A Future Radio-Access Framework

David Astély, *Member, IEEE*, Erik Dahlman, *Member, IEEE*, Pål Frenger, *Member, IEEE*, Reiner Ludwig, Michael Meyer, Stefan Parkvall, *Senior Member, IEEE*, Per Skillermark, and Niclas Wiberg

***Abstract*—This paper discusses the requirements on future radio access and, based on the requirements, proposes a framework for such a system. The proposed system based on orthogonal frequency-division multiplexing supports very low latencies and data rates up to 100 Mb/s with wide area coverage and 1 Gb/s with local area coverage. Spectrum flexibility is identified as one main requirement for future systems, and the proposed framework can be deployed in a wide range of spectrum allocations with bandwidths ranging from 2.5 to 100 MHz. Multihop relaying, useful to extend the range for the high data rates, and multiple-antenna configurations are integral parts of the framework. A packet-centric approach is taken for the dataflow processing, implying that the scheduling mechanism and the retransmission protocol operate on complete packets rather than segments thereof, thus allowing for cross-layer optimization. Finally, numerical evaluations are provided, illustrating the feasibility of future very wideband radio access.**

***Index Terms*—Beyond 3G, packet data, radio access.**

I. Introduction

DEPLOYMENT of third-generation or 3G wireless systems based on the wideband code-division multiple access (WCDMA) and cdma2000 wireless technologies is now gaining momentum around the world. An evolution of the 3G wireless technologies has also already been initiated, e.g., by the introduction of high-speed downlink packet access (HSDPA) [1], [2] and enhanced uplink [3], [4] as extensions to WCDMA, providing peak data rates above 10 Mb/s in the downlink direction and up to 5 Mb/s in the uplink direction. The introduction of HSDPA and enhanced uplink also implies a significant reduction in the delay introduced by the radio-access network. This evolution of 3G wireless technologies will continue in the future, targeting further improvements in system performance and service provisioning, for example the possibility to provide up to 25 Mb/s downlink data rates with wide-area coverage and up to 100 Mb/s in short-range scenarios [5]. As part of this, it is anticipated that the 3G wireless technologies will be extended to support transmission bandwidths up to 20 MHz.

However, in parallel to these activities related to the evolution of current 3G wireless technologies, there is also an increased research effort on future radio access, sometimes referred to as fourth-generation or 4G radio access. Such future radio access is anticipated to take the performance and service provisioning of wireless systems even one step further, providing data rates up to 100 Mb/s with wide-area coverage and up to 1 Gb/s with local-area coverage, fulfilling the requirements for Beyond IMT-2000 systems [6]. Realistically, data rates of that magnitude can only be provided efficiently by the introduction of even wider transmission bandwidth, in the order of 100 MHz. Such a wide transmission bandwidth is not possible to deploy in the frequency bands currently allocated for cellular communication. A prerequisite for a future very wideband radio access is therefore the allocation of new spectrum, to be identified at the World Radio Conference (WRC) to be held in 2007. Leading up to the WRC, there is an intense effort in the research community to develop concepts for and prove the feasibility of very wideband radio access. One such activity is the WINNER project [7], part of the European 6th Framework Program. Other extensive and important research activities are carried out in China [8] and Japan [9].

As already mentioned, the possibility to provide very high data rates is one of the targets for a future very wideband radio access. Another important target is a further reduction of the delay introduced by the radio-access network, targeting the possibility for an end-to-end round-trip time less than 10 ms. The end-to-end round-trip time in fixed broadband networks can be expected to be less than 10 ms, and it is reasonable to set a similar design requirement on wireless systems as similar data rates are targeted. Hence, the radio-access network should contribute at most a few milliseconds to the overall end-to-end delay. A low round-trip time is essential as the end-to-end protocols may otherwise significantly limit the achievable data rate. One example hereof is the congestion control scheme of the transmission control protocol (TCP), where the round-trip time directly affects the achievable data rate [10].

Another identified requirement for future radio access is spectrum flexibility. Fundamentally, this implies that a new radio-access technology should be able to operate in many different spectrum allocations of different sizes and with different duplex properties. In practice, the requirement on spectrum flexibility implies a requirement on bandwidth flexibility, i.e., a future radio-access technology should be able to operate with a wide range of system bandwidths as well as flexibility in the duplex arrangement supported by the radio-access technology.

Assuming a conventional base-station deployment, the provisioning of very high data rates may imply dense, and consequently, costly network deployments. Investigation of new radio-network architectures, such as an extensive use of relaying, is therefore an important part of the research activities.

Manuscript received January 24, 2005; revised August 25, 2005.

D. Astély, E. Dahlman, S. Parkvall, and P. Skillermark are with Ericsson Research, Ericsson AB, SE-164 80 Stockholm, Sweden (e-mail: david.astely@ericsson.com; erik.dahlman@ericsson.com; stefan.parkvall@ericsson.com; per.skillermark@ericsson.com).

P. Frenger and N. Wiberg are with Ericsson Research, Ericsson AB, SE-581 12 Linköping, Sweden (e-mail: pal.frenger@ericsson.com; niclas.wiberg@ ericsson.com).

R. Ludwig and M. Meyer are with Ericsson Research, Ericsson GmbH, D-52134 Herzogenrath, Germany (e-mail: reiner.ludwig@ericsson.com; michael.meyer@ericsson.com).

Digital Object Identifier 10.1109/JSAC.2005.862420

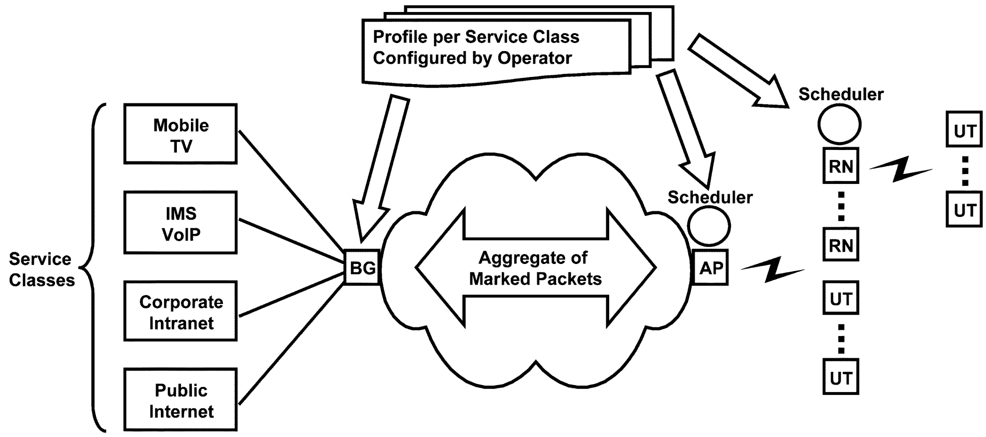

Fig. 1. Key elements of future radio-access networks.

In this paper, we present a framework for a future very wideband radio-access technology. The reasons to develop this basis for a new technology concept are twofold.

1) It will serve as a platform for the development and evaluation of new technology components, e.g., new radio-network solutions such as relaying and new multiple-antenna solutions.
2) It will be used for evaluations, the aims of which are to prove the feasibility of future very wideband radio access.

II. High-Level Overview of Radio-Access Framework

Fig. 1 shows some key elements of a future radio-access network, including the nodes of the radio access: *bearer gateways* (BGs), *access points* (APs), *user terminals* (UTs), and *relay nodes* (RNs) [11]. As already mentioned, relay nodes are envisioned as a potentially cost-efficient means to provide high-data-rate coverage through two-hop relaying (AP-RN-UT).

A central component of the presented radio-access framework is the scheduler located in each AP and RN. A separate scheduler exists for downlink and uplink traffic; although, the uplink could potentially be operated in a contention mode during times of low traffic load. The scheduler decision takes into account the requirements of different services (Mobile TV, IMS[1]-based voice-over-IP (VoIP), etc.), as well as the instantaneous radio-channel conditions. Traffic conditioning and packet marking are performed by the BG, which is where packet traffic enters the radio-access network in the downlink direction. The traffic handling of the BGs and the schedulers in the different APs and RNs are controlled by the network operator through a set of profile definitions.

We refer to the fundamental piece of information to be communicated over the radio interface as a "packet." It is expected that, in the future, transport of Internet protocol (IP) packets will dominate the traffic in wireless systems. Thus, a packet typically corresponds to an IP packet. However, a packet could also correspond to other kinds of information to be communicated over the radio interface, including, e.g., Layer 3 control signaling. A key characteristic of the outlined radio-access framework is that all types of information to be transported over the radio interface are processed in a more or less identical way.

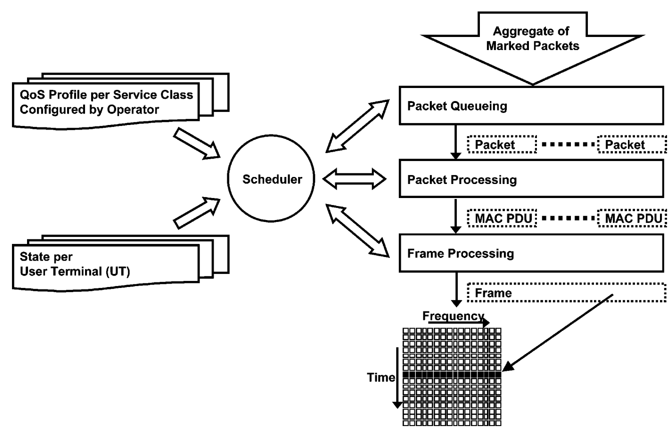

Fig. 2. Overall radio-access processing consisting of a scheduler controlling packet queueing, packet processing, and frame processing operations.

