Abstract—Providing embedded systems with adaptation capabilities is an increasing importance objective in design community. This work deals with the implementation of adaptive compression schemes in FPGA devices by means of a bio-inspired algorithm. A simplified version of an Evolution Strategy using fixed point arithmetic is proposed. Specifically, a simpler than the standard (hardware friendly) mutation operator is designed, modelled and validated using a high-level language. HW/SW partitioning issues are considered and code profiling accomplished to validate the proposal. Preliminary results of the proposed hardware architecture are also shown.

I. INTRODUCTION

JPEG2000, in contrast with the Discrete Cosine Transform (DCT) used in JPEG, is based on the Discrete Wavelet Transform (DWT) [1]. It is a very useful tool for (adaptive) image compression algorithms, since it provides a transform framework that can be adapted to the type of images being handled. This feature allows to improve the performance of the transform according to each particular type of image so that improved compression (in terms of size) can be achieved, depending on the wavelet used.

One of the current design challenges in embedded systems engineering is the implementation of adaptation capabilities. With previous compression algorithms this was not possible (from the point of view of the transform module in the algorithm), but, DWT opens up a possibility for this task to be tackled. Having a system able to adapt its compression performance according to the type of images being handled, may help in, for example, the calibration of image processing systems. Depending on where the system is deployed, certain tunings to the transform coefficients may help in increasing the quality of the transform, and, subsequently, the quality of the compression.

The accepted standard wavelet considered to be the state of the art in compression of photographic images is the hand-designed D9/7 Cohen-Daubechies-Feauveau (9/7-CDF or also D9/7). Therefore, though it has a better coding performance compared to JPEG, this will hold just for the type of images the wavelet is adapted to.

This work deals with the implementation of adaptive wavelet transforms in FPGA devices. An Evolutionary Algorithm (EA), specifically an Evolution Strategy (ES) [2] was chosen as the bio-inspired search algorithm.

Since the intended deployment platform is an FPGA device, a relatively low computing power will be available, what will, undoubtedly, affect the performance of the evolutionary search. However, this will provide the system with adaptation capabilities.

It has to be noted that most previous works on adaptivity in wavelet transforms, deal with adapting the transform on the fly to the local properties of the signal, what implies an extra computational effort to detect its singularities. This classical meaning of adaptive lifting does not apply in this work. However, it refers to the adaptivity of the system as a whole. Therefore, adaptation becomes prior to system operation instead of real time adaptation to the signal at hand.

The structure of the paper is as follows. Sections II and III give an overview of both, DWT and ES. A brief analysis of previously reported works is given in Section IV. Afterwards, the proposed architecture is presented in Section VI, and a simplified ES directed towards an FPGA implementation in Section V. The experimental setup developed and a further insight into the details of the implementation, along with the results obtained can be found in Section VII. Paper is concluded in Section VIII.

II. OVERVIEW OF THE DISCRETE WAVELET TRANSFORM

The DWT is a multiresolution analysis (MRA) tool widely used in signal processing due to its joint time-frequency signal analysis characteristics, that concentrates the signal energy into fewer coefficients to increase the degree of compression when the data is encoded. For a general introduction to wavelet based MRA analysis see [3].

The Lifting scheme (LS), introduced by Sweldens [4], reduces the computational cost of the transform as required by the Fast Wavelet Transform (FWT) algorithm and facilitates the construction of custom wavelets for very specific and
III. BIO-INSPIRED OPTIMIZATION WITH EVOLUTION STRATEGIES

An Evolution Strategy (ES) is one of the fundamental algorithms among Evolutionary Algorithms (EA) that utilize a population of candidate solutions and bio-inspired operators to search for a target solution. ES is primarily used for optimization of real-valued vectors. The algorithm operators are iteratively applied within a loop, where each loop run is called a generation (g), until a termination criterion is met. Variation is accomplished by the so-called mutation operator. For real-valued search spaces, mutation is normally performed by adding a normally (Gaussian) distributed random value to each component under variation (i.e., to each parameter encoded in the individuals). Algorithm 1 shows a pseudo-code description of a typical ES.

The canonical versions of the ES are denoted by \((\mu/\rho, \lambda)\)-ES and \((\mu/\rho + \lambda)\)-ES, where \(\mu\) denotes the number of parents (parent population, \(P_0\)), \(\rho \leq \mu\) the mixing number (i.e., the number of parents involved in the procreation of an offspring), and \(\lambda\) the number of offspring (offspring population, \(P_{\lambda}\)). The parents are deterministically selected from the set of either the offspring, referred to as comma-selection \((\mu < \lambda)\), or both the parents and offspring, referred to as plus-selection. Selection is based on the ranking of the individuals’ fitness \((F)\) taking the \(\mu\) best individuals. Once selected, \(\rho\) out of the \(\mu\) parents (\(R\)) are recombined to produce an offspring individual \((y_1)\) using intermediate recombination, where the parameters of the selected parents are averaged, or randomly chosen if discrete recombination is used. Each ES individual \(a := (y,s)\) comprises the object parameter vector \(y\) to be optimized and a set of strategy parameters \(s\), which coevolve (and are therefore being adapted themselves) with the solution. This is a particular feature of ES called self-adaptation. For a general description of the \((\mu/\rho + \lambda)\)-ES see [2].

