Routing Algorithm with Smart Energy Management on VCSEL Interconnected Networks

I. Gravalos[a], A. Siokis[a], P. Kokkinos[b], E. A. Varvarigos[a,b]
(a): Computer Engineering and Informatics Department, University of Patras, Greece
(b): Computer Technology Institute and Press "Diophantus", Patra, Greece

***Abstract*—Energy consumption and the associated costs constitute a crucial issue concerning the design and operation of data networks and data centers. Energy-awareness is required in all levels, ranging from physical layer to algorithms, protocols and applications. Architecture-wise, a promising solution for tackling the increasing energy requirements is the deployment of optics at both long and shorter distances, including within data centers. Vertical Cavity Surface Emitting Lasers (VCSEL) constitute a popular photonic transmitter technology used in numerous short-range applications, providing also the ability to reduce energy consumption by scaling down the transmission bit rate. In this study we focus on the algorithmic aspects of energy management by proposing an OptiMal EnerGy Aware (OMEGA) routing algorithm to operate in optical networks utilizing VCSEL-based opto-electronic links. The algorithm leverages the capability of VCSELs to adapt the energy dissipation with respect to the transmission bit rate. Simulation results, under various traffic patterns, show that OMEGA balances efficiently the traffic load over the network's links, resulting in high throughput and low energy consumption.**

Keywords*—energy aware routing; energy management; green networks; optical networks; optimal routing.

I. INTRODUCTION

Information and Communications Technology (ICT) sector accounts for an important and rapidly increasing share of global energy consumption. The estimated power consumption of Telecom networks and Data Centers (DC) in 2007 was 293 and 330 billion kWh, respectively, and was predicted to rise to 952 and 1012 billion kWh in 2020 (thus 1964 billion kWh in total) [1]. For comparison, the total energy consumption of the European Union in 2013 was 2798 billion kWh. Power consumption is a critical issue also for High Performance Computing (HPC): studies back in 2010 projected that a 10PF HPC machine in 2012 would require 5MW [2]. K-computer, a top-10 10PF HPC system, which started working in 2011 requires more than double the predicted amount of power [3]. Thus, it is recognized that the issue of energy consumption in all fields of communication and data networks should be very carefully considered.

Optical technology, offering high bandwidth, low loss transmissions and energy efficiency, is a promising solution for reducing the energy requirements in telecoms and datacoms. Optics have already replaced copper-based communication in long-haul telecom systems and are penetrating shorter distances in campus and enterprise LANs (already used for rack-to-rack communication in DC and HPC) [2]. Optics are targeted to be used also for board-to-board, chip-to-chip, and even on-chip communications, leading to 'greener' network architectures. Fig. 1 illustrates the vision for the application of

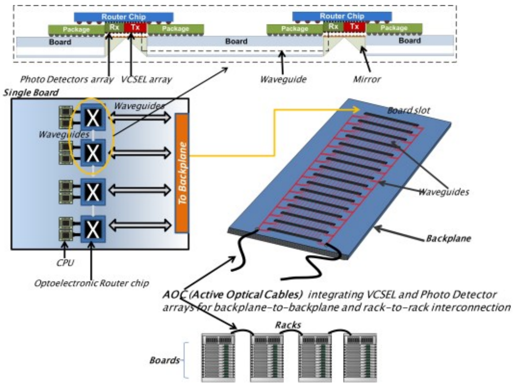


Figure 1: Illustration of the application of photonics in "in-the-box" networks: AOCs, Optical Printed Circuit Boards, Opto-electronic chips.

optics in "in-the-box" networks such as DC and HPC: cabling of racks via Active Optical Cables (AOCs), opto-chips and router chips with embedded photonic transmitters/receivers, coupled to boards with integrated optical waveguides.

However, fiber solution is not a panacea. Hence, regarding energy consumption, the installation of optical technology could benefit consumption but it is not singular approach. The reason is two-fold. First, the space of potential solutions for energy efficient networks is affected by all system levels, ranging from physical layer to algorithms, protocols and applications. Secondly, energy consumption at fiber-based transmissions can be directly influenced by the transmission data rate on each link, as opposed to copper-based communication where the operation of the routing equipment accounts for the largest share of energy consumption. Since the expectation of fiber networks is mainly the bandwidth increase, a respective increase in energy consumption is expected as well.

This is the case regarding optical networks that utilize Vertical Cavity Surface Emitting Lasers (VCSELs). VCSELs is a popular, cost effective photonic transmitter technology that has been established and matured within the datacom industry, serving in data infrastructure links for over a decade. They are easily coupled to fiber and operate at bit rates up to 40 Gb/s with low energy consumption for a distance up to 1000m [4]. From this established foundation, VCSELs are emerging as an enabling technology across a wide range of applications. To name a few: they are deployed in AOCs for rack-to-rack interconnections, they are used for the realization of optical I/O interfaces to ASIC electronic router chips (commercially available) [5], and are also considered to operate in Optical Network Units (ONUs) over next gen passive optical networks for energy efficiency [6]. Furthermore, VCSEL-based links

978-1-5386-1629-1/17/$31.00 ©2017 IEEE

allow the dynamic tuning of the power consumption in relation to the demanded traffic [7].

