# QUADTREE-DECOMPOSED CORDLESS VIDEOPHONES

*J. Streit, L. Hanzo*

Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.
Tel: +44 1703 593 125, Fax: +44 1703 593 045
Email: jss@ecs.soton.ac.uk, lh@ecs.soton.ac.uk
http://rice@ecs.soton.ac.uk

## ABSTRACT

**Arbitrarily programmable, but fixed-rate quad-tree (QT) decomposed, parametrically enhanced videophone codecs using quarter common intermediate format (QCIF) video sequences are proposed as a direct replacement for mobile radio voice codecs in second generation systems, such as the Pan-European GSM, the American IS-54 and IS-95 as well as the Japanese systems. The proposed 11.36 kbps Codec 1 and the 11 kbps Codec 2 are embedded in the adaptively re-configurable wireless videophone Systems 1-4 featured in Table 3 and their video quality, bit rate, robustness and complexity issues are investigated. Coherent re-configurable 16 or 4-level pilot symbol assisted quadrature amplitude modulation (PSAQAM) is used and the system's robustness is improved by a combination of diversity and Automatic Repeat Request (ARQ) techniques. When using a bandwidth of 200 kHz, as in the Pan-European GSM mobile radio system, the number of videophone users supported varies between 3 and 16, while the minimum required channel Signal to Noise Ratio over Gaussian and Rayleigh channels is in excess of 6 and 8 dB, respectively. The salient system features are summarised in Table 3.**

## 1. WIRELESS VIDEO TELEPHONY

In recent years there has been an increasing demand for visual communications while on the move [1]. This treatise [2] is focussed on the design of wireless video telephone systems, suitable for the robust, fixed-rate Quad-Tree (QT) coded transmission of 176×144 pixel Quarter Common Intermediate Format (QCIF) sequences over conventional 2nd generation mobile radio links, such as the Pan-European GSM system, the American IS-54 and the IS-95 systems. A plethora of further video codecs have been proposed for various applications [3], but the most significant advances in the field are hallmarked by the MPEG4 initiative. The QT-coded video stream was source-matched twin-class BCH coded [5] (C1/C2) and transmitted using a re-configurable transceiver, which has a robust but less bandwidth efficient 4-level Quadrature Amplitude Modulation (4QAM) mode of operation or can halve its Bd-rate requirement, if better channel conditions are maintained by invoking the more bandwidth-efficient 4-bit/symbol 16QAM mode [6].

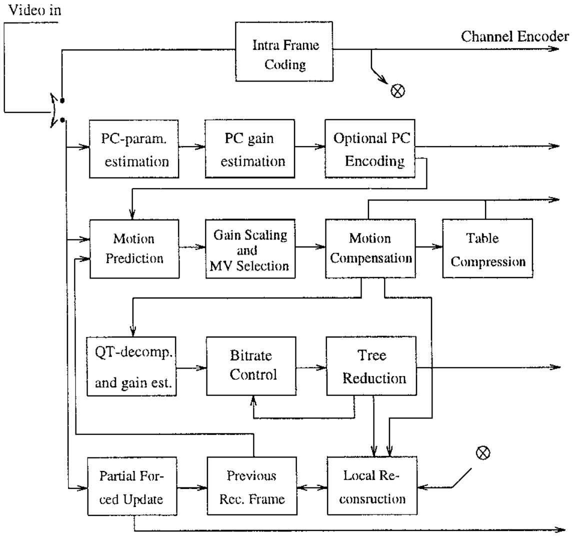

Figure 1: QT Codec Schematic

## 2. QUAD-TREE CODEC DESIGN

### 2.1. QT Codec Outline

The block diagram of the proposed QT codec is shown in Figure 1. For videophony over conventional mobile radio speech channels, such as the Pan-European GSM or the American IS-54 and IS-95 systems, fixed-rate video codecs are required. As it is seen at the top of the codec's schematic, the intra-frame coding mode is invoked at the commencement of communications, during which a low-resolution initial image is transmitted to the decoder in order assist in its start-up phase, as it will be described shortly. In the motion prediction (MP) block of Figure 1 the QCIF frame is first segmented into small, for example 8×8-pixel perfectly tiling blocks. Then each block is slid over a certain motion-velocity and frame-scanning rate dependent search area of the previous reconstructed frame and it is estimated by finding the position of highest correlation, which location each block was deemed to have originated from due to mo-

This treatise is supported by a demonstration package down-loadable from http://rice@ecs.soton.ac.uk portraying various video sequences scanned at 10 frames/s and encoded at fixed bitrates of 6.7, 8, 9.6, 11.4 and 13 kbps.

