System-level analysis of the tradeoffs between power saving and capacity/QoS with DRX in LTE

Giovanni Stea, Antonio Virdis

Abstract— **In an LTE cell, Discontinuous Reception (DRX) allows the central base station to configure User Equipment for periodic wake/sleep cycles, so as to save energy. Several parameters are associated to DRX operations, thus allowing for optimal performance with different traffic profiles (i.e., CBR-like, bursty, periodic arrivals of variable-sized packets, etc.). This work investigates how to configure these parameters and explores the tradeoff between power saving, on one side, and per-user QoS and cell capacity, on the other. Unlike previous work, mostly based on analytical models neglecting key aspects of LTE, our evaluation is carried out using a fully-fledged packet simulator. This allows us to discover previously unknown relationships and to propose configuration guidelines for operators.**

Index Terms—**LTE, DRX, Resource Allocation, Quality of Service, Power Saving, Simulation**

I. INTRODUCTION

THE Long-Term Evolution (LTE) of the UMTS promises ubiquitous, high-speed Internet access. In such systems, a central base station or eNodeB (eNB) shares radio resources among a number of User Equipment (UEs), i.e. handheld devices, laptops or home gateways. Handheld devices are normally battery-powered, hence care must be taken not to waste energy. On the network side, this objective can be aided by properly configuring *Discontinuous Reception* (DRX), which allows UEs to power off the reception/transmission circuitry periodically, waking up for short periods at specific instants. The underlying rationale is that packet transmission/reception is hardly ever continuous over time, hence synchronizing it with wake-up periods is likely to achieve significant energy savings with only a moderate latency increase. The UE DRX is configured by the eNB semi-statically, by tuning several parameters: the *cycle length* and the *on* window and offset within the former; the *inactivity timer*, which prolongs the *on* duration when a packet arrives, thus coping with bursty arrivals; the *short vs. long cycle*, which allows an UE to power down for several short intervals and check for new packets before going to sleep for longer times. These parameters can only be varied with a signaling procedure that takes hundreds of milliseconds, hence cannot follow short-term traffic variations. A more dynamic feature of DRX is instead the *sleep* control message, by which the eNB can send UEs to sleep until their next scheduled wake-up time.

A large number of papers have recently evaluated the performance of DRX under various conditions ([6]-[23]). Most of these studies rely on oversimplified *analytical* models, which unavoidably neglect all the key characteristics of an LTE network. Those few who approach the problem using accurate LTE simulators (e.g., [22]), instead, limit their study to the downlink or to simplified scenarios. This leads to inaccurate conclusions, understating important features of DRX, and generally losing insight on the relationship between network configuration and performance, which is what operators want to know.

While it is fairly obvious that power saving increases latency for the UEs, thus affecting QoS, and – by reducing multiuser diversity (only a subset of UEs is active at any time, in fact) – also affects cell capacity, these relationship depend on a multitude of factors: the traffic profile and requirements, the scheduling employed at the eNB, and – last but not least – the way DRX is configured for each UE. In this work, we tackle the problem of exploring the tradeoff between energy consumption, on one side, and QoS and capacity, on the other, in a network cell employing DRX. We carry out this study via simulation, using a fully-fledged C++ simulator which includes all the layers and functions of LTE, application modeling, and relevant QoS and Quality of Experience (QoE) metrics. We study DRX configuration for several applications: symmetric (VoIP), asymmetric (web browsing) and downlink-only (streaming video).

The rest of the paper is organized as follows: in Section II we report the necessary background on LTE-Advanced and the DRX standards. Section III reports an overview of the related work. We describe our simulation settings in Section IV, and report an extensive performance evaluation in Section V. Finally, Section VI reports conclusive remarks.

II. BACKGROUND ON LTE

Hereafter we describe those aspects of the LTE system which are more relevant to the resource allocation problem in both the downlink and uplink directions. A table of LTE-related acronyms is reported in the Appendix for ease of reference.

In LTE, PDU transmissions are arranged in frames called Transmission Time Intervals, (TTIs), whose duration is 1ms. In the downlink, the eNB allocates a vector of *Resource Blocks* (RBs) to the UEs associated to it on each TTI, by broadcasting the RB allocation map in the Physical Downlink

---

Giovanni Stea and Antonio Virdis are with the Dipartimento di Ingegneria dell'Informazione of the University of Pisa, Largo Lucio Lazzarino 1, 56122 Pisa, Italy (e-mail: g.stea@iet.unipi.it, a.virdis@iet.unipi.it).

Control Channel (PDCCH). Each RB carries a fixed number of symbols, which translate to different amounts of bits depending on the modulation and coding scheme used by the UE. In general, UEs favor more information-dense modulations (e.g., up to 64QAM, which yields 6 bits per symbol) when they perceive a better channel to the eNodeB. The quality of the wireless channel is time-varying, hence UEs report their perceived channel state to the eNodeB as a Channel Quality Indicator (CQI). The latter is an index in a standard table, computed by the UE according to the measured Signal to Noise Ratio (SNR), and determines the modulation that the latter will use (hence, indirectly, the number of bytes per RB) as reported in Table 1. Transmissions are subject to errors, and are therefore protected by a Hybrid ARQ scheme, which allows a configurable number of retransmissions.

TABLE 1 – CQI TABLE.

CQI	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bytes	0	3	3	6	11	15	20	25	36	39	50	63	72	80	93	93

In the uplink, the UE notifies the eNB about its backlog state by issuing quantized *Buffer Status Reports* (*BSRs*). BSRs are transmitted (either alone or trailing a sequence of PDUs) *in band*, i.e. together with the data. Thus, they can only be sent i) when the UE is scheduled, and ii) if there is enough space to do so (a BSR can take up to 24 bits). Therefore, a mechanism is needed that allows a UE to signal a scheduling node that it has switched from empty to backlogged. UEs signal their service requests *out of band*, using a dedicated Random Access Procedure (RAC) and a backoff mechanism to arbitrate collisions. RAC requests are instead responded in-band, by scheduling the UE in a future TTI[1]. The handshake for uplink transmissions, shown in Figure 1, takes five messages: first the UE initiates a RAC request; then, the eNB responds by issuing a short grant, large enough for a BSR; the UE sends its BSR; the eNB sends a larger grant according to some scheduling policy, and finally the UE transmits its data. In some cases (e.g., when uplink traffic is predictable), the eNB may decide to dispense with the middle two interactions, and immediately issue a grant large enough to hold the BSR *and* one or more PDUs in response to the RAC request. This technique, called *bandwidth stealing*, is known to increase the uplink capacity and reduce the latency.

