

UDC 004.724.2 + 004.272.43

 10.25209/2079-3316-2026-17-3-363-414


Parallel Group Operations in Non-Blocking System Area Networks

Viktor Sergeevich **Podlazov**

V. A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russia

 podlazov@ipu.ru

Abstract. This paper considers non-blocking optoelectronic networks in which group operations over a set of numbers are executed during the transmission of one number. A method is proposed to execute group operations using photonic units placed at each multiplexer of the network, along with the circuitry of such executive multiplexers. The new possibility of executing group operations, in parallel across these multiplexers for given sets of users, is described, and procedures for their selection are defined. A bit-by-bit synchrony theorem of group operations is proven. The possibility of designing a non-blocking electronic system area network for any number of users is presented, as well as the possibility of designing a similar optoelectronic network as a tandem network with optoelectronic switches controlled by their analogs from the electronic network. (*Linked article texts in English and in Russian*).

Key words and phrases: non-blocking system networks, distributed group operations, photonic and optoelectronic switches and system networks, photonic multiplexers, tandem non-blocking networks

2020 *Mathematics Subject Classification:* 65Y05; 68Q10

For citation: **Viktor S. Podlazov**. *Parallel Group Operations in Non-Blocking System Area Networks*. Program Systems: Theory and Applications, 2026, **17**:3(72), pp. 363–414. (*In English, in Russian*). https://psta.psir.ru/read/psta2026_3_363-414.pdf

Introduction

In modern computing systems based on system area networks (SANs) being developed, data are transmitted together with commands, forming spatial data flow structures [1]. A particular case is computing systems with distributed parallel data processing during their transmission in the network: a group operation over a set of numbers is executed while one number (the result of this operation) is transmitted via a common channel of the network. A channel is common if users can transmit signals into this channel in parallel with their overlapping, which can be used for “on-the-fly” group computations. A common channel is direct if it has only transmission delays of signals or their combination delays, without any clock delays.

Group operations in a direct channel are called common channel computing (CCC) [2] and can be executed without intermediate buffering (on-the-fly). The simplest examples of CCC operations are summing a set of numbers from different nodes or finding the maximum of them.

CCC operations can be executed during data transmission via optical signals. Special executive units are required for various operations at each network node. They can be built in the element base of optical channel switches in a purely photonic implementation (without electronic control of the units) [3, 4].

Network nodes and executive units exchange optical signals at the highest possible frequency, and the executive units operate only with these signals. The executive units enable arithmetic and logical operations over numbers represented by optical signals without the use of special photonic ALUs as new physical devices [5]. This is the main difference from global practices in the design of photonic processors: computations are organized only using photonic switches based on the two-rail representation of optical signals.

CCC operations can be executed in cyclic sequences of network nodes forming Hamilton graphs [4]. They can be built in a non-blocking SAN with direct channels [4, 6], which has $N_i = m_i$ users and contains a layer of full $m \times m$ switches and $i - 1$ layers of $1 \times m$ demultiplexers and $m \times 1$ multiplexers.

A non-blocking optical network can be created using tandem switching devices consisting of similar electronic and optoelectronic components, as described in [6]. The former are involved in routing and route fixing and determine the configuration of the latter.

Tandem switching devices shall have an identical configuration and operate in the master–slave mode for electronics and optics, respectively. This yields a tandem non-blocking network consisting of duplicate switches for the electronic and optoelectronic networks. Photonic executive units can be placed at the nodes of the second network to execute CCC operations with higher speed and higher noise immunity.

The CCC operations in [4] are executed in a serial-parallel format. First, a Hamilton graph is constructed for a given group of users by transmitting a sequence of their addresses over the network. The operation itself is then executed by sending operands to the executive units in parallel. Due to the serial part of the operation, its execution is accelerated only twofold compared to the conventional method of network addition using a binary tree [4], without accounting for software execution delays.

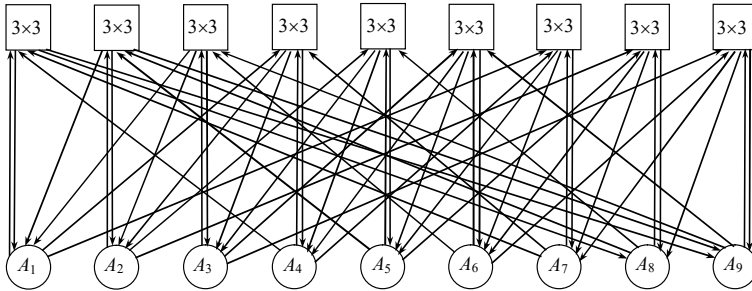
In this paper, we fully parallelize the execution of CCC operations in a non-blocking SAN, thereby significantly reducing their execution time.

To describe the operations, we will distinguish between an organizing user (organizer) and a group of participating users (participants). In a non-blocking SAN, the organizer and participants from any group are connected by trees of direct channels. They consist of $1 \times m$ demultiplexers on the path from the users to the switches and $m \times 1$ multiplexers on the return path to the users.

To execute a CCC operation, conventional multiplexers can be replaced with m -port executive multiplexers consisting of m serially connected executive units. The ports of the executive multiplexers receive the numbers for processing, which are transmitted serially and synchronously bit by bit.

This synchronization is based on the possibility of pre-measuring the signal propagation time between any two users in both directions [7]. (In a non-blocking SAN, these times may differ.) Synchronization is ensured by participating users: they delay the transmission of numbers for different amounts of time, considering the signal propagation delays from the organizer to the executor and back. These delays shall be pre-measured for any source-sink pairs based on the IEEE 1588 standard [8].

The new results presented below consist in parallelizing the stages of a group operation: addressing a group of participants by sending multiple group addresses; the bit-by-bit synchronization of number transmission by all group participants, with the bit-by-bit synchronization of the numbers; and the parallel realization of an operation in the executive multiplexers of a tree subnet.

FIGURE 1. A quasicomplete digraph with $m = 3$

Section 1 describes the design of a non-blocking SAN with direct channels and its optoelectronic implementation. For the sake of completeness, we recall some results from [4, 6] to introduce the basic concepts and justify the possibility of tree channels between the organizer of a group operation and its participants. A novel result of Section 1, presented for the first time, is a bit-by-bit synchronization method for numbers transmitted by group participants, in network multiplexers. In Section 2, we provide the structure of a common channel for executing CCC operations in a photonic implementation. Here, some background from [3, 4] is added to justify the circuitry of Section 3. Next, Section 3 describes the structure of the executive multiplexers and the parallel implementation of synchronous group CCC operations on them in a nonblocking SAN, based on the bit-by-bit synchronization theorem for transmitted numbers from Section 1. In Section 4, we estimate the parallel execution times of group operations and compare them with the serial-parallel execution scheme [6]. Finally, Section 5 proposes a methodology for building non-blocking optoelectronic networks based on Mach-Zehnder interferometers for the use of photonic executive units, which partially repeats a corresponding section of [6].

1. Non-blocking SAN based on the Quasicomplete Graph Topology

In this section, SANs are built in the form of non-blocking, self-routing distributed switches based on switches with the quasicomplete digraph topology. For these networks, the number of users can be scaled to any degree through internal cascading. A quasicomplete digraph can be represented as a bipartite digraph with $N_1 = m^2$ nodes in each part (Figure 1). The connections in it are defined by different incidence tables for arcs from users and to users (Table 1). They exist for any m .

TABLE 1. The incidence table for a quasicomplete digraph with arbitrary $m \pmod{N_1}$

Switches	Channels from users				Channels to users			
1	1	2	...	m	1	$1 + m$	...	$1 + m(m - 1)$
2	1	2	...	m	2	$2 + m$	...	$2 + m(m - 1)$
...	1	2	...	m	...	...	...	...
m	1	2	...	m	m	$m + m$	...	$m + m(m - 1)$
$m + 1$	$m + 1$	$m + 2$	...	$m + m$	1	$1 + m$	...	$1 + m(m - 1)$
$m + 2$	$m + 1$	$m + 2$	...	$m + m$	2	$2 + m$	...	$2 + m(m - 1)$
...	$m + 1$	$m + 2$	...	$m + m$	...	...	...	...
$m + m$	$m + 1$	$m + 2$	...	$m + m$	m	$m + m$	...	$m + m(m - 1)$
...	...	...	...	...	...	...	...	...
$m^2 - m + 1$	$m_2 - m + 1$	$m_2 - m + 2$	...	m_2	1	$1 + m$	...	$1 + m(m - 1)$
...	$m^2 - m + 1$	$m_2 - m + 2$	...	m_2	2	$2 + m$	...	$2 + m(m - 1)$
$m^2 - 1$	$m_2 - m + 1$	$m_2 - m + 2$	...	m_2	...	...	...	...
m^2	$m_2 - m + 1$	$m_2 - m + 2$	...	m_2	m	$m + m$	...	$m + m(m - 1)$

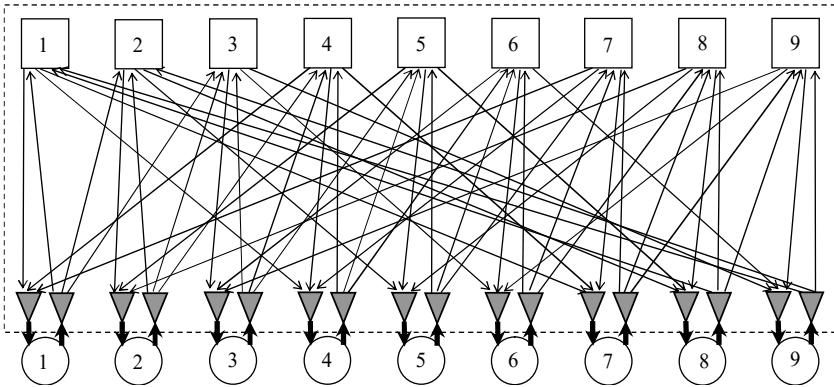


FIGURE 2. Switch $S_1(9, 3)$ with the quasicomplete graph topology with $m = 3$

To connect 1-port users, $1 \times m$ demultiplexers and $m \times 1$ multiplexers are used, respectively (Figure 2). This yields a non-blocking, self-routing distributed switch $S_1(N_1, m)$. In this switch, routing is performed from the source based on the addresses (numbers) of the output ports of the demultiplexers and switches. Regardless of the order of packets, conflict-free packet transmission is carried out via direct channels in a single hop from the sources to the sinks, i.e., with maximum speed.

The internal scaling method [6, 9] allows one to build, from non-blocking switches $S_i(N_i, m)$ with $N_i = m_{i+1}$, a non-blocking switch $S_{i+1}(N_{i+1}, m)$ with $N_{i+1} = m_{i+2}$. The switch $S_{i+1}(N_{i+1}, m)$ is obtained from N_1 switches $S_i(N_i, m)$ and a layer consisting of N_{i+1} $1 \times m$ demultiplexers and $m \times 1$ multiplexers. Each switch $S_i(N_i, m)$ is split into parts with m ports, and in each part, a separate switch $S_1(N, m)$ is built across all switches; the latter switch has the original (local) numbering in each part and end-to-end numbering across all parts. An $m \times m$ switch is denoted by $S_0(m, m)$. Tables 2–4 provide examples of incidence tables for switches $S_1(4, 2)$, $S_2(8, 2)$, and $S_3(16, 2)$ based on the incidence tables for switches $S_0(2, 2)$, $S_1(4, 2)$, and $S_2(8, 2)$, respectively.

Figures 3, 4 and 5 show the diagrams of the corresponding switches as digraphs.

TABLE 2. The incidence table for a switch $S_1(4, 2)$

$S_1(4, 2)$	Arcs from users		Arcs to users	
1	1	2	1	3
2	1	2	2	4
3	3	4	3	1
4	3	4	4	2

 TABLE 3. The incidence table for a switch $S_2(8, 2)$

$S_1(4, 2)$	Arcs from users				Arcs to users			
1	1	2	5	6	1	3	5	7
2	1	2	5	6	2	4	6	8
3	3	4	7	8	3	1	7	5
4	3	4	7	8	4	2	8	6

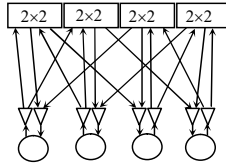
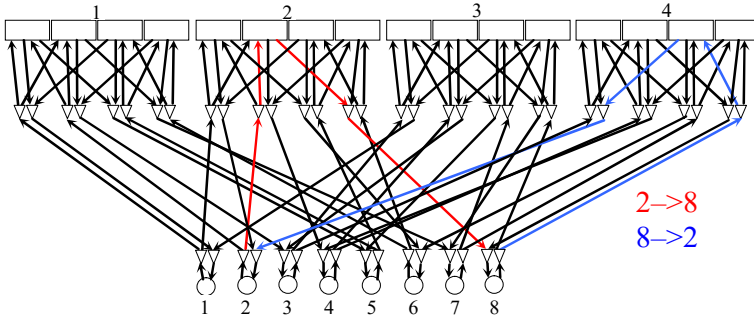
 TABLE 4. The incidence table for a switch $S_3(16, 2)$

$S_2(6, 2)$	Arcs from users						Arcs to users					
1	1	2	5	6	13	14	1	3	5	7	13	15
2	1	2	5	6	13	14	2	4	6	8	14	16
3	3	4	7	8	15	16	3	1	7	5	15	13
4	3	4	7	8	15	16	4	2	8	6	16	14

Routing with route fixing in the switch $S_{i+1}(N_{i+1}, m)$ is performed as follows. The source and sink numbers are used in the incidence table to find the switch $S_i(N_i, m)$ that includes both numbers: from the source and to the sink.

