Chapter 10: DIPLOIDY

10.1 Motivation
Dominance and diploidy can be implemented in genetic algorithms. Most of the research works in genetic algorithms make use of haploid genomes, which contain one allele at each locus. But in nature, many biological organisms, including humans, have diploid genomes having two alleles at each locus and even some organisms have multiploid genomes having two or more alleles at each locus. Diploidy has gained focus of many researchers so as to explore its utilization and apply its biological advantage over haploidy in evolutionary computation. The research reviewed various implementations of diploidy in different applications and then implemented it on the test problem to study the utilization of diploidy in genetic algorithms.

10.2 Haploid and Diploid Genome – Brief Concepts

Ploidy is the complete set of chromosomes in a cell. There are two types of cells in the body - haploid cells and diploid cells. The difference between haploid and diploid cells is related to the number of chromosomes that the cell contains. Diploid cells reproduce by mitosis making daughter cells that are exact replicas. Diploid cells contain two complete sets (2n) of chromosomes. Humans are diploid. A human somatic cell contains 46 chromosomes: 2 complete haploid sets, which make up 23 homologous chromosome pairs.

A haploid cell, on the other hand, is that cell which contains only one set of chromosomes in it. Haploid cells are a result of the process of meiosis, a type of cell division in which diploid cells divide to give rise to haploid germ cells. A haploid cell will merge with another haploid cell at fertilization. Haploid cells have half the number of chromosomes (n) as diploid - i.e. a haploid cell contains only one complete set of chromosomes. Haploid cells are found in various algae, various male bees, wasps and ants. Haploid cells should not be confused with monoploid cells as the monoploid number refers to the number of unique chromosomes in one biological cell. Male bees, wasps, and ants are haploid organisms because of the way they develop from unfertilized, haploid egg cells.
However, many organisms have more than two sets of homologous chromosomes and are called polyploid. Polyploidy refers to a state where there are multiple sets of chromosomes are present. This is commonly seen in plant cells but not in animal cells.

Chromosomes are composed of genes which may take some value called allele. One or more chromosomes combine to form total genetic prescription for construction and operation of some organism. In natural systems, the total genetic package is called genotype and organism formed by interaction of total genetic package with its environment is called phenotype. Each individual’s genotype consists of a set of chromosomes which could be decoded to generate phenotype. In artificial genetic systems, strings correspond to chromosomes and structure corresponds to genotype. Structure decodes to form particular parameter set or solution alternative which corresponds to phenotype.

The simplest genotype in nature is haploid. Haploid genome has single chromosome which is subjected to different genetic operations to find the solution. Diploid genome has two chromosomes that consist of two set of alleles representing different phenotypic properties. Each allele represents the phenotypic property of a particular gene.

There exist all types of organisms in nature – haploid, diploid and multiploid. Genotype of diploid organisms contains double the amount of information for same function than the haploids. This leads to lots of redundant information which is eliminated by the use of genetic operator – Dominance. At a locus, one allele takes precedence over other alleles. The allele which is expressed is dominant and others are recessive. Dominant alleles are expressed by capital letters and recessive ones in small letters in the phenotype.

Dominance can be referred to as genotype to phenotype mapping or genotype reduction mapping [Goldberg 1989]. It could be represented as:

\[
\begin{align*}
\text{AbCDe} & \rightarrow \text{ABCDe} \\
\text{aBCde} &
\end{align*}
\]

Dominant gene is expressed both in Homozygous as well as Heterozygous state. Recessive gene is expressed only in homozygous state. It is represented as:
Diploidy and Dominance clearly state that double information in genotype is reduced by half in its phenotypic representation. Existence of redundant information in chromosomes and then its elimination leads to a thought provoking question. Why does nature keep double information in genotype and utilizes half of the information in phenotype? At first, this redundancy of information seems to be wasteful. But, it is hard fact that nature is not spendthrift. There must be some good reason behind the existence of diploidy and dominance in nature and keeping redundant information in genotype.

Diploidy introduces the concept of genetic memory. Diploidy provides a mechanism for remembering alleles and allele combinations that were previously useful and that dominance provides an operator to shield those remembered alleles from harmful selection in current hostile environment. Dominance acts as a genetic operator that protects remembered alleles from fast destruction [Goldberg 1989]. This memory of diploidy stores the information regarding multiple solutions, but only one dominant solution is expressed in phenotype. Redundant information is carried along to next generation. Any change in dominance leads to utilization of the information stored in genetic memory that can be tested and expressed in further generations. Dominance itself evolves. Dominance or non-dominance of a particular allele is itself under genetic control.