Semi-Persistent Scheduling (SPS, [4]) can be used for uplink transmissions of periodic, low-bandwidth traffic, e.g., VoIP. It consists in the eNB issuing periodic grants to the UEs, which can then transmit without the need for signaling or handshake in the pre-assigned TTIs. A periodic grant can be revoked *explicitly*, via a specific message, or *implicitly*, after the UE fails to exploit it for a given number of consecutive times. Note that, under SPS, the periodic grant also sets – once and for all – the *format* of the uplink transmission, thus preventing link adaptation. Hence, variations in the channel quality (which are unavoidable, especially in the long term) may increase the Block Error Rate (BLER) or force the eNB to overdimension the periodic grant, thus reducing the efficiency.

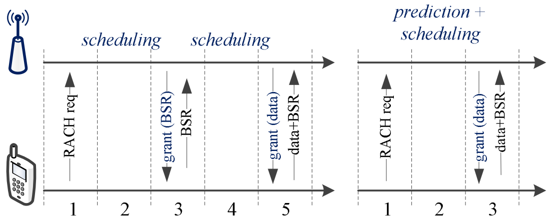

Figure 1 – Handshake for scheduling of uplink UE traffic: standard (left) and using Bandwidth Stealing (right).

A. *Discontinuous Reception (DRX)*

Under DRX[2], the UE periodically wakes up to monitor the PDCCH for a period of time, set by the *On Duration Timer* (ODT), in a cycle whose length and offset are called *DRX Cycle* (DC) and *DRX offset* (DO) respectively. If scheduled during its *on* phase, the UE stays awake until either the ODT expires, or another timer, called *Inactivity Timer* (IT), expires, whichever occurs last. The IT is re-scheduled on each correct reception, and its purpose is to delay the *sleep* phase so that a burst of packets at the end of an *on* phase can be received correctly. Note that the IT must be *at least* one TTI, and that it *prolongs* the duty cycle without altering the period, as shown in Figure 2.

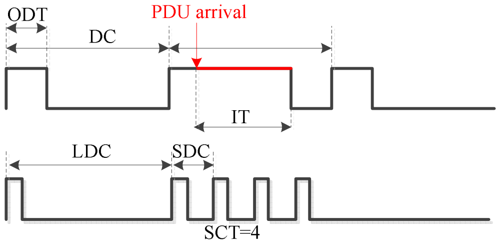

Figure 2 – Basic mechanisms for DRX. Inactivity timer (top) and long/short cycles (bottom)

Some traffic scenarios are characterized by periods of regular transmission, followed by periods of little or no activity (e.g. VoIP with Voice Activity Detection). To handle these cases two types of DRX Cycle are defined: *Long DRX Cycle* (LDC) and *Short DRX Cycle* (SDC). Normally the LDC is followed. When the UE is *on* and is scheduled for a new transmission, it switches to SDC, i.e. to shorter cycles, for a number of consecutive times, known as *Short Cycle Timer* (*SCT*). The SCT is reset each time the UE is scheduled, hence the UE returns to LDCs after receiving no packets for SCT×SDT TTIs. Finally, the LTE standard allows one to asynchronously turn off the UE. This is done via a *DRX-Command* MAC control element (*DCE*), i.e. a MAC header sent within a standard packet. The latter stops *both* the ODT and the IT, thus sending the UE to sleep until the next wake-up time. If short/long cycles are configured, the SCT is restarted and the SDC will be used for the next cycles.

---

[1] The standard also defines a *Dedicated Scheduling Request* mode, whereby UEs issue scheduling requests using in-band *dedicated* resources. DSR is increasingly inefficient as the number of UEs grows large, hence it is scarcely used in practice and will not be considered further in this work.

[2] The acronym DTX, which stands for *Discontinuous Transmission*, is sometimes used to refer to DRX in the uplink. In fact, there is only *one* mechanism in the standard, which affects both directions at the same time, and it goes by the name of DRX, which we will stick to henceforth.

All the above parameters are configurable through the Radio Resource Control (RRC) protocol. However, RRC signaling takes tens of TTIs and occupies downlink resources, which makes it unfeasible for short-term adjustments. In other words, it is not a task to be performed to cope with instantaneous queue variations, rather it should be employed at larger timescales (i.e., seconds or more).

III. RELATED WORK

The last few years have witnessed an increasing interest on DRX for LTE, as testified by a large number of papers appeared in (mostly) conferences and journals. Some of them propose DRX-based solutions, i.e. scheduling ([6]) or extensions for newer LTE deployments, e.g., Carrier Aggregation [7] or TTI-bundling [8], hence are only marginally related to the object of this paper. Works on DRX *evaluation*, instead, such as [9]-[23], deal with one or more of the following:

1. Modeling DRX using analytical techniques (e.g., Markov or Semi-Markov) ([9]-[16]);
2. proposing adaptive techniques for setting some DRX parameters (e.g., [17]-[19]);
3. evaluating the performance of VoIP or HTTP traffic under DRX (e.g., [17]-[23], [11]).

Most of the above works do not simulate an LTE system at all (e.g., [9]-[16]). Rather, they simulate *their own analytical models*, neglecting the LTE protocol stack, MAC-level fragmentation and reassembly, (almost always) H-ARQ, CQI reporting, link adaptation, resource contention through scheduling, random access for the uplink channel etc. Those few who take up scheduling (e.g., [10]) model contention as a probability distribution, without any validation, or consider single-slot systems. We claim that neglecting the above essential features of LTE leads to unreliable results, and we provide evidence to back up this claim in this paper. The only performance analyses (that we are aware of) carried out using a fully-fledged LTE simulator are works [20]-[22], whose shortcomings we describe later on.

Works that propose configuration of DRX parameters often neglect important features, mostly concentrating on long/short cycles: for instance, none deal with *de-synchronization* of UEs *on* phases through *DO* selection, which we will show to play a fundamental role in preserving capacity. Few investigate the *on duration*, which is instead fundamental for VoIP applications. None, finally, investigate using *DCE messages*, whose saving potential is indeed remarkable. Few works deal with assessing the impact of *cell load* on DRX configuration under credible conditions. MAC-level contention reduces the likelihood that a single UE is scheduled in its *on* phase. Therefore, DRX parameters (e.g., the *on* duration) should be set based on the cell load, if QoS is to be preserved. Providing guidelines for the best energy-QoS tradeoff at various cell loads for real-life applications is in fact the purpose of this paper.

