Realizing Dynamic Optical Networking

Mario Baldi

Synchrodyne Networks, Inc.
baldi@synchrodyne.com
Also with Turin Polytechnic
Computer Engineering Department

Yoram Ofek

Synchrodyne Networks, Inc.
ofek@synchrodyne.com

Electronic asynchronous packet switching has been developed since the early 1960s and now, at the dawn of the 21st century, it is a well-understood technology. On the other hand, all-optical packet switching, a major candidate for the realization of dynamic optical networking, is a new technology, currently under-going research and development. In this paper we show that the transition from electronic to optical is not as straightforward as one would like to think.

The main thesis of this paper is that the transition to dynamic all-optical switching will require changing the current asynchronous packet switching paradigm. This is shown by focusing the analysis on the simple problem of optical memory, while not considering the more complex problems associated with optical packet header processing and switching as if they had been resolved – which is indeed not the case.

In order to evaluate optical memory, various comparative measures are used. For example, much more glass (Silicon) is needed for optical memory versus electronic memory, in fact a factor of 1 million! If we assume shrinking of Silicon circuitry by a factor of two every 18 months, the transition to optical packet switching will translate into reversing Moore's law by 30 years.

It will be shown that a *common time reference* (CTR™) is needed for realizing the optical random access memory (O-RAM) required for dynamic optical networking. Furthermore, it is shown that by using *coordinated universal time* (UTC a.k.a. GMT) CTR™ can be globally distributed, then it is possible to (globally) distribute the O-RAM and, consequently, to realize a new dynamic optical networking architecture called *Fractional Lambda Switching* (FλS™), that additionally, does not require optical header processing.

I. INTRODUCTION

The more optical transmission systems are deployed, the greater the bottlenecks in electronic processing and switching may become. Consequently, a solution to these bottlenecks is sought in the optical domain where no electronic conversion is necessary. The following definitions are applied throughout this paper.

Definition 1: *Dynamic optical switching* – Each data unit is transmitted end-to-end over an arbitrary topology network through optical fibers and switches with no conversion to electronics such that each data unit on a given optical channel may be treated differently by the network's switches.

Definition 2: *Static optical switching* – All data units that are transmitted end-to-end over an arbitrary topology network through optical fibers and switches with no conversion to electronics such that all data units on a given optical channel are treated the same way by the network's switches.

Dynamic optical networking implies *dynamic optical switching* that is clearly much closer to current packet and circuit switching than *static optical switching*. *Static optical switching* requires a separate optical channel—a.k.a. wavelength or lambda (λ)—from each source to each possible destination. Consequently, if each source is also a destination, n sources require (n^2-n) optical channels. This n square requirement is a major limit to the scalability of *static optical switching* and hence a compelling reason for resorting to *dynamic optical switching*.

Section II shows that mapping the current asynchronous packet switching model to all-optical networks is highly unadvisable, if at all possible. The importance of time for dynamic optical networking is discussed in Section III, while Section IV shows how time can be used to enable a synchronous switching model suitable for all-optical realization.

Asynchronous packet switching is based on four components: transmission, memory, processing, and switching. In order to use this paradigm to achieve dynamic optical switching, the four components should be implemented in the optical domain. Optical transmission is widely deployed, optical switching is commercially viable, optical processing is still very difficult, and optical memory using optical fibers is feasible. However, Section II.B shows that optical memory, which is well understood, is a major implementation obstacle hindering the realization of all-optical asynchronous packet switching regardless of whatever progress in optical processing takes place.

Various memory reduction techniques can alleviate the optical memory problem of packet switching. Such techniques fall under the two following categories:

(1) Randomized-based techniques, e.g., "hot potato", deflection or convergence routing. The network has unpredictable delay/throughput performance and consequently it may not be suitable for streaming media applications.

(2) Temporal-based techniques. Scheduled transmission provides predictable delay and throughput performance. Consequently, these techniques are suitable for streaming media applications, which are anticipated to fill up high capacity optical networks.

Section III analyses some temporal-based techniques, which use different time measurement methods, while highlighting implications of their optical realization. As explained in more detail in Section III, the temporal-based switching paradigm underlying SONET requires a limited amount of memory due to SONET's byte-by-byte channel multiplexing. Since byte-by-byte de-multiplexing into multiple STS-1 frames cannot be done in the optical domain, it is necessary to separately switch each incoming byte from input to output. This will require, for OC-192 channels, byte-by-byte optical processing and switching times that are well below 100 picoseconds, which is far beyond current technology. Consequently, SONET switching is not considered a viable alternative for dynamic optical networking.

Section IV presents an easily realizable all-optical switching solution that has low optical memory requirements and is suitable for streaming media applications, both interactive and non-interactive, on a global scale. This solution, which is called Fractional Lambda Switching (FλS), leverages on a Common Time Reference (CTR) with Pipeline Forwarding (PF) to provide the required switching granularity with predictable performance guarantees. Furthermore, thanks to the CTR with PF, this solution substantially reduces — possibly eliminating — the need for optical header processing; hence, *dynamic optical networking*, for the first time, becomes an impending reality. Since the paper focuses on showing that FëS enables the realization of *dynamic optical networking*, FëS performance issues are the subjects of separate publications.

II. COMPARATIVE EVALUATION OF OPTICAL MEMORY

A. Model

This work is based on the following assumptions that are derived from what is known in physics today, not from what may be discovered in the future.

Assumption 1: *Optical memory is implemented with optical fiber, made from silicon, with a diameter of 125mm.*

Thus, optical memory is actually a delay line or a pipeline with continuously streaming bits. This will have some major consequences that will be discussed later.

