PBD: Packet Buffer DVFS

Eyal-Itzhak Nave and Ran Ginosar

Electrical Engineering Department

Technion – Israel Institute of Technology

Haifa, Israel

eyal.nave@intel.com; ran@ee.technion.ac.il

***Abstract*—DVFS circuits are applied to network-controllers in-order to reduce power dissipation in networks. When receive and transmit buffers are relatively empty, energy is saved by lowering the packet processing rate. Simulations show an average energy savings of 29% across various traffic loads and 19% across various congestion levels.**

I. INTRODUCTION

A study [1] conducted in 2006 estimates that, in the U.S. alone, energy consumption of networked systems approaches 150 TWh, with an associated cost of around 15 billion dollars. The generation of electric power produces more pollution than any other single industry in the United States [2], contributing to acid rain, urban smog, global climate change and significant health risks. The wide spread of mobile devices (smartphones, tablets, etc.) which are network-connected demands longer battery life and less heat dissipation, while data centers growth (as a result of cloud computing architectures, web databases and social network) struggle with challenges of cooling the data center and lowering electricity high costs.

Dynamic Voltage and Frequency Scaling (DVFS) is a common technique for balancing performance and power consumption in processors and digital logic. Examples of DVFS implementations on multi-core architectures can be found in [3] and [4]. In [3] the authors coupled the multi-core architecture trend (48 IA cores) with a new message-passing protocol to create a "data center on a die". Power was kept at a minimum by transmitting dynamic, fine-grained voltage-change commands over the network to an on-die voltage-regulator controller (VRC) for the 8 voltage islands. Further power savings were achieved through active frequency scaling at the tile granularity using 28 frequency islands. In [4], DVFS was employed on a 2 Tb/s 6x4 mesh network for a Single-Chip Cloud Computer (SCC) with 32 IA cores. The chip featured DVFS to minimize total power consumption. The IA cores transmitted voltage change commands over the Network on Chip (NoC) to an on-die VRCthat was addressable by all cores. Software running on IA cores could independently modulate the voltages and frequency enabling workload-aware DVFS. In [5] NoC architecture was combined with a globally-asynchronous locally-synchronous (GALS) paradigm as a natural enabler for DVFS mechanisms, enabling unit-level (vs. global) scaling. ALPIN, an "Asynchronous Low Power Innovative NoC" circuit, was designed to demonstrate different adaptive design techniques aiming at reducing both dynamic and static power consumption in a 65 nm CMOS technology. The authors presented efficient techniques from system level (SystemC Transaction Level Modeling (TLM)) down to the physical level (such as the VDD hopping technique) proposing also architectural and design level power optimizations. In [6] the costs of integrated on-chip voltage regulators, a key component for efficient voltage transitions in DVFS, were modeled and a CMP system with on-chip voltage regulators was analyzed. Their conclusion was that on-chip regulators could significantly improve DVFS effectiveness and lead to overall system energy savings (21%) in a CMP, but architects must carefully account for overheads and costs, such as the regulator power loss. An accurate model for calculating these overheads was presented in [7], where explicit formulations were given for the various sources of DVFS energy and delay overheads. In our work, we have tried to use these formulations, but the need for characterization of the converter inductor current throughout the transition time ($I_L(t)$) for these calculations drove us towards an empirical solution by extracting these overheads using Virtuoso simulation.

In 2006 the IEEE 802.3 Working Group started an effort to improve the energy efficiency of Ethernet. This effort became IEEE P802.3az Energy Efficient Ethernet (EEE) resulting in IEEE Std 802.3az-2010, which was approved September 30, 2010. EEE uses a Low Power Idle mode to reduce the energy consumption of a link when no packets are being sent. The standard defines mechanisms to stop transmission when there is no data to send and to resume it quickly when new packets arrive. This is done by introducing the concept of Low Power Idle (LPI), which is used instead of the continuous IDLE signal when there is no data to transmit. LPI defines large periods over which no signal is transmitted and small periods during which a signal is transmitted to refresh the receiver state to align it with current conditions. In [8], results from a simulation-based performance evaluation show how packet coalescing can be used to improve the energy efficiency of EEE, if a careful analysis of the trade-offs between energy consumption and network performance is considered. However, EEE is limited in its use to wired network systems using the IEEE 802.3 Ethernet protocol (of the 2^{nd} OSI model layer – "Data Link"). Our PBD method, as opposed to EEE, is not bounded to a specific network protocol of any OSI layer, as long as packet buffers are being used to store the packets before processing. Moreover, it can be applied for either wired or wireless networks (IEEE 802.11).

