Low Rate Service Multiplexing using FlexE for 400ZR Interfaces

OIF-400ZR-MUXING-01.0

September 18, 2020

White Paper

OIF

www.oiforum.com

The OIF is an international non profit organization with over 100 member companies, including the world's leading carriers and vendors. Being an industry group uniting representatives of the data and optical worlds, OIF's purpose is to accelerate the deployment of interoperable, cost-effective and robust optical internetworks and their associated technologies. Optical internetworks are data networks composed of routers and data switches interconnected by optical networking elements.

With the goal of promoting worldwide compatibility of optical internetworking products, the OIF actively supports and extends the work of national and international standards bodies. Working relationships or formal liaisons have been established with CFP-MSA, COBO, EA, ETSI NFV, IEEE 802.3, IETF, INCITS T11, ITU SG-15, MEF, ONF.

For additional information contact:
OIF
5177 Brandin Ct, Fremont, CA 94538
510-492-4040 F info@oiforum.com

www.oiforum.com

Working Group: **PLL Optical**

TITLE: **Low Rate Service Multiplexing using FlexE for 400ZR Interfaces**

SOURCE:

TECHNICAL EDITOR	WORKING GROUP CHAIR
Sebastien Gareau	David Ofelt
Ciena Corporation	Juniper Networks
385 Terry Fox Dr, Ottawa	1133 Innovation Way
Address 2:	Sunnyvale, CA 95033
Phone: 613-670-4539	+1-408-745-2945
Email: sgareau@ciena.com	Email: ofelt@juniper.net

ABSTRACT: White paper addresses 400ZR applications requiring multiplexing of lower rate Ethernet signals (e.g. 100GE) and how to leverage FlexE for this service multiplexing. Example profiles are provided to guide interoperability of common scenarios.

1.1 Table of Contents

1.1 Table of Contents 4
2 LIST OF FIGURES 5
3 LIST OF TABLES 5
4 DOCUMENT REVISION HISTORY 6
5 INTRODUCTION 7
5.1 Application Overview 7
5.2 Multiplexing Application Requirements 8
6 TECHNOLOGY OVERVIEW 9
6.1 400ZR Implementation Agreement Overview 9
6.2 FlexE Implementation Agreement Overview 10
7 ARCHITECTURE 12
7.1 FlexE+400ZR Architecture 12
7.2 Alarms, Defects and Consequent Actions 13
8 PROFILES 13
8.1 4x100GE->400ZR 14
8.2 2x200GE->400ZR 15
9 SUMMARY 17
10 REFERENCES 17
10.1 Normative references 17
10.2 Informative references 17
11 APPENDIX A: GLOSSARY 17
12 APPENDIX B: OPEN ISSUES / CURRENT WORK ITEMS 18

2 List of Figures

Figure 1– Transport Box Application Example 7
Figure 2– Integrated Optics Application Example 8
Figure 3 – 400ZR IA Reference Application Diagram 9
Figure 4 – 400ZR Frame FlexE Payload 10
Figure 5 – General FlexE structure from IA 11
Figure 6 – FlexE overhead and payload structure from IA 11
Figure 7 – 400GBASE-R Mux Functions from IA 12
Figure 8 – FlexE+400ZR Functional Datapath 13
Figure 9 – FlexE calendar slots 14
Figure 10 – 4x100GE/400ZR Profile 15
Figure 11 – 2x200GE/400ZR Profile 16

3 List of Tables

Table 1 - 4x100GE/400ZR Profile 14
Table 2 - 2x200GE/400ZR Profile 16

4 Document Revision History

Working Group: PLL Protocol

SOURCE:	Editor(s)'s Name	Working Group Chair
	Sebastien Gareau	David Ofelt
	Ciena Corporation	Juniper Networks
	385 Terry Fox Dr	1133 Innovation Way
	Ottawa, Ontario	Sunnyvale, CA
	K2K 0L1	95033
	Phone: 613-670-4539	Phone: +1-408-745-2945
	Email: sgareau@ciena.com	Email: ofelt@juniper.net

DATE OF APPROVAL: September 18, 2020

---

5 Introduction

The current scope of the 400ZR Implementation Agreement includes mapping of 400G Ethernet signals, as defined in [IEEE 802.3]. While the 400G Ethernet standard has been published and available for a few years, availability of 400GE interfaces in the market is based on switch silicon and switch/router products. Lower rate services, such as 100G Ethernet, are quite prevalent and still used in huge volumes in Data Centers. As with any new technology introduction, the transition to 400GE clients/services will be gradual. Network operators are seeing power and density benefits of new 400ZR based interfaces/modules, and see value in deploying such interfaces with Ethernet service rates other than 400G. This white paper focuses specifically on FlexE multiplexing of lower rates service, such as 100G and 200G Ethernet. The technology explored in this white paper is not limited, and could also address other rates (e.g. 10G Ethernet) or applications (e.g. bonding) beyond the scope of this white paper.

