The European IST Project DAVID: a Viable Approach towards Optical Packet Switching

L. Dittmann[1] (editor), C. Develder[2], D. Chiaroni[3], F. Neri[4], F. Callegati[5], W. Koerber[6], A. Stavdas[7], M. Renaud[3], A. Rafel[8], J. Solé-Pareta[9], W. Cerroni[5], N. Leligou[7], Lars Dembeck[6], B. Mortensen[1], M. Pickavet[2], N. Le Sauze[3], M. Mahony[10], B. Berde[3], G. Eilenberger[6]

***Abstract*— In this paper promising technologies and a network architecture are presented for future Optical Packet Switched networks. The overall network concept is presented and the major choices are highlighted and compared with alternative solutions. Both long and shorter term approaches are considered as well as both the WAN and MAN parts of the network. The results presented in this paper were developed in the frame of the research project DAVID (Data And Voice Integration over DWDM) project, funded by the European Commission through the IST-framework.**

***Index Terms*—- IP-over-WDM, Optical Packet Switching, SOA based technology, Optical metro networks**

I. INTRODUCTION

In the future communication network the optical technology is expected to play a much stronger role not only for transmission but also for network operation. For the optical network a number of stages are foreseen with optical packet switching as the final goal.

In the short term a relatively static usage of wavelength channels is envisaged, and thus the potential of optical switching technologies that allow fast dynamic allocation of WDM channels are not exploited. Optical Packet Switching (OPS) is a longer-term strategy for network evolution that exploits fast switching techniques to provide greater bandwidth efficiency, flexibility, functionality and offer finer granularity.

With the progress done in electronics we can naturally have the following reflection: why do we need optics when electronics are already present today cost effectively with respect to the proposed optical solutions?

One response can be the following: By increasing the bit rate the first limitations of electronics will be the power consumption and the impedance adaptation. Today this problem is overcome by using a high level of parallelism to process at a low bit rate. However if this technique is very powerful for basic functions there are major drawbacks in case of high capacity routers: the complexity of systems is increased and the performance is degraded.

Complexity can be associated to high cost and at the functional level, the increase of the complexity pushes the constraints on the scheduler part becoming thus the real bottleneck. By exploiting a high level of parallelism, the switch becomes a network of small switches difficult to manage. The packet loss rate is impacted and the latency is degraded mainly because of the crossing of several FIFOs.

Optics can fulfill this demand with less power consumption with higher robustness and simpler structure. The simplicity comes by exploiting the WDM dimension. Studies show that cost of optics is very close to the cost of electronic systems but still higher. Packaging and integration will be then the key factors which will position the optical technology as a winner technology in the future, and some constructors are working in that direction which is fundamental to go into a product.

II. The IST "DAVID" project

The DAVID project aims at proposing a viable approach towards OPS, by developing networking concepts and technologies for future optical networks, including traffic studies and control aspects. Finally, a proof-of-concept will be delivered through a demonstrator. Partners contributing to the DAVID project are [6]Alcatel SEL (D), [3]Alcatel CIT (F), [1]Research center COM (DK), [7]National Technical University of Athens (G), [2]Ghent University (B), OPTO+ (F) (now part of Alcatel CIT), [5]University of Bologna (I), [10]University of Essex (UK), Laboratoire de Recherche Informatique d'Orsay (F), [4]Politecnico di Torino (I), Institut National des Telécommunication (F), [8]BTexact Technologies (UK) and [9]Universitat Politecnica de Catalunya (E) (for the last project year Telenor (N) and Telefonica (E) have joined the project to work on concept optimization, exploitation, manageability and benchmarking).

The network architecture proposed covers both metropolitan area networks (MAN) and the wide-area network (WAN), each having a distinct structure, as illustrated in figure 1. In both domains, fixed-length packets are used in a slotted (synchronous) mode of operation.

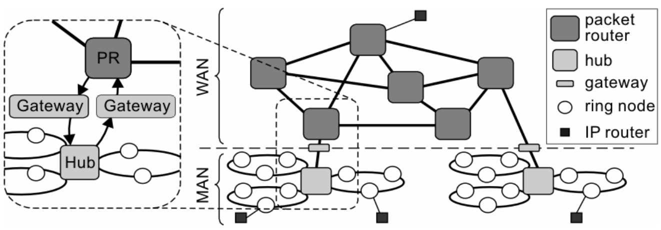

Fig. 1. The DAVID network architecture

The metro network comprises one or more unidirectional optical physical rings interconnected by a Hub, as discussed below. A ring within the MAN will consist of one or more fibers, each operated in DWDM regime, where each wavelength will be used to transport optical packets (consisting of an optical header and a payload part) having a fixed duration in time (a so-called time slot). As such, the adopted ring architecture is using both WDMA and TDMA.

The Ring Node puts the optical packets (containing client layer traffic, e.g., IP traffic) on the ring, using a MAC protocol to decide which timeslot at what wavelength to use. Through the use of this MAC protocol contention is solved and the optical path within the MAN is kept buffer-less.

The Hub will have connection points to the MAN rings and towards a WAN Optical Packet Router through a Gateway. The function of the Hub thus is to interconnect rings: note that also the connection towards the WAN can be seen from a logical viewpoint as an extra ring to switch traffic to and from. Contentions between packets flowing between MAN and WAN will be solved in the Gateway.

In the WAN, a meshed network will be formed, interconnecting Optical Packet Routers (OPRs). A link is a set of one or more fibers, each carrying multiple wavelengths. As in the MAN, the wavelengths will be used to transport optical packets of fixed time duration. The bit rate in the WAN will typically be higher carrying aggregated traffic.

III. THE CONCEPT OF THE DAVID PROJECT – MOTIVATION AND CURRENT STATUS

The main motivation for migrating to optical switching is clearly to match the switching technology to the huge bandwidth capacity of WDM transmission. Moreover, Optical Packet Switching (OPS) promises to take full advantage of available resources mainly because, compared to optical circuit switching, OPS is able to harness traffic at a much finer granularity. An optical circuit (i.e. a single wavelength) always uses a dedicated bandwidth of 2.5, 10, 40 Gbit/s or possibly even higher. This may result in a very poor filling of this bandwidth, if no efficient grooming is used. Grooming multiple client traffic streams may solve the problem, but at the price of a large number of conversions from and to electrical client layers. Optical

Packet Switching alleviates this problem by providing smaller granularity access to the optical layer (on a packet-by-packet basis), thus avoiding the costly electro-optical conversions. In addition, the packet switching approach has also advantages with respect to resilience, as it is easier to share resources in protection schemes when a packet approach is taken compared to circuit approaches due to the use of logical paths [15] that can be created with using bandwidth.

To demonstrate the advantages brought forward by OPS, and to investigate its (technological) feasibility, several OPS test networks have been developed and evaluated [2] [3] [4] [7] [11].

Towards the realization of OPS, there are two principal approaches: employing fixed-length optical packets or variable length optical packets such as e.g. Optical Burst Switching (OBS) [16]. For fixed length OPS, the choice remains either to operate the network synchronously using a time-slotted approach, or running it in an asynchronous manner [14]. Even though unslotted operation simplifies the implementation (e.g., avoidance of synchronization and packet alignment stages), the drawback is that link throughput is lower because contention is more likely to occur. Within DAVID, we have opted for a slotted, fixed-length packet approach, both in MAN and WAN. The advantages of fixed-length packets are numerous since all the telecom techniques (like traffic shaping, load balancing, flow control mechanisms, and traffic differentiation) can be adopted to really support a multi-CoS environment. An example of optical fixed packet size engineering can be found in [17]. The length of a time slot was set at 1μs for the DAVID network (both MAN and WAN).

