

Design Model of Three-Stage Folded Clos Network with Decoupled First Stage Guaranteeing Admissible Blocking Probability

EIJI OKI^{1,*}, RYOTARO TANIGUCHI¹, KAZUYA ANAZAWA^{2,1}, AND TAKERU INOUE²

¹Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

²NTT Network Innovation Laboratories, NTT Corporation, Kanagawa, Japan

*oki@i.kyoto-u.ac.jp

Compiled September 26, 2025

Some data center networks have implemented optical circuit switching (OCS) to replace electrical packet switching, achieving reduced power consumption, lower latency, and increased capacity. This paper focuses on designing an OCS network to maximize the switching network size, defined as the number of terminals connected under the constraints of a fixed number of identical $N \times N$ switches and a specified maximum admissible blocking probability. An existing design for a three-stage folded Clos network encounters a limitation where the network size ceases to grow as the number of switches increases beyond a certain threshold. To address this, we propose a design model for a three-stage folded Clos network with the decoupled first stage, named 3dF, to maximize network switching size while guaranteeing an admissible blocking probability. The 3dF model introduces an input-output layer that combines the first- and second-stage switches into a single, large switch and decouples the switching functions originally handled by a single switch at the first stage into two distinct switches. We formulate this model as an optimization problem and employ an exhaustive search-based algorithm to identify the structure with the largest switching network size in non-increasing order while ensuring that specified constraints, such as blocking probability, are satisfied. We derive theoretical results for three-stage folded Clos variants and, through numerical analysis, show that the 3dF design achieves a larger switching network size than other variants under the SNB condition or a blocking probability guarantee.

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

1. INTRODUCTION

A data center network, consisting of routers and switches, plays a crucial role in handling massive traffic generated by emergent applications such as artificial intelligence (AI). Optical circuit switching (OCS) substitutes electrical packet switching to enable reduced power consumption, lower latency, and increased capacity [1–4]. By reserving a dedicated path for the full duration of data transfer, OCS supports dependable communication. Several data center networks have implemented OCS [1, 5, 6]. The Clos network architecture, featuring a multi-stage design that interconnects switches in a hierarchical fashion, is commonly used in data centers [7–12].

A folded Clos network integrates the input and output stages of a traditional three-stage Clos architecture [9]. This consolidation reduces the number of stages, simplifying the overall design and decreasing the complexity of the switching network. The folded Clos network architecture typically consists of two layers: the input-output layer and the intermediate layer. In this

configuration, the transmitter and receiver pair are connected within the input-output layer, giving rise to the term *folded*.

Mano *et al.* [13, 14] developed the twisted-folded Clos network, which mitigates certain switch-port constraints by introducing a *twisting* mechanism in the connections between the switches in the input-output layer and those in the intermediate layer. The authors formulated a twisted-folded Clos network design model using a limited number of identical $N \times N$ switches, satisfying the strict-sense non-blocking (SNB) condition, to maximize the network size, specifically the number of terminals connected. They focused on a two-stage twisted-folded Clos network (2TF) design.

The SNB condition in the 2TF configuration eliminates blocking but may still limit the potential size of the switching network. Allowing for a certain blocking level in the network design could introduce more flexibility. Taka *et al.* [15–17] investigated how to design twisted-folded Clos networks with a specified admissible blocking probability, aiming to expand the network size. The admissible blocking probability defines the probability that a

terminal's connection request is blocked due to internal congestion. Striking a balance between increasing the network size and maintaining an acceptable rate of successful connection requests involves trade-offs. Therefore, addressing the design issues that increase the switching network size, ensuring a given admissible blocking probability, is essential [18].

Jajszczyk [19] explored expanding a two-stage folded Clos network (2F) into a three-stage folded Clos network (3F) architecture to enlarge the switching network size. The expansion focused on two strategies: the input-output layer and the intermediate layer. The study assumed that the switch sizes for 3F could be customized, demonstrating that both expansion methods are isotopic. However, as noted in [13, 14], a more practical approach for data center networks is considering a fixed number of identical $N \times N$ switches when designing the switching network.

In this context, Taniguchi *et al.* [20] developed a design model for a three-stage twisted-folded Clos network (3TF) architecture, guaranteeing an admissible blocking probability to extend the switching network size beyond that of the 2TF model. 3TF is one of the folded-type Clos network architectures, where the first and second stages are twisted-folded. Using a limited number of identical $N \times N$ switches, they adopt expanding the input-output layer for 3TF—one of the two expansion strategies discussed in [19]—to further enlarge the switching network size.

Oki *et al.* [21] investigated the other expansion strategy of 3TF, which expands the intermediate layer. The work revealed that the input-output layer expansion strategy outperforms the intermediate expansion strategy under a limited number of identical $N \times N$ switches regarding the network switching size. Thus, in the following, we focus on the input-output layer expansion strategy to expand the network switching size of a folded-type Clos network.

However, 3TF still faces an inherent *scalability ceiling*, despite successfully extending the switching network size beyond that of the 2TF model [20, 21]. Further expansion becomes infeasible once the number of available $N \times N$ switches exceeds a certain threshold. This limitation arises because the twisted-folded topology, while effective in certain conditions, introduces structural inefficiencies in the first-stage design, ultimately constraining the network's growth potential. To address this bottleneck and achieve further scalability, a question arises: can we develop a new design model for a three-stage folded Clos network to outperform the 3TF network size by resolving this limitation?

This paper proposes a design model for a three-stage folded Clos network with the decoupled first stage, named 3dF, for application in OCS, ensuring an admissible blocking probability to maximize the network switching size. The proposed design model uses identical switches with the size $N \times N$, which is the same assumption as that of [15–17, 20]. The 3dF structure consists of input-output and intermediate layers. In 3dF, the first- and second-stage switches are combined into a single, large switch in the input-output layer. As a baseline, we introduce a three-stage folded Clos network (3F) architecture. Unlike the 3TF and 3F structures, 3dF introduces a distinct architectural approach by restructuring the first stage, which enhances network scalability by enabling a more balanced distribution of switching functions across layers, leading to a more efficient design. In 3dF, the input-output layer decouples the two switching functions—originally handled by a single switch at the first stage in 3TF and 3F—into two distinct switches. In this design, we do not adopt the rearrangement of existing optical links, as our focus is on preserving communication stability. We formulate

the proposed design model for 3dF as an optimization problem. To solve the design problem, we employ an exhaustive search-based algorithm [20, 21]. Rather than finding a feasible solution for every combination of natural number decision variables, this algorithm identifies the structure with the largest switching network size in non-increasing order while ensuring that specified constraints, such as blocking probability, are satisfied. Our study provides a rigorous theoretical analysis deriving theorems on the switching network size of three-stage folded Clos variants: 3TF, 3F, and 3dF. Such a formal analysis of switching network size has not been thoroughly explored in previous studies, providing new insights into the scalability and structural properties of these network architectures. Numerical results show that the proposed 3dF design model has a larger switching network size than 3TF and 3F under the SNB condition or a blocking probability guarantee.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 describes the assumptions used for designing a switching network and defines the design problem. Section 4 describes Two-stage folded-type Clos network structures. Section 5 presents a design model for 3TF. Section 6 describes the proposal design model for 3dF, including its power management. Section 7 analyzes three-stage folded Clos variants regarding the switching network size. Section 8 describes numerical results. Section 9 concludes this paper.

2. RELATED WORK

The design of Clos networks has been extensively studied across various domains, including packet switching architectures [10, 11, 22], neural networks [23], computing clusters [24], and data center infrastructures [25–27], all of which build upon the foundational principles established by Clos [7, 9]. A notable variant, folded Clos network architectures [23–27], which integrate the input and output layers, has gained prominence due to its symmetric structure and operational efficiency.

In data centers, packet-switching folded Clos network architectures are widely adopted [28–31]. The study in [28] explored load-balancing strategies for folded Clos systems, while [29] introduced a traffic rebalancing algorithm. Research in [30] focused on defining reliability metrics within folded Clos, and [31] investigated network designs with randomly connected switch levels. In contrast, our work focuses on OCS within folded Clos network architectures for data centers, which differs from packet switching.

In circuit switching, once a connection is established between a terminal pair, the links involved are exclusively occupied, indicating that they cannot be shared with other terminal pairs, unlike in packet switching. A blocking probability, which refers to the probability that a connection request is blocked due to unavailable internal resources, is a critical issue in circuit-switched networks. Lee *et al.* [32] examined blocking probability in Clos networks. Yang *et al.* [33] specifically analyzed the three-stage unfolded Clos network, assuming independent link states between adjacent switch layers. Abramov *et al.* [34] applied an Erlang fixed point approximation to estimate blocking in circuit-switched switching networks, while Głabowski and Sobieraj [35] introduced a reservation algorithm to study blocking. Araujo *et al.* [36, 37] utilized artificial neural networks to estimate blocking probability. Lin *et al.* [38] developed non-blocking criteria and probabilistic models for wavelength-selective Clos networks. However, despite these efforts, they did not establish a theoretical upper bound on blocking probability. This paper

deals with a design model that guarantees admissible blocking probability, providing a theoretical upper bound.

Allowing a certain admissible blocking probability in Clos networks increases the switching network size compared to that with the SNB condition. Along this direction, Studies [15–17, 20] address developing design models to enlarge the switching network while guaranteeing an admissible blocking probability. Taka *et al.* [15–17] developed a theoretical upper bound for blocking probability, enabling the design of 2TF to guarantee the admissible blocking probability, whereas Mano *et al.* [13, 14] originally developed a design model for 2TF with the SNB condition. By extending these studies for 2TF, Taniguchi *et al.* [20] addressed designing 3TF, guaranteeing an admissible blocking probability, to have a larger switching network size than 2TF. The developed design model of 3TF, however, encounters a limitation where the network size ceases to grow as the number of available $N \times N$ switches increases beyond a certain threshold. This limitation prevents further expansion of the switching network. Our work focuses on relaxing the restriction on switching network size by introducing a design model of 3dF, allowing for an increase in the number of available $N \times N$ switches.

Recent studies [1, 2, 6, 39–43] have examined the role of optical switching technology in data center networks, particularly focusing on OCSes. OCSes are recognized for their potential to reduce latency compared to conventional electrical packet switches [1].