5.1 Application Overview

Various application examples are considered for multiplexing low-rate services onto a 400ZR line interface. The first set of applications depicted in Figure 1 below, are more akin to traditional transport applications. Existing or legacy 100G Ethernet router/switch ports, typically using QSFP28 optical pluggable modules, can interface to a transport compact DCI box hosting a 400ZR module and interface. A multiplexing function shown as green in the figure can be implemented in various locations:

- Inside an enhanced 400ZR type module
- Inside a device (gearbox, retimer, FEC translation, FPGA) that sits in front of 400ZR module.

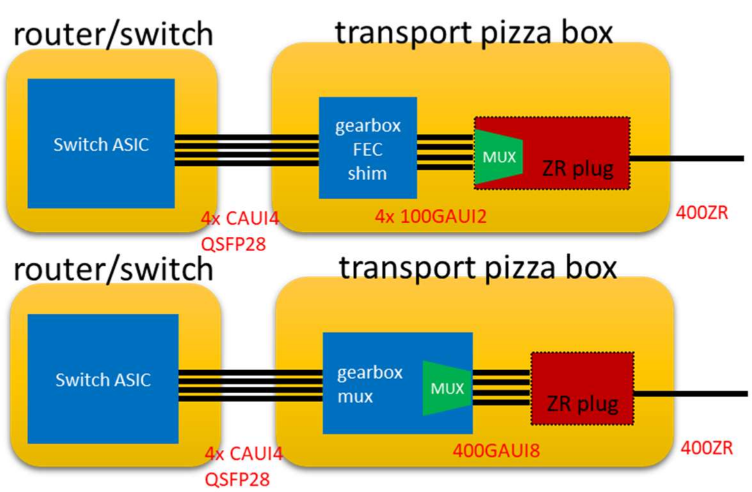

Figure 1– Transport Box Application Example

A second set of applications depicted below in Figure 2 are based on integrated optics, also referred to as IP over DWDM. The optical coherent modules are plugged directly into the Ethernet switch or router. A multiplexing function shown as green in the figure can be implemented in various locations.

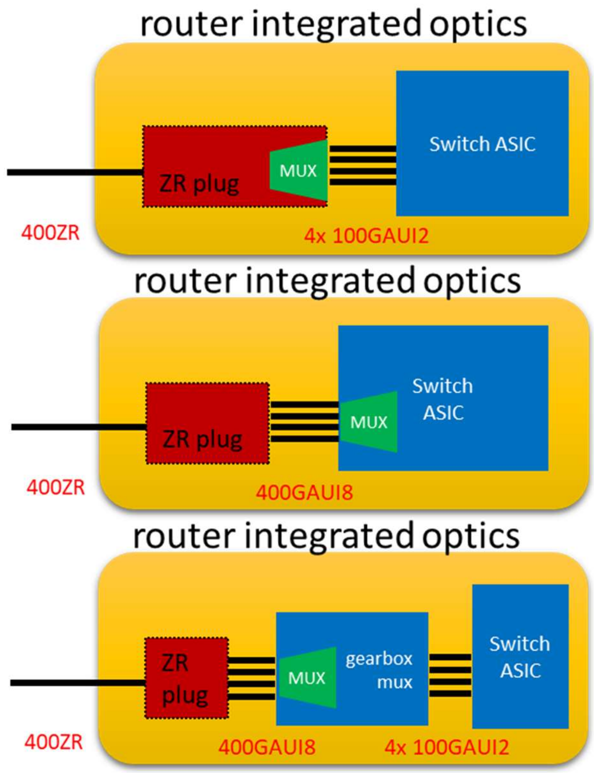

Figure 2– Integrated Optics Application Example

This section is not meant to provide all exhaustive possibilities of applications and implementations of low-rate Ethernet multiplexing. All applications are interoperable on the line side 400ZR interface.

