An Experimental Multi-Gb/s WDM Broadband Network*

Charles Barry, Michael Hickey, Ciro Noronha, Pierluigi Poggiolini and Leonid Kazovsky
Department of Electrical Engineering
Stanford University
Stanford, CA 94305-4055

Abstract

STARNET is an optical broadband network architecture based on a physical passive star topology. Over a single physical network, STARNET offers all users both a moderate-speed packet switched subnetwork and a high-speed WDM circuit switched subnetwork. Based on these two data transport facilities, several topological and protocol solutions are available to the users. As a result, STARNET supports traffic of widely different speed and continuity characteristics. This paper describes an ongoing experimental effort aimed at the realization of a STARNET prototype at the Optical Communication Research Laboratory of Stanford University. The prototype nodes are FDDI and SONET OC-48 data rate compatible. The first four network layers will be fully implemented.

1 Introduction

Future networks are expected to provide a wide variety of services, such as distributed multimedia, file transfer, remote logins, etc. Packet switching and circuit switching have both been considered to provide these services. The prevailing approach is that of packet switching, because packet switching is effective in handling very different data rates and permits efficient processing of bursty data.

Wavelength Division Multiplexing (WDM) passive star solution is a promising option for the physical implementation of broadband optical networks. WDM offers a potentially very large number of independent optical channels which enables *concurrent* access to the network by many users. This results in high overall network throughput. However, realizing a packet mode of transmission in this intrinsically circuit switched environment represents a challenging task. To this end, different media access strategies and node configurations have been devised.

In this paper we describe the implementation of an optical network architecture, STARNET, based on a passive WDM star, designed to take advantage of multiple channels without requiring extreme hardware and protocol complexity. This result is achieved by simultaneously providing two logical subnetworks, a packet switched subnetwork and a circuit switched subnetwork, over the same physical interconnect. The circuit switched subnetwork enables arbitrary subsets of STARNET end-users to set up not only point-to-point links but more advanced multihop logical topologies as well. STARNET was first proposed in [1].

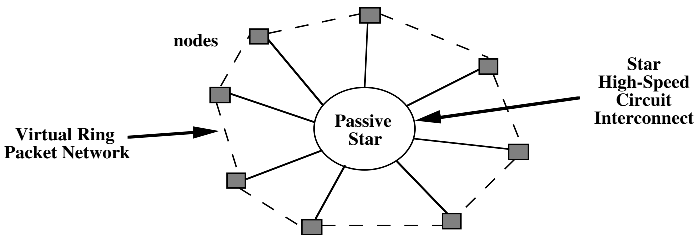

Figure 1: Each STARNET node communicates on two concurrent logical subnetworks over the WDM physical star. One subnetwork carries moderate speed packet switched data; the other carries high speed circuit switched data.

The experimental realization of STARNET is currently in progress. The initial prototype will be a 2-node network, serving high performance workstations. The first four network layers, from the physical layer upward, are being developed.

The paper is organized as follows. In Section 2 the node structure is described and the basic network architecture is introduced. Section 2.1 describes the implementation of the physical layer of the node and the current status of the physical layer implementation is reported. Section 3 describes the approach taken at the realization of the electronic interface between the physical layer and the computer bus. In Section 4 topological and optimization issues are discussed. In Section 5 a simulation of the network is presented.

2 The Basic Node Configuration

Each node of STARNET is optically connected to all the other nodes through a passive optical star, as shown in figure 1.

Each STARNET node is equipped with *one* laser transmitter. In normal operation the laser source is tuned to a unique fixed wavelength so that a *comb* of light carriers is formed. Using a combined multiplexing strategy, the node transmits *two independent data streams*, stream 'P' and stream 'C'. The 'P' stream carries the packet switched traffic; the 'C' stream carries the circuit switched traffic.

Each node has two receivers; one tunable and one fixed. The *tunable* receiver can be tuned to receive any transmitter and decodes the 'C' stream only. The *fixed* receiver is per-

*This work was partially supported by ONR under contract number N00014-91-J-1857 and by NSF under grant number ECS-9111766. C. Noronha is supported by FAPESP under grant 89/1658. C. Barry is supported by Digital Equipment Corporation.