The connection between the users is established through the selected switch. For example, source 2 and sink 8 in Table 3 are connected via a switch $S_1(4, 2)$ with number 2.

Routing within the switch $S_i(N_i, m)$ is performed using the wormhole technique, i.e., by the output port numbers of the demultiplexers from the sources and further by the port numbers in the switch $S_{i-1}(N_{i-1}, m)$.

FIGURE 3. Switch $S_1(4, 2)$ FIGURE 4. Switch $S_2(8, 2)$

According to the figures, the return paths between two users may not coincide either in terms of 2×2 switches ($m \times m$ in the general case) or in terms of the channels in each direction (e.g., the paths between users 2 and 8 in Figure 4).

In the switches $S_i(N_i, m)$ under consideration, group routing can be performed by establishing a connection between an organizing user (organizer) and a group of participating users (participants). For this purpose, a group address $\otimes$ is employed for output ports to connect an input port and all output ports.

An example of such an address is the number $m + 1$. In this case, the organizer designates a group by sending a group-addressed packet containing $i + 1$ symbols, each specifying the number of an output port (or all of them) in each demultiplexer along the paths to the switches. As an illustration, consider the switch $S_3(16, 2)$ in Figure 5.

If organizer 2 sends the packet $(\otimes, \otimes, 1, 1)$, then the group of participants 5, 6, 7, 8 will be selected. If the packet $(\otimes, \otimes, 2, 2)$ is sent, the group 9, 10, 11, 12 will be selected.

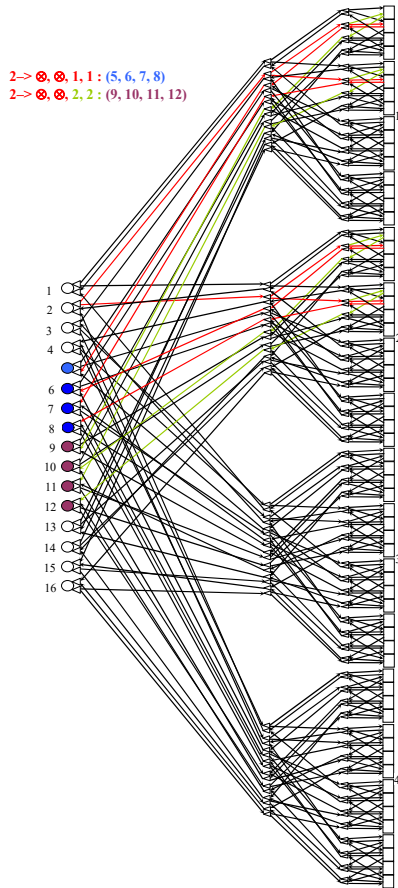


FIGURE 5. Switch $S_3(16, 2)$

Similarly, the packet $(\otimes, 1, \otimes, 1)$ selects the group 7, 8, 11, 12, and so on. The output port numbers do not pass through to the output of the corresponding demultiplexer, being used only to route connections between inputs and outputs.

A group-addressed packet also contains the organizer's address. In the general case, when a packet contains k symbols $\otimes$ ($1 \leq k \leq \log_m N_i$), a group of mk participants is selected, and its composition depends on the set of individual numbers in the train. By serially sending at most $\log_m N_i$ group-addressed packets with different numbers of the symbols $\otimes$, it is possible to form a group of participants of almost any composition.

Group participants transmit acknowledgment packets to the organizer. These packets have arbitrary content and are transmitted by each participant until they are ready to execute a group operation. On their way to the organizer, various acknowledgment packets overlap and become distorted, but do not disappear until their transmission is complete. When all acknowledgment packets have reached the organizer and new ones do not arrive, all group participants are ready to execute a group operation.

During the execution of a group operation, numbers shall be transmitted synchronously bit by bit (i.e., arrive at the inputs of multiplexers in all intermediate layers simultaneously). Such group synchronization is achieved as follows.

We denote by $T(I, J)_U$ and $T(J, I)_D$ the transmission time from organizer I to participant J and in the opposite direction, respectively. The maxima of $T(I, J)_U$ and $T(J, I)_D$ over all I and J will be designated by T_U and T_D , respectively. Let organizer I send a command to start a group operation at a time instant $\Theta(I)$, and let participant J receive it at the time instant $\Theta(J, I) = \Theta(I) + T(I, J)_U$.

To achieve group synchronization, an arbitrary participant J delays the start of a group operation (the transmission of the corresponding word) by the time $T_U + T_D - (T(I, J)_U + T(J, I)_D)$, i.e., begins it at the time instant $\Omega(I, J) = \Theta(I) + T(I, J)_U + (T_U + T_D - (T(I, J)_U + T(J, I)_D))$.

A group operation is executed during the transmission of words through the switches and demultiplexer-multiplexer layers to the organizer without any clock delays. Let $\Omega(I, J)^j$ denote the arrival time of a word from participant J at a certain multiplexer in layer j . It is defined as $\Omega(I, J)^j = \Omega(I, J) + (T(J, I)_D - T(I, J)_D^j)$, where $T(I, J)_D^j$ specifies the signal propagation time from this multiplexer to organizer I . If the path to organizer I from participant L also passes through this multiplexer, then $T(I, J)_D^j = T(I, L)_D^j$.

THEOREM 1 (on Group Synchronization). *The arrival times of words at any multiplexer coincide on the paths to the organizer from the corresponding group participants.*

PROOF. It is straightforward to establish that

$$\begin{aligned}
 \Omega(I, J)^j &= \Theta(I) + T(I, J)_U + T_U + T_D - T(I, J)_U - T(J, I)_D \\
 &\quad + T(J, I)_D - T(I, J)_D^j \\
 &= \Theta(I) + T_U + T_D - T(I, J)_D^j \\
 &= \Theta(I) + T_U + T_D - T(I, L)_D^j.
 \end{aligned}$$

This equality implies the identical arrival times of words at any multiplexer on the paths to the organizer from any group participants.

In particular, for $j = i = \log_m N_i$, we have $T(I, J)_D^i = T(I, L)_D^i = 0$. Hence, the arrival time of the resulting words at the organizer's multiplexer, $\Omega(I, J)^I = \Theta(I) + T_U + T_D$, is independent of the group of participants in the operation. $\square$

2. Group Arithmetic Based on Photonic Logic Circuits

This section considers the execution of group CCC operations under the assumption that the participants are arranged in series along a common channel composed of photonic executive units [4]. To execute CCC operations, logical variables are transmitted via the channel in the two-rail form (along two lines, 0 and 1). Value 0 is transmitted by a signal in line 0 in the absence of a signal in line 1, and value 1 is transmitted by a signal in line 1 in the absence of a signal in line 0. Lines 0 and 1 pass through the converter C_i of each network node i . Each converter executes the unary operation O_i of transforming an input logical variable x_i to an output logical variable $y_i = O_i x_i$. The operation O_i consists of the following transformations: repetition $y_i = x_i$ (“=”), negation $y_i = \neg x_i$ (“−”), change into the value $y_i = 0$ (“0”), and change into the value $y_i = 1$ (“1”).

The converter is used in four fixed states for which the operation O_i is considered to be executed. Each state is defined by some configuration of a pair of fiber sections, given by the value of the two-rail control variable u_i . The circuits in Figure 1 show the use of such pairs of converters for implementing some logical functions for the two-rail variables x and y .

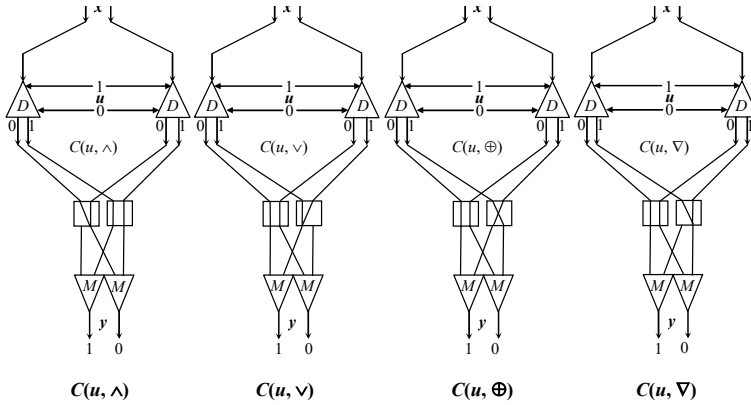


FIGURE 6. The circuits of logical functions from Table 5 for two-rail variables

For such circuits, we will follow the notation given at the bottom of Figure 6.

A photonic 1×2 demultiplexer D and a photonic 2×1 multiplexer M are used in Figure 6. They are assembled from several thin-film layers of metamaterials with electro-optical or magneto-optical properties [10, 11]. An optical signal is used to control switching. In the demultiplexer, it is converted into an internal electronic or magnetic control signal and supplied to a guide layer, which specifies two propagation directions for the information signal received at its optical input. As a result, the input optical signal is routed to one of the two optical outputs. In this paper, a two-rail control variable is used; therefore, other single-layer devices with two outputs (demultiplexer) or inputs (multiplexer), such as gallium arsenide films [8], can be selected. They allow building photonic logical elements directly controlled by light.

Table 5 provides the truth tables of the logical operations used in the circuits in Figure 6.

Obviously, the operation $\aleph$ is defined only by the pair of states of the converter C_i , which are selected by the value of the variable u_i . As a result, it is possible to execute the binary operation $y_i = x_i \aleph u_i$ (any of the sixteen logical operations) on a pair of converters.

TABLE 5. Logical functions by converter states for the adder

AND (logical conjunction)				OR (logical disjunction)				XOR (addition mod 2)				Initial Run			
$\Re = \wedge$				$\Re = \vee$				$\Re = \oplus$				$\Re = \nabla$			
x	u	O	y	x	u	O	y	x	u	O	y	x	u	O	y
0	0	“0”	0	0	0	“ $_$ ”	0	0	0	“ $_$ ”	0	0	0	“ $_$ ”	0
0	1	“ $_$ ”	0	0	1	“1”	1	0	1	“ $\neg$ ”	1	0	1	“0”	0
1	0	“0”	0	1	0	“ $_$ ”	1	1	0	“ $_$ ”	1	1	0	“ $_$ ”	1
1	1	“ $_$ ”	1	1	1	“1”	1	1	1	“ $\neg$ ”	0	1	1	“0”	0

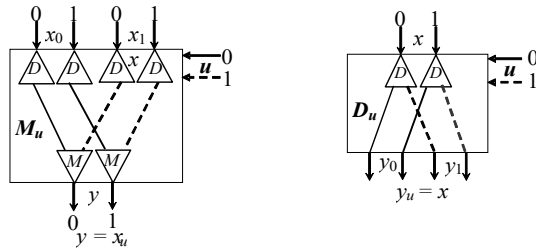


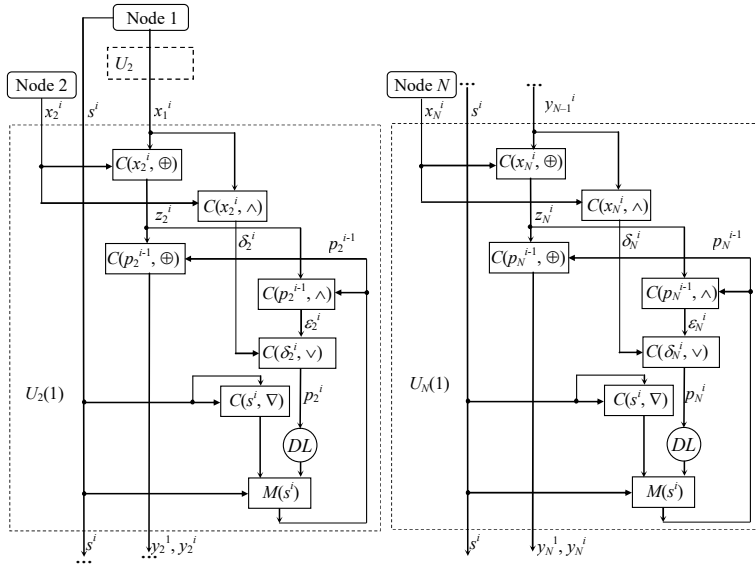
FIGURE 7. The circuits of a controlled photonic multiplexer and demultiplexer

Figure 7 also shows the circuits of a multiplexer (M_u) and demultiplexer (D_u) for two-rail variables, controlled by the variable u .

When nodes are placed serially, it is possible to sum numbers at the nodes or select the maximum of them during the transmission of one number via the common channel. For this purpose, the numbers are transmitted as binary ones serially bit-by-bit. To execute their addition, the numbers are transmitted with the least significant bits first; to find their maximum, with the most significant bits first.

The coordinating user triggers the command of the operation to be executed.

This command synchronizes the start of the distributed operation r for all nodes. Each node has a set of executive units, $U(r)$, configured to

FIGURE 8. The N -node distributed adder

execute individual operations. The operation command selects a particular executive unit. Each executive unit consists of circuits implementing the necessary logical operations in a definite sequence specified by their interconnections.

Figure 8 shows the units $U_2(1)$ and $U_N(1)$ at the second and N^{th} nodes of the cyclic sequence that sum N numbers. The initial first node of the network does not use its unit $U_1(1)$. It outputs slot X_1 , containing a binary multi-bit number starting from the least significant bits (on the two-rail value line), and slot S of the same bit length (on the two-rail synchronization line).

In slot X_1 , the bits are transmitted by the variables x_1^i ; in slot S , by the variables s_i , where i denotes the bit number. The first bit of slot S has value 1 ($s_1 = 1$), and the other bits are 0 ($s_i = 0$ for $i > 1$). The dots on the lines indicate the nodes of light signal separation by branching optical lines. The unit $U_2(1)$ of the distributed adder is connected to the first and second network nodes.

