Impact of gate fan-in and fan-out limits on optoelectronic digital circuits

Lianhua Ji and V. P. Heuring

The impact of gate fan-in and fan-out limits on digital circuit delay is discussed with a set of benchmark circuits. This research presents the advantages of exploiting the ability of optoelectronic gates to perform both logic operations and optical interconnections with systematic optimization. It is possible for gate-level optical interconnected optoelectronic circuits to compete with their pure silicon counterparts in terms of the combinational circuit delay and system clock rate. © 1997 Optical Society of America

1. Introduction

Numerous digital optical computing architectures have been proposed and demonstrated in recent years.[1–5] However, there is still considerable debate as to the usefulness of these architectures in general purpose digital computing. The main issue is the overall computing power of an optical processor compared with its VLSI-based electronic counterpart. We show that, in certain digital computing applications, optoelectronics can compete with electronics in computing speed by systematic optimization based on optoelectronic (OE) devices.

In this research we explore the capability of performing with OE devices both logic operations and optical interconnects to achieve an overall circuit delay smaller than that of electronic devices. In a digital computer the circuit delay determines the clock rate of the system. As shown in Fig. 1, the minimum system clock cycle must be equal to or larger than T_{delay}, the longest circuit delay among all the signal paths in the system. T_{delay} is the sum of the combined circuit delay and interconnection latency. With the optical inputs and outputs, the advantage of a smaller interconnection latency is obvious for OE circuits because the interconnection latency of electronic circuits is determined by the RC value of metal lines, which is a constant and cannot be improved by miniature feature sizes. Thus, if the combined circuit composed of OE circuits can be made equal to or smaller than electronic circuits in size, the system clock rate can be increased.

We show that such a goal can be achieved with high-fan-in and high-fan-out OE gates with gate-level optical interconnection. We also show that it is feasible to fabricate OE gates with desired fan-in and fan-out requirements by using existing technology. Numerical comparison of the best-case circuit delays of OE and electronic circuits is used to evaluate the potential speed of the hybrid computing system, consisting of electronic arithmetic and logic unit (ALU) and OE control circuits. The target applications are those functional units in a superscalar digital computing system. A common characteristic is that these units require less processing operation but have heavy fan-in and fan-out; circuit delay is critical as well.

In our research we use standard VLSI computer-aided design (CAD) tools[6,7] to calculate the lower-bound circuit depth of benchmark[8] circuits as a function of gate fan-in and fan-out. The optimal gate fan-in and fan-out requirements are determined on the basis of the statistical profile of the calculated circuit depth versus the fan-in and fan-out limits of the logic gates. We then discuss the feasibility of high fan-in and fan-out OE gates and the OE circuits. We compare the speed of OE and electronic circuits by mapping benchmarks to their OE and electronic implementations with optimized circuit delays. Finally, the impact of high-fan-in and high-fan-out OE circuits is evaluated at the system level. Because the chosen benchmark circuits and the simulators are those commonly used in industrial and electronic VLSI-design-tools research groups, we believe the data and conclusions of this research are closer to the

The authors are with the Optoelectronic Computing Systems Center, Department of Electrical and Computing Engineering, Campus Box 425, University of Colorado at Boulder, Boulder, Colorado 80309-0525.

Received 15 January 1997; revised manuscript received 5 February 1997.

0003-6935/97/173927-14$10.00/0

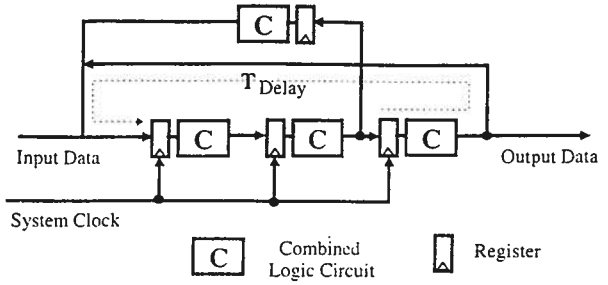


Fig. 1. Digital system clock rate and circuit delay.

situation in real digital computing than are those based on arbitrary assumptions. We hope that our results provide guidance to device developers by identifying, from the circuit designer's point of view, the most desirable characteristics of OE logic elements.

This paper consists of five parts: the statistical relation between minimum circuit depth and gate fan-in and fan-out limits; a feasibility evaluation of high-fan-in and high-fan-out OE gates on the basis of present OE devices technology; a circuit-delay performance comparison of OE and electronic circuits; the rule of OE circuits in computer system architecture; and the most promising near-future applications of OE circuits and the potential computational power of a hybrid superscalar computer.

2. Circuit Depth versus Gate Fan-in and Fan-out Limit

The speed of digital computing hardware is determined ultimately by the delay characteristics of its functional units and interconnections. Thus a smaller circuit and shorter interconnection delay translate to increased system clock rates. A short circuit delay also reduces the access time to the memory hierarchy, which determines the real execution time of a computation task. To sharpen the focus, we concentrate our discussion on the basic combined circuit-delay characteristics.

A. Definition of Combined Circuit Depth and Delay

We define circuit delay as the time interval between the arriving inputs and the appearance of valid signals at the outputs. Figure 2 shows the factors that contribute to circuit delay: T_g, the intrinsic gate delay; T_f, the fan-out delay; and the depth of the circuit's critical path, which is 4 in Fig. 2. The fan-out delay defined here is the total effect of the fan-out number and the connection capacitance. The critical path, represented by the thicker darker lines in Fig. 2, is defined as the propagation delay from the time the input signals are applied to the time all output signals have become valid. Since a computer may consist of millions of gates, it is impossible to perform critical path-delay optimization with a map of the whole computer as a single circuit.

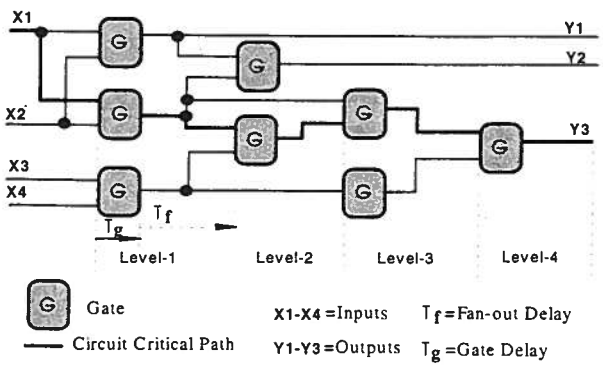


Fig. 2. Circuit depth and delay.

The standard design flow is set up to partition a computer into several smaller circuits and optimize each circuit delay individually. Because the critical path of a subcircuit may not be part of the critical path of the whole system, certain trade-offs can be made between the local (subcircuit) and the global (system) critical path delays. One trade-off strategy is finding such a gate net list whose total output delay T_{op} is the smallest:

$$T_{op} = \min\left\{T_{\text{net\_}k}, [\text{net\_}k \in O, T_{\text{net\_}k} = \sum_{p\in\Omega} t(\text{net\_}k, p)]\right\}.$$

In this equation, $T_{\text{net\_}k}$ is the sum of the output delays of net list k, net_k is one member of O, which is the set of all candidate net lists, p is the index of the pth output, Ω is the set of all paths of outputs, and t(net_k, p) is the longest input-to-output delay of the pth output within net list k. After the optimal net list is found, the circuit delay can be determined by the critical path delay as

$$T_{\text{circuit}} = \max\left\{T_p, [p \in \Omega, T_p = \sum t(p, i)]\right\},$$

where T_{circuit} is the circuit delay, T_p is the delay of output p, and $t(p, i)$ is the delay of ith level along path p. In summary, the *circuit* delay is defined as the *critical path* delay; and at the same time the sum of all *noncritical* path delays of that circuit is the *smallest* among all alternative circuits.

Circuit depth (or circuit level) is defined as the maximum number of gates along each signal path from input to output. The circuit depth is 4 in Fig. 2, which is equal to the number of levels within the critical path. We define the boundary of each circuit level as beginning at the input of the source gate and ending at its destination gate. From circuit input to output, the levels are labeled as lv1 to lv4. The circuit delay can be represented as

$$T_{\text{circuit}} = DT, \tag{1}$$

where D is the circuit depth, and T is the average delay in each circuit level, which is the sum of the gate's intrinsic delay and the signal-propagation delay between two gates.