Hence, in this study, we exploit this latter attribute in optical networks whose links are opto-electronic and based on VCSELs. To this end, we induce the cost minimization problem to a multicommodity flow problem and we propose a routing algorithm, called the OptiMal EnerGy Aware (OMEGA) scheme, which aims to distribute predetermined flows among several paths in order to achieve the minimum aggregate flow at each link (and thus the minimum respective data rate) that minimizes the total energy consumption. Provided that the energy dissipation at each link is not linear [7], we present a constrained nonlinear optimization problem, for which OMEGA obtains the optimal solution based on the optimal routing and flow deviation concept [8]. It optimally load balances traffic among multiple paths, leading to better utilization of the network resources, while it also decides on suitable bit rates for each transmitter in order to operate with minimum energy dissipation. Our simulation results indicate that OMEGA achieves better energy efficiency and higher throughput compared to other (energy-aware or unaware, shortest and non-shortest path) routing algorithms or other load balancing methods. The comparisons exhibit that OMEGA outperforms its competitors for a variety of realistic HPC-application traffic patterns and network topologies.

The remaining of this paper is organized as follows. In Section II we outline the basic ideas of energy aware algorithms and comment on related work. In Section III we give a detailed description of the considered energy model for VCSELs. In Subsection IV.A we formulate the problem of Energy Aware Optimal Routing and in Subsection IV.B we provide a scheme, namely the OMEGA algorithm, for its solution. Finally, in Section V we describe the simulation setup and present the results obtained. Finally, Section VI presents our conclusions.

II. RELATED WORK

Energy awareness in the operation of networks has lately been a research subject of high importance. There is a large literature on energy consumption awareness in wireless and mobile ad-hoc networks [9] and in developing energy efficient protocols in wireline networks [10]. Likewise, energy consumption has attracted great interest in optical networks since it can influence the operation of WDM devices and corresponding switching technologies [11].

There are two main approaches presented in the literature [12], [13] for achieving energy efficiency in data networks. The first one is based on powering on and off network components, creating corresponding transmission and idle periods. The second approach focuses on adapting the link rates to the network load. Regarding the latter perspective, recent studies on energy-aware traffic engineering under a variety of assumptions on the theoretical energy profile [14], [15] have highlighted the potential of important energy savings. These studies conclude that the effectiveness of energy-aware routing on reducing energy consumption depends on: (i) network topology and traffic conditions and (ii) the device technology that corresponds to different power models. Authors in [7], [16] consider VCSEL-based opto-electronic links and explore the energy savings achieved, by scaling down the supply voltage of the link components when the required rate is less than the maximum link rate supported.

In this paper we opt for the second approach because: (i) the energy dissipation on each link (considering VCSELs and photodiodes) can be accurately obtained (representing real networks) and exponentially depends on the respective data rate (as shown in Section IV), (ii) the large amount of external traffic as well as the frequent data exchange in HPCs and cloud networks limit the deployment of the on/off approaches. Furthermore, the proposed scheme does not preclude the former approach, but rather can be applied complementary (after an efficient network establishment) to further reduce the total energy consumption of the network.

III. VCSEL ENERGY CONSUMPTION MODEL

As a fundamental architecture of optical interconnected systems, an opto-electronic link consists of the transmitter, the receiver and the optical channel. Considering a passive channel, the total energy consumption depends on the transmitter (Tx) and the receiver (Rx). In particular, the energy absorbing components that operate at the transmitter are: a *laser source* that we will assume to be implemented by a VCSEL, whose operation is to convert 0s and 1s into low and high intensities, respectively, and a *VCSEL driver* that modulates the driving current to the VCSEL, based on the input bit patterns. Respectively, at the receiver the power consumption is due to the *photodetector* that converts the optical bit stream back into electrical current signals and is implemented by a photodiode, the *transimpedance amplifier* (*TIA*) that converts the current signals fed by the photodiode to amplified voltage signals, and finally the *clock and data recovery circuit* (*CDR*) [7], [16] An optical link with the aforementioned components is depicted in Fig. 2.

