



ELSEVIER

1 March 2001

OPTICS
COMMUNICATIONS

Optics Communications 189 (2001) 39–46

www.elsevier.com/locate/optcom

Comparison of two approaches for implementing free-space optical interconnection networks

Ben Layet, John F. Snowdon *

Department of Physics, Heriot-Watt University, Edinburgh EH14 4AS, UK

Received 7 September 2000; received in revised form 26 October 2000; accepted 12 December 2000

Abstract

A particular design choice in the implementation of free-space optical interconnection networks (e.g. photonic backplanes) based on cascaded image-relay lenses is investigated. In these systems, a communication link can be implemented either by a single hop between source and destination nodes with the signal remaining in the optical domain through many image-relay stages, or by multiple hops between adjacent nodes with the signal undergoing optical–electrical conversion and vice versa at intermediate nodes (which act as repeaters). These two approaches place different demands on the optical system and the optoelectronic interface. We compare the raw bandwidth-per-link available in two example networks (the mesh and the completely connected network) using a model of the bandwidth and power consumption of an optoelectronic data channel and considerations on the aggregate bandwidth of the optoelectronic interface chip. We find that the single-hop approach provides a higher bandwidth-per-link. For example, the single-hop bandwidth-per-link is three times greater than the multiple-hop value for a mesh network of 49 nodes and for a completely connected network of 13 nodes. The advantage can increase further as the network size grows. The methodology is also applicable to the investigation of other implementation choices in optoelectronic interconnects. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Optical interconnects; Optical backplane; Photonic backplane; Free-space optics; Optoelectronic interconnects; Optical computing

1. Introduction

The development of optoelectronic interconnect technologies as alternatives to conventional electronic solutions is a continuing trend. Optoelectronic communications links are being developed for use over shorter and shorter distances. In the

next few years it is likely that board-to-board (photonic backplane), interchip and, perhaps, even intrachip communication will be enhanced by optoelectronic technologies [1]. Free-space optical interconnection networks are particularly attractive for connecting many nodes in a complex topology, where a node may be a board or a chip. Potential applications occur in multiprocessor computing systems and in switching systems. Several architectures exploiting this technology have been designed and various demonstrator systems constructed [2–5]. These systems are generally

* Corresponding author. Tel.: +44-131-451-3652; fax: +44-131-451-3136.

E-mail address: J.F.Snowdon@hw.ac.uk (J.F. Snowdon).

based on an optical system, sometimes referred to as an optical bus, that is comprised of many image-relay stages in a linear (or ring) physical topology. It should be noted that, despite the linear structure, such optical “buses” can support arbitrary logical topologies [5]. This is particularly important for high-dimensional networks that cannot be easily be designed as a free-space system by a direct mapping of the logical topology into 3D space. Multiple instances of optical buses can be used, of course, to further expand the system possibilities.

One important choice in the implementation of an optical bus is the manner in which each logical network link is supported. Consider, that, in general, a link is between a pair of nodes that are not adjacent in the linear physical topology of the bus. One implementation approach is to form such links from multiple hops between physically adjacent nodes. This has the advantage of simplifying the optical system design and assembly. The disadvantage is that the entire bandwidth of the bus passes through the optoelectronic interface at each node. An alternative approach is to use a single hop to form each (physically) long-distance link, with the signal remaining in the optical domain throughout. Since a high-performance free-

space optical system can carry more parallel signals than the optoelectronic interface, this method has the potential to exploit the capacity of the optical system more fully. However, since the signal beams travel further through the optics, therein suffering greater loss and becoming more aberrated, the channel bit rate must be lower. Consequently, it is not clear which approach provides the greatest aggregate bandwidth. The main purpose of this paper is to investigate this point.

