



NON-BLOCKING FAULT-TOLERANT TWO-STAGE DUAL PHOTON SWITCHES

E.A. Barabanova, K.A. Vytovtov, and V.S. Podlazov

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

✉ elizavetaalex@yandex.ru, ✉ vytovtov_konstan@mail.ru, ✉ podlazov@ipu.ru

Abstract. This paper further develops the theory of constructing a fundamentally new class of system area networks – non-blocking dual photon networks with static self-routing. These networks have scalability, high speed inherent in photon systems, and complexity comparable to a full switch. The use of an extended scheme basis (dual photon switches and separate photon multiplexers and demultiplexers) allows balancing the scalability-speed and complexity-speed ratios. This paper proposes a method for constructing a two-stage fault-tolerant dual network with the indicated properties based on networks with the quasi-complete graph and quasi-complete digraph topologies and the invariant extension method with internal parallelization.

Keywords: photon switch, dual switch, photon multiplexers and demultiplexers, multi-stage switch, conflict-free self-routing, non-blocking switch, static self-routing, quasi-complete digraph, quasi-complete graph, switching properties, direct channels, scalability, speed and fault tolerance.

INTRODUCTION

The papers [1–4] proposed a technique for constructing non-blocking photon switches with static self-routing for optical supercomputer systems and communication systems.

Below, we formulate and solve the problem of constructing non-blocking self-routing photon switches with $(\sigma - 1)$ -channel fault tolerance. It is provided by replacing the dual quasi-complete digraph topology [1–3] with the dual quasi-complete graph one, in which the total number of channels can be exchanged for the number of different channels between users. This problem is solved in the same way: using the method of internal parallelization developed earlier in [4].

Non-blocking networks with static self-routing are a fundamentally new class of system area networks that achieve these properties by using a new type of non-blocking dual switches and the invariant extension method based on networks with the quasi-complete digraph and graph topologies [4, 5].

This assertion rests on the fact that no such networks with high scalability and acceptable complexity have been constructed so far. Modern literature [6–13]

widely describes system area networks with the fat tree structure (particularly reconfigurable Clos networks), the generalized hypercube structure, the multidimensional torus structure, and system area networks with a hierarchy of complete graphs or digraphs.

Fat-tree networks are reconfigurable networks [6, 7] with conflict-free transmission only according to predetermined schedules for specific packet permutations. In the case of arbitrary permutations, these networks turn out to be blocking; permutations in them are implemented in several jumps between network nodes. Networks with the generalized hypercube structure are not even reconfigurable [8, 9]. They can be made such by increasing the number of channels in some dimensions. For arbitrary permutations, networks with the multidimensional torus structure cannot transmit packets over direct channels [10, 11]. They implement permutations in several jumps between network nodes. Networks with a hierarchy of complete graphs or digraphs have similar properties [12, 13].

In the papers [1–4], the method of constructing non-blocking photon switches with static self-routing was based on four fundamental principles:

1. Applying a four-channel switch of a new structure, which is dual by the conflict resolution approach. It combines the bus method (separation of conflicting signals to different cycles in one channel) and the switch method (separation of conflicting signals to different channels).
2. Assuming the parallel transmission of signaling and control information for switches at different frequencies for each data bit. This assumption eliminates the problem of synchronizing signals from different channels.
3. Cascading switches so that the I th channel of the J th switch on one stage is connected to the J th channel of the I th switch on the next stage. With such exchange links, the previous and next stages must include the same number of switches, each having the same number of channels. This method allows constructing multichannel switches with a small number of stages.
4. Balancing the speed and complexity of a multistage switch using the invariant extension method of system area networks [5], which preserves the non-blocking property and speed of the switch with an increase in the number of its channels. This method involves an extended scheme (circuit) basis consisting of both $p \times p$ switches for p channels and pairs of $1 \times p$ multiplexers M_p and $p \times 1$ demultiplexers D_p , where $p \geq 2$.

In Section 1, the works on photonics close to the problem under study are considered; see [14–23]. Section 2 describes a technique for constructing non-blocking photon switches, used by the authors in the previous works [1–4]. Section 3 presents a switch with the dual quasi-complete graph topology as the basis for constructing non-blocking self-routing switches with channel fault tolerance. In Section 4, two-stage non-blocking self-routing switches with one-channel and two-channel fault tolerance are constructed; in addition, their characteristics – speed and complexity – are estimated. In the Conclusions, the results are summarized, and their possible development and generalization for constructing four-stage and eight-stage non-blocking fault-tolerant switches of high scalability are discussed.

1. RELEVANT RESEARCH INTO PHOTONICS: A SURVEY

Currently, a promising direction in the development of high-performance supercomputers is using photonic technologies [14, 15]. Much attention is paid

to the design of photon switches that can significantly accelerate information transmission and processing [16–23]. In the paper [16], photon switches were constructed using such well-known topologies as Clos, Benes, the matrix switch, Dragonfly, etc. A disadvantage of these circuits is external electronic control: on the one hand, it allows for non-blocking switching; on the other hand, it restricts the speed of optical circuits on average to several tens of nanoseconds [17, 18]. For example, an externally controlled matrix switch with dimensions 240×240 was presented in [23] based on the MEMS technology. Its speed is 400 ns. Another problem in implementing optical switches is the need to use optical buffer devices [18]. In [22], a method was proposed for transmitting and processing information, which reduces the size of the optical buffer by 4.2 times and increases the speed by 2.6 times. In [18], a new principle of optical switching—segment switching – was proposed. It differs from channel switching: when establishing a connection, not the entire route from the source to the sink is reserved in advance, but its smaller segments. This approach improves network performance and reduces the number of conflicts.

In [14–23], individual components of photon networks were presented, but without any method for combining them into networks with specified properties. On the contrary, in [1–4], a technique for constructing non-blocking photon switches with static self-routing was described.