Another diffuse shortcoming is the use of unrealistic traffic and application models. For instance, Poisson traffic, which hardly matches any real-life application, dominates the analyses (e.g.,[10],[13],[14]). With HTTP traffic, only the downlink leg is considered ([11], [18]-[19]), hence the delay and energy consumption are underestimated. Moreover, works assessing VoIP performance (e.g., [20]-[23]), even when they use realistic traffic profiles (e.g., including *Voice Activity Detection*, VAD), assume *zero jitter in the downlink*, and place perfectly periodic sources directly on the eNB. Instead, packets get to the eNB after traversing access and core networks, which do add jitter. Jitter, in turn, thwarts the periodic nature of the DRX cycles: the net effect, as we will see later on, is that *on* durations need to be increased to compensate for jitter. Thus, power saving and QoS results obtained under the zero-jitter hypothesis are inflated.

IV. MODELING ASSUMPTIONS AND SIMULATION SCENARIOS

Our evaluation is carried out using a system-level simulator, comprising more than 100k lines of object-oriented C++ code, which includes all the layers of the protocol stack, from the physical to the application layer. Protocol layers and functions are conform to the Release 8 standard. Each simulation run lasts for 200s, with a warm-up time of 20s where statistics are not collected. Hereafter, we describe the modeling of the eNodeB, of the UEs and of the application traffic.

A. eNodeB model

A single LTE cell is simulated, with an eNodeB equipped with an omnidirectional antenna at its center and a variable number of UEs experiencing varying channel conditions. The RLC layer at the eNodeB is configured with the *Unacknowledged Mode*, with a fixed PDU size of 40 bytes. The system bandwidth is set to 5Mhz in order to approach cell saturation with a manageable number of UEs (we expect full-spectrum simulations to yield qualitatively similar results). The physical layer is a two-state Markov chain, with a 0.5 state transition probability. In one state, the CQI remains constant, in the other a new CQI is extracted from a uniform distribution, so as to simulate channel variation.

We assume that the eNB is equipped with a MaxC/I scheduler, which achieves the maximum cell throughput. The scheduler is made DRX-aware, meaning that it only schedules UEs in the *on* phase, but does not otherwise exploit energy efficiency considerations (e.g., by prioritizing those UEs which are nearest to their sleep period). A comparative study of MAC schedulers under DRX is left for future study.

B. UE power model

As for the UE power model, we rely on the RF modem consumption model in [23]. It is based on four different states and four different transitions, each one with its power consumption value, reported in Figure 3. The *Light Sleep* state represents the RRC_CONNECTED state. It is used for short inactivity periods, when the UE powers down some of its circuitry. *Deep Sleep* represents the RRC_IDLE state, used for longer inactivity periods wherein the UE powers down more hardware. In the *Active - No Data* state the UE has the whole circuitry powered up but does not send/receive any data. In the other *Active* substates (i.e, RX, TX, RX+TX) the UE receives, sends, or receives and sends data from/to the eNB. Note that power consumption is different whether the UE is receiving, transmitting, or both. While the receiving consumption is fairly independent of the UE channel quality, the transmission one does depend on it, since a center-cell UE will use less power than a border-cell UE for the same PDU. The power consumption used in the model represents that of a border-cell UE.

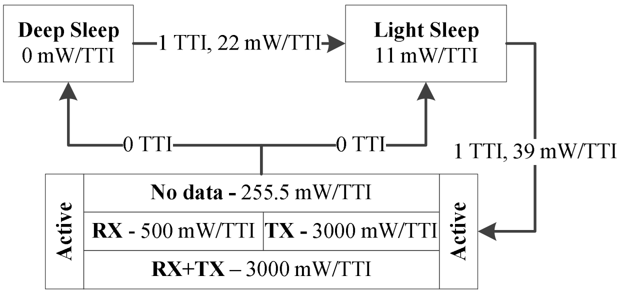

Figure 3 – Power consumption model

C. Application models

As far as application traffic is concerned, either the source of downlink flows or the destination of uplink flows are located beyond a core network. The latter introduces a variable delay modeled with a Laplacian distribution (min 0 ms, mean 80 ms, max 120 ms). We describe in detail the models used for VoIP traffic, web traffic and video traffic.

1) Voice over IP

Voice over IP is modeled according to the *VoIP ns-2 application* [25]. The employed codec is the GSM AMR Narrow Band (12.2 kbit/s) with VAD (no packets are sent during silences). The talkspurts and silence period durations are distributed according to Weibull functions. Header compression is employed. The set of parameters is summarized in Table 2.

TABLE 2 – VOIP MODEL PARAMETERS

Talkspurt duration (Weibull distribution)	Shape scale	1.423 0.824
Silence duration (Weibull distribution)	Shape scale	0.899 1.089
Codec Type	GSM AMR Narrow Band (12.2 kbps) w. VAD	
VAD Model	One-to-one conversation	
Header Compression	Active (RTP+UDP+IP headers = 6 bytes)	
Packet length	32 bytes/frame + 6 bytes Hdr + 1 byte RLC	

At the sender side, we allow for application-induced *frame aggregation* of up to two voice frames into the same RTP message. As far as performance metrics are concerned, we compute the Mean Opinion Score (MOS) [26], which predicts the quality experienced by human users by combining losses and mouth-to-ear delays in a codec-specific formula. The MOS ranges from 1 (unintelligible) to 5 (perfect), and a MOS above a 2.5 threshold in at least 80% of the talkspurts is considered acceptable for the employed codec. Mouth-to-ear delays are accounted for by including the application layer, i.e. encoding/packetization delays and playout buffer delays and losses. Playout buffering is in fact a major source of delay and losses, and cannot be neglected. The receiver employs an *optimal* playout buffer [25], whose performance upper bounds that of any real-life playout buffer. As shown in [25], optimal buffering allows one to discount buffering-induced MOS degradations, while maintaining a good degree of realism at the same time.

2) Video on Demand

Video on Demand (*VoD*) traffic is modeled by a streaming source that generates packets according to a pre-encoded MPEG4 trace file ([2]) whose parameters are summarized in Table 3**.** The frame type (I-frame, P-frame or B-frame), is carried in the packet, so as to enable a correct frame loss accounting (i.e., the loss of an I-frame determines the loss of the whole Group of Picture that relies on it for decoding).