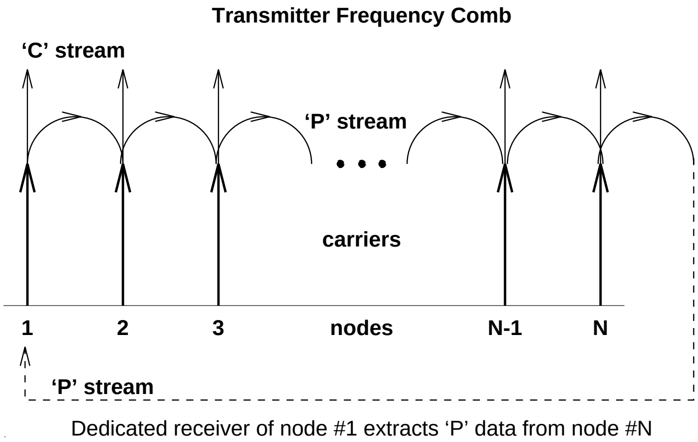

Figure 2: Each transmitter multiplexes the 'C' stream and the 'P' stream onto the outgoing light signal; the 'P' stream is looped from node N to node 1, forming a logical ring.

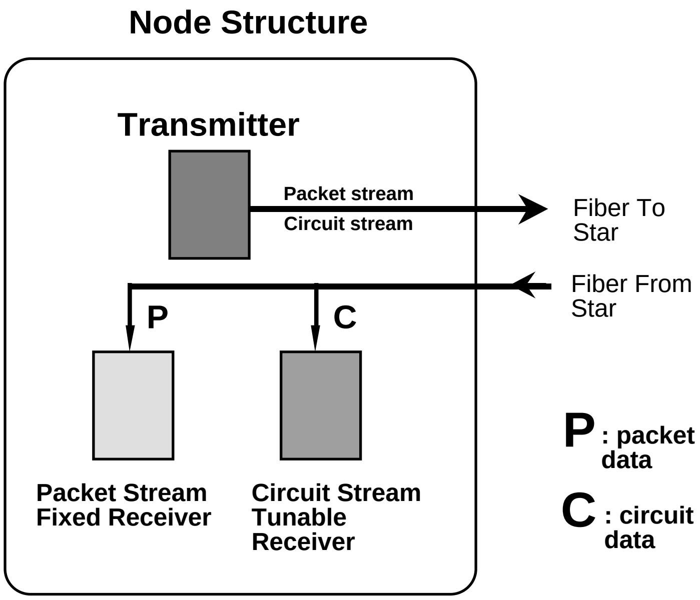

Figure 3: Block diagram of the STARNET transceiver.

manently tuned to the *previous node* in the frequency comb and receives the 'P' stream only. STARNET exploits this feature to form a logical 'chain' along the frequency comb. Equipping the first node of the chain with a receiver that decodes the 'P' stream of the last node in the chain forms a logical unidirectional ring as shown in figure 2.

2.1 Node Implementation

The experimental STARNET prototype will interconnect DECstation™ 5000/240 workstations. In the following, we briefly discuss the experimental implementation of the STARNET transmitter and receiver subsystems. A detailed description can be found in [1].

Fig. 3 shows the block diagram of a STARNET node transceiver. The transmitter uses a 25 mW, 1319 nm Nd:YAG laser. The transmitter multiplexes two independent data streams, the 'C' stream and the 'P' stream on the light carrier. The multiplexing is accomplished with a combined phase/amplitude $LiNbO_3$ optical modulator. The 'C' stream is PSK modulated at up to 2.488 Gb/s (SONET OC-48 rate). Low modulation depth ASK is used to transmit the the 'P' stream at a line rate of 125 Mb/s (provided by commercial FDDI hardware in the workstation).

The tunable 'C' stream receiver is a synchronous PSK heterodyne receiver which receives the 2.488 Gb/s 'C' stream data. The fixed receiver extracts data from the 125 Mbit/s ASK-modulated 'P' stream. The 'P' stream data is then passed to commercial FDDI electronics in the workstation.

The PSK receiver has been demonstrated with pseudo random bit streams at 2.488 Gb/s. The ASK receiver has been demonstrated with FDDI data at 125 Mb/s. Measurements show that the system power budget is 30 dB with both receivers operating simultaneously at better than a 10^{-9} bit error ratio (BER). Detailed results on the combined ASK-PSK modulation technique and its performance are presented in [3].

