

# Using Coordinated Transmission with Energy Efficient Ethernet

Pedro Reviriego, Ken Christensen, Alfonso Sánchez-Macián, Juan Maestro

► **To cite this version:**

Pedro Reviriego, Ken Christensen, Alfonso Sánchez-Macián, Juan Maestro. Using Coordinated Transmission with Energy Efficient Ethernet. Jordi Domingo-Pascual; Pietro Manzoni; Sergio Palazzo; Ana Pont; Caterina Scoglio. 10th IFIP Networking Conference (NETWORKING), May 2011, Valencia, Spain. Springer, Lecture Notes in Computer Science, LNCS-6640 (Part I), pp.160-171, 2011, NETWORKING 2011. <10.1007/978-3-642-20757-0\_13>. <hal-01583421>

**HAL Id: hal-01583421**

**<https://hal.inria.fr/hal-01583421>**

Submitted on 7 Sep 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Using Coordinated Transmission with Energy Efficient Ethernet

Pedro Reviriego<sup>1</sup>, Ken Christensen<sup>2</sup> Alfonso Sánchez-Macián<sup>1</sup> and Juan Antonio Maestro<sup>1</sup>

<sup>1</sup> Universidad Antonio de Nebrija, Calle Pirineos 55, 28040 Madrid, Spain  
{previrie, asanche, jmaestro}@nebrija.es

<sup>2</sup> University of South Florida, 4202 East Fowler Avenue, ENB 118, Tampa, FL 33620, USA  
christen@csee.usf.edu

**Abstract.** IEEE 802.3az Energy Efficient Ethernet (EEE) supports link active and sleep (idle) modes as a means of reducing the energy consumption of lightly utilized Ethernet links. A link wakes-up when an interface has packets to send and returns to idle when there are no packets. In this paper, we show how Coordinated Transmission (CT) in a 10GBASE-T link can allow for key physical layer (PHY) components to be shutdown to further reduce Ethernet energy consumption and enable longer cable lengths. CT is estimated to enable an additional 25% energy savings with a trade-off of an added frame latency of up to 40  $\mu$ s, which is expected to have a negligible impact on most applications. The effective link capacity is approximately 4 Gb/s for symmetric traffic and close to 7 Gb/s for asymmetric traffic. This can be sufficient in many situations. Additionally a mechanism to switch to the normal full-duplex mode is proposed to allow for full link capacity when needed while retaining the additional energy savings when the link load is low.

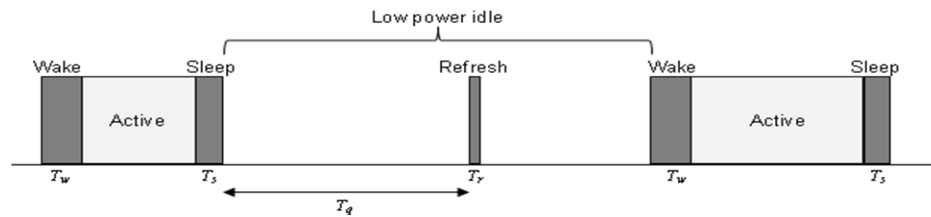
**Keywords:** Ethernet, Energy Efficiency, IEEE 802.3az, 10GBASE-T.

## 1 Introduction

The reduction of energy consumption in computer networks has been a topic of interest for the networking community in the last decade [1]. This has led to a number of initiatives aimed at making networks more energy efficient. Among the different networking technologies, Ethernet was identified as one of the candidates where large savings could be obtained. The potential savings are in the range of many TWh [2]. To materialize those savings the Energy Efficient Ethernet (EEE) initiative was launched to standardize mechanisms that would improve the energy efficiency of Ethernet devices. These efforts have led to the recent approval of the IEEE 802.3az Energy Efficient Ethernet (EEE) standard [3].

The EEE standard defines two modes of operation for a physical layer (PHY) device, active and sleep (aka Low Power Idle (LPI)). When there are frames to transmit the device is in the active mode and when there are no pending frames the device can enter the sleep mode reducing its energy consumption substantially. If

frames arrive while the device is in the sleep mode a transition back into the active mode is initiated. The transitions to wake and sleep a device take place in a small time (in the range of microseconds). This minimizes the delay that a frame experiences while it waits for the link to activate. Finally in the sleep mode periodic refresh periods are scheduled during which signals are transmitted to keep the receiver adapted to the current channel's conditions and synchronized with the transmitter. The operation of EEE is illustrated in Fig. 1.



