TSG-RAN Working Group 1
Boston
January 15-19

TSGR1#18(01)0080

Agenda item: AH24, HSDPA

Source: Lucent Technologies.

Title: A^2IR – An asynchronous and adaptive Hybrid ARQ scheme for HSDPA

Document for: Discussion and decision

1 Introduction

Two contributions comparing the performance of Incremental Redundancy (IR) versus Chase combining (see [1][2]) were presented at TSGR1#17. In [1], it is argued that for lower order modulation and low code rates, the benefits of IR over Chase combining are not significant. In [2], it is argued that the link level performance of IR is significantly better than Chase combining for higher channel code rates and for higher order modulations. It is also concluded that there is no significant performance difference between IR and Chase combining for low coding rates and smaller modulation sizes.

An asynchronous and adaptive IR (A^2IR) scheme for HSDPA was proposed in [3]. The A^2IR scheme uses a variable TTI [4] with fixed size code blocks. This feature of the scheme allows operation in Chase combining mode for low data rates (using lower coding rates and lower order modulations) and operation in IR mode for high data rates (using higher coding rates and higher order modulations). The link level throughput results for the A^2IR scheme can be found in [5]. This contribution provides further details on the operation of the A^2IR scheme.

2 Asynchronous and Adaptive Operation

2.1 Asynchronous operation

In [6] and [7], an N-channel stop-and-wait (SAW) approach that only provides asynchronous operation within a SAW channel is opted as the favourable solution. We call this scheme a “partial asynchronous” scheme because the asynchronous operation is not allowed across SAW channels.

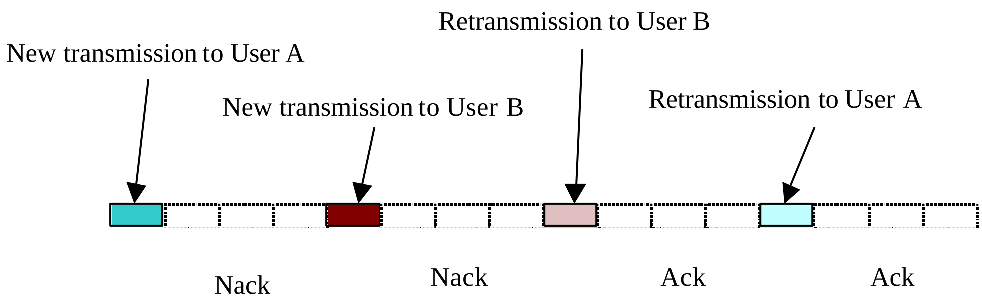

Figure 1. An example of “partial asynchronous” operation [6]

The N-channel Stop-&-Wait as shown in Figure 1 does not provide the flexibility of a full asynchronous operation. The retransmissions are restricted to happen at fixed times (only at integer multiple of N intervals for a given channel). This restriction on retransmissions will interact badly with the scheduler and limit the achievable multi-user diversity gains [9]. The restriction on retransmission timing can be removed by explicit indication of the N-channel identifier through $\log_2(N)$ bits [8].

An equivalent example of fully asynchronous operation is depicted in Figure 2. Note that the retransmission to user A can now be performed in the 6th TTI while in case of "partial asynchronous" operation, the retransmission to user A has to be delayed to 13th TTI. Therefore, the code block transmission times in a partial asynchronous scheme will be larger compared to a full asynchronous scheme. Moreover, with fully asynchronous operation, a large amount of data can be sent to a user when the channel conditions are favorable (Figure 2), thus fully exploiting the multi-user diversity gains [9].

