HARQ Timing and Protocol Considerations for IEEE 802.16m

Document Number: C80216m-08/362r2
Date Submitted: 2008-05-07

Source: Hujun Yin (hujun.yin@intel.com) Intel Corporation
Yujian Zhang (yujian.zhang@intel.com) Intel Corporation
Yuan Zhu (yuan.zhu@intel.com) Intel Corporation
Yuval Lomnitz (yuval.lomnitz@intel.com) Intel Corporation
Tom Harel (tom.harel@intel.com) Intel Corporation
Huaning Niu (huaning.niu@intel.com) Intel Corporation
Xiangying Yang (xiangying.yang@intel.com) Intel Corporation

Re: IEEE 802.16m-08/005: Call for Contributions on Project 802.16m System Description Document (SDD)
HARQ timing and protocol structure

Venue: Macau, China

Base Contribution:

Purpose: To be discussed and adopted by TGm for 802.16m SDD

Outline

- HARQ overview and executive summary
- HARQ timing process
- HARQ overhead analysis
  - Overhead of normal allocation
  - Overhead of persistent allocation
- Dynamic interference
- PHY considerations
- ARQ/HARQ coupling
- Summary and recommendations

HARQ in IEEE 802.16e

- For both DL and UL, **asynchronous** HARQ is used.
  - Retransmission timing is flexible
  - Control signaling is needed for both the initial transmission and retransmissions
- Basic fields for HARQ allocation
  - CID
  - Resource allocation
  - MCS/mode indication
- Additional HARQ information
  - SPID: 2 bits (redundancy version)
  - ACID (HARQ CH ID): 4 bits (HARQ process number)
  - New data indicator: 1 bit

HARQ Classification

HARQ types	Retransmission timing	Retransmission resource allocation and MCS
Synchronous non-adaptive	Predetermined delay	Predetermined MCS Predetermined resource allocation
Synchronous adaptive	Predetermined delay	Adaptive MCS Flexible resource allocation
Asynchronous non-adaptive	Variable delay	Predetermined MCS Flexible resource allocation
Asynchronous adaptive	Variable delay	Adaptive MCS Flexible resource allocation

HARQ Comparison

	Synchronous Non-adaptive	Synchronous Adaptive	Asynchronous Non-adaptive	Asynchronous Adaptive
Retransmission scheduling flexibility	**Restricted in both time and resource/MCS allocation** • Difficult to handle urgent data • Fragmented resource allocation • Limited retransmission diversity • Limited link adaptation	**Restricted in time and flexible in resource/MCS allocation** • Difficult to handle urgent data • Limited time diversity	**Restricted in retransmission MCS and flexible in time and resource allocation** • Limited link adaptation	**Flexible in both time and resource/MCS allocation**
Signaling overhead	**Low** • Full signaling for initial transmission • No signaling for re-transmission	**High** • Full signaling for initial transmission • Full or partial signaling for re-transmission	**Moderate** • Full signaling for initial transmission • Partial signaling for re-transmission	**High** • Full signaling for initial transmission • Full or partial signaling for re-transmission
Robustness	**Low** • Signaling only available in initial transmission	**High**	**High**	**High**
Delay	**Minimum**	**Minimum**	**Longer** • Can be minimum if follow minimum latency allocation	**Longer** • Can be minimum if follow minimum latency allocation
Power Saving	**Friendly**	**Friendly**	**Less friendly**	**Less friendly**
Notes	**Low flexibility, low overhead**	**Moderate flexibility, high overhead**	**Moderate flexibility, moderate overhead**	**High flexibility, high overhead**

Control Overhead/Scheduling Flexibility Consideration

- HARQ control overhead depends on scheduling method
  - Asynchronous HARQ has higher overhead than synchronous HARQ with normal scheduling (5~20%), which translates to 1~4% additional system overhead assuming 20% total control overhead
  - Asynchronous HARQ can accommodate more persistent scheduling than synchronous HARQ with the flexibility to reschedule retransmissions and resulting in lower overall overhead (5 times reduction)
- Synchronous HARQ pre-schedule retransmissions in fixed time/frequency
  - If first HARQ transmission collides with an interfering HARQ transmission, then retransmissions will collide repeatedly, resulting in poor performance
  - The interference issue will be particularly significant with downlink directional transmission and uplink transmission
- HARQ retransmission flexibility over weights control overhead saving due to considerations of
  - Persistent scheduling
  - Downlink directional transmission
  - Uplink transmission