3 Network Interface Issues

3.1 Workstation/STARNET Interface

The interconnection of two workstations using STARNET prototype optics and electronics is shown in Fig. 4. Each experimental STARNET node is attached to a DECstation 5000/240. Each workstation connects to a STARNET transceiver via electronic hardware modules which are installed in the DECstation's TURBOchannel™ I/O bus. A custom designed electronics module interfaces the workstation to the 2.488 Gbps circuit switched subnetwork. An off-the-shelf FDDI electronics module interfaces to the 125 Mbps packet switched subnetwork.

Real-time voice and video functionality is provided by additional TURBOchannel modules. Moreover, the workstation provides connections to SCSI storage devices, an Ethernet port, etc. Consequently, a rich environment of high and low bandwidth bursty and continuous traffic is available for the study of the performance of the STARNET architecture.

In the following we detail some of the important issues that are addressed by the network to workstation interface.

3.2 High Speed Network Interface

The STARNET high speed subnetwork transceiver provides *full-duplex* communication at a sustained rate of 2.488 Gb/s. Although the TURBOchannel I/O bus has a *half-duplex* burst rate of 800 Mbps, a reasonable estimate of the *sustained full-duplex* I/O bandwidth available to the high speed subnetwork is 200-250 Mb/s since the I/O bandwidth is shared with other devices on the bus (e.g. FDDI, Ethernet, SCSI, etc.). In order to take advantage of the 2.488 Gbps capacity, the high speed interface has been designed to support multihop *packet switching*. In Section 4 we explain how this additional capability is used to fully exploit the high speed link.

Routing of incoming packets is done on a packet by packet basis. The headers of incoming packets are checked if they are destined for the receiving node, in which case they are queued for transfer to the node's local memory over the I/O

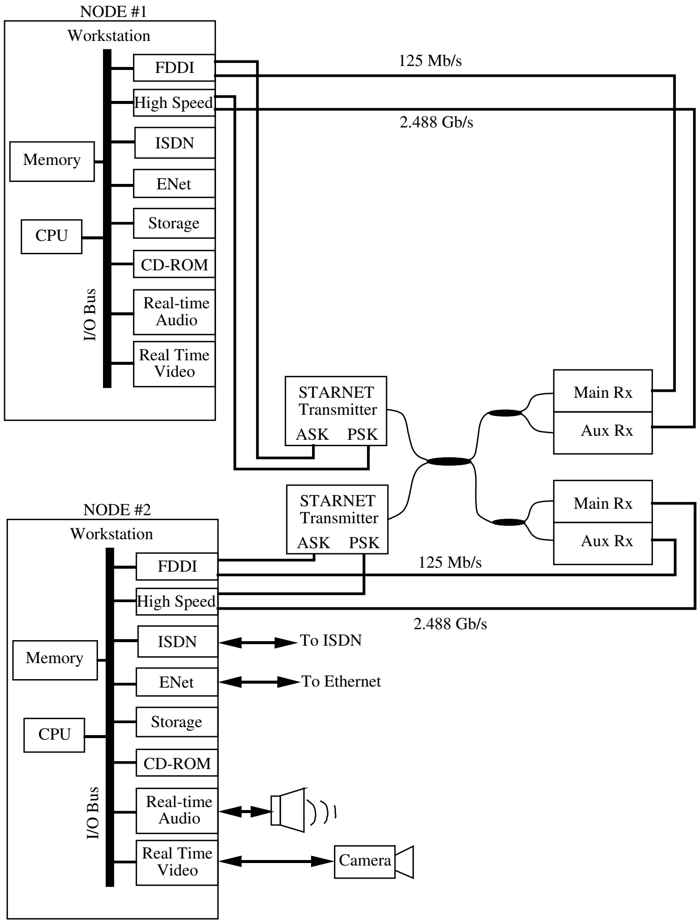

Figure 4: Interconnection of two STARNET nodes.

bus. Packets not destined to the workstation are forwarded to the output link without intervention of the node's CPU. The result is a high speed 2×2 multihop *packet switched* architecture over the circuit switched subnetwork.