In the case of electronic circuits, both T and D are functions of the fan-in and fan-out characteristic of gates. Optimal circuit design must take several optimization steps to minimize the circuit delay. These steps include logic simplification at the logic

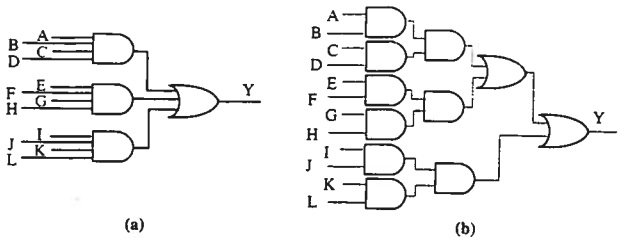


Fig. 3. Circuit depth and fan-in limits: (a) maximum fan-in of 4, circuit depth of 2; (b) maximum fan-in of 2, circuit depth of 4.

level, technology mapping at the circuit level, and routing optimization at the physical-layout level. As explained in Subsection 3.A, T is largely independent of the actual number of fan-ins and fan-outs for OE circuits with OE gates. Thus the impact of fan-in and fan-out on OE circuit delay can be evaluated with circuit depth. The actual circuit delay can be calculated with Eq. (1), with the given level delay T as the scale factor. In Subsection 2.B, we examine the statistical relation of circuit depth and the fan-in and the fan-out limits of logic gates. Our goal is finding the required gate fan-in and fan-out limits by which the circuit depth (therefore the circuit delay) can be reduced to near its theoretical minimum. The CAD tool used here is SIS (Sequential Interactive Synthesis), Version 1.2 from the University of California, Berkeley.[6]

B. Impact of Gate Fan-in on Circuit Depth

We define the theoretical minimum circuit depth as the number of circuit levels after logic minimization and mapping to primary gates (AND and OR) with infinite fan-in and fan-out. According to this definition the theoretical minimum circuit depth is 2 for all logic circuits because any logic function can be expressed in the form of sum of products, and without fan-in limits all the product terms can be implemented by AND gates and summed by one OR gate. The critical path thus includes only one AND gate and one OR gate, corresponding to a circuit depth of 2. In the real world the assumption of infinite fan-in and fan-out is always invalidated. The circuit depth is a function of both gate fan-in and fan-out limits. We examine first the impact of fan-in limit on circuit depth. When the variable number of a product term is larger than the fan-in limit of the AND gate, one must implement this product term by cascading more than one AND gate. As a result the circuit depth increases. Figure 3 shows the impact of gate fan-in on circuit depth, with two implementations of the function:

$$Y = A \cdot B \cdot C \cdot D + E \cdot F \cdot G \cdot H + I \cdot J \cdot K \cdot L.$$

When the maximum fan-in is 4, the circuit depth is 2, as in Fig. 3(a). When the maximum fan-in is reduced to 2, the circuit depth increases to 4, as shown in Fig. 3(b). For circuits with multiple inputs and a single output, the circuit level can be estimated approximately by

$$\text{Level} = \lfloor \log_{n\text{-fin}}[\max\{P\}] \rfloor + \lfloor \log_{o\text{-fin}}[S] \rfloor, \qquad (2)$$

where n-fin is the fan-in limit of AND gates, o-fin is the fan-in limit of OR gates, $\max\{P\}$ represents the maximum number of inputs of all product terms, and S is the number of product terms. According to Eq. (2), the circuit level reduces logarithmically with an increasing fan-in limit of the gates, that is, Level $\propto$ 1/log[maximum fan-in number].

Table 1. Simulation of the Circuit Depth Versus the Gate Fan-In Number[a]

Circuit	Input Number	Output Number	Circuit Depth versus Gate Fan-In Limit 2	3	4	5	6	7	8	Required Maximum Fan-Out
5xp1	7	10	8	5	5	4	3	3	3	53
b12	15	9	7	4	4	4	3	3	2	24
bw	5	28	6	4	3	2	2	2	2	88
clip	9	5	9	6	5	5	5	3	3	118
con1	7	2	5	3	3	2	2	2	2	6
duke2	22	29	9	6	5	4	4	4	4	168
ex5	8	63	5	4	3	3	3	3	3	138
inc	7	9	7	5	4	3	3	3	3	39
misex1	8	7	6	4	4	3	2	2	2	28
misex2	25	18	6	4	3	3	3	3	3	27
misex3c	14	14	7	8	5	5	5	5	4	197
rd53	5	3	7	4	3	3	3	3	3	30
rd73	7	3	6	5	5	4	4	3	3	123
rd84	8	4	7	6	6	5	5	4	4	264
sao2	10	4	6	5	4	4	4	4	4	41
squar5	5	8	5	4	3	3	3	3	2	21
average	9.6	13.5	6.6	4.7	4.1	3.6	3.3	3.1	2.9	85

[a]No fan-out limit was used.

It is difficult to derive a closed-form expression to estimate the circuit depth based on the fan-in limit with circuits that have multiple inputs and multiple outputs. Although the functionality can still be implemented on the basis of a group of multiple input and a single output circuits, this approach requires a considerable number of gates. One way to reduce the gate count is through multilevel logic optimization. The principle of this approach is the introduction of intermediate variables shared by multiple outputs. However, there is a penalty in circuit depth because of the introduction of intermediate variables.

We simulated 16 benchmark circuits with multilevel logic optimization and technology-independent decomposition algorithms[9] to avoid an unreasonably high gate count in the OE implementation. The simulation results allow us to quantify statistically the trend of circuit depth versus fan-in limit for multiple-input and multiple-output combined logic circuits. Table 1 shows the simulated values of circuit depth against the fan-in limit for each benchmark. The last column in Table 1 lists the required maximum fan-out when the maximum fan-in equals 8. The last row of Table 1 shows the arithmetic mean of the circuit depth. As the fan-in limit in-

Table 2. Simulation of the Circuit Depth Versus the Gate Fan-In Number[a]

Circuit	Circuit Depth/Maximum Fan-Out Limit						
	Fan-In 2	Fan-In 3	Fan-In 4	Fan-In 5	Fan-In 6	Fan-In 7	Fan-In 8
5xp1	10/16	8/3	8/4	7/4	7/4	6/4	6/5
b12	10/11	7/14	6/5	5/10	6/10	6/10	5/10
bw	9/6	8/7	7/6	7/4	5/5	5/5	5/5
clip	12/6	9/6	9/5	8/5	8/5	8/6	8/5
con1	6/4	4/5	5/4	3/4	3/4	3/4	3/4
duke2	13/18	10/15	9/10	8/12	8/18	8/18	8/18
ex5	10/5	8/6	7/7	7/5	7/6	8/5	8/6
inc	10/17	8/9	7/16	7/13	8/13	6/13	6/13
misex1	8/5	6/14	6/4	6/12	5/12	5/12	5/12
misex2	8/11	7/9	6/10	6/10	5/10	5/10	5/9
misex3c	14/20	10/23	10/21	10/21	10/21	9/20	9/20
rd53	10/3	8/3	7/3	6/4	6/3	6/4	6/4
rd73	13/6	10/5	9/4	10/5	9/5	7/5	7/4
rd84	13/10	10/10	10/9	11/9	9/9	9/9	9/9
sao2	12/4	8/6	8/5	8/5	8/5	8/4	8/4
squar5	8/14	7/18	6/16	6/16	5/16	5/14	4/14
average	10.4/9.1	7.7 /9.6	7.4 /8.7	7.1 /8.6	6.9 /9.1	6.6 /9.1	6.1 /8.9

[a]Fan-out was limited.

creases from 2 to 8, the average circuit depth decreases from 6.6 to 2.9, a factor of 2.3. This gain in circuit-depth improvement is less than the expected gain of $\log_2[8]/\log_2[2] - 1 = 3$, from Eq. (2). This 70% margin is caused by the multiple outputs of the circuits. Equation (2) applies only to multiple-input–single-output circuits.

C. Impact of Gate Fan-out on Circuit Depth

In this subsection we examine the impact of the gate fan-out limit on circuit depth with a given gate fan-in limit. The fan-out limit defined in this paper is the maximum number of inputs that a logic gate can drive. For the electronic gate the fan-out limit is determined by the current drive ability. The output current must be large enough to charge the load gate(s') input capacitor(s) and wire capacitor within a desired time. For OE circuits the maximum fan-out is limited by the ratio of the gate's output optical power to the minimum switching power of the gate's input, adjusted for any losses between the optical output and input; thus this is referred to as input sensitivity. After logic minimization the circuits might require a large fan-out that is beyond the maximum fan-out limit of the available gates. We use two techniques to deal with the gate fan-out constraint when the logic function is mapped to the gate net list: One is the introduction of an extra buffer stage(s) for use in clock distribution; the other is the modification of the structure of the gate net list by the insertion of intermediate nodes. Both methods cause the circuit depth to increase. To evaluate the circuit-depth increase caused by limiting fan-out, we simulated the same benchmarks using fan-out optimization.

Table 2 lists the calculated circuit depth and its maximum fan-out for each benchmark after fan-out optimization, based on the given fan-in limit. With the same gate fan-in limit, the maximum fan-out for each individual benchmark fluctuates rapidly. However, the statistical maximum fan-out changes little around 9 when fan-in changes from 2 to 8, as shown by the last row in the table.