**Fig. 1.** Transitions between the active and low-power modes in Energy Efficient Ethernet.  $T_w$  wake up time,  $T_s$  sleep time,  $T_r$  refresh time and  $T_q$  quiet time.

The current EEE standard covers the most common types of Ethernet PHYs including 100BASE-TX, 1000BASE-T, and 10GBASE-T. These are commonly used in PCs and servers and use Unshielded Twisted Pairs (UTP) as the transmission media. An important point to note is that the PHY power consumption increases significantly with link speed. For example the consumption of a 1000BASE-T PHY can be around 0.5 Watts while for a 10GBASE-T the consumption can be around 4 Watts [4]. Most Network Interface Cards (NICs) support several speeds and when the link is setup the maximum speed supported by both ends is selected [5]. This means that if we were to optimize the energy consumption for one of the speeds, the highest one should be a priority as it is likely to be selected and it is the one that consumes more. For UTP PHYs this speed would be 10 Gb/s and corresponds to the 10GBASE-T standard. 10GBASE-T PHYs are currently used mostly in servers but as in previous UTP PHY generations as the technology matures it is expected that they will be widely adopted in desktop and laptop computers (see [6] for more details). This transition will be accelerated by EEE that significantly reduces power consumption, making it easier to integrate 10GBASE-T PHYs in laptops and other battery run computers. Therefore given its high power consumption and its expected wide adoption in coming years, in the rest of the paper the analysis focuses on 10GBASE-T links.

The 802.3an 10GBASE-T standard [7] supports full duplex communication at 10 Gb/s over up to 100m of UTP cable. Transmission takes place over four cable pairs in parallel. This means that in each pair in addition to the signal from the remote transmitter, echo and both Near End crosstalk (NEXT) and Far End crosstalk (FEXT) signals are received. All those interferences must be cancelled in order to properly receive the remote signal. To that end 10GBASE-T PHYs use advanced signal processing techniques and therefore the circuitry devoted to echo and NEXT cancellation is a significant part of the PHY area and power consumption [8].

EEE has standardized a sleep mode for 10GBASE-T that enables the independent management of each link direction. That is a link can be in sleep mode in one direction while it is actively sending data in the other direction. This is a useful

feature as in many cases the traffic on the links is asymmetric. The values of the transition time for wake ( $T_w$ ), sleep ( $T_s$ ) and also the duration of the refresh ( $T_r$ ) and quiet ( $T_q$ ) periods are shown in Table 1. It can be observed that the duration of the quiet period is much larger than that of the refresh period to maximize energy savings. The transition times are small in terms of their impact on frame delay but large when compared to the frame transmission time. For example a 1500 byte frame takes 1.2  $\mu$ s to transmit at 10 Gb/s which is significantly less than the wake or sleep transition times. This can lead to large energy overheads due to mode transition when the link is awoken as soon as a frame arrives [9].

**Table 1.** EEE parameters for 10GBASE-T.

$T_w$	$T_s$	$T_q$	$T_r$
4.48 $\mu$ s	2.88 $\mu$ s	39.68 $\mu$ s	1.28 $\mu$ s

To minimize the overheads associated with transitions, coalescing of packets can be used such that the link is activated only when a group of frames is ready for transmission. This minimizes the number of mode changes and can significantly reduce energy consumption as discussed in [2].

In this paper, a Coordinated Transmission (CT) scheme is proposed to further reduce the energy consumption of 10GBASE-T PHYs that implement the EEE standard. The objective is to coordinate transmissions at both ends of the links such that a) at a given point in time only one end is transmitting and b) transmissions can only start at some predetermined points in time.

Objective (a) enables the PHYs to save significant energy as the echo and NEXT cancellation circuits could be put into a low power mode. Additional savings can also be obtained in other PHY elements. An additional benefit is that longer cables could be used as there are less interferences in the channel. Objective (b) enables larger savings beyond the PHY as other system elements can be put in low power mode even if their wake up time is larger than that of the PHY.

The proposed CT scheme reuses EEE signalling and therefore minimizes the additional complexity needed to implement it. It also provides mechanisms to signal transitions between CT and normal mode which will be referred to as Uncoordinated Transmission (UCT) mode in the rest of the paper. This enables a link to operate in the traditional UCT mode if the load in both directions is high.

In the rest of the paper the Coordinated Transmission (CT) is described (section 2). Its benefits in terms of energy savings are discussed in section 3. Then performance simulation results of the proposed scheme are reported in section 4. The paper ends with some conclusions in section 5.

