VLSI systems for image compression. A power-consumption/image-resolution trade-off approach

Javier Bracamonte,* Michael Ansorge and Fausto Pellandini

Institute of Microtechnology, University of Neuchâtel
Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland

ABSTRACT

Low power consumption is a requirement for any battery powered portable equipment. When designing ASICs for image and video compression, emphasis has been placed mainly on building circuits that are fast enough to satisfy the high data throughput associated with image and video processing. The imminent development of portable systems featuring full multimedia applications, adds the low-power constraint to the design of VLSI circuits for this kind of applications. Several techniques as lowering the supply voltage, architectural parallelization, pipelining etc., have been proposed in the literature to achieve low-power consumption. In this paper we report a VLSI circuit featuring a power management user-controllable technique that trades image quality for power consumption in a transform-based algorithm.

1. INTRODUCTION

For data compression purposes, lowering the high frequency content of an image has proved being a successful method for reducing the bit rate (number of bits per pixel). In zonal coding [1] for example, a unitary transform, such as the Discrete Cosine Transform DCT, is applied on 2-D blocks of pixels of the original image. By selecting only a few of the coefficients on a predefined zone on the frequency domain, one can trade reconstructed image quality for bit rate. A similar principle is used on the JPEG baseline algorithm,[2] where the 2-D transform operation is followed by a quantization of the coefficients. Those coefficients that are visually significant are quantized with a relatively short quantization step, while those that are less important are coarsely quantized.

In the methods mentioned above, and in general in any image compression scheme, the effort is aimed at keeping the distortion of the reconstructed image as low as possible for a given bit rate. In certain applications however, the best image quality might not be always required (e.g., image browsing, surveillance). For these cases, several trade-off techniques have been proposed in the literature, to reduce the bit rate, while keeping the quality of the reconstructed images acceptable for the application. In this paper this trade-off is reformulated in terms of image resolution and power consumption for portable equipment applications. That is, when the full image quality is not required, the user can trade image resolution for power consumption.

Power management techniques represent a key strategy in the design of portable equipment.[3] Power management allows the reconfiguration of a system in order to save power, while still operating in a consistent manner with the application. The basic principle is to power down a whole system (or parts of it), or stop the clock to a module, when it is idle for a significant period of time.

This paper is organized as follows: Section 2 describes a power management strategy and its application on a well-known transform-based algorithm. The VLSI implementation of the adapted algorithm is discussed in section 3. The results are reported in section 4, and finally the conclusions are given in section 5.
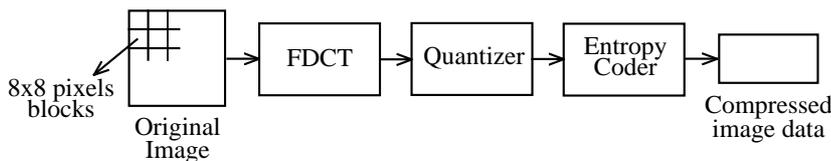
2. POWER MANAGEMENT STRATEGY

Figure 1: JPEG baseline algorithm

Transform-based image coding algorithms have received a good acceptance in the definition of image compression standards. For example standards JPEG, MPEG-1, MPEG-2 and H.261, are all DCT-based. In these algorithms, the DCT is used to map blocks of 8x8 pixels into another block of 8x8 coefficients in the frequency domain. A 64 elements normalization matrix defines the quantization steps by which the DCT coefficients will be quantized. Due to the energy packing efficiency of the DCT and to the coarse value of the normalization matrix on the high frequency region, many of the coefficients will become zero after quantization. A long sequence of zero values can then be effectively compressed by run-length coding, followed by entropy coding to improve the final compression ratio. The block diagram of the JPEG baseline algorithm is shown in Figure 1.