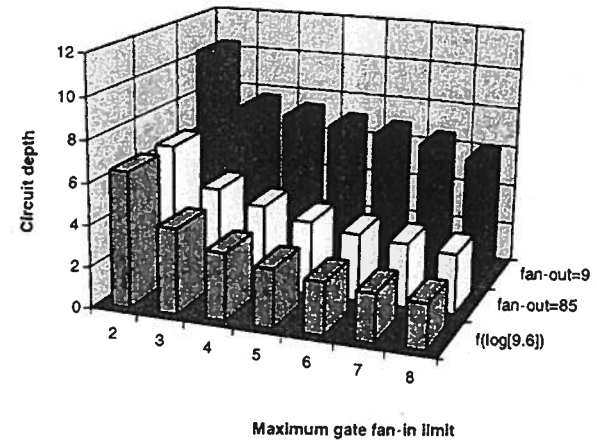
Figure 4 shows the statistical relation between fan-in and circuit depth with and without fan-out. Figures 4(a) and 4(b) represent the absolute value and relative improvement, respectively, of circuit depth. In Fig. 4(b) the circuit-depth improvement is normalized to the worst case, where the fan-in and fan-out limits are 2. The third trace in both figures is the expected lower bound of the circuit depth versus the gate fan-in limit, which is derived from a parallel-prefix circuit model.[10] This lower bound is calculated from

$$\min - \text{depth} = 2 \log_x(N),$$

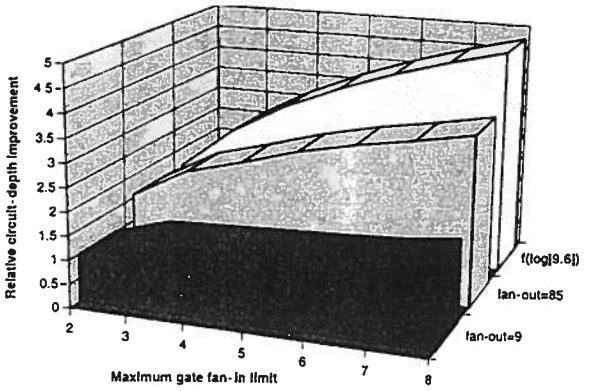
where x is the maximum fan-in limit and N is the number of function input variables. The value of $N = 9.6$ that is used in the plots is the arithmetic mean number of inputs of benchmarks (see Table 1).

According to Fig. 4, the circuit depth decreases in approximately logarithmical fashion with the maximum gate fan-in in both small and large fan-out cases. The value of the optimal fan-in limit is approximately 6 to 8. Beyond 8, the circuit-depth improvement caused by increased gate fan-in is trivial.

A large gate fan-out limit is also important in achieving a smaller circuit depth. The circuit depth with a fan-out limit of 85 is close to the calculated lower circuit-depth bound in Fig. 4(a). However, there is little relative circuit-depth improvement as the gate fan-out limit is reduced to 9. The importance of this relation can be explained by the numerical values of the circuit depths in the two cases. With a gate fan-out limit of 85 and fan-in limit of 8,



(a)



(b)

Fig. 4. (a) Circuit depth versus the fan-in limit; (b) normalized circuit-depth improvement versus the fan-in limit.

the average circuit depth is 2.9. When the gate fan-out limit is reduced to 9, the circuit depth increases to 6.1 at a gate fan-in limit of 8. This corresponds to a relative increase in circuit depth of 6.1/2.9, a factor of 2.1.

With these figures of 8 and 85 as optimum values of OE gate fan-in and fan-out limits, we are in a position to compare electronic and OE circuit delays. The crux of our argument is that circuit delay in electronics is greater than the delay in equivalent OE circuits because the fan-in and fan-out limits imposed on electronic gates are smaller than those that are achievable in OE gates. In Section 3 we show these limits are achievable with OE NOR gates with free-space interconnects. The fan-in and fan-out of submicrometer complementary metal-oxide semiconductor (CMOS) gates are comparatively low owing to constraints of noise margin and transistors' current drive ability.[11] In the electronic CAD world, designers usually limit themselves to a gate fan-in limit of 2–3 and a fan-out of <8. The relative OE and electronic circuit delays then can be estimated by

$$\frac{D_E T_E}{D_{OE} T_{OE}} = \frac{10.4 T_E}{2.9 T_{OE}} = 3.59 \frac{T_E}{T_{OE}},$$

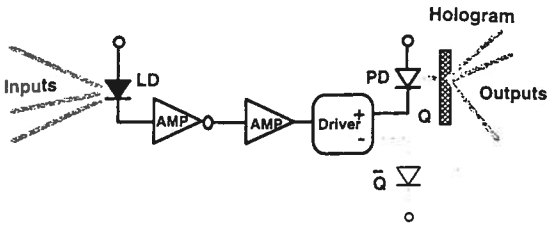


Fig. 5. Circuit diagram of an OE NOR gate. Amp, amplifier.

for electronic circuits with a gate fan-in of 2 and

$$\frac{D_E T_E}{D_{OE} T_E} = \frac{7.7 T_E}{2.9 T_{OE}} = 2.66 \frac{T_E}{T_{OE}},$$

for electronic circuits with a gate fan-in of 3, where D represents the circuit depth, T is the average level delay as defined in Eq. (1), and the subscripts E and OE represent electronic and OE circuits. According to the above equations, the actual value of the relative circuit delay is determined by the ratio of average delays between two gates of electronic to OE circuits. If $T_E \approx T_{OE}$, the expected circuit-delay speed-up for the OE circuit will be between 266% and 359%.

3. Feasibility of High-Fan-in and High-Fan-out Optoelectronic Gates

As mentioned in Subsection 2.C, the optimal gate fan-in and fan-out limits are 8 and 85 to take full advantage of the reduction of circuit depth. In this section we discuss the feasibility of high-fan-in and high-fan-out OE gates on the basis of present existing OE integration technologies. First, some definitions:

- OE gate: An OE gate is a logic element with optical input(s) and output(s). It is possible for us to have an OE gate perform any logic function by inserting a corresponding electronic digital circuit between its optical input and output. We restrict our discussion to primitive NOR gates that consist of a photodetector, analog amplifiers, a current driver, and a laser. Figure 5 shows a typical circuit diagram of an OE NOR gate.[12] The hologram in Fig. 5 should not be counted as a part of a NOR gate. We included it in the figure to show readers how a large fan-out can be implemented. We can implement the logic OR operation by connecting the output laser to the negative output of the driver. We restrict our discussion to NOR gates because they are logically complete, that is, any arbitrary logic function can be completed with NOR gates only.
- OE circuit: An OE combined circuit makes use of OE gates and optical interconnections to perform the desired logic function. We assume in this research that an OE circuit contains only NOR gates connected optically with lenses or other optics such as holograms.

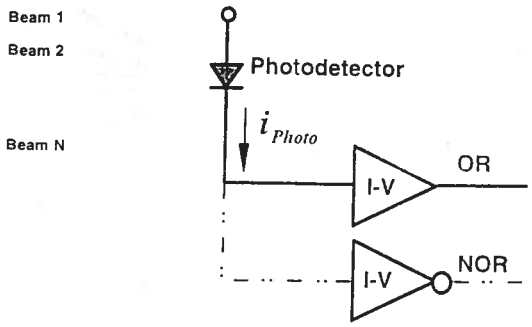


Fig. 6. Optical wired-OR and wired-NOR gates.

A. Fan-in Constraint on Optoelectronic Gates

Figure 6 explains how a multiple-input NOR operation can be performed with a single photodetector. We perform the logical NOR operation by ensuring the presence of a certain amount of photocurrent generated by a detector. The threshold is set so that the photocurrent generated by one single incident input beam is large enough to switch the output laser from logic 1 to 0. Because the NOR gates are logic complete, any function can be constructed exclusively from them.

The most attractive feature of OE NOR gates is their simple hardware when they deal with multiple inputs. For OE NOR gates a single detector is shared for all inputs, whereas for electronic gates at least one transistor must be added for each additional input. Another potential advantage of an OE NOR gate is that its intrinsic gate delay is less independent of the active input number. This advantage results from the built-in gain of the internal amplifier (see the explanation in the fan-out discussion in Subsection 3.C). It is possible for a high-gain and high-bandwidth amplifier to amplify the weakest input signal (a fan-in of 1) to a strong output with an output rise time of less than 100 ps. Although the rise time of an amplified input signal can be improved with higher fan-in, the relative improvement may be negligible owing to the dominant gate internal delay, which can be a few hundred picoseconds. On the basis of such an argument we define the OE gate's intrinsic delay as the internal circuit delay with an optical fan-in = 1 and assume that the interconnection delay is independent of both fan-in and fan-out.

