Abstract
Genetic Algorithm (GA), a stochastic optimization technique, doesn’t ensure optimal solution every time. Nowadays there is a need to improve the performance of each and every application so that the time required for obtaining quality solution can be minimized. This paper gives a brief overview of theoretical advances and computing trends, particularly population diversity in PGA (Parallel GA) and provides information about how various authors, researchers, scientists have parallelized GA over various parallel computing paradigms viz., Cluster, MPP (Massively Parallel Processing), GPGPU (General purpose Graphics Processing Units), Grid, Cloud, Multicore/HPC to ensure more optimal solution every time with efficacy and efficiency.

Keywords:
Genetic Algorithm (GA), Parallel GA (PGA), General Purpose Graphics Processing Unit (GPGPU), Massively Parallel Processor (MPP), Population Diversity, Cloud, Grid, Cluster, HPC

1. INTRODUCTION

The GA is one of the most important soft computing tools used for solving many optimization problems. Being stochastic (soft computing tool) in nature, it is non-deterministic process. Therefore solution evolved using GA may or may not be optimized i.e. it doesn’t ensure optimal solution every time. This is called as premature convergence problem (converge too early giving suboptimal solution). Besides this there are various problems associated with GA viz. diversity problem (either increase population size or have more than one population to get more diversified search space to find global optimum solution), population size problem (increase in population size may deteriorate the performance of GA), curse of dimensionality problem (increases in dimension of search space deteriorates performance of GA).

Also, there are many real-life problems which when run on serial machines may take days or weeks of computing time to solve. Although, by using finite number of computing resources in parallel, one cannot lower the intrinsic time complexity of a problem but parallelism often reduces time to reasonable levels. This is useful (or important) in industrial or commercial setting where the time to get the solution is critical for decision making.

To address these problems efficiently, one can take advantage of parallel computing environment viz., Cluster, MPP, GPGPU, Grid, Cloud, Multicore or HPC and can exploit their functionality to solve GA effectively with more speed up. Apart from this diversity - where extra population (called as reserve population) is used to provide additional diversity to the main population through crossbreeding - also helps to get the optimal solution efficiently.

This paper provides brief literature review of implementation of PGA on various parallel computing paradigms. In the remnant of this paper we briefly introduce various parallel computing paradigms available. Thereafter we discuss recent work, with relevance to diversity in PGA, carried out by various researchers over different parallel computing paradigm. Finally, we discuss the finding of this review work.

2. PGA OVER COMPUTING PARADIGMS

Literature survey given in [1], [2] are on Advances, Computing Trends, application and Perspective of Parallel Genetic Algorithm (PGA) [3]. In computing trends the important issues are architecture of computing paradigm, OS, topologies and programming language, facilitated with set of special system calls or libraries like - Linda, OpenMP\(^a\), OpenCL\(^d\), Express MPI, PVM (Parallel Virtual Machine), POSIX threads, C sharp, CUDA\(^e\) (Compute Unified Device Architecture (CUDA)), Brooks\(^c\), OpenCL\(^e\) (Open Computing Language), OpenGL\(^f\) (Open Graphics Library) and Java threads on SMP (Symmetric multiprocessing) machines, available to exploit the functionality. Some parallel programming options are available in footnote\(^g\).

The performance of GA is optimized by parallelizing it. The Way in which GAs/PGAs are parallelized depends upon following parameters [4]:

- Population Initialization
- Single or multiple subpopulation(s)
- Migration policy, migration rate
- Method of selection
- Fitness evaluation
- Crossover
- Mutation
- Survival selection method

e. https://developer.nvidia.com/opencl
f. https://www.opengl.org/
g. http://wotug.ukc.ac.uk/parallel/
h. http://www.openmp.org
Depending on how each of these elements is implemented, there are different methods of PGAs viz. Master-Slave parallelization (distributed fitness evaluation)- with two versions a) Synchronous b) Asynchronous, Static subpopulations with migration, Static overlapping subpopulations (without migration), Massively parallel GAs, Dynamic demes (dynamic overlapping subpopulations). Parallel steady-state GAs, Parallel messy GAs, Hybrid methods (e.g. static subpopulations with migration, with distributed fitness evaluation within each subpopulation).

The computing paradigms considered for survey are Cluster, MPP, Grid, GPGPU, cloud computing and Multicore/ HPC.

2.1 GA OVER CLUSTER AND MPP

Cluster (a set of loosely connected or tightly connected computers) is generally deployed for improving the performance and availability over that of a single computer. It is much more cost-effective than single computers of comparable speed or availability.