0-7803-3157-5/96 $5.00 © 1996 IEEE

tion translation. The corresponding coordinates referred to as motion vectors (MV) are then used in the motion compensation (MC) process of Figure 1 to appropriately position each incoming block, which are then subtracted from the previously reconstructed frame in order to generate the so-called motion compensated error residual (MCER). The MCER is QT-decomposed to the required resolution under the joint action of the Bitrate Control and Tree Reduction blocks, which will be highlighted at a later stage. In the reconstruction process the MVs assist at both the local and the distant decoder in an inverse fashion in order to appropriately update the reconstruction frame buffers. The eye and lip representation quality of the QT codec can be improved by the optional Parametric Coding (PC) arrangement of Figure 1.

Due to transmission errors the encoder's and decoder's previous reconstructed frame buffers may become misaligned, which leads to prolonged artifacts at the output of the decoder. This effect can be remedied using Partial Forced Updates (PFU) in order to identically replenish the encoder's and decoder's previous reconstructed frame buffers. In our proposed codecs we were constrained to a simple PFU technique using the down-scaled and partially overlaid 4-bit encoded 8×8-bit block averages, which was applied only to a fraction of the blocks in each frame due to the tight bit rate budget available. Although the PFU process partially overwrites the previous reconstructed buffers at both the local and remote decoder by the above mentioned very crude image estimate, therefore slightly impairing the codec's error-free performance, in case of high channel bit error rates (BER) it has an error mitigating effect by gradually replenishing both buffers' contents.

The number of 8×8-pixel PFU blocks per QCIF frame depends on the target bit rate and it is automatically determined by the proposed re-configurable codec. Consequently the frequency of partially forced updating a certain block is also bit rate dependent, although higher prevailing BER values would require more frequent updates, irrespective of the bit rate budget. In our 11.36 kbps QT videophone codec 20 randomly scattered 8×8 blocks out of the 396 blocks per frame are updated, which requires 80 bits/frame. This implies that the PFU frequency of each specific block is about 1/(2 s), which is equivalent to updating the same block about every 2 s or 20 frames. During PFU both the local and remote reconstructed buffers' contents are scaled by 0.7 and the 0.3-scaled 4-bit quantised block averages are superimposed, allowing a non-destructive gradual replenishment to take place.

**The motion compensation** (MC) scheme determines a 4-bit motion vector (MV) for each of the 8×8 blocks within a search window of 4 × 4 = 16 pixels using full-search. The potential gain of MC is assessed in terms of Motion Compensated Error Residual (MCER) energy reduction and the gains in the subjectively important eye and mouth region may be augmented by a factor of two. A bit-rate dependent number of 'motion-active' blocks is then subjected to full motion compensation, while for the 'motion-passive' blocks frame differencing is employed. Each of the 396 blocks would require a 9-bit identifier, leading to a total of $9 + 4 = 13$ bits per active vector. Typical motion activity rates around 60 MV would use most of the available bit rate budget. In order to accommodate around 60 active MVs within a budget of 500 bits, we assign a 1-bit motion-activity flag for each vector and hence create a MV activity table of 396 bits. This motion activity table can be compressed by about a factor of two using the run-length based 'Table Compression' algorithm of Figure 1, when aiming for a target bit rate budget of around 1000 bits/frame or 10 kbps. The final bit allocation scheme of our prototype codec will be revisited in Table 4 after unravelling further details of the QT codec of Figure 1.

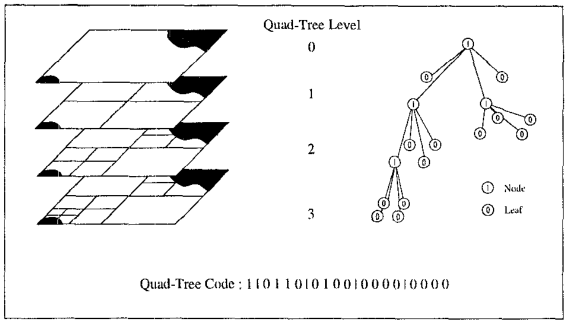

Figure 2: Regular decomposition example and the corresponding quadtree

## 2.2. Quadtree Decomposition

In Quad-Tree [4] (QT) coding the MCER is decomposed in variable size sectors characterized by similar features, for example by similar grey levels. Explicitly, the MCER residual frame portrayed in the centre of Figure 3 is described in terms of two sets of parameters, the structure or spatial distribution of similar regions and their grey levels.