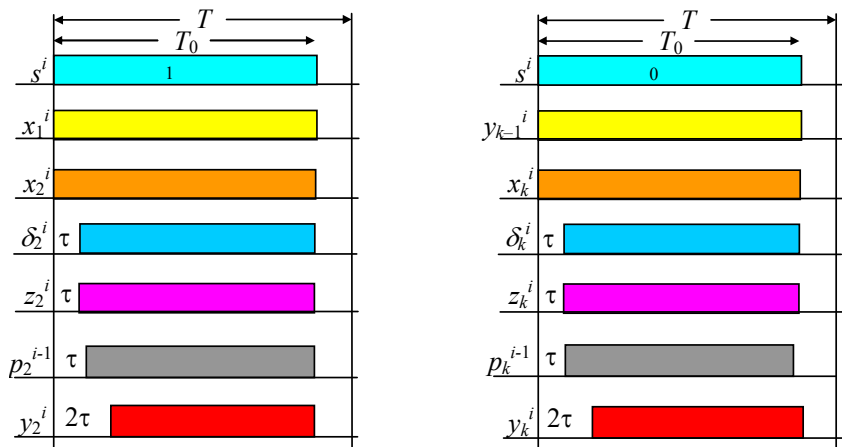


FIGURE 9. The timing diagram of the distributed adder

The circuits in Figure 8 are based on the well-known single-bit adder, supplemented with a circuit to form the zero carry value from the previous bit for the first bit and a circuit to carry the value to the next bit through a one-bit delay line DL .

Any k th node of the network ($2 < k \leq N$) transmits to its unit $U_k(1)$ slot X_k , containing the values of the variables x_k^i for each bit; the other input of this unit receives slot Y_{k-1} , containing the output bits y_{k-1}^i from the unit $U_{k-1}(1)$. In addition, this unit receives slot S from the executive unit of the previous node.

Figure 9 presents the timing diagram of the units $U_2(1)$ and $U_k(1)$ with the following notation: T is the bit period; T_0 is the duration of active signals in it; τ is the time to tune the converter or multiplexer by the control variable u . Clearly, the output variable y_2^i appears in all bits with a delay of 2τ and is defined by signals of duration $T_0 - 2\tau$. Note that the right-hand diagram is valid for any unit $U_k(1)$, $k > 2$. In other words, the output variable y_k^i appears with a delay of 2τ with respect to the start of the synchronization signals s_i for any k and i (for any number of nodes and in all bits) and is given by signals of duration $T_0 - 2\tau$.

Along with the above CCC addition, the first node can also set the maximization operation (find the largest number of those placed at

bits (on the two-rail value line), and slot S of the same bit length (on the two-rail synchronization line). In slot X_1 , the bits are transmitted by the variables x_1^i ; in slot S , by the variables s_i , where i denotes the bit number. The first bit of slot S has value 1 ($s_1 = 1$), and the other bits are 0 ($s_i = 0$ for $i > 1$). Any other node of the network ($1 < k \leq N$) transmits to its unit $U_k(2)$ slot X_k , containing the values of the variables x_k^i for each bit (on the value line). In addition, this unit receives slot S from the executive unit of the previous node.

Figure 10 (left) shows the circuit of the unit $U_2(2)$ connected to the first and second network nodes of the distributed maximizer. The dots indicate the nodes of light signal separation by branching optical lines.

Figure 10 (right) presents the circuit of the unit $U_k(2)$ ($k > 1$), actually repeating that of the unit $U_2(2)$ with x_1 replaced by y_k — 1 and 2 by k in the subscripts.

Figure 11 shows the timing diagram of forming different variables for the j th and $(j + 1)$ th bits with the following notation: τ is the tuning time of the converters and multiplexer; T and T_0 are the durations of the bit and the active signal in it. Obviously, the active signal is shortened by a time not exceeding 3τ , which is independent of j .

3. Group Arithmetic Based on Photonic Multiplexers

Group operations are executed in parallel across the group participants in special executive multiplexers EM_j , which are connected in parallel to each standard multiplexer in any layer j ($1 \leq j \leq i$) of the switch $S_i(N_i, m)$. Figure 12 shows the circuitry of EM_j with m inputs.

The demultiplexer-multiplexer layers are numbered from the users to the switches inclusive. By assumption, the switches are composed of demultiplexers and multiplexers, see the circuitry in Figure 13.

Each participant transmits slots S and X of the same format as described in Section 2. They are propagated through the demultiplexer layers without modification, and bit-by-bit superposition occurs in the executive multiplexer layers. According to the layer numbering, we have $Y_{1,k} = X_k$ ($k = 1, 2, \dots, m$). In each multiplexer, slots S_j and $Y_{j,k}$ are supplied to the executive units U_k (see the previous section) to execute a group operation, with the bit-by-bit synchronization of words, “on-the-fly”.

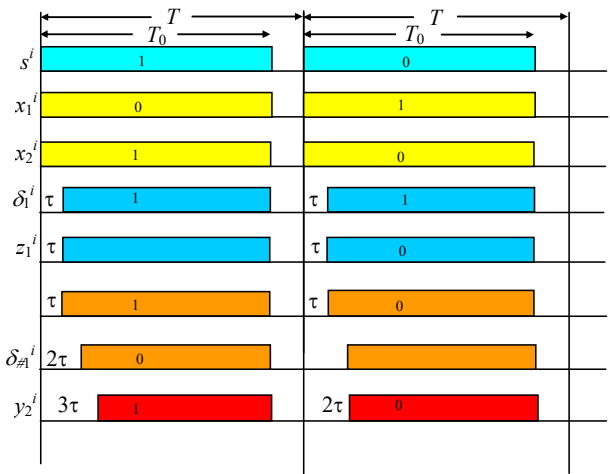


FIGURE 11. Timing diagram of a two-node “maximizer” for the j th bit

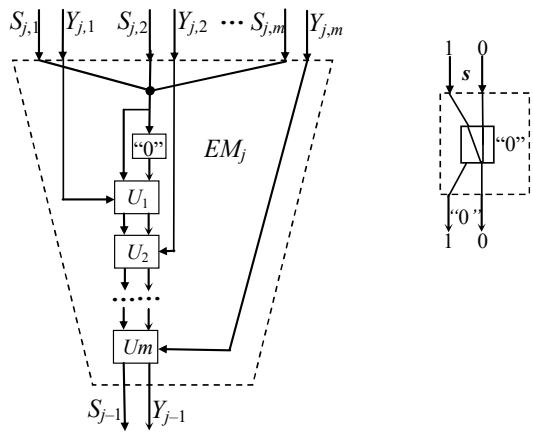


FIGURE 12. Executive multiplexer

The bits of the input slots S_j and Y_j in each unit U_k are used as control variables, while the bits of the output slots Y_{j-1} are interpreted as the result of the operation executed in the corresponding multiplexer EM_j . For each bit, the initial value of the result is generated in the additional unit “0” (see the right-hand side of Figure 12).

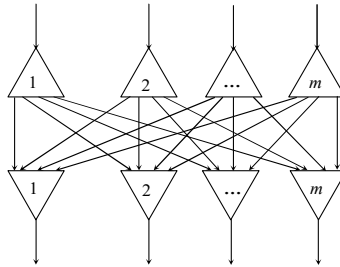


FIGURE 13. An $m \times m$ switch consisting of demultiplexers and multiplexers

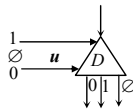


FIGURE 14. A two-layer photonic demultiplexer

Note that the lengths of the synchronization lines and the value lines from the inputs (S_j, Y_j) to outputs (S_{j-1}, Y_{j-1}) of the multiplexer EM_j shall have equal length to preserve the synchronization conditions for the bits. This requirement is easy to satisfy when designing the multiplexers EM_j : the synchronization and result lines passing through them introduce neither combinational nor clock delays, but only signal propagation delays. Unfortunately, the multiplexers EM_j in Figure 12 operate correctly only in the presence of two-rail signals at all inputs $Y_{j,k}$. If signals are absent at even one input, the signals at the output $Y_{j-1,k}$ will be undefined. This drawback can be eliminated using two-layer demultiplexers and multiplexers [10, 11] in logical function circuits. These devices have two layers with controllable refraction of the information signal's propagation. The two-rail signals “0” and “1” specify a change in the propagation direction in the first and second layer, respectively. The original propagation direction is preserved in both layers in the absence of a two-rail signal. As a result, the information signal can be routed to one of three outputs (Figure 14); here, the symbol $\emptyset$ indicates the absence of a signal, and the information signals are represented by a single rail.

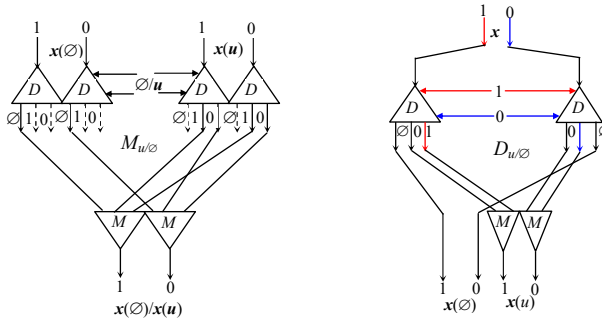


FIGURE 15. Universal controllable demultiplexer $D_{u/\emptyset}$ and multiplexer $M_{u/\emptyset}$

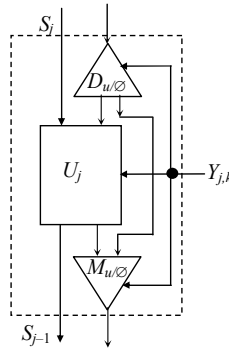


FIGURE 16. The scheme of bypassing the executive unit $U_{k,j}$ in the multiplexer EM_j

This basis can be employed to build a universal controllable demultiplexer $D_{u/\emptyset}$ and a universal controllable multiplexer $M_{u/\emptyset}$. (The circuitries in Figure 15 are analogous to those in Figure 7.)

It can be used in the multiplexer EM_j to bypass the conventional executive unit $U_{j,k}$ (Figure 16).

4. Comparative Characteristics of Parallel and Serial Group Operations

Let us compare the parallel method for executing CCC group operations with the serial method described in [4], in terms of quality characteristics.

We will estimate the complexity of a network with group operations as the sum of the complexities of the switch $S_i(N_i, m)$ and the executive units $U(r)$. In this case, an $m \times m$ switch has an estimated complexity of m^2 ; a demultiplexer/multiplexer, an estimated complexity of m ; and an executive unit U , an estimated complexity of πm , where $\pi < m$.

The complexity ω_i of the switch $S_i(N_i, m)$ is given by the recurrence formula $\omega_i = N_2\omega_{i-1} + 2N_im$, with $\omega_2 = N_2 + 2mN_2 = m^2(1 + 2m)$. Consequently, we obtain $\omega_i = 2m_i(m_{i-1} + m_{i-2} + \dots + 1)$.

For the serial method, the complexity $\Delta_{\otimes, i}$ of the network is estimated as $\Delta_{\otimes, i} = \omega_i + \pi m_{i+1}$.

For the parallel method, the complexity of each EM_j is estimated as πm^2 , so the entire network has the estimated complexity $\Delta_{\otimes, i} = \omega_i + \omega_i \pi m^2 / 2$.

Now, we estimate the execution time of a parallel group operation. It consists of the group addressing time T_{GA} , which includes the organizer's times to send r group addresses ($1 \leq r \leq \log_m N_i$) and wait for the acknowledgements from all participants (to confirm their readiness to execute the operation). Supposing the immediate readiness of all participants and the address transmission time T_A , we have $T_{GA} \leq rT_A + T_U + T_D$.

The execution time of an operation, T_{OP} , consists of the time to send the command, the synchronization delay, and the time to transmit the number, with delivery of the result to the organizer. Supposing that transmitting the command and the number takes times T_I and T_N , respectively, we obtain $T_{OP} \leq T_U + T_I + T_U + T_D + T_N + T_D$. Therefore, the execution of a parallel group operation $\Theta \otimes$ takes the time $\Theta_{\otimes} = T_{GA} + T_{OP} \leq 3(T_U + T_D) + T_A \log_m N_i + T_I + T_N$. Let the signal propagation time be $T_P = T_U = T_D$. In the final analysis, we arrive at the expression $\Theta_{\otimes, I} \leq 6T_P + T_A \log_m N_i + T_I + T_N$.

Using the same notation, the execution time of a serial group operation, $\Theta_{\otimes, i}$, is estimated as $\Theta_{\otimes, i} \leq N_i T_P + N_i T_A + T_I + T_N$ [4].

We will compare the serial and parallel methods by the minimum value of the criterion ρ , known as the complexity-to-speed ratio. For the serial method, it is defined as $\rho_{\otimes, i} = \Delta_{\otimes, i} \Theta_{\otimes, i}$; for the parallel method, as $\rho_{\otimes, i} = \Delta_{\otimes, i} \Theta_{\otimes, i}$.