Algorithm 1 \((\mu/\rho + \lambda)\)-ES

1: \(g \leftarrow 0\)
2: Initialize \(P_0^{(g)} \leftarrow \{(y_m,s_m), m = 1, \ldots, \mu\}\)
3: Evaluate \(P_0^{(g)}\)
4: while not termination condition do
5: \hspace{3em} for all \(l \in \lambda\) do
6: \hspace{6em} \(R \leftarrow\) Draw \(\rho\) parents from \(P_{\lambda}^{(g)}\)
7: \hspace{6em} \(r_1 \leftarrow\) recombine \((R)\)
8: \hspace{6em} \(y_1, s_1 \leftarrow\) mutate \((r_1)\)
9: \hspace{6em} \(F_1 \leftarrow\) evaluate \((y_1)\)
10: \hspace{3em} end for
11: \(P_{\lambda}^{(g)} \leftarrow\) \{(\(y_l,s_l\), \(l = 1, \ldots, \lambda\)\}
12: \(P_{\mu}^{(g+1)} \leftarrow\) selection \(P_{\lambda}^{(g)}, P_{\mu}^{(g)}, \mu, +\)
13: \(g \leftarrow g + 1\)
14: end while

IV. PREVIOUS WORK ON EVOLUTIONARY WAVELETS DESIGN

This work is a continuation of [6], which uses the original idea of combining the lifting technique with EA for designing wavelets proposed by Grasemann and Mikkulainen [7]. Their original contributions are two: the use of a Genetic Algorithm (GA) to encode wavelets as a sequence of lifting steps and the proposal of an idealized version of a transform coder to save time in the complex evaluation method used (which involved computing a number of times the Peak Signal to Noise Ratio (PSNR) for one individual combined with other individuals from each of one of the parallel subpopulations): they propose using only a certain percentage of the largest coefficients (which involves a previous ordering stage) for reconstruction. The GA used had parallel evolving populations (coevolutionary GA). The evaluation consisted of 80 runs, each of which took approximately 45 minutes on a 3 GHz Xeon processor. The results obtained in this work outperformed the considered state-of-the-art wavelet for fingerprint image compression, the D9/7 wavelet used by the FBI, in 0.75 dB. The type of images used to adapt the wavelet to is the set of 80 images from the FVC2000 fingerprint verification competition [8].

Works reported by Babb, Moore et. al. can be considered the current state of the art in the use of EC for image transform design [9], [10]. After using a GA algorithm in their previous works, the authors finally propose the use...
of a ES, outperforming their previous results, but keep on encoding wavelets as filters for the FWT, instead of the LS. The milestones followed in their research are summarized on the next list.

1) Evolve the inverse transform for digital photographs under conditions subject to quantization.
2) Evolve matched forward and inverse transform pairs.
3) Evolve coefficients for three and four level MRA transforms.
4) Evolve a different set of coefficients for each level of MRA transforms.

The algorithms reported are highly computationally intensive (the training runs were done using the Arctic Region Supercomputer Center (ARSC), Fairbanks, Alaska). Although the work by Grasemann and Miikkulainen was done on an accessible computer, both training times and computing resources needed in both cases, show the complexity of the algorithms developed. These approaches are highly unfeasible for an implementation in an FPGA.

V. PROPOSED SIMPLIFIED ES FOR FPGA IMPLEMENTATION

This work proposes using an ES as the search algorithm encoding the individuals (wavelets) using the LS. This is a combination of the original proposals analysed in Section IV. However, a standard ES is highly computationally intensive for an FPGA implementation, as these analysed works show. Therefore, the whole evolutionary process has to be down-calculated in complexity. In the decisions made to simplify the algorithm are presented. They are summarized here:

1) Single evolving population opposed to the parallel populations of the coevolutionary genetic algorithm reported in [7].
2) Use of uncorrelated mutations with one step size [2] instead of the overcomplex method reported in [9], [10].
3) Evolution of one single set of coefficients for all MRA levels.
4) Ideal evaluation of the transform. Since doing a complete compression would turn to be an unsustainable amount of computing time, the simplified evaluation method reported in [7] was further improved. For this work, all wavelet coefficients \( d_j \) are zeroed, keeping only the trend level \( s_j \) of the transform from the last iteration of the algorithm. For 2 levels of decomposition, this is an idealized 16:1 compression ratio.