HARQ PHY and MAC Protocol Consideration

- H-ARQ Incremental Redundancy (IR)
  - Chase Combining (CC) is a special case of IR
  - Improvement based on 802.16e HARQ-IR & CTC
- Coupled ARQ/HARQ operations
  - HARQ performs most error correction and feedbacks
  - ARQ maintains E2E reliability with minimal overhead
  - Modify ARQ suitable for coupling with HARQ operation
  - Robust and quick method to allow correcting HARQ feedback and residual error

HARQ Timing and Protocol Recommendations

- Support HARQ flexibility to schedule retransmission at different time/frequency slots
  - Persistent scheduling
  - Downlink directional transmission
  - Uplink transmission
- Constrain maximum retransmission delay
  - Limit overall transmission latency
  - Save power and memory
- Support flexible HARQ IR by extending 16e FEC
- Support HARQ/ARQ coupling for fast HARQ error recovery and reduced overhead

Proposed SDD Text for HARQ

Insert the following text into MAC Layer clause (Chapter 10 in [IEEE 802.16m 08/003r1])

------------------------------ Text Start ------------------------------

10.x.x.x HARQ timing and protocol

Support N-process Stop-And-Wait HARQ protocol. Support asynchronous HARQ in both downlink and uplink. The retransmissions can be scheduled in time/frequency slots different from initial transmission. The maximum HARQ retransmission delay is bounded. MCS adaptation for retransmissions is for FFS.

10.x.x.x HARQ/ARQ interactions

ARQ and HARQ-ARQ SAP allow coupled HARQ-ARQ operation. HARQ performs most error correction and feedback. ARQ maintains E2E reliability with minimal overhead, and allows robust and quick correction of HARQ feedback and residual error

------------------------------ Text End ------------------------------

Proposed SDD Text for HARQ

Insert the following text into Physical Layer clause (Chapter 11 in [IEEE 802.16m-08/003r1])

------------------------------ Text Start ------------------------------

11.x.x.x HARQ packet encoding

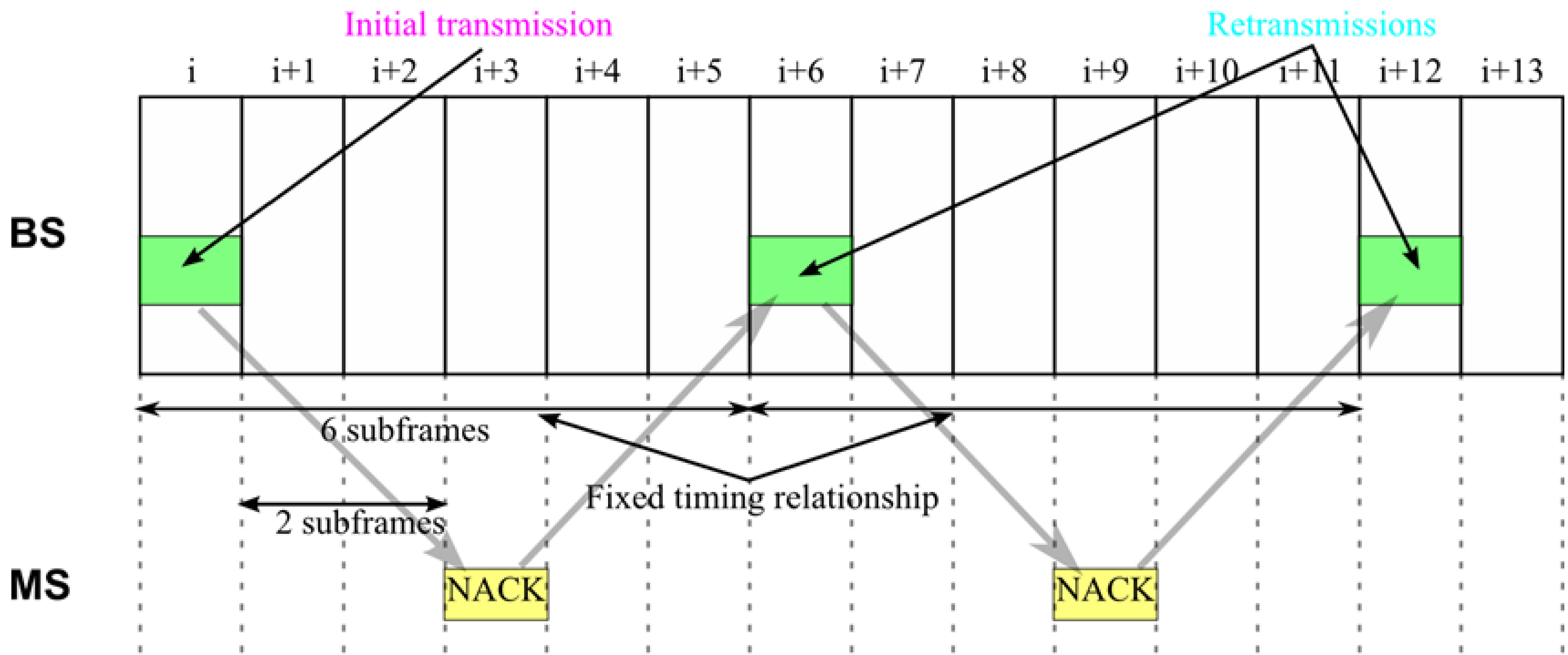
Support incremental redundancy (IR) for HARQ packet encoding. Support Chase Combining (CC) as a special case of IR.

