

**Extended Compact Genetic Algorithm in C++:  
Version 1.1**

**Fernando G. Lobo, Kumara Sastry, and  
Georges R. Harik**

**IlliGAL Report No. 2006012  
March, 2006**

Illinois Genetic Algorithms Laboratory  
University of Illinois at Urbana-Champaign  
117 Transportation Building  
104 S. Mathews Avenue Urbana, IL 61801  
Office: (217) 333-2346  
Fax: (217) 244-5705

# Extended Compact Genetic Algorithm in C++: Version 1.1

Fernando G. Lobo<sup>1</sup>, Kumara Sastry<sup>2</sup>, Georges R. Harik<sup>2</sup>

<sup>1</sup> DEEI-FCT, University of Algarve  
Campus de Gambelas, 8000-117 Faro, Portugal

<sup>2</sup>Illinois Genetic Algorithms Laboratory (IlliGAL)  
Department of Industrial and Enterprise Systems Engineering  
University of Illinois at Urbana-Champaign, Urbana IL 61801

March 26, 2006

## Abstract

This report provides documentation for version 1.1 of the extended compact genetic algorithm (ECGA). Version 1.1 uses Mersenne Twister for the pseudo random number generator and is compliant with GCC 3.4 and 4 series.

## 1 Introduction

In this report we briefly describe how to download compile and run the extended compact genetic algorithm (ECGA) described in Harik's paper (Harik, 1999). It also explains how to modify the objective function that comes with the distribution of the code. The source is written in C++ but a knowledge of the C programming language is sufficient to modify the objective function so that you can try the ECGA on your own problems.

In version 1.1, we replace the pseudo random number generator of version 1.0 (Lobo & Harik, 1999) with Mersenne Twister (Matsumoto & Nishimura, 1998). We also make the code compliant with GCC 3.4 and 4 series.

## 2 How to download the code?

The code is available from `ftp://ftp-illigal.ge.uiuc.edu/pub/src/ECGA/ECGA_1.1.tgz`. After downloading it, uncompress and untar the file by typing

```
tar zxvf ECGA_1.1.tgz
```

At this point you should have in your directory the following files:

DISCLAIMER	chromosome.cpp	inputfile	objfunc.cpp	random.hpp
Makefile	chromosome.hpp	intlist.cpp	objfunc.hpp	subset.cpp
README	ecga.cpp	intlist.hpp	parameter.hpp	subset.hpp
cache.cpp	ecga.hpp	main.cpp	population.cpp	utility.cpp
cache.hpp	gene.cpp	mpm.cpp	population.hpp	utility.hpp
gene.hpp	mpm.hpp	random.cpp		

### 3 How to compile the code?

We assume you have a C++ compiler properly installed on your computer. We have compiled the code using GNU C++ and tested under the Linux operating systems. For windows operating systems, we have used the GNU C++ compiler included in MinGW (<http://www.mingw.org>). To compile the code, type `make` on the UNIX shell prompt. After that, you should have an executable file called `ecga`.

### 4 How to run the code?

The executable `ecga` needs two arguments: the name for an input file and the name for an output file. The ECGA reads its parameters from the input file and stores the results of the run in the output file. A sample input file is provided as an example with the distribution of this code. The filename is called `inputfile` and its contents (along with line numbers) is shown below:

---

```
1  #
2  # Sample parameter file.
3  # Don't change the order of the lines in this file.
4  #
5  BEGIN
6  chromosome_length      40
7  seed                   0.254534
8  population_size        1000
9  probability_crossover   1
10 tournament_size        16
11 learn_MPM               on
12 stop_criteria            allele_convergence
13 stop_criteria_argument   0
14 #
15 # reporting flags
16 #
17 report_pop              off
18 report_string            on
19 report_fitness           on
20 report_MPM               on
21 END
```

---

The `ecga` skips all the lines until it reaches the word `BEGIN` (line 5 in the example above). Then it starts reading the parameters in a predefined order. The program doesn't do any fancy parsing on the input file. This means that after the word `BEGIN`, **you should not change the order of the lines in the input file** because otherwise the `ecga` will get totally confused. The input file is straightforward to understand. Below is an explanation of each of its lines:

**Line 6** indicates that the problem length (# of bits) is **40**.

**Line 7** indicates that **0.254534** is the seed to initialize the pseudo random number generator. **The value for the seed must be a number between 0 and 1.**

**Line 8** indicates that the population size is **1000**. See (Goldberg, Deb, & Clark, 1992; Harik, Cantú-Paz, Goldberg, & Miller, 1999; Pelikan, Sastry, & Goldberg, 2003; Sastry & Goldberg, 2004) for guidelines for setting population sizing in ECGA.

**Line 9** says that the probability of crossover is **1**. That is, the whole population is regenerated after each generation cycle.

**Line 10** indicates that the tournament size is **16**. The only selection method implemented is the tournament selection without replacement (Goldberg, Korb, & Deb, 1989; Sastry & Goldberg, 2001).

**Line 11** indicates that the ECGA learns the marginal product model (MPM) every generation. You can set this option **on** and **off**. If set to **on**, you get the normal ECGA. If set to **off**, you get the compact GA (Harik, Lobo, & Goldberg, 1998).

**Line 12** indicates that the ECGA stops when the population has fully converged. That is, when the population consists of  $n$  copies of the same individual, where  $n$  is the population size. Besides the **allele\_convergence**, you can also choose the **max\_generations** option.

**Line 13** indicates the maximum number of generations in case you choose the **max\_generations** option on the previous line (Line 12). If you choose the **allele\_convergence** option, then this parameter is irrelevant.