TABLE 7. The values of γ_i for $m = 8$ and $\pi = 3$

i	2	3	4	5
N_i	64	512	4096	32768
$V = 10$ GB/c, $L = 0,1$ m	1,04	1,21	7,04	49,6
$V = 10$ GB/c, $L = 1$ m	0,90	1,12	7,04	53,5
$V = 100$ GB/c, $L = 1$ m	0,73	1,04	7,04	56,0
$V = 100$ GB/c, $L = 0,1$ m	0,90	1,12	7,04	53,5

TABLE 8. The values of γ_i for $m = 4$ and $\pi = 3$

i	3	4	5	6	7
N_i	64	256	1024	4096	16384
$V = 100$ GB/c, $L = 0,1$ m	0,61	1,71	6,59	25,1	97,4
$V = 100$ GB/c, $L = 1$ m	0,56	1,71	6,89	27,1	107,7
$V = 10$ GB/c, $L = 1$ m	0,61	1,71	6,59	25,1	97,4
$V = 10$ GB/c, $L = 0,1$ m	0,66	1,72	6,11	21,7	79,1
$V = 1$ GB/c, $L = 0,1$ m	0,68	1,72	5,98	20,71	73,6
$V = 1$ GB/c, $L = 1$ m	0,66	1,72	6,11	21,7	79,1

Consider the values $\gamma_i = \rho_{\otimes,i}/\rho_{\#}, i = \Delta_{\otimes,i}\Theta_{\otimes,i}/\Delta_{\#,i}\Theta_{\#,i}$ for different values of several parameters: the number of users N_i , the demultiplexer parameter m , the physical dimensions L of the switch (in m), and the transmission rate V (in Gb/s). In this case, the times T_P , T_A , T_I , and T_N are calculated using the formulas $TP = (i - 1)L/C$ and $TA = T_I = T_N = 64L/V$, where C is the speed of light. The results are presented in [Tables 7 and 8](#) for the largest group of participants.

According to the tables, for systems with tens and hundreds of users, the serial and parallel methods are roughly comparable, while for systems with thousands and tens of thousands of users, the parallel method is many times better by the above criterion. This advantage is due to the smaller time required to form groups of participants and the faster transmission of the operation result from the participants to the organizer via shorter parallel paths. A particularly noticeable gain is observed in terms of the execution times of group operations (delays, see [Table 9](#)).

TABLE 9. The values of $\Theta_{\otimes,i}/\Theta_{\#,i}$ for $m = 4$

i	3	4	5	6	7
N_i	64	256	1024	4096	16384
$V = 1 \text{ GB/c}, L = 1 \text{ m}$	4,5	13,2	43	147	515
$V = 1 \text{ GB/c}, L = 0,1 \text{ m}$	4,4	12,9	43	150	540
$V = 10 \text{ GB/c}, L = 1 \text{ m}$	3,7	11,8	43	162	625
$V = 10 \text{ GB/c}, L = 0,1 \text{ m}$	4,4	12,9	43	150	540
$V = 100 \text{ GB/c}, L = 1 \text{ m}$	2,9	10,9	43	169	674
$V = 100 \text{ GB/c}, L = 0,1 \text{ m}$	3,7	11,8	43	162	625

5. Non-blocking SANs Based on Optoelectronic Switches

This section is almost identical to the corresponding section in [6]. The difference lies in the placement of photonic executive units. In [6], they are located between the users and the nonblocking network. In this work, they are located within the network itself—in executive multiplexers connected in parallel with conventional multiplexers. An optical network for CCC can be constructed using optoelectronic switches based on Mach–Zehnder interferometers (MZI) [6–17].

Such switches can be used to assemble an optoelectronic 4×4 switch [15] with data transmission by optical signals and electronic control by a network node. Currently, the speed of such switches reaches up to 100 Gbps, and they have a fairly high packing density [15–17].

A non-blocking optical network can be based on tandem switching devices (multiplexers and demultiplexers) consisting of similar electronic and optoelectronic components. The former are involved in routing and route fixing, as well as in setting the configuration of the latter. Tandem switching devices should have an identical configuration and operate in the master-slave mode for electronics and optics, respectively.

The resulting non-blocking tandem network consists of duplicates of electronic and optoelectronic networks. In the multiplexers of the second network, it is possible to place photonic executive nodes (Sections 2 2 and 3) to execute CCC operations with higher speed and higher noise immunity.

Conclusions

In this paper, we have considered a method and circuitry for building non-blocking system area networks (SANs) to execute group operations in the network's direct channels. Group operations are executed as common channel computing (CCC) operations with the highest speed (without clock delays, on-the-fly). The SANs can have electronic and optoelectronic (tandem) implementation. The executive units of CCC operations can have both electronic and photonic implementations.

The novelty of this work lies in the parallelization of the entire group operation, which is achieved through a combination of the following factors.

The first factor is the parallel addressing of the group of participants by sending multiple group addresses.

The second one is the bit-by-bit synchronization of number transmissions by all group participants. And the third is the parallel execution of the operation in the executive multiplexers of a tree subnetwork connecting the group participants and the operation organizer.

A distinctive feature of this work is the combination of four factors.

First, it is the purely photonic implementation of executive units in the multiplexers of the system network.

The second feature is the internal scaling method for non-blocking SANs with the digraph structure.

The third feature is to embed executive units in multiplexers during the parallel realization of group operations.

The final feature is the tandem organization of an optoelectronic non-blocking network.

Note that the purely photonic implementation of executive multiplexers is the main innovation of this work, which ensures minimal execution times for group operations.

References

- [1] S. M. Abramov, S. A. Stepanenko. *On approaches to developing software for a photonic computer*, Nacional'nyj Superkomp'yuternyj Forum (NSKF-2022) (Rossiya, Pereslavl'-Zalesskij, IPS imeni A.K. Ajlamazyana RAN, 29 noyabrya–02 dekabrya 2022 goda), 2022 (in Russian). [↑364](#)
- [2] I. V. Prangishvili, V. S. Podlazov, G. G. Stecyura. *Local Microprocessor-Based Computer Networks*, Glava shestaya, Nauka, M., 1984, 175 pp. (in Russian). [↑364](#)
- [3] V. S. Podlazov. “Distributed arithmetic in a common channel based on photonic switches”, *Trudy 14-go Vserossijskogo soveshchaniya po problemam upravleniya*, VSPU 2024 (Moskva, Rossiya, 17–20 iyunya 2024 g.), IPU RAN, M., 2024, ISBN 978-5-91450-276-5, pp. 2541–2546 (in Russian). [↑364](#), [366](#)
- [4] V. S. Podlazov. “Group computing in non-blocking system area networks”, *Program Systems: Theory and Applications*, **16**:5(68) (2025), pp. 3–42 (Svyazannye teksty stat'i na anglijskom i na russkom yazykax). [doi](#) [URL](#) [↑364](#), [365](#), [366](#), [373](#), [382](#), [383](#)
- [5] S. A. Stepanenko. *Optical logic element (variants)*, Patent No 2 723 906 C1, opublikovano 18.06.2020, Byul. No 17, 55 pp. (in Russian). [URL](#) [↑364](#)
- [6] V. S. Podlazov. “Multichannel non-blocking system area network with direct channels”, *Program Systems: Theory and Applications*, **13**:4(55) (2022), pp. 47–76 (in Russian). [doi](#) [URL](#) [↑364](#), [366](#), [368](#), [385](#)
- [7] G. G. Stecyura. “Decentralized autonomic synchronization of interaction processes of mobile objects”, *Problemy upravleniya*, 2020, no. 6, pp. 46–56 (in Russian). [doi](#) [↑365](#), [385](#)
- [8] A. Scarbro. *Sub-nanosecond network synchronization*, An introduction to White Rabbit, ISSUE2.docx, ARMMS, 2022, 50 pp. [URL](#) [↑365](#), [374](#), [385](#)
- [9] M. F. Karavai, V. S. Podlazov. “An invariant extension method for system area networks of multicore computational systems. An ideal system network”, *Autom. Remote Control*, 2010, no. 12, pp. 166–177. [DOI](#) [↑368](#), [385](#)
- [10] K. Vytovtov, E. Barabanova, S. Zouhdi. “Optical switching cell based on metamaterials and ferrite films”, *2018 12th International Congress on Artificial Materials for Novel Wave Phenomena (Metamaterials)* (Espoo, Finland, 27 August 2018–01 September 2018), IEEE, 2018, ISBN 1-5386-4702-8, pp. 424–426. [doi](#) [↑374](#), [381](#), [385](#)
- [11] E. A. Barabanova, K. A. Vytovtov, T. T. Nguyen. “The control system elements of the new generation optical switching cell”, *Journal of Physics: Conference Series*, **1368**:2 (2019), id. 022002, 10 pp. [doi](#) [↑374](#), [381](#), [385](#)
- [12] M. R. Shcherbakov, L. Sheng, V. V. Zubyuk, A. Vaskin, P. P. Vabishchevich, G. Keeler, T. Pertsch, T. V. Dolgova, I. Staude, I. Brener, A. A. Fedyanin. “Ultrafast all-optical tuning of direct-gap semiconductor metasurfaces”, *Nature Communications*, **8** (2027), id. 17, 6 pp. [doi](#) [↑385](#)

- [13] A. I. Stanley, G. Singh, J. Eke, H. Tsuda. “Mach–Zehnder interferometer: A review of a perfect all optical switching structure”, *Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing, ICRCWIP-2014* (Jaipur, India, 16–17 January 2015), Springer, New Delhi, 2015, ISBN 978-81-322-2638-3, pp. 415–426.  ^{↑385}
- [14] G. Singh, R. P. Yadav, V. Janyani. “Modeling of a high performance Mach-Zehnder interferometer all optical switch”, *Optica Applicata*, **42**:3 (2012), pp. 613–625.  ^{↑385}
- [15] W. M. J. Green, M. Yang, S. Assefa, J. Van Campenhout, B. G. Lee, C. V. Jahnnes, F. E. Doany, C. L. Schow, J. A. Kash, Y. A. Vlasov. “Silicon electro-optic 4×4 non-blocking switch array for on-chip photonic networks”, *Proceedings Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2011* (Los Angeles, California, United States, 6–10 March 2011), Optica Publishing Group, 2011, ISBN 978-1-55752-906-0, id. OThM1. 
^{↑385}
- [16] T.-H. Yen, Y.-J. Hung. “Fabrication-insensitive CWDM (de)multiplexer based on cascaded Mach–Zehnder interferometers on silicon-on-insulator”, *Journal of Lightwave Technology*, **39**:1 (2020), pp. 146–153.  ^{↑385}
- [17] Y. Gui, B. M. Nouri, M. Miscuglio, R. Amin, H. Wang, J. B. Khurgin, H. Dalir, V. J. Sorger. “100 GHz micrometer-compact broadband monolithic ITO Mach–Zehnder interferometer modulator enabling 3500 times higher packing density”, *Nanophotonics*, **11**:17 (2022), pp. 4001–4009.  ^{↑385}

Received
approved after reviewing
accepted for publication
published online

15.01.2026;
13.04.2026;
14.04.2026;
03.10.2026.

Recommended by

PhD, S. A. Romanenko

Information about the author:



Viktor Sergeevich Podlazov

Doctor of Technical (Engineering) Science, assistant professor. Research interests: interconnect architectures and routing in supercomputer systems



0000-0002-9175-1138

e-mail:

podlazov@ipu.ru
podlazov@gmail.com

The author declare no conflicts of interests.

УДК 004.724.2 + 004.272.43

 10.25209/2079-3316-2026-17-3-363-414


Параллельные групповые операции в неблокируемых системных сетях

Виктор Сергеевич Подлазов 

Институт проблем управления имени В. А. Трапезникова РАН, Москва, Россия

 podlazov@ipu.ru

Аннотация. Рассмотрены неблокируемые оптоэлектронные сети, в которых выполняются групповые операции над множеством чисел за время передачи одного числа. Предложен способ выполнения групповых операций фотонными блоками, размещенными в каждом мультиплексоре сети, и схемотехника таких исполнительных мультиплексоров. Описана новая возможность выполнения групповых операций параллельно по исполнительным мультиплексорам для заданных множеств абонентов и определены процедуры их выделения. Доказана теорема об условиях поразрядной синхронности групповых операций. Представлена возможность построения неблокируемой электронной системной сети на любое число абонентов и возможность построения аналогичной оптоэлектронной сети как тандемной сети, в которой оптоэлектронные переключатели управляются их аналогами из электронной сети. (*Связанные тексты статьи на английском и на русском языках*)

Ключевые слова и фразы: неблокируемые системные сети, групповые операции, фотонные и оптоэлектронные переключатели и коммутаторы, гамильтоновы циклы узлов сети, тандемные неблокируемые сети

Для цитирования: Подлазов В.С. *Параллельные групповые операции в неблокируемых системных сетях* // Программные системы: теория и приложения. 2026. Т. 17. № 3(72). С. 363–414. (Англ.+русс.) https://psta.psir.ru/read/psta2026_3_363-414.pdf

Введение

В настоящее время развиваются вычислительные системы на основе системных сетей, в которых данные передаются вместе с командами, образуя spatial data flow структуры [1]. Их частным случаем являются системы с распределенной параллельной обработкой данных в процессе их передачи по сети, в которых выполняется групповая операция над множеством чисел, передаваемых во время передачи по общему каналу сети одного числа, формирующего результат операции. Канал является общим, если абоненты могут вести передачу сигналов в него с их наложением, которое можно использовать для вычислений «на лету». Общий канал является прямым, если в нем имеются только задержки распространения сигналов или их комбинационные задержки и отсутствуют какие-либо тактовые задержки.

Групповые операции в прямом канале называются «Вычисления в общем канале» (ВОК) [2] и они могут выполняться без буферизации промежуточных значений («на лету»). Простейшими примерами операций ВОК являются суммирование множества чисел от разных абонентов или нахождение из них максимального числа.

Операции ВОК можно выполнять при передаче данных оптическими сигналами. Для выполнения разных операций при каждом узле сети необходимо иметь специальные исполнительные блоки. Их можно строить в элементной базе коммутаторов оптических каналов в чисто фотонной реализации (без электронного управления блоками) [3, 4].

