High Speed Packet Access Evolution – Concept and Technologies

J. Peisa, S. Wager, M. Sågfors, J. Torsner, B. Göransson, T. Fulghum, C. Cozzo and S. Grant
Ericsson Research

***Abstract*—In this paper we present the main concepts of High Speed Packet Access Evolution currently being standardized in 3GPP. In general HSPA Evolution consists of introduction of MIMO, higher order modulation, and protocol optimizations and optimizations for voice over IP. We describe these improvements in detail and show that HSPA Evolution can reach performance comparable to those of Long Term Evolution of UMTS Terrestrial Radio Access Network in a 5MHz deployment.**

***Index Terms*—WCDMA, HSPA, HSDPA, UMTS**

I. INTRODUCTION

The demands of packet data applications have resulted in several improvements over the original WCDMA release 99. The downlink was improved in Release 5 with High Speed Packet Data Access (HSDPA), which provided high speed shared channel with fast link adaptation and scheduling, hybrid ARQ, and a short 2 ms transmission time interval (TTI). The corresponding uplink enhancements were done in Release 6 with Enhanced Dedicated Channel (E-DCH), which allow Node B controlled scheduling of the uplink transmission, Hybrid ARQ, short 2ms TTI, and fast inter-cell interference suppression.

Recently work has started to evolve the UTRAN with Long Term Evolution (LTE), providing a new air interface. LTE will supporting flexible spectrum allocation from 1.5 MHz up to 20 MHz, and will provide significant performance improvements in application performance and system capacity.

For operators with existing HSPA deployments, the possibility to evolve HSPA should provide an easy way to update the system. For this reason work on HSPA Evolution has started in 3GPP. In this paper we will look at the main technologies used to evolve the HSPA system, compare the reachable performance to the performance goals of the LTE in 5 MHz bandwidth, and summarize the concept of HSPA Evolution.

The performance of selected technical improvements to the HSPA system has been evaluated in [4]. In this paper we collect the various individual improvements to provide a comprehensive overview of HSPA Evolution.

TABLE I
SUMMARY OF PERFORMANCE GOALS FOR UTRAN LONG TERM EVOLUTION

Performance metric	Goal
Peak data rate	100 Mbps downlink, 50 Mbps uplink
Control plane latency	100 ms idle to active 50 ms dormant to active
Control plane capacity	At least 200 users in active state
User plane latency	5 ms for small IP packet
Average user throughput	Downlink 3-4 times HSDPA Uplink 2-3 times enhanced uplink
Spectrum efficiency	Downlink 3-4 times HSDPA Uplink 2-3 times enhanced uplink

II. PERFORMANCE GOALS

The performance goals for the LTE of UTRAN have been agreed upon in 3GPP [1] and are summarized in Table I. In addition to the performance goals listed in Table I, LTE should work with high mobility, provide large coverage, allow flexible spectrum allocation, easy migration, and coexistence with earlier UTRAN releases. One of the design goals for LTE work is to minimize the complexity by not allowing redundant mandatory features or options.

The performance of HSPA Evolution should target LTE performance with two major differences: the bandwidth is limited to 5 MHz, and the transmission time interval is 2 ms compared to 1 ms (with possibility to use only half a TTI transmission for small amounts of data) for LTE.

The longer TTI will not allow HSPA Evolution to reach the User Plane latency target of LTE. The LTE target is 5 ms one way delay from the UE to the edge of the radio access network. As a rough estimate for the effect of the TTI on the latency, it is possible to assume that the UTRAN network delays for HSPA Evolution are the same as for LTE. Then the impact of the TTI on latency can be roughly estimated as

$$\begin{aligned} Latency\_TTI = {} & TTI\ alignment\ delay \\ & + TTI\ dependant\ TX\ processing \\ & + TTI \\ & + TTI\ dependant\ RX\ processing. \end{aligned}$$

On average the TTI alignment is 0.5*TTI. For TTI dependant processing delay it is possible to assume that the processing delays is in general less than one TTI, resulting in the total dependency on the TTI as $Latency\_TTI = 1.5 – 3.5\ TTI$

1550-2252/$25.00 ©2007 IEEE

Based on this we can expect that the user plane latency target for HSPA Evolution should be 2.25 – 5.25 ms longer than the LTE target, which could correspond to a possible user plane latency target of 10 ms for HSPA Evolution.