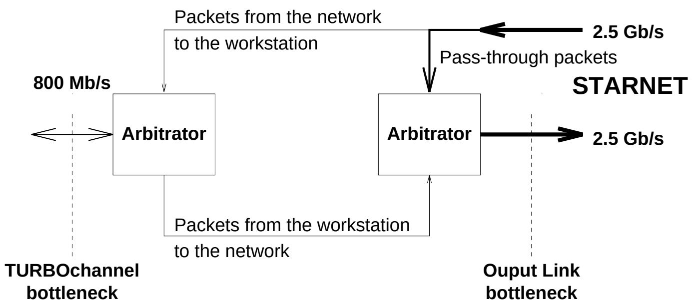

Figure 5: Bottlenecks in the high-speed network interface.

3.3 Channel Bandwidth Allocation

There are two bottlenecks associated with the high-speed network interface: the TURBOchannel and the 2.488 Gb/s output link. The TURBOchannel bandwidth must be shared between the workstation's input and output data streams; the output link must be shared between the workstation's output stream and the pass-through traffic. The allocation

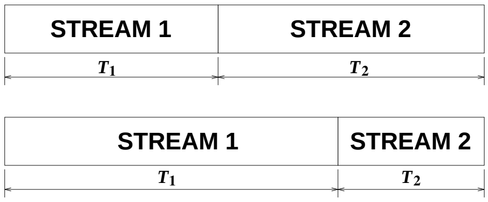

Figure 6: Dynamic bandwidth allocation in the high-speed network interface.

of bandwidth at each bottleneck is done in hardware by two arbitrators, as shown in figure 5. The arbitrators operate as follows: a two-slot cycle is formed, and each of the time slots is allocated to one of the competing streams. By adjusting the time slot sizes, it is possible to control the amount of bandwidth allocated to each stream; this adjustment can be done dynamically by software. The process is illustrated in figure 6. The granularity in the time slot sizes (T_1 and T_2 in figure 6) is one packet, and if one stream is idle, then the whole bandwidth is available to the other stream.

The bandwidth allocation policy described is implemented by hardware in the node interface. Research is still ongoing on the optimal settings of the allocation policy as a function of the traffic requirements.

4 Network Control, Routing and Reconfiguration

Since the high-speed circuit switched subnetwork uses *tunable* receivers, the high speed subnetwork topology is reconfigurable. This feature can be exploited to optimize the overall network throughput.

In its simplest instance, the logical topology of a multihop arrangement is a single ring. Notice however that this ring can be formed using the *tunable* receivers and is completely independent of the ring packet switched subnetwork that utilizes the *fixed* receivers. Depending upon the traffic requirements, more than one high speed ring can be established among *different* subsets of nodes. Furthermore, not *all* the nodes need to be hooked up to one of these multihop virtual subnetworks.

At the transport layer, the model of operation of the network is virtual circuits, which provides reliable end-to-end data streams. Multiple virtual circuits will be multiplexed over each physical link. To avoid the need for resequencing buffers, all packets from a given virtual circuit will follow the same route.

Finally, in a general picture, the multihop virtual subnetworks need not be confined to the use of the high speed links only: they can take advantage of the lower speed packet network to realize more complicated virtual topologies.

STARNET can therefore implement any mesh topology where a ring Hamiltonian Path[1] can be identified. This mesh topology is formed by mapping the Hamiltonian Path onto the packet switched subnetwork, and using the circuit

[1] A *Hamiltonian Path* is a path that visits each node of the network exactly once.

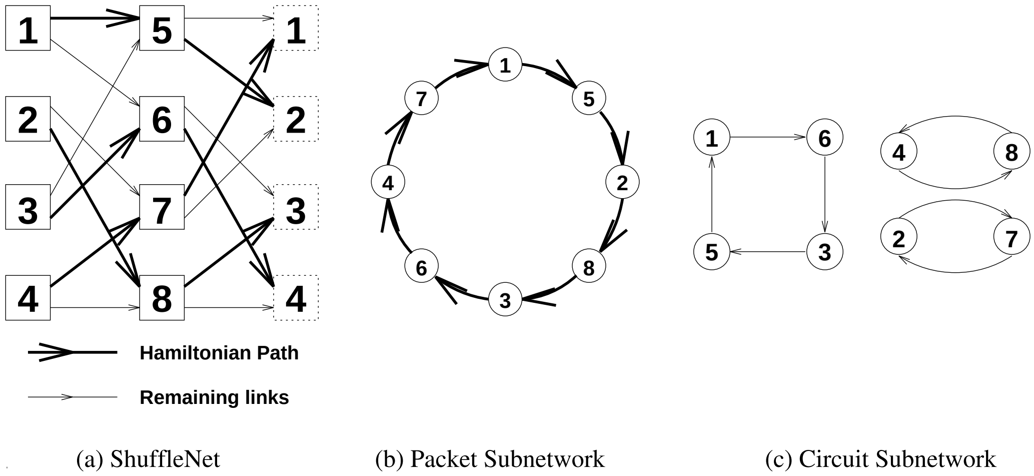

Figure 7: Mapping the ShuffleNet onto STARNET.

