Hybrid ARQ Based on Rateless Coding for Wireless Mesh Networks

GTTI 2009 - Sessione Reti

Emanuele Goldoni, Matteo Lanati, Pietro Savazzi, Lorenzo Favalli
Dipartimento di Elettronica
Università degli Studi di Pavia
Via Ferrata 1, 27100 Pavia, Italy
Email: {name.surname}@unipv.it

***Abstract*—This paper shows an incremental redundancy hybrid automatic repeat request (HARQ) scheme, based on rateless coding, for possible application to a distributed wireless network. The proposed approach is demonstrated to be effective under energy consumption and rate constraints. This objective is pursued by increasing the average throughput, limiting the length of re-transmitted packets thanks to rateless coding. Performance are evaluated in terms of average energy consumption, comparing the presented solution with a HARQ based on traditional coding techniques. Simulation results are obtained using a network simulation model for rateless code transmissions applied to mesh systems.**

I. INTRODUCTION

The design of transmission techniques for wireless radio systems is challenged not only by the need to provide higher and higher throughputs to data intensive applications but also by the need of preserving as much as possible battery life. Under different constraints this is true for different topologies of radio systems from mobile to the newer field of sensor networks.

Throughput and energy efficiency are obviously not independent as we may measure the energy per useful bit delivered to the final user. Although this may not be the only source for power consumption [1], it is commonly assumed that transmission accounts for the largest part of the power budget and in particular it has been stressed [2] that the largest impact is related to the specific medium access (MAC) procedures: properly governing the active and sleep operating modes of the battery operated devices is the first step to minimize the usage of the radio front end.

One final remark must be made on the adaptability of the techniques to the actual transmission conditions. Reliable transmission must be ensured also in channels with widely varying qualities and adaptability may be achieved by adaptive modulation and coding techniques. Most recent wireless standards make use of these techniques and introduce error control via hybrid automatic repeat request (HARQ) protocols that mix ARQ mechanism with forward error correction (FEC) protection coding, in order to avoid waste of resources during retransmissions. As a matter of fact, protecting user information with a FEC code allows to minimize the number of retransmissions while an ARQ strategy permits to limit the redundancy introduced with respect to a FEC-only solution. The fact that a proper mix of FEC and ARQ may outperform each of the single schemes alone is well known [3]. The improved energy efficiency of these techniques has also been demonstrated in several works [4], [5].

As an alternative, a more recent approach providing higher flexibility has been based on fountain codes (FC) [15]. In this type of codes, a source having to transmit a certain amount of data, divided in N packets, sends these packets as a continuous stream that flows through the channel, from which the analogy with a water fountain. The receiver needs to collect at least N encoded packet, not necessarily in the same order as they are produced, to correctly decode the source. A fountain code is rateless because the number of encoded packets is not fixed and the coding rate of the transmission depends on the number of packets collected. Studies on the energy efficiency of such codes, have been performed for example in [6], [7], [8].

To the best of our knowledge, the joint exploitation of HARQ and fountain codes (FC) has been exploited in very few works [9], [10], [13]. In this paper we study the effectiveness of the joint HARQ-FC technique in the scenario of a wireless mesh network based on on a OFDM-TDM access scheme similar to that of most present and upcoming standards.

Mesh networks are commonly built based on the 802.11 standard which offers high flexibilty at the expenses of degradation in performance. Synchronous mesh networks have also been studied and issues related outlined [11], [12].

We focus on a generic fountain code, using a simplified cluster network model [14] that makes use of link channel capacities to evaluate an hybrid ARQ scheme based on rateless coded message retransmissions. The better throughput obtained, with respect to those obtained by more traditional schemes, may be easily viewed as a power consumption gain.

In the next section, a brief description of Hybrid ARQ techniques is presented, while the subject of section III is the simulation model of the proposed scheme. Finally performance measured by simulation results and conclusions are given in the last two sections of this paper.