Consequently, the corresponding total power consumed by an opto-electronic link is given by [7], [16]:

$$P_L = P_V + P_{VD} + P_{PH} + P_{TIA} + P_{CDR} \,, \qquad (1)$$

which adds the energy consumptions of the respective components (as explained in what follows) comprising a link. The individual power consumption at each component are analyzed below:

a) The total power consumed by the VCSEL depends on a current threshold I_t, above which it can be stimulated and emit light. The light intensity then depends on the modulation current I_d that is fed by the driver. Thus, the power consumption of a VCSEL is:

$$P_V = (I_t + \varepsilon I_d)(V_t + V_d + V_{dd} - V_{tn}) \qquad (2)$$

where ε is the switching factor, V_{dd}^2 is the supply voltage, V_t is

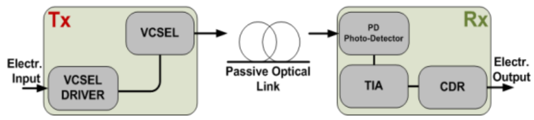


Figure 2. Architecture of a VCSEL-based opto-electronic link.

the threshold voltage, V_d is the voltage drop on the resistance, while the remainder of the subtraction corresponds to the minimum source-drain voltage required for the gate to ensure saturation point.

b) The power consumed by the VCSEL driver is dynamic and accrues to the charging of the inverter chain for each transmission. A VCSEL driver consists of a set of cascaded inverters, each of which has size γ times larger than the previous one. The total power dissipated at the driver stages can be modeled as:

$$P_{VD} = \varepsilon C_{VD} V_{dd}^2 T_{LR} \, , \quad (3)$$

T_{LR} is the transmission link bit rate and C_{VD} is the total VCSEL drive capacitance of the inverters (sum of the input and output capacitance), given by:

$$C_{VD} = C_L - C_{in} + \sum_{\gamma=0}^{k-1} (C_{in} + C_{out}) \varepsilon^{\gamma}, \quad (4)$$

where C_L is the load capacitance of the inverter chain, C_{in} and C_{out} are the input and output capacitances of the minimum sized inverters, respectively.

c) The photodetector at the receiver is responsible for converting the optical signal into photon current. In order to assure successful detection, the photodetector must operate at the minimum receiver sensitivity power R_{min}, which is proportional to the transmitted bit rate. Therefore, the VCSEL power consumption depends on the receiver's needs for a given T_{LR} and V_{dd}. Given that the photodetector's power dissipation P_{PH} (<1mW) is much lower than that of the other components, we will consider it negligible and will ignore it.

d) The total power dissipated at TIA can be calculated as:

$$P_{TIA} = I_{bias} V_{dd} + I_d V_{dd} + \varepsilon(\alpha\beta)^2 R_f \quad (5)$$

However, the dissipation in TIA is dominated by the first term of the right hand side, and hence we simplify the equation to

$$P_{TIA} = I_{bias} V_{dd} \, , \quad (6)$$

where I_{bias} is the bias current of the internal amplifier, and T_{LRmax} is the maximum bit rate that assures the correct functionality of the TIA.

e) Finally, the power consumed by the CDR unit is given by:

$$P_{CDR} = \varepsilon C_{CDR} V_{dd}^2 T_{LR}, \quad (7)$$

where C_{CDR} is the capacitance of the CDR unit.

IV. ENERGY AWARE ROUTING IN VCSEL NETWORKS

A. *Energy Aware Optimal Routing Problem*

The total power consumption at each opto-electronic link depends directly on the transmission bit rate at which the individual components operate. In particular, the supply voltage at each of the aforementioned components of a link can be restricted proportionally when the transmission bit rate required on the link is less than its capacity [7]. We will see that significant energy savings can be obtained by scaling properly the transmission bit rate on each link of a path, while satisfying the network traffic demands in terms of delay and throughput constraints. We assume the existence of a central controller that enables dynamic power management of all channels in the network (similar hypotheses were made in [7], [16]). This implies that the electrical part of the opto-electronic router chips (such as [5]) should allow the reconfiguration of the characteristics of the embedded optical links as described in Section III. This assumption is valid under the Software Defined Networks (SDN) paradigm [17] that promises centralized control on the network, currently under full development and deployment in various networks (optical, IP, wireless, datacenters, etc.).

Considering the (energy) cost function of Eq. (1) and a given set of traffic flows $F = \{F_{s,d}\}$ comprising $F_{s,d}$ flow units (e.g., bits/sec) from an origin node s to a destination node d, our objective is to find the set of paths $P_{s,d}$ (and associated traffic flows) that should be used for routing this traffic in order to minimize the total energy consumption over the total network traffic F. The idea is that distributing the total (s,d) traffic among several paths from s to d, and jointly optimizing power consumption for all source destination pairs (s,d) in the network, the amount of total flow per link is reduced and thus the required bit rate at a VCSEL could also be reduced, decreasing the total power dissipation. Hence the problem can be formulated as:

$$\begin{aligned} &\text{minimize } \sum_{(i,j)} P_L^{ij}(T_{LR}^{ij}) \\ &\text{subject to } \sum_{p \in P_{s,d}} f_p = F_{s,d} \end{aligned} \quad (8)$$