Assumption 2: *The speed of light in optical fiber is 200,000 kilometers per second*[1].

Assumption 3: *Currently, there is no practical all-optical wavelength conversion from any λ to any other λ.*

[1] Recent publications on "stopping the light" (see [1][2]) describe a complex physical experiment realized at near absolute zero temperature – exploiting the revesible property of quantum phase.

In the context of this work the following two definitions are applied:

Definition 3: *Bulk Optical Memory (BOM)*. An optical fiber capable of storing a predefined amount of bits encoded within an optical signal. The access to such bulk optical memory is strictly sequential, or first-in-first-out (FIFO).

For example, a 1 Km fiber is a bulk optical memory that can store 50,000 bits at 10 Gb/s.

Definition 4: *Optical Random Access Memory (O-RAM)*. An optical memory wherein each of the stored data units can be accessed at any predefined time[2], independent of the order in which they had entered the O-RAM.

For example, 1 Km of fiber at 10 Gb/s tapped (e.g., with 1-by-2 optical switches) at regular 10 meter intervals is an O-RAM capable of storing up to 50,000 bits encoded in data units of 500 bits each. The data units can be accessed in any order through the taps, however, since the data units are travelling at the speed of light the access to a data unit at a different time will be done through a different tap.

B. Device Level Analysis of Bulk Optical Memory

The physical dimension of optical memory is compared with that of electronic memory (i.e., a Dynamic RAM (DRAM)) with equivalent capacity.

B.1 Raw Material

A synchronous DRAM chip capable of storing 256 Mbits is manufactured with state of the art technology on a $10\cdot10^{-3}\cdot10\cdot10^{-3}\cdot0.5\cdot10^{-3}=50\cdot10^{-9}$ meter^3 (or $50\cdot\mathbf{10^{-6}}$ Liter) silicon chip.

A 256 Mbit optical memory for an optical signal encoded at 10 Gb/s (thus, each bit is stored in $2\cdot10^{-2}$ meters of fiber) is realized with a $256\cdot10^{6}\cdot2\cdot10^{-2}$ =5,120,000 meter fiber. Since the fiber diameter, core and cladding, is $125\cdot10^{-6}$ meter, the total volume is $\pi\cdot(125\cdot10^{-6}/2)^2\cdot5{,}120{,}000=62.8\cdot10^{-3}$ meter^3 (or 62.8 Liters).

Hence, the step from DRAM to optical memory corresponds to a decrease of more than 1,000,000 folds in the information density. Consequently, if we assume shrinking of silicon circuitry by a factor of two every 18 months (as stated by Moore's law), the transition to optical memory translates into going back 30 years!

B.2 Packaging

The above mentioned 256 Mbits synchronous DRAM chip is contained in a 66 pin package whose volume is $10\cdot10^{-3}\cdot23\cdot10^{-3}\cdot1\cdot10^{-3}= 230\cdot10^{-9}$ meter^3 (or $230\cdot10^{-6}$ Liter).

The fiber realizing an equivalent optical memory can, for example, be rolled on a number of spools. For the sake of comparison, the packaging of Corning fiber is considered here. The longest spools Corning sells are of

25.2 Km. Hence, building a 256 Mb optical memory requires 5,120/25.2=204 spools which occupy a volume of 767 Liters (vs. $230 \cdot 10^{-6}$ Liter of DRAM) and weigh 461 Kg.

C. Sub-system Analysis of Bulk Optical Memory

The device level comparison between electronic memory and optical memory clearly indicates that in order to provide the required memory capabilities, optical packet switches are going to be many orders of magnitude larger than conventional electronic packet switches. Even though the previous section could provide strong enough motivation to proceed to the analysis techniques with smaller memory requirements (i.e., to proceed to Section IV), we concluded that a sub-system analysis is important in order to understand the temporal dimension in realizing optical memory. This section first analyzes sub-system BOM and then Section D analyzes sub-system O-RAM.

Table 1 shows the amount of memory per DWDM channel on state-of-the-art terabit packet switches (see [3][4] for more details). Various physical dimensions of the BOM required to implement the same amount of buffers for one 10 Gb/s DWDM channel are shown in Table 2. The required optical fiber length is given in the second column of Table 2, while the third column shows the required optical fiber volume. The volume and weight of the optical fiber rolled on 25.2 Km spools (as sold by Corning) are provided in the last two columns of Table 2.

Product	**Memory per channel [MB]**
Optera packet solution (Nortel)	22
Terabit switch router system (Avici)	64 - 128
M160 (Juniper)	128
20000 Terabit Network Router (Pluris)	128
GSR12016 (Cisco)	128 - 512
Aranea-1 (Charlottes' Web)	1000
Average	***256***

Table 1: Memory per optical channel in current terabit routers

Note that if the switches in Table 1 have 1000 DWDM optical channels the numbers in Table 2 will increase 1000 fold. For example, the memory for an all-optical switch with 256 MB of optical memory per channel will be 3,686 tons. A network with 2,000 of such switches would weigh more than the Great Pyramid of Khufu in Giza, which weighs 6 million tons.