The overall radio-access processing consists of two main steps (see also Fig. 2).

1) *Packet processing*, including channel coding and the retransmission protocol. For each frame, the packet processing outputs a sequence of coded and multiplexed packets, referred to as a medium access control (MAC) *protocol data unit* (PDU). The packet processing is described in more detail in Section IV.
2) *Frame processing*, including different multiplexing and modulation steps and mapping to the basic physical resource. We assume orthogonal frequency-division multiplexing (OFDM)-based transmission, which implies that the basic physical resource can be expressed as a unit in the time/frequency grid. The frame processing is described in more detail in Section III.

The packet and frame processing is controlled by the scheduler. The scheduling function is further elaborated upon in Section V.

[1]IP multimedia subsystem (IMS) is a standardized architecture for providing voice and multimedia services using IP, including functions for roaming, charging, and bearer control.

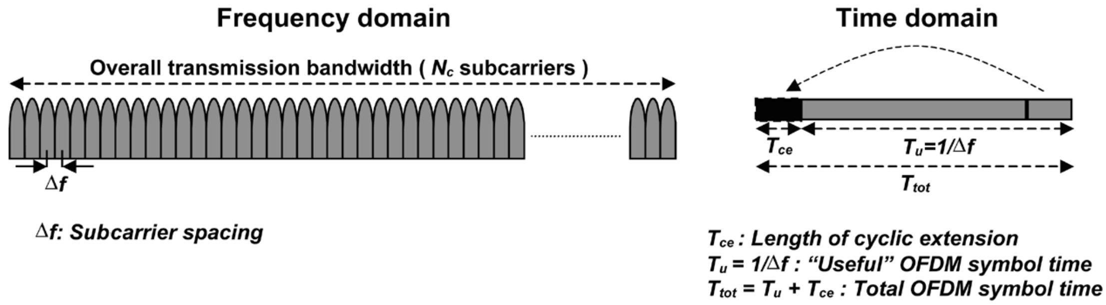

Fig. 3. Downlink transmission scheme—OFDM.

III. Frame Processing

This section describes the frame-processing part of the radio-access framework. The frame processing accepts MAC PDUs from the packet processing (see Section IV for details), carries out modulation and multiplexing, and maps the symbols to the basic physical resource, i.e., a unit in the time/frequency grid.

A. Downlink Transmission

For the downlink, the radio-access framework is based on conventional OFDM with cyclic extension [12]; see also Fig. 3. The selection of OFDM is much related to the fact that we are targeting system and transmission bandwidths much wider than, e.g., the transmission bandwidth of current 3G wireless technologies.

1) A wide bandwidth is typically associated with substantial radio-channel frequency selectivity. Due to a narrow bandwidth per subcarrier in combination with a cyclic extension, OFDM is inherently robust to such radio-channel frequency selectivity.
2) State-of-the-art radio-access technologies, such as the HSDPA extension of WCDMA [1], [2], achieve substantial system performance benefits from the use of channel-dependent scheduling in the time domain. With an increased system bandwidth, the fast time-domain channel variations utilized by the channel-dependent scheduling are reduced, due to increased frequency diversity. Thus, in order to fully benefit from channel-dependent scheduling, a future very wideband radio-access technology should allow for scheduling also in the *frequency domain*. OFDM transmission straightforwardly supports such frequency-domain scheduling by the dynamic allocation of different sets of subcarriers for transmission to different user terminals.

Regarding the cyclic extension, our current assumption is a single cyclic-extension length of approximately 4 μs, which covers the expected delay spread in most cellular scenarios for unicast transmission [13]. As an alternative, one could consider the use of a limited set of different cyclic extension lengths to be used in different scenarios.

1) One could consider the use of a significantly shorter cyclic extension, with a corresponding lower overhead, in short-range low-dispersive environments.
2) One could also consider the use of a significantly longer cyclic extension, e.g., for broadcast scenarios, where the cyclic extension should not only cover the time dispersion of the radio link, but also the difference in timing of the received signals from multiple cells.

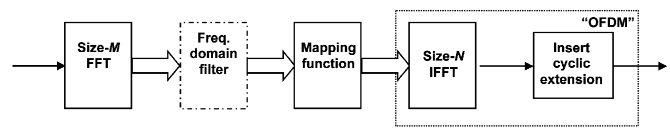

Fig. 4. Uplink transmission scheme—OFDM with precoding.

B. Uplink Transmission Scheme

An important requirement for any uplink transmission scheme is to allow for power-efficient user-terminal transmission. This speaks against the use of conventional OFDM with its inherent large peak-to-average power ratio (PAPR), at least in large-cell high-transmit-power scenarios. At the same time, it is well known that the impact of radio-channel frequency selectivity can also be handled by the introduction of advanced signal processing at the receiver side, e.g., the use of frequency-domain equalization [14], albeit at the cost of additional receiver complexity. For the uplink, this additional receiver complexity is typically less of an issue, compared to the downlink.

Another desirable characteristic of an uplink transmission scheme is the possibility for uplink intracell orthogonality, i.e., the option to operate without interference between simultaneous uplink transmissions within the same cell. One straightforward way to achieve uplink orthogonality is time-domain scheduling, i.e., the introduction of an intracell TDMA component. Time-domain scheduling is, e.g., part of the enhanced uplink extension of WCDMA [3], [4]. However, in case of a very wide overall system bandwidth, such as 100 MHz, being able to separate users in the time domain only would imply a very high instantaneous per-user transmission bandwidth. In many situations, a user terminal would not be able to utilize such a wide bandwidth efficiently, as the available transmit power, rather than the bandwidth, would limit the achievable data rates. Thus, in addition to separation in the time domain, an uplink transmission scheme should support separation also in the frequency domain, i.e., an intracell frequency-division multiple access (FDMA) component, including support for dynamic variations of the instantaneous transmission bandwidth.

A transmission scheme satisfying the above requirements is illustrated in Fig. 4. In principle, the transmission scheme can be described as conventional OFDM with fast Fourier transform (FFT)-based precoding, where the OFDM modulation is

TABLE I
MAIN PARAMETERS OF OFDM MODULATION FOR DIFFERENT SYSTEM BANDWIDTHS

System bandwidth [MHz]	2.5	5.0	10	20	40	80	100
Number of used subcarriers, N_c	96	192	384	768	1536	3072	3840
Subcarrier spacing, Δf [kHz]	24.4						
Number of OFDM symbols per frame	16			8			
Number of subcarriers per chunk	8			16			
Frame duration [ms]	0.72			0.36			
Chunk bandwidth [kHz]	195			390			
Number of chunks per frame	12	24	48	48	96	192	240

essentially the same as for the downlink (see Fig. 3) including the same subcarrier spacing and the same length of the cyclic extension. Selecting the precoding as a size M FFT and mapping the outputs of the FFT to consecutive inputs of the size N OFDM-modulator $(M \leq N)$ results in a low-PAPR signal, where the bandwidth of the transmitted signal depends on the size of the FFT. As illustrated in Fig. 4, additional spectrum shaping by means of frequency-domain filtering can be introduced to further reduce the PAPR of the transmitted signal.

Clearly, by making the precoding transparent, the transmission scheme illustrated in Fig. 4 also allows for conventional OFDM transmission. The use of conventional OFDM for the uplink can be useful in cases where the need for power-efficient uplink transmission is less of an issue, for example in case of short-range low-power communication. Another case when the power-amplifier efficiency is less critical is uplink transmission from relay nodes to the access point. Hence, with the structure in Fig. 4, the most suitable transmission scheme can be selected based on the instantaneous radio conditions. Note that Fig. 4 is a conceptual description. Depending on the pulse shape chosen, an implementation in the time-domain may or may not be preferred.

It can also be noted that, by assuming a slightly different mapping function that distributes the FFT output evenly to the OFDM subcarriers instead of a mapping to consecutive OFDM carriers, the uplink transmission scheme of Fig. 4 would implement interleaved FDMA [15] and the related variable spreading and chip repetition factor CDMA (VSCRF-CDMA) transmission scheme proposed in [16]. These transmission schemes essentially satisfy the same requirements—low PAPR for high power-amplifier efficiency and possibility for uplink orthogonality—as the transmission scheme of Fig. 4. The main difference is whether mapping is done to consecutive subcarriers, as is assumed in this paper, or to distributed equally spaced subcarriers, as is the case for interleaved FDMA and VSCRF-CDMA. Equally spaced mapping can provide additional diversity in frequency-selective fading channels when the transmission is only using a fraction of the overall bandwidth but has stricter requirements on time alignment among the terminals.

