CHAPTER 1

INTRODUCTION

1.1 SURFACE ENGINEERING

The process of surface treatments, more formally surface engineering, tailors the surfaces of engineering materials to

- Control friction and wear,
- Improve corrosion resistance,
- Vary appearance, like color and roughness and
- Reduce cost.

Ultimately, the functions and/or service lives of the materials can be improved.

Common surface treatments can be divided into two major categories:

1. Treatments that cover the surfaces.
2. Treatments that alter the surfaces

1.1.1 Treatments that cover the surfaces

Some times a work piece needs to be built up or material needs to be added to a work piece instead of removing materials from work piece. Surface
coating serves the above purpose and it can broadly be classified as Organic coating and Inorganic coatings.

The organic coatings apply paints, cements, laminates, fused powders, lubricants, or floor toppings on the surfaces of materials.

The inorganic coatings perform electroplating, autocatalytic plating (electro less plating), conversion coatings, thermal sprayings, hot dipping, hard facings, furnace fusing, or coat thin films, glass, ceramics on the surfaces of the materials.

1.1.2 Treatments that alter the surfaces

i) Hardenings: Selective hardenings can be done by flame, induction, laser or electron beam.

ii) High Energy Treatments: Common high-energy treatments include ion implantation, laser glazing/fusion, and electron beam treatment.

iii) Thin Diffusion Treatments: Thin diffusion processes include Ferritic-nitrocarb, boronizing, and other high temperature reaction processes, e.g., TiC, VC.

iv) Heavy Diffusion Treatments: Heavy diffusion processes include carburizing, nitriding, and carbonitriding.

v) Special Treatments: Some special treatments such as cryo, magnetic, and sonic treatments, affect not only the surfaces but also the bulk materials.

Here, diamond incorporated electro composite comes under treatment that covers the surfaces.
1.2 COATINGS

1.2.1 General

A coating is a thin film that is applied to an object to protect it or change its appearance. It may be applied as liquids, gases or solids. Coatings can be used for different purposes. Some objects are plated to increase their sturdiness and provide hard shell for whatever it is coated on. Some are coated to avoid corrosion and few are coated just to offer an attractive finish.

1.2.2 Types of coatings

There are various types of coatings which are listed below.

- Metals
- Alloys of metals
- Semi conductors
- Metal oxides
- Dielectrics
- Ceramics
- Polymers

1.2.3 Techniques of deposition of coatings

Technologies involved in can be classified under two categories.

1.2.3.1 Deposition by wet techniques

- Electro deposition
- Electroless deposition
- Electrophoretic deposition
• Immersion plating
• Anodization

1.2.3.2 Deposition by dry techniques

• Evaporation
• Sputtering
• Chemical vapour deposition
• Ion plating
• Ion implantation

1.3 ELECTRODEPOSITION

1.3.1 General

Electroplating has, over recent decades, evolved from an art to an exact science. This development is responsible for the ever increasing number and varied applications of this branch of practical science and engineering. Electroplating is a form of surface modification, in which metal is deposited electrochemically from a solution of its own ions on the substrate by passing direct current.

Plating is done in a non metallic plating tank, which is filled with electrolyte containing the ions of the metal to be plated. An electroplating set up essentially consists of direct current source such as a rectifier or a regulated power supply unit connected in series through an ammeter to the plating cell comprising the bath, an anode and a cathode immersed in it. The current enters and leaves the electrolyte via the anode and cathodes. The anode is usually the metal to be plated (assuming that the metal will corrode in the electrolyte).
The cathode is the work piece, the substrate to be plated. This is connected to the negative terminal of the power supply. The power supply is well regulated to minimize ripples as well as to deliver a steady predictable current, under varying loads such as those found in plating tanks. Figure 1.1 is a simple electro plating setup describing the main components involved in it.

![Figure 1.1 Electroplating Setup](image)

As the current is applied, positive metal ions from the solution get attracted to the negatively charged cathode and deposit on it. The metal from the anode dissolves into the solution and balance the solution equilibrium.

