Project	**IEEE 802.16 Broadband Wireless Access Working Group** <http://ieee802.org/16>
Title	**Efficient CQI Feedback for 802.16m MIMO Transmission**
Date Submitted	**2008-09-05**
Source(s)	Eldad Zeira eldad.zeira@interdigital.com Ron Murias InterDigital Communications, LLC
Re:	SDD Session 56 Cleanup, in response to the call for PHY details: "Any parts of Section 11 (PHY) that are incomplete, inconsistent, empty, TBD, or FFS."
Abstract	The contribution suggests an efficient CQI feedback mechanism and proposes modifications to the SDD.
Purpose	To be discussed and adopted by TGm
Notice	*This document does not represent the agreed views of the IEEE 802.16 Working Group or any of its subgroups*. It represents only the views of the participants listed in the "Source(s)" field above. It is offered as a basis for discussion. It is not binding on the contributor(s), who reserve(s) the right to add, amend or withdraw material contained herein.
Release	The contributor grants a free, irrevocable license to the IEEE to incorporate material contained in this contribution, and any modifications thereof, in the creation of an IEEE Standards publication; to copyright in the IEEE's name any IEEE Standards publication even though it may include portions of this contribution; and at the IEEE's sole discretion to permit others to reproduce in whole or in part the resulting IEEE Standards publication. The contributor also acknowledges and accepts that this contribution may be made public by IEEE 802.16.
Patent Policy	The contributor is familiar with the IEEE-SA Patent Policy and Procedures: <http://standards.ieee.org/guides/bylaws/sect6-7.html#6> and <http://standards.ieee.org/guides/opman/sect6.html#6.3>. Further information is located at <http://standards.ieee.org/board/pat/pat-material.html> and <http://standards.ieee.org/board/pat>.

Efficient CQI Feedback for 802.16m MIMO Transmission

Eldad Zeira, Ron Murias

InterDigital Communications, LLC

Introduction

Modern communication standards adapt the link to the channel by the appropriate selection of block size, forward error correction, rank and other transmission parameters. Channel quality information (CQI) feedback, transmitted by the MS, has a key role in the link adaptation and must be provided on a timely basis. As system bandwidth grows, and with it the number of resource blocks, so does the amount of information in the CQI which can therefore create a significant uplink overhead. It is therefore important to find means to reduce the number of bits without significantly degrading the quality of CQI information.

One of the well known ways to reduce the amount of information is for the MS to indicate only the best resource blocks and the wideband average. This technique is called best-M reporting and has been included in 802.16e (with M=5). For each of those M RB's reported, MS must indicate which RB it is and the value of the CQI for the RB.

The work on downlink MIMO and uplink control is not yet done and some key decisions have not yet been made. Looking back to 802.16e and other standards we have assumed that:

- CQI reporting of wideband average is required
- Resource blocks may be grouped for the purpose of CQI reporting (and in the analysis have assumed such grouping of 2 resource blocks per group)
- CQI reporting of the Best-M groups is required
- A reporting opportunity (e.g. Fast Feedback Channel) exists for each MS at some subframes

The latest revision of the SDD [2] now requires that both wideband and subband CQI's be sent and that a compression technique be used to reduce its overhead:

> *11.8.2.2.3.1 CQI feedback*
>
> *For CQI feedback, the mobile station measures the downlink pilot channel, computes the channel quality information (CQI), and reports the CQI on the uplink feedback channel. Both wideband CQI and subband CQI may be transmitted by a mobile station. Wideband CQI is the average CQI of a wide frequency band. In contrast, sub-band CQI is the CQI of a localized sub-band.*
>
> *11.9.1.1 Channel quality feedback*
>
> *Channel quality feedback provides information about channel conditions as seen by the MS. This information is used by the BS for link adaptation, resource allocation, power control etc. Channel quality measurement includes narrowband and wideband measurements. CQI feedback overhead reduction is supported through differential feedback or other compression techniques. Examples of CQI include Physical CINR, Effective CINR, band selection, etc. Channel sounding can also be used to measure uplink channel quality.*

In a previous contribution [3] we have introduced a Haar based compression as a basis for the transmission of the CQI values of best-M resource blocks or groups. This contribution expands the concept to the compression of the full band achieving additional savings and higher throughput. It is shown for single word MIMO but can be easily extended to multiple codewords should TGm decide to include them for SU-MIMO.