C. Spectrum Flexibility

As already mentioned, spectrum flexibility, in practice bandwidth and duplex flexibility, is an important requirement for a future radio-access technology. The radio-access framework presented in this paper supports a flexible system bandwidth for both uplink and downlink by adjusting the number of OFDM subcarriers, from 96 to 3840, as illustrated in Table I. In this way, the bandwidth can be varied from 2.5 MHz up to at least 100 MHz. Note that different system bandwidths can very well be used for uplink and downlink. The selection of the subcarrier spacing, which is 24.4 kHz in Table I, is a compromise between the overhead from the cyclic prefix, which calls for a smaller subcarrier spacing, and the influence from phase noise in the oscillators and the possibility to operate in high Doppler scenarios; both call for a larger subcarrier spacing.

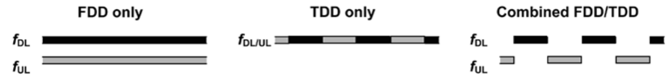

Fig. 5. Duplex arrangements of future radio access (from terminal point of view).

As part of the fulfillment of the requirement on spectrum flexibility, we further believe that a future radio access should support operations with different duplex arrangements; see also Fig. 5:

1) *frequency-division duplexing (FDD) only*—overlapping downlink and uplink transmission in different frequency bands;
2) *time-division duplexing (TDD) only* —nonoverlapping downlink and uplink transmission in the same frequency band;
3) *combined FDD/TDD*—nonoverlapping downlink and uplink transmission in different frequency bands.

Note that these duplex-arrangement characterizations should be seen from a user-terminal point of view. From a system point of view, combined FDD/TDD would typically imply simultaneous downlink and uplink transmission, although to/from different user terminals. An example of a system currently utilizing a combined FDD/TDD-based duplex arrangement is second-generation GSM.

An FDD-only duplex arrangement allows for the highest data rate for a given transmission bandwidth and user-terminal peak power. However, it requires a user-terminal duplex filter for each frequency band. The support for a FDD-only duplex arrangement may thus be most relevant for high-end terminals and perhaps only for a limited set of frequency bands. The need for duplex filters can be avoided with TDD-only and combined FDD/TDD duplex arrangements. A TDD-only duplex arrangement has the benefit that only a single frequency band is needed. At the same time, it suffers from potential problems with direct terminal-to-terminal and access-point-to-access-point interference, which may limit the application of the radio access in case of large-cell deployments.

D. User Multiplexing and Multiple Access

The OFDM framework provides a time/frequency grid, as illustrated in Fig. 6. In the case of multiple transmit antennas, there is one such grid for each antenna.

The proposed concept provides user multiplexing and link adaptation in the time, frequency, and spatial domains. The minimum duration for transmission and resource assignment is denoted a *frame*, consisting of eight OFDM symbols (16 OFDM symbols for the more narrow transmission bandwidths). The frame durations are chosen to be short enough to allow retransmissions without causing unreasonable packet delays, in order

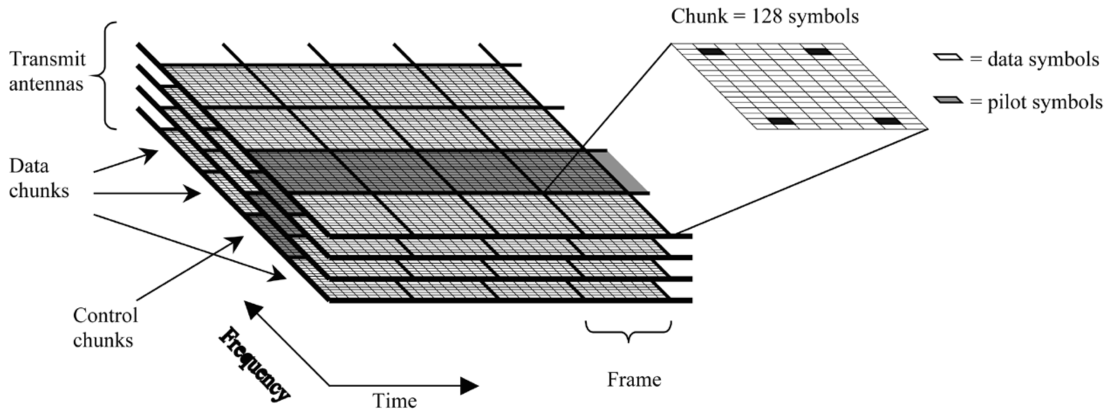

Fig. 6. Fundamental OFDM transmission resource (not all subcarriers are shown) and its division into data chunks and control chunks. Pilot symbol insertion in a chunk is also shown.

to support real-time applications as well as high throughput for TCP-based applications.

The frame is further subdivided in the frequency domain into *chunks*, consisting of 16 adjacent subcarriers (eight subcarriers for the more narrow transmission bandwidths), which are the smallest units for resource allocation and link adaptation. The chunk bandwidths are chosen to be narrow enough that the channel quality is relatively constant within each chunk, even in rather challenging propagation conditions, while at the same time being large enough to keep the control-signaling overhead reasonable. For the more narrow bandwidths, the chunk bandwidth is reduced and the frame length increases compared to the larger bandwidths. This maintains a constant chunk payload while allowing one or two chunks to be allocated for control signaling without a significant reduction in bandwidth also for the more narrow bandwidths.

In Table I, we show the frame duration and chunk sizes for the different system bandwidths that are possible within the proposed concept. A chunk consists of 128 symbols in the time and frequency domains. In case of multiple antenna transmission, a chunk may have multiple layers, each one consisting of 128 symbols.

In the downlink, the chunks can be independently assigned for transmission to different user terminals (or relay nodes) to support channel-dependent frequency-domain scheduling. This means that the set of chunks transmitted to a particular receiver may be fragmented in the frequency domain. Chunks can also be transmitted with different beams or different antennas, potentially with multiple overlapping layers, thereby supporting scheduling and multiplexing also in the spatial domain.

The uplink follows the same frame and chunk structure as the downlink, except that terminals always transmit on contiguous sets of chunks, to support FFT-based precoding for single-carrier modulation; see Fig. 4. Terminals are kept orthogonal by means of network-controlled resource scheduling in the time and/or frequency domain. Contention-based access is used for initial resource requests but can also be applied to user data in the case of low network load.

E. Modulation

The frame-processing functions provide transmission of channel-encoded bit sequences, or MAC PDUs, one for each intended receiver. The bits of each MAC PDU are modulated and spread, potentially FFT precoded, and mapped to the chunks assigned to the corresponding receiver.

The modulation schemes include BPSK, QPSK, 16-QAM, and 64-QAM to efficiently support different signal-to-noise ratios. The modulated symbols are then spread with a three-dimensional linear dispersion code (LDC) [17], generating one or more virtual antenna streams. Each such stream is then mapped onto the physical antennas with a linear beamforming function utilizing available channel knowledge to realize directivity.

In the downlink, the signal constellation and LDC can be selected independently for each chunk, providing frequency-domain link adaptation. The LDC may be used to realize open-loop transmit diversity, for example with orthogonal space–time block codes, as well as multiple-input/multiple-output (MIMO) spatial multiplexing [17]. In the general case, the LDC is chosen to obtain a suitable blend of diversity and multiplexing. Furthermore, with the beamforming functionality and spatial scheduling, several spatially separated users may be served in parallel by transmitting different chunk layers on different beams. MIMO spatial-multiplexing transmission may not only be done with the LDC, but also in the form of per-antenna-rate control (PARC) [18], in which case a single user receives a number of parallel chunks transmitted from different antenna.

Pilots are needed to estimate the channel, both for demodulation and for quality reporting, used for link adaptation. In the downlink, pilots are inserted in the frequency/time grid by reserving a number of symbols within each chunk; see Fig. 6. Different pilot patterns can be used to different users, to support different propagation conditions. For instance, a higher Doppler frequency requires a denser pilot pattern in the time domain, while a smaller coherence bandwidth requires a denser pilot pattern in the frequency domain. In the uplink, the pilots are inserted in the time domain to preserve the desired low peak-to-average power ratio.

F. Control Signaling

In the downlink, in addition to chunks for user data, some chunks are preallocated for control signaling, as illustrated in Fig. 6. These control chunks carry scheduling and transmission-format information, i.e., information indicating the intended receiver as well as the transmission format (signal constellation, LDC, and pilot pattern) used for each data chunk. The location and format for the control chunks are fixed or potentially assigned on a cell basis (using some kind of system information signaling). The latter allows for the use of a frequency reuse larger than one for the control chunks, leading to more robust control signaling.

Frequency-domain scheduling and link adaptation with a resolution of a single chunk can provide high performance gains even in severe fading conditions. The amount of control signaling needed for such a fine resolution will be significant, however, and in many scenarios a more coarse allocation is acceptable. To allow high-resolution scheduling and link adaptation in the frequency domain in scenarios where it is useful, but avoid the signaling costs in cases when it is not needed, a compression technique such as run-length encoding can be used for the chunk control information. In this way, a number of adjacent chunks to the same receiver with the same transmission format can be assigned using a single assignment code word together with a chunk run-length indicator.