---

[2] In general RAM can be accessed *at any time.*

Buffer Size [MBytes]	Fiber Length [Km]	Fiber Volume [Lt]	Spool Volume [Lt]	Spool Weight [Kg]
22	3,520	43	527	317
128	20,480	251	3,068	1,843
256	40,960	503	6,136	3,686
1,000	160,000	1,963	23,969	14,400

Table 2: Physical dimensions of a BOM for a single 10 Gb/s optical channel

Obviously, weight is not the only realization constraint; fiber length is another. Amplification is required to compensate the attenuation introduced by transmission over long fibers. Moreover, the signal degenerates due to the distortion introduced along the fibers and by the amplifiers; hence, after traveling a given distance it is necessary to regenerate the signal. Regeneration is an electrical, not optical, process and encompasses re-amplification, re-timing, and re-shaping that are performed by repeaters. As shown in Table 2, the fiber length for a 256 Mbyte buffer is 40,960 Km (the same as the earth circumference). Thus, given the transmitted optical signal's finite power budget, the BOM will require multiple electronic regenerations. Table 3 shows the number of amplifiers and repeaters required for the implementation of a number of BOM configurations for a 10 Gb/s optical channel. The figures in Table 3 are devised assuming that amplification is performed every 80 Km and regeneration every 480 Km. Table 3 also shows the total attenuation of the required length of fiber assuming a 0.2 dB/Km nominal attenuation, as specified by some Corning products.

Memory Size [MBytes]	Number of Repeaters	Number of Amplifiers	Fiber Attenuation [dB]
22	7	37	704
128	42	214	4,096
256	85	427	8,192
1,000	333	1,667	32,000

Table 3: Amplification, regeneration, and fiber attenuation[3] of a BOM for a single 10 Gb/s optical channel

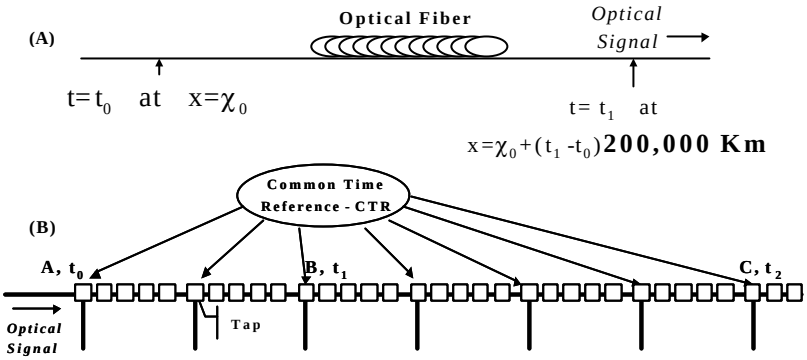
D. Sub-system Analysis of Optical Random Access Memory – O-RAM

Random access implies that at any given time any part of the memory can be accessed. Since the stored optical bits travel along the optical fiber at the speed of light, as shown in Figure 1(A), at any given time the bits are in another position along the optical fiber. Consequently, in order to access such optical memory there are two basic requirements, as shown in Figure 1(B):

1. *Infinite number of taps:* a tap realized by a 1-by-2 switch enables the light that is stored in the fiber to either continue along the fiber or be switched out of the fiber.

[3] With Raman amplifiers it may be possible to perform regeneration every 2,000 Km.

2. *Common Time Reference (CTR)*: a precise knowledge of time is required in order to access a given sequence of bits at a given spot along the optical fiber. This knowledge of time should be based on the same reference along the optical fiber, which is why, this timing requirement is called Common Time Reference or CTR.



Realistic realization:
periodic (equally spaced) taps = pipeline forwarding

Figure 1: Optical random access memory – O-RAM (B) – is time dependent (A)

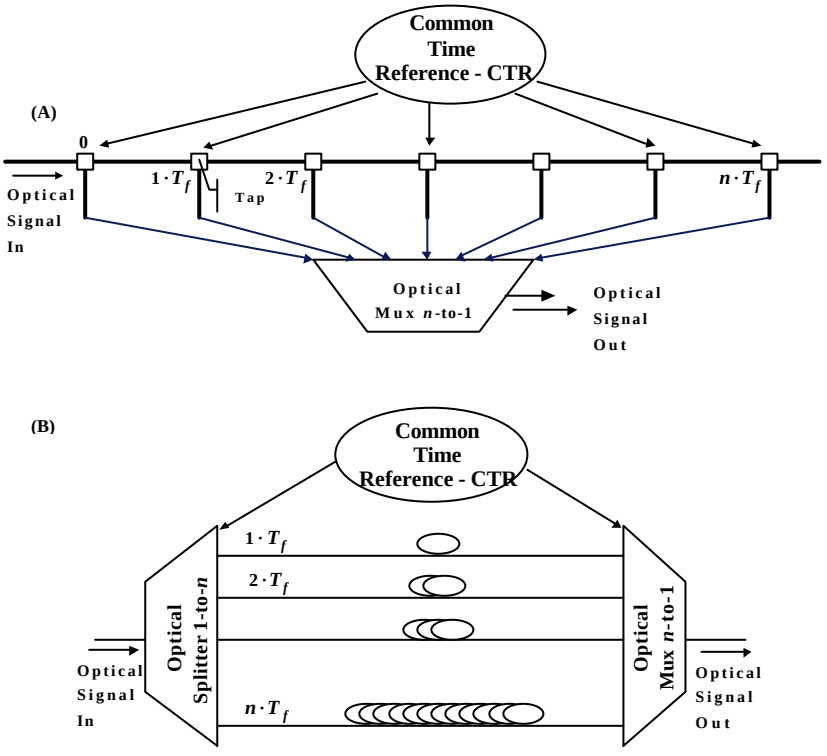


Figure 2: CTR with (A) linear O-RAM (B) parallel O-RAM

Since an infinite number of taps is not feasible, two types of O-RAM can be found in the literature: (1) linear delay line or linear O-RAM and (2) parallel delay line or parallel O-RAM, shown in Figure 2(A) and Figure 2(B), respectively.