These simplifications yielded very positive results, but, since there were still some complex operations around in the algorithm, the complexity relaxation before doing the hardware implementation was taken even further, so that a trade-off between performance and size is observed. They are summarized below:

1) Uniform random distribution. Instead of using Gaussian distributions for the mutation of the object parameters, a Uniform distribution was tested, for being simpler to implement in HW.

2) MAE as evaluation figure. PSNR is the figure of merit more widely used for image processing tasks. But, as previous works show for image filter design problems [11], using MAE gives almost identical results, because the interest is in relative comparisons among population members.

VI. ARCHITECTURE MAPPING. HW/SW PARTITIONING

Typical implementations of evolutionary optimization engines in FPGAs place the EA in an embedded processor. With this approach, some degree of performance is sacrificed to gain flexibility in the system (needed to fine tuning the algorithm), so that modifications may be easily done to the (software) implementation of the EA (which is, of course, much easier than changing its hardware counterpart). Table I shows the partitioning resulting from applying this design philosophy. According to Algorithm 1, each of the EA operators are shown in the table together with further actions to be accomplished: recombination (of the selected parents), mutation (of the recombinant individuals to build up a new offspring population), evaluation (of each offspring individual) and selection (of the new parent population).

When a new offspring population is ready, each of its individuals is sequentially sent to the hardware module responsible of its evaluation. This comprises the computation of the fitness as the Mean Absolute Error (MAE) as shown in (1), where \( R, C \) are the rows and columns of the image and \( I, K \) the original and transformed images respectively. To tackle it, the following sequence of operations has to be done: Forward Wavelet Transform (fWT), Compression (C), Inverse Wavelet Transform (iWT) (wavelet transform) and MAE figure computation (fitness computation). Once each offspring individual has been evaluated, the population is sorted according to the result (sorting population). At this stage, the microprocessor may close the evolutionary loop creating the new parent population. Afterwards, recombination and mutation are applied and a new offspring population will be available to be evaluated.

\[
MAE = \frac{1}{RC} \sum_{i=0}^{R-1} \sum_{j=0}^{C-1} |I(i,j) - K(i,j)|
\]  

Fig. 2 shows the proposed conceptual architecture capable of hosting such a system. The functions to be implemented in hardware work as attached peripherals to the microprocessor embedded in the FPGA.
Since the LS was proposed, several hardware implementations have been reported both for ASICs and FPGAs (JPEG2000 adopted LS). This means that good results centred on exploiting LS features to obtain fast implementations have already been done. But the objectives at this stage of the work are just directed to prove and validate the concepts and the feasibility of the system as a whole. Therefore, the implementation of the Wavelet Transform is a direct, algorithmic mapping of the LS to its hardware equivalent VHDL description (i.e., no hardware optimizations at the level of data dependencies are accomplished).

Taking advantage of the LS features, the fWT and iWT can be computed by just doing a sign flip and a reversal in the order of operations (P and U stages), so both modules are sharing hardware resources in the FPGA. Compression block is simple, since it only needs to substitute the fWT result by zeros for each datum of the details bands. Therefore, it is working in parallel to the fWT. In a similar manner, the Fitness module computes the difference image as each pixel is produced by the iWT.

The fWT/iWT module is build up by applying the sequence of P, U stages dictated by the LS. To mimic the high level modelling of the algorithm (see Section VII), 6 stages have been implemented (3 P and 3 U), each one containing 4 filter coefficients, as seen in Fig. 3, which is enough to implement the most common wavelets utilized at present. Section VII shows the first preliminary results of the implementation.

The Block RAM modules (BRAM) embedded in the FPGA are used as data memory for the wavelet transform module. During evolution, it hosts the training image so that the highest memory bandwidth possible is achieved. It has been over-dimensioned to host up to 4 256x256 pixels (8 bpp) images to speed up the test phase. Therefore, when evolution has finished, this extra memory can be used to load the test images from system memory. In this phase, one of the four sub-banks is used for the actual image being tested, and the other three are loaded with the following test images in the meanwhile, acting as a multi ping-pong memory.

VII. EXPERIMENTAL SETUP AND RESULTS

Before accomplishing the hardware implementation, modelling and simulation of the algorithm was done with Python computing language [12], together with its numerical and scientific extensions, NumPy and Scipy [13], as well as the plotting library, Matplotlib [14].

As shown on [15], [16], for 8 bits per pixel (bpp) integer inputs from an image, a fixed point fractional format of Q2.10 for the lifting coefficients and a bit length in between 10 and 13 bits for a 2 to 5 level MRA transform for the partial results is enough to keep a rate-distortion performance almost equal to what is achieved with floating point arithmetic. Therefore, fixed-point binary arithmetic (16 bits fractional part) was modelled with integer types, defining the required quantization/dequantization and bit-alignment routines to mimic hardware behaviour. Table II gathers all the information related to the proposed ES, where each $P_i$, $U_i$ is made up of 4 coefficients and $k_i$ are single scaling coefficients. D9/7 wavelet is defined by $(P_1, U_1, P_2, U_2, k_1, k_2)$.