The effect of the smaller bandwidth can (in principle) be straightforwardly estimated by assuming a similar spectrum efficiency for LTE and HSPA Evolution. This results in peak throughput targets of 25 Mbps in downlink and 12.5 Mbps in the uplink.

All other targets from [1] should in principle apply unmodified for HSPA Evolution as well as for LTE.

III. TECHNIQUES

In this section we examine the techniques used to achieve targets derived in Section II.

A. MIMO

Increasing data rates can be achieved by transmitting multiple parallel transport blocks to a single user. This is often referred to as Multiple Input Multiple Output (MIMO). The preferred use for MIMO is in conditions with favorable signal-to-noise ratio distribution and for channels with favorable correlation properties, e.g., small cells or indoor deployments. The receiver has the possibility to separate the multiple data streams by using the channel properties and knowledge of the coding scheme. In order for the receivers to solve this task it is necessary to standardize the used multi-layer transmission scheme. The MIMO scheme chosen for HSPA evolution is based on the multiple codeword principle allowing for transmitting two separately encoded streams to a UE. This will facilitate the use of successive interference cancellation receivers which is expected to boost the performance compared to linear receivers such as e.g. MMSE based receivers. Hence the data on each stream is separately encoded, modulated and spread. The up to 15 spreading codes (of spreading factor 16), available for HSDPA, are reused over both streams. Before transmitted on the antennas, the modulated and spread signal is spatially weighted (pre-coded) using an unitary transform. The weights are taken from the same codebook as used for closed-loop transmit diversity mode 1.

The link adaptation is a mix of spatial and temporal adaptation. For each TTI, or rather at a rate set by the network, the UE reports the number of streams, the spatial weight (pre-coding index) and transmission rate that it prefers. This information is used in the Node B scheduler to select a suitable transport format. To inform the UE about the actual parameters of a transmission, the downlink control channel, HS-SCCH, has been adapted to incorporate MIMO information. At each TTI when a particular user is scheduled, the Node B includes the used pre-coding weight, the number of streams and the modulation used on each stream to the first part of the HS-SCCH. For MIMO transmission, two versions of the second part of the HS-SCCH exist. One is used for single stream transmission containing the transport block size and HARQ processing information. When transmitting two transport blocks (dual stream) the same information about the second transport block is also signaled. Since the information about number of streams is allocated in the first part, the UE can use this information when detecting the second part.

HARQ is operating independent between the streams. Thus, each stream is separately (not)-acknowledged and retransmission of not yet detected blocks can be done independently between the two streams.

In order to introduce MIMO, the uplink signaling has been modified slightly. The Release 5/6 HS-DPCCH has been expanded to accommodate the new MIMO signaling. Since two transport blocks can be transmitted, each with its own HARQ process, the HARQ ACK/NACK field of the HS-DPCCH has to be expanded to accommodate signaling for the second stream. In Release 5/6 the ACK bit is repeated to 10 bits. For MIMO, pure repetition would give a code with Hamming distance 5. However, if the ACK/NACK is jointly encoded, a code with Hamming distance 6 can be found [14].

The CQI evaluation is slightly modified compared to Release 5. The main reason for this is the code-reuse interference term that is present when transmitting two streams. Since this term will be dependent on the actual channel realization and code allocation, only the UE can estimate the influence on the experienced SINR. However, since the UE is unaware of the actual code allocation it will be given at the moment the CQI is estimated, two different types of CQI has been defined.

The Type A CQI report is referred to as the preferred single/dual stream CQI, while Type B reports are single stream CQI reports. The network can configure the reporting period of these two CQI reports e.g. every second or third report may be of Type B. If the network receives Type A reports indicating that the user prefers dual stream transmission while the network may not have sufficient resources (power and/or codes) to support dual stream transmission, the information in Type B reports can be used to select the proper modulation and coding format.