The advantage of decoupling the fan-in and fan-out from the interconnection delay is that it simplifies the logic synthesis. On the other hand, the delay performance of an electronic gate degrades linearly with increased fan-ins. The reason for the degradation of an electronic gate's intrinsic delay is the linearly increasing output RC constant. Figure 7 shows the logic circuits of the electronic version of NOR gates and their equivalent ac circuits. In Fig. 7(a) the time constant is $C_L(r2 + r4)$ for two inputs, where C_L is the equivalent output capacitance of a NOR gate, and $r2$ and $r4$ are the pull-up transistor's output resistance. C_L includes the output capacitors of the pull-up transistors, the wire capacitor, and the input capacitor of

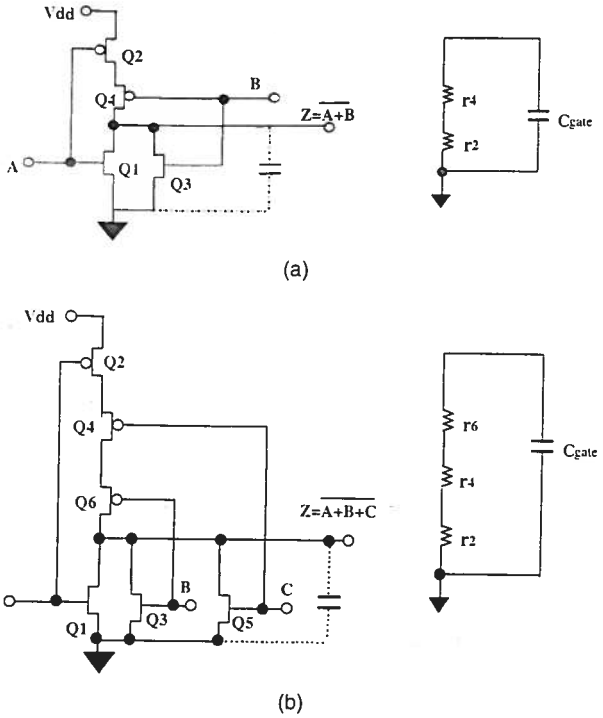


Fig. 7. (a) Two-input electronic OR gate and its equivalent circuit. (b) Three-input electronic OR gate and its equivalent circuit.

the destination gate. Compared with the other two components, the effect of the output capacitors of the pull-up transistors is negligible. Thus we can assume C_L to be a constant, independent of fan-in. When inputs increase to 3, the time constant becomes $C_L(r2 + r4 + r6)$, as shown in Fig. 7(b). If the transistors are of the same size, the output resistance of each transistor is the same, $r2 = r4 = r6 = r$. Thus the time constant is $C_L r$, which linearly increases with the number of fan-ins. Because at least one transistor has to be added to each additional input, the size of a gate increases with fan-in. The spacing between two gates will increase owing to increased gate size, increasing the length of interconnection. As longer wire length means larger capacitance, the connection delay also increases.

The fundamental constraint on the OE gate fan-in limit is on the optical side. The main concerns are optical interconnect loss and cross talk from the optical routing. Cross talk affects the signal-to-noise ratio, which determines system reliability. The interconnect loss determines the maximum gate fan-out at the system level because the photocurrent generated by the detector is proportional to the intensity of incident light.

Stirk[13] calculated the theoretic lower bound of the optical OR logic's bit-error rate (BER) caused by the cross talk from parallel fan-in. His work shows that the BER degradation caused by the shot noise of optical sources is trivial for optical OR–NOR operation if the signal's ON–OFF ratio is high enough. For example the BER can be as low as $<10^{-50}$ with respect to a fan-in of <25 and signal ON–OFF (contrast) ratio

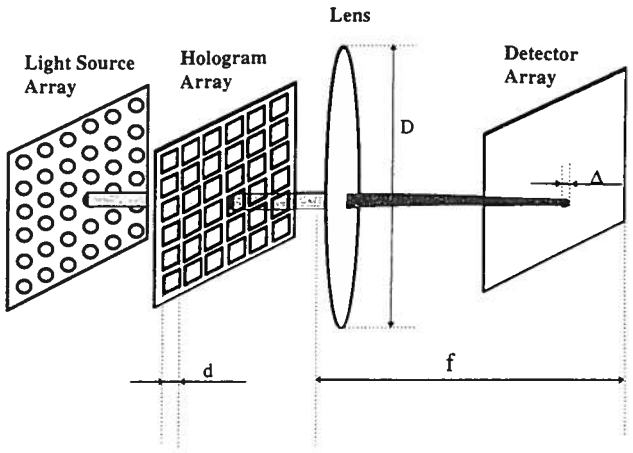


Fig. 8. Photodetector size and optical interconnection loss.

of >32. In the case of a free-space interconnect, the signal-to-noise ratio at the detector end must include the factors of distortion of the hologram, wavelength change of the laser, and assembly misalignment. Morozov[14] calculated the contrast ratio and interconnect efficiency as functions of hologram diameter and detector size. According to his data, a contrast ratio of >200 is practical when the dimensions of hologram D are chosen correctly. For example, if the detector size is $\Delta = 75$ μm and the hologram diameter is 245 μm, the contrast ratio is >200 and the optical connect efficiency is >90%.

The size of the active area of the photodetector is important for the reduction of the power loss in optical routing. The brightness theorem[15] states that no passive linear optical system can increase the light intensity (watts per steradian per unit area per unit bandwidth) of an optical beam. This theory imposes a size limit on the detector for a desired optical routing loss. The actual value of this minimum detector size is dependent on the focusing quality of the optics and on the size of the gate arrays. Figure 8 shows a typical optical computing system that consists of two OE gate arrays. The routing is carried by the holograms and the lens. As shown in Fig. 8, the array size is $N \times N$, the diameter of the lens is D, the focal length of the lens is f, and the diameter of the hologram is d. We define the minimum detector diameter Δ as the resolution of the hologram, which means that >90% power can be picked up by the detector. The diameter Δ can be calculated by

$$\Delta = \frac{2.44 f \lambda}{d}.$$

Substituting d for the diameter of lens D, as $d = D/N$, we have

$$\Delta = 2.44 \lambda \frac{f}{D} N. \tag{3}$$

Equation (3) shows that the detector size is proportional to the array size. If we assume that $f/D = k = 1$, $\lambda = 0.8$ μm, and $N = 32$, the minimum detector diameter is 62.5 μm. Such a large detector-size requirement makes it difficult to realize high fan-in with directly charging OE gates.

B. Fan-out Constraint on Directly Charging Optoelectronic Gates

We define directly charging OE gates as the class of OE logic elements that uses a photocurrent to complete logic switching.[16–18] The typical circuit structure of this class consists of a photodetector and a logic gate, with the photocurrent of the detector charging the gate input capacitor to the switching voltage. The gate's intrinsic delay is determined mainly by the input capacitor charging time, which is usually much longer than the electron transit time inside the transistors.

The switching time of a directly charging gate is related to the time needed to charge the gate input capacitor to the switching voltage, that is,

$$T = VC/I, \tag{4}$$

where V is the switching voltage of the gate, C is the equivalent input capacitance of the OE gate comprising the detector and the transistor's gate-source capacitance, and I is the photocurrent. The switching voltage is a fixed value for a given technology. T is proportional to the ratio of C to I. The most promising high-speed detector seems to be the metal–semiconductor–metal (MSM) detector because it can be fabricated with the same process as that for the metal–semiconductor field-effect transistor integration and it can work at gigahertz speed. Many gigabit per second OE gate arrays with MSM detectors have been reported.[19–21] However, the capacitance of MSM detectors is large because of the minimum size required by the brightness theorem. A typical value is 90 fF for an 80 μm × 80 μm MSM detector.[22] If we assume the transistor gate-source capacitor is 10 fF, the total input capacitance C of an OE gate is ~100 fF. Thus a significant photocurrent is required to charge the detector capacitor when T is of the order of nanoseconds. For example, when $T = 1$ ns, $V = 1$ V, and C = 100 fF, the required photocurrent is 0.1 mA, which can be translated to an optical input power of 0.5 mW, assuming the responsivity of the detector is $\eta = 0.2$.[23]

Regardless of the optical interconnection loss, the input optical power incident upon each destination OE gate is one Nth of the optical output power of the source OE gate, where N is the fan-out number. Thus the minimum optical power requirement of an OE gate linearly increases with the number of fan-outs. As mentioned in Subsection 2.C, an average fan-out of 85 is desired to achieve the theoretical minimum circuit depth, and this number can be as large as 263 for an individual circuit (see circuit rd84 in Table 1). When the fan-out is $N = 85$, the corresponding optical output power of an OE gate is 42.5 mW, without taking the optical routing loss into account. Such high power dissipation may prohibit the integration of even a 32 × 32 OE array on a single

chip. When the optical interconnection loss and the laser quantum efficiency are taken into account, the situation becomes even worse.