D.1 Linear O-RAM

As shown in Figure 2(A), a linear delay line O-RAM is realized by inserting taps at predefined equally spaced intervals into the optical fiber. An optical signal propagating along the fiber can be fetched at regular intervals; we call a *time frame* (TF) the interval between taps with duration T_f. A data unit that was sent into the fiber at time $t=0$ can be fetched from the fiber at times: $t=1 \cdot T_f$, $2 \cdot T_f$, $3 \cdot T_f$, and so on. The physical dimensions of a linear delay line at 10 Gb/s are shown in Table 2. Table 4 provides figures of the power budget related to tap points and the size of the components used to implement them. Each tap point can be implemented by means of either a coupler or a 1x2 switch. Table 4 shows the overall volume occupied by the components, both couplers and switches, implementing the required tap points for an optical packet switch operating with 8 μseconds time frames.

A coupler splits the power of the incoming signal between the output signals; thus, the higher the power of the tapped signal, the more the signal continuing through the optical memory is attenuated. Switches introduce a fixed insertion loss. The last two columns of Table 4 shows the overall attenuation resulting from the switches implementing the tap points and total attenuation due to both switches and fiber, respectively, assuming a 0.6 dB insertion loss per switch (a value published by JDS Uniphase for some of its components). The attenuation introduced by either couplers or switches implementing tap points results in a need for a larger number of amplifiers and regenerators.

Buffer Size [MBytes]	Number of Taps	Tap Coupler Volume [Lt]	Switch Volume [Lt]	Tap Insertion Loss [dB]	Total Attenuation (incl. Fiber) [dB]
22	2,200	0.8	4	1,320	2,024
128	12,800	4.5	22	7,680	11,776
256	25,600	9.0	44	15,360	23,552
1,000	100,000	35.3	173	60,000	92,000

Table 4: Linear O-RAM with 8 μsecond time frames

D.2 Parallel O-RAM

Parallel O-RAMs are more widely known as fiber delay lines (FDLs). As shown in Figure 2(B), fibers of different lengths are deployed to delay data units for different amounts of time. The delay experienced by data units in a parallel O-RAM has predefined granularity with the T_f as the basic time unit, which in turn determines the optical memory granularity – e.g., 10 KB. The number of parallel fibers needed to realize a parallel O-RAM depends on the optical memory size divided by the granularity. The length of each fiber is determined by its delay such that the first fiber delays by $1 \cdot T_f$, the second fiber delays by $2 \cdot T_f$, and so on.

Optical Memory Size [MB]	Fiber Length [Km]	Fiber Volume [m^3]	Spool Volume [m^3]	Spool Weight [Ton]
22	3,873,760	48	580	349
128	131,082,240	1,609	19,637	11,797
256	524,308,480	6,434	78,543	47,187
1,000	8,000,080,000	98,176	1,198,438	720,002

Table 5: Parallel O-RAM with 10 Gb/s optical channel delay line

Table 5 shows the physical dimensions of the fiber needed to implement a few optical memory configurations for a single 10 Gb/s DWDM channel, assuming that the optical memory has a 10 KB granularity. Note that the spool weight of only 10 DWDM channels is more than the weight of the Great Pyramid of Khufu in Giza, which weighs 6 million tons. The memory for an optical packet switch with 1,000 DWDM channels will weigh 100 Khufu pyramids. Obviously, given such "funny" numbers, parallel O-RAM is not a practical optical memory approach.

E. Discussion

The current asynchronous packet switching paradigm is a well-established networking paradigm with realistic implementation based on optical transmission and electronic switching. The all-optical realization of this paradigm presents multiple challenges. This section analyzed the optical memory challenge and showed that even though it is well understood and in principle feasible, its implementation is not realistic because of the optical memory's huge physical size.

It was further proved that optical random access memory (O-RAM) is time dependent and requires a Common Time Reference – CTR. As will be shown in Section IV, a CTR is not only required for realizing O-RAM, but it is also essential for minimizing the amount of memory required per optical channel. Furthermore, by using time it is possible to eliminate the need for optical header processing, which is the main remaining challenge for the realization of dynamic all-optical switching.

III. TIME AND DYNAMIC OPTICAL NETWORKING

This section studies the relationship between time measurements and scheduling in communications networks, in general, and dynamic optical networking, in particular. The broad approach is taken in order to provide the rationale for the new dynamic optical networking architecture proposed in Section IV. At the end of this section, we discuss how SONET uses time and its scheduling differs from the one described here, thus making SONET not suitable for dynamic optical networking.

A. Time Measurement

Measuring time between two events in the same location is performed locally by counting periodic rotations of various sorts. In ancient era the time was measured by counting the earth rotations, or, as some argued, the sun rotations around the earth. Since then, the measurement of time has improved dramatically.

Measuring time between two events at two physically separated points is not simple, and it has been a subject of extensive pursuit since time immemorial. The necessary condition for such time measurement is to have a *common time reference*. Again, in ancient era various celestial bodies were used, via complex measurements and computations, as a common time reference.

