mode allows the

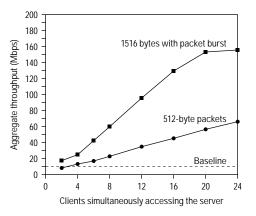


Figure 5. Full Load Performance with Dual Server Access

In this test series, up to 24 clients, 12 per server, simultaneously accessed two high-speed servers through the switch. Packet sizes were identical to the first test set.

switch to connect to backbone networks of unlimited size, reduces the need for a large switch database, and reduces the frequency of flooding within the workgroup, preserving bandwidth for useful data traffic.

BRASICA's Congestion Control Features Congestion control is one of the most important issues in a switching architecture. There are two causes of congestion in a switch:

- Internal resource congestion (memory bandwidth, bus bandwidth, address look-up, and so on) on the switch as traffic load increases
- Over-subscription of a link due to the bursty nature of LAN traffic and the resulting retransmission of dropped packets

Distributed Packet Buffering Controls Internal Congestion

Internal resource congestion is handled in the BRASICA chip set by breaking packet buffering bottlenecks to ensure that there is no performance degradation as network traffic increases. On the LinkSwitch 1000, for example, each group of four 10 Mbps Ethernet ports share access to 128 KB of buffering controlled by a BRASICA/MAC chip. The 100 Mbps Ethernet ports have dedicated access to 128 KB buffering in the LinkSwitch 1000 or 256 KB buffering in the LinkSwitch 3000. The ports are connected by the high-speed 800 Mbps D-bus and 125 Mbps STATBus. The STATBus monitors each port and determines its status (for example, whether address lookup request is required, status of per port buffering, and so on).

Packets are forwarded across the D-bus between ports based on each port's status and the bandwidth allocation algorithm embedded in the BRASICA/CNTRL chip. Because buffering is distributed, packets—even multicast and broadcast packets—cross the bus only once. This differs from switch designs based on a central memory architecture, which require packets to cross the bus into memory at least twice: once on packet receive and once on packet transmit (multicasts usually require attempts to access a single, perhaps lowerspeed destination port, and exceeds the capacity of the available buffers. High-speed ports, new bandwidth-intensive applications, and the changing traffic patterns of clientserver networking are continuing to increase switch traffic load and link over-subscription problems.

Over-subscription of the available buffer typically causes packets to be dropped. Packet loss recovery has a very noticeable effect on the application layer response times seen by users, while packet retransmission worsens the congestion by adding to network traffic.

Adding more buffering has been the traditional solution for reducing the number of dropped packets, but the question is, how

many more transmit cycles, once per outgoing port). This distributed memory architecture prevents bottlenecks even when higherspeed plug-in options are added, such as another 100 Mbps Ethernet port (copper or fiber) or a 155 Mbps ATM port. Scalability is designed in to ensure that there is no compromise on throughput as the potential load increases.

BRASICA's buffer management scheme is also highly efficient and cost

effective. While most switches use fixed 2 KB buffers for packet reception (to ensure that maximum size 1.5 KB Ethernet packets can be accommodated), the BRASICA/MAC chip uses a dynamic buffering algorithm that scales to exactly match the size of the received packet into memory. This scheme guarantees that expensive buffer memory is used efficiently on smaller packets, while enough memory is always available to effectively buffer larger packet sizes.

IFM Controls Link Over-Subscription and Packet Loss

Over-subscription of a link happens when bursty traffic on several different switch ports

Performance tests showed dramatic improvement over 10 Mbps Ethernet networks. At full capacity, the LinkSwitch 1000 achieved 86 and 150 Mbps throughput respectively with one and two high-speed servers. much buffering is sufficient? LAN traffic simulation models have shown that packet loss rates are relatively insensitive to buffer sizes. In other words, holding the switch architecture and other design parameters constant while just increasing buffer size results in a less than proportional reduction in the number of dropped packets, but adds significantly to product cost.

A transmission delay on a congested 10 Mbps

port can reach hundreds of milliseconds. In that same time, a 100 Mbps port can receive approximately 10,000 packets—one every 6.7 microseconds.