At each electrode, there are more than one reaction occurring. They are as follows:

At the anode

\[ M^{n+} \rightarrow M^{(n+1)+} + e^- \] (oxidation)

\[ M^{n+} \rightarrow M + ne^- \] (metal dissolution)

\[ 2H_2O \rightarrow O_2 + 4H^+ + 4e^- \] (oxygen evolution)

\[ M \rightarrow MO \] (passivation)

If metal dissolution is the major reaction, it is said to have good efficiency.
At the cathode

\[ M^{n+} + ne^- \rightarrow M \]  
(metal deposition)

\[ 2H_2O +2e^- \rightarrow H_2 + 2OH^- \]  
(hydrogen evolution)

\[ M^{(n+1)+} + e^- \rightarrow M^{n+} \]  
(reduction)

If the deposition process is the major process it is said to have a good current efficiency.

Electro Plating is governed by Faraday's Laws which are:

1. The weight of a substance formed at an electrode is proportional to the amount of current passed through the cell.
2. The weight of the metal deposited or dissolved for the same quantity of electric current is passed through different electrolytes is directly proportional to the electro chemical equivalent of the metal deposited or dissolved.

If the mass of the metal deposited is known, the thickness can be calculated for objects of known surface area, since Mass deposited = Density of the metal × Surface area × thickness. The current efficiency = (mass of deposited metal / theoretical mass) x 100.

1.3.2 Features of Electroplating

Some of the salient features of electroplating process are listed below.

- Simple and cheap surface modification technique
- Any substrate could be electroplated
- Substrate properties are unaffected
- Smooth and fine grained coating
- Nearly uniform coating
- Thickness of the deposit could be controlled
- Corrosion resistant coating could be produced on reactive metals
- Surface properties can be tailor made to suit the specific requirement
- Suitable for small and big jobs
- Applicable both for short run and mass production
- Multi layer coating systems are possible
- Compositionally modulated alloys could be produced
- Coatings can be produced both at room temperature and elevated temperatures

1.3.3 **Properties achieved by Electroplating**

Many desirable properties can be achieved by electroplating process. They are as listed below.

- Improved appearance
- Corrosion resistance
- Resistance to wear and abrasion
- Increased electrical conductivity
- Improved frictional properties
- Specific metallic appearance
- Resistance to high temperature
- Specific magnetic properties
• Production of complicated shape articles with close dimensions
• To increase dimension in repair and maintenance work

1.3.4 Applications of Electro deposition

Even though the electro deposition technique has certain drawbacks such as slow rate of deposition, high waste disposal cost, use of toxic chemicals, dimensional change and post machining / grinding requirement, it finds its place in many applications. Some of the area of application of electroplating is listed below as Table 1.1.

**Table 1.1 Some of the application of electrodeposition.**

<table>
<thead>
<tr>
<th>Area</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>House hold articles</td>
<td>Decorative items, Air-conditioners, Jewelry, Presentation items / Novelties, Holograms</td>
</tr>
<tr>
<td>Transport</td>
<td>Bicycle, Bus, Car, Train, Plane, Helicopter</td>
</tr>
<tr>
<td>Entertainment</td>
<td>Radio, Tape recorder, Television, Musical instruments</td>
</tr>
<tr>
<td>Electronics</td>
<td>Control systems, Radar instruments, Computer</td>
</tr>
<tr>
<td>Defence / Aerospace</td>
<td>Missiles, Remote controlled helicopter, Fighter plane</td>
</tr>
<tr>
<td>Printing</td>
<td>Bank note press, Coins</td>
</tr>
<tr>
<td>Communication</td>
<td>Telephone wires</td>
</tr>
</tbody>
</table>

Apart from the above areas, electroplating is also used in electrical industries, power stations, solar energy conversion systems and nuclear energy applications.
1.4 COMPOSITES

Composites are materials made up by the combination of two or more materials and possess unique properties that individual materials do not have. They involve two or more constituents- the matrix and the distributed phase. The matrix can be of plastic or metal. Distributed phase may contain fibers or particles.

Composites can be classified with respect to the type of matrix and distributed phase as listed below.

i) Plastic matrix composites
ii) Metal matrix composites
iii) Particle reinforced composite.
iv) Fiber - reinforced composite.
v) Structural composite.

Plastic matrix composites can easily be formed. It will have light weight, high specific strength and suitable for ambient temperature applications but not suitable for high temperature applications.

Metal matrix composites are formed on all substrates over which electroplating is done. Any metal or alloy which can be obtained by electrolysis can be used as a matrix. Particles should be insoluble and wetted by the plating solution and it should be available in the powder or fiber form.