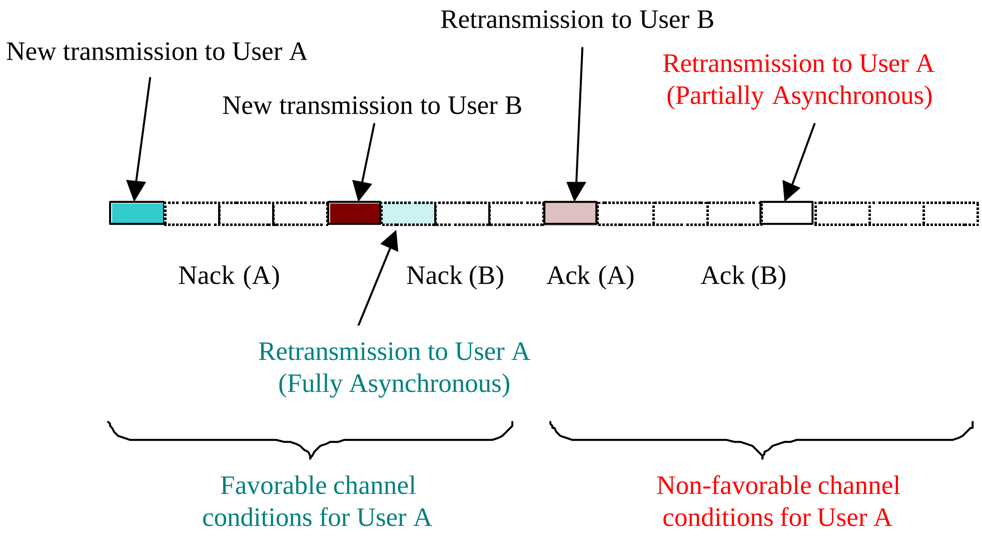

Figure 2. An example of fully asynchronous operation (A^2IR)

The A^2IR scheme assumes a fixed round trip time in order to associate ACK/NACK with the code block transmission. Therefore, ACK/NACK in A^2IR requires a single bit indication. The A^2IR allows N-channel operation through the use of code block identifiers (CBI) (see section 6) that enables fully asynchronous operation. Therefore, the entire HS-DSCH capacity can be assigned to a single user.

2.2 Adaptive operation

The power available for HS-DSCH is continuously changing (on a slot-by-slot basis) due to large variations in the power used by power-controlled circuit switched users. The C/I seen by a user is also varying quickly due to varying interference from neighboring cells and/or large changes in channel quality due to fading etc. Furthermore, with asynchronous IR operation, the time between the transmissions/retransmissions could be longer because a retransmission to a user can be preempted by a transmission/retransmission to another user. Therefore, it is highly likely that the channel conditions, available power and code space are different between transmissions/retransmissions that need to be IR/combined. Under these conditions non-adaptive schemes that do not allow change of MCS on retransmissions will have to abort transmission that will result in degraded system performance. Note that even if the number of codes available for HS-DSCH changes comparatively slowly, a small change in the number of codes makes it impossible to perform a retransmission at the same MCS for non-adaptive HARQ schemes.

Similarly with fast cell site selection (within a Node B or between Node Bs), the channel conditions, available power and the code space will be different in the new cell. With N-channel HARQ, it is highly likely that some of the code blocks will have pending recovery while FCS is performed. Therefore, a true HARQ scheme should be able to do IR/combining across transmissions/retransmissions from different cell sites.

This contribution presents a HARQ scheme that allows IR/combining across transmissions/retransmissions at different MCS (within the same cell or across cell sites).

3 Modulation and coding schemes

3.1 Encoding and sub-block formation

An example of code block encoding and sub-block (unit for incremental redundancy operation) formation is depicted in Figure 3. The code block of size 5120 bits (consisting of 16 transport blocks of size 320 bits) is rate 1/5 Turbo coded to 25600 bits. The sub-blocks for 480 and 960 Kb/s contain the entire sequence of 25600 coded bits i.e., a single sub-block transmission at one of these rates allows 1/5 rate decoding. The subsequent sub-block transmissions at one of these data rates are simply repetitions of the same sequence of bits.