where T_{LR}^{ij} is the total flow in bits/sec on link (i,j), $P_L^{ij}(T_{LR}^{ij})$ is the power consumption on link (i,j), and f_p is the flow of path p. For the total transmission rate T_{LR}^{ij} of all the flows on link (i,j) we have

$$T_{LR}^{ij} = \sum_{\substack{all \ p \ s.t. \\ (i,j) \in p}} f_p \quad (9)$$

We assume that this transmission bit rate T_{LR}^{ij} between two nodes i, j is generated by configurable and identical VCSELs, in which case the consumption *function* $P_L^{ij}(T_{LR}^{ij})$ is the same for all links (i,j), given by Eq. (1), but its value depends on the transmission rate on that link. Thus by substituting Eq. (9) to Eq. (8) we obtain:

$$P(F) = \sum_{(i,j)} P_L^{ij} \Big(\sum_{\substack{all \ p \ s.t. \\ (i,j) \in p}} f_p \Big)$$

where F is a vector containing the flows f_p.

Furthermore, by obtaining the optimal bit rates for each link we can adjust the supply voltage of a VCSEL to the one required for successful communication between two nodes and thus save significant transmission energy. As mentioned in [7][16], the supply voltage varies in proportional values to the bit rate variations. In particular, with a bit rate of 5Gb/s a supply voltage of 0.9V will be required. If the bit rate is doubled to 10Gb/s the supply voltage must also be doubled. Hence, we further simplify Eq. (1) by substituting the supply voltage as follows:

$$V_{dd} = \frac{V_{dd(\max)}}{T_{LR(\max)}} * T_{LR}^{ij} \quad (10)$$

where $T_{LR(\max)}$ is the maximum transmission rate of the VCSEL and $V_{dd(\max)}$ is the respective supply voltage for $T_{LR(\max)}$. Substituting Eqs. (2), (3), (5), (6), (7), (10) into Eq. (1), the cost function can be rewritten as:

$$P_L^{ij}(T_{LR}^{ij}) = (I_t + \varepsilon I_d)\left(V_t + V_d + \frac{V_{dd(\max)}}{T_{LR(\max)}} T_{LR}^{ij} - V_{tn}\right) + \varepsilon C_{VD}\left(\frac{V_{dd(\max)}}{T_{LR(\max)}}\right)^2 (T_{LR}^{ij})^3 + I_{bias}\frac{V_{dd(\max)}}{T_{LR(\max)}} * T_{LR}^{ij} + \varepsilon C_{CDR}\left(\frac{V_{dd(\max)}}{T_{LR(\max)}}\right)^2 (T_{LR}^{ij})^3 \quad (11)$$

B. OptiMal EnerGy Aware (OMEGA) Routing Algorithm

To obtain the optimal solution for the multicommodity problem formulated in (8), we will rely on the following general condition for optimality:

Lemma 1: If f: $R^n \to R$ is a differential convex function on the R^n and X a convex subset of R^n, the $x^* \in X$ is an optimal solution of the general minimization problem in the form of (8) if and only if $\nabla f(x^*)^T(x - x^*) \geq 0, \forall\, x \in X$. For proof we refer the reader to [8].

As one can see, Eq. (11) is a convex monotonically increasing function of T_{LR}^{ij} that increases sharply as T_{LR}^{ij} approaches the maximum capacity of the link (*i,j*), since the second derivatives P_{ij}'' exist and are positive in [0,$T_{LR(max)}$). The partial derivative of *P* is given by:

$$\frac{\partial P(F)}{\partial f_p} = \sum_{(i,j) \in p} (P_L^{ij})' \quad (12)$$

where the derivatives $(P_L^{ij})'$ are evaluated at the aggregated flows corresponding to *F*. Defining the first derivative $(P_L^{ij})'$ of the link energy cost with respect to its flow T_{LR}^{ij} as the (*first derivative*) *energy length* of link (*i,j*), then (12) provides the *energy length* of that path *p*. Hence Lemma 1 can be applied to (8), which provides that:

$$\frac{\partial P(F)}{\partial f_p}(f_p - f_p^*) \geq 0$$

This, along with the requirement of $f_p^* > 0$, implies that the necessary and sufficient condition for optimality is that only paths of minimum first derivative length must have positive flows.

C. Feasible Solution

Provided the abovementioned induction we observe that, for each (*s, d*) pair, if flow traverses a non-optimal path then a portion of corresponding flow could be redirected to the minimum first derivative path in order to come closer to the optimal solution. This can be shown by observing that if *F* is a feasible path flows set and *ΔF* is a corresponding portion shift, then the scalar β given by $G(\beta) = P(F + \beta\Delta F)$ has first derivative around $\beta = 0$:

$$G'(\beta) = \sum_{\substack{for\ all \\ s,d\ pairs}} \sum_{p \in P_{s,d}} \frac{\partial P(F)}{\partial f_p} \Delta F$$

Moreover, if *ΔF* is positive for minimum paths and negative to all other paths while maintaining the flow conservation for each (*s, d*) pair we will obtain

$$G'(\beta)|_{\beta=0} < 0$$

which implies that the objective function can be reduced by a shifting in direction *ΔF*.