In Fig. 1 the two approaches are illustrated for a single link in an arbitrary network. The link, between the nodes *a* and *d*, is implemented by a single hop in Fig. 1(a), and by multiple hops in Fig. 1(b). In both cases the routers and optoelectronic interfaces at *a* and *d* are involved in supporting the link. In the multiple-hop case the optoelectronic interfaces at *b* and *c* act as repeaters whereas in the single-hop case they are not used (and are free to support other links). In neither case are the routers at *b* and *c* involved. Hence the router usage is the same in both cases and need not be considered in this paper. We focus on the capabilities of the optoelectronic interface and the optical bus.

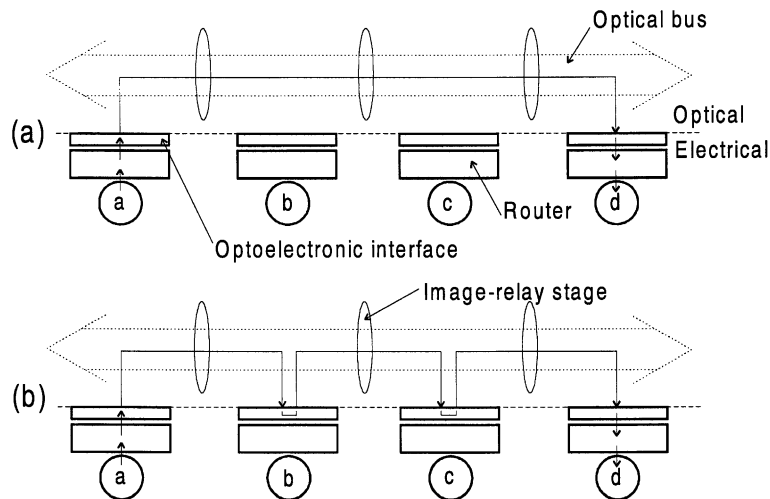


Fig. 1. Four nodes in an arbitrary network using an optical bus for interconnection, showing a single link between nodes *a* and *d* implemented by (a) a single hop involving several image-relay stages, and (b) multiple hops each involving one image-relay stage, with the optoelectronic interfaces at *b* and *c* acting as repeaters. Signal flow from *a* to *d* is indicated by arrows. Above the dashed line these represent optical signals, below electrical signals.

In order to compare the single-hop and multiple-hop approaches a model of an optoelectronic data channel is described and general expressions are derived for the raw aggregate bandwidth of the optoelectronic interface in terms of the performance of the constituent transceivers. Our arguments are based on the physical characteristics of the interconnect and, for the sake of simplicity, no attempt is made to account for protocol and control overheads, which in practice will consume some of the raw bandwidth. The model is used to compare the alternative approaches in the contexts of the mesh network and the completely connected network.

2. Model of an optoelectronic data channel

In this paper a single-bit optoelectronic data channel is considered to be implemented by the combination of a multiple-quantum-well modulator to transmit the signal, imaging optics to focus the beam onto the detector, and a transimpedance-amplifier receiver to amplify the detected photocurrent to a logic-level signal. Since it is generally the case that the modulator is capable of faster operation than the receiver, the raw bandwidth of the data channel is determined by the receiver bandwidth. The power consumption of the data channel is equal to the transceiver (i.e. modulator and receiver) power dissipation.

The receiver performance depends on the area (capacitance) of the detector and the power of the incident beam. These parameters are mainly determined by the optical system. In this paper we assume a system based on bulk lenses although other imaging components, such as microlens arrays, would also be an option. In a coherently illuminated image-relay system the diffraction-limited spot size is independent of the number of (identical) image-relay stages, but, due to random manufacturing errors in the lenses, the geometrical spot size (i.e. the spot size due to aberration) grows as the square root of the number of stages [6]. This assumes a well-corrected lens design, and also that a perfect alignment exists between lenses, otherwise some magnification of the spots may occur. The latter assumption dictates that our results in-

dicates the best possible performance that can be achieved by careful alignment (either manually or by active-alignment techniques). A practical technique sometimes used in the assembly of high-performance multielement lenses is to allow longitudinal adjustment of one element in order to adjust the effective focal length to the design value. Assuming that this technique is used, our detailed analysis of a five-element bulk lens design used in a $4f$ set-up shows that the distortion in the system also grows as the square root of the number of lenses [7]. (This also relies on the assumption of perfect alignment between lenses.) The required diameter of the detector is thus estimated as,