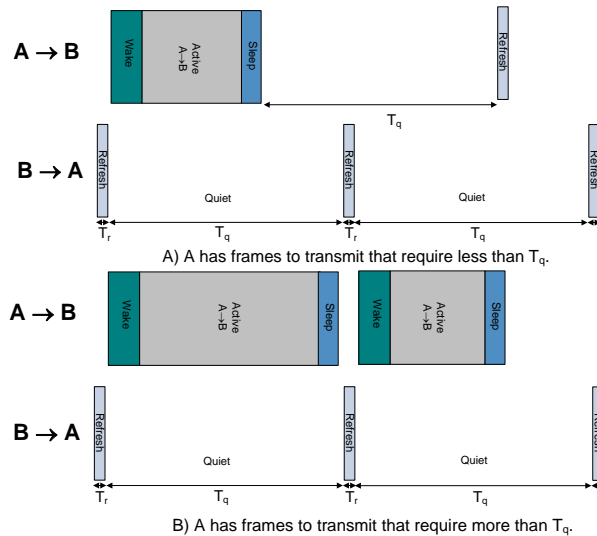
## 2 Coordinated Transmission (CT)

In this section, first the CT scheme is introduced and described in detail. Also some examples are used to illustrate how the scheme operates. Then, a mechanism to switch between CT and UCT modes is discussed.

## 2.1 Coordinated Transmission Scheme

As described in the introduction the objectives of the CT scheme are a) to ensure that only one link end transmits at a given point in time and b) to provide predetermined time instants at which transmission can take place.

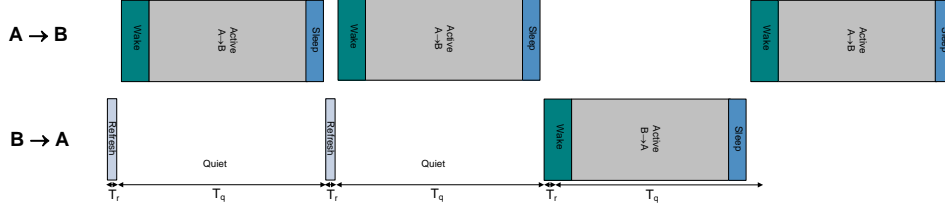
This is achieved by using the EEE refresh periods to coordinate transmissions. The refresh periods are needed to ensure that the receiver is kept aligned to the channel conditions and with the transmitter. Therefore the refresh signal must be, in any case, transmitted each  $40.96 \mu\text{s}$  as per the parameters in Table 1. This means that one link end can at most transmit for  $40.96 \mu\text{s}$  as otherwise its transmission would overlap with the refresh signal in the other direction and objective a) will not be achieved. This leads naturally to using the EEE refresh signal as a means to grant permission to transmit until the next refresh signal. That is a link end can start its transmission when it receives an EEE refresh signal from the other end. Since EEE refreshes are sent periodically a link end is guaranteed to receive a refresh signal each  $40.96 \mu\text{s}$  thus ensuring a low waiting time for transmission. Two examples of this scheme are illustrated in Fig. 2 where obviously the transmission has to be limited to avoid overlapping with the next refresh signal as discussed before. In the first example, link end A has frames to transmit that require less than  $T_q$  seconds. In the second, link end A has frames to transmit that require more than  $T_q$  seconds and therefore has to activate the link two times to send all the frames.



**Fig. 2.** Examples of transmissions with the proposed scheme.

The described scheme assumes that refresh periods occur frequently in both link directions. This is not true when one of the link ends has frames to transmit in every refresh period. In that case that link direction transitions from sleep to active and back to sleep in each period and therefore never spends sufficient time in the sleep mode to send a refresh signal. This issue can be solved by using also sleep transitions in the receive direction to grant permission to transmit in the same way as an EEE refresh

signal. In Fig. 3, the case in which link end A has data to transmit in every refresh period and then B also starts to transmit data after a sleep period is illustrated.