As indicated before, the header contents will be used to control the routing of the optical packet. Different techniques to attach the header to a packet can be used: transmission on a separate wavelength, subcarrier multiplexing (SCM) (e.g., CORD [9], OPERA [7]), serial transmission of header and payload on the same wavelength (e.g., WASPNET [3], KEOPS [1]). As this header is decoupled from the payload, its bit rate can be substantially lower than that of the payload – thereby allowing to keep the header processing in the electrical domain, as explained before. The advantage of having out-of-band headers on a separate (control) wavelength as already adopted in [18] and [16], is the capability to separate the switching plane from the control plane within the nodes together with a considerable reduction of O/E/O conversion component numbers in WAN core nodes. This is especially a viable approach in ring topologies, as the synchronization between the control and data channels is reasonably easy to maintain, as opposed to the complexity this encompasses when applied in a meshed network. Therefore, we only use the out-of-band header transportation in the metro ring network. In the DAVID backbone WAN, headers will be transported in-signal. Both approaches are shown in figure 2. Within DAVID, electronics are still used for header processing while the payload is switched transparently in the optical domain.

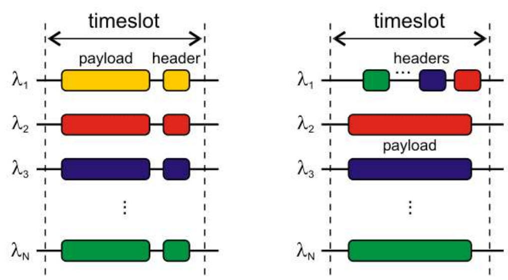

Fig.2. Illustration of out-of-band header transportation on a separate wavelength (right) and in-band headers (left), adopted in the DAVID metro and backbone network, resp.

For the optical switching fabric, there are basically two options. A first is to employ a wavelength routing switch, e.g. using an Array Waveguide Grating (AWG), where the desired output is reached by using the appropriate wavelength. This approach has been taken in WASPNET [3] and OPERA [7], but was also subject of study within KEOPS [2]. The alternative is to use a space switch such as a broadcast-and-select architecture, e.g. chosen in KEOPS. Also in DAVID, a broadcast-and-select architecture, which ensures non-blocking performance, is chosen, using SOA technology [21], as illustrated in figure 3.

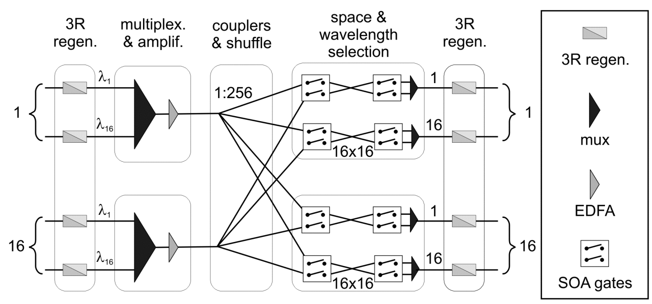

Fig. 3 Structure of the broadcast and select switch matrix adopted in DAVID.

A major issue in every packet switched network is contention resolution. In electronic routers this problem is tackled using RAM, which is unfeasible in the optical domain. Since all-optical buffers today are technologically hard to realize, there seems to be a consensus that they should be avoided as much as possible or at least be limited to a minimum. As stated before, the MAN is completely bufferless in the optical domain. In the WAN, a shared recirculating FDL buffer is used to help solving contention where exploitation of the wavelength domain does not suffice.

Regenerators are used to ensure sufficient cascadability: an end-to-end path will of course pass through several packet switches, each one of them degrading the signal. For this, 2R or 3R all-optical regenerators have been proposed (e.g. within KEOPS), based on SOAs, which have also been successfully employed for wavelength conversion using interferometric devices based on cross-phase modulation (XPM) (e.g. in KEOPS [1], OPERA [7]).

IV. Metropolitan Area Network

The DAVID MAN consists of a number of unidirectional slotted WDM rings of metropolitan dimensions, which collect traffic from several ring nodes. These nodes provide an electro/optical interface to edge routers/switches at the end of access networks via a variety of legacy interfaces (e.g. Gigabit Ethernet in business areas, PONs in mixed or residential areas, cable head-ends, or any other legacy system).

The WDM rings are interconnected to other rings via a bufferless Hub, and to a mesh of packet-switched OPRs in the core creating the complete optical WAN. The rings can be either physically disjoint, or be obtained by partitioning the optical bandwidth into disjoint portions.

The use of a Hub-node that is in control of the resources makes the DAVID MAN different from other optical ring networks like e.g. the HORNET (and without any limiting relation between node counts and the number of wavelength)

The hub node is used to forward optical packets between ring networks as well as to interconnect the metro area to the backbone through an electronic Gateway. The hub is an SOA based optical packet switch capable to cope with a very high level of traffic (TeraBit/s). The lack of real optical memories is compensated through the use of an extended multi-ring MAC protocol. The optical hub is thus buffer-less and its structure is similar to the one of the Optical Packet Router in the backbone but with reduced targeted final capacity. The main difference between the hub and the OPR is at the control level: the optical hub is configured by a controller which exploits the control channels of each connected ring network, in order to calculate the switching permutation.

The Hub comprises synchronization stages, a space switching stage, a wavelength switching stage, and regeneration stages if required (depending on the power budget). Each WDM channel operates at 10 Gbit/s that with 32 wavelengths per ring and a channel spacing of 100 GHz, occupy 24 nm of bandwidth per ring; this corresponds to a reasonable optical bandwidth for the introduction of a SOA-based technology. The maximum capacity of one ring becomes 320 Gbit/s.

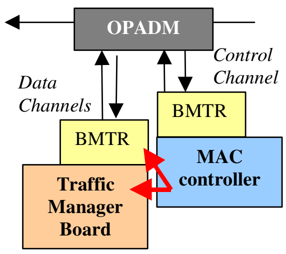

Fig. 4: Generic structure of the Ring Access Node

Ring access nodes in the DAVID MAN are composed of an electronic part and an optical part (Fig. 4). The electronic part realizes the adaptation with client layers, which. is performed in the Traffic Manager Board. Specific Burst Mode Transceivers are used to send/receive optical packets to/from the optical packet ring networks.

At the optical level, two Optical Packet Add/Drop Multiplexers (OPADMs) are currently proposed in the DAVID project to propose a progressive introduction of optical packet technologies.

Targeting a short/medium term approach, a first proposal was made to limit the use of advanced optical technologies and use commercial and mature ones instead. Based on passive structures as described in [42], the architecture uses optical couplers and off-line optical filters to minimize physical issues when cascading the OPADMs (Fig. 5a).

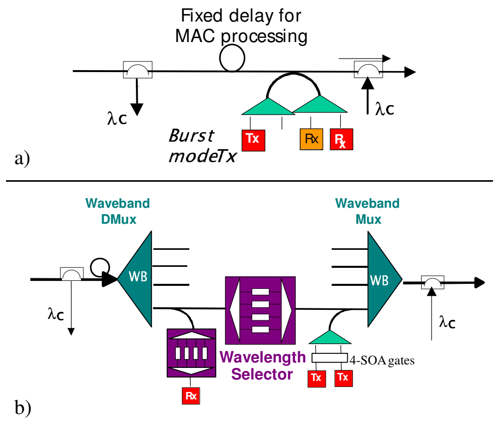

Fig. 5. Structures of the a) Passive OPADM for short-term application and b) active OPADM for longer-term applications.

A second OPADM structure is also considered as a longer-term approach allowing this erasing function (Fig. 5b). It exploits wavelength bands of 4 wavelength each allowing a very flexible design of the OPADM with parallel small add/drop functions per waveband, and limits the tunability range for the transceivers. In addition, this limits the constrains on the wavelength selector component which relies on advanced optical technology.