Figure 3: Global distribution of UTC through: GPS, GLONASS, Galileo, or TWTFT

Today, a common time reference has been established by the *time-of-day* international standard that is called *Coordinated Universal Time* or *UTC* (a.k.a. *Greenwich Mean Time* or *GMT*). Specifically, time is measured by counting the oscillations of the cesium atom in multiple locations. In fact, 9,192,631,770 oscillations of the cesium atom define one UTC second. As shown in Figure 3, UTC is available everywhere around the globe from several distribution systems, such as, GPS (USA satellites system) [5], GLONASS (Russian Federation satellites system) [6], and in the future by Galileo (European Union and Japanese satellites system) [7]. There are other means for distribution of UTC, such as, CDMA cellular phone systems and a TWTFT (Two-Way Satellite Time and Frequency Transfer) [8] technique based on general purpose communications satellites.

B. Scheduling

Scheduling requires the ability to measure time. We consider scheduling with two time measurement methods:

1. Scheduling with local time based measurements. The delay between nodes cannot be measured, and therefore, the scheduling is based on local time. This method is used in circuit switching (e.g., SONET), where the local clock accuracy is established by international standards: *Stratum 1, 2, 3, and 4 clocks.* In packet networks scheduling with local time is used by a number of queuing schemes, such as, for example, Weighted Fair Queuing (WFQ).
2. Scheduling with UTC-based measurements. The delay between nodes can be measured by using UTC and scheduling can be based on UTC. Scheduling with UTC implies no clock slips or drifts, and consequently, very simple implementation. UTC-based scheduling, possibly with very low accuracy, is used in every day life.

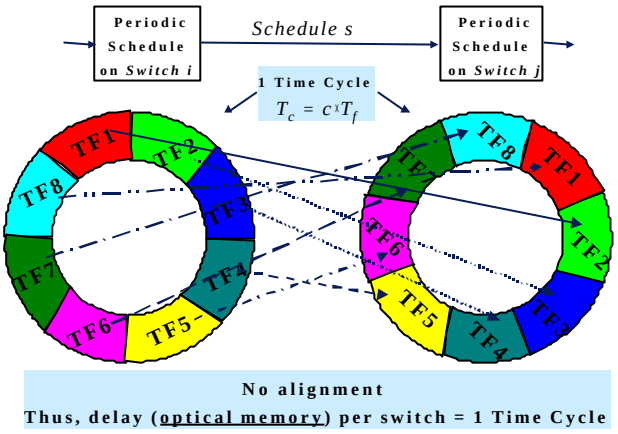


Figure 4: Local time based-scheduling

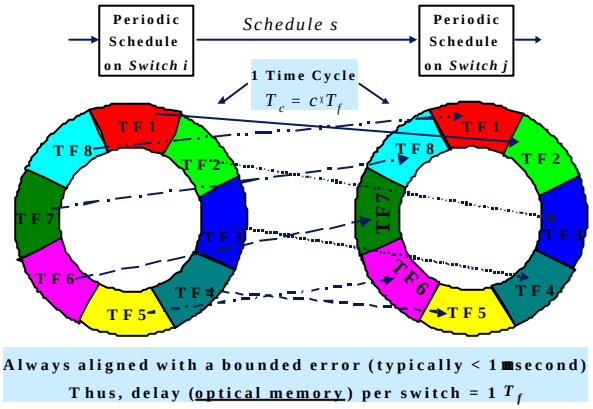


Figure 5: UTC-based scheduling

Figure 4 and Figure 5 are examples of the above two scheduling methods (without loss of generality, the propagation delay between *Switch i* and *Switch j* was ignored). In these examples, scheduling is periodic and time is divided into time frames (TF) of predefined duration T_f. For example, a time frame of 10 µseconds is obtained by dividing one UTC second by 100,000. For periodic scheduling c time frames are grouped into *time cycles*; for example, c=1,000 time frames of 10 µseconds create a 10 milliseconds time cycle.

C. Per Switch Delay and Optical Memory

Let's assume that two neighboring switches, *Switch i* and *Switch j*, perform a given task — e.g., switching or transmiting data units — during predefined time frames according to a schedule, *Schedule s*. *Schedule s* repeats every time cycle, T_c, where $T_c = c \times T_f$. In the examples in Figure 4 and Figure 5, $T_c = 8 \times T_f$, and *Schedule s* on *Switch i* during time frame *k*, is scheduled on *Switch j* during time frame $(k+1) \bmod 8$.

When the scheduling on *Switch i* and *Switch j* is based on local time, the delay between *Schedule s* on *Switch i* and on *Switch j* is not known, and consequently, the delay between TF *k* and TF $(k+1) \bmod c$ is not known. Since the schedule repeats every time cycle, the maximum delay between a TF on *Switch i* and the corresponding TF on *Switch j* is one time cycle, T_c.

When the scheduling on *Switch i* and *Switch j* is based on UTC, the delay between *Schedule s* on *Switch i* and on *Switch j*, is known, and consequently, the delay between TF *k* and TF $(k+1) \bmod c$ is known. As a result, the maximum time between the execution of the aforementioned task in *Switch i* and in *Switch j* is one time frame – T_f (which results from *quantization delay* – because the actual data unit propagation delay between the two switches is not an integer number of time frames). Since data units need to be stored while waiting for the task execution in *Switch j*, the time between the two task exectutions determines the amount of (optical) memory required within the switches.

D. SONET

SONET switches operate according to a reoccurring schedule that, as was mentioned before, is based on a local clock; consequently, data traversing a SONET switch are delayed up to a whole time cycle. Due to byte-by-byte channel multiplexing, the SONET time cycle is the time between the transmission of two successive bytes of the same channel. For example, the time cycle — hence the scheduling delay — of an STS-1 switch with OC-48 interfaces is 125/48 = 2.6 μs.