$$d_{\text{det}}(m) = \sqrt{d^2 + ma_1^2} + \sqrt{m}l_1 \quad (1)$$

where d is the diffraction-limited spot size, a_1 is the geometrical spot size after one image-relay stage, l_1 is the distortion after one image-relay stage, m is the number of image-relay stages. If the beam is not significantly clipped by apertures within the lens then the diffraction-limited spot size can be equated to the waist size of the input Gaussian beam from the source laser. In this case, the smallest spot size for a given m is ensured by an optimum choice of input Gaussian beam waist; too large or too small a waist results in a large spot. (A small waist produces a highly divergent beam that experiences large aberrations.) In the optical bus the number of image-relay stages in the signal path can be variable, requiring a compromise in the choice of the waist size. For a particular lens [8], our modelling shows that a 95% encircled energy waist size of $d = 18.4 \mu\text{m}$ produces spots that are within 15% of the minimum size for each value of m from 1 to greater than 10. The 0.95 percentile values of the other parameters are $a_1 = 9.8 \mu\text{m}$ and $l_1 = 9.6 \mu\text{m}$. The transmittance of the optics is estimated at 0.9 per image-relay stage, which includes bulk absorption and surface reflections in the two lenses in the $4f$ relay and also in a polarisation beamsplitter (which is required for beam combination) [8].

In this paper we use the transimpedance-receiver model developed in Ref. [9], in which the bandwidth of the two-beam transimpedance front end is related to the circuit parameters and the peak

incident optical powers on the two detectors, P_{high} and P_{low} , by,

$$B = \left[\frac{3V_{\text{min}}}{2S(P_{\text{high}} - P_{\text{low}})} \left(C_{\text{F}} + \frac{C_{\text{IN}}}{A + 1} \right) + \frac{3(C_{\text{IN}} + C_{\text{L}})}{Wg_{\text{m}}} \right]^{-1} \quad (2)$$

where W is the transistor width, C_{IN} , C_{F} and C_{L} are the input, feedback and load capacitances (each having constant terms and terms proportional to W), g_{m} is the transconductance per unit width, A is the ratio of the transconductance to the output conductance, S is the photodiode responsivity and V_{min} is the voltage swing required to drive the post-amplifier. Values for most of the parameters, for a 0.5- μm CMOS technology, are also given in Ref. [9]. The photodetector capacitance (which contributes to the parameter C_{IN}) is 0.15 fF/ μm^2 . (The optics model determines the required photodetector area.) The optics model and the modulator model together determine P_{high} and P_{low} .

For the sake of simplicity, we use this expression for the front-end bandwidth to estimate the bandwidth of the entire receiver. Expanding the capacitances into zeroth and first-order terms in W and applying the binomial expansion to Eq. (2) yields,

$$B = W \frac{1}{c} - W^2 \frac{a}{c^2} + O(W^3) \quad (3)$$

where,

$$a = \frac{3V_{\text{min}}}{2S(P_{\text{high}} - P_{\text{low}})} \left(C_{\text{F0}} + \frac{C_{\text{IN0}}}{A + 1} \right) + \frac{3(C_{\text{IN1}} + C_{\text{L1}})}{g_{\text{m}}} \quad (4)$$

$$c = \frac{3(C_{\text{IN0}} + C_{\text{L0}})}{g_{\text{m}}} \quad (5)$$

The extra single digits in the capacitance subscripts denote coefficients of zeroth or first-order terms (e.g. $C_{\text{IN}} = C_{\text{IN0}} + C_{\text{IN1}}W$). For small transistor sizes, the bandwidth depends linearly on the transistor size. However, as the transistor size grows, the slope of the curve decreases, which indicates that to increase the bandwidth further a