When fibers or particles are dispersed in a metal matrix, fibre-reinforced metal composites (FRMC) or Particle dispersed metal composites (PDMC) will be formed. There are various methods of preparation of metal...
matrix composites namely Powder metallurgy, metal spraying, internal oxidation, Co-precipitation, electrodeposition and electroless deposition.

1.5 COMPOSITE COATING

1.5.1 General

Composite coatings help to get the synergistic properties of two or more material that single material does not posses for the specific application. The interest of electrolytic and electroless composite coatings has increased substantially during the last decade. This interest is spurred by the wear and abrasion resistance, high temperature and oxidation resistance, corrosion resistance and many other desired properties made possible with these coatings.

Co-deposition of a homogeneously dispersed second phase material in the form of a particulate material, whisker, fibre in a metal matrix offers surface coatings with improved or sometimes completely new engineering properties. Unique functional properties of these coatings are derived by the presence of particles dispersed in the bulk of the metallic matrix and their proportions. The reason for the wide adaptation of this technology by the industry is not only their unique properties but also their longer service life and cost effectiveness compared to conventional coatings.

In the past, these coatings were developed primarily to induce micro porosity in the top chromium layer with a view to improve the corrosion resistance of decorative nickel-chromium layers for the automotive industry. Today because of their unique properties micro structural stability, protective performance and high temperature resistance their application area has been extended to transportation, machine constructions, machine tools, aerospace, nuclear engineering and many such high tech areas as structural materials.
The composites are dispersion hardened by particle reinforcement or fibre reinforcement. Fine inert particles or filamentary reinforcements with high temperature strength, combined with a soft conventional metal material can produce composites that outperform super alloys.

The composites in the high performer arena today are Particle dispersed metal composites (PDMC), Fibre reinforced metal composites (FRMC) and Electroless composites. Particle dispersed metal composite coatings are used to

i) Increase wear, abrasion and creep resistance of metal and alloys
ii) Impart dry self lubrication
iii) Increase high temperature strength and oxidation resistance
iv) Increase corrosion resistance
v) Produce heat treatable alloys
vi) Produce nuclear coatings

There are various techniques to develop these composite coating namely physical vapor deposition (PVD), chemical vapor deposition (CVD), electro and electro less plating. Among the above processes, Electroplating process is simple and less expensive.

PVD involves the formation of a coating on the substrate by physically depositing atoms, ions, or molecules of a coating species. CVD involves the formation of a coating by the reaction of the coating substance with the substrate. Both PVD and CVD must be carried out in vacuum atmosphere, which involves high cost wherein Electroplating process is simple and less expensive.
Electrodeposition technique has been used as a method to produce variety of metal composite coatings since 1963. Particles to be dispersed in a metallic matrix are kept suspended in the bath by agitation and during electroplating become incorporated in the deposit. The bath compositions commonly used for composite plating are identical to conventional plating baths used for deposition of the pure metals.

Metal matrix composite consists of at least two constituents, the matrix and the homogenously dispersed second phase in a compact metallic form each making property contribution to the final product. Generally, they impart high hardness, wear resistance, lubricity or oxidation resistance. Those with highest hardness and wear resistance are produced by incorporating diamond powder, silicon carbide, tungsten carbide etc., in a metal matrix preferably nickel the most preferred engineering material. The advantage of nickel is that it can be easily electroplated as a thin composite coating on any substrate like steel or aluminum with which the tool can be easily fabricated.

Various types of electro composite coating based on nickel metal matrix used are

- Ni-diamond composite coating,
- Ni-alumina composite coating,
- Ni-CBN composite coating,
- Ni-SiC composite coating,
- Ni-Cobalt-diamond Composite coating

1.5.2 Nickel-diamond electro composites

Electro composites offer unique synergistic properties which can not be obtained by any single metallic coating. They enhance the wear resistance
and hardness of the base metal thus offering increased life. Diamond and diamond based composite coatings are considered as super abrasive materials. Due to their superior tribological properties, these materials are industrially important. Though the exceptional properties of diamonds are well known, it is difficult to make any component using only diamond powder. This has necessitated the growth of composite market throughout the world. Composite diamond coating has the following advantages like

i) Exceptional wear resistance

ii) Excellent hardness

iii) Enhanced corrosion resistance

iv) Perfect conformity to complex geometries.

v) Increased thermal transfer since diamond acts as a heat sink.