Diploidy increases diversity in GAs by allowing recessive genes to survive in a population and become active at some later time when changes in the environment make them more desirable. One drawback of diploidy is that the mechanics of a diploid GA requires twice as much computational effort as the mechanics of a haploid GA because we have twice as many alleles to deal with [Ryan et al. 2003].

Goldberg took a step toward solving dynamic fitness problems by using diploidy to create individuals in changing, often oscillating, fitness environments [Goldberg 1989]. As predicted, the diploid populations were able to adapt to changing environments more readily than haploid populations, a result that many people attribute to the act of calling upon once-successful traits that were cached in combinations of recessive alleles that could re-emerge at
random times and potentially exploit different environments. In an experiment, Goldberg showed that how diploidy and dominance permit alternate solutions to be held in abeyance – shielded against over selection.

### 10.3 Historical Survey

In 1987, D.E. Goldberg and R.E. Smith used diploid representations and a dominance operator in GA’s to improve performance of non-stationary problems in function optimization. They used three schemes: a simple haploid GA, a diploid GA with a fixed dominance map (1 dominates 0) and a behavior of genetic algorithm with dominance and diploidy and applied them to a 17-object, blind, nonstationary 0-1 knapsack problem where the weight constraint is varied in time as a periodic step function [Goldberg et al. 1987a].

They found diploidy and dominance to be effective mechanisms of inducing a form of long term distributed memory within a population of structures. They also examined the schema representation of these mechanisms. They proved the superiority of diploidy over haploidy in a nonstationary knapsack problem by performing computational experiments and concluded that both diploid schemes were better than haploid scheme in satisfying switching requirements of the nonstationary environment. The diploid scheme with the evolving dominance map was better than the diploid scheme with the fixed map in performance and speed.

In 1992, R.E. Smith & D.E. Goldberg extended their research and examined the effects of diploid representations and dominance operators in genetic algorithms applied to nonstationary search problems. Analytical and experimental evidence showed that a diploid GA maintained extra diversity at loci where alternative alleles were emphasized in the recent past [Smith et al. 1992]. In effect, diploidy embodied a form of temporal memory that is distributed across the population. Analysis also revealed that abeyant recessives are sensitive to past environmental conditions, and can therefore act as a form of distributed, probabilistic memory of environmental conditions that occur periodically. Experiments showed that an adaptive dominance map is necessary to effectively exploit the advantages of diploidy. Analysis of and experiments with the triallelic adaptive dominance scheme had shown the importance of a dominance shift operator in such schemes. The dominance shift rate is a control on probabilistic memory in a diploid GA, with recessive alleles retained longer and recalled less quickly for lower dominance shift rates. Experiments also show that the diploid
GA can sustain rates of dominance shift that are much higher than those typically used for mutation. They concluded that diploidy and dominance are naturally inspired additions that improve the performance of GAs in a broad class of search, optimization, and learning tasks.

In 1994, F. Greene used diploid/dominance in genetic search. Diploid chromosomes were computed separately and were evaluated to produce two intermediate phenotypes. Dominance was implemented by mapping two sub-phenotype fitness values to scalar fitness value. Mapping function was called dominance map or dominance function and fitness was referred to as sub-fitness [Greene 1994]. Experiment was performed using diploid chromosome C++ object compatible to Genitor and arithmetic crossover was implemented on it and they identified the changing global optima. This method was problem independent and could be readily extended to a polyploidy dominance.

In 1996, Calabretta et al. compared the behavior of haploid and diploid populations of ecological neural networks in fixed and changing environments. They showed that diploid genotypes were better than haploid ones in terms of fitness and diploid genotypes retained better changes in environment. They analysed the effect of mutation on both type of populations. Mutations, which are mostly deleterious, had larger effects on diploids than on haploids, so more individuals are negatively affected by mutations in diploids than in haploids [Calabretta et al. 1996]. This led to pushing average fitness down in diploids but it also caused peak fitness to go up. Hence, they concluded that diploids had lower average fitness but higher peak fitness than haploids.

In 2002, Yilmaz S. Ayse, Annie S. Wu proposed a new diploid scheme without dominance on integer representation. Earlier, diploidy and dominance were considered together and were mainly applied for binary representation and stationary problems. In their research, they evolved all diploid individuals without using any haploid stage and compared performance of diploid GA with haploid GA for both stationary and non-stationary environments [Ayse et al. 2002]. They considered TSP as their test problem. They compared simple haploid GA and diploid GA in terms of computational overhead, solution quality and storage overhead and found that simple haploid GA outperformed diploid GA.