To support precoding, the CQI reports also contain PCI (precoding control indication) which indicates the preferred precoding vector for the primary stream. Due to the construction of the codebook, the weight for the secondary stream is well defined. The weights are taken from the closed-loop transmit diversity mode 1 alphabet and consists of phase shifts applied to the second antenna. The possible entries are given by $w \in \left\{ \frac{1+j}{2} \quad \frac{1-j}{2} \quad \frac{-1+j}{2} \quad \frac{-1-j}{2} \right\}$. From this, it is seen that if e.g. the first entry is preferred for the primary stream, only one other entry exists giving an orthogonal weight for the secondary stream.

The possible gains achieved from MIMO can be found in Figure 1 where the 90th percentile throughput is depicted. Further explanations and simulation assumptions are given in the next section.

B. Higher order modulation

Release 6 HSPA systems support the use of 16QAM in the downlink and QPSK in the uplink. These modulation schemes may provide high enough data rates given the received symbol SNRs of macro cell environments, however, for indoor or small-cell system deployments, higher SNRs and higher order modulation (HOM) can be supported.

Modulation and coding scheme (MCS) tables determine the best combination of modulation and coding rate for a given SNR. With existing MCS tables, high symbol SNRs may "max out" the choice of MCS, giving the highest order modulation with the least amount of coding. As a result, these high SNR systems become peak rate limited. Besides MIMO, another means to increase this peak rate is to extend the MCS tables into higher SNRs with the introduction of even higher order modulations: 64QAM in the downlink and 16QAM in the uplink. While HOM can be used in conjunction with MIMO, it is important in its own right in those cases where deployment of MIMO systems is prohibited by physical, zoning, or budgetary limitations at the transmitter.

In the downlink, the MCS table would be modified to include 64QAM. While the present downlink MCS table has a maximum rate entry of QPSK modulation with rate $r = 1$ coding (4 information bits per symbol) at 14.5 dB SNR, 64QAM with various coding rates would be added to extend the table up to 64QAM with $r = 1$ coding (6 information bits per symbol) at 22.5 dB SNR. A proposed MCS table can be found in the appendix of [6].

As presented in [6] and [7], an "SNR/rate-lookup" simulation was used to evaluate the impact on throughput of introducing 64QAM and MIMO into the HSPA downlink. This method allows an expedient yet accurate assessment of the radio link in question. Essentially, this simulation consists of generating the appropriate channel realization, determining a G-RAKE [8] or MIMO G-RAKE [9] combining weight solution as a function of this channel realization, analytically calculating the output symbol SNR for the each stream in the receiver, looking up the maximum supportable rate of the output SNR from the MCS table, and averaging this maximum supportable rate over many channel realizations (assumes ideal link adaptation).

While the MCS table gives us the modulation and coding scheme which is best for a given operating condition, it does not give us the *transport* format, which includes the number of spreading codes employed. Throughout, we will assume that 15 spreading codes are available at spreading factor 16, and all codes will be used by the UE. Further, we assume 80% of the downlink power to be devoted to the high-speed data channel, and 20% devoted to other (pilot channels, control channels, and voice channels). We assume ideal timing recovery, delay, channel, and impairment covariance estimation. As we are logging the maximum supportable information rate for a given realization, ideal link adaptation is also assumed.

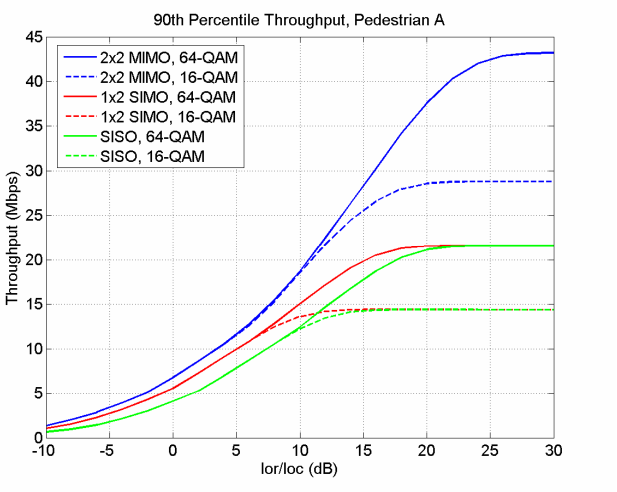

Figure 1. 90th percentile throughput in Pedestrian A channel for HOM and MIMO.