vi) Applicability to all common metals and alloys

vii) Coverage of entire surfaces or selected critical areas

1.5.3 Methods of co-deposition of Nickel-diamond electro composites

Nickel-diamond electro composites can be produced in two ways

i) Conventional electro co-deposition (CECD)

ii) Sedimentation co-deposition (SCD)

In Conventional electro co-deposition (CECD) method using vertically positioned anode and cathode the particles are continuously stirred to keep them in suspension allowing the formation of a composite deposit on the cathode during electrolysis. But in the case of Sedimentation co-deposition (SCD) method the electrodes are positioned horizontally and the particles are periodically stirred by vigorous agitation and allowed to settle on
the horizontal cathode during electrodeposition. The advantage of SCD technique is that a higher incorporation can be obtained for the given amount of powder concentration in solution in comparison with the CECD technique, and that a stratified incorporation by periodic sedimentation is possible. SCD technique is possible on horizontal near flat cathodes.

1.6 MODELING TECHNIQUES USED IN THIS INVESTIGATION

Two modeling techniques and two optimization techniques were employed for the prediction of both volume fraction of diamond deposition in the nickel-diamond metal matrix and hardness of the composite coating. They are as follows.

i) Regression Modeling
ii) Artificial Neural Network (ANN) modeling
iii) Genetic Algorithm (GA) optimization
iv) Particle swarm optimization (PSO)

1.7 REGRESSION MODELING

Regression model is a statistical tool. Regression analysis is for the prediction or estimation of the value of a variable from a known value of other variable to which it is related i.e) the regression analysis helps to estimate (predict) the value of one variable from the given value of another.

Regression means to return or to go back. So it implies the act of returning to or going back to. Natural phenomena generally have a tendency to return to normal. In statistics the term ‘regression’ is used to denote backward tendency which means going back to average or normal.
The term regression was first used in the study of heredity. But now the concept of regression is extended to other spheres of phenomena which have a tendency to regress or set back to the normal or the general average. The concept of regression is very helpful in the study of correlation.

Regression shows a relationship between the average values of two variables. Thus regression is very helpful in estimating and predicting the average value of one variable for a given value of the other variable. The estimate or prediction may be made with the help of a regression line which shows the average value of one variable x for a given value of the other variable y. The best average value one variable associated with the given value of the other variable may also be estimated or predicted by means of an equation and the equation is known as regression equation.

There are two types of variables in regression analysis namely independent variable and dependent variable. The variable whose value is influenced or is to be predicted is called Dependent variable, whereas the variable which influences the value or is used for prediction is called Independent variable. The independent variable is also known as Regressor or Predictor or Explanator while the dependent variable is called as Regressed or Explained variable. Regression analysis can be extended to three or more variables. If two variables are taken into account, it is called as simple regression. The tool of regression when extended to three or more variables, it is called as multiple regressions.

The importance of regression modeling is summarized as below:

- Regression analysis provides an useful tool to estimate an unknown value of one variable for a given values of another variables
Regression analysis is used in industries to control product quality.

It provides an accurate basis for the comparison of variables and significant relationships.

The purpose of regression analysis is to discover a predictive variable.

1.8 ARTIFICIAL NEURAL NETWORK (ANN) MODELING

1.8.1 General

The development of artificial neural network was initially motivated by research into biological nervous systems which consist of densely connected networks of neurons. In the human nervous system, there are over 100 types of neurons. A simplified diagram of one type of neuron is presented in Figure 1.2.

![Biological neuron](image)

Figure 1.2 Biological neuron

The neuron is essentially a chemical processing plant. Packets of chemicals are transported through the dendritic tree to the synaptic bulb.
When the synaptic bulb is sufficiently charged, it releases a packet of chemicals across the synaptic gap to the body of the neuron. Depending on the nature of the synaptic gap, this either increases or decreases the activity in the neuron. The size of the synaptic gap determines the magnitude of the influence. When the neuron is sufficiently charged up, it releases a packet of chemicals through the axon to its dendritic tree. Synapses interact locally in a non-linear fashion. The neuron is sensitive to both the frequency and arrival time of pulses. An individual biological neuron has minimal computational ability. Interesting computational properties only emerge when neurons are combined together in various ways. An artificial neural net is a network of artificial neurons (variously referred to as "Processing Elements", "PEs", "Nodes", or "Units") which have several input paths and one output path. A typical PE is shown in Figure 1.3.

