



# **Migrating to an Efficient Ethernet Transport Solution Based on OTN**

**White Paper**



**19 February 2008**

**Document Number 308951**

**Revision 2.1**



**INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH CORTINA SYSTEMS® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT.**

**EXCEPT AS PROVIDED IN CORTINA'S TERMS AND CONDITIONS OF SALE OF SUCH PRODUCTS, CORTINA ASSUMES NO LIABILITY WHATSOEVER, AND CORTINA DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO THE SALE AND/OR USE OF CORTINA PRODUCTS, INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.**

Cortina products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Cortina Systems® and the Cortina Systems logo are the trademarks or registered trademarks of Cortina Systems, Inc. and its subsidiaries in the U.S. and other countries. Other names and brands may be claimed as the property of others.

Copyright © 2005–2008 Cortina Systems, Inc. All rights reserved.

## Contents

<b>1.0</b>	<b>Abstract</b> .....	<b>5</b>
<b>2.0</b>	<b>Introduction</b> .....	<b>6</b>
<b>3.0</b>	<b>Current Standardized Options for Packet over OTN</b> .....	<b>7</b>
<b>4.0</b>	<b>A New Proposal for High-Rate Packet Transport</b> .....	<b>7</b>
4.1	A Smooth Migration Path.....	8
<b>5.0</b>	<b>Rate Adaptation Overview</b> .....	<b>10</b>
5.1	Standard Rate Adaptation Mode.....	10
5.2	Proprietary Rate Adaptation Mode.....	10
5.3	PCS Block Handling.....	10
5.4	Flexible Application Support.....	11
<b>6.0</b>	<b>Implementation Considerations</b> .....	<b>12</b>
<b>7.0</b>	<b>Summary</b> .....	<b>13</b>
<b>8.0</b>	<b>Key Terminology</b> .....	<b>13</b>

## Tables

1	Packet Mapping Procedures Defined for SDH/OTN Transport Layer.....	7
2	Proposed Mapping Configurations.....	8

## Figures

1	Integration Options for the 10 GE Rate Adaptation Function.....	9
2	Idle Removal Algorithm.....	10
3	Unknown PCS Block Processing.....	11
4	Proposed Implementation Partitioning.....	12

## Revision History

<b>Revision 2.1</b> <b>Revision Date: 19 February 2008</b>	
New Corporate Logo.	

<b>Revision 2.0</b> <b>Revision Date: 26 December 2006</b>	
<b>Page #</b>	<b>Description</b>
N/A	First release of this document from Cortina Systems, Inc.

---

**Revision 001**  
**Revision Date: 01 June 2004**

<b>Page #</b>	<b>Description</b>
N/A	Initial version

---

## 1.0 Abstract

Demand for transparent transport of Ethernet client traffic over optical transport lines is growing rapidly, driven by continued expansion of enterprise LAN systems. Existing data client mapping schemes are not sufficient to effectively address the requirements imposed by the growing percentage of Ethernet clients on the Optical Transport Network (OTN) infrastructure.

This paper proposes a method for mapping a 10 Gigabit Ethernet LAN signal into a true OPU2 (ITU-T G.709) payload container. The adapted client signal retains its 10GBASE-R characteristics, except for the adapted data rate. Thus, the proposed mapping serves as an effective and cost efficient migration path to an OTN-based Ethernet transport implementation with significant advantages over existing solutions.

---

## 2.0 Introduction

In the past, most telecommunication/broadcasting services successfully converged on unified packet/IP based communication protocols. However, due to their LAN heritage, these protocols lacked many management capabilities typical to public networks. For example, end-to-end monitoring, bandwidth management, and protection switching capabilities had to be added. While new technologies like Multi Protocol Label Switching (MPLS) and Resilient Packet Rings (RPR) were developed to focus on issues close to the service interface (for example, bandwidth management and QoS), there is still no consistent solution for packet protocols that meets the more stringent requirements of the transport core (for example, carrier's carrier environments and network resilience).

Employing OTN (ITU-T recommendation G.709) as the transport layer is a compelling option because it precisely addresses core transport-related Operations Administration Maintenance & Provisioning (OAM&P) tasks, while keeping all other service related management functions (for example, bandwidth management, aggregation, and end-to-end monitoring) at the level of packet technologies mentioned above. OTN provides a cost efficient transport layer that supports Dense Wavelength Division Multiplexing (DWDM) technologies while simultaneously leveraging many SDH/SONET concepts. This allows OTN to serve as a converged transport layer for new packet-centric and Time Division Multiplexing (TDM)-based legacy networks.

