



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Circuits, Systems, and Signal Processing

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa49021>

Paper:

Torquato, M. & Fernandes, M. (2019). High-Performance Parallel Implementation of Genetic Algorithm on FPGA.
Circuits, Systems, and Signal Processing
<http://dx.doi.org/10.1007/s00034-019-01037-w>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

High-Performance Parallel Implementation of Genetic Algorithm on FPGA

Matheus F. Torquato · Marcelo A. C. Fernandes

Received: date / Accepted: date

Abstract Genetic Algorithms (GAs) are used to solve search and optimization problems in which an optimal solution can be found using an iterative process with probabilistic and non-deterministic transitions. However, depending on the problem's nature, the time required to find a solution can be high in sequential machines due to the computational complexity of genetic algorithms. This work proposes a full parallel implementation of a genetic algorithm on field-programmable gate array (FPGA). Optimization of the system's processing time is the main goal of this project. Results associated with the processing time and area occupancy (on FPGA) for various population sizes are analyzed. Studies concerning the accuracy of the GA response for the optimization of two variables functions were also evaluated for the hardware implementation. However, the high-performance implementation proposes in this paper is able to work with more variable from some adjustments on hardware architecture. The results showed that the GA full parallel implementation achieved throughput about 16 millions of generations per second and speedups between 17 and 170000 associated with several works proposed in the literature.

Keywords Parallel implementation · FPGA · Genetic algorithms · Reconfigurable computing

Matheus F. Torquato
College of Engineering, Swansea University, Swansea, SA2 8PP, Wales, UK
ORCID: 0000-0001-6356-3538
E-mail: m.f.torquato@swansea.ac.uk

Marcelo A. C. Fernandes
Department of Computer Engineering and Automation, Federal University of Rio Grande do Norte (UFRN), 59078-970, Natal, RN, Brazil
Tel.: +55-84-3215-3771
ORCID: 0000-0001-7536-2506
E-mail: mfernandes@dca.ufrn.br

1 Introduction

In the last years, the increasing number of critical applications involving real time systems in conjunction with the growth of integrated circuits density and the continuous reduction in the power supply voltages transformed the development of new suitable computational solutions an even harder task to achieve. Due to the intense demand in the electronics goods market for high processing speeds at smaller time frames, without neglecting the energy savings, the technology industry has faced an extremely competitive and challenging scenario in terms of designing hardware solutions to meet this constantly growing demand. One way found by researchers and developers to address such demands is by using algorithm parallelization techniques. Parallel processing is used to manipulate data concurrently, so that while computing one section of the algorithm, other stations perform similar operations on another set of data [26]. Combining the hardware implementation with the parallelization of algorithms is often a satisfactory solution for high performance and higher speed applications when compared to sequential solutions.

The Field Programmable Gate Arrays (FPGAs) are reconfigurable hardware devices suited to this scenario due to the nature of its architecture. Given that FPGAs are huge configurable gates, they can be programmed to operate as multiple parallel paths in hardware. In this way, there is a real parallelization in which the running operations do not need to compete for the same resources since each one will be executed by different gates [12]. The increasing density and price reduction of FPGAs expand the opportunities for developers and researchers to use higher density FPGA devices for hardware implementations [14] considering the use of such devices is advantageous since the development time and costs are significantly reduced [29].

The convergence among genetic algorithms, parallelization techniques and reconfigurable hardware implementation results in this work which presents a proposal of parallel implementation of a genetic algorithm on FPGA. This paper focuses on high-performance and critical applications that require nanoseconds time constraints to be satisfied. On the other hand, in applications where processing speed is not the critical factor or it is less limiting than the necessity for low power consumption, it is possible to decrease the energy utilization by reducing the clock cycles rate, considering that the dynamic power utilization is diminished when an operating frequency lower than the maximum theoretical one is used [30]. Applications that process a large flow of data can be benefited and accelerated by this implementation here developed. Some applications examples are: data mining, tactile internet, massive data processing and bioinformatics.

1.1 Related Work

Genetic algorithms and Artificial Intelligence (AI) have long been used in applications of the most diverse areas to optimize and find satisfactory solutions

in computing, engineering and other fields. More recently, a wider range of applications and variations of genetic algorithms such as parallel and distributed applications, hardware implementations, new proposals for genetic operators, and hybrid (software and hardware) implementations of genetic algorithms have been observed within the research scenario.

In [6, 7], it is proposed an implementation of a customizable Intellectual Property Core (IP Core) for FPGA that implemented a general-purpose genetic algorithm. In this work, the authors have focused on the genetic algorithm programmability implemented in the IP core. The customization could be done regarding population size, generation number, crossover and mutation rates, random number generators seeds and the fitness function. One of the work's highlights is the support for a multiple of these functions. The proposed IP can be programmed with up to eight fitness functions which could be synthesized in conjunction with the GA and implemented in the same FPGA device. The proposed core also has additional input/output ports that allow the user to add further fitness functions that have been implemented on a second FPGA device or some other external device. The implementation utilized 13% of the available logical cells of a Xilinx Virtex II Pro (xc2vp30-7ff896). However, since a trade-off between performance and flexibility exists, and once the authors focused on flexibility over performance, the speedup over analogous software implementation was only of $\times 5.16$.

Hardware Genetic Algorithms implementations can also be observed in [23, 31, 33, 17, 22] and [32]. The work detailed in [33] showed the OIMGA which its strategy was to retain only the ideal individual from the population making the memory requirements drastically reduced. The paper [23] presented a compact implementation of a genetic algorithm on FPGA that represented the population of chromosomes as a vector of probabilities. The work focused on the lower consumption of memory, power and space resources in hardware, but it was not fully implemented on FPGA as it used a software written in C++ to compute the values from the fitness function. The work [31] proposed a high-speed GA implementation on FPGA. The implementation was based on the HGA proposed by [27], the first known GA implementation on FPGA, and the authors claimed that the developed system surpassed any existing or proposed solution according to their experiments. The P-HGAv1, version developed by [31] of the HGA claimed to be parametric, have low silicon requirements and support multiple fitness functions. Although the authors have focused on the speed of the algorithm and reached a time of 0.021 milliseconds for each GA generation, this speed may not be compatible with real-time applications that require low latency. In [17, 22] and [32], it is developed a modular implementation of a GA for FPGA. However, the implementation does not use the full-parallel strategy and spend several clocks for each generation. Basically, in each generation it is necessary to read and write individuals from/to a memory bank, in other words, this work spends more than N clocks for N individuals per generation.

The works presented by [19], [13] and [25] showed applications for ground mobile robots using GAs, where these first two were embedded implementations on FPGA. The work [19] developed, according to the authors, the

first GA-based hardware implementation of a simultaneous localization and mapping (SLAM) system for ground robots. The authors achieved significant hardware acceleration compared to software implementation by exploiting the pipelining and parallelization capabilities of reconfigurable hardware. In this project the GA's genes that made up the population represented possible robot movements based on the previous position. Later, In the work developed by [13], the goal was to determine the optimal movements considering various aspects such as route tracking and low energy consumption, avoiding obstacles collision. The authors pointed out that the implementation was suitable for real-time use and stated that all GA stages have been implemented in hardware modules. The solution presented in this work offered a convergence time of less than 2 milliseconds, it used 17124 out of the 17600 (97%) Lookup tables (LUTs) available in the FPGA, but the frequency obtained after the synthesis process was not informed. [25] developed a genetic algorithm with a coevolutionary strategy for global trajectory planning of several mobile robots. According to [15], co-evolution is the process of mutual adaptation of two or more populations simultaneously, and it was used to reflect the fact that all species are simultaneously co-evolving in a given physical environment. The implementation of [25] promised an improvement in the genetic operators of conventional GAs and proposed a new operator of genetic modification, but these developments were not implemented in hardware.

The implementations seen in [18], [28] and [4] were GA applications in digital signal processing and control systems embedded on FPGA. The work [18] presented a real-time GA for adaptive filtering application with all modules implemented in hardware such as fitness function, selection, crossover, mutation and random number generator functions. The implementation was designed in hardware and after its synthesis, a rate of 320 thousands generations per second was achieved. Meanwhile, [28] proposed a GA for multi-carrier dynamic systems based on filter banks. The authors of [4] proposed a design and an implementation of a PID controller based on GA and FPGA. The researchers stated that the design method of the intelligent PID controller based on FPGA and GA was successfully verified and had some advantages such as flexible design, automatic online tuning, high reliability, short technical development cycle and high execution speed. For this case, each GA chromosome was coded with the controller set of gains K_p , K_i and K_d . Details of FPGA area occupancy and obtained throughput were not reported.