A power management (PM) strategy has been studied for its application in the realization of a single chip VLSI circuit that performs the sequential JPEG coding algorithm. Under this PM scheme, the circuit features four different modes of operation as shown in Figure 2: a) Mode DC, b) Mode 4, c) Mode 16 and d) full JPEG mode. Each dot in Figure 2 represents a coefficient in the frequency domain. With the top-left point representing the DC coefficient.
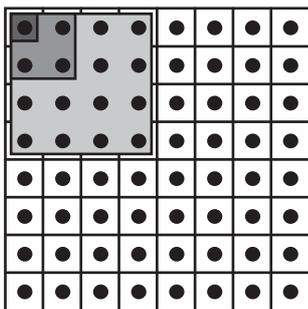

Figure 2: PM modes

In Mode DC, only the DC coefficient is selected and coded. In this mode a rough blocky version of the image is obtained. Mode 4, commands the encoder to select and code only the 4 lower frequency coefficients. In this mode, the image is intelligible and a particular image could be easily recognized. However the image still presents blocking effects especially at the edges of objects. In Mode 16, the encoder selects and codes only the 16 lower frequency coefficients. In this mode the images are relatively of a good quality. The blocking effects on the edges of objects is strongly attenuated, with respect to Mode 4. The full JPEG mode, allows the circuit to give full spectral resolution.

When the full image quality is not required to be transmitted, the user can select the mode of operation by means of a knob on the encoder. As soon as the number of coefficients corresponding to a mode of operation has been computed, the control unit powers down the DCT unit and the quantization unit. A multiplexer allows to set the values of all the non-computed coefficients to zero, in order to generate a fully JPEG-compliant bit-stream. This scheme saves power by avoiding unnecessary computation. It can be interpreted as a hardware adaptive zonal coding embedded in the baseline JPEG algorithm. Since in JPEG, the DCT is the most expensive operation in terms of computational power, the gain can be very significant. In full JPEG mode the circuit works as a regular JPEG coder system.

It is worth noting that in a baseline JPEG encoder, when the best image quality is not required, one can scale up the normalization matrix to improve the compression ratio. This bit rate reduction is achieved by the increase of the number of 2-D DCT coefficients that become zero after quantization. In a regular baseline JPEG coder circuit with no PM strategy, that would imply a waste of power, since the DCT unit must always compute the
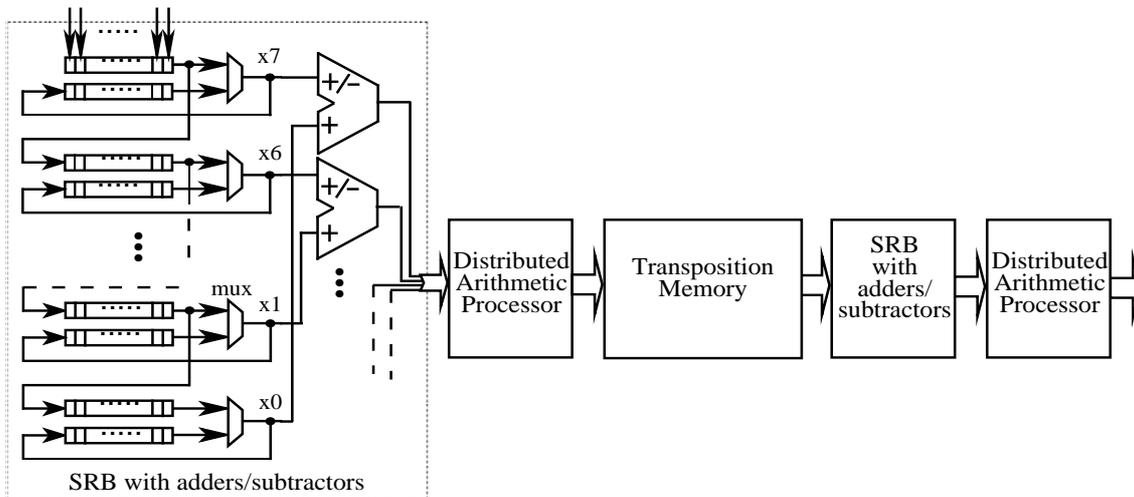
Figure 3: Architecture of the 2-D DCT

64 2-D DCT coefficients regardless of the scaling factor of the quantization matrix.