In the rest of this contribution we:

- Present best-M compression scheme (copied from [3])
- Present the full-band compression scheme
- Show its performance

CQI compression for 802.16m

Haar Compression for Best-M

Haar compression is based on the Haar wavelet transform that encodes an input stream in multiple steps in varying levels of details. A brief explanation of the Haar compression method can be found in [1] and in Appendix B.

Haar compression followed by quantization belongs to the class of lossy compression methods, and is recognized as an effective and low complexity compression/decompression means for processing 1-Dimensional or 2-Dimensional data. Here we present Haar for M values of 3, 4, 5 and 7.
For a system with N_{RG} resource groups and M values of main interest, the procedures of proposed Haar-based compression of CQI information are:

1. Pick the Best-M CQIs
2. Calculate the average CQI of the remaining blocks
3. Create a vector with the following format, where the Best-M CQI values are reported in the same order as their relevant blocks:

   For M=3 →
   $$\mathbf{y} = \lfloor CQI_1 \quad 0 \quad CQI_2 \quad 0 \quad CQI_3 \quad 0 \quad CQI_{Avg} \quad 0 \rfloor \qquad \textbf{(1)}$$
   For M=4 →
   $$\mathbf{y} = \lfloor CQI_1 \quad CQI_2 \quad CQI_3 \quad 0 \quad CQI_4 \quad 0 \quad CQI_{Avg} \quad 0 \rfloor \qquad \textbf{(2)}$$
   For M=5 →
   $$\mathbf{y} = \lfloor CQI_1 \quad CQI_2 \quad CQI_3 \quad CQI_4 \quad CQI_5 \quad 0 \quad CQI_{Avg} \quad 0 \rfloor \qquad \textbf{(3)}$$
   For M=7 →
   $$\mathbf{y} = \lfloor CQI_1 \quad CQI_2 \quad CQI_3 \quad CQI_4 \quad CQI_5 \quad CQI_6 \quad CQI_7 \quad CQI_{Avg} \rfloor \qquad \textbf{(4)}$$

   For M=5, the total number of CQI values involved in steps 1 and 2 is 6. On the other hand, since the length of the input vector for Haar compression has to be 2^m, the vector **y** has two zeros inserted at locations 6 and 8. As such, assuming 5 bit/CQI, the vector will contain 30 bits worth of information. Similarly for M=3, M=4 and M=7, the vector **y** contains 20, 25 and 40 bits of information, respectively.
4. Apply the Haar transform iteratively (see appendix) or as expressed in Equation 5.

   $$\mathbf{y}_3 = \mathbf{y}\mathbf{W}_8 \qquad (5)$$

   where $\mathbf{W}_8$ is the compression operator (see Appendix A).
5. As a result of the zero insertions in step 3, after the compression the last four, three and two elements of the compressed vector $\mathbf{y}_3$ become non-relevant for transmission and can be dropped without any loss of information, for M=3, M=4 and M=5 respectively.
6. Quantize and send the remaining elements of the vector.
7. Send some indication (e.g. a label) of the location of the Best-M.

Simulation results indicate that each element of the compressed vector has a statistical distribution that can be exploited to optimize the quantization process.

Table 1 shows an example of the predefined offset values and required number of quantization bits for each element of the compressed vector. Each element of the compressed vector is represented by a fixed offset value and a Q bit binary word ($0 \rightarrow 2^{Q}-1$). As shown in the table, a higher number of bits is only required for the first two elements of the vector that carry more information than the others. The remaining elements can be represented by a fewer number of bits.