Most of the existing parallel implementations of PGAs use either cluster or MPP (large number of processors (or computers) to perform a set of coordinated computations in parallel.) for execution. A communication delay should be taken into consideration while implementing GA on cluster and MPP. Most of the worlds supercomputers are clusters and rest are MPPs [54]. MPI is very commonly used library for Clusters and MPPs environments. Cluster has various disadvantages over grid such as centralized control, less security, no direct access to distributed heterogenous resources and complex access to remote data sources. The advantage of Cluster and MPPs is, it consist of homogenous nodes, hence parallel algorithms can be easy to implement on them.

Most famous model for PGAs running over clusters is the island model [5], [6]. The [7] [8] and [9] reports the PGA implemented on cluster.

2.2 GA OVER GRID

Grid (composed of many loosely coupled computers networked together to perform large tasks) is important and quickly evolved parallel paradigms. It is widely accepted distributed form of computing [10], [11], [12], [13], [14]. MPI is very commonly used library for grid environment. It is more loosely coupled, heterogeneous, and geographically dispersed form of distributed computing than cluster. Cluster has many advantages over grid such as no centralized control, less security, easy access to distributed heterogeneous resources, and easy and reliable access to remote data sources and service to any available application.

GA on grid is called as Grid Oriented Genetic Algorithm (GOGA). This term was first introduced by [15], [16]. Coarse grained, medium grained and fine grained GAs is implemented on grid.

2.3 GA OVER GPGPU

Today, the cheapest and highly available computing power is GPU. Many PGAs are implemented on GPGPU. GAs can be parallelized over GPGPU by parallelizing the genetic operators such as random number generator, selection, crossover, mutation. Some implementation guidelines are given in [17], [18], [19]. Some of the parallel hybrid GA [20], steady GA, binary GA, real coded GA [21], cellular GA [22] etc. are implemented over GPGPU. Many GAs / PGAs are yet to be implemented on GPGPU, so researchers have huge scope in this area and can further exploit the functionality of GAs using GPU. Based on the search space complexity of problems, it is possible to provide diversity in search space. For such problems use of GAs / PGAs on GPGPU is not a good option. Many real life problems, those support the SIMD (Single Instruction Multiple Data) (Flynn’s taxonomy) architecture, can be effectively solved over GPGPU. CUDA and OpenCL are programming options available for programming with GPGPU.

Problems solved by PGA on GPGPU are Medical Image Registration [23], Feature Selection, Electrical Circuit Synthesis and Data Mining [24], SAT Problems [25], Function Optimization [26], Benchmark Problems [27], [28], [29], Texture-Rendering [30], One-MAX Problem [31], Quadratic Assignment Problems [32], Non-convex Mixed Integer Non-Linear Programming (MINLP) and Non-convex Non Linear Programming (NLP) Problems [33], Cellular Automata Rules Acceleration [34], Stereo Matching [35], Data Mining[36], Drug discovery [37], Gaming Application [38], Image Matching[39], Gaming-Sudoku Solution [40] , Task Scheduling and Load Balancing Based on Workload and Case Study of Financial Option Pricing Problem [41], Multi-objective evolutionary algorithm Test Suite Minimization [42], Sketching Interface [43], Daily Activity Plans [44] etc.

2.4 GA OVER CLOUD

Today is the fast growing era of computing and supporting software for parallelization. Buying latest configuration PC or high performance system will be absolute in a year or two. Further, it is not possible to invest in all latest computing infrastructures and softwares. Therefore, cloud computing (use of computing resources such as hardware and software, which are delivered as a service over a network) is a good solution for on demand computing power (PaaS- Platform as a service) and on demand software support (SaaS – Software as a Service). The availability of computing infrastructure and software on cloud is quite easily available. Due to computational intensive and parallel nature of PGAs, it is hard to implement and optimize the performance on cloud. Developing PGA on cloud will require knowledge of GA as well as of cloud. Many issues related to cloud like security, bandwidth etc. are still under discussion.

Zhao et al., (2011) implemented PGA on prototype of cloud called Hadoop. Hadoop is an open source Cloud computing. The Function Optimization using PGA on Cloud Computing gives speed-up but with short communication delay and it is suitable to solve large dimension problems. Comparing traditional PGAs on cloud with - HPC, cluster, Grid - PGA on cloud is simple, easy to implement and easy to extend to solve larger-scale problems [45], PGA on cloud will not be effective because of its distributed environment, as performance depends on network delay (Internet bandwidth).