Besides of the baseline sequential mode, the JPEG standard defines other modes of operation[4] (e.g., progressive and hierarchical) which are not addressed in this paper and that represent potential power-saving schemes, especially for the decoder. It is important to note however, that the VLSI implementation of these JPEG modes comes at the expense of additional power-consuming hardware and increased system complexity with respect to the baseline mode.

3.   VLSI IMPLEMENTATION

The PM strategy described in the preceding section can be applied to trade image quality for power consumption both at the JPEG coder and decoder. Applications that use the PM scheme on the encoder or the decoder are independent, and no information overhead is required to indicate the mode in which the images are coded. From a VLSI architectural point of view, the solution for both cases is very similar, and thus, we will limit the discussion to describe the VLSI circuit of the encoder.

The forward 2-D DCT circuit on Figure 1 is implemented as a sequence of two 1-D DCTs. The first 1-D DCT operating over the rows of an input of 8x8 pixels and the second operating over the columns of the intermediate result. This reduces the hardware complexity compared to the case of a straightforward 2-D DCT implementation.

The VLSI architecture of the 2-D DCT is shown in Figure 3. Each 1-D DCT is executed by a Distributed Arithmetic Processor (DAP).[5] The 1-D transform of an 8 elements input vector is obtained by applying this vector eight consecutive times to the DAP. The Shift Register Bank (SRB) provides the storage and the sequencing required to this purpose, and for each iteration the DAP uses a different set of coefficients (addresses a different look-up-table). The quantizer in Figure 1 is implemented with a serial-parallel multiplier. The bit-serial input being the output of the 2-D DCT and its parallel input being the output of a ROM containing the inverse of the normalization coefficients. Details regarding the VLSI implementation of the different modules of Figure 1 are given on reference 6.

The basic module of the PM architecture is shown in Figure 4. The multiplexer at the output is required to keep the bit-stream uninterrupted when the unit is powered off. It sets to zero the value of all the non-computed

coefficients. One or several large pMOS transistors are used to power the unit on or off.
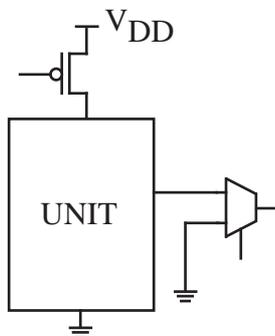

Figure 4: PM Unit

The PM strategy is dynamic[7] in the sense that only certain modules of the encoder (or decoder) can be powered off. Referring to Figure 3, due to the storage function and inherent power-dependent condition, the RAM that executes the matrix transposition must remain powered on independently of the PM mode. The same is true for the Huffman coder that must release a complete decodable bit-stream. On the other hand, in Mode 4 for example, 60 of the 2-D DCT coefficients are not going to be evaluated. This leaves place for powering off, both DAPs and the quantization unit. In Mode 4, the first DAP is powered off during 48 word cycles (active only during 16 word cycles). The second DAP is powered off during 60 word cycles (active only during 4 word cycles). The quantization unit can be equally powered off during 60 word cycles. A word length is equal to the number of bits required to represent a 2-D DCT coefficient, which in our circuit is 12 bits. Thus, one word cycle is equal to 12 clock cycles.

4. RESULTS

Figure 5: Original image

To illustrate the image quality obtained with the different modes of operation, the image Lena (256x256 pixels, 8 bits/pixel) shown in Figure 5 was JPEG-coded in each of the four PM modes. Figure 6 shows the images after they have been JPEG-decoded.

When the JPEG coder circuit is operating on full JPEG mode, it executes 128 scalar products per block of 8x8 pixels (64 scalar product each DAP of the 2-D DCT) and 64 multiplications corresponding to the 64 normalization of the DCT coefficients. For a typical image of 256x256 pixels, that means a total of 131,072 scalar products and 65,536 multiplications.

When the coder circuit is set in Mode DC, for the same image size, the number of operations is reduced to 9,216 scalar products and 1,024 multiplication. That means that the 2-D DCT unit is powered off about 93% of the time. Figure 6(a) shows the quality of an image that has been coded in Mode DC.