C. Feasibility of High-Fan-in and High-Fan-out Optoelectronic Gates with a Built-In Amplifier

One option for solving the fan-out-capability and power-dissipation conflict is the insertion of a high-gain and high-bandwidth analog amplifier between the detector and the logic switching element, as shown in Fig. 4. A transimpedence amplifier converts the weak photocurrent to a voltage swing. The output of the amplifier is then digitized to turn the laser on and off.

An OE gate with a built-in amplifier has a power-dissipation performance superior to that of a directly charging OE gate. The reason is explained with the following example: If we assume an optical output power of the OE gate of 2 mW, an optical routing loss of 50%, and a fan-out of 85, the photocurrent generated in each destination gate is $(2 \times 0.5 \times 0.2)/85 = 2.35$ μA. When the amplifier is designed to have a front-end amplifier with $1k$ transimpedence and is cascaded through three stages of voltage amplifiers with a voltage gain of 10, the output voltage swing is $2.35\ \mu A \times 1\ k \times 10^3 \approx 2.35$ V. Such a voltage swing is large enough to switch the laser driver.

To evaluate the integration feasibility of an OE gate array, we designed a four-stage transimpedence amplifier with Vitesse Semiconductor Corporation's 1-μm HGaAs3 process. There are several reasons for our selection of the GaAs process instead of the standard silicon process. The first is that Vitesse's 1-μm HGaAs3 process is the only accessible fabrication choice for a test chip with gigabit/second detectors and an amplifier array. The second is that GaAs is the most promising current technology for monolithic integration of an OE gate consisting of OE gates and lasers, and the silicon CMOS will be the mainstream technology for electronic circuits. The last reason (it may be arguable) is that we believe Vitesse's 1-μm HGaAs3 technology is not necessarily better than submicro or deep-submicro CMOS technology in terms of large digital system performance. A SPICE simulation shows the power dissipation for each amplifier stage to be <2 mW.

If we assume a 30% electrical-to-optical conversion loss in a laser diode, the 2-mW optical output power requires an electrical power of $2/0.3 = 6.7$ mW, resulting in a total gate power dissipation of 14.7 mW. Since the power dissipation of 1000 OE gates is less than 14.7 W, it is practical to integrate a 32×32 array onto a single chip. The power dissipation can be reduced further if high-speed transistors, such as high electron mobility transfers, and current-mode amplifiers are used.[24]

High detector capacitance is not a serious problem for the speed of the OE gate with a built-in amplifier. The toggling speed of a digital gate is determined by the signal rise–fall time. Unlike the directly changing gates whose rise–fall time is determined by detector capacitance and photocurrent, the rise–fall time of OE gates with built-in amplifiers is a function of the amplifier's bandwidth. The bandwidth of an amplifier is determined by its input RC constant, and the relation between an amplifier's bandwidth and the signal rise time is expressed as[25]

$$T_r = 2\pi RCK, \tag{5}$$

where T_r is the 10%–90% rise time, R is the input resistance of the amplifier, C is the input capacitance of the amplifier, and $K = 0.338$–0.35 is an input signal shape-dependent constant. Although the detector capacitance is still involved in determining the gate operation speed, a short rise–fall time is practical because the R can be designed to be very small. The typical value of R is between 50 and 100 Ω for a transimpedance amplifier. Thus, with a MSM detector of 80 μm × 80 μm at the input end,[26] a rise–fall of less than 50 ps is achievable. Because photocurrent is not involved in determining the speed as it is in directly charging gates and a low-power amplifier is possible, the OE gates with a built-in amplifier have better speed–power performance than the former and are more tolerant of high detector capacitance.

One of the drawbacks of OE gates with a built-in amplifier is the increased intrinsic gate delay because amplifier delay is added. However, this extra amplifier delay is acceptable. According to our SPICE simulation, the 50%:50% delay from the input to the output of a four-stage cascaded amplifier is ~320 ps. With a detector delay of a few tens of picoseconds and a laser switching delay of 100–200 ps, a 500-ps intrinsic gate delay is feasible. As shown in Fig. 8, the OE circuit produced by a NOR gate with a 500-ps intrinsic delay is 2 times faster than a 0.7-μm CMOS circuit.

Low density is another drawback of OE gates with a built-in amplifier. However, low density may not necessarily be a critical issue that determines the usefulness of optoelectronic integrated circuits (OEIC's). First, OEIC's suffer less circuit degradation from the low circuit density as a result of the speed-of-light signal-propagation delay. Second, the circuit density can be increased by use of the advances in semiconductor technology. For example, when the cutoff frequency of a transistor is 150 GHz, a voltage gain of 30 is practical for a single amplifier stage. Instead of four stages of amplifiers, two stages of amplifiers are enough to amplify the input optical signal to the required voltage level at the output of the second amplifier. We also argue that there will be sufficient semiconductor real estate to permit an amplifier in each OE gate. It is a constraint that VLSI designers have had to consider in the past; now, however, most researchers think in terms of "extra silicon is free." We hope that the same thing will happen with OEIC's.

Table 3. Circuit-Delay Comparison between 0.7-μm CMOS and OE Circuits

	0.7-μm CMOS		OE Implementation		
Circuit	Delay (ns)	Gate Count	Delay (ns)	Gate Count	Speedup (%)
5xp1	3.43	127	2.00	171	171
b12	2.28	82	1.33	110	171
bw	3.37	183	1.33	255	253
clip	4.52	177	2.00	346	226
con1	1.57	25	1.33	26	118
duke2	3.81	576	2.67	630	143
ex5	3.46	477	2.00	591	173
inc	3.55	113	2.00	124	177
misex1	2.92	74	1.33	78	219
misex2	2.7	169	2.00	90	130
misex3c	4.73	427	2.67	666	1.77
rd53	2.44	62	2.00	76	1.22
rd73	4.12	133	2.00	314	2.06
rd84	3.97	148	2.67	620	1.49
sao2	3.23	135	2.67	211	1.2
squar5	3.3	68	1.33	70	2.47
average	3.34	186	1.96	274	170

4. Delay Performance Comparison of Optoelectronic and Electronic Combined Circuits

We compare the best-case delay performances of OE and electronic implementations of combined benchmark circuits using the CAD tool SIS-II for circuit-delay optimization. The electronic implementation makes use of 0.7-μm L_{eff} (effective channel length) CMOS standard cells, which include inverters, AND's, NAND's, OR's, NOR's, XOR's, XNOR's, AOI's, and OAI's. The intrinsic gate delay and fan-out delay parameters of individual cells are from the Motorola data book.[27] The OE implementation makes use of NOR and OR gates with a fan-in of 8 and a maximum fan-out of 85. In OE simulation we assume the delay between two gates to be 667 ps, composed of a 500-ps gate intrinsic delay and a 167-ps optical signal-propagation delay. These are conservative assumptions since the 500-ps gate intrinsic delay is feasible for an OE gate with a fan-out of 85, as discussed in Subsection 3.C, and the 167-ps connection delay corresponds to a physical signal path length of 5 cm.

Table 3 and Fig. 9 show the simulated circuit delays of electronic and OE implementations for each benchmark. The arithmetic mean of the circuit delay is represented in the last row of Table 3. This value is 3.34 ns for 0.7-μm CMOS circuits and 1.96 ns for OE circuits. This is an average speed-up of 170% for OE circuits over their electronic counterparts. OE circuit delay can be improved further with reductions in both gate intrinsic delays and connection delays. Optimization of the connection delay can be achieved if the optical interconnect length is shortened. The amount of connection delay is a function of the detailed optical system architecture and is not discussed further here. The impact of the intrinsic delay of an OE gate on the circuit delay is shown in Fig. 9. The expected circuit delay is represented by filled triangles. The solid and dashed horizontal curves represent the best-case circuit delay of two electronic circuits based on 2- and 0.7-μm CMOS technologies. Figure 9 shows that the circuit delay scales down along with the OE's intrinsic delay. It also indicates that the OE circuits can be faster than submicrometer CMOS circuits.