TABLE 3 – VOD TRACE STATISTICS

Target rate	1Mbps
Min frame size	8 Bytes
Max frame size	29088 Bytes
Mean frame size	4.167116 Kbytes
Mean bit rate	1.000108 Mbps
Peak bit rate	6.981120 Mbps

3) HTTP

The HTTP model is an extension of the Empirical Web traffic model originally implemented in ns-2. This simulates Web traffic based on a set of CDF (Cumulative Distribution Function) data derived from live tcp dump traces. The communication is composed of page requests of fixed size, each one followed by one main object plus zero or more embedded objects. The time between two consecutive page request is called *reading time*. The time between two consecutives object downloads is called *server response time*. The set of parameters is summarized in Table 4**.** The key performance indicator is the *Page Delay* i.e. the time needed to receive a full page, including all the embedded objects.

TABLE 4 – HTTP TRAFFIC MODEL

Inter Page Time (exponential distribution)	Avg.	25
Objects per Page (truncated Pareto distribution)	Avg. shape	6.64 2
Bytes per Object (truncated log normal distribution)	Avg. Std.	6.17 2.36
Request Size	constant	320
Inter-Object Time (double exponential distribution)	Avg.	0.13

V. PERFORMANCE ANALYSIS

We present here performance results relate to the above three traffic types: VoIP, VoD and Web, along with guidelines on how to set the DRX parameters for each type of traffic.

A. VoIP

1) Downlink

We analyze the *downlink* (DL) part of a VoIP communication (i.e. the flow having the UE as a sink). We first consider the impact of the DO. A wise choice is to minimize the amount of UEs that concur for downlink resources at any TTI, which can be obtained by minimizing the overlap of their *on* phases as follows:

$$DO_i = \left(DO_{i-1} + ODT\right) \bmod LDC$$

Such *Minimum Overlap* solution is compared with a *fixed* and a *random* DO schemes. The first one makes two groups, one with DO=0 and one with DO=LDC/2, whereas the second assigns the DO randomly when the UE joins the cell. Figure 4 is a scatterplot of the MOS of each UE (i.e., each UE corresponds to a dot), with 100 to 400 UEs, under the three above DO selection schemes. As the figure shows, the *fixed* solution leads to poor MOS performance, already with 100 UE (hence is not plotted at higher loads), while the *random* and *Minimum Overlap* show better results. Note that while the *average* MOS value of the last two solution is similar, UEs are less scattered with *Minimum Overlap*, i.e., the performance is more predictable.

Given that traffic is CBR during talkspurts, under reasonable DRX settings (i.e. a LDC or SDC matching the period and a reasonably low ODT) it is hardly likely that more than one packet will be received on each DRX period, barring severe jitter conditions. We can thus safely send an UE to sleep using DCE every time it is sent a VoIP packet. This cuts down the *on* phase, whatever the ODT and IT values. DCE messages are piggybacked within a MAC PDU, hence have negligible to null cost in terms of resources. Figure 5 shows the power saved using the DCE, in various configurations. Sensible reduction are obtained even for ODT=1, since the IT is bypassed (recall that the IT cannot be null). The saving increases with the LDC, and decreases with the load. The latter effect is justified by the fact that a higher load implies a reduced chance of being scheduled (and, thus, sent to sleep) *early* in the *on* phase. In Figure 6 we show the effects of the DCE on MOS for two load scenarios (100 and 400 UEs), two ODT (1,5), two LDC (20, 40) with/without the DCE. The figure shows that the MOS is not affected by whether DCE is used or not (minor differences are observable for LDC=40). In this case the DRX cycle is twice the period, making it highly likely that *two* VoIP packets will be available at the beginning of each *on* phase. If those packets are not transmitted within the same TTI, the DCE may delay the second by one cycle (by sending the UE to sleep after the first one). However, even in that case, the MOS reduction is negligible, *because the added jitter is easily absorbed by the receiver playout buffer*. For this reason in the following we will always use the DCE for the DL traffic.

We now analyze the impact of ODT and LDC. The power consumption of a UE is proportional to the duty cycle ODT/LDC. The LDC has the highest impact, especially when using DCE, as the actual duration of the *on* phase may vary. In Figure 7 and Figure 8 we show the MOS reduction with various DRX configurations, with respect to the maximum achievable MOS. We separately show the effects of varying the ODT with a constant LDC (Figure 7) and vice-versa (Figure 8), in a scenario with 100 UEs. It can be seen that increasing the LDC affects the MOS more than decreasing the ODT. In fact, the DRX further delays packets when the LDC is larger than the VoIP period, i.e., when more than one VoIP packet is sent in an *on* phase. Increasing the ODT, instead, increases the chances of being scheduled when the competition is higher. For the above reasons, the practical guidelines for configuring DRX in downlink VoIP flows are the following:

- Always use the DCE, and send UEs to sleep as soon as they receive a packet;
- set the LDC according to the desired basic trade-off between power and QoS, regardless of the cell load;
- set the ODT based on the cell load, i.e. increase it as the number of UE increases, to compensate for a reduced scheduling probability.

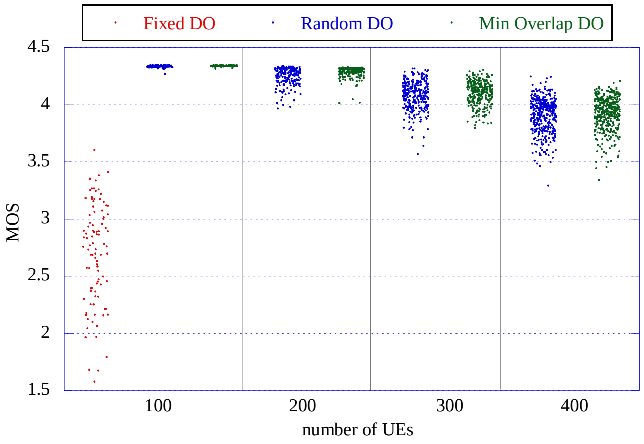

Figure 4 – MOS of VoIP conversation as a function of the no. of UEs for various DO selection strategies.