While these studies primarily focus on integrating OCSes into data center architectures and optimizing network operations, they do not specifically address the challenge of expanding the switching network size within an OCS-based framework. For instance, Wang *et al.* [39] developed TopoOpt, a distributed direct-connect framework for deep neural network training that leverages OCSes to optimize computation, communication, and network topology. However, their study does not tackle fundamental design constraints related to switch scalability. Similarly, Liang *et al.* [41] introduced NegotiaToR, a reconfigurable optical data center network architecture that dynamically adapts to traffic patterns, yet their work does not provide a systematic design model for network size expansion.

Our work distinguishes itself by addressing this specific limitation: we introduce a design model that enables an increase in the number of available $N \times N$ switches, overcoming the scalability ceiling observed in the prior model of 3TF [20]; the 3TF model extends the switching network size beyond that of the 2TF model [15–17] while guaranteeing an admissible blocking probability, though it still faces inherent scalability limitation. By introducing 3dF, our approach provides a theoretical foundation for expanding OCS network size while maintaining an admissible blocking probability. This advancement has direct implications for OCS-enabled data center applications, as it allows for greater scalability in interconnect designs, complementing existing optical switching frameworks such as those explored in [42, 43].

3. ASSUMPTIONS AND PROBLEM DEFINITION

We describe the assumptions used in this paper as follows. We design a Clos network for a data center using identical switches whose size is $N \times N$. The number of switches is at most a . Terminals, such as top-of-rack (ToR) or aggregation switches, in a data center network can establish direct connections with optical circuit switching.

A request is routed through switches in each layer without

Table 1. List of given parameters.

Parameters	Descriptions
N	Number of ports in a switch.
a	Number of available switches.
p	Probability that a request is active.
ϵ	Admissible blocking probability in the network.

disrupting the current optical connections. In this work, we do not employ the rearrangement of existing optical links, as our focus is on preserving communication stability. A transmitter linked to a terminal generates a request, which is active with a probability of p and inactive with a probability of $1 - p$. The generation of these requests follows an independent and identically distributed (i.i.d.) model. The destination of each request is not predetermined and can be arbitrary.

The given parameters are listed in Table 1. The parameter ϵ represents the admissible blocking probability. A network designer sets ϵ considering the quality of service. $f(N, a, p, \epsilon)$ denotes the maximum switching network size, i.e., the maximum number of terminals connected to the switching network, given N, a, p , and ϵ .

Given the above assumption, we define a design problem (DP) to maximize the switching network size.

Problem DP. Given a type of Clos network, the number of identical $N \times N$ switches, a , the request probability generated by each terminal, p , and the admissible blocking probability, ϵ , find the network structure to maximize the switching network size.

4. TWO-STAGE FOLDED-TYPE CLOS NETWORK STRUCTURES

We describe the essences of 2F and 2TF in Sections 4.A and 4.B, respectively. A folded-type Clos structure reduces the number of stages, streamlining the design and reducing the complexity of the switching network.

A. Two-stage folded Clos network (2F)

2F comprises $k + m$ switches, each equipped with N input ports and N output ports, as shown in Fig. 1 [13, 14]. The input-output layer has k switches, labeled as S_i^1 , where i ranges from 1 to k . Each switch in this layer connects to n transmitters, denoted by t_j^i with $i \in [1, k]$ and $j \in [1, n]$, as well as n receivers, denoted by r_j^i with $i \in [1, k]$ and $j \in [1, n]$. In the intermediate layer, there are m switches labeled as S_i^2 , with $i \in [1, m]$. One side of each switch in the input-output layer is connected to one side of each switch in the intermediate layer via v links.

Designing 2F specifies the values of n, k, m , and v to maximize the switching network size of nk , given a total of a available $N \times N$ switches, where $k + m \leq a$ [13, 14].

B. Two-stage twisted-folded Clos network (2TF)

Studies in [13, 14] introduced 2TF to expand the switching network size compared to 2F, while maintaining the key benefit of 2F—the folded structure reduces the number of stages, simplifying the architecture and decreasing the switching network's

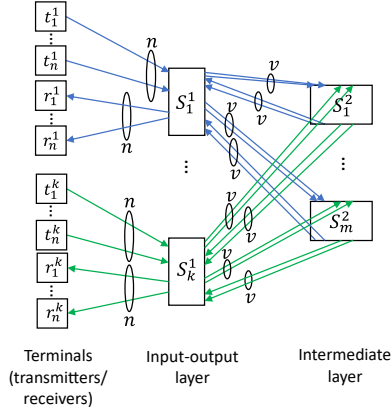


Fig. 1. Structure of 2F.

complexity. The 2F design faces constraints in scaling the number of ports in the input-output layer, which limits the overall number of switching ports. The 2TF approach addresses this limitation by introducing the concept of *twisting*.

In the same way as 2F, 2TF comprises $k + m$ switches, each equipped with N input ports and N output ports, as shown in Fig. 2 [13, 14]. The input-output layer has k switches, labeled as S_i^1 , where i ranges from 1 to k . Each switch in this layer connects to n transmitters, denoted by t_j^i with $i \in [1, k]$ and $j \in [1, n]$, at one side as well as n receivers, denoted by r_j^i with $i \in [1, k]$ and $j \in [1, n]$, at the other side. In the intermediate layer, there are m switches labeled as S_i^2 , with $i \in [1, m]$. One side of each switch in the input-output layer is connected to one side of each switch in the intermediate layer via v links for incoming connections. The other side of each switch in the input-output layer is connected to the other side of each switch in the intermediate layer via v links for outgoing connections.

Designing 2TF specifies the values of n, k, m , and v to maximize the switching network size of nk , given a total of a available $N \times N$ switches, where $k + m \leq a$ [13–17].

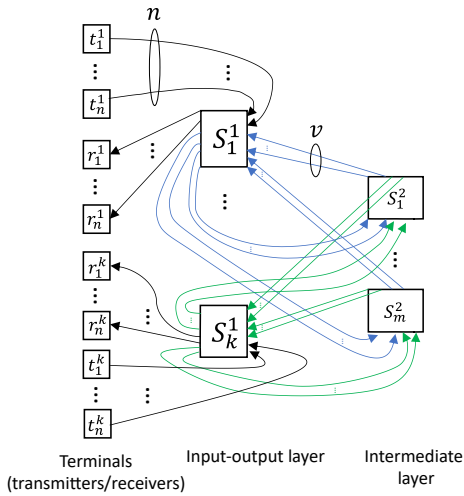


Fig. 2. Structure of 2TF.

5. DESIGN FOR THREE-STAGE TWISTED-FOLDED CLOS NETWORK (3TF)

This section describes an existing design model for 3TF [20]. 3F is a variant of a three-stage folded-type Clos structure. This section focuses on explaining the 3TF design model, whereas we describe the 3F design model in Appendix A.

A. Structure of 3TF

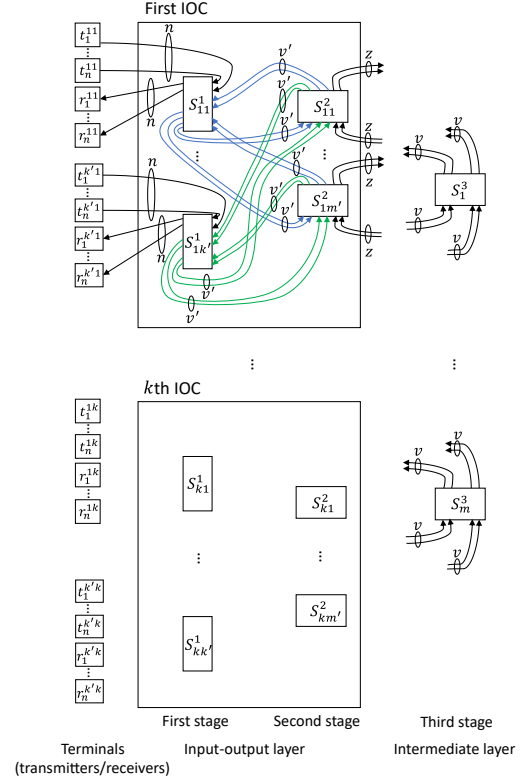


Fig. 3. Structure of 3TF.

Fig. 3 shows the 3TF structure consisting of input-output and intermediate layers [20]. 3TF is composed of $k(k' + m') + m$ switches, each equipped with N input and N output ports. The input-output layer encompasses the first and second stages, while the intermediate layer constitutes the third stage. The input-output layer includes k switching-network components, each with $k' + m'$ switches, and the intermediate layer contains m switches. We refer to a switching network component in the input-output layer as an IOC. The IOCs and third-stage switches are interconnected following the twisted-folded Clos network configuration, as are the switches in the first and second stages. $nk'k$ transmitters and $nk'k$ receivers are connected across all first-stage switches. The total switching network size for 3TF is given by $nk'k$. $S_{i_1 i_2}^1$ refers to the i_2 th first-stage switch in the i_1 th IOC, where $i_1 \in [1, k]$ and $i_2 \in [1, k']$. Similarly, $S_{i_1 i_2}^2$ refers to the i_2 th second-stage switch in the i_1 th IOC, where $i_1 \in [1, k]$ and $i_2 \in [1, m']$. The i th third-stage switch is denoted as S_i^3 , where $i \in [1, m]$. Transmitters and receivers are denoted by $t_j^{i_1 i_2}$ and $r_j^{i_1 i_2}$, respectively, with $j \in [1, n]$, $i_1 \in [1, k']$, and $i_2 \in [1, k]$. The switch $S_{i_1 i_2}^1$ connects n transmitters and n receivers. Each $S_{i_1 i_2}^1$ is linked to $S_{i_1' i_2'}^2$ via v' outgoing and v' incoming links, where

Table 2. List of decision variables for designing 3TF, 3F, and 3dF.

Variables	Descriptions
n	Natural number variable that is the number of transmitters or receivers connected to a first-stage switch.
n_1^{snb}	Natural number variable that is the number of ports satisfying the SNB condition in an IOC.
n_2^{snb}	Natural number variable that is the number of ports satisfying the SNB condition in the network.
k'	Natural number variable that is the number of first-stage switches in an IOC.
k	Natural number variable that is the number of IOCs in the network.
m'	Natural number variable that is the number of second-stage switches in an IOC.
m	Natural number variable that is the number of third-stage switches in the network.
v'	Natural number variable that is the number of links from (to) a first-stage switch outgoing to (incoming from) a second-stage switch, respectively.
v	Natural number variable that is the number of links from (to) an IOC outgoing to (incoming from) a third-stage switch, respectively.
z	Natural number variable that is the maximum number of output (input) ports at a second-stage switch outgoing to (incoming from) third-stage switches, respectively.
η	Real number variable that is the upper bound of blocking probability in each IOC.