5.2 Multiplexing Application Requirements

The multiplexing scheme described in this white paper is predominately addressed by DCI (Data Center Interconnect) applications, and shares similar requirements as discussed in other OIF projects such as 400ZR and FlexE Implementation Agreements.

- The multiplexing scheme should not increase the interface rate.
- Only Ethernet clients are considered. Other multi-service clients and rates are outside the scope of consideration.
- Timing transparency for Synchronous Ethernet or [IEEE 1588] Precise Timing Protocol is not a requirement.
- The scheme should be compatible with existing and future Ethernet PMDs that IEEE can develop based on OIF Implementation Agreements.

6 Technology Overview

6.1 400ZR Implementation Agreement Overview

The 400ZR IA [OIF 400ZR] enables interoperable, cost-effective, 400Gb/s implementations based on single-carrier coherent DP-16QAM modulation, low power DSP supporting absolute (Non-Differential) phase encoding/decoding, and a Concatenated FEC (C-FEC) with a post-FEC error floor <1.0E-15. A figure below taken from the 400ZR IA illustrates the application and interfaces.

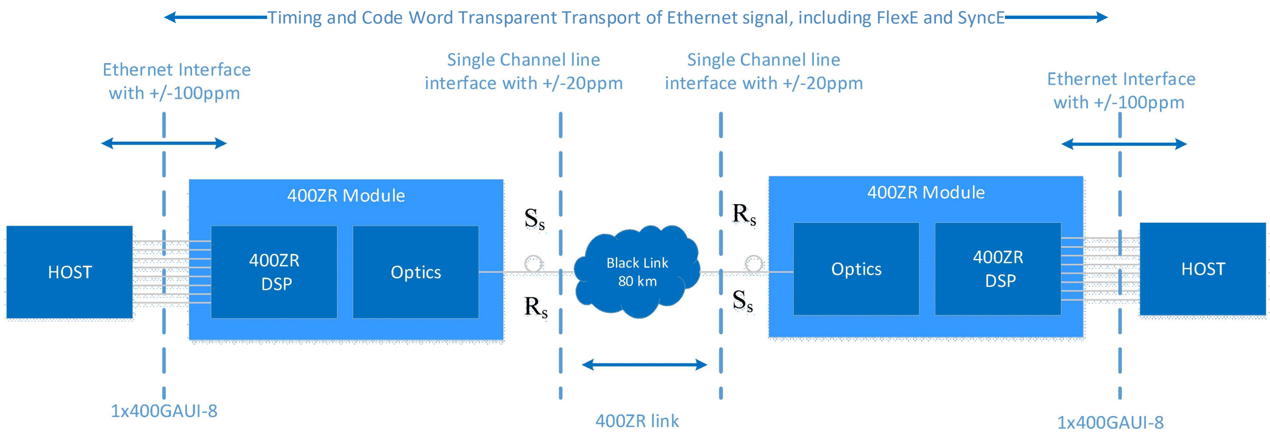

Figure 3 – 400ZR IA Reference Application Diagram

The only client currently specified in the IA is 400G Ethernet, however as illustrated in the reference figure, provisions have been made to enable FlexE as well. The 400ZR mapping process is asynchronous, preserves the client signal timing and based on a simplified Generic Mapping Procedure (GMP) [ITU-T G.709]. The GMP process is used to rate-adapt the client signal (400GE) to a 400ZR frame structure which is based on a local clock reference with tolerance of ± 20ppm. The mapping processes defined in the 400ZR IA are not packet aware and support any signal that maintains a legal 400G Ethernet PCS structure [IEEE 802.3] clause 119.

The 400ZR interface itself may operate and behave as a 400GBASE-R Ethernet PHY. The current ZR frame format does not segment the payload and does not support the concept of multiplexed structures in its overhead. A 400ZR frame maps 1:1 a 400G Ethernet client signal. This white paper complements the 400ZR IA by using FlexE to provide a multiplexed/channelized structure (payload and control overhead) inside the payload area of a 400ZR frame.