In the second OPADM approach the payload receivers are tunable. The number of data transceivers can be less than the number of wavelength channels available on the ring (often one single data transceiver is considered). In this approach the ring nodes must also have a selective erasure capability, in order to remove packets from the ring that are being received.

Medium Access Control

The rings are shared media, requiring a Medium Access Control (MAC) protocol to arbitrate access to its slots, in order to regulate both time and wavelength dimensions. The overall system works as a combined wavelength/time/space distributed multiplexer. Contention and collision is avoided by an allocation algorithm and intelligent operation in the ring nodes using the control channel.

The slots in the control channel have a locked timing relationship to the data multi-slots and arrive earlier at each node by a fixed amount of time allowing for processing the contents of the control slot (and tuning time in case of tunable transceivers). Only the control channel is converted to the electrical domain for processing at each ring node, while the bulk of user information remains in the optical domain until its final destination in the end ring. The control slot information includes the state (empty or used) of the data slots, and the destination address of the corresponding data packets.

The inlet-outlet Hub allocation algorithm works as follows: a measurement cycle is defined, during which the Hub monitors the use of the slots allocated to any ring pair *(i,j).* The monitoring of ring-to-ring traffic can be based either upon measurements of the load on the different rings at the Hub, or upon explicit reservations issued by ring nodes. At the end of the measurement cycle, the Hub issues a new set of switching permutations, to be used for the coming measurement cycle. The order in which permutations are generated may be of importance to smoothen as much as possible the flow entering each ring.

The Hub acts as a non-blocking switch that is re-configured in every time slot and can exploit wavelength conversion to solve contention.

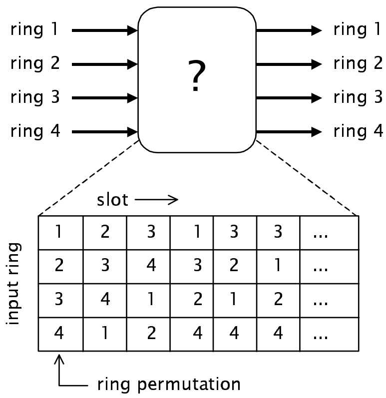

Fig. 6. Scheduling at the Hub.

The computation of the sequence of permutations operated by the Hub is a scheduling problem [22][23], as shown in figure 6. Several approaches can be envisaged to solve this problem, ranging from complex optimizations to simple heuristics.

Given this Hub behaviour, each multi-slot traverses a sequence of rings, e.g. as illustrated in figure 7. Nodes of ring *x* transmit data to be received by nodes of ring *y* (steps 2 to 4). Ring *x* can be viewed as the "upstream" ring, where transmissions occur, while ring *y* can be viewed as the "downstream" ring, where receptions occur.

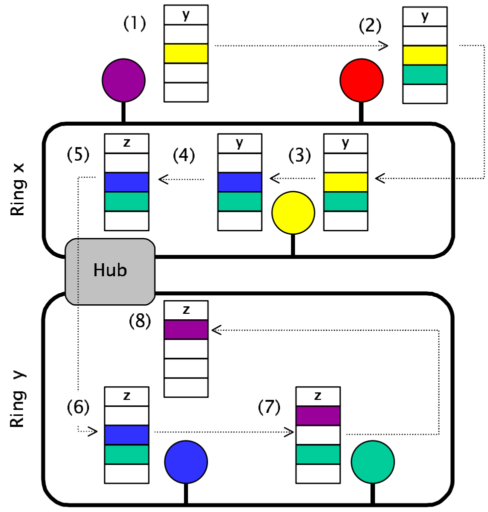

Fig. 7. Multi-slot forwarding in the MAN. Colors in slot represent packet destinations.

A simple scheduling algorithm

A simple greedy approach to compute the scheduling at the Hub is described in this section. The algorithm is run at the Hub in a centralized fashion, and can be suited for best-effort services with no strict QoS guarantees. Only unicast traffic is considered.

Assume that, for each ring pair (upstream ring, downstream ring), an indication of the "urgency" of serving that pair is available at the Hub. This information may come from measurements at the Hub, or from node reservations, or it can be obtained by other ways (see next section on fairness control). All ring pairs are sorted by decreasing urgency.

The permutation provided in outgoing multi-slots is computed as follows. Since it must be a permutation, each ring can appear one time at most as upstream ring and one time at most as downstream ring. All ring pairs (ordered as above) are sequentially scanned, and the current pair is selected for the permutation if the upstream ring was not previously selected as an upstream ring, and the downstream ring was not previously selected as a downstream ring either.

Other scheduling algorithms were studied in the project (see, e.g., [25]).

Fairness control and QoS

The proposed empty-slot operation can exhibit fairness problems under unbalanced traffic. This is particularly true in the ring topology, in which, as already mentioned, upstream nodes have generally better access chances than downstream nodes.

Credit-based schemes can enforce throughput fairness [27], such as e.g. the Multi-MetaRing [26], which is basically a generalization of the token-ring technique: a control signal or message, called SAT, is circulated in store-and-forward mode from node to node along the ring. A node forwarding the SAT is granted a transmission quota: the node can transmit up to Q packets before the next SAT reception. When a node receives the SAT, it immediately forwards the SAT to the next node on the ring if it is satisfied (hence the name SAT), i.e. if no packets are waiting for transmission or if Q packets were transmitted since the previous SAT reception. If the node is not satisfied, the SAT is kept at the node until one of the two conditions above is met. To be able to provide the full bandwidth to a single node, the quota Q must be at least equal to the number of data slots contained in one ring latency.

Another, novel QoS-sensitive MAC mechanism for slotted WDM rings based on class reservations is presented in [28]. This mechanism presents satisfactory performance with fast adaptation to a changing traffic mix, allowing for high efficiency. Other approaches, aiming at delay and bandwidth guarantees, are also being studied within the DAVID project.

V. Wide Area Network (WAN)

The architecture of the WAN network is assumed to be a mesh of high capacity optical packet routers (OPRs). The mesh could rely on a virtual topology based on lightpaths in a wavelength-routed network. The OPRs are used only where high capacity is required and where the traffic can be aggregated and conditioned properly.

This overall WAN architecture is well suited to the use of GMPLS for network control, with a hierarchy ranging from conventional electrical MPLS to optical MPLS and MPλS. As a result of this choice it is possible to guarantee scalability to the hierarchy according to the needs in terms of level of aggregation and capacity, support for Quality of Service (QoS) at the network level, and the tools for traffic engineering in the network.

It has already been outlined that for the DAVID network a fixed length slotted packet format has been chosen. However the format is different and traffic aggregation can be performed by the gateway.

Two alternatives are available when considering the management of packet slots resulting from segmentation of the same client burst. The first places a header per slot and treats any slot independently and the second treats the slots as a whole with just one header on the first slot. It results into a sort of trade-off between a purely slotted, ATM-like network, that has already been studied in the past with reference to optical packet switches [2][11] and an IP-based network with variable length packets. Both options are investigated within the DAVID project, the former being called Fixed Length Packet (FLP) and the latter Slotted Variable Length Packet (SVLP).

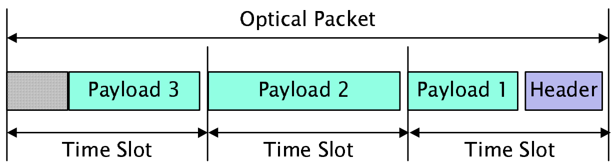

Fig.8. Example of slotted variable length packet.