$i_1 \in [1, k]$, $i_2 \in [1, k']$, and $i'_2 \in [1, m']$. Additionally, $S_{i_1 i_2}^2$ is connected to S_i^3 through v outgoing and v incoming links, where $i_1 \in [1, k]$, $i_2 \in [1, m']$, and $i \in [1, m]$. The switch $S_{i_1 i_2}^2$ supports up to z output ports to the third-stage switches and up to z input ports from the third-stage switches. The total number of output ports from each IOC to the third-stage switches is capped at zm' , and the same applies to the input ports from the third-stage switches to each IOC.

Table 2 lists the decision variables for designing three-stage folded Clos variants: 3TF, 3F, and 3dF. We explain the reasoning behind ordering the decision variables as listed in Table 2. In the 2TF design model, n and k determine the switching network size nk , making them the primary decision variables. The variables m and v define the network structure and are included as additional decision variables. Based on this, the fundamental ordering is n , k , and v . For three-stage folded Clos variants, the variables n_1^{snb} and n_2^{snb} originate from n , while k' corresponds to k , and m' corresponds to m , with v' related to v . We group decision variables based on their association with n , k , and m . Within the n -related group, we order n , n_1^{snb} , and n_2^{snb} sequentially from the terminal stage to the intermediate layer via the

input-output layer. Similarly, within the k -related and m -related groups, we order k' and k , as well as m' and m , and v' and v . The decision variables z and η are placed at the end, following the order of natural number and real number variables.

B. Formulation of 3TF design model

The DP of 3TF with blocking probability guarantee is given by:

$$\max \quad nk'k \quad (1a)$$

$$\text{s.t.} \quad n + v'm' \leq N \quad (1b)$$

$$v'k' + z \leq N \quad (1c)$$

$$vk \leq N \quad (1d)$$

$$vm \leq zm' \quad (1e)$$

$$\sum_{w=n_1^{\text{snb}}+1}^n \binom{n}{w} p^w (1-p)^{n-w} \leq \eta \quad (1f)$$

$$2 \left\lfloor \frac{n_1^{\text{snb}}-1}{v'} \right\rfloor + 1 \leq m' \quad (1g)$$

$$1 - (1 - \eta) \times \left(1 - \sum_{w=n_2^{\text{snb}}+1}^{nk'} \binom{nk'}{w} p^w (1-p)^{nk'-w} \right) \leq \epsilon \quad (1h)$$

$$2 \left\lfloor \frac{n_2^{\text{snb}}-1}{v} \right\rfloor + 1 \leq m \quad (1i)$$

$$k(k' + m') + m \leq a \quad (1j)$$

$$\eta \leq \epsilon \leq 1 \quad (1k)$$

$$n, n_1^{\text{snb}}, n_2^{\text{snb}}, k', k, v', v, m', m, z \in \mathbb{N} \quad (1l)$$

$$\eta \in \mathbb{R}^+, \quad (1m)$$

where $\mathbb{N}$ denotes a set of natural numbers, and $\mathbb{R}^+$ denotes a set of non-negative real numbers. Eq. (1a) serves as the objective function to maximize the switching network size of 3TF, which corresponds to maximizing the number of terminals connected to the input and output ports of the switching network. Eq. (1b)–Eq. (1d) define the constraints on the number of available ports for switches in the first, second, and third stages, respectively. Eq. (1e) ensures that the total number of output (or input) ports from (or to) an IOC to (or from) third-stage switches is at least vm . Eq. (1f) demonstrates that the blocking probability in the first stage is bounded by η . Eq. (1g) specifies the SNB condition in this stage. Similarly, Eq. (1h) shows that the blocking probability in the third stage is restricted by ϵ , with Eq. (1i) detailing the corresponding the SNB condition. Eq. (1j) ensures that the number of switches in the network does not exceed a . Eq. (1k) defines the relationship between η and ϵ , ensuring that η is at most ϵ . Finally, Eq. (1l) and Eq. (1m) define the types of decision variables.

6. PROPOSED DESIGN MODEL FOR THREE-STAGE FOLDED CLOS NETWORK WITH DECOUPLED FIRST STAGE (3dF)

A. Structure of 3dF

Fig. 4 shows the 3dF structure consisting of input-output and intermediate layers. 3dF differs from 3TF and 3F (see Appendix A) in the first stage, whereas other structures in 3TF, 3F, and 3dF are similar. In contrast, 3dF devises the input and output layer, decoupling the $n \times v'm'$ and $v'm' \times n$ switching functions—both originally implemented in a single switch at

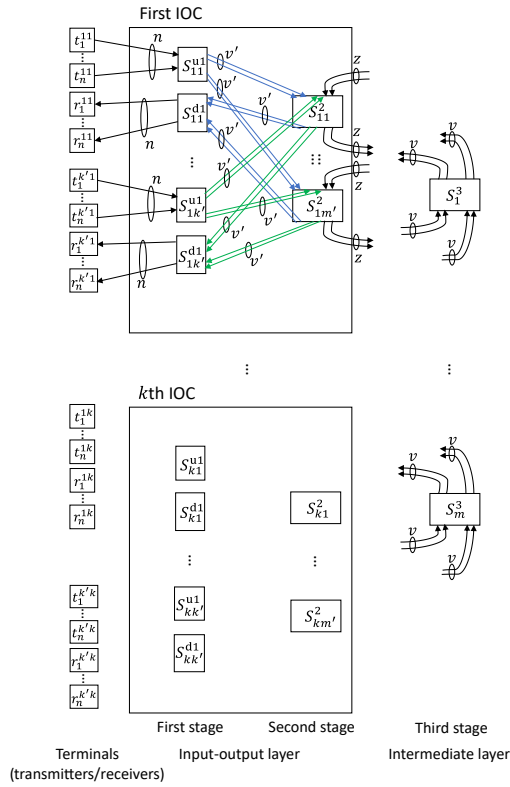


Fig. 4. Structure of 3dF.

the first stage in 3TF and 3F—into two distinct switches. One switch is connected to n transmitter on one side and is connected to m' switches in the second stage with v links for each. The other switch is connected to n receivers on one side and is connected to m' switches in the second stage with v links for each. 3dF comprises $k(2k' + m') + m$ switches, each of which has N input ports and N output ports. The input-output layer corresponds to the first and second stages, and the intermediate layer corresponds to the third stage. The input-output layer comprises k switching-network components, each of which, referred to as an IOC, has $2k' + m'$ switches, and the intermediate layer comprises m switches.

By decoupling the $n \times v'm'$ and $v'm' \times n$ switching functions into two distinct sets of switches, the 3dF architecture enhances the switching network size by allowing independent handling of input and output OCSes. In contrast, in 3TF, the first-stage switch simultaneously handles both directions of OCSes, which can create a bottleneck in switching ports as the number of terminals increases. By separating these functions in 3dF, each first-stage switch focuses on a specific task—handling OCSes in one direction—thereby eliminating the switching-port bottleneck and scaling the switching network size.

B. Formulation of 3dF design model

The DP of 3dF with blocking probability guarantee is given by:

$$\max \quad nk'k \quad (2a)$$

$$\text{s.t.} \quad n \leq N \quad (2b)$$

$$v'm' \leq N \quad (2c)$$

$$k(2k' + m') + m \leq a \quad (2d)$$

$$\text{Eq. (1c) – Eq. (1i), Eq. (1k), Eq. (1m).} \quad (2e)$$

Eq. (2b) and Eq. (2c) in 3dF replaces Eq. (1b) in 3TF. Eq. (2d) indicates that the number of switches in 3dF does not exceed a , which also differs from 3F. Other constraints are the same as those of 3F.

C. Algorithm designing 3dF

To solve the DP of 3dF in Eq. (2a)–Eq. (2e), we apply an algorithm based on an exhaustive search for the integer decision variable, Algorithm 1, which is based on the algorithm developed for 3TF in [20]. Instead of searching for a feasible solution for every combination of the natural number decision variables, the algorithm exhaustively explores the structure with the largest switching network size in non-increasing order. Once a feasible solution with a larger switching network size is identified, any solution with a smaller one is no longer optimal. Therefore, once a feasible solution is found, it is guaranteed optimal, and the algorithm terminates. We accelerate this process by ordering the decision variables to reduce unnecessary searches, which allows us to efficiently prune the search space by eliminating combinations that do not satisfy the constraints in the design problem. The order in which the decision variables are searched is different from the order in which they are listed in Table 2. By executing the search in Algorithm 1, we derive the optimal structure for 3dF. The overall computational time complexity for solving the DP of 3dF is $O(N^8 a^2)$ if $N \leq a$, and $O(N^5 a^5)$ otherwise.

Algorithm 1 handles a set of Eq. (1c) and Eq. (1e), and a set of Eq. (1f), Eq. (1h), and Eq. (1k), eliminating z and η as follows [20]. First, we integrate Eq. (1c) and Eq. (1e) into one constraint:

$$v'k' + \left\lceil \frac{vm}{m'} \right\rceil \leq N, \quad (3)$$

where $z = \left\lceil \frac{vm}{m'} \right\rceil$ is determined by v , m , and m' . Second, since the left-hand side of Eq. (1h) is a non-decreasing function regarding η , we can get an optimal solution by substituting the smallest η satisfying Eq. (1f) into Eq. (1h) and Eq. (1k). Thus, instead of Eq. (1f), we can use the following relationship:

$$\sum_{w=n_1^{nb}+1}^n \binom{n}{w} p^w (1-p)^{n-w} = \eta. \quad (4)$$

Given the nature of the problem, finding the exact optimal solution requires an exhaustive search, as no alternative approach has been found to derive the optimal structure while ensuring optimality. Moreover, we have not identified an effective method to significantly accelerate this process without compromising the solution's optimality. While heuristic or approximation techniques might be explored in future research to provide near-optimal solutions with reduced computational complexity, such approaches would inherently trade off optimality for computational efficiency, which may not be suitable for scenarios requiring strict guarantees.

D. Power Management in 3dF

This section describes a power-management example in 3dF. An OCS connection is established between a transmitter and a receiver on top-of-rack (ToRs) and aggregation switches in a data center network, passing through five optical switches, four optical patch panels between adjacent stages, and optical fibers.