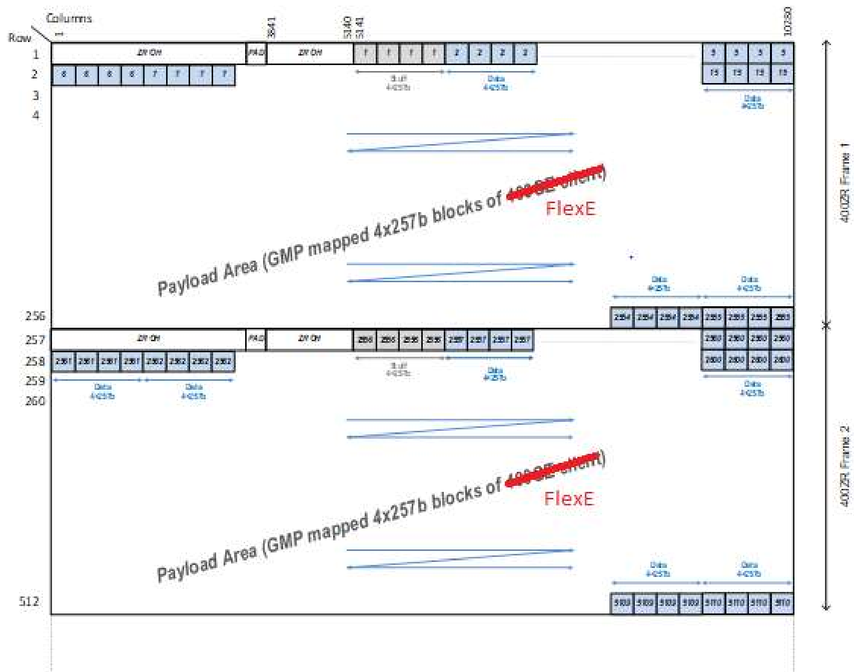

Figure 4 – 400ZR Frame FlexE Payload

The host interface to a 400ZR pluggable module cannot distinguish between a 400G Ethernet client and a FlexE client. As such, no extensions to the 400ZR IA are needed to identify the payload type.

6.2 FlexE Implementation Agreement Overview

The Flexible Ethernet IA [OIF FlexE] provides a generic mechanism for supporting a variety of Ethernet MAC rates that may or may not correspond to any existing Ethernet PHY rate. This includes MAC rates that are both greater than (through bonding) and less than (through sub-rate and channelization) the Ethernet PHY rates used to carry FlexE. While FlexE can enable a variety of applications and use cases, this white paper will focus on multiplexing applications where FlexE is used to map 4x100G or 2x200G Ethernet/FlexE clients into a 400G group, consisting of a single 400G PHY in the group. In this instance the 400G PHY is a 400ZR interface, which acts like a 400GBASE-R PHY. Nothing prevents FlexE from mapping other sized clients (e.g. 10GE) or mapping into larger sized groups (e.g. bonding), however these applications are being the scope of this white paper. More details on the specific client mappings are found in the profiles section.

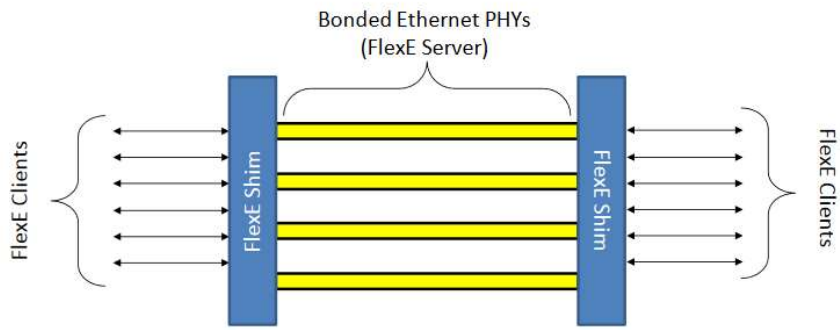

Figure 5 – General FlexE structure from IA

FlexE clients are mapped into a FlexE group, through a FlexE shim. The FlexE shim is a 66b block-based structure with overhead and a payload area that is segmented into calendar slots. The FlexE group in the application described in this white paper compromises four 100G Instances, each of which has 20 calendar slots and can support up to 10 clients.

FlexE overhead is located in a periodically placed 66b ordered set. It contains OAM and multiplex structure information for the application. It is completely dissassociated from any other 400ZR or Ethernet PHY related overhead.