However, since the path costs depend on flow values, the minimum path length generally changes after each flow redirection. Hence, the formulated problem can be optimally solved by methods such as the Frank-Wolfe (flow deviation) or the steepest descent. In this case, given an initial (non-optimal) vector *F* of flow allocations for all *(s, d)* pairs, the optimal solution can be obtained by iteratively shifting portions of flows β along the minimum paths, obtaining new values as:

$$F = F + \beta(F^* - F)$$

where $\beta \in [0,1]$. The iterations continue until further flow redirections cannot improve the overall cost of (8). The value of β can be obtained by estimating the second order Taylor approximation of $G(\beta) = P(F + \beta(F^* - F))$ around β=0, deriving :

$$\beta = \min[1, -\frac{\sum_{(i,j)}(T_{LR}^{ij*} - T_{LR}^{ij})P_{ij}'}{\sum_{(i,j)}(T_{LR}^{ij*} - T_{LR}^{ij})^2 P_{ij}''}]$$

where, P_{ij}' and P_{ij}'' (the first and second derivatives of P_L^{ij}, estimated at f_{ij}), are given by:

$$P_{ij}' = (I_t + \varepsilon I_d)\frac{V_{dd(\max)}}{T_{LR(\max)}} + 3\varepsilon C_{VD}\left(\frac{V_{dd(\max)}}{T_{LR(\max)}}\right)^2 (T_{LR}^{ij})^2 + I_{bias}\frac{V_{dd(\max)}}{T_{LR(\max)}} + 3\varepsilon C_{CDR}\left(\frac{V_{dd(\max)}}{T_{LR(\max)}}\right)^2 (T_{LR}^{ij})^2$$

$$P_{ij}'' = 6\varepsilon C_{VD}\left(\frac{V_{dd(\max)}}{T_{LR(\max)}}\right)^2 T_{LR}^{ij} + +6\varepsilon C_{CDR}\left(\frac{V_{dd(\max)}}{T_{LR(\max)}}\right)^2 T_{LR}^{ij}$$

Apparently β cannot take large values because optimality constraint will be violated (to see that otherwise, the upper part of Eq. (13) is a good approximation only for βsmall enough). The power minimization algorithm at each step tries to shift a portion β of the flow from a non-shortest energy length path to the shortest one for each communication pair (*s,d*). In this way the flow is balanced between the links of a given topology, thus obtaining the minimum transmission rates (thus minimum energy consumption) with respect to the traffic entering the network.

V. Performance Evaluation

In order to assess the performance of the proposed optimization technique we performed a number of simulation experiments with respect to the total energy dissipation on the network, the total traffic losses and the standard deviation of the total load for each link over varying traffic loads. As total energy dissipation we define the cumulative energy of all network's links, as estimated by Eq. (1), which is necessary for each link to transmit at appropriate data rates in order to satisfy network's flow demands. The total power dissipation is metered in Watts with respect to network's traffic load given in Gb/s. Traffic losses may occur when the total flows from all the paths traversing a link exceed the maximum link capacity. The load deviation is used as a metric of the load balancing that each routing algorithm achieves and exhibits how load ranges between links.

A. Simulation Model and Network Topologies

The simulation environment was implemented in OMNET++ and consists of (a) 16 nodes placed on a mesh 4x4 topology and (b) a larger network with 30 nodes randomly connected by 37 bidirectional links. Mesh-like architectures are popular in several HPC systems, such as IBM, CRAY supercomputers as well as Fujitsu's K-computer. For example, 16 nodes in a CRAY system [18] correspond to 4 blades, each hosting 4 nodes (with a single rack hosting 24 blades or 96 nodes). Each node hosts the processing elements and routes traffic through proprietary CRAY router chips. In our case, we assume that each node's routing element is equipped with VCSEL transmitters and PhotoDetectors (PD) that are assigned to corresponding communication links (such opto-electronic routing elements have been presented, see [5]). Thus, a VCSEL-to-PD connection comprises a node-to-node communication link. The maximum link capacity is defined to match the maximum VCSEL data rate at 10Gb/s. Consequently, in case the total flow assigned at a link exceeds the maximum capacity, the respective link dissipation will be estimated according to the maximum bit rate since the surplus flow is assumed to be lost.