Table 1 – Quantization example for different M values

Elements of the Vector	Elements of the Compressed Vector d	Offset Value set Value	Quantization Range	Number of Quantization bits per element	Total Number of Quantization bits
M=3	$y_3(8)$	3	3 → 16	4	N_{Haar}=11
	$y_3(7)$	-1	-1 → 2	3	
	$y_3(6)$	-1	-1 → 1	2	
	$y_3(5)$	0	0 → 2.5	2	
M=4	$y_3(8)$	3	3 → 19	4	N_{Haar}=15
	$y_3(7)$	1	1→ 5	3	
	$y_3(6)$	1	1 → 8	3	
	$y_3(5)$	0	0 → 2.5	3	
	$y_3(4)$	0	0 → 2	2	
M=5	$y_3(8)$	5	5 → 24	4	N_{Haar}=18
	$y_3(7)$	2	2→ 9	4	
	$y_3(6)$	-1	-1 → 1	3	
	$y_3(5)$	0	0 → 2.5	3	
	$y_3(4)$	-2	-2 → 2	2	
	$y_3(3)$	-2	-2 → 2	2	
M=7	$y_3(8)$	7	7 → 31	4	N_{Haar}=21
	$y_3(7)$	-1	-1→ 2	3	
	$y_3(6)$	-1	-1 → 1	3	
	$y_3(5)$	-1	-1 → 3	3	
	$y_3(4)$	-2	-2 → 2	2	
	$y_3(3)$	-2	-2 → 2	2	
	$y_3(2)$	-2	-2 → 2	2	
	$y_3(1)$	0	0 → 4	2	

Using the quantization guidelines above for best-M Haar we have computed examples of required number of bits for different methods. All methods (except Haar) assume fixed 5b per CQI value and therefore do not guarantee equal performance (we have that below)

Table 2 – Example for number of bits for various schemes

Scheme	Location Information	Other information	Signaling Cost (bits)	Number of bits for N_{RG}=25 and			
				M=3	M=4	M=5	M=7
Full Feedback	Inherent	All, quantized	$5N_{\text{RG}}$	125	125	125	125
Best-M Average	Transmitted	Average of best-M & average of rest	$5+\left\lceil \log_2 \left\lceil \frac{N_{\text{RG}}}{M} \right\rceil \right\rceil + 5$	22	24	26	29
Best-M Individual	Transmitted	Quantized value of all best-M and average of the rest	$5M+\left\lceil \log_2 \left\lceil \frac{N_{\text{RG}}}{M} \right\rceil \right\rceil + 5$	32	39	46	59
Haar Best-M Individual	Transmitted	Compressed best-M individual	$N_{\text{Haar}}+\left\lceil \log_2 \left\lceil \frac{N_{\text{RG}}}{M} \right\rceil \right\rceil$	23	29	34	40

Haar Compression for Full-Band Reporting

The general mechanism for Full-band CQI reporting can be summarized as follows: the MS computes CQI values and performs compression on the whole CQI vector. The compressed information is than sent incrementally according to the channel condition, MS speed and the requested granularity of the reporting by the BS. At the BS the received vector is incrementally decompressed using always the same decompression matrix, providing improving reconstructions of the whole of the CQI vector.

The size of the compression/decompression matrices is determined from the number of N_{RG} resource groups. For a system with e.g. N_{RG} =25 resource groups, the size of the compression/decompression matrices will be 32×32. The remaining 7 unused places in the input vector are filled by zeros. The zeros are needed to be spread across the vector to balance the weight of the vector.

Figure 2 shows incremental update of the full-band Haar compression/decompression process. As shown, in successive reporting opportunities the BS improves the quality of its information regarding the CQI values. Thus decompression with two coefficients yields only information about the average of the lower and upper bands but by taking more coefficients into consideration finer resolutions become available.