An example of QT decomposition is portrayed in Figure 2 in a so-called top-down approach, where the MCER residual is initially decomposed into four sections, since the MCER is too inhomogeneous to satisfy the so-called similarity criterion. Then two of the quadrants become sufficiently homogeneous, while two of them need a number of further decomposition steps in order to fulfil the similarity criterion. This decomposition can be continued until the MCER is actually resolved into pixels, if a high resolution is required and the bit rate budget allows it.

Pursuing the so-called top-down QT decomposition approach of Figure 2 we are now ready to derive the variable-length QT-code given at the bottom of the Figure. The inhomogeneous QCIF MCER constitutes a so-called node in the QT structure of Figure 2, which upon decomposition into four smaller quadrants generates four further nodes that are subjected to a similarity test. Explicitly, if all pixels of a quadrant deviate from the mean $m$ by less than the similarity threshold $\sigma$, then they are deemed to be a homogeneous 'leaf node', which does not require further decomposition and similarity tests. Hence they are represented simply by the mean value $m$. In our example a four-level decomposition was used (0-3), but the number of levels and/or the similarity threshold $\sigma$ can be arbitrarily adjusted in order to achieve the required image quality and/or bit rate target. In contrast, if the pixels constituting the current node or quadrant deviate from the mean $m$ by more than the similarity threshold $\sigma$, it becomes a node instead of a leaf and it has to be further decomposed in order to achieve homogenity.

The derivation of the QT-code now becomes explicit from Figure 2, where each parent node is flagged with a binary one classifier, while the leaf nodes are denoted by a binary zero classifier and the flags are read from top to bottom and left to right. It can be inferred from Figure 2 that the location and size of 13 different blocks can be encoded using a total of 17 bits. A typical segmentation of a frame is exemplified in Figure 3, where the QT structure is portrayed with and without the overlaid MCER residual frame and the original video frame. In the eye and lip regions a more stringent similarity match was required than in the background, which led to a finer QT decomposition. Another design alternative is to use a more coarse QT decomposition and employ a more sophisticated, higher resolution QT block description, which will be addressed in the next Section.

### 2.3. Quadtree Block Intensity Match

As mentioned above, in case of coarse QT decomposition, or if high image quality requirements must be met, the simple block average representation of the QT decomposed blocks can be substituted by more sophisticated techniques, such as vector quantisation (VQ), discrete cosine transformed (DCT) or subband (SB) decomposed representations. Naturally, increasing the number of hierarchical levels in the QT decomposition leads to blocks of different sizes and applying VQ or DCT to blocks of different sizes increases the codec's complexity. Hence the employment of these schemes in case of more than 2-3 hierarchy levels becomes impractical.

In order to explore the range of design trade-offs we first studied the performance of a low-complexity zero-order mean- or average-intensity model, which was then compared to a first-oder luminance intensity profile, stretching over the block, which corresponds to a luminance plane sloping in both $x$ and $y$ directions.

#### *2.3.1. Zero-order Block Intensity Match*

In order to design the appropriate mean intensity quantisers, we evaluated the PDF of the average values $m$ of the variable sized MCER blocks portrayed in Figure 4. In accordance with our expectations, for various block sizes quantisers having different mean values and variances were required, since the mean of the MCER blocks towards the top of the QT, namely at QT Levels 2-4, which cover a large area is more likely to be close to zero than the mean of smaller blocks, which exhibited itself in a more highly peaked and hence less spread PDF. Observe that for clarity of visualisation the $x$ and $y$ scales of the Level 5-7 PDFs are about an order of magnitude augmented. The mean of the smallest blocks at QT Levels 5-7 tends to fluctuate over a wide range, yielding a near-uniform PDF for Level 7. Table 1 summarizes the intervals over which the block means fluctuate as the block size is varied. Therefore, it was necessary to design separate quantisers for each QT hierarchy level.