Figure 1 shows the 90th percentile throughput of the downlink for 1000 realizations of the Pedestrian A dispersive channel [10], for systems employing techniques for peak rate improvement. For purposes of comparison, SISO and SIMO systems are shown employing the existing MCS table (16QAM). We can see that, as expected, the peak rate with the addition of 64QAM increases to 21.6 Mb/s as compared to the 16QAM case. The use of multiple receiver branches (SIMO vs. SISO) extends the benefits of HOM into lower SNRs. MIMO with 16QAM doubles the peak rate to 28.8 Mb/s, while the combination of MIMO and 64QAM increases the peak rate to 43.2 Mb/s: essentially giving a similar data rate as Long Term Evolution (LTE) in a 5 MHz bandwidth. While the combination of MIMO and 64QAM are not proposed is the next 3GPP release, the combination is being considered for future releases.

In the uplink, the introduction of 16QAM allows the peak data rate to reach about 11.5 Mbps, featuring an increase of 100% in peak rate compared with the enhanced uplink in release 5, where the highest peak rate with QPSK is 5.74 Mbps with coding rate equal 1. In the 3GPP study item on uplink HOM, 8PSK was also considered, however only 16QAM is part of the work item for introduction in release 7 because of the higher peak rate and better performance for all rates when HOM is used.

Link-level simulations are shown here to illustrate the data rates achievable with 16QAM. In order to reach such high rates in dispersive channels, the receiver needs to suppress the interference generated by the code multiplexed channels (spreading factor is equal 2 or 4). The receiver used is the GRake receiver [8] with 2 antennas. The simulator models inner loop power control and HARQ with a maximum of 4 transmission attempts. The TTI is 2ms. The power settings are β_c=0.446, β_{ec}=0.233, and β_{ed}=1.16 for the DPCCH, E-DPCCH and E-DPDCH, respectively. Transport block sizes from 6 000 to 22 000 bits are simulated. The receiver assumes knowledge of the channel coefficients. More simulation details can be found in [11].

Figure 2 shows the throughput for the pedestrian A channel at 3 km/h as a function of E_c/N_0, where E_c is the total receiver chip energy and N_0 is the noise power spectral density. For rates lower than 4 Mbps, QPSK is more power efficient than 16QAM. As the rate increases, 16QAM outperforms QPSK. In the region between 4 Mbps and 5.5 Mbps, although both modulations can reach those rates, 16QAM uses more realistic coding rates and is more power efficient. Note that at 5 Mbps, the coding rate is 0.86 for QPSK and 0.43 for 16QAM. Rates higher than about 5.7 Mbps can be reached only with 16QAM.

The high rates provided by 16QAM are reached at high power and therefore the likely scenario for uplink HOM is when the user experiences favorable channel conditions.

The impact on 3GPP specifications for supporting uplink HOM concerns mostly the physical layer and is expected to be minor. Changes will involve support for additional signaling and larger transport block sizes. The need for an improved phase reference to enable high rate transmission with 16QAM is currently under discussion in RAN1[12]. In Layer 2 specifications, the transport block size tables will be modified.