With the described transmission scheme, each active receiver continuously monitors the control chunks to determine whether any transmission is directed to it. When this is the case, the receiver determines the locations and transmission formats of the used data chunks and processes the signal appropriately. In the receiver, frame processing performs demodulation, whereas packet processing (described in Section IV) decodes each packet. With a traditional receiver architecture, frame processing generates soft bits which are delivered to packet processing as a noisy version of the MAC PDU. In an advanced receiver, frame and packet processing may be more tightly coupled.

IV. Packet Processing

The packet-processing part of the radio-access framework acts on top of the frame-processing functionality described in the previous section and can be compared with conventional wireless link layers. Traditionally, at wireless link layers a combination of forward error correction (FEC) and automatic repeat request (ARQ), denoted as hybrid ARQ (HARQ), is used to achieve an essentially error-free transmission. We follow this approach and describe in this section the particular characteristics of the proposed concept in order to meet the requirements anticipated for future radio-access networks.

In the design, we have followed two guiding principles. First, the concept should be *optimized for packet traffic*, since we assume that the majority of future traffic is IP based. Independent of the service used by the end-user, ultimately packets need to be transmitted. The second guiding principle is *simplicity*. According to this principle, the complexity and flexibility should only be increased if justified by significant gains, since complexity leads to increased cost in development and operation.

In addition to these two design principles, we have considered three additional requirements for the link-layer design. The first requirement emerges from the guideline simplicity: all packets should be treated using the same structure. Irrespective of the application, the packet processing is identical, even for system-internal control-plane messages. This means that no specific HARQ strategies are applied, depending on the application type or traffic class. To treat all packets identically is possible in the proposed radio-access framework, as we assume very short frame length and very fast retransmissions. Thus, several link-layer retransmissions are possible even for applications with strict delay requirements such as VoIP. Note also that at such data rates it is pointless to distinguish between nonreal time and real-time applications since high-speed TCP connections have much stricter delay requirements than, e.g., VoIP. However, although the packet processing is identical for all packets, it is still possible to differentiate between packets with different priorities based on packet-based markings. This is further discussed in Section V.

Second, it is required that the link-layer protocol provides *scalability* over a wide range of application data rates ranging from a few kilobytes per second for VoIP up to more than 100 Mb/s for file downloads, i.e., the protocol should operate efficiently in terms of resource consumption and overhead in different regimes. An important design choice in this context is the sequence number space that determines the size of the protocol's sequence number field(s), which in turn can greatly contribute to protocol overhead. However, the choice of the sequence number space and the maximum bit rate of the radio access set a lower limit on the size of the link layer retransmission unit. In addition, low bit-rate applications often use small packets (e.g., VoIP) while high bit-rate applications benefit from using large packets (e.g., bulk data transfers). Thus, finding the "right" size of the retransmission unit between these different constraints becomes a challenge.

The third requirement is *support for multihop operation*. Since we envision that relaying functionality may be an important component of future radio-access networks, the link layer should offer optimized support for such scenarios. Some alternative solutions use a single link layer ARQ protocol for each hop and sometimes even an additional ARQ protocol layer on top. This may lead to inefficient protocol interactions.

In the remainder of this section, we describe how we have addressed these design guidelines and requirements in the proposal for the packet processing.

The main feature of the concept is that packets are mapped one-to-one to a retransmission unit. Therefore, we denote our proposal as *packet-centric processing*. Another important contribution is a novel relay ARQ protocol, which operates across all involved nodes. It is embedded in the packet-centric concept and allows that the ARQ state can be shared easily between the involved nodes and enables efficient packet transmission in relaying environments.

A. Packet-Centric Processing

A key design guideline for the proposed scheme has been the optimization toward the transport of packets. As part of this, an attempt has been made to keep packets as integral units as far down in the protocol stack as possible, which allows us to keep the focus on packets as long as possible. Consequently, there is a one-to-one mapping of packets (e.g., IP packets or control messages) to retransmission units of the radio link control (RLC) protocol. These are also denoted as RLC PDUs (see Fig. 7). Each RLC PDU is characterized by a sequence number,

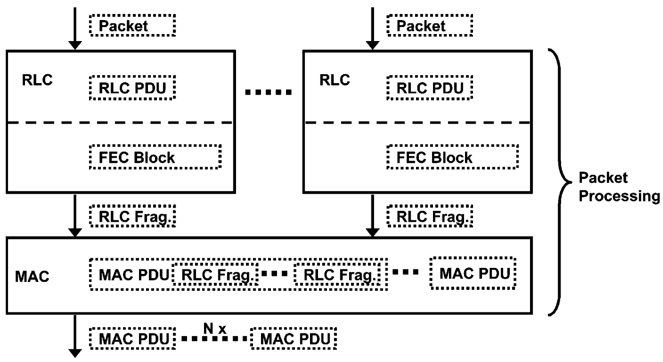

Fig. 7. Packet processing.

the corresponding payload, and a cyclic redundancy check (CRC). Note that in the proposed concept retransmission units have variable sizes depending on the packet size they carry. The next step in the packet-centric design is to map one RLC PDU to exactly one FEC block, i.e., once again a one-to-one mapping. Thus, the receiver can decode each packet (or RLC PDU) independently.

The advantages of our packet-centric design are the following. First, the one-to-one mappings from packets down to FEC blocks enables the scheduler to decide about the transmission of packets. This optimizes the delivery of packets since their reception is of interest and not the reception of parts thereof. In addition, if scheduling decisions are done for packets, it is straightforward to consider any related attributes (e.g., priorities) of these packets in the decision. Second, investigations have shown that the RLC overhead consisting of header overhead and padding overhead can be decreased compared with existing solutions based on fixed sized PDUs by 5%–15% depending on the application specific packet size distributions. The gain stems mainly from the elimination of padding overhead, since packets are not segmented. Third, the packet-centric approach decreases the required amount of RLC PDUs compared to the case when packets are segmented into relatively small fixed-sized PDUs. This reduces the RLC processing demands and permits a smaller sequence number space. Fourth, instead of applying the commonly used solution of segmenting a packet into fixed-sized PDUs, we use rate matching techniques to not only adapt to the instantaneous radio-channel conditions, but also to generate, from a packet, variable-sized fragments, named RLC fragments. As a result, the protocol can adapt to different operational situations, e.g., different bandwidths or varying traffic loads. Finally, the approach leads to a scalable protocol, because the retransmission-protocol overhead is kept constant for a packet, regardless of its packet size. This is in contrast to protocols that perform segmentation, where the overhead increases with increasing packet size. High-speed applications (e.g., file download) typically make use of large packets, while lower speed applications (e.g., VoIP) often create smaller packets. For retransmission protocols, the more demanding applications are the high-speed applications because they lead to large send and receiver buffers, require a larger sequence-number space, and correspondingly create larger ARQ feedback messages. The observation that high-speed applications typically use large packets, and the choice to tie the ARQ sequence numbers to packets lead to a concept that scales well in the critical scenarios for the retransmission protocol due to the inherent characteristic that high-speed applications send large packets.

However, note that significant performance gains of the packet-centric concept compared with existing link layers cannot be expected, because these are already highly optimized and there is little room for performance improvements. The main motivation for the proposed concept lies in its simplified structure and its potential for an efficient cross-layer design by keeping packets as integral units.

A concrete proposal for the realization of the HARQ scheme is to use a selective-repeat retransmission protocol in combination with turbo codes and incremental redundancy. Turbo codes have been selected because of their strong error-correction capabilities for large code blocks. It is believed that this property fits well to the packet-centric design with its tendency to large FEC blocks. The reason to apply incremental redundancy is that the amount of coded information can be flexibly selected. This is required since not only packets vary in size, but also the available radio resources vary over time due to channel variations and multiuser scheduling. Therefore, an efficient mechanism is required to fit the payload into the available resources of the next frame. Based on the current channel conditions, the frame processing provides a certain amount of capacity. A certain code rate is selected to match the quality of the channel to allow for a successful transmission with a certain probability. Then, the rate-matching process selects the corresponding amount of bits out of the FEC block. This selected group of coded bits is called an RLC fragment. Note that at high data rates, often more than one RLC fragment can be transmitted in one frame. For example, assuming a code rate of 2/3, a frame of 0.36 ms at a link speed of 100 Mb/s can convey 3000 bytes, i.e., two packets of 1500 bytes. This can be used to exploit multiuser diversity or to send RLC fragments for two RLC PDUs. One or more RLC fragments of a single user are then grouped to form a MAC PDU. Together with MAC PDUs of other users, this is then handed over to the frame processing described above.

It is well known that soft combining is beneficial and thus also applied here. Therefore, packet-control information needs to be sent out-of-band to allow the receiver to put the received information at the right decoder position. This packet-related control information is denoted as an RLC *fragment header* and comprises the PDU sequence number, the RLC fragment version number, and the payload length, i.e., the packet size. Without this control information, the receiver is not able to use the received payload information. Therefore, it needs to be protected by a separate channel code and a CRC. Note that for every RLC fragment sent in one frame, separate control information is sent in order to allow independent decoding for each individual PDU.