II. HYBRID ARQ

HARQ merges two well-known techniques employed in case wrong bits are received: forward error correction (FEC) protection coding and ARQ retransmission scheme. These two approaches are usually seen as alternative solutions to the same problem of guaranteeing error free communications. As a matter of fact, FEC can reduce the use of the channel because it relies on redundancy and computational capabilities of a node, on the contrary, ARQ heavily exploits transmission resources and it can easily adapt to varying conditions. Effectiveness of FEC has always been a milestone in energy-constrained scenarios, since computation is less expansive than other activities involving the radio interface, but recently researchers started to consider the benefits that a retransmission scheme could introduce.

A direct implementation of HARQ consists in adding to each packet some bits of redundancy, with the possibility of retransmission in case errors occur. This is a short description of Type I HARQ, while Type II HARQ can be further subdivided in Chase Combining (CC), or diversity-combining, and Incremental Redundancy (IR), or code-combining. In CC, when the sender receives a negative acknowledgement behaves as in the Type I case, it simply retransmits the whole packet, applying the same FEC protection. The difference with respect to the basic implementation consists in the fact that the receiver does not discard erroneous blocks, they are rather stored in a buffer to be compared with replicated ones, so it is possible to take advantage of energy and diversity gain. In IR, each retransmission is different from the previous one because additional redundant information is given, for example, simply varying the puncturing scheme. In this case the block is self-decodable because systematic bits are included, however it is possible also to add only redundancy bits to save much more resources, adopting a so-called Type III scheme [21].

Starting from this short description, it is clear that the main HARQ advantage is the possibility to allow a higher block error rate, using high modulation order and code rate, which reflects a relevant throughput gain. The counterpart is the need for a reliable feedback channel and variable delays for handling error events.

The retransmission scheme adopted in our HARQ version is derived from those proposed for HSDPA (High Speed Downlink Packet Access) [22] UMTS (Universal Mobile Telephone System) evolution and WiMax. In particular, it is not based on Selective Repeat, but it guarantees the same level of resiliency through multiple Stop and Wait (SaW) instances. The maximum number of SaW processes assigned to a user should be enough to cover with a reasonable margin the round trip time. Consequently, at least the first acknowledgement can come back to the sender before the pool of available instances is empty, allowing a continuous transmission if the total traffic volume is low. Since we are supposing to deal with a time-division system, we fixed the average round trip time to six time slots: a maximum number of eight parallel transmissions should be enough to avoid the starvation of the source. Figure 1 can help in understanding how HARQ retransmission scheme works. A new SaW process is is started each time a user is given the possibility to transmit. This is particularly useful in case of continuous transmission in order to avoid resource wasting, since the instance just begun is still waiting for the acknowledgment. For example, in the figure we can see that the first process, P1, is assigned to user number 1 (SAW1), the second process P2 is assigned to the second user (SAW2) and so on. Later on, when the user 2 is given again the possibility to transmit, the data block is assigned to the first process available, P5 (SAW2/P5) When a positive acknowledgement is received (the block marked with POS ACK in the receiving side of Figure 1), the associated SaW process can be reused for a new transmission. If the acknowledgement is negative (marked as NEG ACK in Figure 1), the erroneous block of data is retransmitted by the same process, the first time the user has the possibility to use the channel. For instance, the block transmitted by the first user using the first process (SAW1/P1) receives a negative acknowledgment, so it is retransmitted using the same process P1.

An external observer looking at this retransmission scheme as a black box system can see the same behaviour of a fixed-size moving transmission window, whose goal is to avoid ambiguity in attribution of feedback messages. The advantage is that SaW processes are easier to manage with respect to a buffer keeping track of what should be sent again.