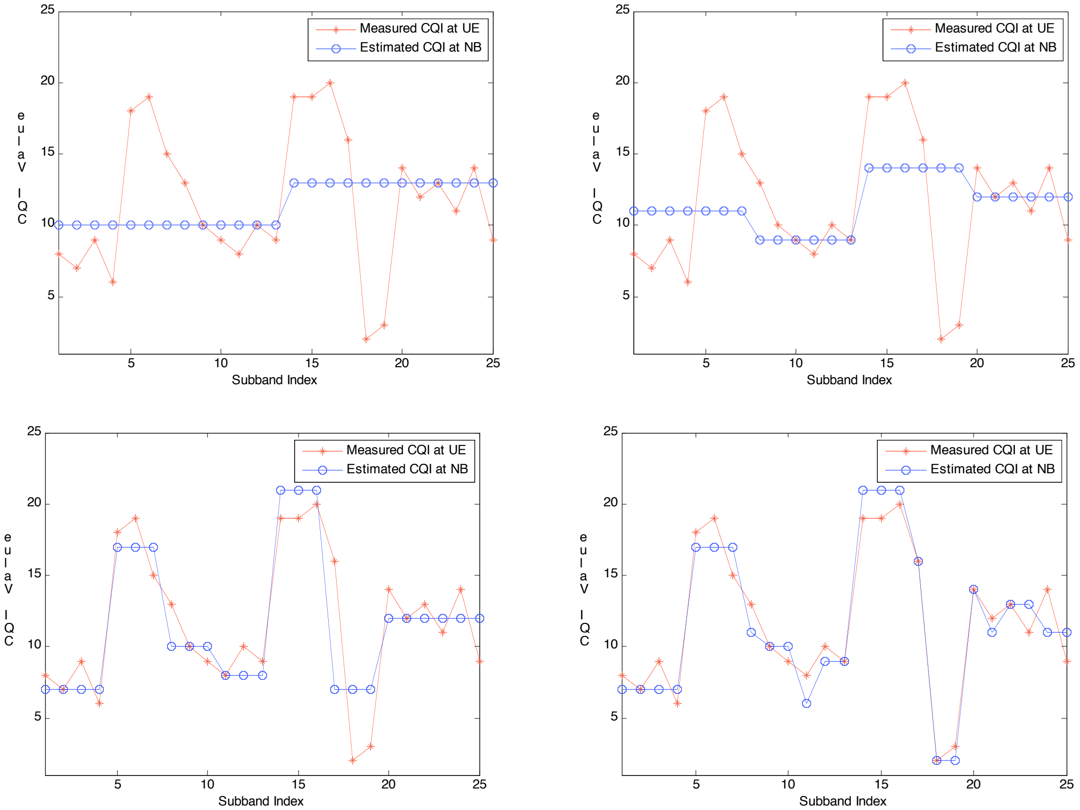


Figure 1 – Incrementally reconstructed CQI vector using Full-band Haar. Clockwise from top left: Using 2, 4, 8 and 16 coefficients

As location information isn't required for full band reporting, the BS does not need to receive the whole vector before it starts decompressing the information. This reduces reporting latency and improves tracking.

There are several benefits using the proposed full-band Haar CQI reporting that can be listed as:

- Compared to Best-M methods, there is a significant saving by not requiring sending the label and average information.
- Compared to schemes such as bitmapping, it provides significantly more accurate representation of the channel and consequently more reliable scheduling can be achieved.
- Gradual update is possible. In other words, it is not necessary to receive the whole set of coefficients at the BS to start updating the scheduler. The BS can update the scheduler per reception of each element. Thus, the update rate could be faster.
- By using incremental update, the system can be easily adapted to various MS conditions or a given CQI budget.
- Full-band Haar is low complexity. The whole matrix calculations rely only on basic shift and addition/subtraction operations. Also, a significant number of matrix elements are zero which contributes to more savings in computations.

Performance Evaluation

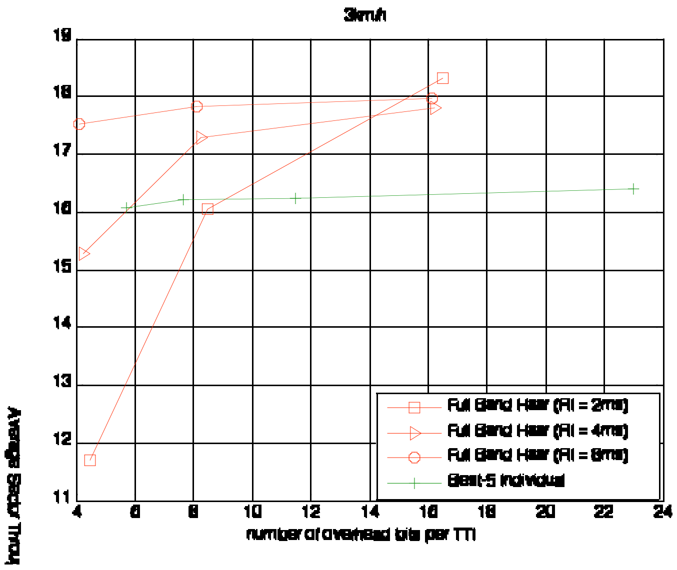
The average sector throughput performance of Best-5 Individual and Full-band Haar with 25 resource groups and varying number of coefficients sent is evaluated under equal overhead conditions and compared to Best-M individual. Figures … show their average sector throughputs performance of the system for UE speeds of 3km/h and 15km/h, respectively.