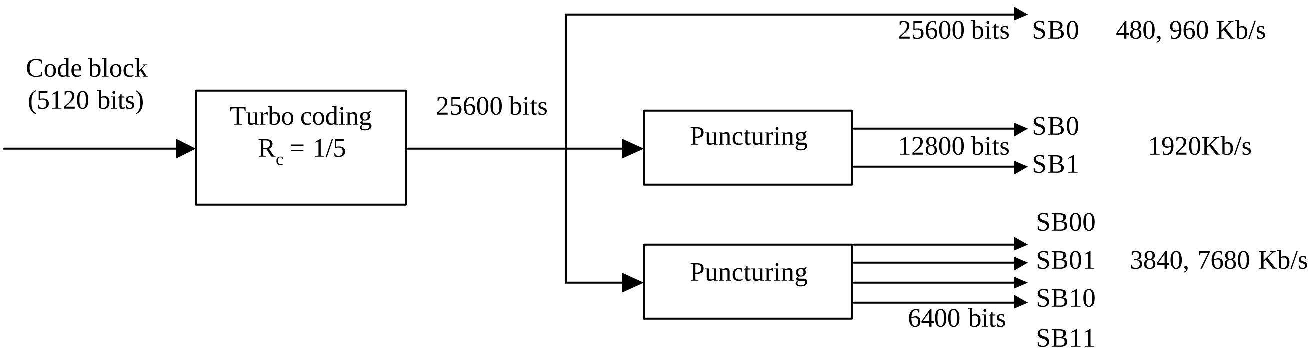

Figure 3. Code block encoding and sub-block formation

For data rates of 1920 Kb/s, two different sub-blocks are obtained by puncturing the 25600 coded bits. The two sub-blocks contain complementary bits. Therefore, the first sub-block transmission will provide coding rate of 0.4 and the first and second sub-block transmissions together provide 1/5 coding rate. If more than two sub-blocks need to be transmitted, the subsequent sub-blocks transmissions will simply be the repetitions of these two types of sub-blocks. For 3840 and 7680 Kb/s, 4 different types (complementary) of sub-blocks of size 6400 bits are obtained by complementary puncturing of 25600 coded bits.

Note that the code block encoding and sub-block formation strategy used here allows for Chase combining at lower coding rates and lower order modulations and the use of incremental redundancy (IR) at higher coding rates and higher order modulations.

An example of mapping of sub-blocks to slots for code blocks of size 5120 bits and different MCS types is shown in Table 1.

Table 1. Mapping of sub-blocks to slots for 5120 bits code block [assumes 20 channelization codes at SF=32]

MCS	Data Rate [Kb/s]	sub-block size [bits]	Modulation	Repetition factor	Modulated symbols	Symbols after spreading	Number of slots/sub-block
4	480	25600	QPSK	2	25600	40960	16
5	960	25600	QPSK	1	12800	20480	8
6	1920	12800	QPSK	1	6400	10240	4
7	3840	6400	QPSK	1	3200	5120	2
8	7680	6400	16-QAM	1	1600	2560	1

3.2 Data Rates

Table 2 and Table 3 summarise different modulation and coding schemes supported in A^2IR. Note that different code block sizes (number of transport blocks) can be selected for the same data rate depending upon the data backlog in the user's buffer in order to minimise frame-fill inefficiency. The data rates lower than 60 Kb/s are achieved by repeating the 60 Kb/s code sub-blocks using the A^2IR scheme. The A^2IR scheme also provides the flexibility to switch to non-MIMO (MIMO) MCS for retransmissions when the original transmission was at MIMO (non-MIMO). For example, in Table 3, the transmissions/retransmissions for a code block of size 10240 bits can occur at any of MIMO MCS (MC8-9) or non-MIMO MCS (MCS4-7) while allowing IR/combining across all these MCS. Note that the MIMO data rates are given for illustration purpose and further simulations are needed in order to tune the rates.

Table 2. Data Rates for non-MIMO case [assumes 20 channelization codes of SF=32]

MCS	Data Rate [Kb/s]	Modulation	Effective coding rate [actual coding + repetition]	Transmission Time Interval (TTI) [number of slots]			
				16 Transport blocks per TTI [code block = 5120 bits]	8 Transport blocks per TTI [code block = 2560 bits]	4 Transport blocks per TTI [code block = 1280 bits]	2 Transport blocks per TTI [code block = 640 bits]
1	60	QPSK	0.0125				16
2	120	QPSK	0.0250			16	8
3	240	QPSK	0.0500		16	8	4
4	480	QPSK	0.1000	16	8	4	2
5	960	QPSK	0.2000	8	4	2	1
6	1920	QPSK	0.4000	4	2	1	
7	3840	QPSK	0.8000	2	1		
8	7680	16-QAM	0.8000	1			

Table 3. Data Rates with 2 receive antennas [assumes 20 channelization codes of SF=32]