III. PROBLEM FORMULATION

A. Rateless Coding Modelling

Rateless codes may be used as an alternative of Hybrid ARQ schemes [17], especially for communication systems with delay and power consumption constraints. Fountain [18] and Raptor codes [19] are the representatives of such techniques available in literature, and are considered rateless, in the sense that the number of encoded symbols that can be generated from the source message is potentially limitless. Furthermore, in presence of good channel conditions, a little redundancy may be enough to recover the transmitted message. Considering the transmissions of coded messages over a block fading channel, the realized capacity of a sequence $\vec{h} = (h_1, h_2, ..., h_n)$ of channel state information, may be computed as [17]:

$$C(\vec{h}) = \frac{1}{n}\sum_{i=1}^{n} \log_2(1 + \gamma |h_i|^2) \tag{1}$$

where γ is the transmitted SINR, including the fading multiplicative noise, while $C(\vec{h})$ may be viewed as the transmission rate supported by $\vec{h}$. When using a codeword of length n to transmit k symbols of information, the theoretical lower limit of the outage probability may be given by [17]:

$$P_{out} = P\left(\frac{k}{n} > C(\vec{h})\right) \tag{2}$$

In order to evaluate an incremental redundancy ARQ based on rateless coded message transmissions, eq. (1) and (2) may be used in a cluster traffic simulator. Whenever the recovery of the received codeword fails, the transmitter uses the channel state information to compute the additional transmission redundancy required to correctly decode the message, using

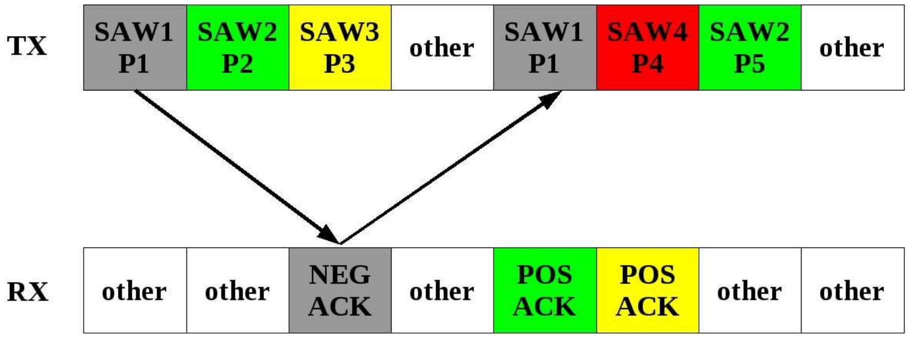

Fig. 1. Hybrid ARQ retransmission scheme

eq. (1). The lower bound of eq. (2) is also employed as actual simulation method for computing the probability of decoding error.

B. Simulator Description

In our work we are interested in studying energy and throughput performances of wireless nodes forming a cluster headed by a super node which has the function of collecting data, for example because it is the gateway to Internet. Our remarks starts from the comparison between a traditional ARQ, implemented by the Chase Combining algorithm associated to a turbo code protection scheme, and an Incremental Redundancy solution based on Fountain Codes.

The simulator consists of four basic blocks as shown in Fig. 2.

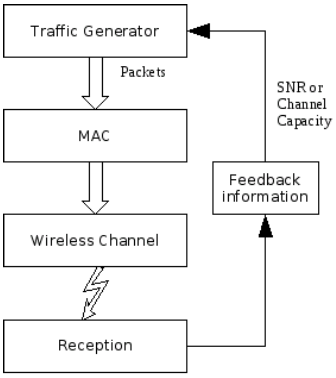

Fig. 2. Cluster simulator scheme

First of all, a traffic generator produces packets according to a profile defined at the beginning of the simulation. In this particular case, in order to stress the entire system, we consider that all users have backlogged traffic so that every time a user is given the possibility to transmit, his buffer has queued packets. Packets are then delivered to the second block, which summarises the functions of the MAC (Medium Access Control) layer. This component has been simulated starting from some assumptions that, in our opinion, simplify the analysis shown in the next section. In particular, we refer to a MAC providing access to the channel on the basis of a contention-free mechanism, sharing resources according to a TDMA (Time Division Multiple Access) scheme.

We choose to implement a TDMA-based MAC layer whose slot assignment is managed by the gateway. Here, nodes are hierarchically grouped into smaller clusters prior to network operations and a TDMA scheme is adopted for intra-cluster communication. Sychronization among the cluster heads is not explicitly considered in this work and interaction among the different clusters is only tacken into account as a general interference level included in the definition of equations 1 and 2. Within each cluster a special node, called cluster head, centralizes all control operations and collects data from local nodes before sending them to the network gateway.