For the Best-5 individual we have assumed a total of 46b for 5 best CQI values, a label for their position and an average of the rest of the CQI values. These 46b where distributed between varying number of reporting opportunities thus creating a varying overhead in bits per reporting interval.

For the full-band Haar, for each overhead load we have used 3 different reporting intervals (RI=2, 4 and 8 reporting opportunities, a reporting opportunity occurs once per 2 sub-frames or 1ms) with different number of coefficients reported in each case. For each case we used 5b for the first coefficient and 4b for each of the rest. Note that further reduction is bit count is possible for higher coefficients which we haven't done. Table 3 below defines this reporting scheme.

Table 3 – Full Band Haar reporting scheme used in simulation

[redacted]	[redacted]	Length of reporting interval in reporting opportunities		
		2	**4**	**8**
# Coefficients reported per reporting interval	Total bit count & Quantization	CQI load bits / reporting opportunity		
2	9=1x4+5	4.5	-	-
4	17=3x4+5	8.5	4.25	-
8	33=7x4+5	16.5	8.25	4.125
16	65=15x4+5	-	16.25	8.125
32	129=31x4+5	-	-	16.125



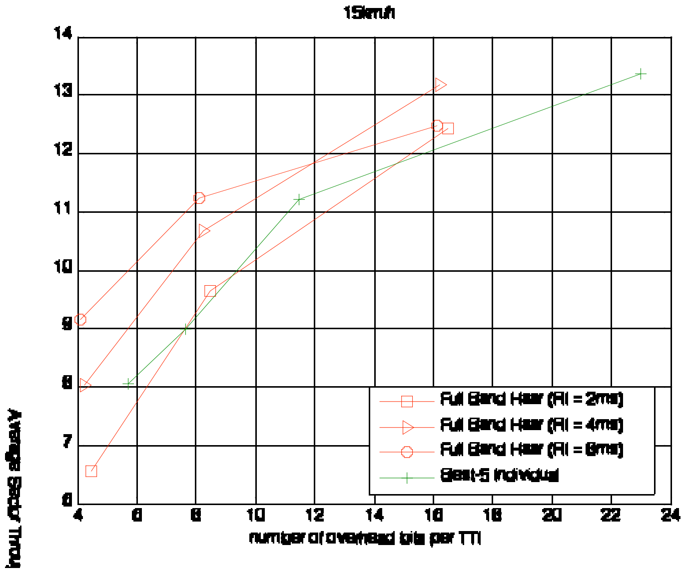


Figure 2 – Sector throughput for Best-5 individual and various full-band Haar strategies

As can be seen, the full band Haar compression outperforms Best-M individual per equal overhead loading. Moreover for low mobile speeds it achieved performance that cannot be achieved by Best-M alone at any loading.

Conclusions:

In this contribution we have presented 2 CQI compression schemes based on the Haar transform followed by an appropriate quantization. The full band Haar achieves a significant reduction in overhead per given performance or can be used to increase performance per given overhead load. This overhead reduction is achieved due to avoiding the necessity to send best-M location label and full-band average. Moreover the depicted results do not show further quantization optimization which can be made with full band Haar.

See appendix for text proposal.

References:

[1] Colm Mulcahy, “Image compression using the Haar wavelet transform”, Spelman Science and Mathematics Journal, Vol. 1, No 1, April 1997.
[2] 802.16m SDD 80216m-08_003r4
[3] “CQI Feedback for 802.16m MIMO Transmission”, C80216m-UL_ctrl-08_022

Appendices

Appendix A: Text Proposal

11.9.1.1 Channel quality feedback

Channel quality feedback provides information about channel conditions as seen by the MS. This information is used by the BS for link adaptation, resource allocation, power control etc. Channel quality measurement includes narrowband and wideband measurements. CQI feedback overhead reduction is supported through differential feedback or other compression techniques. Examples of CQI include Physical CINR, Effective CINR, band selection, etc. Channel sounding can also be used to measure uplink channel quality.