progressively greater penalty is paid in power-dissipation and circuit area. (These are proportional to the transistor size.) Since the receiver should be suitable for a smart-pixel environment, in which there are severe constraints on both of these parameters, it is sensible to select a receiver design that operates in or near the linear region of the curve. Setting the ratio of the second term to the first term to be a small number, h ($= 0.1$ for example), and ignoring higher order terms, gives, $W_{\text{des}} = hc/a$. Hence,

$$B_{\text{des}} = W_{\text{des}}(1 - h)/c = h(1 - h)/a \quad (6)$$

In our model the value of a is found to increase from 0.084 to 0.21 ns as the number of image-relay stages between source and detector increases from 1 to 10. Thus, the receiver bandwidth varies from 1.07 to 0.43 GHz over the same range. The receiver power dissipation can be estimated from the steady-state currents that flow from the power supply in the analogue front-end and post-amplifier stages. A simple model of the FET in saturation (see, e.g., Ref. [10]) can be used. The power dissipation is proportional to the transistor width. If the following constants are assumed for a generic 0.5 μm CMOS technology: $V_{\text{DD}} = 5$ V, $V_{\text{T}} = 0.8$ V and (the process transconductance parameters) $k'_n = 3.3$, $k'_p = 120 \mu\text{A}/\text{V}^2$, then the constant of proportionality is 1.11 mW/ μm .

The MQW modulator is assumed to have a reflectivity of 0.6 in the highly reflecting state and of 0.2 in the low state. The power dissipation of the modulator driver is estimated using the model described in Ref. [11]. The driver circuit consists of an appropriately sized transistor to drive the device plus (if necessary) a superbuffer to drive the transistor. For a fixed optical power in the read beam the power dissipation is given by,

$$P_{\text{MOD}} = P_{\text{T0}} + BP_{\text{T1}} \quad (7)$$

where P_{T0} and P_{T1} are constants. Using the values for the technology constants that are given in the same paper (which are based on 0.5- μm CMOS with hybrid integration of MQW modulators), and assuming the read signal power is 1 mW, one finds that P_{T0} is 1.65 mW and P_{T1} is 12.4 mW/GHz.

3. Aggregate bandwidth of the optoelectronic interface

The optoelectronic interface between a node and the optical interconnect supports a large number of optoelectronic data channels. In practice, the total number may be limited by device yield issues or power-dissipation constraints, for example. In order to examine these two cases we begin by deriving expressions for the (raw) aggregate bandwidth in terms of the individual channel bandwidths and either the total number of transceivers or the total transceiver power dissipation. Thus, the aggregate bandwidth is obtained either per transceiver or per unit of power dissipation, which permits the comparison of various network implementations that is presented in the following section.

In general the various transceivers in an optoelectronic interface may operate at different bit rates due to differences in the path through the optical system of the associated beams (which affect the spot size and power). Thus, in order to proceed with our analysis, we group the transceivers into L groups, with all the transceivers in a group identical. In group i , there are T_i transceivers with beam paths comprised of m_i image-relay stages. If each group is used to support a single logical link in the network topology then, because each network link should be able to support the same bandwidth, we have,

$$T_i B(m_i) = T_j B(m_j), \quad 1 \leq i \leq L, \quad 1 \leq j \leq L \quad (8)$$

where the bandwidth of a single channel $B(m)$ is explicitly shown as a function of the beam path. The parameters in the previous section that depend on m are the beam powers P_{high} and P_{low} and (through the beam area) the input capacitance C_{IN} .

(i) *Number-of-transceivers-limited aggregate bandwidth:* The aggregate bandwidth is given by,

$$B_{\text{SPA}} = \sum_{i=1}^L T_i B(m_i) \quad (9)$$

By using the requirement that all logical links operate at the same bandwidth (Eq. (8)) we find,

$$B_{\text{SPA}} = LT \left[\sum_{i=1}^L \frac{1}{B(m_i)} \right]^{-1} \quad (10)$$

where T is the total number of transceivers.