Algorithm 1. Algorithm to solve DP of 3dF**Input:** N, a, p, ϵ **Output:** $n, n_1^{\text{snb}}, n_2^{\text{snb}}, k', k, v', v, m', m$

```

1:  $\text{nk}'k\_array \leftarrow [[1, 1, 1], [1, 1, 2], \dots, [N-1, N-1, N]]$ 
2: Sort in descending order concerning the value of the multiplication of three values in the array with  $\text{nk}'k\_array$ .
3: for  $\text{nk}'k\_array$  do
4:   for  $v$  do
5:     if Eq. (1d) is satisfied then
6:       for  $n_1^{\text{snb}}$  do
7:         if Eq. (4) and Eq. (1k) are satisfied then
8:           for  $n_2^{\text{snb}}$  do
9:             if Eq. (1h) is satisfied then
10:              for  $m$  do
11:                if Eq. (1i) is satisfied then
12:                  for  $m'$  do
13:                    if Eq. (2d) is satisfied then
14:                      for  $v'$  do
15:                        if Eq. (2b), Eq. (2c), Eq. (1g), and Eq. (3) are satisfied then
16:                          return  $n, n_1^{\text{snb}}, n_2^{\text{snb}}, k', k, v', v, m', m$ 

```

Our example assumes that we use a transceiver with a power budget of 8 dB [44, 45], an optical switch such as a software-defined optical automatic patch panel with an insertion loss of at most 1 dB [46], an optical patch panel with an insertion loss of at most 0.1 dB [47], a fiber with a propagation-loss coefficient of 0.35 dB/km [48], where the total length of optical fibers in a data center is at most 3 km. The total loss of an OCS connection is 6.45 dB at maximum, which is lower than the 7.0 dB transceiver budget.

We note that the power-management discussion in this section for 3dF also applies to 3TF and 3F, as an OCS connection between a transmitter and a receiver passes through five optical switches, four optical patch panels between adjacent stages, and optical fibers in the same way as 3dF. Due to structural differences, the number of accommodated terminals differs among 3TF, 3F, and 3dF, leading to variations in the total power consumption of the entire switching network. However, the discussion in this section focuses on the optical power level per connection, which remains unaffected by the difference in the number of terminals. Since each OCS connection involves the same number of optical components in 3TF, 3F, and 3dF, their optical power characteristics remain identical.

7. ANALYSIS OF THREE-STAGE FOLDED CLOS VARIANTS

We provide analysis deriving theorems to compare our considered three-stage folded Clos variants: 3dF, 3TF, and 3F. $f_{3\text{TF}}(N, a, p, \epsilon)$, $f_{3\text{F}}(N, a, p, \epsilon)$, and $f_{3\text{dF}}(N, a, p, \epsilon)$ denotes the switching network size maximized by solving each corresponding DP, respectively.

Theorem 1. Given N, a, p , and ϵ for 3TF and 3F, The switching network size of 3TF is equal to or larger than that of 3F; we have:

$$f_{3\text{TF}}(N, a, p, \epsilon) \geq f_{3\text{F}}(N, a, p, \epsilon). \quad (5)$$

Proof. The DP of 3TF in Eq. (1a)–Eq. (1m) differs from that of 3F in Eq. (22a)–Eq. (22d) in that: constraint Eq. (1b) in 3TF replaces constraints Eq. (22b) and Eq. (22c) in 3F; other constraints are the same.

Using Eq. (22b) and Eq. (22c), we have:

$$n + v'm' \leq \left\lfloor \frac{N}{2} \right\rfloor + \left\lfloor \frac{N}{2} \right\rfloor \leq N, \quad (6)$$

which is equivalent to Eq. (1b). Thus, satisfying constraints Eq. (22b) and Eq. (22c) in 3F ensures satisfying constraint Eq. (1b) in 3TF. This indicates that all the feasible solutions in 3F are included in those of 3TF, i.e., 3TF has a more flexible design than 3F. $\square$

Theorem 2. $a_{3\text{dF}}$ and $a_{3\text{TF}}$ denotes the number of available switches for 3dF and 3TF, respectively. We assume that $a_{3\text{dF}} - a_{3\text{TF}} \geq N(N-1)$. Given $N, p, a_{3\text{dF}}$ and $a_{3\text{TF}}$, and ϵ for 3dF and 3TF, respectively, the switching network size of 3dF is equal to or larger than that of 3TF; we have:

$$f_{3\text{dF}}(N, a_{3\text{dF}}, p, \epsilon) \geq f_{3\text{TF}}(N, a_{3\text{TF}}, p, \epsilon). \quad (7)$$

Proof. The DP of 3dF in Eq. (2a)–Eq. (2e) differs from that of 3TF in Eq. (1a)–Eq. (1m) in that: constraints Eq. (2b), Eq. (2c), and Eq. (2d) in 3dF replaces constraints Eq. (1b) and Eq. (1j) in 3TF; other constraints are the same.

Using Eq. (1b), we have:

$$n < n + v'm' \leq N \quad (8a)$$

$$v'm' < n + v'm' \leq N, \quad (8b)$$

leading to satisfying Eq. (2b) and Eq. (2c), respectively. Thus, satisfying constraint Eq. (1b) in 3TF ensures satisfying constraint Eq. (2b) and Eq. (2c) in 3dF. In addition, since $k \leq N$ and $k' \leq N-1$, $a_{3\text{dF}} - a_{3\text{TF}} \geq N(N-1)$ ensures that satisfying Eq. (1j) in 3TF ensures Eq. (2d) in 3dF. This indicates that, under the condition of $a_{3\text{dF}} - a_{3\text{TF}} \geq N(N-1)$, all the feasible solutions in 3TF are included in those of 3dF, i.e., 3dF has a more flexible design than 3TF. $\square$

Theorem 3. Given N, a, p , and ϵ for 3TF, 3F, and 3dF, where $a \geq N(4N-3)$, the switching network size of 3dF is equal to or larger than those of 3TF and 3F; we have:

$$f_{3\text{dF}}(N, a, p, \epsilon) \geq f_{3\text{TF}}(N, a, p, \epsilon) \geq f_{3\text{F}}(N, a, p, \epsilon). \quad (9)$$

Proof. $f_{3TF}(N, a, p, \epsilon) \geq f_{3F}(N, a, p, \epsilon)$ holds by Theorem 1.

Next, we prove $f_{3dF}(N, a, p, \epsilon) \geq f_{3TF}(N, a, p, \epsilon)$ with $a \geq N(4N - 3)$. From constraints Eq. (1b)–Eq. (1m) and Eq. (2c)–Eq. (2e) of DP for 3TF and 3dF, respectively, we have $k \leq N$, $k' \leq N - 1$, $m \leq N(N - 1)$, and $m' \leq N$. Thus, the left-hand side of Eq. (1j) is:

$$\begin{aligned} k(k' + m') + m &\leq N \{(N - 1) + N\} + N(N - 1) \\ &= N(3N - 2). \end{aligned} \quad (10)$$

The left-hand side of Eq. (2d) is:

$$\begin{aligned} k(2k' + m') + m &\leq N \{2(N - 1) + N\} + N(N - 1) \\ &= N(4N - 3). \end{aligned} \quad (11)$$

If $a \geq N(3N - 2)$ for 3TF and $a \geq N(4N - 3)$ for 3dF, each value of a is sufficiently large in Eq. (1j) and Eq. (2d), respectively. Since a is only used in one constraint, if a is equal to or larger than the above bound, $N(3N - 2)$ for 3TF or $N(4N - 3)$ for 3dF, the optimal objective value of DP reaches the upper bound; it cannot be maximized any more. Thus, we find the upper bound of f_{3dF} for $a \geq N(4N - 3)$ and that of f_{3TF} for $a \geq N(3N - 2)$. Given $a \geq N(4N - 3)$, we have:

$$f_{3TF}(a) = f_{3TF}(a - N(N - 1)). \quad (12)$$

With Theorem 2, we prove Theorem 3. $\square$

Theorem 4. Given N , a_1 , p , and ϵ for 3TF and 3F, when $a_1 \geq N(3N - 2)$, each solution (k, k', m, m') of DP for 3TF and 3F denotes $(k_{3TF}^*, k_{3TF}'^*, m_{3TF}^*, m_{3TF}'^*)$ and $(k_{3F}^*, k_{3F}'^*, m_{3F}^*, m_{3F}'^*)$, respectively. a_1 and a_2 denote the number of available switches. When

$$k_{3TF}^*(k_{3TF}'^* + m_{3TF}'^*) + m_{3TF}^* \leq a_2 \leq N(3N - 2), \quad (13)$$

we have:

$$f_{3TF}(N, a_2, p, \epsilon) = f_{3TF}(N, a_1, p, \epsilon). \quad (14)$$

When

$$k_{3F}^*(k_{3F}'^* + m_{3F}'^*) + m_{3F}^* \leq a_2 \leq N(3N - 2), \quad (15)$$

we have:

$$f_{3F}(N, a_2, p, \epsilon) = f_{3F}(N, a_1, p, \epsilon). \quad (16)$$

Proof. As $(k_{3TF}^*, k_{3TF}'^*, m_{3TF}^*, m_{3TF}'^*)$ and $(k_{3F}^*, k_{3F}'^*, m_{3F}^*, m_{3F}'^*)$ are the solutions of DP for 3TF and 3F, respectively, we have:

$$\begin{aligned} &f_{3TF}(N, k_{3TF}^*(k_{3TF}'^* + m_{3TF}'^*) + m_{3TF}^*, p, \epsilon) \\ &= f_{3TF}(N, a_1, p, \epsilon) \end{aligned} \quad (17)$$

$$\begin{aligned} &f_{3F}(N, k_{3F}^*(k_{3F}'^* + m_{3F}'^*) + m_{3F}^*, p, \epsilon) \\ &= f_{3F}(N, a_1, p, \epsilon). \end{aligned} \quad (18)$$

Since $f_{3TF}(N, a_2, p, \epsilon)$ and $f_{3F}(N, a_2, p, \epsilon)$ are non-decreasing functions with respect to a_2 in addition to Eq. (17) and Eq. (18), Theorem 4 is proved. $\square$

Theorem 5. Given N , a_1 , p , and ϵ for 3dF, when $a_1 \geq N(4N - 3)$, the solution (k, k', m, m') of DP denotes (k^*, k'^*, m^*, m'^*) . a_1 and a_2 denote the number of available switches. When

$$k^*(2k'^* + m'^*) + m^* \leq a_2 \leq N(4N - 3), \quad (19)$$

we have:

$$f_{3dF}(N, a_2, p, \epsilon) = f_{3dF}(N, a_1, p, \epsilon), \quad (20)$$

where $f_{3dF}(N, a_1, p, \epsilon)$ is defined in Theorem 3.

