# Time and Statistical Complexity of Proposed Evolutionary Algorithm in Artificial Neural Networks

**G. Y. Sagar[1], R. Sudheer Babu[2], K. Siva Rama Krishna[3], Bhallamudi Ravi Krishna[4], G. V. R. Sagar[5]**
[1]Professor, Department of Statistics, College of Natural and Computational Sciences, Gambella University, Ethiopia.
[2, 5]Associate Professor, Department of Electronics and Communication Engineering, G. Pulla Reddy Engineering College (Autonomous), Kurnool, A.P., India
[3]Associate Professor, Department of Computer Science Engineering (AIML), Geethanjali College of Engineering & Tech, Hyderabad, T.S., India
[4]Associate Professor, Department of Artificial Intelligence and Data Science, Vignan Institute of Technology and Science, Hyderabad, T.S., India

**Corresponding Author:**
sagar.ece@gprec.ac.in

**Abstract**
The important issue in Evolutionary Algorithms (EAs) analysis, is time-complexity. Here to obtain the mean hitting time of EA the concept of take-overtime is considered. The time complexity of the EA such as the takeover time is considered, i.e. the concept of the takeover time is generalized rather than a selection of operator alone. This generalization is applied to benchmark problems like N-Bit parity. For various input sizes N, the time complexity in terms of number of generations is estimated. An empirical model is also generated for proposed EA using statistical tool.

**Keywords:** Evolutionary Algorithms; Take-Over Time; Wide-Gap Problem; ANOVA.

## 1. Introduction

Evolutionary algorithms (EAs) are adaptive search algorithms. In general solving the a few EA optimization problems are very hard. In that cases, wide-gap technique is used to avoid the long gap between generations i.e. exponential generations to find the global optima of EA. But it is very hard to get the solution for hard problems which can be solved by a proper acceptable selection pressure and carefully attention on mutation etc. The other part of the work emphasis on adapting the selection pressure [1], [2] for wide-gap problem that utilizes the mean first hitting time of the EA. There are two methods of selection pressures are considered, the first is a truncation selection which is taken selection I and the second is a tournament selection named as selection II.

## 2 Selection

On completion of crossover and mutation, form the new population of parent and off-spring and assigned the survival probabilities to each individuals in the population $\Omega t$ [3]. Then, selected the few individuals based on the fitness and their probability for the next generation $\Omega t_{+1}$. These two selection schemes are selected in the analysis of time complexity of the EA.

### 2.1. Truncation selection

2N individuals are formed on combination of Parents and offspring. These are based on their fitness in descending manner. Then these N individuals are selected to the next generation.

### 2.2. Tournament selection

In this methodology, p individuals are grouped and make them arranged ‘r’ number of groups from 2N individuals (both parent and off-spring population).Select the two individuals from each group and ‘r’ number of tournament are arranged from which the best one is selected based on Hamming distance method. More details given in [8]. The Hamming distance is measured based on mean square error between two individuals X and Y and the distance between the neurons is given by the following equation

$$H(X,Y) = \sum_{i=1}^{n} |S_i - S_i'|$$

Where X = ($s_1$, $s_2$,$s'_n$)Y = ($s'_1$,$s'_2$,…..$s'_n$)1.1
Therefore for a given maximum fitness f, an individual x* = ($s^*_1$,.$s^*_n$) is an optimum if and only if, for all x that satisfy H(x, x*) = 1, f(x*) > f(x) holds.
Therefore, the probability of individuals from the ‘r’ number of groups is given as

$$P(a_1') = 1/r^p(((r-1)+1)^2 + (r-1)^2)$$

The last offspring or $r^{th}$ group gives the optimum offspring and is the best solution.

These two selection criterions are used with mutation in EA. To compare these selection methods for estimation of the time complexity of EA using Take over Time the mean first hitting time is estimated which is illustrated in the next section.