**Fig. 3.** Example of transmission with the proposed scheme showing alternating transmission between link ends.

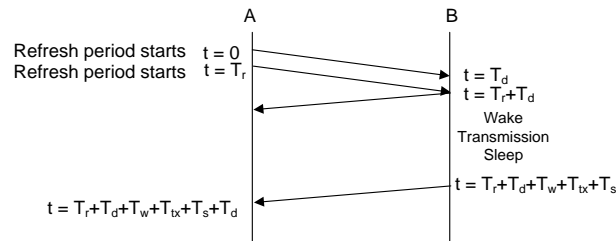
In the discussion so far, it has been implicitly assumed that the link delay is negligible and therefore the same signal is present at both link ends at the same time. If that is not the case the scheme will not work. For example if the link delay was  $50 \mu\text{s}$  then if one link end sends a refresh at time  $0 \mu\text{s}$  and the other at time  $20 \mu\text{s}$  those would arrive at  $50 \mu\text{s}$  and  $70 \mu\text{s}$  respectively enabling transmissions that can overlap. The delay of 100m of UTP cable is approximately  $0.5 \mu\text{s}$  which is significantly smaller than the refresh period or the transition times. Therefore the scheme will work but care should be taken to account for this extra delay. This implies reducing the transmission time by another  $1 \mu\text{s}$  to avoid overlapping at the remote end with the start of the next refresh signal in the other direction. This is illustrated in Fig. 4. Which shows the detailed timing at both link ends for a complete transmission period starting after a refresh is received at B.  $T_d$  corresponds to the link propagation delay and  $T_{tx}$  to the actual data transmission time. To avoid the next refresh overlapping with the end of the transmission the following inequality must be met

$$T_r + T_d + T_w + T_{tx} + T_s + T_d < T_r + T_q \quad (1)$$

which can be solved in terms of  $T_{tx}$  as

$$T_{tx} < T_q - T_d - T_w - T_s - T_d. \quad (2)$$

Using the values in Table 1 and assuming the worst case link delay of  $0.5 \mu\text{s}$  we obtain a value for the data transition time ( $T_{tx}$ ) of up to  $31.82 \mu\text{s}$ . In a practical implementation there will be a short time in between the end of the reception of the refresh or sleep signal and the start of the transmission and there should also be some margin in between the end of the transmission and the next refresh period. If we allow for an additional  $0.25 \mu\text{s}$  for each, the transmission time would be up to  $31.32 \mu\text{s}$ , that would enable the transmission of twenty six 1500 byte frames. This means that roughly 75% of the period can be used for actual data transmission. The performance of the proposed scheme will be further discussed in section 4.



**Fig. 4.** Detailed transmission time diagram for the proposed scheme.

## 2.2 Signaling Mechanisms for Coordinated Transmission

In the previous subsection the CT scheme was described assuming that both link ends support CT and have somehow agreed to use it. Now the mechanisms that allow link ends to signal support for CT and to switch between CT and UCT modes are described.

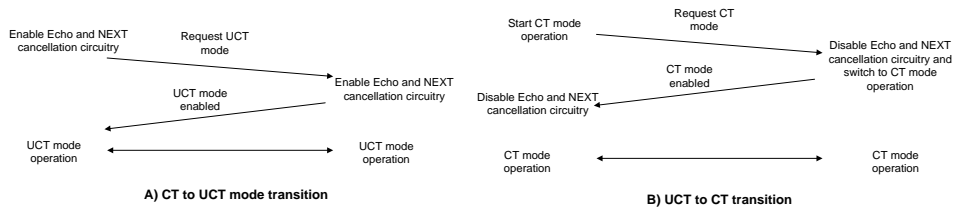
To better understand the mechanisms let us first review what exactly they are needed for. When the link is setup we need to convey whether a link end supports CT. This would enable link ends to know if the remote end also supports CT and therefore if it could be used. Then during link operation we need a mechanism that allows to switch between CT and UCT modes. This is needed as CT effectively reduces the link capacity such that when the traffic load is high we may want to switch to UCT even if the energy consumption is larger. As the pattern of use of networks is such that the utilization is low most of the time [10] most links would only need to switch to UCT during a small fraction of time. There is one case where this second mechanism is not needed, that is when CT is used to extend the reach of 10GBASE-T. In that case the link would always operate in CT mode.

The Ethernet standard defines a signalling procedure called Auto-Negotiation that is used when the link is setup to exchange the capabilities of both link ends [11]. During this phase the information is encoded in link pulses so that the exchange can take place even though the link has not been established. Auto-Negotiation is used for example to agree on the speed to use if multiple speeds are supported as discussed before. The support of the CT mode can be easily added to the information exchanged during Auto-negotiation by using a bit in the 10GBASE-T Auto-Negotiation control register defined in the standard. Currently only one bit is used in that register to signal support for fast retrain in 10GBASE-T. Therefore the use of Auto-negotiation is proposed to signal support for the CT mode at link setup.