B. Traffic Patterns

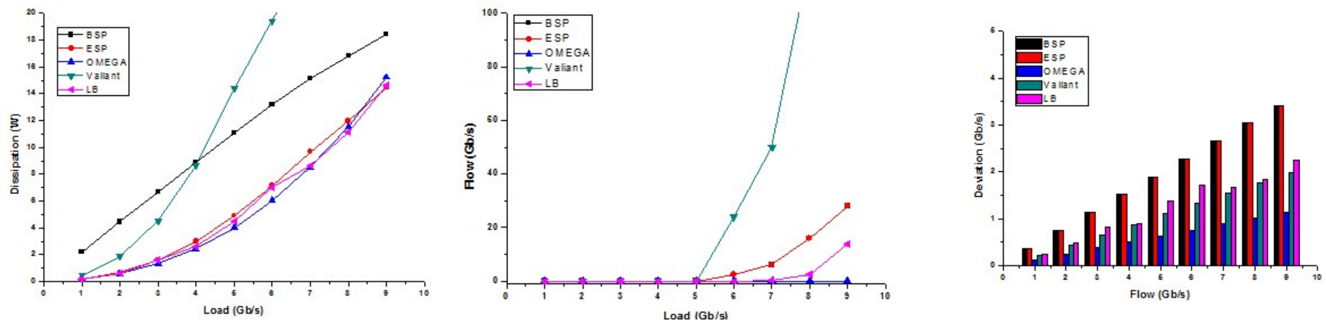
The traffic patterns we used in the simulation scenarios correspond to 2 HPC application-based traffic patterns (FFTW [19], SuperLU [20]) and one synthetic pattern (bit complement). The 2 HPC application profiles were obtained by using the IPM (Integrated Performance Monitoring) tool which profiles performance aspects and resource utilization of a parallel program, maintaining a low-overhead. The FFTW is an implementation of the discrete Fourier transform (DFT). Its behavior closely resembles Uniform Random Traffic (URT) where each node communicates with all the other nodes (equally likely) using one-to-one communication (not broadcasting). Hence, the generated traffic L_n of node n is equally distributed among the individual flows of the corresponding communication pairs. Thus, every node n sends L_n/N units of traffic to every other node. The SuperLU is a general purpose library for the direct solution of large, sparse, non-symmetric systems of linear equations on high performance machines. The SuperLU is data intensive only locally. Finally, the bit complement is a permutation traffic pattern in which each source sends all of its traffic to a single destination (computed by complementing the bits of the source address). It is a traffic pattern that, among others, is typically relied upon to demonstrate poor performance [21].

C. Alternative Approaches

In order to further assess the performance of the proposed OptiMal EnerGy Aware routing algorithm over VCSELs interconnects, we performed comparisons against four well known routing algorithms: (a) the basic shortest path algorithm (BSP), i.e. the minimum hop routing algorithm, which is energy agnostic (in the sense that no voltage scaling is performed based on the link load), (b) an energy-aware variation of it, called Energy Shortest Path (ESP). ESP determines the routing paths in means of minimum hops, similarly to BSP, but it also additionally adapts the voltage V_{dd} on each link to suitable values with respect to instantaneous aggregate bit rates, as explained earlier in Section III, in order to further lower energy consumption. In BSP, the value of V_{dd} is fixed and set to 1.8V, which is the respective value to assure successful transmission at data rates of 10Gb/s. We also used 2 algorithms performing load-balancing: (c) Valiant's algorithm [21] and (d) a shortest path balancing algorithm (LB). In the former, every packet sent from some source to some destination node, is first sent from the source to a randomly chosen intermediate terminal node and then from the latter to destination. This is an effective way to randomize routing for worst case traffic patterns, as it converts a traffic pattern into an "average" traffic pattern through randomization. For mesh/torus topologies, both the intermediate and destination nodes are reached through shortest path/dimension order routing [21]. Thus, in terms of flows, regardless of the original traffic pattern, each one of the 2 phases of Valiant's algorithm appears to be URT, leading eventually to twice the link loads of URT. LB algorithm follows a similar approach, but load balances traffic using only the shortest paths (in means of minimum hops) for the communication pairs. The voltage scaling is applied in those algorithms as well.

D. Simulation Results

As expected, the proposed algorithm achieves the optimal load balance among the networks' links with respect to the energy dissipation (depicted in Fig. 3). This means that the flows accrued at each link are kept low, thus requiring low pick data rates. As a result, the network resources are utilized properly in order to serve the requested load, while maintaining the total consumption to low levels. The following results exhibit the superiority of the proposed method, since the rest of the algorithms lack the optimal flow

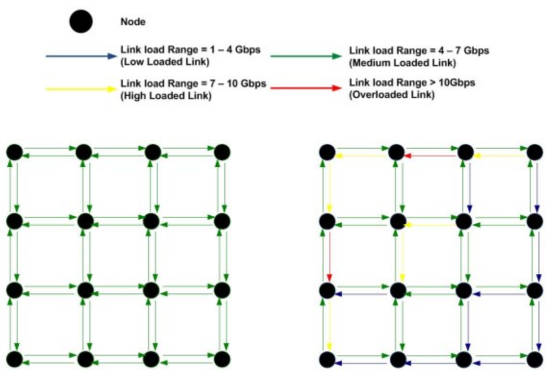


(a) (b) (c)

Figure 3: Total network (a) energy dissipation, (b) flow losses and (c) standard link load deviation of the 4x4 Mesh topology over FFTW/URT.

allocation per link, with respect to the energy dissipation.