To sum up, for the various tests run, the set of parameters used correspond to a $(10/5, 70)$-ES with a varying initial $\sigma = \{0.1, \ldots, 1.5\}$.

The experiments reported in this work have also used, as in [7], the first set of 80 images of the FVC2000 fingerprint verification competition. Images were black and white, sized 300x300 pixels at 500 dpi resolution. One random image was used for training and the other 79 for testing purposes.

A. Algorithm optimization results

All the results obtained are compared with the D9/7 transform implemented in fixed point arithmetic and evaluated with
Fig. 4. Result of a typical evolution run

Fig. 5. Tests of the best evolved wavelet. The best individual (and D9/7 for comparison) is exercised for each of the 79 images of the test set

the proposed method. Although evolution used MAE as the fitness function, for comparison purposes, the results shown in this Section are given as the standard PSNR. As a reference, for the fixed point D9/7 wavelet the average PSNR measured (with the proposed evaluation method) for the 79 test images is 28.74 dB.

Fig. 4 shows how the algorithm behaves during a typical run. As it can be seen, around generation 100 the performance of the evolved wavelet is already similar to the D9/7. In Fig. 5 the best evolved wavelet is compared with the D9/7 for the whole image test set, showing how the algorithm was able to evolve a solution yielding 29.80 dB, that outperforms the standard D9/7 by an average of 1.06 dB. Results correspond to $\sigma = 1.0$ and a randomly seeded population. Fig. 6 shows a comparison of the performance of the best evolved wavelet against a fixed point implementation of D9/7 for a test image, where some artefacts can be noticed in the D9/7 result. Besides the error images show how the error introduced by the best evolved wavelet is lower, indicating a more accurate reconstruction of the original image.

B. HW/SW partitioning validation

The model developed to validate the algorithm has been profiled. Table III shows profiling results for 500 generations for each EA operator. Table I is repeated (for clarity) adding extra columns with the result values. Absolute values are not of real interest (although NumPy routines are highly optimized, a C implementation would be faster), since what is being checked is the relative amount of time spent in each phase so that design partitioning is validated as a whole. As expected, most of the time is consumed evaluating the individuals. In each generation, 20.479 ms ($= 1433.56/(500$ generations * 70 individuals + 2 transforms)) are needed to compute a single wavelet transform (forward or inverse). Obtained results validate the design partitioning proposed except for the selection operator, which is low enough to be implemented in SW. The reason to choose a HW implementation for it, is that it can be applied as results are produced by the fitness computation module, saving extra time. In contrast, the simulation of the Python model runs on a single processor thread. Therefore, all operators are applied sequentially. But in the hardware implementation, some operators can be easily applied in parallel. For this reason, and depending on the scope of the system (see Section VIII), some other operator will probably benefit from being implemented in hardware, as, for example, the mutation. Besides, the subset of the C-language used to program the PowerPC processor in the FPGA, impose restrictions that will probably make that the percentage of the time each operator takes to compute increases.

C. Preliminary Hardware results

The prototype platform selected is an ML507 development board, which contains a Xilinx Virtex 5 XC5VFX70T FPGA device with an embedded PowerPC processor, responsible of running the ES. Table IV shows the preliminary implementation results for an over-dimensioned datapath of 32 bits, using 16 bits for the fractional part representation. This implementation is directed towards a system level functional validation in the FPGA, yielding higher area results than expected for the final system.

The current hardware, non-optimized, implementation yields one result each clock cycle. For a 256x256 pixels image, with a clock frequency of 100 MHz, the computation time of a wavelet transform is approximately 0.65 ms. This is a speed-up factor of around (20.5/0.65) 31 times, what would turn into 45 seconds to do all the transforms required by a population of 70 individuals during 500 generations.
Fig. 6. Transform performance. (a) is the original fingerprint image; (b) shows the result of applying the D9/7 Fixed Point implementation wavelet and (c) the equivalent for the best evolved individual; (d) shows the difference image for the D9/7 result and (e) for the best evolved wavelet.

**TABLE IV**

<table>
<thead>
<tr>
<th>Module</th>
<th>Resources</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>fWT/iWT</td>
<td>3651 / 11200 (12%)</td>
<td>181 MHz</td>
</tr>
<tr>
<td>Fitness function</td>
<td>25 / 11200 (&lt;1%)</td>
<td>330 MHz</td>
</tr>
<tr>
<td>Population sorting</td>
<td>984 / 11,200 (8%)</td>
<td>265 MHz</td>
</tr>
</tbody>
</table>

**REFERENCES**