MCS	Data Rate [Kb/s]	# of transmit antennas	Modulation	Effective coding rate [actual coding + repetition]	Transmission Time Interval (TTI) [number of slots]	
					32Transport blocks per TTI [code block = 10240 bits]	16 Transport blocks per TTI [code block = 5120 bits]
4	480	1	QPSK	0.1		16
5	960	1	QPSK	0.2	16	8
6	1920	1	QPSK	0.4	8	4
7	3840	1	QPSK	0.8	4	2

8	7680	2	QPSK	0.8	2	1
9	15360	2	16-QAM	0.8	1	

3.3 Sub-block starting rates

The starting rate is selected based on the link quality feedback from the receiver. In order to benefit from HARQ operation, the starting rate is always selected to be greater than or equal to the supportable rate determined by the mapping from link quality feedback. The larger the difference between the starting rate and the supportable rate better the throughput due to improved IR/combining granularity. However, the delays are directly proportional to the rate difference because the average number of transmission attempts needed to recover a code block increases as the starting rates become more and more aggressive relative to the supportable rate. The starting rates corresponding to different supportable rates are given in Table 4. The supportable rate would be determined based on mapping of link quality feedback to rate.

Table 4. A^{2}IR sub-block starting rates

Supportable rate [Kb/s]	Sub-block starting rate [Kb/s]
15	60
30	120
60	240
120	480
240	960
480, 960	1920
1920	3840
3840	7680
7680	7680

3.4 IR/combining across different MCS

The A^{2}IR scheme provides the flexibility to adapt the sub-block transmission rate according to the most recently received link quality feedback from the receiver. Figure 4 shows an example of A^{2}IR operation where the sub-blocks from the same code block of size 5120 bits are transmitted at different modulation and coding schemes.

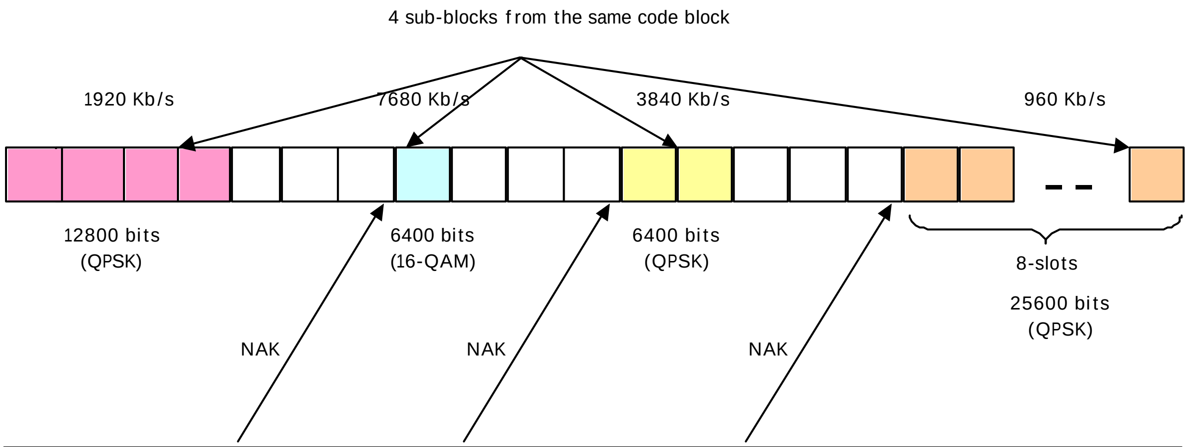

Figure 4. Sub-blocks from the same code block are transmitted at 4 different MCS.

The fact that all sub-blocks are derived from the same coded sequence of 25600 bits makes it possible to perform IR/combining across these sub-blocks in order to recover the original code block as depicted in Figure 5. The first sub-block contains half of the coded bits i.e., 12800 provides a coding rate of 0.4. The second sub-block contains 6400 bits that are complementary to 12800 bits received in the first sub-block. The first and second sub-blocks jointly provide a coding rate of 0.266. The third sub-block contains 6400 bits that are complementary to 12800 bits received in the first sub-block and 6400 bits received in the second sub-block. The fourth sub-block transmission at 960 Kb/s provides all the 25600 coded bits. The bits from first, second and third sub-blocks also provide the entire sequence of 25600 coded bits. Therefore, the first three transmissions can be Maximal ratio combined with the 4th sub-block transmission.