3Com's patented Intelligent Flow Management (IFM), designed into the BRASICA chip, is an active congestion control mechanism that ensures zero packet loss during periods of temporary overload due to link over-subscription. When IFM detects that a port's buffer is almost full, it throttles incoming traffic by sending a "jam signal." End-stations respond by backing off and scheduling packet transmission at a later time. The process is repeated until there is sufficient buffering or until traffic reaches a predefined

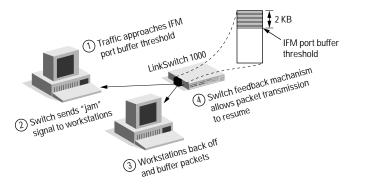


Figure 6. Active Congestion Control Technique with IFM

threshold (this prevents IFM from excessively throttling a segment's traffic).

This ability to tune the level of segment throttling is one of the main differences between IFM and other vendors' congestion control techniques. IFM is fully compatible with existing desktops and cables; no changes to workstations are required. Figure 6 demonstrates how IFM works.

BRASICA's Multimedia Networking Support Under full load, traditional Ethernet networks are subject to high latency (the delay in packet transfer time) and jitter (the variation in packet transfer delay). Real-time multimedia applications, which require regular, predictable data delivery, are especially sensitive to latency and jitter. 3Com's Priority Access Control Enabled (PACE) technology, built into the BRASICA chip, significantly reduces switch latency and enables latency- and jitter-sensitive multimedia traffic to flow through an existing switched Ethernet network along with normal data traffic.

PACE is designed to work on switched desktop links. When PACE is configured on the switch, its Interactive Access protocol bounds access latency from the end-station to the switch to a maximum of 5 milliseconds. The protocol actually allows network managers to program latency on a point-topoint link so that each link can perform at top efficiency, which measurably improves packet transmission and arrival time.

PACE's implicit class of service (COS) programming algorithm, based on MAC addressing, gives multimedia traffic priority access to needed bandwidth. COS differentiates high-priority multimedia packets from data packets and tunnels them through the switch. This tunneling allows high-priority traffic to jump past lower-priority traffic, guaranteeing fast, predictable multimedia traffic transmission.

PACE's patented technology has been endorsed by other network and multimedia equipment vendors, including AMD, Apple, AT&T, Dell, InSoft, Matra, Novell, Oracle, PictureTel, Silicon Graphics, Star Graphics, and Sun Microsystems.

BRASICA's VLAN Capability

Virtual LANs group network components such as workstations and servers into a single virtual network or workgroup, although these components may physically reside on different LAN segments. BRASICA's built-in VLAN capability is implemented as separate broadcast domains. Because only devices within the domain communicate with one another, VLANs provide security and preserve bandwidth within the workgroup. Users assigned to the same broadcast domain communicate at wire speeds with low latencies, and with no routing bottlenecks, at prices much lower than router-based solutions. VLANs also keep network management simple by not requiring separate logical subnets for each switch port.

BRASICA's VLAN feature can be configured in either port or desktop mode. In port mode, the switch is configured to put ports into VLANs. In desktop mode, the switch automatically configures end-stations connecting to its ports to the appropriate VLAN based on the end-stations' MAC addresses. VLAN capability spans multiple switching devices, so changes are reconfigured automatically, or can be handled through 3Com's Transcend graphical drag-and-drop network management software. And again, since VLAN support is built into silicon, there is no degradation in switching performance when the capability is enabled, unlike softwarebased solutions. BRASICA's VLAN capability will be enabled by software upgrade in the second quarter of 1996.

BRASICA's Distributed RMON Capability Switches provide maximum bandwidth within workgroups by segmenting network traffic. However, as a result, network monitoring traffic is stopped at the switch interface, so switched networks begin to lose the ability and visibility to gather valuable diagnostic information. RMON MIB support built into the

BRASICA chip adds RMON processing power to each switch. Because this capability is distributed, the RMON processing power scales as switches are added to the network.

RMON spans a wide range of statistical monitoring, from physical link port utilization and error type statistics to tracking and monitoring host-matrix conversations. This compre-

hensive monitoring provides network managers with valuable network analysis, design, and capacity planning information on even the most complex networks. Because the capability is built into silicon, it provides realtime, wire-speed remote monitoring on all ports with no degradation in switching performance.