2. A METHOD FOR CONSTRUCTING NON-BLOCKING PHOTON SWITCHES

In the papers [1–4], one of the circuits of a one-chip dual 4×4 switch is a two-stage circuit of four demultiplexers and four multiplexers with feedback links through the delay lines. The switch stages are interconnected by exchange links. If the complexity of multiplexers M_4 and demultiplexers D_4 is measured by the number of switching points (equal to 4), then the switching complexity of the switch is $S_1 = 32$. The combination of two control frequencies uniquely determines the demultiplexer's mode in which the information signal can be directed to one of the four outputs.

The signals from the demultiplexer's outputs are supplied to the multiplexer's inputs. One of them is passed to the output, and the rest return to their delay lines DL_δ with a duration of δ cycles (signals). The switch implements dynamic signal delay using the feedback links through DL_δ , which requires a higher period to transmit the signals.



The dual switch SS4 provides non-blocking with static self-routing under an appropriate length δ of the delay line. The value δ depends on the number of the stage in which the switch SS4 is used.

For the first stage, $\delta = 1$. Let four signals of duration T_0 be simultaneously supplied to the inputs of the switch SS4, all received in one cycle. With dynamic signal delay at the switch outputs, there are four possible options of signal placement: one at each output, two signals in a row at two outputs, one and three signals in a row at two outputs, and four signals in a row at one output. As a result, the switch SS4 will be non-blocking for any input traffic when the period T_1 of information signals is four cycles. Here, the underline denotes the values obtained in [1–3].

Therefore, the non-blocking self-routing switch SS4 has the following performance characteristics: the signal period $T_1 = 4 = \underline{N}_1$ cycles, the number of channels $\underline{N}_1 = 4$, and the switching complexity $\underline{S}_1 = 32 = \underline{N}_1^{5/2}$.

The papers [1–3] considered the two-stage 16×16 switch S_{216} with exchange links, consisting of four switches SS4 on each stage. The first stage includes DL1, and the second stage includes DL0 (no delay lines). For an arbitrary permutation of packets, S_{216} turned out to be a non-blocking self-routing switch with the following performance characteristics: the number of channels $\underline{N}_2 = 16$, the signal period $T_2 = 4 = \underline{N}_2^{1/2}$ cycles, and the switching complexity $\underline{S}_2 = 2 \cdot 4 \cdot 32 = 256 = \underline{N}_2^2$.

The papers [1–3] considered the four-stage 256×256 switch S_{4256} with exchange links, consisting of 16 switches S_{216} on each stage. It consists of four stages of switches SS4. The first stage includes DL1, the second stage includes DL4, the third stage includes DL15, and the fourth stage includes DL0 (no delay lines). For an arbitrary permutation of packets, S_{4256} turned out to be a non-blocking self-routing switch with the following performance characteristics: the signal period $T_4 = 49 \approx 3\underline{N}_4^{1/2}$ cycles, the number of channels $\underline{N}_4 = 256$, and the switching complexity $\underline{S}_4 = 2 \cdot 16 \cdot 256 = 8192 = \underline{N}_4^{1.625}$. Note the large digit period (low speed) and small complexity of this switch.

In the papers [1–3], the “speed–complexity” ratio was balanced using the invariant extension of small-period switches. In particular, the switch S_{216} was extended through external multiplexers M4 and demultiplexers D4. This method of increasing the number of channels can be called the external parallelization of a non-blocking network. As a result, the non-blocking self-routing switch S_{364} was constructed, consisting of 16 switches S_{216} and 64 demultiplexers D4 and multiplexers M4. This switch has the follow-

ing performance characteristics: the number of channels $\underline{N}_3 = 64$, the signal period $T_3 = 4 = \underline{N}_3^{1/3}$ cycles, and the switching complexity $\underline{S}_3 = 16 \cdot 256 + 4 \cdot 128 = 4608 = \underline{N}_3^{2.028}$.

In the paper [4], a non-blocking switch with the quasi-complete digraph topology was used not only to extend two-stage switches but immediately to construct them. The resulting switches have a higher speed (a smaller signal period) and more channels than those proposed in [1–3].

A switch with the quasi-complete digraph topology is constructed from dual $p \times p$ switches (DS p) together with $1 \times p$ demultiplexers (D p) and $p \times 1$ multiplexers (M p). The non-blocking switch SF N_1 with $N_1 = p^2$ channels consists of N_1 switches DS p , demultiplexers D p , and multiplexers M p . Its circuits for $p = 2$ and $p = 4$ are shown in Fig. 1 and 2, respectively. The values obtained below will be given without the underline.

The switch SF N_1 has the switching complexity $S_1 = 2pN_1 + 2p^2N_1 = 2p^3(p+1)$ and the channel complexity $L_1 = 2pN_1 = 2p^3$ (the number of channels without internal channels DS p). In Table 1, their values are expressed by an exponential relationship through the number of channels. In what follows, the complexity specified in this way will be called the exponential complexity.

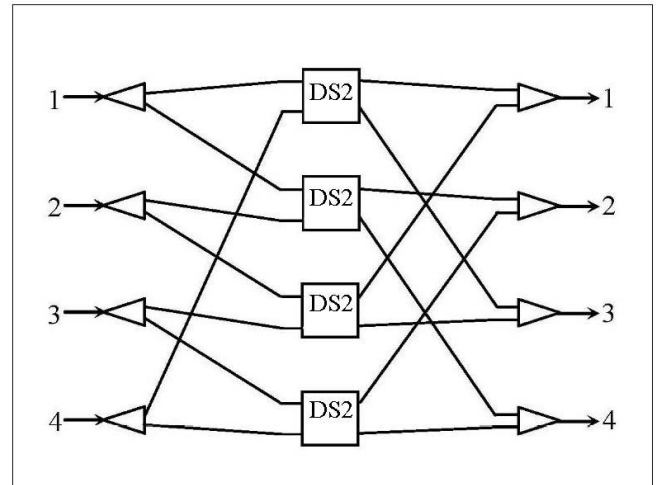


Fig. 1. Dual switch SF4 with quasi-complete graph topology. Boxes correspond to switches DS2, and triangles to demultiplexers D2 and multiplexers M2.