The works [9] and [16] presented parallel and distributed implementation of GA using FPGAs. The paper [9] proposed a solution for parallel genetic algorithms in multiple FPGAs. Using multiple populations in parallel GAs was based on the idea that population isolation can maintain greater genetic diversity, while communication between them can cause GAs to work together to find good solutions. The implementation of [9] was applied to three different benchmarks, including the traveling salesman problem, and the authors stated from the experimental results that in a configuration of 4 FPGAs an average acceleration of 30 times over a multi-core processor GA was achieved. [16] introduced GRATER, an automated design workflow for FPGA accel-

ators that leveraged imprecise computation to increase data-level parallelism and achieved higher computational throughput. In this work the main idea was establishing a negotiation among circuit that involved area, energy and performance in exchange for precision reduction. This was achieved through an imprecise implementation of specific hardware blocks such as adder and multiplier, since hardware area reduction resulted in better data parallelism utilization and, therefore, increased the yield. Also in [16], genetic programming was used to evolve variants of the input kernel until one was found with ideal assignments which reduced the synthesized kernel area while still stochastically satisfying the output desired quality. The synthesis results in an Altera Stratix V FPGA showed that the reduced area of the approximate kernels produced a gain of $\times 1.4$ to 3.0 higher with less than 1% of quality loss compared to benchmarks.

Lastly, recent works, [20,10,1,2,24], have been proposing generic implementations relative to GA parameters, similar to the proposals [6] and [7]. In [20] it proposed an implementation of the genetic algorithm (called HGA) on reconfigurable hardware using the hardware-software (HW/SW) co-design methodology. Part of the GA algorithm is executed by a software embedded on CPU, the other is on a dedicated hardware (both are embedded on a FPGA). Several approaches of the high-speed general purpose hardware to accelerate genetic algorithms are proposed in [10,1,2,24]. These approaches improve the GA parameter configuration on hardware architecture, on the other hand, reduce the parallelization of the hardware design and decrease the speedup.

It is essential to observe that the works presented in the literature propose the solutions based on software and hardware on FPGA. This kind of the answer increase de flexibility but decrease the throughput. Thus, differently, of the papers in the literature, this work proposed a high-performance parallel implementation of GA. The implementation uses the full-parallel strategy in all GA operations in order to maximize the throughput.

1.2 Paper organization

In section 2 a theoretical foundation about the genetic algorithms will be explored. Section 3 will present a detailed description of the architecture development and implementation, describing the various hardware modules used to construct the parallel genetic algorithm. Later, in section 4, a careful analysis of the results obtained from the implementation described in the previous section will be performed. Simulation results, synthesis in the FPGA and the validation of the proposed architecture will be presented. The analysis will be carried for parameters such as occupation area and sampling frequency, taking into account different configurations of the proposed architecture embedded in the reconfigurable hardware. Following, Section 5 shows a comparison of the obtained with other similar works found in the state of the art. Finally, Section 6 will present the conclusions about the work.

2 Genetic Algorithms

The GAs are used to solve search and optimization problems where an optimal solution can be found through an iterative process in which the search starts from an initial population and then, combining the best representatives of it, obtains a new population that replaces the previous one [11].

GA is an iterative algorithm that is started from a population of N chromosomes randomly created. N is even, in the case of this proposed work, in order to facilitate the implementation. In every k -th iteration, also called generation or epoch, the N chromosomes are evaluated, selected, recombined and mutated to form a new population also of N chromosomes, that is, the entire population of parents is replaced by the new offspring. Then, the new population is used as input to the algorithm's next iteration (generation), and this procedure of population updating is repeated K times, where K is the GA generations number.

The Algorithm 1 displays the pseudo code of a GA. This code details all the variables and procedures that will be used in the implementation to be presented in the following sections. The variable $x_j[m](k)$ represents the j -th chromosome of m bits in the k -th generation and $\mathbf{X}[m](k)$ is a vector that stores all the N chromosomes, that is,

$$\mathbf{X}[m](k) = \begin{bmatrix} x_1[m](k) \\ \vdots \\ x_N[m](k) \end{bmatrix}. \quad (1)$$

After the initialization process, the fitness function, called FF (Line 4 of Algorithm 1), calculates the fitness value of the N chromosomes $x_j[m](k)$ of the population. This operation is applied to all chromosomes and results in a respective value $y_j[a](k)$ for each j -th chromosome, where b is the number of bits representing the fitness value. The better the value $y_j[a](k)$ of the chromosome $x_j[m](k)$, the more likely it is to continue in the new generations. The fitness values of all N individuals are stored in

$$\mathbf{Y}[a](k) = \begin{bmatrix} y_1[a](k) \\ \vdots \\ y_N[a](k) \end{bmatrix}. \quad (2)$$

After calculating the fitness value of each j -th chromosome of the k -th generation, the selection operation is performed. In GAs, the selection's purpose is to highlight the chromosome $x_j[m](k)$ alongside its respective fitness values, $y_j[a](k)$, in order to produce better populations. There is a great variety of selection methods in the literature and among them it can be mentioned: the method of selection by ranking, by tournament, roulette selection and elitism. The tournament selection method used in this implementation is one of the most used [22, 32] and it makes a competition between two or more randomly chosen chromosomes from the population stored in $\mathbf{X}[m](K)$. This

Algorithm 1 Genetic algorithm.

```

1: Initialise  $\mathbf{X}[m](k)$  with random values
2: for  $k \leftarrow 1$  to  $K$  do
3:   for  $j \leftarrow 1$  to  $N$  do
4:      $y_j[a](k) \leftarrow FF(x_j[m](k))$ 
5:   end for
6:   for  $j \leftarrow 1$  to  $N$  do
7:      $w_j[m](k) \leftarrow SF(\mathbf{Y}[a](k), \mathbf{X}[m](k))$ 
8:   end for
9:   for  $i \leftarrow 1$  to  $N/2$  do
10:     $\begin{bmatrix} z_{2i-1}[m](k) \\ z_{2i}[m](k) \end{bmatrix} \leftarrow CF\left(\begin{bmatrix} w_{2i-1}[m](k) \\ w_{2i}[m](k) \end{bmatrix}\right)$ 
11:   end for
12:   for  $v \leftarrow 1$  to  $P$  do
13:      $x_v[m](k) \leftarrow MF(z_v[m](k))$ 
14:   end for
15: end for

```

competition consists of comparing the strength (fitness), $y_j[a](k)$, of all participating chromosomes and the one who holds the best respective value in $\mathbf{Y}[a](K)$, proceeds in the algorithm to pass their genes forward. The selection function, called here SF (Line 7 of the Algorithm 1), has the vectors $\mathbf{Y}[a](k)$ and $\mathbf{X}[m](k)$ from the k -th generation as its inputs and, for each input value, it outputs the variable $w_j[m](k)$ that can assume the value of any of the N chromosomes stored in $\mathbf{X}[m](k)$. All $w_j[m](k)$ values are grouped in

$$\mathbf{W}[m](k) = \begin{bmatrix} w_1[m](k) \\ \vdots \\ w_N[m](k) \end{bmatrix} \quad (3)$$

in order to be used in the crossover stage.

The crossover stage in the k -th generation occurs after the selection of the most fit chromosomes in the population (stored in $\mathbf{W}[m](k)$) by the selection function and aims to originate new chromosomes of which will, after the mutation stage, compose the next GA generation updating the vector $\mathbf{X}[m](k)$. There are several crossover techniques presented in the literature and the strategy adopted in this implementation was the single point crossover. The crossover operation, called here CF (Line 10 of the Algorithm 1), has as input pairs of elements from vector $\mathbf{W}[m](k)$ of the k -th generation and as output, pairs of

$$\mathbf{Z}[m](k) = \begin{bmatrix} z_1[m](k) \\ \vdots \\ z_N[m](k) \end{bmatrix} \quad (4)$$

which stores the chromosomes after crossingover, that is, the new k -th offspring.

The last GA's step is the mutation operation that changes the value of a group P chromosomes, in order to provide greater diversity to the population

avoiding its solution to stabilise in local minimums. The mutation rate, MR , is the parameter responsible for controlling the amount of mutated chromosomes. Normally, the MR ranges from 0.1% to 2%. The P value can be easily calculated by the expression

$$P = \lceil N \times MR \rceil. \quad (5)$$

The mutation operation, referred here as MF , is presented in the Line 13 of the Algorithm 1 and detailed in Equation 6.

$$x = (\neg z \wedge Rand) \vee (z \wedge \neg Rand). \quad (6)$$

Where z is the chromosome to be mutated and $Rand$ is a random variable. The result of this exclusive OR operation between z and $Rand$ ($Z \oplus Rand$) results in x .

3 Hardware Proposed

Figure 1 presents the general architecture of the proposed GA hardware implementation. The entire algorithm was developed using a parallel architecture focusing on accelerating the processing speed, taking advantage of the available hardware resources, similarly to [21]. The Figure details in block diagram the main subsystems of the proposed implementation, which in turn were encapsulated in order to make the general visualization of the architecture less complex. It is possible to observe a population of N chromosomes of m bits in which $x_j[m](k)$ represents the j -th chromosome of the population in the k -th generation, according to the Algorithm 1. Each j -th chromosome $x_j[m](k)$ is stored in a m -bit register, called here RXj whose value is updated by the new population of N chromosomes produced after the processes of selection, crossover and mutation. This updating process occurs every time the synchronization module, called here $SyncM$, enables the registers to store new values.