To solve the compute intensive optimization problems without hiring the computing infrastructure, cloud is good option. Some of optimization problems which are solved using PGAs are Resource Scheduling [46], Scheduling HPC...
Applications [47], Task Scheduling [48], [49], [50], Performance Improvement of Cloud Storage [51], Power Management in Cloud [52], Clustering composite SaaS components [53] etc.

2.5 GA OVER MULTICORE AND HPC SYSTEM

Many GAs are implemented on multicore system but do not care about their CPU/core utilization. Nowadays multicore processors are getting cheaper and are easily available. One cannot ignore their importance anymore. Multicore systems have multiple processing cores on same chip while multiprocessor systems have multiple chips inside a system. PGAs can be implemented over multicore architecture using MPI or OpenMP. MPI performs better on the distributed memory architecture while OpenMP performs better on the shared memory architecture [54]. The OpenMP, POSIX threads and Java threads are also the programming option available for programming on Multicore and HPC. PGA’s parallel algorithmic design need to be considered and explored on multicore and HPC computing paradigm. The challenges of GA over HPC facing are fault tolerance, scalability, load balancing, data storage and multicore architectures.

Zheng et al. (2011) explained the PGA with perspective of architectures-multicore and many core systems. In their work, they have shown Architecture-based performance evaluation of GAs on Multi/Many-core Systems. They described PGA on thread organization, memory hierarchy, core utilization, execution time and solution quality. They considered the material-salve, island and cellular models of PGA for solving non linear programming problem [55]. Cristea showed the conception and design of PGA on HPC and performance analysis models [56]. Zhuang et al. proposed PGA for IC chip planning on HPC [57]. Dunlop et al. uses GA on automatic benchmark tuning tool [58].

PGAs are implemented over Clusters, MPPs, Grids, GPGPU, cloud computing, Multicore and HPC. The GAs implemented over distributed environment such as cluster, MPPs, grid and cloud computing is also called PGAs, because of parallel hardware used. The design of algorithm needs to be tuned with the hardware selected for implementation. The parameters to be considered are network delay, system configuration of each node in distributed system. Network delay plays important role on the performance of algorithm in case of distributed system. The programming on such parallel hardware is called loosely coupled programming.

In case of Multicore, HPC and GPGPU, GAs on them are also called PGAs; here also the design of algorithm needs to be tuned with the hardware selected for implementation. While designing PGA on Multicore, HPC and GPGPU, the number of cores can be taken into consideration to maximize the CPU or GPU utilization. The programming on such parallel hardware is called as tightly coupled programming. No parameters like network delay, system configuration of each node need to be considered in tightly coupled system. PGAs have capacity to run in parallel on many computing paradigms but finding suitable hardware and software will give optimal solution with optimal resource utilization.

3. LITERATURE REVIEW

Recently, PGAs have received more attention from the research community due to changing computing infrastructures. Also, PGAs have many interesting unique features like parallelizable operators, ability to search large dimension search space etc. which require in-depth analysis [59]. PGA is intended to provide following benefits [60]:
- Reduces the time to locate a solution
- Reduces the number of function evaluations
- Explores the large populations size over the parallel platforms used for running the algorithms
- Improves the quality of the solutions
- Solves the large scale, large dimensions problems with efficacy and efficiency

The GAs with multiple populations but normal selection, crossover and mutation with exchange of some individuals are also called as PGAs. The migrations rate, migration destination type were important parameters introduced in PGAs [61], [62], [63], [64]. The interdeme / interpopulation communications result in significant improvements in the quality of the solutions, which results in reductions in deme/population size and the computational effort, however with more processors, additional communication is required. Thus there is a tradeoff between increasing communication cost and decreasing computations [65], [66].

Ursem (2000) [67] proposed the Multicannical GA (MGA) which spread different population through different peaks. In the fitness landscape, a hill-valley detection procedure and a migration policy are used. The MGA was very complex and requires more tuning time.


Fernandes, Tavares and Rosa (2000) [69] proposed combined strategy of outbreeding and a varying population size. Fernández (2001) [70] experimented multipopulation parallel genetic programming. Again Fernandes, Tavares and Rosa (2001) proposed the importance of non random mating called as assortative or dissipative mating in GA. They tested algorithms by varying population size scheme on royal road problem.

Rivera (2001) [71] investigated how to implement PGAs for getting quality solutions efficiently. Rivera reviewed the state-of-the-art in PGAs, parallelization strategies, emerging implementations and relevant results.

Alba and Troya (2001) [72] proposed a common framework for studying PGAs on cluster of workstations. It shows linear and even super-linear speedup. Alba and Troya (2002) [73] bring some uniformity in comparison, and knowledge exchanges among the traditionally opposite kinds of serial and parallel GAs. They analyzed the properties of steady-state, generational and cellular GAs and extended the idea to distributed model ring of GA islands.