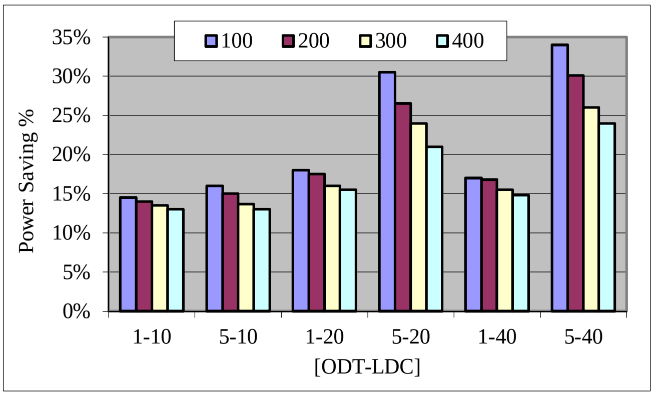

Figure 5 – Power saving reduction when using DCE with respect to normal DRX with the same parameters

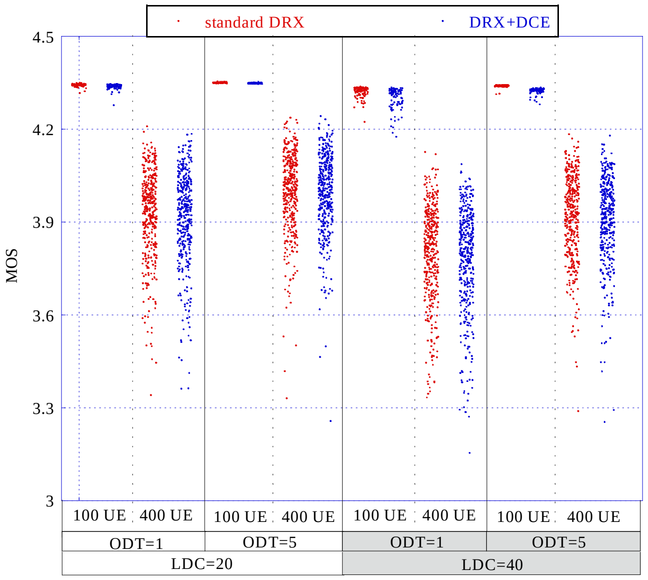

Figure 6 – MOS of VoIP conversation in case of standard DRX and DRX with DCE. ODT={1,5}, LDC={20,40}. For each ODT-LDC configuration the number of UEs is 100 (left) and 400 (right)

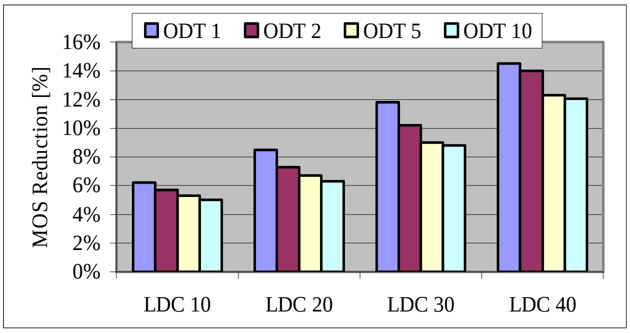

Figure 7 – MOS Reduction as a function of the ODT. The reduction is computed with respect to the maximum achievable MOS value.

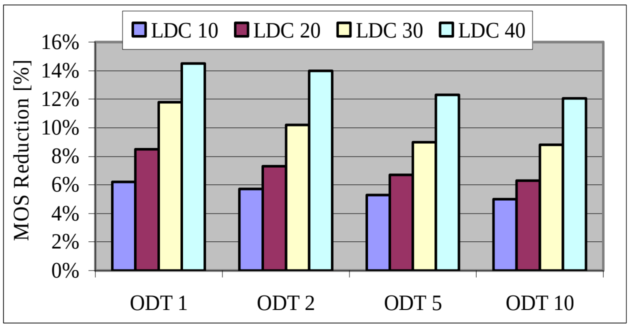

Figure 8 – MOS Reduction as a function of the LDC. The reduction is computed with respect to the maximum achievable MOS value.

2) *Uplink*

In the uplink direction, the UE signals the presence of new traffic to the eNB via RAC requests. The success probability of RAC requests decreases with resource contention, i.e., cell load. Hence the system capacity depends sensibly on the efficiency of the RAC mechanism. First of all we analyze the impact of *Bandwidth Stealing*. Figure 9 shows the MOS of 100 to 250 UEs with several {ODT, LDC} values. As the figure shows, BS does not improve the MOS sensibly, unless the cell is saturated (i.e., in the rightmost part of each {ODT, LDC} column), in which case saving the uplink resources otherwise reserved for BSR transmission becomes significant. BS instead reduces power consumption, as shown in Figure 10. The higher savings are obtained at low ODTs (1-5 ms), where UEs are highly likely to complete the RAC handshake in 3ms. For large ODTs (e.g., 10ms) there is practically no difference. Another possible solution is to use SPS, as explained in Section II. SPS is well suited for periodic traffic such as VoIP: at the beginning of a talkspurt the UE requests a grant via a random access procedure. The eNB serves the request, and - at the beginning of the next *on phase* - allocates a periodic grant with size large enough to transmit the number of packets arriving in a cycle, *given the current CQI* at the time of decision.

On one hand, SPS allows more conservative DRX configuration than BS. In fact, an ODT of 1 is enough to cope with periodic grants in the steady state (i.e., after the beginning of a talkspurt), whereas RAC-based scheduling (even with BS) requires UEs to stay on for 3 TTIs at least. However, SPS is inefficient at the cell capacity level, since it books resources based on the CQI at the onset of a talkspurt (which may be lower than the average for that UE), whereas RAC-based scheduling always uses recent CQIs. This inefficiency is multiplied by the number of packets that a periodic grant should accommodate, hence is more visible with larger LDCs. Figure 11 shows that – as the cell load increases – SPS achieves worse MOS than BS. Decreasing the handshake time from 3 to 1 TTI yields the saving shown in Figure 12. As can be seen, SPS is beneficial only with short LDC cycles, e.g. equal to a VoIP period. If longer cycles are to be used (e.g., to reduce the power consumption further) BS is instead more advisable.