As already mentioned above, both the wavelength domain and FDLs are used to solve the contention problem in DAVID, and is sketched in figure 9. A recirculating buffer is chosen, because this avoids extra switching stages/components outside the switching matrix (the buffer block only contains fiber loops).

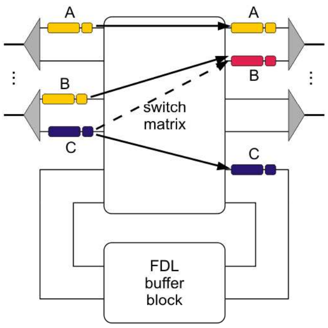

Fig.9. Illustration of contention resolution approaches in the DAVID network.

To accommodate for packet B, the wavelength domain is exploited. Packet C is redirected to the optical buffer to solve the contention.

VI. Technology and Basic Building Blocks

In the creation of an optical packet switch network both optical and electronic technologies play an important role and a seamless interaction between the two technologies is needed.

1. Optical technology

The first challenge is probably at the concept level that must be attractive enough in terms of performance and cost to force a standard on an optical packet format. Linked to the packet format, two building blocks are mandatory: the interface responsible for the creation of optical packets and the burst mode transceiver, at the edges of the optical network. Finally, at the optical technology level, the challenges are quite important since the building of large optical switches or advanced optical packet add/drop multiplexers requires a high level of integration of the technology and a low cost packaging to be competitive with respect to electronics.

Due to the asynchronous nature of the packet stream at the reception side, a specific receiver has to be designed able to receive packets that can experience packet-by-packet power variations and packet-by-packet phase fluctuations. Figure 10 shows a 10 Gbit/s burst mode transceiver capable to cope with 14 dB of power variations (electrical) between consecutive packets.

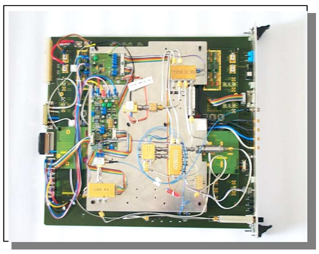

Figure 10 : Photo of a burst mode transceiver including the mux and demux part.

For the WAN the basic building block is an optical packet space switch offering capacities of several

Terabit/s. To realize such a switch, the DAVID project focuses on SOA technology, which today represents an interesting alternative because it is the only technology capable of providing all the required specifications (i.e., high ON/OFF ratio, fast switching time, low polarization sensitivity, WDM compatibility simple to implement, robustness (exploitation of the carrier density)). As a first step of integration, the objective is then to build and test a SOA gate arrays module. Figure 11 shows a highly integrated 32 SOA module [3], including its driving electronics and exploiting self-aligned flip chip assembly of SOA arrays on silicon sub-mounts [46].

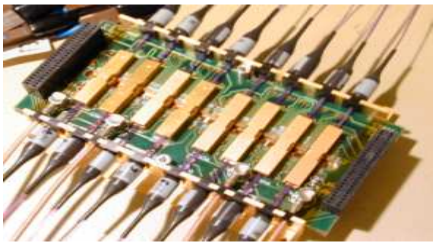

Figure 11 : Photo of a 32 SOA module

In addition to this, and in the perspective of an all-optical packet switching concept (as proposed in the DAVID project), regeneration is mandatory to enable the cascading of optical packet routers. The SOA-based Mach-Zehnder Interferometer (MZI) has been demonstrated as a key element for the regeneration [38]. Figure 12 shows a 3R-regenerator architecture validated at 40 Gb/s with integrated SOA-based MZI.

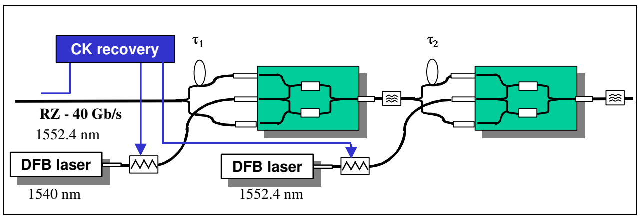

Figure 12. 3G-regenerator architecture using SOA-based MZI

In the case of a MAN using a basic optical packet add and drop structure as it is proposed in the frame of the DAVID project for a cost saving introduction of an optical technology, there is no need for a specific development. We thus need only Burst Mode Transmitters, based on traditional low cost Integrated Laser Modulators (ILM) gated by a simple SOA gate, and Burst Mode receivers. The burst mode transmitter has been experimentally validated showing no significant penalty with respect to the ILM only considering transmission distances compatible with metro needs [40].

However if we want to increase the performance of the concept by introducing more flexibility at the optical level, then tunable elements are mandatory like e.g. a monolithically integrated wavelength selector as experimentally validated in [41].

2. Electronic technology and building blocks

Necessary switching and routing functions, such as header recognition and delineation, data processing, table lookup or bit level synchronization, for instance, but also many interface operations (e.g. between non-optical legacy networks), cannot solely (or only with insufficient performance) be carried out by pure optical means, at least in the near and mid term future. So, even in an optical network, electronics will be widely deployed in addition to advanced optical switching technologies. The project integrates the advantages of both technologies in a complementing manner and functional split.

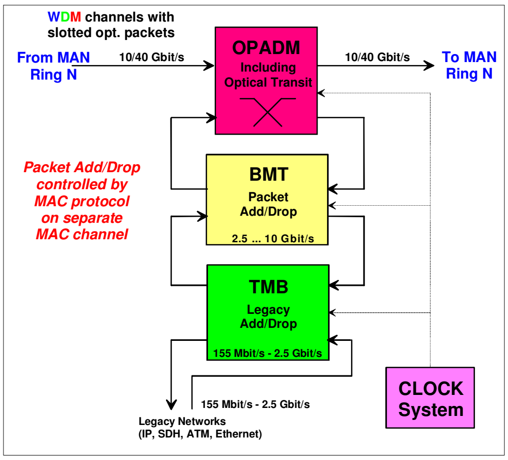

Fig. 13 Basic Outline of a DAVID Metro Ring Node

The main electronic blocks of the Metro Ring Node, that is the Traffic Manager Board (TMB; including MAC controller) and the Burst Mode Transceiver (BMT), take over all functions needed to interconnect the OPADM and the client layer, which are listed next:

Traffic from client layer to OPADM (DAVID Packet Add)

- Overhead processing of legacy traffic (e.g. IP headers). L3 routing functions, if required.

- Aggregation of legacy traffic with common attributes as destination, CoS (Class of Service), QoS (Quality of Service), etc.;
- Burstification to build packets according to the DAVID packet format and segmentation (if required) to optimize the filling ratio of the optical packet payloads;
- Queing/buffering in various queues to differentiate packets according to different possible destinations and/or priorities etc.;
- MAC to control the access to/from the optical Metro ring (separate MAC channel; MAC controller); separate transport of packet overhead in the MAN;
- Load balancing to spread the load over accessible wavelengths and/or optical bands (MAC controlled selection of transmission wavelength at or after E/O conversion);

Traffic from OPADM to client layer (DAVID Packet Drop)

- Burst mode reception of optical packets at the BMT after O/E conversion; bit level synchronization and clock recovery; preamble detection; drop function towards TMB (including bitrate adaptation);
- Extraction of client traffic from the payloads of the optical packets; recovery of e.g. IP packets (using appropriate buffer tools); reassembly (if required);
- Addressing of the client traffic (e.g. IP packets) to the correct destination port/legacy interface;
- Transmission to the client layer (synchronized with client layer, if required; using e.g. FIFO buffers).

1) Functional Overview of the BMT

The BMT (Burst Mode Transceiver) is the intermediate block between optical network components (e.g. OPADM) and TMBs and presently is designed to process one single 10 Gbit/s data channel (wavelength) running with a continuous stream of bursts (slots). For that purpose some specific functions have to be implemented in a BMT:

Optical/Electrical Conversion (O/E, E/O) at the input/output of the data channel (NRZ signal format). High input receiver sensitivity is prerequisite to cope with possibly low optical power at specific nodes.