switched subnetwork to complete the mesh. Figure 7 shows how to map the ShuffleNet topology onto STARNET.

The challenge is then to design and investigate the proper reconfiguration/routing algorithms that maximize network throughput, given the current traffic requirements. Those algorithms will identify the network topology (i.e., the tuning of the PSK receivers) and the multihop virtual circuit routing. As indicated before, a virtual circuit can use either the lower-speed ring packet network or the circuit interconnection on its trip from the source to the destination.

4.1 The Operation of the WDM Network

The first point to be addressed is how tunable receivers should be used in a network. A tunable receiver can have the following applications:

- switching;
- altering the topology to cope with congestion; and
- altering the topology to cope with failures.

We will concern ourselves here with the first two categories.

Switching: Tunable devices (e.g. lasers) are required to build tunable WDM receivers are not yet readily available that can be tuned in time intervals comparable with the packet transmission time. For instance, the (transmission time for a 53-byte ATM packet over a SONET OC-48 (2.488 Gbps) link is just 170ns. The problem is exacerbated by the need to establish phaselock and synchronization of the serial bit stream. Therefore we will consider virtual circuit switching whereby the tuning times for the reconfigurable links are comparable to the transmission time for the virtual circuit connection (milliseconds to seconds).

Congestion: In STARNET, connections assume the form of a virtual circuit which are established between the source and the destination; further, all the packets belonging to one connection use the same path. This is important for isochronous traffic, like voice and video. In such a network, congestion assumes the following forms:

- loss of packets because one of the existing connections increases its rate; or
- blocked connections because a node along the virtual path cannot accept them.

The first form of congestion is dealt with by call policing (i.e., ensuring that a connection, once established, conforms to an agreed parameter profile). In the STARNET prototype, we will not address the problem of call policing.

The second form of congestion can be handled efficiently by a reconfigurable network. By reconfiguring the interconnection pattern between the nodes, new paths can be created and the new call accepted. It is necessary to specify how congestion is to be detected, and how the topology is changed, without the need for a central controller. In other words, a distributed reconfiguration algorithm must designed to realize the potential gains that may arise from the possibility to reconfigure the network.

The following two sections detail the algorithms we used in the simulations to accept connections (allocate bandwidth), and (re)configure the network to route the connection.

4.2 The Bandwidth Allocation Algorithm

The bandwidth allocation algorithm decides whether or not a given call can be accepted on a link. We assume that the packet arrivals to the virtual circuit (VC) form a Poisson process with rate λ. We also assume that the arrival process to a buffer is a Poisson process with rate $N\lambda$, where N is the number of VCs using the link associated to that buffer. We model the buffer as an $M/M/1/K$ queue, and limit the input rate $N\lambda$ so that the maximum acceptable loss probability is not exceeded.

4.3 The Routing/Reconfiguration Algorithm

The routing algorithm that we have chosen in the simulations is as follows. If a tunable receiver is *idle*, and if tuning the receiver will either enable acceptance of a new call or reduce the number of hops taken by the virtual circuit from source to destination, then the network is reconfigured

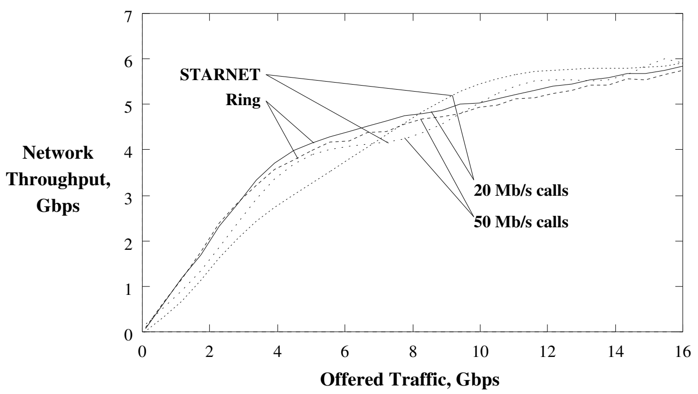

Figure 8: Comparison of the throughput of STARNET and a ring network under virtual circuit operation.

and the virtual circuit is established. Otherwise, the call is routed over the existing links or dropped.