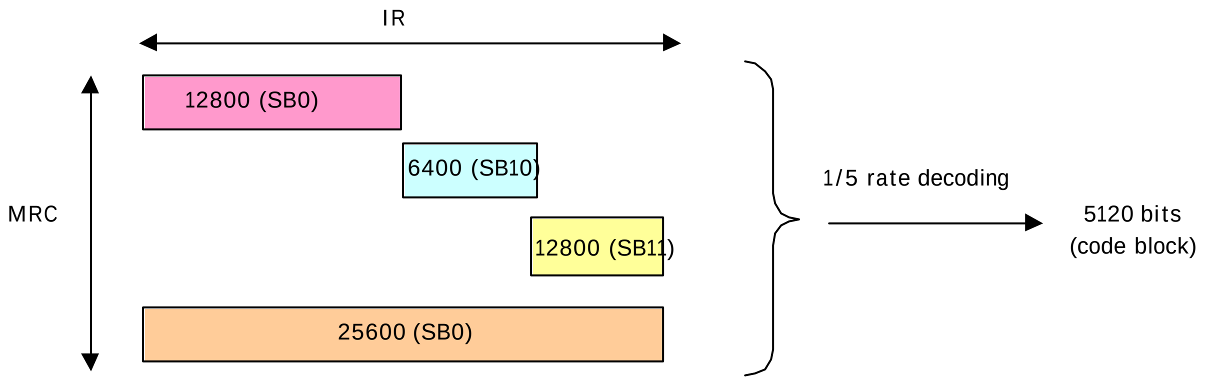

Figure 5. IR/MRC of sub-blocks received at different MCS.

4 Operation across non-homogeneous channel structures

4.1 Dynamically changing channel structure within the same cell

The power available for HS-DSCH is continuously changing (on a slot-by-slot basis) due to large variations in the power used by power-controlled circuit switched users. The C/I seen by a user is also varying quickly due to varying interference from neighboring cells and/or large changes in channel quality due to fading etc. Furthermore, with asynchronous IR operation, the time between the transmissions/retransmissions could be longer because a retransmission to a user can be preempted by a transmission/retransmission to another user. Therefore, it is highly likely that the channel conditions, available power and code space are different between transmissions/retransmissions that need to be IR/combined. The A^{2}IR scheme allows for IR/combining from transmissions at different MCS under different power and code profiles. For example, in Figure 4 and Figure 5, the 4 sub-block transmissions could have been performed under different power and code profiles.

4.2 Operation with fast cell site selection

With fast cell site selection, the outstanding code blocks can be recovered by retransmissions from a cell different from where the original transmissions were made. However, the channel conditions, number of codes and the available power allocated for HS-DSCH may be different in different cells.

The A^{2}IR scheme provides considerable flexibility by allowing IR/combining of transmissions from different cells under different MCS, power and code profiles. For example, in Figure 4 and Figure 5, the 4 sub-block transmissions could potentially come from 4 different cell sites.

5 Operation in pure Link Adaptation (LA) mode

A^{2}IR can be operated in the pure LA mode by making the starting rates equal to the supportable rate determined based on link quality feedback. . The LA operation would be useful for UEs that choose not to implement the HARQ functionality.

6 Parallel channel operation

The use of a code block identifier (CBI) allows for parallel multiple code blocks transmissions from the same user. Note that the 2-bit CBI field is used to identify the code block at the physical layer. With a 2-bit CBI, up to 4 code blocks can be transmitted in parallel as shown in Figure 6. Note that the CBIs can be reused as soon as they become available. In Figure 6, code block A2 (with CBI=2) is ACK'ed after only one sub-block (A21 in slot 2) transmission. The same CBI is used for another code block transmission in slot 6.

The parallel transmissions of code blocks allow for exploiting of multi-user diversity [9] because a large amount of data can be sent to a user when the channel conditions are good.