The BS may indicate to the MS:

- A set of sub-carriers for full band reporting of CQI (e.g. 10MHz) and their grouping into resource groups
- A reporting interval, defined in terms of the number of reporting messages (could be one or greater)
- The number of coefficients

The MS shall, for each reporting interval:

- Perform full-band Haar compression as specified in 11.9.1.1.1
- Quantize the coefficients. The exact quantization scheme is for FFS.
- Distribute the quantized coefficients between the reporting messages and transmit then to the BS. Details TBD.

The BS may extract the full CQI information after all reporting messages have been received. Alternatively the BS may extract inexact information when each message is received.

11.9.1.1.1 Haar Compression

Given a vector c of N_{RG} CQI values, pad it with zeros to the nearest power of 2, denote the padded vector c_1.

The compression is then mathematically defined by the right multiplication $c_1\mathbf{W}$ of the padded CQI vector with matrix W, where W is a square matrix the size of the padded CQI vector defined as below.

Editorial note: sizes 8 and 16 provided, size 16 with zeros omitted for clarity. Size 32 not provided – cannot be handled by Equation Editor

$$\mathbf{W}_8 = {}^1\!/_8\begin{bmatrix} 1 & 1 & 2 & 0 & 4 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & -4 & 0 & 0 & 0 \\ 1 & 1 & -2 & 0 & 0 & 4 & 0 & 0 \\ 1 & 1 & -2 & 0 & 0 & -4 & 0 & 0 \\ 1 & -1 & 0 & 2 & 0 & 0 & 4 & 0 \\ 1 & -1 & 0 & 2 & 0 & 0 & -4 & 0 \\ 1 & -1 & 0 & -2 & 0 & 0 & 0 & 4 \\ 1 & -1 & 0 & -2 & 0 & 0 & 0 & -4 \end{bmatrix}.$$

$$W_{16} = \frac{1}{16}\begin{bmatrix}
1 & 1 & 2 & & 4 & & & & 8 & & & & & & & \\
1 & 1 & 2 & & 4 & & & & -8 & & & & & & & \\
1 & 1 & 2 & & -4 & & & & & & & & & & & \\
1 & 1 & 2 & & -4 & & & & & & & & & & & \\
1 & 1 & -2 & & & & & & & & & & & & & \\
1 & 1 & -2 & & & & & & & & & & & & & \\
1 & 1 & -2 & & & & & & & & & & & & & \\
1 & 1 & -2 & & & & & & & & & & & & & \\
1 & -1 & & 2 & & & 4 & & & & & & & & & \\
1 & -1 & & 2 & & & 4 & & & & & & & & & \\
1 & -1 & & 2 & & & -4 & & & & & & & & & \\
1 & -1 & & 2 & & & -4 & & & & & & & & & \\
1 & -1 & & -2 & & & & 4 & & & & & & & 8 & \\
1 & -1 & & -2 & & & & 4 & & & & & & & -8 & \\
1 & -1 & & -2 & & & & -4 & & & & & & & & 8 \\
1 & -1 & & -2 & & & & -4 & & & & & & & & -8
\end{bmatrix}$$

Appendix B: Haar Compression

The main idea is to shift the weight and importance of the vector elements to the first element of the vector. The process can be explained by an example as follows. Let the input vector **y** be:

$$\mathbf{y} = \begin{bmatrix} 12 & 3 & 28 & 14 & 5 & 11 & 2 & 9 \end{bmatrix}. \tag{a-6}$$

Since the vector has 23 elements, the transformation takes 3 steps of sum and difference operations as follows: first, group the elements of the vector y in groups of 2's.

$$\mathbf{y} = \begin{bmatrix} \underbrace{12 \quad 3} & \underbrace{28 \quad 14} & \underbrace{5 \quad 11} & \underbrace{2 \quad 9} \end{bmatrix} \tag{a-7}$$

Find the sum and the difference terms for each group and divide the results by two. The results are now in a new vector $\mathbf{y}_1$. As shown in Figure a-1, the first four elements of the vector $\mathbf{y}_1$ are called "Approximate" and the last four elements are called "Detail" coefficients. Steps 2 and 3 are similar to step 1, with the only difference being that they apply only on the "Approximate" coefficients, while the "Detail" coefficients are maintained to the end. As shown in Figure a-1, the final compressed vector is comprised of one "Approximate" coefficient along with seven "Detail" coefficients.