The OTN payload rates were defined to match the Synchronous Digital Hierarchy (SDH) signals STM-16 to STM-256, assuming any required rate matching for packet traffic clients would be handled by appropriate pre-processing. The market, however, took a somewhat different turn.

Volume-driven per-port costs and other market realities favored the 10 Gigabit Ethernet LAN PHY signal, designated as 10GBASE-R. Specified by IEEE 802.3\* for LAN applications, 10GBASE-R traffic is now a large and growing WAN transport client. Unfortunately, the 10.3125 Gbps 10GBASE-R signal does not fit into the standard 9.995 Gbps OPU2 payload container.

A straightforward solution to provide 10GBASE-R service interfaces is to simply increase the OPU2 container rate appropriately. This approach neither affects the management layer, nor requires new silicon components because it can be easily realized by slightly overclocking the existing OTN mapping devices and PMA components. However, because the resulting OPU2-like payload runs at 10.3125 Gbps it does not support convergence with TDM-based clients, nor does it provide any convincing argument for switching and aggregation at the OTN layer. These limitations are acceptable for point-to-point transport infrastructure, where this approach is already deployed today. However, they pose a serious obstacle to interfacing with OTN network equipment that supports electrical multiplexing (OPU1/2/3 - 2.5/10/40 Gbps).

The alternative approach proposed here retains the benefits of a standard OTN signal such as frame structure and line speed, while opening a path to a cost efficient mapping solution for 10 GE LAN client signals.

### 3.0 Current Standardized Options for Packet over OTN

As OTN was defined “on top” of the SDH/SONET stack, options for mapping packet traffic over SDH/SONET networks also apply to OTN-based transport networks. For packet bandwidths in the range of 2.5 Gbps to 10 Gbps, most methods allow replacing SDH containers (VC-4-Xc) with OTN containers (OPUk and OPUk-Xv). With several well-defined SDH options available, new OTN-specific schemes were not required or developed. (In principle, G.709 also proposes direct mapping of ATM into OPUk-Xv, but no such implementation is known to the author.)

Popular mapping solutions are summarized in Table 1 (SDH notation used for simplicity only). POS-based implementations (option 1) suffer from missing L2 support. Options 2 and 3 do not support expansion to 10 Gbps. Assuming a public network implements all multiplexing/aggregation and service related functions on the packet layer (L2 and L3), only options 4 and 5 are of long term interest. Network interfaces based on GFP-F mapping (option 4) and on a 10GBASE-W interface (option 5) suffer from limited availability and increased costs compared to those based on a 10GBASE-R interface.

The limitations of these formally specified mapping options can be overcome by extending use of the 10 Gbps 10GBASE-R signal from LAN to WAN applications. This new option also provides additional benefits. For example, because 10GBASE-R is bit-transparent at the PCS layer, the mapping scheme uses the same PCS as deployed in the LAN. This allows future L2 extensions (such as Ethernet packet preamble usage) to be efficiently ported from LAN applications to the WAN.

**Table 1 Packet Mapping Procedures Defined for SDH/OTN Transport Layer**

Option	Packet Client		Mapping Procedure	Predominant Container	Transparency	Notes
	Type	Rate				
1	IP	Up to 10 Gbps	POS	VC-4-Xc VC-12-Xc	L3	L2 (Ethernet) terminated.
2	IP / Ethernet	Up to 100 Mbps	AAL5/ATM (RFC1483)	VC-4-Xc VC-12-Xc OPUk-Xv	L2/L3	Very complex solution. Expansion to 10 Gbps not practical.
3	Ethernet	Up to 1Gbps	GFP-T	VC-4-Xc VC-12-Xc	L2/L3	More than L2 transparency, for 8b/10b encoded signals. Rather complex. Does not support 64b/66b encoded signals such as 10 GE.
4	IP / Ethernet	Up to 10 Gbps	GFP-F	VC-4-Xc VC-12-Xc OPUk-Xv	L2/L3	L2 transparency does not support proprietary L2 extensions. Complex implementation.
5	Ethernet	10 Gbps	WIS 10GBASE-W	VC-4-64c	(L1)/L2/L3	As the VC-4-64c container carries a down paced “LAN” type PCS signal, it could be viewed as L1 transparent. Relatively easy and cost efficient to implement.