A 10 Gb/s link in which only one link end can transmit a given time instant can at most carry 10 Gb/s in both directions compared to 20 Gb/s when operating in UCT mode. This means that when the link traffic load is high we should be able to switch to UCT mode operation. This requires coordination between both ends of the link. To ensure that there are no overlapping transmissions during a mode change the scheme shown in Fig. 5 could be used. To switch to UCT mode, the requesting end first enables the echo/NEXT cancellation circuitry and then sends the request. The other end upon reception enables the cancellation circuitry and starts operating in UCT mode. Also a message is sent back to the other end that upon its reception enables the UCT mode of operation. To go back to CT mode the requesting end starts to schedule its sending as if it were already in CT mode and sends a request message to the other end. Upon its reception the other end disables the cancellation circuitry and also starts to schedule sending as in CT mode. Also a message is sent back to the other end that upon its reception can disable the cancellation circuitry. In the transition from UCT to CT we need to be careful to avoid overlapping transmissions and to ensure that the refresh periods in both directions do not take place at the same time. To ensure that, the end that received the request should stop transmitting and go to the sleep mode when it receives the request. Then wait for the other end to either a) stop transmitting or b) send a refresh. At that point wake and start CT operation by sending a message back to the other end confirming the transition to CT mode and any further frames

waiting for transmission. This will ensure that the refresh periods are not aligned when entering the CT mode as assumed in the previous subsection.

The message exchange required for the proposed mode switching mechanism can be done above the PHY layer using Ethernet data frames but since the mechanism is heavily related to the PHY it would be better to perform the exchange using PHY layer mechanisms. This can be done using the auxiliary bit defined in the 10GBASE-T standard [7]. This bit provides a channel of approximately 3Mb/s for signalling at the PHY layer so that mode transitions could be done in a short time. The use of this channel for signalling of different advanced PHY features has been proposed by some PHY vendors [12], [13]. In fact since the channel is fully devoted to control signalling, the status of each link ends in terms of the mode of operation (CT or UCT) could be periodically transmitted over this channel to ensure that even in the event of data lost a consistent state would be reached.



**Fig. 5.** Message exchanges to change mode of operation.

Finally it is worth noticing that the link remains operational during transitions so the effect on frame delay should be small. The decision of when to switch modes can be based on buffer occupancy or on measured load on the link. The study of those algorithms is left for future work.

### 3 Benefits of Coordinated Transmission

The use of CT provides benefits in terms of additional energy savings and also in extended reach, those are discussed in the rest of this section.

#### 3.1 Energy Savings

The use of CT would only be of interest if it can provide substantial energy savings. To understand the potential savings a basic understanding of how 10GBASE-T works is needed. In 10GBASE-T full duplex transmission takes place over four Unshielded Twisted Pairs (UTP) that are bundled in the same cable. This simultaneous transmission causes that in addition to the remote signal, echo and both near end and far end crosstalk (NEXT and FEXT) are received on each pair. The receiver must remove all those unwanted signals and for that echo and crosstalk cancellation circuits are used.

In Fig. 6 a block diagram of the circuitry used in each of the four pairs in a 10GBASE-T transceiver [8] is shown. In CT mode the Echo Cancellers and the



NEXT cancellers are not needed since there is no transmission when the device is receiving. These elements are an important part of a 10GBASE-T PHY [14]. Echo cancellation is typically done both in the analog domain to reduce the requirements on the Analog to Digital Converter (ADC) and then in the digital domain to eliminate the rest of the echo [8]. This means that in CT mode energy savings would be obtained both in the analog and digital circuitry. NEXT cancellation is typically done in the digital domain. Additional savings maybe obtained in the ADCs as less precision is needed if there is no echo at the input.

Therefore if a PHY operates in CT mode all the circuitry related to echo and NEXT cancellation can be put in low power mode. The energy savings that could be obtained will vary for different implementations but they will be substantial. From the information in [8], savings of 20-30% seem a reasonable estimate.

Additional benefits can be obtained in other PHY elements. For example, since there is no echo, the requirements of the timing recovery are relaxed. This means that some extra power could be saved on the timing recovery circuitry and/or that transitions could be made faster. The main point is that during wake up the timing recovery circuitry must compensate the deviation between the clocks that occurred during the time that the PHY was in low power mode. This is much harder to do if there is echo in addition to the remote signal. Also if echo is present, the adjustment must be done slowly so that the local transmission is not disturbed by the timing recovery and the echo is properly cancelled. Another element that could benefit from the CT mode is the Low Density Parity Check (LDPC) code decoder that could possibly be switched to a simplified mode to save power as the Signal to Noise Ratio (SNR) will probably be larger than needed. This may be easier to do than the reduction of precision in the ADC.