In 2003, Robert Schafer investigated the ability of diploid GA to approach a class of problems in which fitness landscape is reciprocally affected by population itself. He used GA
to create entire systems in states of dynamic equilibrium, in which maintaining diversity of individuals in a population is as important to the solution as is maintaining high fitness. In his study, a GA protocol was presented as a tool to approach dynamic systems having reciprocal individual-environment interaction and then applied on a model problem in which a population of simulated creatures lived and metabolized in a three-gas atmosphere. The resulting complex system was judged in terms of fitness of the individuals and equilibrium of the environment itself and revealed the ability of GA with diploidy and dominance to create dynamically stable systems without a loss of fitness [Schafer 2003].

In 2007, Shengxiang Yang proposed an adaptive dominance learning scheme for diploid genetic algorithms in dynamic environments. In this scheme, the genotype to phenotype mapping in each gene locus was controlled by a dominance probability. Dominance probability vector was learnt towards best individual during the searching progress and adapted to dynamic environments [Yang 2007]. The proposed dominance scheme was experimentally compared to two other schemes for diploid genetic algorithms, namely, Ng-Wong dominance scheme and additive diploidy scheme. Experimental results validated the efficiency of the dominance learning scheme. It outperformed the other two investigated dominance change schemes. The parameter sensitivity analysis showed that the learning rate had a significant effect on the dominance learning scheme and setting in the range of (0:25; 0:5) emerged as a good choice for the scheme. Out of the two schemes compared with Dominance learning scheme, additive diploidy scheme proved to be better than the Ng-Wong dominance scheme [Ng et al. 1995].

In 2009, Dan Simon utilized diploidy and dominance in genetic algorithms (GAs) to improve performance in time-varying optimization problems. He used the scaled One Max problem to provide additional theoretical basis for the superior time-varying performance of diploid GAs. He suggested an intuitive way to implement diploidy and presented some mathematical analysis of fitness proportional selection to justify its use in time-varying problems. An extension of the classical schema theorem for diploid GAs was used and GA with selection and mutation only (no crossover) was taken into consideration. The analysis confirmed that diploidy increases diversity, and provided some quantitative results for diversity increase as a function of the GA population characteristics. The derivation of a diploid version of the classical schema theorem exhibited additional diversity provided by diploidy, which is more pronounced as the fitness ratio of high-fit to low-fit allele’s increases [Simon 2009].
10.4 Simulation and Analysis

To explore the benefits of diploid genome and dominance, genetic algorithm was developed in MATLAB for Travelling Salesman Problem as Test problem. The algorithm was coded using both haploid and diploid genome set and was implemented first for 10 cities. The cost of different paths was computed for fifteen consecutive runs and then compared. The implementation was then carried out for 50 cities and the results were compared. It was observed that in majority of runs both in case of 10 cities and 50 cities, diploid genome resulted in better results than haploid genome. The cost of path of final result using diploid genome was found to be less than that computed with haploid genome [Kumar et al. 2010]. Moreover, computational time was also found to be less in case of diploid chromosomes. Comparison of cost and time for different cases is illustrated in following figures.

![Plot of Points of 10 Cities for the TSP Route](image)

Figure 10.1 Plot of Points of 10 Cities for the TSP Route
Figure 10.2: Hamiltonian Path and Cost incurred using haploid genome for TSP of 10 cities

Figure 10.3: Hamiltonian Path and Cost Incurred Using Diploid Genome for TSP of 10 Cities
Figure 10.4 Comparison of Cost Incurred for Haploid and Diploid Genome for 10 Cities TSP

Figure 10.5 Comparison of Time Taken by Genetic Algorithm for Haploid and Diploid Genome for 10 Cities TSP
Figure 10.6 Comparison of Cost Incurred for Haploid and Diploid genome for 50 Cities TSP

Figure 10.7 Comparison of Time Taken by Genetic Algorithm for Haploid and Diploid Genome for 50 Cities TSP
By comparing haploid and diploid implementation of genetic algorithm, it has been shown that the genetic algorithm with diploid chromosomes performs better than the genetic algorithm with haploid chromosome. The experimental results show that the diploid GA can achieve faster response and is easy to implement. In continuation with the research work, it is proposed to develop a genetic algorithm using crossover probabilities and different crossover points to evaluate the performance in each case.

10.5 Summary

In nature, genome of biological organisms can be haploid-single set of chromosomes, diploid- double set of chromosomes or multiploid. Diploidy and multiploidy have led to dominance of genes in the genotype and subsequently changes in the phenotypes of organisms. Lots of research work is carried out to establish the utility of diploidy in nature-inspired algorithms. The chapter explores research work related to diploidy and dominance. It has been seen in past researches that diploid genome is better than haploid genome – both in nature as well as in genetic algorithms. The proposed research works tests the performance of genetic algorithms on standard travelling salesman problem (TSP) using MATLAB code and the simulation results confirm the benefits of diploid genome over haploid- both in terms of cost and time.