Given that the implementation optimizes two-variables functions, each register RXj stores the values of both binary inputs for the fitness function using bits concatenation for such storage. The first $\frac{m}{2}$ bits represent the first input of the fitness function, $px_j[\frac{m}{2}](k)$, while the second block of $\frac{m}{2}$ bits stores the second input for the fitness function, $qx_j[\frac{m}{2}](k)$. Thus,

$$x_j[m](k) = px_j \left[\frac{m}{2} \right] (k) \| qx_j \left[\frac{m}{2} \right] (k) \quad (7)$$

where $\|$ is the concatenation operator.

The initial population of the algorithm is randomly chosen. All random values from the present implementation is generated by pseudo random number generators based on Linear Feedback Shift Register (LFSR) [5] and [8]. 32 bits independent LFSRs based on the polynomial $r^{32} + r^{22} + r^2 + 1$ [8] were used. Each generator is characterised as $CCLFSRlj$ whose CC , l and j are labels for its position in the circuit. Every k -th generation a random variable

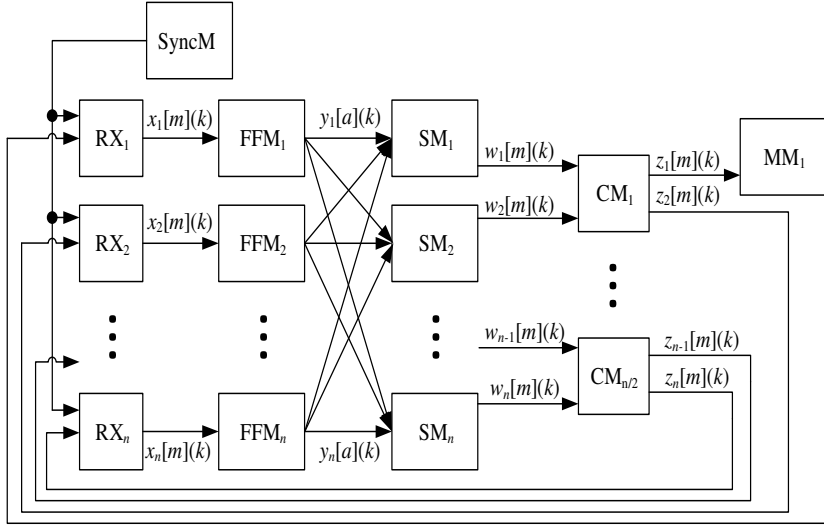


Fig. 1 General architecture of the proposed parallel genetic algorithm implementation.

of 32 bits, called here $CCr_{lj}[32](k)$, is produced by each LFSR. To avoid the same sequence of values, each generator $LFSRCCl_j$ has a different initial value of 32 bits, called $CCseed_{lj}[32]$.

The notation used in the following diagrams will be in the $x[m](c)$ form, where x is the variable, m is the bit word width and c represents the generation of the genetic algorithm ranging from 1 to K . In some cases only the bracketed notation, $[m]$, will be shown to represent the amount of bits transferred on the bus.

The implementation consists of five main modules called: Fitness Function Module (FFM), the Selection Module (SM), Crossover Module (CM), Mutation Module (MM) and Synchronization Module (SyncM). Each module has its specific implementations that will be detailed in the following sections.

3.1 Fitness Function Module - FFM

The Fitness Function Module (Figure 2) has the purpose of calculating the fitness value of each j -th chromosome from a fitness function $f(\cdot)$. The proposed structure has N FF modules and each j -th module, called here FFM_j , is associated to an individual $x_j[m](k)$ and generates as output in every k -th generation a fixed-point fitness value expressed by

$$y_j[a](k) = FFM_j(x_j[m](k)), \quad (8)$$

where a represents the bit width (equivalent to the 4 line of the Algorithm 1).

Not only for the Fitness Function Module, but for all other stages, the proposed architecture is capable of solving one or two variable problems. Regardless of the case, the user will not need to make any adjustments to the

input data. The difference between these two options reflects only on how the data is manipulated by the subsequent modules, but this does not change the performance of the system and is done invisibly to the operator.

Figure 2 details the operation of the j th FFM. The FFM input value, $x_j[m](k)$ stored in the RX_j register, is divided into two halves of $\frac{m}{2}$ bits, $px_j[\frac{m}{2}](k)$ and $qx_j[\frac{m}{2}](k)$, by the bit splitters $FFMDIV1_j$ and $FFMDIV2_j$ so that it is thus possible to operate each variable independently in the case of two variables problems.

After split, the variable $px_j[\frac{m}{2}](k)$ is directed to the ROM memory $FFMROM1_j$ which implements the α function through a Look-Up Table (LUT) and the variable $qx_j[\frac{m}{2}](k)$ is directed to $FFMROM2_j$ which implements the β function in the same fashion.

After this, both values are added by the $FFMADD_j$ adder resulting in the $\delta_j[d](k)$ variable, where

$$\delta_j[d](k) = \alpha(px_j)[c](k) + \beta(qx_j)[c](k). \quad (9)$$

The variable $\delta_j[d](k)$ is then directed to the LUT $FFMROM3_j$ where the γ function will be implemented, hence

$$y_j[a](k) = \gamma(\delta_j[d](k)). \quad (10)$$

In general, the FFM shown in Figure 2 is able to solve any one or two variables problem in the format

$$y_j[a](k) = \gamma(\alpha(px_j[c](k)) + \beta(qx_j[c](k))). \quad (11)$$

Other expressions are possible, but it is necessary to change the hardware structure of the FFM.

3.2 Selection Module - SM

The selection module (SM) implements the tournament selection method, as mentioned in Section 2, by doing a competition between two chromosomes. Similarly to the FFM, there are N SMs for a group of N chromosomes. As detailed in Figure 3, each j -th SM, here called SM_j , has as input the N fitness values, $y_j[a](k)$, and N chromosomes, $x_j[m](k)$, from the k -th generation (equivalent to the 7 line of the Algorithm 1).

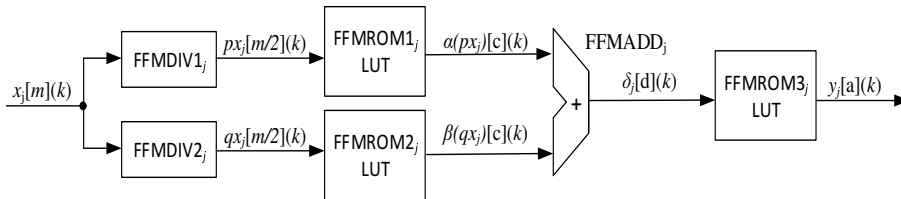


Fig. 2 Fitness function module - FFM.

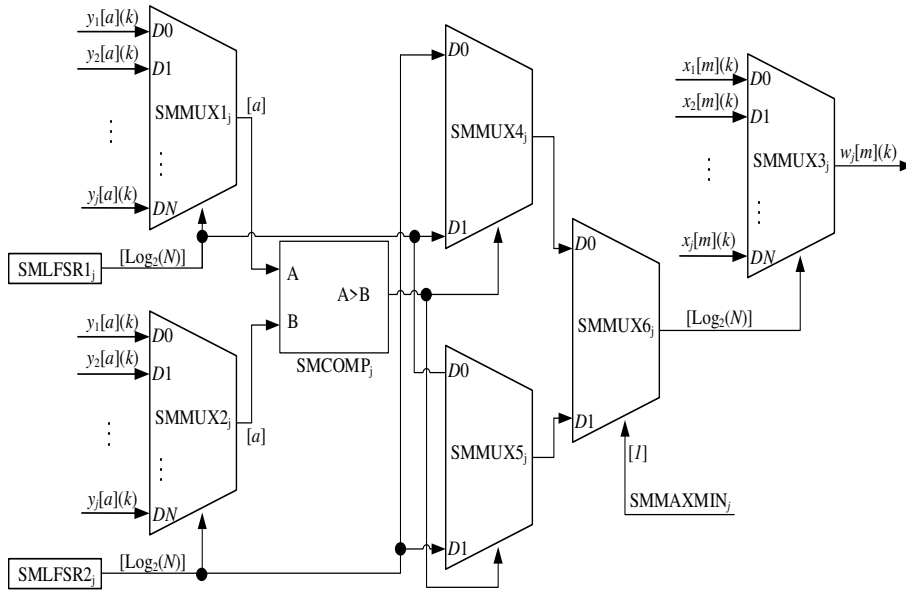


Fig. 3 Selection module - SM.

Each j -th SM has two random generators called SMLFSR1 j and SMLFSR2 j . In addition to the random generators, this module is formed by three N input multiplexers called here SMMUX1 j , SMMUX2 j and SMMUX3 j , a m bits comparator, called SMCOMP j and three two-input multiplexers, called SMMUX4 j , SMMUX5 j and SMMUX6 j .

The SMMUX1 j and SMMUX2 j multiplexers are driven by the SMLFSR1 j and SMLFSR2 j generators output signal, ($SMr_{1j}[32](k)$ and $SMr_{2j}[32](k)$), respectively. As shown in Figure 3 the output signal of each generator ($SMr_{1j}[32](k)$ and $SMr_{2j}[32](k)$) is truncated in the most significant $\lceil \log_2(N) \rceil$ bits in order to match the population size. The SMMUX1 j and SMMUX2 j multiplexers select one fitness value each, which is related to its correspondent chromosome by its index value.