## 1. Take over time

The take over time of a desired selection method is the number of generations/offspring's needed for the individuals and to fill the population under pure selection operator. Let us select an iterated selection process which is initialized by any selection process on population (P). In each offspring, the selection method selects a number of individuals from the existing population (P) to form the new population. The population in the next generation has the same size (P). This proceeds for the 't' number of generations of EA. The algorithm is shown in Fig. (1).

- Initialize the subspace with population p
- At $t^{th}$ generation apply the mutation operator by adding random number generated by Gaussian distribution with probability of $p_x$ of the individual $x_t$.
- Then an offspring individual $x_t^m$ is obtained.
- Evaluation the fitness of $x_t^m$
- If $f(x_t^m) \geq f(x_t)$, then
- Set $x_{t+1} = x_t^m$ else
    - Set $x_{t+1} = x_t$
    - Set t = t+1
- Repeat the procedure until some stopping criterion is

**Fig.1:** Algorithm for Takeover Time.

The performance of an EA on a problem can be measured by first hitting time of the EA and its expectation is called the take over time.

### 3.1. First hitting time

The crossovers on EA is given as L:{Lt, t=0, 1,….} (Lt ∈ Ωm) and a subspace $Q_j$ (∈ Ω) is obtained by decomposing or partition the population (P) then the first hitting time ' to' of Q is defined as

$$\tau = \min\{t \geq 0; L_\tau \in Q\} \quad (1)$$

Let Q is the subset of populations for an EA whose elements all contain the best individuals or global optimum x* [5], [6], [7]. Therefore, the first hitting time of EA is given as

$$\tau = \min\{t \geq 0; x^* \in Q_\tau\}$$
(2)

Where x* is the best global individuals and $Q_t$ is the population of the EA at the $t^{th}$ generation. Therefore, the estimation of the first hitting time is called the takeover time and is given as

$$\bar{\tau} = E[min\{t \geq 0; x^* \in \Omega_\tau\} \quad (3)$$

Consider the input size (m) of N-bit Parity problem for subset sum problem to estimate the take over time.

### 3.2. Wide-gap problem

It is the mean first hitting time and exponential function of the problem size then the resulting problem is called as wide-gape problem [7]. Let a problem with two subspaces $Q_j$ and $Q_{j+1}$ is considered such that the mean first hitting times of the EA starting from any $a \in Q_j \; and \; b \in Q_{j+1}$ satisfy that $E[\tau|x^0 = \alpha] - E[\tau|x^0 = \rho]$ is an exponential function having a size N then the given problem is a wide-gap problem.
In a finite search space Ω and a fitness function f takes a limited number of values, then arrange the function f is arranged in descending order $f_{max} = f_o > f_1 > \ldots\ldots > f_l = \min$. By the value of f, the whole space is subdivided into l+1 subspace.

$$\forall i \in \{0,1, \ldots\ldots l\}: Q_i = \{x \in \Omega; f(x) = f_i\}$$
(4)

Now if for any problem , there are two subspaces $Q_j$ and $Q_{j+1}$ such that the first hitting time of the EA with Gaussian mutation starting from any $a \in Q_j \; and \; b \in Q_{j+1}$ satisfies that, the number of generations is an exponential to the size m of the problem, then problem is set to be a wide-gap problem.

#### 3.2.1. Algorithm for wide-gap problem

Validation of a wide-gap problem is done as follows

- Initialize the search space of subspace Ωm of the problem, let x be the individual.
- Two subspaces $Q_j$ and $Q_{j+1}$ are selected from decomposed l+1 subspace from Ωm.
- The takeover time of number of generations is calculated by using any mutation and selection process
- The mean first hitting time of the EA starting with the pair of solution from $Q_j$ and $Q_{j+1}$ is estimated.

- The mean difference between the first hitting time of subspaces $Q_j$ and $Q_{j+1}$ calculated, if this is exponentially large, then it is a wide-gape problem.
- It has a unique global optimum at x* form so and local optimum at x' from $s_1$.