### 4.0 A New Proposal for High-Rate Packet Transport

Assuming 2.5 Gbps to 40 Gbps will be an attractive interface range between core transport networks, and also assuming L2/L3 packet-based service processing elements such as routers are not substantially re-defined, it makes sense to develop a cost efficient and scalable L2 mapping solution for Ethernet clients.

By supporting a clear decoupling between the packet layer and the transport layer, the solution would meet several key requirements:

- Allow proprietary add-ons on the packet layer without affecting the transport layer
- Allow future Ethernet protocol extensions to be smoothly leveraged from the LAN domain without touching the transport infrastructure
- Allow packet processing equipment to be physically separated from network termination equipment (NTE), by providing a client format that can be carried over a short serial link
- Allow aggregation, concatenation, and multiplexing on the electrical layer by maintaining the WAN data rate and protocol integrity, avoiding higher costs of performing these functions in the optical layer

All these requirements could be met by deploying a standard OTN transport platform that provides OPUk / OPUk-Xv payload containers to carry the packet-client signal in a bit-transparent manner.

To keep costs low and to provide upwards compatibility with future developments in the LAN space, physical encoding and packet delineation should be identical to LAN applications. Thus, the standard 66b/64b encoding as described by IEEE 802.3ae-2002, clause 49, is proposed as the PCS layer.

All mappings and the underlying OTN network should support asynchronous client mapping to maximize clocking options. The proposed rates are listed in [Table 2](#).

**Table 2 Proposed Mapping Configurations**

MAC Rate <sup>1</sup>	PCS Rate <sup>2</sup>	Transport Container
2.42129 Gbps	2.48832 Gbps	OPU1
X*2.42129 GBit/s (1<X<256) <sup>3</sup>	X*2.48832 GBit/s (1<X<256)	OPU1-Xv
9.6932 Gbps	9.995276 Gbps	OPU2
X*9.6932 GBit/s (1<X<256)	X*9.995276 GBit/s (1<X<256)	OPU2-Xv
38.933 Gbps	40.1505 Gbps	OPU3
X*38.933 GBit/s (1<X<256)	X*40.1505 GBit/s (1<X<256)	OPU3-Xv
1. Mac-to-PCS or XGMII rate (± 20 ppm) 2. PCS-to-PMA or XSBI rate (± 20 ppm) 3. Unlike with VC-4 containers, more than 4 members per group are highly unlikely here for practical reasons		

The unshaded options in [Table 2](#) are relatively easy to implement, for example, by a standard LAN PHY/MAC clocked at the corresponding speed in combination with an off-the-shelf OTN framer. The shaded options, however, are significantly more complex. If the system is partitioned into separate OTN NTE and packet processing elements, either their interface would have to run at a variable rate, or an oversubscription approach could be considered, where their interface runs at maximum rate and the OTN NTE performs the adaptation to the lower rate concatenate group.

## 4.1 A Smooth Migration Path

Currently, the most compelling data rate for a migration to the proposed mapping is 10 Gbps. Lower rates are either carried fairly efficiently within the SDH stack (for example, using GFP in a VC-4-Xv) or aggregated into a 10 Gbps signal by a packet processing

element (i.e. router/switch) before transport. 40 Gbps solutions are currently discussed as the transport option, but no effort to support such a rate for LAN applications is known to the author.

Assuming the predominant 10 Gbps Ethernet interface is 10GBASE-R, an efficient solution must be found that adapts a 10GBASE-R client into the proposed signal. Further, this adaptation function should implement maximum transparency to support compliance and adaptability of equipment when deployed in a future network running on the proposed mapping.