Finally, SMMUX3 j selects the chromosome associated with the best fitness function value from the output of SMMUX6 j which selects whether the goal is to maximize or minimize the evaluation function through the SMMAXMIN j variable.

3.3 Crossover Module - CM

The crossover module detailed here in this section implements single point crossover. The architecture proposed here contains $\frac{N}{2}$ crossover modules and each one consists of four bit splitters, two identical crossover submodules, and two concatenators. Similarly to the FFM described in Section 3.1, the CM also

has chromosome splitters in order to manipulate the two variables stored in $w[m](k)$ independently.

As seen in Figure 4, the two input chromosomes, $w_{j-1}[m](k)$ and $w_j[m](k)$, are split into two halves, each. The first $w_{j-1}[m](k)$ half is sectioned by the splitter CMDIV1_j which is renamed $pw_{j-1}[\frac{m}{2}](k)$ and the second half of that same variable is sectioned by the splitter CMDIV2_j and becomes $qw_{j-1}[\frac{m}{2}](k)$. The same happens with the chromosome $w_j[m](k)$ which is sectioned into $pw_j[\frac{m}{2}](k)$ and $qw_j[\frac{m}{2}](k)$ through the divisors CMDIV3_j and CMDIV4_j , respectively.

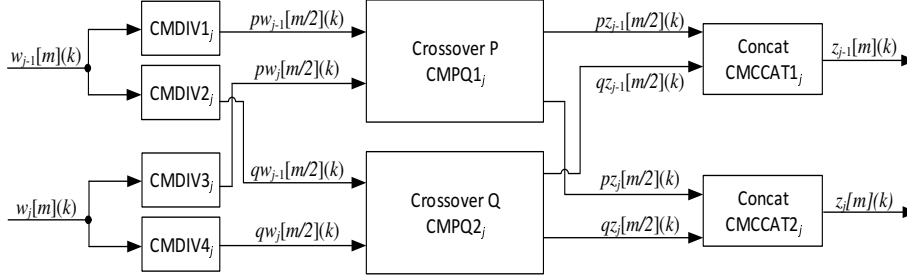


Fig. 4 J -th Crossover module (CM_j).

Separating the variables of each chromosome, they are forwarded to the CM submodules CMPQ1_j and CMPQ2_j so then the crossing is performed. This is conducted in such a way that the crossover is performed between similar variables, that is, the first variable $pw_{j-1}[\frac{m}{2}](k)$ from the chromosome $w_{j-1}[m](k)$ will be crossed with the first variable $pw_j[\frac{m}{2}](k)$ from the chromosome $w_j[m](k)$.

In the case of single variable problems the system works in an equivalent way. Only the least significant half of the variables $w_{j-1}[m](k)$ and $w_j[m](k)$ will contain useful data and only block CMPQ2_j will handle nonzero data.

Figure 5 presents in detail the circuit of the j -th crossover submodule named CMPQ1_j . It is composed by a $\frac{m}{2}$ -input MUX, called here CMPQMUX_j , whose purpose is to randomly select one of the $\frac{m}{2}$ possible cutting points. The selection of each CMPQMUX_j is controlled by the pseudo random number generator CMPQLFSR1_j whose output signal, $\text{CMPQr}_{1i}[32](k)$, is truncated in the $\lceil \log_2(\frac{m}{2} + 1) \rceil$ more significant bits before entering the MUX selector.

The selection of the CMPQ1_j cut-off point is done relying on the mask originated from the constant $2^{\frac{m}{2}} - 1$. This constant creates a vector of 1s of the size of the chromosome to be crossed, in the case $\frac{m}{2}$. Then, a random and zero-padding right shift is performed according to CMPQLFSR1_j value. This displacement will transform the vector of 1s into a vector of 0s and 1s concatenated and still of size $\frac{m}{2}$. This mask and its inverse will be responsible for carrying out the crossover operation aided by the AND and OR logic gates shown also in Figure 5.

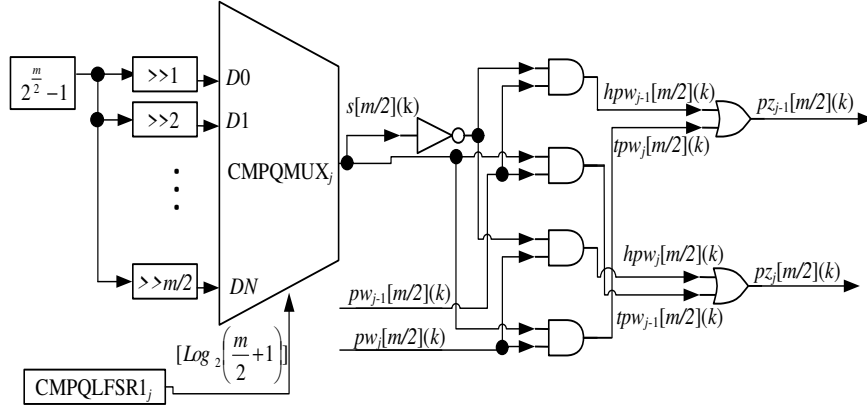


Fig. 5 J -th Crossover submodule (CMPQ1 $_j$).

Equations 13 and 14 exemplify a case where $m = 20$ and CMPQMUX_j shifts the value of $2^{\frac{m}{2}} - 1$ three times

$$2^{\frac{m}{2}} - 1 = 1111111111; \quad (12)$$

$$s_i \left[\frac{m}{2} \right] (k) = 0001111111; \quad (13)$$

$$\neg s_i \left[\frac{m}{2} \right] (k) = 1110000000. \quad (14)$$

In each k -th generation, the two entries of the module CMPQ1_j , the variables $pw_{j-1}[\frac{m}{2}](k)$ and $pw_j[\frac{m}{2}](k)$ are divided into head

$$hpw_{j-1} \left[\frac{m}{2} \right] (k) = \neg s_j \left[\frac{m}{2} \right] (k) \wedge pw_{j-1} \left[\frac{m}{2} \right] (k) \quad (15)$$

$$hpw_j \left[\frac{m}{2} \right] (k) = \neg s_j \left[\frac{m}{2} \right] (k) \wedge pw_j \left[\frac{m}{2} \right] (k) \quad (16)$$

and tail

$$tpw_{j-1} \left[\frac{m}{2} \right] (k) = s_j \left[\frac{m}{2} \right] (k) \wedge pw_{j-1} \left[\frac{m}{2} \right] (k), \quad (17)$$

$$tpw_j \left[\frac{m}{2} \right] (k) = s_j \left[\frac{m}{2} \right] (k) \wedge pw_j \left[\frac{m}{2} \right] (k), \quad (18)$$

where $s[\frac{m}{2}](k)$ is the CMPQMUX output. After this step, the crossover will be performed by concatenating the head of parent 1, $hpw_{j-1}[\frac{m}{2}](k)$, with the parent's tail 2, $tpw_j[\frac{m}{2}](k)$, and the parent head 2, $hpw_j[\frac{m}{2}](k)$, with the parent's tail 1, $tpw_{j-1}[\frac{m}{2}](k)$, thus giving rise to two new chromosomes of the new population,

$$pz_{j-1} \left[\frac{m}{2} \right] (k) = hpw_{j-1} \left[\frac{m}{2} \right] (k) \vee tpw_j \left[\frac{m}{2} \right] (k). \quad (19)$$

and

$$pz_j \left[\frac{m}{2} \right] (k) = hpw_j \left[\frac{m}{2} \right] (k) \vee tpw_{j-1} \left[\frac{m}{2} \right] (k). \quad (20)$$

For the CMPQ2_j submodule the equivalent happens. In this case, the input values will be $qw_{j-1}[\frac{m}{2}](k)$ and $qw_j[\frac{m}{2}](k)$ and the outputs will be $qz_{j-1}[\frac{m}{2}](k)$ and $qz_j[\frac{m}{2}](k)$.

After the similar variables have been crossed within each submodule, they are directed to the outputs of each respective CMPQs where the concatenators CMCCAT1_j and CMCCAT2_j will give rise to new individuals (chromosomes) from the population by concatenating both the parts forming them, $pz[\frac{m}{2}](k)$ and $qz[\frac{m}{2}](k)$ (Figure 4).

It is important to emphasize that after $\frac{N}{2}$ CMs have performed their operations, N new chromosomes that will form a new population will have been created. Some of these individuals will pass through the MM (to be described in Section 3.4) before the start of the next generation, but always at the end of each iteration, N new individuals will have been created so that the GA population will always remain with N chromosomes.

3.4 Mutation Module - MM

As in the Algorithm 1 in Line 13 the mutation operation will be performed on a group of P individuals, that is, there are P mutation modules and each j -th module, MM_j , changes the value of the chromosome to be mutated through an XOR operation with a number created randomly by an associated generator called MMLFSR_j (Figure 6). The P MM will modify the first P individuals of the population as shown in Figure 1.