Extend 802.16e CTC coding scheme to support HARQ IR with lower mother code rate, finer rate and code block granularity

------------------------------ Text End ------------------------------

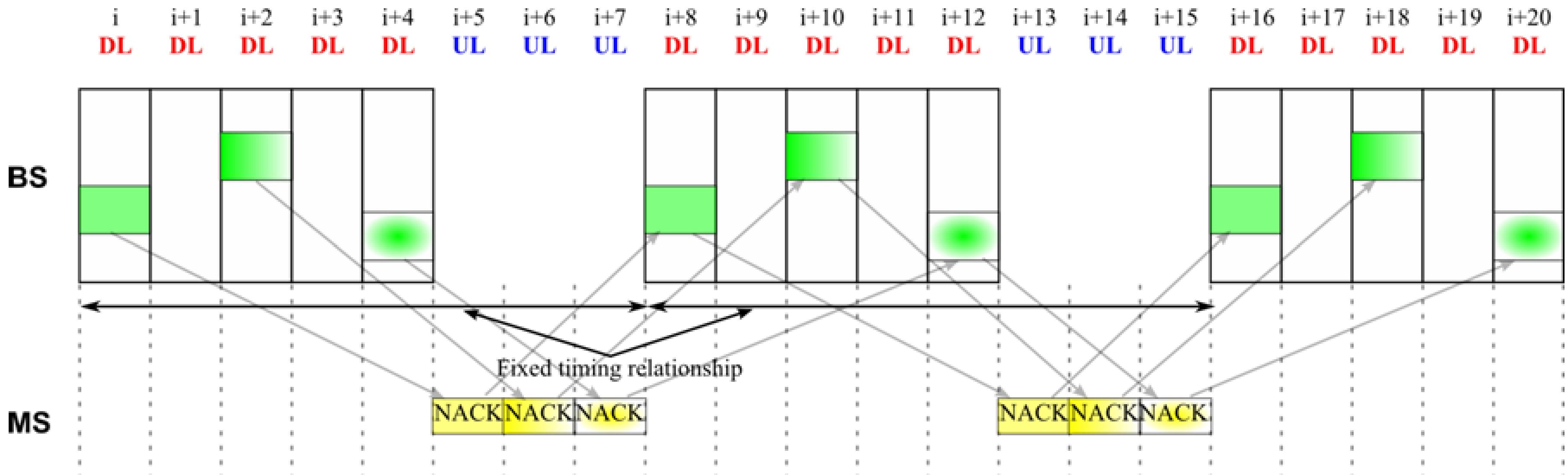
HARQ Timing Process

Synchronous HARQ Process (FDD)