(ii) *Power-dissipation-limited aggregate bandwidth:* The power dissipation of the transceivers is given by,

$$P_{\text{SPA}} = \sum_{i=1}^L T_i (P_{\text{T0}} + P_{\text{T1}} B(m_i) + P_{\text{R1}} W(m_i)) \quad (11)$$

By using Eq. (6) to express W in terms of B , and, by using Eqs. (8) and (9), we obtain the following expression for the aggregate bandwidth,

$$B_{\text{SPA}} = P_{\text{SPA}} \left[\frac{P_{\text{T0}}}{L} \sum_{i=1}^L \frac{1}{B(m_i)} + P_{\text{T1}} + \frac{P_{\text{R1}}}{L(1-h)} \sum_{i=1}^L c(m_i) \right]^{-1} \quad (12)$$

4. Comparing the single-hop and multiple-hop approaches

In order to compare the single-hop and the multiple-hop approaches we examine the relative raw bandwidth available to each logical network link in the contexts of two network topologies: the 2D-mesh and the completely connected network (CCN) [12]. As stated in Section 1, we ignore protocol and control overheads which would simply reduce somewhat the actual data bandwidth. The topology, the network size, n , and the choice of either the single-hop or multiple-hop implementation determine the values of L and m_i to be used in Eqs. (10) and (12) for the aggregate bandwidth. These values are derived below.

4.1. Mesh network

Conceptually, the nodes in a 2D-mesh network can be pictured as lying on a square grid, with connections only between nodes that are immediately horizontally or vertically adjacent to one another. This topology can be sensibly embedded in a linearly structured optical bus in a row-by-row (or column-by-column) fashion. Thus, of the four

possible neighbours of a node in the logical mesh topology, two remain physically adjacent in the mapping onto the linear structure, whilst the distance to the other two becomes equal to the number of nodes in a row (or column). Hence, to evaluate the single-hop aggregate bandwidth one should use, $L = L_{sh} = 4$, $m_1 = m_2 = 1$ and $m_3 = m_4 = \sqrt{n}$, (which assumes that a link between m th nearest neighbours uses m image-relay stages). The bandwidth available to each logical link is the aggregate value divided by L .

In the multiple-hop case, by definition all physical connections are to adjacent nodes. Therefore, $m_i = 1$, for all i . However, in addition to the four possible links that are logically connected to a node, $\sqrt{n} - 1$ links between other nodes use the optoelectronic interface as a repeater (in the same manner as, in Fig. 1, the optoelectronic interfaces of b and c are used by the link between a and d). Hence, to evaluate the multiple-hop aggregate bandwidth one should use $L = L_{mh} = 4 + 2(\sqrt{n} - 1)$ and $m_i = 1$. The links using the optoelectronic interface as a repeater are counted doubly because each requires a signal to be received and retransmitted in both directions (for a bidirectional link) whereas each logically connected link requires a single optical signal reception and transmission. Again the bandwidth available to each logical link is the aggregate value divided by L .

The ratio of the single-hop bandwidth-per-link to the multiple-hop bandwidth-per-link in the mesh topology is plotted against network size in Fig. 2. Note that the longest optical path, for the largest network shown, is 16 image-relay stages. Separate curves are plotted for the cases of an equal transceiver power dissipation and an equal numbers of transceivers. Also plotted is the ratio $L_{mh}:L_{sh}$, which is equal to the single-hop to multiple-hop bandwidth ratio in the idealised case in which the aggregate bandwidth of the optoelectronic interface is fixed. For small network sizes all three curves are very similar; the difference in the single-hop and multiple-hop performance is largely due to the difference in the number of links that share each optoelectronic interface. As the network size grows the effect of longer optical signal paths in the single-hop approach (which reduce the bandwidth of 1-bit data channels) causes the modelled