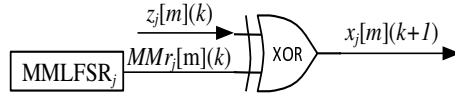


Fig. 6 Mutation module - MM.

The output of the j -th, MM_j , module in every k -th generation can be expressed by

$$x_j[m](k+1) = (\neg z_j[m](k) \wedge \text{MMr}_j[m](k)) \vee (z_j[m](k) \wedge \neg \text{MMr}_j[m](k)) \quad (21)$$

where $\text{MMr}_j[m](k)$ represents the pseudo random number generated by j -th MMLFSR_j .

In the case of single-variable problem optimization, this mutation operation will possibly assign non-zero values to the $\frac{m}{2}$ unused bits of the mutated chromosome. However, this will not be a problem since these $\frac{m}{2}$ bits will be zeroed when passing through the FFM in the following generation.

3.5 Synchronization Module - SyncM

Finally, the last module is the synchronization module. It aims to enable the registers, responsible for storing the population chromosomes of the genetic algorithm, to receive new values. These new values result from the mutation and crossover processes of the previous generation and are stored in the RX registers to initiate a new iteration of the algorithm.

This module contains a counter, a constant value and a comparator as shown in Figure 7. The variable *enable* is enabled when the comparison returns a true value, that is, when the counter value matches the value stored in the constant. The *SyncVal[2]* value is obtained according to the implemented design, and it is adjusted according to the delay that the system needs to perform all its operations and provide a new set of chromosomes. The output value of this module is a boolean value, and the values of the counter and constant output are 2-bit values. This number was chosen because it was the maximum delay found in the implementation for an entire generation, a delay for each ROM of the FFM.

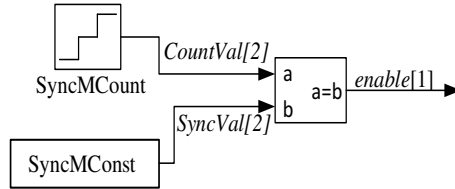


Fig. 7 Synchronization module - SyncM.

In all the tests performed in this work the GA operations were performed at a sampling rate (or clock frequency)

$$R_c = \frac{3}{T_g} \quad (22)$$

where T_g is the time for each k -th generation be finished.

Although R_c is the maximum possible sample rate to operate the system and T_g is the minimum equivalent time, the equation 22 divides these values by 3 since only every three clocks a new population is originated in the GA, since there are two delays in the architecture between the beginning of the k generation and the end of it. Thus, these two delays caused by the LUTs contained in the FFM (Section 3.1) make the frequency $\frac{3}{T_g}$ the one which will process the population $k + 1$ from an earlier population k . The Equation 22 can be rewritten as

$$R_g = \frac{R_c}{3} \quad (23)$$

where R_g is the number of generations per second ($R_g = \frac{1}{T_g}$).

Generally, if the architecture contained any η components that caused system delays, the sample rate R_c of this system would be

$$R_c = \frac{\eta}{T_g}. \quad (24)$$

4 Results

Aiming to validate the proposed implementation of the GA on FPGA, simulations, analyses and syntheses were performed in the optimisation of different functions for various population sizes. The first function, called here F1, used in the tests to validate the proposal was an one variable function expressed as

$$f(x) = x^3 - 15x^2 + 50, \quad (25)$$

The second function, called here F2, was

$$f(x, y) = 8x - 4y + 1020, \quad (26)$$

and, lastly, the last function, here called F3, was the function

$$f(x, y) = \sqrt{x^2 + y^2}. \quad (27)$$

This work has implemented and synthesised on FPGA the three functions previously mentioned for populations of size $N = 4$, $N = 8$, $N = 16$, $N = 32$ and $N = 64$ and for chromosomes with size $m = 20$, $m = 22$, $m = 24$, $m = 26$, and $m = 28$ bits.

It is important to emphasise that these functions were chosen for comparison reasons, since they have already been used in the state of the art in previous works that will be shown next. However, the implementation proposed is capable of implementing any function in the format shown by Equation 11 requiring only the modification of the values stored in the memories.

All results were obtained using the development platform and a FPGA Virtex 7 xc7vx550t-1ffg1158. The Virtex 7 FPGA used has 86,600 slices that group 692,800 flip-flops, 554,240 logical cells that can be used to implement logical functions or memories and 2,880 DSP cells with multipliers and accumulators.

As previously mentioned, three different functions were minimised for the validation of the implementation. The first one was the function F1 presented in Equation 25 and shown in Figure 8.

This function was chosen because it was previously used by [31] to validate its proposal which developed a high-speed Genetic Algorithm on FPGA. Regarding the implementation of this function as described in Equation 11, the following associations can be made:

$$\alpha(px) = 0, \quad (28)$$

$$\beta(qx) = (qx)^3 - 15(qx)^2 + 50, \quad (29)$$

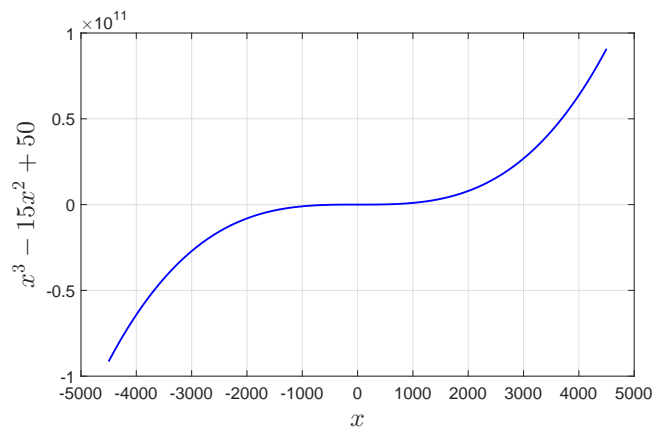


Fig. 8 Fitness function F1: $f(x) = x^3 - 15x^2 + 50$.

$$\gamma(\delta) = \delta, \quad (30)$$

Therefore, F1 can be represented by

$$y(px, qx) = 1 * ((qx)^3 - 15(qx)^2 + 50 + 0). \quad (31)$$

The second function is presented in Figure 9 and has been previously used in [6, 7] to validate the implementation of a customizable GA IP core for general purposes.

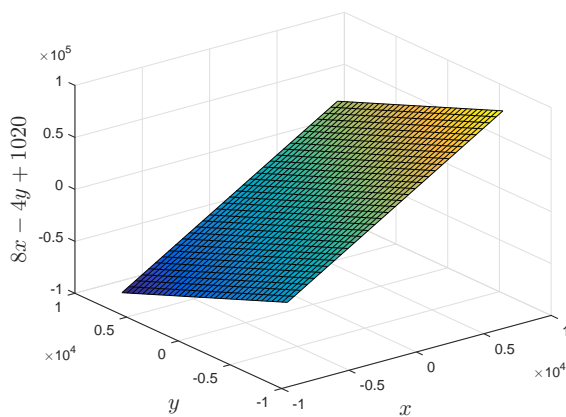


Fig. 9 Fitness function F2: $f(x, y) = 8x - 4y + 1020$.

Regarding the implementation of F2 as described in Equation 11, the following associations can be made:

$$\alpha(px) = 8(px), \quad (32)$$

$$\beta(qx) = -4(qx) + 1020, \quad (33)$$

$$\gamma(\delta) = \delta, \quad (34)$$

Therefore, F2 can be represented by

$$y(px, qx) = 1 * (8(px) - 4(qx) + 1020). \quad (35)$$

Finally, Figure10 shows the last function used to validate the proposal presented here. This function could be seen previously with the same use in [9] and [25]. Both works use GA, but only [9] implements the algorithm on FPGA.

Similarly, the F3 in the parameters of the equation 11, can be seen as follows:

$$\alpha(px) = (px)^2, \quad (36)$$

$$\beta(qx) = (qx)^2, \quad (37)$$

$$\gamma(\delta) = \sqrt{\delta}, \quad (38)$$

Therefore, F3 can be represented by

$$y(px, qx) = \sqrt{(px)^2 + (qx)^2}. \quad (39)$$

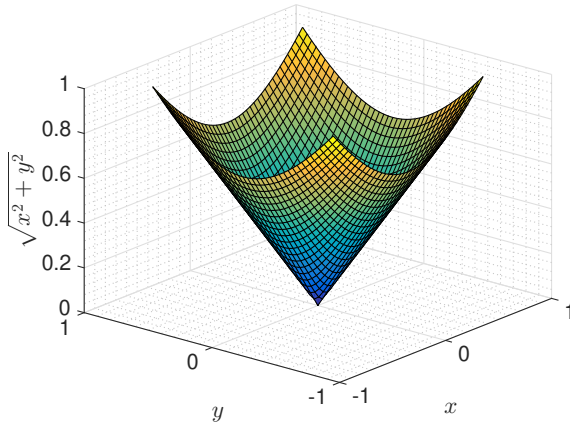


Fig. 10 Fitness function F3: $f(x, y) = \sqrt{x^2 + y^2}$.