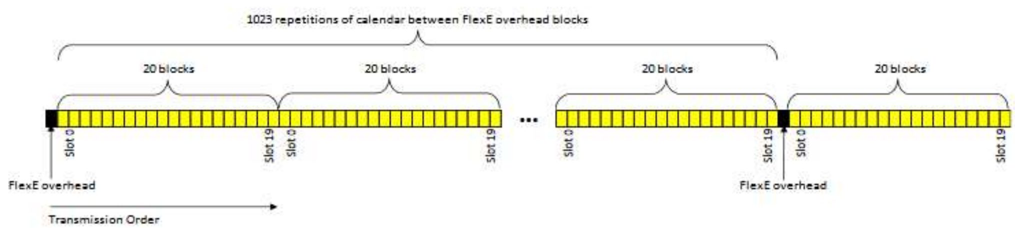

Figure 6 – FlexE overhead and payload structure from IA

FlexE IA revision 2.0 added the support of 400GBASE-R PHY, which are fully compatible to 400ZR mapping processes as explained in previous sections. The FlexE shim essentially performs the multiplexing processes required to support 4x100G and 2x200G Ethernet applications over 400ZR coherent optics. The FlexE mapping procedures are asynchronous, and rate adapt the clients with idle insertion/removal. The Ethernet clients mapped using FlexE are not timing transparent.

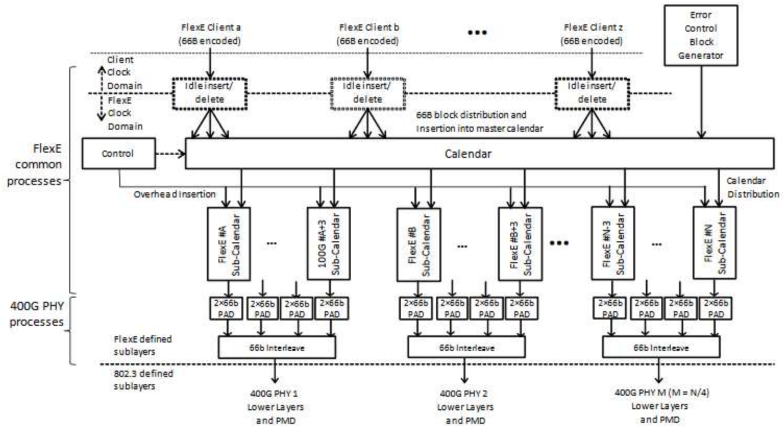

Figure 7 – 400GBASE-R Mux Functions from IA

7 Architecture

7.1 FlexE+400ZR Architecture

FlexE and 400ZR are complementary technologies, where FlexE provides the multiplexing and 400ZR provides an interface behaving like a standard Ethernet PHY. FlexE is designed to work with any type of Ethernet PHY at 50G, 100G, 200G and 400G. The 400ZR IA contains a detailed datapath diagram which shows the full signal flow through different mapping stages. The 400ZR host interface expects a fully standard 400GBASE-R PMA sublayer, which FlexE can provide. The functional block diagram can be augmented with FlexE processes as shown in Figure x below—some details were abstracted in this figure, but can be found in FlexE and 400ZR IAs. Some implementations might choose to skip intermediate functions (for example 400GE FEC encode/decode) when co-locating FlexE and 400ZR functional logic.

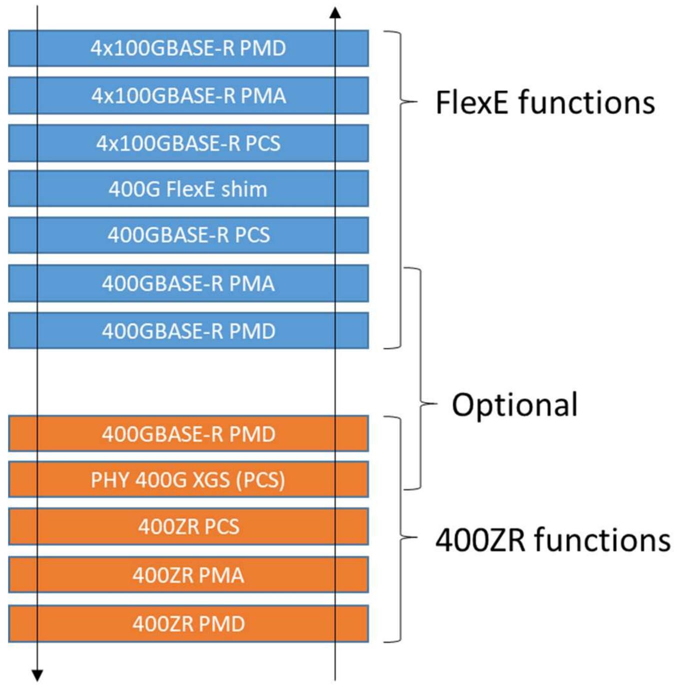

Figure 8 – FlexE+400ZR Functional Datapath

7.2 Alarms, Defects and Consequent Actions

Alarms, defects and consequent actions are specified in their respective Implementation Agreements—there are no news alarms or defects defined for the multiplexing application described in this white paper. When using FlexE over a 400ZR optical interface, one can expect the same behavior as if running with a standard Ethernet PHY as defined in [IEEE 802.3]. FlexE equipment specifications are additionally captured in [ITU G.8023].