\[
I = W_0 + W_1 X_1 + W_2 X_2 + W_3 X_3
\]

\[
F(I) = \frac{1}{1+e^{-I}}
\]

(non-linear transfer function)

Figure 1.3 A typical processing element in a neural net

Comparing this with Figure 1.2, pulses have been converted to pulse rates or frequencies (X’s). The effects of the synaptic gap on the internal activation of the neuron are modeled by weights (W’s) which are multiplied by the frequencies (X’s). These individual contributions are summed together to form the internal activity (I) in the neuron. A constant input called the
"Bias" is used to simulate thresholding effects in the neuron and simplify the mathematics. A picture of a neuron is actually a graphical short-cut for writing out a formula. Each neuron consists of two parts: A way to compute the internal activation, and a way to transform this to an output. The internal activation is usually the sum of the input times its associated weight. Weight zero is always associated with the "bias", and input zero is always equal to one.

1.8.2 Multi-Layer Perceptron:

Processing elements can be connected in different ways. One of the most popular architectures is the multi layer perceptron (MLP). This uses PEs of the type shown as Figure 1.3 which was connected and shown as Figure 1.4.

![Multi layer perceptron architecture](image)

Figure 1.4 Multi layer perceptron architecture

This graphical representation of a neural net could also be written as a formula involving many summations and transfer functions. This formula maps input vectors presented at the input buffer to output values. One of the important aspects of a multi layer perceptron is that it can synthesize any function to any desired level of accuracy given enough hidden layer PEs.
However, for real world data, which is typically incomplete and noisy, there are practical limitations to achieving good accuracy and there are trade-offs between the accuracy and the generalization ability of the formula. Just as with linear regression, solving for the parameters or weights in this formula requires data. The data must consist of a set of input records and corresponding "target" output records which are representative of problem data domain. This set of correspondences between input and output data provide historical examples which the neural net training algorithm uses to learn its weights.

### 1.8.3 Training a multi-layer perceptron

There are methods available for training a multi layer perceptron network. Most of them are based on a technique called “gradient back propagation”. An objective function is specified which is a measure of how closely the outputs of the network, match the target outputs in the training set of data. Back-propagation is a method for assigning responsibility for mismatches to the each of the processing elements in the network; this is achieved by propagating the gradient of the objective function back through the network to the hidden units. Based on the degree of responsibility, the weights of each individual processing element are modified iteratively to improve the objective function. Two types of learning rules are used: (i) An adaptive gradient learning rule which is a form of back-propagation and is widely applicable, and (ii) a Kalman learning rule which is applicable to regression type problems in which the number of inputs is not too large. This second learning rule is especially effective for noisy behavioral problems such as stock market prediction because of its inherent ability to suppress noise.
1.8.4 Neural net construction

Many training schemes assume a fixed architecture for the neural net. In other words, the number of hidden units is fixed in advance. Predict uses a constructive method for determining a suitable number of hidden nodes. This constructive method is referred to as "Cascade Learning". This method is loosely characterized by the following: (a) Hidden PEs are added one or a few at a time; (b) New hidden PEs have connections from both the input buffer and the previously established hidden nodes. It is common for Cascade Learning architectures to have direct connections from the input buffer to the output units.

1.8.5 Training the artificial neural network

The neural network has to be first trained and then tested to use for a particular application. The training can be done using a computer. The parameters like type of network, input range, training function, performance function, number of layers and transfer function used to create the network are to be selected. The training data are to be normalized to get better results. Then the developed network is to be trained with some stopping criteria. The training of network would stop after reaching the set stopping criteria.

1.8.6 Validating the artificial neural network

The trained network is to be checked for validation. Once the network was trained such that the maximum error for any of the training data was less than allowable error, the weights and the threshold values are automatically saved by the program and the network can be checked with any set of inputs. As the input values from the validation experiments are given to the ANN program, the program predicts the required output. To validate the
results of the Artificial Neural Network analysis, fresh set of data or same data used for training may be used.