**Line 17** indicates that the population should not be stored in the output file at the end of each generation. You can set this option **on** or **off**. If set to **on**, you should be careful as the output file size can easily become quite large.

**Line 18** indicates that the best chromosome of every generation is stored in the output file. you can set this option **on** or **off**.

**Line 19** indicates that the best fitness and the average fitness of the population at the end of each generation is stored in the output file. you can set this option **on** or **off**.

**Line 20** indicates that the MPM—including the greedy search steps taken in the construction of the MPM—for each generation is stored in the output file. you can set this option **on** or **off**.

Now you are ready to go ahead and run the ECGA. At the prompt, type

```
ecga inputfile outputfile
```

Population statistics are displayed on the screen at the end of each generation. The same information is also sent to the **outputfile**. In addition, the **outputfile** also shows the different MPM structures that the ECGA finds during its MPM search.

The objective function that comes with the distribution of the code is the same test problem described in the original ECGA report (Harik, 1999). The test problem is a concatenation of 10 copies of a 4-bit trap function. The trap function has fitness  $u$  if the string has  $u$  ones, except when the string is 0000, in which case the fitness is 5. Thus, for each of the 10 blocks, the global optima is at 0000, each with fitness 5, and the local (deceptive) optima is at 1111, and has a fitness of 4. Therefore, for the overall problem, the optimal solution is the string with all zeros and has a fitness of 50.

## 5 How to plug-in your own objective function?

The code for the objective function is in the function `objective_func()` in the file `objfunc.cpp`. This is the only function that you need to rewrite in order to try your own problem. The function header is as follows:

```
double objective_func(char *chrom, int lchrom)
```

It takes as argument a character string of 1s and 0s, and the string length. The function returns a real number: the objective function of the string. In the current implementation, the `ecga` assumes that the problem is binary coded.

## 6 About the C++ code

The implementation of the ECGA doesn't use advanced features of the C++ language such as templates and inheritance. This means that you don't need to be a C++ expert in order to modify the code. In fact, you can modify the code and plug-in your own objective function using the C programming language alone. Next, we give brief description of the source files. Each `.cpp` file has a corresponding `.hpp` file, except `main.cpp`. The `.hpp` files are the header files and contain the definitions of the various classes. The `.cpp` files contain the actual implementation.

`gene.cpp` contains the implementation of the class `gene`. A gene has a locus and an allele.

`chromosome.cpp` contains the implementation of the class `chromosome`. A chromosome is an array of genes.

`population.cpp` contains the implementation of the class `population`. A population is an array of chromosomes. Selection operators, population statistics, and stopping criteria are implemented here.

`objfunc.cpp` contains the code for the objective function. If you want to try the ECGA on your own problem, you should modify the function `objective_func()` contained in this file.

`utility.cpp` contains utility functions and procedures.

`intlist.cpp` implements a list of integers.

`subset.cpp` contains operations that can be done on a subset structure of an MPM.

`mpm.cpp` contains operations that can be done on an MPM.

`cache.cpp` implements a cache used for speedup-up the MPM search.

`random.cpp` contains subroutines related to the pseudo random number generator.

`ecga.cpp` contains the main loop of the ECGA.

`main.cpp` contains the `main()` function and the initialization procedures.

## 7 Disclaimer

This code is distributed for academic purposes only. It has no warranty implied or given, and the authors assume no liability for damage resulting from its use or misuse. If you have any comments or find any bugs, please send an email to `kumara@illigal.ge.uiuc.edu`.

## 8 Commercial use

For the commercial use of this code please contact Prof. David E. Goldberg at `deg@uiuc.edu`

## Acknowledgments

This work was also sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant FA9550-06-1-0096, the National Science Foundation under grant ITR grant DMR-03-25939 at the Materials Computation Center. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the National Science Foundation, or the U.S. Government.

We thank Prof. David E. Goldberg for encouraging us to write this report.

## References

- Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6, 333–362. (Also IlliGAL Report No. 91010).
- Goldberg, D. E., Korb, B., & Deb, K. (1989). Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, 3(5), 493–530. (Also IlliGAL Report No. 89003).
- Harik, G. (1999, January). *Linkage learning via probabilistic modeling in the ECGA* (IlliGAL Report No. 99010). Urbana, IL: University of Illinois at Urbana-Champaign.
- Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1999). The gambler’s ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, 7(3), 231–253. (Also IlliGAL Report No. 96004).
- Harik, G., Lobo, F., & Goldberg, D. E. (1998). The compact genetic algorithm. *Proceedings of the IEEE International Conference on Evolutionary Computation*, 523–528. (Also IlliGAL Report No. 97006).
- Lobo, F. G., & Harik, G. R. (1999, June). *Extended compact genetic algorithm in C++* (IlliGAL Report No. 99016). Urbana, IL: University of Illinois at Urbana-Champaign.
- Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1), 3–30. <http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/emt.html>.
- Pelikan, M., Sastry, K., & Goldberg, D. E. (2003). Scalability of the Bayesian optimization algorithm. *International Journal of Approximate Reasoning*, 31(3), 221–258. (Also IlliGAL Report No. 2001029).

- Sastry, K., & Goldberg, D. E. (2001). Modeling tournament selection with replacement using apparent added noise. *Intelligent Engineering Systems Through Artificial Neural Networks*, 11, 129–134. (Also IlliGAL Report No. 2001014).
- Sastry, K., & Goldberg, D. E. (2004). Designing competent mutation operators via probabilistic model building of neighborhoods. *Proceedings of the Genetic and Evolutionary Computation Conference*, 2, 114–125. Also IlliGAL Report No. 2004006.