Сетевые узлы и исполнительные блоки обмениваются оптическими сигналами максимально возможной частоты, и исполнительные блоки работают только с этими сигналами. Исполнительные блоки позволяют выполнять арифметические и логические операции над числами, представленными оптическими сигналами, без использования специальных фотонных АЛУ как новых физических устройств [5]. В этом и состоит главное отличие от мировой практики построения фотонных процессоров, т. к. вычисления организуются только с использованием фотонных коммутаторов на основе парафазного представления оптических сигналов.

Операции ВОК можно выполнять в циклических последовательностях сетевых узлов, образующих гамильтоновы графы [4]. Их можно строить в неблокируемой системной сети с прямыми каналами [4, 6], которая имеет $N_i = m^i$ абонентов и содержит слой полных коммутаторов $m \times m$ и $i - 1$ слоев демультиплексоров $1 \times m$ и мультиплексоров $m \times 1$.

Неблокируемую оптическую сеть можно создать на основе тандемных коммутирующих устройств, состоящих из аналогичных электронных и оптоэлектронных компонентов, введенных в статью [6]. Первые участвуют в прокладке и фиксации маршрутов, и задают конфигурацию вторых.

Тандемные коммутирующие устройства должны иметь идентичную конфигурацию и работать в режиме *master-slave* (электроника-оптика). Таким образом создается тандемная неблокируемая сеть, состоящая из копий коммутаторов электронной и оптоэлектронной сетей. В узлах второй сети можно разместить фотонные исполнительные блоки для выполнения операций ВОК с большим быстродействием и с большей помехозащищенностью.

Выполнение операций ВОК в работе [4] имеет последовательно-параллельный формат. Сначала строится гамильтонов граф на заданной группе абонентов посредством посылки по сети последовательности их адресов. Затем выполняется сама операция посредством параллельной посылки операндов в исполнительные блоки. Наличие последовательной части операции позволяет достичь только двукратного ускорения выполнения операции по сравнению с традиционным способом сетевого сложения по двоичному дереву [4] без учета в нем задержек программного выполнения.

В данной работе ставится и решается задача полностью распараллелить выполнение операций ВОК в неблокируемой сети и тем самым многократно снизить время их исполнения.

Для описания операций будем различать абонента-организатора и группу абонентов-участников. В неблокируемой сети организатор и участники из любой группы связаны встречными деревьям прямых каналов. Они состоят из демультиплексоров $1 \times m$ на пути от абонентов к коммутаторам и из мультиплексоров $m \times 1$ на обратном пути к абонентам.

Для выполнения операции ВОК вместо обычных мультиплексоров можно использовать m -портовые исполнительные мультиплексоры, состоящие из m последовательно соединенных исполнительных блоков. На порты исполнительных мультиплексоров поступают обрабатываемые числа, передаваемые последовательно синхронно по разрядам.

Организация указанной синхронизации базируется на возможности предварительного измерения времени прохождения сигналов между любыми двумя абонентами в обоих направлениях [7], так как в неблокируемой сети они могут быть разными. Синхронизация осуществляется посредством задержки передачи чисел абонентами-участниками на разное время, учитывающее задержки прохождения сигналов от организатора к исполнителю и обратно. Эти задержки должны быть измерены заранее для любых пар источник-приемник на основе стандарта IEEE 1588 [8].

Новые результаты, представленные в статье, состоят в распараллеливании этапов групповой операции, а именно: адресация группы участников посредством посылки нескольких групповых адресов, поразрядная синхронизация передачи чисел всеми участниками группы с поразрядной их синхронизацией, параллельная реализация операции в исполнительных мультиплексорах древовидной подсети.

В разделе 1 описывается построение неблокируемой сети с прямыми каналами и ее оптоэлектронную реализацию. Для полноты изложения раздел содержит повторы из статей [4, 6], чтобы ввести основные понятия и обосновать возможность иметь древовидные каналы между организатором групповой операции и ее участниками. В разделе 1 впервые рассматривается способ поразрядной синхронизации в мультиплексорах сети чисел, передаваемых участниками группы. В разделе 2 описывается структура общего канала для выполнения операций ВОК в фотонной реализации. Раздел 2 содержит повторы из статей [3, 4], чтобы обосновать схмотехнику раздела 3. В разделе 3 описывается структура исполнительных мультиплексоров, и параллельная реализация на них синхронных групповых операций ВОК в неблокируемой сети на основе теоремы о поразрядной синхронизации передаваемых чисел из раздела 1. В разделе 4 даются оценки времен параллельного выполнения групповых операций и сравнение результатами их последовательно-параллельного выполнения в статье [6]. В разделе 5 предлагается методика построений неблокируемых оптоэлектронных сетей на основе интерферометров Маха-Зандера для применения фотонных исполнительных блоков, который частично повторяет раздел из статьи [6].

1. Неблокируемая системная сеть на основе топологии квазиполного оргфа

В данном разделе самомаршрутизируемых строятся системные распределенных сети в виде коммутаторов на базе неблокируемых коммутаторов с топологией квазиполных оргфов. Для них возможна любая степень масштабирования числа абонентов посредством внутреннего каскадирования.

Квазиполный оргграф может быть представлен как двудольный оргграф с $N_1 = m^2$ узлами в каждой доле (рисунок 1). Соединения в нем задаются разными таблицами инцидентности для дуг от абонентов и дуг к абонентам (таблица 1). Они существуют при любых m .

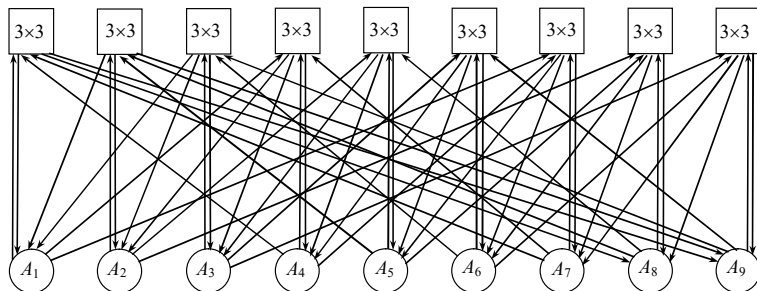


Рисунок 1. Квазиполный оргграф при $m = 3$

Таблица 1. Таблица инцидентности для квазиполного орграфа при произвольном m по $\text{mod } N_1$

Коммутаторы	Каналы от абонентов				Каналы к абонентам			
1	1	2	...	m	1	$1 + m$	...	$1 + m(m - 1)$
2	1	2	...	m	2	$2 + m$	...	$2 + m(m - 1)$
...	1	2	...	m	...	...	...	...
m	1	2	...	m	m	$m + m$	...	$m + m(m - 1)$
$m + 1$	$m + 1$	$m + 2$	...	$m + m$	1	$1 + m$	...	$1 + m(m - 1)$
$m + 2$	$m + 1$	$m + 2$	...	$m + m$	2	$2 + m$	...	$2 + m(m - 1)$
...	$m + 1$	$m + 2$	...	$m + m$	...	...	...	...
$m + m$	$m + 1$	$m + 2$	...	$m + m$	m	$m + m$	...	$m + m(m - 1)$
...	...	...	...	...	...	...	...	...
$m^2 - m + 1$	$m_2 - m + 1$	$m_2 - m + 2$	...	m_2	1	$1 + m$	...	$1 + m(m - 1)$
...	$m^2 - m + 1$	$m_2 - m + 2$	...	m_2	2	$2 + m$	...	$2 + m(m - 1)$
$m^2 - 1$	$m_2 - m + 1$	$m_2 - m + 2$	...	m_2	...	...	...	...
m^2	$m_2 - m + 1$	$m_2 - m + 2$	...	m_2	m	$m + m$	...	$m + m(m - 1)$

Для подсоединения 1-портовых абонентов используются демультиплексоры $1 \times m$ и мультиплексоры $m \times 1$ соответственно (рисунок 2). Таким

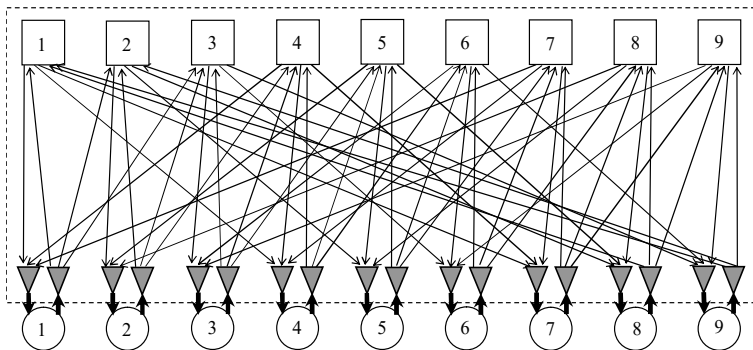


Рисунок 2. Коммутатор $K_1(9, 3)$ с топологией квазиполного орграфа при $m = 3$

образов образуется неблокируемый самомаршрутизируемый распределенный коммутатор $K_1(N_1, m)$. В нем маршрутизации осуществляется от источника по адресам (номерам) выходных портов демультиплексоров и коммутаторов. Бесконфликтная передача пакетов на произвольной их перестановке осуществляется по прямым каналам за один скачок от источников к приемникам, т.е. с максимальным быстродействием.

Существует [6, 9] метод внутреннего масштабирования, который позволяет строить из неблокируемых коммутаторов $K_i(N_i, m)$ с $N_i = m^{i+1}$ неблокируемый коммутатор $K_{i+1}(N_{i+1}, m)$ с $N_{i+1} = m^{i+2}$. Коммутатор $K_{i+1}(N_{i+1}, m)$ строится из N_1 коммутаторов $K_i(N_i, m)$ и слоя из N_{i+1} демультиплексоров $1 \times m$ и мультиплексоров $m \times 1$. Каждый коммутатор $K_i(N_i, m)$ разбивается на доли по m портов, и в каждой доле по всем коммутаторам строится отдельный коммутатор $K_1(N, m)$ с исходной нумерацией в каждой доле и сквозной нумерацией по всем долям. При этом коммутатор $m \times m$ обозначается как $K_0(m, m)$.

В таблицах 2, 3, 4 приводятся примеры построения таблиц инцидентности коммутаторов $K_1(4, 2)$, $K_2(8, 2)$ и $K_3(16, 2)$ из таблиц инцидентности коммутаторов $K_0(2, 2)$, $K_1(4, 2)$ и $K_2(8, 2)$ соответственно.

На рисунках 3, 4 и 5 представлены схемы соответствующих коммутаторов как орграфов.

ТАБЛИЦА 2. Таблица инцидентности коммутатора $K_1(4, 2)$

$K_1(4, 2)$	Дуги от абонентов		Дуги к абонентам	
1	1	2	1	3
2	1	2	2	4
3	3	4	3	1
4	3	4	4	2

 ТАБЛИЦА 3. Таблица инцидентности для неблокируемого коммутатора $K_2(8, 2)$

$K_1(4, 2)$	Дуги от абонентов				Дуги к абонентам			
1	1	2	5	6	1	3	5	7
2	1	2	5	6	2	4	6	8
3	3	4	7	8	3	1	7	5
4	3	4	7	8	4	2	8	6

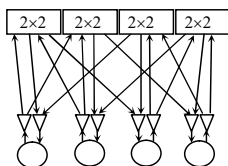
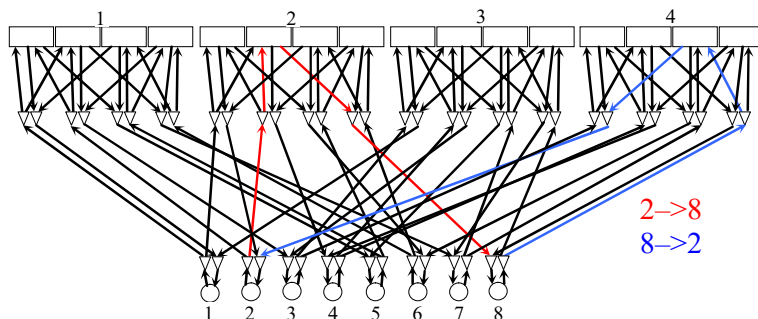
 ТАБЛИЦА 4. Таблица инцидентности для неблокируемого коммутатора $K_3(16, 2)$, получаемого из коммутатора $K_1(4, 2)$

$K_2(6, 2)$	Дуги от абонентов						Дуги к абонентам					
1	1	2	5	6	13	14	1	3	5	7	13	15
2	1	2	5	6	13	14	2	4	6	8	14	16
3	3	4	7	8	15	16	3	1	7	5	15	13
4	3	4	7	8	15	16	4	2	8	6	16	14

Маршрутизация с фиксацией соединений в коммутаторе $K_{i+1}(N_{i+1}, m)$ осуществляется следующим образом. По номерам источника и приемника в таблице инцидентности ищется коммутатор $K_i(N_i, m)$, в который входят оба номера: от источника и к приемнику.

Соединение между абонентами прокладывается через выбранный коммутатор. Например, в таблице 3 источник 2 и приемник 8 соединяются через коммутатор $K_1(4, 2)$ с номером 2.

Маршрутизация в коммутаторе $K_i(N_i, m)$ осуществляется методом червоточины — по номерам выходных портов демультиплексоров от источника и далее по номерам портов в выбранном коммутаторе $K_{i-1}(N_{i-1}, m)$.