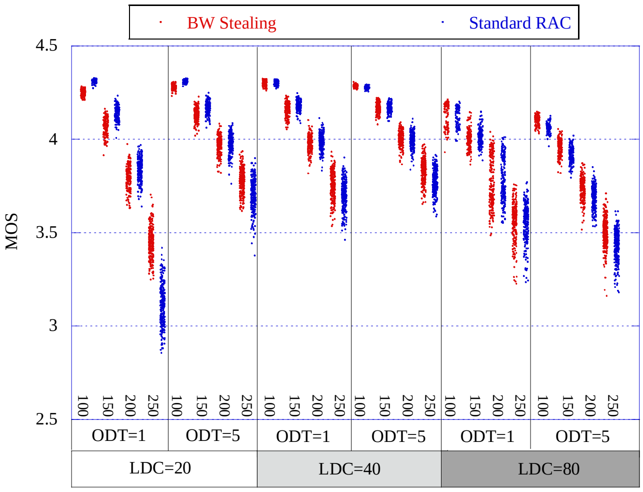

Figure 9 - MOS comparison of BW Stealing vs. Standard RAC. For each {ODT,LDC} pair, the number of UEs goes from 100 (left) to 250 (right)

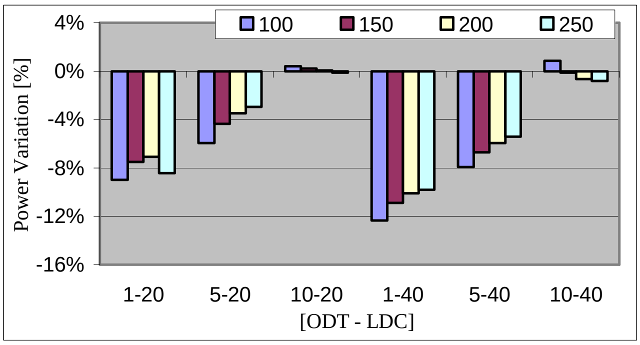

Figure 10 - Consumption reduction from Standard RAC to BW Stealing

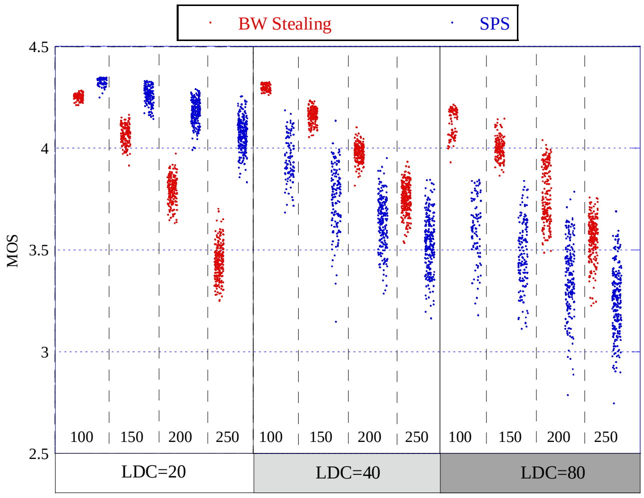

Figure 11 - MOS BW Stealing vs SPS. For each configuration of LDC the #UEs goes from 100 to 250

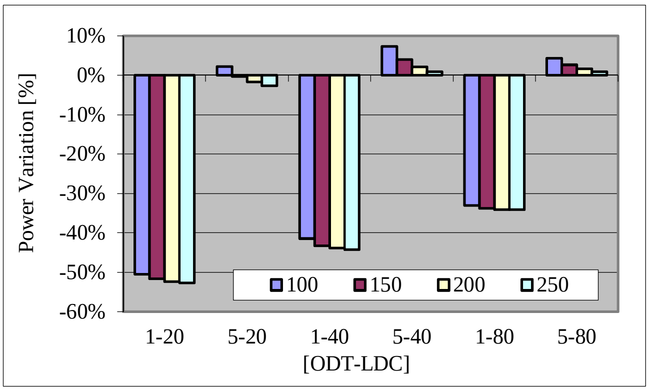

Figure 12 - Consumption reduction from BW Stealing to SPS

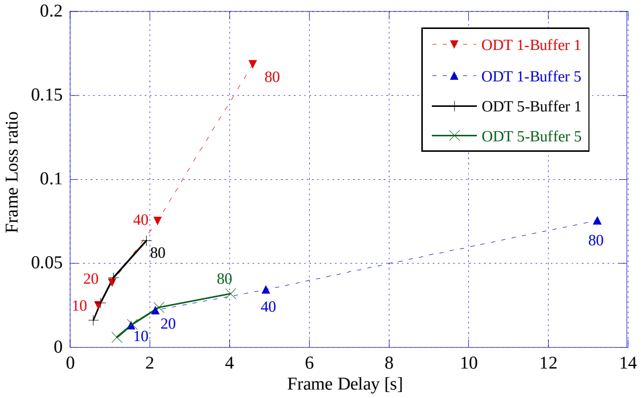

Figure 13 – VoD Frame Loss and Frame Delay as a function of LDC. For each curve LDC goes from 10 to 80. Some LDC values are reported besides the markers for explanation.

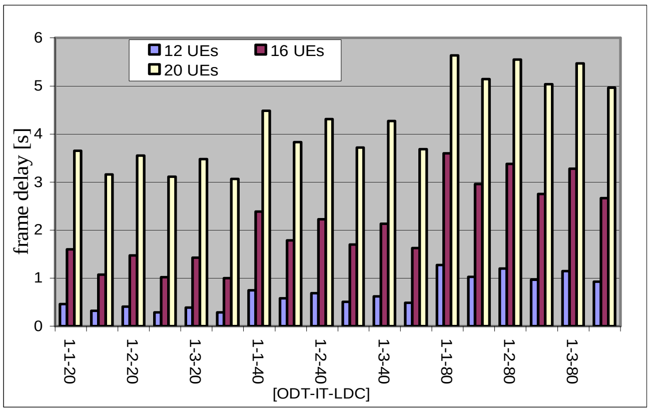

Figure 14 - VoD Frame Delay

B. *VoD*

VoD has constant inter-arrival times and variable-size frames, leading to bursty traffic when large frames occur. DRX adds delay to packets, hence it may require significant buffering at the eNB to prevent losses induced by overflows.

In Figure 13 we show how the LDC affects the *frame delay* and *frame loss*, which are however correlated. We let the LDC range from 10 to 80 ms, and we join the different (delay loss) points in curves, for two values of the buffer and ODT. As the LDC increases, two different effects can be observed, depending on the buffer size: with a small buffer (1MB, red and black curves) the frame loss grows with the LDC more rapidly than the delay, whereas with a large buffer (5MB) the LDC affects the frame delay more than the losses, which is expectable. Increasing the ODT (black and green curves) dampens these effects, meaning that the same trend can be observed, only the losses and delay grow at a much smaller rate with the LDC. In the following we will assume 5MB buffers, so as to focus our attention on frame delay. Figure 14 shows the effects of DRX parameters on the frame delay. Again the LDC has the highest impact. Minor improvements may be obtained by increasing ODT or IT, with the former being more effective in reducing the delay, but more costly in terms of power. In Figure 15 and Figure 16 we show the variation of delay and power consumption when varying the ODT from 1 to 5, while keeping the IT constant, and vice versa. The above effect is more evident as the load increases: the IT is triggered only when the UE is scheduled, thus UEs with low channel quality will be less likely to activate it. With a longer LDT (80ms) this effects tends to decrease with the distance between the DO of two consecutives UE groups. The DO needs to be set in order to avoid overlap of *on phases* even when they are extended due to IT. This can be achieved by spreading the DOs of different UEs as far as possible within the cycle.