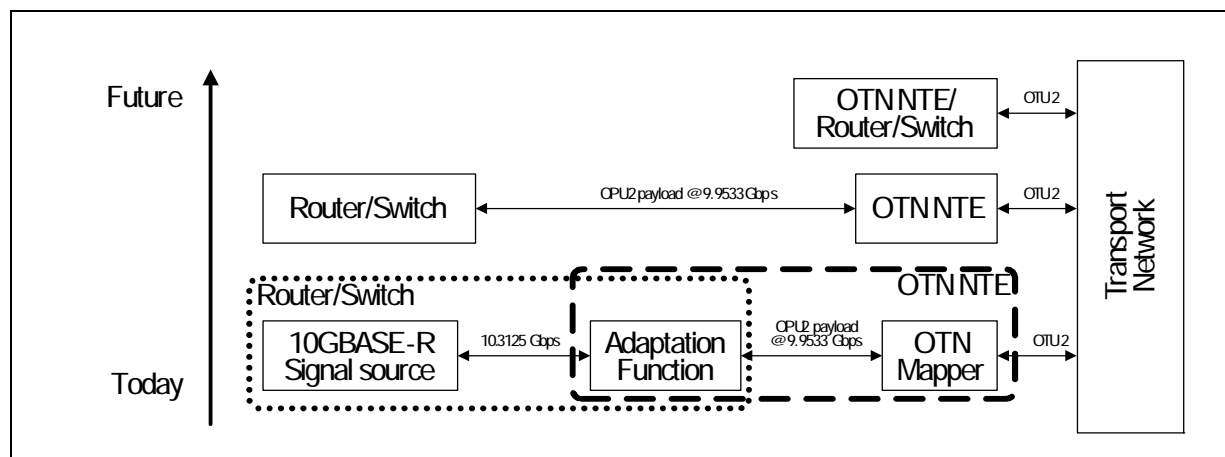
For initial development and short term solutions, the function can be implemented in a medium-size FPGA that is largely autonomous and independent of any adjacent management layers. The implementation does not require any costly additional components when implemented as part of an OTN NTE client port. A longer term migration path is shown in Figure 1.

Today, both OTN NTE and router/switch equipment supporting 10 GE are widely available. The adaptation function between might be added to both types of equipment. However, practical considerations such as cheaper router/NTE interfaces and the current commercial model of operating such transport networks make it highly likely the adaptation will take place in the OTN NTE. This partitioning also makes it easier to implement and manage 10GBASE-R into OPU1 or OPU1-Xv oversubscription.

First generation equipment applying the proposed mapping will likely retain the current router/NTE partitioning, as it requires only slight modifications to the clocking and interface configuration of such elements. One drawback of this solution is it does not efficiently support flexible oversubscription as this would require a variable rate interface between the devices. Currently, there is little interest in merging transport and router/switch management functions into one unified layer.

Eventually, the advent of automatic switching technologies like ASTN/ASON will drive development of a single equipment type that merges these functions and significantly reduces hardware costs.

**Figure 1 Integration Options for the 10 GE Rate Adaptation Function**



## 5.0 Rate Adaptation Overview

The proposed adaptation function will operate on unscrambled 66-bit PCS blocks (refer to IEEE 802.3ae-2002, clause 49) to allow maximum transparency on both the MAC and XGMII layers. Two rate adaptation options, standard and proprietary, provide the flexibility for inter-operability and product differentiation. For example, this approach supports proprietary use of ordered sets and/or preamble bytes.

### 5.1 Standard Rate Adaptation Mode

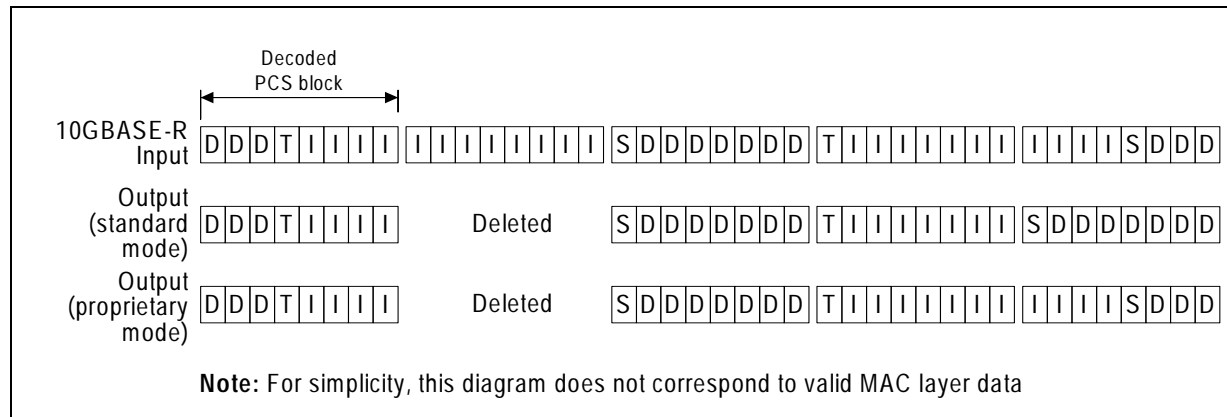
**Standard mode** provides efficient support for applications with IEEE 802.3ae traffic patterns. Idle sequences are removed or added according to the standard as shown in Figure 2. Either a complete PCS transmit/receive block containing 8 idle symbols is removed, or half a transmit/receive block (4 symbols) is removed and the subsequent XGMII symbols reassigned to new PCS blocks.

