to minimize multi-hop communications, so that most messages traverse only a single link between "nearest neighbor" nodes. Hop latency and congestion are reduced at the expense of more programming effort and inefficient memory use (read-only data replicated at every node to avoid communication). These mitigating factors will have less impact with advances in processor power, memory access (SDRAM, RAMBUS, use of fast cache[6,7]), and shifts to shared memory multiprocessing, as indicated above. For these reasons, it is therefore important to develop alternative interconnection media for high-performance multiprocessor and embedded systems.

Fiber optics is an attractive interconnect alternative with high transmission capacity over long distances, light weight, and no EMI and ground bounce effects. Most significant is the absence of capacitance and transmission line effects which limit electronic fanout. Fanout is limited simply by the optical power required to achieve error-free transmission. Passive optical star couplers provide broadcast capability analogous to electronic data busses, and could support thousands of nodes for typical 1 GHz transceivers.[8] In addition, wavelength division multiplexing (WDM) creates multiple logical busses, one per system wavelength, for a single passive star.[8] Messages can be simultaneously transmitted on each wavelength without interference. Thus, an optical WDM bus provides high bandwidth, one-hop communication among many nodes (high fanout), and high concurrency (simultaneous transactions on different WDM channels). Fundamentally, WDM optics enables high connectivity routers by source routing, which decouples node fanout (degree) from physical pinout restrictions. A single optical cable can connect a node to multiple destinations, which are chosen by transmitter wavelength selection. Fanout occurs elsewhere in the network by passive optic components (star couplers, wavelength filters). Thus, optics decouples fanout from the pinout and wire density limitations[1] of electronic VLSI technology. In the bus topology, the number of logical busses can be increased without increasing the physical pinout of the optoelectronic transceivers. If multiple wavelengths are received at a node, however, local electronic pinout and wiring density can become an issue.

Fig. 1 shows an example WDM bus using wavelength tunable transmitters and fixed wavelength receivers. The advantages of this configuration are fast WDM tuning (transmitter tuning is currently faster than for receivers) and no requirement for input transmitters to track WDM filter tuning changes.

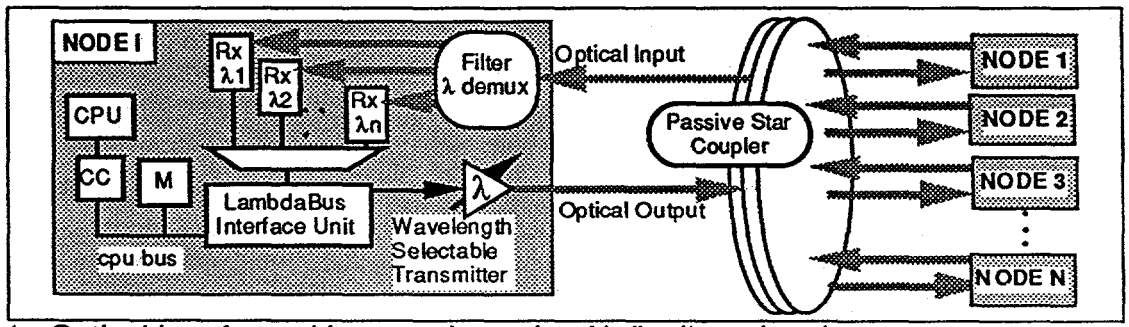
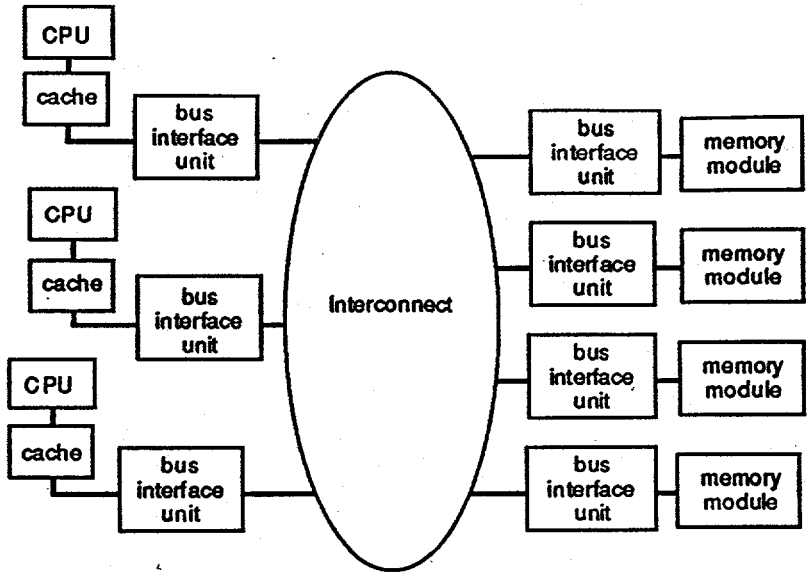
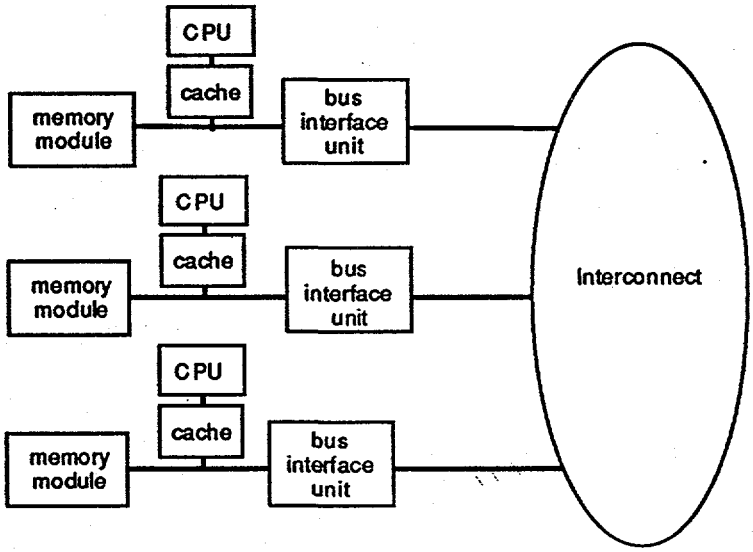