As pointed out earlier, byte-by-byte de-multiplexing of STS-N frames into multiple STS-1 frames cannot be done in the optical domain. Consequently, in order to implement SONET-based dynamic optical switching, each incoming byte must be independently switched from input to output. This requires, for OC-192 channels, byte-by-byte optical processing and switching time well below 100 picoseconds, which is far beyond current technology.

In order to overcome the picosecond accuracy requirements, a SONET look-alike might be devised in which the multiplexed one byte slot size is increased by a factor of *x*. However, since SONET scheduling uses local time measurements, this will imply a factor of *x* increase in the time cycle, and consequently, in the per switch delay and cost optical memory.

Note that increasing the slot size of SONET by a factor of x will anyway not eliminate the need for optical processing of the overhead information, such as, time multiplexing pointers. These pointers are needed since local time measurements on different switches are contiguously drifting from one another.

IV. FλS: THE NETWORK IS THE MEMORY

This section presents *fractional lambda switching* (FλS™) and clarifies why it is a viable dynamic optical networking solution. FλS™ *dynamically* allocates fractions of an optical channel (i.e., lambda) over predefined routes in the network. Each lambda fraction, or *fractional lambda pipe* (FλP™), is equivalent to a leased line in circuit switching.

A. The Memory is the Network

There are several ways to explain the operation principles of FλS; in the context of this paper, the most natural way appears to be describing FλS as an optical memory that is distributed over the network, as shown in Figure 6. More specifically, such a network can be viewed as a mesh of linear O-RAMs (see Figure 2(A)) where the taps are optical switches.

In fact, as described in Section II.D.1, a linear O-RAM is implemented by connecting a sequence of optical switches via fixed length fibers, each switch providing access to a stored data unit at a given time. (In general an optical packet consists of a predefined number of data units.) Propagation through a fixed length fiber requires a constant time: one time frame – T_f.

Distributing the optical memory over the network requires the following changes:

1. The propagation delay between two adjacent taps or optical switches should be an integer number of time frames.
2. The 1-by-2 taps or optical switches are replaced with N-by-N optical switches in order to enable the realization of an arbitrary topology.

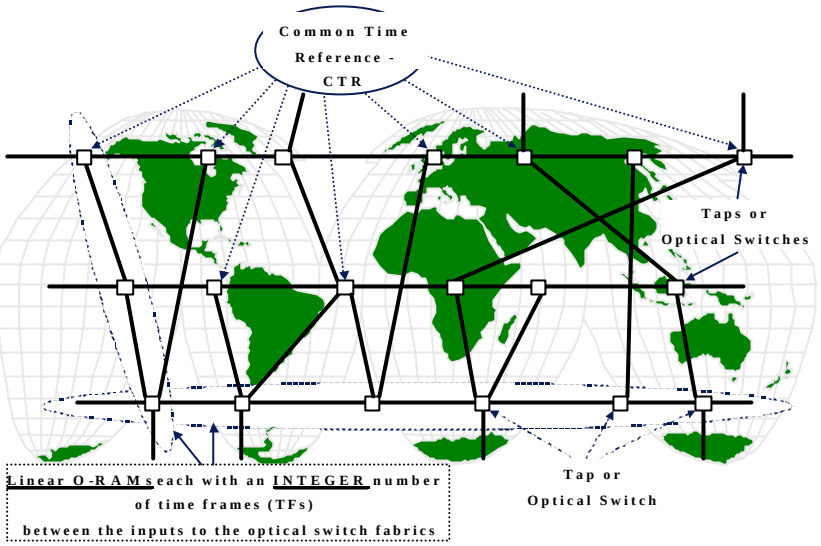


Figure 6: A mesh of linear O-RAMs forming an FλS™ network

As depicted in Figure 2(A), the operation of the optical switches implementing a linear O-RAM requires a Common Time Reference (CTR™) in order to guarantee proper operation. Providing a CTR in one box is simple, but doing it over a network is less obvious. However, since UTC is available globally, it can be used as CTR™ for FλS™ networks. As discussed in Section III.A, UTC is distributed on a global scale by various systems (see Figure 3).

B. Realization of Dynamic Optical Switching with F lS Ô

B.1 F l S Principles of Operation

Fractional lambda (λ) *switching* (FλS) is based on **Pipeline Forwarding (PF Ô) of time frames**, whose basic operation principle is shown in Figure 7. PF provides FëS with the necessary features that enable its all-optical implementation, such as little or possibly no demand for memory and header processing. PF is based on the following principles:

1. **Switching of time frames**: (*i*) each time frame has a predefined duration and contains a *payload* with a predefined size (i.e., number of bytes), (*ii*) between two successive payloads there is a *safety margin* or *idle time* of a predefined duration, and (*iii*) the payload of a time frame is switched as a whole from input to output.
2. **Idle time or safety margin between two successive payloads** is used to change the switching matrix of the optical switch so that the payload of each time frame of a given optical channel can be switched to a different output.