In order to achieve the best possible performance, a two-stage quantiser training procedure was contrived. First the unquantised mean values for each QT-decomposed block size were recorded using a training sequence in order to derive an initial set of quantisers. During the second, true training stage this initial set of quantisers was then used tentatively in the QT codec's operation in order to record future unquantised block averages generated by the codec operated at a rate of 1000 bits/frame, which was the average rate for the 2nd generation mobile videophone systems as well as for the adaptive multimode treminals of the near future. This two-stage approach was necessary to obtain realistic training data for the Max-Lloyd training of the final quantisers.

We then generated codebooks for a range of different number of reconstruction levels and endeavoured to determine the best distribution of coding bits between the QT-code and block intensity encoding. Figure 5 characterises the codec's performance for various quantisers ranging from two to 64 level schemes, while generating 1000 bits/frame. Observe in the Figure that the two-level and four-level quantisers were found to have the best performance, since the adaptive bit rate control algorithm restricted the number of hierarchy levels in the QT decomposition process, when more bits were allocated to quantise the block averages. Therefore it was disadvantageous in terms of both objective and subjective quality to finely resolve the block luminance intensity at the cost of reducing the QT resolution. A more efficient exploitation of the higher number of QT block description bits may be in this case to have a more elaborate QT block intensity model, such as the first-order model or the previously mentioned DCT-, VQ- or SBC-based QT block representations, if a higher complexity is acceptable.

#### *2.3.2. First-order Intensity Match*

In order to verify this hypothesis we then embarked on studying the performance of the $n^{th}$-order luminance-intensity surface defined by Equation 1:

$$b(x,y) = a_0 + a_{x1}x + a_{y1}y + \ldots + a_{xn}x^n + a_{yn}y^n \qquad (1)$$

where, in order to describe the luminance $b(x,y)$ of the pixel at coordinates $x$ and $y$, the coefficients $a_{xn}$ and $a_{yn}$ must be known. In order to explore the potential of this technique, we used a low-complexity linear approach and fixed $n$ to 1, which leads to Equation 2:

$$b(x,y) = a_0 + a_{x1}x + a_{y1}y. \qquad (2)$$

In case of $n = 1$ a block's luminance is approximated by a plane sloping in both $x$ and $y$ directions. The squared error of this linear approximation is given by Equation 3. The constants $a_0$,$a_{x1}$ and $a_{y1}$ are determined by setting the partial derivative of Equation 3:

$$e = \sum_{x=1}^{b}\sum_{y=1}^{b}(b(x,y) - (a_0 + a_{x1}x + a_{y1}y))^2 \qquad (3)$$

with respect to all three variables to zero and solving the resulting three dimensional problem.

In order to assess the performance of this scheme Max-Lloyd quantisers were designed for each of the constants and the PSNR performance of a range of quantisers using different number of quantisation levels was tested. We found that the increased number of quantisation bits required for the three different coefficient quantisers was too high to facilitate a sufficiently fine QT-based MCER decomposition. Hence the PSNR performance of the first order scheme became inferior to that of the zero-order model under our specific bit rate constraints. Table 2 reveals that the PSNR

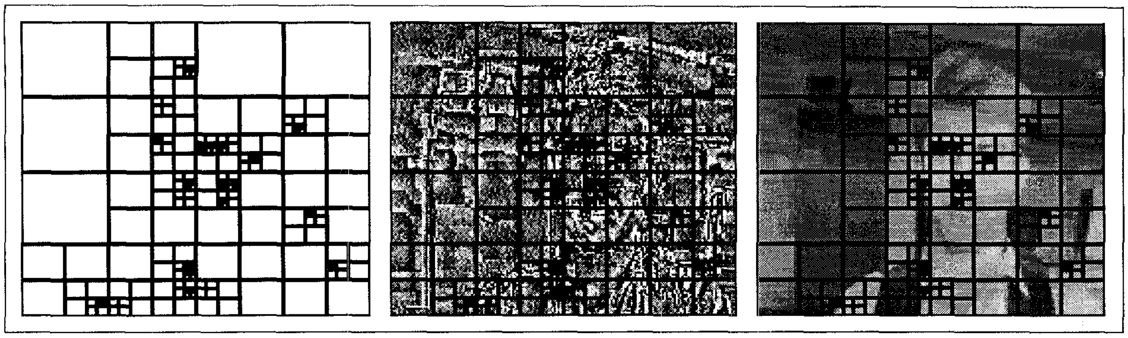

Figure 3: Quad-tree segmentation example with and without overlaid MCER residual and original video frame