### 5.2 Proprietary Rate Adaptation Mode

**The Proprietary mode** delivers additional transparency at the expense of reduced error robustness and an inferior rate adaptation algorithm. Proprietary mode rate adaptation is also performed by idle symbol removal/insertion (see Figure 2). However, in proprietary mode, only quantities of 8 Idle symbols are removed to avoid breaking up following PCS transmit/receive blocks of potentially proprietary content. As the probability of 8 consecutive idle symbols decreases with increasing packet traffic, this mode will have less room for rate adaptation, (i.e. smaller margins) and may lead to larger FIFO sizes.

Depending on the implementation, rate adaptation could include flow control by generation of pause packets towards the 10GBASE-R interface. This is discussed in more detail under Section 6.0, *Implementation Considerations*, on page 12.

**Figure 2 Idle Removal Algorithm**



### 5.3 PCS Block Handling

The standard and proprietary modes differ in how unknown PCS block types are processed (see Figure 3 on page 11). In standard mode, invalid blocks are converted into valid blocks containing 8 error symbols. In proprietary mode, unknown PCS blocks are passed through

unmodified. This delivers maximum transparency, however error robustness is reduced as damaged PCS blocks are not detected at the point of occurrence, but at the final receiving 10GBASE-R node.

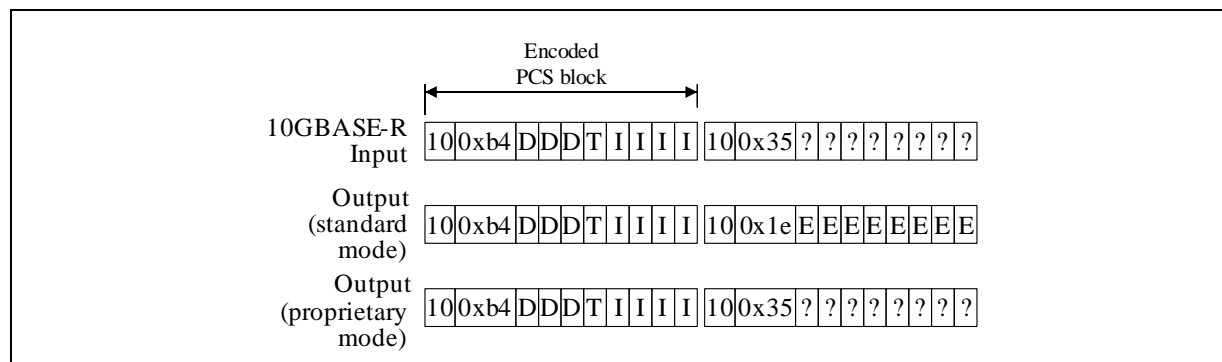
## 5.4 Flexible Application Support

The proposed procedure is not limited to OPU2 transport containers. It can be applied to any rate listed in [Table 2, Proposed Mapping Configurations, on page 8](#). Hence, it provides flexibility for future enhancements towards lower and higher speed classes.

The proposed mapping can be summarized as follows:

- Scrambling and 66b/64b PCS layer encoding are performed as in a standard 10 GE LAN port
- Idle symbols are encoded as defined in IEEE 802.3ae-2002, as part of standard-compliant PCS blocks
- To avoid data corruption, the packet rate on the 10BASE-R interface must be limited to match the maximum OTN payload capacity. Typically, this is 97% of a 10 GE LAN capacity in case of OPU2 mapping. This can be controlled either by the network management layer, by configuration of the interpacket gap (IPG) stretch parameter on the MAC layer, or by applying pause frame-based flow control, as outlined below.