Giacobini (2003) [74] proposed a theoretical study of the selection pressure in asynchronous cellular (fine-grained) evolutionary algorithms (eEAs).

Xiao and Armstrong (2003) [75] proposed a model of parallel evolutionary algorithms (EAs) called a specialized island model (SIM).

Gagné (2003) [5] says that the classic master slave distribution model was superior to the island-model when
exploiting Beowulf and networks of heterogeneous workstations. They identified the key features of a good computing system for evolutionary computation viz. transparency, robustness and adaptivity.

Yang (2003) [76] proposed Primal-Dual Genetic Algorithm (PDGA) which adopted complimentary and dominance mechanism for diversity. PDGA solved complex dynamic problems efficiently than traditional SGA.

Ochoa and Jaffe (2005) [77] proposed the idea related to Assortative Mating GA (AMGA) and applied it to dynamic optimization problems. Assortative GAs (very similar to AMGA) are used to solve a dynamic knapsack problem. Again Ochoa and Jaffe (2006) [78] proposed a study on the error threshold of replication in GAs with different mating strategies.

Russell (1998) [79] investigated that the Outbreeding increases the diversity. The inbreeding decreases the genetic diversity in a population. The mating of dissimilar (not from one population) individuals, called as dissortative mating (negative assortative mating), does not give the diversity but this type of mating results in highly distinct cluster of similar genotypes. In EAs, selection pressure and genetic diversity are important parameters. EAs need to provide selection pressure and genetic diversity with delicate equilibrium between exploration and exploitation. Therefore nonrandom mating naturally came in EAs in order to provide genetic diversity and avoid premature convergence.

The crossover (recombination or mating) gives exploitation and mutation is responsible for exploration. The choice of exploitation, exploration, selection pressure and genetic diversity are of prime concern depending on the problem characteristics like static and dynamic, unimodal and multimodal, linear and non linear, separable and non separable, Scalable and non scalable and the dimensionality of the problem [80].

Most of the PGAs concentrated on multiple demes with common fitness function for all populations. But there are some PGAs, with multiple demes and multiple fitness functions, implemented on multicore/HPC, GPGPU and even on cloud which gave birth to the concept of diversity. The algorithm is said to be parallel only when its design is parallel and not because of parallel hardware or parallel programming used for implementations. But many researchers used the latter concept of parallelism.

**PGAs AND DIVERSITY**

Many researchers have been working on the issue of maintaining population diversity, avoiding premature convergence and improving the convergence speed. Research in the area of evolutionary algorithm (EA) is mainly categorized into following five categories: 1) Dynamic operator and parameter control, 2) Specialized selection operators, 3) Specialized variation operators 4) Memory-based algorithms and 5) Complex population structures [100].

In some of the multiple demes based PGAs, the concept of additional objective (Fitness function) is added for diversity in search space. These PGAs are parallel by design and provide diversity in search space and quality of solutions irrespective of other kind of parallelism. When attempts are made to get the quality solutions through diversity as a second objective, it is called as niching techniques. Niching techniques in GAs are given in [81]. Niche PGAs can provide both multiple high-quality solutions for multimodal problems and techniques for improving diversity in population [2].

The literatures based on diversity as second objectives in PGAs are as follows,

In GA, the concept of reserve population with diversity as an objective, called DPGA, was experimented by Park and Ryu, (2007) [82]. The DPGA outperforms on stationary optimization problems.

Park, Choe, and Ryu (2007) [83] has proposed DPGA-ED, an improved version of binary representation based DPGA. The reserve population in DPGA-ED is different than reserve population in DPGA, as it evolves itself. The DPGA-ED outperforms on nonstationary optimization Problems. Park, Choe, and Ryu (2008) [84] has proposed DPGA2 for Nonstationary Optimization Problems. The improvement in DPGA2 is that there are two reserve populations, which provide controlled inflow information to main population with survival selection. DPGA2 shows the performance without relying on prior knowledge of nonstationary optimization problems.

Junhua and Ming (2008) [85] proposed DSGA called Dual Species GA. In this algorithm one subpopulation work on local exploitation and other subpopulation work for global exploration with migration between them for optimal solution.

Fernandes and Rose (2008) [86] proposed Variable Dissortative Assortative Mating GA (VDMGA) based on mating strategies. VDMGA performs crossover when the hamming distance between the chromosomes is below a threshold value. VDMGA maintain genetic diversity and can tackle static and dynamic optimization problems. The VDMGA falls under the category of Dynamic operator and parameter control.