| Level | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Block Size | 176× 144 | 88× 72 | 44× 36 | 22 × 18 | 11× 9 | 5/6 × 4/5 | 2/3 × 2/3 |
| Max. Range +/- | 0.22/0.36 | 0.57/0.79 | 2.89/2.49 | 8.48/7.74 | 20.59/20.69 | 57.59/50.39 | 97.26/86.62 |

Table 1: Maximum Quantiser Ranges at Various Hierarchy Levels

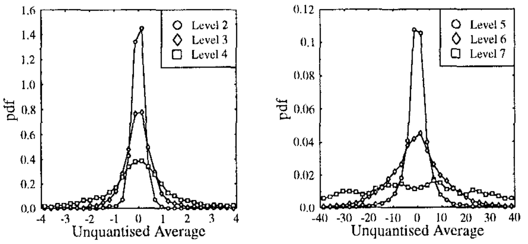

Figure 4: Probability Density function of the block averages at Levels 2 to 7

| Approximation | 8 kb/s | 10 kb/s | 12 kb/s |
|---|---|---|---|
| Zero-order PSNR (dB) | 27.10 | 28.14 | 28.46 |
| First-order PSNR (dB) | 25.92 | 27.65 | 27.82 |

Table 2: PSNR Performance comparison for zero- and first-order QT models at various constant bit rates

performance of the first-order model is at least 0.5 dB worse than that of the simple zero-order model, hence we opted for the zero-order model.

### 2.4. QT Decomposition and Bit Allocation Protocol

Having assessed the potential of a number of different approaches to contriving an appropriate fixed, but programmable bit allocation scheme we finally arrived at Algorithm 1, which allowed us to eliminate the specific leaves from the QT that resulted in the lowest decomposition gains, while providing a powerful cost-gain quantised bit rate control protocol.

The achievable PSNR performance of the proposed QT codec at various bit rates is characterised by Figure 6 in case of the Miss America sequence and a higher activity clip, which we referred to as the 'Lab Sequence'. Two QT codecs, Codec 1 and 2, having rates of 11.36 and 11 kbps

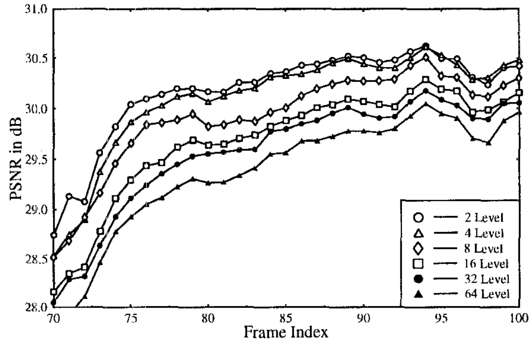

Figure 5: Performance for various quantisers at a constant bit rate

at 10 frames/sec were then designed and embedded in Systems 1-4 of Table 3 in order to assess their overall performances. The bit allocation scheme of Codec 1 is summarised in Table 4.

## 3. SYSTEM PERFORMANCE AND CONCLUSIONS

Four re-configurable QT-based wireless videophone systems characterised by Table 3 have been contrived. The proposed video codec can replace the speech codec of conventional second generation wireless systems, such as for example the GSM, IS-54 and IS-95 systems. The video PSNR versus channel SNR performance of Systems 1-4 was evaluated over both Additive White Gaussian Noise (AWGN) and Rayleigh channels and the minimum channel SNR values required were also summarised in the Table.