In this paper we investigate the potential for power saving in computer networks. As opposed to modern processors, which handle quite well idle (or low utilization) periods, network controllers consume the same power whether they are idle or fully utilized. This problem stems from the desire to maintain open network links even when they are not utilized, thus preventing shutdown of relevant components. However, power consumption during low utilization periods can be reduced by lowering the voltage and frequency, which enable maintaining the links open and even handling low traffic loads. Higher voltage and frequency should be used only when needed, i.e. higher traffic loads which require higher processing rate from the network controller.

A network controller contains both a transmitter (TX unit) and a receiver (RX unit). Each one of these units contains a packet buffer which stores the packets before processing for either transmission or acceptance. We claim that the amount of packets stored in the packet buffer (i.e. packet buffer size) is a good indicator of the unit's work-load because it is directly affected by the number of packets awaiting processing. Our simulation results show an average energy savings of 29% across various traffic load levels and 19% across various congestion levels.

The rest of this paper is organized as follows: Section II describes the proposed "Packet Buffer based DVFS" (PBD). Section III describes the simulation and compares energy saving results. Sections IV and V compare energy consumption and saving of PBD across various traffic loads and various network congestion levels, respectively. Section VI summarizes this work and offers conclusions.

II. PBD: Packet Buffer DVFS

The DVFS work-point (frequency and voltage) at any specific time point is adjusted to the work-load at that time. We apply the same reasoning to network controllers, where the work-load is the number of packets in its transmit and receive buffers. In this work we examine a two-states (high / low) DVFS system: When using PBD, if the packet buffer is filled above a predefined threshold, a high power mode is used, enabling higher performance at the cost of higher power consumption. Once the packet buffer is emptied below a predefined threshold, a low power mode is selected, enabling lower power consumption at the cost of lower performance.

Two research questions are examined in this paper:

1. What is the expected energy saving of PBD in a network system?
2. How do network system conditions (traffic load, transmission rates and network congestion) affect PBD energy saving? What conditions maximize energy saving?

III. Simulation Modeling and "Behavior"

This section compares energy consumption of PBD and existing network systems. A configurable DVFS simulation environment was developed, simulating different traffic patterns transmitted via a TCP session with different network conditions and comparing energy consumption of different DVFS policies. We assume separate DVFS work-points (High/Low) and transmission rates for TX and RX. We further assume that the switching time between power states is negligible. The simulated system is schematically described in Fig. 1.

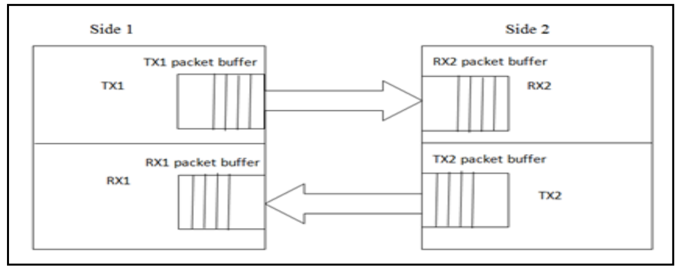

Figure 1. simulation architecture
one port, separate voltage & frequency domains for RX/TX

PBD behavior during a simulation run can be observed in Fig. 2. At time units 6, 55, 63, 77, 162 and 301 the packet buffer size crosses the high threshold (20), which causes a transition to high power mode. At time units 51, 60, 72, 159 and 297 the packet buffer size reaches or crosses the low threshold (10), affecting a transition to low power mode.

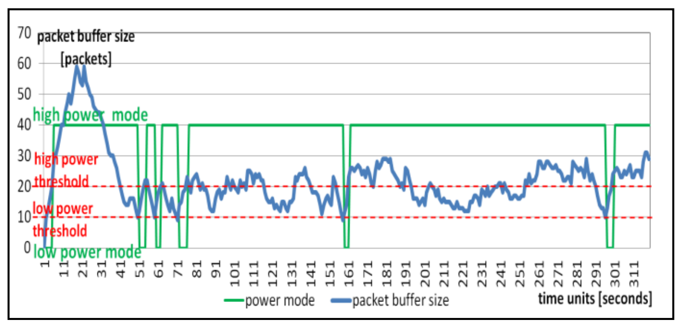

Figure 2. Packet buffer simulation

When summing up the energy dissipated during the packet transmission sessions in both TX and RX of both sides, energy consumption with PBD is 4,695J, compared to 5,217J without DVFS, saving 522J or 10% energy.

IV. Comparison of Energy Saving Across Various Traffic Loads

DVFS is adjusted according to the amount of packets that need to be transmitted or received (traffic load). Therefore, we investigate the effect of different traffic load levels on DVFS and energy savings. Packet arrival rate is modeled as a Poisson distributed stochastic process. The probability that k packets will arrive in a single time-unit ($\Delta t=1$) is

$$P(k,\lambda) = (\lambda^k e^{-\lambda})/k! \qquad (1)$$

where λ is the expected the number of packets arriving per time unit. . We use a benchmark of packets arriving during

1600 seconds at a random Poisson distribution rate. In addition to varying mean arrival rate λ, we also adjust the transmission rate to maintain a stable queue system to 3λ and λ at high and low power modes, respectively. We assume the combined effect of double frequency (from 250 MHZ to 500MHZ) and higher voltage (from 1V to 1.2V) causes the processing rate to triple.