Therefore, the mean first hitting time starting from x* is zero and x' is an exponential. The probability of x*, such that x' mutates to x* by a Gaussian mutation is (1/2m) m and it is exponentially nearer to zero. Hence, the population has only feasible solutions through this evolutionary process and the total set of the feasible populations is represented by E. This population is again sub divided into some subsets.

## 2. Comparative analysis of EA on two selection problems

In this section, first a combination of Gaussian mutation and truncation selection is applied on the instance of subsets of EA to get the mean first hitting time. This subset contains the single layer feed-forward networks. Let the initial population be considered as $P_o \in E_k(k = 0,1, \dots, m-1)$, when apply the mutation and truncation selection on EA, it shows that, this selection process always retain the worst individual and in one generation the probability of reaching $E_i(i = 0,1, \dots, k-1)$ is not smaller than K/m. Therefore, the mean first hitting time from $E_k$ to $E_o$, gives a logarithmic response in the order of $mln(m) + l$ where l is the ratio in terms of problem size and population.

Now, the second combination as mutation and tournament selection is applied on the instances of subsets of the problem with the same initial population and let p be the transaction probability from state $E_i$ to $E_j$ where (i = 0,1,.. m-1). The EA with this selection process apparently exhibits an exponential behavior with approximately same mean first hitting time and is in the order of $mlnln(m) + l$ where l is the ratio in terms of problem size and population [2] [3].

But the tournament selection process eliminates the problem having sufficiently higher selection pressure and very large population but it fails in lower selection pressure. So, for initial state or beginning (to get the path quickly) lower selection pressure is used i.e. truncation selection and latter or to search along the path a high selection pressure like tournament selection was used. The result of the both the selection pressures of subset problem on pseudo Boolean logic like N-bit parity is discussed.

### 4.1. Computational time complexity of proposed evolutionary algorithm based on selection methods

Some optimization problems are very difficult to get the solutions using the EAs and they take very long duration to reach the global optima. In the above discussion take-over time is expressed in terms of generation. A correct selection pressure is opted with mutation operator, on N-Bit Parity problem and to obtain the first hitting time of EA. Here, at the beginning a selection I (truncation selection) with mutation is used and the mean first hitting time is generated and this result are almost approaches to the order of $mln(m) + l$ . In the truncation selection, it hold back the best and worst individuals in the union population. Secondly, tournament selection with mutation operator is applied and the mean first hitting time is approaches to the order of $mlnln(m) + l$ . In this selection process select the seed individuals from the best or worst individual and copies this seed individual P times and to fill the population of the next generation. So in this selection method apparently exhibits an exponential behavior. The tournament selection problem solves the problem with sufficiently higher selection pressure and large population but it fails at lower selection pressure. The results are validated using N-Bit parity with m is the number of inputs to the problem and it varies from 4 to 60. The time complexity shows and experimental results in Fig (2) for selection I. This is averaged over 10 independent trial runs and Fig (3) shows the experimental results showing exponential behavior of selection II (tournament selection) and the number of generations is averaged over 10 independent trial runs. For first N = 2 to 22 the experimental results are obtained using Mathlab software and the rest of the results are generated using the Design Expert statistical tool. The analysis and the empirical statistical model are explained in the next section. The entire results are approaches to the theoretical results.

## 3. Analysis of statistical model of proposed EA

In science and Engineering the Statistically designed experiments acts an important role. The experimentation is an application of treatments to experimental units, and then measurement of one or more responses. In an experiment, some inputs (m) transform into an output that has one or more observable response variable 'number of generations'. All the results and conclusions can be obtained by experiments. In order to get an objective conclusion an experimenter has to plan and design the experiment, and make analyze the results. There are plenty types of experiments in real-world situations and problems. A statistical model is developed for proposed EA is using observation and experimental results. This analysis gives the mean first hitting time for any input size of the problem. The statistical analysis is obtained using Design Expert (version-8) statistical software tool.