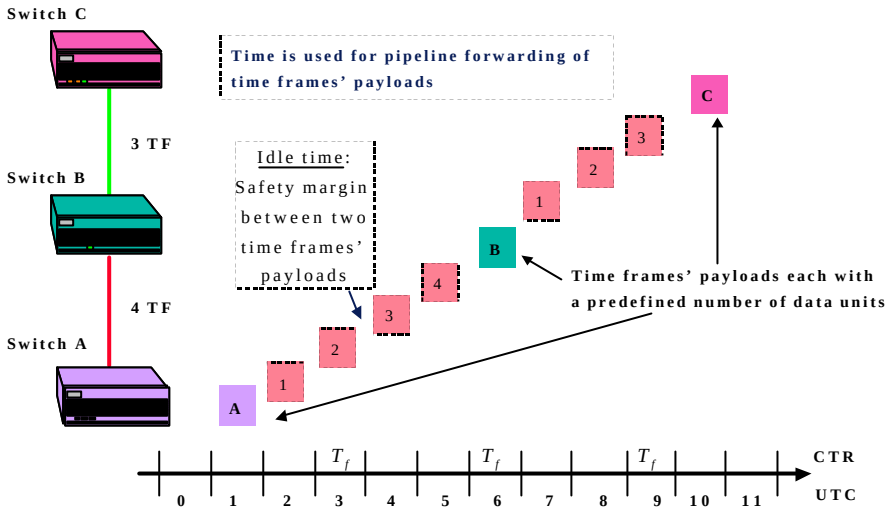


Figure 7: Pipeline Forwarding (PF™) of time frames

3. **CTR is UTC** and is coupled to all the (optical) switches. The UTC second is divided into a predefined number of equal duration time frames—T_f—as shown in Figure 8. Time frames are grouped into *time cycles* and time cycles are grouped into *super cycles*, wherein the super cycle is equal to one UTC second, as shown in Figure 8.

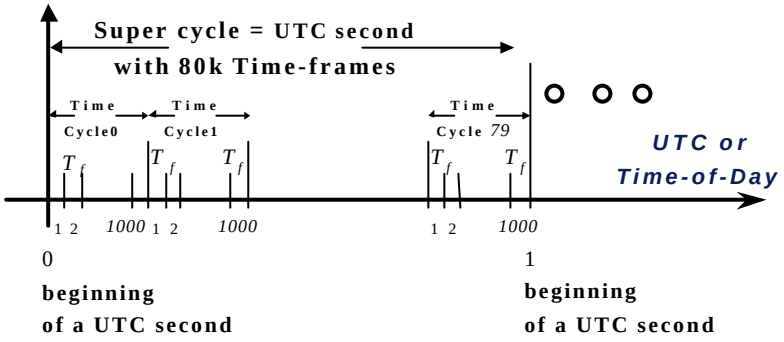


Figure 8: CTR™ with time frames T_f = 12.5 µs

4. **Alignment of all received time frames to UTC**, such that the delay between inputs of adjacent *optical switching fabrics* (and consequently, between the inputs of any two optical switching fabrics) after alignment is an *integer number of time frames*, as shown in Figure 6. The alignment operation is performed before the optical switching, as shown in Figure 9.

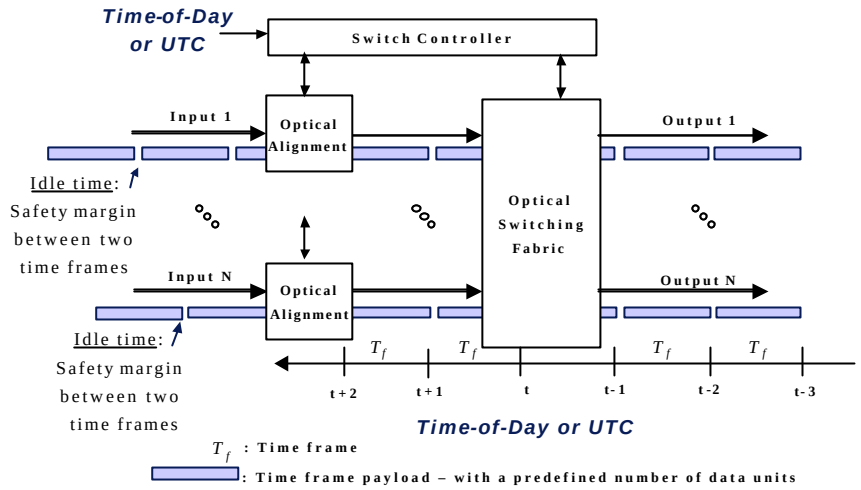


Figure 9: Incoming time frames are aligned with UTC before reaching the optical switching fabric

5. **Periodic switching pattern through the optical switching fabric**, which repeats every time cycle or super cycle. In other words, during every time frame within a time cycle (or super cycle) the optical switching fabric has a predefined input/output configuration and the sequence of input/output configurations repeats every time cycle (or super cycle). This implies that for every time frame within a time cycle (or super cycle) the time frame payload is switched to a predefined output.

B.2 Pipeline Forwarding (PF) over a Fractional λ Pipe (FλP)

A fractional λ pipe (FλP), *p*, is defined along a path of successive FλS switches: $S_p(1)$, $S_p(2)$, … , $S_p(k)$ such that, the forwarding of a time frame along *p* has a predefined schedule. More specifically, let

(i) the delay, in time frames, between successive inputs of the optical switching fabric along *p* be $d_{1,2}$, $d_{2,3}$, … , $d_{k-1,k}$,

(ii) the time cycle and super cycle duration be *c* and *s* time frames, respectively, and

(iii) the scheduling per FλP repeat itself every *c* (*s*) time frames.

Then, a time frame forwarded along path *p* from $S_p(1)$ at t_0 will be forwarded by $S_p(2)$ at $(t_0+d_{1,2}) \bmod c$, by $S_p(3)$ at $t_0+d_{1,2}+d_{2,3} \pmod{c\ (s)}$, and so on; and will reach the last switch of *p*, $S_p(k)$, at $t_0+d_{1,2}+d_{2,3}+\ldots+d_{k-1,k} \pmod{c\ (s)}$.