As Fig. 3a depicts, for the FFTW/URT case the energy aware approaches clearly outperform the basic shortest path routing and Valiant's algorithm, while OMEGA achieves significant energy savings of almost 18,5% compared to ESP even when the load increases enough for the most links on the ESP to get saturated (around 5 Gb/s as shown in Fig. 3b). That is also the exact case for the BSP, which is overlapped by ESP in Fig. 3b. OMEGA routing achieves 70% energy savings compared to Valiant's algorithm. The LB approach performs well at light loads (experiencing losses for loads of 8Gb/s and more), but OMEGA still achieves 7% better savings. In addition, the load deviation of the links from the total average link load is much lower due to the proposed technique (Fig. 3c), thus achieving zero flow losses (Fig. 3b) and utilizing better the available network resources. OMEGA's optimal load balancing among several available paths for each communication pair reduces the bandwidth utilization of each link, even when the total load of the network is high. This equilibrium is illustrated in Fig. 4, assuming a mesh network topology, which exhibits the link loads at the point where the shortest path algorithm starts saturating. Since link loads depend on the routing algorithm, BSP and ESP accrue the same link loads. Energy savings of ESP is a result of voltage scaling due to bit rate variations.



(a) (b)

Figure 4: Link loads for (a) OMEGA and (b) shortest path

Similar results are obtained for the SuperLU traffic pattern depicted in Fig. 5. The somewhat increased locality of traffic compared to FFTW/URT lends to minimum hop routing. OMEGA routing performs marginally better than LB algorithm in terms of energy consumption. Since both algorithms in this case use shortest paths, the small differences in deviation from the mean value of the demanded channel bandwidths in Fig. 5c indicate that channel loads are marginally better balanced in OMEGA.

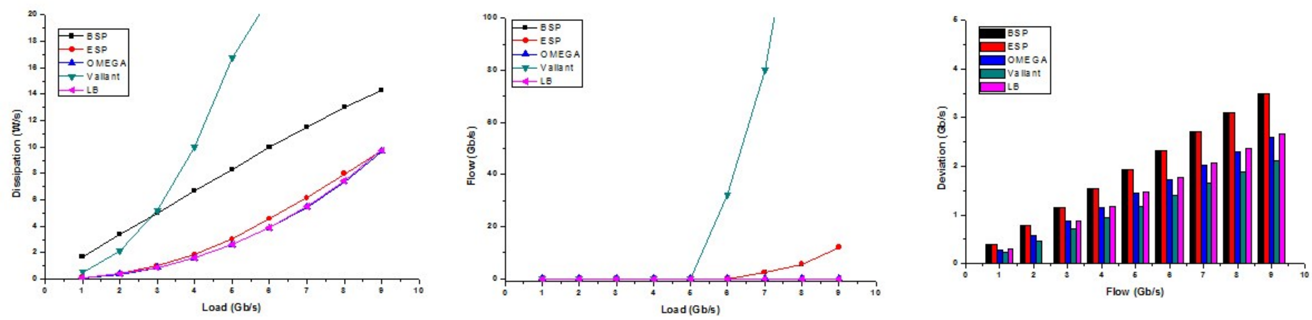
For the bit complement traffic pattern (corresponding results are not shown due to space constraints) all examined shortest path algorithms, using load balancing or not, perform poorly. BSP and ESP saturate the network for injection bandwidth of 1.66 Gb/s, while LB saturates for 3.22 Gb/s. Valiant's algorithm exhibits identical behaviour as in Fig. 5, saturating the network for load of 5 Gb/s. OMEGA routing for 5Gb/s of traffic yields 30% energy savings compared to Valiant's algorithm.

Similar results were obtained for the random network topology. Despite the sparse network connectivity, the proposed OMEGA algorithm can achieve less energy consumption and better link utilization than the other algorithms considered, since the load is optimally distributed and as a result links become saturated much slower. In particular, the obtained results showed energy savings of up to 76%, 20.5% and 11% compared to BSP, ESP and LB, respectively.