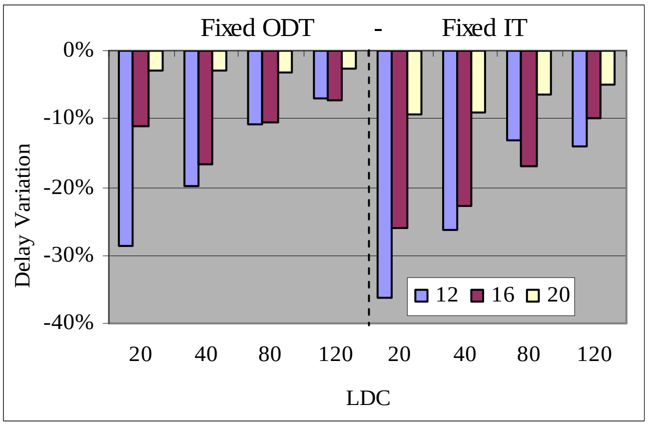

Figure 15 – Variation of the VoD delay as a function of DRX settings

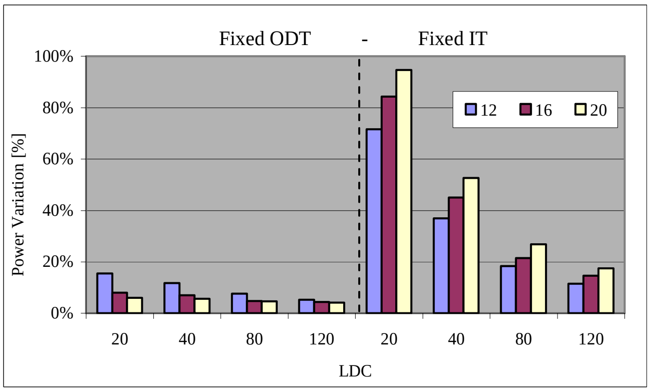

Figure 16 – Variation of per-UE power consumption when increasing IT from 1 to 5, while keeping ODT constant to 1, and vice-versa, with VoD

C. *HTTP*

HTTP is characterized by small sporadic packets in the UL (page requests) and bursts in the DL (object downloads), which implies that long periods of inactivity alternate with bursts of resource requests. In fact, the mechanism of short/long cycles has been envisaged to cope with these situations, hence we configure DRX with SDC and LDC.

The LDC has the highest impact on the page delay. The latter is slightly dependent from the system load (Figure 17), as the LDC affects only the reaction time of the system after a period of inactivity. Suitable values for the LDC are from to 160 to 1024ms: larger values introduce high delay without decreasing the power consumption significantly. The SCT should be set large enough to cope with the delay between the page request and the first object download, and between two consecutives object downloads. The negative effects of a too small SCT increase with the LDC. Increasing the IT should be preferred over ODT to manage resource competition among UEs when the cell load grows, as shown in Figure 18: this way the *on phase* is extended only when a download is in progress, and not when LDT is active. The energy cost of increasing the ODT is sensibly higher with respect to IT (see Figure 19).

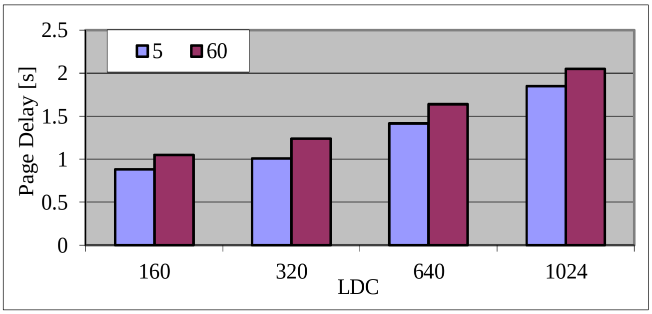

Figure 17 – HTTP page delay as a function of LDC and system load. SCT=2, SDC=20, IT=1, ODT=1

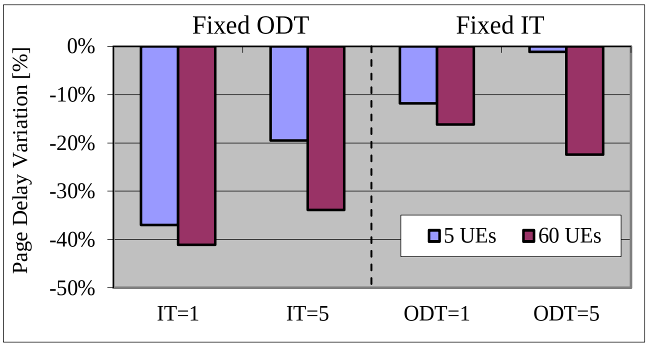

Figure 18 – Page delay variation when increasing IT from 1 to 5, while keeping ODT fixed to 1, and vice-versa. SDC=20, SCT=2

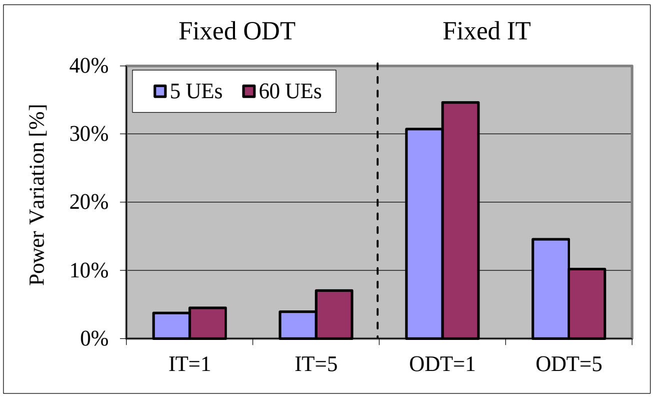

Figure 19 – Variation of the per-UE power consumption when increasing IT from 1 to 5, while keeping ODT constant to 1, and vice-versa. SDC=20, SDT=2

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have analyzed the effect of DRX on the QoS and power consumption of UEs, with VoIP, VoD and HTTP traffics. The evaluation has been carried out by simulation, analytical modeling being out of the equation due to the intricacies of the LTE environment. For each type of traffic, the specific DRX mechanisms which are more suitable for that type of traffic have been identified, and the parameters have been tuned accordingly so as to trade QoS for power saving.