Рисунок 3. Коммутатор $K_1(4, 2)$ Рисунок 4. Коммутатор $K_2(8, 2)$

Представленные рисунки показывают, что встречные пути между двумя абонентами могут не совпадать ни по коммутаторам 2×2 ($m \times m$ в общем случае) ни по каналам в каждом направлении (например, пути между абонентами 2 и 8 [на рисунке 4](#)).

В рассматриваемых коммутаторах $K_i(N_i, m)$ можно осуществлять групповую маршрутизацию, осуществляя соединение между абонентом-организатором и группой абонентов-участников. Для этого в демультимплексах можно использовать групповой адрес выходных портов, обозначаемый $\otimes$ и задающий соединение между входным портом и всеми выходными портами.

Таким адресом может быть, например, номер $m + 1$. В этом случае организатор выделяет группу посредством послышки пакета групповой адресации, содержащего $i + 1$ символов, каждый из которых задает номер выходного порта (или всех их) в каждом демультимплексе на путях к коммутаторам. В частности, если в коммутаторе $K_3(16, 2)$ [на рисунке 5](#) организатор 2 посылает пакет $(\otimes, \otimes, 1, 1)$, то выделяется группа участников $\{5, 6, 7, 8\}$. Если же посылается пакет $(\otimes, \otimes, 2, 2)$, то выделяется группа $\{9, 10, 11, 12\}$.

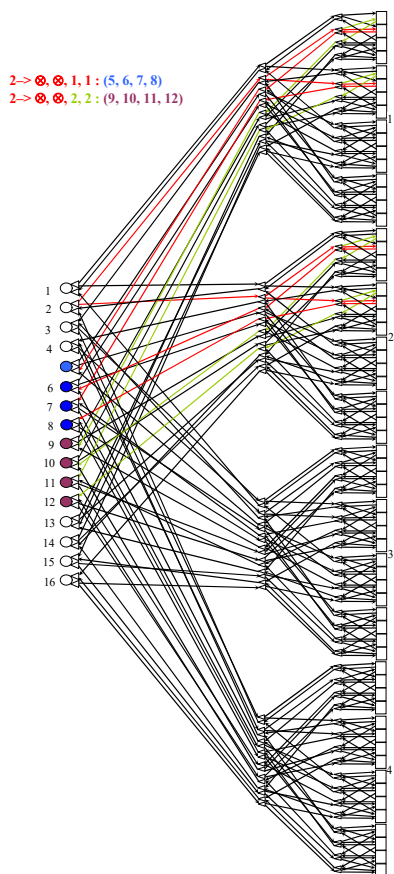


РИСУНОК 5. Коммутатор $K_3(16, 2)$

Аналогично посредством пакета $(\otimes, 1, \otimes, 1)$ выделяется группа $\{7, 8, 11, 12\}$ и т.п. Номера выходных портов не проходят на выход соответствующего демультиплексора, а используются только для прокладки соединений входа и выходов.

Пакет групповой адресации содержит также адрес организатора. В общем случае при использовании в пакете k символов $\otimes$ ($1 \leq k \leq \log_m N_i$) выделяется группа из m^k участников, состав которых зависит от набора индивидуальных номеров в пути. При последовательной посылке не более $\log_m N_i$ пакетов групповой адресации с разным количеством символов $\otimes$ можно выделить группу участников практически любого состава.

Участники группы передают пакеты подтверждений организатору. Эти пакеты имеют произвольное содержание и передаются каждым участником до момента наступления его готовности к выполнению групповой операции. На пути к организатору разные пакеты подтверждений накладываются и искажаются, но не исчезают до конца их передачи. Прекращение поступления пакетов подтверждений к организатору фиксирует момент готовности всех участников группы к выполнению групповой операции.

Во время выполнения групповой операции числа должны передаваться синхронно по разрядам, т.е. так чтобы они приходили на входы мультиплексоров во всех промежуточных слоях одновременно. Достижение такой групповой синхронизации основывается на следующих обозначениях.

Время передачи от организатора I к участнику J обозначается как $T(I, J)_U$, а время передачи от участника J к организатору I — как $T(J, I)_D$. Максимальное время $T(I, J)_U$ по всем I и J обозначается как T_U , а максимальное время $T(J, I)_D$ по всем J и I — как T_D . Пусть организатор I посылает команду на начало групповой операции в момент $\Theta(I)$, и участник J принимает ее в момент $\Theta(J, I) = \Theta(I) + T(I, J)_U$.

Для достижения групповой синхронизации произвольный участник J задерживает начало группой операции (передачу слова в ней) на время $T_U + T_D - (T(I, J)_U + T(J, I)_D)$, т.е. начинает ее в момент

$$\Omega(I, J)^j = \Theta(I) + T(I, J)_U + (T_U + T_D - (T(I, J)_U + T(J, I)_D)).$$

Групповая операция реализуется за время передачи слов через коммутаторы и слои демультиплексоров-мультиплексоров к организатору без тактовых задержек. Обозначим $\Omega(I, J)^j$ момент прихода слова от участника J к некоторому мультиплексору слоя j . Он задается как $\Omega(I, J)^j = \Omega(I, J) + (T(J, I)_D - T(I, J)_D^j)$, где $T(I, J)_D^j$ задает время прохода сигнала от данного мультиплексора к организатору I . Если через данный мультиплексор проходит путь к организатору I также и от участника L , то $T(I, J)_D^j = T(I, L)_D^j$.

ТЕОРЕМА 1 (о групповой синхронизации). *Моменты прихода слов на любой мультиплексор совпадают на путях к организатору от соответствующих участников группы.*

ДОКАЗАТЕЛЬСТВО. Нетрудно убедиться, что

$$\begin{aligned}
 \Omega(I, J)^j &= \Theta(I) + T(I, J)_U + T_U + T_D - T(I, J)_U - T(J, I)_D \\
 &\quad + T(J, I)_D - T(I, J)_D^j \\
 &= \Theta(I) + T_U + T_D - T(I, J)_D^j \\
 &= \Theta(I) + T_U + T_D - T(I, L)_D^j.
 \end{aligned}$$

Последнее равенство и означает совпадение моментов прихода слов на любой мультиплексор на путях к организатору от любых участников группы.

В частности, при $j = i = \log N_i$ имеем $T(I, J)_D^i = T(I, L)_D^i = 0$. Это означает, что момент прихода результирующих слов к мультиплексору организатора $\Omega(I, J)^i = \Theta(I) + T_U + T_D$ не зависит от состава группы участников операции. $\square$

2. Групповая арифметика на основе фотонных логических схем

В данном разделе рассматривается выполнение групповых операций ВОК в предположении последовательного размещения участников вдоль общего канала, составленного из фотонных исполнительных блоков [4]. По каналу логические переменные передаются в парафазном виде, т.е. по двум линиям 0 и 1. Значение 0 передается сигналом в линии 0 при отсутствии сигнала в линии 1, а значение 1 – сигналом в линии 1 при отсутствии сигнала в линии 0. Линии 0 и 1 проходят через конвертор C_i каждого i -го узла сети. Он выполняет унарную операцию O_i преобразования входной логической переменной x_i в выходную логическую переменную $y_i = O_i x_i$. Операция O_i состоит из следующих преобразований: повторения $y_i = x_i$ («=»), отрицания $y_i = \neg x_i$ («¬»), перевода в значение $y_i = 0$ («0») и перевода в значение $y_i = 1$ («1»).

Конвертор используется в 4-х фиксированных состояниях, для которых операция O_i считается выполненной. Каждое из них задается некоторой конфигурацией пары отрезков волоконных линий связи, задаваемых значением парафазной управляющей переменной u_i . На рисунке 6 представлены схемы использования таких пар конверторов для реализации некоторых логических функций для парафазных переменных x и y .

Для этих схем в дальнейшем используются обозначения, приведенные внизу рисунок 6.

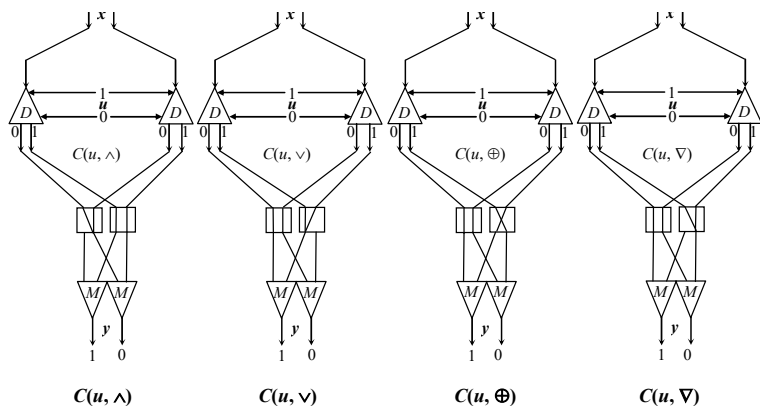


РИСУНОК 6. Схемная реализация логических функций из таблицы 5 для парафазных переменных

На рисунке 6 используются фотонные демультиплексор $D 1 \times 2$ и мультиплексор $M 2 \times 1$, содержащие пленочный слой метаматериалов с электрооптическими или магнитооптическими свойствами [10, 11].

Для управления коммутацией используется оптический сигнал. В демультиплексоре он преобразуется во внутренний электрический или магнитный управляющий сигнал, который подается на управляющий слой, задавая два направления распространения оптического сигнала, поступившего на информационный вход. В результате входной оптический сигнал направляется на один из двух оптических выходов.

В работе используются одна управляющая парафазная переменная, поэтому можно использовать и другие однослойные устройства на два выхода (демультиплексоры) или входа (мультиплексоры), например, пленки арсенида галлия [12]. Они позволяют строить фотонные логические элементы, управляемые светом напрямую.

В таблице 5 приводятся таблицы истинности логических операций, используемых в схемах на рисунке 6.

Видно, что операция $\aleph$ задается только парой состояний конвертора C_i , которые выбираются по значению переменной u_i . Как результат, на паре конверторов можно выполнить бинарную операцию $y_i = x_i \aleph u_i$ — любую из 16 логических операций.

ТАБЛИЦА 5. Логические функции по состояниям конвертора для сумматора

Функция «И»				Функция «ИЛИ»				«Сложение по mod2»				«Начальный запуск»			
$\mathbb{N} = \wedge$				$\mathbb{N} = \vee$				$\mathbb{N} = \oplus$				$\mathbb{N} = \nabla$			
x	u	O	y	x	u	O	y	x	u	O	y	x	u	O	y
0	0	«0»	0	0	0	«=»	0	0	0	«=»	0	0	0	«=»	0
0	1	«=»	0	0	1	«1»	1	0	1	«¬»	1	0	1	«0»	0
1	0	«0»	0	1	0	«=»	1	1	0	«=»	1	1	0	«=»	1
1	1	«=»	1	1	1	«1»	1	1	1	«¬»	0	1	1	«0»	0

На рисунке 7 представлена также схемная реализация для парафазных переменных мультиплексора (M_u) и демультиплексора (D_u), управляемых переменной u .

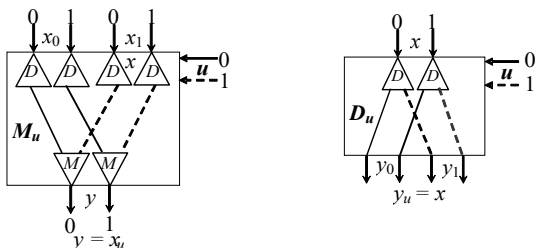


РИСУНОК 7. Схемные реализации управляемых фотонных мультиплексора и демультиплексора

При последовательном размещении узлов можно за время передачи по общему каналу одного числа образовать в нем сумму размещенных по узлам чисел или выделить из них максимальное число. Для этого числа передаются как двоичные числа последовательно по разрядам. Для суммирования они передаются младшими разрядами вперед, а для нахождения максимума – старшими разрядами.

Абонент-организатор, выдает команду выполняемой операции.

Эта команда синхронизирует для всех узлов начало выполнения распределенной операции r . При каждом узле имеется набор исполнительных блоков $U(r)$, настроенных на выполнение отдельных операций.

Команда операции выбирает конкретный исполнительный блок. Каждый исполнительный блок состоит из схем, выполняющих необходимые логические операции в определенной последовательности, заданной их межсоединениями.

На рисунке 8 представлены блоки $U_2(1)$ и $U_N(1)$ при втором и N -ом узлах, выполняющие операцию суммирования N чисел. первый узел не использует свой блок $U_1(1)$.

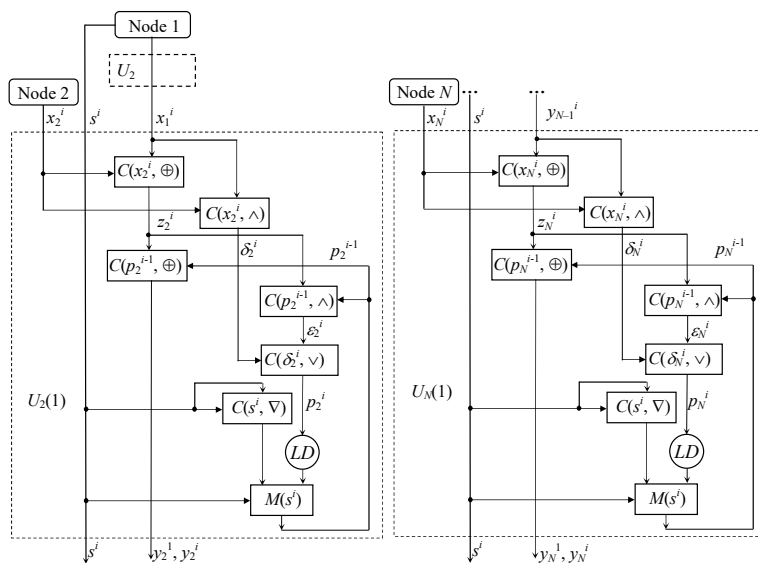


Рисунок 8. N -узовый распределенный сумматор

Он выдает по парафазной линии значений слот X_1 , содержащий двоичное многоразрядное число, начиная с младших разрядов, а по парафазной синхролинии – слот S той же разрядности.