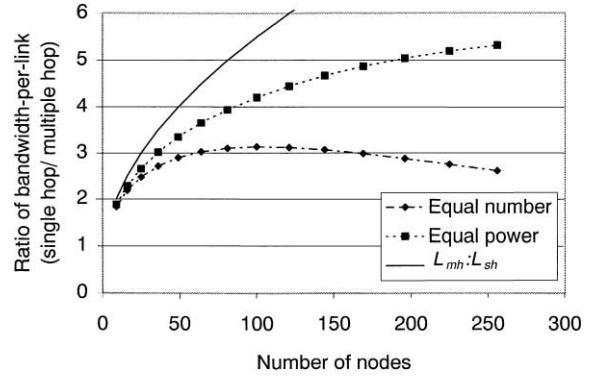


Fig. 2. Ratio of the single-hop to multiple-hop raw bandwidth-per-logical-link against network size assuming a mesh topology. The dotted lines show results from the full model. The solid line shows the idealised case of a fixed optoelectronic-interface aggregate bandwidth.

bandwidth ratios to grow more slowly (than $L_{mh}:L_{sh}$). In the equal-transceiver-power-dissipation case the ratio falls less quickly, because as the single-hop data channels slow down they also consume less power so more of them can be used. Obviously, this is not permitted when a fixed number of transceivers is assumed.

4.2. Completely connected network

In the CCN each node is connected to all other nodes, with a dedicated link available for each connection. Hence, the mapping of the CCN onto the optical bus is trivial; the ordering of the nodes is entirely arbitrary. Assuming the bus is connected in a ring, the length of the links varies from 1 to $n/2$. Therefore, in the single-hop case, $L = n - 1$ and $m_i = \lceil i/2 \rceil$.

In the multiple-hop case, $m_i = 1$, for all i . In a ring, there is a choice of two routes for each link. To evaluate the number of links that pass through the optoelectronic interface at each node we assume that all links use the shortest distance around the ring. If n is odd then the longest link is $(n - 1)/2$. For a pair of nodes separated by this distance there are no cases of a link connected to one node using the optoelectronic interface of the other as a repeater (it would have to be longer than the longest link). If the separation of the nodes is reduced by one then there is one such case. The

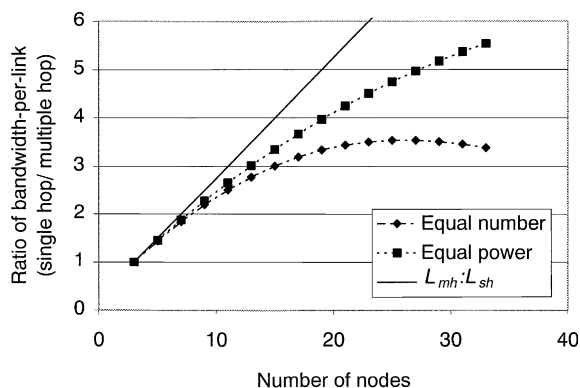


Fig. 3. Ratio of the single-hop to multiple-hop raw bandwidth-per-logical-link against network size assuming a completely connected network topology. The dotted lines show results from the full model. The solid line shows the idealised case of a fixed optoelectronic-interface aggregate bandwidth.

number of cases increases by one each time the separation decreases by one. Therefore the number of links, L_{rpt} , that use the optoelectronic interface of a node as a repeater is given by,

$$L_{\text{rpt}}(n \text{ odd}) = \sum_{i=1}^{((n-1)/2)-1} i = \frac{(n-3)(n-1)}{8} \quad (13)$$

Therefore, the value of $L = (n-1) + 2L_{\text{rpt}} = (n^2 - 1)/4$.

In Fig. 3 the bandwidth-per-link in the two cases is compared for the CCN in the same manner as for the mesh network above. Again, and for the same reasons, for small network sizes all three curves are very similar. The advantage of the single-hop approach is evident. This advantage grows much more rapidly with network size in the CCN because of the longer links required to implement this higher connectivity topology. Again, the longest optical path, for the largest network shown, is 16 image-relay stages.