The next issue to discuss is the situation that the predicted amount of coded data needed for a successful decoding at the receiver does not fit into the next frame. Nowadays, an IP packet size of 1500 bytes is fairly common, but with increasing data rates, it can be expected that deployed maximum transfer units (MTU) will grow correspondingly. For example, gigabit ethernet has an MTU size of 9000 bytes. It is obvious that even in a high data rate system it cannot be assumed that a 9000-byte packet always fits into one frame, e.g., due to poor channel conditions requiring a low code rate. As already stated and opposed to traditional concepts, link layer segmentation is not used to obtain PDUs of the size that radio resources are able to convey. Instead, we rely again on rate matching. A too large code block is rate-matched to such an extent that it fits into the assigned resources. In this case, the sender can already predict that the decoding is probably not going to be successful, in particular, if the number of transmitted bits is smaller than the PDU size (i.e., code rate > 1). In order to save processing power at the receiver, it is proposed to use a complete flag (CF) to signal for each PDU whether the decoder should start the decoding process. In case the code-rate transmitted is not sufficient, the decoding process waits for the next RLC fragment. To reduce delays in this case, the sender does not await a negative acknowledgment for this PDU. Instead, it sends the next RLC fragment for this PDU as soon as the scheduler allows it, preferably already in the next frame. Once the amount of coded data sent is deemed to be appropriate for successful decoding, the sender sets the CF to indicate that the decoder should now decode the code block. Note that for large packets the transmission could take several frames.

Furthermore, the packet-centric approach includes another optimization. Since each RLC fragment corresponds to one packet, such an RLC fragment can be potentially large. In cases when transmission errors occur, it is not always required to retransmit a similar amount of redundancy to the receiver. It could be that the reception had been almost successful, but unfortunately, a traditional negative acknowledgment (NACK) signal cannot express this. Therefore, it is proposed to provide more fine-grained ARQ feedback by using one or a few intermediate NACK signals, which indicate the appropriate amount of redundancy that is required to decode a FEC block successfully. The appropriate NACK level can be determined based on decoder internal metrics. For example, if two NACK signals are used, the signal NACK1 could indicate that no useful information was received. In that case, a complete retransmission is appropriate. The NACK2 signal could indicate that only a small amount of redundancy is required, e.g., only 30%. It is expected that this *rich ARQ feedback* approach, which fits nicely to incremental redundancy, helps to reduce the retransmission overhead significantly. A detailed study remains to be done on a suitable number of required NACK levels and the respective gains, but the benefit of reduced retransmission data seems to outweigh the additional signaling costs in the status messages.

To summarize, the packet-centric packet transmission concept provides powerful means to handle varying packet sizes and changing channel conditions. This is the result of a tight integration of handling packets, retransmissions, and channel coding in one functional block. The feature that basically packets are visible at the scheduler allows for resource efficient channel-dependent scheduling while at the same time taking packet-based markings of different service classes into account. This will become more evident in Section V.

B. Relay ARQ for Multihop Operation

As already stated above, 4G access systems will most likely support a relaying mode to ensure coverage for the very high data rates envisioned to be supported with 4G radio access, without an unreasonable deployment effort. We assume a link-layer relaying approach instead of a physical-layer repeater solution. This means that a relay node forwards successfully received link-layer PDUs to the next node. Alternative ARQ solutions spanning across two or more hops often use layered ARQ protocols to achieve a reliable data transfer. Disadvantages are partly doubled functionality, an increased overhead, and potentially harmful protocol interactions. Therefore, a novel relay ARQ concept is proposed in which the ARQ process spans across all involved nodes (see Fig. 8). The same PDU sequence number is used for the transmission of a packet across all nodes. An example for the message exchange in a two-hop scenario is also shown in Fig. 8. Each PDU propagates from node to node. The respective receiving node attempts to decode a PDU. If this is not successful, it sends a status message with a NACK signal for the respective PDU, which will trigger a retransmission. If the packet reception is successful, the relay node takes over the transmission responsibility for the subsequent hop. In a status message it signals a so-called relay ACK (RACK) to the preceding node. The RACK is a new, third status signal in addition to conventional ACK and NACK signals. If a RACK signal for a certain PDU is received, the sender knows that another node has now the responsibility for this PDU, but it does not remove it from the send buffer yet, since it serves as a fallback node if the relay node fails to deliver the data due to extraordinary failure cases (e.g., a node failure). Once the PDU has arrived at the ultimate receiver, it signals an ACK. Only this node is allowed to generate an ACK. Upon reception of the ACK, the relay node updates its state variables and removes the PDU from its buffer. In addition, it propagates the ACK signal to the sender node, which can then remove the PDU as well. Note that the description refers to an example for two hops, but the concept can be applied to an arbitrary number of hops. In particular, it needs to be emphasized that even in a scenario with more than two hops, the two different ACK signals, ACK and RACK, are still sufficient. The number of ACK signals required does not grow with an increasing number of hops. More details about the relay ARQ concept including a performance evaluation can be found in [19]. There, it is shown that the proposal provides performance gains in most scenarios compared to an approach involving layered ARQ protocols.

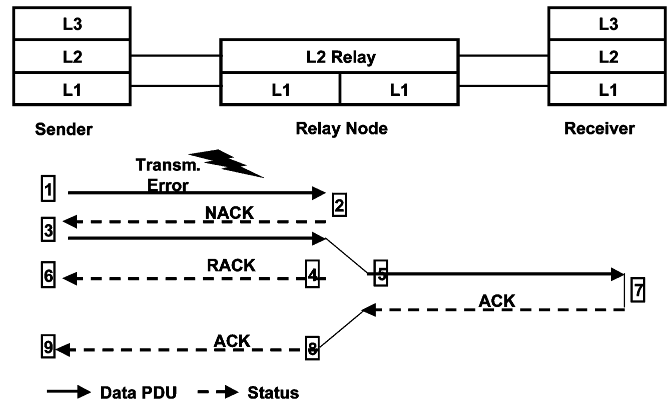

Fig. 8. Relay ARQ message sequence.

A feature of the proposed relay ARQ concept is that error recovery is always performed at the link where the error occurred. Such local error recovery is faster compared to peer-to-peer retransmissions. In addition, since protocol layering is avoided, our concept eliminates the potential problem of concurrent retransmissions that may occur with layered ARQ solutions. These advantages can be achieved at small expenses since the approach can be based on the packet-centric packet processing and requires only the introduction of the novel status signal RACK.

C. Status Format

The proposed concepts of relay ARQ and rich ARQ feedback have both an impact on the format of the status messages. Compared to traditional concepts, additional feedback information is required: first, the different NACK levels in case of rich ARQ feedback and then the RACK signal in case of relay nodes. Although the rich ARQ feedback and the appropriate number of NACK levels has not been evaluated yet, an exemplary deployment could be to spend two bits per PDU in the status message, i.e., four different states per PDU. In a single-hop scenario, this would allow expressing in addition to the ACK signal three different NACK signals. Relay nodes could signal ACK, RACK, and two NACK signals.

V. Service Differentiation, Packet Queueing, and Scheduling

Having described the packet and frame processing in the previous two sections, we now describe the other two components shown in Fig. 2, the packet-queueing process and the scheduler. However, we first briefly describe our general approach toward service differentiation.

Future mobile broadband access networks need to efficiently carry traffic from multiple services. This is depicted in Fig. 1. A key question then is how to achieve service differentiation, i.e., how to allocate network resources to the traffic associated with the different services. For this purpose, we have adopted the differentiated services concept [20] into our framework proposal. In that concept, traffic from one or more service(s) can be classified into a single service class. A network operator then controls the allocation of network resources per service class through profile definitions. An example of such a profile is the triplet ⟨ *priority, minimum bit rate, maximum bit rate* ⟩ that is associated with the traffic aggregate of a specific service class. After the classification process, the traffic of each service class is conditioned to ensure that it complies with the corresponding profile definition, in particular, the defined maximum traffic rate. This can be achieved through delaying (shaping) or dropping (policing) packets. Finally, packets are marked with a service-class identifier and optionally also with a marking that identifies a packet as in-profile or out-of-profile. The latter marking can then be used by the packet queueing function described as follows to preferentially drop out-of-profile packets or to let them pass when capacity is available. An example of an out-of-profile packet is a packet associated with a traffic aggregate currently entering the network at a rate above the minimum but below the maximum bit rate.

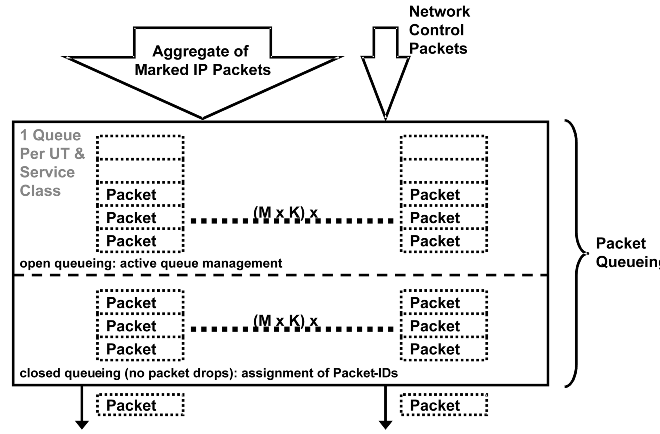

Fig. 9. Packet queueing process (M service classes and K active UTs).