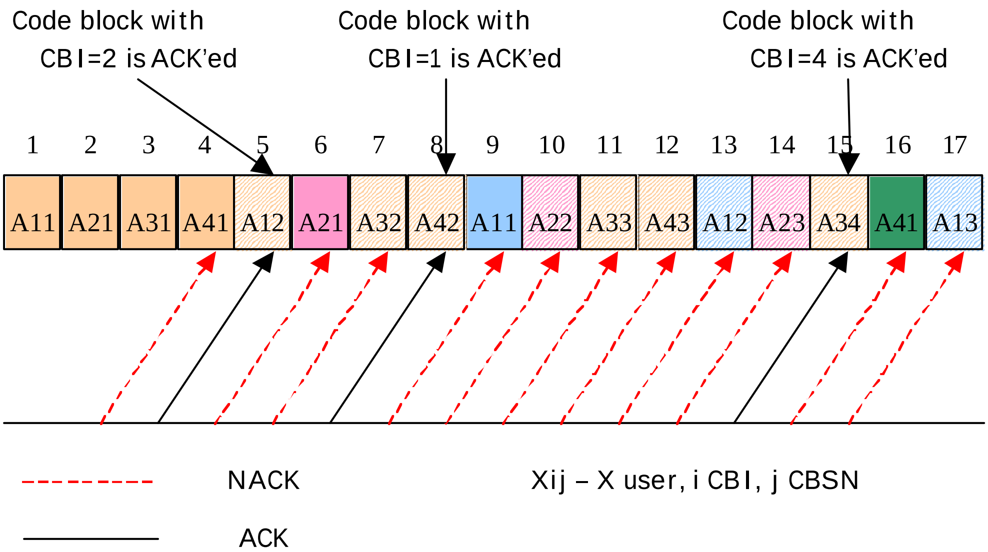

Figure 6: A^2IR operation with multiple channels.

When sub-block sequence numbers are used in conjunction with code block identifiers, the identifier ambiguity problem can be solved. For example, when code block with CBI of 2 is ACK'ed in slot 5, it is possible that a NACK was misinterpreted as an ACK. Then the receiver will be waiting for sub-block A22 but on receiving sub-block A21, it will know (through sub-block start sequence number) that this is a new sub-block on channel 2 (CBI of 2) and will not try to combine it with previously received A21.

7 Early ACK

The Early ACK feature of the A^2IR scheme can reduce the transmission times and improve throughput for cases where the sub-block transmission time is greater than the feedback delay.

An example of Early ACK operation is depicted in Figure 7. The transmission rate for user A is 960 Kb/s that requires transmission over 8 slots for code block size of 5120 bits. The ACK/NACK feedback delay is assumed to be 4-slots (i.e., the ACK/NACK feedback for transmission in slot n is received in slot n+4). In this example, a NACK is received after the first sub-block transmission to user A. The transmitter scheduled a 2-slot transmission (at rate 3840Kb/s) for user B during the user A idle period (due to ACK/NACK feedback). After receiving the NACK for first sub-block to user A, the transmitter starts another 8-slot transmission to user A. But the receiver successfully decodes the code block after 2 slots have been transmitted and sends back an ACK. The transmitter aborts transmission of the sub-block after receiving the ACK in slot 5 of second sub-block. In this way, 3-slots worth of transmission time is saved compared to a scheme where all the 8-slots were used for sub-block transmission. The transmitter can use the newly available slots for transmitting to another user (For example, user B in Figure 7).

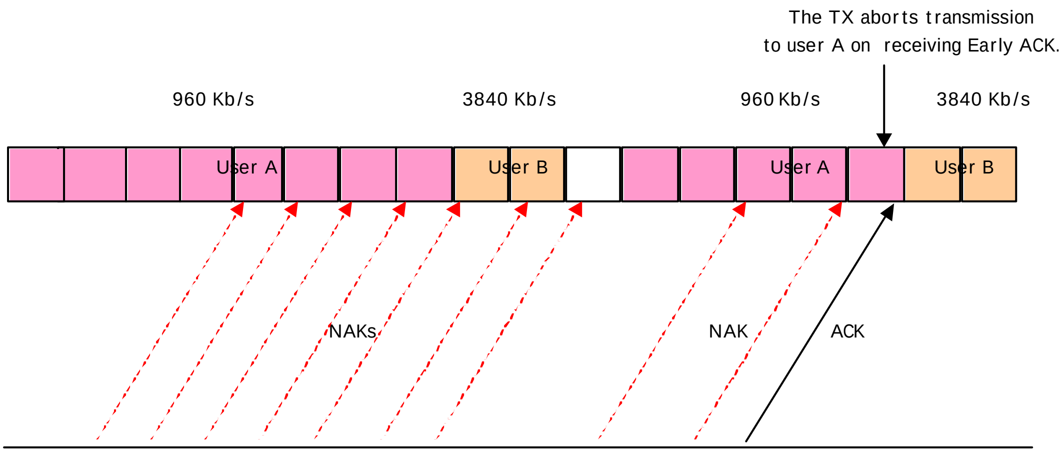

Figure 7. Early ACK operation

8 Receiver Operation

Figure 8 shows the receiver operation for the proposed A^{2}IR scheme. A brief description of the flow chart follows.