### 5.1. Analysis of variance (ANOVA)

Statistical analysis (regression and ANOVA analysis) of the responses are carried out to estimate the coefficients of the polynomial equation of the response by regression and to check the significance of the regression coefficients of independent variables and interaction variables by ANOVA [9] [10]. Analysis of variance (ANOVA) table is used to determine the significance of the first degree, second degree, and cross-product terms of the polynomial. In this case the adequacy of the model is confirmed when the model probability "prob > F" is less than 0.05. Regression is a procedure which selects, from a certain class of functions, the one which best fits a given set of empirical data generated from the time complexity results. The class selected are usually from nonlinear regression and the parameters are called regression parameters. The important task is to get the good estimators of the regression coefficients (with expected values and variances as low as possible), to be used for predicting values of output results when new observations are considered. The significance is a test to determine if there is a linear relationship between the number generations and any of the regressor variables like input size of the N-Bit Parity. This procedure is often thought of as an overall or global test of model adequacy.

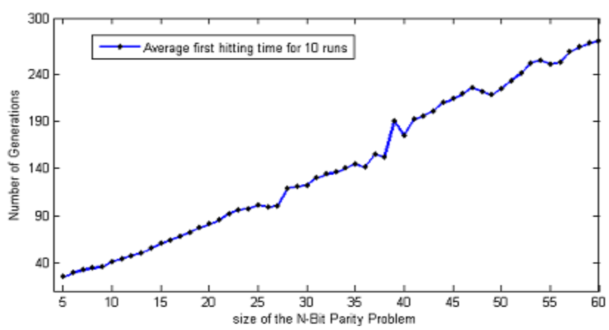


**Fig. 2:** The Average First Hitting Time of EA with Selection I.

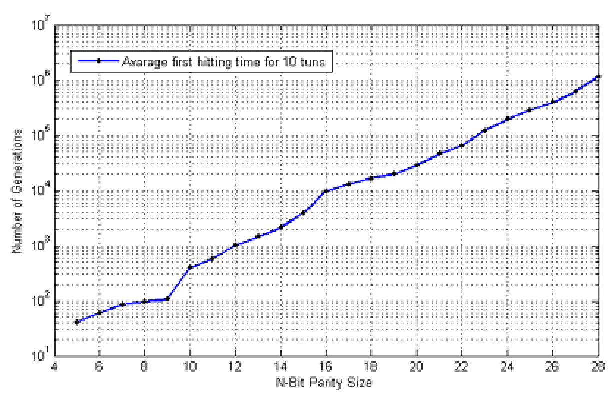


**Fig. 3:** The Average First Hitting Time of EA with Selection II.

In statistics, the number of degrees of freedom (df) is the number of values in the final calculation of a statistic that are free to vary. F-test is mostly used to compare the models that have been fitted to data set and to identify the statistical model that best fits the population from which the data is sampled.

In statistical significance testing, the p-value represents the probability of obtaining a test statistic at least as extreme as one that was actually observed, assuming that the null hypothesis is true. A closely related expectation value with average number of times in multiple testing, one expects to obtain a test statistic as minimum as extreme as the one that was originally observed. Assuming that the null hypothesis is true. When the tests are statistically independent, the product of the number of tests is E-value and the p-value. The lower the p-value, the less likely the result is, if the null hypothesis is true, and consequently the more 'significant' the result is in the sense of statistical significance. One often accepts the alternative hypothesis, if the p-value is less than 0.05 or 0.01, corresponding respectively, to a 5% or 1% chance of rejecting the null hypothesis when it is true.