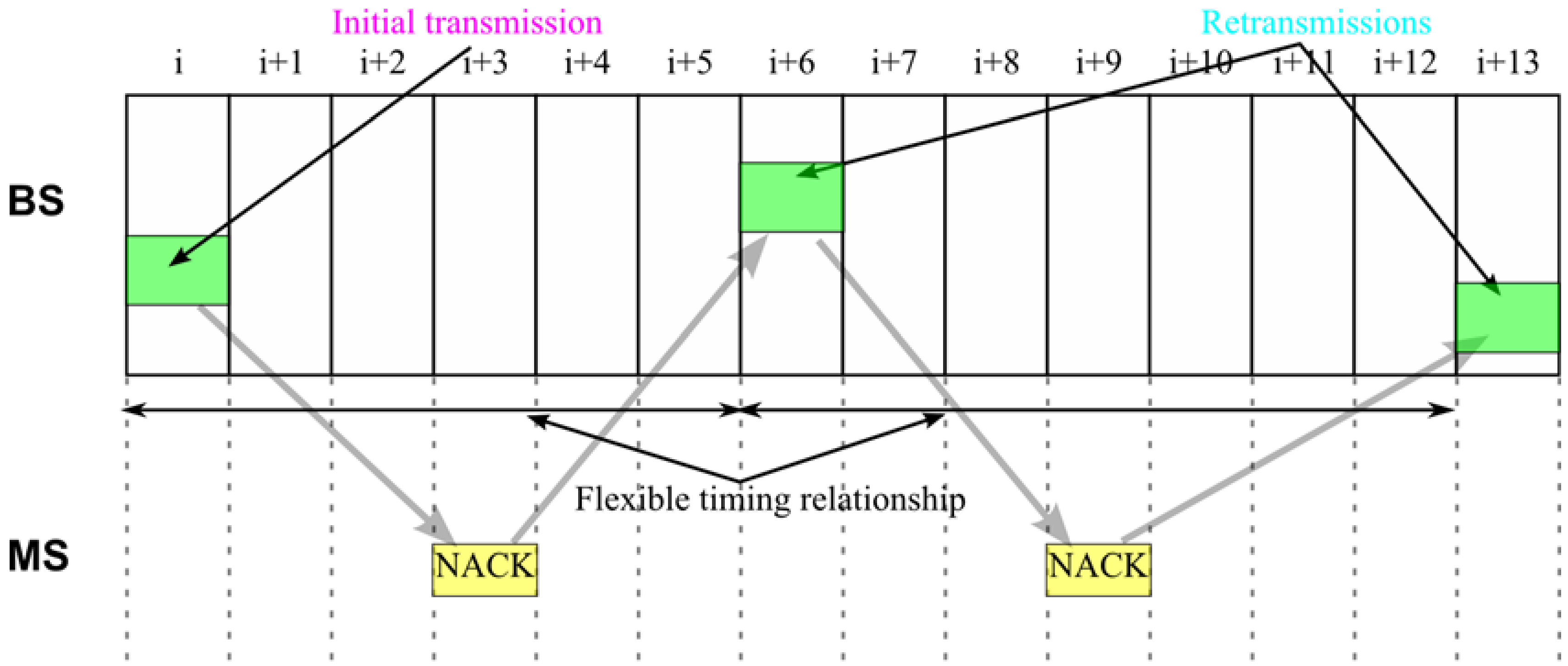
- Transmission-ACK/NACK delay: 2 subframes
- NACK-Retransmission delay: 2 subframes

Synchronous HARQ Process (TDD 5:3)



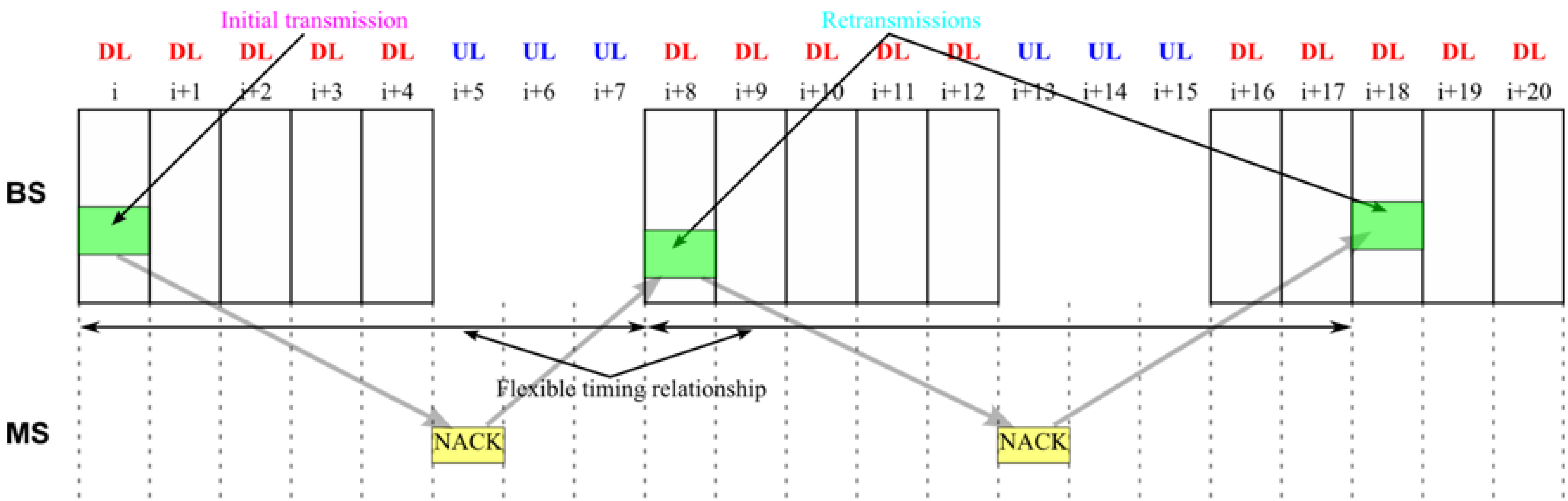
- Transmission-ACK/NACK delay: max(2 subframes, associated UL subframe)
- NACK-Retransmission delay: max(2 subframes, associated DL subframe)