For the sake of completeness, in Figure 2 the physical layer is lacking, however, for ease of comparison withe results reported in [14] we assume that there is an underlying OFDM (Orthogonal Frequency Division Multiplexing) system made up of 48 subchannels. Each node can be assigned one ore more subchannels, in order to increase the coding rate without damaging the efficiency, that is the ratio between payload and redundancy.

Besides all the tasks previously mentioned, MAC has also to segment packets in smaller blocks to fit the radio channel for transmission. There is a significant difference if a FC approach is used instead of a traditional HARQ one.

In the traditional case both the data block size and the coding rate of the turbo code have fixed values. Precisely, data blocks are 200 bits wide, while there are two possible coding rates, namely 1/3, and 2/3. Table I shows the minimum SNR ensuring a 10^{-6} BER with the modulation and coding scheme specified in that particular row. As we can see, if the feedback information on the channel shows good transmission conditions, more blocks are coded as a single entity and then sent together to optimise resource exploitation. The same table also shows the number of orthogonal OFDM subchannels used for the transmission. Supposing that a coding rate equal to 2/3 is a good compromise between protection and efficiency, if the channel's conditions get worst and we would like to enforce the transmission switching to a stronger protection (rate equal to 1/3), we are doubling the number of redundancy bits. In order to maintain payload size, since splitting would have been a complication of the protocol, we decided to assign two

TABLE I
MODULATION AND CODING SCHEME

E_b/N_0	Coding Rate	OFDM Sub channel(s)	No. of data blocks	Modulation
-0.52	$1/3$	2	1	QPSK
3.15	$2/3$	1	1	QPSK
4.7	$1/3$	2	2	16QAM
6.7	$2/3$	1	2	16QAM
7.35	$1/3$	2	3	64QAM
8.65	$2/3$	1	3	64QAM

TABLE II
CLUSTER ENVIRONMENT SPECIFICATIONS

Average path loss:	$128.1 + 37.6 * log_{10}(d)$; d in Km
Slow fading, mean value:	0 dB
Slow fading, variance:	5 dB
Number of sub channels per time slot:	48
Time slot duration:	0.5 ms

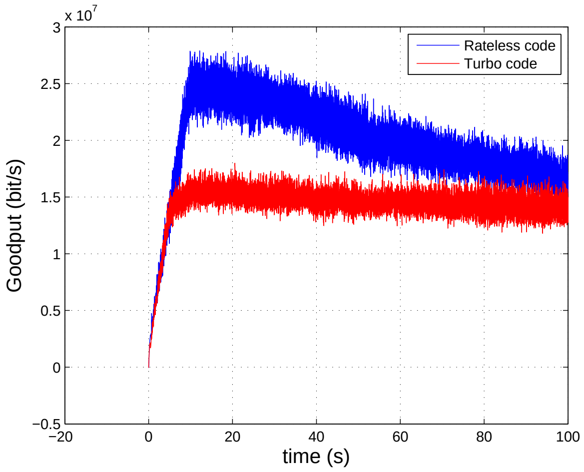

Fig. 3. Goodput versus time in the cluster

sub channels to carry the extra redundancy. In case of error, the retransmission involves the entire block, but during the decoding phase the new information is combined with the old one previously stored. The SNR needed for a correct decoding is reduced, according to the CC scheme.

On the contrary, using HARQ based on a FC, the ratio between payload and redundancy is decided every time slot, depending on the rate estimated for a successful transmission at the current channel conditions. The feedback information is a measure of the channel capacity and the coding rate is fixed in that particular time slot in order to minimise the redundancy needed for a successful reception. If the rate requested is lower than one third, throughput can benefits from a more robust modulation, if available. On the contrary, if channel state suddenly gets worse and parity bits can not guarantee a successful decoding, a retransmission is triggered. The erroneous data block, including its redundancy, is kept in a buffer and the retransmission has to be intended in an IR sense, since it does not include the payload, but only extra parity bits. The goal is to combine the stored information with the new one, increasing the overall coding rate of corrupted data. The unfilled space in the time slot used for retransmission is then used for new data, protected according to the feedback information.