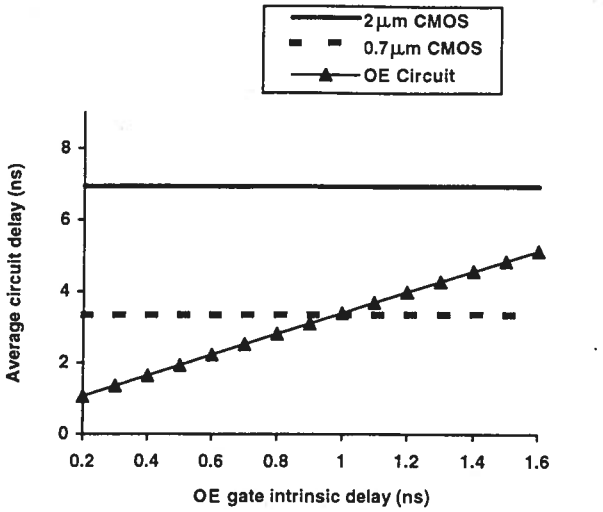


Fig. 9. OE circuit delay and gate intrinsic delay.

5. Role of Optoelectronic Circuits in Computer System Architectures

To evaluate the impact of OE circuits on the computing architecture, let us first look at the relative strengths of OE circuits over electronic circuits from a system designer's viewpoint. The four strengths of OE circuits are described below.

A. Reduced Logic-Operation Delay

As we saw in Section 2, circuit depth can be reduced by exploitation of the high-fan-in and high-fan-out ability. In Section 3, we show that an intrinsic delay of a few hundreds of picoseconds is achievable. Thus it is possible for OE circuits to perform certain logic operations with less delay than that encountered with electronic circuits. According to the simulation of benchmark circuits, there is an average speedup of 170% for OE circuits over 0.7-μm CMOS circuits, which we discuss in Section 4.

B. Reduced Global Interconnection Delay

Here we define the global interconnection delay as the signal-propagation latency between two function units. The flexibility of the three-dimensional interconnection and the superiority of the speed-of-light propagation delay provided by the OE gates' optical input and output make the delay in OE circuits much less sensitive to the physical length of the interconnection than in the case of electrical interconnection.

In the case of a pure electronic computer, the connection delay is the main constraint on system speed. The increased chip size and the limitations of a two-

Table 4. Power Distribution of Function Units in a Processor

Processor	Speed	Distribution Areas: Input-Output (%)	Cache (%)	Controller (%)	Clock (%)	Function Unit (%)	Interconnection Units (%)	Total
DEC VAX	100 MHz	8.0	22.7	19.7	18.4	17.2	13.5	16.3 W
MIPS	300 MHz	17.8	48.8	6.2	12.0		4.8	115 W

dimensional connection are two major restricting factors. According to the CMOS scaling theory,[28] the intrachip connection delay per unit length is determined by the RC value of the metal trace instead of by the electrical signal-propagation speed. For example, the capacitance of an on-chip metal trace is $C_{wire} = \varepsilon_{ox} WL/t_{ox}$, where W and L are the length and width of the trace, respectively, and ε_{ox} and t_{ox} are the permittivity and the thickness of the field oxide (SiO_2), respectively. For a typical 0.8-μm process (with $W = 1.2$ μm, $t_{ox} = 0.33$ μm), the capacitance is approximately 1 pF/cm. The output resistance of the gate at the switching point is calculated as $R = V_{0.5}/I_D$, where $V_{0.5}$ is 50% of the logic high voltage (power supply voltage) and I_D is the constant drain of current by the transistors. If we assume values of $I_D = 1$ mA (N-type metal-oxide semiconductor field-effect transistor with a channel width of 10 μm and a channel length of 1.0 μm) and $V_{0.5} = 2.5$ V, then the connection delay is 2.5 ns/cm, a factor of 75 larger than the optical connection delay. Thus the length of the interconnection becomes the primary driving force of the VLSI circuit.

Although the minimized transistor size reduces the connection lengths of a functional unit, the connection lengths between units tend to increase. This occurs as more and more functional units are integrated onto a single chip. The integration scale is so large these days that it overtakes the rate of the downscaled components; the chip size turns out to be bigger and bigger. As a result, the global connection delay has now become the bottleneck in the system clock rate, as was expected 10 years ago.[29]

Two-dimensional layout imposes another constraint on global connection lengths. The clock fan-in tree is usually located at the center of a microprocessor chip to reduce clock skew. Thus the other function circuits have to be located around the clock fan-out tree. In the worst case, this path length can be as large as the sum of the chip edges when two functional circuits are at the opposite corners. The situation becomes even worse when the fan-out number increases and the interconnection must be implemented in parallel, as is the case with data and control buses used to connect ALU's, register files, caches, and their interface circuits. Multiple metal wires have to be used to interconnect these functional units, which are located at different positions. The wire capacitance is huge compared with the gate input capacitance.

The connection of the OE circuit is determined by the speed of light. As we assumed in the fan-in and fan-out discussion, i.e., that the optical-to-electrical or electrical-to-optical conversion delay was included in the OE gate intrinsic delay, the interconnection delay for the OE circuit includes only the signal-propagation time from the source to its destination. When the average distance between two OE functional units is assumed to be 5 cm, the average connection delay between the units is only 167 ps as a result of an optical signal-propagation speed of 33 ps/cm. On the other hand, the intrachip connection delay of 0.7-μm CMOS circuits is approximately 2.5 ns/cm. A size of a few centimeters squared is very common nowadays for high-end microprocessor chips, which implies an interconnection delay between functional units of a few nanoseconds for an electrical connection. The importance of the small global connection delay is multifold: First, we wish to increase system clock rate; second, it decouples the clock rate and the connection length requirements, allowing more room for exploiting the spatial parallelism; finally, it provides a solution to the problem of relieving the power-dissipation constraint.

C. Relieved Power-Dissipation Constraint

Power dissipation is a fundamental constraint on the speed of an electronic computer. Downscaled VLSI technology relies on the reduction of wire length to reduce the connection delay. This approach results in an increased circuit density and therefore increased power density. For reliability and packaging considerations, there is a limit on the power density and the maximum power dissipation of a single chip. Table 4 lists the power distribution of units in two typical processors.[30,31] It shows that only 5% to 30% of the total power is related directly to ALU's. Most of the power is consumed by the functional units that handle global interconnection, such as the cache manager, the input–output drivers, the clock fan-out tree, and the instruction decoder. This power distribution indicates that it is the intrachip global interconnection that slows down the system clock rate. Today, even in the personal computer the processor must be cooled by fans mounted directly on it.

The highest power dissipation of a signal chip processor is 115 W (see Table 4). With such high power dissipation, a massively parallel computing system based on the VLSI chip faces the challenge of chip cooling. The strength of a high-fan-in and high-fan-out OE circuit is its tolerance to a longer physical interconnect length. This strength makes OE circuits

able to implement the parallel architectures much more easily than can its pure electronic counterpart.

D. Potential for Reduced Hardware Complexity

Hardware complexity savings is a potential strength of gate-level optically interconnected OE circuits. The speed–power advantage of optical interconnects has been studied by many researchers in the past.[32] Ozaktas and Goodman[33] studied the implications of optical interconnection for high-speed digital computing. Feldman[34] calculated the power-consumption break-even length for optical interconnects versus electrical wires.

E. Reduced Hardware Complexity from Simultaneous Logic Operation and Optical Interconnection

We now illustrate one feature of optoelectronics not yet stressed: that the gate-level optically interconnected OE, NOR, and OR gates can perform both logic operation and optical interconnection at the same time. This combination of logic operation and interconnection not only provides an optimized system lock rate but may also possibly lead to a reduction in hardware complexity because no extra hardware is required to implement the interconnect between grains.

Let us compare the hardware complexities of two implementations: One is a hybrid of an electronic circuit for logic operation and optics for interconnection, and the other is a gate-level optically interconnected OE circuit. The hardware complexity is defined as the total number of gates needed to perform both logic operations and interconnects. In the following comparison we use well-accepted power–speed criteria to partition the hybrid implementation.

According to Feldman,[34] the power–speed performance of the optical interconnection is better than that of the electrical connection when the connection length is >1 mm for a single fan-out. This break-even connection length is even shorter when the modulation speed goes higher, when the fan-out increases, or both. Considering the gigahertz speed and the high-fan-out applications, we assume the electronic circuits are partitioned into 0.5 mm × 0.5 mm grains. In such an area, 500 equivalent two-input NAND gates can be integrated by use of a 0.7-μm CMOS gate-of-sea design. By Rent's rule,[35] the number of external interconnects to or from other grains is kN^p, where k is the average number of inputs and outputs for each gate, p is an empirical constant with typical values of 0.5–0.7, and N is the total gate count of the grain. If the average fan-out is 2, fan-in plus fan-out is $k = 4$. In the worst case the required external interconnection number is $4 \times 500^{0.7} = 310$. Since each interconnect requires at least one laser–detector pair with the driver circuit, the hardware complexity for external connection is equal to 310 OE gates. Thus the hardware complexity of the hybrid implementation is 810 gates, the same as 310 OE gates and 500 electronic gates.