Proof. As (k^*, k'^*, m^*, m'^*) is the solution of DP for 3dF, we have:

$$\begin{aligned} &f_{3dF}(N, k^*(2k'^* + m'^*) + m^*, p, \epsilon) \\ &= f_{3dF}(N, a_1, p, \epsilon). \end{aligned} \quad (21)$$

Since $f_{3dF}(N, a_2, p, \epsilon)$ is a non-decreasing function with respect to a_2 in addition to Eq. (21), Theorem 5 is proved. $\square$

Theorems 1–5 offer the following insights.

- When comparing 3TF and 3F, Theorem 1 indicates that choosing 3TF is advantageous regarding the switching network size.
- Studies [20, 21] observed that 3TF encounters a limitation where its size can no longer expand due to the port limitation of the first-stage switches when the number of available $N \times N$ switches exceeds a certain threshold. In such a case, Theorem 2 suggests that 3dF can expand the switching network size by increasing the number of available switches as it relaxes the port limitation of the first-stage switches.
- Theorem 3 suggests that, when we have a sufficiently large number of available switches, 3dF can have a larger switching network size than 3TF and 3F.
- Theorems 4 and 5 indicate that, for 3TF, 3F, and 3dF, we find the ranges of a_2 , which are given by Eq. (13), Eq. (15), and Eq. (19), respectively. a_2 indicates the number of available switches. These ranges determine the upper bound of the switching network size, f_{3TF} , f_{3F} , and f_{3dF} , respectively. From Eq. (13), Eq. (15), and Eq. (19), the corresponding lower limits of a_2 for 3TF, 3F, and 3dF are given as $k_{3TF}^*(k_{3TF}'^* + m_{3TF}'^*) + m_{3TF}^*$, $k_{3F}^*(k_{3F}'^* + m_{3F}'^*) + m_{3F}^*$, and $k^*(2k'^* + m'^*) + m^*$, respectively. Then, to achieve the upper bound of the switching network size, it is sufficient for the number of available switches to be at least the corresponding lower limit of a_2 for 3TF, 3F, and 3dF, respectively.

8. NUMERICAL RESULTS

This section provides a detailed numerical comparison among three-stage folded Clos variants: 3dF, 3TF [20], and 3F. For reference, we also present the performance results of 2TF and 2F [14]. Our approach involves implementing algorithms based on an exhaustive search to determine the network structure for each model, with the 3dF design model specifically executing Algorithm 1. These algorithms are implemented in C++ and run on a hardware platform with an AMD EPYC Rome 7502P 32-core CPU with 128 GB of memory for computation.

We examine a scenario in which a service provider aims to design a data center network that maximizes the switching network size while ensuring either an admissible blocking probability ($\epsilon > 0$) or meeting the SNB condition ($\epsilon = 0$), given a specified number of $N \times N$ switches. Sections 8.A and 8.B present performances of network switching size and computation time, respectively.

A. Network switching size

Fig. 5 illustrates how network switching size depends on N under the SNB condition ($\epsilon = 0$). We compare the switching network sizes of 3dF, 3TF, 3F, 2TF, and 2F for different values from 10 to 50, with $a = 10000$. a is set to a sufficiently large value across all our examined N in this comparison such that the restriction of a can be eliminated, whereas we focus on the

effect of different N . Since we consider the SNB condition, p does not affect the results. To clarify the tendency of switching network size for 2TF and 2F, we also show the results using the logarithmic scale in Fig. 6. Increasing the number of stages from two to three results in a larger switching network size. 3dF has a larger network switching size than 3TF and 3F.

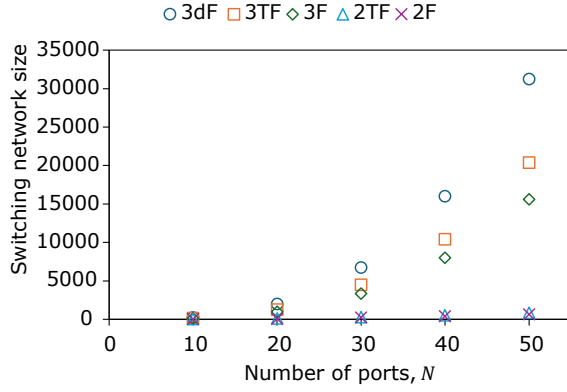


Fig. 5. Dependency of switching network size on N with $a = 10000$ and $\epsilon = 0$.

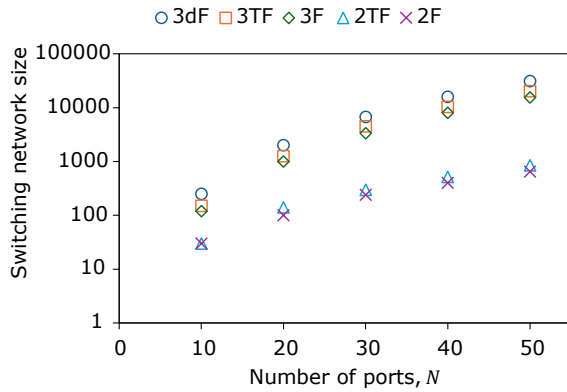


Fig. 6. Dependency of switching network size on N with $a = 10000$ and $\epsilon = 0$.

We observe the dependency of switching network size on a with $N = 40$ and the SNB condition ($\epsilon = 0$), as shown in Fig. 7. Given the SNB condition, p does not affect the results. We focus on the behavior of the switching network size for three-stage variants: 3dF, 3TF, and 3F. In contrast, the switching network sizes of the two-stage variants, 2TF and 2F, are limited as a increases compared to the three-stage variants. It is observed that the switching network size of 3TF is larger than that of 3F for any value of a , consistent with Theorem 1. As a increases, the switching network sizes of 3dF, 3TF, and 3F increase with each value of a at maximum. However, we observe that the switching network size of 3F saturates at $a = 1999$, and that of 3TF saturates at $a = 2359$, whereas the switching network size of 3dF increases with $a = 3999$ at maximum. We observe that, for 3TF, 3F, and 3dF, each saturated value of a matches the lower bound of the a_2 region given by Theorem 4 for 3TF and 3F and Theorem 5 for 3dF. Theorems 4 and 5 provide the saturated region of a that achieves the corresponding upper bound of the switching network size, f_{3TF} for 3TF and f_{3dF} for 3dF, respectively, as observed in Fig. 7. When we compare the

upper bounds of switching network sizes, f_{3dF} , f_{3TF} , and f_{3F} of 3dF, 3TF, and 3F, respectively, with a sufficiently large value of a , we observe that $f_{3dF} \geq f_{3TF} \geq f_{3F}$. This observation follows from Theorem 3.

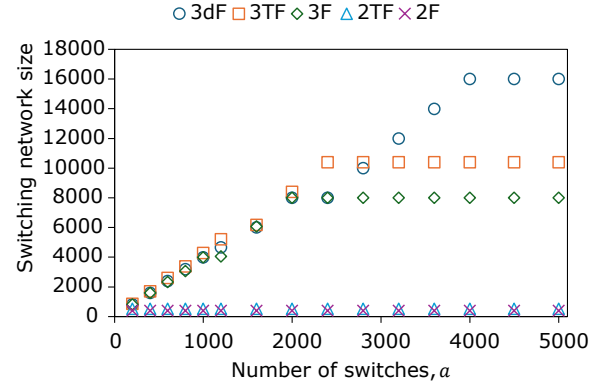


Fig. 7. Dependency of switching network size of 3dF on a with $N = 40$ and $\epsilon = 0$.

Fig. 8 depicts the changes in the switching network size of 3dF when the number of available switches, a , varies with different values of N under the SNB condition ($\epsilon = 0$). As a increases, the switching network size increases for all values of N . The larger N is, the larger the switching network size at which 3dF saturates. With a sufficiently large value of a , a larger N results in a larger switching network size. A larger switch size, N , allows for a 3dF design with more relaxed port constraints, leading to a larger switching network size.

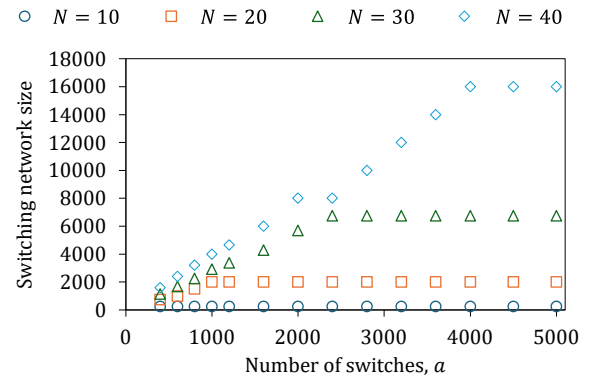


Fig. 8. Dependency of switching network size on a with different values of N under $\epsilon = 0$.

Fig. 9 shows how the switching network size varies with the admissible blocking probability ϵ for $N = 40$, $a = 5000$, and $p = 0.5$. For reference, we also include the switching network size under the SNB condition ($\epsilon = 0$). The 3dF model has a larger network switching size than other design models across all tested values of $\epsilon \geq 0$. For each model, as ϵ increases, the switching network size also increases. Notably, the difference in network switching sizes between $\epsilon = 0$ and $\epsilon = 0.1$ for 3dF is more pronounced than for other design models.

We investigate the switching network size dependent on p with $N = 40$, $a = 4000$, and $\epsilon = 0.001$, as shown in Fig. 10. For reference, we also include the switching network size under the SNB condition ($\epsilon = 0$). As p increases, the switching network

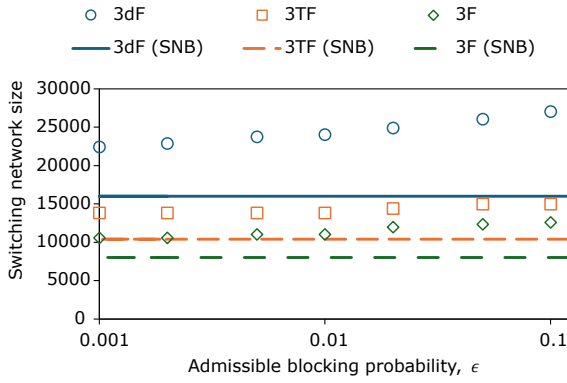


Fig. 9. Dependency of switching network size on ϵ with $N = 40$, $a = 4000$, and $p = 0.5$.

size of each design model becomes large, and the difference from that of the SNB condition ($\epsilon = 0$) becomes large. This is because the constraint of the blocking probability guarantee gets more relaxed with a small value of p . The switching network size of 3dF is larger than those of other design models across our examined values of p .