### 5.2. Empirical analysis of time complexity results of EA

The future mean first hitting time for any input size N of problem can be predicted by this empirical model. Here, Analysis of variance (ANOVA) test [9] [10] is performed on experimental results (number of generations) of N Bit Parity problem for different input sizes. For selection I the results are shown in Fig (4) and Fig (5) for selection II. The top dashed line signifies the maximum allowable variance, bottom dashed line gives the minimum possible variance, the thick black line shows the average variance of generations and the red dots represents the experimental values. The ANOVA process results in a model in terms of input size (N) and is given by equations (5) and (6). Based on these results, for any input size of the problem, mean first hitting time (□) in terms of generations are easily obtained. For selection I the empirical model for mean first hitting time (□) is

$$\tau = 9.31665E + 005 - 5.05079E + 005 * N + 1.06031E +005 * N^2 - 11096.74810 * N^2 + 616.91088 * N^4 -17.46219 * N^3 + 0.19914 * N^6$$

(5)

For selection II the model for mean first hitting time (□) is

$$\tau = 77.66326 - 5.84296 * N + 0.42205 * N^2 - 4.50359E - 003 * N^2 \quad (6)$$

Could occur due to noise. Values of "Prob > F" less than 0.0500 indicate that model terms are significant. Values greater than 0.1000 indicate that the model terms are not significant. In this case A, $A^2$, $A^5$, $A^6$ are significant model terms. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model. The "Lack of Fit F-value" of 0.39 implies the Lack of Fit is not significant relative to the pure error. There is an 86.90% chance that a "Lack of Fit F-value" this large could occur due to noise. On-significant lack of fit is good so we want the model to fit.

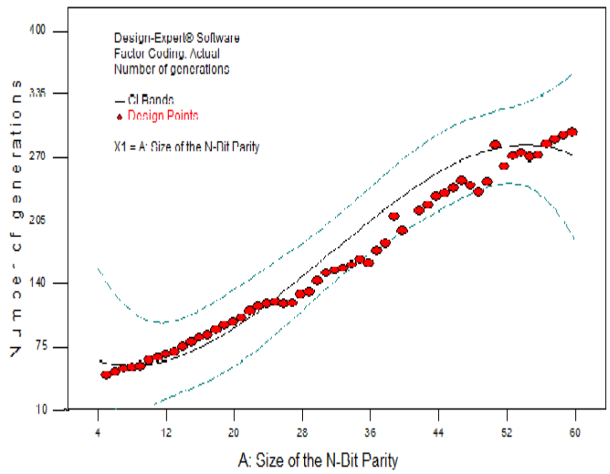


**Fig. 4:** ANOVA Result of Selection I (Truncation).

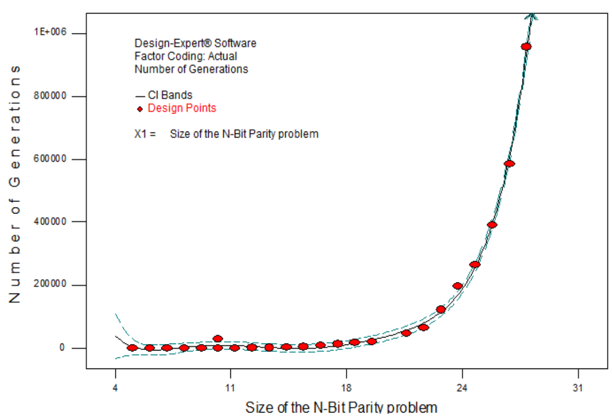


**Fig. 5:** ANOVA Result of Selection II (Truncation).

**Table 1:** ANOVA Result of Selection I

| Analysis of variance table [Partial sum of squares - Type III] | | | | | | |
|---|---|---|---|---|---|---|
| Source | Sum of Squares | Degree of freedom (df) | Mean Square | F Value | p-value Prob > F | |
| Model | 409312.8 | 3 | 136437.6 | 20.18415 | < 0.0001 | significant |
| A-Size of the N-Bit Parity | 129510.7 | 1 | 129510.7 | 19.15941 | < 0.0001 | |
| A^2 | 257.4605 | 1 | 257.4605 | 0.038088 | 0.8460 | |
| A^3 | 12454.77 | 1 | 12454.77 | 1.84252 | 0.1805 | |
| Residual | 351501.2 | 52 | 6759.639 | | | |
| Cor Total | 760814 | 55 | | | | |