| Feature | System 1 | System 2 | System 3 | System 4 |
|---|---|---|---|---|
| Video Codec | Codec 1 | Codec 2 | Codec 1 | Codec 2 |
| Video rate (kbps) | 11.36 | 11 | 11.36 | 11 |
| Frame Rate (fr/s) | 10 | 10 | 10 | 10 |
| C1 FEC<br>C2 FEC<br>Header FEC | BCH(127,71,9)<br>BCH(127,71,9)<br>BCH(127,50,13) | BCH(127,50,13)<br>BCH(127,50,13)<br>BCH(127,50,13) | BCH(127,71,9)<br>BCH(127,71,9)<br>BCH(127,50,13) | BCH(127,50,13)<br>BCH(127,50,13)<br>BCH(127,50,13) |
| FEC-coded Rate (kbps) | 20.32 | 27.94 | 20.32 | 27.94 |
| Modem | 4/16-QAM | 4/16-QAM | 4/16-QAM | 4/16-QAM |
| ARQ | No | No | Yes | Yes |
| User Signal. Rate (kBd) | 18 or 9 | 12.21 or 24.75 | 18 or 9 | 12.21 or 24.75 |
| System Signal. Rate (kBd) | 144 | 144 | 144 | 144 |
| System Bandwidth (kHz) | 200 | 200 | 200 | 200 |
| No. of Users | 8 or 16 | 5 or 11 | 6 or 14 | 3 or 9 |
| Eff. User Bandwidth (kHz) | 25 or 12.5 | 40 or 18.2 | 33.3 or 14.3 | 66.7 or 22.2 |
| Min. AWGN SNR (dB) 4/16QAM | 7.5/13 | 7.5/12 | 6/12 | 6/11 |
| Min. Rayleigh SNR (dB) 4/16QAM | 20/20 | 15/18 | 8/14 | 8/14 |

Table 3: Summary of System Features

**Algorithm 1** *This algorithm adaptively adjusts the required QT resolution, the number of QT description bits and the number of encoding bits required in order to arrive at the target bit rate.*

```
1. Develop the full tree from minimum to maximum number of QT levels (eg 2-7).
2. Determine the MSE gains associated with all decomposition steps for the full QT.
3. Determine the average decomposition gain over the full set of leaves.
4. If the potentially required number of coding bits is more than twice the target number of bits for the frame, then delete all leaves having less than average gains and repeat Step 3.
5. Otherwise delete leaves on an individual basis, starting with the lowest gain leaf, until the required number of bits is attained.
```

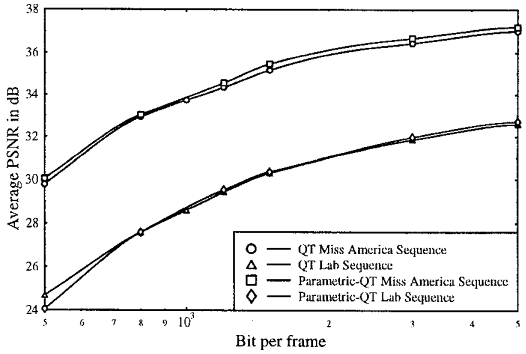

Figure 6: PSNR versus bit rate performance of the proposed QT codec

| Parameter | MC | PFU | QT | Total |
|---|---|---|---|---|
| No. of bits | < 500 | 80 | 566 | 1136 |

Table 4: Bit allocation for the 11.36 kbps Codec 1

## 4. ACKNOWLEDGEMENT

The financial support of the EPSRC, UK in the framework of the contract GR/J46845 is gratefully acknowledged.

## 5. REFERENCES

[1] Advanced Communications Technologies and Services (ACTS), Workplan, DGXIII-B-RA946043-WP, Commission of the European Community, Brussels, 1994

[2] **J. Streit, L. Hanzo**: Quadtree-based reconfigurable cordless videophone systems, to appear in IEEE Tr. on Video Techn., 1995

[3] Special Issue on Image Sequence Compression, IEEE Tr. on Image processing, Sept. 1994, Vol. 3, No. 5, Guest Editors: B. Girod et al

[4] **Eli Shustermann and Meir Feder**: Image Compression via Improved Quadtree Decomposition Algorithms, IEEE Transactions on Image Processing, Vol 3, No 2, March 1994, pp 207-215

[5] **K.H.H. Wong, L. Hanzo**: Channel Coding, pp 347-489, Chapter 4 in R. Steele (Ed.) Mobile Radio Communications,IEEE Press-Pentech Press, London, 1992

[6] **W.T. Webb, L. Hanzo**: Modern quadrature amplitude modulation: Principles and applications for fixed and wireless channels, IEEE Press-Pentech Press, 1994, ISBN 0-7273-1701-6, p 557