Clock and Data Recovery (CDR) to extract the bit clock of incoming 10 Gbit/s bursts (slots), amplify and equalize the signal amplitude, detect start of slot, and finally to synchronize overhead and payload bits with the extracted clock. Especially for clock recovery and signal amplitude equalization new fast electronic approaches (e.g. in SiGe technology) are required as conventional PLL and AGC techniques, respectively, prove to be too slow for such applications.

Framer Board to implement the slot/packet Add/Drop functionality. It is connected to the DAVID network by means of 10 Gbit/s SERDES interfaces and to the TMB via 2.5 Gbit/s Add/Drop parallel interfaces.

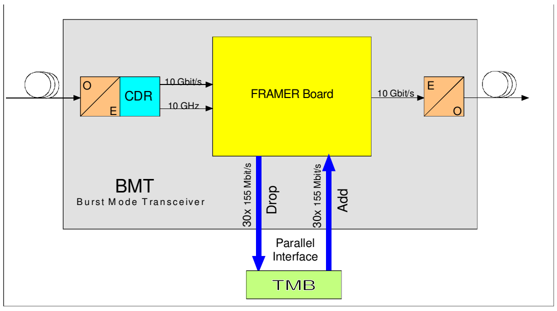

Fig. 14 Burst Mode Transceiver (BMT) connected with TMB

Regarding the MAN part of the DAVID network, one additional BMT per node is needed to receive the MAC channel. Buffering and priority scheduling of slots to be added is managed by the MAC controller and performed in the TMB.

2) Functional Overview of the TMB

A number of major functions the TMB needs to use in a Metro ring node can be outlined:

- **Buffer controller** stores the packets going towards the MAN ring according to priority and destination. It keeps the MAC controller informed about packets in the buffer each time a packet is stored in the buffer. The MAC controller informs the buffer controller when it must forward a packet towards the BMT and Metro ring.
- **MAC controller** terminates and regenerates the MAC packets travelling on the MAC channel.
- **Legacy networks controller** stores packets destined for the legacy network interfaces (one or more). Flow control from any legacy network controller has to be taken into account for this purpose.

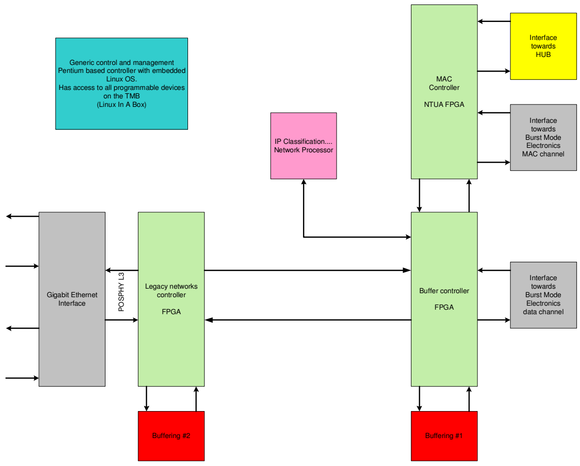

Fig. 15 Functional blocks of the TMB

VII. TRAFFIC AND PERFORMANCE ASPECTS

The traffic sources relevant to the DAVID project are traffic multiplex collected from legacy networks by routers and/or switches that interface with the backbone for high speed and long distance site to site interconnection. It is unlikely that the single application or the single user terminal will have direct access to the network. As a consequence traffic analysis and modeling should be kept as general as possible, by means of a few, general theoretical traffic models.

MAN performance

The main issues to address regarding the performance of the DAVID Metro Network are the achievable *throughput* and the *fairness*. As usual in multi-access protocols, some overhead is necessary to perform access control. The more the throughput is close to 1 (the ideal value) the better the MAC protocol. This is not enough because we also want the MAC protocol to share as evenly as possible the bandwidth between the nodes. For a multi-ring architecture several solutions have been proposed and investigated in the project and here, as an example, we present results referring to the multi-SAT access control scheme described in the previous section.

In figure 16 the throughput per node is plotted against the number of nodes active on the ring for the case of a MAN with 16 logical rings with 10 nodes and 16+1 wavelengths per ring. The packet slot is set to 1 µs and the size of the rings is such that the round trip time (RTT) is equal to 0.5 ms (that is 500 slots). Two cases have been simulated with the SAT quota set at Q=100 and Q=500. The traffic scenario consisted of IP traffic sources generating fix length optical packets according a Bernoulli distribution. The traffic is intra-ring with a load of 0.7 per ring, and node unbalanced: all nodes send packets only to node 0 of ring 0, except node 0 of ring 0 that sends packets uniformly to the rest of nodes of ring 0. As expected without fairness control (solid line), the bandwidth utilisation is *unfair*: upstream nodes (node 1, 2, 3, 4, and partially the node 5) use all empty-slots (the relative throughput is 1) and the downstream nodes (node 6, 7, 8, 9 and 10) cannot transmit (the used bandwidth is 0). By introducing the fairness control the bandwidth utilisation becomes *fair* both for the case Q=500 (triangle markers) and Q=100 (square markers).

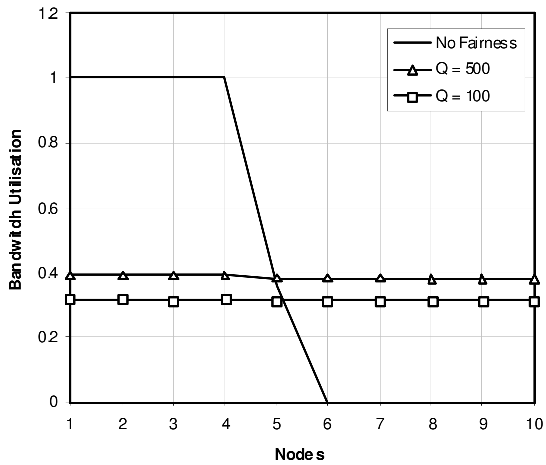

Fig. 16. Intra-ring traffic; relative throughput per node for total load on the ring of 0.7, without fairness control (solid line) and with SAT for two values of Q.

The fairness problem for inter-ring communications is addressed in Figure 17. Here an unbalanced inter-ring traffic matrix is considered where about 30% of the traffic is intra-ring while the remaining 70% is evenly spread among the remaining rings, and per-ring uniformly distributed among the nodes.

Under the traffic conditions used to obtain Figure 17, also the benefits of exploiting the Spatial Reuse in the MAC protocol was studied varying the number rings from 4 to 16. Figure 18 shows the throughput with (dashed line) and without (solid line) exploiting the Spatial Reuse for a 4 rings MAN. As expected, the higher the percentage of intra-ring traffic the more the gain of performance due to the Spatial Reuse. This is more evident in small networks (low number of rings).

a)

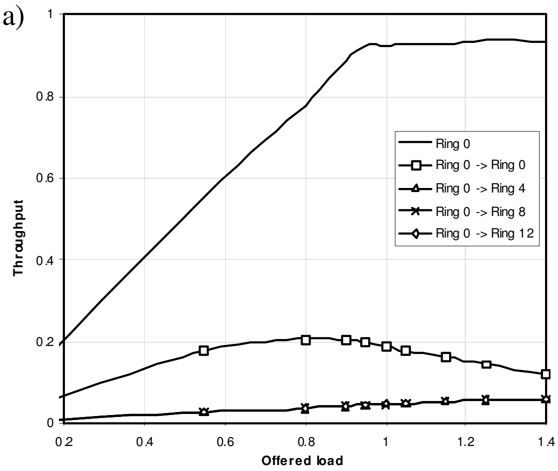

Fig. 17. Inter-ring MAN throughput with unbalanced traffic;
overall ring 0 throughput (solid line) and inter-ring throughput (marked lines) as a function of the total traffic load.