5. Discussion

The results presented above are useful to compare single-hop and multiple-hop implementations of an optical bus in cases where the optoelectronic interface design is significantly influenced by the properties of the optoelectronic devices and asso-

ciated circuitry (including the photoreceivers), through either power-dissipation restrictions or yield issues. In addition, circuit area concerns are largely captured by the power-dissipation case, as both these parameters are proportional to transistor size. In cases where the optoelectronic interface design is strongly influenced by logic circuitry (in smart-pixel systems, perhaps) or drivers for electrical lines going off-chip, for example, the relative ease of building the appropriate optical system may be more important in determining which implementation is chosen. However, in practice, the power consumption, area usage and yield issues of the optoelectronic devices and associated circuitry are rarely secondary concerns.

Of course, our results rely on the accuracy and relevance of the underlying models. In the optics model, random errors are assumed to be dominant. However, for large systems (i.e. with many image-relays stages) and for systems based on lenses with significant residual aberrations in their design, systematic errors, which grow linearly with the number of stages, will tend to become dominant. This implies that the advantage of the single-hop approach that we have shown becomes harder to reach in practice for larger systems, as it becomes more challenging to eliminate systematic errors. Our electrical model is based on 0.5- μm CMOS technology. We have not considered scaling to smaller feature-size technologies but do not expect to see any fundamental changes since the scaling is likely to have a similar affect on both the single-hop and multiple-hop approaches. In any case, regardless of these limitations, we believe that this paper demonstrates a useful methodology for analysing optical bus-type systems.

In summary, we have demonstrated that the raw bandwidth of an optical bus formed from a cascade of image-relay lenses is greater for single-hop implementations of network links than for multiple-hop implementations for a large range of network sizes, for two very different network topologies and under two alternative assumptions on the limit of the optoelectronic interface aggregate bandwidth. In practice, the performance advantage must be weighed against the more complex optical system and system alignment that is required for a single-hop approach.

Acknowledgements

Funding from EPSRC through the OSI programme is acknowledged.

References

- [1] D.A.B. Miller, Motivations for optical interconnects to silicon chips, in: R.A. Lessard, T. Glastian (Eds.), *Optics in Computing 2000*, SPIE vol. 4089, 2000, p. 486.
- [2] J.A.B. Dines, J.F. Snowdon, M.P.Y. Desmulliez, D.B. Barsky, A.V. Shafarenko, C.R. Jesshope, J. Parallel Distrib. Comput. 41 (1997) 120–130.
- [3] Y. Liu, B. Robertson, G.C. Boisset, M.H. Ayliffe, R. Iyer, D.V. Plant, *Appl. Opt.* 37 (1998) 2895–2914.
- [4] S. Araki, M. Kajita, K. Kasahara, K. Kubota, K. Kurihara, I. Redmond, E. Schenfeld, T. Suzuki, *Appl. Opt.* 35 (1996) 1269–1281.
- [5] T.H. Szymanski, H.S. Hinton, *Appl. Opt.* 35 (1996) 1253–1268.
- [6] W. Smith, *Modern Optical Engineering – The Design of Optical Systems*, second ed., McGraw-Hill, New York, 1990, pp. 486–487 (Chapter 14).
- [7] B. Layet, J.F. Snowdon, *Modelling of free-space image-relay systems for optical interconnection*, unpublished.
- [8] D.T. Neilson, S.M. Prince, D.A. Baillie, F.A.P. Tooley, *Appl. Opt.* 36 (1997) 9243–9252.
- [9] M. Forbes, *Electronic design issues in high-bandwidth parallel optical interfaces to VLSI circuits*, Ph.D. Thesis, Heriot-Watt University 1999, Chapter 4 (available online: <http://www.phy.hw.ac.uk/resrev/SPOEC/thesis.pdf>).
- [10] A.S. Sedra, K.C. Smith, *Microelectronic Circuits*, fourth ed., Oxford University Press, New York, 1998 (Chapter 5).
- [11] G.I. Yayla, P.J. Marchand, S.C. Esener, *Appl. Opt.* 37 (1998) 205–227.
- [12] V. Kumar et al., *Introduction to Parallel Computing*, Benjamin/Cummings, Redwood City, 1994 (Chapter 2).