Once the packet traffic has been classified, conditioned, and marked, there are two more functions involved in providing service differentiation: the packet queueing process and the scheduler described in the remainder of this section. Note that in our radio-access framework neither the packet processing nor the frame processing is involved in the service differentiation. That is, both processes treat every packet in exactly the same way, independent of whether a packet belongs to a delay-sensitive or a throughput-sensitive application. We believe that we can afford this simplification due to the extremely low delays and high bit rates that we foresee for future radio access. We further believe that this conscious design choice can greatly simplify the design and operation of the packet and frame processing.

The packet queueing process is depicted in Fig. 9. Although it performs a rather simple function, it is an essential component of service differentiation. In our framework proposal, we assume that separate queues are allocated for each active UT and for each service class. A destination field in the packet, and the packet's service-class marking, is used to place the packet into the corresponding queue. Each queue comprises a closed and an open part. Once a packet has progressed to the closed part it will not be dropped except in an abnormal situation. For example, a timer associated with each packet could trigger an exception-handling routine if a packet remains queued in the closed part for too long. Note that, in our framework, the packet processing may not drop a packet, again except in an abnormal situation.

Packets in the closed part of their respective queue get a unique identifier assigned. The packet identifier is required by the scheduler to associate a particular packet with an RLC PDU in its interactions with the packet processing and ultimately with a set of chunks in its interactions with the frame processing. Recall from the packet processing section that one packet is always mapped to one RLC PDU.

Packets in the open part of their respective queue may get dropped during transient phases of congestion. However, with proper network dimensioning, the dropping of in-profile packets should be an exception. An operator is assumed to have access to various packet-related statistics maintained in real-time, including statistics related to the dropping of in-profile packets. In that way, an operator can constantly monitor the utilization

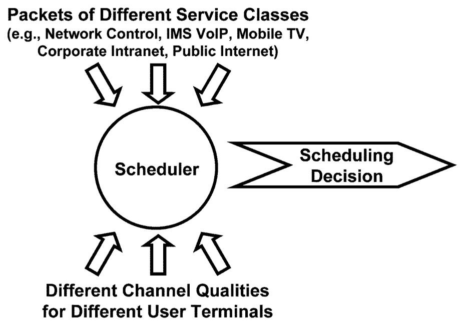

Fig. 10. Cross-layer scheduler.

of network resource in order to ensure proper and timely dimensioning of the network as traffic load increases over time. Still, a considerable fraction of out-of-profile packets can enter the packet queues. If capacity is available, those packets will be forwarded to the packet processing to ensure high network utilization. Otherwise, such packets need to be dropped in order to fulfill the minimum bit-rate requirements of the different service classes. Active queue management schemes are widely used already today in packet routers to control the packet dropping per queue [21]. The choice of the right queue management scheme and its configuration used per service class could be part of the mentioned profile definitions. This is an important choice of operation an operator has to make. In general, larger queues are appropriate for traffic where the corresponding application is not sensitive to delay and jitter (e.g., file transfer or e-mail). The same is true for traffic originating from applications that try to comply with a certain average bit rate. In that case, the queue is required to absorb short-term bursts. However, low latency traffic (e.g., VoIP) is best served with small queues, which—with proper dimensioning and setting of the service class profiles—should be mostly empty.

Our concept proposal aims at a common design for both the uplink and the downlink scheduler. One downlink scheduler and one uplink scheduler is assumed to be located in each AP and each RN. Both schedulers communicate with their corresponding packet queueing, packet processing, and frame processing functions (see Fig. 2) in the same way, using the same kind of information. The only difference between the uplink and the downlink scheduler is that for the uplink scheduler, this communication is assumed to be carried out of band. Thus, the following description applies likewise to the downlink and the uplink scheduler, and for brevity we only refer to both as "the scheduler." The description of an uplink contention mode that could potentially be operated during times of low traffic load is outside the scope of this paper.

The scheduling function is commonly seen as a function of the MAC layer. However, the scheduler in our concept proposal could rather be seen as a cross-layer scheduler that schedules between packets from different service classes while at the same time taking into account the instantaneous radio-channel conditions (channel-dependent scheduling), as shown in Fig. 10. As shown in Fig. 2, we envision the scheduler as *the* central function that controls the packet queueing, packet processing, and frame processing functions "clocked" on a cycle determined by the frame duration of roughly 0.3 ms. For example, the following control could be executed by the scheduler for each frame.

1) The scheduler receives from the frame processing function quality reports for each chunk in the next frame that needs to be scheduled. The scheduler combines this feedback with the profiles defined for each service class and some moving average statistics about the bit rate that has recently been scheduled per service class (e.g., to ensure that minimum bit rates are met) as well as the amount of radio resources that have recently been scheduled per UT within one service class (e.g., to ensure weighted fairness).
2) For each frame, the scheduler decides whether the packet processing function needs a new packet from the packet queueing function. If a new packet is needed, the scheduler identifies which packet gets forwarded.
3) For each frame, the scheduler identifies for which RLC PDU(s) a corresponding RLC fragment gets transmitted in the next frame. In addition, the scheduler determines the size of each RLC fragment.
4) For each frame, the scheduler identifies to the frame processing function the chunks to be used for each MAC PDU, generated by packet processing function, and which chunk-transmission formats to be used per chunk. Thus, the scheduler effectively controls the link adaptation.

A key feature of a scheduler that operates according to this approach is that it effectively supports work-conserving preemption of packet transmissions. This means that a higher priority packet can preempt the transmission of a lower priority packet going to the same node (AP, RN, or UT). For example, the scheduler can preempt the transmission of a large low-priority packet that needs to be sent across multiple frames, in favor of a short higher priority packet. Once the smaller packet has been transmitted, the scheduler can resume the transmission of the larger packet. Note that the packet-centric design of the packet processing function explained in Section V is the key feature that enables the scheduler to perform work-conserving preemption of packet transmissions.

VI. Performance Evaluations

This section presents results of some initial performance evaluations carried out based on the outlined radio-access framework. The intention with the evaluations is to get an initial understanding of coverage and capacity for very wideband radio access as well as an indication for how some multiple-antenna techniques can be used to enhance the performance.

A. Assumptions for Evaluation

These initial evaluations have focused on the downlink in a suburban environment with an access-point antenna height of 30 m and an average rooftop height of 15 m. We assume a transmission bandwidth of 80 MHz (3072 subcarriers and a subcarrier spacing $\Delta f = 24.4$ kHz; see Table I), a cyclic prefix of approximately 4 μs, and an FDD-only duplex arrangement at 5 GHz. Overhead, in terms of pilots and protocol signaling, is accounted for as 20% of the data rate.

Users are assumed to be located outdoors and the user-terminal speed is Rayleigh distributed with mean 3 km/h. The COST 231-Walfisch-Ikegami path loss model [22] is used with, as in [23], the free-space loss at 5 GHz. A break-point distance of 30 m is considered so that the path loss as a function of distance d in meters is modeled as

$$L(d) = \begin{cases} 46.4 + 20\log_{10} d & d \leq 30 \text{ m} \\ 19.8 + 38\log_{10} d & d > 30 \text{ m} \end{cases} \quad [\text{dB}]. \tag{1}$$

The shadow fading is modeled as a log-normally distributed random variable with a standard deviation equal to 8 dB and intersite correlation equal to 0.5. To model the spatial and temporal characteristics of the channel, the suburban macro spatial channel model of 3GPP [27] is taken as the starting point. This is a ray-based channel model with correlation between delay spread, angle spread, and shadow fading. However, since the 3GPP channel model is only applicable for a transmission bandwidth of 5 MHz, the intracluster delay-spread extension proposed in [23] is used. This extension groups the subpaths of each path into three groups with different delays. A consequence of this is that the channel for each link comprises 18 resolvable paths in the time domain. Note that the same channel model is used also for interference from sectors of the same and other sites in order to capture the properties of the interference in the frequency and space domain.

A deployment with three-sector sites is studied, and the access point antenna elements are vertically polarized with an element-pattern gain of 18.5 dBi, a half-power beam width of 63°, and a front–back ratio of 30 dB. The terminal antennas are omnidirectional with 0 dBi gain and no body loss is accounted for. Uniform linear arrays are considered, both at access points and user terminals. The antenna spacing is half a wavelength at both ends, except when spatial multiplexing is used. In this case, the element separation is four wavelengths at the access point but still half a wavelength at the user terminal. The access point output power per sector is 80 W (49 dBm), which corresponds to 30 dBm/MHz, independent of the number of transmit antennas. At the terminals, a noise-power spectral density of -167 dBm/Hz is assumed, corresponding to a noise figure of 7 dB.

User terminals connect to the sector cell with the lowest path loss, including shadowing, and a hard handover algorithm with a 3 dB handover margin handles user mobility. In case of downlink beamforming, the beam with the lowest path loss is used for cell selection. All users under consideration are assumed to have full buffers and, for simplicity, pure time-domain round-robin scheduling is considered. A natural continuation would accordingly be to introduce channel-dependent scheduling in the time and frequency domains.