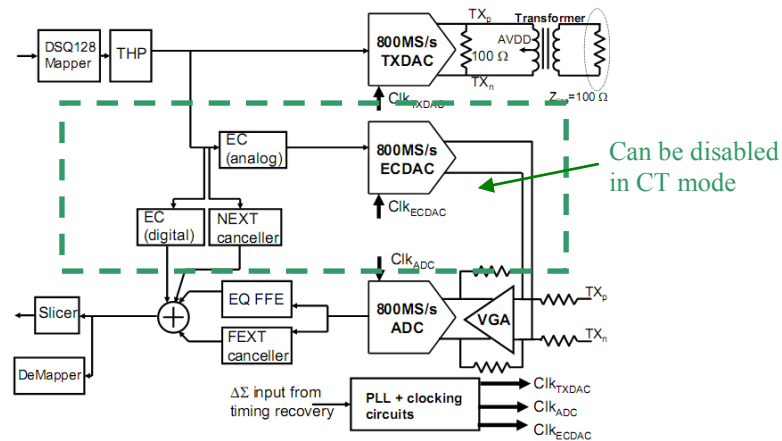


Fig. 6. 10GBASE-T circuitry for each twisted pair (from [8]).

It is expected that since link utilization is typically low, in many cases the IEEE-PHYs will stay in the LPI mode most of the time. Therefore it is interesting to also consider the savings that CT enables in LPI mode. When we are in LPI mode the proposed coordination transmission mechanism enables us to know when the other

end may start transmitting. This in turn enables us to power down all the receiver circuitry as we are sure that no alert/wake signal will be received. Without CT, some receiver elements must be active at all times to detect the alert/wake signal from the other end as that can start at any point in time. Additionally during the refresh periods some of the energy savings of CT described for the active mode also apply. This means that additional significant savings can be achieved in LPI idle mode when CT is used. A first estimate for the savings in LPI mode can also be in the order of 20% of the LPI mode power consumption.

Finally, as CT enables a link end to know in advance when the other end can start transmitting, additional savings beyond the PHY could be achieved [15]. For example if link end A schedules a transmission for a complete interval between refresh periods (approximately 40  $\mu$ s), then no frames will be received during that time. This would enable components with transition times larger than the PHY to enter the low power mode and still be awake on time to process the data. This could also be done using extended wake timers in the PHY as proposed in [16] but that would increase the PHY power consumption as each EEE mode transition will involve a larger overhead.

### **3.2 Extended Reach**

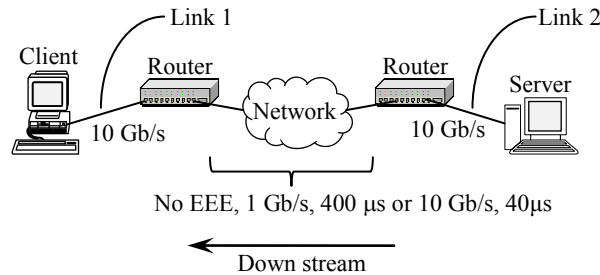
Another interesting option is to use Coordinated Transmission to extend the reach of a 10GBASE-T link beyond the standard 100 m. The extended reach can be obtained as there is no echo at the receiver such that the signal can be further amplified before the ADC resulting in a better SNR. Also we get rid of the quantization noise from the cancellers and any residual echo/NEXT after cancellation. The number of meters that could be extended will depend on the PHY architecture and is difficult to predict; this requires further work. In this case obviously transitions to UCT mode will not be allowed and the PHYs will always operate in CT mode. This option would provide a capacity in between 1 Gb/s and 10 Gb/s as CT transmission uses some of the capacity for coordination. For symmetric traffic the effective capacity of CT would be approximately 4 Gb/s as discussed in the next section.

## **4 Performance Evaluation**

As mentioned in the introduction, CT introduces an additional frame delay. This is the time that a link end has to wait until it receives a refresh from the other end or the other end stops its transmission if it was transmitting. In the worst case the link end would have to wait for the time between two refresh periods that is approximately 40  $\mu$ s. This delay is very small compared with the typical delay of Internet connections or with the requirements of multimedia applications such as voice over IP, both in the order of tens of ms. This means that for most users the impact of the added delay will be negligible.