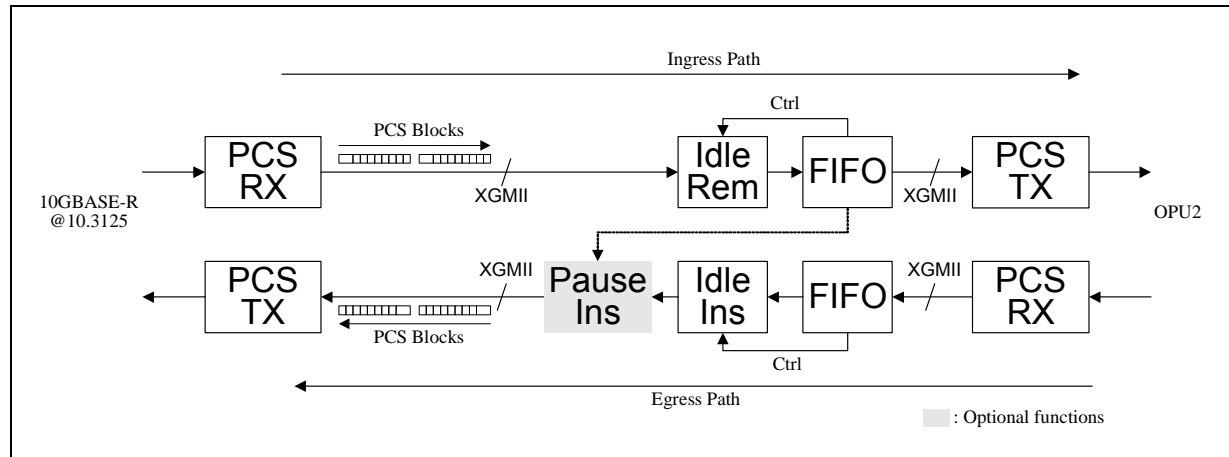
**Figure 3 Unknown PCS Block Processing**



## 6.0 Implementation Considerations

As both interfaces of the adaptation function run on the same data format, a highly symmetrical implementation can be used for the OTN ingress and egress path (see Figure 4). Both ports are equipped with a PCS RX/TX block that delineates/serializes, (de-)scrambles, and optionally decodes/encodes the data stream. Decoded PCS blocks are passed from/to the inside through an interface resembling the XGMII as defined in IEEE 802.3ae-2002, clause 46.

**Figure 4 Proposed Implementation Partitioning**



Rate adaptation is performed by idle removal/insertion as described in Section 5.0, *Rate Adaptation Overview*, on page 10. Typically, this process is controlled by the fill level of the adjacent data FIFOs. The size  $S_F$  of the FIFO depends on the difference of the two port rates ( $f_{10GBASE-R}$  and  $f_{OTN}$ ), the maximum transmission unit (MTU) size, and a margin  $S_M$ . This margin is required to average out the packet bandwidth distribution on the 10GBASE-R input uniformity in the event of high traffic load:

**Equation 1**

$$S_F = \frac{MTU(f_{10GBASE-R} - f_{OTN})}{f_{10GBASE-R}} + S_M$$

Determination of  $S_M$  is not covered in this document as it requires consideration of many system and network level parameters.

For applications where it is not possible to limit the link utilization through port configuration, the 10GBASE-R interface must be paced down by inserting pause packets (IEEE 802.3ae-2002, Annex 31b) into the stream towards this port. Here, two strategies might be deployed. First a classical closed loop control can be implemented to feed back the ingress path FIFO fill state to the 10GBASE-R transmitting MAC. This option has the drawback that with increased loop latency (due to the physical distance between the NTE and the connected router/switch) FIFO size must increase and thus overall latency increases.

To avoid this, pause frames may also be inserted at a periodic rate, directly reflecting the frequency offset ( $f_{10GBASE-R} - f_{OTN}$ ) which is a constant that is known and traffic pattern independent. As this is an open loop control, loop latency does not impact the system performance and the available FIFO size.

## 7.0 Summary

OTN offers an efficient transport technology for packet-centric networks, but currently lacks a unified mapping definition that matches packet client rates with transport container capacities. The author proposes a simple, yet highly transparent rate adaptation function that covers a wide range of transport capacities using standard PCS encoding as defined by IEEE 802.3ae-2002, and a relatively simple modification to existing 10GBASE-R (10GE LAN) ports. The realistic and achievable migration path outlined here points the way to an efficient and cost-effective solution to the growing demand for transparent transport of 10 GE packet client traffic over an OTU2-based optical network link.

## 8.0 Key Terminology

ASON	Automatically Switched Optical Network
ASTN	Automatically Switched Transport Network
ATM	Asynchronous Transfer Mode
DWDM	Dense Wavelength Division Multiplexing
GFP	Generic Framing Procedure Provisioning
LCAS	Link Capacity Adjustment Scheme
MPLS	Multi Protocol Label Switching
MTU	Maximum Transmission Unit
NTE	Network Termination Equipment
OAM&P	Operations Administration Maintenance & Provisioning
OPU2	Optical Payload Unit
OTN	Optical Transport Network
PCS	Physical Coding Sub-layer
RPR	Resilient Packet Rings
SDH	Synchronous Digital Hierarchy
TDM	Time Division Multiplexing



**For additional product and ordering information:**

[www.cortina-systems.com](http://www.cortina-systems.com)