To sum up, simple energy-aware shortest path routing strategies achieve relatively low energy losses for low loads, but they do not perform well in terms of throughput for certain traffic patterns and they tend to saturate the network early. On the other hand, load balancing traffic throughout the network topology (as in Valiant's algorithm) achieves good performance for adversarial traffic patterns where minimum hop routing performs poorly, destroying however any locality of the traffic (also performing poor when the traffic is already balanced). Another important observation is that such "blind" traffic load balancing all over the network is also prohibitive from an energy consumption perspective for VCSEL based networks, since it keep all links loaded. Thus, a routing optimization with respect to energy consumption, while respecting the constraint that all generated traffic must reach the respective destinations as in OMEGA, yields the optimal results in terms of both energy reservations and throughput.

VI. CONCLUSIONS

We considered a VCSEL-based optical network infrastructure, supporting dynamic reconfiguration of the bit rate and energy footprint of its links, and proposed an optimal routing strategy using a VCSEL related energy model as a cost function. Our results showed energy consumption improvements up to 63.5% and 18.5% compared to basic and energy aware shortest path algorithms and up to 70% compared to an energy-aware version of Valiant's algorithm, while achieving higher throughput than all of them. The respective improvements in energy consumption were 76% and 20.5% for a small network with randomly interconnected nodes.



(a) (b) (c)

Figure 5: Total network (a) energy dissipation, (b) flow losses and (c) standard link load deviation of the 4x4 Mesh topology over SuperLU.

REFERENCES

[1] Make IT Green: Cloud Computing and its Contribution to Climate Change. Greenpeace International, 2010.

[2] M.A. Taubenblatt, "Optical Interconnects for High-Performance Computing," Journal of Lightwave Technology, 30(4), pp. 448-457, 2012.

[3] http://www.top500.org/system/177232#.U3mxKnbm5GkP.

[4] Moser, et al. "Energy-efficient oxide-confined high-speed VCSELs for optical interconnects. "*SPIE OPTO*. International Society for Optics and Photonics, 2014.

[5] K. Hasharoni et al. "A high end routing platform for core and edge applications based on chip to chip optical interconnect," Optical Fiber Communication Conference. Optical Society of America, 2013.

[6] E. Wong, M. Mueller, P.I. Dias, C. A. Chan, M. C. Amann, "Energy saving strategies for VCSEL ONUs.," *Optical Fiber Communication Conference*. Optical Society of America, 2012.

[7] X. Chen, L.-S. Peh, G.-Y Wei, Y.-K. Huang, P. Prucnal, "Exploring the design space of power-aware opto-electronic networked systems," Proceedings - International Symposium on High-Performance Computer Architecture, pp. 120-131, 2005

[8] D. P. Bertsekas, R. G. Galager, "Data Networks," 1987

[9] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials* 15(1), pp. 167-178, 2013

[10] J. Chabarek, et al., "Power awareness in network design and routing," *Proceedings - IEEE INFOCOM*, art. no. 4509688, pp. 1130-1138, 2008

[11] R.S. Tucker, et al. "Evolution of WDM optical IP networks: A cost and energy perspective," *Journal of Lightwave Technology*, 27 (3), pp. 243-252, 2009

[12] S. Nedevschi, L. Popa, G. Iannaccone, D. Wetherall, S. Ratnasamy, "Reducing Network Energy Consumption via Sleeping and Rate-Adaptation," *Proc. 5th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*, San Francisco, pp. 323-336, 2008

[13] R. Bolla, R. Bruschi, , F. Davoli, F. Cucchietti, "Energy efficiency in the future internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures," IEEE Communications Surveys and Tutorials, 13 (2), 5522467, pp. 223-244, 2011

[14] J. C. C. Restrepo, C. G. Gruber, and C. M. Machuca, "Energy profile aware routing," Communications Workshops, 2009. ICC Workshops 2009. IEEE International Conference on. IEEE, 2009.

[15] A. P. Bianzino, et al. "Energy-aware routing: a reality check," IEEE GLOBECOM Workshops (GC Wkshps), 2010.

[16] A. K. Kodi, A. Louri, "Energy-efficient and bandwidth-reconfigurable photonic networks for high-performance computing (HPC) systems," IEEE Journal on Selected Topics in Quantum Electronics, 17 (2), 5510030, pp. 384-395, 2011.

[17] N. McKeown, et al., "Openflow: Enabling innovation in campus networks," SIGCOMM Computer Communication Review, Vol. 38, No. 2, pp. 69–74, 2008.

[18] A. Bland, et al. "Jaguar: The world's most powerful computer," *Memory (TB)* 300.62, 362, 2009.

[19] M. Frigo, S. G. Johnson. "The design and implementation of FFTW3." *Proceedings of the IEEE* 93.2, pp. 216-231, 2005.

[20] X. Li, and J. W. Demmel. "SuperLU_DIST: A scalable distributed-memory sparse direct solver for unsymmetric linear systems," *ACM Transactions on Mathematical Software (TOMS)* 29.2, pp. 110-140, 2003.

[21] W. J. Dally and B. Towles, "*Principles and Practices of Interconnection Networks*," Morgan Kaufmann Publishers, 2004.