Table 1

Performance characteristics of switch SF N_1

p	2	4	6	8
N_1	4	16	36	64
S_1	$N_1^{2.79}$	$N_1^{2.33}$	$N_1^{2.24}$	$N_1^{2.19}$
L_1	N_1^2	$N_1^{1.75}$	$N_1^{1.69}$	$N_1^{1.67}$

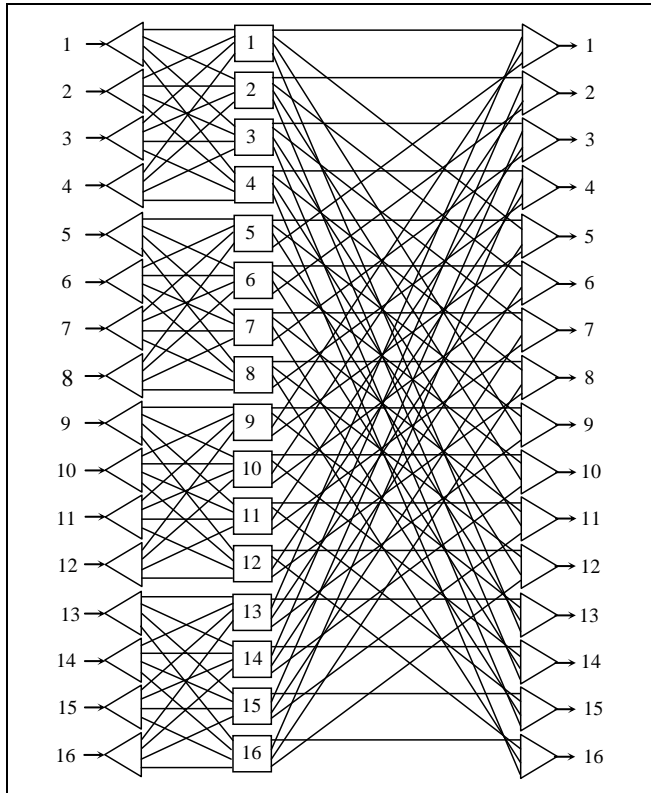


Fig. 2. Dual switch SF16 with quasi-complete digraph topology. Boxes correspond to switches DS4, and triangles to multiplexers M4 and demultiplexers D4.

From non-blocking dual switches SF4, it is possible to construct a two-stage network N_216 with exchange links [5]. It consists of two stages with four SF4 on each stage. Unfortunately, the network N_216 is a blocking network: implementing an arbitrary permutation in N_216 is not reduced to implementing permutations on switches SF4 of the first and second stages. Signal conflicts can occur on a stage of multiplexers M2; see the gray segment in Fig. 5 of [5]. Signals with the same cycles can conflict at different multiplexer inputs.

The invariant internal parallelization method [4] transforms the network N_216 into the non-blocking self-routing switch S_216 with the signal period $T_2 = 2$ cycles.

For any $p > 2$, the same approach yields the non-blocking self-routing switch S_2N_2 with $N_2 = p^4$ channels and the digit period $T_2 = p$. The switching complexity of the switch S_2N_2 is given by the recursive formula $S_2 = N_1S_1 + pN_1S_1$; the channel complexity, by the recursive formula $L_2 = N_1L_1 + pN_1L_1$. The performance characteristics of the switch S_2N_2 are presented in Table 2.

Note that the switching and channel complexity of the switch S_2N_2 is significant-

ly less than that of the switch SFN_1 (see Table 1). In other words, internal parallelization opens up the possibility of reducing the exponential complexity of a two-stage switch compared to the original switch SFN_1 .

In the paper [4], the invariant extension method with external parallelization of non-blocking switches was also used through additional compound multiplexers Md and demultiplexers Dd with $d = p^\alpha$ ($1 \leq \alpha \leq 4$). It yielded a set of non-blocking self-routing switches with the highest scalability.

Well, the papers [1–4] considered various approaches to constructing non-blocking photon switches of high scalability. However, the fault tolerance of such networks was not touched upon. The photon switches [1–4] have only one physical path of information transmission between any “input – output” pair. In view of this feature, the development of new switching circuits with many such channels will increase their fault tolerance and is a topical problem.

In this paper, we construct non-blocking self-routing photon switches with $(\sigma - 1)$ -channel fault tolerance. It is provided by replacing the dual quasi-complete digraph topology with the dual quasi-complete graph one, in which the total number of channels can be exchanged for the number of different channels between users. This problem is solved using the method of internal parallelization.

3. A NON-BLOCKING SELF-ROUTING SWITCH WITH THE QUASI-COMPLETE GRAPH TOPOLOGY

Dual switches DSp , together with demultiplexers Dp and multiplexers Mp , can be used to construct a non-blocking self-routing switch with the quasi-complete graph topology for N_1 channels – the dual switch $SQG(N_1, p, \sigma)$ – in which $N_1 = p(p - 1)/\sigma + 1$ and σ specifies the number of duplex channels between any two users through different switches DSp .

The switch $SQG(N, p, \sigma)$ is isomorphic to such a mathematical object as an incomplete balanced sym-

Table 2

Performance characteristics of switches S_2N_2

p	$N_2 = p^4$	$T_2 = p$	S_2	L_2
2	16	2	$576 = N_2^{2.29}$	$224 = N_2^{1.95}$
3	81	3	$7776 = N_2^{2.04}$	$2187 = N_2^{1.75}$
4	256	4	$51\,200 = N_2^{1.96}$	$11\,264 = N_2^{1.68}$
5	625	5	$225\,000 = N_2^{1.91}$	$40\,625 = N_2^{1.65}$
6	1296	6	$762\,048 = N_2^{1.89}$	$116\,640 = N_2^{1.63}$
7	2401	7	$2\,151\,296 = N_2^{1.87}$	$285\,719 = N_2^{1.61}$
8	4096	8	$5\,308\,416 = N_2^{1.86}$	$622\,592 = N_2^{1.60}$



metric block design $B(N, p, \sigma)$ [24]. It contains N blocks and N elements arranged so that each block contains exactly p different elements, each element appears in exactly p different blocks, and each pair of elements appears in exactly σ blocks. The block design $B(N, p, \sigma)$ sets the maximum value of N for given values of p and σ .