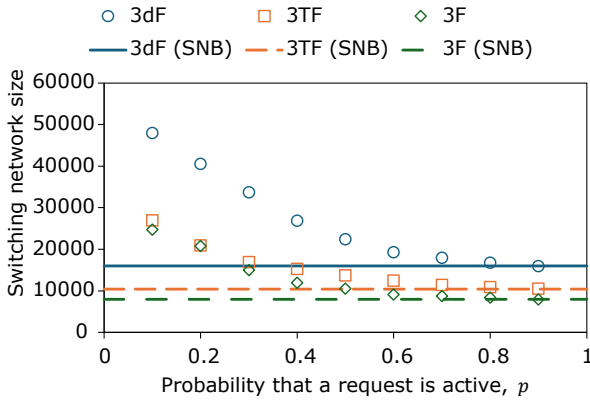


Fig. 10. Dependency of switching network size on p with $N = 40$, $a = 4000$, and $\epsilon = 0.001$.

Fig. 11 depicts the changes in the switching network size of 3dF when the admissible blocking probability of ϵ varies with different values of the request arrival rate of p under $N = 40$ and $a = 4000$. For reference, we also include the switching network size under the SNB condition ($\epsilon = 0$). A smaller p and a larger ϵ further relax the constraint in Eq. (1h), expanding the feasible region of the 3dF structure under the blocking probability guarantee and resulting in a larger switching network size. As ϵ increases from 0.001 to 0.1, the switching network size continues to increase for $p = 0.4, 0.5, 0.6$, and 0.7 , which is consistent with the observation in Fig. 9. When $p = 0.3$, the switching network size increases as ϵ grows from 0.001 to 0.01. However, for $\epsilon > 0.01$, the increase in ϵ no longer leads to a larger switching network size. Under the condition of $p = 0.3$ and $\epsilon \leq 0.1$, further relaxation of the constraint in Eq. (1h) does not contribute to an increase in the switching network size.

B. Computation time

We investigate the computation time dependent on N with $a = 5000$ under the SNB condition ($\epsilon = 0$), as Fig. 12. The compu-

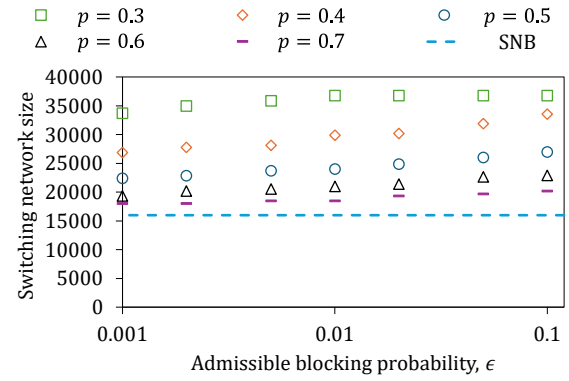


Fig. 11. Dependency of switching size of 3dF on ϵ with different values of p under $N = 40$ and $a = 4000$.

tation times of the three-stage variants, 3dF, 3TF, and 3F, increase more rapidly than those of two-stage variants, 2TF and 2F. The 3dF design model requires approximately five days to obtain the optimal network structure at $N = 50$. We observe that the computation times for 3dF and 3TF are similar, with no significant differences. We note that a service provider aims to develop a data center network that maximizes the switching network size while ensuring the desired blocking quality before commencing service. The design process for the network structure is allowed to extend over a few weeks.

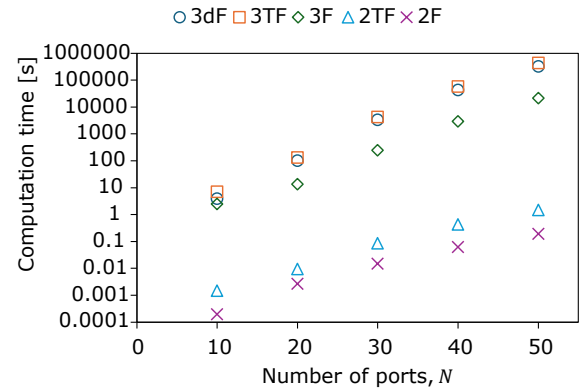


Fig. 12. Dependency of computation time on N with $a = 10000$ and $\epsilon = 0$.

Fig. 13 shows the computation time as a function of a with $N = 40$ under the SNB condition ($\epsilon = 0$). We observe that, across the examined values of a , the computation times for the 3dF and 3TF design models are higher than those for other design models, with similar computation times between 3dF and 3TF. Each three-stage design model has more decision variables than each two-stage design model. Focusing on the computation time behavior of the 3dF model as a increases, we observe that the computation time decreases. A higher a value relaxes the constraint on the total number of switches being less than a , making it easier for the algorithm's conditional branching to satisfy this requirement. As this condition is more easily met, the algorithm reaches a solution faster. Additionally, as the number of *for* loop iterations required to find the solution increases, the reduction in computation time due to the increase in a becomes more pronounced. Consequently, we observe a decrease

in computation time for the 3dF design model as a increases.

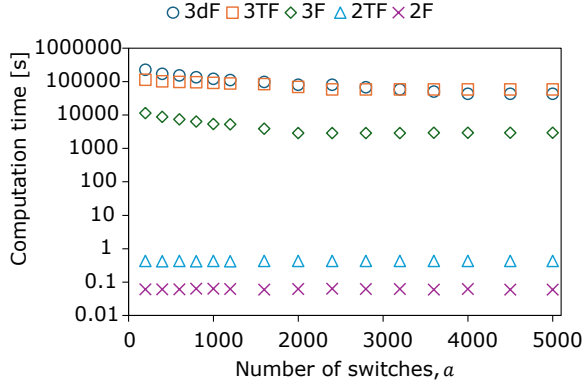


Fig. 13. Dependency of computation time on a with $N = 40$ and $\epsilon = 0$.

Fig. 14 depicts the changes in computation time of 3dF when the number of available switches, a , varies with different values of N under the SNB condition ($\epsilon = 0$). The larger N is, the longer the computation time is; a larger N increases the feasible solution region for designing 3dF. When $N = 20, 30$, and 40 , as a increases, the computation time tends to decrease. This behavior can be explained in the same way as in Fig. 13. A larger a relaxes the constraint in Eq. (2d). Relaxing this constraint with a larger a allows the algorithm to reach a solution faster. In contrast, when $N = 10$, as a increases, the computation time increases. In the case of $N = 10$, the feasible solution region for designing 3dF is already significantly restricted. Increasing a does not sufficiently relax the constraint in Eq. (2d) enough to broaden the feasible region. Instead, the increase in a results in more possible solutions that need to be checked within the already narrow space, which leads to longer computation time.

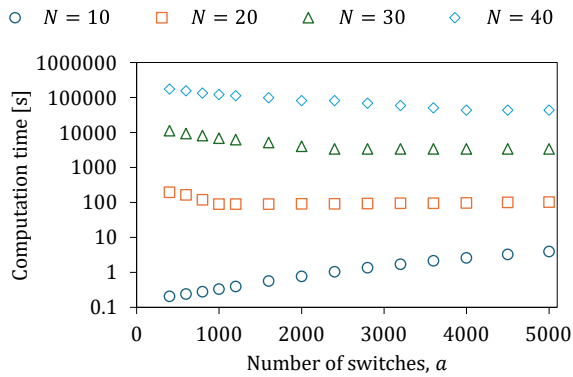


Fig. 14. Dependency of computation time of 3dF on a with different values of N under $\epsilon = 0$.

Fig. 15 presents the computation time dependent on ϵ with $N = 40$, $a = 4000$, and $p = 0.5$. a is set to a sufficiently large value. For reference, we also include the computation time under the SNB condition ($\epsilon = 0$) for each design model. In the three-stage variants, the computation time with $\epsilon > 0$ is larger than that of the SNB condition ($\epsilon = 0$). The design with $\epsilon > 0$ includes the extra process to check whether Eq. (4) and Eq. (1k) are satisfied, whereas that of the SNB condition does not. This causes a clear difference in computation time between $\epsilon > 0$ and

the SNB condition. The computation times for the three-stage variants, 3dF, 3TF, and 3F, depend on ϵ when $\epsilon > 0$, but no clear trend with respect to ϵ is observed for each variant. The computation time varies depending on how quickly the solution with the largest switching network size is found under a predetermined search order in their respective algorithms. Different values of ϵ influence their feasible solution regions, leading to variations in computation time. We observe that the computation time of 3dF is shorter than that of 3TF. In each algorithm to design 3dF and 3TF, instead of searching for a feasible solution for every combination of the natural number decision variables, the algorithm exhaustively explores the structure with the largest switching network size in non-increasing order. Once a feasible solution is found, it is guaranteed optimal, and the algorithm terminates. Since 3dF has a larger switching network size than 3TF, the structure of 3dF is found faster than that of 3TF. In contrast, the computation time of 3F is shorter than that of 3dF. Constraint Eq. (22b) in the 3F design is stricter than constraint Eq. (2b) in the 3dF design. The possible range of n in 3F is approximately half that of 3dF. The size of nk/k_array in 3dF presented in line 1 of Algorithm 1 is reduced to a smaller size in 3F, which reduces the number of searching interactions. This results in faster computation in 3F than in 3dF.

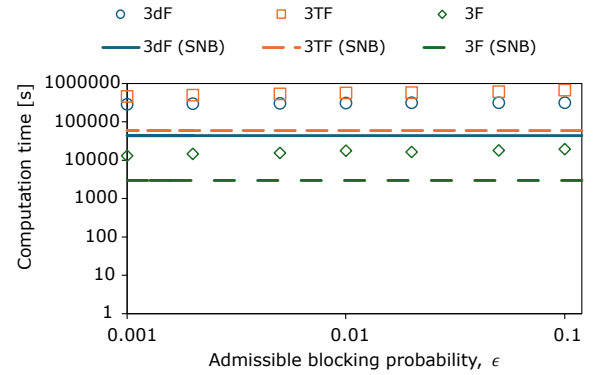


Fig. 15. Dependency of computation time on ϵ with $N = 40$, $a = 4000$, and $p = 0.5$.