This work can be extended in at least two directions. The first one is evaluating DRX policies for machine-to-machine traffic. The latter is characterized by a highly sporadic traffic pattern, although it may impose stricter requirements in terms of QoS (especially for real-time M2M applications) and reliability, making the search for an optimal trade-off an hard task. Furthermore, the present work has given the scheduler for granted, whereas our preliminary results show that a DRX-aware scheduler might strike a better compromise between power consumption and QoS. We are actively pursuing this line of research at the time of writing.

REFERENCES

[1] http://www.3gpp.org

[2] http://trace.eas.asu.edu/cgi-bin/show.cgi?name=Jurassic_Park_1_DVD

[3] H.Holma, A.Toskala, “LTE for UMTS - OFDMA and SC-FDMA Based Radio Acces”, John Wiley & Sons, 2009

[4] Medium Access Control (MAC) Protocol Specification, 3GPP TS 36.321 (rel 9)

[5] Physical Layer Procedures, 3GPP TS 36.213 (rel 9)

[6] H. Bo, T. Hui, C. Lan, Z. Jianchi, “DRX-Aware Scheduling Method for Delay-Sensitive Traffic”, IEEE Communications Letters, Vol. 14, No. 12, December 2010

[7] C: Zhong, T. Yang, L. Zhang, J. Wang, “A new discontinuous reception (DRX) scheme for LTE-Advanced Carrier Aggregation Systems with Multiple Services”, 2011

[8] S. Fowler, “Study on Power Saving Based on Radio Frame in LTE Wireless Communication System Using DRX”, 2011

[9] L. Zhou, H. Xu, H. Tian, Y. Gao, L. Du, L. Chen, “Performance Analysis of Power Saving Mechanism with Adjustable DRX Cycles in 3GPP LTE”, IEEE 68thVehicular Technology Conference, VTC Spring, September 2008, pp. 1-5

[10] V. Mancuso, S. Alouf, “Analysis of power saving with continuous connectivity”, Computer Networks 56 (2012) 2481–2493

[11] V. Mancuso, S. Alouf, N. C. Fofack, “Analysis of power saving and its impact on web traffic in cellular networks with continuous connectivity”, Pervasive and Mobile Computing, 2012

[12] Y. Wen, J. Liang, K. Niu, W. Xu, “Performance Analysis and Optimization of DRX Mechanism in LTE“, in Proc. of IC-NIDC 2012

[13] S. Jin, D. Qiao, “Numerical Analysis of the Power Saving in 3GPP LTE Advanced Wireless Networks”, IEEE Trans. On Vehicular Technology, Vol. 61, No.4, May 2012

[14] K. Zhou, N. Nikaein, T. Spyropoulos, “LTE/LTE-A Discontinuous Reception Modeling for Machine Type Communications”, IEEE Wireless Communication Letters, 2012

[15] S. Baek , B. D. Choi, “Analysis of Discontinuous Reception (DRX) with both Downlink and Uplink Packet Arrivals in 3GPP LTE”, proc. of QTNA 2011, Aug. 23-26, 2011, Seoul, Korea

[16] Y. Y. Mihov, K.M. Kassev, B.P. Tsankov, “Analysis and Performance Evaluation of DRX Mechanism for Power Saving in LTE”, IEEE 26th Convention of Electrical and Electronics Engineers in Israel, 2010

[17] L. Liu, X. She, L. Chen, “Multi-User and Channel Dependent Scheduling Based Adaptive Power Saving for LTE and Beyond System”, Proc. of 16th APCC, 2010

[18] J. Wigard, T. Kolding, L. Dalsgaard, C. Coletti, “On the User Perfomance of LTE UE Power Savings Schemes with Discontinuous Reception in LTE”, 2009

[19] T. Kolding, J. Wigard, L. Dalsgaard, “Balancing Power Saving and Single User Experience with Discontinuous Reception in LTE”, 2008

[20] K. Aho, *et al*., “Tradeoff Between Increased Talk-time and LTE Performance”, 9th International Conference on Networks, 2010

[21] K. Aho, T. Henttonen, L. Dalsgaard, “Channel Quality Indicator Preamble for Discontinuous Reception”, 2010

[22] K. Aho, T. Henttonen, J. Puttonen, L. Dalsgaard, T. Ristaniemi, “User Equipment Energy Efficiency versus LTE Network Performance”, IARIA Int. J. on Advances in Telecommunications, vol 3 no 3 & 4, 2010

[23] M. Polignano, D. Vinella, D. Laselva, J. Wigard, T.B. Sorensens, “Power savings and QoS Impact for VoIP Application with DRX Feature in LTE”, 2011

[24] 3GPP - TS 36.300 Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2

[25] A. Bacioccola, C. Cicconetti, G. Stea, “User level performance evaluation of VoIP using ns-2", in *Proc. NSTOOLS'07*, Nantes, France, Oct. 22, 2007.

[26] ITU-T Recommendation G.107. The Emodel, a computational model for use in transmission planning. Dec. 1998

VII. APPENDIX

TABLE 5 – LTE-RELATED ACRONYMS USED IN THE PAPER

Acronym	Definition
BLER	Block Error Rate
BSR	Buffer Status Report
CQI	Channel Quality Indicator
DC	DRX Cycle
DCE	DRX-Command MAC Control Element
DO	DRX Offset
DRX	Discontinuous Reception
DSR	Dedicated Scheduling Request
eNB	Evolved Node-B
H-ARQ	Hybrid Automatic Repeat reQuest
IT	DRX Inactivity Timer
LDC	DRX Long DRX Cycle
LTE	Long-term Evolution
MaxC/I	Maximum Carrier over Interference
ODT	DRX On Duration Timer
PDCCH	Physical Downlink Control CHannel
PDU	Protocol Data Unit
RAC	Random Access Procedure
RB	Resource Block
RLC	Radio Link Control
RRC	Radio Resource Control
SCT	DRX Short Cycle Timer
SDC	Short DRX Cycle
SNR	Signal to Noise Ratio
SPS	Semi-Persistent Scheduling
TTI	Transmission Time Interval
UE	User Equipment