Step 1:	y_1=	7.5	21	8	5.5	4.5	7	-3	-3.5
Step 2:	y_2=	14.25	6.75	-6.25	1.25	4.5	7	-3	-3.5
Step 3:	y_3=	10.5	3.75	-6.25	1.25	4.5	7	-3	-3.5

Figure a-1 - 3 step compression of the vector y, Note: Green cells are approximate coefficients, rest are detail coefficients.

In an abstract form, the successive averaging and differencing steps involved in the compression process can be mathematically expressed as

$$\mathbf{y}_3 = \mathbf{y}\mathbf{W}_8 \tag{a-8}$$

where

$$\mathbf{W} = \begin{bmatrix} \tfrac{1}{8} & \tfrac{1}{8} & \tfrac{1}{4} & 0 & \tfrac{1}{2} & 0 & 0 & 0 \\ \tfrac{1}{8} & \tfrac{1}{8} & \tfrac{1}{4} & 0 & -\tfrac{1}{2} & 0 & 0 & 0 \\ \tfrac{1}{8} & \tfrac{1}{8} & -\tfrac{1}{4} & 0 & 0 & \tfrac{1}{2} & 0 & 0 \\ \tfrac{1}{8} & \tfrac{1}{8} & -\tfrac{1}{4} & 0 & 0 & -\tfrac{1}{2} & 0 & 0 \\ \tfrac{1}{8} & -\tfrac{1}{8} & 0 & \tfrac{1}{4} & 0 & 0 & \tfrac{1}{2} & 0 \\ \tfrac{1}{8} & -\tfrac{1}{8} & 0 & \tfrac{1}{4} & 0 & 0 & -\tfrac{1}{2} & 0 \\ \tfrac{1}{8} & -\tfrac{1}{8} & 0 & -\tfrac{1}{4} & 0 & 0 & 0 & \tfrac{1}{2} \\ \tfrac{1}{8} & -\tfrac{1}{8} & 0 & -\tfrac{1}{4} & 0 & 0 & 0 & -\tfrac{1}{2} \end{bmatrix}. \tag{a-9}$$

Therefore, the decompression can be easily implemented by

$$\mathbf{y} = \mathbf{y}_3\mathbf{F} \tag{a-10}$$

$$\mathbf{F} = \mathbf{W}^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}. \qquad \text{(a-11)}$$

It is worth mentioning that due to the particular value of the coefficients for the compression and the decompression, all the required multiplications can be performed by simple shift-add functions to reduce the involved complexity of matrix multiplication.

Appendix C: Simulations Assumptions

A system-level simulation using a proportional fair scheduler was performed to evaluate the CQI reporting schemes in a 10 MHz system. In the downlink transmission RB grouping is assumed, where one LLRU contains 2 PRUs. In the simulation a CQI granularity of 20 MCS levels is used. The impact of CQI measurement delay and errors is considered. The simulation parameters are listed in Table B-1.

Table B-1 – Simulation parameters

Parameter	Assumption
Cellular Layout	Hexagonal grid, 19 cell sites, 3 sectors per site
Inter-site distance (ISD)	500m
Number of transmit antennas at Node B	1
Number of receive antennas	2
Distance-dependent path loss	$L=I + 37.6\log_{10}(.R)$, R in kilometers I=128.1 – 2GHz
Lognormal Shadowing	Similar to UMTS 30.03, B 1.41.4
Shadowing standard deviation	8 dB
Penetration Loss	20dB
Channel model	Typical Urban (TU)
Antenna pattern (horizontal) (For 3-sector cell sites with fixed antenna patterns)	$A(\theta) = -\min\left[12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_m\right]$ θ_{3dB} = 70 degrees, $A_m = 20$ dB
BS Antenna Gain plus cable loss	15 dBi
Carrier Frequency	2.0 GHz
System Bandwidth	10 MHz
RB bandwidth	196kHz
UE speeds of interest	3km/h, 15 km/h
Maximum Node B transmission power	35 dBm
UE Traffic Model	Full Buffer
Noise Figure	9dB
Thermal noise density	-174 dBm
Scheduler	Proportional Fair
HARQ	Asynchronous (Chase combining)
CQI measurement error	Gaussian zero-mean error model
CQI averaging window	1 frame (= 8 sub-frames) - FDD
CQI feedback delay	4 sub-frames (FDD)
Target BLER	10%