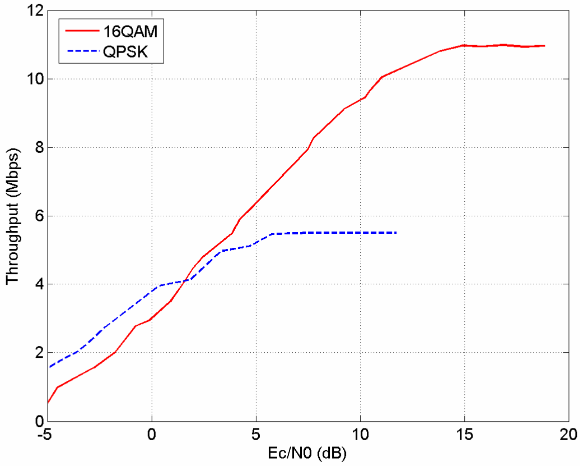

Figure 2. Throughput as a function of E_c/N_0 in Pedestrian A channel for uplink HOM.

C. Continuous Packet Connectivity

The increasing deployment of applications, which send small packets either continuously (e.g. Voice over IP) or intermittently (e.g. presence information of the messaging applications, push email) require efficient support for continuously connected applications. Both the battery life of the mobile terminal and the capacity of the cellular system will benefit from the introduction of discontinuous transmission and reception in active state. In addition the possibility to transmit small packets without out-band control signaling used normally to schedule transmissions on HSDPA ("HS-SCCH less operation") will improve the system capacity. Both the discontinuous transmission and reception and the "HS-SCCH less operation" can only be used with HSPA (i.e. they do not apply to the Release 99 dedicated channels).

Discontinuous transmission and reception can be applied in either uplink or in downlink.

The downlink discontinuous reception is configured by the RNC, but can be turned on and off by the base station with physical layer signaling. It allows the UE to restrict the downlink reception times. Note that the UE is also allowed to receive continuously; discontinuous reception is not required even if it is configured. If the UE has not been scheduled for a predetermined time, the UE is not required to receive downlink channels except at designated time instances and in some special cases (e.g. reception of the HARQ ACK/NACK of an uplink transmission). The downlink discontinuous reception will allow the UE to enter power saving mode between the downlink receptions, and thus reduce the UE battery consumption in the active state.

In the uplink, the RNC can configure the UE to transmit discontinuously the power control commands (on DPCCH). The base station can also enable the uplink discontinuous transmission with a physical layer command. The UE will autonomously enter the discontinuous transmission mode, in which there are two transmission cycles UE_DTX_cycle_1 and UE_DTX_cycle_2. The latter is used whenever there is no uplink data transmission activity, whilst the former is applied depending on the duration of E-DCH inactivity. This allows uplink DPCCH transmission to adapt to the data transmission activity. Similar to the downlink discontinuous reception, the uplink discontinuous transmission will allow the UE to enter power saving mode between uplink transmissions. In addition, the reduced power control transmission will reduce the uplink interference, and increase the system capacity.

The uplink discontinuous transmission will only impact the transmission of the power control commands. At any point of the time, the UE is able to transmit higher layer data (e.g. signaling or user plane data). The actual transmissions are followed by a preamble and a postamble to maintain correct power control level. However, it is also possible to configure discontinuous reception, for example, if the base station would like to share the receiver units between several discontinuously transmitting users. In that case the possible time instances for initial uplink transmission after the UE has entered discontinuous transmission are restricted.

The "HS-SCCH-less transmission" allows initial transmission of small packets without scheduling the UE with the HS-SCCH in advance. The retransmissions are still scheduled and the transmission formats are restricted to four different transport block sizes.

If the UE does not detect HS-SCCH, it will try to decode the packet received on HS-DSCH. If the reception is successful, the UE will send a HARQ ACK. If the reception does not succeed, the UE will store the received soft bits, but will not send a NACK. For subsequent retransmissions, both ACK and

NACK signals are transmitted normally, and thus the retransmissions do not require blind detection. For simplicity the number of retransmissions is limited to a maximum of two retransmissions.

The packets are protected with a 24-bits long UE specific CRC. In addition, the UE continues to attempt reception of the legacy HS-SCCH in the configured HS-SCCH set.