1.9 GENETIC ALGORITHM (GA)

1.9.1 General

Genetic algorithms are computerized search procedures based on the mechanics of natural genetics and natural selection that can be used to obtain global and robust solutions to optimization problems. Genetic algorithms are computational optimization schemes with an unconventional approach. Genetic algorithms combine survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. In every generation, a new set of artificial strings is created using bits and pieces of the fittest of the old; an occasional new part is tried for measure. Genetic algorithms differ from conventional optimization and search procedures in several fundamental ways as follows: (i.) GA work with a coding of solution set, not the solution themselves; (ii) GA search form a population of solutions not a single solution; (iii) GA use payoff information (fitness function) of derivatives or other auxiliary knowledge; (iv) GA use probabilistic transition rules, not deterministic rules. Genetic Algorithms normally begin with a population of strings created randomly. There after, each string in the population is evaluated. The population is then operated by three main operators. (i) Reproduction (ii) Cross over (iii) Mutation. The population is further evaluated and tested for termination. If the termination criteria are not met, the population is again operated by the three operators and evaluated further. This procedure is continued till the termination criterion is met. One cycle of these operators and evaluation procedure is known, as a generation in GA terminology and the flow chart describing the above is shown as Figure.1.5.
1.9.2 Problem formulation (Objective function)

For any optimization problem, the problem formulation is the first and foremost step. Hence the objective function of the optimization problem is to be defined. Genetic algorithms are naturally suitable for solving maximization problems. But minimization problems are usually transformed into maximization problems by using suitable transformation. So generally fitness function $F(x)$ is derived from the objective function. Then constraints if any are to be fixed.
To solve any problem using Genetic Algorithm, variables involved in the problem must be coded in some string structure. Binary coded strings having 1’s and 0’s are mostly used.

1.9.3  GA Optimization

Basic steps of GA are (1) Initialization (2) Population evaluation (3) Reproduction (4) Cross over (5) Mutation.

1.9.3.1  Initialization

The first step in GA is to initialize the population. The initialization can be executed with either a randomly created population (or) a well adapted population.

1.9.3.2  Population Evaluation

Two important issues in the evaluation process of genetic search are population diversity and selective Pressure. i) Population diversity means that the genes from the already discovered good individuals are exploited while premising the new areas of the search space continue to be explored. ii) Selective pressure is the degree to which the better individuals are favored.

1.9.3.3  Reproduction

Reproduction is the first operator applied on population. Reproduction selects good strings in a population and forms a mating pool. There exist a number of reproduction operators in GA. But the average strings are picked from the current population and their multiple copies are inserted in the mating pool in a probabilistic manner. The “Roulette- wheel” method can be used for the selection of chromosomes for parents to cross over.
1.9.3.4 Cross – Over

After the reproduction phase is over, the population is enriched with better individuals. Reproduction makes clones of good strings, but does not create new ones. Swapping of bits would take place in between two randomly chosen parent strings so that fitness value may be improved further. Single point crossover method is illustrated in Figure 1.6, where the two parents are cut at their center points, and the child chromosome is given the same combination of genes as the first parent in its outer half chromosome and the same genes as the second parent in the inner half chromosome. When crossover is not conducted, the parent is copied into the next generation. This procedure continues until the population is reached.

![Image of single point crossover](image)

**Figure 1.6 Cross-Over**

1.9.3.5 Mutation

Mutation as it can occur in nature will also be taken into account. But the probability of its occurrence is less. Mutation occurs only if the random number generated is less than the mutation probability. Mutation
operator alters a chromosome locally to hopefully create a better string. Mutation of a bit involves flipping it, changing 0 to 1 and vice versa. The flow chart describing the Genetic algorithm procedure involving all the above detailed operations is shown as Figure 1.7.

![Flow chart describing the genetic algorithm procedure.](image)

Figure 1.7 Flow chart describing the genetic algorithm procedure.
1.10 PARTICLE SWARM OPTIMIZATION (PSO)

1.10.1 General

Particle Swarm Optimization is an evolutionary computation technique. The particle swarm idea originated as a simulation of a simplified social system, the graceful but unpredictable choreography of a flock of birds. The word ‘swarm’ is used after a paper by Millonas, who developed several models in artificial life and considered certain principles in swarm intelligence. The selection of the term ‘particle’ comes from mechanics and is justified by the fact that positions and velocities are applied to the population elements, despite them being considered to have zero mass and volume. Kennedy and Eberhart’s first original idea was to simulate the social behaviour of a flock of birds in their endeavor to reach, when flying through the field (search space), their unknown destination (fitness function), e.g. the location of food resources. In PSO, each problem solution is a bird of the flock and is referred to as a particle. In this algorithm, birds evolve in terms of their individual and social behavior and mutually coordinate their movement towards their destination.