We investigate the dependency of the computation time on p with $N = 40$, $a = 4000$, and $\epsilon = 0.001$, as shown in Fig. 16. Similar to the trend observed in Fig. 15, the computation time of 3dF is shorter than that of 3TF across our examined values of p . We observe the behavior of the computation time for 3dF as follows. As p gets close to zero, the computation time decreases; since the switching network size becomes large, Algorithm 1 terminates faster. On the other hand, as p gets close to one, the computation time also decreases and gets close to that of the SNB condition ($\epsilon = 0$). A large p makes the situation close to the SNB condition; the constraints with $\epsilon > 0$ checked in Algorithm 1 become close to those of the SNB condition.

Fig. 17 depicts the changes in computation time of 3dF when the admissible blocking probability of ϵ varies with different values of the request arrival rate of p under $N = 40$ and $a = 4000$. Regarding the impact of the two parameters, ϵ and p , on the computation time of 3dF, the trends observed in Fig. 17 are consistent with those in Figs. 15 and 16. With respect to ϵ , the computation time depends on ϵ for different values of p when $\epsilon > 0$. The trend of the computation-time dependency on ϵ varies across different values of p . Different values of ϵ influence their feasible solution regions, leading to variations

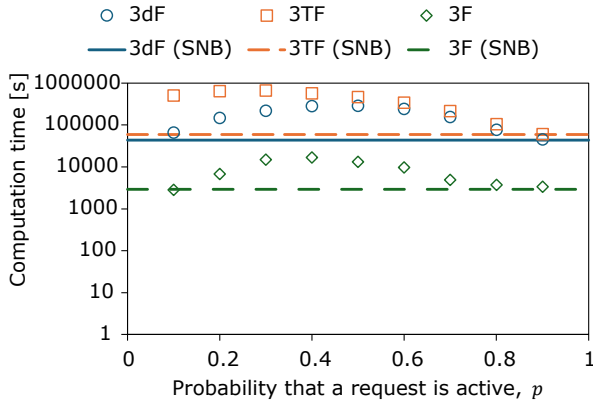


Fig. 16. Dependency of computation time on p with $N = 40$, $a = 4000$, and $\epsilon = 0.001$.

in computation time. With respect to p , the computation time is highest at $p = 0.5$ across most of the examined range of ϵ and decreases as p deviates from 0.5 in either direction. The underlying reasons for these trends are explained in Figs. 15 and 16.

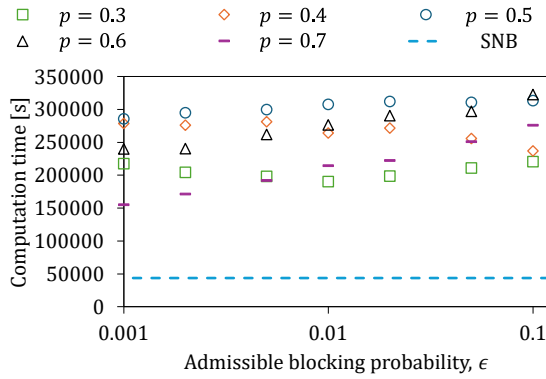


Fig. 17. Dependency of computation time of 3dF on ϵ with different values of p under $N = 40$ and $a = 4000$.

9. CONCLUSIONS

This paper proposed a three-stage enhanced folded Clos network (3dF) design model for OCS, ensuring an admissible blocking probability to maximize the network switching size. The proposed design model uses identical switches with the size $N \times N$. In 3dF, the first- and second-stage switches are combined into a single, large switch in the input-output layer. Differing from the 3TF and 3F structures, 3dF adopts a distinct architectural design that decouples the first stage, allowing for a more balanced distribution of switching functions across the layers. This restructuring improves network scalability and efficiency by eliminating a bottleneck in the number of ports in a first-stage switch. In 3dF, the input-output layer decouples the two switching functions into two distinct switches. We formulated the proposed 3dF design model as an optimization problem. To solve the design problem, we employed an exhaustive search-based algorithm, which identifies the structure with the largest switching network size in non-increasing order while ensuring that specified constraints are satisfied. This study provides a

detailed theoretical analysis of the switching network size for three-stage folded Clos variants. This type of formal analysis has not been thoroughly conducted in previous research, offering our insights into the scalability and structural properties of these network architectures. Numerical results demonstrated that the proposed 3dF design model has a larger switching network size than 3TF and 3F under the SNB condition or a blocking probability guarantee. The proposed 3dF design model for OCS can be applied to a data center network that directly links terminals while preserving blocking quality.

We presented the model designed for OCS-based data center networks operating at the circuit level rather than the packet level. Our approach is particularly well-suited for workloads with stable or predictable traffic patterns, such as those often encountered in high-performance computing environments. For instance, as noted in [43], the authors observed that traffic between aggregation switches in a data center, corresponding to terminals defined in this paper, tends to be stable, suggesting that OCS networks can effectively accommodate such traffic. However, data center networks can exhibit high diversity in traffic characteristics, such as packet size, interarrival times, and burstiness. Our model is thus more suitable for applications with stable traffic patterns, but further research may be needed to assess its performance for highly variable or bursty workloads.

A. DESIGN FOR THREE-STAGE FOLDED CLOS NETWORK (3F)

A. Structure of 3F

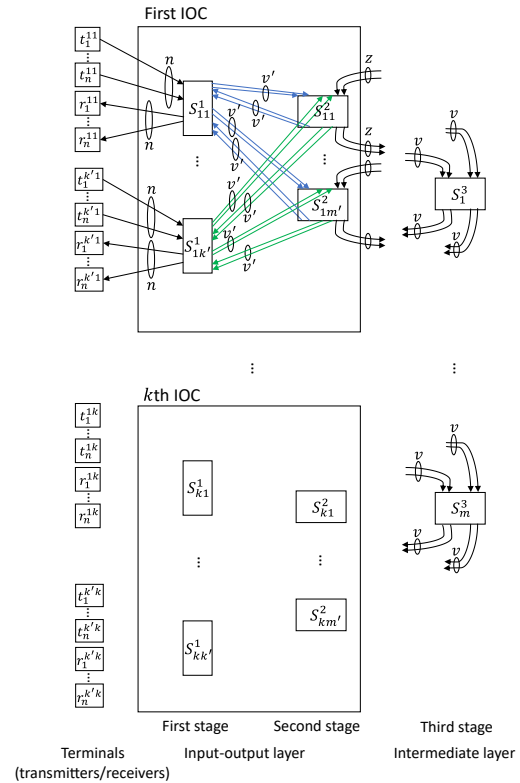


Fig. 18. Structure of 3F.

Fig. 18 shows the 3F structure consisting of input-output and intermediate layers. Similarly to 3TF described in Section 4.A, 3F

comprises $k(k' + m') + m$ switches, each of which has N input ports and N output ports. The input-output layer corresponds to the first and second stages, and the intermediate layer corresponds to the third stage. The input-output layer comprises k switching-network components, each of which, referred to as an IOC, has $k' + m'$ switches, and the intermediate layer comprises m switches. How to connect k' switches and m' in an IOC of 3F is different from that of 3TF.

B. Formulation of 3F design model

The DP of 3F with blocking probability guarantee is given by:

$$\max \quad nk'k \quad (22a)$$

$$\text{s.t.} \quad n \leq \left\lfloor \frac{N}{2} \right\rfloor \quad (22b)$$

$$v'm' \leq \left\lfloor \frac{N}{2} \right\rfloor \quad (22c)$$

$$\text{Eq. (1c)} - \text{Eq. (1m)}. \quad (22d)$$

Eq. (22a)–Eq. (22d) can be explained in the same way as Eq. (1a)–Eq. (1m) in 3TF. Eq. (22b) and Eq. (22c) in 3F replaces Eq. (1b) in 3TF. Other constraints are the same as Eq. (1c)–Eq. (1m).