In the switch representation, blocks are interpreted as switches DS_p ; elements, as users of degree p (with p duplex ports); the entry of an element into a block, as a connection of each switch DS_p to each user through a duplex channel. In this case, a single-port user is connected to different switches DS_p through a pair of demultiplexers D_p and multiplexers M_p . Figure 3 shows the circuit of the switch $SQG(7, 4, 2)$ with duplex channels and a combined pair of demultiplexers D_p and multiplexers M_p in one duplex channel splitter (hub).

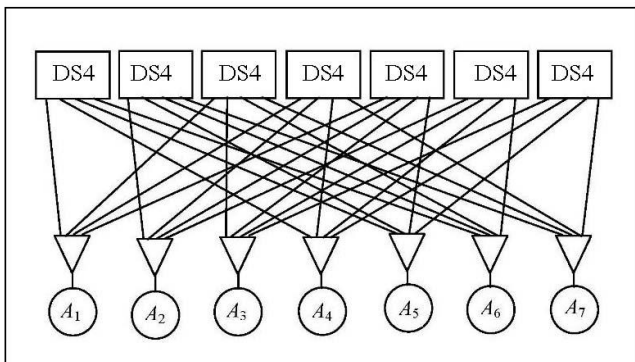


Fig. 3. Non-blocking network based on switch $SQG(7, 4, 2)$ with two different channels between any users.

To construct a multichannel non-blocking network, we use the switch $SQG(N, p, \sigma)$ with dividing each duplex channel into two simplex ones: the input channel from the demultiplexers D_p and the output channel from the multiplexers M_p . Figure 4 shows the circuit of the switch $SQG(7, 4, 2)$ in this format. The $(\sigma - 1)$ -channel fault tolerance is understood as the presence of σ different paths through different switches DS_p between the input demultiplexers D_p and the output multiplexers M_p .

In the general case of a more complex network based on the switch $SQG(N, p, \sigma)$, its $(\sigma - 1)$ -channel fault tolerance is interpreted as the presence of σ different paths through different segments of the network between the input demultiplexers D_p and the output multiplexers M_p .

Unfortunately, quasi-complete graphs do not exist for all values of the parameters p and σ ; for each pair

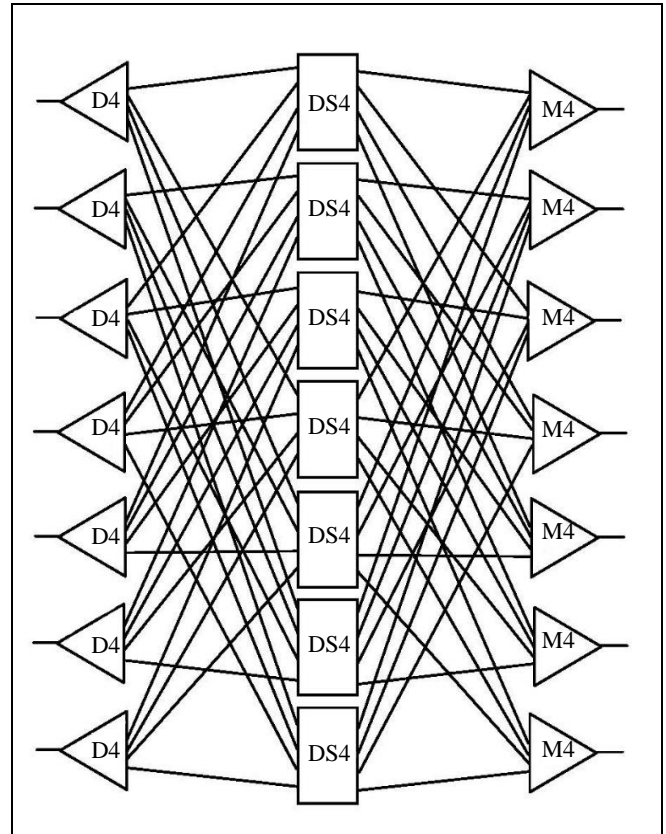


Fig. 4. Non-blocking self-routing switch $SQG(7, 4, 2)$ with two different paths between any input demultiplexers $D4$ and output multiplexers $M4$.

of p and σ , they have to be constructed by enumeration. Table 3 presents the values of N quasi-complete graphs for small values of these parameters. Empty cells indicate graphs not existing by definition. The dashes in the cells correspond to the graphs that cannot exist according to the theory. Finally, the crossed-out values denote the graphs that have not yet been constructed.

Table 3

Parameters of switch $SQG(N, p, \sigma)$

	p									
σ	3	4	5	6	7	8	9	10	11	12
1	7	11	21	31	–	57	73	91	111	133
2	4	7	11	16	–	–	37	–	56	–
3	3	5		11	15		25	31	45	

The need to construct fault-tolerant networks requires some effective filling of empty cells in Table 3. For this purpose, one-extended switches $SQG(N^*, p, \sigma/\sigma + 1)$ were constructed in [25], in which a small part of the users are connected through $(\sigma + 1)$ differ-

ent paths, and the rest through exactly σ different paths. The numbers N and N^* of nodes in the above-mentioned block designs are given in Table 4. (The latter numbers are underlined.)

The switches $SQG(N_1, p, \sigma)$ have one layer of output multiplexers Mp , $V_1 = (p + 1)N_1$ in total.