All inputs and outputs are optical in the OE implementation. The external connection complexity is zero. Moreover, it is possible to realize the logic of 500 two-input NAND gates with 310 NOR and OR gates for certain functions by use of the OE circuit. One example is shown in Fig. 3, where the functions of 11 two-input AND and OR gates can be performed by a total of four NOR and OR gates.

6. Application Examples

In this section, we use three examples to evaluate the speed potential of high-fan-in and high-fan-out OE circuits at the circuit, functional-unit, and system levels.

A. Cache Memory Address Decoder and Comparator

Cache memory access time is the fundamental bottleneck in computer speed. The address-decoder and comparator delays are two dominant contributors to access time. The decoder delay is caused by the circuit depth of the decoding logic and heavy fan-out. Since the output of the address decoder, also called the word line, fans out to a block of memory cells, the output capacitance causes a significant fan-out delay. Wada *et al.*[36] calculated the optimized address-decoder delay on the basis of the 0.8-μm CMOS generic process with the gate intrinsic delay of 122 ps. His calculation shows the address-decoder delay of a 32K cache memory is approximately 5–8 ns. The circuit depth of the decoding logic is $\log_{fin}(W)$, where fin is the fan-in of gates and W is the width of the address in bits. With a fan-in of 8, the OE address decoder can decode a 64-bit-wide address with a circuit depth of 2, corresponding to a decoding delay of less than 1.4 ns.

Comparators are basic components in cache manager circuits for hit detection. The problem of an electronic comparator is the fan-in capacitance and the circuit depth of comparing logic. Since all the outputs of the same tag bit are connected to the comparator, the fan-in capacitance is large—thereforethe fan-in delay. The logic expression to compare two n-bit words, A and B, is

$$f = (\overline{a_0} \cdot \overline{b_0} + a_0 \cdot b_0) \cdot (\overline{a_1} \cdot \overline{b_1} + a_1 \cdot b_1) \cdots (\overline{a_{n-1}} \cdot \overline{b_{n-1}} + a_{n-1} \cdot b_{n-1}),$$

or

$$f = \overline{\left(\overline{(a_0 + b_0)} + \overline{(\overline{a_0} + \overline{b_0})}\right)} + \overline{\left(\overline{(a_1 + b_1)} + \overline{(\overline{a_1} + \overline{b_1})}\right)} + \cdots + \overline{\left(\overline{(a_{n-1} + b_{n-1})} + \overline{(\overline{a_{n-1}} + \overline{b_{n-1}})}\right)},$$

where $A = a_{n-1}a_{n-2} \times \ldots \times a_1a_0$ and $B = b_{n-1}b_{n-2} \times \ldots \times b_1b_0$. As the memory space be-

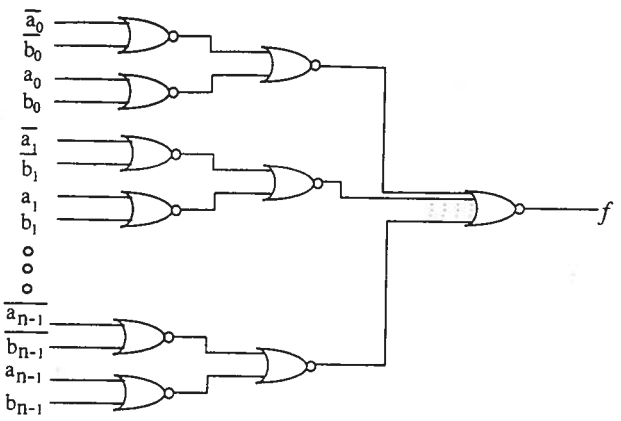


Fig. 10. OE comparator.

comes bigger and bigger, the width of the tag file increases. The limited fan-in number of electronic gates again tends to increase the circuit depth of the comparator. When the width of the tag is equal to or less than 32 bits and there is an OE gate fan-in of 8, an OE comparator has a circuit depth of 3, corresponding to a delay of <2 ns. Figure 10 shows the logic circuit of the OE comparator.

B. Microinstruction Decoder

With holograms that store next-state and microinstruction output, OE NOR gates can also be used to construct finite-state machines, as with the OE controller proposed by Heuring and Morozov (see Ref. 1). Figure 11 shows the circuit of a microinstruction decoder. Dual-rail instruction codes and addresses are the inputs of the first-stage NOR gate array (at the left side). The outputs of the first stage are routed to the second NOR gate array, where the output of each gate represents a MIN term. The output of the second array illuminates the hologram array on the right side to generate the next microinstruction address and control word. The individual bits of the control word are routed to appropriate devices, while the next microinstruction addresses are fed back to the first stage to select the next microinstruction. For simplicity the timing circuit is not shown in Fig. 11. The timing can be realized with embedded latches in NOR gates or by optical clock gates.

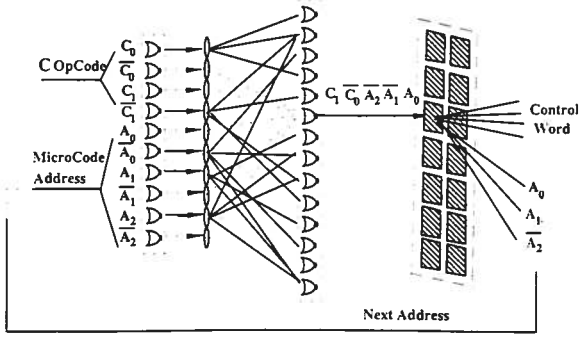


Fig. 11. OE microinstruction decoder.

We have demonstrated a 300-MHz counter using four OE NOR gates and optical clock gating that has a structure that is simple but similar to that shown in Fig. 11 (see Ref. 12). Since the NOR's are built from discrete detectors, amplifiers, and lasers, the intrinsic gate delay is comparatively large at 1.1 ns. The gate delay can be optimized with an integrated version of the NOR gate array. As the circuit depth is only 2 for an OE instruction decoder, it is practical to push the clock rate to 500 MHz and beyond.

C. Hybrid Electronic–Optoelectronic Superscalar Computer

The impact of the high-fan-in and high-fan-out OE circuit on system performance is evaluated with a hybrid computer. The proposed computer consists of electronic-version ALU's and an OE-version cache manager, bus controller, and instruction decoder. The electronic parts also include photodetectors and lasers for optical interconnection. The system clock rate is the only criterion in the evaluation.

For simplicity, we assume the electronic ALU's consist of pipelined integer and floating-point units. In our example in Subsection 6.A. of application circuits we saw that the decoder, the cache manager, and the interrupt controller could be constructed with a circuit depth <3, corresponding to an average circuit delay of 1.3–1.8 ns. If all components are assembled within a volume of 10 cm × 10 cm × 10 cm, the longest connection delay between units is $<\sqrt{3} \times$ 10 cm/30 cm ns = 0.577 ns. Thus the maximum delay is 2.377 ns between the inputs to the OE circuits and the inputs of the destination unit.

In the electronic part, it is practical to assign the pipeline clock cycle to 2.377 ns when the depth of the pipeline is 2 for the integer unit and 4 for the floating-point unit. A 1.2-ns delay, 32-bit integer adder[37] and a 9-ns floating-point multiplier[38] have been built. The maximum clock rate is 420 MHz for the proposed system. The peak performance is 420 million instructions per second (MIPS) and 420 million floating-point instructions per second (MFLOPS). As for the system architecture, the superscalar computer is preferred by the OE circuit because it exploits spatial parallelism, the strength of an OE circuit. For example, when the ALU number is increased from 2 to 4, the throughput will be linearly improved, although the connection length may be extended to accommodate more circuits. If we assume the volume is also doubled, i.e., to 20 cm × 20 cm × 20 cm, the worst-case interconnection delay is $\ll\sqrt{3}$ × 20 cm/30 cm ns = 1.154. At this time the clock cycle becomes 2.954 ns, which represents a degradation of 24%. However, this degradation will be overcompensated for by the increased parallelism. Since the number of processors is doubled, the peak performance is 677 MIPS and 677 MFLOPS.

7. Conclusion

The high-fan-in and high-fan-out capability of OE gates can be exploited to reduced circuit depth. Simulation shows that the optimal fan-in and fan-out

limits of an OE gate are 8 and 85, respectively, to take full advantage of circuit-depth reduction. The integrated high-fan-in and high-fan-out OE gate is feasible for current semiconductor technology. A 500-ps intrinsic OE gate delay is achievable on the basis of current OE integrated technology. Compared with 0.7-μm CMOS electronics, the circuit delay in OE implementation can be 1.7 times smaller than its electronic counterpart. The interconnection delay of an OE circuit is less sensitive to its physical length. This interconnection delay advantage is desired for optimizing system architecture by exploitation of spatial parallelism. The small interconnection and circuit delays also allow the replacement of a larger high-power VLSI chip with several small low-power OE circuits without sacrificing speed. Several applications for OE circuits have been proposed and evaluated at the circuit, function-unit, and system levels. Theoretical analysis and experimental results show that a 420-MHz clock rate is practical.