To this end, each bird keeps track of its coordinates in the problem space and aims at a specific direction: the best solution (best local position) it has achieved so far. Birds also communicate among them and are able to identify the bird in the best position. In a coordinated way each bird evolves by changing its velocity so that it accelerates towards both its best position and the best position obtained so far by any bird in the flock (best global position). This enables each bird to explore in the search space from its new location. The process is repeated until the best bird reaches certain desired location. It is worth noting here that, according to the description, the process involves not only intelligent behavior but also social interaction. This way,
birds learn both from their own experience (local search) and from the group experience (global search).

PSO exhibits common evolutionary computation features including:

i) Initialization with a population of random solutions

ii) Search for optima by updating generations

iii) Particles evolution through the problem space by following some specific strategies.

Thus, the process initially starts with a group of particles, which have been randomly generated, representing different solutions of the problem. The $i^{th}$ particle is represented by its location in an $s$-dimensional space, where $s$ corresponds to the number of variables of the problem. Any set of values of the $s$ variables, determining the particle location, represents a candidate solution for the optimization problem.

PSO shares with other evolutionary techniques that it does not guarantee the global optimum. But, on the other hand, PSO does not need specific operators (such as crossover and mutation in the case of Genetic Algorithms, or pheromone updating in Ant Colony Optimization, amongst others), since particles update themselves with internal velocity. They also have memory and receive information only from the best particle in history, which is a simpler mechanism of information transmission than those used in Genetic Algorithms, for example. Particles try to converge to the best solution quickly, but PSO’s main drawback is that it is difficult to keep good levels of population diversity and to balance local and global search. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles.
Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called pbest. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbors of the particle. This location is called lbest. When a particle takes all the population as its topological neighbors, the best value is a global best and is called gbest.

After finding the two best values, the particle updates its velocity and positions with following equations (1.1) and (1.2).

\[
V[] = w \cdot v[] + c1 \cdot \text{rand()} \cdot (pbest[] - \text{present}[]) + c2 \cdot \text{rand()} \cdot (gbest[] - \text{present}[]) \tag{1.1}
\]

\[
\text{Present}[] = \text{present}[] + v[] \tag{1.2}
\]

where

w is the inertia weight
v[] is the particle velocity,
Present[] is the current particle (solution).
pbest[] the particle’s best
gbest[] the global best.
rand() is a random number between (0, 1).
c1, c2 are learning factors.

The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its pbest and lbest locations. Acceleration is weighted by a random term, with separate
random numbers being generated for acceleration toward pbest and lbest locations.

1.10.2 PSO parameter control

In PSO, followings are the parameter need to be tuned in. (i) The number of particles (ii) Dimension of particles (iii) Range of particles (iv) Vmax (v) Learning factors (vi) Stopping condition.

1.10.3 PSO Algorithm

1.10.3.1 Initialization of position and velocity vectors of the swarm:

Each decision variable represents a parameter to be optimized in the model. The initial positions of all particles have to be generated randomly within the limits specified for each decision variable. Each particle is initialized with random position vectors, $X$, and random velocity vectors, $V$, for all $i = 1, 2, \ldots, N$.

1.10.3.2 Initial evaluation of fitness function

The fitness of each particle is evaluated using the objective function of the problem. Then the best value out of all the fitness values is searched. Two “best” values are defined, the global best and the particle best. The global best $g_{best}$ is the highest fitness value among all the particles in an entire run (best solution so far) and the particle best $p_{best}$ is the highest fitness value of a specific particle up to the current iteration.
1.10.3.3 Modification of each searching point

Using the global best of each particle up to the current iteration, the searching point of each particle is modified using the change in velocity as given by equation (1.1) and the positions of the particles are updated using equation (1.2).

1.10.3.4 Evaluation of fitness function

After modification of the particle positions, the fitness of each particle is evaluated using the objective function of the problem.

1.10.3.5 Updating the global and the particle bests

The gbest and pbest values have to be updated according to the new fitness values. If the best fitness value of a particular particle in the swarm is better than the current gbest, then gbest is to be changed to the value of the searching point of the corresponding particle contributing to this best fitness value. Similarly the local best of other particles in the population should be changed accordingly if the present fitness function value is better than the previous.

1.10.3.6 Termination criteria

Repeat steps (3) to (6) in the PSO algorithm until either the pre-set maximum number of iterations is reached or no significant improvement is observed over a pre specified number of iterations.