Table 4

Switch $SQG(N, p, \sigma)$ and one-extended block designs of switches $SQG(N^*, p, \sigma|\sigma + 1)$

σ	p									
	3	4	5	6	7	8	9	10	11	12
1	7	11	21	31	<u>39</u>	57	73	91	<u>95</u>	133
2	4	7	11	<u>15</u>	<u>21</u>	<u>27</u>	37	<u>42</u>	<u>51</u>	<u>63</u>
3	3	5	<u>7</u>	11	15	<u>19</u>	<u>23</u>	<u>29</u>	<u>36</u>	<u>43</u>

4. A NON-BLOCKING SELF-ROUTING TWO-STAGE FAULT-TOLERANT SWITCH

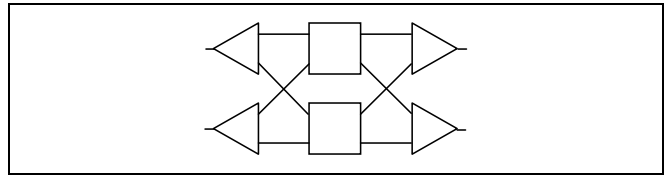
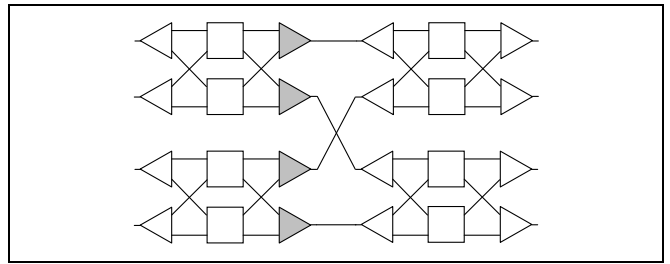
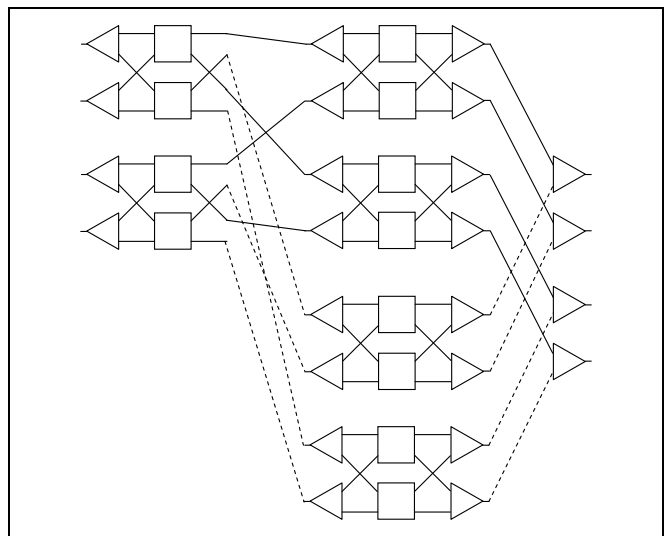
We introduce the notion of p -partitions of packets transmitted through some cross-section of a network under a permutation. All packages are divided into groups of variable composition, each containing at most p packages. For a common permutation of packets, a 1-partition occurs at the input and output of the network. A transmission in which a 1-partition occurs at the network input and a p -partition on some cross-section will be called a p -permutation. A transmission in which a p -partition occurs at the network input and a 1-partition on a given cross-section will be called an inverse p -permutation.

For the dual switch $SQG(N_1, p, \sigma)$, this cross-section is through the inputs of the output multiplexers and called the output cross-section. According to the property of the dual switch DSp , a p -partition occurs on the output cross-section of the dual switch $SQG(N_1, p, \sigma)$ for any traffic. In particular, a direct p -permutation occurs under any permutation in the dual switch $SQG(N_1, p, \sigma)$.

Lemma 1. The dual switch $SQG(N_1, p, \sigma)$ is non-blocking under any inverse p -permutation and preserves $(\sigma - 1)$ -channel fault-tolerance.

P r o o f. The second assertion is based on the non-blocking property of the switch $SQG(N_1, p, \sigma)$ on an arbitrary 1-permutation and the fact that an inverse p -permutation consists of sparse 1-permutations separated by different cycles. The number of different paths between any source and sink is preserved in each such permutation. ♦

Multichannel fault-tolerant networks will be further constructed on an example of the switches $DS2$ and $SQG(2, 2, 2)$ (the minimal switch with single-channel fault-tolerance, see Fig. 5). First, a blocking two-stage four-channel N_24 network with exchange links is constructed, as shown in Fig. 6. It turns out to be blocking due to possible conflicts in the multiplexers $M2$ of the first stage, shaded in the figure. One-channel fault tolerance is also violated on them. With two copies of the second stage and internal parallelization, the network N_24 is transformed into a non-blocking self-routing switch S_24 (Fig. 7).


 Fig. 5. Two-channel non-blocking switch $SQG(2, 2, 2)$ with $T_2 = 2$ and one-channel fault tolerance.

 Fig. 6. Two-stage network N_24 with exchange links.

 Fig. 7. Four-channel non-blocking self-routing switch S_24 with one-channel fault tolerance.



This transformation is carried out by removing all multiplexers M2 from the first stage and dividing their input channels between two copies of the second stage. The cut multiplexers M2 combine the same-named output channels of the copies of the second stage, forming the first dimension circuit. Sparse direct p -permutations with disjoint inputs-outputs occur at the inputs of its switches SQG(2, 2, 2). By Lemma 1, they are implemented without conflict. In addition, any two paths on the first stage are divided between the two copies of the second stage, which preserves one-channel fault tolerance.

As a result, the switch S_{24} is non-blocking with static self-routing. It has one-channel fault tolerance: all paths between the input demultiplexers D2 and the output multiplexers M2 pass through different copies of the second stage of the network N_{24} in the first dimension circuit since $p = \sigma = 2$.

In the general case ($p > 2$ and $\sigma = 2$), the switches DS_p and $SQG(N_1, p, 2)$ are used. The minimal non-trivial dual graph of the switch $SQG(N_1, p, 2)$ is the dual switch $SQG(4, 3, 2)$; see Fig. 8.

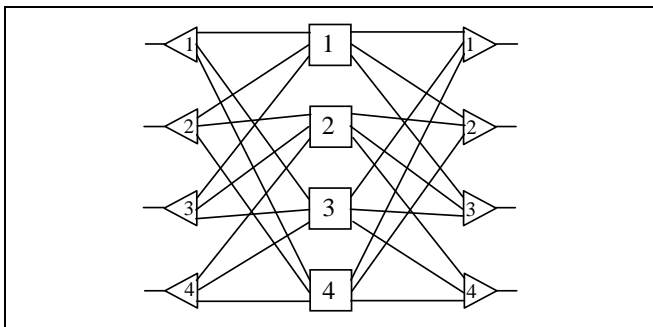


Fig. 8. Non-blocking self-routing dual switch SQG(4, 3, 2) with one-channel fault tolerance.