The third block takes into account the transmission of data blocks over a block fading wireless channel, where fading attenuation values are updated every time slots, introducing a correlation coefficient. This model follows the specifications [25] [26], reported in Table II.

Finally, the last block has to decide whether the received codewords are correct or erroneous and it has also to provide the feedback information. In the traditional HARQ scheme simulation, the decision process needs to know the probability of error, analytically evaluated considering the modulation used and a corrective factor due to error protection coding. The SNR represents the feedback information and it is used as index for Table I.

In the HARQ scheme simulation based on FC, the error probability is not needed, instead the realized capacity of the wireless medium is calculated by using Equation 1. If the rate used for transmission is lower than the capacity, it is assumed that the applied redundancy was sufficient to ensure a correct reception.

IV. SIMULATION RESULTS

The two HARQ approaches have been simulated in the same channel conditions reported in Table II. The input traffic represents a challenging situation, that is 200 nodes entering in the network in the first ten seconds of simulations, trying to support a source originating traffic at a constant rate of 350 Kb/s. Each node is at least 150 metres away from the gateway, and this can be considered a long distance related to the power levels involved. We assume a simple model where the power consumption of the radio circuitry is 25 mW when transmitting or receiving data, while the node consumes a negligible amount of power in sleep state.

The first aspect we consider is the total goodput of nodes making up the cluster. In Figure 3 results for the turbo codes refer to the traditional Chase Combining HARQ scheme, while the other one shows HARQ implementing rateless coding. Fountain code outperforms the other implementation, especially when all users are entered the system and are originating traffic. This is due to efficient use of bits in a time slot, introducing only the needed redundancy. The fixed scheme employs a minimum coding rate of 2/3 while the proposed one sends only the redundancy strictly needed. In other words, if the feedback information shows channel conditions good enough, data can be sent with very low protection, or even unprotected.

Same conclusions can be drawn looking at the statistics of transmission delays in Figure 4. Here, we show the cumulative density distribution of time intervals between the beginning of a data block transmission and its correct reception. The great

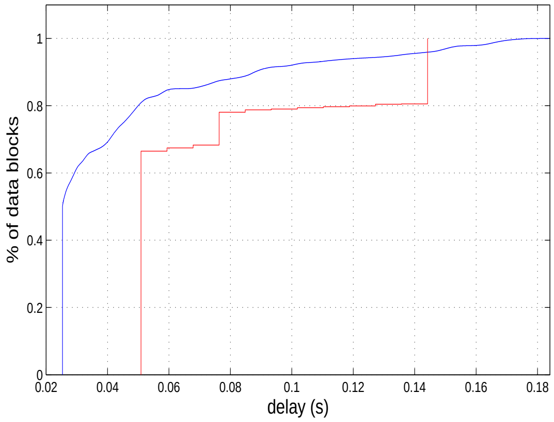

Fig. 4. CDF of transmission delays

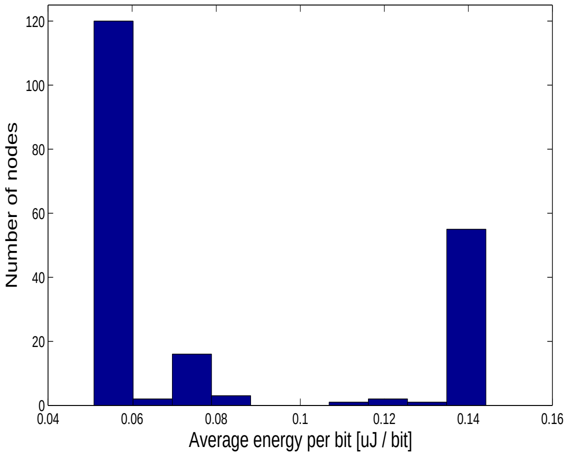

Fig. 5. Average energy per bit (Turbo coding)

majority of data blocks reduces this amount of time, in fact in more than 90% of cases the red line lies below the blue one. The reason is simply that no time is wasted in transmitting an unnecessary protection.