Asynchronous HARQ Process (FDD)



- Transmission-ACK/NACK delay: 2 subframes
- NACK-Retransmission delay: $\geqslant$2 subframes

Asynchronous HARQ Process (TDD 5:3)



- Transmission-ACK/NACK delay: max(2 subframes, associated UL subframe)
- NACK-Retransmission delay: ≥max(2 subframes, associated DL subframe)

HARQ Signaling Overhead Analysis

Signaling Overhead Comparison: Assumptions

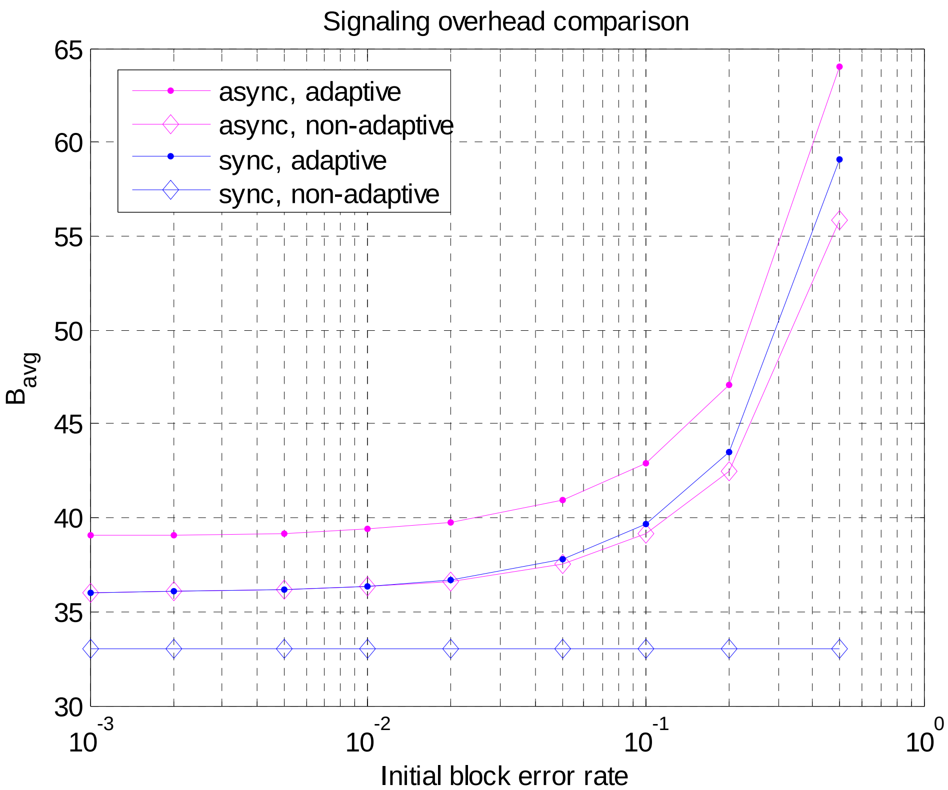
- Focus on overhead for DL HARQ
  - Overhead of UL HARQ would be similar
- Assumptions
  - Since there is no definition of related fields in 802.16m, signaling overhead is estimated
  - Resource allocation: highly dependent on bandwidth and actual schemes used. Assuming 10 MHz system bandwidth, using 12 bits for resource indication.
  - HARQ process number: 3 bits
  - Modulation and coding: 5 bits
  - Redundancy version: 2 bits
  - New data indicator: 1 bit

HARQ Control Signal

Field	Asynchronous				Synchronous			
	Adaptive		Non-adaptive		Adaptive		Non-adaptive	
	Init. Tx	Re. Tx	Init. Tx	Re. Tx	Init. Tx	Re. Tx	Init. Tx	Re. Tx
Resource allocation	12	12	12	12	12	12	12	0
HARQ process number	3	3	3	3	0	0	0	0
Modulation and coding	5	5	5	0	5	5	5	0
Redundancy version	2	2	0	0	2	2	0	0
New data indicator	1	1	0	0	1	1	0	0
CID	16	16	16	16	16	16	16	0
Total	39	39	36	31	36	36	33	0