The proposed scheme also reduces link capacity. To evaluate this effect, CT has been implemented in the ns-2 simulator and tested in several experiments. The network configurations used are shown in Fig. 7. The link delays for Link 1 and Link 2 are 0.5  $\mu$ s that corresponds to the propagation delay of approximately 100m of UTP

cable. The intermediate link, models a network that limits the transfer capacity in the 1 Gb/s case and does not in the 10 Gb/s case. The first configuration tries to model a situation in which the bandwidth is limited by the Internet connection having a lower speed than the LAN as it commonly occurs.



**Fig. 7.** Configurations for the evaluation experiments.

In the first set of experiments a large FTP transfer takes place between the client and the server. The traffic in this case is highly asymmetric with large data frames in one direction and small acknowledgement frames (ACKs) in the other. Therefore the link utilization should be high in the data direction and much smaller in the ACK direction. The results in terms of link utilization are shown in Table 2 (FTP in one direction). When the middle link operates at 10 Gb/s, the link utilization in the data direction is approximately 66% of its capacity and the use of the UCT mode could improve performance by getting close to 100% utilization. Therefore this would be a case in which switching to UCT mode using the proposed mechanism could be desirable to reduce the transfer time for large files. When the middle link operates at 1 Gb/s and a large FTP transfer takes place between the client and the server, the results show how the CT mode is capable of performing the FTP at the same speed as a UCT link (1 Gb/s or 10% utilization). In this case the UCT mode will not give a performance advantage and energy can be saved by using the CT mode. This would be the case when the transfer speed is limited by the network but also when it is limited by the client or server. As discussed in [17] in many cases the transfer speed may be limited by the processor, the operating system, and/or the application software such that the full link speed can not be achieved in any case.

In the second set of experiments a one way 1 Gb/s Constant Bit Rate source is used. The results show that the proposed scheme is able to carry this traffic with no performance lost when the middle link operates at 1 Gb/s and 10 Gb/s. In this case the transmissions are always triggered by the refresh periods in the other direction. This CBR example models the transmission of video or other applications that can only send data as it is generated in the remote end and therefore cannot take advantage of the full link speed.

Finally, the FTP experiments were re-run with simultaneous FTP transfers in both directions. The results show that when the middle link operates at 10 Gb/s the total link utilization is higher than for the case of one direction only. In this case CT reaches a total utilization of close to 8 Gb/s (4 Gb/s in each direction) compared to around 7 Gb/s in the asymmetrical case. This is so because in this case each time the

link is activated transmission takes place for the maximum allowed time in CT (31.2  $\mu$ s). Therefore the overhead of EEE mode transition is lower than in the asymmetric case in which in one direction the link is activated to send ACKs that require less than 2  $\mu$ s to be sent. For this configuration the use of UCT transmission would increase the performance and therefore a switch to the UCT mode could be done to reduce transfer time. When the middle link operates at 1 Gb/s, the CT mode is sufficient to perform the FTP file transfers with no performance degradation.

The experiments conducted have served as an initial validation of the proposed scheme. The results show how the CT mode operation can achieve a total link utilization of close to 7 Gb/s for asymmetric traffic and approximately 8 Gb/s (4 Gb/s in each direction) for symmetric traffic. This would be sufficient in many environments where speed is limited by the network connection or by the system hardware and/or software. The same applies to applications that do not require or cannot use the full link speed, like for example multimedia applications. In any case a fast mechanism to switch to UCT mode is provided that enables the use of the complete link capacity when required.

**Table 2.** Results from ns-2 experiments.

	Config (10 Gb/s, 40 $\mu$ s delay)		Config (1 Gb/s, 400 $\mu$ s delay)	
	Link Utilization		Link Utilization	
	downstream	upstream	downstream	upstream
FTP in one direction	66%	3.3%	10%	0.5%
CBR source	10%	0%	10%	0%
FTPs in both directions	38.9%	39.1%	10%	10%

## 5 Conclusions

In this paper a mechanism to achieve additional savings in 10GBASE-T links that implement the IEEE 802.3az Energy Efficient Ethernet (EEE) standard has been proposed. The mechanism coordinates transmission over the link such that at any given time instant only one link end is transmitting. To do so existing EEE signaling is used. The proposed scheme enables large additional savings in the range of 20-30% in the Physical layer devices (PHYs) as many components such as the echo cancellation circuitry can be put in low power mode. This savings are trade off for additional link delay and reduced link capacity. In the worst case, the proposed approach adds 40  $\mu$ s of delay for frame transmission that is negligible compared to the typical Internet connection delays and the requirements of most end users applications. The effective capacity of the link is close to 7 Gb/s for asymmetric traffic and approximately 4 Gb/s for symmetric traffic which would be sufficient when the network connection or the computer performance limit the speed. For cases in which the full link speed is needed sporadically, a fast mechanism to switch to normal full-duplex (or UCT) transmission mode has also been described. In this way most of the energy savings would be retained while minimizing the impact on performance.