The two-stage network N_2N_2 with exchange links and $N_2 = N_1^2$ is constructed by analogy. Then it is transformed into a non-blocking self-routing switch S_2N_2 using internal parallelization over p copies of the second stage, which form the first dimension circuit. The switch S_2N_2 is one-channel fault-tolerant: it has p paths in the first dimension circuit, and $p \geq 2$. For $p = 3$, the switch S_{216} is shown in Fig. 9.

The switching complexity of a dual $p \times p$ switch has the formula $S_0 = 2p^3$. Then the switching complexity of the switch $SQG(N_1, p, \sigma)$ is given by $S_1 = N_1S_0 + 2pN_1$; its channel complexity, by $L_1 = 2pN_1$.

The switch S_2N_2 uses all the multiplexers of the switches $SQG(N_1, p, \sigma)$ without adding new ones, and their input and output channels are redirected only.

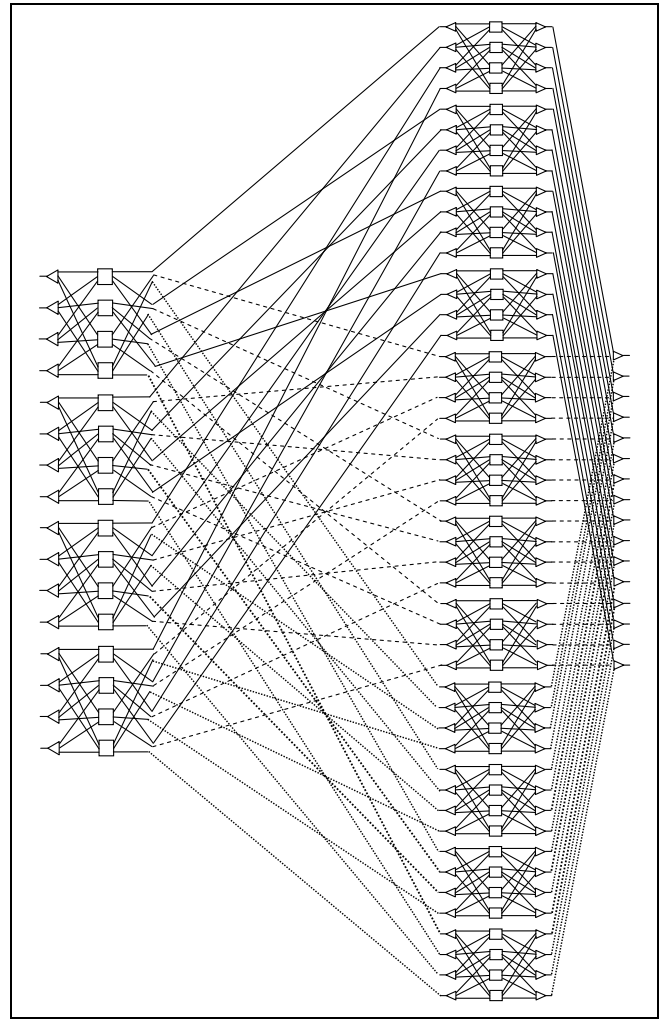


Fig. 9. Non-blocking self-routing switch S_{216} with one-channel fault tolerance.

Hence, we conclude that by construction, the switch S_2N_2 has the switching complexity $S_2 = N_1S_1 + pN_1S_1$ and the channel complexity $L_2 = N_1L_1 + pN_1L_1$. The performance characteristics of the switch S_2N_2 for different σ are presented in Tables 5 and 6.

We compare the complexities of the switch S_2N_2 with one-channel fault tolerance and a switch with the complete graph topology and duplicated channels, e.g., for $p = 4$ ($N_2 = 49$). The latter switch has the switching complexity $S_2 = N_2^{2.35}$ and the channel complexity $L_2 = N_2^{2.17}$. Obviously, their switching complexities are commensurable, but the channel complexity of the switch S_2N_2 is significantly smaller. In addition, the digit period of the switch with the complete graph topology is four times smaller. However, it seems too early to finalize the conclusions.

Generally, the first dimension circuit is defined as follows: p copies of the second stage of the network N_2N_2 , with the inputs of the cut multiplexers Mp connected to them and the outputs combined by these multiplexers Mp , form the first dimension circuit.

Table 5

**Performance characteristics of switches S_2N_2
with one-channel fault tolerance**

p	N_1	$N_2 = N_1^2$	$T_2 = p$	S_2	L_2
2	2	4	2	$144 = N_2^{3.58}$	$56 = N_2^{2.9}$
3	4	16	3	$1536 = N_2^{2.65}$	$432 = N_2^{2.19}$
4	7	49	4	$9800 = N_2^{2.37}$	$2156 = N_2^{1.97}$
5	11	121	5	$43\,560 = N_2^{2.23}$	$7865 = N_2^{1.87}$
6	15	225	6	$132\,300 = N_2^{2.18}$	$20\,250 = N_2^{1.84}$
7	21	441	7	$395\,136 = N_2^{2.12}$	$52\,479 = N_2^{1.79}$
8	27	729	8	$944\,784 = N_2^{2.09}$	$110\,808 = N_2^{1.77}$

The switch S_2N_2 contains only one circuit of the first dimension, whose outputs are the outputs of the switch itself.

To provide two-channel fault tolerance, we use graphs of $SQG(N_1, p, 3)$. The minimal graph of $SQG(N_1, p, 3)$ is $SQG(3, 3, 3)$, and the minimal nontrivial graph is $SQG(5, 4, 3)$; see Fig. 10.

In the general case ($p \geq 3$ and $\sigma = 3$, see Table 4), the network N_2N_2 with $N_2 = N_1^2$ is constructed by analogy. Then, as described above, it is transformed into a non-blocking self-routing switch S_2N_2 using internal parallelization with p copies of the second stage. The switch S_2N_2 has two-channel fault tolerance.