When misconnecting a standard 400GE/400ZR interface with a FlexE/400ZR interface, the 400ZR interface will be clean of alarms and defects. However, the FlexE will alarm the misconnection (e.g. loss of FlexE frame). Is it expected that a switch/router 400GE port will also alarm the misconnection and detect the problem with the block sequence based on the state machine defined in figure 119-5 in [IEEE 802.3].

When experiencing a line fault, the line 400ZR interface will forward LF to the FlexE shim, which will then alarm the appropriate loss of FlexE frame.

8 Profiles

The FlexE mechanism operates using a calendar which assigns 66B block positions on sub-calendars on each FlexE Instance of the FlexE Group to each of the FlexE Clients. As specified in the FlexE-02.0 IA [OIF

FlexE], two calendar configurations are supported. For the purpose of this white paper, the FlexE group will use a static calendar configuration A for mapping and demapping from the FlexE group. Calendar B can be optionally provisioned with identical configurations as A.

8.1 4x100GE->400ZR

In the case of mapping 4x100GE FlexE clients into a 400G FlexE group, each group will occupy twenty calendar slots as show in Figure 10 taken from the FlexE-02.0 IA [OIF FlexE].

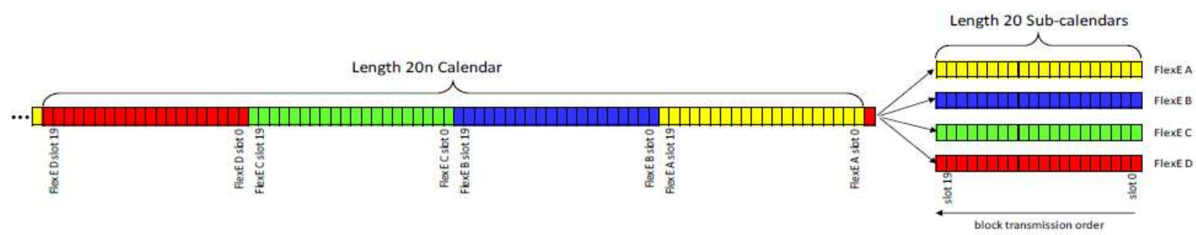

Figure 9 – FlexE calendar slots

Each 100GE client will map to fixed contiguous calendar slots as follows and as illustrated above. Each instance will have a fixed PID, 1-4. The first 100G will always occupy slots 0-19, follow by the next 100G, and so on. Table 1 illustrates the intended configuration.

<table>
<tr><th>Client</th><th>100G Flex Instance</th><th>Calendar Slot</th><th>PID</th><th>GID</th></tr>
<tr><td>100GE a</td><td>1</td><td>0-19</td><td>1</td><td rowspan="4">1</td></tr>
<tr><td>100GE b</td><td>2</td><td>20-39</td><td>2</td></tr>
<tr><td>100GE c</td><td>3</td><td>40-59</td><td>3</td></tr>
<tr><td>100GE d</td><td>4</td><td>60-79</td><td>4</td></tr>
<tr><td colspan="5">Note – GID can be user provisioned.</td></tr>
</table>

Table 1 - 4x100GE/400ZR Profile

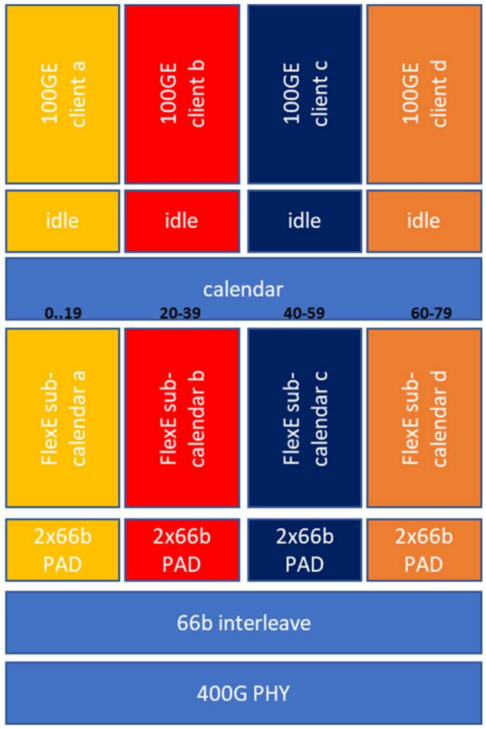

Figure 10 – 4x100GE/400ZR Profile

The use of management channels is optional for this application. The sync messaging channel is unused.