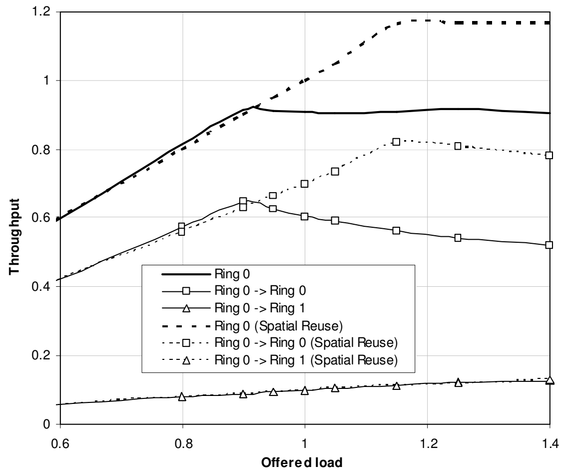

Fig. 18. Inter-ring MAN throughput under unbalanced traffic, overall ring 0 throughput:
with Spatial Reuse (dashed line), without Spatial Reuse (solid line), as a function of the total traffic load.

Results regarding the performance of MAC protocols able to differentiate the QoS between traffic streams and to guarantee some given performance to real time traffic are currently under investigation, as variation of the basic protocols here described.

OPR performance

Within DAVID the effectiveness of congestion resolution schemes in the OPR has been analyzed for FLP and SVLP. The OPR control logic performs the following function:

- *forwarding*: decide first to which fiber the packet has to be sent, in accordance with the routing table;

- *wavelength allocation*: select the wavelength, among the *w* available on the fiber, in order to minimize congestion (at most *w* packet may be transmitted on a given fiber at a given instant);
- *delay allocation*: in case no wavelength is available choose the fiber delay line to sent the packet.

The results regarding wavelength allocation presented in figure 19 refer to the algorithms studied in [33], applied to the DAVID scenario. Without loss of generality the average packet (train of slots) length is assumed to be unit and the slot is set to 0.2 (on average 5 slots per packet). The switching matrix has M=4, w=16 and no re-circulation of packets in the buffer are allowed. The curves show that there is an optimal value of the delay unit (D), placed around 1, and that the wavelength allocation algorithms have a huge influence on performance. A purely random or round-robin choice of the output wavelength bring to a very high packet loss probability. Smarter algorithms, such as that choosing the wavelength with the shortest queue (Minl) or that with the minimum gap between packets (Ming), improve the performance of several orders of magnitude at the expense of a very limited increase in the computational complexity [33].

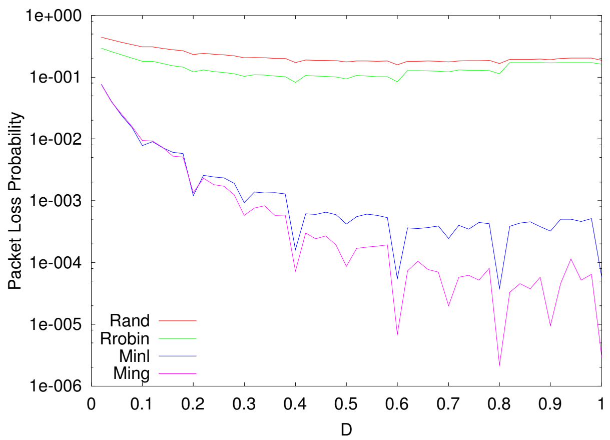

Fig. 19. SVLP: performance comparison of several wavelength allocation algorithms, as a function of the increasing rate of the FDLs, for a switching matrix with M=4, w=16, B=4 and load 0.9.

Regarding the delay allocation, the first issue to address is whether a F-FDL or an I-FDL architecture can be adopted. F-FDL appears simpler to control and, in principle, should perform similarly to I-FDL because, by using re-circulation of packets, variable delays are achievable. Unfortunately this is only partially true. Fixed the parameters B and k (number of wavelength per FDL), the product $B \times k$ puts an upper limit on the number of packet that may present in the buffer at a given time. Therefore $P = B \times k$ measures the number of "places" in the buffer. Obviously a packet needing a long delay will circulate in the buffer more than once and therefore use buffer places. Therefore for a given P we expect better performance from a I-FDL buffer because packets will re-circulate less and use less places. This is confirmed by figure 20 for FLP, where is shown that to obtain the same performance a F-FDL buffer would require many more wavelengths, meaning, at the least, a larger hardware cost. For SVLP again the analysis is more complex but the results are similar.

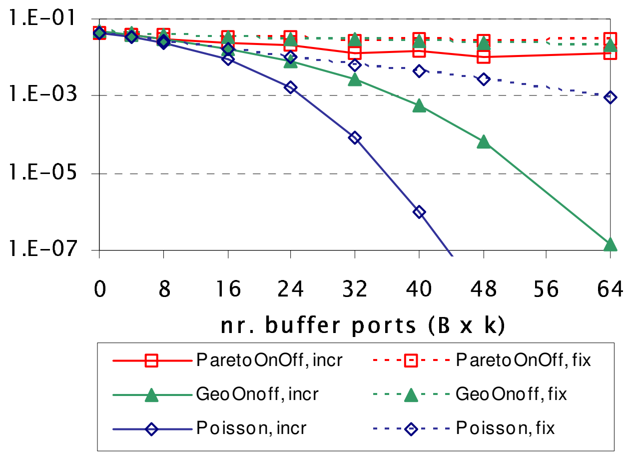

Fig. 20. FLP: performance comparison of F-FDL and I-FDL optical buffers, as a function of the number of buffer places B × k, for a switching matrix with M=6 and w=32.

VIII. Network Management Issues

The work carried out on network management stresses the service-centric perspective of an optical transport network, which is tailored to a packet transporting and/or switching capability while preserving circuit-based service provisioning. The focus has been put on four main aspects: (i) provide a centralized functional model, (ii) define a control channel, (iii) performance monitoring and network resilience, and (iv) quality of service impact on management.

For the functional model of an Optical Packet Transport Network (OPTN) as proposed by the DAVID consortium, two approaches can be taken: the first is the optical transparent version, meaning that the payload is transparently switched; the second is the opaque approach implying that the payload of each packet at every wavelength is terminated at each single node. The former requires advanced optical components like e.g. fast optical switches and hence the latter is considered to suit best at least in the short term (or intermediate) solution.

The OPTN is able to support wavelengths as well as optical packets. Therefore the architecture specified in the ITU-T Recommendation G.872 [1] has been taken as a starting point to construct a model incorporating the extra dimension of multiplexing introduced by sub-dividing each wavelength into a stream of packets. Optical Packet Switching (OPS) capability can be modeled as a new independent layer or as a so-called sub-layer, which amounts to enhancing one of the three defined layers in [43]. The latter is preferred because the existing architecture can be maintained avoiding additional efforts in defining new management interfaces. Extending G.872 to OPS can be done by adding two sub-layers within the Optical Channel layer (OCh): an Optical Packet (OPA) sub-layer which has an ingress-egress view within the OPTN where any of them can be an inter-domain access point, and an Optical Packet Section Monitoring sub-layer, which performs section monitoring between two adjacent OPTN nodes. These functional layered models require a centralized network management system. The effects a decentralised system would have on the layered model, e.g. using a separate protocol for exchange and update of routing information's to set-up and release paths across the OPTN, is not subject of discussion in this paper.