In Mode 4, The most computational intensive part of the JPEG circuit is powered off 84% of the time. The images present a blocky effect at the edges of the objects, but they are easily recognizable. In this mode the peak signal-to-noise ratio (PSNR)[†] of the JPEG-decoded image Lena, shown in Figure 6(b), is 25.91 dB.

In Mode 16 the most computational intensive part of the JPEG circuit is powered on only 38% of the time. The decoded images are relatively of a good quality. In this mode, the PSNR of the JPEG-decoded image Lena, shown in Figure 6(c), is 30.30 dB.

In full JPEG mode, no power-saving is made and from an algorithmic point of view, the circuit works as a regular JPEG system. For comparison, the PSNR of Lena in full JPEG mode is 32.70 dB, the decoded image is shown in Figure 6(d).

[†]If $f(i,j)$ and $\hat{f}(i,j)$ represent the pixels of the original and the coded-decoded NxN pixels image respectively, then the $PSNR$ is given by: $PSNR = 20\log_{10}(\frac{255}{RMSE})$, where $RMSE = \sqrt{\frac{1}{N\cdot N}\sum_{i=1}^{N}\sum_{j=1}^{N}[f(i,j) - \hat{f}(i,j)]^2}$.
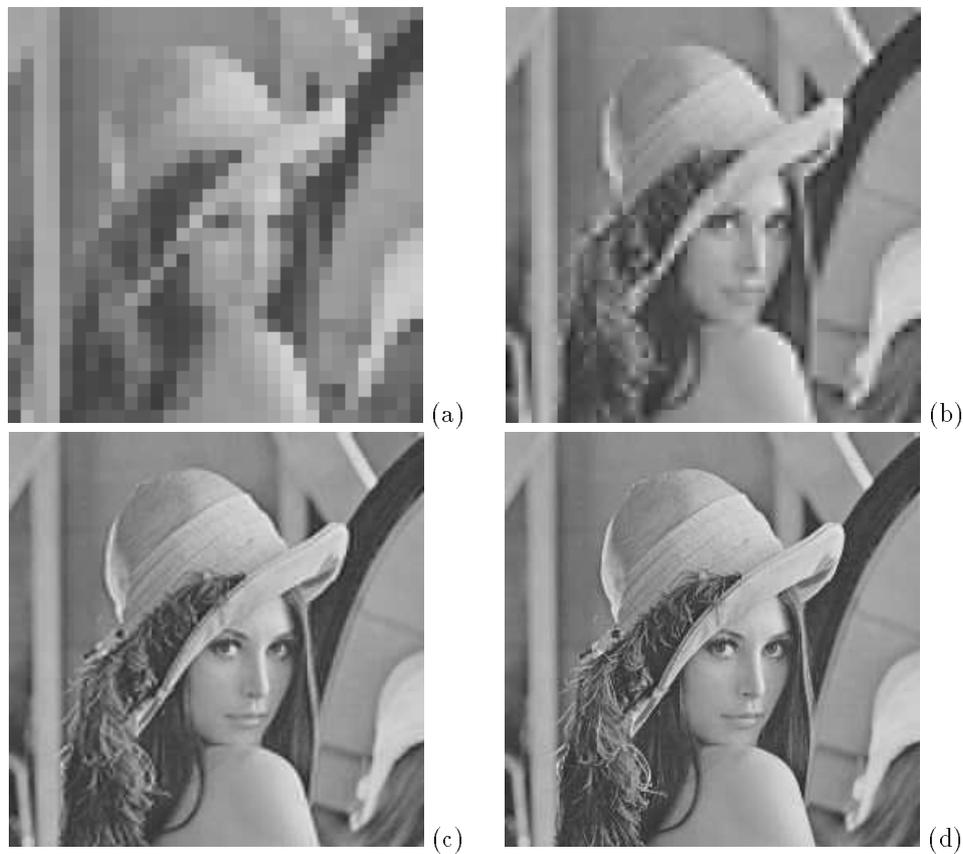
Figure 6: Image quality for the different modes of operation. (a) Mode DC, (b) Mode 4, (c) Mode 16, and (d) Full JPEG mode.