В слоте X_1 разряды передаются переменными x_1^i , а в слоте S – переменными s^i , где i – это номер разряда. Первый разряд слота S имеет значение 1 ($s_1 = 1$), а остальные его разряды – значения 0 ($s_i = 0$ при $i > 1$). Точки на линиях обозначают узлы разделения световых сигналов посредством ветвления оптических линий. Блок $U_2(1)$ распределенного сумматора подсоединяется к первому и второму сетевым узлам.

Схемы на рисунке 8 получены на основе известной схемы одноразрядного сумматора, к которой добавлена схема формирования нулевого

значении переноса из предыдущего разряда для первого разряда и схема переноса значения в следующий разряд через линию задержки LD длительности в один такт.

Любой k -й узел сети ($2 < k \leq N$) передает в свой блок $U_k(1)$ слот X_k , содержащий значения переменных x_k^i для каждого разряда, на другой вход которого поступает слот Y_{k-1} , содержащий выходные разряды y_{k-1}^i из блока $U_{k-1}(1)$. Кроме того, в этот блок поступает слот S , прошедший через исполнительный блок предыдущего узла

На рисунке 9 представлена временная диаграмма работы блоков $U_2(1)$ и $U_k(1)$ с использованием следующих обозначений.

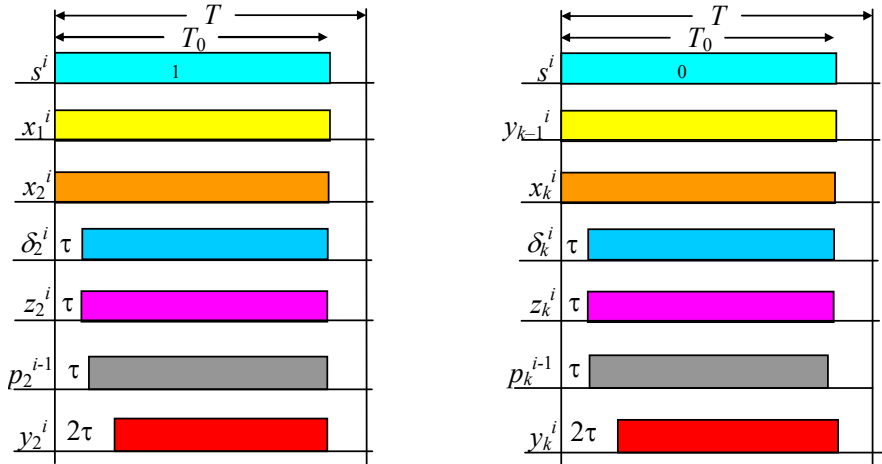


Рисунок 9. Временная диаграмма распределенного сумматора

Период разряда задается как T , длительность активных сигналов в нем – как T_0 , а время настройки конвертора или мультиплексора по управляющей переменной u – как τ . Видно, что выходная переменная y_2^i во всех разрядах появляется с задержкой на 2τ и задается сигналами длительности $T_0 - 2\tau$.

Следует отметить, что правая диаграмма справедлива для любого блока $U_k(1)$ при $k > 2$. Это означает, что выходная переменная y_k^i появляется с задержкой на 2τ относительно начала синхросигналов s^i при любых k и i , т.е. при любом числе узлов и во всех разрядах чисел и задается сигналами длительности $T_0 - 2\tau$.

Первый узел сети выдает по парафазной линии значений в блок $U_2(2)$ слот X_1 , содержащий двоичное многоразрядное число, начиная со старших разрядов, а по парафазной синхролинии – слот S той же разрядности. В слоте X_1 разряды передается переменными x_1^i , а в слоте S – переменными s^i , где i – это номер разряда. Первый разряд слота S имеет значение 1 ($s_1 = 1$), а остальные его разряды – значения 0 ($s^i = 0$ при $i > 1$).

Любой другой узел сети ($1 < k \leq N$) передает по линии значений в свой блок $U_k(2)$ слот X_k , содержащий значения переменных x_k^i для каждого разряда. Кроме того, в этот блок поступает слот S , прошедший через исполнительный блок предыдущего узла.

На рисунке 10 (слева) представлена схема блока $U_2(2)$, подключаемого к первому и второму сетевым узлам распределенного «максимизатора». Точки на линиях обозначают узлы разделения световых сигналов посредством ветвления оптических линий.

На рисунке 10 (справа) приводится схема блока, $U_k(2)$ ($k > 1$), которая повторяет схему блока $U_2(2)$ с заменой x_1 на y_{k-1} и заменой индекса 2 на индекс k .

На рисунке 11 приводится временная диаграмма формирования разных переменных для разрядов j и $j + 1$, где τ обозначает время

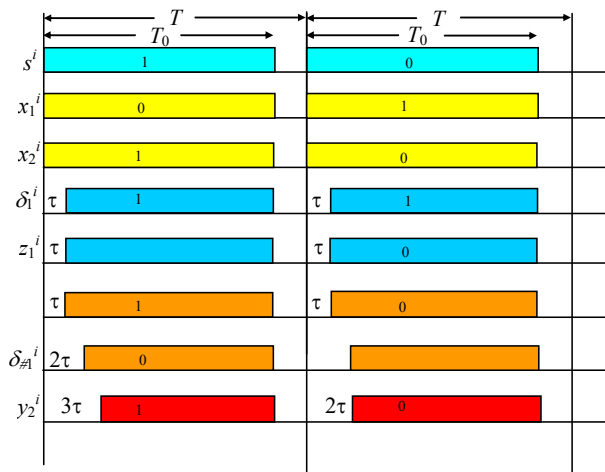


Рисунок 11. Временная диаграмма двухузлового «максимизатора» для j -го разряда

настройки конвертеров и мультиплексора, а T и T_0 – длительности разряда и активного сигнала в нем. Видно, что активный сигнал укорачивается на время не большее 3τ , которое не зависит от значения j .

3. Групповая арифметика на основе фотонных мультиплексоров

Групповые операции выполняются параллельно по участникам группы в специальных исполнительных мультиплексорах $ИМ_j$, включаемых параллельно каждому стандартному мультиплексору в любом слое j ($1 \leq j \leq i$) коммутатора $K_i(N_i, m)$. Схема мультиплексора $ИМ_j$ на m входов представлена на рисунке 12.

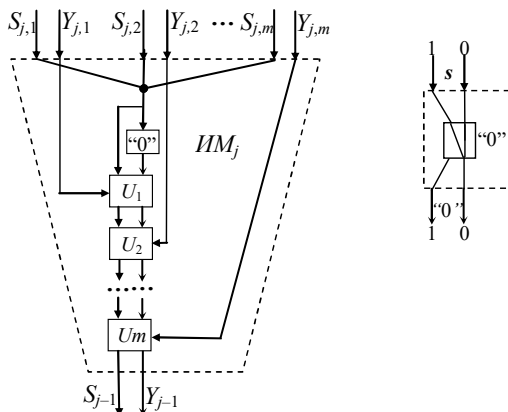


Рисунок 12. Исполнительный мультиплексор

Слои демультиплексор–мультиплексор нумеруются от абонентов до коммутаторов включительно. При этом предполагается, что коммутаторы составлены из демультиплексоров и мультиплексоров по схеме на рисунке 13.

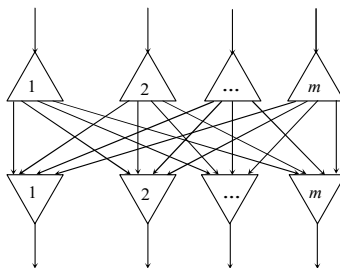


Рисунок 13. Коммутатор $m \times m$ из демультиплексоров и мультиплексоров

Каждый участник передает слоты S и X , имеющие тот же формат как в разделе 2. Они распространяются по слоям демультиплексоров без

изменений, а в слоях исполнительных мультиплексоров происходит их поразрядное наложение. Согласно нумерации слоев имеем $Y_{1,k} = X_k$ ($k = 1, 2, \dots, m$). В каждом мультиплексоре слоты S_j и $Y_{j,k}$ поступают на исполнительные блоки U_k , рассмотренные в предыдущем разделе для выполнения групповой операции синхронно по разрядам слов «на лету».

Значения разрядов входных слотов S_j и Y_j в каждом блоке U_k используются как управляющие переменные, а значения разрядов выходных слотов Y_{j-1} трактуются как значения результата операции, выполненной в данном мультиплексоре IM_j . Начальное значение результата операции для каждого разряда формируется в дополнительном блоке «0», схема которого приведена справа на рисунке 12.

Здесь необходимо отметить, что длины линий синхронизации и значений со входов (S_j, Y_j) мультиплексора IM_j на его выходы (S_{j-1}, Y_{j-1}) должны иметь одинаковую длину, чтобы не нарушать условия синхронизации разрядов. Это требование просто выполнить при построении мультиплексоров IM_j , так как проходящие через них линии синхронизации и результата не вносят ни комбинационных, ни тактовых задержек, а только задержки распространения сигналов.

К сожалению, мультиплексоры IM_j по схеме рисунок 12 работают правильно, только если парафазные сигналы имеются на всех входах $Y_{j,k}$. При отсутствии сигналов хотя бы на одном входе значения сигналов на выходе $Y_{j-1,k}$ становятся неопределенными.

Для устранения этого недостатка можно в схемах логических функций использовать двухслойные демультиплексоры и мультиплексоры из [10, 11]. Они имеют два слоя с управляемым преломлением распространения информационного сигнала. Парафазный сигнал «0» задает изменение направления распространения в одном слое, а парафазный сигнал «1» – в другом слое. Отсутствие парафазного сигнала сохраняет исходное направление распространения в обоих слоях.

В результате информационный сигнал может проходить на один из трех выходов (рисунок 14).

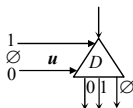


Рисунок 14. Двухслойный фотонный демультиплексор

На рисунке 14 знак $\emptyset$ обозначает отсутствие сигнала, а информационные сигналы представлены только одной фазой.

На его основе можно построить универсальный управляемый демультиплексор $D_{u/\emptyset}$ и управляемый мультиплексор $M_{u/\emptyset}$ (рисунок 15 – аналогии рисунка 7).

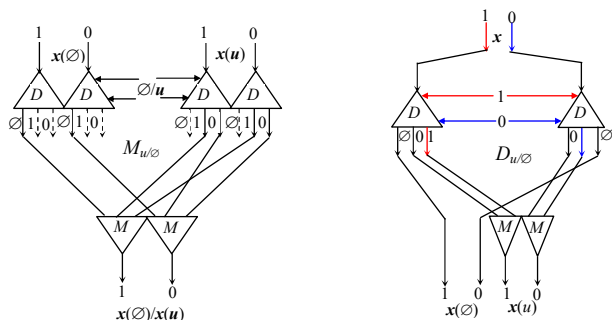


Рисунок 15. универсальный управляемый демультиплексор $D_{u/\emptyset}$ и управляемый мультиплексор $M_{u/\emptyset}$

Его можно использовать в мультиплексоре IM_j для обхода традиционного исполнительного блока $U_{j,k}$ (рисунок 16).

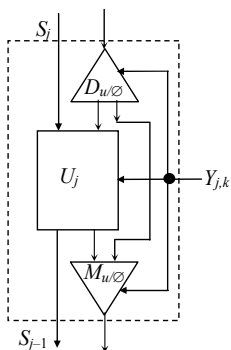


Рисунок 16. Схема обхода исполнительного блока $U_{k,j}$ в мультиплексоре IM_j

4. Сравнительные характеристики параллельной и последовательной групповых операций

Оценим характеристики параллельного способа выполнения групповых операций ВОК в сравнении с последовательным способом в [4].

Сложность сети с групповыми операциями будем оценивать как сумму сложности коммутатора $K_i(N_i, m)$ и сложности исполнительных блоков $U(r)$. При этом сложность коммутатора $m \times m$ оценивается как m^2 , демультимплексора и мультимплексора – как m , а сложность исполнительного блока U – как πm , где $\pi < m$.

Сложность ω_i коммутатора $K_i(N_i, m)$ оценивается как $\omega_i = N_2 \omega_{i-1} + 2N_i m$ при $\omega_2 = N_2 + 2mN_2 = m^2(1 + 2m)$. В результате имеем $\omega_i = 2m^i(m^{i-1} + m^{i-2} + \dots + 1)$.

Для последовательного способа сложность $\Delta_{\otimes, i}$ сети оценивается как $\Delta_{\otimes, i} = \omega_i + \pi m^{i+1}$.

Для параллельного способа сложность каждого IM_j оценивается как πm^2 , тогда сложность $\Delta_{\otimes, i}$ всей сети оценивается как $\Delta_{\otimes, i} = \omega_i + \omega_i \pi m^2 / 2$.

Оценим время выполнения параллельной групповой операции. Она состоит из времени групповой адресации T_{GA} , которая состоит из послышки организатором r групповых адресов ($1 \leq r \leq \log_m N_i$) и ожидания подтверждений участников готовности к выполнению операции. Если принять, что все участники готовы сразу и передача адреса занимает время T_A , то $T_{GA} \leq rT_A + T_U + T_D$.