D. Enhanced CELL_FACH operation

To get the full advantage of the increased bit rates offered by MIMO and Higher Order Modulation, focus in HSPA evolution is also to improve signaling and state transition performance in the system, following the analogy: "What is the use of a highway if it takes too long to reach it"? Analysis in [13] has shown that the setup delays and channel switching times in WCDMA considerably affect the user perceived performance, and that the use of HSPA for transport of the signaling messages can significantly decrease the delays. Another observation is that the control signaling and synchronization overhead of moving a user to CELL_DCH state, where dedicated or shared channels with high bit rates are available, is significant for small amounts of data. As a result of these findings, a new work item was started in 3GPP to enable the use of HSPA also in the CELL_FACH state, where UEs with lower activity typically camp, to improve packet data and signaling performance when common channels are used.

In Enhanced CELL_FACH operation, HSDPA is activated also for users in CELL_FACH. Connected UEs in CELL_FACH monitor the HSDPA control channels for scheduling information with their user specific identity (H-RNTI), much in the same way as in CELL_DCH. The difference is that in CELL_FACH, no dedicated uplink channel exists. This means that continuous transmission of Channel Quality Indicator (CQI) and HARQ feedback will not be supported, and that link adaptation and HARQ need to be modified. Link adaptation will be based on measurements provided in the Radio Resource Control protocol (RRC) on the random access channel in the uplink, and HARQ will be replaced by blind repetition on MAC. Simulations in [5] have shown that with link adaptation based on initial CQI and fixed number of repetitions, average throughputs in the order of 300-500 kbps can be achieved.

Another feature of Enhanced CELL_FACH is the target to have the same layer 2 header format as in CELL_DCH, described in section E. This means that data transmission can continue uninterrupted on the user specific H-RNTI during the channel switch procedure between CELL_FACH and CELL_DCH, and provides a significant improvement in the user perception compared to the current solution, where data transmission is suspended during the channel switch.

The work item also includes an option to transmit paging messages on HSDPA. Most aspects and definitions of the paging procedure remain as specified in Rel-6, e.g. usage of paging indication, paging occasions, paging groups, paging message format etc. When the paging indication has been given for a certain paging group, the UEs of that group start to monitor the HSDPA control channels for a common H-RNTI used to schedule the paging message. The benefit with this approach is that it can be extended to support also data transmission to users in the cell paging state (CELL_PCH), without the need of first moving the users to CELL_FACH with the Cell Update Procedure. This way, latency and especially control signaling overhead for delivering background traffic to inactive UEs is reduced.

E. Layer 2 protocol enhancements

It is known from work on HSDPA (e.g.[2]) that the Acknowledged Mode RLC downlink peak data rate is limited by the RLC PDU size, the RLC round trip time (RTT) and the RLC window size. For reasonable RLC PDU sizes, such as 320 or 640 bit, the RLC protocol can not sustain the peak data rate of the physical layer in HS-DSCH, let alone the peak data rate targeted for HSPA evolution. There are several solutions to solve this protocol stalling problem, such as increasing the RLC PDU size or sequence number space. However, a slightly different approach, Flexible RLC, has been adopted for Rel-7. The Flexible RLC concept is based on the Packet Centric RLC concept presented in [3] with necessary modifications to fit the specific properties of HSPA evolution. In [3] it was proposed that the RLC PDU size could be variable such that the PDU size is selected to exactly match the incoming SDU size (e.g. IP packet). This allows for high peak data rates since the IP packet size for high data rate applications typically is large and results in low header overhead and completely eliminates the padding that is present in Rel-6 due to the fixed RLC PDU size. The difference in overhead between fixed and flexible RLC PDU sizes is shown in Figure 3 with the RLC header sizes taken from the Rel-6 RLC protocol. As can be seen, the flexible RLC solution provides lower overhead for all SDU sizes. The relative improvement in terms of padding is particularly large for small SDU sizes when an SDU segments unfavorably into a fixed RLC PDU.