Non-persistent Scheduling Overhead

- Average total number of DL control signaling bit (B_{avg})
  - Block error rate of initial transmission is P, after 1st retransmission is P^2, after 2nd retransmission is P^3
  - Maximum number of transmissions is 4
- $B_{avg} = (1-P)B_{init} + P(1-P^2)(B_{init} + B_{re}) + PP^2\,(1-P^3)((B_{init} + 2B_{re}) + PP^2P^3((B_{init} + 3B_{re}) = B_{init} + (P + P^3 + P^6)B_{re}$



Overhead Comparison

Initial block error rate	0.001	0.002	0.005	0.01	0.02	0.05	0.1	0.2	0.5
async+adaptive over sync+adaptive (%)	8.3	8.3	8.3	8.3	8.3	8.3	8.3	8.3	8.3
async+non-adaptive over sync+non-adaptive (%)	9.2	9.3	9.6	10.0	11.0	13.8	19	28.6	69.3
async+adaptive over sync+non-adaptive (%)	18.3	18.4	18.8	19.4	20.5	24.1	30	42.8	93.9

- For typical initial block error rate (10%)
  - The gain of synchronous vs. asynchronous is 8.3% for adaptive case, and 19% for non-adaptive case.
  - The gain of sync+non-adaptive vs. async+adaptive is 30%

Persistent Scheduling Overhead

- Persistent synchronous HARQ
  - Initial transmissions and retransmissions are persistently scheduled
  - Resources freed due to early HARQ termination are dynamically scheduled
- Persistent asynchronous HARQ
  - Initial transmissions are persistently scheduled
  - Re-transmissions are dynamically scheduled
- Assumptions
  - In period *T*, there are *M* resource units for scheduling
  - Maximum *N* HARQ transmissions (including initial transmission)
  - Block error rate of initial transmission is *P*
  - For typical scenario (e.g. $P=0.1$), only 1st retransmission likely to occur, ignore 2 or more retransmissions
  - The overhead of persistent allocation is ignored
- Notations
  - *As, Ad, At*: number of persistent allocations, dynamic allocations and total allocations during period T.
  - *B:* total signaling overhead.

	Synchronous	**Asynchronous**
A_s	M/N	$M/(1+P)$
A_d	$M/(1+P) – M/N$	0
A_t	$M/(1+P)$	$M/(1+P)$
B	$M(1/(1+P) – 1/N)B_{sync,ini}$	$MPB_{async,re}/(1+P)$

Signaling Overhead Comparison: Persistent Scheduling

- P = 0.1

N	2	3
Overhead of sync+adaptive over async+adaptive (%)	356.9	543.1
Overhead of sync+non-adaptive over async+non-adaptive (%)	426.9	641.6
Overhead of sync+non-adaptive over async+adaptive (%)	318.8	489.5

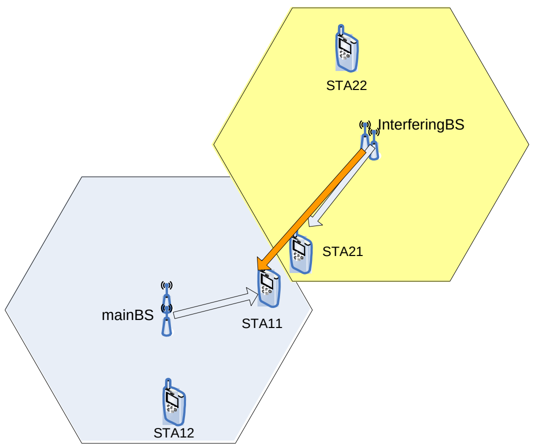
- Synchronous HARQ incurs SIGNIFICANT signaling overhead!

Signaling Overhead Comparison: Summary

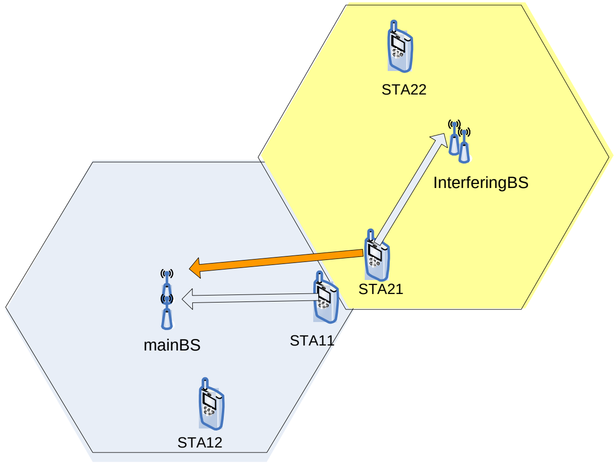
- For *non-persistent* scheduling
  - The gain of synchronous HARQ over asynchronous HARQ is ***not*** significant.
    - Assume initial block error rate is 0.1.
    - Signaling increase of asynchronous HARQ over synchronous HARQ is: 8.3% for adaptive case, and 19% for non-adaptive case
    - Assume 20% of scheduling overhead, 19% of additional overhead translates to 4% reduction in resource available for data, which may be compensated by adaptation gain
- For *persistent* scheduling:
  - The gain of asynchronous HARQ over synchronous HARQ is ***significant***.
    - Assume initial block error rate is 0.1 and maximum number of transmissions is 3.
    - Signaling increase of synchronous HARQ over asynchronous HARQ is
      - 543% for adaptive case
      - 641.6% for non-adaptive case
      - 489.5% for asynchronous+adaptive HARQ over synchronous+non-adaptive HARQ
- HARQ retransmission flexibility is more important than control overhead especially for persistent scheduling