**Table 2:** ANOVA Result of Selection II

| Analysis of variance table [Partial sum of squares - Type III] | | | | | | |
|---|---|---|---|---|---|---|
| Source | Sum of Squares | Degree of freedom (df) | Mean Square | F Value | p-value Prob > F | |
| Model | 1.23E+12 | 6 | 2.05E+11 | 1220.845 | < 0.0001 | significant |
| A-Size of the N-Bit Parity problem | 3.54E+09 | 1 | 3.54E+09 | 21.06151 | 0.0003 | |
| A^2 | 1.49E+09 | 1 | 1.49E+09 | 8.864329 | 0.0085 | |
| A^3 | 28578189 | 1 | 28578189 | 0.17009 | 0.6852 | |
| A^4 | 4.27E+08 | 1 | 4.27E+08 | 2.544031 | 0.1291 | |
| A^5 | 9.32E+09 | 1 | 9.32E+09 | 55.47513 | < 0.0001 | |
| A^6 | 2.73E+09 | 1 | 2.73E+09 | 16.25983 | 0.0009 | |
| Residual | 2.86E+09 | 17 | 1.68E+08 | | | |
| Lack of Fit | 2.47E+09 | 16 | 1.54E+08 | 0.39457 | 0.869 | Not significant |
| Cor Total | 1.23E+12 | 23 | | | | |

## 4. Conclusion

For the proposed EA the time complexity analysis is verified that for N bit parity wide-gap problem has only few optima's or solutions are present. It is advised to use the lower selection pressure at the staring of the evolution so that the path can be found quickly and so truncation selection is used. For bigger or huge populations, the number of generations is exponential and to searching along the path, so tournament selection is used. The Analysis of variance (ANOVA) test on time complexity (generations) analysis is used and a statistical model for selection I and II are generated. The significance of the models are evaluated for real world problems based on F-value, p-test and degree of freedom. The proposed EA on both connection weights and architectural behavior is significant.

## References

[1] Chen et al, "New Approach for analyzing average time complexity of EAs on Unimodal problems", IEEE Transactions on systems, Man, and Cybernetics – Part B, Cybernetics, Vol.39, No.5, October – 2009.

[2] He, X. Yao, Drift analysis and average time complexity of evolutionary algorithms, Artif. Intell. 127 (1) (2001) 57_85.https://doi.org/10.1016/S0004-3702(01)00058-3.

[3] J. Garnier and L. Kallel, "Statistical distribution of the convergence time of evolutionary algorithms for long path problems," *IEEE Trans. Evol. Comput.*, vol. 4, no. 1, pp. 16–30, Apr. 2000.https://doi.org/10.1109/4235.843492.

[4] Tianshi Chen et al., "Choosing selection pressure for wide-gap problems, Theoretical Computer Science", 411 (2010) 926_934https://doi.org/10.1016/j.tcs.2009.12.014.

[5] J. He, X. Yao, Drift analysis and average time complexity of evolutionary algorithms, Artif. Intell. 127 (1) (2001) 57_85.https://doi.org/10.1016/S0004-3702(01)00058-3.

[6] S. Khuri, T. Bäck, J. Heitkotter, An evolutionary approach to combinatorial optimization problems, in: D. Cizmar (Ed.), Proc. 22nd Ann. ACM Comput. Sci. Conf., ACM Press, New York, 1994, pp. 66_73.https://doi.org/10.1145/197530.197558.

[7] J. He, C. Reeves, X. Yao, A discussion on posterior and prior measures of problem diffciulties, in: Proc. PPSN IX Workshop on Evolutionary Algorithms Bridging Theory and Practice, 2006.

[8] T. Jansen, K.A.D. Jong, I. Wegener, On the choice of the offspring population size in evolutionary algorithms, Evol. Comput. 13 (4) (2005) 413_440.https://doi.org/10.1162/106365605774666921.

[9] Davison, A. C., 2003. Statistical Models. New York: Cambridge University Press.https://doi.org/10.1017/CBO9780511815850.

[10]Rice, John A., 1995. Mathematical Statistics and Data Analysis. Belmont: Duxbury Press.