Fig. 1: Optical bus for multiprocessing using N distributed nodes.

The performance, cost, and complexity of optical bus implementations vary dramatically with the required bandwidth per wavelength $B\lambda$, number of system wavelengths Λ, number of wavelengths detected at each node R_Λ, and method for



(a)



(b)

Fig. 2: Simulated system in dance hall configuration (a), and SMP of interest with main memory distributed among processing nodes (b). Systems with only three processing elements are shown.

Optical Interconnect: The optical interconnect is operated as a set of Λ independent, parallel busses with one bus per system wavelength. Every node has the capability of transmitting on all system wavelengths, but transmits a message on only one wavelength at a time. Wavelength-selectable optoelectronic transmitters of this type have been demonstrated with tuning times of a few nsec.[8,13] Messages can be simultaneously transmitted on different wavelengths without interference, so that the maximum concurrency of the interconnect is Λ. Outgoing messages from a given node are queued at the bus interface unit, and are sequentially transmitted on a first in, first

Acknowledgments

We are grateful to E.D. Brooks III for useful discussions and for assistance with the Cerberus simulator, and to R.J. Sherwood for assistance in coding and porting sections of the bus simulator.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

[1] The observing platform generates a sequence of "data cubes", or three-dimensional data sets comprised of layers of two-dimensional image pixels, each layer representing a different spectral window. If each image layer contains 10^3x10^3 pixels each 12 bits deep and there are 10^3 layers (one for each spectral bin), each cube will contain 1.2x10^{10} bits of data in every hyperspectral snap-shot of a finite area of the earth's surface. An imaging rate of 1 data cube/ sec. yield s a sensor data flow of 12 Gb/s.
[2] Amdahl/Case rule of thumb, in J.L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach* (Morgan Kaufman, San Mateo 1990), pg. 17
[3] S. Zai and C. Ranson, "Summit: a 1 GByte/sec Multiprocessor System Bus", and M. Galles, "The Challenge Interconnect: Design of a 1.2 GBs Coherent Multiprocessor Bus", both in *Proc. Hot Interconnects* (IEEE, August 1993).
[4] W.J. Dally, *IEEE Trans. Comput.* C-39, 775 (1990).
[5] D.E. Lenoski and W.-D. Weber, Scalable shared-memory Multiprocessing (Morgan Kaufmann, San Francisco, 1995).
[6] D. Barringer, R. Leong, and P. Novell, "Modernize your Memory Subsystem Design", *Electronic Design*, pp. 83-92 (February 5, 1996).
See also D. Bursky, "Memories hit new highs and clocks run jitter free", *Electronic Design*, pp. 79-93 (February 19, 1996).
[7] J. Handy, The Cache Memory Book (Academic, New York, 1993).
[8] C.A. Brackett, "Dense wavelength division multiplexing networks: principles and applications," *IEEE J. Selec. Area Commun.* SAC-8, 948 (1990).
[9] E.D. Brooks III, T.S. Axelrod, and G.A. Darmohray, "The Cerberus Multiprocessor Simulator", in *Parallel Processing for Scientific Computing* (G. Rodrigue ed., SIAM), p. 384 (1989).
[10] J.E. Hoag, "The cache group scheme for hardware controlled cache coherence", M.S. Thesis (University of California-Davis,; March 1991), UCRL-LR-106975.
[11] "National Technology Roadmap for Semiconductors", Semiconductor Industry Association (1994).
[12] L.M. Censier and P. Feautrier, "A new solution to coherence problems in multicache systems," *IEEE Trans. Comput.* C-27, 1122 (1978).
[13] H. Kobrinski et al., "Fast wavelength switching of laser transmitters and amplifiers", IEEE J. Selec. Area Commun. SAC-8, 1190 (1990).