HARQ Additional Considerations

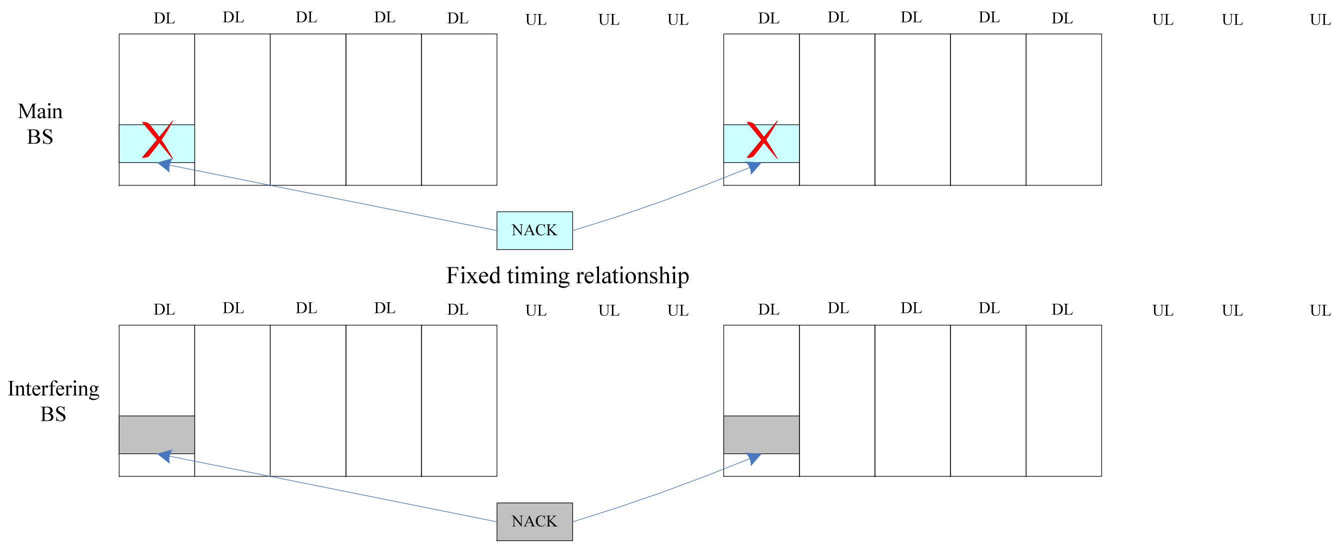
Sync HARQ Interference Example



DL interference with beamforming



UL interference with power control



Retransmission collides with Sync HARQ.

Interference Issue with Sync HARQ

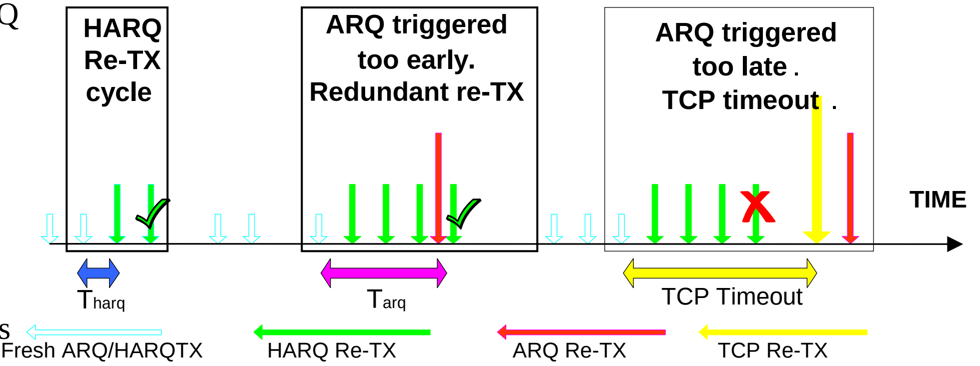
- Synchronous HARQ pre-schedule retransmissions in fixed time/frequency slots
- If the first transmission collide due to interference, then retransmissions will collide repeatedly
  - HARQ performance will be dramatically reduced
  - The interference issue will be particularly significant with downlink directional transmission and uplink transmission
- HARQ retransmission should be able to change sub-frame or sub-channel
  - With downlink beamforming, the re-transmissions should be on different sub-frames to maintain the beamforming vector