Finally, we can also compare the two HARQ implementations from the power consumption point of view, showing some statistics regarding the energy per bit needed to receive the payload stored in a data block.

The average energy per bit is 0.1187 μJ for the Turbo coding HARQ, while this value is reduced to 0.083 μJ in the other case. A more detailed statistic distribution is represented in Figures 5 and 6 for Turbo coding and Rateless coding, respectively. As already seen for the interarrival time, switching from the legacy implementation to the proposed one, a limited number of users has poorer performances, but the vast majority reduces the energy needed. A clever usage of redundancy guarantees a 20-30 % of energy reduction.

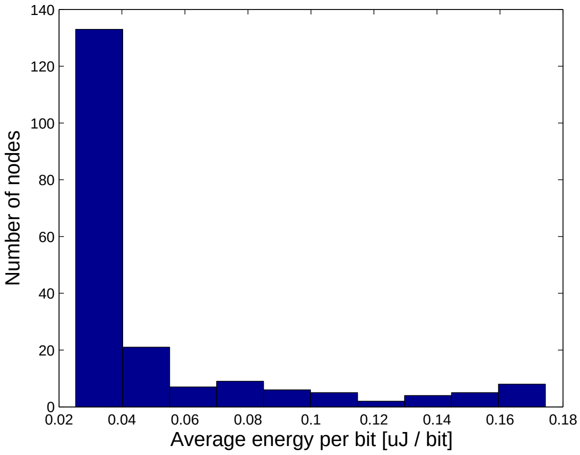

Fig. 6. Average energy per bit (Rateless coding)

V. CONCLUSIONS

Simulation results show that the novel proposed HARQ scheme may be applied to wireless mesh networks, obtaining a substantial gain with respect to a more traditional approach. Obtaining a better goodput means more robust performance from the point of view of energy consumption constraints, making the proposed approach useful for wireless applications. Ongoing works are devoted to try using network coding theoretic analysis in order to compute asymptotic performance limits of the proposed technique.

REFERENCES

[1] A.J. Goldsmith, S.B. Wicker, “Design challenges for energy-constrained ad hoc wireless networks,” IEEE Wireless Communications, Vol. 9, No. 4, Aug. 2002, pp. 8 - 27.

[2] M. Stemm, R. H. Katz, “Measuring and reducing energy consumption of network interfaces in hand-held devices,” IEICE Transactions on Communications, vol. E80-B, no. 8, pp. 112531, 1997.

[3] S. Lin, D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1983.

[4] T. Ji, W.E. Stark, “Turbo-Coded ARQ Schemes for DS-CDMA Data Networks Over Fading and Shadowing Channels: Throughput, Delay and Energy Efficiency,” IEEE J.Select. Areas Commun., vol. 18, Aug. 2000, pp. 1355-1364.

[5] I. Staojev, O. Simeone, Y. Bar-Ness, D.H. Kim, “On the energy efficiency of Hybrid-ARQ protocols in fading channels,” IEEE ICC ’07, June 24-28, 2007, pp. 3173-3177

[6] D.Starobinski, W.Xiao, X.Qin, A.Trachtenberg, “Near-optimal data dissemination policies for multi-channel, single radio wireless sensor networks,” IEEE INFOCOM, 2007.

[7] N. Rahnavard, F. Fekri, “CRBcast: a collaborative rateless scheme for reliable and energy-efficient broadcasting in wireless sensor networks,” Fifth International Conference on Information Processing in Sensor Networks, 2006, (IPSN 2006), 19-21 April 2006, pp. 276 - 283.

[8] A.W. Eckford, J.P.K. Chu, R.S. Adve, “Low-Complexity Cooperative Coding for Sensor Networks using Rateless and LDGM Codes,” IEEE ICC 2006, June 2006 Vol. 4, pp. 1537-1542.

[9] P. Casari, M. Rossi, M. Zorzi, “Towards Optimal Broadcasting Policies for HARQ based on Fountain Codes in Underwater Networks,” in, Proc. of WONS 2008, Garmish-Partenkirken, Jan. 23-25, 2008, pp. 11 - 19.