A. *Energy Consumption*

Fig. 3 compares the energy consumption with and without PBD across various traffic loads and transmission rates (the horizontal axis indicates λ).

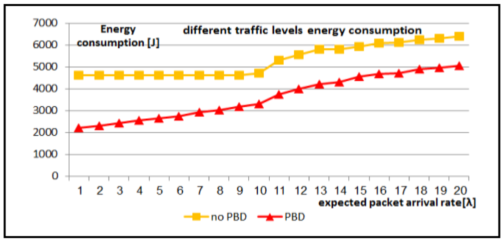

Figure 3. Various traffic levels energy consumption

Energy consumption using PBD increases with traffic load at an average rate of additional 158J per unit increase of λ. In high traffic loads higher performance is required to serve the packet buffer. Therefore, the network controller uses high power mode and its associated high transmission rate at the cost of higher power/energy consumption. Higher transmission rates result in more packets being transmitted and during non-network-congested periods, these packets are delivered successfully. However, during network congestion periods, some packets are lost. Each lost packet as a result of timeout requires the network controller to retransmit the lost packet as well as all the packets that have been transmitted after the lost packet (according to TCP guidelines). The more packets that are transmitted in parallel (at 3λ transmission rate), the more packets that are required to be retransmitted in case of a lost packet due to timeout. Every packet retransmission consumes energy in the transmission circuits in the network controller and postpones shutting this circuits down (namely switching to low power mode). The more packets that are retransmitted, the more energy that is dissipated. When the network is not congested, high power mode consumes more energy in order to enable higher performance so that more packets may be transmitted to the other side faster. When the network is congested, lost packets prevent the actual progress of the transmission task, as the transmitted packets do not reach the other side, causing the performance advantage of high power mode to be wasted.

For low traffic load levels up to λ=9, the energy consumption when not using PBD is roughly constant at 4618J. For high traffic load levels (above λ=10), the energy consumption increases as traffic load levels increase (roughly at the same rate as when using PBD). The energy consumption at the low traffic load levels ($\lambda \leq 9$) is roughly constant because using high power mode (the default without PBD) provides ample performance to complete the processing of all the packets in the packet buffer in the same minimal required time (1,605 seconds for transmitters and 1,602 seconds for receivers). The sum of all the active duration periods of the four components of the "network system" described in Fig. 1 (TX1, TX2, RX1, RX2) is 6,414 seconds (≈ 1 hour and 47 minutes). The energy consumption ($E=P^*\Delta T$) is the same, because only high power mode (P=0.72W) is used (for all four components) for the same total amount of time (ΔT=6,414 seconds) for all low traffic load levels ($\lambda \leq 9$).

However, for higher traffic load levels, $\lambda \geq 10$, the performance provided by high power mode, when not using PBD, is insufficient to transmit all the packets arriving at the average λ rate during 1,600 seconds, especially when packets are lost during congested network periods and significant retransmission of packets is required, resulting in higher energy consumption. Therefore, in higher traffic loads, the trend of energy consumption is similar to that of PBD with increasing energy consumption as traffic load increases.

B. *Energy Saving*

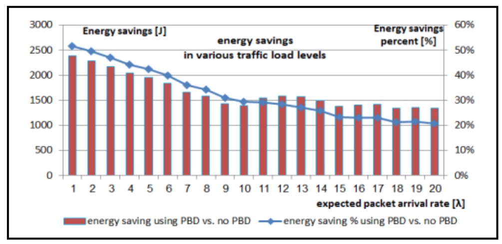

Figure 4. Various traffic levels energy saving

According to Fig. 4, the energy saving trend of PBD vs. no PBD decreases at low traffic load levels (below λ=10), increases at 10<λ<15, and becomes roughly stable afterwards. The maximum energy saving of ~2.4KJ, 52% of the energy consumed with no PBD, is achieved at λ=1. As traffic load increases the energy saving decreases. This is because high traffic loads fill the packet buffer above the threshold causing the network controller to operate at high power mode. The packet transmission rate at high power mode is three times the average packet arrival rate, enabling the network controller to empty the packet buffer below the low power mode threshold every once in a while, resulting in a transition to low power mode and saving energy. The total energy saving during low power mode is at least 1.3KJ.

The slight increase in energy saving when 10<λ<15 of up to 1.6KJ in Fig. 4 is a result of the ability of the network controller to transition to low power mode which is especially useful at higher traffic loads when the network is congested and packets get lost. At low power mode packets get transmitted at a lower transmission rate of λ packets/second (vs. 3λ), therefore, fewer packets get lost and fewer packets need to be retransmitted, saving energy. The same absolute level of energy saving presents a lower percentage at higher energy levels, shown as a continually decreasing trend. To summarize, energy savings percentage is between 21% and 52%, with a median of 29%.