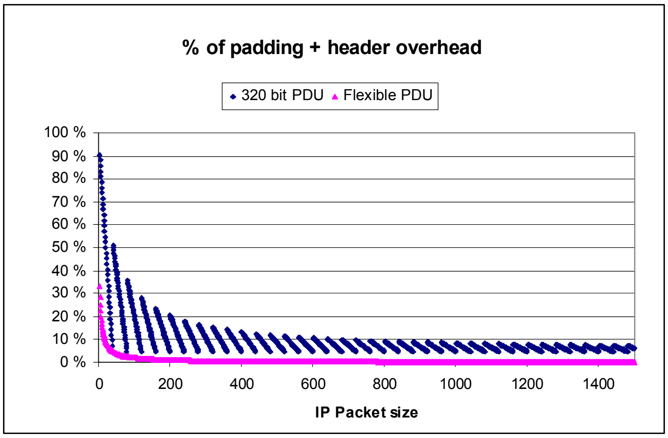

Figure 3. The sum of the padding and header overhead relative to the IP packet size for fixed and flexible RLC PDU sizes.

In HSPA evolution the RLC segmentation possibility is kept, and the RLC SDUs are segmented if the SDU size exceeds an upper limit. The main reason for this is to increase the retransmission efficiency in cases of MAC HARQ failure.

Segmentation has been introduced in MAC such that the RLC PDUs are segmented into MAC PDUs with size that is adapted to the momentary radio conditions. It would appear natural that RLC directly creates RLC PDUs with a size adapted to the radio conditions. This is, however, not feasible since the momentary radio conditions are not known in RLC.

Rel-7 also introduces the possibility to multiplex data from different radio bearers in the same TTI through MAC multiplexing which increases the resource efficiency for mixed service scenarios.

IV. CONCLUSION

We have presented a detailed overview of the HSPA Evolution, which consists of a number of enhancements. The enhancements include Multiple Input Multiple Output (MIMO), Higher order modulation, Control channel improvements, and Protocol enhancements. With these improvements it should be possible to reach performance close to the performance goals of the Long Term Evolution. However, due to smaller bandwidth and longer transmission time interval, some LTE targets will not be reachable.

REFERENCES

[1] 3GPP TR 25.913, “Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)”.

[2] H. Ishii et al. “Effects of UE capabilities on high speed downlink packet access in WCDMA systems”, Proceedings of the VTC 2004 Spring.

[3] H. Ekström et al. “Technical solutions for the 3G long-term evolution”, IEEE Communications Magazine, v. 44, March 2006, pp. 38 – 45.

[4] E. Dahlman et al. “The 3G Long-Term Evolution – Radio Interface Concepts and Performance Evaluation”, Proceedings of the VTC 2006 Spring.

[5] R1-063339, “Further analysis of HSDPA in CELL_FACH state”, Nokia, 3GPP TSG-RAN WG1 Meeting #47

[6] Ericsson, “64QAM for HSDPA – Link-Level Simulation Results,” R1-062264, 3GPP TSG-RAN WG1 Meeting #46, September, 2006.

[7] Ericsson, “64QAM for HSDPA – Performance with Realistic Algorithms,” R1-063499, 3GPP TSG-RAN WG1 Meeting #47, November, 2006.

[8] G. E. Bottomley, T. Ottoson, and Y.-P.E. Wang, “A generalized RAKE receiver for interference suppression,” *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1536-1545, August 2000.

[9] S. J. Grant, K. J. Molnar, and G. E. Bottomley, “Generalized RAKE receivers for MIMO systems,” *Proceedings of the VTC 2003 Fall.*

[10] 3rd Generation Partnership Project, “ UE Transmission and Reception,” TS 25.101 v7.4.0, June, 2006.

[11] R1-062266, “16QAM for HSUPA – Link-Level Simulation Results,” Ericsson, 3GPP TSG-RAN WG1 #46

[12] R1-070415, “Enhanced Phase Reference for 16QAM’, Qualcomm, Ericsson, 3GPP TSG-RAN WG1 #47-bis.

[13] 3GPP TR 25.815, “Signalling enhancements for Circuit-Switched (CS) and Packet-Switched (PS) Connections”

[14] R1-062879, “Coding for ACK/NACK in MIMO operation”, Ericsson, 3GPP TSG-RAN WG1 Meeting #46bis.