## References

1. Gupta, M and Singh, S.: Greening of the Internet Proc. of ACM SIGCOMM pp 19-26, Karlsruhe, Germany, August 2003.
2. Christensen, K., Reviriego, P., Nordman, B., Bennett, M., Mostowfi, M., and Maestro, J. A.: IEEE 802.3az: The Road to Energy Efficient Ethernet, IEEE Communications Magazine, vol. 48, issue. 11, November 2010, pp: 50 – 56.
3. IEEE Std 802.3az: Energy Efficient Ethernet-2010.
4. Aquantia to Deliver World's First Quad 10GBASE-T Solution, press release, May 2009.
5. Gunaratne, C., Christensen, K., Nordman, B., and Suen, S.: Reducing the Energy Consumption of Ethernet with Adaptive Link Rate (ALR), IEEE Transactions on Computers, vol. 57, issue 4, April 2008 pp.: 448 – 461.
6. Van Babel, N., Done, D., Flatman A., and McConnell, M.: Short Haul 10Gbps Ethernet Copper PHY Call for Interest, IEEE 802.3 November 2005, (available online in [http://grouper.ieee.org/groups/802/3/cfi/1105\\_1/10G\\_Short\\_Haul\\_CFI.pdf](http://grouper.ieee.org/groups/802/3/cfi/1105_1/10G_Short_Haul_CFI.pdf)).
7. IEEE Std 802.3an: 10GBASE-T-2006.
8. Gupta, S., Tellado, J., Begur, S., Yang, F., Balan, V., Inerfield, M., Dabiri, D., Dring, J., Goel, S., Muthukumaraswamy, K., McCarthy, F., Golden G., Jiangfeng, W.; Arno, S., and Kasturia, S.: A 10 Gb/s IEEE 802.3an-Compliant Ethernet Transceiver for 100m UTP Cable in 0.13 $\mu$ m CMOS, IEEE International Solid-State Circuits Conference, 2008. pp. 106-599.
9. Reviriego, P., Hernández, J.A., Larrabeiti D., and Maestro, J.A.: Performance Evaluation of Energy Efficient Ethernet, IEEE Communication Letters, vol 13 no 9, September 2009, pp 697-699.
10. Odlyzko, A.: Data Networks are Lightly Utilized, and will Stay that Way, Review of Network Economics, Vol. 2, No. 3, pp. 210-237, September 2003.
11. IEEE 802.3 Carrier Sense Multiple Access with Collision Detection (CSMA/CD) access method and physical layer specifications, 2008.
12. Diab, W.: Method And System For Asymmetric Transition Handshake In An Energy Efficient Ethernet Network US Patent Application 2009/0154467, Filed on October 2, 2008, Published on June 18, 2009.
13. Lee H. W. and McConnell S. M.: 10GBASE-T link speed arbitration for 30m transceivers, US Patent Application 2008/0219289, Filed on Nov 13, 2007, Published on Sep 11, 2008.
14. Chen, J. and Parhi K.K.: Further cost reduction of adaptive echo and next cancellers for high-speed Ethernet transceivers, IEEE Workshop on Signal Processing Systems (SiPS) 2008. pp. 227-232.4.
15. Dove, D.: Energy Efficient Ethernet: A Switching Perspective, presentation at the IEEE 802.3az May 2008 meeting, (available online in [http://www.ieee802.org/3/az/public/may08/dove\\_02\\_05\\_08.pdf](http://www.ieee802.org/3/az/public/may08/dove_02_05_08.pdf)).
16. Wael, D.: LLDP use in EEE, presentation at the IEEE 802.3az January 2009 meeting, (available online in [http://www.ieee802.org/3/az/public/jan09/diab\\_02\\_0109.pdf](http://www.ieee802.org/3/az/public/jan09/diab_02_0109.pdf)).
17. Bencivenni, M, Carbone, A., Fella, A., Galli, D., Marconi U., Peco, G., Perazzini, S., Vagnoni, V., and Zani S.: High Rate Packet Transmission on 10 Gbit/s Ethernet LAN using Commodity Hardware, Proceedings of the 16th IEEE-NPSS Real Time Conference, 2009.