References

1. V. P. Heuring and V. N. Morozov, "Optically controlled digital optical matrix processor," in *Photonics for Computers, Neural Networks, and Memories*, W. J. Miceli, J. A. Neff, and S. T. Kowel, eds., Proc. SPIE **1773,** 201–207 (1993).
2. G. C. Marsden, A. V. Krishnamoorthy, S. C. Esener, and S. H. Lee, "Dual-scale topology optoelectronic processor," Opt. Lett. **16,** 1970–1972 (1991).
3. P. S. Guilfoyle, R. S. Rudokas, R. V. Stone, and E. V. Roos, "Digital Optical Computer II (DOC II): 'Performance Specifications'," in *Optical Computing*, Vol. 6 of 1991 Technical Digest Series (Optical Society of America, Washington, D.C., 1991), pp. 203–206.
4. T. Main, R. J. Feuerstein, and H. F. Jordan, "Implementation of a general purpose stored program digital optical computer," Appl. Opt. **33,** 1619–1623 (1994).
5. V. N. Morozov, "Parallel optoelectronic computer: An example of high efficiency free-space utilization of global interconnects," Tech. rep. 94-15 (Optoelectronic Computing Systems Center, University of Colorado at Boulder, Boulder, Colo., 1994).
6. E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. R. Stephan, R. K. Brayton, and A. Sangiovanni-Vincetelli, "SIS: A system for sequential circuit synthesis," UCB/ERL report M92/41 (University of California, Berkeley, 1992).
7. H. Savoj, H. Y. Wang, and R. K. Brayton, "Improved scripts in MIS-II for logic minimization of combinational circuits," in *Proceedings of IEEE/ACM International Workshop on Logic Synthesis* (Institute of Electrical and Electronics Engineers, New York, 1991).
8. S. Yang, *Logic Synthesis and Optimization Benchmarks User Guide, Version 3.0* (MCNC, Research Triangle Park, N.C., 1991).
9. H. Touati, H. Savoj, and R. K. Brayton, "Delay Optimization of Combinational Logic Circuits by Clustering and Partial Collapsing," in *Proceedings of the IEEE International Conference on Computer-Aided Design*, S. Goto and L. Trevillyan, eds. (IEEE Computer Society, Los Alamitos, Calif., 1990).
10. B. Sugla and D. A. Carlson, "Extreme area-time trade-off in VLSI," IEEE Trans. Comput. **39,** 251–257 (1990).
11. K. J. Singh and A. Sangiovanni-Vincenteli, "A heuristic algorithm for the fanout problem," *Proceedings of the Design Automation Conference* (Institute of Electrical and Electronics Engineers, New York, 1990), pp. 357–360.
12. V. P. Heuring, L. H. Ji, R. J. Feuerstein, and V. N. Morozov, "Toward a free-space parallel optoelectronic computer: a 300-Mhz optoelectronic counter using holographic interconnects," Appl. Opt. **33,** 7579–7587 (1994).
13. C. W. Stirk, "Bit error rate of optical logic: fan-in, threshold, and contrast," Appl. Opt. **31,** 5632–5641 (1992).
14. V. Morozov, J. Neff, A. Fedor, and H. J. Zhou, "Analysis of a 3-D Computer Optical Scheme with Bi-Directional Interconnects," in *Optical Computing*, Vol. 10 of 1995 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1995), pp. 49–51.
15. W. T. Welford and R. Winston, *The Optics of Nonimaging Concentrators* (Academic, New York, 1978).
16. J. I. Song, Y. H. Lee, J. Y. Yoo, J. H. Shin, A. Scherer, and R. E. Leibenguth, "Monolithic arrays of surface emitting laser NOR logic devices," IEEE Photon. Technol. Lett. **5,** 902–904 (1993).
17. D. Z. Tsang, "Free-space board-to-board optical interconnections," in *Optical Enhancements to Computing Technology*, J. A. Neff, ed., Proc. SPIE **1563,** 66–71 (1991).
18. A. H. Sayles and J. P. Uyemura, "An optoelectronic CMOS memory circuit for parallel detection and storage of optical data," IEEE J. Solid-State Circuits **26,** 1110–1115 (1991).
19. R. Mactaggart, M. Bendett, and S. S. Tatlor, "A complete 400-Mb/s burst-mode data OEIC receiver," IEEE J. Solid-State Circuits **28,** 1018–1022 (1993).
20. A. Yariv, "Semiconductor photodiodes," in *Optical Electronics* (CBS College Publishing, New York, 1985), pp. 367–376.
21. S. Kawanishi, Y. Yamabayashi, T. Takada, H. Takara, M. Saruwatari, and K. Nakagawa, "2 Gb/s operation of an optical-clock-driven monolithically integrated GaAs D-flip-flop with metal-semiconductor-metal photodetectors for high-speed synchronous circuits," IEEE Photon. Technol. Lett. **4,** 160–162 (1992).
22. Y. H. Lee, J. I. Song, M. S. Kim, C. S. Shim, B. Tell, and R. E. Leibenguth, "Active optical NOR logic device using surface-emitting lasers," IEEE Photon. Technol. Lett. **4,** 479–482 (1992).
23. O. Wada, H. Hamaguchi, M. Makiuchi, T. Kumai, M. Ito, K. Nakai, T. Horimatsu, and T. Sakurai, "Monolithic four-channel photodiode/amplifier receiver array integrated on a GaAs substrate," J. Lightwave Technol. **LT-4,** 1694–1703 (1986).
24. C. Toumazou, F. J. Lidgey, and D. G. Haigh, "Analogue IC design: The current node approach," in *IEE Circuit and System Series* (Peter Peregrinus Ltd., London, 1990), Vol. 2.
25. H. W. Johnson and M. Graham, "A note about 3-dB and RMS frequencies," in *High-Speed Digital Design* (Prentice-Hall, New Jersey, 1993), pp. 8–10.
26. E. Sano, "A device model for metal-semiconductor-metal photodetectors and its applications to optoelectronic integrated circuit simulation," IEEE Trans. Electron Devices **37,** 1964–1968 (1990).
27. *H^{4}C Series Design Reference Guide* (Motorola, Schaumberg, Ill., 1995).
28. N. H. E. Weste and K. Eshraghian, "Scaling of MOS-transistor Dimensions," in *Principles of CMOS VLSI Design* (Addison-Wesley, Reading, Mass., 1993), Chap. 4.13, pp. 250–256.
29. S. Kohyama, *Very High Speed MOS Devices* (Oxford U. Press, New York, 1989).
30. R. W. Badeau, R. L. Bahar, D. Bernstein, L. L. Biro, "A 100-Mhz macropipelined vax microprocessor," IEEE J. Solid-State Circuits 27, 1585–1597 (1992).
31. N. P. Jouppi, P. Boyle, J. Dion, "A 300-Mhz 115-W 32-b Binary ECL Microprocessor," IEEE J. Solid-State Circuits **28,** 1152–1165 (1993).
32. D. A. B. Miller, "Optics for low-energy communication inside digital processors: quantum detectors, sources, and modulators

as efficient impedance converters," Opt. Lett. **14,** 146–14 8 (1989).
33. H. M. Ozaktas and J. W. Goodman, "Implications of interconnection theory for optical digital computing," Appl. Opt. **31,** 5559–5567 (1992).
34. M. R. Feldman, "Optical interconnections for VLSI computational system using computer generated holography," Ph.D. dissertation (University of California at San Diego, San Diego, Calif., 1989), Chap. 2.
35. R. L. Landman and R. L. Russo, "On a pin versus block relationship for partitions of logic graphs," IEEE Trans. Comput. **20,** 1469–1479 (1971).
36. T. Wada, S. Rajan, and S. A. Przybylski, "An analytical access time model for on-chip cache memories," IEEE J. Solid-State Circuits **27,** 1147–1155 (1992).
37. M. Susuki, M. Ohkubo, and T. Shinbo, "A 1.5-ns 32-b CMOS ALU in Bouble pass-tran," IEEE J. Solid-State Circuits **28,** 1145–1149 (1993).
38. L. R. Tate, R. J. Niescier, A. C. Hu, J. Scorzelli, W. Leung, C. H. Tzinis, P. J. Robertson, and A. Baca, "32 bit GaAs HFET IEEE float point multiplier," in *Proceedings of IEEE GaAs IC Symposium* (Institute of Electrical and Electronics Engineers, New York, 1992), pp. 85–88.