Park (2010) [87] has proposed binary, real valued and ordered based DPGA with adaptive diversity control for Unimodal and multimodal optimization Problems. Real valued DPGA for Unimodal functions with large D is fluctuating. Real valued DPGA for multimodal functions with large D is better. Real valued DPGA for multimodal functions with small D is fluctuating. Binary valued DPGA for Unimodal functions with large D is better. Binary valued DPGA for multimodal functions with large D is better. Binary valued DPGA for multimodal functions with small D is fluctuating. It is observed that, in DPGA, by controlling the crossbreeding probability between main and reserve population the diversity can be adjusted base on requirements of problems modality and other features.

Umbarkar and Joshi in [88] compared the performance serial DPGA with Multithreaded Parallel DPGA (MPDPGA). The performance of MPDPGA with DPGA was in terms accuracy, number of generations and execution time on multicore system.

The fitness calculation of reserve population, to add diversity in search space, is comparison of two chromosomes; usually compute by using Hamming distance or Euclidian distance, which is an additional computational overhead, results in increased search time. The main population evolution, reserve population(s) evolution and crossbreeding are the three important processes of DPGA.

Umbarkar and Joshi in [89] compared the performance of DPGA with simple GA and OpenMP GA for multimodal
function optimization. DPGA shows the performance over simple GA and programming parallel OpenMP GA.

The DPGA, DPGA-ED, DPGA2, DSGA and MPDPGA falls under the category of complex population structures, where it works on the concept of dual populations to get the optimal solution.

The other population diversity based works are - Classical EP (CEP), Improved Fast EP (IFEP) [90], Adaptive EP with Lévy Mutation (ALEP) [91], Island-model GA (IMGAs) [92], Restricted Truncation Selection (RTS) [93], Real Coded Memetic Algorithm (RCMA) with Crossover Hill Climbing (XHC) [94], Comprehensive Learning Particle Swarm Optimizer (CLPSO) [95], RCMA with Adaptive Local Search (LSRCMA) [96], Differential Evolution with Neighborhood Search (NSDE) [97], Covariance Matrix Adaptation Evolution Strategy (CMAES) [98]. Diversity Guided Evolutionary Programming (DGEP) [99].

4. DISCUSSION

The computing power is increasing over the time. It is moving from loosely coupled architecture to tightly coupled architecture or hybrid architecture. This computing power can be exploited optimally by using various programming languages mentioned above. Therefore the algorithm designers have challenges in understanding the latest computing and programming architectures. So it is need of time to redesign the algorithm for new computing paradigms like cloud, HPC and GPGPU.

In case of MPP, Grid and cluster- fine grained, medium grained and coarse grained GAs are explored and superliner speedup is observed with some of the applications. The GPGPU is latest computing platform, where GPUs are used for processing Single Instruction Multiple Data (SIMD) kind of part of algorithm. It is observed that the PGAs which support SIMD architecture when implemented on GPGPU give better speedup.

GA on cloud is prototypic concept of GA on cluster. It is helpful to solve the optimization problem in which processing of big data is needed. It is observed that the GAs are used to solve the optimization problems of cloud but GAs are not implemented on cloud to solve optimization problems.

In case of HPC/Multicore systems many challenges like proper utilization of cores, memory, and cache are yet to be addressed. Selection of proper programming language, algorithmic design (based on the architecture), use of niching technique, diversity etc. also play critical role in obtaining optimal solution efficiently.

Survey shows that, PGA over parallel platforms helps in reducing the function evolution, improve the quality of solution, explore the large population size, and solve the large scale, large dimensions problems efficiently.

Some authors proposed multiple populations based PGA and the importance of communication delay in performance is recognized.

PDGA, VDMGA, DPGA, DPGA-ED, DPGA2 and MPDPGA works on the concept of diversity in population(s). DPGA has extra population with diversity as objective, results in increased in computational efforts.

The diversity based EAs are available in the literature but not tried to parallelize them on today’s parallel computing paradigms such as multi-core system, GPU etc. The reasons for not being exposed for parallelization could be:

- The additional population or objective for diversity adds complexity to the EA.
- Researchers have concentrated more on altering the EA to improve the performance rather than modifying it for implementation over parallel computing system.

ACKNOWLEDGMENT

We express our sincere thank to all the authors, whose papers, in the area of evolutionary algorithm over various parallel computing paradigms, are published in various conference proceedings and journals.

REFERENCES


[41] R. Bogdanski, L. Peter, T. Becker and X. Yao, “Improving Scheduling Techniques in Heterogeneous Systems with