Transcend network management software integrates 3Com's built-in features such as SmartAgent[®] gauges, action on event, and auto-calibrate with RMON alarms to automate monitoring and recovery tasks. For example, a per-port broadcast throttling feature works with RMON to cost-effectively monitor and prevent network downtime due to network broadcasts storms. A SmartAgent gauge can be set up on each switching port. Once a predefined broadcast threshold is exceeded, the gauge's action-on-event feature can be sent to automatically disable the affected port for a set duration and alert the central management workstation. When the broadcast storm subsides, the gauge automatically reenables the port. This kind of automatic network monitoring and recovery feature greatly enhances the proactivity and effectiveness of the network management software.

Summary

3Com's BRASICA chip set is optimized for Ethernet–Fast Ethernet workgroup and floor aggregation switching. ASIC-based switches bring the cost per port down to the range of high-end repeating hubs, and within the

RMON capability scales as switches are added to the network because distributed RMON processing power is built into each switch. budgets of small and large LAN environments. Integrating what was formerly software-based functionality directly onto the chip set offloads the processing from the switch's CPU and ensures the highest level of switching performance at all times. The BRASICA chip set is the foundation for 3Com's Ethernet and Fast Ethernet switching

products, including LinkSwitch 1000 and 3000 workgroup switches.

Independent performance testing demonstrates the BRASICA-based LinkSwitch 1000's ability to forward packets at full wire speeds, under increasing traffic load, with zero packet loss. Built-in distributed packet buffering, Intelligent Flow Management, and PACE-based multimedia features effectively eliminate workgroup congestion while reducing the product cost. An ATM-ready downlink port and built-in value-added network management features such as virtual LAN and advanced RMON support combine to lower the cost of product ownership and protect your network investment.



3Com Corporation

P.O. Box 58145 5400 Bayfront Plaza Santa Clara, CA 95052-8145 Phone: 800-NET-3Com or 408-764-5000 Fax: 408-764-5001 World Wide Web: http://www.3com.com

3Com ANZA

ANZA East Phone: 61 2 9937 5000 Fax: 61 2 9956 6247 ANZA West Phone: 61 3 9653 9515 Fax: 61 3 9653 9505

3Com Asia Limited

Taiwan Phone: 886 2 377 5850 Fax: 886 2 377 5860 *Thailand* Phone: 622 231 8151 2 Fax: 622 231 8121

3Com Benelux B.V.

Belgium, Luxembourg Phone: 32 2 716 4880 Fax: 32 2 716 4780 Netherlands Phone: 31 34025 5033 Fax: 31 34025 4630

3Com Canada

3Com European HQ

Phone: 44 1628 897000 Fax: 44 1628 897041 3Com France Phone: 33 1 69 86 68 00 Fax: 33 1 69 07 11 54

3Com GmbH

Germany Phone: 49 89 627320 Fax: 49 89 62732233 *Berlin* Phone: 49 30 3498790 Fax: 49 30 3498799

Poland Phone: 48 22 6451351 Fax: 48 22 6451352 Switzerland Phone: 41 31 9984555 Fax: 41 31 9984550

3Com Ireland

Phone: 353 1 820 7077 Fax: 353 1 820 7107

3Com Japan

Phone: 81 3 3345 7251 Fax: 81 3 3345 7261

3Com Latin America

3Com Northern Latin America Miami, Florida Phone: 305-261-3266 Fax: 305-261-4901 *Venezuela* Phone: 582 261 0710 Fax: 582 261 5257

3Com Mediterraneo

3Com Middle East

Phone: 971 4 349049 Fax: 971 4 349803

3Com Nordic AB

3Com South Africa

Phone: 27 11 803 7404/5 Fax: 27 11 803 7411

3Com UK Ltd.

© 3Com Corporation 1996. All rights reserved. 3Com is a publically owned corporation (NASDAO.COMS). 3Com, EtherLink, LinkBuilder, NETBuilder II, Parallel Tasking, SmartAgent, and Transcend are registered trademarks of 3Com Corporation. BRASICA, FMS, LinkSwitch, PACE, and SuperStack are trademarks of 3Com Corporation. 3Com registered trademarks are registered in the United States and may or may not be registered in other countries. Pentium is a registered trademark of Intel. Excel is a trademark of Microsoft. NetWare is a registered trademark of Novell, and Perform3 is a trademark of Novell. Lotus Notes is a registered trademark of Lotus Development Corporation. Other brand and product names may be trademarks or registered trademarks of their respective owners. All specifications are subject to change without notice.