Additionally, the Optical Supervisory Channel (OSC) in [1] needs to be extended to both new optical packet sub-layers where processing is done in the electronic domain. In case 3R-regeneration is done electrically the Tandem Connection Monitoring (TCM) and Digital Channel (DCH) sub-layers can not be eliminated, since it will also be necessary in the optical packet network, either on a link basis (span by span; TCM) or end-to-end (DCH). Only if all-optical 3R-regeneration can be deployed, these TCM and DCH sub-layers are redundant and this functionality could now be modeled as part of the adaptation function between OCh and OPA sub-layers.

Network Resilience is a critical functional area that needs fully exploitation of the control channel to detect network failures and undertake recovery actions such as e.g. re-route traffic crossing the failed network element. The detection of failures is part of Performance Monitoring that needs to be extended compared to circuit-switched WDM networks because of the specific equipment used in the OPTN (e.g. SOAs, burst transmitters). The service recovery response is dependent on its place of occurrence: in the WAN, which is a meshed network, classical GMPLS schemes such as local or path protection can be deployed, while in the MAN ring-based schemes are more appropriate. Within the DAVID project, different protection strategies are discussed and evaluated in terms of both cost and complexity.

The OPTN is aimed at transporting different types of services with guaranteed Quality of Service (QoS), and therefore its impact onto service management is an important topic also addressed. There exist a legal contract frame between a service provider and a customer called Service-related level Agreement Hierarchy (SAH) that sets QoS expectations in terms of network levels like throughput, loss rate, delays, etc., and times of availability, method of measurement, pricing, as well as the consequences when QoS levels are not achieved. DAVID network encompasses both packet and circuit based services where customers expect end-to-end, measurable, formalized, and guaranteed services that can be offered through a SAH. QoS dimensions are abstract characterizations of requirements and QoS management is the concrete realization of the required quality level for services in a real system. This is commonly translated into a Service Level Agreement (SLA) and a Service Level Specification (SLS) that fall into the frame of the SAH.

IX. Prospective issues

The scope of the IST-DAVID project is twofold: Firstly, to present and validate a concept allowing dynamically reconfigurable optical packet networks. Secondly, to identify and present advanced concepts having the potential to alleviate more general and complex optical packet switching problems and to study alternative technologies and implementation methods. However, these potential solutions are naturally not mature yet and hence they require further investigation. In this context, the prospective studies address two thematic clusters. The first cluster aims at identifying novel transportation and signaling formats as well as alternative information processing methods. The second cluster aims at investigating alternative networking concepts and system architectures.

Within the first theme, the issue of identifying the most suitable transportation and signaling format is central. This format not only has a direct impact on the attainable level of integration between the IP and the WDM control planes but also it determines the amount of information that could be exchanged between nodes as well as the frequency of updating this information. The latter two issues are also related to the more general problem of "processing bottleneck". This has its origin in the fact that the currently available electronic systems are struggling to cope with information processing requirements at the transmission line rate of 10 Gb/s, for capacities approaching or exceeding few hundreds of Gigabit/s.

The IST-DAVID tackles the "processing bottleneck" problem from different perspectives. The straightforward approach is to develop 40 Gb/s electronic processing systems based on SiGe technology needed for any function that requires bit level synchronization (e.g. clock and data recovery, CDR), as well as for table lookup, header processing, burstification of data and the implementation of various control elements. Alternatively, still under the framework of information processing based on electronic devices, two approaches are examined. When multi-level (*M-ary*) coding principles based on a combination of wavelength and time are adopted, electronic processing could be carried out at a lower speed without compromising the overall information transfer rate. At the same time, the subsequent multi-wavelength label/header allows forming an out-of-band signaling mechanism compatible with GMPLS requirements [1, 2]. The second approach is complementary to the previous one demonstrating a N x 40 Gb/s transportation and switching scheme where each single wavelength 40 Gb/s high speed channel is decomposed into a group of WDM channels adopting the bit-parallel transmission principle [3].

X. Conclusion

This paper has been titled "a Viable Approach towards Optical Packet Switching" because the DAVID project is not only about developing a final solution but also to provide a path towards it. The project has taken a very pragmatic approach in which optical and electronic technologies are combined in a way that exploits the different technologies in the best way in the current state. This holds true for both the shorter and the longer-term approaches.

The project has been strongly focused on the concept and the migration path, but is full aware that a number of things need to be further enhanced to enable a cost effective introduction of optical packet switching. This specifically concerns the optical component technology that today mainly has been driven by the transmissions applications rather than networking. The project has identified that such initiatives have to be taken to improve the level of integration and reduce the cost in packaging, etc.

With this in mind the DAVID project believes that the optical packet switching option is viable and that the DAVID concept is a strong candidate. The approach has been carefully analyzed in a holistic way both theoretically and through a concept demonstrator. Throughout the project the concepts of both the WAN and the MAN parts have been adjusted according to results and observations. In this way the DAVID project provides both a viable and a validated approach towards optical packet networks.

ACKNOWLEDGMENT

The authors would like to thank all people involved in the project. We acknowledge the support of the European Commission for its financial support through the IST-project DAVID.

REFERENCES

[1] L. Roberts and C. Crump, "US Internet IP Traffic Growth", Caspian Networks, http://www.caspiannetworks.com/library/presentations/traffic/Internet_Traffic_081301.ppt, Aug. 2001.

[2] C. Guillemot et al., "Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach", J. of Lightwave Techn. Vol. 16, No. 12, December 1998, pp. 2117-2134.

[3] D. Hunter, M. Nizam, M. Chia, I. Andanovic, K. Guild, A. Tzanakaki, M. O'Mahony, J. Bainbridge, M. Stephens, R. Penty, I. White, "WASPNET: A Wavelength Switched Packet Network ", IEEE Comm. Mag. March 1999, pp. 120-129

[4] D. Wonglumson, I. M. White, S. M. Gemelos, K. Shirkande, L. G. Kazovsky, "HORNET- A Packet-Switched WDM Network: Optical Packet Transmission and Recovery", IEEE Phot. Techn. Lett., Vol. 11, No. 12, December 1999, pp. 1692-1694.

[5] D. Chiaroni et al., "Physical and Logical Validation of a Network Based on All-Optical Packet Switching Systems", J. of Lightwave Techn., Vol. 16, No. 12, December 1998, pp. 2255-2264.

[6] D.K. Hunter, I. Andonovic, “Approaches to Optical Internet Packet Switching”, IEEE Comm. Mag., September 2000, pp. 116-122.

[7] A. Crean, M. D. Vaughn, R. Gaudino, M. Shell, D. L. Blumenthal, “OPERA: An Optical Packet Experimental Routing Architecture with Label Swapping Capability”, J. of Lightwave Techn., Vol. 16, No. 12, December 1998, pp. 2135-2145.

[8] M. D. Vaughn, D. L. Blumenthal, “All-Optical Updating of Subcarrier Encoded Packet Headers with Simultaneous Wavelength Conversion of Baseband Payload in Semiconductor Optical Amplifiers”, IEEE Photon. Lett., Vol. 9, 1997, pp. 827-829.

[9] I. Chlamtac, et. al., “CORD: Contention Resolution by Delay Lines”, IEEE J. Selected Areas in Commun., Vol. 14, No. 6, June 1996, pp. 1014-1029.

[10] D.K. Hunter, M.C. Chia and I. Andonovic, “Buffering in Optical Packet Switches”, J. Lightwave Tech., Vol. 16, No. 12, December 1998, pp. 2081-2094.

[11] F. Masetti et al., “High Speed, High Capacity ATM Optical Switches for Future Telecommunication Transport Networks”, IEEE J. Selected Areas in Commun., Vol. 14, June 1996, pp. 979-998.