Table 6

**Performance characteristics of switches S_2N_2
with two-channel fault tolerance**

p	N_1	$N_2 = N_1^2$	$T_2 = p$	S_2	L_2
3	3	9	3	$864 = N_2^{3.08}$	$243 = N_2^{2.50}$
4	5	25	4	$5000 = N_2^{2.65}$	$1100 = N_2^{2.18}$
5	7	49	5	$17\,640 = N_2^{2.52}$	$3185 = N_2^{2.07}$
6	11	121	6	$71\,148 = N_2^{2.31}$	$10\,890 = N_2^{1.94}$
7	15	225	7	$201\,600 = N_2^{2.26}$	$26\,775 = N_2^{1.88}$
8	19	361	8	$467\,856 = N_2^{2.22}$	$54\,872 = N_2^{1.85}$

We compare the complexities of the switch S_2N_2 with two-channel fault tolerance and a switch with the complete graph topology and tripled channels, e.g., for $p = 4$ ($N_2 = 25$). The latter switch has the switching complexity $S_2 = N_2^{2.55}$ and the channel complexity $L_2 = N_2^{2.34}$. Obviously, their switching complexities are commensurable, but the channel complexity of the switch S_2N_2 is significantly smaller. In addition, the digit period of the switch with the complete graph topology is four times smaller.

We draw a given cross-section in the switch S_2N_2 after the dual switches $SQG(N_1, p, \sigma)$, i.e., at the input of

the output multiplexers layers. Then the following lemma holds for any values p and σ .

Lemma 2. The dual switch S_2N_2 has a direct p -permutation on the indicated cross-section. It is a non-blocking switch with static routing on any reverse p -permutation for any p and has $(\sigma - 1)$ -channel fault tolerance.

P r o o f. The first assertion is based on using the dual switch $SQG(N_1, p, \sigma)$ and Lemma 1 for it. The second assertion is based on the non-blocking property of the switch $SQG(N_1, p, \sigma)$ and the fact that an inverse p -permutation consists of sparse 1-permutations separated by different cycles. The union of the sparse alternative permutations through the cut multiplexers M_p generates no conflicts.

The property of $(\sigma - 1)$ -channel fault tolerance follows from the fact that all paths between the input demultiplexers

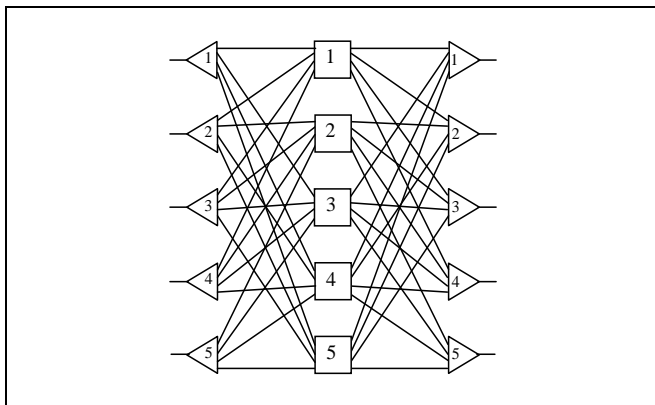


Fig. 10. Non-blocking self-routing dual switch $SQG(5, 4, 3)$ with two-channel fault tolerance.



D_p and the output multiplexers M_p pass through p copies of the second stage of the network N_2N_2 in the first dimension and $p \geq \sigma$. On the other hand, only σ different paths come from any source in the original dual switch $SQG(N_1, p, \sigma)$, and their failure will break the network connectivity. ♦

CONCLUSIONS

This paper has proposed a method for constructing a new type of non-blocking self-routing photon networks in the form of a two-stage switch with $(\sigma - 1)$ -channel fault tolerance. The new switch differs from those proposed in the authors' previous works by a new property called channel fault tolerance. This property has been achieved by replacing the basic dual switch with the quasi-complete digraph topology from [5] with a dual switch with the quasi-complete graph topology. Due to such a replacement, two or three communication channels are physically implemented between each "input-output" pair, in contrast to one channel in the original photon switches [1–4]. Each additional channel can be used in the event of a failure of the main channel. With the increase in the fault tolerance of photon switches, their channel and switching complexities have grown expectedly. This paper has presented expressions for calculating both performance characteristics depending on the number of channels.

In addition, the proposed designs of fault-tolerant two-stage switches are poorly scaled by external invariant parallelization with additional multiplexers and demultiplexers [5], which was used in [4] to scale a two-stage switch without channel fault tolerance based on the quasi-complete digraph topology. For example, the switch S_{216} with one-channel fault tolerance (Fig. 12) can be extended using multiplexers M_3 and demultiplexers D_3 only up to the switch S_{220} and, using the repeated extension, to the switch S_{225} .

Therefore, the following problem arises to improve photon networks of large dimensions: construct a non-blocking self-routing fault-tolerant multistage switch of high scalability. This problem can be solved by extending such a switch into four-stage and eight-stage non-blocking switches based on the development and application of a generalized internal parallelization method.

Note that the exponential switching complexity can be significantly reduced by increasing the number of network stages and using a generalized internal parallelization method for each stage. This approach to scaling fault-tolerant photon switches is a continuation of this study and will be considered separately.