The sub-block sequence number (SBSN) can be used to distinguish between NEW and CONTINUE sub-blocks. When the receiver is waiting for a NEW encoded sub-block and receives a NEW encoded sub-block, it attempts to decode it. If the decoding is successful it transmits an ACK and then waits for the next NEW encoded sub-block. If the decoding is unsuccessful, it transmits a NACK, stores the received encoded sub-block for subsequent combining with redundant information, and waits to receive a CONTINUE encoded sub-block. If the receiver receives a CONTINUE encoded sub-block while waiting for one, it attempts decoding by combining the received encoded sub-block with the stored information. If the decoding is successful, it transmits an ACK and waits for the next NEW encoded sub-block, while if decoding is unsuccessful, it transmits a NACK and then waits for the next CONTINUE encoded sub-block.

For the case, when a receiver receives a NEW encoded sub-block while waiting for a CONTINUE, the receiver abandons recovery of the previous code block and moves on to attempt to decode the code block corresponding to the new received encoded sub-block.

When a receiver receives a CONTINUE encoded sub-block while waiting for a NEW, the receiver just discards the received encoded sub-block and continues to wait for a NEW encoded sub-block.

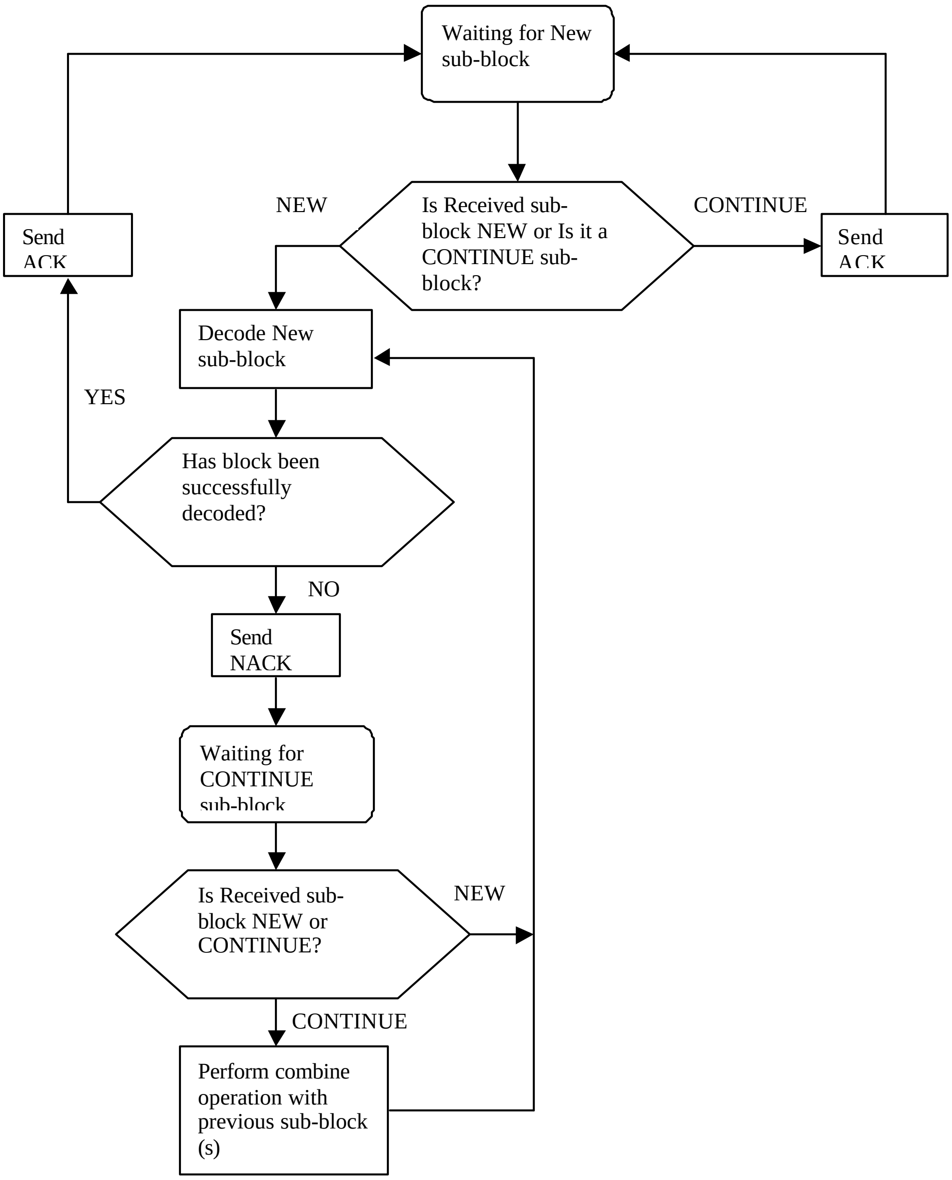

Figure 8. Receiver operation for the A^2IR scheme.

9 Signalling requirements

The control fields needed for A^2IR operation are shown in Table 5.

Table 5. Control fields for A^2IR

Control Field	**Number of bits**
Code Block Identifier (CBI)	2
Sub-block Sequence Number (SBSN)	2
Code Block Size (CBS)	2
	Total bits = 6

10 Concluding Remarks

The detailed description of the asynchronous and adaptive IR scheme (A^2IR) is provided. The key features and benefits of the A^2IR scheme are summarised below:

?? The A^2IR scheme provides the full asynchronous operation (and scheduling flexibility) in order to fully exploit the multi-user diversity gains.

?? The scheme provides the flexibility to perform IR/Combining across different rates (different MCS).