REFERENCES

1. K.-I. Sato, "Optical switching will innovate intra data center networks [invited tutorial]," *J. Opt. Commun. Netw.* **16**, 1–23 (2024).
2. M. Taubenblatt, P. Maniotis, and A. Tantawi, "Optics enabled networks and architectures for data center cost and power efficiency," *J. Opt. Commun. Netw.* **14**, A41–A49 (2022).
3. Y. Shen, W. Wang, T. Liu, and J. Zhang, "Energy-efficient scaling of active electrical/optical switches in hybrid packet/circuit switched data center networks," in *Int. Conf. Opt. Commun. Netw.*, (2021), pp. 1–3.
4. K. Anazawa, T. Inoue, T. Mano, H. Nishizawa, and E. Oki, "Efficient fiber-inspection and certification method for optical-circuit-switched datacenter networks," *J. Opt. Commun. Netw.* **16**, 788–799 (2024).
5. H. Liu, Urata *et al.*, "Lightwave fabrics: At-scale optical circuit switching for datacenter and machine learning systems," in *Proc. ACM SIGCOMM 2023 Conf.*, (2023).
6. Y. Mori and K.-I. Sato, "High-port-count optical circuit switches for intra-datacenter networks," *J. Opt. Commun. Netw.* **13**, D43–D52 (2021).
7. C. Clos, "A study of non-blocking switching networks," *Bell Syst. Tech. J.* **32**, 406–424 (1953).
8. A. Jajszczyk, "Nonblocking, repackable, and rearrangeable Clos networks: fifty years of the theory evolution," *IEEE Commun. Mag.* **41**, 28–33 (2003).
9. W. Kabaciński, *Nonblocking Electronic and Photonic Switching Fabrics* (Springer, 2005).
10. E. Oki, Z. Jing, R. Rojas-Cessa, and H. Chao, "Concurrent round-robin-based dispatching schemes for Clos-network switches," *IEEE/ACM Trans. Netw.* **10**, 830–844 (2002).
11. E. Oki, N. Yamanaka, K. Nakai, and N. Matsuura, "Multi-stage switching system using optical WDM grouped links based on dynamic bandwidth sharing," *IEEE Commun. Mag.* **41**, 56–63 (2003).
12. E. Oki, H. Taka, and T. Inoue, "Enhancing capacity of optical circuit switching clos network in data center: Progress and challenges," in *2024 33rd Int. Conf. Compu. Commun. and Netw. (ICCCN)*, (2024), pp. 1–9.
13. T. Mano, T. Inoue, K. Mizutani, and O. Akashi, "Increasing capacity of the Clos structure for optical switching networks," in *2019 IEEE Global Commun. Conf.*, (2019), pp. 1–6.
14. T. Mano, T. Inoue, K. Mizutani, and O. Akashi, "Redesigning the non-blocking Clos network to increase its capacity," *IEEE Trans. Netw. Serv. Manag.* **20**, 2558–2574 (2023).
15. H. Taka, T. Inoue, and E. Oki, "Design of twisted and folded Clos network with guaranteeing admissible blocking probability," *IEEE Netw. Lett.* **5**, 265–269 (2023).
16. H. Taka, T. Inoue, and E. Oki, "Twisted and folded Clos-network design model with two-step blocking probability guarantee," *IEEE Netw. Lett.* **6**, 60–64 (2024).
17. H. Taka, T. Inoue, and E. Oki, "Design model of twisted and folded Clos network with multi-step grouped intermediate switches guaranteeing admissible blocking probability," *J. Opt. Commun. Netw.* **16**, 328–341 (2024).
18. E. Oki, H. Taka, and T. Inoue, "Toward increasing switching capacity of twisted and folded clos network guaranteeing admissible blocking probability," in *2024 24th Int. Conf. Transparent Opt. Netw. (ICTON)*, (2024), pp. 1–4.
19. A. Jajszczyk, "On combinatorial properties of broadband time-division switching networks," *Comput. Netw. ISDN Sys.* **20**, 377–382 (1990).
20. R. Taniguchi, T. Inoue, K. Anazawa, and E. Oki, "Optical circuit switched three-stage twisted-folded Clos-network design model guaranteeing admissible blocking probability," *J. Opt. Commun. Netw.* **16**, 1104–1115 (2024).
21. E. Oki, R. Taniguchi, K. Anazawa, and T. Inoue, "Expanding optical-circuit-switching multi-stage networks to ensure admissible blocking probability in data centers," in *Int. Conf. Comput. Netw. Commun. (ICNC)*, (2025, invited paper, to appear), pp. 1–5.
22. F. Chiussi, J. Kneuer, and V. Kumar, "Low-cost scalable switching solutions for broadband networking: the atlanta architecture and chipset," *IEEE Commun. Mag.* **35**, 44–53 (1997).
23. R. Hojabr, M. Modarressi, M. Daneshmand, A. Yasoubi, and A. Khonsari, "Customizing clos network-on-chip for neural networks," *IEEE Trans. Comp.* **66**, 1865–1877 (2017).
24. X. Yuan, "On nonblocking folded-Clos networks in computer communication environments," in *2011 IEEE Int. Parallel and Distributed Processing Symp.*, (2011), pp. 188–196.
25. P. Wang, H. Xu, Z. Niu, D. Han, and Y. Xiong, "Expeditus: Congestion-aware load balancing in Clos data center networks," *IEEE/ACM Trans. Netw.* **25**, 3175–3188 (2017).
26. W. Seher and T. Charles Clancy, "Load balancing in data center networks with folded-Clos architectures," in *Proceedings of the 2015 1st IEEE Conf. Netw. Softwarization (NetSoft)*, (2015), pp. 1–6.
27. W. Seher and C. Clancy, "Flow optimization in data centers with Clos networks in support of cloud applications," *IEEE Transactions on Netw. Serv. Manag.* **14**, 847–859 (2017).
28. W. Seher and T. C. Clancy, "Load balancing in data center networks with folded-Clos architectures," in *Proc. the 2015 1st IEEE Conf. Netw. Softwarization*, (IEEE, 2015), pp. 1–6.
29. S. Ohta and D. Miyamoto, "Implementation and experimental evaluation of the rebalancing algorithm for folded Clos networks," in *2022 32nd Int. Telecommun. Netw. Appl. Conf.*, (2022), pp. 312–315.
30. D. Li, M. Xu, Y. Liu, X. Xie, Y. Cui, J. Wang, and G. Chen, "Reliable multicast in data center networks," *IEEE Trans. Comput.* **63**, 2011–2024 (2013).
31. C. Camarero, C. Martinez, and R. Beivide, "On random wiring in practicable folded clos networks for modern datacenters," *IEEE Trans. Parallel Distrib. Syst.* **29**, 1780–1793 (2018).
32. C. Y. Lee, "Analysis of switching networks," *Bell Syst. Tech. J.* **34**, 1287–1315 (1955).
33. Y. Yang and N. H. Kessler, "Modelling the blocking behaviour of Clos networks," *DIMACS Ser. Discret. Math. Theor. Comput. Sci.* **42** (1998).
34. V. Abramov, S. Li, M. Wang, E. W. M. Wong, and M. Zukerman, "Computation of blocking probability for large circuit switched networks," *IEEE Commun. Lett.* **16**, 1892–1895 (2012).
35. M. Głabowski and M. Sobieraj, "Point-to-group blocking probability in switching networks with multi-service sources and bandwidth reservation," in *2014 9th International Symposium on Communication Systems, Networks and Digital Sign (CSNDSP)*, (2014), pp. 93–98.
36. D. R. B. Araújo, C. J. A. Bastos-Filho, and J. F. Martins-Filho, "Artificial neural networks to estimate blocking probability of transparent optical networks: A robustness study for different networks," in *2015 17th Int. Conf. Transparent Opt. Netw. (ICTON)*, (2015), pp. 1–4.
37. D. R. B. de Araujo, C. J. A. Bastos-filho, and J. F. Martins-filho, "Methodology to obtain a fast and accurate estimator for blocking probability of

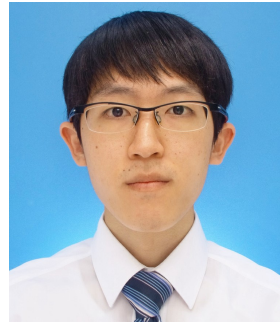
- optical networks," *J. Opt. Commun. Netw.* **7**, 380–391 (2015).
38. J. Lin, T. Chang, Z. Zhai, S. K. Bose, and G. Shen, "Wavelength selective switch-based Clos network: Blocking theory and performance analyses," *J. Light. Technol.* **40**, 5842–5853 (2022).
 39. W. Wang, M. Khazraee, Z. Zhong, M. Ghobadi, Z. Jia, D. Mudigere, Y. Zhang, and A. Kewitsch, "TopoOpt: Co-optimizing network topology and parallelization strategy for distributed training jobs," in *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*, (2023), pp. 739–767.
 40. Y. Tang, H. Guo, Y. Zhu, T. Yuan, X. Gao, C. Wang, and J. Wu, "Effectively reconfigure the optical circuit switching layer topology in data center network by OCBridge," *J. Light. Technol.* **37**, 897–908 (2019).
 41. C. Liang, X. Song, J. Cheng, M. Wang, Y. Liu, Z. Liu, S. Zhao, and Y. Cui, "Negotiator: Towards A simple yet effective on-demand reconfigurable datacenter network," in *Proceedings of the ACM SIGCOMM 2024 Conference, ACM SIGCOMM 2024, Sydney, NSW, Australia, August 4-8, 2024*, (ACM, 2024), pp. 415–432.
 42. M. Murakami, H. Kubokawa, K. Sugiura, E. Oki, S. Okamoto, and N. Yamana, "Online parameter tuning of the flow classification method in the energy-efficient data center network "holst"," *J. Opt. Commun. Netw.* **12**, 344–354 (2020).
 43. L. Poutievski *et al.*, "Jupiter evolving: Transforming google's datacenter network via optical circuit switches and software-defined networking," in *SIGCOMM*, (2022), pp. 66–85.
 44. "JNP-QSFP-100G-PSM4," (Accessed: 1-Nov.,-2024). <https://edgeoptic.com/products/juniper/compatible-juniper-100g-optics/jnp-qsfp-100g-psm4/>.
 45. "QSFP-100G-PSM4," (Accessed: 1-Nov.,-2024). <https://edgeoptic.com/products/arista/qsfp-100g-psm4-2/>.
 46. "G4 NTM," (Accessed: 1-Nov.,-2024). <https://www.telescent.com/products>.
 47. "IP-FT902B fiber optic patch panels," (Accessed: 1-Nov.,-2024). <https://infilinktechnologies.com/product/ip-ft902b/>.
 48. "OS2 singlemode optical fiber (global)," (Accessed: 1-Nov.,-2024). <https://www.panduit.com/content/dam/panduit/en/products/media/0/10/310/6310/100656310.pdf>.



Eiji Oki is a Professor at Kyoto University, Kyoto, Japan. He was with Nippon Telegraph and Telephone Corporation (NTT) Laboratories, Tokyo, from 1993 to 2008, and The University of Electro-Communications, Tokyo, from 2008 to 2017. From 2000 to 2001, he was a Visiting Scholar at Polytechnic University, Brooklyn, New York. His research

interests include routing, switching, protocols, optimization, and traffic engineering in communication and information networks.

Ryotaro Taniguchi is pursuing the M.E. degree in the Graduate School of Informatics, Kyoto University, Kyoto, Japan. He received the B.E. degree from the Undergraduate School of Electrical and Electronic Engineering, Kyoto University, Japan, in 2024. His research interests include network design, optimization, and algorithm design in data centers.



Kazuya Anazawa is a researcher at Nippon Telephone and Telegraph (NTT) Corporation Network Innovation Laboratories, Japan. He received his B.E. (2016) and M.E. (2018) degrees in computer science and engineering from the University of Aizu. In 2018, he joined NTT Network Innovation Laboratories. His research interests include optical network design and autonomous control of optical networks. He serves as a technical co-lead of open optical & packet transport network operating system (OOPT-NOS) group in Telecom Infra Project.



Takeru Inoue is a Distinguished Researcher at Nippon Telegraph and Telephone Corporation (NTT) Laboratories, Japan. He received the B.E. and M.E. degrees in engineering science and the Ph.D. degree in information science from Kyoto University, Japan, in 1998, 2000, and 2006, respectively. In 2000, he joined NTT Laboratories. From 2011 to 2013, he was an ERATO Researcher with the Japan Science and Technology Agency, where his research focused on algorithms and data structures. His research interests widely cover algorithmic approaches in communication networks.