REFERENCES

1. Barabanova, E.A., Vytovtov, K.A., and Podlazov, V.S., Multi-stage Switches for Optical and Electronic Supercomputer Systems, *Proceedings of the 8th National Supercomputer Forum (NSCF-2019)*, Pereslavl-Zalessky, 2019. URL: http://2019.nscf.ru/TesisAll/02_Apparatura/037_BarabanovaE_A.pdf. (In Russian.)
2. Barabanova, E.A., Vytovtov, K.A., Vishnevsky, V.M., and Podlazov, V.S., The New Principle for the Construction of Optical Information Processing Devices for Information-Measuring Systems, *Sensors and Systems*, 2019, no. 9, pp. 3–9. (In Russian.)
3. Barabanova, E., Vytovtov, K., Podlazov, V., and Vishnevskiy, V. Model of Optical Non-blocking Information Processing System for Next-generation Telecommunication Networks, *Proceedings of the 22nd International Conference on Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2019)*, Moscow, 2019. Communications in Computer and Information Science, vol. 1141, Cham: Springer, pp. 188–198. DOI: 10.1007/978-3-030-36625-4_16.
4. Barabanova, E.A., Vytovtov, K.A., and Podlazov, V.S., Two-Stage Dual Photon Switches in an Extended Scheme Basis, *Control Sciences*, 2021, no. 1, pp. 69–81.
5. Karavai, M.F. and Podlazov, V.S., An Invariant Extension Method for System Area Networks of Multicore Computational Systems. An Ideal System Network, *Automation Remote Control*, 2010, vol. 71, no. 12, pp. 2644–2654.
6. Pipenger, N., On Rearrangeable and Non-Blocking Switching Networks, *J. Comput. Syst. Sci.*, 1978, vol. 17, pp. 307–311.
7. Scott, S., Abts, D., Kim, J. and Dally, W., The Black Widow High-radix Clos Network, *Proc. 33rd Intern. Symp. Comp. Arch. (ISCA'2006)*, Boston, 2006. https://www.researchgate.net/publication/4244660_The_Black_Widow_High-Radix_Clos_Network.
8. Gu, Q.P., and Tamaki, H., Routing a Permutation in Hypercube by Two Sets of Edge-Disjoint Paths, *J. of Parallel and Distributed Comput.*, 1997, vol. 44, no. 2, pp. 147–152.
9. Lubiw, A. Counterexample to a Conjecture of Szymanski on Hypercube Routing, *Inform. Proc. Let.*, 1990, vol. 35(2), pp. 57–61.
10. Alverson, R., Roweth, D., and Kaplan, L., The Gemini System Interconnect, *Proc. 18th IEEE Symposium on High Performance Interconnects*, 2009, pp. 3–87.
11. Ajima, Y., Inoue, T., Hiramoto, S., and Shimiz, T., Tofu: Interconnect for the K Computer. URL: https://www.researchgate.net/publication/265227674_Tofu_Int_erconnect_for_the_K_computer.
12. Arimili, B., Arimili, R., Chung, V., et al., The PERCS High-Performance Interconnect, *Proc. 18th IEEE Symposium on High Performance Interconnects*, 2009, pp. 75–82.
13. De Sensi, D., Di Girolamo, S., McMahon, K.H., Roweth, D., and Hoefler, T., An In-Depth Analysis of the Slingshot Interconnect, *arXiv: 2008.08886v1*, August 20, 2020. URL: https://www.researchgate.net/publication/343786515_An_In-Depth_Analysis_of_the_Slingshot_Interconnect.
14. Stepanenko, S., Structure and Implementation Principles of a Photonic Computer, *EPJ Web of Conferences* 224, vol. 04002, 2019. DOI: <https://doi.org/10.1051/epjconf/201922404002>

15. Babae, A., Momeni, A., Abdolali, A., and Fleury, R., Parallel Optical Computing Based on MIMO Metasurface Processors with Asymmetric Optical Response, *Phys. Rev. Applied*, 2021, vol. 15, article no. 044015.
16. Cheng, Q., Glick, M., and Bergman, K., Optical Interconnection Networks for High Performance Systems, in *Optical Fiber Telecommunications*, vol. VII, 2019, pp. 785–825.
17. Maniotis, P., Dupuis, N., Kuchta, D.M., Taubenblatt, M.A., and Lee, B.G., Intra-Node High-Performance Computing Network Architecture with Nanosecond-Scale Photonic Switches, *Journal of Optical Communications and Networking*, 2020, vol. 12(12), pp. 367–377.
18. Duro, J., Petit, S., Góme, M.E., and Sahuquillo, J., Segment Switching: A New Switching Strategy for Optical HPC Networks, *IEEE Access*, 2021, vol. 9, pp. 43095–43106.
19. Lee, B.G., Dupuis, N., Pepeljugoski, P., et al., Silicon Photonic Switch Fabrics in Computer Communications Systems, *Journal of Lightwave Technology*, 2015, vol. 33(4), pp. 768–777.
20. Krishnamoorthy, A.V., Ho, R., Zheng, X., et al., Computer Systems Based on Silicon Photonic Interconnects, *Proceedings of the IEEE*, 2009, vol. 97(7), pp. 1337–1361.
21. Cheng, Q., Rumley, S., Bahadori, M., and Bergman, K., Photonic Switching in High Performance Datacenters, *Optics Express*, 2018, vol. 26(12), article no. 16022.
22. Meyer, H., Sancho, J.C., Mrdakovic, M., et al., Optical Packet Switching in HPC. An Analysis of Applications Performance, *Future Generation Computer Systems*, 2018, vol. 82, pp. 606–616.
23. Seok, T.J., Kwon, K., Henriksson, J., et al., Wafer-Scale Silicon Photonic Switches beyond Die Size Limit, *Optica*, 2019, vol. 6(4), pp. 490–494.
24. Hall, M., *Combinatorial Theory*, Waltham: Blaisdell Publishing Company, 1967.
25. Karavai, M.F. and Podlazov, V.S., Expanded Block-Diagrams for Ideal System Area Networks, *Control Sciences*, 2012, no. 4, pp. 45–51. (In Russian.)

This paper was recommended for publication by V.G. Lebedev, a member of the Editorial Board.

Received March 11, 2021, and revised May 13, 2021.

Accepted May 13, 2021.

Author information

Barabanova, Elizaveta Aleksandrovna. Dr. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, ✉ elizavetaalex@yandex.ru

Vytovtov, Konstantin Anatol'evich. Dr. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, ✉ vytovtov_konstan@mail.ru

Podlazov, Viktor Sergeevich. Dr. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, ✉ podlazov@ipu.ru

Cite this article

Barabanova, E.A., Vytovtov, K.A., Podlazov, V.S. Non-blocking Fault-Tolerant Two-Stage Dual Photon Switches. *Control Sciences* **4**, 67–76 (2021). <http://doi.org/10.25728/cs.2021.4.7>

Original Russian Text © Barabanova, E.A., Vytovtov, K.A., Podlazov, V.S., 2021, published in *Problemy Upravleniya*, 2021, no. 4, pp. 82–92.