?? The scheme also allows the flexibility of selecting different code block sizes for the same data rate (to avoid frame-fill inefficiency).

?? The scheme provides a transparent IR/combining operation when the channel conditions, codes and/or power allocated to HS-DSCH changes in the same cell.

?? The scheme allows for transparent fast cell site switching operation i.e., any modulation and coding scheme (supported by channel conditions, available power and number of codes in the new cell) can be used to recover outstanding (ongoing transmission) code blocks in the new cell.

?? The scheme provides a transparent IR/combining operation across MIMO and non-MIMO rates i.e., the retransmissions could be at non-MIMO (MIMO) rate when the original transmission was performed at MIMO (non-MIMO) rate.

?? The scheme can be operated in pure rate controlled (link adaptation) mode. This feature is useful for UEs that choose not to implement the HARQ functionality.

?? The Early ACK feature of the A^2IR scheme can reduce the transmission times and improve throughput for cases where the sub-block transmission time is greater than the feedback delay.

11 References

[1] "Performance Comparison of Hybrid-ARQ Schemes", TSG-RAN #17(00) 1396, Motorola.

[2] "Performance Comparison of Chase Combining and Incremental Redundancy", TSG-RAN #17(00) 1428, Ericsson.

[3] "Asynchronous and Adaptive IR (A^2IR) for HSDPA", Lucent Technologies, TSG-RAN #17(00) 1382, Stockholm, Sweden, November 2000.

[4] "Variable TTI for HSDPA", TSG-RAN #18(01) 0079, Lucent Technologies.

[5] "Throughput Results for A^2IR for HSDPA", TSG-RAN #18(01) 0082, Lucent Technologies.

[6] "Discussion on topics and issues raised on fast Hybrid ARQ proposals for HSDPA", TSG-RAN #18(01) xxxx, submitted to TSG_RAN_WG1 mailing list, Ericsson.

[7] "Clarifications on Dual-Channel Stop-and-Wait HARQ", TSG-RAN #18(01) 0048, Motorola.

[8] "Considerations on HSDPA HARQ concepts", TSG-RAN #18(01) 0007, Nokia.

[9] David Tse, "Multi-user Diversity in Wireless Networks: From Information Theory to Architecture", MIT EECS Colloquium, November 13, 2000.