[10] M. Rossi, G. Zanca, L. Stabellini, R. Crepaldi, A.F. Harris, M. Zorzi, “SYNAPSE: A Network Reprogramming Protocol for Wireless Sensor Networks Using Fountain Codes,” in proc. of SECON ’08, san Francisco, USA, June16-20, 2008, pp. 188 - 196.

[11] G. K, Venkatesan, K. Kulkarni, “Wireless backhaul for LTE - requirements, challenges and options,” in 2nd International Symposium onAdvanced Networks and Telecommunication Systems, 2008. ANTS ’08, Dec. 15-17, 2008 pp. 1-3.

[12] H.Y. Wei, S. Ganguly, R. Izmailov, Z. J. Haas, “Interference-Aware IEEE 802.16 WiMax Mesh Networks,” in Proc. of 61st IEEE Vehicular Technology Conference (VTC 2005 Spring), Stockholm, Sweden, May 29-June 1, 2005.

[13] A.F. Molisch, N.B. Mehta, J.S. Yedidia, Zhang Jin, “Performance of Fountain Codes in Collaborative Relay Networks,” IEEE Transactions on Wireless Communications, Vol. 6, No. 11, Nov. 2007, pp. 4108 - 4119.

[14] L. Favalli, M. Lanati, P. Savazzi, “Hybrid ARQ Based on Rateless Coding for UTRAN LTE Wireless Systems,” Wireless Communications: 2007 CNIT Thyrrenian Symposium, 2007, Springer Verlag.

[15] M. Luby, “LT Codes”, in Proc. of 43rd IEEE Symposium on Foundations of Computer Science (FOCS), November 16-19, 2002.

[16] H. Ekström, A. Furuksär, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, M. Wahlqvist, “Technical Solutions for the 3G Long-Term Evolution,” IEEE Communications Magazine, Vol. 44, Issue 3, March 2006, pp. 38-45

[17] J. Castura,“Rateless coding over fading channels,” IEEE Commun. Letters, Vol. 10, No. 1, Jan. 2006. pp. 46-48.

[18] D.J.C. MacKay, “Fountain codes,” IEE Proc Commun, Vol. 152, No.6, December 2005, pp. 1062-1068.

[19] A. Shokrollahi, “Raptor codes,” IEEE Trans. Information Theory, Vol. 52, No.6, June 2006, pp. 2552-2567.

[20] 3GPP TR 25.814, v. 7.1.0, ”Physical layer aspects for evolved universal terrestrial radio access (UTRA),” release 7. http://www.3gpp.org.

[21] Y. Wang, L. Zhang, D. Yang, Performance analysis of type III HARQ with turbo codes, Proc. of 57th IEEE Vehicular Technology Conference (VTC-Spring 2003), Seoul, ROK, 22-25 Apr. 2003, Vol. 4, pp. 2740-2744.

[22] 3GPP, High Speed Downlink Packet Access (HSPDA); Overall Description , (Release 5), TS 25.308, Mar. 2002.

[23] S. Ganeriwal, R. Kumar, M. B. Srivastava, “Timing-sync protocol for sensor networks,” In SenSys ’03: Proceedings of the 1st international conference on Embedded networked sensor systems, pages 138–149, New York, NY, USA, 2003. ACM.

[24] M. Maroti, B. Kusy, G. Simon, and A. Ledeczi, “The flooding time synchronization protocol,” In SenSys ’04: Proceedings of the 2nd international conference on Embedded networked sensor systems, pages 39–49, New York, NY, USA, 2004. ACM.

[25] Recommendation ITU-R M.1225, Guidlines for evaluation of radio transmission technologies FOR IMT-2000 (Question ITU-R 39/8)

[26] W.S. Jeon, D.J. Jeong, B. Kim, “Packet Scheduler for Mobile Internet Services Using High Speed Downlink Packet Access,” IEEE Trans. on Wireless Commun., Vol. 3, No. 5, September 2004, pp. 1789-1801.