The parameters used in the experiments were based on configurations of previous experiments found in the literature together with some empirically

obtained configurations. For the number of generations k , it was experimentally observed that for all evaluated functions, the minimum value sought was obtained before the 100 GA generations were reached. This number is in agreement with what was found in the literature, as can be seen in [31], for example. Thus, $k = 100$ was adopted as default value for the optimization experiments performed here.

Similarly, the GA population sizes to be implemented and synthesised in the FPGA were determined. As seen in [21] the population size was $N = 32$. In [27] the population had size $N = 16$ and in [5] GA was implemented with populations of sizes $N = 64$ and $N = 128$. Thus, the architecture of the proposal presented here was implemented for the five population sizes already mentioned before. The aim behind these different sizes of N was to compare how much this parameter influences the convergence, speed and area occupation in the FPGA.

Finally, for the same purpose of comparing how much a parameter influences certain convergence and synthesis characteristics, the bit length m varied for all population sizes as also quoted previously. The Figures 11 and 12 picture the operation of the proposed GA for the fitness functions F1 and F3, respectively.

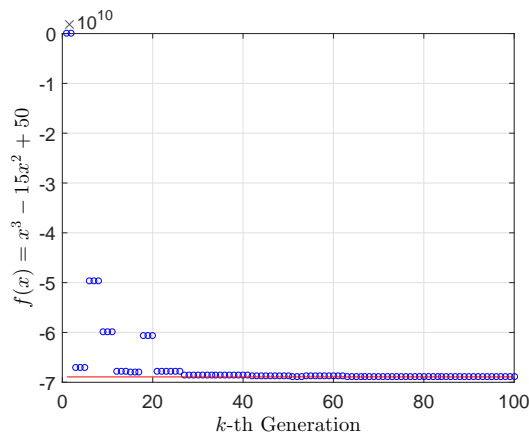


Fig. 11 Optimising F1.

The fitness function 1 (Equation 25) was minimized using the GA with a population size of $N = 32$ and $m = 26$. Thus, $\lceil \frac{m}{2} \rceil$ bits were used for each variable from the fitness function. Given that this is a single-variable problem, the function made use of only $\lceil \frac{m}{2} \rceil$ bits. For the minimization shown in Figure 11 the range of the F1 was of $f(-2^{12})$ to $f(2^{12} - 1)$. Thus, the minimum possible value in the range is $f(-2^{12}) = -6.8971 * 10^{10}$. As depicted, it is noticed that the global minimum was reached approximately in the half of the 100 generations, thus proving the functionality of the proposed system.

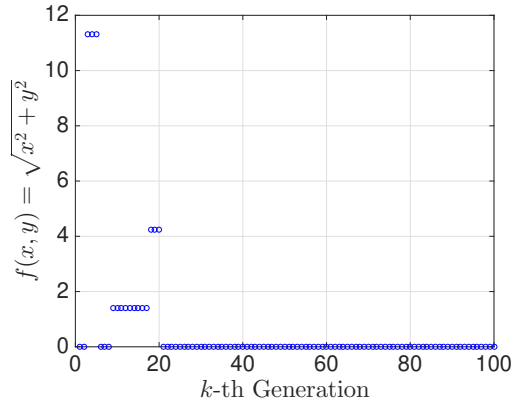


Fig. 12 Optimising F3.

Similarly, the fitness function F3 (Equation 27) was also minimized, as shown in Figure 12 but with a population size of $N = 64$ and $m = 20$. In this case, $f(x, y)$ only allows results greater than or equal to zero when working in the domain of real numbers, so the smallest possible value is zero. The parallel Genetic Algorithm implemented on FPGA proposed herein has managed to minimize F3 in a little over 20 iterations of the algorithm. This is not a fixed value, since the GA is a stochastic algorithm, but with this value it is possible to have an idea of the number of generations required for convergence.

Both results were obtained from the average of multiple results. It is also important to emphasize that parameters such as the range of values to be calculated, bit width (m), decimal precision and the possibility of exploring negative numbers are all parameters of the LUT (Section 3.1) and configurable by the user. As already mentioned, the option of maximizing or minimizing the function to be optimized is also another configurable variable.

The Table 1 presents the synthesis results in the target FPGA for various population sizes and $m = 20$. It is clear in all scenarios that the area of occupation, clock frequency (or sampling rate), R_c , consequently the number of generations per second, R_g , are parameters considerably sensitive to the population size, N . Here, the R_g represents the number of generations performed in the GA per second, which can also be interpreted as some possible solutions which the system provides in that interval. Equation 22 states that this number is equal to $\frac{R_c}{3}$, that is, the clock frequency (or sampling rate) divided by three. This is explained because the system generates two delays when placing two ROM memories in series in the FFM described in Section 3.1. Consequently, a new GA population is generated only after three system clocks.

The clock is the maximum frequency the system performs when implementing this architecture, and it does not take into account the delays required to generate a new population. The clock represents only the hardware speed for

that specific implementation, so it is $3\times$ faster than the number of generations performed in the GA per second.

Table 1 GA synthesis on FPGA for $m = 20$.

N	Registers flip-flops	Logic cells (LUTs)	Clock frequency (R_c) (MHz)	Generations per second (R_g) $\times 10^6$
4	457	592(1%)	50.28	16.76
8	839	1,558(1%)	49.32	16.44
16	1,616	4,400(1%)	49.32	16.44
32	3,225	15,908(4%)	48.51	16.17
64	6,598	58,875(16%)	34.56	11.52

The area occupation spent in registers (Figure 13), presented in the second column of the Table 1, is due to the storage of population values in RX (Figure 1) and the pseudo random number generators, mainly. This occupation increases linearly according to N , since the larger the population, the greater the number of RX registers required, as well as operations that require the pseudo random number generators. Figure 13 shows this growth graphically with a linear interpolation.

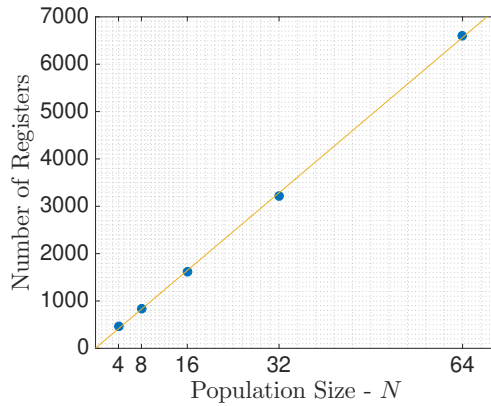


Fig. 13 Registers' occupation in the FPGA varying with N .

The logical cells (LUTs) occupation, presented in the third column of the Table 1, was increasing and not linear with N , as can be seen in Figure 14. This nonlinear growth is caused by the selection module (Subsection 3.2) that for each j -th module, SM_j , there are three N inputs multiplexers ($SMMUX1_j$, $SMMUX2_j$ and $SMMUX3_j$).

According to [3], each Virtex 7 logical cell can construct four 1-input MUXs, thus to build a N -inputs multiplexers, approximately $\frac{N}{4}$ logical cells are re-

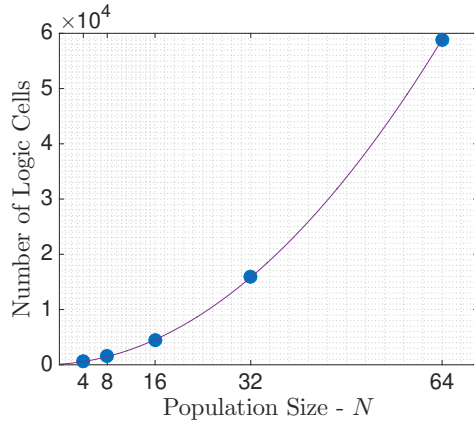


Fig. 14 FPGA LUTs occupancy varying with N .

quired, totalling approximately $\frac{3N}{4}$ cells for each SM_j (SMMUX 4_j , SMMUX 5_j and SMMUX 6_j) have not been considered). Since there are N SM modules, there are approximately $\frac{3N^2}{4}$ logical cells for each bit of the input bus of the MUX. Thus, the exponential growth result from the use of the logical cells is explained.

In this context, it is important to note that implementation with $N = 64$ individuals does not reach even one fifth of the FPGA cells (around 16% of Virtex 7). This is a positive indicator for implementations with larger populations.

Finally regarding the table, the last two columns show the Clocks and the number of generations performed in the GA per second for each value of N and there is speed reduction according to the population growth. Theoretically, if all the modules were independent (specific for each individual) this reduction should not happen, however, it is observed that in the selection modules, SM_j (Figure 3.2), there is a dependency between the N chromosomes (due to information sharing) causing a join in the circuit and thus, an increase in processing time. On the other hand, it is also observed that the reduction rate is not linear, which favours the implementation. Another important information to note is that even with the reduction, each GA generation of 64 chromosomes is generated in $T_g \approx 87$ ns, in other words $R_g = 11.52 \times 10^3$ generations to every 1 ms. This result has a very significant impact and makes the use of GA possible in several real-time embedded applications such as robotics, telecommunications and others.