The capacity allocated per FλP is determined by the number of allocated time frames in every time cycle (or super cycle). For example, if *c*=100 and one time frame per time cycle is allocated for an FλP, then the capacity of the FλP is 1/100 of the optical channel capacity.

Due to Pipeline Forwarding (PF), if all FλPs are unchanged, the switching and forwarding operations along all FλPs at all time frames are known in advance in a reoccurring order every time cycle or super cycle. Consequently, all switches know in advance the switching and forwarding operations that are to be performed

during every time frame. The deterministic operation of PF makes FλS suitable for interactive and non-interactive multimedia applications [9] and group communications [10] on a global scale.

C. CTR/UTC accuracy

The CTR accuracy, and hence UTC accuracy, requirement has a direct impact on cost, stability, and implementation complexity. The idle time between the payloads of two successive time frames serves two objectives: (*i*) to allow enough time for the optical switching fabric to be reconfigured, and (*ii*) to provide a time frame delimiter. When a time frame delimiter exists, the UTC accuracy requirement is $\pm \frac{1}{2} T_f$ (i.e., UTC±1/2·(10 μseconds to 125 μseconds)). The reason for such a relaxed requirement is that the UTC is not used for detecting the time frame boundaries, as they are detected by the idle times. Consequently, the only function of UTC is enabling the correct mapping of the incoming time frames from the optical channel to the CTR time frames. It is easy to show that up to $\frac{1}{2} T_f$ timing error can be tolerated while maintaining the correct mapping of time frames.

Today, a time card with 1 pps (pulse per second) UTC from GPS with accuracy of 10-20 ns is available from multiple vendors. The card is small and costs $100-200. By combining UTC from GPS with local Rubidium or Cesium clocks it is possible to have a correct UTC (±1 μsecond) without an external time reference from GPS for days (with Rubidium clock) and months (with Cesium clock).

V. DISCUSSION

The evaluation of the optical memory required has shown that, due to optical memory limitations, the asynchronous packet switching paradigm could not be realized in the optical domain. Fractional λ switching (FλS™) was proposed as a switching paradigm that minimizes the need for optical memory. Moreover, FλS reduces the complexity of switching and eliminates the need for header processing, whichis a major open problem. Thus, FλS is as realizable as static whole λ switching and consequently dynamic all-optical networking with FλS is viable with state of the art optical components.

FλS uses a global *common time reference* (CTR™), which is realized with UTC (Coordinated Universal Time), to implement *pipeline forwarding* (PF™) of time frames, virtual containers of 5-20 Kbytes each. Pipeline forwarding, over a meshed FλS network, requires that the delay between any two switching fabric inputs be an integer number of time frames, which is realized with an *alignment to CTR* operation before each switching fabric input. Since the time frame boundaries are explicitly identified, a relaxed CTR accuracy of less than one half of a time frame suffices.

Dynamic optical networking with FλS: (*i*) provides scalable switching with minimum complexity (i.e., Banyan network) – thereby solving the switching bottleneck, (*ii*) provides minimum complexity aggregation and grooming in

the time domain – thereby solving the link bottleneck at the edges of the network, and (*iii*) is compatible with current public standards, such as IP/MPLS and related protocols.

The efficient and deterministic bandwidth provisioning of FλS enables the optical core to be extended towards the edges of the network in the metro and enterprise, thus confining costly header processing to the low capacity periphery. The fractional λ pipes (FλPs) realized in FλS networks have the same deterministic characteristics as leased lines in SONET and circuit emulation in ATM. Consequently, FλS eliminates the need for SONET that, as was discussed, cannot be implemented in the optical domain.

REFERENCES

[1] L.V. Hau, S.E. Harris, Z. Dutton, and C.H. Behroozi, "Light speed reduction to 17 metres per second in an ultracold atomic gas," *Nature* 397: 594-598 (1999).

[2] C. Liu, Z. Dutton, C.H. Behroozi, and L.V. Hau, "Observation of coherent optical information storage in an atomic medium using halted light pulses," *Nature* 409: 490-493 (2001).

[3] "The ATM & IP Report," Vol. 7 No. 1, December 1999.

[4] "The ATM & IP Report," Vol. 7 No. 3, March 2000.

[5] National Institute of Standards and Technology (NIST), "GPS Data Archive," USA, `http://www.boulder.-nist.gov/timefreq/service/gpstrace.htm`

[6] Russian Federation Ministry of Defense - Coordination Scientific Information Center, "Global Navigation Satellite System – GLONASS," Russian Federation, `http://www.rssi.ru/SFCSIC/english.html`

[7] European Union, "Transport-Satellite Navigation," Union Policies, Brussels, Belgium, June 2001, http`://europa.eu.int/scadplus/leg/en/lvb/l24205.htm`

[8] National Physical Laboratory, "Two-Way Satellite Time and frequency Transfer (TWTFT)," Teddington, Middlesex, UK, `http://www.npl.co.uk/-npl/ctm/twstft.html`

[9] M. Baldi, Y. Ofek, "End-to-end Delay Analysis of Videoconferencing over Packet Switched Networks," IEEE/ACM Transactions on Networking, Vol. 8, No. 4, Aug. 2000, pp. 479-492.

[10] M. Baldi, Y. Ofek, B. Yener, "Adaptive Group Multicast with Time-Driven Priority," IEEE/ACM Transactions on Networking, Vol. 8, No.1, Feb. 2000, pp. 31-43.