[12] D. K. Hunter, W. D. Cornwell, T. H. Gilfedder, A. Franzenm I. Andonovic, “SLOB: A Switch with Large Optical Buffers for Packet Switching”, J. Lightwave Technol., Vol. 16, No. 10, October 1998, pp. 1725-1736.

[13] S. Yao, B. Mukherjee, S. Dixit, “Advances in Photonic Packet-Switching: An Overview”, IEEE Comm. Mag., February 2000, pp. 84-93.

[14] D. Cotter, A. D. Ellis, “Asynchronous Digital Regeneration and Networks”, J. Lightwave Techn., Vol. 16, No. 12, December 1998, pp. 2068-2080.

[15] D. Colle, A. Groebbens, P. Van Heuven, S. De Maesschalck, M. Pickavet, P. Demeester, “Porting MPLS-recovery techniques to the MPλS paradigm”, Special Issue on ‘Protection and Survivability’ of Optical Networks Magazine, Vol. 2, No. 4, Jul./Aug. 2001.

[16] C. Qiao, M. Yoo, “Optical Burst Switching (OBS) – a new paradigm for an optical internet”, Special Issue on WDM Networks, J. High Speed Networks, Vol. 8, No.1, January 1999, pp.69-84.

[17] F. Callegati, “Which Packet Length for a Transparent Optical Network?”, SPIE Symposium on Broadband Networking Technologies, Dallas, USA, November 1997.

[18] J. Turner, “Terabit Burst Switching”, http://www.arl.wustl.edu/~jst/, July 1998

[19] K.Kitayama, K.Onohara, M.Murata, “Capability of optical code-based MPLS (OC-MPLS)”, in Proceedings of 6th IFIP Working Conference on Optical Network Design and Modelling (ONDM2002), Torino, Italy, February 2002.

[20] D. Chiaroni, N. Le Sauze, T. Zami, J.-Y. Emery, “Semi-conductor optical amplifiers: a key technology to control the packet power variation”, in Proceedings of 27th European Conference on Optical Communication (ECOC 2001), Amsterdam, The Netherlands, October 2001.

[21] D. Chiaroni, et al., “First demonstration of an asynchronous optical packet switching matrix prototype for MultiTerabitclass routers/switches”, in Proceedings of 27th European Conference on Optical Communication (ECOC 2001), Amsterdam, The Netherlands, October 2001.

[22] B. Hajek, T. Weller, “Scheduling Non-Uniform Traffic in a Packet-Switching System with Small Propagation Delay”, IEEE/ACM Trans. on Networking, Vol. 5, No. 6, Dec. 1997, pp. 813-823.

[23] C.S. Chang, W.J. Chen, H.Y. Huang, “Birkhoff-von Neumann Input Buffered Crossbar Switches”, in Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel, pp. 1614-1623, March 2000.

[24] R.E. Tarjan, *Data Structures and Network Algorithms*, Society for Industrial and Applied Mathematics, Pennsylvania, Nov. 1983.

[25] A. Bianco, G. Galante, E. Leonardi, F. Neri, and M. Rundo, “Access Control Protocols for Interconnected WDM Rings in the DAVID Metro Network”, in Proceedings of 2001 Tyrrhenian Intern. Workshop on Digital Communications (IWDC 2001), Taormina, Italy, Sept. 2001.

[26] I. Cidon, Y. Ofek, “MetaRing - a Full-Duplex Ring with Fairness and Spatial Reuse”, IEEE Trans. on Communications, Vol. 41, No.1, Jan. 1993, pp. 110-120.

[27] A. Marsan, A. Bianco, E. Leonardi, A. Morabito, F. Neri "All optical WDM Multi-rings with Differentiated QoS", IEEE Comm. Mag., Feb. 1999, Vol. 37, No. 2, pp. 58-66.

[28] J.D. Angelopoulos, N. Leligou, H. Linardakis, A. Stavdas "A QoS-sensitive MAC for slotted WDM metropolitan rings", in Proceedings of 6th IFIP Working Conference on Optical Network Design and Modelling (ONDM2002), February 2002, Torino, Italy.

[29] F. Callegati, H. C. Cankaya, Y. Xiong, M. Vandenhoute, "Desing issues for optical IP routers", IEEE Communication Magazine, Vol. 37, N. 12, pp. 124-128, Piscataway N.J., USA, Dec.1999.

[30] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott, "Optical routing of asynchronous, variable length packets", IEEE Journal on Selected Areas in Communications, Vol. 18, No. 10, pp. 2084-2093, October 2000.

[31] An Ge, F. Callegati, L. Tamil, "On Optical Burst Switching Protocols and Traffic Self-similarity", IEEE Communications Letters, Vol. 4, N.3, pp. 98-100, Piscataway N.J., USA, Feb. 2000.

[32] H. Christiansen, *Using OPNET to compare and analyse different traffic-bundling schemes*, Proceedings of OPNETWORK 2001, Washington, D. C., USA, August 2001.

[33] F. Callegati, W. Cerroni, G. Corazza "Optimization of Wavelength Allocation in WDM Optical Buffers", Optical Networks Magazine, Vol. 2, No. 6, pp. 66-72, 2001.

[34] K. Thompson, G. J. Miller, and R. Wilder, Wide-area internet traffic patterns and characteristics, IEEE Network, vol.11, n. 6, pp. 10-23, November-December 1997.

[35] F. Masetti et al, "Design and Implementation of a Multi-Terabit Optical Burst/Packet Router prototype", Postdeadline paper, OFC 2002

[36] A. Dupas, et al, "10 Gbit/s RZ asynchronous packet mode receiver with high input power dynamic for future optical packet switching systems", Postedeadline paper, Photonics In Switching, Monterey CA, June 2001, Post Deadline session

[37] N. Sahri et al " A highly integrated 32-SOA gates optoelectronic module suitable for IP multi-terabit optical packet routers ", Postdeadline paper, OFC 2001.

[38] B. Lavigne et al. "Cascade of 100 optical 3R regenerators at 40 Gbit/s based on all-active Mach-Zehnder interferometers", regular paper, We F2.6 , ECOC 2001.

[39] D. Chiaroni et al., " Semiconductor Optical Amplifiers : a key technology to control the packet power variation", regular paper paper, ECOC 2001.

[40] N. Le Sauze et al., " DBORN: a shared WDM Ethernet bus architecture for optical packet metropolitan networks", accepted paper for PIS 2002, July 2002

[41] S. Khalfallah et al. "First Optical Packet Switching Demonstration with Sixteen-Channel InP Monolithically Integrated Wavelength Selector Module", postdeadline paper, ECOC 2001.

[42] N. Le Sauze, et al, "A novel, low cost optical packet metropolitan ring architecture", Postdeadline paper, ECOC 2001.

[43] ITU_T Recommendation G.872:" Architecture of Optical Transport Networks".

[44] M. Renaud, "Components for advanced WDM networks", invited paper, ECOC'98

[45] M. Renaud, D. Keller, N. Sahri, S. Silvestre, D. Prieto, F. Dorgeuille, F. Pommereau, J.Y. Emery, E.Grard, H.P. Mayer, "SOA-based optical network components", Invited paper, ECTC'2001, Lake Buena Vista, Florida, May 29-June 1, 2001.

[46] D.Keller, P.Aribaud, D.Auvinet, M.DiMaggio, F.Dorgeuille, C.Drocourt, D.Legat, V.Dhalluin, F.Pommereau, C.Porcheron, M.Prunaire, P.Piatti, JC.Remy, M.Renaud, N.Sahri, S.Silvestre, S.Squedin, A.Göth, J.Scherb, "Hybridization of SOAs on Si platform for routing applications", invited paper, IPRM'2002.