The Figure 15 represents the influence of the bit width m on the Clock for a GA with $N = 32$. It is noted a decrease of the processing speed with the increase of the number of bits, however this fall is not significant. The clock variation is only slightly more than 1 MHz when the implementation is compared using $m = 20$ with the implementation using $m = 28$. The results

shown in the Figure 15 suggests a linear regression expressed as

$$\hat{R}_g = (-0.136m + 18.87) \times 10^6 \quad (40)$$

where the variable \hat{R}_g is an estimation of R_g . The linear regression R-squared was of $R^2 = 0.9314$.

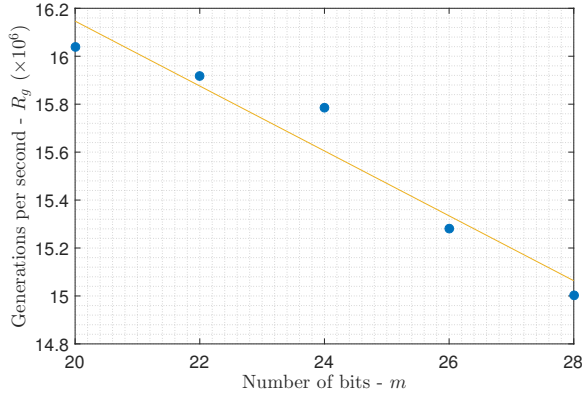


Fig. 15 Decrease of R_g when varying m .

The last Figure 16 illustrates the relationship between the increase of LUTs used in the FPGA and the variation of the bit width m for three different population sizes. A larger difference is observed between the quantities of LUTs used in $m = 28$, mainly due to the nonlinear growth of these components comparing to N . However, as already seen in Figure 15 the increase of m is also a factor responsible for slowing the processing speed, R_g .

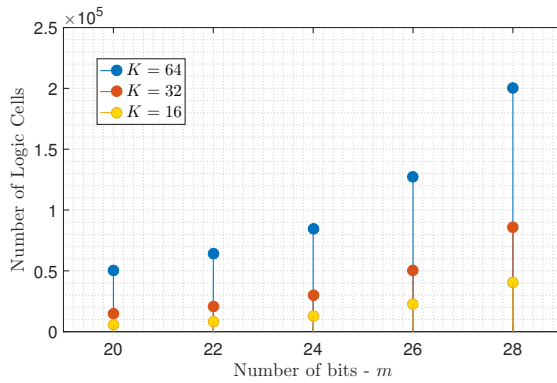


Fig. 16 Relation between the LUTs usage with the increase of m .

Analyzing the synthesis results, it was noticed that different fitness functions such as F1, F2 and F3 did not result in significant differences in the LUTs consumption and Registers in the FPGA, as well as no significant differences were observed in R_g . This result was already expected, since the only variation that occurs when changing the fitness function is the content of the FFM LUTs. Thus, it is possible to extend this thinking and assert that the values of the Table 1 are true for any other function, in the parameters of Equation 11, using $m = 20$ bits.

5 Comparisons with state of the art works

Following, comparisons of the results obtained by the proposed implementation with equivalent results found in works belonging to the state of the art are presented. The comparisons which will be shown below and which are summarized in the Table 2 were made with the greatest similarity of parameters as possible. The table presents a column that presents the comparative references, the next two columns show the parameters of the GAs compared, then the times obtained by the works of the state of the art are shown and, finally, the results obtained by the implementation presented here and the respective speedups are displayed.

Table 2 Comparative table with state of the art works.

Reference	N	K	Reference time	Obtained time	Speedup
[31]	32	100	2.19 ms	6.18 μ s	354
[5]	32	60	1.702 ms	3.71 μ s	459
[6, 7]	32	32	7.29 ms	1.98 μ s	3683
[33]	64	500	0.8 s	43.40 μ s	18432
[17]	64	500	941.17 μ s	29.7 μ s	31
[22, 32]	64	1024	1ms	58.48 μ s	17
[18]	16	256	0.8 ms	19.45 μ s	41
[21]	64	500	7.47 s	43.40 μ s	172119
[20]	64	1000	1.45 s	92 μ s	15760
[10]	32	1000000	8 s	68.49 ms	116
[2]	20	380	74 ms	22.76 μ s	3251
[24]	20	60	793.23 μ s	3.6 μ s	220

The system presented by [31], a high-speed implementation of GA on FPGA, demonstrated a run-time of 2.19 ms for a GA implemented on FPGA with $K = 100$ generations and a population of size $N = 32$. For the same settings, the system proposed here achieved a time of $\approx 6.18 \mu$ s, which proves to be $\approx 354\times$ faster.

Similarly, the implementation presented by [5] also presented an GA on FPGA with population size $N = 32$, chromosome size of $m = 16$ and $K = 60$ generations. The implementation validated its proposal with the traveling salesman problem and resulted in a running time of 1.702 ms. Although a test in the same parameters of [5] has not been performed here, a comparison can

still be made due to the versatility of problems solved by different LUT as shown in the FFM. A GA with $K = 60$ generations, $N = 32$ chromosomes and $m = 20$ bits can be solved in $\approx 3.71\mu\text{s}$ in the work presented here, meaning a time $\approx 459\times$ faster than in [5].

In similar fashion, the work of [6, 7] presented a highly programmable GA IP core on FPGA. For a setting of $K = 32$ and $N = 32$ the authors stated a speedup of $5.16\times$ over an equivalent software implementation which achieved a running time of 37.615 ms . In a comparison, the implementation presented here performed the equivalent situation in a time of $\approx 1.98\mu\text{s}$, which represents a speedup of $\approx 19007\times$ over the serial implementation shown in [6, 7]. In other words, a time $\approx 3683\times$ less than the GA IP core proposed by [6, 7]. The flexibility proposed by [6, 7] decreases the parallelization on the hardware implementation and reduces the speedup.

The clock achieved in the work proposed in [17] was about $\frac{1}{9.8\text{ ns}} = 102\text{ MHz}$ with $m = 15$ bits. However, the hardware implementation needs to read and write the RAMs, sequentially, for saving GA variables. In other words, the implementation is similar to Algorithm 1, and for N individuals it is needed to execute three times the N -sized loop (Lines 3-5, 6-8 and 9-11). Based on the architecture proposed in [17], the value of $R_g = \frac{102}{3N} = 531.25 \times 10^3$ generations per second for $N = 64$. In [22, 32] the GA executes $K = 1024$ generations at 1 ms with a clock frequency of 40 MHz ($R_g = 1 \times 10^6$ generations per second), $N = 64$ chromosomes and $m = 10$ bits. Using Equation 15, the implementation presented here can perform the equivalent of $\hat{R}_g = 16.83 \times 10^6$ generations per second and $\hat{R}_g = 17.51 \times 10^6$ for $m = 15$ and $m = 10$ bits, respectively.

The work [18] presented a real-time GA for adaptive filtering application with all modules implemented in hardware. The implementation achieved a rate of $R_g = 320 \times 10^3$ generations per second for a population of $N = 16$ and $m = 42$ bits (seven parameters with 6 bits). Using Equation 40, it can be estimated that the implementation, proposed here, has about $\hat{R}_g = 13.158 \times 10^6$ generations per second for $N = 32$ and $m = 42$ bits. Similarly to the architecture shows in work [17], the research [21] achieved about $R_g = 67$ generations per second (189000 clock pulses at 12.5 MHz).

The implementation here proposed can also be compared to the work published in [33]. As already mentioned in Section 1.1, this article presents the OIMGA, an implementation of a monogenetic FPGA algorithm that retains only the best chromosome of the generation. In one of the tests performed to validate the implementation, the authors optimized a one variable function with a population of $N = 64$ chromosomes in ≈ 0.8 seconds. In a scenario where the proposed parallel GA take this time to solve the same function it would process $K = 9.2 \times 10^6$ generations. Of course, this value is unreasonable for that function. As shown previously in the results, $K = 100$ generations was the default value to optimize functions of one or two variables, thus, even if the number of generations needed to optimize the same function was $K = 500$ (a generous estimate), the time resulting from the implementation proposed by [33] would still be $\approx 18432\times$ higher.

The architecture shown in [20] executes $K = 1000$ generations in 1.45 s with a clock frequency of 50 MHz, $N = 20$ chromosomes, $m = 16$ bits and this corresponds to the $R_g = 689.65$ generations per second (72740 clock cycles per generation). In [10], $K = 1 \times 10^6$ generations are processed in about 8 s ($R_g = 125 \times 10^3$ generations per second) with $N = 32$ chromosomes and, $m = 32$ bits. With similar parameters and using Equation 15, the work here proposed can achieve $\hat{R}_g = 14.5 \times 10^6$ generations per second. A comparison among several metaheuristics algorithms on hardware is studied in [2], and the GA performed $R_g = 5.13 \times 10^3$ generations per second (9740 clock cycles per generation) for $K = 380$, $N = 20$, $m = 16$, and a clock of 50 MHz. Lastly, a fully customizable hardware implementation is proposed in [24]. This work uses the same strategy of the works presented in [6, 7, 10, 2], where it allows the setup of several GA parameters on hardware after the FPGA synthesis. For the same parameters used in [10] ($N = 20$, $m = 16$, and a clock of 50 MHz), and $K = 60$ the work [24] archived $R_g = 75.64 \times 10^3$ generations per second (661 clock cycles per generation). The GA implementation here proposed can execute $\hat{R}_g = 16.69 \times 10^6$ generations per second for $N = 32$ and $m = 16$.