A rate 1/3 turbo code [24] is used together with rate matching to obtain code rates in the range 1/3 to 8/9. In combination with BPSK, QPSK, 16-QAM, and 64-QAM modulation, this implies that data rates between 18.4 and 291.2 Mb/s are achievable. In case of spatial multiplexing with four antennas, the corresponding data rates are 73.6 and 1164.8 Mb/s. For simplicity, no frequency-domain link adaptation is considered. All chunks use the same modulation format and are assigned equal transmit power. An error-free channel-quality measurement with no delay is used for adaptive modulation and coding (AMC). The AMC is configured to maximize the data rate under the constraint that the estimated block-error probability (BLEP) is below 10%. Hybrid ARQ type I is used, and the results do thus not illustrate the benefits of more efficient retransmission schemes, like HARQ type II with incremental redundancy.

Link-level performance is modeled by calculating an effective SINR and then mapping this effective SINR to a BLEP as outlined in [25]. However, instead of using the exponential function to form the effective SINR from the set of post-receiver-processing SINRs, the average mutual information for bit-interleaved coded modulation (BICM) with LogMax demodulation as given in [26] is used. Note that imperfections, such as estimation errors, and the effects of delay spread greater than the cyclic prefix are neglected.

In addition to transmission with a single antenna, downlink beamforming with a fixed multibeam antenna and MIMO spatial multiplexing, herein exemplified with per-antenna rate control (PARC), are considered. The multibeam antenna uses four antennas to generate eight beams covering each sector, and the beam with the lowest path loss is selected for downlink transmission. For PARC, one individually rate-controlled stream is transmitted from each of the four antennas. All terminals have the same number of antennas and employ minimum mean-square error (MMSE) combining on a per-carrier basis. Since the spatial properties of the interference are modeled, the benefits of suppression of interference from other sectors are included. For the PARC transmission, successive interference cancellation after channel decoding is employed, and, at each stage, streams decoded are cancelled. Streams not decoded are suppressed just like interference from other sectors. In the evaluations, it is further assumed that if decoding of one stream fails, all streams fail, independently on the decoding order of the streams.

B. Single Link Throughput Versus Distance

To assess the possible coverage, the performance of a single link is first evaluated. For this purpose, a single sector of a single site is considered and the data rate averaged over shadow fading and angle with respect to the access point for different distances is evaluated. The average user link-layer data rate for the case when the user is alone in the system is depicted in Fig. 11. The figure indicates that for short distances all techniques reach a data rate that equals the maximum available data rate, and the throughput is hence modulation limited. With multistream MIMO transmission, it is possible to increase the data rate significantly compared with single-stream transmission and, with four antennas, the local area target data rate of 1 Gb/s is achieved. Fig. 11 also indicates the possible coverage gains obtained by downlink beamforming and terminal receive diversity as compared to the case with single antenna transmission and reception. Furthermore, as expected, downlink beamforming is preferable to multistream MIMO transmission at large distances due to the array gain and the low SINR. Moreover, the considered lowest code rate may also penalize the multistream transmission at low SINRs, where lower rates could be beneficial. Note that because only a single sector of a single site is considered, the benefits of macrodiversity with respect to coverage may not be seen in an evaluation of this kind.

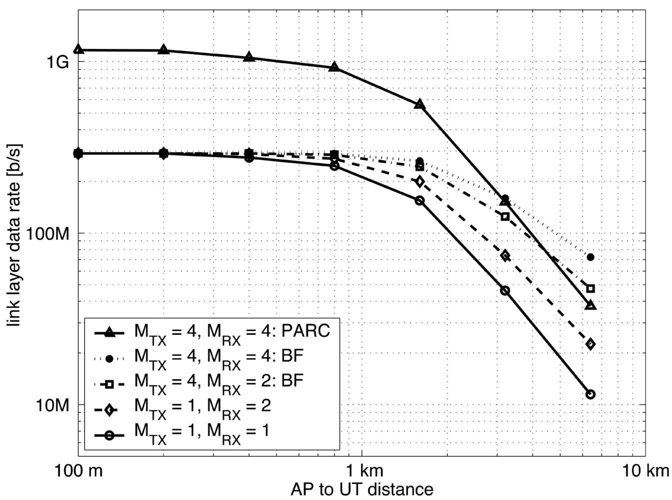

Fig. 11. Single-sector single-user link-layer data rate as function of distance for single-antenna transmission, downlink beamforming (BF), and PARC.

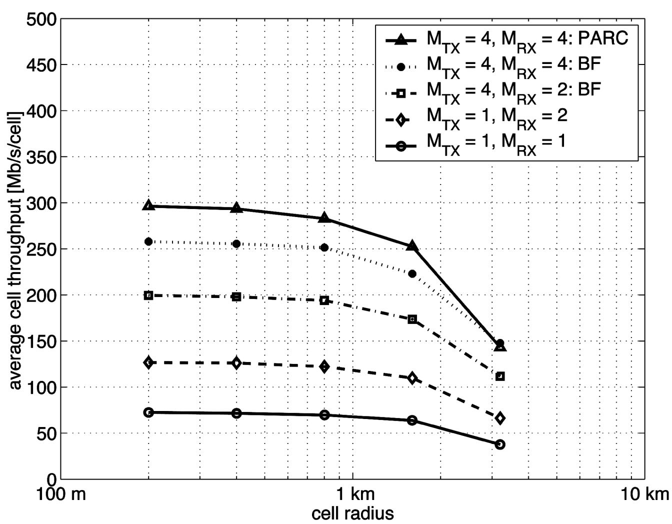

Fig. 12. Average cell throughput versus cell radius.

C. Multiple Users in Multicell Deployment

The performance in a highly loaded multicell environment is next considered, and Fig. 12 depicts the average cell throughput as a function of the cell radius. Note that for the considered deployment, the site-to-site distance equals three times the cell radius; whereas, the cell range is two times the radius. The performance of the various techniques is ordered in essentially the same way as in the single-link case. The cell-throughput figures range from around 80 Mb/s/cell up to about 300 Mb/s/cell. The distributions of the user data rates for cell radius of 800 m are given in Fig. 13 and, as compared with single-antenna transmission, downlink beamforming and receive diversity do not only increase the user data rates uniformly, but also reduce the variation of data rates of different users. The results in Fig. 13 further indicate that even in the multicell environment, the user data rates may be modulation limited for a nonnegligible fraction of the users when combining techniques such as beamforming and receive diversity. Furthermore, in the specific scenario studied, downlink beamforming in combination

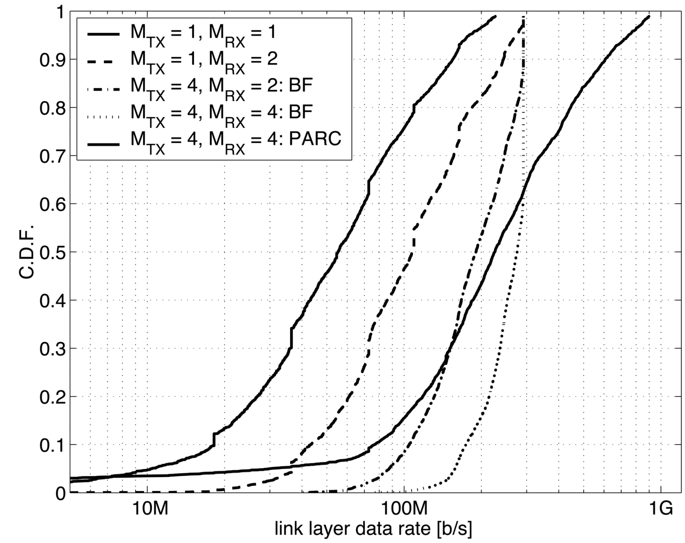

Fig. 13. Link-layer data-rate distributions for cell radius of 800 m.

with two-branch terminal receive diversity provide a data rate of 100 Mb/s or more to at least 90% of the users. The results also indicate that downlink beamforming outperforms MIMO spatial multiplexing transmission with PARC when it comes to serving the noise- and interference-limited users. This was also observed in the single link case, and it is envisioned that an adaptive scheme, which selects one out of a set of multiple antenna transmission methods based on, e.g., channel and interference conditions, could be beneficial.

Finally, continuous refinements of the performance evaluations and the scenario considered alongside the development of the framework itself are expected. Recall that downlink overhead for signaling and pilots for demodulation and measurements were coarsely modeled and that measurement delays and imperfections as well as estimation errors were not accounted for. One may, therefore, argue that the results are optimistic. On the other hand, the results do not include benefits from features and functionality such as HARQ type II with incremental redundancy, channel-dependent scheduling in time and frequency domains, and frequency-domain link adaptation.

VII. Summary and Conclusion

Research on future radio access systems, supporting up to approximately 100 Mb/s with wide-area coverage and up to 1 Gb/s with local-area coverage, is currently gaining momentum around the world. In this paper, a framework for the design of such a packet-data system is presented. The framework allows for flexible spectrum utilization through the support of a wide range of bandwidths, up to 100 MHz, as well as different duplex arrangements.

OFDM-based transmission is proposed due to the high bandwidths supported and the possibility for the scheduler to exploit channel variations, not only in the time domain but also in the frequency domain. Advanced multi-antenna techniques such as MIMO spatial multiplexing are an integral part of the proposed framework, required to support the highest data rates. Simulations show that with the proposed techniques, peak data rates up to 1 Gb/s and a cell throughput of 200–300 Mb/s can be supported in a macrocell deployment.