8.2 2x200GE->400ZR

In the case of 2x200GE mapped into a 400G FlexE group. Each 200GE client will occupy 2x 100G FlexE Instances in a fixed configuration as follows:

Each 200GE client will map to fixed contiguous calendar slots as follows and as illustrated if figure x. Each instance will have a fixed PID, 1-4. The first 100G will always occupy slots 0-19, follow by the next 100G, and so on.

Client	100G Flex Instance	Calendar Slot	PID	GID
200GE a	1	0-19	1	1
	2	20-39	2	
200GE b	3	40-59	3	
	4	60-79	4	
Note – GID can be user provisioned.				

Table 2 - 2x200GE/400ZR Profile

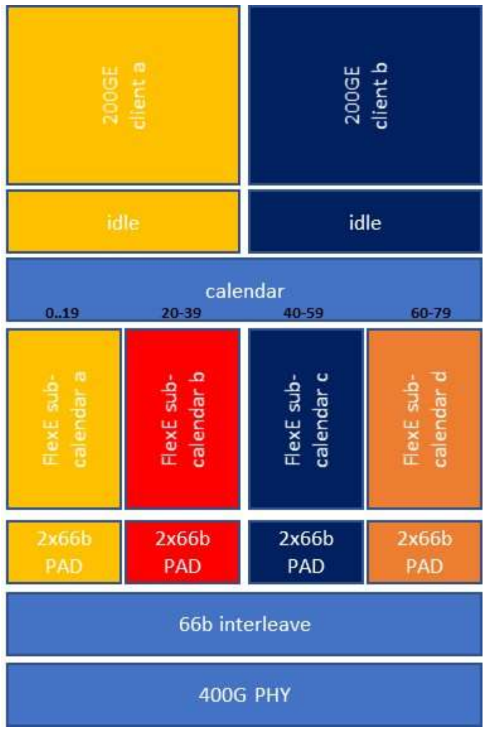

Figure 11 – 2x200GE/400ZR Profile

The use of management channels is optional for this application. The sync messaging channel is unused.

9 Summary

This white paper addresses 400ZR applications requiring multiplexing of lower rate Ethernet signals (e.g. 100GE) and describes how to leverage FlexE for this service multiplexing. Example profiles are provided to guide interoperability of common scenarios.

10 References

10.1 Normative references

[IEEE 802.3] IEEE Standard for Information Technology – Telecommunications and Information Exchange Between Systems – Local and Metropolitan Area Networks – Specific Requirements Part 3: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications.

[IEEE 1588] IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems.

[ITU-T G.709] Recommendation ITU-T G.709/Y.1331 (2019), Interfaces for the optical transport network.

[ITU-T G.709.3] Recommendation ITU-T G.709.3/Y.1331.1 (2019), Flexible long-reach interfaces.

[ITU-T G.8023] Recommendation ITU-T G.8023 (2018), Characteristics of equipment functional blocks supporting Ethernet physical layer and Flex Ethernet interfaces.

[OIF 400ZR] Optical Interworking Forum, OIF (2020), 400ZR Implementation Agreement 1.0.

[OIF FlexE] Optical Interworking Forum, OIF (2019), Flex Ethernet Implementation Agreement 2.0.

10.2 Informative references

11 Appendix A: Glossary

ASIC	application specific integrated circuit
DCI	data center interconnect
DWDM	dense wavelength division multiplexing
FEC	forward error correction
FPGA	field-programmable gate array
GMP	generic mapping procedure
IP	internet protocol
LF	local fault
MUX	multiplexing

PHY	physical layer
PMA	physical medium attachment sublayer
PMD	physical medium dependent sublayer
QAM	quadrature amplitude modulation
QSFP	quad small form-factor pluggable optics transceiver

12 Appendix B: Open Issues / current work items