6 Conclusion

After the presentation of the results in Section 4 it can be affirmed that the implementation proposal was in fact validated and fulfilled with its objective of being a parallel implementation of high-performance of a GA. The synthesis results confirmed that the present proposed parallel implementation of GA on FPGA is able to optimize a wide range of functions in a viable time for critical applications that require short time constraints or a large amount of data to be processed in a short interval.

Comparisons with other implementations found in the literature in Section 5 reinforce the high speed achieved by the implementation developed. This enables the use of this system in a commercial context for applications such as tactile internet, robotics, real-time applications and medical applications. In addition, this system has proven to be an acceleration tool for any hardware system that makes use of genetic algorithms.

As well as the high-performance achieved, the small area consumption of the implementation developed here is a notorious feature. This makes it possible for other systems to also be embedded in the FPGA, since the on-board GA occupies less than $\frac{1}{5}$ of the Virtex 7 logic cells used as a test. This logical cells low consumption feature is essential for applications where the area is the biggest constraint as spatial applications, for example.

The experiments carried out proved that the sizes of N tested are sufficient to solve most of the practical problems as the literature confirms. It has been found that the duration in iterations (k) of GA does not need to be greater than a few hundred. It has been proven that a few hundred generations or even $k = 100$ is a reasonable number of generations for a GA. The parameter m proved to be of great importance, since it directly affects the GA convergence

speed, the area occupied on the FPGA, the response precision, as well as the achieved R_g .

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001.

References

1. Alinodehi, S.P.H., Moshfe, S., Zaeimian, M.S., Khoei, A., Hadidi, K.: High-speed general purpose genetic algorithm processor. *IEEE transactions on cybernetics* **46**(7), 1551–1565 (2016)
2. Ameer, M.S.B., Sakly, A.: Fpga based hardware implementation of bat algorithm. *Applied Soft Computing* **58**, 378–387 (2017)
3. Chapman, K.: Multiplexer design techniques for datapath performance with minimized routing resources. Application Note: Spartan-6 Family, Virtex-6 Family, 7 Series FPGAs (2014)
4. Chen, Y., Wu, Q.: Design and implementation of pid controller based on fpga and genetic algorithm. In: *Electronics and Optoelectronics (ICEOE), 2011 International Conference on*, vol. 4, pp. V4–308. IEEE (2011)
5. Deliparaschos, K., Doyamis, G., Tzafestas, S.: A parameterised genetic algorithm ip core: Fpga design, implementation and performance evaluation. *International Journal of Electronics* **95**(11), 1149–1166 (2008)
6. Fernando, P., Sankaran, H., Katkoori, S., Keymeulen, D., Stoica, A., Zebulum, R., Rajeshuni, R.: A customizable fpga ip core implementation of a general purpose genetic algorithm engine. In: *Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE International Symposium on*, pp. 1–8. IEEE (2008)
7. Fernando, P.R., Katkoori, S., Keymeulen, D., Zebulum, R., Stoica, A.: Customizable fpga ip core implementation of a general-purpose genetic algorithm engine. *IEEE Transactions on Evolutionary Computation* **14**(1), 133–149 (2010). DOI 10.1109/TEVC.2009.2025032
8. Goresky, M., Klapper, A.: Pseudonoise sequences based on algebraic feedback shift registers. *IEEE Transactions on Information Theory* **52**(4), 1649–1662 (2006)
9. Guo, L., Funie, A.I., Thomas, D.B., Fu, H., Luk, W.: Parallel genetic algorithms on multiple fpgas. *ACM SIGARCH Computer Architecture News* **43**(4), 86–93 (2016)
10. Guo, L., Funie, A.I., Xie, Z., Thomas, D., Luk, W.: A general-purpose framework for fpga-accelerated genetic algorithms. *International Journal of Bio-Inspired Computation* **7**(6), 361–375 (2015)
11. Holland, J.H.: *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press (1975)
12. Instruments, N.: *Understanding parallel hardware: Multiprocessors, hyperthreading, dual-core, multicore and fpgas* (2011). URL <http://www.ni.com/tutorial/6097/en/>
13. Ionescu, L.M., Mazare, A., Lita, A.I., Serban, G.: Fully integrated artificial intelligence solution for real time route tracking. In: *2015 38th International Spring Seminar on Electronics Technology (ISSE)*, pp. 536–540. IEEE (2015)
14. Jewajinda, Y., Chongstitvatana, P.: Hardware architecture and fpga implementation of a parallel elitism-based compact genetic algorithm. In: *TENCON 2009-2009 IEEE Region 10 Conference*, pp. 1–6. IEEE (2009)
15. Koza, J.R.: Genetic evolution and co-evolution of computer programs. *Artificial life II* **10**, 603–629 (1991)
16. Lotfi, A., Rahimi, A., Yazdanbakhsh, A., Esmaeilzadeh, H., Gupta, R.K.: Grater: An approximation workflow for exploiting data-level parallelism in fpga acceleration. In: *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1279–1284. IEEE (2016)
17. Mengxu, F., Bin, T.: Fpga implementation of an adaptive genetic algorithm. In: *2015 12th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–5. IEEE (2015)

18. Merabti, H., Massicotte, D.: Hardware implementation of a real-time genetic algorithm for adaptive filtering applications. In: Electrical and Computer Engineering (CCECE), 2014 IEEE 27th Canadian Conference on, pp. 1–5. IEEE (2014)
19. Mingas, G., Tsardoulas, E., Petrou, L.: An fpga implementation of the smg-slam algorithm. *Microprocessors and Microsystems* **36**(3), 190–204 (2012)
20. Nambiar, V.P., Balakrishnan, S., Khalil-Hani, M., Marsono, M.N.: Hw/sw co-design of reconfigurable hardware-based genetic algorithm in fpgas applicable to a variety of problems. *Computing* **95**(9), 863–896 (2013)
21. Nedjah, N., de Macedo Mourelle, L.: An efficient problem-independent hardware implementation of genetic algorithms. *Neurocomputing* **71**(1), 88–94 (2007)
22. Noraini, M.R., Geraghty, J.: Genetic algorithm performance with different selection strategies in solving tsp. *World Congress on Engineering 2011 Vol II* (2011)
23. Oliveira, T.C., Júnior, V.P.: An implementation of compact genetic algorithm on fpga for extrinsic evolvable hardware. In: *Programmable Logic, 2008 4th Southern Conference on*, pp. 187–190. IEEE (2008)
24. Peker, M.: A fully customizable hardware implementation for general purpose genetic algorithms. *Applied Soft Computing* **62**, 1066–1076 (2018)
25. Qu, H., Xing, K., Alexander, T.: An improved genetic algorithm with co-evolutionary strategy for global path planning of multiple mobile robots. *Neurocomputing* **120**, 509–517 (2013)
26. Rodriguez, A., Moreno, F.: Evolutionary computing and particle filtering: A hardware-based motion estimation system. *IEEE Transactions on Computers* **64**(11), 3140–3152 (2015)
27. Scott, S.D., Samal, A., Seth, S.: Hga: A hardware-based genetic algorithm. In: *Third International ACM Symposium on Field-Programmable Gate Arrays*, pp. 53–59 (1995). DOI 10.1109/FPGA.1995.241945
28. Sehatbakhsh, N., Aliasgari, M., Fakhraie, S.M.: Fpga implementation of genetic algorithm for dynamic filter-bank-based multicarrier systems. In: *Design & Technology of Integrated Systems in Nanoscale Era (dtis), 2013 8th International Conference on*, pp. 72–77. IEEE (2013)
29. Tirumalai, V., Ricks, K.G., Woodbury, K.A.: Using parallelization and hardware concurrency to improve the performance of a genetic algorithm. *Concurrency and Computation: Practice and Experience* **19**(4), 443–462 (2007)
30. Uyemura, J.P.: *Introduction to VLSI circuits and systems*. Wiley India (2002)
31. Vavouras, M., Papadimitriou, K., Papaefstathiou, I.: High-speed fpga-based implementations of a genetic algorithm. In: *Systems, Architectures, Modeling, and Simulation, 2009. SAMOS'09. International Symposium on*, pp. 9–16. IEEE (2009)
32. Yan-cong, Z., Jun-hua, G., Yong-feng, D., Huan-ping, H.: Implementation of genetic algorithm for tsp based on fpga. In: *Control and Decision Conference (CCDC), 2011 Chinese*, pp. 2226–2231. IEEE (2011)
33. Zhu, Z., Mulvaney, D.J., Chouliaras, V.A.: Hardware implementation of a novel genetic algorithm. *Neurocomputing* **71**(1), 95–106 (2007)