A key characteristic of the proposed framework is the packet-centric processing, implying that packets are not segmented as in existing cellular systems. Instead, the scheduling mechanism and the retransmission protocol both operate on complete packets instead of segments thereof. This allows for a simple, yet efficient structure, able to scale over a large range of varying payload sizes without excessive overhead. Multihop relaying is an integral part of the retransmission protocol as the use of relays is likely to be required to support wide-area coverage of data rates in the order of 100 Mb/s in a cost-efficient manner.

The development of an integrated framework allows for studies of the interaction between the different building blocks used and ensures a joint optimization of the component solutions used. We believe that the proposed framework provides a suitable platform for development and evaluation of novel techniques, supporting the research on high-performance future radio access.

REFERENCES

[1] 3GPP, "High speed downlink packet access (HSDPA); Overall description; Stage 2," TR 25.308.

[2] S. Parkvall *et al.*, "The high speed packet data evolution of WCDMA," in *Proc. IEEE PIMRC*, Sep. 2001, pp. G-27–G-31.

[3] 3GPP, "FDD enhanced uplink; Overall description; Stage 2," TR 25.309.

[4] S. Parkvall, J. Peisa, J. Torsner, M. Sågfors, and P. Malm, "WCDMA enhanced uplink—Principles and basic operation," in *Proc. Veh. Technol. Conf.-Spring*, May 2005, pp. 1411–1415.

[5] 3GPP, "Proposed study item on evolved UTRA and UTRAN," RP-040 461, 2004.

[6] "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000," ITU, ITU-R M.1645.

[7] [Online]. Available: https://www.ist-winner.org/

[8] X. Yu, G. Chen, M. Chen, and X. Gao, "Toward beyond 3G: The FUTURE project in China," *IEEE Commun. Mag.*, pp. 70–75, Jan. 2005.

[9] S. Abeta, H. Atarachi, and M. Sawahashi, "Broadband packet wireless access incorporating high-speed IP packet transmission," in *Proc. Veh. Technol. Conf.-Fall*, Sep. 2002, pp. 844–848.

[10] M. Mathis *et al.*, "The macroscopic behavior of the TCP congestion avoidance algorithm," *Computer Commun. Rev.*, vol. 27, no. 3, pp. 67–82, Jul. 1997.

[11] P. Beming, M. Cramby, N. Johansson, J. Lundsjö, G. Malmgren, and J. Sachs, "Beyond 3G radio access network reference architecture," in *Proc. Veh. Technol. Conf.-Spring*, May 2004, pp. 2047–2051.

[12] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications—Theory and Applications of OFDM*. New York: Springer-Verlag, 2004.

[13] L. J. Greenstein, V. Erceg, Y.-S. Yeh, and M. V. Clark, "A new path-gain/delay-spread propagation model for digital cellular channels," *IEEE Trans. Veh. Technol.*, vol. 46, no. 5, pp. 477–485, May 1997.

[14] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.

[15] U. Sorger, I. De Broeck, and M. Schnell, "Interleaved FDMA—A new spread-spectrum mulitple-access scheme," in *Proc. IEEE ICC*, Jun. 1998, pp. 1013–1017.

[16] Y. Goto, T. Kawamura, H. Atarachi, and M. Sawahashi, "Variable spreading and chip repetition factors (VSCRF)-CDMA in reverse link for broadband wireless access," in *Proc. PIMRC*, Sep. 2003, pp. 254–259.

[17] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1828, Jul. 2002.

[18] S. T. Chung, A. Lozano, and H. C. Huang, "Approaching eigenmode channel capacity using V-BLAST with rate and power feedback," in *Proc. Veh. Technol. Conf.-Fall*, Oct. 2001, pp. 915–919.

[19] H. Wiemann, M. Meyer, R. Ludwig, and C. O. Pae, "A novel multi-hop ARQ concept," in *Proc. Veh. Technol. Conf.-Spring*, May 2005, pp. 3097–3101.

[20] S. Blake *et al.*, "An architecture for differentiated services," IETF, RFC 2475, Dec. 1998.

[21] B. Braden *et al.*, "Recommendations on queue management and congestion avoidance in the Internet," IETF, RFC 2309, Apr. 1998.

[22] "Digital mobile radio toward future generation systems," COST 231, Final Rep. EUR 18 957, 1999.

[23] D. Baum *et al.*, "An interim channel model for 4G systems, extending the 3GPP spatial channel model (SCM)," in *Proc. Veh. Technol. Conf.-Spring*, May 2005, pp. 3132–3136.

[24] 3GPP, "Multiplexing and channel coding (FDD)," TS 25.212.

[25] 3GPP, "Feasibility study for OFDM for UTRAN enhancement (release 6)," TR 25.892, v 1.1.0.

[26] J. F. Cheng, "Coding performance of various type-I HARQ schemes with BICM," in *Proc. ISIT* , 2004, pp. 319–319.

[27] 3GPP, "Spatial channel model for multiple input multiple output (MIMO) simulations," TR 25.996, V6.1.0.

David Astély (S'95–M'99) received the M.S. degree in electrical engineering and the Ph.D. degree in signal processing from the Royal Institute of Technology, Stockholm, Sweden, in 1995 and 1999, respectively.

During 1998, he held a visiting research position at the Brigham Young University, Provo, UT, and from 1999 to 2001, he was with Nokia Networks. Since 2001, he has been with Ericsson Research, Stockholm, Sweden, working in the area of advanced antenna systems for 3G and beyond.

Dr. Astély received an IEEE Signal Processing Society Best Young Author Paper Award in the area of signal processing for communications in 2000.

Erik Dahlman (M'04) received the M.S. and Ph.D. degrees from the Royal Institute of Technology, Stockholm, Sweden, in 1987 and 1992, respectively.

Since 1993, he has been with Ericsson Research, Stockholm, Sweden, where he currently holds a position as Expert in Radio Access Technologies. He was deeply involved in the development and standardization of 3G radio access technologies, first in Japan and later on within the 3GPP standardization body. More recently, he has been involved in the high-speed packet-data evolutions of the WCDMA 3G standard. His current research interests are focused on the evolution of 3G systems and radio-access technologies applicable to future wireless systems.

Dr. Dahlman was awarded the Vehicular Technology Society Jack Neubauer Best System Paper Award in 1999.

Pål Frenger (S'94–M'99) received the M.S. degree in electrical engineering and the Lic.Eng. and Ph.D. degrees in signals and systems from Chalmers University of Technology, Göteborg, Sweden, in 1994, 1997, and 1999, respectively.

In 1999, he joined Ericsson Research, Stockholm, Sweden, and since 2005, he has been with Ericsson Research, Linköping, Sweden. He has been working with algorithm development for both wireline (xDSL) systems, as well as wireless (WCDMA, HSDPA, 3GPP-long term evolution, and 4G) systems.

Reiner Ludwig received the Diploma and Ph.D. degree in computer science from the University of Technology, Aachen, Germany, in 1994 and 2000, respectively.

He pursues his research with Ericsson Research, Herzogenrath, Germany, where he joined in 1994. From 1997 to 1999, he worked at the University of California at Berkeley as a Guest Researcher on behalf of Ericsson Research. In his research, he focuses on cross- layer issues that arise in the design of wireless packet-based networks. This includes user-plane interactions of end-to-end functions with wireless link and physical-layer functions, and control-plane mechanisms for providing quality-of-service.

Michael Meyer received the Diploma and Ph.D. degree in electrical engineering from the University of Paderborn, Paderborn, Germany, in 1991 and 1996, respectively.

In 1996, he joined Ericsson Research, Herzogenrath, Germany. He holds the position of Senior Specialist and is leading a performance analysis team. Currently, he is involved in the concept development for evolved 3G systems. His research interests include the design and analysis of link-layer protocols for wireless systems.

Stefan Parkvall (S'92–M'96–SM'05) received the M.S. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden, 1991 and 1996, respectively.

He joined Ericsson Research, Stockholm, in 1999 and is currently a Senior Specialist in adaptive radio access, working with research on and standardization of future cellular technologies. His previous positions include being an Assistant Professor of Communication Theory at the Royal Institute of Technology, Stockholm (1996-2001), and a Visiting Researcher at the University of California, San Diego (1997-1998).

Per Skillermark received the M.S. degree in applied physics and electrical engineering from Linköping University, Linköping, Sweden, in 1999.

Since 1999, he has been with Ericsson Research, Stockholm, Sweden, working with radio resource management for 3G and beyond 3G systems. His research interests include design, analysis, and modeling of radio networks.

Niclas Wiberg received the M.S. degree in computer engineering and the Ph.D. degree in information theory from Linköping University, Linköping, Sweden, in 1990 and 1996, respectively.

From 1996 to 1998, he was an Assistant Professor of Information Theory at Linköping University. Since 1998, he has been with Ericsson Research, Linköping, where he holds a position of Senior Specialist in radio network modeling. His current interests include algorithm development for radio resource management and development of radio system simulation tools, within the context of 3G and beyond 3G systems.