The layout of the JPEG coder circuit featuring the four PM modes is shown in Figure 7. The area of the chip is 4.6 x 3.1 mm^2 $\approx$ 14.5 mm^2. It was implemented in the 1.2μm 5V CMOS CMN12 process from VLSI Technology Inc. On the left part of the layout are located the SRBs, the quantizer and the two DAPs. All these circuits were built from a full-custom library, a fact that is reflected by the very small size of the modules. The RAM for executing the matrix transposition is located at the bottom-left corner of the circuit. The Huffman coder was build entirely with standard cells. At a clock frequency of 36 MHz this circuit is able to process 25 CIF (Common Intermediate Format: 352 x 288 pixels) images per second. Thus, as a regular JPEG chip (full JPEG mode) it is suitable for motion JPEG (MJPEG) or for the non-recursive path of the H.261 low-bit rate video coder.
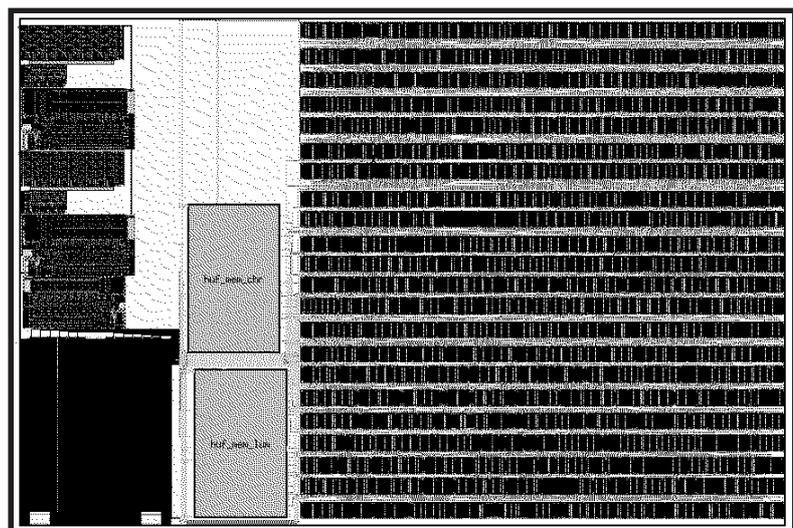

Figure 7: Layout of the JPEG coder circuit with 4 PM modes

5. CONCLUSIONS

Power management techniques play an important role to reduce power consumption in portable equipment. They are currently being used intensively in the design of the latest microprocessors and computer systems. In this paper a PM strategy that allows trading image resolution for power consumption was described. The power-saving technique makes the trade-off at the algorithm level, and thus an important reduction of power consumption is achieved. Depending on the power down mode of operation, the most computationally intensive part of a JPEG coder in powered off up to 93% of the time. An area-efficient single chip VLSI circuit implementing JPEG coding featuring four power management modes was also described.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] M. Rabbani and P.W. Jones, *Digital Image Compression Techniques*. Vol. TT 7, SPIE Optical Engineering Press, Bellingham, WA, USA, 1991.

[2] W.B. Pennebaker and J.L. Mitchell, *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York, USA, 1993.

[3] E. Harris, S. Depp, W. Pence, S. Kirkpatrick, M. Sri-Jayantha and R. Troutman, "Technology directions for portable computers", *Proc. IEEE*, Vol. 83, No. 4, pp. 636-658, April 1995.

[4] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards. Algorithms and Architectures*. Kluwer Academic Publishers, Boston, MA, USA, 1995.

[5] U. Sjöström, *"On the design and implementation of DSP algorithms: An approach using wave digital state-space filters and distributed arithmetic"*. Ph.D Thesis, University of Neuchâtel, Switzerland, 1993.

[6] J. Bracamonte, M. Ansorge and F. Pellandini, "Design methodology for VLSI implementation of image and video coding algorithms", *Fourth Bayona Workshop on Intelligent Methods in Signal Processing and Communications*, Bayona-Vigo, Spain, June 24-26, 1996.

[7] A. Bellaouar and M.I. Elmasry, *Low-Power Digital VLSI Design. Circuits and Systems*. Kluwer Academic Press, Boston, MA, USA, 1995.