Время выполнения самой операции T_{OP} складывается из послышки команды, задержки синхронизации и передачи числа с доставкой результата организатору. Если принять, что передача команды и числа занимает времена T_I и T_N , то $T_{OP} \leq T_U + T_I + T_U + T_D + T_N + T_D$. В результате выполнение параллельной групповой операции $\Theta_{\#}$ занимает время $\Theta_{\#} = T_{GA} + T_{OP} \leq 3(T_U + T_D) + T_A \log_m N_i + T_I + T_N$. Примем, что время распространения сигнала $T_P = T_U = T_D$. Окончательно имеем соотношение $\Theta_{\#, i} \leq 6T_P + T_A \log_m N_i + T_I + T_N$.

Время выполнения последовательной групповой операции $\Theta_{\otimes}$ в тех же обозначениях оценивается [4] как $\Theta_{\otimes, i} \leq N_i T_P + N_i T_A + T_I + T_N$.

Будем сравнивать последовательный и параллельный способ по минимальному критерию ρ , задаваемый как отношение сложность/быстродействие. Для последовательного способа он задается как $\rho_{\otimes, i} = \Delta_{\otimes, i} \Theta_{\otimes, i}$, а для параллельного способа – как $\rho_{\otimes, i} = \Delta_{\otimes, i} \Theta_{\otimes, i}$.

Рассмотрим отношения $\gamma_i = \rho_{\otimes, i} / \rho_{\#, i} = \Delta_{\otimes, i} \Theta_{\otimes, i} / \Delta_{\#, i} \Theta_{\#, i}$ при разных значениях таких параметров как: число абонентов N_i , параметр демультимплексора m , физические размеры коммутатора $L(m)$ и скорость передачи V (Гб/с). При этом времена T_P , T_A , T_I , T_N рассчитываются

Таблица 7. Значения γ_i при $m = 8$ и $\pi = 3$

i	2	3	4	5
N_i	64	512	4096	32768
$V = 10 \text{ Гб/с}, L = 0,1 \text{ м}$	1,04	1,21	7,04	49,6
$V = 10 \text{ Гб/с}, L = 1 \text{ м}$	0,90	1,12	7,04	53,5
$V = 100 \text{ Гб/с}, L = 1 \text{ м}$	0,73	1,04	7,04	56,0
$V = 100 \text{ Гб/с}, L = 0,1 \text{ м}$	0,90	1,12	7,04	53,5

Таблица 8. Значения γ_i при $m = 4$ и $\pi = 3$

i	3	4	5	6	7
N_i	64	256	1024	4096	16384
$V = 100 \text{ Гб/с}, L = 0,1 \text{ м}$	0,61	1,71	6,59	25,1	97,4
$V = 100 \text{ Гб/с}, L = 1 \text{ м}$	0,56	1,71	6,89	27,1	107,7
$V = 10 \text{ Гб/с}, L = 1 \text{ м}$	0,61	1,71	6,59	25,1	97,4
$V = 10 \text{ Гб/с}, L = 0,1 \text{ м}$	0,66	1,72	6,11	21,7	79,1
$V = 1 \text{ Гб/с}, L = 0,1 \text{ м}$	0,68	1,72	5,98	20,71	73,6
$V = 1 \text{ Гб/с}, L = 1 \text{ м}$	0,66	1,72	6,11	21,7	79,1

по формулам $T_P = (i - 1)L/C$ и $T_A = T_I = T_N = 64L/V$, где C – это скорость света. Результаты расчетов приводятся в таблицах 7 и 8 для самой большой группы участников.

Таблицы показывают, что для систем с десятками и сотнями абонентов последовательный и параллельные способы примерно сопоставимы, а для систем с тысячами и десятками тысяч абонентов параллельный способ по выбранному критерию во много раз лучше. Это преимущество обеспечивается меньшим временем формирования групп участников и более быстрым прохождением результата операции от участников к организатору по более коротким параллельным путям. Особенно заметно это преимущество по отношениям времен выполнения групповых операций – задержек (таблица 9).

ТАБЛИЦА 9. Значения $\Theta_{\otimes,i}/\Theta_{\#,i}$ при $m = 4$

i	3	4	5	6	7
N_i	64	256	1024	4096	16384
$V = 1 \text{ Гб/с}, L = 1 \text{ м}$	4,5	13,2	43	147	515
$V = 1 \text{ Гб/с}, L = 0,1 \text{ м}$	4,4	12,9	43	150	540
$V = 10 \text{ Гб/с}, L = 1 \text{ м}$	3,7	11,8	43	162	625
$V = 10 \text{ Гб/с}, L = 0,1 \text{ м}$	4,4	12,9	43	150	540
$V = 100 \text{ Гб/с}, L = 1 \text{ м}$	2,9	10,9	43	169	674
$V = 100 \text{ Гб/с}, L = 0,1 \text{ м}$	3,7	11,8	43	162	625

5. Неблокируемые системные сети из оптоэлектронных переключателей

Данный раздел почти повторяет аналогичный раздел из [6]. Отличие состоит в местах размещения фотонных исполнительных узлов. В [6] они размещаются между абонентами и неблокируемой сетью. В данной работе они размещаются в внутри самой сети – в исполнительных мультиплексорах, включенных параллельно обычным мультиплексорам.

Оптическую сеть для ВОК можно построить из оптоэлектронных переключателей на базе интерферометров Маха-Зандера (MZI) [6–17]. Из таких переключателей можно собрать оптоэлектронный коммутатор 4×4 [15], в котором передача данных через коммутатор ведется оптическими сигналами, а управление коммутатором осуществляется узлом сети электронным образом. В настоящее время быстродействие таких переключателей доходит до 100 Гбит/сек и они имеют достаточно высокую плотность упаковки [15–17].

Неблокируемую оптическую сеть можно создать на основе тандемных коммутирующих устройств (мультиплексоров и демультиплексоров, состоящих из аналогичных электронных и оптоэлектронных компонентов. Первые участвуют в прокладке и фиксации маршрутов и задают конфигурацию вторых. Тандемные коммутирующие устройства должны иметь идентичную конфигурацию и работать в режиме master(электроника)-slave(оптика).

Таким образом создается тандемная неблокируемая сеть, состоящая из копий электронной и оптоэлектронной сетей. В мультиплексорах второй сети можно разместить фотонные исполнительные узлы (разделы 2 и 3) для выполнения операций ВОК с большим быстродействием и с большей помехозащищенностью.

Заключение

Рассмотрена методика и схемотехника построения неблокируемых системных сетей, обеспечивающих параллельное выполнение групповых операций в прямых каналах сети. Групповые операции выполняются как операции вычислений в общем канале (ВОК) без тактовых задержек (налету) с наибольшим быстродействием. Системные сети могут иметь электронную и оптоэлектронную (тандемную) реализацию. Исполнительные блоки выполнения операций ВОК могут иметь как электронную, так и фотонную реализацию.

Новизна данной работы состоит в распараллеливании всей групповой операции, которое достигается посредством сочетания следующих факторов.

Во-первых, это параллельная адресация группы участников за счет посылки нескольких групповых адресов.

Во-вторых, это поразрядная синхронизация передачи чисел всеми участниками группы.

В-третьих, это параллельная реализация операции в исполнительных мультиплексорах древовидной подсети, связывающей участников группы и организатора операции.

Отличительной особенностью данной работы является сочетание четырех факторов.

Во-первых, это чисто фотонная реализация исполнительных блоков в мультиплексорах системной сети.

Второй особенностью является способ внутреннего масштабирования неблокируемых системных сетей с оргграфовой структурой.

Третьей особенностью является способ использования исполнительных блоков в мультиплексорах при параллельном выполнении в них групповых операций.

Последней особенностью является тандемная организация оптоэлектронной неблокируемой сети.

Необходимо отметить, что чисто фотонная реализация исполнительных мультиплексоров является главной новацией данной работы, которая обеспечивает минимальные времена выполнения групповых операций.

Список использованных источников

- [1] Абрамов С. М., Степаненко С. А. *О подходах к разработке программного обеспечения для фотонной вычислительной машины*, Национальный Суперкомпьютерный Форум (НСКФ-2022) (Россия, Переславль-Залесский, ИПС имени А.К. Айламазяна РАН, 29 ноября–02 декабря 2022 года).– 2022.
↑390
- [2] Прангишвили И. В., Подлазов В. С., Стецюра Г. Г. *Локальные микропроцессорные вычислительные сети*, Глава шестая.– М.: Наука.– 1984.– 175 с.
↑390
- [3] Подлазов В. С. *Распределенная арифметика в общем канале на основе фотонных коммутаторов // Труды 14-го Всероссийского совещания по проблемам управления*, ВСПУ 2024 (Москва, Россия, 17–20 июня 2024 г.), М.: ИПУ РАН.– 2024.– ISBN 978-5-91450-276-5.– С. 2541–2546. ✨ ↑390, 392
- [4] Подлазов В. С. *Групповые вычисления в блокируемых системных сетях // Программные системы: теория и приложения*.– 2025.– Т. 16.– № 5(68).– С. 3–42. ↑390, 391, 392, 399, 408, 409
- [5] Степаненко С. А. *Оптический логический элемент (варианты)*, Патент № 2 723 906 С1, опубликовано 18.06.2020, Бюл. № 17.– 55 с. ↑390
- [6] Подлазов В. С. *Самомаршрутизируемая блокируемая системная сеть с прямыми каналами: сложность и быстродействие. // Программные системы: теория и приложения*.– 2022.– Т. 13.– № 4(55).– С. 47–76. ↑390, 392, 394, 411
- [7] Стецюра Г. Г. *Децентрализованная автономная синхронизация процессов взаимодействия мобильных объектов // Проблемы управления*.– 2020.– № 6.– С. 46–56. ↑391, 411
- [8] Scarbro A. *Sub-nanosecond network synchronization*, An introduction to White Rabbit, ISSUE2.docx.– ARMMS.– 2022.– 50 pp. ↑391, 411
- [9] Каравай М. Ф., Подлазов В. С. *Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем. Идеальная системная сеть // Автоматика и телемеханика*.– 2010.– № 12.– С. 166–177. ↑394, 411
- [10] Vytovtov K., Barabanova E., Zouhdi S. *Optical switching cell based on metamaterials and ferrite films // 2018 12th International Congress on Artificial Materials for Novel Wave Phenomena (Metamaterials)* (Espoo, Finland, 27 August 2018–01 September 2018).– IEEE.– 2018.– ISBN 1-5386-4702-8.– Pp. 424–426.
↑400, 407, 411
- [11] Barabanova E. A., Vytovtov K. A., Nguyen T. T. *The control system elements of the new generation optical switching cell // Journal of Physics: Conference Series*.– 2019.– Vol. 1368.– No. 2.– id. 022002.– 10 pp. ↑400, 407, 411
- [12] Shcherbakov M. R., Sheng L., Zubuyuk V. V., Vaskin A., Vabishchevich P. P., Keeler G., Pertsch T., Dolgova T. V., Staude I., Brener I., Fedyanin A. A. *Ultrafast all-optical tuning of direct-gap semiconductor metasurfaces // Nature Communications*.– 2027.– Vol. 8.– id. 17.– 6 pp. ↑400, 411

- [13] Stanley A. I., Singh G., Eke J., Tsuda H. *Mach-Zehnder interferometer: A review of a perfect all optical switching structure* // *Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing, ICRCWIP-2014* (Jaipur, India, 16–17 January 2015), New Delhi: Springer.– 2015.– ISBN 978-81-322-2638-3.– Pp. 415–426.  [↑411](#)
- [14] Singh G., Yadav R. P., Janyani V. *Modeling of a high performance Mach-Zehnder interferometer all optical switch* // *Optica Applicata*.– 2012.– Vol. **42**.– No. 3.– Pp. 613–625.  [↑411](#)
- [15] Green W. M. J., Yang M., Assefa S., Van Campenhout J., Lee B. G., Jahnes C. V., Doany F. E., Schow C. L., Kash J. A., Vlasov Y. A. *Silicon electro-optic 4 × 4 non-blocking switch array for on-chip photonic networks* // *Proceedings Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2011* (Los Angeles, California, United States, 6–10 March 2011).– Optica Publishing Group.– 2011.– ISBN 978-1-55752-906-0.– id. OThM1.  [↑411](#)
- [16] Yen T.-H., Hung Y.-J. *Fabrication-insensitive CWD (de)multiplexer based on cascaded Mach-Zehnder interferometers on silicon-on-insulator* // *Journal of Lightwave Technology*.– 2020.– Vol. **39**.– No. 1.– Pp. 146–153.  [↑411](#)
- [17] Gui Y., Nouri B. M., Miscuglio M., Amin R., Wang H., Khurgin J. B., Dalir H., Sorger V. J. *100 GHz micrometer-compact broadband monolithic ITO Mach-Zehnder interferometer modulator enabling 3500 times higher packing density* // *Nanophotonics*.– 2022.– Vol. **11**.– No. 17.– Pp. 4001–4009.  [↑411](#)

Поступила в редакцию	15.01.2026;
одобрена после рецензирования	13.04.2026;
принята к публикации	14.04.2026;
опубликована онлайн	03.10.2026.

Рекомендовал к публикации

к.ф.-м.н. С. А. Романенко

Информация об авторе:



Виктор Сергеевич Подлазов

Д. т. н., гл.н.с. Института проблем управления им. В.А. Трапезникова РАН. Научные интересы: архитектуры интерконнекта и маршрутизация в суперкомпьютерных системах



0000-0002-9175-1138

e-mail: podlazov@ipu.ru
podlazov@gmail.com

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*