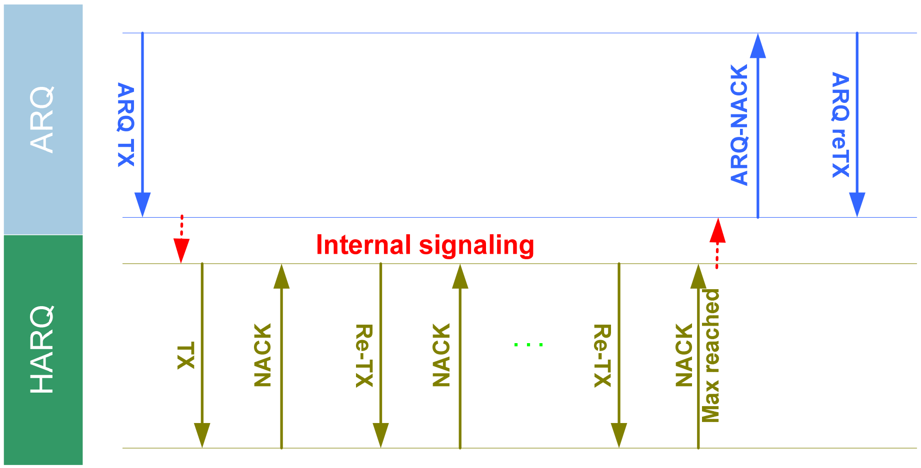
HARQ PHY Concepts

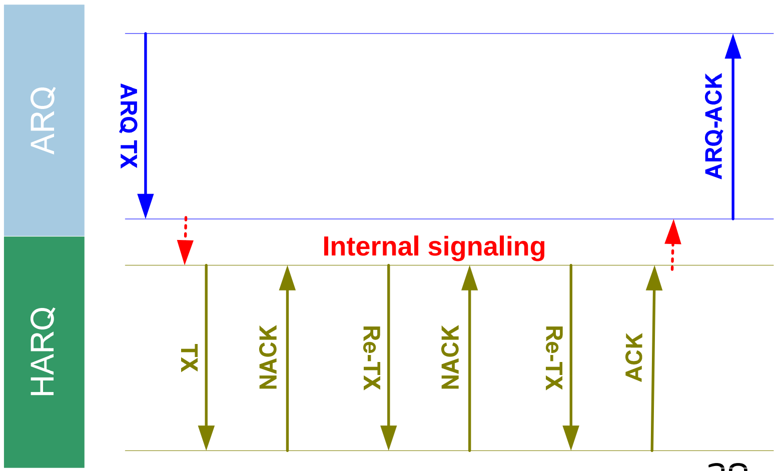
- Although H-ARQ PHY and FEC is not discussed here, some PHY concepts are relevant to the H-ARQ protocol, and possible concepts are listed below to provide background
- Support H-ARQ Incremental Redundancy (IR)
  - Chase Combining (CC) is a special case of IR
- Consider 802.16 HARQ-IR scheme & CTC as base, with some modifications, e.g.
  - Constellation re-arrangement for HARQ re-transmission of same coded bits
  - Finer granularity of block size
  - Lower mother code rate
  - Rate matching (for variable pilot overhead, shortened sub-frame, small packet efficiency). Hence definition of code rate may change to continuous scale (specify #resources,modulation, and #bits, not MCS)
  - Support of variable re-transmission size

ARQ/HARQ Coupling Operations

- Without coupling
  - waste resources: redundant retransmission, extra feedback
  - Latency: conservative ARQ retry timer leads to slow ARQ retransmission
  - UL signaling overhead for ARQ feedback message
  - ARQ still necessary for good TCP performance and HARQ feedback errors
- Basic coupled HARQ-ARQ operations
  - ARQ ACK is internally triggered by HARQ ACK
  - ARQ NACK is internally triggered by HARQ reaching its max retransmission count.
  - NACK→ACK error need to be considered.







HARQ-ARQ Coupling Recommendations

- HARQ performs most error correction and feedbacks
- ARQ maintains E2E reliability with minimal overhead
- HARQ-ARQ SAP allows coupled HARQ-ARQ operations
- Modify ARQ suitable for coupling with all HARQ operation modes
- Robust and quick method to allow correcting HARQ feedback and residual error