Note that this algorithm that we use is suboptimal in the sense that the routing is evaluated on a connection by connection basis rather than on an overall network average traffic matrix.

5 Network Simulation

In the simulations we wish to characterize the total throughput of a moderate size (8 node) STARNET network versus the offered traffic (connection request rate) in the cases of moderate and high bandwidth connections. In addition we compare the performance of STARNET to a ring network operating at 2.488 Gbps in virtual circuit mode[2]. The assumptions used in the simulations are as follows: The model of operation is virtual circuit (VC), and all the packets belonging to a VC use the same route; each station maintains a table of all the VC numbers that it is carrying, in order to find the correct output link for each packet it receives; the number of packets transmitted on a VC has a geometric distribution, and the last packet transmitted frees the links; the packet arrival process is Poisson; the packet arrival time is exponentially distributed; the arrival of connection requests to a node is a Poisson process; each node has two output buffers of size 128, one for each of the transmit streams; the target packet loss probability is set to 10^{-9}; and a new connection is accepted only if a path with enough capacity can be found to the destination. In the simulations we assumed uniform addressing of connection destinations, and fixed the connection duration to 10 packets per connection. We ran separate simulations for average connection bandwidths of 20 and 50 Mbps.

5.1 Simulation Results

Simulation results are shown in Fig. 8. The most important feature to note is that the performance of STARNET using the proposed routing/reconfiguration algorithm is approximately the same as that for a ring network. This indicates that a routing/reconfiguration policy based on locally optimizing on a call by call basis is inefficient. It is likely that the throughput could be improved by using an algorithm that estimates the actual traffic parameters over a larger time scale and uses these global estimates to block reconfiguration of the network for calls that would result in overall sub-optimal topologies; these calls would be forced to take multihop routes without reconfiguration of the network. An additional is that the overall network throughput is only weakly dependent upon the connection bandwidth, as one would expect.

[2]Note that this is not the same as a token ring because we allow multiple concurrent transmissions over different links of the ring.

6 Conclusions

STARNET is an optical broadband LAN based on a physical passive star topology. STARNET offers all users two data transport facilities: a moderate speed (125 Mb/s) packet network and a high speed (2.488 Gb/s per node) WDM circuit interconnect, both running on the same physical network. The moderate-speed packet switched subnetwork is used for control, optimization, maintenance and exchange of low and medium speed data. The high-speed circuit switched subnetwork provides high speed links that can be used for node-to-node connections, broadcasting, or to build broadband multihop topologies among arbitrary subsets of nodes. In addition, the use of the packet switched subnetwork in the multihop topology allows reconfiguration of the high speed network.

A STARNET prototype is being built at the Optical Communication Research Laboratory of Stanford University. The network operates at FDDI and SONET OC–48 data rates (the packet subnetwork and the circuit subnetwork, respectively). The first four network layers will be built, including both (optical and electronic) hardware and software.

Computer simulations of network performance indicate that a routing/reconfiguration algorithm based on global knowledge of the traffic is needed to efficiently control the network.

References

[1] Poggiolini, P., T., Kazovsky, L., G.: "STARNET: an integrated services broadband optical network with physical star topology", *SPIE meeting, Conference 1579 'Advanced Fiber Communication Technologies'*, Proc. SPIE 1579, pp. 14-29, Boston, September 1991.

[2] L.G. Kazovsky et al., "WDM Local Area Networks", *IEEE LTS*, May 1992, pp 8-15.

[3] M. Hickey, C. Barry, C. Noronha, and L. Kazovsky, "An experimental PSK/ASK transceiver for a multi-Gb/s coherent WDM local area network," *OFC '93*, paper ThM4, San Jose, CA, Feb. 1993.

[4] Kleinrock:'Queueing Systems: Volume 1: Theory'; John Wiley & Sons, New York, 1975.