V. Comparison of Energy Saving Across various Network Congestion Levels

We now observe the impact of PBD at various levels of network congestion. Ten congestion levels 1-10 were simulated, where 1 is the least congested network level and 10 is the most congested one. A congested network is characterized by lost packets. The more congested the network is, the more packets are lost. A network is usually not congested 100% of the time. We define simulated congestion levels according to both the frequency and length of the congestion periods and the frequency of packet loss in a congested period.

A. Energy Consumption

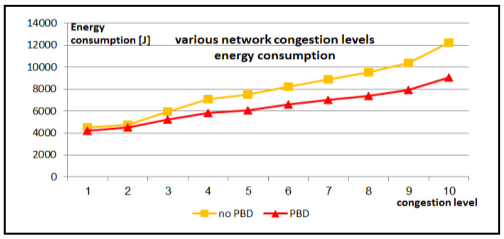

Figure 5. various congestion levels energy consumption

At low congestion levels, there is little benefit in reducing energy consumption by PBD. The energy saving benefit increases with congestion, starting with 200J in low congestion levels and reaching above 3KJ in high congestion levels. Clearly, Fig. 5 shows the major benefit in energy consumption of a PBD controller over existing controllers without PBD, and in particular its advantage in coping well with high congestion networks. When congestion increases to the highest level (10), PBD incurs lower energy consumption of 9KJ whereas no PBD dissipates more than 12KJ, resulting in 3KJ saving, about 25%.

B. Energy Saving

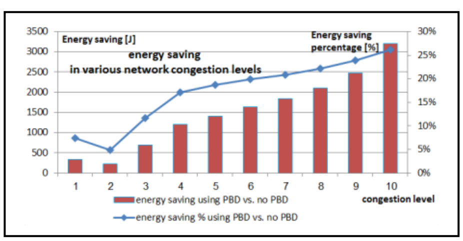

Figure 6. Energy savings at various congestion levels for PBD and no PBD

Fig. 6 highlights the advantage in energy saving achieved with PBD, which increases with congestion. At low congestion levels (1-2) the chart shows a decrease in savings, while increasing at higher levels of congestion, implying that the congestion at levels 1-2 is too low to affect energy consumption.

VI. Summary and Conclusions

We apply DVFS for the first time to network controllers and introduce a novel power management circuit, Packet Buffer DVFS (PBD). The DVFS operating-point (voltage and frequency) is based on work-load as indicated by the number of packets in the receive and transmit buffers. A simulation-based study of energy savings benefits is reported. We compare a network system with and without PBD. The effects of various traffic loads and network congestion levels were examined as well. We show that PBD can achieve up to 52% energy savings compared to standard network controllers, and on average it achieves energy savings of 29% across various traffic loads and 19% across various congestion levels.

References

[1] B. Nordman, “Networks, Energy, and Energy Efficiency,” Cisco Green Research Symposium, March 2008.

[2] C. Van Atten,A. Saha, I. Liberman and L. Reynolds,“Benchmarking air emissions of the 100 largest electric power producers in the United States.” Ceres, Constellation Energy, Entergy, the Natural Resources Defense Council (NRDC), Public Service Enterprise Group (PSEG), 2010.

[3] J. Howard, S. Dighe et al., "A 48-Core IA-32 Message-Passing Processor with DVFS in 45nm CMOS", Proc. Int. Solid-State Circuits Conference (ISSCC), 2010.

[4] P. Salihundam, S. Jain et al., “A 2 Tb/s 6x4 mesh network for a Single-chip Cloud Computer with DVFS in 45 nm CMOS,” IEEE J. Solid-StateCircuits, vol. 46, pp. 757–766, 2011.

[5] E. Beigne, F. Clermidy et al., “An Asynchronous Power Aware and Adaptive NoC based Circuit”, Proc. IEEE Symp. VLSI Circuits, pp 190-191, 2008.

[6] W. Kim, M. Gupta, G.-Y. Wei, and D. Brooks, "System level analysis of fast, per-core DVFS using on-chip switching regulators," Int. Symp. High-Performance Computer Architecture (HPCA), 2008.

[7] J. Park, D. Shin, N. Chang and M. Pedram, "Accurate modeling and calculation of delay and energy overheads of dynamic voltage scaling in modern high-performance microprocessors " Proc. 16th ACM/IEEE Int. Symp.Low Power Electronics and Design (ISLPED), pp. 419-424, 2010.

[8] K. Christensen, P. Reviriego et al., “IEEE 802.3az: The Road to Energy Efficient Ethernet,” IEEE Commun. Mag., 2010.